Systematic Analysis of Structural Data as a Research Technique in Organic Chemistry

FRANK H. ALLEN, OLGA KENNARD,* and ROBIN TAYLOR

Crystallographic Data Centre, University Chemical Laboratory, Cambridge CB2 1EW, Great Britain

Received April 20, 1982 (Revised Manuscript Received November 1, 1982)

Crystallography is the most powerful method available for studying molecular structures. It has become increasingly important in recent years with the introduction of automatic diffractometers and new methods of structure solution and refinement. There are now more than 30000 organocarbon structures (i.e., organics, organometallics, and metal complexes) in the literature, a number that is likely to double within 5–7 years. Unfortunately, this wealth of information has not been greatly exploited; detailed discussions of individual crystal structures are commonplace, but systematic studies of large numbers of related structures are rarely undertaken.

The purpose of this Account is to illustrate how the systematic analysis of crystallographic data can be a versatile research technique in organic chemistry.[1] Examples are taken from many areas, including studies of bonding theories, conformational analysis, reaction pathways and hydrogen bonding. Many of the studies were carried out with the aid of the Cambridge Structural Database (CSD),[2] which is described briefly below.

Cambridge Structural Database (CSD)

CSD contains results of X-ray and neutron diffraction studies of organics, organometallics, and metal complexes. The information stored for each entry (Table I) may be divided into three categories: bibliographic information (BIB),[3] chemical connectivity information (CONN), and crystallographic data (DATA). CSD is fully comprehensive from 1935 onward, and is also a depository for unpublished atomic coordinates.[4] The database is available in 22 countries and on January 1, 1982, contained 31631 studies of 28978 different compounds. About 4000 new entries are added each year.

The information in CSD is searched and analyzed by a system of computer programs;[2] examples of program input and output are given in Figure 1. The two main search programs are BIBSER and CONNSER. BIBSER uses the bibliographic information fields italicized in Table I to locate entries on the basis of their chemical name, molecular formula, chemical class, etc. (e.g., Figure 1a). CONNSER is used to search for compounds containing specific chemical fragments. Program input consists of a coded representation of the desired fragment (e.g., the coding in Figure 1c would be used to find compounds containing the substructure shown in Figure 1b).

Table I
Principal Information Held in the Cambridge Structural Database (CSD)

Bibliographic (BIB)
Compound name(s); *qualifying phrase(s)* [e.g., neutron study, low-temperature study, absolute configuration determined]; *molecular formula, literature citation; chemical class(es)* [e.g., 15 = benzene nitro compounds, 51 = steroids, 58 = alkaloids, etc.]
Chemical Connectivity (CONN)
Chemical structural diagrams are coded in terms of *atom* and *bond properties*. Atom properties: atom sequence number (*n*); element symbol (*el*); no. of connected non-H atoms (*nca*); no. of terminal H atoms (*nh*); net charge (*ch*). Bond properties: pair of atom sequence numbers (*n* = *i*, *j*); bond type for bond *i*–*j* (*bt*); [see Figure 1, b and c].
Crystallographic Data (DATA)
Unit-cell Parameters; space group; symmetry operators; atomic coordinates; accuracy indicators [mean estimated standard deviation of C–C bonds, *R* factor]; evaluation flags [indicating: presence of disorder, presence of errors, method used for data collection, etc.]; comment text [e.g., describing any disorder, or errors in the original reference].

Output from both BIBSER and CONNSER consists of a listing of all structures in the database that satisfy the search criteria, together with the corresponding literature references (e.g., Figure 1, parts d and e). The RETRIEVE program may now be used to create a file of crystallographic data for entries located by the search. This DATA subfile can be processed by the programs PLUTO78, which produces plots and illustrations (e.g., Figure 1f), and GEOM78, which is used for geometrical analysis. GEOM78 will calculate the intra- and/or intermolecular geometries of all entries in a DATA subfile or provide systematic tabulations of selected geometrical parameters for a specific chemical fragment. Figure 1g is a tabulation of *R* factor, five independent bond lengths, and the torsion angle O(1)–C(2)–C(3)–X (X = midpoint of C(4)–C(5)) for the fragment shown

After graduating from Cambridge University, Olga Kennard worked at the Cavendish Laboratory and subsequently joined the Medical Research Council in London. She returned to Cambridge in 1961, as a member of M. R. C. External Staff, to build up an X-ray crystallographic unit in the University Chemical Laboratory. The Cambridge Crystallographic Data Centre (CCDC) was established there, under her direction, in 1965. Olga Kennard's research is concerned with biological structures, particularly oligonucleotides and their interaction with drugs and proteins. She was awarded an Sc.D. (Cantab) in 1971 and the Structural Chemistry Prize of the Chemical Society in 1980.

Frank Allen obtained a Ph.D. from London University and did postdoctoral crystallographic work at the University of British Columbia. He joined the CCDC in 1970 and has been involved in many aspects of its development; he is a member of the Data Commission of the International Union of Crystallography. Current research interests involve applications of the Structural Database in physical organic chemistry.

Robin Taylor obtained a B.A. from Oxford and a Ph.D. in chemical crystallography from Cambridge University. He held postdoctoral positions in York and Pittsburgh before joining the CCDC in 1980. His research interests include the structure and energy of hydrogen-bonded systems and the application of statistical methods to crystallographic data.

(1) This Account complements an earlier paper that was primarily concerned with the chemical interpretation of *individual* structures: Wilson, S. R.; Huffman, J. C. *J. Org. Chem.* **1980**, *45*, 560–566.

(2) 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. *Acta Crystallogr. Sect. B* **1979**, *B35*, 2331–2339 and references therein.

(3) Bibliographic information is also published annually in: "Molecular Structures and Dimensions", Kennard, O., Watson, D. G., Allen, F. H., Bellard, S., Eds.; D. Reidel: Dordrecht, The Netherlands.

(4) Journals involved in this scheme since 1977 include *Chemical Communications*, *Tetrahedron*, and *Tetrahedron Letters*.

0001-4842/83/0116-0146$01.50/0

(a) BIBSER : Sample Questions

(i) Find all steroids :

```
Q *CLASS '51'
```

Note : Chemical class 51 = steroids

(ii) Find all penicillins published between 1970 and 1980 inclusive :

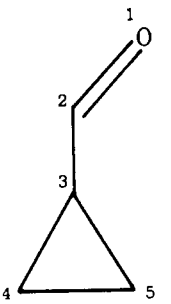
```
Q *COMPND 'PENICILL' and *YEAR '70-80'
```

Note : The string PENICILL will locate penicillin, penicillanic, penicilloic, etc.

(iii) Find all compounds for which absolute configuration has been established by X-ray methods :

```
Q *COMPND 'ABSOLUTE CONFIGURATION'
```

(b) CONNSER : Search Fragment



(c) CONNSER : Search Query [for fragment in (b)]

```
Q Cyclopropyl-carbonyls
C
C Atom properties:
C only nca is specified :
C
AT1  O  1
AT2  C  2
AT3  C  3
AT4  C  2
AT5  C  2
C
C Bond properties :
C node numbers, bond-type,
C and Acyclic/Cyclic flag :
C
BO  1  2  2  A
BO  2  3  1  A
BO  3  4  1  C
BO  4  5  1  C
BO  3  5  1  C
END
```

(d) BIBSER : Sample Output from Question (a,i)

AAMAND 16alpha-Acetyl-3beta-methoxy-CD-cis-D-norandrostane
C21 H34 O2; Class 51
J.Meinwald,A.J.Taggi,P.A.Luhan, A.T.McPhail
Proc.Nat.Acad.Sci.U.S.A., 71, 78, 1974

ABAXES 2alpha-Bromo-17beta-acetoxy-9-methyl-5alpha,9beta,10alpha-estran-3-one
C21 H31 Br1 O3; Class 51
J.C.A.Boeyens,J.R.Bull,J.Floor, A.Tuinman
J.Chem.Soc.,Perkin 1, 808, 1978

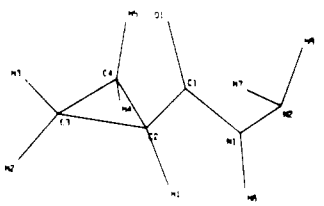
(e) CONNSER : Sample Bibliographic Listing for Search Query (c)

CORAMA (-)-Coronamic acid N-acetate
(absolute configuration)
C8 H13 N1 O3; Class 48,20
A.Ichihara,K.Shiraishi,S.Sakamura, A.Furusaki,N.Hashiba,T.Matsumoto
Tetrahedron Lett., 365, 1979

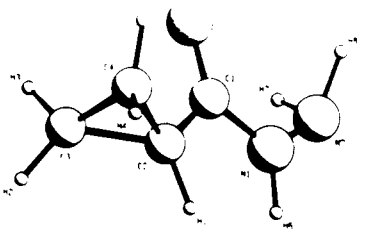
CPCOHA Cyclopropanecarbohydrazide
C4 H8 N2 O1; Class 20,9
D.B.Chesnut,R.E.Marsh
Acta Crystallogr., 11, 413, 1958

CPRPCX10 Cyclopropanecarboxamide
C4 H7 N1 O1; Class 20,1
R.E.Long,H.Maddox,K.N.Trueblood
Acta Crystallogr.,Sect.B, 25, 2083, 1969

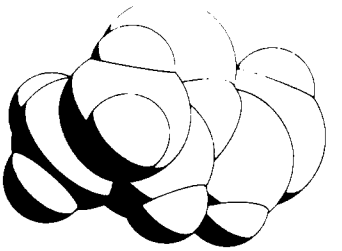
(f) PLUTO78 : Sample Plots for Cyclopropanecarbohydrazide [CPCOHA in (e)]



(i) 'stick' diagram



(ii) 'ball-and-spoke' style with shading and perspective



(iii) 'space-filling' model with shading and perspective

(g) GEOM78 : Example of Geometry Tabulation for Fragment (b)

Notes :

Distances in Angstroms, angle (TAU) in degrees.

TAU is the torsion angle : O(1) - C(2) - C(3) - X(1), where X(1) is the mid-point of the bond C(4) - C(5).

STD. DEVN. is the standard deviation of each sample. The standard deviations of the means are 0.004, 0.006, 0.003, 0.005 and 0.006.

CSD code	*RFACT	C4-C5	C3-C4	C3-C5	C3-C2	C2=O1	TAU
CORAMA	0.059	1.491	1.534	1.503	1.502	1.216	28.04
CPCOHA	0.130	1.478	1.520	1.493	1.478	1.213	4.52
CPRPCX10	0.087	1.467	1.485	1.501	1.484	1.238	-7.74
CPRPCX10	0.087	1.450	1.489	1.493	1.470	1.249	-4.67
DCPEDO	0.047	1.475	1.515	1.510	1.456	1.213	3.19
DMCPRC	0.085	1.477	1.521	1.510	1.456	1.246	-7.30
MBCPCX	0.062	1.488	1.531	1.508	1.480	1.210	30.07
NPCPMK	0.092	1.474	1.513	1.488	1.469	1.235	8.64
PMCPRC10	0.037	1.490	1.548	1.505	1.484	1.205	-21.17
SDPPCX	0.042	1.482	1.533	1.510	1.470	1.201	-5.47
MEAN		1.477	1.519	1.502	1.475	1.223	
STD. DEVN.		0.012	0.020	0.008	0.014	0.018	

Figure 1. Examples of input and output for the CSD program system.

in Figure 1b. The DATA subfile may also be processed by the user's own programs.

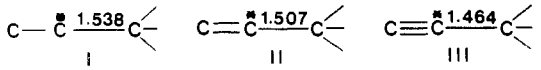
Systematic Analysis of Intramolecular Geometry

Mean Geometries. The results of early X-ray analyses were used to derive mean bond lengths,[5] covalent radii, etc., which were important in the development of structural chemistry and bonding theories. As the number of structural studies increased, it became possible to determine the mean geometries of complete

(5) Sutton, L. E., Ed., "Tables of Interatomic Distances and Configuration in Molecules and Ions"; The Chemical Society: London; Special Publication 11 (1958) and 18 (1965).

chemical residues. Thus, the average dimensions of furanose,[6] pyranose,[7] and nucleic acid base residues[8] have all been determined. Given these mean values, together with their standard deviations, it is easy[8] to derive orthogonal coordinates for "standard residues". Apart from a variety of crystallographic uses, standard geometries are invaluable in model building, parameterization of empirical force fields, and interpretation of new structural data.

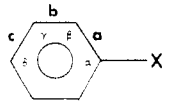
Substituent and Hybridization Effects. Early bond length tabulations[5] were subdivided on the basis of gross differences in hybridization and environment. For example, hybridization changes at C* in I–III cause



I II III

appreciable bond length alterations, which are easily recognized in any individual crystal structure. Smaller changes in electron distribution—due, perhaps, to substitution and slight rehybridization—produce correspondingly smaller geometric variations, which are often close to the error limits of individual structure determinations. These are only discovered by the systematic analysis of many related structures. Studies of this type provide results that can be correlated with theoretical calculations, reactivity, spectral properties, and other physical phenomena. Two examples are discussed below.

Substituent-Induced Ring Deformations in Benzene. Early microwave and X-ray investigations showed that the regular hexagonal geometry of benzene is perturbed when the ring is substituted by strong electron-withdrawing or -donating groups. With reference to IV,



IV

deformations due to electron-withdrawing groups involve an increase in the ipso angle α, a decrease in the ortho angles β, and shortening of bonds a with respect to bonds b. Deformations due to electron-donating groups are in the opposite sense. Only groups with strong electronic effects produce deformations that are consistently larger than the error limits of individual structure determinations.

The angular deformations ($\Delta\alpha$, $\Delta\beta$, $\Delta\gamma$, $\Delta\delta$ from 120°; see IV) in more than 100 mono- and para-disubstituted benzenes have been analyzed. Early results[9,10] showed that $\Delta\alpha$ is the largest deformation, and that its value for a particular substituent X is unaffected by para-substitution; $\Delta\beta$ was found to be $\simeq -\Delta\alpha/2$. Values of α were determined for a variety of substituents (Table II) by averaging relevant fragment geometries. It was recognized that σ-effects predominate in determining $\Delta\alpha$ values.[10,11] This is consistent with the observed correlation between α and the electronegativity (χ) for

Table II
Angular Deformations in Substituted Benzenes (IV)[a,b]

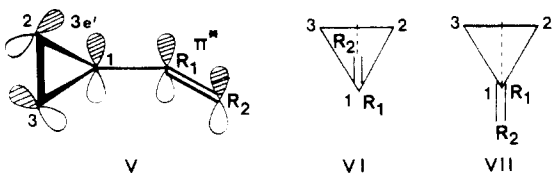
X	α(mean), deg[c]	Linear regression values, deg[d] $\Delta\alpha$	$\Delta\beta$	$\Delta\gamma$	$\Delta\delta$
$N(Me)_2$	117.2	−2.4 (3)	0.6 (2)	1.4 (2)	−1.7 (3)
Ph	117.6	−2.3 (2)	1.0 (1)	0.6 (1)	−0.9 (2)
Me	118.1	−1.9 (2)	1.0 (1)	0.4 (1)	−0.8 (2)
CH=CHR	118.0	−1.8 (2)	0.8 (1)	0.3 (1)	−0.4 (2)
COMe	118.8	−1.0 (2)	0.4 (1)	0.2 (1)	−0.3 (2)
COOH	119.8	0.1 (2)	−0.2 (1)	0.1 (1)	0.2 (2)
OMe	119.9	0.2 (2)	−0.6 (1)	1.1 (1)	−1.1 (2)
CN	121.8	1.1 (2)	−0.8 (1)	0.3 (1)	−0.1 (2)
Cl	121.4	1.9 (2)	−1.4 (1)	0.6 (1)	−0.2 (2)
NO_2	122.1	2.9 (2)	−1.9 (1)	0.3 (1)	0.4 (1)
F	123.4	3.4 (2)	−2.0 (1)	0.5 (1)	−0.4 (2)

[a] Standard deviations in α (mean) are in the range 0.1–0.2°. [b] Substitution reduces perfect D_{6h} ring symmetry to C_{2v}; hence $\Delta\alpha + 2\Delta\beta + 2\Delta\gamma + \Delta\delta = 0$. [c] Reference 11. [d] Reference 13.

first-row functional groups[11] and second-row elements[10] (Figure 2a,b). Later,[12] it was realized that substituents that interact with the benzene π-system also produce significant deformations ($\Delta\gamma$, $\Delta\delta$). Factor analysis showed that *all* substituent-induced deformations can be ascribed to two independent effects, which were interpreted as σ- and π-interactions.[13] Values of $\Delta\alpha$, $\Delta\beta$, $\Delta\gamma$, $\Delta\delta$ for 21 functional groups were derived by linear regression, the geometries of 71 mono- and para-disubstituted benzenes (Table II)[14] being used.

The marked sensitivity of α to σ-effects, evidenced by the α–χ plots, is confirmed by the plot[11,13] of $\Delta\alpha$ against the Taft[15] inductive parameter, σ_I (Figure 2c). The sensitivity of γ to π-effects is shown by the $\Delta\gamma$–σ_R° plot[13] in Figure 2d (σ_R° = Taft resonance parameter[15]). These correlations suggest that Δ-values may be a useful addition to traditional reactivity parameters, since they measure substituent effects on the ring in the ground state, do not depend on the choice of a suitable reaction series, are independent of solvent effects, and can be related to conformational changes.

Conjugation and Hybridization in Three-Membered Rings. Cyclopropane is unusual among cycloalkanes in that it conjugates with π-acceptor substituents,[16] as is shown by spectroscopic results[17] and the stabilization of carbonium ions by cyclopropane. Molecular orbital theory suggests[18] that conjugation involves transfer of electron density from the cyclopropane 3e′ orbitals to the π^* orbitals of the substituent (V). This would be



V VI VII

(6) Arnott, S.; Hukins, D. W. L. *Biochem. J.* **1972**, *130*, 453–465.
(7) Arnott, S.; Scott, W. E. *J. Chem. Soc., Perkin Trans. II* **1972**, 324–335.
(8) Taylor, R.; Kennard, O. *J. Mol. Struct.* **1982**, *78*, 1–28. Taylor, R.; Kennard, O. *J. Am. Chem. Soc.* **1982**, *104*, 3209–3212.
(9) Domenicano, A.; Vaciago, A.; Coulson, C. A. *Acta Crystallogr., Sect. B* **1975**, *B31*, 221–234.
(10) Domenicano, A.; Vaciago, A.; Coulson, A. A. *Acta Crystallogr., Sect. B* **1975**, *B31*, 1630–1641.
(11) Domenicano, A.; Mazzeo, P.; Vaciago, A. *Tetrahedron Lett.* **1976**, 1029–1032.
(12) Domenicano, A.; Vaciago, A. *Acta Crystallogr., Sect. B* **1979**, *B35*, 1382–1388.
(13) Domenicano, A.; Murray-Rust, P. *Tetrahedron Lett.* **1979**, 2283–2286.
(14) A similar independent analysis is described by: Norrestam, R.; Schepper, L. *Acta Chem. Scand. Sec. A* **1981**, *A35*, 91–103. The two sets of angle deformations have a correlation coefficient of 0.966 and a root mean square deviation of 0.22°.
(15) Ehrenson, S.; Brownlee, R. T. C.; Taft, R. W. *Prog. Phys. Org. Chem.* **1973**, *10*, 1–80.
(16) Charton, M. "The Chemistry of Alkenes"; Zabicky, J., Ed.; Interscience, London, 1970; Vol. II, pp 511–610.
(17) Pete, J.-P. *Bull. Soc. Chim. Fr.* **1967**, 357–370.
(18) Hoffmann, R. *J. Chem. Phys.* **1964**, *40*, 2480–2488. Hoffmann R.; Stohrer, W.-D. *J. Am. Chem. Soc.* **1971**, *93*, 6941–6948.

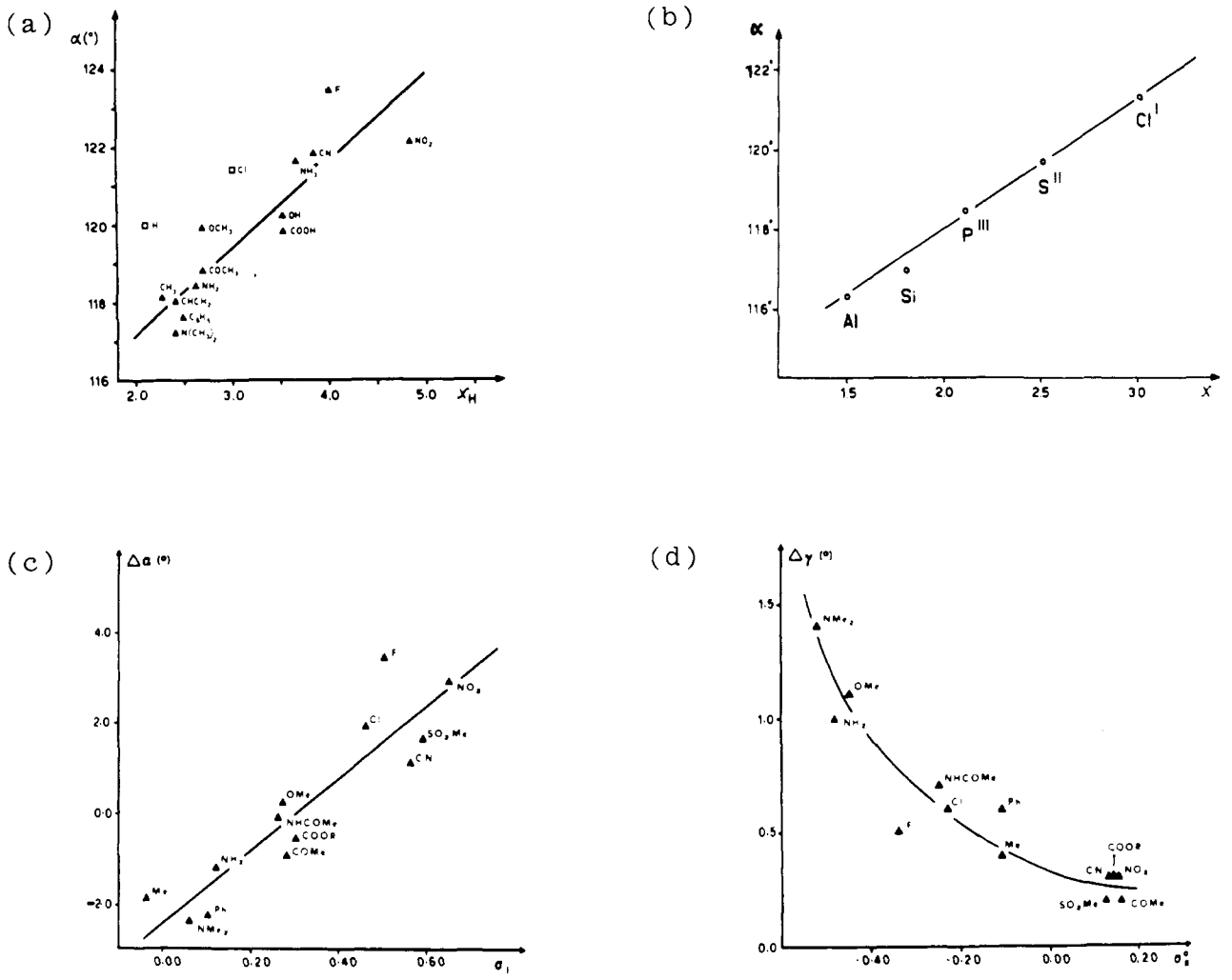
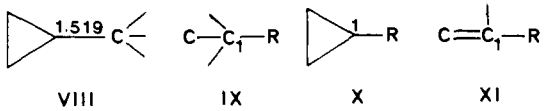


Figure 2. Correlation of substituent-induced angular deformations in benzene with electronegativity (χ) and Taft parameters (σ). The ipso angle (α in IV) is plotted (a) against χ for first-row functional groups (Reprinted with permission from ref 11. Copyright 1976, Pergamon Press Ltd.) and (b) against χ for second-row elements (Reprinted with permission from ref 10. Copyright 1975, International Union of Crystallography). In (c) the deformation $\Delta\alpha$ is plotted against the inductive parameter σ_I (Reprinted with permission from ref 13. Copyright 1979, Pergamon Press Ltd.), while in (d) the deformation $\Delta\gamma$ is plotted against σ_R° (Reprinted with permission from ref 13. Copyright 1979, Pergamon Press Ltd.). We gratefully acknowledge permission to reproduce these plots from Dr. Aldo Domenicano and the copyright holders.

expected to shorten the distal (2–3) bond by δ and increase the vicinal (1–2,1–3) bonds by $\sim\delta/2$. However, the degree of conjugation will depend on the extent of orbital overlap, which is maximized in the cis- and trans-bisected conformations (Newman projections VI and VII, respectively).[19]

In order to quantify these effects, Allen[20] analyzed results from 146 X-ray studies of cyclopropane derivatives. The mean geometry of nonconjugated cyclopropanes was established by using purely C(sp^3) derivatives (VIII). Substituent-induced bond length



VIII IX X XI

asymmetries were consistently observed in the geometries of derivatives involving the π-acceptor substituents —C=O (ketones, acids, esters; see Figure 1g for a typical geometrical tabulation), —C=C, and —CN. It was found that multiple substitution produces net distortions that can be approximated by simple sums of individual substituent effects. The bond length asymmetry parameters (δ in the preceding paragraph) due to some representative substituents are given in Table III. The variation of bond length asymmetries with conformation suggests[20,21] that conjugative overlap of ring and acceptor orbitals is effective for —C=O and C=C conformations within 30° of the bisected positions (VI, VII). These results are in agreement with ultraviolet spectral data.[17]

Table III
Asymmetry Parameter[a] for
Cyclopropane Rings Having π-Acceptor Substituents

subst.	δ, Å	subst.	δ, Å
C=O	−0.026 (5)	N=C	−0.018 (-)
C=C	−0.022 (4)	CN	−0.017 (2)
Ph	−0.018 (-)	N=N	−0.014 (-)

[a] Reference 20.

Cyclopropane is also unusual because its protons exhibit vinylic properties.[16] This is consistent with the

(19) Hoffmann, R. *Tetrahedron Lett.* **1970**, 2907–2909. Hoffmann, R.; Davidson, R. B. *J. Am. Chem. Soc.* **1971**, *93*, 5699–5705.
(20) Allen, F. H. *Acta Crystallogr., Sect. B* **1980**, *B36*, 81–96.
(21) Allen, F. H. In "Molecular Structure and Biological Activity"; Griffin, J. F., Duax, W. L., Eds.; Elsevier Biomedical: New York, 1982; 105–116.

Table IV
Variation in the C(1)-R Bond Length[a] in IX-XI due to Changes in Hybridization (h) and to Conjugative Effects (c)

R		fragment IX	X	XI	$h(\Delta)$[b]	$h(sp^2)$[b]
C sp^3	h	1.538	1.519	1.507	−0.019	−0.031
C=C	h	1.507	1.480	1.472	−0.027	−0.035
	$h + c$		1.470	1.458		
	c		−0.010	−0.014		
C=O	h	1.512	1.489	1.482	−0.023	−0.030
(keto)	$h + c$		1.474	1.460		
	c		−0.015	−0.022		
C=O	h	1.514	1.504	1.497	−0.010	−0.017
(acids,	$h + c$		1.484	1.470		
esters)	c		−0.020	−0.027		

[a] Reference 22. [b] $h(\Delta) = h(X) - h(IX)$, $h(sp^2) = h(XI) - h(IX)$ for each substituent R (see text).

abnormally short exocyclic C–C bond in derivatives of the type VIII (cf. I, II).[20,22] Theoretical models[23] indicate that the C(ring) hybrid involved in the exocyclic C–C bond in VIII is ~sp^2, whereas those involved in the ring σ-bonds are ~sp^5. Hybridization in cyclopropane was studied in detail[22] by comparing the C–R distances in the related fragments IX–XI (R = C(sp^3), vinyl, keto, acid, ester). In those cases where conjugation can occur (i.e., X and XI with R = —C=C or —C=O), two C–R distances were obtained, for (i) conformations outside the ranges established for effective conjugation (this is a measure of hybridization effects (h)) and (ii) conjugative conformations, where C–R is further shortened by π-effects (c). Representative results are given in Table IV. The relative bond length contractions due to hybridization, defined by

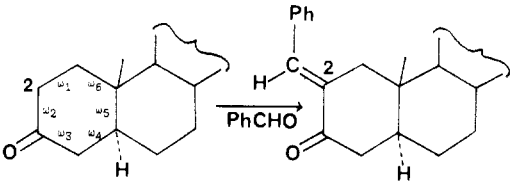
$$h(\Delta) = [\text{R–C(cyclopropane)}]_h - [\text{R–C}(sp^3)]$$

$$h(sp^2) = [\text{R–C}(sp^2)]_h - [\text{R–C}(sp^3)]$$

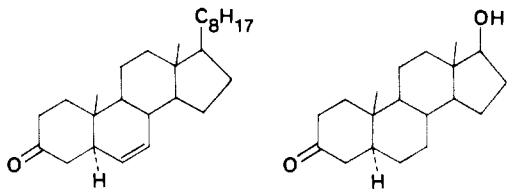
can be used to infer the hybridization of the cyclopropane ring atoms. Values of $h(\Delta)/h(sp^2)$ range from 59 to 77%. The mean, 69%, corresponds to $sp^{2.25}$ hybridization (30.8% s character) in exocyclic bonds; hence, ring σ-hybrids are $sp^{4.20}$ (19.2% s). Table IV also permits comparison of the relative conjugative abilities of cyclopropane and the vinyl group.[16] The ratios $c(\Delta)/c(sp^2)$ average to 71%, which compares well with a UV bathochromic shift ratio of ~60%.[17]

Conformational Analysis. Comparative reviews of steroids,[24] alkaloids and terpenes,[25] and medium rings[26] illustrate the extensive application of crystallographic data to conformational analysis.[27] In the present review, we are particularly concerned with the *systematic* analysis of solid-state conformations. Although some molecules are known to adopt different conformations in solution and in the crystalline state, it has been shown that solid-state conformational data can be successfully related to solution properties, provided that a sufficient number of crystal structures are studied. This is illustrated by the following example.

The relative rates of condensation with benzaldehyde of a series of 3-keto-5α steroids (XII) were determined



XII XIII



XIV XV

by Barton et al.[28] The 2-benzylidene ketone XIII is exclusively formed by elimination of OH^-, the reaction involving rehybridization of C(2) from sp^3 to sp^2. The presence of ethylenic links and/or substituents at positions remote from the reaction center was found to produce large differences in reaction rates. It was therefore inferred that gross conformational changes at remote sites produce small changes—here at C(2)—by "conformational transmission". This influences the rate of condensation, r, which may be expressed as[28]

$$r = r_0 \Pi_i f_i$$

where r_0 is the rate for saturated XII and the f_i are "group rate factors" for remote substituents or double bonds.

The crystal structure[29] of cholest-6-en-3-one (XIV), which has the highest reaction rate (r = 645), shows an unusually flat A ring, with a mean ring torsion angle $\bar{\omega}$ ($=\sum_i \omega_i/6$ in XII) of only 50.4°. The flattening is especially pronounced at C(2), where ω_{12} ($=(\omega_1 + \omega_2)/2$ in XII) is 42.2°. The ring bond angle at C(2), v, is correspondingly high at 114.4 (8)°. In contrast, the structure[30] of 17β-hydroxyandrostan-3-one (XV), which has a much lower reaction rate (r = 188), shows $\bar{\omega}$ = 54.1°, ω_{12} = 51.7°, and v = 110.7°. Clearly, C(2) is more prepared to undergo rehybridization in XIV than in XV.[29] An analysis[21] of several other A-ring conformations confirmed that the reaction rate tends to increase as the ring becomes flattened at C(2). The parameter $\alpha = \omega_{12} - \bar{\omega}$ was used to express the relative puckering in each ring (negative α represents flattening at C(2)). For steroids with r = 33, 188, 235, and 645, the corresponding mean values of α were 2.6, −2.8, −3.3, and −8.3°. Although the data are limited, the r–α correlation is qualitatively acceptable.

Systematic Analysis of Intermolecular Interactions

Hydrogen-Bonding Studies. Systematic analyses of crystal structures have made a significant contribution to our knowledge of hydrogen bonding. Studies of O–H···O and C–H···O bonds are discussed below.

(22) Allen, F. H. *Acta Crystallogr., Sect. B* **1981**, *B37*, 890–900.
(23) Bernett, W. A. *J. Chem. Educ.* **1967**, *44*, 17–24 and references therein.
(24) Altona, C.; Geise, H. J.; Romers, C. *Tetrahedron* **1968**, *24*, 13–32. Duax, W. L.; Norton, D. A. "Atlas of Steroid Structures"; Plenum: London, 1975. Duax, W. L.; Weeks, C. M.; Rohrer, D. C. *Recent Prog. Horm. Res.* **1976**, *32*, 81–116.
(25) Mathieson, A. McL. *Perspect. Struct. Chem.* **1967**, *1*, 41–108.
(26) Dunitz, J. D. *Perspect. Struct. Chem.* **1968**, *2*, 1–70.
(27) The correlation of crystal structure conformations with pharmaceutical activity cannot be covered adequately here. Interested readers are referred to: Griffin, J. F.; Duax, W. L., Eds., "Molecular Structure and Biological Activity"; Elsevier Biomedical: New York, 1982.
(28) Barton, D. H. R.; Head, A. J.; May, P. J. *J. Chem. Soc.* **1957**, 935–944. Barton, D. H. R.; McCapra, F.; May, P. J.; Thudium, F. *J. Chem. Soc.* **1960**, 1297–1311.
(29) Guy, J. J.; Allen, F. H.; Kennard, O.; Sheldrick, G. M. *Acta Crystallogr., Sect. B* **1977**, *B33*, 1236–1244.
(30) Courseille, C.; Precigoux, G.; Leroy, F.; Busetta, B. *Cryst. Struct. Commun.* **1973**, *2*, 441–446.

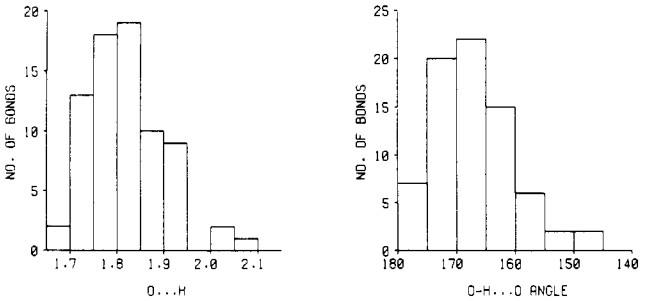
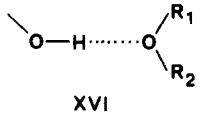


Figure 3. Distribution of O···H distances (Å) and O–H···O angles (°) in O–H···O hydrogen bonds (ref 31).

O–H···O Hydrogen Bonds. The distributions of O···H distances and O–H···O angles were determined in a survey of the neutron diffraction geometries of 74 O–H···O< bonds (Figure 3).[31] The mean values are 1.818 (9) Å and 167.1 (8)°, respectively. The latter is in good agreement with the most probable value predicted by molecular orbital calculations (163°).[32] For hydrogen bonds with O···H < 1.812 Å (the median O···H distance of the sample) the mean O–H···O angle was 168.4 (9)°, compared with 165.8 (12)° for bonds with O···H > 1.812 Å. The difference between these means is statistically significant at the 92.5% level, which suggests that short hydrogen bonds tend to be more linear than long ones. This can be ascribed to the unfavorable van der Waals interaction between the oxygen atoms in short, nonlinear O–H···O bonds.[33]

In the same study,[31] it was found that the proton in XVI tends to lie in, or near to, the plane containing the



XVI

lone-pair orbitals of the acceptor atom (assumed to be the plane bisecting the R_1–O–R_2 angle). However, there was no evidence of a preferred direction within the plane. Apparently, hydrogen bonding does not occur preferentially along the directions which correspond to idealized sp^3 lone-pair orbitals. This finding is consistent with charge density studies,[34] which show the lone-pair electron density of sp^3 oxygen atoms to be rather diffuse, bearing little resemblance to the simple "rabbit-ear" model. The directional influence of the lone pairs may be more important in hydrogen bonds that involve carbonyl acceptors[35] because the angular separation of sp^2 lone pairs is greater than that of sp^3 lone pairs.

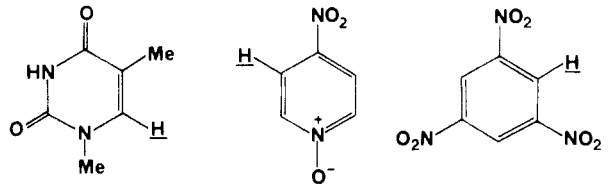
Theoretical calculations[36] suggest that the length of an O–H···O bond should be dependent on its environment. Specifically, hydrogen bonds belonging to infinite ···O–H···O–H··· chains should be stronger, and therefore shorter, than isolated O–H···O bonds (the so-called "cooperative effect"). This is due to the changes that occur in the electron densities at the oxygen and hydrogen atoms upon hydrogen-bond formation. The mean O···H distance of hydrogen bonds belonging to infinite chains was found[31] to be 1.805 (9) Å, compared with 1.869 (23) Å for isolated hydrogen bonds. The difference between these means is statistically significant at the 99.5% level, thus confirming the theoretical prediction. Many of the hydrogen-bond patterns observed in carbohydrate crystal structures can be rationalized by the cooperative effect.[37] For example, all of the hydrogen bonding in sugar alcohol structures is of the favorable infinite chain variety. These structures are ideally suited to this type of arrangement because they possess only hydroxyl groups; i.e., there is a 1:1 ratio of hydrogen-bond donor and acceptor atoms. Thus, it is possible for infinite ···O–H···O–H··· chains to incorporate all of the oxygen atoms in the structure. This is not the case in cyclic pyranoses, where the ring oxygen atom can accept but not donate hydrogen bonds. It therefore acts as a chain stopper and is often completely excluded from the hydrogen-bonding schemes in pyranose crystal structures.[37]

C–H···O Hydrogen Bonds. The hydrogen-bonding schemes observed in the structures of other biologically important molecules (e.g., amino acids, nucleosides) are more difficult to interpret. Nevertheless, some useful results were obtained in a recent survey of hydrogen-bond patterns in amino acid structures.[38] Among these was the observation that short C–H···O contacts are relatively common in amino acid crystal structures, particularly when the C–H group is adjacent to a $^+NH_3$ group. Thus, C–H···O hydrogen bonding may be a significant factor in determining the conformations and crystal-packing arrangements of amino acids. This conclusion was supported by an analysis[39] of the crystallographic environments of 661 (C–)H atoms (i.e., hydrogen atoms covalently bonded to carbon), observed by neutron diffraction in 113 organic crystal structures. Fifty-nine C–H···O contacts (of which 41 were intermolecular) were found with H···O < 2.4 Å and C–H···O > 90°; 2.4 Å is 0.3 Å shorter than the sum of the van der Waals radii of H and O, using the values of Bondi and Kitaigorodsky.[40] In contrast, short C–H···C and C–H···H contacts were found to be extremely rare: it was established that (C–)H atoms have a statistically significant (> 99.9%) tendency to form short intermolecular contacts to oxygen rather than carbon or hydrogen. The proton in the majority of short C–H···O contacts lies within 30° of the plane containing the oxygen lone pairs.[39]

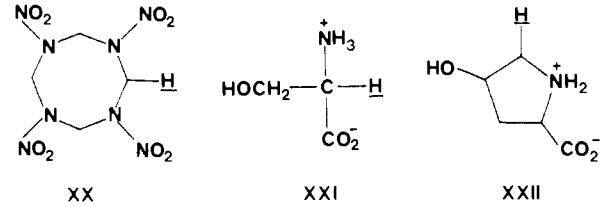
The (C–)H atoms involved in the 10 shortest C–H···O contacts found in the above survey are underlined in XVII–XXVI. Eight of them are immediately adjacent to nitrogen; the other two belong to nitrogen-containing aromatic molecules. Presumably, the inductive effect of nitrogen decreases the electron density at nearby (C–)H atoms, thereby enhancing their ability to form short C–H···O contacts.

Inference of Reaction Pathways. The use of crystallographic information for the inference of reac-

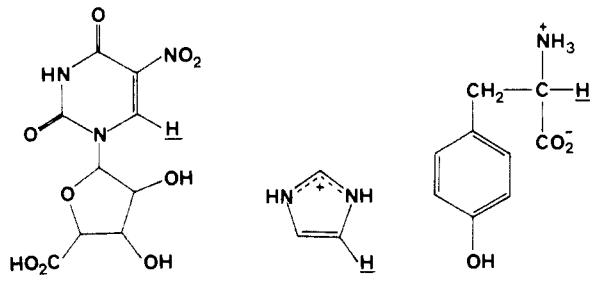
(31) Ceccarelli, C.; Jeffrey, G. A.; Taylor, R. *J. Mol. Struct.* **1981**, *70*, 255–271.
(32) Newton, M. D.; Jeffrey, G. A.; Takagi, S. *J. Am. Chem. Soc.* **1979**, *101*, 1997–2002.
(33) Taylor, R. *J. Mol. Struct.* **1981**, *73*, 125–136.
(34) For example, Diercksen, G. H. F. *Theor. Chim. Acta* **1971**, *21*, 335–367.
(35) Olovsson, I.; Jönsson, P.-G. In "The Hydrogen Bond—Recent Developments in Theory and Experiments"; Schuster, P., Zundel, G., Sandorfy, C., Eds.; Elsevier North-Holland: Amsterdam, 1976; Vol II 416–420.
(36) Del Bene, J. E.; Pople, J. A. *J. Chem. Phys.* **1973**, *58*, 3605–3608.
(37) Jeffrey, G. A.; Lewis, L. *Carbohydr. Res.* **1978**, *60*, 179–182. Jeffrey, G. A.; Takagi, S. *Acc. Chem. Res.* **1978**, *11*, 264–270.
(38) Jeffrey, G. A.; Maluszynska, H. *Int. J. Biol. Macromol.* **1982**, *4*, 173–185.
(39) Taylor, R.; Kennard, O. *J. Am. Chem. Soc.*, **1982**, *104*, 5063–5070.
(40) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441–451. Kitaigorodsky, A. I. "Molecular Crystals and Molecules"; Academic Press: New York and London, 1973; 10–18.



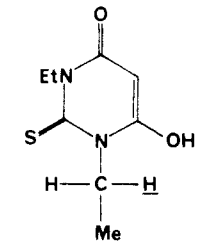
XVII XVIII XIX



XX XXI XXII

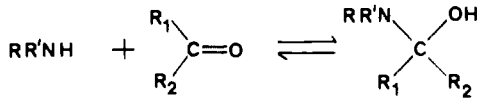


XXIII XXIV XXV



XXVI

tion pathways is illustrated by an early application:[41] the addition of a nitrogen nucleophile to a carbonyl group (XXVII). N···C–O contacts in six crystal



XXVII

structures were examined, the N···C distances ranging from 2.91 Å (i.e., nonbonded N···C=O contact) to 1.49 Å (i.e., covalent N–C–O linkage). It was assumed that each contact represents a point on, or near to, the minimum-energy pathway followed by a nitrogen nucleophile as it approaches a carbonyl group. The longest contact represents the incipient addition reaction, while the shortest contact corresponds to the completed reaction. Between them, the contacts should map out the minimum-energy reaction pathway, provided that the perturbing effects of crystal-packing forces are small enough to average out over the 6 structures.

The N···C–O geometries were described[41] by parameters shown in Figure 4 and tabulated in Table V. The N···C distance (d_1) is inversely correlated with the C–O

(41) Bürgi, H. B.; Dunitz, J. D.; Shefter, E. *J. Am. Chem. Soc.* **1973**, *95*, 5065–5067.

Table V
Geometries of N···C–O Contacts[a]

compound	d_1, Å	d_2, Å	Δ, Å	N···C–O, deg
methadone	2.910	1.214	0.064	105.0
cryptopine	2.581	1.209	0.102	102.2
protopine	2.555	1.218	0.115	101.6
clivorine	1.993	1.258	0.213	110.2
retusamine	1.64	1.38	0.36	110.9
N-brosylmitomycin A	1.49	1.37	0.42	113.7

[a] Reference 41. Parameters are illustrated in Figure 4.

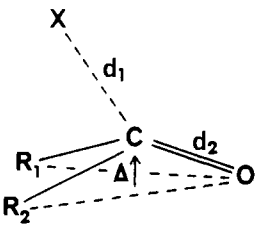


Figure 4. Parameters used to describe geometry of N···C=O and O···C=O contacts (ref 41, 42, 44). X = N or O.

distance (d_2), and the deviation of the carbon atom from the R_1,R_2,O plane (Δ). Thus, as the nitrogen atom approaches the carbonyl group, the C–O distance increases and the carbon atom is progressively displaced from the R_1,R_2,O plane. This can be related to the changes that occur in the molecular orbitals as the reaction proceeds: the π and π^* carbonyl group orbitals develop into an oxygen lone-pair orbital and an empty sp^n orbital on the carbon atom, which overlaps with the nitrogen lone-pair orbital. The N···C–O angles suggest that the nitrogen atom approaches from a constant direction, at ~107° to the C–O bond, i.e., the approach direction is *not* perpendicular to the C–O bond.

These conclusions were supported by a subsequent analysis[42] in which eight more structures were studied. It was found that the deviation of the carbon atom from the R_1,R_2,O plane depends on the nature of the nitrogen atom. Weak nucleophiles (e.g., anilines) induce smaller deviations than strong nucleophiles (e.g., aliphatic amines). Thus, a 2.76 Å (C—aryl)N···C=O contact in 1-*p*-tolyl-1-azacyclooctan-5-one[43] induces a deviation of only 0.016 Å (cf. Table V). Presumably, the energy required to displace the carbon atom from the R_1,R_2,O plane is offset by the attractive N···C interaction, which becomes weaker as the nucleophilicity of the nitrogen atom is reduced.[42] The same principles govern the interaction between oxygen nucleophiles and carbonyl groups.[44]

The crystallographic environments of divalent sulfur groups (R_1–S–R_2) have also been investigated.[45] These groups form short nonbonded contacts to both electrophiles and nucleophiles. The majority of S···electrophile contacts are approximately perpendicular to the R_1,S,R_2 plane. However, most S···nucleophile contacts lie in, or near to, this plane, and at an angle of ~180° to one of the S–R bonds. This suggests that approaching electrophiles interact preferentially with

(42) Dunitz, J. D. "X-Ray Analysis and the Structure of Organic Molecules"; Cornell University Press: Ithaca and London, 1979; 366–379.
(43) Kaftory, M.; Dunitz, J. D. *Acta Crystallogr., Sect. B* **1975**, *B31*, 2912–2914.
(44) Bürgi, H. B.; Dunitz, J. D.; Shefter, E. *Acta Crystallogr., Sect. B* **1974**, *B30*, 1517–1527.
(45) Rosenfield, R. E.; Parthasarathy, R.; Dunitz, J. D. *J. Am. Chem. Soc.* **1977**, *99*, 4860–4862.

the highest occupied molecular orbital, i.e., a sulfur lone-pair orbital nearly perpendicular to the R_1,S,R_2 plane. Conversely, nucleophiles tend to approach along the direction of the lowest unoccupied molecular orbital, i.e., σ^* (S–R_1) or σ^* (S–R_2).[46]

Concluding Remarks

This Account illustrates the variety of chemical phenomena that can be investigated by systematic examination of published crystallographic results. Such analyses can now be carried out with comparative ease by both crystallographers and chemists. Although there is much scepticism about the relevance of crystallographic results to molecular properties in solution, this Account shows that this reservation has only limited justification. The reaction pathway studies, in particular, demonstrate that crystal-packing effects are small enough to permit meaningful analysis of very weak interactions, such as N···C=O and O···C=O.

(46) For other reaction pathway studies, see: Bürgi, H. B. *Inorg. Chem.* **1973**, *12*, 2321–2325. Bürgi, H. B. *Angew. Chem., Int. Ed. Engl.* **1975**, *14*, 460–473. Britton, D.; Dunitz, J. D. *Helv. Chim. Acta* **1980**, *63*, 1068–1073.

The 30 000 organocarbon crystal structures currently available cover an enormous chemical range and include many novel and unusual compounds (since these are the ones most likely to be examined by crystallographic methods). By the end of the century, there will be over 100 000 structures in the public domain. The importance of this information to organic chemistry can scarcely be overstated, because the results of most crystal structure determinations are precise, detailed, and unambiguous. However, studies of individual structures are often of limited value: if an unexpected structural feature is observed, it may not be statistically significant and may well be ascribed to experimental errors or packing effects. When the same feature is consistently observed in a series of related structures, these explanations become untenable and the observation must be rationalized in physicochemical terms. Thus, the *systematic* analysis of large numbers of related structures is a powerful research technique, capable of yielding results that could not be obtained by any other method.

We thank Professors Aldo Domenicano and Jack Dunitz for reading the manuscript prior to submission.

From Crystal Statics to Chemical Dynamics

HANS BEAT BÜRGI*

Laboratory for Chemical and Mineralogical Crystallography, University of Berne, CH-3012 Berne, Switzerland

JACK D. DUNITZ*

Organic Chemistry Laboratory, Swiss Federal Institute of Technology, CH-8092 Zürich, Switzerland

Received May 19, 1982 (Revised Manuscript Received October 6, 1982)

The study of molecular structure involves both static and dynamic aspects, the static aspects dealing with equilibrium arrangements of atoms in molecules, the dynamic ones with the relative motion of atoms during molecular vibrations or transformations, i.e., chemical reactions. The contribution of X-ray crystallography to the static aspects is well-known; in fact, for many years now, crystallography has been the main source of our information about the three-dimensional structure of molecules.[1] Automatic diffractometers and fast, powerful methods for interpreting the diffraction measurements have led to an unprecedented increase in the amount of structural information available.[2] Fortunately, for organic structures at least, this information is stored in computer-readable form in the Cambridge Structural Database (CSD), which has established itself as an essential instrument for answering questions about the systematics of atomic arrangements in molecules.[3]

It is not so well-known that crystallography can also provide important information about the dynamic aspects of molecular structure. Of course, we cannot observe the actual motion of the atoms; an X-ray diffraction experiment takes roughly 10^5–10^6 s, whereas dynamic processes on the molecular level occur on a time scale of 10^{-12} s or less. Thus, diffraction measurements provide at best averaged information, in the form of mean-square vibration amplitudes for the individual atoms (expectation values of the square of the atomic displacements from equilibrium). This is the kind of information that is expressed in vivid, pictorial fashion by the thermal ellipsoids in the computer-drawn ORTEP diagrams that adorn many crystallographic publications.[4] At a given temperature, large extension

Hans-Beat Bürgi was born in Switzerland in 1942. He studied chemistry and received his Ph.D. degree at the ETH, Zürich, working with Jack D. Dunitz at the Laboratory for Organic Chemistry. After 2 years as a postdoctoral student of Larry S. Bartell at the University of Michigan, he joined the Laboratory for Inorganic Chemistry, ETH, Zürich. In 1979 he became Professor of X-ray crystallography at the University of Berne, Switzerland.

Jack D. Dunitz was born in Glasgow in 1923 and studied chemistry at Glasgow University, where he received his B.Sc. and Ph.D. degrees. Following a decade of postdoctoral studies at Oxford, Caltech, NIH, and the Royal Institution, London, he moved in 1957 to the ETH, Zürich, where he has been ever since and where he is Professor of Chemical Crystallography.

(1) J. D. Dunitz, "X-Ray Analysis and the Structure of Organic Molecules", Cornell University Press, Ithaca, NY, 1979.

(2) As of early 1982 the number of published organic crystal structures was more than 30 000 and increasing at the rate of about 4000 per year. Corresponding figures for inorganic structures are not known to us.

(3) For a recent authoritative review, see F. H. Allen, O. Kennard, and R. Taylor, *Acc. Chem. Res.*, preceding paper in this issue. See also the editorial by J. P. Glusker, *Acc. Chem. Res.*, **15**, 231 (1982).

0001-4842/83/0116-0153$01.50/0