is in the C programming language, it is expected that it would be relatively easy to port the software to a wide variety of systems. There are several reasons why this conversion was done:

• To attain hardware independence. Code written in ALPHA BASIC for the Alpha Micro cannot be used on other hardware.

• To permit the use of much more sophisticated word-processing software. The current Alpha Micro systems use word-processing packages that are approximately 10 years old in design, and output is limited to the ASCII character set. It is desirable to permit such document preparation capabilities as generation of mathematical expressions, chemical structures, and tables.

• To improve security. Alpha Micro systems are particularly weak in this area. Although we have had no problems in this regard, improvements would be a wise precaution. The first UNIX-based system was installed for *Analytical Chemistry* in August 1988, followed by a UNIX installation for *Biochemistry* in November 1988. Replacement of all Alpha Micro systems is expected to require 3–4 years.

We have received numerous inquiries about making Peer Review Plus available on personal computers, in particular for IBM PCs and clones and on Apple's Macintosh systems. Because Peer Review Plus software requires a multiuser (the system can be used by more than one user at a time) and multitasking (multiple processes can be simultaneously executed) operating environment—as well as the capability to execute programs in background at predetermined dates and times—these personal computers are not suitable. Moreover, the Peer Review Plus software is relatively complex in operation, in particular the telecommunications aspects, which requires that it be routinely monitored by professional computer staff. Also, we have found that in excess of 100 megabytes of disk capacity is needed to store data for an average size journal. Such storage capacity has only recently become available on personal computers. At such time when personal computers have these capabilities, which may happen with

machines using CPUs based on the Intel 80386 or Motorola 680x0 series processors, we will carefully examine the possibility of making our software available on a wider selection of computers using the UNIX operating system.

REFERENCES AND NOTES

(1) "Edge-notched" cards have sequentially numbered positions marked around the perimeter of the card which can be punched out to indicate a particularly defined characteristic. One card is used for each person (record)—reviewer, author, manuscript, etc. Each hole position is assigned a particular attribute (e.g., position 1 = nucleic acid, position 2 = NMR). If a card is to be classified with a particular attribute, the appropriate hole is notched at the top of the card. To retrieve cards with a specific attribute, the cards are carefully aligned and a "knitting needle" is inserted through the appropriate hole. Cards notched in that position fall out of the stack and are thus retrieved.

(2) Garson, L. R. Computer-aided Selection of Reviewers and Manuscript Control. *Scholarly Publishing*, **1980**(Oct), 65–74.

(3) Neufell, J. L. Personal communication. Systems Analyst, *New England Journal of Medicine*, Waltham, MA, March 17, 1988.

(4) Rice, C. Personal communication. Director of Editorial Office Services, American Physical Society, Ridge, NY, Aug 30, 1989.

(5) The AMOS (Alpha Micro Operating System) operating system imposes a maximum of 512 bytes per record. Consequently, two physical records had to be designated for each logical record.

(6) A 'K' is a unit of 1024 characters, which is 2^{10}. If telecommunications were completely error free, that is, some portions of data did not have to be retransmitted due to errors, one could expect slightly more than 14 512K blocks to be transmitted each minute.

(7) File names on Alpha Micro systems can have a maximum of six characters, followed by a period, followed by an extension of three more characters. By convention, executable files written in BASIC have an extension of RUN (POSTIT.RUN, for example). BASIC source code files have the extension BAS (for example, POSTIT.BAS). Files containing data have the extension DAT (GATHER.DAT).

(8) Information packets consist of 256 bytes of data along with a hash value which are transmitted from one site to the other. The receiving site recalculates the hash value for the data received and compares it to the tranmitted hash value. If the two have identical values, the receiving system transmits a signal to the sending system to send the next packet. If the hash values are not identical, a signal is sent to retransmit the previous packet. The specific communication routines used in our system, written in Assembly language, were developed by Michael Knolls of Nome, AK, under contract to the American Chemical Society. For more details on telecommunication protocols, see reference 9.

(9) Campbell, J. *C Programmer's Guide to Serial Communications*; Howard W. Sams: Indianapolis, IN, 1987.

Using Chemical Bonds To Analyze Data Retrieved from the Inorganic Crystal Structure Database

I. D. BROWN

Institute for Materials Research, McMaster University, Hamilton, Ontario, Canada L8S 4M1

Received August 9, 1989

The programs SINDBAD (batch) and STRUMO (interactive) use a chemical bond model to examine structures retrieved from the Inorganic Crystal Structure Database. The bonding topology at any desired level of bond strength is determined from the bonds (and their valences) calculated directly from the atomic coordinates stored in the database. Hydrogen bonding, disorder, internal stress, defects, and diffusion paths can be examined and modeled by using related techniques.

(1) INTRODUCTION

The four crystal structure databases[1] covering the fields of metals and inorganic, organic, and macromolecular crystals contain data for the 100 000 crystals whose structures are currently known. The data are stored in the form of atomic coordinates, lattice dimensions, and symmetry operators, information that allows one to find the location of all atoms in the crystal. Chemists are more interested in the way atoms bond together, and an array of atoms, whether described by a list of coordinates or displayed on a screen, does not directly provide this information. To appreciate the chemical properties

of the structure, one must know something about the sizes of atoms and the locations of the chemical bonds.

Chemical descriptions of structure are usually based on empirical or highly simplified physical models. Several such models, each with its own advantages and disadvantages, are currently used by solid-state chemists to describe the structures of inorganic compounds. There are models based on chemical bonds and others on the close packing of anions, models that emphasize symmetry relations or that look for structural similarities, and models that use interatomic potentials. Any program to transform the coordinates in the database into
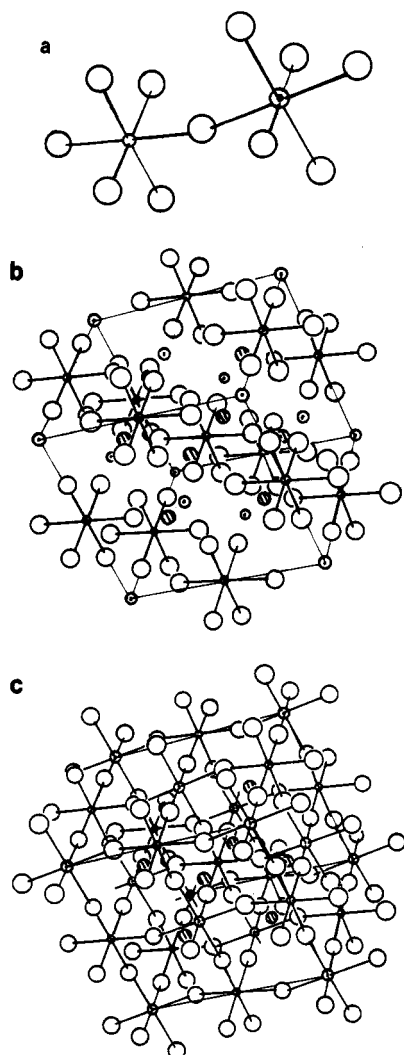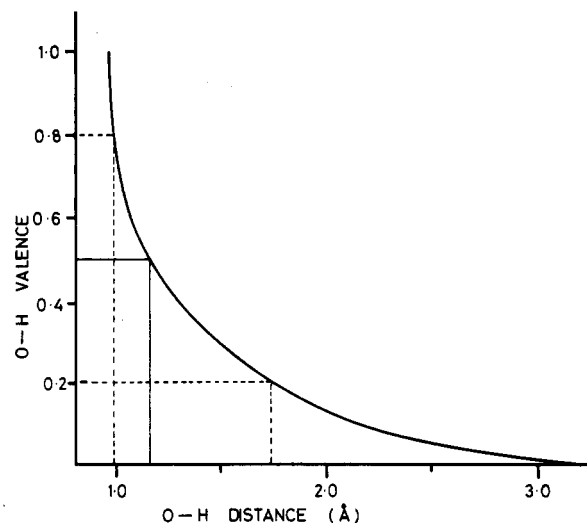
Figure 1. Structure of $BaBiO_3$ showing (a) the geometry around Bi(V) (small circle) and Bi(III) (circle with dot); (b) extended structure showing only bonds stronger than 0.7 vu [bonds to Bi(V)]; (c) extended structure showing all bonds stronger than 0.3 vu (all bonds to Bi).

Figure 2. Variation of bond valence with bond length for H–O bonds. The dotted lines correspond to typical long and short components of the hydrogen bond.

descriptions of the structure by use of such concepts must contain appropriate transformation algorithms. Unfortunately, many chemical concepts are ill-defined, and incompatible models are frequently combined with little regard to logic. Programming a computer to reflect the eclectic thinking of a chemist is not trivial.

This paper describes a suite of programs that transform the atomic coordinates of an inorganic crystal into chemical bonds and use these bonds to examine crystal properties. The computational problem is described in section 2, and recent extensions of the bond model that make it more quantitative for the compounds found in the Inorganic Crystal Structure Database (ICSD)[2] are described in section 3. The programs that are used in this model are described in section 4, where it is shown how they can be used to analyze and model retrieved structures. Some applications are illustrated in section 5.

(2) PROBLEM

In inorganic materials atomic structure is a major determinant of many physical and chemical properties such as lattice dynamics, diffusion and ionic conductivity, and optical, electronic, magnetic, and surface properties. To explore the relation between structure and properties, one needs to model the changes that occur as the chemical composition or atomic positions are varied. There are many possible ways of ex-

ploring these relationships, but the one described here uses the atomic positions to determine the locations of the chemical bonds.

The chemical bond is not a well-defined concept in inorganic chemistry. The difficulty is to know which interatomic distances correspond to bonds. Quite different impressions of a structure can be created depending on how the bonds are selected as shown in Figure 1. Defining the bonds as only strong (covalent) interactions emphasizes the ionic character of a material by breaking the structure into discrete complexes held together by weak electrostatic forces (Figure 1b). Including the weak (ionic) interactions among the bonds emphasizes the connectedness of the structure (Figure 1c).

A practical problem arises from the infinite connectivity usually found in these materials. How much of such a bond network need one consider? Focusing on a small region of the structure highlights the local geometry (Figure 1a), but considering a larger volume emphasizes the overall organization of the structure (Figure 1c).

Another problem that the program must address is that chemists are not well versed in the subtleties of crystallography. An ideal program should present the chemical structure without requiring the user to be familiar with the contents of *International Tables for Crystallography*[3] or even aware of its existence.

The program must therefore meet the following criteria. It must contain algorithms for selecting the interatomic distances that correspond to the user's concept of a bond, it must be able to create the bond network over an appropriate volume to meet the user's needs, and it must not require the user to be familiar with crystallographic symmetry. The next section describes a version of the chemical bond model developed for inorganic materials that allows these criteria to be met.

(3) BOND VALENCE MODEL

In the bond valence model the valence (V_i = formal oxidation number) of an atom (i) is assumed to be distributed between the bonds that the atom forms so that each bond can be assigned a bond strength or *bond valence* (S_{ij}), where

$$V_i = \sum S_{ij} \qquad (1)$$

It is found experimentally that the bond valence is inversely related to the *length* (R_{ij}) of the bond as illustrated for the H–O bond in Figure 2.[4] For most bonds this relationship can be approximated over the range of interest by an equation such as

$$S = \exp((R_0 - R)/0.37) \qquad (2)$$

where R_0 is an empirical constant, values of which have been tabulated for many bonds.[5] Bond valences calculated from experimentally measured distances by using eq 2 are found to obey eq 1 to within 0.1 valence units (vu).[4]

The *valence sum rule* (eq 1) applies to virtually all compounds that contain bonds only between Lewis acids (cations) and Lewis bases (anions) whether or not the bonds are ionic or covalent (both found, for example, in $CsClO_4$). Therefore, without implying that the model is in any way restricted to ionic bonding, one can retain the terminology of the ionic model that usefully distinguishes between the Lewis acid and Lewis base atoms.

Any Lewis acid–base compound can be considered a network of atoms linked by bonds. Since each bond connects a cation at one end to an anion at the other, all the loops in the network will contain an even number of bonds. In graph theory such a network is described as an oriented bipartite graph.[6] The quantitative concepts connected with this network are the valences of the atoms (V_i) and the valences of the bonds (S_{ij}). The valence sum rule (eq 1) is a network equation whose form suggests Kirchhoff's current law for electrical networks. A second law,[7] similar to Kirchhoff's voltage law

$$0 = \sum_{\text{loop}} S_{ij} \qquad (3)$$

is found to hold for most (but not all) Lewis acid–base compounds. In the cases where it does hold (those compounds that do not involve atoms with special stereochemical properties such as stereoactive lone electron pairs), eq 3 together with eq 1 and 2 has been shown to predict bond lengths with an accuracy of about 0.05 Å.

Since most cations form bonds of roughly the same valence in all their compounds, it is convenient to call this quantity the cation's *Lewis acid strength* (S_a). It can be calculated by dividing the cation's oxidation state by its coordination number averaged over all compounds.[8] In a similar manner one can define a *Lewis base strength* (S_b) for an anion, and, since both strengths are estimates of the same bond valence, a bond should only be formed between the anion and cation if

$$S_a = S_b \qquad (4)$$

Equation 4 is known as the *valence matching principle* and is useful in predicting both the structure and the chemical reactivity of a compound: the poorer the valence match between a cation and an anion, the more likely the compound is to be attacked by a molecule, say water, that can form a better match.[9] The Lewis acid strengths of cations correlate with electronegativity,[8] and the Lewis base strengths of anions correlate with the pK_a's, which measure the ability of a base to form a matched bond with the proton, a strong Lewis acid.[9]

Bond valences provide a universal scale by which the strengths of all bonds can be compared. Covalent bonds lie at the strong end of the scale $(S > 0.8 \text{ vu})$, and ionic bonds, hydrogen bonds, etc., lie at the weak end. In the weak limit a bond can approach a valence of 0 (Figure 2), though in practice one can usually ignore bonds with valences less than about 0.03 vu, a necessary approximation since eq 2 does not give $S = 0$ for any finite value of R.

(4) PROGRAMS

The criteria set in section 2 are met by first finding the bonds in the crystallographic asymmetric unit and calculating their valences. Then the symmetry operators of the space group are calculated and used to expand the bonds into a network under the user's direction. The user can select the atom at which the expansion is to begin, the strengths of the bonds that are to be included, and the region into which the network is to be expanded.

The algorithms are incorporated into a system of FORTRAN subroutines that are accessed by either the calling program SINDBAD, designed for processing many entries from the ICSD in batch mode, or the program STRUMO, an interactive program more suited to the examination of individual compounds.

(a) Finding the Bonds. The first step in the analysis of a new entry is to find which atoms are bonded. The algorithms described by Altermatt and Brown[10] make use of the atomic valences (V) given in the ICSD and look for all interatomic distances between cations $(V_c = \text{positive})$ and anions $(V_a = \text{negative})$ having valences between $|V_a|$ and $0.038V_c$. In addition, the program looks for bonds between two anions or between two cations when their atomic valences are less than the maximum values expected for atoms of their group in the periodic table.

Once all the interatomic distances lying within these windows have been found, the program checks that they are chemically reasonable by determining their bond valences and, using eq 2, calculating the bond valence sums around each atom. These are compared to the atomic valences given in the database (cf. eq 1). If the agreement is not satisfactory, either a warning or an error message is printed. Sixty percent of the error-free entries in the ICSD passed the check. A further 30% gave a warning, but in most of these cases the bonds had been calculated correctly. Only for 10% of the entries are the bonds not correctly found, usually because the compounds concerned are not Lewis acid–base compounds of the kind required by the model. Such compounds include polysulfur molecules, boranes, and carbonyls, all compounds for which it is not possible to assign atomic valences in a way that is meaningful for acid–base bonding. In practice, the check may also fail because of poorly determined atomic coordinates or errors in the database. In such cases the program attempts to see if there are identifiable problems with the data (e.g., incorrect symmetry, wrong assignment of hydrogen bonding).

(b) Expanding the Network. Once the asymmetric set of bonds has been found, it is possible to expand the network. The user may elect to expand it using all the bonds or using only those whose valences lie between specified limits. By choosing only bonds with valences greater than, say, 0.8 vu, one can isolate a complex anion or the structural framework of a mineral (compare parts b and c of Figure 1). There are two modes for the expansion. In the first the network is expanded to fill one unit cell. Since any further expansion involves only the trivial operation of translation, this expansion contains all the structural chemistry and is sufficient for many applications. In the second mode the network is expanded around a particular atom to first, second, or any desired order of nearest neighbors, a mode suitable for exploring the bonding environment and atomic properties, such as the NMR chemical shift,[11] that depend on this environment.

Either mode of expansion starts at a user-specified atom, by default the strongest Lewis acid. The program selects from the asymmetric bond set those bonds of the appropriate strength formed by the starting atom. It then proceeds to each of the first neighbors in turn and repeats the selection, multiplying each bond by the appropriate symmetry operator until the expansion is complete. Each atom (node) in the network is identified by the symmetry operator (including translation) that relates it to the corresponding atom in the asymmetric unit. Each bond is identified by its two terminal atoms and is represented by its vector and bond valence. The expansion program does not use the atomic coordinates themselves.

Atoms lying on positions of high symmetry (crystallographic special positions) present a particular problem since it is necessary to know not only their point symmetry group but
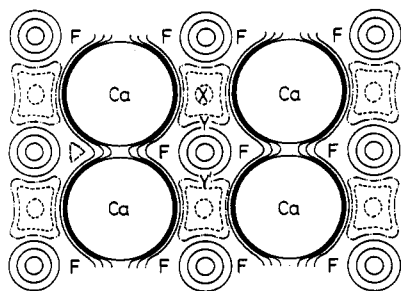
Figure 3. Valence sum map of a (110) section of fluorite. The contour interval is 0.2 vu. Solid contours are greater than 1.0, dashed contour is 1.0, and dotted contours are less than 1.0. X is a cavity that does not provide a good location for an F interstitial; Y and Y' are probable locations of F interstitials (taken from Brown[16]).

Figure 4. Valence sum density maps of forsterite showing possible locations of Mg. The sites occupied in the normal crystal are indicated by crosses. (a) Calculated with $N = 16$ in eq 5 giving an impression of a room temperature atomic density map; (b) calculated with $N = 8$ giving the impression of a high-temperature atomic density map. Note how the empty cavity is incorporated into a diffusion path along the easy diffusion direction, c.

also its orientation. Consequently, the atomic coordinates from the database must be transformed to a predefined setting with a known point group orientation before the bonds are calculated. To do this, the program first generates the space group symmetry operators and special position coordinates in a known order using the algorithm described by Altermatt and Brown.[12] This part of the program effectively reproduces the symmetry tables given in *International Tables for Crystallography*, Vol A.[3]
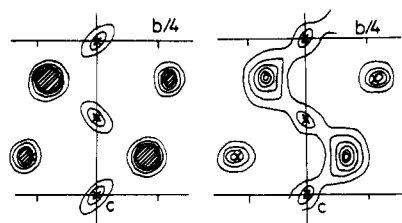
(c) Structure Analysis and Modeling. In a first exploration of a structure one might expand the bond network using only the strong bonds. This will identify the complex ions and determine whether they are discrete or form a framework in one, two, or three dimensions. Since such a complex is usually anionic, the program calculates its Lewis base strength, from which its pK_a can be found as well (by use of eq 4) as the Lewis acid strength of the most likely counterions, a result that can be useful in planning chemical syntheses or postulating mechanisms of geosynthesis.[13] At this stage the anisotropic atomic displacement parameters can be analyzed to determine whether the fragment found, or some part of it, is a rigid body.[14]

Analysis of an inorganic structure cannot be divorced from modeling since many of the properties that one might wish to examine (e.g., diffusion or catalysis) depend on an understanding of *local* atomic environments, rather than the average environments given by diffraction methods. The modeling modules in STRUMO allow the user to explore these local environments.

One may first wish to find which locations in the crystal are accessible to a particular atom to model defects or find diffusion paths. The technique used here is the *valence sum map*,[7,15] calculated as follows: the atom in question is removed from the structure and is placed at an arbitrary point (x) in the crystal. The program calculates all the interatomic distances from x to a distance of about 4 Å. If the atom is in its correct location, the sum of the bond valences corresponding to these distances will be equal to its atomic valence; otherwise, the two numbers will be different. The atom is then moved systematically through the crystal producing a valence sum map whose value (D_x) will be equal to the atomic valence only at those sites (x) in the crystal that are possible locations for the missing atom (Figure 3). At most other places the valence sums will be higher, though inside large cavities they will be lower. Such maps are particularly useful for locating atomic positions and examining possible diffusion pathways.[16,17] By calculation of

$$D^* = D^{-N} \qquad (5)$$

where $N = 16$, they can be made to simulate an atomic density function[7] at room temperature as shown in Figure 4a. Setting $N = 8$ gives the impression of the atomic density at higher temperatures and serves to bring out the diffusion paths. In

cases where the cavity is not too large, these maps can also predict the anisotropy of the Debye–Waller temperature factor as shown by the Mg atom on the c axis in Figure 4a.

When an approximate location has been found for all atoms by using, if necessary, the valence sum map, the structure can be refined. The bonding network is calculated, and the Kirchhoff network equations (eq 1 and 3) are used to predict bond lengths.[7] These in turn are used to refine the atomic coordinates by minimizing the squares of the differences between the predicted bond lengths and those of the model (distance least squares[18]). The lengths can be given unit weights, user-assigned weights, or weights equal to the bond valence, the latter being normally preferred. In some cases the bonds provide more than enough constraints to determine all the free crystallographic parameters, and the modeling can be used to explore the presence of internal stresses;[19] in other cases additional constraints in the form of nonbonded contacts will be needed. The algorithm used to generate these is based on eq 4 of reference 20, and although the algorithm is not entirely satisfactory, the values are not critical since the contact distances are given low weight relative to the bonds. In suitable cases such refinements are able to predict the correct coordinates within a few picometres. A similar technique with a more limited application is the valence refinement. In this case the atomic coordinates of a single atom are varied to give the best fit between the atomic valences and the bond valence sums around the atom and its bonded neighbors.

(d) Graphics. STRUMO comes with its own graphics package, or it can be used to prepare files that can be directly read by CHEM-X. The graphics package is currently limited in scope and in the terminals it supports. It differs from other graphics packages in its ability to define the bonds to be plotted in terms of their bond valences, permitting views of the structure that emphasize either the strong bonding within complexes or the overall bonding connectivity. Figure 1 was prepared with the aid of this package.

(5) APPLICATIONS

(a) Locating Hydrogen Atoms. In many crystal structure analyses the weak scattering of X-rays from hydrogen atoms makes them difficult to locate. In these cases dummy hydrogen atoms are included in the ICSD for completeness, and an attempt is made to identify the anion to which the hydrogen is attached. SINDBAD is designed to check anions whose valence sums are low and, in cases where dummy hydrogen atoms are present, to decide how many hydrogen atoms are attached to each anion. A typical O–H bond has a valence of 0.8 vu, so a deficiency of 1.6 vu on an oxygen atom indicates a water molecule and a deficiency of 0.8 a hydroxyl group.[21]

(b) Assignment of Occupation Numbers and Oxidation States. When a crystallographic site is occupied by two different elements or the same element in two different oxidation states, it is not always easy to determine the degree of occu-

pation by the different species by using X-ray diffraction, particularly when the elements lie close to each other in the periodic table. Differences in the bond lengths that lead to differences in the bond valence sums provide a ready estimate of the distribution. A site occupied by both Al and Si atoms that gives a valence sum of 3.25 would be 25% occupied by Si and 75% by Al. The distribution of valence (electrons) in mixed-valence compounds (vanadium oxides, ceramic superconductors) can be similarly determined with the check that the valence sums around the anions should be close to their known oxidation states.[22] Modeling the structure by using the Kirchhoff equations and distance least squares under the assumption of different oxidation-state distributions can provide additional insights.[19]

(c) Internal Stress and Instabilities. In crystals whose structure is heavily constrained because there are few crystallographic degrees of freedom, the valence sum rule is often not well obeyed. This can be a sign of the presence of internal stresses. The cubic perovskite structure (ABO_3), for example, consists of an infinite array of cubic cages of composition BO_3, each with an A atom at the center. Although many different combinations of elements can appear on the A and B sites, the structure has only one crystallographic degree of freedom, the unit cell length, and this cannot simultaneously satisfy both atom sizes. In most perovskites the size of the A atom does not match the size of the cavity formed by the BO_3 cage, with the result that the bonds around one of the cations (A or B) are too long and around the other too short. This is reflected in valence sums that are too small or too large, respectively. If the deviations are more than about 0.2 vu, the structure collapses, and a new lower symmetry phase is formed.[22] In the case of perovskites the collapsed phases are either ferroelastic or ferroelectric depending on whether the undersize cation is A or B. Calculation of the valence sums in cubic perovskites can draw attention to the internal stresses and give an indication of an impending instability. For example, the perovskite related superconductor $Ba_2YCu_3O_x$ has a perfect match for $x = 6.5$ but is internally stressed in different senses when $x = 6$ or 7.[19]

(d) Modeling Defects. Defects are important structural features that affect such properties as diffusion, mechanical strength, and catalytic and chemical reactivity, but diffraction techniques only reveal the average structure of the whole crystal. Defects are rarely visible and never resolved. The structure of an individual defect can be modeled in STRUMO, starting with the average structure taken from the ICSD. The valence sum map allows one to find the possible locations for a defect in a crystal by finding those sites that have favorable structural chemistry (positions Y and Y′ but not X in Figure 3). Since defects are important in diffusion, an examination of the easy diffusion paths revealed by the valence sum map will help in the selection of appropriate defect sites. The defect is modeled by removing an atom from its normal site and moving it to one of the sites identified from the valence sum map. Calculation of the bond network around the defect and application of the Kirchhoff equations allow the bond lengths to be predicted, and these in turn can be used in a distance least squares refinement to determine the extent of the relaxation around the defect.[16] The success of the refinement process and the agreement between the valence sums and the atomic valences of the refined defect structure indicate its stability.

(e) Modeling Surface Structures. Closely related to the modeling of defects is the modeling of surface structures. The bond valence model has been shown to give correct valence sums for O atoms chemisorbed onto metallic surfaces[23] and has been used in modeling catalytic mechanisms on ceramic materials.[24] The procedures for modeling surfaces, not yet

incorporated into STRUMO, would closely follow those used for modeling defects and could be used to explore the binding of reactants to catalytic surfaces. In this case the valence matching principle could also be used with the base (or acid) strength of the surface atoms to determine the species most likely to bond to them.

(f) Other Potential Applications. A number of bond properties, such as coefficients of thermal expansions[25,26] and force constant,[27] have been shown to correlate directly with bond length or bond valence. This should permit the direct calculation of such lattice properties as thermal expansion and lattice vibrational modes from structures retrieved from the ICSD.

(6) AVAILABILITY

The programs are still under development, but a working version of STRUMO is available from the author. It is written in machine-independent FORTRAN that has been tested on VAX, NORD, and CYBER (with virtual memory) computers (a few well-identified lines of code may have to be changed). The source code for STRUMO comprises 17 files occupying just over 1 MB of memory. SINDBAD employs most of the same code but has a different calling program. Two small data files (73 kB) are also needed, and a noninteractive graphics routine (SIGMA and TEKTRONIX 4000 terminals only) occupying 145 kB is available.

STRUMO is menu driven and has help menus that describe the various features of the program. The program is designed to be self-directing; it can read in crystal structure data from ICSD (CISTI, Ottawa format and the archival Z format), from the terminal, and from data in the Standard Crystallographic File Structure[28] or files produced by SHELX.[29] There is currently no operating manual available.

SINDBAD is command driven and is designed for batch checking of entries from ICSD. It is less well documented and does not include most of the modeling features.

Many parts of the program (e.g., symmetry handling and bond length calculation) have been well tested, and the program has been designed to recover from errors, but no warranty is given that the program will always work as intended.

REFERENCES AND NOTES

(1) *Crystallographic Databases*; Allen, F. H., Bergerhoff, G., Sievers, R., Eds.; International Union of Crystallography: Bonn, Cambridge, and Chester, 1987.
(2) Bergerhoff, G.; Hundt, R.; Sievers, R.; Brown, I. D. *J. Chem. Inf. Comput. Sci.* **1983**, *23*, 66–69.
(3) *International Tables for Crystallography*; Hahn, T., Ed.; Reidel Publishing: Dordrecht, 1983; Vol. A.
(4) Brown, I. D.; Shannon, R. D. *Acta Crystallogr.* **1973**, *A29*, 266–282.
(5) Brown, I. D.; Altermatt, D. *Acta Crystallogr.* **1985**, *B41*, 244–247.
(6) Rutherford, J. S. *Acta Crystallogr.* (in press).
(7) Brown, I. D. *Phys. Chem. Miner.* **1987**, *15*, 30–34.
(8) Brown, I. D. *Acta Crystallogr.* **1988**, *B44*, 545–553.
(9) Brown, I. D. in *Structure and Bonding in Crystals*; O'Keeffe, M., Navrotsky, A., Eds.; Academic Press: New York, 1981; Volume II, Chapter 14, pp 1–30.
(10) Altermatt, D.; Brown, I. D. *Acta Crystallogr.* **1985**, *B41*, 240–244.
(11) Sherriff, B. L.; Grundy, H. D. *Nature* **1988**, *332*, 819–822.
(12) Altermatt, U. D.; Brown, I. D. *Acta Crystallogr.* **1987**, *A43*, 125–130.
(13) Hawthorne, F. C. *Am. Mineral.* **1985**, *70*, 455–473.
(14) Rosenfield, R. E., Jr.; Trueblood, K. N.; Dunitz, J. D. *Acta Crystallogr.* **1978**, *A34*, 828–829.

(15) Waltersson, K. *Acta Crystallogr.* **1978**, *A34*, 901–905.
(16) Brown, I. D. *Solid State Ionics* **1988**, *31*, 203–208.
(17) Garrett, J. D.; Greedan, J. E.; Faggiani, R.; Carbotte, S.; Brown, I. D. *J. Solid State Chem.* **1982**, *42*, 183–190.
(18) Villiger, H. *DLS Manual*; Inst. Krist. Petr. ETH: Zurich, 1969.
(19) Brown, I. D. *J. Solid State Chem.* (in press).
(20) Brown, I. D. *Acta Crystallogr.* **1976**, *A32*, 24–31.
(21) Donnay, G.; Allman, R. *Am. Mineral.* **1970**, *55*, 1003–1015.
(22) Brown, I. D. *Chem. Soc. Rev.* **1978**, *7*, 359–376.
(23) Mitchell, K. A. R.; Schlatter, S. A.; Sodhi, R. N. S. *Can. J. Chem.* **1986**, *64*, 1435–1439.
(24) Ziółkowski, J. *Thermochim. Acta* **1987**, *110*, 333–336.
(25) Kahn, A. A. *Acta Crystallogr.* **1976**, *A32*, 11–16.
(26) Hazen, R. M.; Prewitt, C. T. *Am. Mineral.* **1977**, *62*, 309–315.
(27) Badger, R. M. *J. Chem. Phys.* **1934**, *2*, 128–131; **1935**, *3*, 710–714.
(28) *Acta Crystallogr.* **1988**, *A44*, 232.
(29) Sheldrick, G. SHELX76, Program for Crystal Structure Determination; University of Cambridge.

Problem Solving with the *Beilstein Handbook*[†]

REINER LUCKENBACH* and JOSEF SUNKEL

Beilstein Institute, Varrentrappstrasse 40–42, D-6000 Frankfurt/Main 90, West Germany

Received March 2, 1989

Beilstein is unique among handbooks of organic chemistry in that it provides a collection of critically examined and carefully reproduced data on the known organic compounds. In this respect, *Beilstein* is superior to all other straight bibliographical documentation and series of abstracts. Moreover, the *Beilstein Handbook* is the world's most extensive collection of physical data on organic compounds in printed form. While previous publications mainly concentrated on how to search for specific *compounds* in the *Handbook*, this paper provides some representative examples of how to find specific *information and factual data* for the following topics: nomenclature; physical data of "families" of chemical compounds; stereochemical assignments based on physical data; interpolation and extrapolation of physical data; oxidation reactions of a given substance; development of synthesis strategies for known and unknown compounds; "partial structure retrieval" ("similarity searches").

INTRODUCTION

The "Beilstein Information System" is the collective term for the database Beilstein Online and the printed reference work *Beilstein Handbook of Organic Chemistry*. Together they form the most comprehensive source of data and information on organic compounds to be found anywhere.

The purpose of this paper is to provide some practical hints to optimize the use of the Beilstein Information System. In particular, the use of "intuitive guesswork" to derive information from the data available is illustrated with practical examples. Although all the material presented here is taken from the *Beilstein Handbook*, many of the ideas involved are just as relevant to database searches.

1. THE *BEILSTEIN HANDBOOK*[1-11]

The *Beilstein Handbook of Organic Chemistry*, commonly referred to as *Beilstein*, is a collection of important published data on the preparation and properties of organic compounds. With well over 350 single volumes comprising more than 270 000 printed pages, the *Beilstein Handbook* today constitutes the world's largest collection of physical data on organic compounds in printed form!

The *Handbook* is published in series, each covering the literature of a certain period (see Table I). All volumes from E V onward are published in English.

Each series consists of 27 (nominal) volumes, almost all of them subdivided into several subvolumes in the more recent series, in which the individual compounds are arranged according to the Beilstein System.[1,7,8] This system classifies all organic compounds according to their structure, and each volume corresponds to a particular class of compounds (see Tables II and III).

Table I. Series of the *Beilstein Handbook*

series	abbrev	period of literature covered	color of label on spine
Basic Ser.	H	up to 1910	green
Suppl. Ser. I	E I	1910–1919	dark red
Suppl. Ser. II	E II	1920–1929	white
Suppl. Ser. III	E III	1930–1949	blue
Suppl. Ser. III/IV	E III/IV[a]	1930–1959	blue/black
Suppl. Ser. IV	E IV	1950–1959	black
Suppl. Ser. V	E V	1960–1979	red

[a] Volumes *17–27* of *Suppl. Ser.* III and IV, covering the heterocyclic compounds, are combined in a joint issue.

Table II. Main Divisions of the *Beilstein Handbook*

main division	vol. no.
(1) acyclic compounds	*1–4*
(2) isocyclic (carbocyclic) compounds	*5–16*
(3) heterocyclic compounds	*17–27*

The various *Beilstein* volumes contain the research data reported within the literature over a specific period, critically sifted and correlated in a reliable and logical form and assessed in the light of current chemical knowledge. The editors are careful to point out errors in the published data, to direct attention to the doubtfulness of some published statements, and to check assertions of speculative nature against subsequent findings. This frequently involves citing very recent publications.

To the organic chemist as well as to every scientist working in the field of organic chemistry, *Beilstein* is an indispensable source of information for which there is no substitute. The alert user will find it can save him from embarking on false trails and stimulate him to further research.

For each compound described in *Beilstein* the following aspects are covered: structural formula; compound name(s); molecular formula; constitution and configuration; natural