tion, there are two other short contacts, namely N...O(1) ($II_{0\bar{1}1}$), 3·08$_4$ Å and N...O(1), 2·75$_2$ Å. The atom O(1) belongs to the same molecule while the atom O(1) ($II_{0\bar{1}1}$) lies approximately on the line C(2)–N. The arrangement of this latter type is very similar to that of carbonyl oxygen [O(1'')] around the $-NH_3^+$ group in valine hydrochloride (Parthasarathy, 1966) or in γ-glycine (Iitaka, 1961).

The authors wish to express their gratitude to C. Itoh Electronic Computing Service Co. Ltd. for the use of a CDC 3600 computer. This work was supported by a research grant from the Ministry of Education.

References

BERGHUIS, J., HAANAPPEL, IJ. M., POTTERS, M., LOOPSTRA, B. O., MACGILLAVRY, C. H. & VEENENDAAL, A. L. (1955). *Acta Cryst.* **8**, 478.
BUSING, W. R. & LEVY, H. A. (1964). *Acta Cryst.* **17**, 142.
BUSING, W. R., MARTIN, K. O. & LEVY, H. A. (1962). *ORFLS*. USAEC Report ORNL-TM-305, Oak Ridge National Laboratory, Oak Ridge, Tennessee.
DAWSON, B. (1960). *Acta Cryst.* **13**, 403.
DONOHUE, J. & TRUEBLOOD, K. N. (1952). *Acta Cryst.* **5**, 414, 419.
FRIDRICHSONS, J. & MATHIESON. A. MCL. (1962). *Acta Cryst.* **15**, 569.
IITAKA, Y. (1961). *Acta Cryst.* **14**, 1.
JOHNSON, C. K. (1965). *ORTEP*. ORNL-3794. Oak Ridge National Laboratory, Oak Ridge, Tennessee.
KAYUSHINA, R. L. & VAINSHTEIN, B. K. (1965). *Kristallografiya*, **10**, 833.
KILPATRICK, J. E., PITZER, K. S. & SPITZER, R. S. (1947). *J. Amer. Chem. Soc.* **69**, 2483.
LEUNG, Y. C. & MARSH, R. E. (1958). *Acta Cryst.* **11**, 17.
MATHIESON, A. MCL. & WELSH, H. K. (1952). *Acta Cryst.* **5**, 599.
PARTHASARATHY, R. (1966). *Acta Cryst.* **21**, 422.
PITZER, K. S. (1945). *Science*, **101**, 672.
SASISEKHARAN, V. (1959). *Acta Cryst.* **12**, 941.
WILSON, A. J. C. (1942). *Nature, Lond.* **150**, 151.
WRIGHT, B. A. & COLE, P. A. (1949). *Acta Cryst.* **2**, 129.
ZUSSMAN, J. (1951). *Acta Cryst.* **4**, 493.

Acta Cryst. (1969). B**25**, 2192

Least-Squares Refinement of the Crystal and Molecular Structures of DNA and RNA from X-ray Data and Standard Bond Lengths and Angles

BY STRUTHER ARNOTT, S. D. DOVER* AND A. J. WONACOTT†

M.R.C. Biophysics Research Unit and King's College London Department of Biophysics, 26–29 Drury Lane, London W.C.2, England

(*Received* 16 *October* 1968)

A method is described for producing molecular models (with standard bond lengths and angles) that optimally fit the low resolution X-ray diffraction data characteristically obtained from crystalline fibres. The application of this linked-atom, least-squares method in the production of refined structures for DNA and RNA complementary helices is discussed. Atomic coordinates and molecular parameters are presented for these new standard polynucleotide models that not only have contemporary bonded stereochemistry but fit the X-ray intensities to a substantially better degree. The geometrical properties of these new models are analysed in detail and compared with one another and with the conformations found in monomers. In addition, the results of the refinements allow a decision to be made that the RNA helices are eleven- and not tenfold.

Introduction

The alkali metal salts of native, double-helical polynucleotides exist in a number of different, fully crystalline forms. In the *A* and *B* forms of deoxyribonucleic acid (DNA) the polyanions have very different conformations (also called *A* and *B*) while in the α and β forms of ribonucleic acid (RNA) the molecular conformations are probably the same. Diffraction analyses of all four structures have been published (Fuller, Wilson, Wilkins, Hamilton, & Arnott, 1965; Langridge, Marvin, Seeds, Wilson, Hooper, Wilkins & Hamilton, 1960; Arnott, Wilkins, Fuller & Langridge, 1967*a*, *b*; Arnott, Wilkins, Fuller, Venable & Langridge, 1967).

These crystalline polynucleotides do not occur as single crystals but as ordered arrays of microcrystals in fibres. The number of X-ray diffracted intensities obtainable from such systems is characteristically not more than a few hundreds, and few represent periodicities less than 2·5 Å. To define atomic positions in these molecules it is therefore necessary to supplement the diffraction data with additional stereochemical information, mainly in the form of the expected bond

* Present address: Laboratory of Molecular Biophysics, Department of Zoology, The University, Oxford, England.
† Present address: Department of Biological Sciences, Purdue University, Lafayette, Indiana, U.S.A.

lengths and angles of nucleoside and nucleotide structures. In all previous analyses the combination of diffraction and stereochemical data was achieved by building scale wire models, and adjusting them by systematic trial-and-error to obtain calculated diffraction intensities in good agreement with those observed.

We will discuss later our view that the manual molecular model-building approach is not without some advantages. Nevertheless it has some clear disadvantages.

(1) It is somewhat subjective and the relative weights given to competing diffraction and stereochemical requirements may vary from one structure determination to another. This makes it difficult adequately to compare the different structures, and for the process by which they were obtained to be reproduced in other laboratories.

(2) Because of the physical properties of the models themselves, it is difficult to impose on them precisely the bond length and angles one would wish, and therefore one is in doubt as to how far any particular property of the final model would survive the removal of the inevitable minor stereochemical anomalies.

(3) It is not easy to be sure that the best model obtained is indeed the best possible model, and this be comes particularly important when alternative structures have to be considered. Then, the only way of deciding whether the intensity data support only one of the possibilities is to compare the models of each species that *best* fit the diffraction.

Our successful application of the linked-atom method for refining fibre crystal structures from X-ray data and stereochemical constraints (Arnott & Wonacott, 1966*a*; Arnott, 1968) to a number of polypeptide structures (Arnott & Wonacott, 1966*b*; Arnott & Dover, 1967; Arnott, Dover & Elliott, 1967; Arnott & Dover, 1968) has prompted us to use this method for nucleic acid refinement and thereby to remedy what we consider the deficiencies of the manual model-building approach. We describe below the tactics we have adopted and our experiences during the refinements, and compare the models we obtain with those presented previously. In particular we examine how many of the nucleic acid geometrical features have survived our amelioration of the bonded stereochemistry and refinement of the variable parameters to fit the X-ray data. We discuss also the probable accuracy of the standardized nucleic acid models we have produced.

Molecular and crystal parameters

The structure factor for a crystal containing identical helical molecules is

$$F(hkl)=\sum_m \sum_n G(l\xi n)\exp in(\psi_m+\pi/2)\exp 2\pi i(hx_m+ky_m+lz_m) \quad (1)$$

where

$$G(l\xi n)=\sum_j f_j J_n(2\pi r_j\xi)\exp i(2\pi lZ_j/c-n\theta_j) \quad (2)$$

(ψ_m is related to the azimuthal orientation of the mth molecule, at (x_m, y_m, z_m), in the unit cell; (r_j, θ_j, Z_j) are the real space cylindrical polar coordinates of the jth atom in the repeating unit of the helix; J_n is the Bessel function of the first kind of order $n=(l-mN)/K$, where m is any integer; K and N are respectively the number of helix pitches and the number of helix residues in c).

For the DNA and RNA molecules we will consider here, there is a diad in each helical repeat perpendicular to the helix axis, *i.e.* for every atom at (r_j, θ_j, Z_j) there is an equivalent atom at $(r_j, -\theta_j, -Z_j)$. If we consider two diadically related nucleotides as the helical repeat group, $G(l\xi n)$ becomes

$$\sum_j 2f_j J_n(2\pi r_j\xi)\cos(2\pi lZ_j/c-n\theta_j)\,. \quad (3)$$

(1), (2), (3) help to explain why, in these analyses, it is convenient to segregate molecular and packing parameters and emphasize that for the former we need be concerned only with defining the atomic coordinates of a *molecular* asymmetric unit (in these cases one nucleotide).

In the analysis of the lower resolution X-ray data from crystalline fibres, where supplementation by additional information is required, it is more convenient to treat the atomic coordinates as implicit variables only, and to consider the parameters of the molecular asymmetric unit to be (*a*) bond lengths, (*b*) bond angles (*c*) dihedral angles of the nucleotide, (*d*) three angles to define the orientation of the nucleotide with respect to the helix axis, and (*e*) the distance of some reference point in the nucleotide from the helix axis. The standard bond lengths and angles we have chosen for the sugar-phosphate backbone are shown in Fig. 1(*a*). None of these values is likely to be in error by more than $\pm 0{\cdot}02$ Å and $\pm 3{\cdot}0°$ and most are probably good to $\pm 0{\cdot}1$ Å and $\pm 1{\cdot}5°$. The dihedral angles along the chain are shown in Fig. 1(*b*). In the convention used by us (see also Arnott & Wonacott, 1966*a*) a set of axes is assumed to exist at each chain atom after the first, such that the x-axis lies along the bond between the chain atom i and its predecessor $(i-1)$; the y axis, which is perpendicular to this axis, lies within the obtuse angle formed by atoms $(i-1)$, i, $(i+1)$ and is coplanar with them; the z axis completes a right-handed orthogonal set. The coordinates (X, Y, Z), with respect to the axial set at i, of any atoms can be transformed to refer to the axial set at $(i+1)$ when they become

$$\begin{bmatrix}X'\\Y'\\Z'\end{bmatrix} = \begin{bmatrix}-\cos\varphi_i & -\sin\varphi_i & 0\\ \sin\varphi_i\cos\tau_{i,i+1} & -\cos\varphi_i\cos\tau_{i,i+1} & \sin\tau_{i,i+1}\\ -\sin\varphi_i\sin\tau_{i,i+1} & \cos\varphi_i\sin\tau_{i,i+1} & \cos\tau_{i,i+1}\end{bmatrix}\begin{bmatrix}X\\Y\\Z\end{bmatrix} + \begin{bmatrix}l_{i,\,i+1}\\0\\0\end{bmatrix} \quad (4)$$

where φ_i is the chain angle at i, $l_{i,i+1}$ is the bond length

between i and $(i+1)$ and $\tau_{i,i+1}$, is the angle between the planes formed by the atom triplets $(i-1)$, i, $(i+1)$ and i, $(i+1)$, $(i+2)$. In this convention, therefore, $\tau=0$ represents a *cis* conformation; the direction of positive rotations as one moves down the chain is indicated by the arrows in Fig. 1(*b*). By successive transformations of this kind the positions with respect to the terminal axial set of all chain and pendant atoms can be defined as functions of each l, φ and τ.

For the molecular unit of DNA (RNA) we have chosen a chain [see Fig. 1(*b*)] of 11 atoms with 22 pendant atoms, 18 (17) of which are base atoms which are consid-

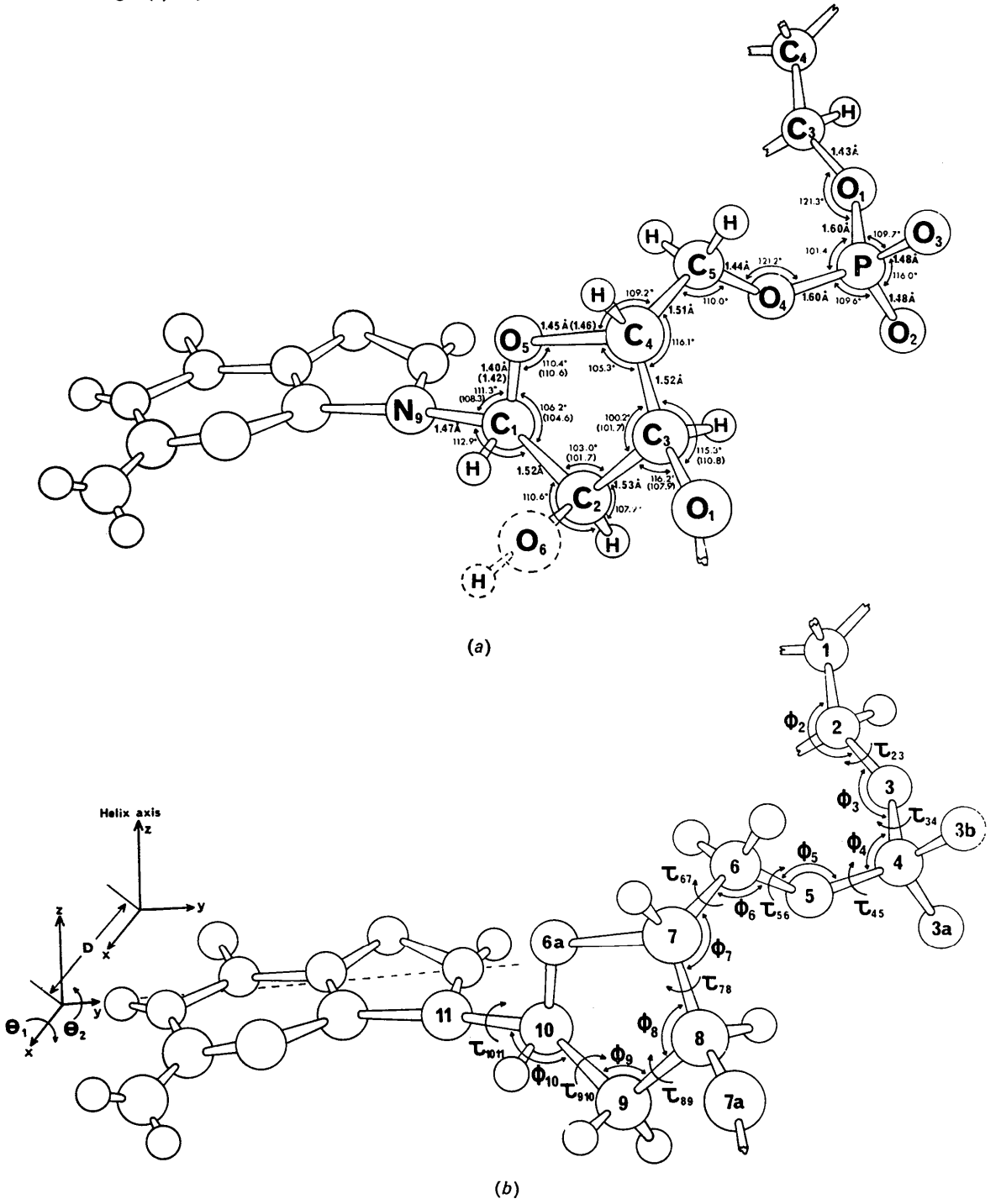


Fig. 1. (*a*) Values of the fixed stereochemical parameters for the C(3)-*endo* sugar conformation. (The bracketed values are for the C(2)-*endo* sugar conformation). (*b*) The nucleotide chain description.

ered pendant on the nitrogen atom N(11). Three corresponding atoms, 1,2,3 and 7, 8, 7*a*, from each of two successive nucleotides are included so that we can later conveniently define the relationships that will ensure that successive helix residues are linked appropriately. The pendant base atoms are the half weight C, N and O atoms of a substituted purine and pyrimidine group together with quarter-weight N and C atoms to represent the $-NH_2$ group of guanine and the methyl group of thymine respectively. In this way we take account of the fact that the Bragg reflexions from these polynucleotides correspond to the statistical diffracting structure that arises from the essentially random sequence of bases.

For the four further parameters required to relate the molecular asymmetric unit to the helix axis, we have chosen not the set of three angles and a distance that would directly relate the terminal set of axes on atom 11 to this axis, but an equivalent set that have become traditional in describing nucleic acid double-helix geometry. These are (see Fig. 2): θ_1 the angle by which the base planes are tilted about the diad axis relating the two glycosidic links in each base-pair; θ_2 the angle by which the base pairs are twisted about an axis perpendicular both to the helix axis and to the diad axis [we have assumed that this axis passes through pyrimidine C(4)]; θ_3 the angle whose variation would result in the bases 'swivelling' about the axis perpendicular to the twist and tilt axes at their intersection; D the distance of the bases from the helix axis, the reference distance being to the twist axis. Provided the rotations are carried out in the order θ_3, θ_2, θ_1 the diadic relationship between the glycosidic links in the base pair is maintained. For these pairs we have chosen standard dimensions that are shown in Fig. 3. It is noteworthy that, in spite of the rigorous demands of contemporary values of covalent and hydrogen bond lengths and covalent bond angles, the nucleic acid purines and pyrimidines lend themselves very readily to the construction of base-pairs with perfect diads relating the glycosidic links and with identical distances between these links, so that all the pairs can replace one another without distorting the regularity of the sugar-phosphate backbone. To preserve our chosen geometry for the base pairs, θ_3 was fixed at $-74{\cdot}5°$ in all our analyses. θ_2 was expected to be near zero but not fixed.

The disposition of the molecules in the various unit cells is shown in Fig. 4, where the variable packing parameters are indicated. In all the refinements we also included a parameter, K, to relate the scale of the observed intensities to those calculated, and a parameter, B, to ensure that the average fall-off of intensity with increasing diffraction angle is similar for the model and the observed data.

The strategy of constrained refinement

Indices of 'goodness of fit'

The aim of these refinements has been to obtain models yielding the best agreement between the calculated (F_m) and observed ($_oF_m$) structure amplitudes. Earlier nucleic acid analyses suggest that good agree-

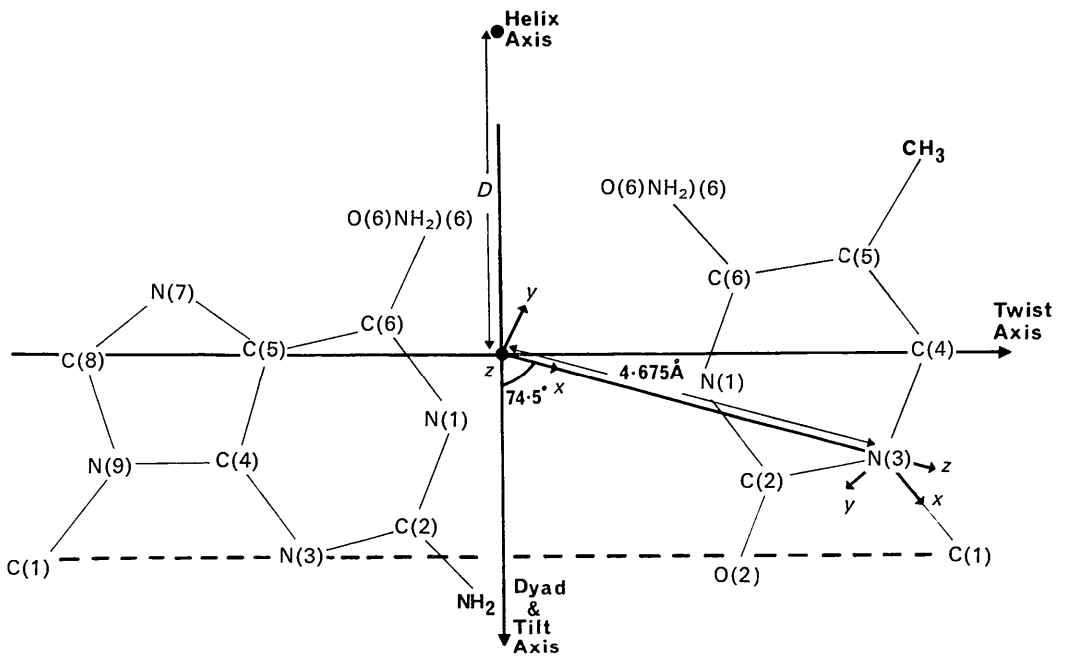


Fig. 2. The relationship of the four parameters θ_1 (tilt), θ_2 (twist), θ_3 (swivel, fixed at $-74{\cdot}5°$) and D to the purine–pyrimidine pairs in DNA and RNA.

ment implies that the traditional crystallographic residual,

$$R = \sum_m |{}_oF_m - F_m| / \sum_m {}_oF_m \,, \qquad (5)$$

normally should be less than 0·40 (Arnott, 1964; Fuller *et al.*, 1965; Arnott *et al.*, 1967*b*). However R is not a convenient function to minimize by least-squares processes and we have therefore defined the best fit as minimum Φ with

$$\Phi = \sum_{m=1}^{M} \omega_m \{ {}_oF_m - F_m(1/K) \exp(-B\varrho^2/4) \}^2 \qquad (6)$$

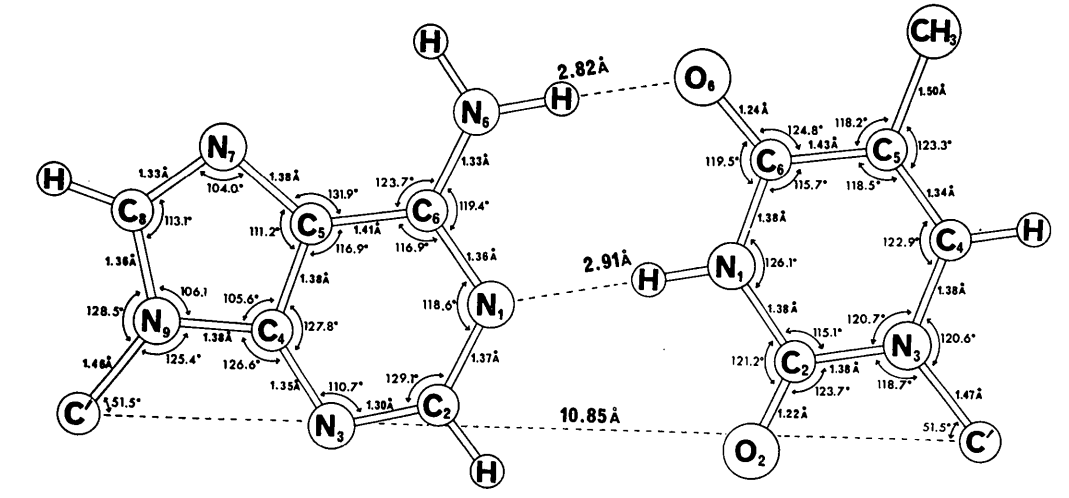
where ω_m is the weight given to the mth reflexion. Strictly ω_m should be related to the reliability of measurement, but following Cruickshank (1960) we have used an analytical expression

$$\omega_m^{-1} = \{1 + a({}_oF_m) + b({}_oF_m)^2\} \,. \qquad (7)$$

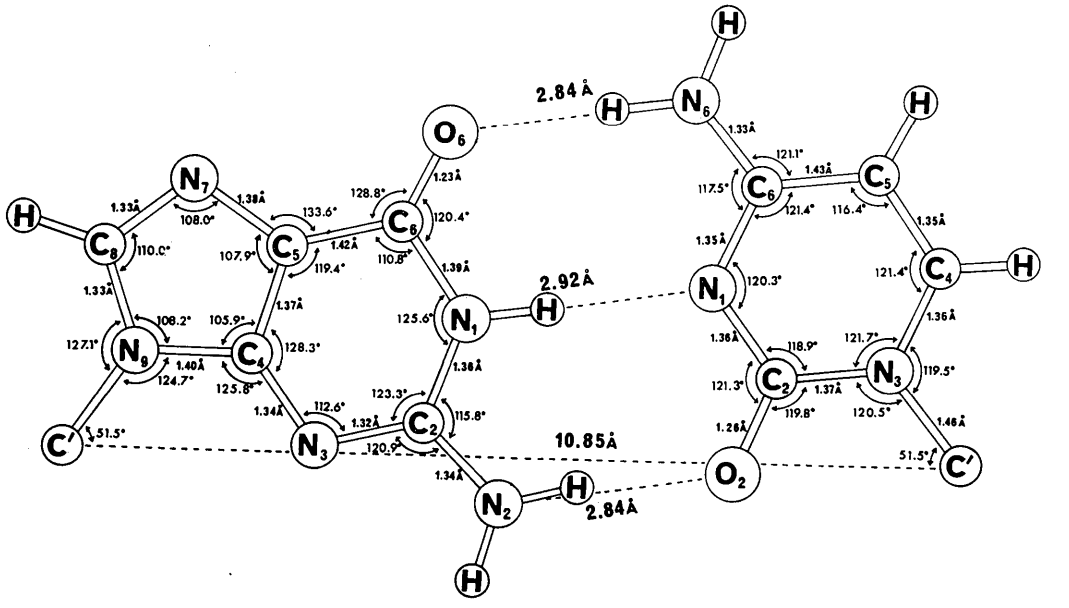
Since the number, M, and the scale, K, of the reflexions is not necessarily the same for each analysis, we later quote

$$\langle \Phi \rangle = K^2 \Phi / M \qquad (8)$$

when comparing the agreement achieved for the different structures.



(*a*)



(*b*)

Fig. 3. The Watson–Crick type base pairs used in this analysis (*a*) adenine–thymine, (*b*) guanine–cytosine.

Constraints

The constraints to be imposed on the molecular models may be either exact or elastic. In the former class we have imposed explicit, exact constraints by maintaining bond lengths (l) and bond angles (φ) as fixed quantities during all the refinements. We have felt justified in doing this because the accuracy with which these qualities are known from accurate single crystal studies is *at least* an order of magnitude better than could be determined from fibre diffraction data.

Once fixed bond lengths and angles have been chosen for the sugar residues, the ring conformation is fixed except that four discrete solutions (Spencer, 1959; Sundaralingam & Jensen, 1965) are possible, *viz.* (*i*) C(2)-*endo*, (ii) C(3)-*endo*, (iii) C(2)-*exo*, (iv) C(3)-*exo*. In (i) C(2) and (ii) C(3) are ~0·5 Å from the C(1)–O(5)–C(4) plane on the same side as C(5), while C(3) and C(2) are ~0·1 Å from the plane on the opposite side from C(5). In the *exo* conformations the named atom is also the one more distant from the C(1)–O(5)–C(4) plane, but in this case on the side distant from C(5). It appears that in *A*-DNA and RNA the conformation is C(3)-*endo* and in *B*-DNA, C(2)-*endo*. The sugar conformational solutions are defined by τ_{78}, τ_{89} and $\tau_{9,10}$ and they therefore also had fixed values. For the values of the sugar parameters we have chosen, the displacements of C(2) and C(3) are: in C(3)-*endo*, −0·15 and 0·47 Å, and in C(2)-*endo*, 0·50 and −0·13 Å respectively.

The linked-atom description of the molecular chain in which the ls, φs and τs are among the parameters, was adopted to facilitate the *explicit* imposition of the constraints discussed above. However, we impose other exact constraints requiring *implicit* relationships among the conformational parameters. These derive from an assumption that all the molecules considered here are regular helices. From the positions of the diffraction maxima one may accurately determine the pitch, c, and from considerations such as the positions of meridional reflexions and the general intensity distribution one may define the number, p, of nucleotide pairs per helix turn. Therefore, if r_n, θ_n and Z_n are the cylindrical polar coordinates of chain atoms, to ensure that successive residues are linked appropriately, the chain parameters have to be such that all $G_h=0$ where

$$G_1=r_1-r_7,\ G_2=r_2-r_8,\ G_3=r_3-r_{7a},$$
$$G_4=\theta_1-\theta_7-360°/p,\ G_5=\theta_2-\theta_8-360°/p,\ G_6=\theta_3-\theta_{7a}-360°/p\ ,$$
$$G_7=Z_1-Z_7-c/p,\ G_8=Z_2-Z_8-c/p,\ G_9=Z_3-Z_{7a}-c/p\ . \quad (9)$$

We have imposed such implicit, exact relationships by minimizing

$$\Theta=\Phi+\sum_1^H \lambda_h G_h \quad (10)$$

in the least-squares fashion previously described (Arnott & Wonacott, 1966*a*) so obtaining the minimum Φ compatible with all $G_h=0$. In minimizing Θ, the Lagrange multipliers, λ_h, are variables as well as the non-constant structure parameters.

Elastic constraints would arise from a foreknowledge of geometrical properties less precise than that available

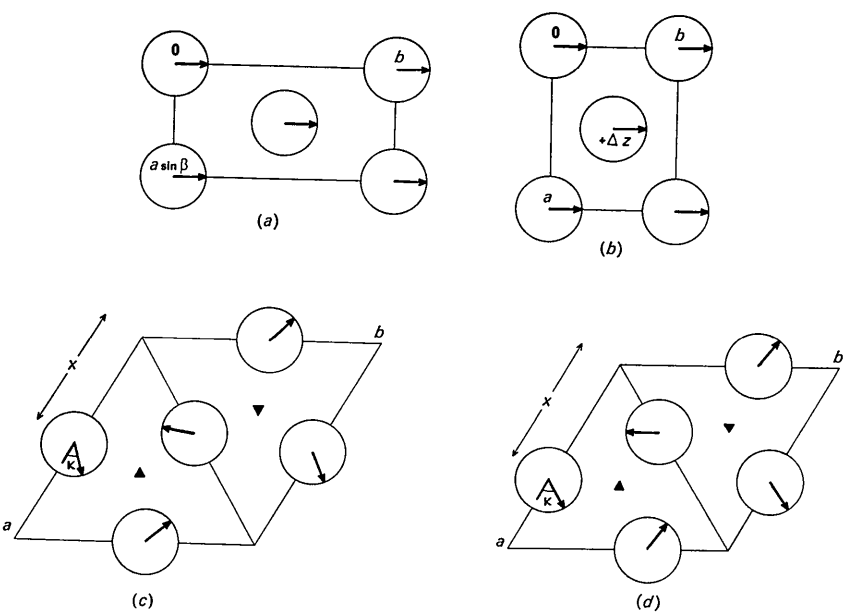


Fig. 4. Views of the unit cell for (*a*) *A*-DNA, (*b*) *B*-DNA, (*c*) α-RNA, (*d*) β-RNA, projected down the *c*(fibre)axis. The heavy arrows indicate the direction of one molecular diad by which the molecular orientation is defined. In *A*-DNA no packing parameters were varied; in *B*-DNA the orientation was fixed but ΔZ was varied; in α- and β-RNA, κ and x were varied. The triad axes (▲) shown in (*c*) and (*d*) are appropriate for elevenfold RNA models; for tenfold models a left-handed screw triad replaces them. For *B*-DNA the refined value of ΔZ is 0·327*c*. For elevenfold RNA models κ and x are 19·7°, 0·541*a* in α and 4·9°, 0·637*a* in β. For tenfold models the corresponding values are 22·3°, 0·539*a* and −0·8°, 0·634*a*.

for bond lengths and angles. For example, the likely range of values for a number of the chain dihedral angles can be deduced from single-crystal analyses of nucleosides and nucleotides and from model-building studies; infrared dichroism in particular absorption bands can give some indication of phosphate group orientation. Ideally, then, one should minimize

$$\Psi = \Theta + \sum k_i(d_i - {}_od_i)^2 \qquad (11)$$

to retain some angle or distance, d_i, near the mean value of its range, ${}_od_i$. The main difficulty arises in deciding the relative weights, k_i, to be assigned to different structure features, and how far the fit with the X-ray data should be sacrificed to achieve reasonable values of each d. Accordingly we have not imposed any elastic constraints but have preferred to use this less exact information to make semi-quantitative tests of the credibility of the models we obtain by minimizing only Θ.

We have made one exception for what we view as less precise stereochemical information, and have, by imposing an exact implicit constraint, required that the phosphate group orientations fall within the limits (usually $\pm 10°$) set by the infrared analyses unless this resulted in a gross worsening of the fit between the observed and model structure amplitudes.

We have not included any of the possible inequality constraints, but between atoms we have assumed that there should rarely be non-bonded distances 0·2 Å less than the sum of the appropriate van der Waals radii (for C, 1·7 Å; for N, 1·5 Å; for O, 1·4 Å) and none less than 0·4 Å below the van der Waals sum. Models where the *intra*molecular non-bonded distances did not all satisfy these criteria would therefore have been rejected. In the event, the only questionable distances discovered were two *inter*molecular contacts in *B*-DNA, and the special circumstances that would justify these are discussed later.

Intensity data

We have obtained the intensity data for *A*-DNA from Fuller (1961), for *B*-DNA from Marvin (1960), and for α- and β-RNA from Arnott, Wilkins, Fuller & Langridge (1967*a*). To make all the refinements comparable, the measured intensites were all reduced to structure amplitudes (${}_oF_m$) by use of the geometrical factors appropriate for crystalline fibre diffractograms (Arnott, 1965). Where reflexion overlap occurred, ${}_oF_m = (\sum f_i^2)^{1/2}$ is composite and was treated as one datum and compared with the corresponding model quantity.

The course of the refinements

Six polynucleotide refinements are reported here: *A*-DNA, B-DNA, α-and β-RNA (for the RNA crystal structures both ten- and elevenfold helical models had to be considered). In all the refinements the variable molecular parameters were τ_{23}, τ_{34}, τ_{45}, τ_{56}, τ_{67}, $\tau_{10,11}$, θ_1, θ_2 and D. θ_2 always had an initial value of zero, but the other variables were assigned starting values derived from the previous models, except in the case of α-RNA, where the starting values were the final ones from the β-RNA study. The other variables were K, B and the relevant packing parameters indicated in Fig. 4.

Because triangles of atoms are rigid bodies in our system only six of the nine G_h need be included when minimizing so that the number of degrees of freedom in each molecular conformation was $9-6=3$, except in the cases where a phosphate orientation constraint was imposed when it was $9-7=2$. With the other variables, every crystal structure had at least 4 degrees of freedom and none more than 7.

In spite of the reasonable first values of the variables, because of our amendments to the bond lengths and angles, the initial models invariably far from satisfied the constraint relationships, $G_h=0$. Even when the required parameter shifts are large and the $G_h \gg 0$, few iterations are required before the minimum Θ is obtained. Although many τ values had to change as much as 20°, and in some cases G_h had to be reduced from 1·5 Å to zero, the number of iterations required before convergence was achieved was rarely more than 6.

Reduction of $\sum_1^H \lambda_h G_h$ to zero and the reduction of Φ occurred continuously and simultaneously during the refinements. The final values of R for *A*-DNA, *B*-DNA, α-RNA11 β-RNA11, α-RNA10, β-RNA10 (0·41, 0·36, 0·37, 0·34, 0·35, 0·33 respectively) are, except in the case of *A*-DNA, better than the values achieved by the trial and error method (0·39, 0·40, 0·45, 0·38, 0·48, 0·41), although the latter method, with its greater molecular flexibility, is (presumably) *potentially* capable of yielding a better fit with the X-ray data. The final values of $\langle\Phi\rangle$ (see Table 1) for the DNA structures ($0{\cdot}28 \times 10^7$ and $0{\cdot}91 \times 107$) are lower than for the RNA structures in token of their better quality data and the fact that only two double helices pass through each unit cell compared with three in the RNA structures. The better quality of the β-RNA compared with the α-RNA diffractograms is also reflected in lower $\langle\Phi\rangle$. For RNA, tenfold helices have worse values of $\langle\Phi\rangle$ ($2{\cdot}77 \times 10^7$ in α, $2{\cdot}71 \times 10^7$ in β) than the elevenfold helices ($2{\cdot}64 \times 10^7$ and $2{\cdot}42 \times 10^7$).

Table 1. *Agreement with the X-ray data*

$R = \Sigma|{}_oF_m - F_m|/\Sigma\,{}_oF_m$ is the conventional crystallographic residual but is not directly related to the quantity minimized in the reported refinements. The quantity minimized is $\Phi = \Sigma\omega_m({}_oF_m - F_m)^2$ and $\langle\Phi\rangle = K^2\Phi/M$ is a measure of Φ on a quasi-absolute scale that allows comparison of different refinements.

Structure	Number of reflexions (M)	R	$\langle\Phi\rangle \times 10^{-7}$
A-DNA	232	0·41	0·28
B-DNA	226	0·36	0·91
α-RNA10	125	0·35	2·77
α-RNA11	125	0·37	2·64
β-RNA10	93	0·33	2·71
β-RNA11	93	0·34	2·42

Stereochemical acceptability

One of the more convincing tests of the credibility of the refined models is the almost complete absence of unacceptably short non-bonded intramolecular contacts Only in the DNA structures were there any such contacts with less than the fully allowed values (defined to be 0·2 Å less than the sum of the van der Waals radii of the atom pair in question). That this is achieved simultaneously with a good fit with the X-ray data is compelling circumstantial evidence of the adequacy of our stereochemical presumptions.

In *A*-DNA the only contacts not fully allowed involve the methyl group of thymine with the base below, particularly if it is a pyrimidine when there are four distances of ~3·1 Å. These short contacts were a feature also of the Fuller *et al.* (1965) model, where they were relieved by postulating that the carbon–methyl bond was bent 3° out of the mean base plane. There is no difficulty involving the methyl group in the *B*-DNA conformation and it may be that this is a factor in determining that *B* is the more usual and native form. It is interesting that in the stable RNA structures which all resemble *A*-DNA, this difficulty is avoided only because uracil replaces thymine.

In *B*-DNA there are only two anomalous intramolecular contacts, *viz.* 2·7 Å between O(4) and sugar C(3) and 2·8 Å between sugar C(2) and pyrimidine C(4) [or purine C(8)]. We regard both as tolerable and in this we are supported by the Haschmeyer & Rich (1967) review of the minimum distances observed in nucleosides and nucleotides: such short distances are observed particularly when at least one of the atoms is part of a rigid ring system.

The only unacceptable *inter*molecular distances are in *B*-DNA where twice in each double helical pitch length (*i.e.* only once every ten nucleotides) there is a short contact of 2·3 Å between the O(2) atoms of neighbouring molecules. We note that this could be relieved by a small local distortion in each helix and that such a distortion might explain the weak meridional reflexions (004, 006) in the *B*-DNA diffractograms that indicate that not all the scattering matter lies on regular tenfold helices.

From studies of the absorption of polarized infrared radiation, information about the orientation of the phosphate group in polynucleotides may be deduced. This is normally expressed as the inclination of the O(2)–O(3) direction and of the bisector of the O(2)–P–O(3) angle to the helix axis. We compare, in Table 2, the orientations in our refined models with the values in infrared experiments. Although we are uncertain of the weight to be given to the infrared data, only in the case of *A*-DNA have we not been able to obtain satisfactory values while fitting the X-ray data. Fuller *et al.* (1965) experienced a similar difficulty and we must conclude either that some of our constraint assumptions are in error or that the infrared evidence for this structure should be reappraised. We have no information at this time that would support the former proposition.

Table 2. *Phosphate group orientation*

Bold figures are values achieved by explicit constraint in the X-ray refinements.

Structure	Angle between O(2)–O(3) line and helix axis		Angle between the bisector of O(2)–P–O(3) and helix axis	
	X-ray	Infrared	X-ray	Infrared
A-DNA	**65°**	55°	31°	65°
B-DNA	**61**	55	70	67
α-RNA11	76	70	**33**	40
β-RNA11	66	70	**33**	40
α-RNA10	72	70	36	40
β-RNA10	71	70	32	40

Comparison with earlier models

In this section we consider mainly the differences in bonded stereochemistry between the models presented in this paper and those published earlier, and go on to discuss to what extent there is agreement between the trial-and-error approach and these linked-atom refinements, in the disposition of the bases and phosphates that are the dominant X-ray scattering groups.

A-DNA

Of all the published structures, that of *A*-DNA by Fuller *et al.* (1965) requires least amendment of the bonded stereochemistry. The main differences are in the deoxyribose ring, where we have amended the angles at C(1) by 2·5° on average, and the C(1)–C(2) bond length that we reduced from 1·56 to 1·52 Å. The chain P–O distances are now 1·60 rather than 1·54 Å.

The refined base parameters θ_1, θ_2 and D are (19·3°, −3·2°, 4·71 Å) instead of (20·0°, −8·0°, 4·25 Å). The new value for θ_2 is substantially nearer zero and therefore represents an improvement: the previous large value was chosen apparently to avoid short inter-base contacts. With bases now 0·36 Å further away from the helix axis, this difficulty does not arise. The cylindrical polar coordinates of the P atom are now (8·79 Å, 68·2°, −4·15 Å) and very close to the Fuller *et al.* (1965) values (8·84 Å, 67·2°, −3·93 Å).

The model we offer (parameters in Table 3, coordinates in Table 4) is therefore very similar to that of Fuller *et al.* (1965). The differences in, say, the P coordinates (−0·05 Å, +1·0°, −0·22 Å) are probably a good indication of the limits of accuracy that can be achieved with fibre diffraction data.

The amelioration of the stereochemistry in this structure has therefore been achieved with little change in the disposition of the main scattering groups and consequently with little change in the fit with the X-ray intensities. R is now 0·41 whereas the value found by Fuller *et al.* was 0·39. Our higher value for R must arise from the more rigorous constraints we have imposed, but does not represent a serious worsening of the X-ray data fit.

B-DNA

Since most of the available accurate single-crystal studies of nucleosides *etc.* were completed subsequent to the development of the model of Langridge *et al.* (1960) for *B*-DNA, there are notable differences with the bonded stereochemistry we have now adopted. The bases themselves are somewhat different but more important is the reduction of the hydrogen bond lengths in the base-pairs that reduces the interglycosidic link distance from 11·1 to 10·85 Å. The six bond angles at C(1) and C(4) in the sugar ring have been changed by 4·8° on average and the ring now has a standard C(2)-*endo* conformation with C(2) 0·51 Å, and C(3) −0·13 Å out of the C(1)–O(5)–C(4) plane; previously the corresponding quantities were 0·19 and −0·10 Å. The chain C–O bond lengths are now 0·05 Å less and the chain P–C lengths 0·06 Å greater.

In spite of the many stereochemical changes the main scattering groups return to more or less their previous positions after the structure refinement. The base parameters (Table 3) are now $\theta_1 = -2{\cdot}1°$, $\theta_2 = 4{\cdot}0°$, $D = -0{\cdot}16$ Å, compared with the previous values of −2°, 5°, −0·63 Å. The new P atom position, previously (9·05 Å, 94·8°, 2·04 Å), is (9·22 Å, 94·4°, 1·90 Å) (see Table 4).

Although our model has many fewer degrees of freedom than Langridge *et al.* allowed, *R* is now 0·36 compared with 0·40 previously. Substantial refinement both of the fixed geometrical features and their arrangement in the molecule has therefore been accomplished.

α- and β-RNA

Both the ten- and elevenfold models for the RNA −0·16 Å, compared with the previous values of–2°, 5°, molecule (Arnott *et al.*, 1967*b*) are somewhat similar to *A*-DNA but the manner in which they were generated from a model of the latter left a number of stereochemical infelicities. The standard bond angles we have chosen are on average 4° different from the previous

Table 3. *Molecular parameters* (*fixed values in bold type*)

	τ_{23}	τ_{34}	τ_{45}	τ_{56}	τ_{67}	τ_{78}	τ_{89}	$\tau_{9,10}$	$\tau_{11,11}$	θ_1	θ_2	θ_3	D
A-DNA	160·7°	64·1°	64·8°	170·2°	−45·0°	**155.3°**	**−37.3°**	**−94.4°**	−80·5°	19·3°	−3·2°	**−74.5°**	4·71 Å
B-DNA	189·5	102·9	41·0	166·0	−33·4	**89.4**	**37.3**	**−153.5**	−142·6	−2·1	4·0	**−74.5**	−0·16
α-RNA11	158·2	67·1	66·5	173·6	−49·7	**155.3**	**−37.3**	**−94.4**	−75·0	13·6	−4·5	**−74.5**	4·41
β-RNA11	158·0	67·2	65·8	174·2	−49·4	**155.3**	**−37.3**	**−94.4**	−75·2	13·3	−3·9	**−74.5**	4·34
α-RNA10	166·0	64·1	103·0	161·7	76·6	**155.3**	**−37.3**	**−94.4**	−76·0	13·4	−0·4	**−74.5**	3·65
β-RNA10	147·0	73·6	98·7	180·1	−83·5	**155.3**	**−37.3**	**−94.4**	−70·2	12·2	2·9	**−74.5**	3·65

Table 4. *Cylindrical polar coordinates for refined DNA models with standard stereochemistry*

		A-DNA			*B*-DNA		
		r(Å)	θ(°)	Z(Å)	r(Å)	θ(°)	Z(Å)
Phosphate	P	8·68	68·7	−4·06	9·22	94·4	1·90
	O(1)	9·52	72·3	−2·82	8·80	96·5	3·41
	O(2)	9·52	68·7	−5·27	10·59	91·2	1·88
	O(3)	7·40	74·0	−4·17	9·12	102·1	1·07
	O(4)	8·58	58·2	−3·58	8·20	86·7	1·49
Sugar	C(5)	9·86	54·5	−3·40	8·06	79·2	2·44
	O(5)	9·36	43·9	−1·85	6·31	67·3	1·69
	C(4)	9·83	45·8	−3·81	7·72	69·7	1·75
	O(1)	9·59	39·4	−5·41	8·80	60·8	0·03
	C(3)	9·03	40·3	−4·12	8·22	69·0	0·32
	C(2)	9·01	32·0	−3·32	6·99	72·8	−0·45
	C(1)	8·74	35·5	−1·92	5·85	67·5	0·37
Purine	O(6), $(NH_2)(6)$	3·32	14·3	−0·20	1·86	153·0	−0·08
	C(6)	4·55	16·5	−0·43	1·45	110·4	0·03
	C(5)	5·27	29·0	−0·89	2·72	95·5	0·09
	N(7)	5·31	43·4	−1·23	3·98	104·6	0·08
	C(8)	6·59	44·7	−1·62	4·91	92·2	0·18
	N(9)	7·31	35·4	−1·56	4·63	76·4	0·25
	C(4)	6·66	26·0	−1·10	3·31	70·3	0·20
	N(3)	7·39	17·0	−0·90	3·21	46·7	0·25
	C(2)	6·82	8·2	−0·46	2·16	29·6	0·18
	N(1)	5·50	5·4	−0·23	0·82	42·8	0·07
Guanine	$(NH_2)(2)$	7·72	0·5	−0·20	2·85	2·4	0·22
Pyrimidine	O(6), $(NH_2)(6)$	3·41	33·4	−0·55	2·85	136·0	−0·06
	C(6)	4·66	34·8	−0·88	2·99	110·4	0·04
	N(1)	5·56	23·3	−0·80	2·35	84·1	0·11
	C(2)	6·86	25·6	−1·13	3·43	67·4	0·22
	O(2)	7·75	18·8	−1·03	3·65	47·5	0·28
	N(3)	7·31	35·4	−1·56	4·63	76·4	0·25
	C(4)	6·67	45·3	−1·66	5·03	92·1	0·18
	C(5)	5·38	47·8	−1·33	4·38	106·6	0·08
Thymine	CH_3	4·94	64·0	−1·40	5·38	120·2	−0·01

ones, and the bond lengths too have been amended, particularly in the sugar ring where the average reduction is the C–C bond lengths in 0·04 Å.

We have not assumed that the molecular conformation is identical in both crystals and have refined both ten- and elevenfold models with both the α and β sets of data. For the elevenfold model the previous values of the base parameters were $\theta_1 = 14{\cdot}0°$, $\theta_2 = 0{\cdot}0°$ and $D = 4{\cdot}25$ Å; we find that these are now 13·6°, −4·5°, 4·41 Å in α-RNA and 13·3°, −3·9°, 4·34 Å in β-RNA. The P atom positions in the new models are also similar, *viz.* (8·83 Å, 70·6°, −3·52 Å) and (8·83 Å, 71·0°, −3·47 Å) respectively; the previous values were (8·84 Å 68·5°, −3·62 Å). The tenfold models of Arnott *et al.* (1967*b*) had $\theta_1 = 10°$ $\theta_2 = 0{\cdot}0°$, $D = 4{\cdot}05$ Å; we find in α-RNA (13·4°, −0·4°, 3·65 Å) and in β-RNA (12·2°, 2·9°, 3·65 Å). The new P atom positions are (8·59 Å, 75.3°, −2.99Å) and (8.80Å, 74.8° −2.81Å) compared with (8.96 Å, 72·5°, −2·67 Å) previously.

The overall fit with the intensities has been substantially improved. The elevenfold models have values of R of 0·37 and 0·34 in α and β-RNA respectively compared with the values 0·45 and 0·38 of Arnott *et al.*; the tenfold models also exhibit an improved fit with the data: $R = 0{\cdot}35$ in α-RNA and 0·33 in β-RNA compared with 0·48 and 0·41 previously.

The refined values of the structure parameters are in Table 3 and the atomic positions in Tables 5(*a*) and 5(*b*).

The molecular and crystal structure of RNA

Removal of the stereochemical anomalies in the RNA models and refinement of the conformations and molecular packing allows us *objectively* to reappraise to what extent the following propositions are true:

(*a*) the RNA helices are elevenfold and not tenfold;
(*b*) the RNA helices are identical in α- and β-RNA;
(*c*) only one set of RNA–RNA interactions is used in the formation of both crystal forms;
(*d*) there is a well-defined system of hydrogen bonds linking the RNA double helices with one another.

(*a*) The arguments that led Arnott and co-workers (Arnott, Wilkins, Fuller & Langridge, 1967 *a*, *b*; Arnott, Wilkins, Fuller, Venable & Langridge, 1967) to indicate a preference for elevenfold helices in α and β-RNA could not be based on the fit with the X-ray data since the only proper basis for such a decision would be a comparison of *optimized* ten- and elevenfold models. The refinements we have conducted above show that in both crystal structures elevenfold helices are favoured when one considers the overall fit with the X-ray intensities. In α-RNA $\langle\Phi\rangle$ is 5% higher for tenfold helices than for eleven, and in β-RNA (which has the better quality data) the preference for elevenfold helices is more marked at 12%. That the distinction is not more marked emphasizes how similar eleven and tenfold helices can be, but since the same bias

Table 5(*a*). *Cylindrical polar coordinates for elevenfold* RNA *models*

		α-RNA			β-RNA		
		r(Å)	θ(°)	Z(Å)	r(Å)	θ(°)	Z(Å)
Phosphate	P	8·82	70·6	−3·53	8·83	71·0	−3·48
	O(1)	9·63	73·7	−2·24	9·62	74·0	−2·18
	O(2)	9·72	70·7	−4·70	9·74	71·1	−4·64
	O(3)	7·59	76·3	−3·67	7·61	76·7	−3·63
	O(4)	8·64	60·5	−3·12	8·63	60·9	−3·08
Sugar	C(5)	9·88	56·1	−2·96	9·87	56·4	−2·90
	O(5)	9·21	45·6	−1·42	9·17	46·0	−1·37
	C(4)	9·77	47·4	−2·73	9·75	47·7	−2·67
	O(1)	9·63	41·0	−4·96	9·62	41·3	−4·90
	C(3)	8·99	42·1	−3·71	8·97	42·4	−3·65
	O(6)	10·24	31·0	−2·86	10·20	31·3	−2·80
	C(2)	8·91	34·0	−2·84	8·87	34·3	−2·80
	C(1)	8·58	37·5	−1·46	8·54	37·9	−1·41
Purine	O(6), $(NH_2)(6)$	3·03	16·1	−0·08	2·96	16·4	−0·09
	C(6)	4·27	18·1	−0·29	4·21	18·4	−0·29
	C(5)	5·05	31·4	−0·63	4·99	31·9	−0·62
	N(7)	5·18	46·5	−0·84	5·13	47·1	−0·83
	C(8)	6·48	47·5	−1·15	6·44	47·9	−1·13
	N(9)	7·14	37·6	−1·15	7·09	38·0	−1·12
	C(4)	6·43	27·9	−0·83	6·37	28·3	−0·80
	N(3)	7·12	18·3	−0·73	7·06	18·5	−0·70
	C(2)	6·52	8·8	−0·41	6·45	9·0	−0·38
	N(1)	5·20	5·9	−0·19	5·13	6·0	−0·18
Guanine	$(NH_2)(2)$	7·41	0·5	−0·26	7·34	0·6	−0·23
Pyrimidine	O(6), $(NH_2)(6)$	3·19	37·2	−0·32	3·13	37·9	−0·32
	C(6)	4·46	37·8	−0·59	4·41	38·4	−0·59
	N(1)	5·31	25·3	−0·58	5·25	25·6	−0·57
	C(2)	6·63	27·5	−0·86	6·57	27·8	−0·83
	O(2)	7·49	20·1	−0·84	7·43	20·4	−0·80
	N(3)	7·14	37·6	−1·15	7·09	38·0	−1·12
	C(4)	6·56	48·1	−1·18	6·52	48·6	−1·16
	C(5)	5·27	51·1	−0·90	5·23	51·7	−0·89

Table 5(*b*). *Cylindrical polar coordinates for tenfold* RNA *models*

		α-RNA			β-RNA		
		r(Å)	θ(°)	*Z*(Å)	*r*(Å)	θ(°)	*Z*(Å)
Phosphate	P	8·59	75·3	−2·99	8·80	74·8	−2·81
	O(1)	9·29	79·7	−1·73	9·48	77·7	−1·44
	O(2)	9·56	75·0	−4·11	9·82	74·7	−3·89
	O(3)	7·32	80·4	−3·29	7·63	80·9	−3·08
	O(4)	8·47	65·3	−2·43	8·51	64·6	−2·46
Sugar	C(5)	9·55	59·7	−2·80	9·55	58·9	−2·86
	O(5)	8·72	49·4	−1·24	8·73	49·8	−1·11
	C(4)	9·36	50·8	−2·52	9·40	50·2	−2·39
	O(1)	9·29	43·7	−4·72	9·48	41·6	−4·43
	C(3)	8·59	45·3	−3·51	8·71	43·8	−3·29
	O(6)	9·72	33·5	−2·56	9·85	33·3	−2·06
	C(2)	8·41	37·0	−2·62	8·53	36·3	−2·27
	C(1)	8·03	41·0	−1·28	8·06	41·4	−1·02
Purine	O(6), $(NH_2)(6)$	2·30	20·8	−0·19	2·30	20·5	−0·25
	C(6)	3·56	21·9	−0·31	3·56	21·9	−0·30
	C(5)	4·42	36·6	−0·63	4·43	36·7	−0·58
	N(7)	4·68	53·2	−0·89	4·68	53·3	−0·85
	C(8)	5·99	41·8	−1·14	6·01	53·0	−1·04
	N(9)	6·57	52·8	−1·05	6·60	42·1	−0·89
	C(4)	5·79	31·6	−0·73	5·80	31·9	−0·60
	N(3)	6·43	20·7	−0·56	6·44	21·0	−0·38
	C(2)	5·78	10·3	−0·26	5·78	10·6	−0·13
	N(1)	4·45	7·0	−0·14	4·45	7·2	−0·08
Guanine	$(NH_2)(2)$	6·66	1·0	−0·05	6·66	1·3	0·12
Pyrimidine	O(6), $(NH_2)(6)$	2·60	46·9	−0·44	2·59	46·7	−0·50
	C(6)	3·88	44·6	−0·65	3·88	44·6	−0·63
	N(1)	4·45	29·3	−0·55	4·65	29·5	−0·47
	C(2)	5·99	31·0	−0·75	6·00	31·3	−0·60
	O(2)	6·81	22·7	−0·64	6·82	23·0	−0·44
	N(3)	6·57	41·8	−1·05	6·60	42·1	−0·89
	C(4)	6·09	53·4	−1·17	6·10	53·6	−1·06
	C(5)	4·82	58·0	−0·97	4·82	58·1	−0·94

occurs in the two independent structure determinations, and since this is in the same direction as all the other arguments advanced by Arnott and co-workers, we feel justified in rejecting all tenfold helical models.

(*b*) The molecular parameters separately determined for the elevenfold helices in α and β-RNA show an almost incredible agreement: the average difference in the variable τs, $\langle|\Delta\tau|\rangle = 0{\cdot}4°$. For the refined tenfold helices $\langle|\Delta\tau|\rangle = 9{\cdot}6°$.

It is clear that since elevenfold helices are more likely, we can assert the identity of the molecular conformation in both crystals.

(*c*) The equidimensional and equilateral triangles *ABC* and *AB'C'* (see Fig. 5) in α-RNA10 and β-RNA11 have sides 21·8 and 22·1 Å; in α-RNA these sides are both 22·1 Å. We can therefore accept that all these triangles are effectively equidimensional. We can determine the mutual orientation of molecules in triangle *ABC* from $\mu = \kappa + \varepsilon \pm 360°n/N$ and in triangle *AB'C'* from $\mu = \mu' - 2\varepsilon - 60° \pm 360°n/N$ (*n* is any integer, *N* is the number of residues per helix pitch). The experimentally determined numbers are in Table 6.

We can conclude that in the unlikely event of the helices being tenfold there are two different RNA–RNA interactions. However if, as is more likely, the helices are elevenfold we have the very interesting result that the mutual orientations of the RNA molecules throughout both crystals are the same to within 1·3°.

Table 6. *Values of the angles defining* RNA *molecular orientations*

The symbols are defined in Fig. 5.

Structure	κ(°)	ε(°)	μ(°)	μ'(°)
α-RNA10	7·7	52·3	24·0	3·4
β-RNA10	24·8	35·2	24·0	2·4
α-RNA11	7·6	52·0	26·9	26·4
β-RNA11	22·4	24·6	24·3	25·9

(*d*) For RNA11, in each pitch length of each molecule, there is one O--O contact of 2·8 Å from O(6) to O(2) of each nearest neighbour. The C–O--O angle is 123°. This contact is therefore well qualified to represent an O–H--O hydrogen bond and, since each molecule has four nearest neighbours, each molecule throughout the crystal is involved in many hydrogen bonds with its nearest neighbours. Although in RNA10 there are two short contacts per pitch length between O(6) atoms of one molecule and O(2) atoms of the next (both are 2·6 Å long and have C–O--O angles of 130° and 150°) these contacts are confined to triangles *AB'C'* and do not therefore provide such an attractive rationale for the mode of packing of RNA helices as do the elevenfold model hydrogen bonds.

The molecular parameters

For the most part the positions of the main scattering groups in our new standard models are not greatly

different from those determined in the earlier trial-and-error studies. Since, however, the overall agreement with the X-ray diffraction is generally better and since all our models possess standard bonded geometry we might expect that we can now attach more significance to the values of τ we obtain. We commented on the insignificant differences in the τ values obtained for α-RNA11 and β-RNA11, but we note that the values for *A*-DNA are also very similar to these. The average discrepancy in the variable τ between *A*-DNA and the RNA11 structures is only 3·4°. Three polynucleotide structures all of which have the same C(3)-*endo* sugar conformation therefore appear also to have the same dihedral angles in the rest of the chain; the only changing features being the tilt of the base-pairs and their distance from the helix axis. *B*-DNA, the only polynucleotide refined that has the C(2)-*endo* sugar conformation, has also a similar set of τ values. The DNA and RNA structures may be compared and contrasted in Fig. 6 and Table 3.

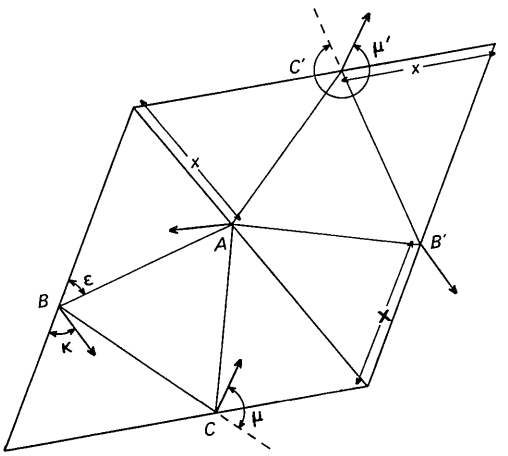


Fig. 5. The packing of RNA molecules. κ defines the absolute orientation in the unit cell and μ defines the mutual orientation of two neighbouring molecules in one nearest neighbour triangle; μ' is the corresponding angle in the second nearest neighbour triangle. ε is determined by the value of x.

One particular dihedral angle, $\tau_{10,11}$, that defines the mutual orientation of the base and sugar has been much discussed (*e.g.* Donohue & Trueblood, 1966; Haschmeyer & Rich, 1967). The usual values found in mononucleotides correspond to the *anti* conformation of Donohue & Trueblood (1960), but distinctive values exist for the different sugar conformations: for C(3)-*endo* rings, the average value of $\tau_{10,11}$ is about 20° higher than for C(2)-*endo* rings. The quantity usually quoted is φ_{CN} (with our choice of fixed stereochemistry, for C(2)-*endo*, $\varphi_{CN}=\tau_{10,11}+64{\cdot}8°$; for C(3)-*endo*, $\varphi_{CN}=\tau_{10,11}+61{\cdot}6°$). For *A*-DNA, α- and β-RNA11, $\varphi_{CN}=$

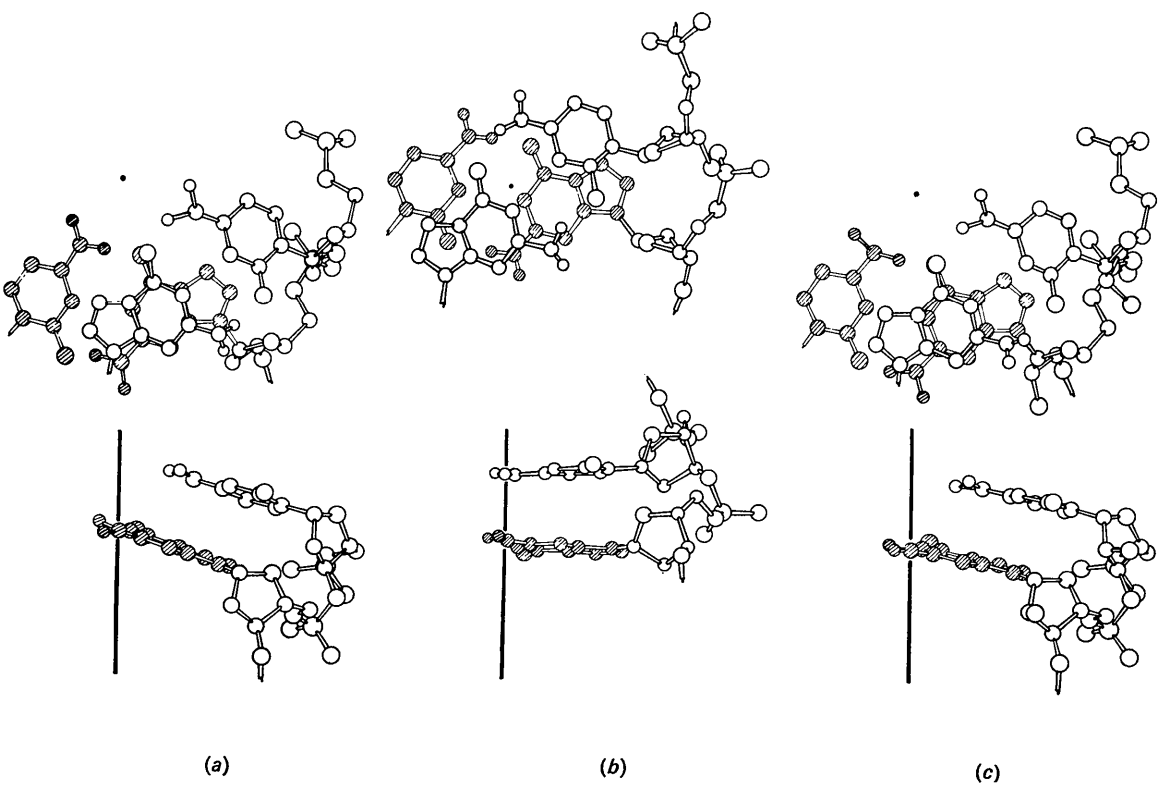


Fig. 6. Part of the molecular structures of (*a*) *A*-DNA, (*b*) *B*-DNA, (*c*) RNA11, seen projected down the fibre axis and down a molecular diad. ● indicates the projected helix axis position.

$-18{\cdot}9°$, $-13{\cdot}4°$, $-13{\cdot}6°$, respectively; these values are very close to the average value found in C(3)-*endo* nucleotide monomers. For C(2)-*endo* mononucleotides φ_{CN} is about 20° lower but it is of interest that for *B*-DNA $\varphi_{CN}(=-80{\cdot}6°)$ is some 16° lower than the lowest value found in these mononucleotides. This low value was a feature of earlier *B*-DNA models criticized by Donohue & Trueblood (1960) and its persistence in the refined models with updated stereochemistry makes us confident of its reality. The range of allowed values for φ_{CN} is fairly large, and of course there is no reason why the polymeric molecules that have simultaneously to satisfy other energy requirements should have exactly the same values as monomers in crystals. However, it might well be that the presence of features like this is partially responsible for the reduced thermal stability of DNA compared with RNA helices.

The angles about C(4)–C(5), τ_{67}, are similar in *A*-DNA, *B*-DNA and RNA ($-45{\cdot}0°$, $-33{\cdot}4°$, $-49{\cdot}6°$ respectively). These values imply that the C(5)–O(4) bond is *gauche* both to C(4)–O(5) and C(4)–C(3). This conformation appears to be the one most frequent in simple nucleotides.

The dihedral angles about C(5)–O(4), τ_{56}, are also similar (170·2°, 166·0°, 173·9°) to one another and their value near 180° confers on the polymeric molecule the extended form that is found in all the 5′-ribose phosphates that have been analysed (Trueblood, Horn & Luzzati, 1961; Kraut & Jensen, 1963; Furberg & Mostad, 1962; Shefter, Barlow, Sparks & Trueblood, 1964). The dihedral angles about the chain P–O bonds appear to be correlated with the sugar pucker. For the structure with C(3)-*endo* sugar conformations (*A*-DNA, RNA) the angles about O(4)–P, τ_{45}, are 64·8°, 66·2° respectively so that O(4)–C(5) is *gauche* to P–O(1), *trans* to P–O(2) and *gauche* to P–O(3). In C(2)-*endo*-containing *B*-DNA, $\tau_{45}=44{\cdot}1°$ implies a similar *gtg* conformation for O(4)–C(5).

As for the angle about P–O(1), τ_{34}, in *A*-DNA and RNA (64·1°, 67·2°) it is again contrived that the appropriate chain C–O [this time C(3)–O(1)] is *trans* to one pendant P–O but *gauche* to the other and to the chain bond P–O(4). In *B*-DNA $\tau_{34}=102{\cdot}9°$ is a similar but not identical conformation.

Finally, τ_{23}, the dihedral angle about C(3)–O(1), has similar values (160·7°, 189·5°, 158·1°) in *A*-DNA, *B*-DNA and RNA. O(1)–P is therefore *trans* to C(3);C(4) and *gauche* to C(3)–C(2). This is the reverse of the situation in cytidine monophosphate (Sundaralingam & Jensen, 1965) the only 3′-phosphate to have been studied in detail.

It is worth mentioning that the base pair twists (θ_2) that we would expect to be near (but, as single-crystal analyses of monomers show, not necessarily equal to) zero, deviate from this by only 3·5° on average among the 6 structures refined. In our base-pair as presently defined, this would imply a reduction (albeit not a serious one) in the linearity of the inter-base hydrogen bonds. However, (*e.g.* for cytosine in Marsh, Bierstedt & Eickhorn, 1962) the exocyclic amino and oxygen groups can be displaced, in opposite directions from the ring plane. It may be that in the polymers such small displacements also exist to return the hydrogen bonds towards their original perfection.

The structures with large values of θ_1 and *D* correlate with the presence of the C(3)-*endo* sugar conformation.

Base overlap

Much of the stability of DNA and RNA helices is alleged to derive from the hydrophobic interactions between stacked bases. We think it important that there should be a clear picture of what the term stacking implies in physical terms. Projected views of two successive base-pairs from the refined models discussed here are shown in Fig. 7. Firstly it is obvious that the stacking in *A*-DNA and RNA is similar but quite different from that in *B*-DNA. Secondly it is clear that there is certainly overlap of the base-pairs as a whole, if we consider not only the heterocyclic rings but the hydrogen-bonded regions as well. However, in every case there is relatively little overlap of the base planes themselves. This lack of overlap is particularly noticeable when a pyrimidine-purine occurs below a purine-pyrimidine pair in *B*-DNA. It is not therefore surprising that DNA (presumably in the *B* conformation) and RNA exhibit different thermal stabilities in solution. That RNA is the more stable may be because its degree of base-overlap is energetically preferable. It is not therefore necessary to invoke the earlier hypothesis of *intra*-molecular hydrogen bonding involving O(6), to account for the higher melting temperature. We must also assume that the absence of O(6) in desoxyribose makes the C(2)-*endo* sugar conformation preferable and the presence of the thymine methyl further supports *B* rather than *A* being the preferred conformation in DNA although *A*-DNA has RNA-like base-overlap.

The accuracy of the linked atom method and comparison with the traditional trial-and-error method

If we assume that in *A*-DNA (where no further improvement in *R* was obtained in these refinements) the difference in the position of the P atom as determined by our linked atom method and that determined by the manual trial-and-error method (0·27 Å) is not significant, a first approximation to the accuracy that can be claimed for these analyses is between a quarter and a third of an ångström. We note that the positions of the bases in *A*-DNA as determined by the two methods differ by a similar figure (0·36 Å). Since however, no non-bonded distances turn out to be more than 0·4 Å less than standard van der Waals distances it is likely that in many parts of the models the atomic positions are accurate to better than 0·2 Å. In addition, if we accept that RNA11 has the same molecular conformation in α and β-RNA we would conclude that the error might be as low as the 0·1 Å difference between

the P atom positions determined in the refinement of RNA11 from these two independent sets of diffraction data. In sum, therefore, we feel that 0·25 Å is likely to be near the *maximum* error in the atomic positions.

The computerized least-squares linked-atom approach we have described has the important virtues of speed, objectivity, explicitly defined (and maintained) constraints and a well-defined end-point of the analysis. In this it is superior to the traditional approach. The felicity of its use should make it, or an analogous method, a minimum requirement in the reporting of crystalline nucleic acid structures. We would emphasize, however, that it should be thought of as complementary to manual model-building analyses, since these are most desirable in providing a visual representation of a model, in obtaining preliminary parameters near enough the final ones for convergence to be assured, and in offering a system with additional flexible parameters that would be necessary in the cases, now fortunately rare, where many stereochemical constraints were available whose precise accuracy was suspect.

Notes added in proof:–

(1) Arnott, Fuller, Hodgson & Prutton (1968) describe a form of β-RNA obtained from a synthetic polyadenylic acid – polyuridylic acid complex. Twice as many data (188) have been obtained and their quality is much improved. $\langle\Phi\rangle$ for RNA10 is $1{\cdot}85\times10^7$ for this system and 33% higher than $\langle\Phi\rangle=1{\cdot}43\times10^7$ for RNA11.

(2) The unit-cell dimensions of *A*-DNA are $a=22{\cdot}2$, $b=40{\cdot}6$, $c=28{\cdot}2$Å, $\beta=97°$; of *B*-DNA are $a=31{\cdot}2$, $b=22{\cdot}7$, $c=33{\cdot}8$Å; of α-RNA are $a=b=44$, $c=30$Å, $\gamma=120°$; of β-RNA are $a=40$, $c=30$Å, $\gamma=120°$. The turn angles and translations per residue in *A*- and *B*-DNA and RNA10 and RNA11 are (32·7°, 2·56Å), (36°, 3·38Å), (36°, 3·00Å) and (32·7°, 2·73Å) respectively.

We are obliged to Professor Sir John Randall for our excellent facilities, to Professor M. H. F. Wilkins, Dr W. Fuller and Dr M. Spencer for discussion and to Dr Fuller and Dr W. J. Pigram for the use of some ancillary computing programs. We thank Mrs Janet Ric-

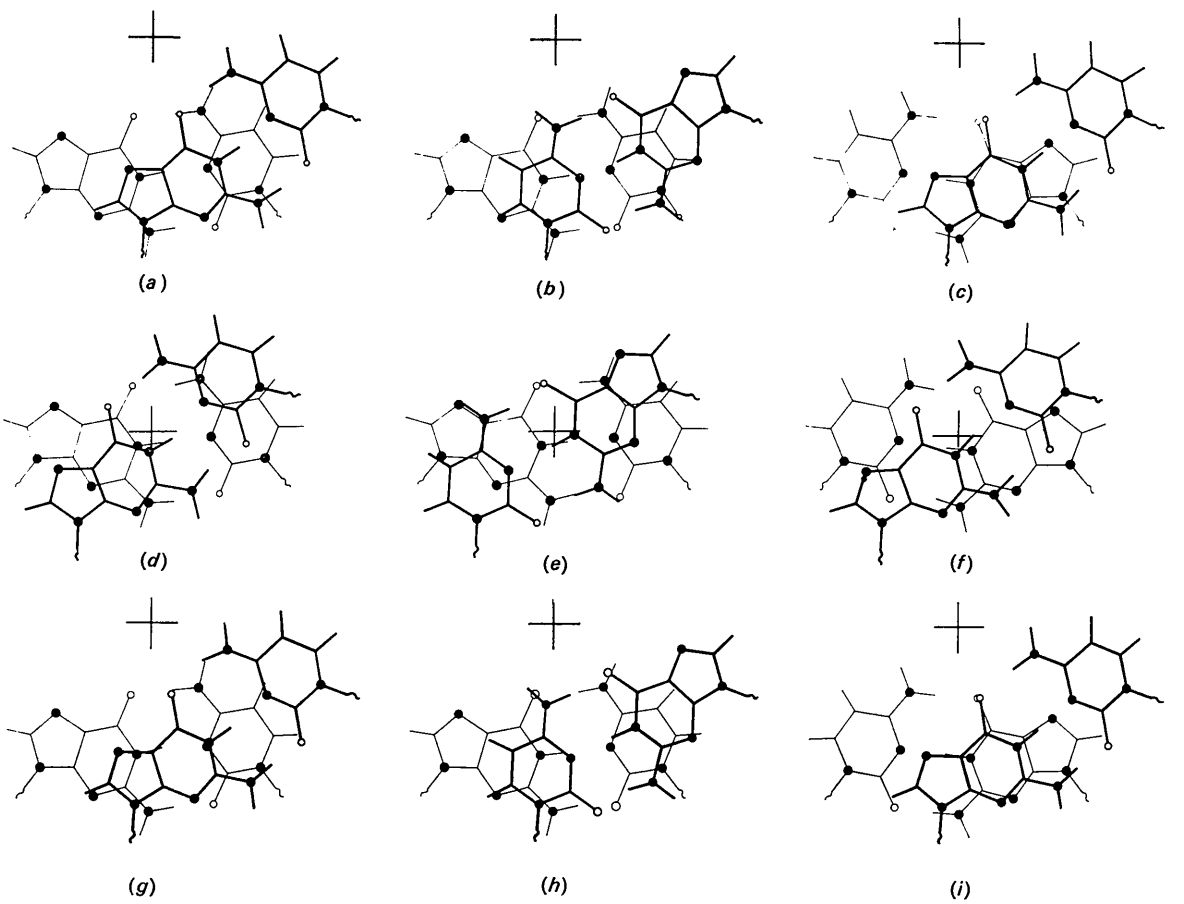


Fig. 7. Base overlap in DNA and RNA. (*a*), (*b*), (*c*) are for *A*-DNA, (*d*), (*e*), (*f*) for *B*-DNA and (*g*), (*h*), (*i*) for RNA. Only guanine–cytosine pairs are shown to typify purines and pyrimidines. (*a*), (*d*), (*g*) show the overlap when a purine follows a purine *and* when a pyrimidine follows a pyrimidine on the same chain; (*b*), (*e*), (*h*) when a purine follows a pyrimidine on the right-hand chain; (*c*), (*f*) and (*i*) when a pyrimidine follows a purine. The crosses are centred on the positions of the helix axes. The crossbars and uprights are each 1 Å.

kard and Mr Z. Gabor for technical assistance and the S.R.C. and M.R.C. for training grants (to A.J.W. and S.D.D. respectively).

References

ARNOTT, S. (1964). *Bull. Inst. Phys.* **15**, 169.
ARNOTT, S. (1965). *Polymer*, **6**, 478.
ARNOTT, S. (1968). *Symposium on Fibrous Proteins (Australia* 1967), p.26. Butterworth.
ARNOTT, S. & DOVER, S. D. (1967). *J. Mol. Biol.* **30**, 209.
ARNOTT, S. & DOVER, S. D. (1968). *Acta Cryst.* B**24**, 599.
ARNOTT, S., DOVER, S. D. & ELLIOTT, A. (1967). *J. Mol. Biol.* **30**, 201.
ARNOTT, S., FULLER, W., HODGSON, A. & PRUTTON (1968). *Nature, Lond.* **220**, 561.
ARNOTT, S., WILKINS, M. H. F., FULLER, W. & LANGRIDGE, R. (1967*a*). *J. Mol. Biol.* **27**, 525.
ARNOTT, S., WILKINS, M. H. F., FULLER, W. & LANGRIDGE, R. (1967*b*). *J. Mol. Biol.* **27**, 535.
ARNOTT, S., WILKINS, M. H. F., FULLER, W., VENABLE, J. & LANGRIDGE, R. (1967). *J. Mol. Biol.* **27**, 549.
ARNOTT, S. & WONACOTT, A. J. (1966*a*). *Polymer J.* **7**, 156.
ARNOTT, S. & WONACOTT, A. J. (1966*b*). *J. Mol. Biol.* **21**, 371.
CRUICKSHANK, D. W. J. (1961). In *Computing Methods and the Phase Problem in X-ray Crystal Analysis*, p. 32. Oxford: Pergamon Press.
DONOHUE, J. & TRUEBLOOD, K. N. (1960). *J. Mol. Biol.* **2**, 363.
FULLER, W. (1961). Ph.D. Thesis, Univ. of London.
FULLER, W., WILKINS, M. H. F., WILSON, H. R., HAMILTON, L. D. & ARNOTT, S. (1965). *J. Mol. Biol.* **12**, 60.
FURBERG, S. & MOSTAD, A. (1962). *Acta Chem. Scand.* **16**, 1627.
HASCHEMEYER, A. E. V. & RICH, A. (1967). *J. Mol. Biol.* **27**, 369.
KRAUT, J. & JENSEN, L. H. (1963). *Acta Cryst.* **16**, 79.
LANGRIDGE, R., MARVIN, D. A., SEEDS, W. E., WILSON, H. R., HOOPER, C. W., WILKINS, M. H. F. & HAMILTON, L. D. (1960). *J. Mol. Biol.* **2**, 38.
MARSH, R. E., BIERSTEDT, R. & EICHHORN, E. L. (1962). *Acta Cryst.* **15**, 310.
MARVIN, D. A. (1960). Ph.D. Thesis, University of London.
SHEFTER, E., BARLOW, M., SPARKS, R. & TRUEBLOOD, K. M. (1964). *J. Amer. Chem. Soc.* **86**, 1872.
SPENCER, M. (1959). *Acta Cryst.* **12**, 59.
SUNDARALINGAM, M. & JENSEN, L. H. (1965). *J. Mol. Biol.* **13**, 930.
TRUEBLOOD, K. N., HORN, P. & LUZZATI, V. (1961). *Acta Cryst.* **14**, 965.

Acta Cryst. (1969). B**25**, 2206

Röntgenstrukturanalyse eines Isomeren von Tris(cyclopentadienyl-rhodium-carbonyl), $(RhC_5H_5CO)_3$

VON E. F. PAULUS

Abteilung für Röntgenstrukturforschung am Max-Planck-Institut für Eiweiss- und Lederforschung München, Deutschland und Anorganisch-chemisches Laboratorium der Technischen Hochschule München, Deutschland

(*Eingegangen am* 1. *August* 1968 *und wiedereingereicht am* 18. *Dezember* 1968)

The crystal-structure of an isomer of tris(cyclopentadienylcarbonylrhodium) has been determined by X-ray methods. The structure has been refined to an R value of 3·6% with 1443 unique reflexions. The crystals belong to the space group $P2_12_12_1$. The rhodium atoms are arranged in a triangle, with rhodium–rhodium distances 2·620, 2·663, 2·705 Å. There are two bridging and one terminal carbonyl group. Each rhodium atom has one π-bonded cyclopentadienyl ring.

Einleitung

Bei der UV-Bestrahlung einer Lösung von

$$RhC_5H_5(CO)_2$$

in Hexan entsteht als Hauptprodukt eine Verbindung der Zusammensetzung RhC_5H_5CO (I), deren Struktur röntgenographisch aufgeklärt wurde und sich als trimer erwies (Mills & Paulus, 1966). Als Nebenprodukt dieser Reaktion tritt eine Verbindung (II) mit der gleichen Bruttozusammensetzung in Form schwarzer Nädelchen auf. Es lag nahe anzunehmen, dass es sich um ein Isomeres der ersteren Verbindung handelt.

Verbindung (I) zeigt im Gegensatz zu Verbindung (II) im IR-Spektrum keine Banden, die auf endständige Karbonylgruppen schliessen liessen. Die Röntgenstrukturanalyse von (I) ergab, dass ein Rhodiumdreiring mit drei völlig identischen Rhodium–Rhodiumbindungen vorliegt. Die Rhodiumatome werden ausserdem noch durch drei Brückenkarbonyle zusammengehalten. Im IR-Spektrum von (II) tritt ausser den Banden für verbrückte Karbonylgruppen bei 1827, 1794 und 1744 cm^{-1} noch eine Bande bei 1973 cm^{-1} auf.

Gegenwärtige Adresse: Farbwerke Hoechst AG, vorm. Meister Lucius & Brüning, Frankfurt (M) – Hoechst, Deutschland.