difference between C(18)–C(19) and C(13)–C(14) is an exception to this.

Unlike the benzene ring discussed above, the naphthalene grouping is significantly non-planar. Least-squares planes calculations are shown in Table 8.

Both the six-membered rings (planes 1,2) making up the naphthalene group are planar to within experimental error. However, the deviations from the least-squares plane for the 10 atoms making up the naphthalene group appear to be significant but the results are not as conclusive as they would be if the thermal motion of the molecule were less. Calculations show that the planes of the two six-membered rings intersect at an angle of 1·8°.

The non-planarity of the naphthalene group does not appear to be the consequence of packing or of molecular configuration. The methyl group C(23) bonded to C(10) is directed away from the plane of the naphthalene group. The c projection of the unit cell is shown in Fig. 4. There are no significantly short intramolecular contacts; there are only two distances between non-hydrogen atoms under 3·0 Å. These are C(8)–C(21) at 2·91 and C(22)–O(1) at 2·98 Å. None of the intermolecular distances are significantly short; the minimum distances between the naphthalene and benzene groups involve hydrogen atoms: C(2)–H(14) $(x-1,y,z)$ at 2·75 Å and C(18)–H(2) $(1-x,y-\frac{1}{2},1-z)$ at 3·00 Å.

The hydrogen atoms, as can be seen in Fig. 2, have all refined to acceptable positions and the C–H distances are listed in Table 4. The average C–H bond length is 1·025 Å, a value which is somewhat shorter than the standard C–H bond length but is typical of X-ray diffraction work. The three methyl groups all have approximately staggered configurations.
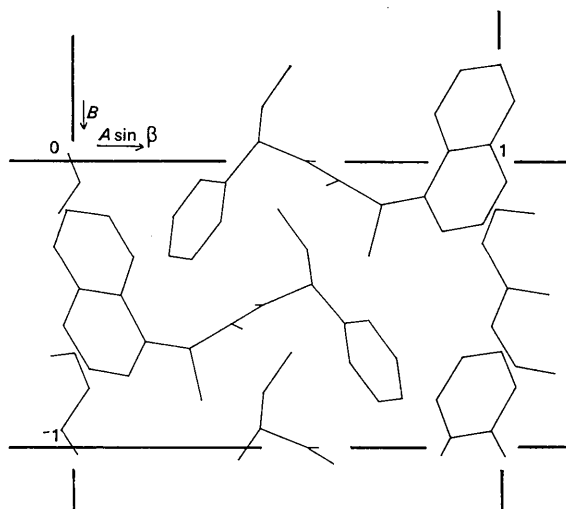


Fig. 4. The c projection of the unit cell.

### References

CERVINKA, Ö. (1966). Collection Czech. Chem. Commun. 31, 1371.

CERVINKA, Ö. & FUSEK (1968). Z. Chem. 8. 145.

CRUICKSHANK, D. W. J. (1957). Acta Cryst. 10, 504.

DREW, M. G. B., TEMPLETON, D. H. & ZALKIN, A. (1969). Acta Cryst. B25, 261.

KARLE, J. & HAUPTMAN, H. (1956). Acta Cryst. 9, 635.

KARLE, J. & KARLE, I. L. (1966). Acta Cryst. 21, 849.

International Tables for X-ray Crystallography (1962). Vol. III, p. 201. Birmingham: Kynoch Press.

MADDOX, H. S. & MADDOX, M. L. (1965). Private communication.

MISLOW, K., JACOBUS, J., DREW, M. G. B., TEMPLETON, D. H. & ZALKIN, A. (1968). J. Amer. Chem. Soc. In the press.

Tables of Interatomic Distances (1965). Special Publication No. 18. London:The Chemical Society.

---

# The Use of a 'Connectivity' or 'Bonding' Array in Molecular Geometry Calculations

By F. H. ALLEN* AND D. ROGERS

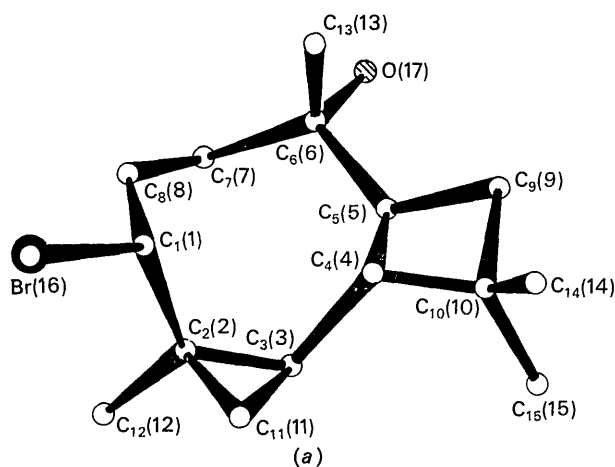Chemical Crystallography Laboratory, Imperial College, London, S.W.7, England

A symmetric array can be formed in computer store from a connectivity or bonding scheme for the crystallographic asymmetric unit. Its application to the systematic evaluation of bond lengths, direction cosines, valence angles and dihedral (torsion) angles is discussed.

### Introduction

The calculation of the various geometrical characteristics of a molecule or crystal structure is lengthy, but

* Present address: Chemistry Department, University of British Columbia, Vancouver 8, Canada

mathematically trivial. Computer programs for this task require either the calculation of all possible interatomic distances less than a specified limit (e.g. Rollett, 1965), a process easy to use but wasteful of computer time (and requiring the user to sift a large volume of output), or the calculation of specified distances, a process wasteful of the user's time and prone to mis-

(a)

| i | $j_1$ | $j_2$ | $j_3$ | $j_4$ | terminator |
|---|---|---|---|---|---|
| 1 | 2 | 8 | 16 | | * |
| 2 | 1 | 3 | 11 | 12 | * |
| 3 | 2 | 4 | 11 | | * |
| 4 | 3 | 5 | 10 | | * |
| 5 | 4 | 6 | 9 | | * |
| 6 | 5 | 7 | 13 | 17 | * |
| 7 | 6 | 8 | | | * |
| 8 | 1 | 7 | | | * |
| 9 | 5 | 10 | | | * |
| 10 | 4 | 9 | 14 | 15 | * |
| 11 | 2 | 3 | | | * |
| 12 | 2 | | | | * |
| 13 | 6 | | | | * |
| 14 | 10 | | | | * |
| 15 | 10 | | | | * |
| 16 | 1 | | | | * |
| 17 | 6 | | | | * |

(b)

values of j, l



(c)

Fig. 1 (a) The structure of humulene bromohydrin showing atomic designations. The reference numbers of the atoms in a possible input list are given in parentheses. (b) Input of bonding information: each atom ($i$) is bonded to atoms $j_1, j_2, j_3$, etc. (c) The bonding array (B) (heavily outlined) for humulene bromohydrin. Both reference numbers and atomic designations are given for clarity.

takes and omissions. The process to be described here is a compromise, ensuring an exhaustive search over one asymmetric unit (or indeed any portion of the total structure specified by the user). Needless calculations are avoided, and input and output are cut to a minimum.

The process was devised originally for the computation of dihedral angles, but is ideally suited for the quick and efficient checking of bond lengths and valence angles during structure refinement. It utilizes a connectivity or bonding array, first suggested by Mooers (1959), and developed by other workers (Lynch, 1968) as an important topographical tool. In particular the array in its present form has been employed (Penny, 1965) in the recognition of certain ring systems in larger chemical units by a computer procedure.

### The connectivity or bonding array

The fractional coordinates, $x, y, z$, and possibly their standard deviations, $\sigma_x, \sigma_y, \sigma_z$, for the $N$ atoms in an asymmetric unit are input to the computer, together with an alphanumeric atom identifier (e.g. C24). They may then be referred to by this identifier, or, more simply, by their position ($1 \to N$) in the input list. An $N \times N$ square is then cleared to zero in computer store, and the connectivity or bonding information is read in. This requires $N$ lines of input, one for each atom treated, each a string of digits of the form:

$$i \ j_1 \ j_2 \ j_3 \ j_4 \ * \qquad (1)$$

indicating that the atom numbered $i$ in the input list is bonded to those numbered $j_1, j_2, j_3, j_4$; any convenient non-numeric character may be used to terminate each line.

To set up the array (B), the cells $j_{1,2,3,4}$ on the $i$th row have unity added into them; the resultant array should be symmetric (i.e. elements $b_{ij}$ should match $b_{ji}$), and all cells should contain zero or unity only. Before proceeding these two requirements should be checked: a non-symmetric array indicates incorrect or incomplete input of bonding information, while entries $> 1$ indicate unwanted duplicates. Other checks may also be added if required, e.g. the number of entries per row or column for carbon atoms should be $\le 4$. An array (B) made up of single-bit locations is compact, but it is then unsafe to add unity to denote each bond, since duplication would cause errors.

Fig. 1(a) shows the structure of humulene bromohydrin (Allen & Rogers, 1966a,b); both the chemical numbering and atomic designations are given, together with their possible reference numbers in an input positions list. Input of bonding information is shown in Fig. 1(b), in the form of statements similar to equation (1) above. Fig. 1(c) shows the resultant array (B), in which both types of atomic designation are given for clarity.

## The systematic calculation of bond lengths, valence angles and direction cosines

The position of the $i$th atom is defined by the vector:

$$\mathbf{r}_i = x_i\mathbf{a} + y_i\mathbf{b} + z_i\mathbf{c}, \qquad (2)$$

where $\mathbf{a}, \mathbf{b}, \mathbf{c}$, are the unit-cell vectors. The vector $\mathbf{r}_{ij}$ between two atoms is simply the difference between the vectors, $\mathbf{r}_i, \mathbf{r}_j$,

$$\mathbf{r}_{ij} = \Delta x_{ij}\mathbf{a} + \Delta y_{ij}\mathbf{b} + \Delta z_{ij}\mathbf{c}, \qquad (3)$$

where $\Delta x_{ij}$ equals $x_i - x_j$, with similar expressions for $y$ and $z$. The square of the magnitude of the bond length $|\mathbf{r}_{ij}|$ is given by the scalar product of equation (3) with itself:

$$|\mathbf{r}_{ij}|^2 = \mathbf{r}_{ij} \cdot \mathbf{r}_{ij} = \Delta x_{ij}^2 a^2 + \Delta y_{ij}^2 b^2 + \Delta z_{ij}^2 c^2$$
$$+ 2ab\Delta x_{ij}\Delta y_{ij} + 2bc\Delta y_{ij}\Delta z_{ij} + 2ac\Delta x_{ij}\Delta z_{ij}. \qquad (4)$$

The direction cosines of the bond, $l_{ij}, m_{ij}, n_{ij}$, referred to axes $\mathbf{a}, \mathbf{b}, \mathbf{c}$, are given by:

$$(l, m, n)_{ij} = (a\Delta x, b\Delta y, c\Delta z)_{ij}/|\mathbf{r}_{ij}|. \qquad (5)$$

An ordered and complete sequence of bond lengths and direction cosines can be obtained by working systematically through the bonding array, reading along each row ($i$) and using equations (4), (5) for each cell ($j$) containing unity. Only the upper right triangular half of the array need be scanned for this purpose: use of the whole array produces each bonded distance twice over. The bond lengths and direction cosines are output, and also stored in similar triangular arrays $(\mathbf{R}, \mathbf{L}, \mathbf{M}, \mathbf{N})$; it should be noted that the full arrays $\mathbf{L}, \mathbf{M}, \mathbf{N}$, are anti-symmetric (*i.e.* $m_{ij} = -m_{ji}$); due allowance must be made for this sign reversal when using these half arrays to read the direction cosines of bonds in the lower triangular half (*i.e.* for vector $\mathbf{r}_{ji}$).

Valence angles, $\theta_{ijk}$, are calculated from

$$\cos\theta_{ijk} = \mathbf{r}_{ij} \cdot \mathbf{r}_{jk}/|\mathbf{r}_{ij}| |\mathbf{r}_{jk}|. \qquad (6)$$

All angles occurring at atom ($j$) can be evaluated together by reading down column ($j$) to locate the first entry ($i$), and then reading on down the column to identify the possible values of ($k$). Cos $\theta_{ijk}$ is computed and output for each indicated combination of $ijk$. The next ($i$) is selected by moving down the column, and is combined with all $k$'s indicated *lower down* the same column, thus avoiding duplication. The process is repeated until all permitted combinations ($ik$) are exhausted for atom ($j$); scanning is then transferred to the next column, $j+1$. In every case cos $\theta_{ijk}$ is stored, together with the integers $i, j, k$, for use in the calculation of dihedral angles.

## The systematic calculations of dihedral (torsion) angles

An attempt will first be made to define and generalize the nomenclature for dihedral, or torsion, angles.

Reference to Fig. 2(*a*) shows two groups of atoms, $ABC$ bonded to $D$, and $FGH$ bonded to $E$ which is below the plane of the paper. The Newman projection (Newman, 1955) down the bond $D \rightarrow E$ is shown in Fig. 2(*b*).

The dihedral or torsion angle ($\omega$) is the angle between say, the projections of $CD$ and $EF$ on a plane perpendicular to the bond $DE$. Thus we may define the angle as:

$$\omega\{C(D, E)F\} = x^0. \qquad (7)$$

Other more general dihedral angles, of use in certain contexts, can be described analogously. Thus the angle between the projections of $AD$ and $CD$ on the plane perpendicular to $DE$ may be written:

$$\omega\{A(D, E)C\} = x^0. \qquad (8)$$
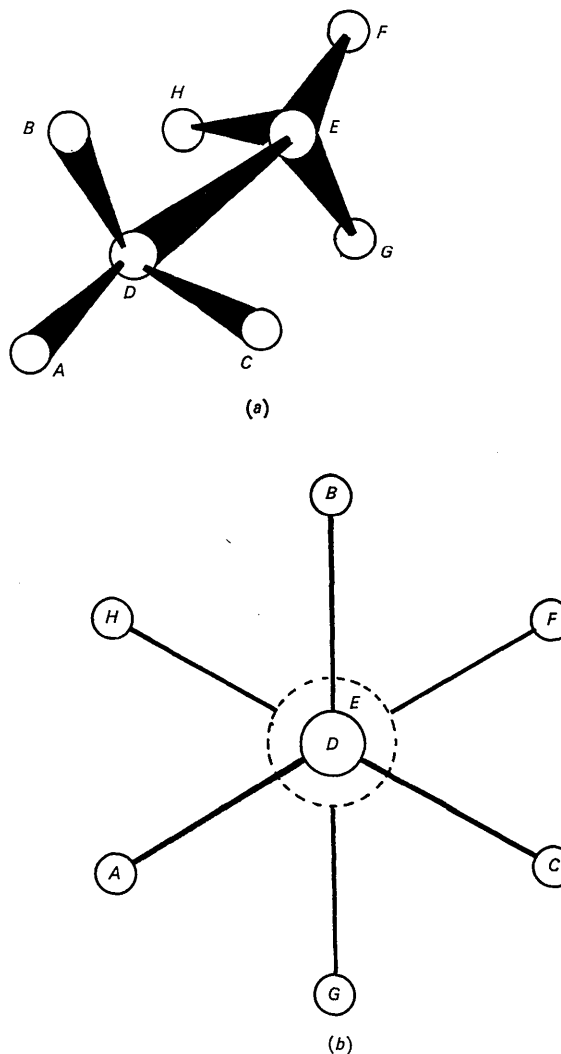




Fig. 2. Nomenclature for dihedral angles (*a*) Atom $E$ is below the plane of the paper. (*b*) Newman projection down the bond $D \rightarrow E$.

Similarly the angle between proj($GE$), proj($EF$), as seen from viewing point $D$ is:

$$\omega\{G(\mathbf{D},E)F\}=x^0 .\tag{9}$$

The letters in parentheses indicate the direction of view down the bond, *i.e.* from $D$ to $E$; letters in Clarendon
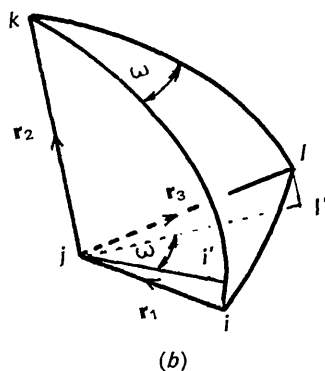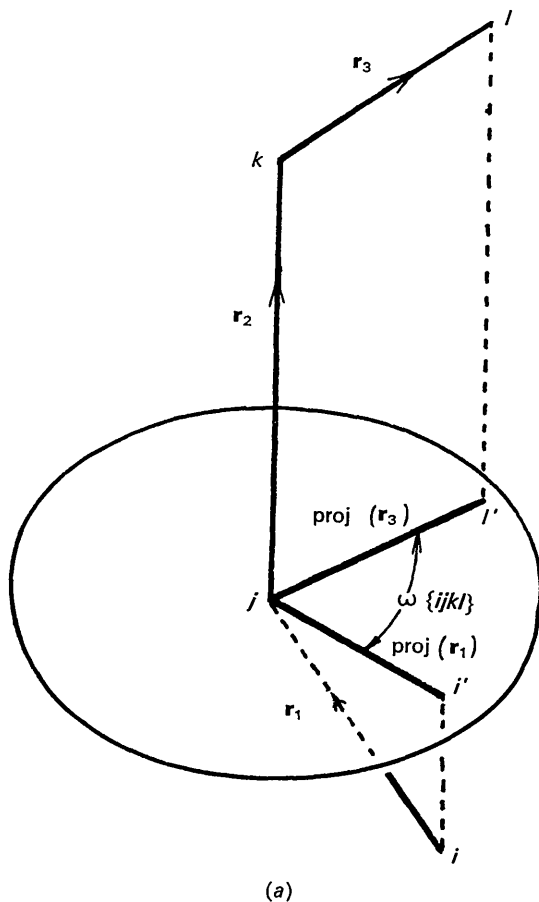


(a)



(b)

Fig. 3. (a) Definition of the dihedral angle $\omega_{\{ijkl\}}$, where $ijkl$ are for example $C(D,E)F$ in equation 7. (b) The vectors in relation to a spherical triangle. The angle $\omega_{\{ijkl\}}$ ($= <i'jl'$) is the 'angle' $ikl$ of the triangle, *i.e.* between the planes $ijk$, $ljk$; the 'sides' of the triangle are the valence angles $<\mathbf{r}_1\mathbf{r}_2$, $<\mathbf{r}_2\mathbf{r}_3$, and the angle $<\mathbf{r}_1\mathbf{r}_3$, whence equation 10 may be derived from the cosine formula for spherical trianges.1

type indicate atoms belonging to the nearer group in the viewing direction. For ordinary dihedral angles, representing twist or torsion about the bond $DE$ [equation (7)] the first quoted atom is a member of the nearer group, and the last of the farther group. For the more general dihedral angles they are both of the nearer group [equation (8)], or both of the farther group [equation (9)]. Furthermore it is often imperative to distinguish between $sp^3$–$sp^3$, $sp^3$–$sp^2$, and $sp^2$–$sp^2$ torsion angles in an organic structure, where the hybridization states referred to are those of atoms $D$ and $E$ respectively. This can be effected by reserving the symbol ($\omega$) for the first mentioned, and referring to the other two as ($\delta$) and ($\tau$) respectively. General adoption of this type of nomenclature saves the reader from continually referring to the structural formula.

In all cases the sense of rotation is defined in terms of the movement required to rotate the first mentioned bond [$CD$ in equation (7), $AD$ in (8), $GE$ in (9)] to overlie the projection of the second bond ($EF$, $DC$, $EF$, respectively). Clockwise rotation of the first-quoted bond is taken as positive, anti-clockwise rotation is taken as negative (Klyne & Prelog, 1960). Thus, with reference to Fig. 2(b), we may write:

$$\omega\{A(\mathbf{D},E)H\}= +60°\tag{7a}$$

$$\omega\{A(\mathbf{D},E)G\} = -60°\tag{7b}$$

$$\omega\{A(\mathbf{D},E)B\} = +120°\tag{8a}$$

$$\omega\{G(\mathbf{D},E)F\} = -120° .\tag{9a}$$

Fig. 3(a) shows the relationship between the three bond vectors involved, *i.e.* $\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3$, represent bonds $CD,DE,EF$, in equation (7), or $AD,DE,DC$, in (8), or $GE,DE,EF$, in (9). Fig. 3(b) shows the vectors in relation to a spherical triangle, whence we may write:

$$\cos(\omega)=[\cos(<\mathbf{r}_1\mathbf{r}_3)$$
$$- \cos(<\mathbf{r}_1\mathbf{r}_2)\cos(<\mathbf{r}_2\mathbf{r}_3)]/\sin(<\mathbf{r}_1\mathbf{r}_2)\sin(<\mathbf{r}_2\mathbf{r}_3),$$
$$\tag{10}$$

where $<\mathbf{r}_1\mathbf{r}_2$ and $<\mathbf{r}_2\mathbf{r}_3$ are the valence angles:

$\theta_{CDE}$ and $\theta_{DEF}$ in equation (7),
$\theta_{ADE}$ and $\theta_{CDE}$ in equation (8),

and

$\theta_{GED}$ and $\theta_{FED}$ in equation (9),

and $<\mathbf{r}_1\mathbf{r}_3$ is the angle between $\mathbf{r}_1,\mathbf{r}_3$ in each case.

The angles $\omega_{\{i(j,k)l\}}$ can be evaluated, and systematically output in a convenient order, using the bonding array as follows. For a given viewing bond $DE$, identified by unity in cell $b_{jk}$, look up and down column ($j$) to identify possible ($i$'s), and backward and forward along row ($k$) to identify possible ($l$'s). Equation (10) is then evaluated for all combinations of:

(i) $i,l(i\neq l)$    *cf.* equation (7),

(ii) $i,i'(i\neq i')$  [both obtained from column ($j$)]
     *cf.* equation (8),

(iii) $l, l'(l \neq l')$ [both obtained from row $(k)$]
    *cf.* equation (9) ,

for any given combination $jk$. The process is repeated systematically for all bonds $(jk)$ specified in the lower triangular half of array (**B**); this avoids duplication since $\omega\{ijkl\}$ is equivalent to $\omega\{lkji\}$ in both magnitude and sign. The sines and cosines of $<r_1r_2$, $<r_2r_3$, required by equation (10), can be obtained from the previously stored valence angles. Cos $(<r_1r_3)$ can be obtained using:

$$\cos (<r_1r_3) = \cos (<r_{ij}r_{kl}) = r_{ij} \cdot r_{kl}/|r_{ij}| |r_{kl}| . \quad (11)$$

The sign of the rotation is given by the sign of the determinant $V$:

$$V = \begin{vmatrix} l_{ij} & m_{ij} & n_{ij} \\ l_{jk} & m_{jk} & n_{jk} \\ l_{kl} & m_{kl} & n_{kl} \end{vmatrix} \quad (12)$$

which is proportional to the volume of the parallelepipedon defined by the vectors $r_1, r_2, r_3$.

The method was tested at the University of London Institute of Computer Science using the I.C.T. Atlas machine; the co-operation of the Staff is gratefully acknowledged. One of us (F.H.A.) thanks the Science Research Council for a Research Studentship.

### References

ALLEN, F. H. & ROGERS, D. (1966a). *Chem. Comm.* p. 582.
ALLEN, F. H. & ROGERS, D. (1966b). *J. Chem. Soc.* In the press.
KLYNE, W. & PRELOG, V. (1960). *Experientia,* 16, 521.
LYNCH, M. F. (1968). *Endeavour,* 27, 68.
MOOERS, C. (1959). *Zator Tech. Bull.* No. 59. Boston: Zator Co.
NEWMAN, M.S. (1955). *J. Chem. Educ.* 32, 344.
ROLLETT, J. S. (1965). *Computing Methods in Crystallography,* p. 25 *et. seq.* Oxford and London: Pergamon.

# The Crystal and Molecular Structure of 6-Mercaptopurine Monohydrate

BY E. SLETTEN*, J. SLETTEN* AND L. H. JENSEN

*Department of Biological Structure, University of Washington, Seattle, Washington, U.S.A.*

The crystal structure of 6-mercaptopurine monohydrate, $C_5H_6N_4OS$, has been determined by X-ray diffraction methods. 2186 independent reflections were measured on a diffractometer with Mo $K\alpha$ radiation. The structure was refined to an $R = 0.038$ by full-matrix least squares. Standard deviations in the bond lengths involving only non-hydrogen atoms are approximately 0.002 Å and for those involving hydrogen atoms, 0.02 Å. The corresponding standard deviations in angles are 0.1° and 1° respectively. The molecule is not strictly planar; the maximum deviation of 0.029 Å is displayed by N(1). The sulfur atom is attached in thion form to the C(6) atom, the C–S bond length being 1.676 Å and the angle N(1)–C(6)–C(5), 110.4°. In the imidazole ring the hydrogen atom is bonded to N(7) rather than to N(9). The water molecule participates in four hydrogen bonds in a nearly tetrahedral arrangement.

## Introduction

A number of sulfur derivatives of nucleic acid components have been shown to possess anticarcinogenic properties. The effect of the drug 6-thioguanine may be a result of the incorporation of this molecule in the nucleic acid as an antagonist to a natural purine base (LePage, 1960), while the powerful drug 6-mercaptopurine is thought to enter the synthesis of nucleic acid at an early stage, probably as a competitor to hypoxanthine in the conversion of inosinic acid to nucleic acid (Skipper, 1954). The activity of the sulfur derivatives of nucleic acid components is very sensitive to

minor structural modification; *e.g.* the purine molecule substituted by sulfur in the 2-position displays no anticarcinogenic effect (Robins, 1964). An accurate structure determination of 6-mercaptopurine monohydrate was undertaken to investigate the packing mode of this type of molecule and to determine to what extent the sulfur substituent alters the geometry of the purine skeleton.

## Experimental

Crystals satisfactory for X-ray work were selected directly from a sample of 6-mercaptopurine obtained from the Sigma Chemical Company, St.Louis, Missouri. The crystals are faintly yellow, with a slightly distorted bipyramidal shape. Preliminary cell dimensions and space-group determination were carried out

---

* Permanent address: Department of Chemistry, University of Bergen, Bergen, Norway.