Acta Cryst. (1978). B34, 2534–2546

Computer Retrieval and Analysis of Molecular Geometry. III. Geometry of the β-1′-Aminofuranoside Fragment

BY PETER MURRAY-RUST

Department of Chemistry, University of Stirling, Stirling, Scotland

AND SAM MOTHERWELL

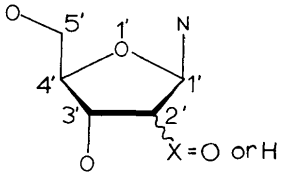
Cambridge Crystallographic Data Centre, University Chemical Laboratory, Lensfield Road, Cambridge, England

(*Received* 17 *January* 1978; *accepted* 9 *March* 1978)

Automatic methods were used to search the Cambridge Crystallographic Data File for all compounds containing a nucleoside fragment. 99 references were found to structures with atomic positions on file and from these the molecular geometry was calculated for 110 occurrences of the β-nucleoside fragment (78 derived from ribose, 19 from 2′-deoxyribose and 13 from arabinose). Factor analysis of the torsion angles showed that there were only three independent factors in the variability of the conformation: two concerned with the ring and one with the C(4′) side chain. Interpretation of the ring factors showed two pathways for conformational interchange: pseudorotation and ring flattening. When the most accurate data (about 55%) were selected, small bond- and torsion-angle variations were seen to be linked to pseudorotation and it was possible to describe these by periodic functions whose constants were determined by regression analysis.

Introduction

The retrieval and analysis of molecular geometry by methods described in the preceding papers (Murray-Rust & Motherwell, 1978; hereafter MM; Murray-Rust & Bland, 1978; hereafter MB) is well suited to investigating the structural units of macromolecules. Any relationships which may exist between the geometrical parameters of a monomer can be of use in model-building techniques, particularly for the interpretation of electron density maps, *e.g.* in $tRNA^{Phe}$ (Jack, Klug & Ladner, 1976). Accordingly we have studied the unit common to all nucleic acids, the β-1′-aminofuranoside fragment (I); this occurs in both ribose and 2′-deoxyribose derivatives.

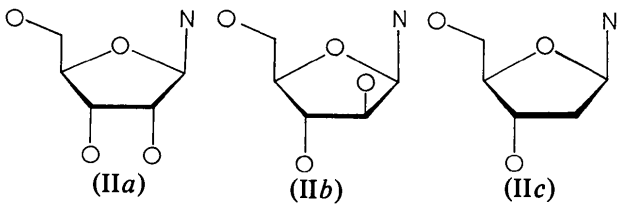


(I)

A discussion of the relevance of our findings to nucleic acid conformation will be published elsewhere and we shall not elaborate this aspect. Here we describe the detailed procedures used for the screening of data and for the discovery of relationships between parameters. The operations described below follow closely the general strategy put forward in MM and MB. They have been divided into two groups: one concerned with searching and screening and the other concerned with analysis of the retrieved molecular geometry. Because of the iterative use of screening in analysis, this division is somewhat arbitrary, and may not always be as clear-cut.

Searching and screening

We wish to retrieve data for compounds which contain any of the three fragments (II); these are, respectively, derivatives of ribose, 2-deoxyribose and arabinose which all contain the sugar as a furanosyl ring. We want only β-nucleotides but will accept both D and L derivatives of the sugars above. (Only D sugars occur in nucleic acids but since enantiomers are isometric, L sugars can be used for the analysis of molecular geometry.) It is important that the searching be totally efficient and that *only* compounds belonging to categories (II) are accepted. Since there will be many data, we can afford to exclude all structures flagged as error sets, *i.e.* containing one or more uncorrected typographical errors. At this stage we retain all error-free sets (in the above sense) although some of these may contain large experimental errors. It is often valuable to postpone the rejection of inaccurate data sets until after preliminary statistical analysis, since if the observed variance is very large [as is found for the torsion angles in (I)] even structures of very limited accuracy can be used. Unless very large files have to be set up, screening on accuracy is therefore usually an integral part of the iterative analysis of molecular geometry.



All the screening processes described in MM were used (Fig. 1). It would have been possible to omit the bibliographic search but its simplicity and efficiency as a pre-screen for the connectivity screen saved considerable computer time. As with the accuracy screen (on raw data) the derived-data screen (on conformation) was delayed until after the formation of the molecular-geometry system file. Because unexpected effects and relationships may be revealed in statistical analysis the system file should be as comprehensive as possible in terms of the number of parameters for each fragment and the number of fragments themselves.

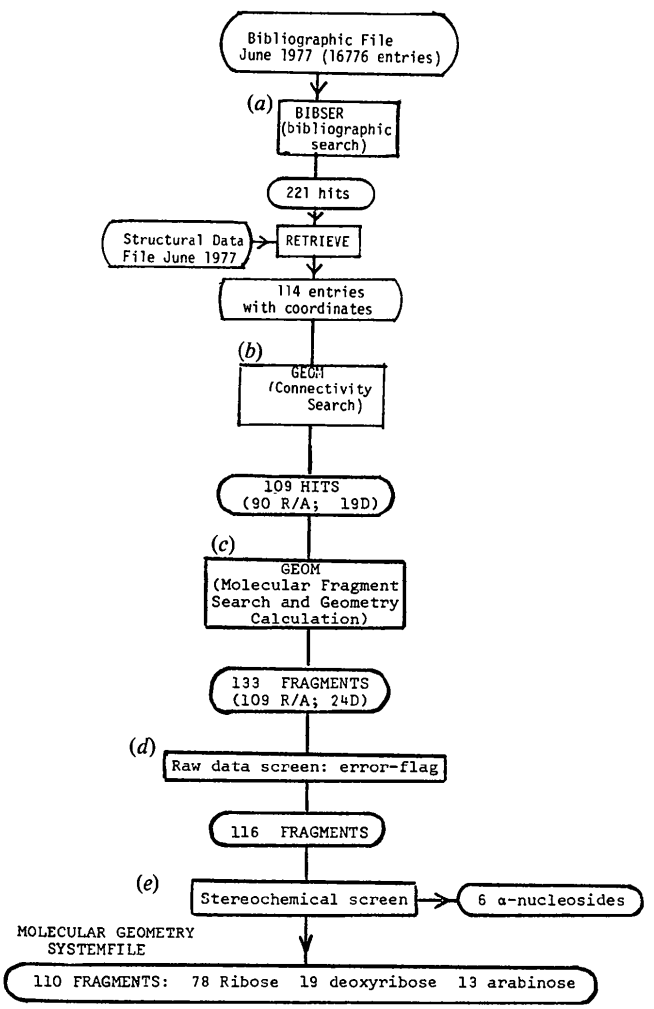


Fig. 1. Progress of the searching and screening of data; (*a*)–(*e*) refer to the text and to Table 1. Abbreviations: *R*, ribose; *A*, arabinose; *D*, 2′-deoxyribose.

(*a*) *Bibliographic search*

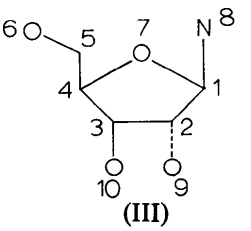
This is very simple (Table 1*a*) since all fragments (II) will occur in class 47 (nucleosides and nucleotides). It lets through about 13% of unwanted structures, for example 2′,3′-dideoxy derivatives, *C*-nucleosides, halogeno-substituted compounds [at C(5′) or C(2′)] and pyranosides. About half the hits (a hit is a

Table 1. *Details of screens*

(*a*) Bibliographic screen (program *BIBSER*). (All nucleosides are given this classification.)

```
Q *CLASS '47'
```

(*b*) Connectivity screen (program *GEOM*)



(III)

```
AT1 C 3
AT2 C 3
AT3 C 3
AT4 C 3
AT5 C 2
AT6 O 1
AT7 O 2
AT8 N 1
AT9 O 1
AT10 O 1
BO 8 1
BO 1 7
BO 7 4
BO 4 3
BO 3 2
BO 2 1
BO 4 5
BO 5 6
BO 2 9
BO 3 10
```

This retrieves ribose and arabinose derivatives.

A similar input with the omission of AT9 and the alteration of AT2 to AT2 C 2 E retrieves 2′-deoxyribose derivatives.

(*c*) Geometry calculation (program *GEOM*).

The atom numbering is shown in (III); thus the quantity A03 is the C(3′)–C(4′)–O(1′) angle and T02 is the C(2′)–C(3′)–C(4′)–O(1′) torsion angle. The 14 bond and 15 torsion angles are input to the systemfile as A01 to A14, and T01 to T15.

```
DEF A01   1 2 3
DEF A02   2 3 4
DEF A03   3 4 7
DEF A04   4 7 1
DEF A05   7 1 2
DEF A06   8 1 7
DEF A07   8 1 2
DEF A08   9 2 1
DEF A09   9 2 3
DEF A10  10 3 2
DEF A11  10 3 4
DEF A12   5 4 3
DEF A13   5 4 7
DEF A14   4 5 6
DEF T01   1 2 3 4
DEF T02   2 3 4 7
DEF T03   3 4 7 1
DEF T04   4 7 1 2
DEF T05   7 1 2 3
DEF T06   8 1 7 4
DEF T07   8 1 2 3
DEF T08   9 2 1 7
DEF T09   9 2 3 4
DEF T10  10 3 2 1
DEF T11  10 3 4 7
DEF T12   5 4 3 2
DEF T13   5 4 7 1
DEF T14   3 4 5 6
DEF T15   7 4 5 6
```

Table 1 (*cont.*)

(*d*) Raw-data screen (using the *SPSS* package). The error flag is listed by *GEOM*; a value of 1 indicates an error set.

```
SELECT IF   (ERR NE 1)
```

(*e*) Stereochemical screen (using the *SPSS* package)
The following torsion angles are enough to define all the configurations (VAR LABELS is an *SPSS* card giving a glossary of variables).

```
VAR LABELS  T05   O(1')–C(1')–C(2')–C(3')
                                  TORSION ANGLE
VAR LABELS  T01   C(1')–C(2')–C(3')–C(4')
                                  TORSION ANGLE
VAR LABELS  T04   C(4')–O(1')–C(1')–C(2')
                                  TORSION ANGLE
VAR LABELS  T06   C(4')–O(1')–C(1')–N
                                  TORSION ANGLE
VAR LABELS  T02   C(2')–C(3')–C(4')–O(1')
                                  TORSION ANGLE
VAR LABELS  T08   O(1')–C(1')–C(2')–O(2')
                                  TORSION ANGLE
VAR LABELS  T10   C(1')–C(2')–C(3')–O(3')
                                  TORSION ANGLE
VAR LABELS  T12   C(2')–C(3')–C(4')–C(5')
                                  TORSION ANGLE
```

The configurations at C(1') to C(4') are found (Q1 to Q4 are all initially set to 1, corresponding to β-D-ribose).

```
IF    (T04 LT T06) Q1 = -1
IF    (T08 LT T05) Q2 = -1
IF    (T10 LT T01) Q3 = -1
IF    (T02 LT T12) Q4 = -1
```

If Q4 is negative, this corresponds to an L sugar and the signs of all torsion angles are changed.

```
DO REPEAT    XT = T01 TO T15
IF           (Q4 LT 0) XT = -XT
END REPEAT
```

Configurations are compared with those in D-ribose

```
SELECT IF  (Q3 EQ Q4)  (rejects xylose and lyxose)
SELECT IF  (Q1 EQ Q4)  (rejects α-nucleosides)
IF         (Q1 EQ Q2)  TYPE = 'ARAB'
IF         (Q1 NE Q2)  TYPE = 'RIBO'
```

file entry fulfilling the conditions of a search question) had no atomic coordinates on the CCDC file. This discrepancy was due in small part to the lag between the CCDC processing the small amount of bibliographic information for a structure and typing and checking the large numbers of atomic coordinates. A more serious and regrettable cause is the preliminary publication of many structures without atomic coordinates.

(*b*) *Connectivity screens*

At this stage a chemical connectivity search (*CONNSER*) including bond type would normally be run. For a fragment as large as (I) where different bond types are not expected, the fragment definition in *GEOM* is as efficient as *CONNSER* and considerable computer time can be saved by proceeding directly to the molecular-geometry calculations. These were carried out (Table 1*b*) for both ribose/arabinose (II*a* and II*b*) and 2'-deoxyribose (II*c*); the hits were subsequently merged. The connectivity screens are almost totally efficient but the specification for AT5 is later shown to be not quite precise enough. Among the 99 hits are some relating to independent studies of the same crystal form of a compound.

(*c*) *Molecular-geometry calculation*

For this study the only desired data required were bond angles and torsion angles. (Bond-length variations were thought to be relatively unimportant at this stage, but might well be included in a further study. Since system files are allowed to be very large it is normally advisable to include all the parameters that might conceivably be used in a subsequent analysis.) All the bond angles and most of the torsion angles were calculated (Table 1); the atoms are referenced by their number in the connectivity search.

(*d*) *Raw-data screen for errors*

Data sets with error flags were rejected automatically (Table 1*d*). It is probable that some of them contain no errors in the coordinates of the atoms in the fragment (I) but since there are enough other data it is simplest to exclude them without further consideration.

(*e*) *Stereochemical screen*

Fragment (I) has four chiral centres and it is essential that all of them have the correct configurations (II*a–c*) or the mirror images. The configurations can be compared with β-D-ribofuranoside by using the ring and substituent torsion angles (Table 1*e*). In some cases the configuration at C(4') was that of an L sugar (although the authors still named the sugar as D) and for these data sets the signs of all torsion angles were changed (to give the isometric enantiomer). If either of the configurations at C(3') or C(1') was opposite to ribose the fragment was excluded; this occurred in six cases of α-nucleosides such as vitamin B_{12}. The 2'-epimer type (coded as variable TYPE) is added to the system file, which allows us to calculate statistics for any required subset of the data for a particular isomer (*e.g.* arabinose derivatives).

Molecular geometry for ribose and arabinose derivatives is merged with that for 2'-deoxy compounds. The system file is now complete and contains the information in Table 2. For the 2'-deoxy derivatives the variables involving O(2') are irrelevant. These are given a flag of 999, recognized later by the MISSING VALUES card, which ensures that statistical analysis involving (say) angle C(1')–C(2')–O(2') is not invalidated by the 2'-deoxy data. Since *SPSS* system files can have 500 variables, it would have been possible to write all bond lengths, angles and torsion angles to file

Table 2. *Original systemfile after searching and screening*

	Variable	Format	Notes
1	NAME	A8	Cambridge reference code uniquely identifying the structure (but not the fragment).
2	A01	F8.1	A01 (see Table 1) representing the C(1′)–C(2′)–C(3′) angle.
⋮	⋮		
15	A14	F8.1	A14: the C(4′)–C(5′)–O(5′) angle.
16	T01	F8.1	T01: the C(1′)–C(2′)–C(3′)–C(4′) torsion angle.
⋮	⋮		
30	T15	F8.1	T15: the O(1′)–C(4′)–C(5′)–O(5′) torsion angle.
31	TYPE	A4	'ARAB', 'RIBO' or 'DEOX'; identifying the sugar.
32	ERR	F1.0	1 represents an error set.
33	RFACTOR	F8.1	Lowest reported *R* factor.
34	AS	F1.0	E.s.d. category (1, 2, 3, or 4).

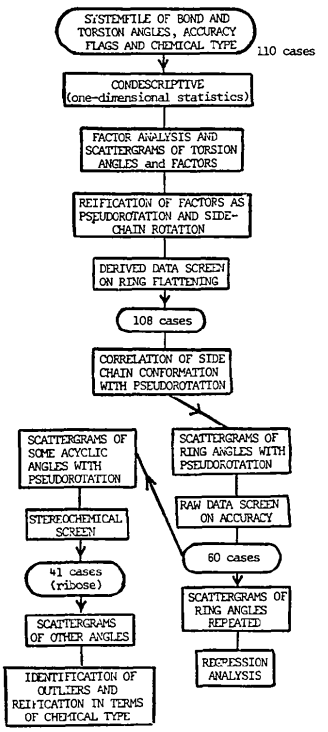


Fig. 2. Progress of the statistical analysis.

in case they might have been used later. In a similar way chemical information (such as NMR parameters) could have been added at this stage.

Subsequent statistical analysis may, of course, reveal an effect for which extra geometrical information is required (such as inertial axes). This necessitates recreation of the system file, and is to be avoided if possible by including all conceivable output from *GEOM*. In particular, if orthogonal atomic coordinates are output to the system file it is possible to calculate any intramolecular distance or angle with *SPSS* (Nie, Hull, Jenkins, Steinbrenner & Bent, 1975).

Statistical analysis of the conformations of nucleosides

There is considerable literature on the conformations of nucleosides and nucleotides which might suggest several models against which to test the data. In the account that follows, however, we have taken a naive approach and assumed no previous knowledge of this literature, preferring to process the data in an objective phenomenological manner. This apparent perversity has the merit of avoiding preconceptions which might result in a biased treatment of the data. In the present study we find an unreported pathway for deformation of nucleosides.

Table 3. *Means, standard deviations and variances of variables*

(*a*) *SPSS* input. Note the exclusion of data from 2′-deoxy compounds related to four angles.

MISSING VALUES	A08, A09, T08, T09 (999)
CONDESCRIPTIVE	A01, TO A14, T01 TO T14
STATISTICS	1, 5, 6

(*b*) Statistical quantities for all 110 fragments

	Variable (p_{ij})	Mean ($\bar{p}_j$)	Standard deviation (σ_j)	Variance
A01	C(1′)–C(2′)–C(3′)	101·6°	1·4°	1·9(°)²
A02	C(2′)–C(3′)–C(4′)	102·6	1·2	1·4
A03	C(3′)–C(4′)–O(1′)	105·0	1·4	2·1
A04	C(4′)–O(1′)–C(1′)	109·4	1·6	2·6
A05	O(1′)–C(1′)–C(2′)	106·4	1·4	1·9
A06	O(1′)–C(1′)–N	108·4	1·2	1·5
A07	N–C(1′)–C(2′)	113·0	3·0	9·0
A08	C(1′)–C(2′)–O(2′)	109·7	3·1	9·4
A09	C(3′)–C(2′)–O(2′)	111·5	3·4	11·7
A10	C(2′)–C(3′)–O(3′)	111·2	3·3	11·0
A11	C(4′)–C(3′)–O(3′)	110·8	2·6	6·8
A12	C(3′)–C(4′)–C(5′)	114·9	2·2	4·7
A13	O(1′)–C(4′)–C(5′)	109·8	1·7	2·8
A14	C(4′)–C(5′)–O(5′)	110·9	2·8	7·6
T01	C(1′)–C(2′)–C(3′)–C(4′)	−3	34	1155
T02	C(2′)–C(3′)–C(4′)–O(1′)	−2	30	906
T03	C(3′)–C(4′)–O(1′)–C(1′)	6	19	356
T04	C(4′)–O(1′)–C(1′)–C(2′)	−8	18	333
T05	O(1′)–C(1′)–C(2′)–C(3′)	7	30	870
T06	N–C(1′)–O(1′)–C(4′)	−130	19	386
T07	N–C(1′)–C(2′)–C(3′)	126	29	856
T08	O(2′)–C(2′)–C(1′)–N	54	210	44379
T09	O(2′)–C(2′)–C(3′)–C(4′)	−111	183	33550
T10	O(3′)–C(3′)–C(2′)–C(1′)	115	37	1373
T11	O(3′)–C(3′)–C(4′)–O(1′)	−120	34	1137
T12	C(5′)–C(4′)–C(3′)–C(2′)	−122	30	906
T13	C(5′)–C(4′)–O(1′)–C(1′)	130	19	371
T14	O(5′)–C(5′)–C(4′)–C(3′)	72	71	5072
T15	O(5′)–C(5′)–C(4′)–O(1′)	−46	71	5045

(*c*) Quantities for 78 ribose fragments for the four angles relating to C(2′)

	Variable	Mean	Standard deviation	Variance
A08	C(1′)–C(2′)–O(2′)	109·5°	2·9°	8·6(°)²
A09	C(3′)–C(2′)–O(2′)	111·6	3·5	12·5
T08	O(2′)–C(2′)–C(1′)–N	121	34	1188
T09	O(2′)–C(2′)–C(3′)–C(4′)	−114	38	1454

For molecules less studied than nucleosides objectivity in approach may be essential, not only to discover relationships but also to make sure none has been overlooked. Schematic progress of the analysis is shown in Fig. 2.

Variability of individual parameters

The means, standard deviations (s.d.'s) and variances of the 29 parameters are listed with the CONDESCRIPTIVE routine (Table 3), and several effects are immediately apparent. The variance in torsion angles is much greater than that expected from experimental errors (which although not known exactly is not likely to exceed 3–4°). The large variance and s.d. of T08 and T09 are a consequence of combining arabinose and ribose derivatives; in subsequent analysis these angles are either excluded or the two sets of data are treated separately. The results for ribose derivatives alone are also given in Table 3.

Torsion angles for acyclic bonds [*e.g.* T14 and T15 which relate to C(4′)—C(5′)] present a problem because of their periodic nature. The difference between values of, say, +180 and −180° will have a serious effect on the statistical analysis. If the distribution is continuous over the whole rotational range, no correlations can easily be carried out. For other distributions it is essential that the ends of the range do not fall near modes. The distributions of T14 and T15 are trimodal and the ends of their ranges have been adjusted to fall between the two smallest modes.

Whilst the variance in torsion angles is clearly structural, the variance in bond angles is due in considerable part to experimental errors. Since structures of low accuracy are still present, standard deviations of up to 2° might be expected. It will be seen in later analysis that structural effects are also important in bond-angle variance, especially when inaccurate data sets are excluded.

Factor analysis of torsion angles

The 15 torsion angles cannot be independent [relationships for five-membered rings have been derived from a geometrical point of view, *e.g.* by Dunitz (1972)]. Factor analysis reveals the dimensionality of the problem and shows the best viewpoints for further investigation of variance. Standardized data (*z*-scores) are computed for the torsion angles

$$z_{ij} = (p_{ij} - \bar{p}_j)/\sigma_j \qquad (1)$$

where p_{ij} and σ_j are listed in Table 3; the z_i are distributed with zero mean and unit variance and all quantities derived in the factor analysis relate to them. To refer factors directly to torsion angles retransformation by (1) would be necessary.

Table 4. *Factor analysis*

(*a*) *SPSS* input. Note the exclusion of T08, T09 belonging to two different populations. For rotation of factor axes NOROTATE is replaced by VARIMAX.

```
MISSING VALUES   T08, T09 (999)
FACTOR           VARIABLES = T01 TO T07,
                   T10 TO T15/TYPE = PA1/ROTATE
                   = NOROTATE/FACSCORE/
                   MINEIGEN = 1·0/
```

(*b*) Eigenvalues of the correlation matrix (*i.e.* absolute values of factors) and the percentage of variance each explains

Factor	Eigenvalue	Percentage of variance
1	7·98	61·4
2	3·03	23·3
3	1·95	15·0
4	0·02	0·1
5	0·009	
6	0·004	
7	0·002	
8	0·002	
9	0·0009	
10	0·0001	
11	0·00007	
12	0·00005	
13	0·00003	

(*c*) The three significant factors (*i.e.* the eigenvalue-weighted eigenvectors). The components (loadings) are shown for the thirteen torsion angles. Lines represent approximate blocks.

Variable	F_1	F_2	F_3
T01	−1·00	0·06	−0·01
T02	0·97	0·22	0·04
T03	−0·66	−0·73	0·13
T04	−0·53	0·84	−0·15
T05	0·93	−0·35	0·06
T06	−0·48	0·86	−0·13
T07	0·93	−0·36	0·06
T10	−1·00	0·04	0·01
T11	0·97	0·23	−0·6
T12	0·98	0·20	−0·04
T13	−0·67	−0·72	−0·13
T14	0·03	0·26	0·96
T15	0·03	0·26	0·97

(*d*) Rotated-factor matrix

The rotation matrix, found by varimax, is:

0·85	−0·52	0·02
0·51	0·83	0·22
−0·13	−0·08	0·97

Postmultiplication of the matrix in (*c*) gives the rotated-factor matrix, which has almost exact block structure.

Variable	G_1	G_2	G_3
T01	−0·82	0·58	−0·01
T02	0·95	−0·32	0·03
T03	−0·96	−0·29	−0·04
T04	0	1·00	0·03
T05	0·61	−0·79	0
T06	0·04	0·99	0·05
T07	0·60	−0·80	0
T10	−0·83	0·56	0
T11	0·95	−0·31	0·01
T12	0·94	−0·34	0·02
T13	−0·96	−0·27	0·05
T14	0·03	0·03	1·0
T15	0·03	0·03	1·0

The analysis is shown in Table 4. Variables T08 and T09 are excluded and the principal components of the distribution found for the other 13 torsion angles. It is strikingly apparent that only three eigenvalues are significant, showing a reduction in dimensionality by 10. (This could perhaps have been predicted from geometrical considerations in the present case, but in general it may be difficult to derive relationships for flexible molecules.) The first three factors are tabulated (Table 4*c*) and we attempt to explain them in chemical terms (reification). As suggested in MB there is often a close parallel with normal coordinates in molecular vibrations and reification of factors is similar to the identification of group frequencies. Thus F_1 and F_2 can be labelled ring-puckering factors and F_3 side-chain libration/rotation if the small coefficients are considered unimportant. Indeed the factors show an almost blocked structure, analogous to the near-independence of group frequencies in many vibrating molecules.

It is possible to *rotate* the factor axes so that the block structure becomes even more apparent. This is a common procedure in the social sciences but it must be used with caution in the present context. The original solutions to the analysis are the p non-zero factors **F** (principal components) where

$$\mathbf{Z}^{\mathrm{T}}_{mn}\mathbf{Z}_{nm} = \mathbf{R}_{mm} = \mathbf{F}_{mp}\mathbf{F}^{\mathrm{T}}_{pm} = (\mathbf{E}_{mp}\boldsymbol{\lambda}^{1/2}_{pp})(\boldsymbol{\lambda}^{1/2}_{pp}\mathbf{E}^{\mathrm{T}}_{pm}). \quad (2)$$

If we *rotate* the p factors using a $(p \times p)$ matrix **A** we can get new factors **G**:

$$\mathbf{G}_{mp} = \mathbf{F}_{mp}\mathbf{A}_{pp} \quad (3)$$

which still satisfy the equation

$$\mathbf{G}_{mp}\mathbf{G}^{\mathrm{T}}_{pm} = \mathbf{R}_{mm} \quad (4)$$

since

$$\mathbf{A}_{pp}\mathbf{A}^{\mathrm{T}}_{pp} = \mathbf{I}_{pp}. \quad (5)$$

Careful choice of **A** [by the varimax or quartimax methods; see, for example, Harman (1967)] will result in many of the factor components being small, thus aiding the interpretation of the factors. However, the new factor axes no longer lie along the eigenvectors (the directions of maximum variance) and the parallel with molecular vibrations is lost. We strongly recommend that rotation is only used in preliminary investigation of the problem; any quantitative results must be derived from unrotated factors parallel to the eigenvectors.

Rotation of the factors is shown in Table 4(*d*), where G_3 is shown to relate solely to T14 and T15. It is clear that this factor corresponds to side-chain conformation alone and is entirely independent of all the ring torsion angles. These other variables, however, are involved in the two factors G_1 and G_2 in a complicated manner and further analysis requires examination of the unrotated *factor scores*. These are the original z-scores (**Z**) referred to the factors **F**(*i.e.* the new axes are formed by transforming the p_j by **E**). The factor scores **S** are defined by:

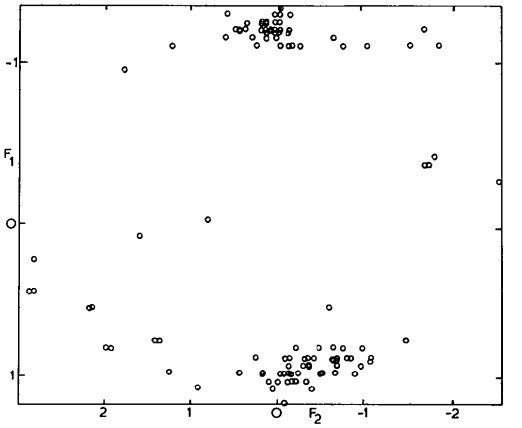


Fig. 3. Scattergram (for all 110 fragments) of the scores on the first two factors. The scales give a mean of zero and a standard deviation of unity for each factor.

$$\mathbf{S}_{np} = \mathbf{Z}_{nm}\mathbf{F}_{mp}. \quad (6)$$

Since the factors include weighting by the eigenvalues, the scores have unit variance (and zero mean). The scores are the best viewpoint for the data in the sense that no other p axes will show as much of the variance of the original m variables. A two-dimensional scattergram of S_1 and S_2 (scores of the n cases on F_1 and F_2) is shown in Fig. 3, which will show if there is any relationship between F_1 and F_2. Further progress is thus dependent on reification of this plot.

Theoretical studies have shown a five-membered ring to have two degrees of freedom for out-of-plane deformation, usually represented by an amplitude ($\tau_{\max}$) and a phase angle (P) (Kilpatrick, Pitzer & Spitzer, 1947). This concept of pseudorotation has been applied to nucleosides by Altona & Sundaralingam (1972) and a scatterplot (in polar coordinates) of $\tau_{\max}$ against P (Fig. 4) is very similar to Fig. 3. Despite their near identity, however, the plots differ importantly in their fundamental construction. The pseudorotational formula is derived for *small* out-of-plane displacements of a *symmetrical* pentagon and is not strictly applicable to nucleosides (although the errors will be small). Fig. 3 is a purely phenomenological description of the data unrelated to chemical theory. In the present case the two methods yield almost identical views of the data and reification of F_1 and F_2 is easily possible in terms of pseudorotation. A smooth pseudorotational pathway (corresponding to constant $\tau_{\max}$) has been superimposed on Fig. 4 and it can be seen that almost all points fall close to it. This further reduction of dimensionality can only be made by reification and is directly related to the pseudorotational concept. There are two points which deviate considerably from the pathway (one corresponds to a virtually flat ribose ring) and the significance of these outliers is discussed later.

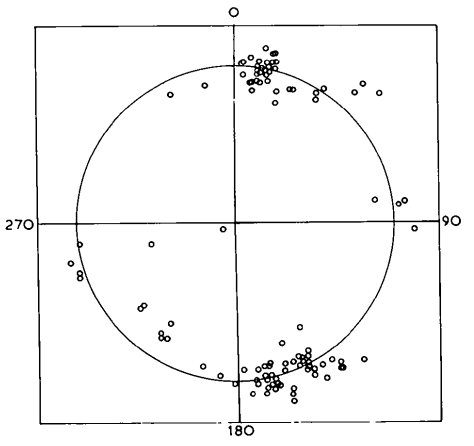


Fig. 4. Pseudorotational plot for all 110 fragments. The radial coordinate is τ_{max} (°), the angular coordinate is P (°), where $\tau_j = \tau_{max} \cos(P + 2\pi j/5)$ and $\tan P = (\tau_2 + \tau_4 - \tau_1 - \tau_3)/2\tau_0(\sin \pi/5 + \sin 2\pi/5)$ (Altona & Sundaralingam, 1972). Note the similarity to Fig. 3, except for changes of scale. The variables are computed by the procedure in Table 5. The circular pathway corresponds to $\tau_{max} \sim 39°$. Note the subsidiary pathway from $P = 90°$ to $P = 270°$ through the origin.

The reification of F_1 and F_2 allows us to make an important conceptual step and replace them by the pseudorotational parameters τ_{max} and P. Since τ_{max} is almost constant we use P alone to describe the ring conformation, excluding the two fragments with highly deviant values of τ_{max}. Similarly we shall replace F_3 with the side-chain conformation, represented by the torsion angle C(3′)–C(4′)–C(5′)–O(5′) (ψ). The analysis may seem to have been slightly circular but it will not always be as easy to replace factors with known variables. In some cases the factors will represent unknown effects and may be used to define new chemical variables. Complicated molecular deformations may be revealed by factor analysis for which no chemical terminology exists at present.

Derived-data screen

We have reduced the ring conformation to a single variable P except for two structures. Since we shall use P as the sole measure of ring conformation in future analysis, these two outliers must be excluded by a *derived-data screen*. The value of τ_{max} is calculated, and cases are rejected if this quantity is too low (Table 5). Since τ_{max} is defined only for the pseudorotational concept it is a derived parameter, as is P. [It would be similarly possible to select particular ring conformations on this basis, *e.g.* C(3′)-*endo* has $0° < P < 36°$.] In all the subsequent analysis the derived-data screen has been applied.

Table 5. *Derived data and derived-data screens*

PHASE is the phase of pseudorotation and TAUMAX is the puckering amplitude (Altona & Sundaralingam, 1972).

```
COMPUTE  DTOR = 3·14159/180
COMPUTE  NUM = T03 + T05 – T02 – T04
COMPUTE  DEN = 2*T01*(SIN(3·14159/5)
               + SIN(3·14159/2·5))
IF       (DEN LT ·0001 AND GT – ·0001) DEN = ·0001
COMPUTE  PHASE = ATAN(NUM/DEN)*180/3·14159
IF       (DEN LT 0) PHASE = PHASE + 180
IF       (PHASE GT 360)PHASE = PHASE – 360
IF       (PHASE LT 0)PHASE = PHASE + 360
COMPUTE  TAUMAX = 0·2*(T01/COS(PHASE*DTOR) +
         T02/COS((PHASE + 72)*DTOR +
         T03/COS((PHASE + 144)*DTOR) +
         T04/COS((PHASE + 216)*DTOR) +
         T05/COS((PHASE + 288)*DTOR))
SELECT IF (TAUMAX GT 25)
```

Correlation of ring and side-chain conformations

Any dependence of side-chain conformation on ring conformation will be of use in predicting the shapes of polynucleotides. Since factors are orthogonal, F_1 and F_3 show no correlation and, similarly, P and ψ show no covariance (Fig. 5). The distribution has approximately five modes and within any of these there is no local covariance of ψ and P. As far as the β-amino-furanoside group (I) alone is concerned, the conformations of side chain and ring must be regarded as independent. [Altona & Sundaralingam (1972) have suggested that when the base type is taken into account there is a correlation between ψ and P.]

Analysis of sub-population

It is easy to repeat the analyses of ring and side-chain conformation using the three subsets of the data corresponding to ribose, arabinose and 2′-deoxyribose

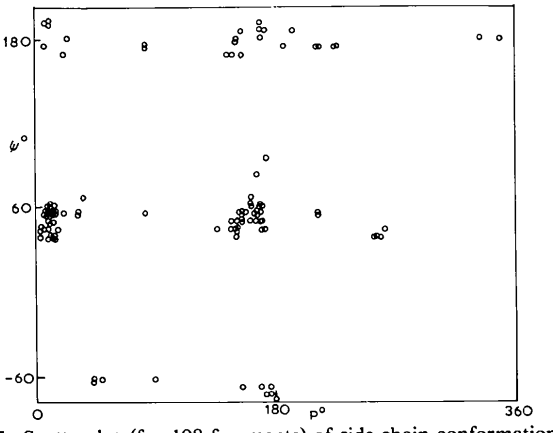


Fig. 5. Scatterplot (for 108 fragments) of side-chain conformation (ψ) against angle of pseudorotation (P).

derivatives. This was done for the ring conformation showing that the pseudorotational pathway itself (*i.e.* the degree of ring pucker) is independent of chemical type, but the position on the pathway (*i.e.* the pseudorotational angle) does show some correlation. The results will be discussed in a chemical context elsewhere.

Variability of bond angles in the furanose ring

Application of factor analysis to bond-angle variation is much less satisfactory than for torsion angles. The question of how many factors are significant is difficult to answer, and reification is almost impossible. There are two reasons for this. Firstly, much of the variance in the data is due to experimental error. Secondly, the variation will be seen later to be due in considerable part to unique variance in particular angles caused by specific chemical substitution patterns.

It is to be expected (Dunitz, 1972) that ring bond angles will change during pseudorotation and this is a useful model to investigate. Moreover, any outliers (whether caused by experimental errors or chemical forces) may become visible. Accordingly, scattergrams (input shown in Table 6) were plotted of ring angles with phase of pseudorotation; three of these are shown in Fig. 6. It is clear that despite the vertical scatter of points considerable variation of the C(4′)–O(1′)–C(1′) angle occurs during pseudorotation but the exact form of the function is not clear. For a symmetrical ring a cosine function of period 180° (in P) is expected with minima when the ring is in the O(1′)-*endo* ($P = 90°$) or O(1′)-*exo* ($P = 270°$) conformations. The plots for O(1′)–C(4′)–C(3′), C(2′)–C(3′)–C(4′), C(1′)–C(2′)–C(3′) (not shown) and O(1′)–C(1′)–C(2′) (not shown) are less clear; although it is certain that the angles change by about 2–3°, the form of the variation is not revealed. There is a scatter of points of about 3–5° in the vertical direction which obscures much of the covariance of bond angle with P. This effect is most marked in the scattergram for the angle C(2′)–C(3′)–C(4′) where any structural variance is almost totally obscured by errors and unique variance.

Selection of accurate data is the obvious way to improve these plots and a raw-data screen (Table 6)

Table 6. *Production of scattergrams of angles with pseudorotational phase*

The first SCATTERGRAM CARD produces 12 plots for all the data; the second includes only accurate structures; the third only accurate ribose structures.

VAR LABELS	AS ACCURACY FLAG FOR AVERAGE E.S.D.
SCATTERGRAM	A01 TO A07, A10 TO A14 WITH PHASE (0,360)
OPTIONS	4,7
*SELECT IF	(AS EQ 1 OR 2 OR RFACTOR LE 0·06)
SCATTERGRAM	A01 TO A07, A10 TO A14 WITH PHASE (0,360)
OPTIONS	4,7
*SELECT IF	(TYPE EQ 'RIBO' AND AS EQ 1 OR 2 OR RFACTOR LE 0·06)
SCATTERGRAM	A08, A09 WITH PHASE (0,360)

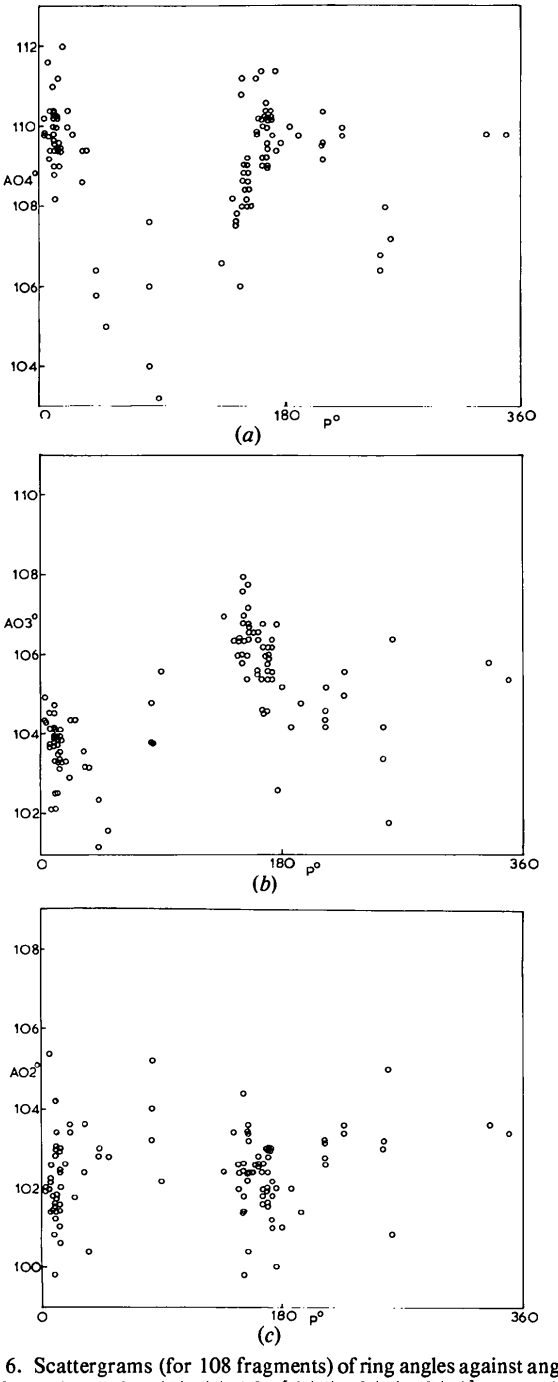


Fig. 6. Scattergrams (for 108 fragments) of ring angles against angle of pseudorotation (P). (*a*) A04 [C(1′)–O(1′)–C(4′)], one point omitted, (*b*) A03 [C(3′)–C(4′)–O(1′)], (*c*) A02 [C(2′)–C(3′)–C(4′)].

was used to select structures with R below 6% or bond-length e.s.d.'s <0·01 Å. The scattergrams were repeated (Fig. 7) and show considerably smaller variance, confirming the cosine nature of the functions for C(4′)–O(1′)–C(1′) and C(3′)–C(4′)–O(1′). Similar effects were observed for C(1′)–C(2′)–C(3′) and O(1′)–C(1′)–C(2′) (neither shown). In contrast

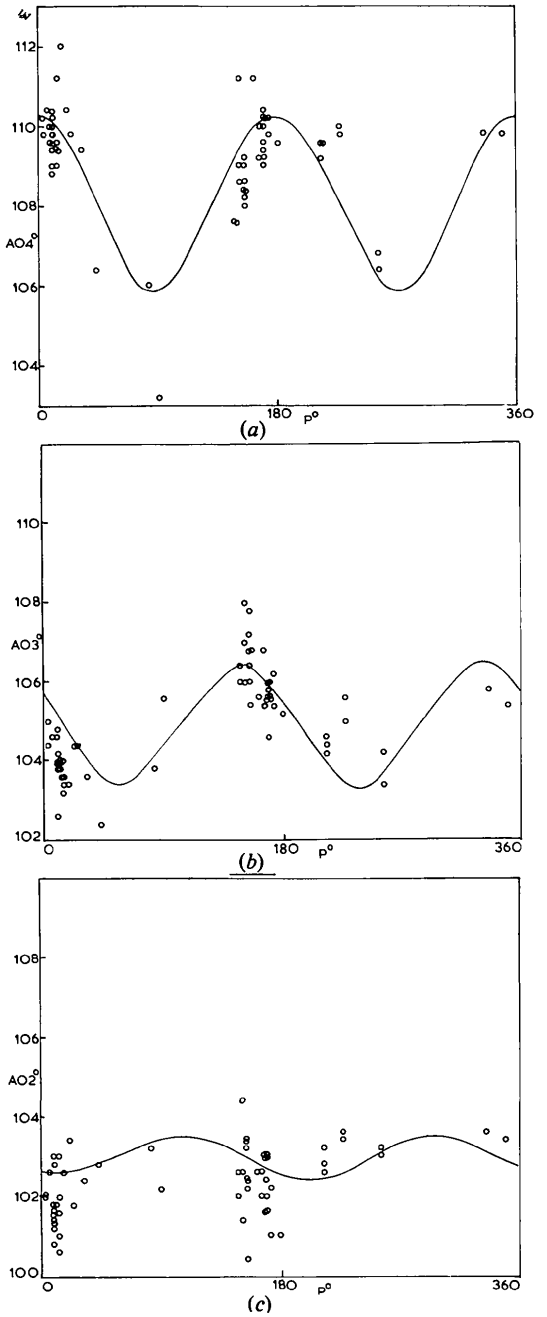


Fig. 7. Scattergrams (for 60 accurate data) of ring angles against angle of pseudorotation (P). (*a*) A04 [C(1′)–O(1′)–C(4′)], (*b*) A03 [C(3′)–C(4′)–O(1′)], (*c*) A02 [C(2′)–C(3′)–C(4′)]. The cosine curves obtained by regression (Table 7) have been added.

there is almost no significant variation in C(2′)–C(3′)–C(4′) even when accurate data are selected (Fig. 7*c*). The improvement in definition allows regression on to cosine curves (Table 7). There is little evidence that much of the scatter is caused by the chemical differences between the three sugars, and it is much more likely to be due mainly to experimental error and unique variance. None of the plots had serious outliers, the greatest deviation from any curve being about 3°.

The analysis of ring angles has shown interesting variations which could not have been reliably obtained by looking at structures in isolation. This synoptic view reveals that ring angles show the variations in Table 7. It is easier to distort the angle at O than those at the C atoms but the particular stiffness of C(2′)–C(3′)–C(4′) is not easily explainable in chemical terms. [We shall see later that the exocyclic angles at C(3′) are much more flexible.]

Variability of exocyclic bond angles

Of the remaining nine bond angles, eight are centred on the C atoms in the ring, and might be expected to vary with pseudorotation. The function will have a period of 360° and may be irregular in form; no attempt has been made to determine it. A further problem is that bond angles at C(2′) are likely to differ considerably between ribose and arabinose, requiring selection when C(1′)–C(2′)–O(2′) and C(3′)–C(2′)–O(2′) are analysed. Because of the improved results for ring angles accurate data were used throughout (Table 6).

Scattergrams are shown in Fig. 8 and the problem of outliers is seen to be important. There is little systematic variation in the N–C(1′)–C(2′) angle (Fig. 8*a*), except for four points about 13° below the curve.

Table 7. *Regression of bond angles on pseudorotational phase*

A model for ring bond-angle variation during pseudorotation is assumed: $A_i = \bar{A}_i + B_i \cos(2P + 72i)$, where $i = 0$ for C(4′)–O(1′)–C(1′), 1 for C(3′)–C(4′)–O(1′), *etc.* Regression analysis [*SPSS* input is shown in (*a*)] gives values for $\bar{A}_i$ and B_i with standard deviations (σ) and the correlation coefficient (r).

(*a*)
```
COMPUTE P1 = COS(2*PHASE*DTOR – 0·8*3·14159)
    :                                   :
COMPUTE P5 = COS(2*PHASE*DTOR – 0·4*3·14159)
REGRESSION VARIABLES = A01 TO A05, P01 TO P05/
        REGRESSION = A01 WITH P1/
            :                   :
        REGRESSION = A05 WITH P5/
```

(*b*)

	Variable	$\bar{A}$	$\sigma(\bar{A})$	B	$\sigma(B)$	r
A01	C(1′)–C(2′)–C(3′)	102·5	0·2	1·5	0·2	0·68
A02	C(2′)–C(3′)–C(4′)	102·8	0·15	0·6	0·2	0·34
A03	C(3′)–C(4′)–O(1′)	104·9	0·1	1·5	0·2	0·76
A04	C(4′)–O(1′)–C(1′)	108·1	0·2	2·1	0·2	0·74
A05	O(1′)–C(1′)–C(2′)	106·1	0·1	1·6	0·2	0·76

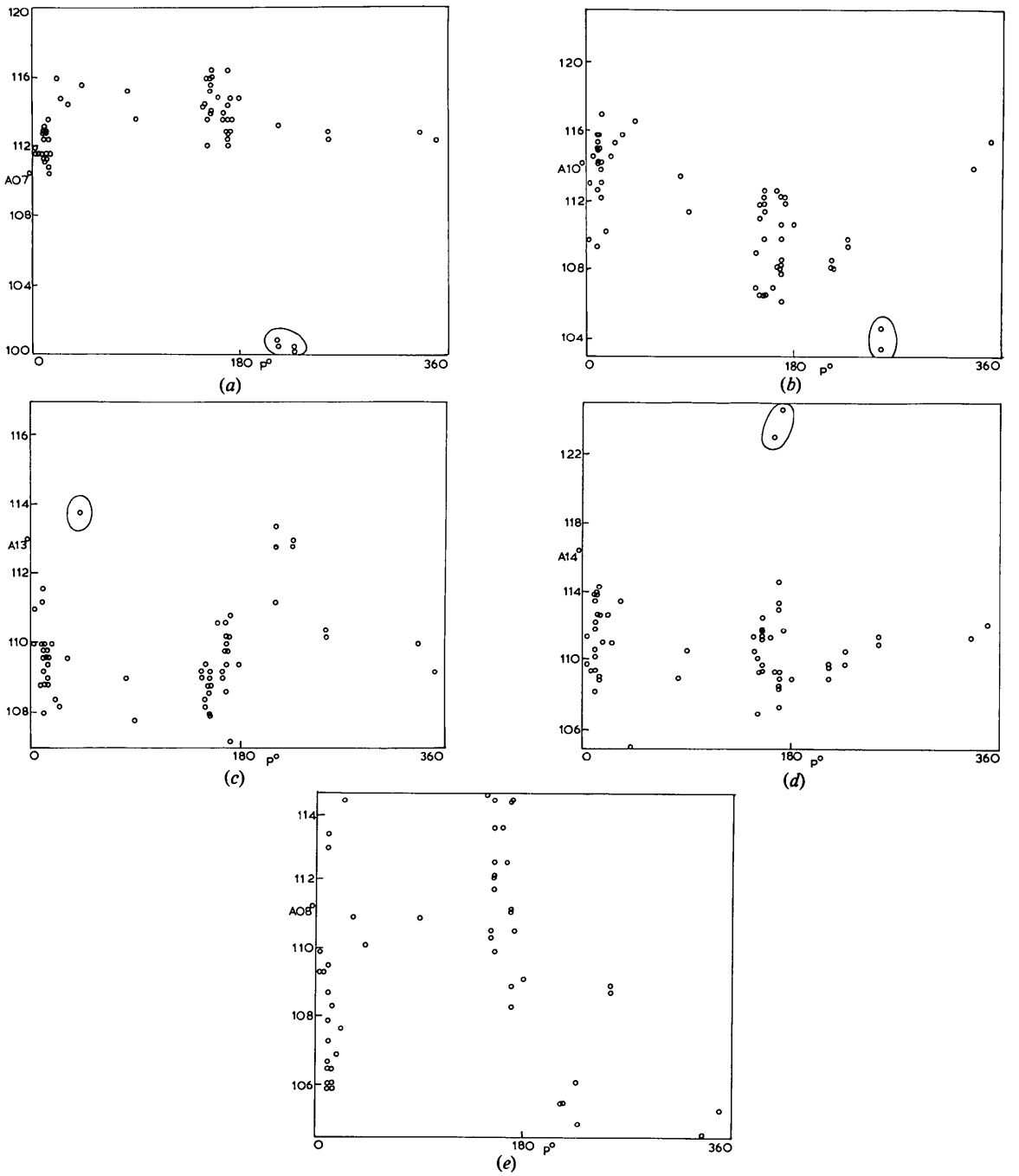
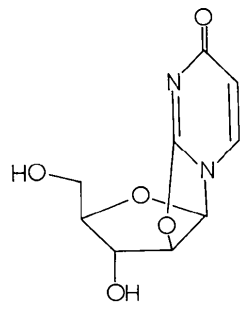


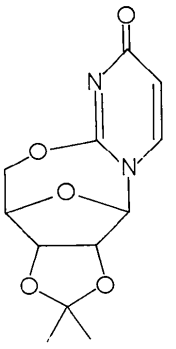
Fig. 8. Scattergrams of exocyclic bond angles with pseudorotation. (*a*) A07 [N–C(1′)–C(2′)] (60 data), (*b*) A10 [C(2′)–C(3′)–O(3′)] (60 data), (*c*) A13 [O(1′)–C(4′)–C(5′)] (60 data), (*d*) A14 [C(4′)–C(5′)–O(5′)] (60 data), (*e*) A08 [C(1′)–C(2′)–O(2′)] (51 data from ribose only). Outliers are circled and discussed in the text.

These are independent measurements for (IV) (AHARFU, CYURID),* with an arabinose ring cyclized to uridine. The five-membered ring is clearly responsible for the dramatic decrease in bond angle. Similar effects due to cyclization occur in the C(3′)–C(2′)–O(2′) angle (not shown), observed in the cyclic 2′,3′-isopropylidene cyclic uracil (V) (AIPCUR01, AIPCUR10); in the C(2′)–C(3′)–O(3′) angle (Fig. 8*b*) also observed in (V); and in the O(1′)–C(4′)–C(5′) angle (Fig. 8*c*) observed in the cyclic 3′,5′-phosphate

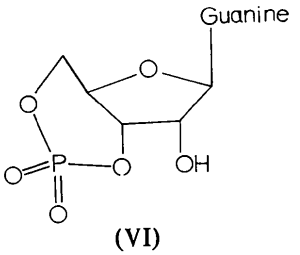
* The Cambridge reference codes, see Table 8.



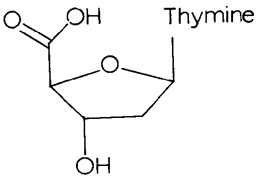
(IV)



(V)



(VI)



(VII)

(VI) (SCGMPT10). [Two outliers were also found in the scattergrams (Fig. 8*d*) for the C(4′)–C(5′)–O(5′) angle but these were due to a 5′-carboxylic acid (VII) (TYMCXA) which should have been excluded by the connectivity search. It is instructive to note that statistical analysis can thus serve in some cases as a final check on the accuracy of searches.]

Several points are apparent from the scattergrams. Regardless of which angle is involved, cyclization can cause changes of over 10°. This is a well-known effect but it serves to emphasize the flexibility of bond angles, often overlooked in model-building. Even without cyclization there is a variation of several degrees in most angles, often in a systematic manner. The exocyclic bond angles at C(2′) and C(3′) change by about 5° during pseudorotation, being largest in the *endo* conformation involving the particular C atom. The bond is then pseudo-equatorial and it would appear that torsional interactions can be reduced by this distortion. A similar effect makes the angle smaller in the *exo* conformation. A cosine curve of appropriate phase is not inconsistent with any of the scattergrams, although it is sometimes poorly defined, but since there is no theoretical basis regression has been omitted. For angles at C(4′) a scatter larger than any experimental error is observed, probably due to steric effects involving the phosphate group; this again emphasizes the flexibility of bond angles. The angle at C(5′) shows a similar scatter and also the decrease caused by cyclization in (VI), but it is uncorrelated with pseudorotation (as might be expected from the independence of ψ and P).

Discussion of results

There have been many studies of nucleoside (and particularly nucleotide) geometry, often concerned with relating the conformation of monomers to polynucleotides. This is not the primary aim of the present work which rather seeks to investigate the variability in geometry of the fragment (I). For this reason no attention has been paid to the types of substituent at 5′, 3′ and 1′ (especially the distinction between purines and pyrimidines). We have also included all recorded compounds containing (I), some of which [*e.g.* (IV)–(VI)] are not relevant to polynucleotide conformation. Nevertheless, there is some overlap with other studies, especially that of Altona & Sundaralingam (1972) who used crystal structure data to map ring and side-chain conformations of nucleosides and nucleotides, and suggested the idea of the rigid nucleotide. Their data set showed a sharply bimodal distribution of the pseudorotation parameter P and they concluded that configurations other than C(2′)-*endo* and C(3′)-*endo* are energetically unfavourable. Our study uses a larger data set, partly because of those structures carried out in the meantime and partly because of a greater chemical variety of compounds containing (I). It is clear that a large part of the pseudorotational pathway is accessible and this must shed doubt on the rigidity of the nucleotide. It is still clear from Fig. 4 that the C(3′)-*endo* and C(2′)-*endo* conformations (P = 18 and 162°) are the most favoured but that distortion from these is not energetically too difficult.

A study of the glycosidic bond length [C(1′)–N] in nucleosides and nucleotides (Lo, Shefter & Cochran, 1975) is relevant to our findings. The idea of deriving reaction coordinates from crystal structures (Bürgi, Dunitz & Shefter, 1973) suggested a possible correlation of the C(1′)–N length with the ring conformation and with the torsion angle about the C(1′)–N bond (χ). Data from 36 fragments with e.s.d.'s in bond lengths <0·012 Å showed a variation of about 0·05 Å in r[C(1′)–N], which correlated well with both conformational parameters. Though the form of the dependence was not explicit, periodic functions similar to those we have derived for bond angles seem appropriate.

Since bond-length variations will cause smaller changes in atomic positions than variations in either bond or torsion angles, the relationships in Figs. 3–8 and the information in Tables 3, 4 and 7 constitute a fairly complete description of the geometry of (I). There are two degrees of freedom, P and ψ, which account for most of the variation in geometry not due to experimental errors. For given values of P and ψ the torsion and bond angles can be calculated with e.s.d.'s.

What are the uses of such a model? Firstly, the reduction in dimensionality can be of value in the interpretation of electron density maps of nucleic acids or

enzyme-bound nucleotides. In particular the variation of conformation can be accurately explored rather than simply fitting a rigid C(3′)-*endo* configuration to the electron density.

Secondly, we have information, such as Fig. 4, about the intrinsic deformability of the molecule. The idea that crystal structures can map dynamic processes of molecules has been fruitful in a number of fields (*e.g.* Bürgi, Dunitz & Shefter, 1973; Bürgi, 1973; Murray-Rust, Bürgi & Dunitz, 1975; Dunitz, 1977; and, most relevantly, Lo, Shefter & Cochran, 1975). The ring torsion angles define two pathways: the pseudorotation already discussed and a ring flapping process involving conformations with $P = 90$ or 270°. This second pathway shows the value of inspecting outliers carefully because it involves the two points in the centre of Fig. 4. (A manual check was made on the original literature for these structures, the only time that a reference was made to coordinates in a journal during the study. There was no evidence for errors, abnormal thermal motion or other crystallographic effects and the structures must be considered to be as reliable as any of the other 108 data.) These two points taken with the four points at $P = 270°$ and four at $P = 90°$ hint at an alternative pathway for ring deformation. Most of the 10 compounds contain an extra ring connecting C(2′) and C(3′) and are thereby prevented from pseudorotation. Any deforming forces appear to be relieved by a flattening of the ring [O(1′)-*endo* ⇄ flat ring ⇄ O(1′)-*exo*]. Additional data will be needed to confirm the exact pathway but it seems clear that a flat ribose ring can be produced by crystal packing forces.

The variation of bond angles with pseudorotation is encouragingly well defined. Mechanical models often lead to the mistaken idea that angle deformation involves much energy but it is clear that most bond angles at saturated C vary by up to 5° during pseudorotation, a process involving little energy change and small barriers. The form of the ring-angle variation has a theoretical geometrical basis (Dunitz, 1972) but the substituent movements must be determined by chemical forces, presumably torsional and non-bonded interactions. Variation at C(2′) and C(3′) can be interpreted (in ribose) by consideration of the O(2′)–C(2′)–C(3′)–O(3′) torsion angle. For tetrahedral C atoms it would be expected to be identical to C(1′)–C(2′)–C(3′)–C(4′) and this angle varies from −40 to 40°. The bond-angle variation at C(2′) and C(3′) results in O(2′)–C(2′)–C(3′)–O(3′) having a range greater by about 10°, suggesting that torsional effects or non-bonded forces push O(2′) and O(3′) further apart. The variation at C(4′) and C(1′) is likely to be dependent on the size and conformation of the substituents at these positions.

A knowledge of the position of a fragment on a pathway would allow complete prediction of its geometry but this problem has not yet been investigated. Some simple rules can be drawn up on the basis of chemical formulae (*e.g.* cyclization at 2′ and 3′ gives $P = 90$ or 270°) but position on the pathway will often be totally due to crystal packing forces. The present data show that these are not unimportant. For example, most 2′-deoxy compounds have conformations in the range $P = 140$–210° but the disodium salt of 2′-deoxyguanosine 5′-phosphate (SDGUNP) has $P = 83°$. Even more striking is cytidine cyclic 2′,3′-phosphate

Table 8. *Bibliography*

The searches were carried out in June 1977 on the version of the BIB file which contained 16 776 entries. The 110 fragments in the *SPSS* system file are referenced by 90 independent publications to each of which is assigned a unique reference code (*e.g.* ADENOS10). In certain cases there have been independent studies of the same crystal (*e.g.* AIPCUR01 and AIPCUR10). This code, together with the shortened reference, uniquely defines each compound.

ABHPTB K.-I.ASAHI, CHEM.LETTERS, ,1197,1973.
ACADOS S.T.RAO, J.AMER.CHEM.SOC., 92,4963,1970.
ADENOS10 T.F.LAI, ACTA CRYST.(B), 28,1982,1972.
ADDSHC K.SHIKATA, ACTA CRYST.(B), 29,31,1973.
ADPOSD M.SUNDARALINGAM, ACTA CRYST., 21,495,1965.
ADPOSM J.KRAUT, ACTA CRYST., 16,79,1963.
ADURPO10 E.SHEFTER, ACTA CRYST.(B), 25,895,1969.
AGQPCD K.AOKI, ACTA CRYST.(B), 32,1454,1976.
AHARFU L.T.J.DELBAERE, ACTA CRYST.(B), 29,2905,1973.
AIPCUR01 P.C.MANOR, BIOCHIM.BIOPHYS.ACTA, 340,472,1974.
AIPCUR10 L.T.J.DELBAERE, ACTA CRYST.(B), 30,1241,1974.
ANIMPH01 K.AOKI, BULL.CHEM.SOC.JPN., 48,1260,1975.
APAPAD10 D.SUCK, ACTA CRYST.(B), 32,1727,1976.
ARADEN10 G.BUNICK, ACTA CRYST.(B), 30,1654,1974.
ARATUR10 W.SAENGER, J.AMER.CHEM.SOC., 94,621,1972.
ARBCYT10 P.TOUGARD, ACTA CRYST.(B), 30,86,1974.
ARFCYT10 J.S.SHERFINSKI, ACTA CRYST.(B), 29,192,1973.
ARFUAD A.K.CHWANG, ACTA CRYST.(B), 30,2273,1974.
ARFUAD01 T.HATA, BULL.CHEM.SOC.JPN., 47,2758,1974.
ASTHYM10 D.SAENGER, ACTA CRYST.(B), 27,2105,1971.
BEURID10 E.A.GREEN, ACTA CRYST.(B), 31,1702,1975.
BREDIN H.YOSHIOKA, TETRAHEDRON LETT., ,4031,1975.
BROXUR10 J.IBALL, PROC.R.SOC.,A, 295,320,1966.
BRURID10 J.IBALL, PROC.R.SOC.,A, 295,320,1966.
CLDOUR D.W.YOUNG, ACTA CRYST.(B), 29,1259,1973.
CLPURB H.STERNGLANZ, ACTA CRYST.(B), 31,2888,1975.
CLURID10 J.W.HAWKINSON, ACTA CRYST.(B), 27,34,1971.
CYTCYP20 C.L.COULTER, J.AMER.CHEM.SOC., 95,570,1973.
CYTIAC M.SUNDARALINGAM, J.MOLEC.BIOL., 13,914,1965.
CYTIAC01 C.E.BUGG, J.MOLEC.BIOL., 25,67,1967.
CYTID110 S.FURBERG, ACTA CRYST., 18,313,1965.
CYURID D.SUCK, ACTA CRYST.(B), 29,1323,1973.
DADPNH10 B.S.REDDY, ACTA CRYST.(B), 31,19,1975.
DAZADN10 S.SINGH, J.AM.CHEM.SOC., 98,825,1976.
DHTHUR10 B.KOJIC-PRODIC, ACTA CRYST.(B), 30,1550,1974.
DHURID01 D.SUCK, ACTA CRYST.(B), 28,596,1972.
DMGUAN10 T.BRENNAN, J.AMER.CHEM.SOC., 94,8548,1972.
DOCYPO M.A.VISWAMITRA, J.AMER.CHEM.SOC., 93,4565,1971.
DOCYPD01 M.A.VISWAMITRA, J.AMER.CHEM.SOC., 93,4565,1971.
DOCYTC E.SUBRAMANIAN, ACTA CRYST.(B), 26,303,1970.
DOURID A.HAHN, ACTA CRYST.(B), 28,2260,1972.
DOXADM D.G.WATSON, ACTA CRYST., 19,111,1965.
DTURID G.H.-Y.LIN, ACTA CRYST.(B), 27,961,1971.
DTURID01 W.SAENGER, ACTA CRYST.(B), 27,1178,1971.
DXCYTD D.W.YOUNG, ACTA CRYST.(B), 31,961,1975.
ERFIMP A.H.-J.WANG, J.AMER.CHEM.SOC., 96,1205,1974.
ESMENM N.NAGASHIMA, ACTA CRYST.(B), 30,1094,1974.
FDOURD R.A.HARRIS, BIOPHYS.J., 4,203,1964.
GUANPH W.MURAYAMA, ACTA CRYST.(B), 25,2236,1969.
GUANSH10 U.THEWALT, ACTA CRYST.(B), 26,1089,1970.
GUOSBH P.TOUGARD, ACTA CRYST.(B), 30,214,1974.
HDTURD10 B.KOJIC-PRODIC, ACTA CRYST.(B), 32,1090,1976.
HICYTM B.KOJIC-PRODIC, ACTA CRYST.(B), 32,1103,1976.
HMETHD THEWALT, ACTA CRYST.(B), 29,1393,1973.
IDOXUR N.CAMERMAN, ACTA CRYST., 18,203,1965.
INOSIN10 A.R.I.MUNNS, ACTA CRYST.(B), 26,1101,1970.
INOSND01 A.R.I.MUNNS, ACTA CRYST.(B), 26,1114,1970.
INOSND10 U.THEWALT, ACTA CRYST.(B), 26,1089,1970.
IURIDN10 A.RAHMAN, ACTA CRYST.(B), 26,1765,1970.
MADENS10 P.PRUSINER, ACTA CRYST.(B), 32,161,1976.
MARAFC G.I.BIRNBAUM, J.AM.CHEM.SOC., 97,5904,1975.
MCYTNS10 E.SHEFTER, CRYST.STRUCT.COMM., 3,209,1974.
MEURID D.A.HUNT, ACTA CRYST.(B), 25,2144,1969.
MEYRID D.SUCK, J.AMER.CHEM.SOC., 94,6520,1972.
MRFPUR T.TAKEDA, ACTA CRYST.(B), 31,1202,1975.
NAINPH10 S.T.RAO, J.AMER.CHEM.SOC., 91,1210,1969.
NEBULR T.TAKEDA, ACTA CRYST.(B), 30,825,1974.
RPPYPY20 P.NARAYANAN, CARBOHYDR.RES., 44,169,1975.
SALCYS C.TAMURA, CHEM.LETTERS, ,1221,1973.
SCGMPT10 A.K.CHWANG, ACTA CRYST.(B), 30,1233,1974.
SDGUNP D.W.YOUNG, ACTA CRYST.(B), 30,2012,1974.
SDGUNP01 T.P.SESHADRI, PRAMANA, 3,218,1974.
SURIDP M.A.VISWAMITRA, ACTA CRYST.(B), 28,1108,1972.
TCYTDH G.H.-Y.LIN, J.AMER.CHEM.SOC., 93,1235,1971.
TEAURP10 C.L.COULTER, ACTA CRYST.(B), 25,2055,1969.
TGUANS10 U.THEWALT, J.AMER.CHEM.SOC., 94,8892,1972.
THIRDS10 W.SAENGER, J.MOLEC.BIOL., 50,153,1970.
THOPAD10 G.KENNARD, J.CHEM.SOC.(B), ,1940,1971.
THPRIB E.SHEFTER, J.PHARM.SCI., 57,1157,1968.
THPYUR P.H.STOTHART, ACTA CRYST.(B), 29,2237,1973.
THYDIN D.W.YOUNG, ACTA CRYST.(B), 25,1423,1969.
TRFBIM P.PRUSINER, ACTA CRYST.(B), 29,2328,1973.
TYMCXA D.SUCK, BIOCHIM.BIOPHYS.ACTA, 361,1,1974.
URARAF01 J.S.SHERFINSKI, ACTA CRYST.(B), 30,873,1974.
URARAF10 P.TOLLIN, ACTA CRYST.(B), 29,1641,1973.
URIDPS10 W.SAENGER, J.AMER.CHEM.SOC., 92,4712,1970.
UROAME K.MORIKAWA, ACTA CRYST.(B), 31,1004,1975.
VIRAZL P.PRUSINER, ACTA CRYST.(B), 32,419,1976.
VIRAZL01 P.PRUSINER, ACTA CRYST.(B), 32,419,1976.
XANTOS G.KOYAMA, ACTA CRYST.(B), 32,969,1976.

(CYTCYP20) which has two molecules in the asymmetric unit. One has a conformation on the pseudorotational pathway at $P = 91°$; the other has the planar ribose ring already discussed. These conformations may differ in energy by up to 5 kcal mol^{-1} which can only be provided by differences in the local packing forces of the two fragments. Only by relation of the crystal environment of each fragment to P and ψ will complete prediction of molecular geometry become possible.

Conclusion

The approach outlined in these three papers is applicable to any molecular fragment provided that enough examples are available on file. [Fragments in totally inorganic structures will have to be retrieved manually; see, for example, Murray-Rust, Bürgi & Dunitz (1978).] When a fragment can show symmetry [for example a phosphate group may show T_d symmetry (or any subgroup of T_d)] the factor analysis involves group-theoretical considerations (to be published later). It is important that the total data file of retrieved fragments is made clear (see Table 8) so that the analysis can be independently verified. Equally important is the publication of the algorithms for screening (particularly on derived data) so that results are seen not to be due to biased rejection of cases.

The present study was accomplished over a few weeks and future analyses will be considerably quicker. Compared with manual methods, which would require several months, the computer retrieval and analysis can be seen as a new tool in crystallography and structural chemistry. As experience is gained and more software is written (particularly for examining crystal environments) major increases in understanding crystal structures and the geometry of molecules will be possible.

References

ALTONA, C. & SUNDARALINGAM, M. (1972). *J. Am. Chem. Soc.* **94**, 8205–8212.

BÜRGI, H. B. (1973). *Inorg. Chem.* **12**, 2321–2327.

BÜRGI, H. B., DUNITZ, J. D. & SHEFTER, E. (1973). *J. Am. Chem. Soc.* **95**, 5065–5066.

DUNITZ, J. D. (1972). *Tetrahedron*, **28**, 5459–5467.

DUNITZ, J. D. (1977). Chemical Society Centenary Lecture.

HARMAN, H. H. (1967). *Modern Factor Analysis*. Univ. of Chicago Press.

JACK, A., KLUG, A. & LADNER, J. E. (1976). *Nature (London)*, **261**, 250–251.

KILPATRICK, J. E., PITZER, K. S. & SPITZER, R. (1947). *J. Am. Chem. Soc.* **69**, 2483–2488.

LO, A., SHEFTER, E. & COCHRAN, T. G. (1975). *J. Pharm. Sci.* **64**, 1707–1710.

MURRAY-RUST, P. & BLAND, R. (1978). *Acta Cryst.* B**34**, 2527–2533.

MURRAY-RUST, P., BÜRGI, H. B. & DUNITZ, J. D. (1975). *J. Am. Chem. Soc.* **97**, 921–922.

MURRAY-RUST, P., BÜRGI, H. B. & DUNITZ, J. D. (1978). *Acta Cryst.* B**34**, 1793–1803.

MURRAY-RUST, P. & MOTHERWELL, S. (1978). *Acta Cryst.* B**34**, 2518–2526.

NIE, N. H., HULL, C. H., JENKINS, J. G., STEINBRENNER, K. & BENT, D. H. (1975). *Statistical Package for the Social Sciences*, 2nd edition. New York, London: McGraw-Hill.

Acta Cryst. (1978). B**34**, 2546–2551

The Camphoroxime System. I. An X-ray Study of (–)-Camphoroxime (m.p. 118°C)

BY F. BAERT AND R. FOURET

Laboratoire de Physique des Solides, Equipe de Dynamique des Cristaux Moléculaires associée au CNRS (N° 465), Université des Sciences et Techniques de Lille I, *Villeneuve d'Ascq* 59650, *BP* 36, *France*

(*Received* 26 *January* 1978; *accepted* 21 *March* 1978)

(–)-Camphoroxime (m.p. 118°C) is monoclinic, $P2_1$, with $a = 12{\cdot}06$ (2), $b = 11{\cdot}81$ (2), $c = 7{\cdot}16$ (2) Å, $\beta = 99{\cdot}81°$, $Z = 4$. The structure was determined from 1491 independent intensities and refined to $R = 0{\cdot}052$. Except for one intramolecular distance, C(6)—C(5), there are no significant differences between the conformations of the two molecules of the asymmetric unit, which are linked by hydrogen bonds.

Introduction

This work is part of a program concerned with the study of enantiomorphic compounds, the phase diagrams of which seem to predict the existence of solid solutions. Such solid solutions are characterized by equilibrium diagrams of three types, named Roozeboom I, II and III. Examples (Baert & Fouret,