

# Prediction of Possible Crystal Structures for C-, H-, N-, O-, and F-Containing Organic Compounds

James R. Holden, Zuyue Du, and Herman L. Ammon\*

Department of Chemistry and Biochemistry, University of Maryland, College Park, Maryland 20742

Received 1 June 1992; accepted 11 November 1992

A procedure is reported for the prediction of dense crystal structures of C-, H-, N-, O-, and F-containing organic compounds in the primitive triclinic, monoclinic, and orthorhombic space groups with  $Z \leq 4$ . The crystal environments of molecules in 242 crystal structures have been analyzed to determine the common coordination sphere patterns. This led to the development of the MOLPAK (MOlecular PAckKing) program, which uses a rigid-body molecular structure probe to build packing arrangements (possible crystal structures) in the various space groups. A MOLPAK search, which involves the investigation of all unique orientations of a central molecule and the construction of the appropriate coordination patterns about the central molecule, provides a 3-D map of minimum unit cell volume as a function of the orientation of the central molecule. MOLPAK uses a repulsion-only potential and a preset threshold to place molecules in contact with each other. The 5–10 smallest volume packing arrangements from a search are subjected to a lattice energy minimization refinement with the WMIN program to yield possible crystal structures. The results are described from the analyses of several known compounds starting with the crystal molecular structures as the MOLPAK search probes in the  $P\bar{1}$ ,  $P2_1$ ,  $P2_1/c$ , and  $P2_12_12_1$  space groups. In addition, several examples are given in which the search probes were created by AM1 geometry optimization of preliminary molecular models. More extensive data are given in supplementary tables. © 1993 by John Wiley & Sons, Inc.

## INTRODUCTION

Over the last several years, procedures have been under development in our laboratories for the prediction of the crystal structures of unknown organic compounds. Portions of this work are summarized in various laboratory reports; an introduction to molecular packing analysis was given in a 1979 report by Holden and Dickinson<sup>1</sup> and an overview of the procedures discussed in this article was published in 1987 by Cromer, Ammon, and Holden.<sup>2</sup> Our interest in crystal structure prediction is a consequence of efforts to develop accurate methods for the prediction of the solid-state (crystal) densities of energetic materials (explosives and propellants). Because of the historic importance of crystal density estimation to our crystal structure prediction research, the following paragraphs introduce density prediction and demonstrate the need for accurate density methods for energetic materials.

The search for new energetic materials would be facilitated by the availability of reliable computational methods for the identification of compounds expected to have significant advantages over materials currently in use. The estimation of properties such as shock sensitivity and detonation performance would enable one to restrict possibly compli-

cated and expensive laboratory syntheses to only the most promising substances. Density is one of the most significant physical properties of an energetic material<sup>3</sup>; important performance parameters such as detonation velocity and pressure increase with density. For example, the Chapman–Jouget pressure behind a detonation front is proportional to the square of the density<sup>4</sup>; thus, a 5% increase in density from 1.9 to 2.0 g cm<sup>-3</sup> would correspond approximately to an 11% increase in the detonation pressure.

Heretofore, the estimation of crystal density from a molecular structure has been accomplished primarily by the summation of appropriate atomic or group volumes to yield a crystal-molecular volume that can be converted readily into an effective crystal density. The available so-called “group or volume additivity” procedures<sup>5</sup> are basically similar and differ mainly in the number and types of crystal structures contained in the data bases used to derive the atomic/group volume parameters. Volume additivity does not specifically recognize crystal packing efficiency\* or molecular conformation and will yield identical calculated densities for conformational isomers and different compounds with the same functional group composition. A case in point involves

\*Cady's procedure<sup>5a</sup> for density estimation included a packing coefficient correction related to the fraction of hydrogen in a compound.

\*Author to whom all correspondence should be addressed.

the observed crystal densities of RDX [1,3,5-trinitro-1,3,5-triazacyclohexane,  $(\text{CH}_2\text{NO}_2)_3$ ],  $\alpha$ -HMX [1,3,5,7-tetranitro-1,3,5,7-tetraazacyclooctane,  $(\text{CH}_2\text{NO}_2)_4$ ], and  $\beta$ -HMX of 1.806,<sup>6</sup> 1.839,<sup>7</sup> and 1.902 g cm<sup>-3</sup>,<sup>8</sup>; a volume additivity calculation<sup>5c</sup> predicts a density of 1.783 g cm<sup>-3</sup> for each.

It occurred to us that density estimations, which take account of molecular structure and conformation and crystal packing efficiency, could be achieved by predicting possible crystal structures; this led to the development of procedures described herein for C-, H-, N-, O-, and F-containing organics. Possible crystal structures could be constructed as more or less random arrays of molecules (Monte Carlo) followed by energy minimization refinement with wide convergence properties (simulated annealing). The central step in the overall procedure reported in this article has similarities to the cluster-building approach of Gavezzotti.<sup>9</sup> Here, we describe the development and details of the MOLPAK program (MOlecular PACKing), which is designed to search for packing patterns of minimum volume for molecules with fixed conformations (i.e., rigid bodies), and the refinement of the crude packing arrangements with the WMIN program<sup>10</sup> to give possible crystal structures.

## METHODS

### Development and Overview of the MOLPAK Program

The unit cell is the basic building block of the crystal lattice and the specification of its size, shape, and contents completely describes the crystal structure. A unit cell usually contains internal symmetry (crystallographic symmetry elements). The complete crystal structure can be fully described by specifying a smaller part of the structure, the asymmetric unit, and the symmetry operations (equivalent positions of a space group) that relate each asymmetric unit to the other asymmetric units in the unit cell. The asymmetric unit contains a single molecule in most crystal structures and we concentrated on this pattern in the development of crystal packing methodology.

The usual way to describe a crystal structure is to specify the unit cell dimensions and the space group and then the orientation and placement of the molecules within the unit cell. This is the order in which the information most often is acquired when a crystal structure is determined by X-ray or neutron diffraction experiments. However, an equivalent way to describe a crystal structure is to place the centroid of one molecule at the origin of an orthogonal coordinate system and specify its molecular surroundings. This alternative is especially powerful in crystals with one molecule per asymmetric unit

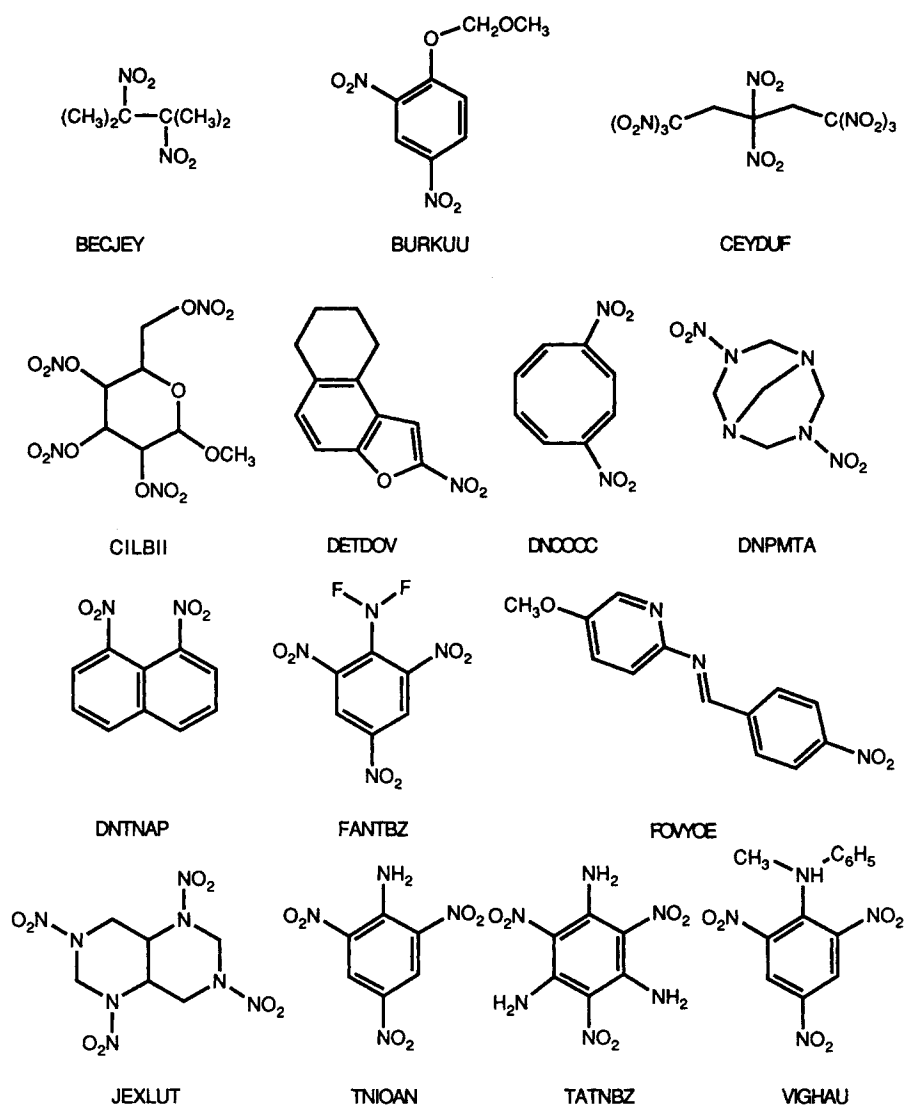
because the orientations of the surrounding molecules are related to the central molecule by the crystallographic symmetry operations. Here, the number of molecules per unit cell (*Z*) is equal to the number of symmetry operations (equivalent positions) that interrelate pairs of molecules in the unit cell.

MOLPAK at present is limited to crystal structures with one molecule per asymmetric unit and to the symmetry operations of inversion through a point, mirror plane, glide plane, two-fold rotation axis, and two-fold screw axis. The program handles the triclinic, the *Z* = 2 monoclinic, and the primitive *Z* = 4 monoclinic and orthorhombic space groups. Seventy-one percent of the known C-H-N-O-F-containing crystal structures with one structural unit per asymmetric unit, given in the January 1992 Cambridge Structural Database,<sup>11</sup> fall within MOLPAK's capabilities.\*† All unit cell symmetry operations involve a change of orientation about a fixed centroid followed by translation of the centroid to a new position. For example, a glide plane is a mirror plane coupled to a translation parallel to the plane and a screw axis is a two-fold rotation plus a translation parallel to the axis. The three types of orientation changes are inversion, plane, and axis. In the following discussion, the term "identity" refers to a translational symmetry operation that does not alter the molecular orientation.

All molecules are equivalent for the structures being considered. That is, any molecule could be the central molecule without changing the structure, and the surroundings of all molecules are equivalent. The "view" in all directions from all identity locations is identical to the view from the central molecule at the origin. The view from a plane location is inverted through a mirror plane and that from an axis is rotated about a two-fold axis in comparison to the view from an identity location. These properties restrict the locations of plane- and axis-related molecules. In the monoclinic space group *Pc*, for example, with equivalent positions of *x*, *y*, *z* and *x*,  $-y$ ,  $1/2 + z$ , the crystal can be envisioned as consisting of 2-D grids of identity and plane (mirror image) molecules parallel to the unit cell *a* and *c* axes. The surround-

\*A search of the January 1992 Cambridge Structural Database for C-H-N-O-F-containing molecules with one molecular unit (i.e., cation-anion and other complexes, solvates, etc. were eliminated) per asymmetric unit and with atomic coordinates given in the data base gave the following information: 17,118 total entries; 16,545 (82.9%) of the total are in triclinic, monoclinic, or orthorhombic space groups; 85.6% of the 16,545 have *Z* = 1, 2, or 4 and are MOLPAK accessible. The four most common of the MOLPAK-accessible space groups are *P2<sub>1</sub>/c* (36.0%), *P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>* (19.5%), *P1* (12.6%), and *P2<sub>1</sub>* (10.8%). Within the MOLPAK-inaccessible 16.4%, the two most common space groups are the monoclinic *C2/c* (5.4%) and orthorhombic *Pbca* (4.8%), both of which have *Z* = 8.

†The Cambridge Structural Database reference codes are used for compound reference in this paper. The structures are shown in Figure 1.



**Figure 1.** Structures and Cambridge Structural Database reference codes for compounds referred to in this article.

ings of all molecules on the identity and plane grids are equivalent and related by a mirror plane normal to the unit cell **b** axis.

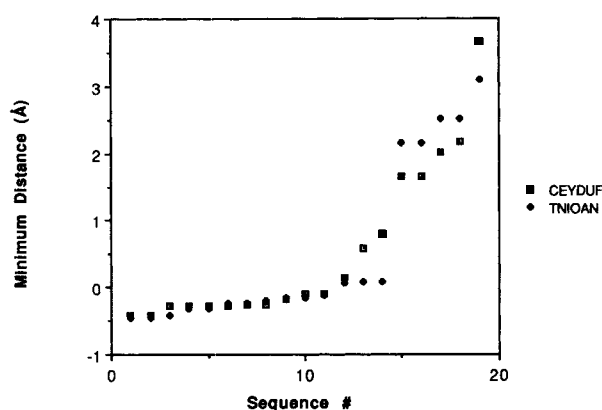
The "coordination sphere" of a molecule consists of molecules that are in contact or close to van der Waals contact with the central molecule. Kitagorodsky<sup>12</sup> defined the packing coefficient (PC = molecular volume/unit cell volume per molecule) as a measure of molecular packing efficiency; there presumably is a relationship between PC and the molecular coordination environment of a central molecule. Because  $\rho_{\text{crystal}} = \text{PC} \times \rho_{\text{molecule}}$  ( $\rho_{\text{crystal}}$  = normal crystal density;  $\rho_{\text{molecule}}$  = molecular mass/molecular volume), the highest-density crystals will be associated with efficient crystal packing and high molecular densities.

### Molecular Packing Investigations

Two programs, MOLCON (MOLEcular CONformation) and MOLPAN (MOLEcular Packing ANALYSIS),

were written to assist the identification and examination of the coordination spheres in the crystal structures of energetic compounds and related materials. A total of 242 unique crystal structures were examined; some of the structures contained two molecules per asymmetric unit for a total of 267 coordination spheres. A complete summary of the MOLPAN analysis is given in Table IS of the supplementary data. The PCs ranged from 0.78–0.63 (the van der Waals radii used were C = 1.70, H = 1.10, N = 1.55, O = 1.52, and F = 1.50 Å<sup>13\*</sup>). Information on the makeup of the coordination spheres was obtained in the following manner. For each structure, the minimum distances between the van der Waals spheres of the central molecule and each of the surrounding molecules were calculated, sorted in as-

\*The H van der Waals radius of 1.10 Å is the average of 1.2<sup>13a</sup> and 1.0 Å<sup>13b</sup>; see also ref. 13c. The C, N, O, and F radii are from ref. 13d.



**Figure 2.** Minimum distance vs. sequence number for the coordination environments of two crystal structures: TNIOAN ( $P2_1/c$ ) and CEYDUF ( $P2_1/a$ ). The six letter codes are Cambridge Structural Database reference codes. This figure illustrates one structure (TNIOAN) with a well-defined sharp break between sequence numbers 14 and 15 and another (CEYDUF) with a less clear break to define the coordination number.

cending order, and plotted as distance vs. sequence number. Typical plots are shown in Figure 2. The distance vs. sequence number curve slowly increases until, in most cases, the occurrence of a sharp break at about sequence 12–14 (see Table I). The average magnitude of the break was about 1.55 Å and the maximum distance from a coordination sphere molecule to the central molecule averaged about 0.42 Å. There were, for example, 152 cases with a break between 14 and 15; the average size of the break and the average maximum distance were 1.74 and 0.43 Å, respectively. There were 42 cases with a break at 12–13 and 27 with a break at 13–14. Only 6 cases had a break at 10 or less and in 38 the break occurred at 15 or more. These findings, together with studies of the patterns of the locations of the centroids of molecules in the coordination spheres of various structure types, provided the background for the de-

velopment of the molecular pattern search techniques described in the next section.

The MOLCON/MOLPAN analysis established 14 as the most probable number of molecules in a coordination sphere.<sup>14†</sup> Further, the composition of the sphere (identity-, plane-, axis-, and center-related molecules) is characteristic of each space group, and the 3-D relationships between the molecules in the coordination spheres fall into patterns and subpatterns. Crystal structures can be obtained by building these patterns around a central molecule. When more than one pattern has been identified for a space group, each of the patterns is treated as a separate case. The methods derived for the  $P\bar{1}$  space group and two of the  $P2_1/c$  subclasses are described below. The characteristic patterns for the four most common organic space groups are illustrated in Figures 3–7, and pattern information for the major primitive triclinic, monoclinic, and orthorhombic space groups with  $Z \leq 4$  is given in Table II.

The coordination number of 14 can be readily seen in the triclinic space group  $P\bar{1}$ . The 14 surrounding molecules (all identities) would consist of 2 along each of the 3 axes (6), 2 along each of the 3 shortest face diagonals (6), and 2 along the shortest body diagonal for a total of 14. A more useful view of this coordination in terms of pattern building shows that the centroids of six molecules lie in a plane surrounding the central molecule with four molecules in offset planes above and below the central plane. In the case of  $P\bar{1}$  (Fig. 3), the 8 molecules above and below the central 6-molecule plane of identities (Is) are inversion-related molecules (Cs). A considerably more complicated pattern is shown in Figure 4 for the Al category  $P2_1/c$  space group. This is clas-

†A crystal packing potential energy approach used to determine the number of molecules in the coordination spheres of 1440 hydrocarbons, oxahydrocarbons, and azahydrocarbons averaged 12.3.

**Table I.** Coordination sphere summary from MOLPAN analysis of 267 crystal structure packing environments.

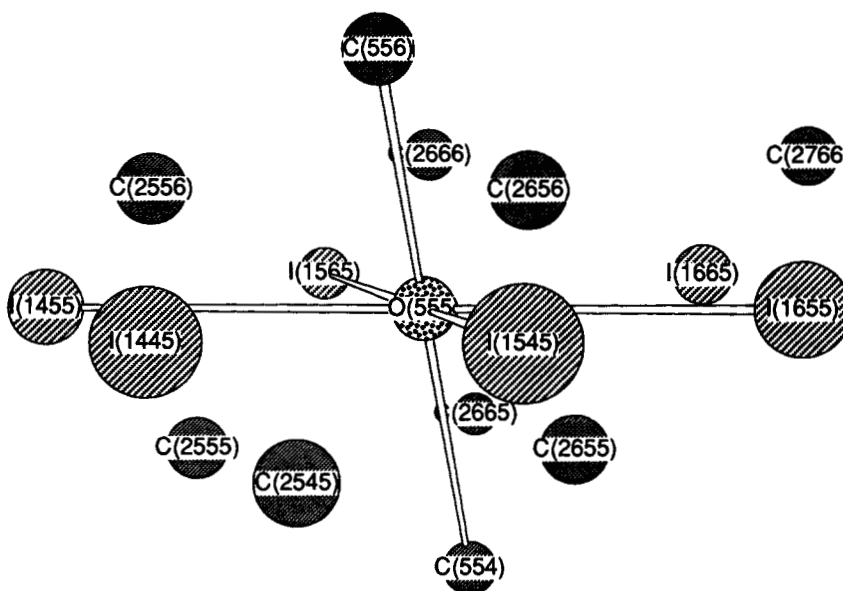
$N^a$	Break location <sup>b</sup>	Break distance (Å) <sup>c</sup>	Molecular separation (Å) <sup>d</sup>
1	8–9	1.80	0.11
1	9–10	2.40	0.09
4	10–11	1.03	–0.12
2	11–12	1.62	0.17
42	12–13	1.26	0.10
27	13–14	1.33	0.30
152	14–15	1.74	0.43
18	15–16	1.18	0.80
20	16–17	1.43	0.97

<sup>a</sup>Number of occurrences.

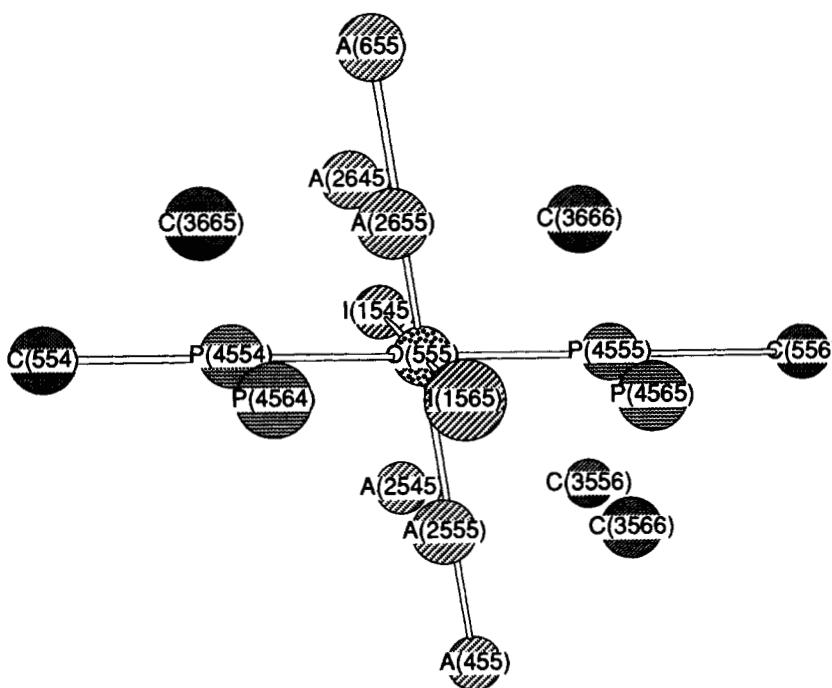
<sup>b</sup>Location of the large distance break.

<sup>c</sup>Average magnitude of the large distance break.

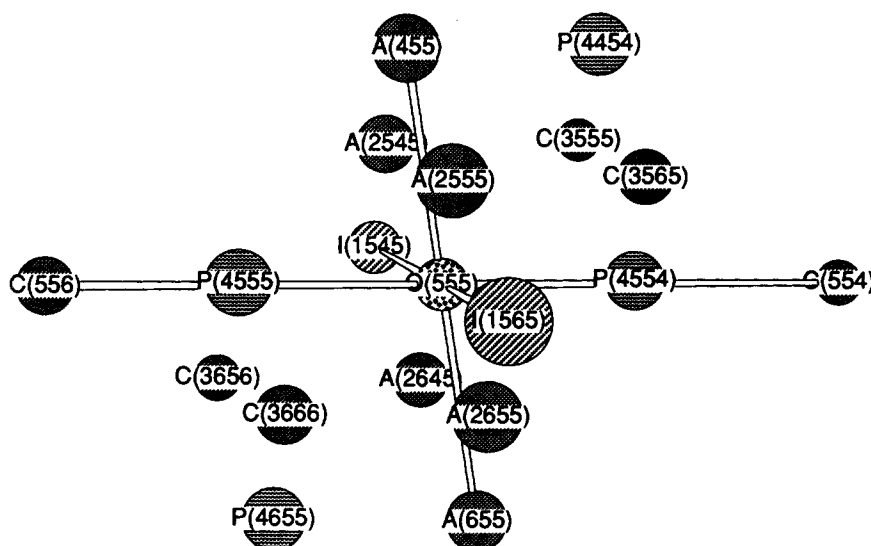
<sup>d</sup>Average maximum distance between the van der Waals envelopes of a coordination molecule and the central molecule. For example, the seventh line of the table indicates that a large break occurs in 152 of the distance vs. sequence number plots between sequence numbers 14 and 15; the average size of the break (distance from 14 to 15) was 1.74 Å and the maximum distance between the van der Waals envelopes of the 14 molecules and the central molecule averaged 0.43 Å.



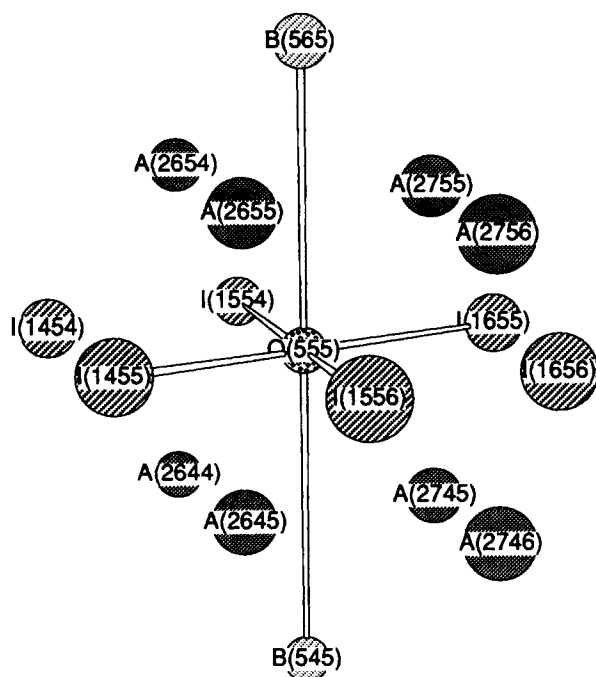
**Figure 3.** Coordination sphere for an AB  $P\bar{1}$  structure TATNBZ. I and C refer to molecules related to the central molecule [O(555)] by identity and center of symmetry operations. The numbers in parentheses are similar to ORTEP ADC codes. The 555 for the central molecule [O(555)] is equivalent to 0,0,0. Positions for the  $\pm c$  axes are indicated as C(556) and C(554). The points that are part of the coordination spheres have four-digit ADCs: a 1 or 2 for the first digit refers to the  $x, y, z$  and  $-x, -y, -z$  symmetry operations, respectively. The C(2545) point is related to the central molecule by  $-x, -1-y, -z$ . The example shows the most common pattern for  $P\bar{1}$  with six Is in a plane (horizontal) and four Cs in parallel planes above and below the I plane. The drawings in Figs. 3–7 were made with the Chem3D program<sup>26</sup> on Macintosh IIfx and IIsi workstations.



**Figure 4.** Coordination sphere for an AI category  $P2_1/c$  structure: VIHGAU. Points along the  $\pm a$  [A(655) and A(455)] and  $\pm c$  [C(556) and C(554)] axes are shown. The I, C, A, and P points associated with four-digit ADCs refer to identity, center, axis ( $2_1$ ), and plane (c glide) symmetry operations; the first digit in the four-digit ADC codes refer to symmetry operations of  $x, y, z, -x, 1/2 + y, 1/2 - z, -x, -y, -z$ , and  $x, 1/2 - y, 1/2 + z$ . This coordination is characterized by Is [I(1565) and I(1545)] along the  $b$  axis and two sets of Ps [P(4555) and P(4554); P(4564) and P(4565)] parallel to  $c$  and in the  $bc$  plane.



**Figure 5.** Coordination sphere for an AK category  $P2_1/c$  structure: FANTBZ. Points along the  $\pm a$  [A(655) and A(455)] and  $\pm c$  [C(556) and C(554)] axes are shown. The I, C, A, and P points associated with four-digit ADCs refer to identity, center, axis ( $2_1$ ), and plane (c glide) symmetry operations; the first digit in the four-digit ADC codes refer to symmetry operations of  $x, y, z, -x, 1/2 + y, 1/2 - z, -x, -y, -z$ , and  $x, 1/2 - y, 1/2 + z$ . This coordination is characterized by Is [I(1565) and I(1545)] along the  $b$  axis and one set of Ps [P(4555) and P(4554)] parallel to the  $c$  axis. The P4454 and P4655 are, respectively, below and above the  $ac$  plane.

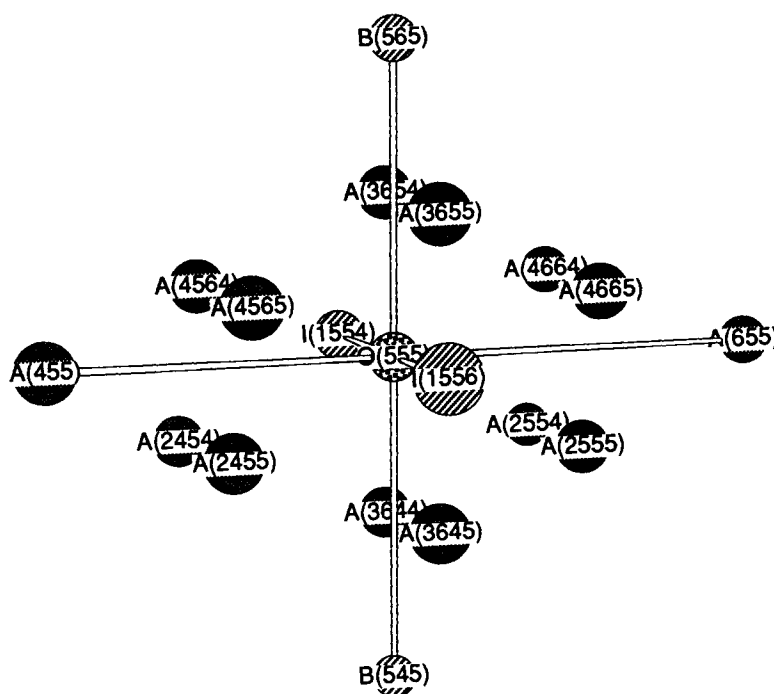


**Figure 6.** Coordination sphere for an AH category  $P2_1$  structure: JEXLUT. Points along the  $\pm b$  [B(565) and B(545)] axes are shown. The I and A points associated with four-digit ADCs refer to identity and axis ( $2_1$ ) symmetry operations; the first digit in the four-digit ADC codes refers to symmetry operations of  $x, y, z$  and  $-x, 1/2 + y, -z$ . This coordination is characterized by Is [I(1655) and I(1455); I(1556) and I(1554)] along the  $a$  and  $c$  axes.

sified as having Is along the unique  $b$  axis and two sets of Ps along the  $b$  axis and parallel to the  $bc$  plane. The AM category  $P2_1/c$  space group (not shown) has Is along  $a$  (or  $c$ ) and two sets of Ps parallel to the  $c$  axis and parallel to the  $ac$  plane.

The construction and subsequent refinement of a 3-D pattern yields a possible crystal structure. Most patterns are formed in three steps. First, a central molecule is placed at the origin of a Cartesian system\* and a line of molecules is established along axis-1 by moving a second molecule (in the same orientation as the central molecule) toward the central molecule until some repulsion criterion is met. The second step establishes a 2-D grid by moving a line of molecules toward the central molecule from the direction of axis-2 to some repulsion threshold. This step may require sliding the line of molecules parallel to axis-1 to locate the minimum distance that yields the repulsion threshold. In the third step, the 2-D grid obtained in step two is placed parallel to the central grid and moved toward the central molecule along axis-3 to find a minimum distance consistent with the repulsion threshold; the minimum axis-3 distance at the repulsion threshold may re-

\*The construction of a packing arrangement is performed in a Cartesian system. The three orthogonal axes will be referred to as axis-1, axis-2, and axis-3.



**Figure 7.** Coordination sphere for an AQ category  $P2_12_1$  structure: CILBII. Points along the  $\pm \mathbf{b}$  [B(565) and B(545)] axes are shown. The I and A points associated with four-digit ADCs refer to identity and axis ( $2_1$ ) symmetry operations; the first digit in the four-digit ADC codes refers to symmetry operations of  $x, y, z, 1/2 + x, 1/2 - y, -z, -x, 1/2 + y, 1/2 - z$ , and  $1/2 - x, -y, 1/2 + z$ . This coordination is characterized by Is [I(1556) and I(1554)] along an axis (c).

quire shifting the upper grid along the axis-1 and axis-2 directions. The orientations of the molecules (I, C, P, or A) used to fill in around the central molecule depend upon the space group pattern under investigation.

The repulsion energy between two (or more) molecules is computed as the sum of all pair-wise atom-to-atom interactions. The interactions are evaluated with a 6-12 function described in the following section. The threshold used for the axis-1 search is the maximum allowable repulsion between any pair of molecules in the structure. Twice the value is used for the step two search because the central molecule in general is in contact with two molecules in a line. In step three, the central molecule usually will be in contact with four molecules in a plane above or below the central plane and thus four times the threshold energy is used in this search. The threshold energy was derived from known structures by searching for values that reproduced the true unit cell volume with the central molecule in the orientation of the known structure. The values were in the range of 0.2–1.0 kcal mol<sup>-1</sup> in most cases, and 0.5 kcal mol<sup>-1</sup> has been used as the molecule...molecule threshold energy in the construction of unknown structures.

MOLPAK uses a fixed structure (rigid body) as the search probe in the construction of a packing motif.

The structure is specified in a Cartesian coordinate system ( $\text{\AA}$ ) with the origin at the centroid of the molecule. This is the reference orientation of the center and identity molecules. A 14-molecule coordination sphere is constructed, around the central molecule, that defines a crystal structure of the specified symmetry type (space group and subclass) with a minimum unit cell volume per molecule. Coordination spheres of all types can be investigated in the same computer run if more than one symmetry type has been specified, but only information for the type corresponding to the minimum volume is saved. This procedure provides the minimum volume (and unit cell constants) for a given orientation of the central molecule. The orientation of the central molecule is altered and the procedure repeated to ultimately produce a 3-D map of minimum unit cell volume as a function of the central molecule's orientation (specified by three Eulerian angles).

As previously indicated, the three types of symmetry operations that alter a molecule's orientation in the triclinic, monoclinic, and orthorhombic space groups are planes, two-fold axes, and inversion centers. These operations involve changing the signs of one, two, or all three of the coordinates of the atomic positions. The unique direction of a plane or axis must be parallel to a cell axis that is perpendicular to the other two axes (e.g., the  $\mathbf{b}$  axis in monoclinic).

**Table II.** Coordination sphere patterns for the primitive triclinic, monoclinic, and orthorhombic space groups with  $Z \leq 4$ 

Code	Sp Grp	Z	Coordination sphere contents
AA	P1	1	6 I in plane-1,2; 4 I above; 4 I below
AB	P1	2	6 I in plane-1,2; 4 C above; 4 C below
AC <sup>a</sup>	Pm	2	6 I in plane-1,3; 4 P <sub>2</sub> - above; 4 P <sub>2</sub> - below
AD	Pc	2	6 I in plane-1,3; 4 P <sub>2</sub> 3 above; 4 P <sub>2</sub> 3 below
AE <sup>a</sup>	P2	2	6 I in plane-1,3; 4 A <sub>2</sub> - right; 4 A <sub>2</sub> - left
AF	P2 <sub>1</sub>	2	6 I in plane-1,2; 4 A <sub>2</sub> 2 right; 4 A <sub>2</sub> 2 left
AG	Pc	2	6 I in plane-1,2; 4 P <sub>2</sub> 3 right; 4 A <sub>2</sub> 3 left
AH	P2 <sub>1</sub>	2	6 I in plane-1,3; 4 A <sub>2</sub> 2 above; 4 A <sub>2</sub> 2 below
AI	P2 <sub>1</sub> /c	4	2 I on axis-2; 4 P <sub>2</sub> 3; 4 A <sub>2</sub> 2; 4 C
AJ	P2/c	4	2 I on axis-2; 4 P <sub>2</sub> 3; 4 A <sub>2</sub> -; 4 C
AK	P2 <sub>1</sub> /c	4	2 I on axis-2; 4 P <sub>2</sub> 3; 4 A <sub>2</sub> 2; 4 C
AL	P2/c	4	2 I on axis-2; 4 P <sub>2</sub> 3; 4 A <sub>2</sub> -; 4 C
AM	P2 <sub>1</sub> /c	4	2 I on axis-1; 4 P <sub>2</sub> 3; 4 A <sub>2</sub> 2; 4 C
AN <sup>a</sup>	P2 <sub>1</sub> /m	4	2 I on axis-1; 4 P <sub>2</sub> -; 4 A <sub>2</sub> 2; 4 C
AO <sup>b</sup>	P2/m	4	2 I on axis-1; 4 P <sub>2</sub> -; 4 A <sub>2</sub> -; 4 C
AP	P2 <sub>1</sub> 2 <sub>1</sub> 2	4	2 I on axis-3; 4 A <sub>1</sub> 1; 4 A <sub>2</sub> 2; 4 A <sub>3</sub> -
AQ	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	4	2 I on axis-3; 4 A <sub>1</sub> 1; 4 A <sub>2</sub> 2; 4 A <sub>3</sub> 3
AR <sup>b</sup>	Pma2	4	2 I on axis-3; 4 P <sub>1</sub> -; 4 P <sub>2</sub> 1; 4 A <sub>3</sub> -
AS <sup>a</sup>	Pmn2 <sub>1</sub>	4	2 I on axis-3; 4 P <sub>1</sub> -; 4 P <sub>2</sub> 1,3; 4 A <sub>3</sub> 3
AT	Pnn2	4	2 I on axis-3; 4 A <sub>3</sub> -; 4 P <sub>2</sub> 1,3; 4 P <sub>1</sub> 2,3
AU <sup>a</sup>	Pmn2 <sub>1</sub>	4	2 I on axis-3; 4 A <sub>3</sub> 3; 4 P <sub>2</sub> 1,3; 4 P <sub>1</sub> -
AV	Pna2 <sub>1</sub>	4	2 I on axis-3; 4 P <sub>1</sub> 2,3; 4 P <sub>2</sub> 1; 4 A <sub>3</sub> 3
AW	Pba2	4	2 I on axis-3; 4 P <sub>1</sub> 2; 4 P <sub>2</sub> 1; 4 A <sub>3</sub> -
AX <sup>a</sup>	Pnc2	4	2 I on axis-3; 4 A <sub>3</sub> -; 4 P <sub>2</sub> 3; 4 P <sub>1</sub> 1,3
AY	Pca2 <sub>1</sub>	4	2 I on axis-2; 4 A <sub>3</sub> 3; 4 P <sub>1</sub> 3; 4 P <sub>2</sub> 1
AZ	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	4	6 I in plane-2,3; 4 A <sub>2</sub> 2; 4 A <sub>3</sub> 3
BA	P2 <sub>1</sub> 2 <sub>1</sub> 2	4	6 I in plane-2,3; 4 A <sub>1</sub> 1; 4 A <sub>3</sub> -
BB <sup>b</sup>	P2 <sub>1</sub> 2 <sub>1</sub> 2	4	6 I in plane-2,3; 4 A <sub>2</sub> 2; 4 A <sub>3</sub> -
BC <sup>b</sup>	P222 <sub>1</sub>	4	6 I in plane-2,3; 4 A <sub>2</sub> -; 4 A <sub>3</sub> 3
BD	Pna2 <sub>1</sub>	4	6 I in plane-2,3; 4 P <sub>1</sub> 2,3; 4 A <sub>3</sub> 3
BE	Pnn2	4	6 I in plane-2,3; 4 P <sub>1</sub> 2,3; 4 A <sub>3</sub> -
BF	Pna2 <sub>1</sub>	4	6 I in plane-1,3; 4 P <sub>2</sub> 1; 4 A <sub>3</sub> 3
BG	Pba2	4	6 I in plane-2,3; 4 P <sub>1</sub> 2; 4 A <sub>3</sub> -
BH	Pca2 <sub>1</sub>	4	6 I in plane-2,3; 4 P <sub>1</sub> 3; 4 A <sub>3</sub> 3
BI <sup>a</sup>	Pnc2	4	6 I in plane-2,3; 4 P <sub>2</sub> 3; 4 A <sub>3</sub> -
BJ <sup>a</sup>	Pmn2 <sub>1</sub>	4	6 I in plane-2,3; 4 P <sub>1</sub> -; 4 A <sub>3</sub> 3
BK <sup>b</sup>	Pma2	4	6 I in plane-2,3; 4 P <sub>1</sub> -; 4 A <sub>3</sub> -

axis-1, axis-2, and axis-3 refer to the  $a$ ,  $b$ , and  $c$  crystallographic axes. The various symbols have the following meaning: plane- $ij$  is a plane containing axis- $i$  and axis- $j$ ;  $P_i$  - is a mirror plane perpendicular to axis- $i$  (- indicates no translation);  $P_i j$  is a glide plane perpendicular to axis- $i$  with an axis- $j$  translation;  $P_i j, k$  is a glide plane perpendicular to axis- $i$  with a diagonal translation along axis- $j$  and axis- $k$ ;  $A_i$  is a two-fold screw axis parallel to axis- $i$  (translation along axis- $i$ );  $A_i$  - is a two-fold rotation axis parallel to axis- $i$  (- indicates no translation); "above" and "below" refer to positions with respect to the plane of Is; "left" and "right" refer to sides of axis-2. AI and AJ are special cases of AK and AL, respectively, with pairs of  $P_2$  3 along axis-3; this allows the MOLPAK search routine to run much faster.

<sup>a</sup>There were no examples of C-, H-, N-, O-, F-containing structures that did not have molecules in special positions. Thus, experimental crystallographic verification of this coordination could not be accomplished.

<sup>b</sup>No examples for these space groups.

Translations associated with these operations also are restricted in that the movement of the centroid along the direction of any unchanged coordinate (sign) must be either 0 or  $\frac{1}{2}$  the unit cell translation. These restrictions often simplify the steps (2 and 3) that involve the placement of lines and 2-D grids in the development of the coordination environment.

### Program Speed Considerations

MOLPAK creates a minimum volume packing arrangement for a given symmetry type (space group) and orientation of the probe structure with respect

to the orthogonal axes. Because the optimum orientation of the probe with respect to the space group symmetry is unknown beforehand, it is necessary to produce a 3-D minimum volume map as a function of the unique set of orientations of the probe. For example, if the probe is oriented in  $10^\circ$  increments about three (Eulerian) axes over a  $180^\circ$  range for each there are 6869 orientations that must be converted into minimum volume packing arrangements. The amount of computer time needed for such a minimum volume vs. orientation search can be significant because of the large numbers of atom-to-atom interactions that require evaluation. Methods



that have been incorporated to reduce computational times are discussed below.

First, the evaluation of the energy between two atoms, which is in the innermost loop of the calculations, should be as fast as possible. MOLPAK uses either a Lennard-Jones 6-12 potential,  $E = C_1/r_{ij}^6 + C_2/r_{ij}^{12}$ , or Buckingham 6-exp expression,  $E = C_1/r_{ij}^6 + C_2 \exp(C_3 r_{ij})$  ( $r_{ij}$  is the atom  $i$  to atom  $j$  distance and  $C_1$ ,  $C_2$ , and  $C_3$  are coefficients that depend upon the atom types). The coefficients were taken primarily from the work of Giglio<sup>15</sup> and Holden and Dickinson<sup>16</sup> with some MOLPAK-specific adjustments. The program allows either potential to be used by direct evaluation or table look-up, but the latter is invariably used because of speed.

Second, the number of interatomic distances and interaction energies that must be evaluated was restricted by the use of only the repulsive part of the potential function. That is, the typical 6-12 potential curve was modified by adding the value of the function at its minimum energy ( $E_{\min}$  at  $d_{\min}$ ) and, in the calculation of the repulsive energy between two molecules, the contribution from the  $i,j$  atom pair was ignored if  $r_{ij}$  was greater than the appropriate  $d_{\min}$ . Further, because distances between atoms are evaluated as the sum of the squares of the distances in the three orthogonal directions, the calculation of a distance was terminated if either the square of a separation in one direction or the sum of the squares in two directions exceeded the appropriate  $d_{\min}^2$ .

A third economy was achieved by eliminating the redetermination of atom separations in one and two directions. For example, when the positions of molecules in a 2-D grid that includes axis-1 and axis-2 are being determined the axis-3 separation between the  $i$ th atom in the central molecule and the  $j$ th atoms in all of the other molecules is the same. Further, if the axis-3 separation is larger than the distance at which the modified potential function becomes zero, then all of these atom-pair interactions can be ignored. The first step, therefore, in finding a 2-D grid of minimum unit area is to generate a table of the squares of axis-3 separations less than the maximum distance required for energy evaluation. Other required parameters are copied to this table. In a similar fashion, when interaction energies between two molecules along axis-1 are evaluated the atomic separations along the axis-2 and axis-3 directions are independent of the distance between the molecular centroids along axis-1. Therefore, a second table is generated that contains the sums of the squares of the atomic separations along the axis-2 and axis-3 directions, which are less than the square of the maximum distance in the appropriate interatomic potential function. This table is used to find the distance between molecules along axis-1 at which the interaction energy reaches the chosen threshold.

A final economy involves monitoring the calculation of the energy sum between a central molecule and another molecule (or molecules) and comparing

the sum with the threshold value after the addition of each atom-atom increment. The accumulation is terminated if the sum exceeds the threshold. In cases in which the tables described in the previous paragraph are used, the square of the contributing interatomic distance is calculated by adding the square of the axis-1 separation to the saved sum of the squares of the separations in the axis-2 and axis-3 directions.

The use of a repulsion-only potential and predetermined docking thresholds are among the most important computational aspects of MOLPAK. The procedure is considerably faster than one that minimizes the energy between molecules because of the need to examine all (or most) atom-to-atom interactions between molecules at each step. In addition, we found from test calculations (in agreement with Kitaigorodsky<sup>12</sup>) that the fine details of an intermolecular fit can be adequately represented by repulsion.

## Packing Procedures

The procedures used to form the 1-, 2-, and 3-D line and grids are described below.

### *Procedure 1: Approach of Two Molecules to Form a Structure Line*

An identity is placed at an arbitrary distance along orthogonal axis-1 from the central molecule. The repulsion energy is calculated. If the energy is less than the threshold, the molecule is moved toward the central molecule in arbitrary equal steps until the energy is greater than the threshold and then backed off by one step. If the energy is greater than the threshold initially, the molecule is moved away in arbitrary equal steps until the energy drops below the threshold. This position, at which the energy is less than the threshold, is the "reference position." The step size is halved and the molecule is moved closer. If the energy is less than the threshold, this new position becomes the reference position. The procedure to refine the reference position is repeated until the step size reaches some preset fraction of its original magnitude. The program default is 1/128. This establishes the repeat distance along one axis (e.g., **a**) of the unit cell. The threshold distance in the opposite direction (i.e.,  $-\mathbf{a}$ ) is the same as that just determined because the roles of the central and moving molecules are simply reversed.

### *Procedure 2: Approach of Two Structure Lines of Molecules to Form a 2-D Grid*

Procedure 1 establishes a translational repeat distance along axis-1 that applies to all features of the structure. Placing a center molecule, for example, automatically establishes a line of center molecules

with this unit translation distance along the axis-1 direction. Similarly, placement of a second line of molecules, parallel to the first, adds a second translation vector that establishes a 2-D grid of translational repeat positions (in the axis-1/axis-2 plane). When the coordination sphere contains six or more identities, the second line is usually a line of identity molecules, not displaced in the axis-3 direction. However, for some space groups the second line is a line of plane- or axis-related molecules, displaced by a preset amount along axis-3. In either case, procedure 2 determines the minimum distance between the lines, consistent with space group symmetry and the established threshold repulsion energy between pairs of molecules.

For orthorhombic space groups, the positions of the molecules in the second line, along the axis-1 direction, are restricted. In this case, procedure 2 determines only the length of a unit cell axis. For monoclinic space groups, procedure 2 determines both the  $\beta$  angle and the length of the **a** (or **c**) axis; both **b** and  $\gamma$  are determined for triclinic space groups. For these space groups, the second line of molecules must be shifted along axis-1 to find the minimum allowed separation along the axis-2 direction. This is done by placing one molecule at various offset distances from the intersection of the second line (of molecules) with axis-2.

First, the centroid of this molecule is placed on (or directly above) axis-2 at an arbitrary distance from the central molecule and a modification of procedure 1 is performed to determine the minimum allowed distance. Although this molecule is part of a second line of molecules, the coordinates of the other molecules are not calculated and stored; rather, atoms in other molecules are created in the distance-energy calculating loop by appropriate translations along axis-1. Twice the specified threshold energy is used because the central molecule is usually in contact with two molecules in the second line.

Next, the nonredundant range of positions of this molecule along the axis-1 direction, from zero to the line translation distance, is covered in a specified number of equal steps (default = 8). The reference location is set equal to that corresponding to the smallest line separation. The step size is then divided by two and the line separation is found at the reference location plus the new step. If this is smaller than the previous minimum, the reference location is reset and the step size is again halved. If not, the line separation is calculated at reference minus new step, the reference is reset or not, but in either case the step size is halved. This search procedure continues until the step size is a specified fraction of the line interval (program default = 1/128).

This procedure defines the second unit cell axis (e.g., **b**) and, if required, the **a/b** angle ( $\gamma$ ).

### *Procedure 3: Approach of Two 2-D Grids of Molecules to Form a 3-D Grid*

Procedures 1 and 2 define a 2-D grid of unit cell translation positions; this procedure completes the determination of the 3-D grid or unit cell. The procedure is simply a 2-D version of procedure 2 in which a 2-D grid, parallel to the central molecule-containing 2-D grid of molecules, is located along axis-3. The repulsion energy threshold used is four times that of the procedure 1 value because the central molecule, in known structures, most often is in contact with four other molecules in planes above and below the central plane.

For monoclinic and orthorhombic space groups, the positions of the molecules of the 2-D grid, relative to its intersection with axis-3, are restricted, and procedure 3 determines only the length of a unit cell axis. However, for the two triclinic space groups the grid must be shifted in both the axis-1 and axis-2 directions to determine two angles in addition to the length of an axis. Here, the range of nonredundant locations for the centroid of one molecule of the 2-D grid is the parallelogram defined by the repeat intervals of the 2-D grid, with one corner on axis-3.

Determination of the minimum allowed distance by movement of the 2-D grid along axis-3 toward the central molecule is similar to the placement of the molecule line along axis-2 in procedure 2. Again, the positions of the atoms of the other molecules of the grid are not calculated and stored but generated in the energy-distance calculation loop. Placement of the 2-D grid in the axis-1 and axis-2 directions is a 2-D equivalent of the line placement in procedure 2. The range is initially covered by calculations at a specified number of equal intervals along both the axis-1 and axis-2 directions (program default = 8). The intervals are then halved as in procedure 2.

### **Examples of Space Group Search Procedures**

Examples of three MOLPAK space group search procedures are given in this section. Two of the three are for  $P2_1/c$ , the most important of the space groups for light atom-containing organics.

#### *Triclinic, $P\bar{1}$ (AB group)*

A coordination sphere diagram is given in Figure 3. The **a** axis length is determined by procedure 1. The **b** axis length and the angle  $\gamma$  between **a** and **b** are determined by procedure 2; an adjustment of the atoms of the second line along the **a** axis direction is required. The remaining cell parameters are determined by two applications of procedure 3. First, a grid of center-related molecules is located long + axis-3, with adjustments along the axis-1 and axis-2 directions. Then, an identical grid is located along - axis-3 and suitably adjusted. The length and di-

rection of the **c** axis of the unit cell is the vector from a center-related molecule on the lower grid to a center-related molecule on the upper grid.

#### *Monoclinic, $P2_1/c$ (AI subgroup)*

A coordination sphere diagram is given in Figure 4. The **b** axis length is determined by procedure 1 along orthogonal axis-1. The **c** axis and the **P** molecule offset (along **b**) are then determined by procedure 2 with a line of **P** molecules parallel to axis-1 (the separation of the **I** and **P** lines is along axis-2). Positioning a line of **P** molecules against the line of **I**s specifies the length of the **c** axis because the distance from the **I** to **P** lines is equal to  $c/2$ .

The **a** axis length and  $\beta$  angle are determined by two applications of procedure 3 along  $\pm$ orthogonal axis-3. The plane consists of alternate rows of **A** and **C** molecules. Because the unit cell has a screw axis, the **A** molecules are halfway along the **b** axis direction and the vector between each **A** and **C** pair is the same as that between the central molecule and a **P**. The **A** (or **C**) position need only be adjusted along the axis-2 (**c** axis) direction. A vector from an **A** molecule on the minus side to an **A** molecule on the plus side determines the **a** axis.

#### *Monoclinic, $P2_1/c$ (AK subgroup)*

A coordination sphere diagram is given in Figure 5. This unit cell is determined by an iterative procedure in which the off-set of the glide plane molecules along **b** is arbitrarily set and then refined. The **b** axis length is determined by procedure 1 along orthogonal axis-1. The range of possible **b** off-sets for the **P** glide plane molecules is from zero to the **b** axis length. The range is covered by a number of equal steps (default = 8) and then refined by halving the step size to a specified fraction of the **b** length (default =  $1/128$ ).

With the off-set arbitrarily set at the first step, a tentative value of the axis length is determined by applying procedure 2 along axis-2 with a line of **P** molecules along axis-1 (the **b** axis). However, in this case the **P** molecule off-set along **b** is not adjusted. The corresponding values of **c** and  $\beta$  are determined by two applications of procedure 3 along  $\pm$ axis-3. The plane for the +axis-3 direction consists of alternate lines of **A** and **C** molecules. Because the space group contains a screw axis, the **A** molecules lie halfway along the **b** axis. The vector between each pair of **A** and **C** molecules is the same as that between the central molecule and the **P** molecule.

The plane for the -axis-3 direction by application of procedure 3 also consists of the same lines of **A** and **C** molecules. However, each placement of this second **A-C** plane specifies a tentative **c** axis, the vector between an **A** molecule on the plus side and

**Table III.** WMIN energy coefficients.

	A (kcal mol <sup>-1</sup> )	B (kcal mol <sup>-1</sup> )	C (Å)
C	19.012	290.0	1.80
H	6.710	43.49	1.87
N	20.722	150.0	1.805
O	28.866	390.0	1.98
F	17.606	296.5	2.08

This is a general-purpose set of coefficients that does not differentiate between different bonding situations (e.g., the same values are used for N in  $-\text{NH}_2$  and  $-\text{NO}_2$ ). An alternative set of coefficients has been developed that does recognize an atom's bonding environment; however, the relative order of WMIN lattice energies for the several possible crystal structures for a compound is the same from calculations with the general and specific coefficient sets.

an **A** molecule on the minus side. Therefore, tentative locations of other **I** and **P** molecules are also generated and included in the energy calculation. This procedure determines the minimum volume cell corresponding to the first tentative of the **P** molecule off-set. The off-set is stepped, then refined to the value that produces the smallest unit cell volume.

#### **Trial Structure Refinement**

It is necessary to refine an initial MOLPAK packing arrangement to yield a possible "crystal structure" because of the repulsion-only force field used in the MOLPAK procedure and rather approximate information available on the molecular orientation. The WMIN crystal structure (or packing arrangement) refinement<sup>10</sup> involves a minimization of the calculated lattice energy ( $W_c$ ) with the atom-to-atom potential function shown below; the A, B, C terms are energy coefficients,  $qs$  are atomic charges,  $rs$  are distances between atoms, and  $n$  and  $N$  refer to atoms in one and all unit cells, respectively.<sup>10</sup> Because of initial problems encountered with obtaining a workable set of atomic charges ( $qs$ ), we eliminated the Coulombic term and developed a set of A and B coefficients (Table III) to compensate for the chargeless function.\*  $W_c$  is minimized as a function of the

\*The coefficients of Mirsky<sup>17</sup> and/or Williams and coworkers<sup>18</sup> were used in some of the initial WMIN calculations. In addition, attempts were made to determine atomic charges for the lattice energy calculations with the procedure of Cox and Williams,<sup>19</sup> together with coefficient development. These results were less successful than the use of A and B coefficients developed by fitting the chargeless form of  $W_c$  to a number of crystal structures. Thus, these A and B coefficients contain hidden charge components. Group-specific coefficients have been developed for  $\text{C}=\text{O}$ ,  $\text{C}-\text{H}$ ,  $\text{NH}_2$ ,  $\text{NO}_2$ ,  $\text{NH}_2$ ,  $\text{N}-\text{NO}_2$ ,  $\text{C}=\text{O}$ ,  $\text{NO}_2$ ,  $\text{NF}$ , and  $\text{NF}$  at this writing. The inclusion of the Coulombic term in the crystal structure refinement step is under active investigation. An alternative to WMIN is the PCK/ME procedure,<sup>20</sup> which uses an atom-centered multipole charge model for the Coulombic term; see also the LOPAS procedure in ref. 25.

unit cell parameters (e.g.,  $a$ ,  $b$ ,  $c$ ,  $\alpha$ ,  $\beta$ ,  $\gamma$  in triclinic) and

$$W_c = (1/2Z) \sum_i^n \sum_{j \neq i}^N (q_i q_j / r_{ij} - A_i A_j / r_{ij}^6 + B_i B_j \exp[(C_i + C_j) r_{ij}])$$

orientation and translation parameters for the rigid body structure probe.

## DISCUSSION AND EXAMPLES

The MOLPAK procedure for the construction of reasonable packing arrangements has been tested and validated with a number of C-H-N-O-F-containing compounds with known crystal structures; all structures contained one or more nitro groups because of our interest in energetic materials. The rigid body search probes used were taken from the published crystal structures; the X—H distances were altered to N—H = 1.013, C(sp<sup>2</sup>)—H = 1.084, and C(sp<sup>3</sup>)—H = 1.098 Å. Each of the compounds has been tested in the  $P\bar{1}$ ,  $P2_1$ ,  $P2_1/c$ , and  $P2_12_12_1$  space groups with complete MOLPAK searches for the optimum packing arrangements followed by WMIN refinement. These data are summarized in Table IIS (supplementary data); four examples are given here in Tables IV–VII, one for a compound in each of the four key space groups.

The results in Table VI for VIHGAU (known space group  $P2_1/c$ ) are discussed in detail here. The first row labeled “obsd” lists the experimental unit cell parameters and cell volume per molecule. The second “refd” row lists similar information obtained from the WMIN refinement of the unit cell parameters ( $a$ ,  $b$ ,  $c$ ,  $\beta$ ), molecular orientation, and position. The  $W_c$  value of  $-34.72$  kcal mol<sup>-1</sup> is the WMIN-derived lattice energy. Differences between the obsd and refd unit cell data presumably reflect imperfections in the WMIN force field. Data in the third “calcd” row were obtained from the complete MOLPAK/WMIN procedure in the true space group

( $P2_1/c$ ) for the compound; the 10 best packing arrangements from a global 10° MOLPAK search were refined by 2° searches followed by WMIN refinement. This line gives the unit cell and  $W_c$  data for the MOLPAK/WMIN packing arrangement that yielded the lowest lattice energy. Lines 4–6 labeled “ $P\bar{1}$ ,” “ $P2_1$ ,” and “ $P2_12_12_1$ ” provide similar data for the best packing arrangement found in the indicated space groups. The lowest value of  $W_c$  presumably indicates the best (correct) crystal structure from among the space groups investigated.

It is interesting that the  $W_c$  for the correct space group of  $-34.71$  kcal mol<sup>-1</sup> is only about 0.7 kcal mol<sup>-1</sup> lower than that for the next best  $P\bar{1}$  arrangement at  $-33.99$  kcal mol<sup>-1</sup>. The majority of our calculations have indicated that there is relatively little difference, energetically, between the observed (true) and other (plausible) crystal structures. Even smaller  $W_c$  differences are found between the correct and next best arrangements in many of the other test cases and, in a number of other examples, the  $W_c$  for an incorrect space group is lower than that for the correct system. Such is the case for DNTNAP (Table VIII), in which a  $P2_1/c$  arrangement has a lower  $W_c$  of  $-25.94$  kcal mol<sup>-1</sup> than the correct  $P2_12_12_1$  structure with  $W_c$  of  $-25.13$  kcal mol<sup>-1</sup>. These discrepancies probably point to inadequacies in the WMIN refinement step, although the existence of unknown polymorphism is a possibility.

## Potential for Crystal Structure Prediction

The procedures described in the preceding sections hold promise for *ab initio* crystal structure predictions. The crystal structure prediction strategy has three parts. First, an accurate molecular model (search probe) must be created. At present, we construct the structure with the MacroModel graphics program,<sup>21</sup> followed by a semiempirical molecular orbital geometry optimization (e.g., AM1 procedure in the MOPAC package<sup>22</sup>). Second, the probe is used by the MOLPAK procedure to locate a number of

**Table IV.** Unit cell data and lattice energies from MOLPAK/WMIN calculations in the triclinic space group  $P\bar{1}$  for BECJEY.

Sp Grp	$a$	$b$	$c$	$\alpha$	$\beta$	$\gamma$	$V/Z^a$	$W_c^b$
$P\bar{1}$								
obsd <sup>c</sup>	6.300	6.380	11.851	79.27	81.00	61.28	409.20	
refd <sup>d</sup>	6.313	6.509	11.460	82.83	81.98	62.10	411.13	$-24.37$
calcd <sup>e</sup>	6.313	6.509	11.459	82.85	82.01	62.10	411.15	$-24.37$
$P2_1$	6.325	12.061	6.461	90	118.76	90	432.08	$-22.13$
$P2_1/c$	10.719	6.629	11.784	90	101.22	90	410.68	$-24.39$
$P2_12_12_1$	6.804	7.904	15.752	90	90	90	423.57	$-22.71$

<sup>a</sup>Unit cell volume per molecule.

<sup>b</sup>WMIN lattice energy (kcal mol<sup>-1</sup>).

<sup>c</sup>Experimental cell parameters (lengths are in Å, angles in °).

<sup>d</sup>Cell parameters from WMIN refinement of observed crystal structure.

<sup>e</sup>Cell parameters for the best packing arrangement from a complete MOLPAK/WMIN search in  $P\bar{1}$ .

**Table V.** Unit cell data and lattice energies from MOLPAK/WMIN calculations in the monoclinic space group  $P2_1$  for JEXLUT.

Sp Grp	<i>a</i>	<i>b</i>	<i>c</i>	$\alpha$	$\beta$	$\gamma$	<i>V/Z</i>	<i>W<sub>c</sub></i>
$P2_1$								
obsd	7.471	10.777	7.642	90	103.84	90	597.43	
refd	7.382	10.716	7.722	90	104.66	90	590.98	−46.81
calcd	7.382	10.716	7.722	90	104.66	90	590.99	−46.81
$P\bar{1}$	13.713	7.187	6.590	106.84	83.20	98.92	612.23	−44.41
$P2_1/c$	10.978	9.417	12.017	90	79.84	90	611.43	−43.86
$P2_12_12_1$	14.888	10.688	7.617	90	90	90	605.99	−44.42

**Table VI.** Unit cell data and lattice energies from MOLPAK/WMIN calculations in the monoclinic space group  $P2_1/c$  for VIHGAU.

Sp Grp	<i>a</i>	<i>b</i>	<i>c</i>	$\alpha$	$\beta$	$\gamma$	<i>V/Z</i>	<i>W<sub>c</sub></i>
$P2_1/c$								
obsd	12.654	7.371	15.083	90	101.76	90	344.32	
refd	12.772	7.344	15.021	90	100.39	90	346.47	−34.71
calcd	12.773	7.344	15.021	90	100.38	90	346.50	−34.71
$P\bar{1}$	10.963	8.800	8.194	99.23	72.34	112.29	348.00	−33.99
$P2_1$	8.305	9.032	10.248	90	70.20	90	361.61	−32.42
$P2_12_12_1$	15.067	13.492	6.976	90	90	90	354.53	−32.49

**Table VII.** Unit cell data and lattice energies from MOLPAK/WMIN calculations in the orthorhombic space group  $P2_12_12_1$  for CILBII.

Sp Grp	<i>a</i>	<i>b</i>	<i>c</i>	$\alpha$	$\beta$	$\gamma$	<i>V/Z</i>	<i>W<sub>c</sub></i>
$P2_12_12_1$								
obsd	15.503	13.600	7.052	90	90	90	371.71	
refd	15.292	13.446	7.091	90	90	90	364.50	−44.90
calcd	15.293	13.445	7.091	90	90	90	364.50	−44.91
$P\bar{1}$	14.653	7.712	7.035	84.37	81.53	101.66	381.67	−39.46
$P2_1$	6.998	15.331	11.231	90	38.75	90	377.11	−41.67
$P2_1/c$	15.407	6.967	13.857	90	93.11	90	371.31	−43.34

**Table VIII.** Unit cell data and lattice energies from MOLPAK/WMIN calculations in the orthorhombic space group  $P2_12_12_1$  for DNTNAP.

Sp Grp	<i>a</i>	<i>b</i>	<i>c</i>	$\alpha$	$\beta$	$\gamma$	<i>V/Z</i>	<i>W<sub>c</sub></i>
$P2_12_12_1$								
obsd	11.375	14.974	5.388	90	90	90	229.43	
refd	11.391	15.295	5.356	90	90	90	233.29	−25.13
calcd	11.391	15.296	5.356	90	90	90	233.31	−25.13
$P\bar{1}$	11.444	5.396	8.416	67.10	85.43	77.01	233.25	−24.49
$P2_1$	5.999	15.417	5.263	90	103.09	90	237.04	−24.71
$P2_1/c$	15.950	6.086	10.356	90	113.03	90	231.29	−25.94

possible packing arrangements for a particular space group. The MOLPAK search produces a 3-D map of minimum cell volume as a function of the orientation of the rigid-body probe. A Eulerian rotation matrix is used to generate all unique orientations of the probe structure for subsequent packing; the range for each of the Eulerian angles spans  $180^\circ$  and the step increment typically is  $10^\circ$ . Five to 10 of the best (smallest volume) unique minimum volume arrangements are individually refined by a MOLPAK search over a  $\pm 10^\circ$  Eulerian range ( $2^\circ$  step) centered on the appropriate Eulerian orientation from the previous

$10^\circ$  search, and each of these  $2^\circ$  arrangements is then refined with WMIN to yield a final (possible) crystal structure. Third, the NIST\*LATTICE program<sup>23</sup> is used to obtain reduced cell parameters from the WMIN-refined values for these structures; this determination is an essential finishing touch for the triclinic and monoclinic space group searches. With the present MOLPAK capabilities, the most important space groups for possible crystal structure generation are  $P\bar{1}$ ,  $P2_1$ ,  $P2_1/c$ , and  $P2_12_12_1$ .

The determination of the correct crystal structure for a compound depends upon a degree of similar-

**Table IX.** Unit cell data and average coordinate differences for the observed and predicted crystal structures of five nitro-containing compounds with AM1-optimized search probes.

	<i>a</i>	<i>b</i>	<i>c</i>	$\alpha$	$\beta$	$\gamma$	$\Delta x^a$	$\Delta y^a$	$\Delta z^a$
DETDOV, C <sub>12</sub> H <sub>11</sub> NO <sub>3</sub> , P $\bar{1}$									
obsd <sup>b</sup>	7.291	7.395	9.724	82.54	78.85	87.43			
refd <sup>c</sup>	7.379	7.377	9.435	88.74	79.99	86.02	0.109	0.011	0.029
calcd <sup>d</sup>	7.378	7.396	9.513	86.96	79.62	86.32	0.074	0.044	0.053
FOVYOE, C <sub>13</sub> H <sub>11</sub> N <sub>3</sub> O <sub>3</sub> , P2 <sub>1</sub>									
obsd <sup>b</sup>	3.857	19.542	8.066		89.37				
refd <sup>c</sup>	3.845	19.435	8.094		92.69		0.086	0.054	0.125
calcd <sup>d</sup>	3.801	20.034	8.205		91.80		0.118	0.062	0.153
DNPMTA, C <sub>5</sub> H <sub>10</sub> N <sub>6</sub> O <sub>4</sub> , P2 <sub>1</sub> /c									
obsd <sup>b</sup>	9.345	8.284	11.566		105.60				
refd <sup>c</sup>	9.278	8.460	11.412		104.18		0.046	0.086	0.109
calcd <sup>d</sup>	9.693	8.218	11.960		108.04		0.230	0.221	0.128
DNTNAP, C <sub>10</sub> H <sub>6</sub> N <sub>2</sub> O <sub>4</sub> , P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>									
obsd <sup>b</sup>	11.375	14.974	5.388						
refd <sup>c</sup>	11.391	15.295	5.356				0.013	0.117	0.035
calcd <sup>d</sup>	11.722	15.391	5.287				0.175	0.137	0.062
DNCOOC, C <sub>8</sub> H <sub>6</sub> N <sub>2</sub> O <sub>4</sub> , Pna2 <sub>1</sub>									
obsd <sup>b</sup>	11.783	15.996	4.717						
refd <sup>c</sup>	12.213	16.032	4.395				0.450	0.254	0.123
calcd <sup>d</sup>	12.079	16.201	4.478				0.350	0.403	0.132

<sup>a</sup> $\Delta x$ ,  $\Delta y$ , and  $\Delta z$  are the average coordinate differences (Å) between the obsd vs. refd and obsd vs. calcd crystal structures for C, N, and O atoms (H atoms were eliminated in the comparison).

<sup>b</sup>Experimental (X-ray) cell parameters (lengths in Å, angles in °).

<sup>c</sup>Cell parameters from WMIN refinement of observed crystal structure.

<sup>d</sup>Cell parameters for the MOLPAK/WMIN solution with an AM1 geometry-optimized search probe.

ity between the structure (conformation) of the MOLPAK search probe and the structure in the crystal. Examples of successful crystal structure predictions, starting with AM1 geometry-optimized search probe structures, are given in Table IX (Table IIIS of the supplementary data gives the X-ray and predicted atom coordinates) for a dinitrocyclooctatetraene (DNCOOC, Pna2<sub>1</sub>), a dinitronaphthalene (DNTNAP, P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>), a bicyclic nitramine (DNPMTA, P2<sub>1</sub>/c), a nitrostilbene analog (FOVYOE, P2<sub>1</sub>), and a nitrobenzofuran (DETDOV, P $\bar{1}$ ). Rigid-body, least-squares comparisons<sup>24\*</sup> of the AM1 and actual crystal molecular structures gave rms deviations for DNCOOC (14 C, N, O atoms), DNTNAP (16 C, N, O atoms), DNPMTA (15 C, N, O atoms), FOVYOE (19 C, N, O atoms), and DETDOV (16 C, N, O atoms) of 0.116, 0.076, 0.202, 0.126, and 0.053 Å, respectively. These procedures will fail to produce the correct crystal structure if the search probe and crystal conformations differ substantially because of current limitations in MOLPAK that require a conformationally rigid probe.

### Program Availability

We at present have MOLPAK and WMIN programs running on DEC and IBM platforms and interactive programs to create DEC VAX/VMS COM files and

IBM/CMS EXEC files for automatically performing the complete sequence of second and third steps; UNIX (IBM RS6000) and ULTRIX (DEC) versions are in preparation. MOLPAK may be obtained from the authors. The input is minimal: a compound code, title, space group symbol, model coordinate file name, Eulerian angle ranges and steps for probe rotation in the first (e.g., 10°) and second (e.g., 2°) MOLPAK searches, and the number of the best packing arrangements from the initial MOLPAK search to be elaborated. The final output is condensed to a small table that provides, for each initial MOLPAK 10° map structure selected for further elaboration, information on the initial orientation of the probe, initial and final (WMIN refined) unit cell volumes and lattice energies, and the reduced unit cell parameters.

### CONCLUSIONS AND FUTURE DIRECTIONS

In this article, we describe the development and testing of the MOLPAK/WMIN procedure for the prediction of crystal structures for C-H-N-O-F-containing compounds in the primitive triclinic, monoclinic, and orthorhombic space groups with  $Z \leq 4$ . Plausible packing arrangements are constructed from information for the 3-D molecular structure and possible space groups. The correct crystal structure usually is obtained if the search and refinement steps are restricted to a single (correct) space group. In these cases, the structure is correctly indicated by the low-

\*The comparisons were performed with the BMFIT (best molecular fit) program described in ref. 24.

est WMIN lattice energy ( $W_c$ ). The problem of the correct crystal structure (space group) not having the lowest  $W_c$  is encountered more frequently if a number of possible space groups are examined. In almost all cases, the lattice energy differences are small (e.g.,  $\leq 1$  kcal mol<sup>-1</sup>) between the correct structure and a plausible incorrect structure. An important area for further work is the modification of the WMIN force field to improve the sensitivity of the refinement step. This may be accomplished by including a Coulombic term with atom charges obtained, for example, from the MOPAC-ESP<sup>22</sup> procedure or with an alternative refinement procedure such as LOPAS.<sup>25\*</sup>

Although the current MOLPAK code can fail to identify the correct packing arrangements in crystals with strong intermolecular hydrogen bonding, the procedure has been successful with several nitroanilines and amides. In keeping with the MOLPAK logic of only using the repulsive part of the intermolecular potential, this could be addressed by including special hydrogen atom parameters that are small (or negative) and utilized only in interactions such as O—H $\cdots$ O or N—H $\cdots$ O.

The use of conformationally flexible MOLPAK search probes is under investigation. Each of the preliminary MOLPAK coordination geometries (possible crystal structures), obtained with a rigid probe and the repulsion-only potential, will be subjected to further refinement that permits conformational flexibility. This fine-tuning would utilize a force field with both attractive and repulsive components and, in the interest of computational speed, involve only the central and 14 surrounding coordination molecules.

The extension of the MOLPAK procedure to  $Z = 8$  space groups such as Pbca and C2/c will be undertaken. These procedures will be considerably more complicated than many of the existing  $Z = 4$  search patterns and iterative in nature similar to the P2<sub>1</sub>/c AK option.

Finally, MOLPAK will be extended to general organic compounds and not limited on C-H-N-O-F-containing molecules with nitro groups. An extension of the data base may uncover other (at present unknown) coordination sphere geometries. A case in point is BURKUU (triclinic, P $\bar{1}$ ,  $Z = 2$ ) whose 14 molecule coordination sphere differs from the AB arrangement illustrated in Figure 3 (six Is in a plane with planes of four Cs above and below the central I plane). BURKUU shows a plane of four Is (at  $\pm a$  and  $\pm c$ ), planes of four Cs above and below the four I plane and two Is along the *ab* plane diagonal (at [1,1,0] and [-1,-1,0]).

It has not escaped our notice that the present procedure could be utilized to solve X-ray or neutron diffraction crystal structures. In the case of a diffraction investigation, the crystal system and space group are known from the Laue symmetry, unit cell parameters, and pattern of systematic absences in the diffraction intensities. The correct crystal structure from a number of MOLPAK-generated possibilities could be selected from a comparison of the predicted and experimental cell parameters, as well as from the lattice energies of the various packing arrangements. The packing/refinement approach is not yet a cost-effective alternative to direct methods for solving the majority of crystal structures but could be a viable alternative when these procedures fail to solve the phase problem. The present MOLPAK procedure cannot readily deal with crystals with more than one structural unit per asymmetric unit or with molecules for which a good preexisting model of the structure is unavailable. The procedure could be used to solve structures from powder diffraction data; these data sets frequently have an insufficient number of resolved diffraction intensities for use with normal structure solution techniques. A comparison of the experimental (e.g., X-ray) and predicted cell parameters could serve as a figure-of-merit to point up which of the packing arrangements is correct and, in addition, one could perform structure factor calculations for a small number of strong, low-order reflections with crystal coordinates from the various packing arrangements and use something analogous to the crystallographic residual (*R* factor) to indicate the correct solution.

This work was supported by the Office of Naval Research. The authors thank Dr. R.S. Miller for continued interest and encouragement.

## References

1. J.R. Holden and C. Dickinson, *Prediction of Crystal Density through Molecular Packing Analysis*, Report NSWC MP 79-185, Naval Surface Weapons Center, Silver Spring, MD, 1979.
2. D.T. Cromer, H.L. Ammon, and J.R. Holden, *A Procedure of Estimating the Crystal Densities of Organic Explosives*, Report LA-11142-MS, Los Alamos Scientific Laboratory, Los Alamos, NM, 1987.
3. R.J. Spear and W.S. Wilson, *J. Energetic Mater.*, **2**, 61 (1984).
4. M.J. Kamlet and S.J. Jacobs, *J. Phys. Chem.*, **48**, 23 (1968).
5. (a) H.H. Cady, *Estimation of the Densities of Organic Explosives from Their Structural Formulas*, Report LA-7760-MS, Los Alamos Scientific Laboratory, Los Alamos, NM, 1979; (b) C.M. Tarver, *J. Chem. Eng. Data*, **24**, 136 (1979); (c) A. Immirzi and B. Perini, *Acta Cryst.*, **A33**, 216 (1977); (d) D.A. Cichra, J.R. Holden, and C. Dickinson, *Estimation of Normal Densities of Explosive Compounds from Empirical Atomic Volumes*, Report 79-273, Naval Surface Weapons Center, Silver Spring, MD, 1980; (e) J.R. Stine, *Prediction of*

\*The LOPAS acronym refers to "Local Orbitals for Polyatomic Systems."<sup>25</sup>

- Crystal Densities of Organic Explosives by Group Additivity*, Report LA-8920, Los Alamos National Laboratory, Los Alamos, NM, 1981.
6. C.S. Choi and E. Prince, *Acta Cryst.*, **B28**, 2857 (1972).
  7. H.H. Cady, A.C. Larson, and D.T. Cromer, *Acta Cryst.*, **16**, 617 (1963).
  8. N.G. Zhitomirskaya, L.T. Eremko, N.I. Golovina, and L.O. Atovmnyan, *Bull. Acad. Sci. USSR, Div. Chem. Sci.*, **36**, 525 (1987).
  9. A. Gavezzotti, *J. Am. Chem. Soc.*, **113**, 4622 (1991).
  10. W.R. Busing, *WMIN, A Computer Program to Model Molecules and Crystals in Terms of Potential Energy Functions*, Report ORNL-5747, Oak Ridge National Laboratory, Oak Ridge, TN, 1981.
  11. Cambridge Structural Database, v. 4.6, Cambridge Crystallographic Data Center, Cambridge, UK, 1992.
  12. A.I. Kitaigorodsky, *Molecular Crystals and Molecules*, Academic Press, New York, NY, 1973.
  13. (a) L. Pauling, *The Nature of the Chemical Bond*, 3rd ed., Cornell University Press, Ithaca, NY, 1960, p. 260; (b) W.H. Baur, *Acta Cryst.*, **B28**, 1456 (1973); (c) A. Bondi, *J. Phys. Chem.*, **68**, 441 (1964); (d) T.R. Stouch and P.C. Jurs, *J. Chem. Inf. Comp. Sci.*, **26**, 4 (1986).
  14. G. Filippini and A. Gavezzotti, *Acta Cryst.*, **B48**, 230 (1992).
  15. E. Giglio, *Nature*, **222**, 339 (1969).
  16. J.R. Holden and C. Dickinson, *J. Phys. Chem.*, **81**, 1505 (1977).
  17. K. Mirsky, In *Computing in Crystallography*, H. Schenk, R. Olthof-Hazkamp, H. van Koningsveld, and G.C. Bassi, Eds., Delft University Press, Delft, The Netherlands, 1978, pp. 169–182.
  18. T.L. Starr and D.E. Williams, *Acta Cryst.*, **A33**, 771 (1977); S.R. Cox, L. Hsu, and D.E. Williams, *Acta Cryst.*, **A37**, 293 (1981).
  19. S.R. Cox and D.E. Williams, *J. Comp. Chem.*, **2**, 304 (1981).
  20. E. Kober, In *Tenth Annual Working Group Institute on the Synthesis of High Energy Density Materials*, S. Iyer, Ed., U.S. Army Armament Research, Development and Engineering Center, June 1991, p. 95; J.P. Ritchie, E.M. Kober, and A.S. Copenhaver, unpublished.
  21. C. Still, Macromodel, Columbia Chemistry Molecular Modeling System, v. 3.0, 1990, Columbia University, New York, 1990.
  22. M.J.S. Dewar, E.G. Zoebisch, E.F. Healy, and J.J.P. Stewart, *J. Am. Chem. Soc.*, **107**, 3902 (1985); J.J.P. Stewart, MOPAC, A General Molecular Orbital Package, v. 6.0, QCPE program 455, 1990.
  23. V.L. Karen and A.D. Mighell, NIST\*LATTICE, A Program to Analyze Lattice Relationships, Technical Note 1290, National Institute of Standards and Technology, Gaithersburg, MD, 1991.
  24. S.C. Nyburg, *Acta Cryst.*, **B30**, 251 (1974).
  25. A.B. Kunz and J.M. Vail, *Phys. Rev.*, **B38**, 1058 (1988); A.B. Kunz, *Int. J. Quantum Chem. Symp.*, **24**, 607 (1990); A.B. Kunz, *Theor. Chem. Acta*, to appear.
  26. *Chem3D, the Molecular Modeling System*, v. 3.0, Cambridge Scientific Computing, Cambridge, MA 1990.