Anisotropic Thermal-Parameter Refinement of the DNA Dodecamer CGCGAATTCGCG by the Segmented Rigid-Body Method

By Stephen R. Holbrook

Laboratory of Chemical Biodynamics and Department of Chemistry, University of California, Berkeley, CA 94720, USA

Richard E. Dickerson

Molecular Biology Institute, University of California, Los Angeles, CA 90024, USA

and Sung-Hou Kim

Laboratory of Chemical Biodynamics and Department of Chemistry, University of California, Berkeley, CA 94720, USA

Abstract

A structure-factor least-squares refinement of the deoxyoligonucleotide (CGCGAATTCGCG)$_2$ has been conducted using a model with constraints and restraints on the positional parameters and a segmented rigid-body representation for the anisotropic temperature factors. The macromolecule was divided into subgroups each of which was treated as a rigid body in terms of both positional and thermal parameters. For each subgroup, the thermal parameters determined were elements of translation, libration and correlation (TLS) tensors. This segmented rigid-body model of thermal motion has not previously been applied to the refinement of a macromolecular crystal structure. The anisotropic thermal-parameter refinement has significantly reduced the classical R factor as judged by the Hamilton test. The resulting difference Fourier map has a considerably lower noise level allowing fifteen additional low-occupancy water positions to be identified. In addition, analysis of the anisotropic thermal parameters has revealed new information about the local mobility of the groups in the oligonucleotide. Thus, the method of segmented rigid-body anisotropic temperature-factor refinement appears to be uniquely suited to macromolecules, especially nucleic acids, where high-resolution data are usually unavailable.

Introduction

The combination of high thermal motion and static disorder in the loosely packed unit cells of macromolecular crystals usually limits the resolution of observable X-ray diffraction. Generally, individual atoms are not completely resolved in macromolecular electron density maps, but can be identified by their locations within resolved groups of atoms. During crystallographic refinement of a model structure, the lack of high-resolution data and the large percentage of weak reflections results in a low ratio of the number of observations to parameters to be refined. To overcome the effects of errors in the observations and approximations made in data reduction this overdetermination ratio must necessarily be as large as possible for accurate structural determination.

The number of parameters to be refined for a macromolecule can be reduced by incorporating information about the geometry of the constituents of the macromolecule obtained from high-resolution crystal structures of these constituents. Three approaches have been used to incorporate this information during refinement of the positional parameters: (1) constraints, (2) restraints, and (3) combined constraints and restraints. The first method uses rigid constraint conditions, where bond distances, bond angles, and some dihedral angles are maintained strictly at canonical values (Scheringer, 1963). A refinement procedure where all positional parameters, except flexible dihedral angles, were constrained has been developed for the refinement of protein structures (Diamond, 1971). In the second method, information about the molecular geometry of the constituents is incorporated by non-rigid restraint conditions. These conditions are enforced by least-squares minimization of the difference in distance between observed and canonical values specifying bond lengths, angles, and any fixed torsion angles within the constituents. This least-squares minimization is done in conjunction with structure-factor least-squares refinement and the deviation of geometry from canonical values is adjusted by varying the ratio of weights applied (Konnert, 1976; Konnert & Hendrickson, 1980). The third method, which has been referred to as *CORELS* (COnstrained-REstrained Least Squares), uses a combination of

constraints and restraints in order to achieve the advantages of both methods (Sussman, Holbrook, Church & Kim, 1977). The combination of both constraints and restraints allows sufficient flexibility to permit rapid convergence while keeping the number of variable parameters at a minimum.

A similar variety of approaches have been used for refinement of the *isotropic* thermal parameters of macromolecules. For structures where high-resolution data are available, independent isotropic atomic thermal parameters may be assigned and refined. This has been done for rubredoxin at 1·2 Å resolution (Watenpaugh, Margulis, Sieker & Jensen, 1978) and the Z-DNA hexanucleotide, CGCGCG, at 0·9 Å resolution (Wang, Quigley, Kolpak, Crawford, van Boom, van der Marel & Rich, 1979) as well as other structures.

In most cases, however, atomic resolution data are not available and either constraint or restraint conditions must be applied to the thermal parameters for a meaningful refinement. Hendrickson & Konnert (1979) and Konnert & Hendrickson (1980) have developed a widely used procedure in which the isotropic atomic temperature factors are restrained to vary smoothly along the chain of covalently bonded atoms. An alternate approach, which has been used successfully in several crystallographic refinements (Sussman, Holbrook, Warrant, Church & Kim, 1978; Shoham, Yonath, Sussman, Moult, Traub & Kalb, 1979; Shakked, Rabinovich, Kennard, Cruse, Salisbury & Viswamitra, 1983), is to constrain rigid groups of atoms within the macromolecule to a single isotropic temperature factor. This is in close analogy to the use of constraints on the positional parameters in the CORELS procedure.

For *anisotropic* thermal-parameter refinement, Konnert & Hendrickson (1980) have extended their restraint procedure to allow refinement of the parameters of the anisotropic thermal ellipsoids of the individual atoms. This method results in a very large number of parameters to be refined, although the 'effective' number of variables is reduced by the restraint conditions. This procedure has been used in the refinement of the carp muscle calcium binding protein (Konnert & Hendrickson, 1980).

In this paper we introduce a new approach to determination of *anisotropic* parameters of thermal motion for macromolecules. In analogy to the use of constrained rigid bodies in positional and isotropic thermal refinement we have adapted the 'TLS', or rigid-body, model of anisotropic thermal motion originally developed for small molecules to the refinement of macromolecules. In the 'TLS method' (Schomaker & Trueblood, 1968), the motion of a rigid body is described by three tensors: T, L, and S which correspond to translation, libration, and their average quadratic correlation respectively. The T and L tensors are symmetric 3×3 matrices and the S tensor

is in general a 3×3 non-symmetric matrix when an arbitrary origin is specified. A shift in origin changes the S and T matrices while L is left unchanged. The anisotropic thermal parameters of individual atoms can be calculated from these 'TLS' tensors. These parameters reflect only the rigid-body motions of translation and libration and neglect the effect of uncorrelated internal vibrations of the atoms in the group. Since the T and L tensors are symmetric, there are 21 tensor elements per rigid body, of which 20 are independent parameters. As there are 6 anisotropic thermal parameters needed per atom in the atomic ellipsoid representation, a rigid group of four or more atoms in the 'TLS' model will be specified by fewer thermal parameters. The reduction in the total number of thermal parameters in the TLS model will increase linearly with the size of the rigid subgroup. This is precisely what is necessary for the determination of anisotropic thermal parameters in macromolecules.

We have, therefore, incorporated refinement of the parameters of the TLS tensors into the CORELS procedure for refinement of large structures. The structure is divided into a series of constrained groups (which usually corresponds to the constrained groups used in positional refinement) to be *anisotropically* refined in terms of their TLS thermal parameters. We refer to this division of constrained groups as the *segmented rigid-body model of thermal motion.*

Using this method, we have refined the anisotropic thermal parameters of the self-complementary deoxydodecanucleotide CGCGAATTCGCG. This crystal structure was previously solved and refined with isotropic thermal parameters at 1·9 Å resolution (Wing, Drew, Takano, Broka, Takana, Itakura & Dickerson, 1980). We chose this structure for the initial trial of the segmented rigid-body method for anisotropic refinement because of its biological importance as the first example of a B form DNA double helix, the unavailability of atomic resolution data, and the 'natural' division of oligonucleotides into subgroups which can be treated as rigid bodies.

Methods

A detailed discussion of the CORELS method of structure-factor least-squares refinement using constrained and restrained positional parameters and constrained group isotropic thermal parameters has previously been published (Sussman, Holbrook, Church & Kim, 1977). The quantity minimized in this procedure is a sum of terms containing the structure-factor differences and the deviations of the restrained distances (between constrained groups) from ideal values:

$$Q = W_f \sum_h [W_h (F_{o,h} - F_{c,h})^2]$$
$$+ W_d \sum_l [W_l (D_{o,l} - D_{c,l})^2]. \qquad (1)$$

In the above equation F_o and F_c are the observed and calculated structure factors of the reflections (**h**). D_o and D_c are the observed (canonical) and calculated distances between atoms for each of the restraint conditions. As the structure is divided into a series of constrained groups, within which all bond distances and angles are fixed, the normal-equation matrix can be constructed of derivatives of the group positional and thermal parameters obtained according to the chain rule:

$$(\partial F_c/\partial P_n)_B = \sum_a \sum_b [(\partial F_c/\partial X_{a,b})(\partial X_{a,b}/\partial P_n)] \quad (2)$$

$$(\partial F_c/\partial P_n)_X = \sum_a [(\partial F_c/\partial B_a)], \quad (3)$$

where a is an index over all the atoms in a group or subgroup, b runs from 1 to 3 and n is an index referring to the group parameters of which there are six for group positions and one per isotropic subgroup.

The chain rule as stated above can also be used to formulate derivatives of the constrained groups in the **TLS** model of anisotropic motion (Pawley, 1964, 1966).

$$(\partial F_c/\partial P_{\text{tls}}) = \sum [\sum (\partial F_c/\partial U_{ij})(\partial U_{ij}/\partial P_{\text{tls}})]. \quad (4)$$

From the parameters of the **T**, **L** and **S** matrices which are the variables refined *versus* the structure-factor data, the individual atomic anisotropic temperature factors can be calculated (Schomaker & Trueblood, 1968) for purposes of structure-factor calculation or graphic representation by a program such as *ORTEP* (Johnson, 1965).

$$U_{ij} = (G_{ijkl})(L_{kl}) + (H_{ijkl})(S_{kl}) + (T_{ij}). \quad (5)$$

The tensors G_{ijkl} and H_{ijkl} are of the form given by Schomaker & Trueblood (1968) and are functions of the atomic coordinates. The partial derivatives of the above equation are used to form the product derivatives used in the least-squares matrix.

$$(\partial U_{ij}/\partial T_{mn}) = \delta_{im}\delta_{jn} \quad (6)$$

$$(\partial U_{ij}/\partial L_{mn}) = G_{ijmn} \quad (7)$$

$$(\partial U_{ij}/\partial S_{mn}) = H_{ijmn}. \quad (8)$$

The above equations have been incorporated into the *CORELS* reciprocal-space least-squares refinement program so that the structure being refined may be divided into any number of rigid subgroups for the purpose of anisotropic (**TLS**) thermal refinement. In order to represent best the internal flexibility within the DNA dodecanucleotide refined here, the structure was divided into subgroups as illustrated in Fig. 1. The segmentation used for the **TLS** refinement was not arbitrary, but was that which best reproduced the observed atomic thermal ellipsoids in a test structure, deoxycytidine 5'-phosphate (Viswamitra, Reddy, Lin & Sundaralingam, 1971). The details of this partition-

ing test have been discussed elsewhere (Holbrook & Kim, 1984).

Rigid-body refinement of the DNA dodecamer

Crystals of the self-complementary DNA dodecanucleotide CGCGAATTCGCG (space group $P2_12_12_1$, $a = 24\cdot87$, $b = 40\cdot39$, $c = 66\cdot20$ Å) contain 486 non-hydrogen atoms and 80 ordered water molecules per asymmetric unit (Wing, Drew, Takano, Broka, Takana, Itakura & Dickerson, 1980; Dickerson, Drew & Conner, 1981). A total of 5691 data have been collected to a resolution of $1\cdot9$ Å. The structure has been previously refined by the procedure of Jack & Levitt (1978) which simultaneously minimizes conformational energy and structure-factor differences. This method applies no geometrical constraints, but the energy minimization is equivalent to using a restrained geometry. The total number of parameters used in this refinement, including the individual atomic temperature factors, was therefore four times the number of atoms or 2264. The ratio of structure-factor data to parameters varied was less than $2\cdot5$ for this model. One problem with refinement of the unrestrained individual atomic thermal parameters is that at this resolution the atoms are not resolved and the electron density is smeared over the entire group. It is not surprising then that a significant degree of random noise is introduced into these values.

Calculation of the classical R factor from the atomic isotropic model refined by the Jack–Levitt
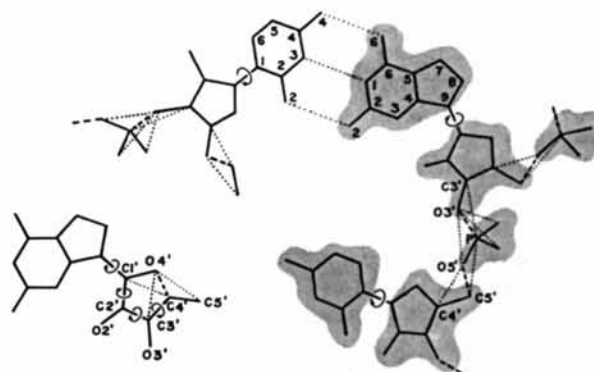

Fig. 1. Definition of the constraints and restraints used in the positional- and thermal-parameter structure-factor least-squares refinement. The dotted lines represent restraints specifying bond angles and hydrogen-bond distances, the dashed lines restraints representing covalent bonds. Groups of atoms connected by solid lines are treated as rigid (constrained) bodies. Variable dihedral angles within constrained groups are signified by arrows drawn around bonds. The sugar-pucker flexibility is handled by a combination of restraints and variable torsion angles within the ribose as shown in the inset in the lower left of the figure. The thermal subgroups for both isotropic and anisotropic rigid group refinement are shown by separate, shaded regions. The O(3') atom is arbitrarily included in the ribose and the O(5') in the phosphate group.

method was 26·4% for the entire 5691 data. In order to follow as closely as possible the previous refinement of the dodecamer structure for purposes of comparison, we decided to use only the data of resolution higher than 8 Å. This set of data contained 5599 reflections and the initial R value for the atomic isotropic model was 25·2%.

For the *CORELS* isotropic and anisotropic refinement the dodecanucleotide model was divided into phosphate, ribose and base subgroups as discussed under *Methods* and illustrated in Fig. 1. Each subgroup was initially assigned a group isotropic temperature factor corresponding to the average of the atomic isotropic temperature factors of the atoms in the group. These group isotropic temperature factors were then refined by the structure-factor least-squares method while holding the positional parameters constant. The 80 water molecules previously located were included in this isotropic refinement as O atoms, with individual isotropic thermal parameters. Unit weights were applied to the structure factors for the entire course of the refinement. After seven cycles of least-squares refinement all of the isotropic thermal parameters had converged. One water molecule was removed from the model because its temperature factor exceeded 250 Å^2, an arbitrary cutoff point.

Next, the positional parameters of each constrained group were refined for five cycles subject to the constraint and restraint conditions shown in Fig. 1. Each constrained group had 3 variable translational parameters, 3 variable rotational parameters, and the number of variable internal dihedral angles indicated in Fig. 1. The restraints specifying base–base hydrogen bonds were assigned zero weight so as not to impose undue restrictions. The positions of the water molecules were held fixed.

The group isotropic thermal parameters which had not been varied during positional refinement were then allowed to refine for an additional three cycles (holding positions constant). At the end of these cycles one additional water molecule was removed from the model due to a temperature factor exceeding 250 Å^2, leaving a total of 78 waters.

Finally, the constrained anisotropic TLS thermal parameters were varied, while holding the positional parameters constant. Again each phosphate, ribose, *and base was considered a rigid group with 20* independent variable thermal parameters or a total of 1400 parameters for the entire dodecamer. A total of eleven cycles resulted in convergence of both the thermal-parameter shifts and the R factor. Several additional cycles of isotropic refinement of the water temperature factors were conducted during the anisotropic refinement of the dodecamer itself. In retrospect, the same result for the anisotropic refinement may have been obtained in fewer cycles, but we utilized only fractions of the calculated shifts in order to assure convergence.

Table 1. *Refinement results*

Model	No. of atoms	No. of data	No. of variables	R	Correlation coefficient
(A) Brookhaven (atomic isotropic)*					
All data	566	5691	2264	28·6%	0·8983
(No data < 8Å)	566	5599	2264	25·2	0·9604
(No data < 2σ)	566	2728	2264	19·5	0·9592
(B) CORELS group isotropic†					
(No data < 8Å)	564	5599	149	25·3	0·9619
(C) CORELS group anisotropic‡					
(No data < 8Å)	563	5599	1478	20·6	0·9776
(No data < 2σ)	563	2728	1478	14·9	0·9785

* The structure factors used were calculated from the coordinates and compared to observed amplitudes deposited at the Brookhaven Protein Data Bank.

† The starting coordinates were those deposited in the Brookhaven Protein Data Bank. The group isotropic thermal parameters and rigid-body positional parameters were refined in alternate cycles as indicated.

‡ The starting values of the 'TLS' parameters were calculated from the final group isotropic B values. During this step the water molecules were refined isotropically. The values shown are the results of several cycles of anisotropic refinement.

Results

The results of the isotropic and anisotropic thermal refinements are summarized in terms of number of variables, observations, R factors, and correlation coefficients in Table 1. Fig. 2(a,b) shows the R factor as a function of both $(\sin\theta)/\lambda$ and amplitude for both the isotropic and the anisotropic models. These plots show improved agreement of F_o with F_c for all resolution blocks except the very highest resolution. This is probably due to the low quality of data available at high resolution where many F_o values are very weak.

The final group isotropic B values are given in Table 2. The R factor of 25·3% for the model with group isotropic thermal parameters is not significantly different from that of 25·2% calculated using the individual atomic isotropic temperature factors despite the fact that the former model has fewer than 10% of the variable thermal parameters of the latter. The average atomic isotropic results agree with those obtained here for group isotropic temperature factors in that the phosphates have larger B values than the riboses which in turn have larger B values than the bases, but the ratios of the average atomic isotropic values differ from the ratios of the group isotropic values we have determined.

The constrained–restrained positional refinement resulted in a small average shift in position of 0·11 Å (r.m.s. shift 0·15 Å). The only positional shifts greater than 0·5 Å were for C(5') of cytosine 1 (0·88 Å), O(5') of the phosphate of residue 2 (0·63 Å), and C(5') of guanosine 10 (0·52 Å). The sugar pucker remained quite similar to that observed in the model refined by the Jack–Levitt method, with many of the riboses adopting the C(1')-*exo* conformation. We tried a variety of weights for the restraint conditions down to the point of nearly relaxing them completely and

Table 2. *Final isotropic B values* (Å^2)

| Residue	B atomic isotropic			B group isotropic			B equivalent		
	Phosphate	Ribose	Base	Phosphate	Ribose	Base	Phosphate	Ribose	Base
Strand one									
Cyt 1		42·6	29·6		40·5	27·2		45·2	34·0
Gua 2	58·8	36·8	22·0	75·0	36·0	19·1	100·3	31·5	24·2
Cyt 3	47·9	41·6	24·0	43·9	35·3	21·6	50·0	38·2	34·0
Gua 4	44·0	46·0	28·6	39·5	46·0	25·3	46·5	44·7	38·0
Ade 5	53·7	47·9	24·0	56·8	39·1	21·8	70·3	51·3	32·8
Ade 6	42·2	36·8	20·2	38·5	29·9	18·2	42·9	35·3	22·1
Thy 7	41·3	33·8	24·8	41·8	30·8	23·1	41·5	35·4	27·6
Thy 8	44·6	37·4	25·5	51·5	34·6	25·7	59·0	38·4	28·7
Cyt 9	51·7	50·3	31·9	51·3	48·7	27·9	68·1	45·6	34·3
Gua 10	60·1	51·0	32·0	63·8	53·9	25·5	86·1	55·2	31·8
Cyt 11	58·6	47·0	25·5	57·4	45·4	24·0	61·7	44·8	27·9
Gua 12	52·2	45·3	36·5	62·7	42·0	34·4	62·3	45·5	38·4
Strand two									
Cyt 1		49·6	43·6		60·4	46·2		66·4	49·3
Gua 2	53·7	44·2	32·6	69·0	36·2	28·3	96·5	46·1	33·3
Cyt 3	43·4	34·4	27·3	39·8	27·0	23·0	46·5	29·5	26·0
Gua 4	49·7	36·0	22·2	43·1	30·2	20·9	50·0	31·6	25·5
Ade 5	45·2	45·6	30·2	45·0	41·0	23·8	50·8	44·6	36·9
Ade 6	44·9	47·8	21·9	43·1	45·4	18·4	54·2	47·4	23·7
Thy 7	49·2	44·9	29·1	49·0	37·7	24·1	62·4	40·6	31·4
Thy 8	44·9	39·7	27·4	45·0	32·2	25·0	52·0	34·0	27·7
Cyt 9	56·2	50·9	28·2	53·8	42·6	28·5	62·6	41·5	35·2
Gua 10	56·9	45·0	25·4	55·3	45·2	22·5	72·0	54·0	30·1
Cyt 11	50·6	49·3	28·7	47·0	47·8	23·6	59·6	43·0	32·8
Gua 12	46·4	44·2	24·4	51·1	37·4	22·0	53·8	42·8	30·0
Averages	50 (5)	44 (5)	28 (5)	51 (10)	40 (8)	25 (6)	61 (16)	43 (8)	31 (6)
Ratios	1·80	1·57	1·00	2·04	1·61	1·00	1·95	1·36	1·00

(a)

(b)
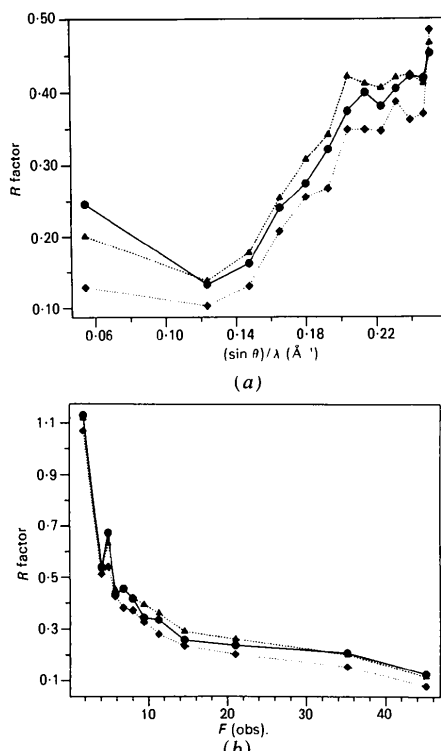
Fig. 2.(a) The classical R factor is plotted *versus* (sin θ)/λ for blocks containing approximately equal numbers of data sorted according to the diffraction angle. The plots are based on data between 1·9 and 8·0 Å resolution and include data of intensity less than 2σ. The diamonds are from the anisotropically refined data set, the triangles are from the data after group isotropic and positional refinement, and the circles plot the original structure-factor data as deposited in the Protein Data Bank. (b) The R factor *versus* average amplitude of diffraction intensity. The symbols are as in (a).

the results of the positional refinement were generally unchanged.

The final R value for the anisotropic refinement was 20·6% when the data less than 2σ were included and 14·9% when these reflections were omitted. Application of the analytical form of the Hamilton R-factor test (Pawley, 1970) shows that the final R value obtained by anisotropic TLS refinement is significantly improved from both the starting atomic isotropic model ($R = 25\cdot2\%$) and the group isotropic model ($R = 25\cdot3\%$) at the 99·99% confidence level. We are therefore justified in drawing conclusions about the anisotropic modes of thermal vibration from the results of the TLS refinement.

Individual anisotropic thermal parameters were calculated for each atom from the final TLS matrices using equation (5). The equivalent B values obtained by averaging over the diagonal elements of the U_{ij} for each atom and then over all atoms in the subgroup are shown in Table 2 and may be compared to their isotropic counterparts.

A stereo *ORTEP* drawing of a single base pair of the double-helical molecule is shown in Fig. 3(a-c) with the electron density distribution according to: (a) the individual atom isotropic model, (b) the constrained group isotropic model, and (c) the segmented rigid-body anisotropic model. It can be seen that the group isotropic model (b) reproduces well the atomic isotropic model (a) even though it is determined by far fewer parameters. The amplitudes of the atomic ellipsoids determined from the segmented rigid-body anisotropic model (c) also follow the trend observed for the two isotropic models, (a) and (b), supporting the validity of this method.

Another useful way of analyzing the results of the anisotropic refinement is by examining the eigenvectors and eigenvalues of the libration and translation matrices. An example of these eigenvectors is shown in Fig. 4(a,b) together with the atomic thermal ellipsoids as calculated from the **TLS** tensors. Some general tendencies for local mobility in the double helix of this dodecamer can be discerned by this method.

A detailed discussion of our interpretation of the mobility of the subgroups in the crystal structure of (CGCGAATTCGCG)$_2$ in terms of the eigenvectors and eigenvalues of translation and libration has been published elsewhere (Holbrook & Kim, 1984).*

Difference Fourier maps

To assist us in evaluation of the rigid-body anisotropic refinement, we have computed difference Fourier maps using structure factors calculated from the model with atomic isotropic thermal parameters (isotropic map) and from the model with anisotropic thermal parameters (anisotropic map). The critical factor of the difference maps, the root-mean-square electron density or background noise level, has been reduced in the anisotropic map by 22% from 0·065 to 0·051 e Å^{-3}. The largest peak in the isotropic map is 0·342 e Å^{-3} while that of the anisotropic map is only 0·232 e Å^{-3}. It is clear from these values that the anisotropic refinement has improved the quality of the map considerably and thereby provided information unavailable in the isotropic map (see below).

Representative sections of the two maps are shown in Fig. 5(a,b). These maps have been contoured at the same level for comparison. In Fig. 5(b) the major peak labelled 'A' in the anisotropic map also appears in the isotropic map. However, in the isotropic map the higher overall noise level obscures the possible significance of this peak. From the fact that this peak is retained and strengthened *versus* the background in the anisotropic map and from its distance to the hydrogen-bond-forming atoms O(1P), O(5') and water, this peak may be identified as due to a low-occupancy bound water molecule. From such a search we have located 15 additional 'water' positions from the 50 highest peaks in the anisotropic map which are at proper distances from hydrogen-bond-forming atoms in the existing model. Such information is characteristic of the improved map from the anisotropic refinement.

---

* Atomic coordinates, temperature factors and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 7BNA, R7BNASF), and are available in machine-readable form from the Protein Data Bank at Brookhaven or one of the affiliated centers at Cambridge, Melbourne or Osaka. The data have also been deposited with the British Library Lending Division as Supplementary Publication No. SUP 37014 (microfiche). Free copies may be obtained through The Executive Secretary, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.
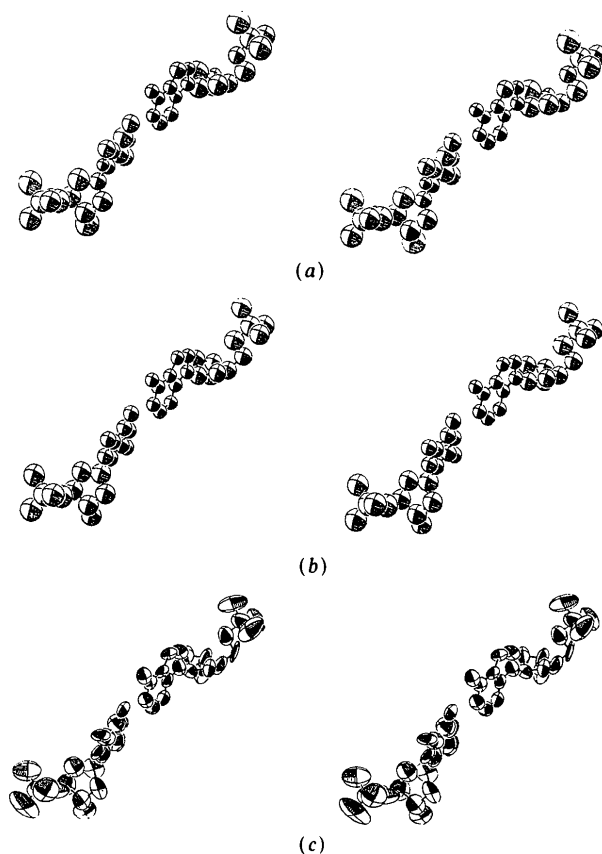

(a)

(b)

(c)

Fig. 3.(a) A stereoview of a typical base pair (A6:T19) of the dodecamer CGCGAATTCGCG with the temperature factors calculated from the individual atom isotropic model. (b) A stereodrawing of the same base pair as in (a) with all atoms within a group constrained to one group isotropic thermal parameter as calculated by *CORELS*. (c) A stereodrawing of the atomic thermal ellipsoids of the base pair of (a) and (b) as calculated from the parameters of the TLS matrices of the subgroups. The parameters of the TLS matrices were determined from the diffraction data using the program *CORELS*.
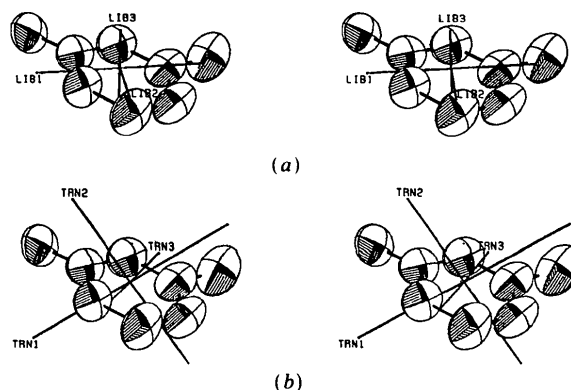

(a)

(b)

Fig. 4.(a) A stereodrawing of a nucleotide base (cytosine 15) with solid lines indicating the principal eigenvectors of libration as extracted from the TLS matrices. The length of the eigenvectors is proportional to the eigenvalue of libration around that axis. (b) A stereodrawing of a nucleotide base (cytosine 15) with the eigenvectors of translation indicated by solid lines. The length of the eigenvectors is proportional to the magnitude of the translational eigenvalue.

Examination of the anisotropic difference Fourier map shows no significant peaks occurring at the location of the two water molecules which were removed during our refinement due to a shift in isotropic temperature factor above 250 $Å^2$. This justifies our removal of these positions in the list of bound water molecules.

Discussion

We have demonstrated that the **TLS** method of representing rigid-body anisotropic motion is useful in the refinement of a macromolecular structure. Significantly improved agreement between observed and calculated structure factors and new information about the anisotropy of the 'thermal' motion of subunits in the structure has been obtained from this refinement. By combining the **TLS** parameters with the use of constraints and restraints on the positional parameters we have tried to maximize the amount of information obtained from limited available data.

Larger structures, such as transfer RNA, for which only medium-resolution data are available (2·7 Å) may be refined anisotropically by defining the thermal subgroups as entities larger than was done for the dodecanucleotide where relatively high-resolution data were collected. Even large structural elements such as the double-helical stems of transfer RNA, α helices in proteins, or even entire protein domains could be specified as rigid units when only low-resolution data are available, or used as an intermediate model to extend the phasing to higher resolution (Sussman & Podjarny, 1983).

The segmented rigid-body method of anisotropic refinement seems especially well suited to oligo- and polynucleotide crystal structures because they can be conveniently subdivided into relatively rigid subgroups (phosphate, ribose, base) which can be expected to undergo concerted translational and librational motion with little independent atomic anisotropic motion. The side-chain flexibility of many of the amino acids of proteins inherently allows for more independent atomic motion by rotation about single bonds. At low to medium resolution this may not be a problem if the thermal subgroups are chosen appropriately, but at high resolution it may be necessary to introduce additional parameters to allow for libration around internal axes of torsional freedom (Dunitz & White, 1973; Trueblood, 1978). We are considering this modification for future work.
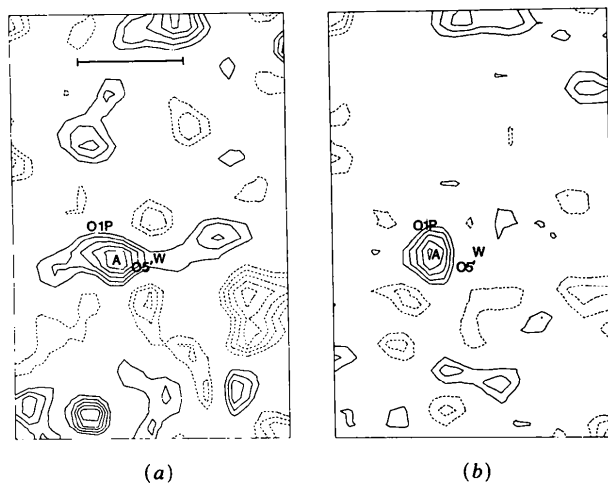
Fig. 5.(a) A section of the difference Fourier map calculated from the model of the dodecamer with atomic isotropic thermal parameters. The view is down the Z crystallographic axis showing the XY plane. Areas with electron density greater than $0·1$ e $Å^{-3}$ are contoured, with subsequent contour lines at $0·04$ e $Å^{-3}$. The scale of the map is indicated by the bar near the top of the figure which corresponds to 5 Å. The distances of peak 'A' from the potential hydrogen-bonding atoms $O(1P)$, $O(5')$ (both of cytidine 15), and W (a previously located water ligand) are 3·0, 3·4, and 2·5 Å respectively. (b) The same section of the difference Fourier map as in (a) calculated from the dodecamer model with anisotropic thermal parameters. The view and contour level are the same as in (a).

References

DIAMOND, R. (1971). *Acta. Cryst.* A27, 436–452.

DICKERSON, R. E., DREW, H. R. & CONNER, B. N. (1981). *Biomolecular Stereodynamics*, Vol. 1, edited by R. H. SARMA, pp. 1–34. New York: Adenine Press.

DUNITZ, J. D. & WHITE, D. N. J. (1973). *Acta Cryst.* A29, 93–94.

HENDRICKSON, W. A. & KONNERT, J. H. (1979). *Biomolecular Structure, Conformation Function and Evolution*, Vol. I, edited by R. SRINIVASAN, pp. 43–57. New York: Pergamon Press.

HOLBROOK, S. R. & KIM, S.-H. (1984). *J. Mol. Biol.* 173, 361–388.

JACK, A. & LEVITT, M. (1978). *Acta Cryst.* A34, 931–935.

JOHNSON, C. K. (1965). *ORTEP.* Report ORNL-3794. Oak Ridge National Laboratory, Tennessee.

KONNERT, J. H. (1976). *Acta Cryst.* A32, 614–617.

KONNERT, J. H. & HENDRICKSON, W. A. (1980). *Acta Cryst.* A36, 344–350.

PAWLEY, G. S. (1964). *Acta Cryst.* 17, 457–458.

PAWLEY, G. S. (1966). *Acta Cryst.* 20, 631–638.

PAWLEY, G. S. (1970). *Acta Cryst.* A26, 691–692.

SCHERINGER, C. (1963). *Acta Cryst.* 16, 546–550.

SCHOMAKER, V. & TRUEBLOOD, K. N. (1968). *Acta Cryst.* B24, 63–76.

SHAKKED, Z., RABINOVICH, D., KENNARD, O., CRUSE, W. B. T., SALISBURY, S. A. & VISWAMITRA, M. A. (1983). *J. Mol. Biol.* 166, 183–201.

SHOHAM, M., YONATH, A., SUSSMAN, J. L., MOULT, J., TRAUB, W. & KALB, A. J. (1979). *J. Mol. Biol.* 131, 137–155.

SUSSMAN, J. L., HOLBROOK, S. R., CHURCH, G. M. & KIM, S.-H. (1977). *Acta Cryst.* A33, 800–804.

SUSSMAN, J. L., HOLBROOK, S. R., WARRANT, R. W., CHURCH, G. M. & KIM, S.-H. (1978). *J. Mol. Biol.* **123**, 607-630.

SUSSMAN, J. L. & PODJARNY, A. D. (1983). *Acta Cryst.* B**39**, 495-505.

TRUEBLOOD, K. N. (1978). *Acta Cryst.* A**34**, 950-954.

VISWAMITRA, M. A., REDDY, B. S., LIN, G. H.-Y. & SUNDARALINGAM, M. (1971). *J. Am. Chem. Soc.* **93**, 4565-4573.

WANG, A. H.-J., QUIGLEY, G. J., KOLPAK, F. J., CRAWFORD, J. L., VAN BOOM, J. H., VAN DER MAREL, G. & RICH, A. (1979). *Nature (London)*, **282**, 680-686.

WATENPAUGH, K. D., MARGULIS, T. N., SIEKER, L. C. & JENSEN, L. H. (1978). *J. Mol. Biol.* **122**, 175-190.

WING, R., DREW, H., TAKANO, T., BROKA, C., TANAKA, S., ITAKURA, K. & DICKERSON, R. E. (1980). *Nature (London)*, **287**, 755-758.

Real-Space Crystal Structure Solution.
Crystal and Molecular Structure of Laminarabiose Octaacetate, C28H38O19

BY S. PÉREZ AND C. VERGELATI

Centre de Recherche sur les Macromolécules Végétales CNRS,* BP 68, 38402 Saint-Martin D'Hères, France

AND V. H. TRAN†

Laboratoire de Chimie, Groupe Macromolécules Végétales, Département de Recherche Fondamentale, Centre d'Etudes Nucléaires, 85X, 38041 Grenoble, France

Abstract

The elucidation of a crystal structure on the basis of complete minimization of intramolecular and intermolecular energy of packing has been applied to laminarabiose octaacetate. [Crystal data: $M_r = 678.6$, orthorhombic, $P2_12_12_1$, $a = 10.379 (2)$, $b = 22.943 (7)$, $c = 14.599 (4)$ Å, $Z = 4$, $V = 3476.4$ Å^3, $D_m = 1.30 (1)$, $D_x = 1.297$ Mg m^{-3}, λ(Cu $K\alpha$) = 1.54178 Å, μ(Cu $K\alpha$) = 0.91 mm^{-1}, $F(000) = 1432$, room temperature, final $R = 0.095$ for 1660 observed independent reflexions.] The two D-glucose residues have the $^4C_1$ pyranose conformation and are β-(1 → 3) linked. The conformational angles φ and ψ at the glycosidic linkage have the values -81.1 and $134.8°$ respectively. The present work establishes in an unambiguous manner the foundation of conformational analysis theory applied to the elucidation of crystal structures. It appears that intermolecular interactions in the crystal can be treated in good approximation with the intramolecular potential functions. When coupled with appropriate X-ray data, crystalline conformations may be deduced without highly refined potential functions. As compared to what is faced in polymer crystallography, where the number of packing parameters is reduced, finding a good starting point may represent the essential difficulty associated with the proposed methodology. Once a reasonable location

has been found for the molecule, the minimization procedure allows one to handle a reasonable number of parameters and yet yields only a few possible structural models. At this stage, the best packing models have to be checked against X-ray structure amplitudes.

Introduction

The prediction of crystal structure on the basis of non-bonded potential energy has received relatively little attention (Williams, 1969; Coiro, Giglio & Quagliata, 1972; Simonetta, 1974; Dauber & Hagler, 1980). This method may be of special use in the case of small molecules (Zugenmaier & Sarko, 1972; Leser & Rabinovich, 1978) or for rigid molecules with few rotatable substituents (Williams, 1972; Ramachandran & Shamala, 1976) and in the area of polymer structure determination (Kitaigorodskii, 1973; Zugenmaier & Sarko, 1976; Smith & Arnott, 1978) where the lack of diffraction data makes it imperative to predict suitable structural models from the stereochemistry of the repeating units.

In the field of biological molecules, with the exclusion of globular proteins, it is becoming clear that, owing to either the poor quality or the dimensions of the single crystals which can be grown, only a small number of reflexions can be obtained experimentally. In this particular situation, neither conventional direct methods, nor the methods that have been developed for protein crystallography can be applied with success.

* Laboratoire Propre du CNRS, associé à l'Université Scientifique et Médicale de Grenoble, France.

† Present address: Laboratoire de Stockage et de Conservation des Denrées Alimentaires, INRA, Centre de Recherche de Nantes, Chemin de la Géraudière, 44072 Nantes, France.