Acta Cryst. (1996). D**52**, 299–314

DNA–Drug Refinement: a Comparison of the Programs *NUCLSQ*, *PROLSQ*, *SHELXL*93 and *X-PLOR*, Using the Low-Temperature d(TGATCA)–Nogalamycin Structure

GEERTRUI S. SCHUERMAN,[a,b] CAMILLA K. SMITH,[c] JOHAN P. TURKENBURG,[c] ANNE N. DETTMAR,[c] LUC VAN MEERVELT[a] AND MADELEINE H. MOORE[c]*

[a]*Laboratorium voor Macromoleculaire Structuurchemie, Department Scheikunde, Katholieke Universiteit Leuven, Celestijnenlaan 200F, B-3001 Leuven, Belgium,* [b]*Laboratorium voor Analytische Chemie en Medicinale Fysicochemie, Faculteit Farmaceutische Wetenschappen, Katholieke Universiteit Leuven, Van Evenstraat 4, B-3000 Leuven, Belgium, and* [c]*Department of Chemistry, University of York, Heslington, York, YO1 5DD, England. E-mail: moore@yorvic.york.ac.uk*

(Received 19 June 1995; accepted 1 September 1995)

Abstract

In an earlier study [Smith, Davies, Dodson & Moore (1995). *Biochemistry*, **34**, 415–425] the crystal structure of the d(TGATCA)–nogalamycin complex was determined to 1.8 Å and refined with *PROLSQ* to $R = 19.5\%$ against 4767 reflections with $F > 1\sigma(F)$. A low-temperature crystallographic study on this complex has now been performed. Native data collection at liquid-nitrogen temperature (120 K) improved the resolution to 1.4 Å. The structure has now been refined against these new diffraction data in the resolution range 8–1.4 Å using *NUCLSQ*, *PROLSQ*, *SHELXL*93 and *X-PLOR*, in order to determine to what extent the resulting DNA conformation and associated solvent structure would differ and to examine the suitability of these programs for the refinement of oligonucleotide structures. With the advent of more DNA–protein structure determinations, it is of interest to see how well the protein-refinement packages, *PROLSQ* and *X-PLOR*, and the small-molecule program, *SHELXL*93, are able to accommodate DNA. Comparisons are made between the dictionaries, weights and restraints used and the final models obtained from each program. Although the final R values, using all data in the resolution range 8.0–1.4 Å, from *PROLSQ* (22.8%), *SHELXL*93 ($R1 = 21.7\%$ after isotropic refinement) and *X-PLOR* (24.4%) are higher than the R value from the *NUCLSQ* refinement (21.2%), the root-mean-square deviations between the four final models are very small. Using this high-quality 8.0–1.4 Å data set neither the dictionary nor the refinement program leave an imprint on the final fully refined complex. Likewise, the helical parameters and backbone conformation including sugar-puckering modes are not influenced by the refinement procedure used. Although a different number of water molecules is found in each refinement, varying from 62 (*X-PLOR*) to 86 (*NUCLSQ*), the first hydration sphere is well conserved in all four models.

1. Introduction

Refinement of the crystal structures of biological macromolecules cannot be performed using the measured data alone when, owing to limited resolution of the data observed, the data:parameter ratio is insufficient. Therefore, the experimental data must be supplemented with restraints and/or constraints; including stereochemical, thermal and, where necessary, restraining non-crystallographic symmetry. Various refinement packages have important differences in terms of minimization algorithms used, schemes to restrain geometry, methods of weighting geometric to data terms, ways of scaling calculated and observed structure factors, modelling bulk solvent *etc.*, hence the results obtained may differ.

This may be a problem when analysing the fine structural details of macromolecules and is pertinent to X-ray studies on nucleic acids. For any sequence of bases along DNA, the global conformation of the double stranded form of the polymer is usually predefined in one of the helical families. Sequence-dependent structural variations provide the required specificity for DNA recognition by proteins and drugs. For example, it is of interest to look at the fine details of how different anthracycline antibiotics, like nogalamycin, exhibit sequence-preferential DNA binding. The aim of these studies is, by investigating the nature of non-covalent interactions, to interpret the importance of the different substituents on anthracyclines in relation to varying degrees of cytotoxicity and cardiotoxicity exhibited by these closely related molecules. Also solvent is believed to play an important role in determining the DNA conformation and its recognition characteristics (Otwinowski *et al.*, 1988; Jochimiak, Haran & Sigler, 1994; Gewirth & Sigler, 1995). The solvent structure is difficult to determine, even at atomic resolution, due to disorder displayed by individual water molecules.

A good quality data set should result in the correct structure whichever suitably modified refinement pro-

tocol is used. If the solvent structure is biologically relevant, it ought to be well ordered. Therefore, at least the same first-shell water structure should consistently be found. The quality and completeness of the low-temperature 1.4 Å data were ideal for a detailed comparison of different refinement programs, assessing the suitability of these protocols for the refinement of oligonucleotides by using the d(TGATCA)–nogalamycin complex as a test case. The four programs used were *NUCLSQ* (Westhof, Dumas & Moras, 1985), *PROLSQ* (Hendrickson & Konnert, 1981; Collaborative Computational Project, Number 4, 1994), *SHELXL*93 (Sheldrick, 1993) and *X-PLOR* (Brünger, 1992*a*). A comparison is made between the dictionaries used, the restraints applied and the final models. The suitability and specific features of each program are discussed.

In the field of DNA crystallography the most widely used refinement program is *NUCLSQ*, a restrained least-squares method for nucleic acids originally adapted from the protein-refinement program *PROLSQ*. As more DNA–protein structures are being determined, protein-refinement programs are adapted to accommodate nucleic acids. Hence, it is important to test that DNA structures refined using protein packages are directly comparable with oligonucleotide structures resulting from refinement using the more specific program *NUCLSQ*.

In the d(TGATCA)–nogalamycin complex, nucleotides are labelled in the 5′ to 3′ direction from T1 to A6 for strand 1 and T7 to A12 for strand 2. The two nogalamycins are labelled NOG1 corresponding to the drug intercalated at the d(C5pA6/T7pG8) step and NOG2 for the drug at the d(T1pG2/C11pA12) step. Atomic nomenclature used for nogalamycin is consistent, for the chromophore, with previous work (Williams *et al.*, 1990) and is given in Fig. 1.

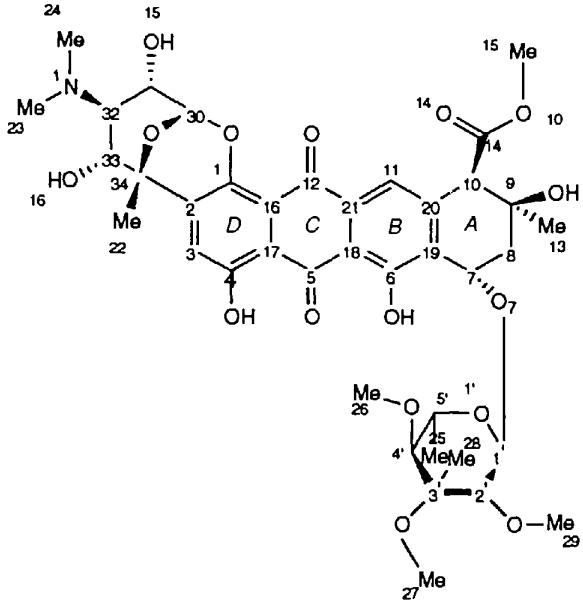


Fig. 1. Molecular formula and nomenclature of nogalamycin $C_{30}H_{49}NO_{16}$. The anthraquinoline chromophore comprises four fused rings labelled *A* to *D*. The nogalose sugar is attached to the only saturated ring *A*, and the aminoglucose sugar is fused to ring *D*.

2. Methods

2.1. *Crystallization and data-collection*

Nogalamycin was purchased from Sigma Chemicals Ltd. Synthesis and purification of the hexanucleotide d(TGATCA) were described previously (Smith *et al.*, 1995). Crystals were grown in sitting drops by vapour diffusion from an initial solution containing 1.5 m*M* oligonucleotide, 2.7 m*M* nogalamycin, 5% HMD (1,6-hexanediol), 1 m*M* spermine and 17 m*M* $MgCl_2$ in pH 6.5 sodium cacodylate buffer equilibrated against 200 µl 100% MPD. The complex crystallized in space group $P4_12_12$ with unit-cell dimensions $a = b = 37.29$ and $c = 71.12$ Å. One single crystal was used to collect a complete data set to 1.4 Å, using cryocrystallographic techniques, at the SRS, Daresbury, station PX9.6. These data were collected at a wavelength of 0.898 Å, using a 30 cm MAR research image plate. The crystal, of dimensions 0.3 × 0.2 × 0.1 mm, was transferred directly from the mother liquor to a cold nitrogen stream at a temperature of 120 K, using a rayon loop. The mother liquor contained a sufficient amount of HMD to act as a suitable cryoprotectant. Three data sets with different crystal-to-detector distances, exposure times and oscillation ranges were necessary in order to measure all reflections within the linear response of the detector. Data were processed using *DENZO* (Otwinowski, 1991) and reduced using the *CCP*4 programs *ROTAVATA* and *AGROVATA* (Collaborative Computational Project, Number 4, 1994). The quality of the data as a function of resolution is presented in Table 1.

2.2. *Refinement*

The starting model for each 1.4 Å refinement was identical to that used for the 1.8 Å d(TGATCA)–nogalamycin structure (Smith *et al.*, 1995), as built into the initial single isomorphous replacement with anomalous scattering (SIRAS) map. Oligonucleotide coordinates from the d(TGATCA)–epiadriamycin complex (Langlois d'Estaintot, Gallois, Brown & Hunter, 1992) were taken from the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977) and used as a guide to manually build the DNA in the initial map. In a similar way, coordinates for nogalamycin (Arora, 1983), obtained from the Cambridge Structural Database (Allen, Kennard & Taylor, 1983), were incorporated into the initial model. The temperature factors of the starting model for the d(TGATCA)–nogalamycin *NUCLSQ*, *PROLSQ* and *SHELXL*93 refinements reported here, were those of the coordinates from the database. In the *X-PLOR*

Table 1. *A summary of the data quality and completeness*

$$R = \sum_{hkl} |I - \langle I \rangle| / \sum_{hkl} \langle I \rangle.$$

Resolution (Å)	Reflections measured	Unique reflections	Percentage complete	Multiplicity %	$I > 3\sigma(I)$	R_{fac}	R_{cum}
4.14	4388	386	85.8	10.2	97.2	0.032	0.032
3.03	7575	644	99.1	11.3	98.5	0.033	0.033
2.51	7717	800	100.0	9.6	97.6	0.033	0.033
2.19	5187	916	99.7	5.5	97.8	0.030	0.033
1.96	5742	1043	99.4	5.5	97.3	0.039	0.033
1.80	6291	1126	99.1	5.5	92.7	0.046	0.033
1.67	6764	1230	98.9	5.6	91.4	0.054	0.034
1.56	7108	1300	98.2	5.6	89.4	0.071	0.034
1.47	7467	1395	98.1	5.5	81.3	0.124	0.035
1.40	4641	1466	78.7	4.7	80.8	0.213	0.036

refinement, where positional parameters and B values are refined at separate stages, global temperature factors were initially estimated for base atoms (10 Å^2), sugar atoms (15 Å^2), phosphate atoms (20 Å^2) and drug atoms (12 Å^2).

For cross-validation purposes, the entire data set (9852 reflections) was separated into a working set containing 90% (8867) and a reference set containing 10% of the reflections (984), randomly sampled throughout the resolution range. Refinement against F was performed with the 90% working set of the 8–1.4 Å resolution data using $F > 2\sigma(F)$ (Table 2). Applying the low-resolution cut-off from ∞ to 8 Å exluded approximately 50 reflections and the 2σ cut-off resulted in the omission of a further 200 reflections. No $\sigma(F)$ cut-off was applied in *SHELXL*93 as this would reduce the advantage of refining against F^2. In each refinement the same reference set was used to monitor R_{free} values as a diagnostic for overfitting of the observations (Brünger, 1992*b*, 1993). Once the overall R and R_{free} values had converged, refinement was completed using 100% data (9852 reflections) with no $\sigma(F)$ cut-off applied.

NUCLSQ, *PROLSQ* and *SHELXL*93 use restrained least-squares methods (Hendrickson & Konnert, 1981). *X-PLOR* uses simulated annealing and conformational search procedures to minimize a combined energy function, consisting of an X-ray data term and an empirical energy term, alongside conjugate-gradient methods.

In order to be able to compare the different restraints used in the programs, the overall weighting of these restraints has to be considered. The weighting schemes are quoted for each program separately in the sections below.

In all of the programs covalent bond lengths, bond angles and chiral centres were restrained to specified target dictionary values. R.m.s. differences between the different dictionary target bond lengths were insignificant; with an average of 0.012 Å. *SHELXL*93 has the added possibility to restrain values to be similar without having to specify a target value. This is very useful for oligonucleotides, as good distance restraints are available for the bases, but for the sugars and the phosphates it is preferable to assume that chemically equivalent 1,2 and 1,3 distances are equal without specifying actual target values. In this way the effect of pH on the protonation state of the phosphates and hence the P—O distances do not need to be predicted. In all programs, planarity restraints for the bases and the aromatic systems of the nogalamycin were applied, whereas all other torsion angles remained unrestrained.

In the *NUCLSQ*, *PROLSQ* and *X-PLOR* refinements, solvent positions were identified using the *CCP*4 peak search program *WATPEAK*. *SHELXL*93 automatically picks peaks for possible solvent atoms from difference-density calculations using the criteria that at least one hydrogen bond is made to an electronegative atom and that the distance to its nearest neighbour is not less than 2.3 Å. From these peaks, listed by *WATPEAK* or *SHELXL*93, solvent molecules were selected by careful inspection of the electron-density maps. In this procedure, the following strict criteria for the identification and confirmation of water sites were applied: (i) the density distribution around the peak in the $F_o - F_c$ map had to be spherical; (ii) the peak height needed to be greater than 3.5 standard deviations in the initial stages; (iii) plausible hydrogen-bonding partners with reasonable geometry had to lie within 2.3 and 3.3 Å; and (iv) an acceptable thermal parameter, *i.e.* less than 50 Å^2, had to be obtained upon subsequent refinement. If such water molecules were not in at least 2σ $2F_o - F_c$ density after refinement, they were removed. In the later stages, the contour levels were reduced to 3σ for the $F_o - F_c$ map and 1.25σ for the $2F_o - F_c$ map. Only solvent sites which were regarded as fully occupied, were included in the refinement, as it is these ordered water molecules especially that are expected to be conserved throughout.

In all refinement procedures water molecules were added in small numbers, as it was seen that adding large numbers created extra density for waters that were not necessarily correct *i.e.* either the $2F_o - F_c$ density was not apparent or the B values increased to > 80 Å^2 on subsequent refinement, or both. As many as 20–25 waters were included in the first addition. The number of solvent molecules included decreased as refinement

Table 2. *A summary of residuals, data: parameter ratios and final temperature factor information for the NUCLSQ, PROLSQ, SHELXL93 and X-PLOR refinements*

R_{cryst} and R_{cryst_w}, are the respective R factors calculated using 90% data, (i) after positional and temperature-factor refinement before the addition of solvent, and (ii) after positional and temperature-factor refinement including all possible water molecules; R_{anis}, is the R value after anisotropic temperature-factor refinement in *SHELXL93*, using 90% data. R_{all} is the final R value calculated using 100% data (9852 reflections). R_{free} is the final free R value calculated using 10% data (967 $2\sigma F$ reflections). The total number of non-H atoms includes the atoms of the double conformations where they apply. Min. $B_{complex}$, Max. $B_{complex}$, Ave. $B_{complex}$, are the minimum, maximum and average temperature factors of the DNA-drug complex atoms excluding solvent; Ave. B_{water}, is the average temperature factor for solvent alone.

	NUCLSQ	*PROLSQ*	*SHELXL93*	*X-PLOR*
R_{cryst} (%)	30.2	31.4	29.4	31.9
R_{cryst_w} (%)	20.9	22.4	21.7	24.1
R_{anis} (%)	—	—	15.6*	—
R_{all} (%)	21.2	22.8	16.0*	24.4
R_{free} (%)	28.1	27.4	26.9	27.0
$R_{free-anis}$ (%)	—	—	25.2*	—
R_{free}-R_{cryst_w} (%)	7.2	5.0	5.2	2.9
$R_{free-anis}$-R_{anis} (%)	—	—	9.6*	—
No. waters	86	77	66	62
Total No. non-H atoms	438	432†	421†	418†
No. 90% $2\sigma(F)$ reflections	8682	8679	8830	8630
Data:param. (isotropic)	5.0	5.0	5.2	5.1
Data:param. (anisotropic)	—	—	2.3	—
Min. $B_{complex}$ (Å^2)	4.6	4.0	7.2	5.3
Max. $B_{complex}$ (Å^2)	33.5	49.9	30.9	62.7
Ave. $B_{complex}$ (Å^2)	13.1	12.4	16.8	11.3
Ave. B_{water} (Å^2)	30.0	27.8	29.3	26.8

* After anisotropic refinement. † An extra three atoms for *PROLSQ* and *SHELXL93* and an extra four atoms for *X-PLOR* refinement are included for the double conformations modelled in each case.

Table 3. *Weights on the restraints (in bold type) used in the NUCLSQ and PROLSQ refinements of the d^5(TGATCA)-nogalamycin complex and final r.m.s. deviations of their respective models*

	NUCLSQ		*PROLSQ*	
	σ	R.m.s.d	σ	R.m.s.d
Bonding distances				
Sugar/base/nog 1–2 distances (Å)	**0.025**	0.045	**0.020**	0.020
Sugar/base/nog 1–3 distances (Å)	**0.050**	0.075	**0.040**	0.050
Phosphate 1–2 distances (Å)	**0.025**	0.075	**0.020**	0.023
Phosphate 1–3 distances (Å)	**0.050**	0.112	**0.040**	0.054
Planar restraints				
Deviation from plane (Å)	**0.030**	0.056	**0.020**	0.031
Chiral centres				
Deviation of chiral volume (Å^3)	**0.100**	0.175	**0.150**	0.129
Non-bonded restraints				
van der Waals contacts (Å)	**0.063**	0.100	**0.500**	0.231
Hydrogen-bonded contacts (Å)	**0.063**	0.192	**0.500**	0.273
Temperature factors				
Sugar/base/nog/bonds (Å^2)	**7.500**	3.611	**2.000**	2.815
Sugar/base/nog/angles (Å^2)	**7.500**	4.452	**2.500**	4.134
Phosphate bonds (Å^2)	**7.500**	4.818	**2.000**	1.756
Phosphate angles (Å^2)	**7.500**	6.395	**2.500**	2.612
Restraints against excessive shifts				
Positional parameter shifts (Å)		**0.350**		**0.300**
Thermal parameter shifts (Å^2)		**0.350**		**3.000**

progressed, until as little as one to five were added in the final stages of refinement. No metal ions were located in the structure. The criteria which needed to be fulfilled were that (i) an exceptionally low B value had to be obtained on refinement and (ii) the atom suspected as being a metal ion was coordinated to four or more neighbouring atoms at distances of < 2.7 Å.

Structure factors, $2F_o - F_c$ and $F_o - F_c$ difference-density maps were calculated using the *CCP*4 programs *SFALL*, *FFT* and *EXTEND*. All maps were calculated using all data. Examination of the maps was performed using either *TURBO FRODO* (Jones & Cambillau, 1989) on an SGI or *FRODO* (Jones, 1978) on an ESV.

A summary of relevant information pertaining to all four refinements is given in Table 2.

2.2.1. *NUCLSQ*. The program *NUCLIN* uses a dictionary of ideal bond lengths and angles and a dictionary of weights and σ's on the restraints to set up the stereochemical restraints for the refinement of nucleic acids in *NUCLSQ*. The dictionary for nogalamycin was made by calculating 1–2 and 1–3 distances and chiral volumes from the structure by Arora (1983), to which planar groups were added. In *NUCLSQ* there was also the possibility of explicitly restraining sugar puckering. The global weight applied to the structure factors remained at 0.01 throughout refinement. Distance, angle, planar, chiral volume and van der Waals restraints are weighted by $1/\sigma^2$. A list of σ's used during the *NUCLSQ* refinement together with the r.m.s. deviations of the final model is given in Table 3.

Refinement of positional and thermal parameters prior to the addition of any solvent resulted in an R value of 30.2% (R_{free} = 37.3%). The refinement converged at R = 20.9% (R_{free} = 28.1%) for 8682 $2\sigma F$ reflections between 8.0 and 1.4 Å with the inclusion of 86 solvent molecules treated as O atoms. Refinement using all 9815 reflections resulted in a final overall R = 21.2%.

There was some 1.5σ $2F_o - F_c$ electron density indicating another possible conformation for the methyl ester group at C10 of nogalamycin NOG2, and eventually also for T7 O5′, but as *NUCLSQ* is not well adapted for handling static disorder, this was not attempted.

2.2.2. *PROLSQ*. The program *PROTIN* uses the standard dictionary of bonded 1–2 and 1–3 distances, chiral volumes and planar groups for each nucleotide created in *QUANTA* (Molecular Simulations, Burlington, MA), to set up the stereochemical restraints used in *PROLSQ* refinement. A dictionary for nogalamycin was created based on the Arora (1983) structure, in the same way as for *NUCLSQ*.

In *PROLSQ* a 'scip' value is applied which determines the relative weighting of the X-ray term to the geometry term. A 'scip' value of 0.5 was judged to be optimal throughout the *PROLSQ* refinement with respect to providing the lowest values of both R and R_{free}. The weights on distance, planar, chiral volume, temperature-factor and van der Waals restraints in *PROLSQ* are $1/\sigma^2$. The values of σ used for the d(TGATCA)–nogalamycin refinement are given in Table 3. Most of the default values were maintained, except for the σ's on the thermal-factor restraints which were considered to be too stringent for the refinement and therefore were increased from 1.0 and 1.5 Å^2 to 2.0 and 2.5 Å^2, respectively.

In contrast to *NUCLSQ*, *PROLSQ* allows the monitoring of R_{free} after each cycle of refinement, therefore R_{free} was used as an additional criterion for solvent fitting. If the R_{free} was seen to fluctuate throughout a set of cycles, it was considered to be indicative of a combination of correct and incorrect changes made to the model which had resulted in no global improvement. Inspection of electron-density maps allowed discrimination between good and bad changes. This was facilitated by carrying out model adjustments in small steps. The initial R factor converged to 31.4% and the R_{free} to 36.8%, before the addition of water molecules. At the end of solvent addition, 77 waters were included in the final model. The R factor converged to 22.4% and the R_{free} to 27.4%, using 8679 $2\sigma F$ reflections (90% data). Refinement was completed to convergence using all 9813 reflections which gave an R factor of 22.8%.

A double conformation for the methyl ester group at C10 of NOG2 was visible in the density. On inclusion and subsequent refinement, an improvement in the density was observed.

2.2.3. *SHELXL93*. Bond lengths and angles within the bases were restrained to target values as specified by Taylor & Kennard (1982). All other chemically equivalent bond lengths and angles, including glycosidic linkages and the sugar–phosphate backbone, were restrained by similar distance restraints which do not require specification of an actual target value.

In *SHELXL*93, each reflection is weighted individually by 'w', which depends on F_o^2, $\sigma(F_o)^2$ and F_c^2, and hence was updated throughout refinement. Restraints are weighted individually by $1/\sigma^2$. σ's for both target and similar distance restraints were set at 0.03 Å for bonds and 0.05 Å for angles. After the first cycle in *SHELXL*93 the chiral volumes of sugars and of nogalamycin were tabulated, and only those which deviated significantly from the ideal, were restrained during subsequent refinement. After adding water molecules antibumping restraints were applied to prevent the solvent molecules from coming closer than 2.3 Å. Regions with diffuse solvent were modelled using Babinet's principle (Langridge *et al.*, 1960; Driessen *et al.*, 1989).

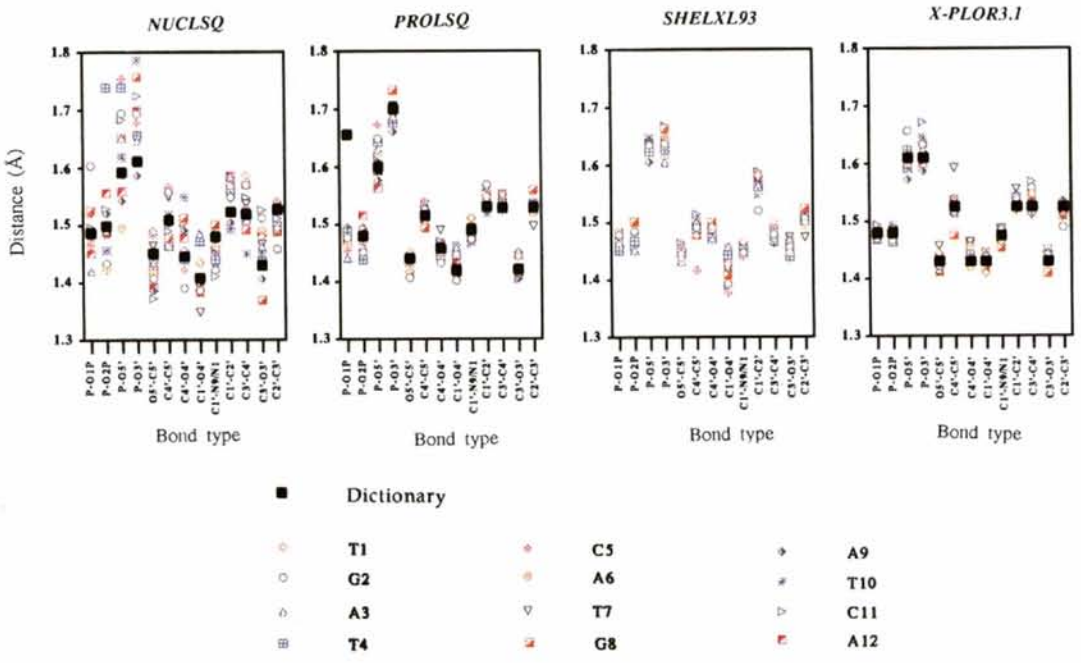
In order not to lose the gain in experimental information by refining against F^2 in *SHELXL*93 no σ cut-off was applied to the 8.0–1.4 Å data. R indices based on F^2 (wR_2) are larger than those based on F. Hence, for comparison with the other programs a conventional R index (R_1) based on observed F values larger than $4\sigma(F_o)$ is also quoted. After the addition of 66 water molecules, R_1 dropped from 29.4% ($wR_2 = 65.5\%$) to 21.7% ($wR_2 = 53.73\%$) and R_{1free} dropped from 34.1 to 26.9%. Subsequent to isotropic refinement, all H atoms were added at calculated positions and anisotropic refinement was performed. This yielded an $R_1 = 15.6\%$ ($wR_2 = 40.1\%$) and $R_{1free} = 25.2\%$ for 8830 reflections in the 8.0–1.4 Å resolution range. The final R_1 value for all 9813 reflections was 16.0% ($wR_2 = 40.9\%$).

A double conformation for the methyl ester group of NOG2 (named O14B, O10B and C15B) was clearly present, and was already incorporated during the isotropic refinement. During the anisotropic refinement, modelling of a double conformation for the methyl ester group of NOG1 was also attempted, but this did not improve the map, neither did a double conformation for T7 O5′. Both similarity and rigid bond restraints on the U_{ij} values during anisotropic refinement could not prevent thermal ellipsoids of some atoms from becoming non-positive definite. An additional restraint on the U_{ij} components of all atoms to approximate isotropic behaviour was necessary. No additional $F_o - F_c$ density was observed for possible waters after anisotropic temperature factor refinement in *SHELXL*93.

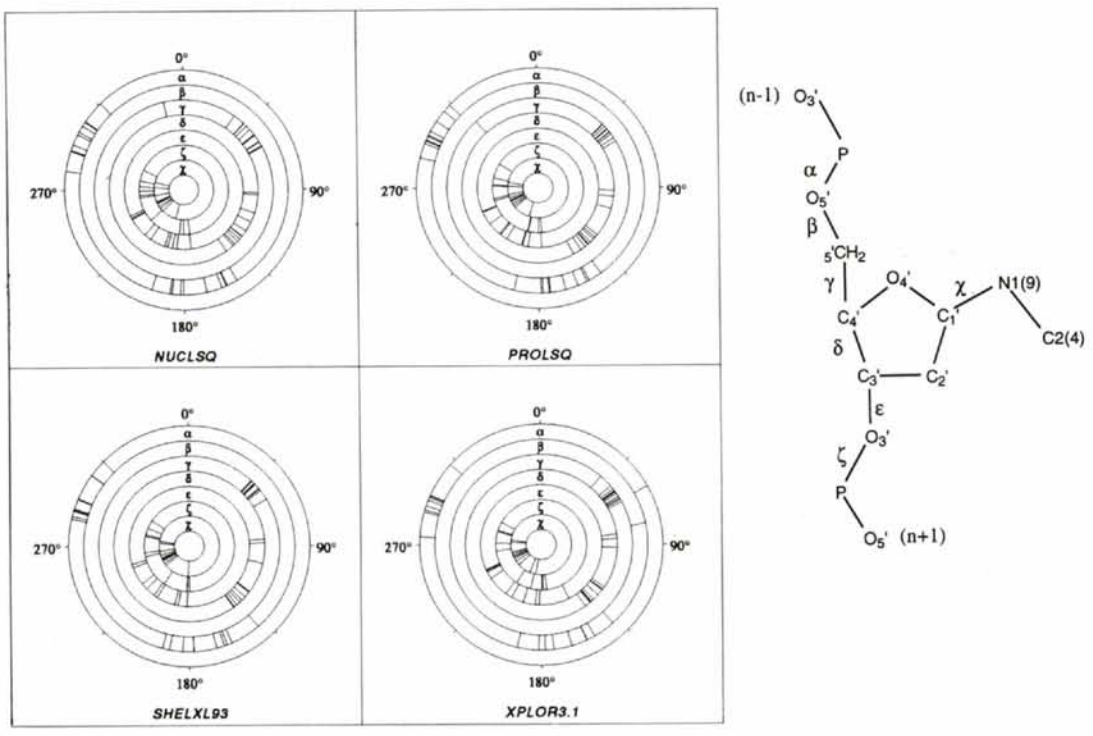
In order to obtain estimated standard deviations on atomic positions, bond lengths and angles, five cycles of full-matrix least-squares refinement were performed. Each cycle involved the refinement of a block of parameters. The structure was divided into three overlapping blocks of xyz and U_{ij} parameters for the complex and a fourth block containing the same parameters for the water molecules. In the fifth cycle all xyz but no U_{ij} values were refined. The maximum standard deviations on bond lengths and angles in the final model were 0.026 Å for the C5′–O5′ distance of T7 and 4.22° for the O1P–O2P angle of T10 in the DNA, and 0.053 Å for the O10B-C14 distance and 7.47° for the C15B–C14 angle in NOG2.

2.2.4. *X-PLOR*. *X-PLOR* version 3.1 provides standard topology (toph11.dna) and parameter (param11.dna) files for nucleic acids. Independent entries into both topology and parameter files for the drug were generated from nogalamycin coordinates (Arora, 1983) by *XPLO2D* (Kleywegt, 1995). The coordinates of the d(TGATCA)–nogalamycin complex were prepared for *X-PLOR* refinement by the addition of polar H atoms, which is a standard protocol in *X-PLOR*. The B values for base, sugar, phosphate and nogalamycin atoms were set initially at 10, 15, 20 and 12 Å^2 respectively.

The overall weight 'W_A' for the X-ray pseudo-energy term relative to the geometric energy term was determined before each run with the *CHECK* stage of *X-PLOR*. The individual contributions to the empirical



(*a*)



(*b*)

Fig. 2. Comparison of the fine structural details of DNA in the d(TGATCA)–nogalamycin complex models derived from *NUCLSQ*, *PROLSQ*, *SHELXL*93 and *X-PLOR* refinement. (*a*) Dictionary and final bonded distances within the sugar–phosphate backbone of all models. (*b*) Distribution and definition of sugar–phosphate torsion angles. Ranges typical of B-DNA are highlighted in orange. The torsion wheels were provided by the Nucleic Acid Database (Berman *et al.*, 1992).

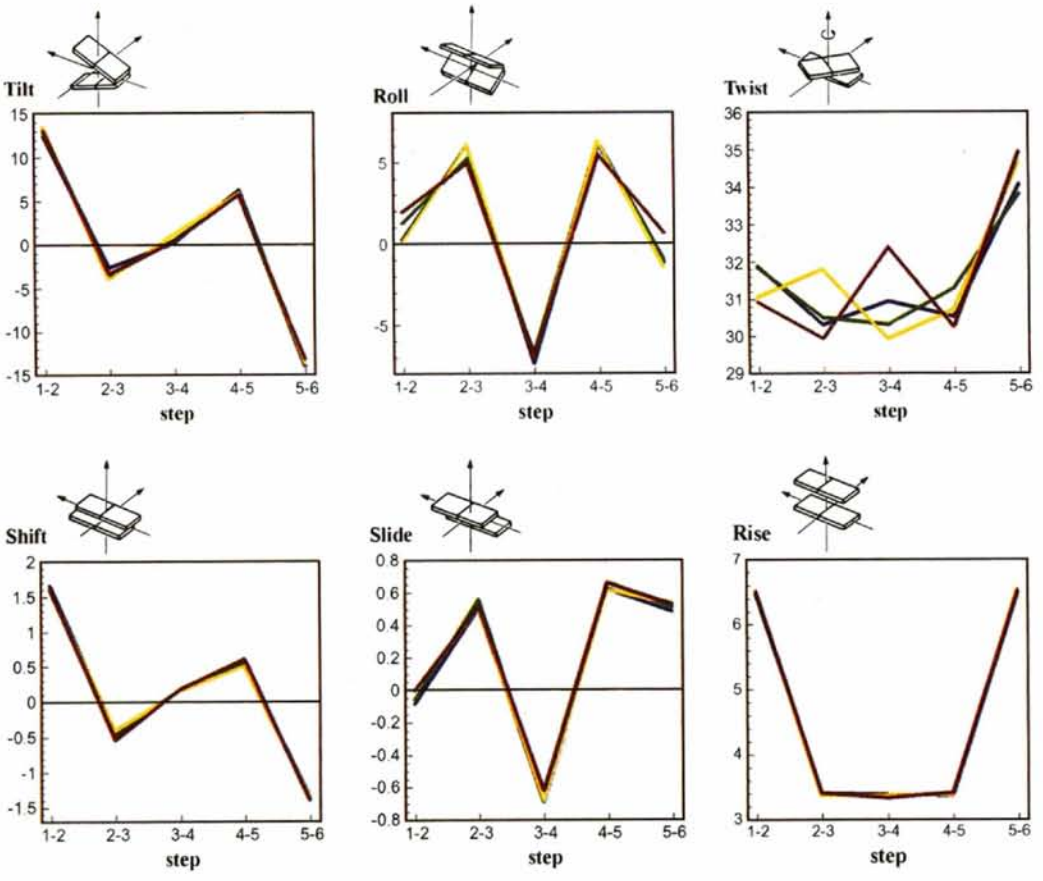
energy term from distance, planar, chiral volume and van der Waals restraints are weighted by type-based force constants.

The SIRAS model was first subjected to normal positional refinement followed by simulated annealing, which reduced the R factor to 33.3% and the R_{free} to 36.6%. Simulated annealing was performed with the slow-cooling protocol (Brünger, Krukowski & Erickson, 1990) starting at an initial temperature of 3000 K and descending in steps of 25 K to 300 K, with 50 cycles of refinement performed at each step. In contrast to the other programs, positional parameters and temperature factors are refined in two separate stages. The refined model was refined by conjugate-gradient methods, with 60 steps for coordinates and 15 steps for temperature factors in each cycle. Overall and individual B-factor refinement followed the simulated-annealing stage which yielded an R value of 31.9% and R_{free} value of 35.6%. In the next cycles, solvent molecules were added and the structure was further refined using conventional conjugate-gradient positional and isotropic B-value refinement. As in *PROLSQ*, the R_{free} was calculated in each cycle and used as an additional criterion for solvent fitting. 62 water molecules were located in the final model. The overall R factor converged to 24.1% using 8630 reflections and R_{free} to 27.0%. The overall R factor using all 9813 data was 24.4%.

Incorporation of double conformations for the methyl ester group at C10 of NOG2 and the O5′ of T7 in refinement resulted in an improvement of the density.

The final coordinates from refinement have been deposited together with observed intensities in the Brookhaven Protein Data Bank.*

* Atomic coordinates and observed intensities have been deposited with the Protein Data Bank, Brookhaven National Laboratory (References: 224D and R224DSF). Free copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: JN0018).



(*c*)

Fig. 2 (cont.). (*c*) Helical parameters for the models resulting from refinement with *NUCLSQ* (blue), *PROLSQ* (green), *SHELXL*93 (purple) and *X-PLOR* (yellow). The values for the rotational parameters, tilt, roll and twist are given in °, those for the translational parameters shift, slide and rise in Å. Parameters are defined according to the EMBO Cambridge Workshop (1989) convention and calculated with the program *RNA* (Babcock & Olson, 1993) using Cartesian coordinate frames.

Table 4. *Least-squares superposition of the final d^5(TGATCA)-nogalamycin models from the NUCLSQ, PROLSQ, SHELXL93 and X-PLOR 1.4 Å refinements*

DNA, superposition of the two DNA strands. N1, superposition of the nogalamycin intercalated between T7-G8/C5-A6. N2, superposition of the nogalamycin intercalated between T1-G2/C11-A12. COMP, superposition of the whole d^5(TGATCA)-nogalamycin complex. The superpositions exclude atoms in double conformations and the following atoms from the *X-PLOR* model (see text) P(T10), O1P(T10), O2P(T10), O2P(C11), C26(N13). These positions deviate from the other models by more than 1 Å.

	PROLSQ				*SHELXL93*				*X-PLOR*			
	DNA	N1	N2	COMP	DNA	N1	N2	COMP	DNA	N1	N2	COMP
NUCLSQ												
R.m.s. displacement (Å)	0.14	0.11	0.15	0.14	0.17	0.11	0.18	0.16	0.14	0.11	0.16	0.14
Ave. displacement (Å)	0.10	0.09	0.12	0.10	0.13	0.10	0.13	0.13	0.11	0.09	0.12	0.11
Max. displacement (Å)	0.81	0.46	0.48	0.82	0.79	0.28	0.59	0.79	0.60	0.48	0.59	0.60
PROLSQ												
R.m.s. displacement (Å)					0.12	0.12	0.12	0.12	0.12	0.10	0.10	0.12
Ave. displacement (Å)		—			0.10	0.10	0.10	0.10	0.09	0.10	0.09	0.09
Max. displacement (Å)					0.55	0.52	0.29	0.56	0.76	0.30	0.24	0.76
SHELXL93												
R.m.s. displacement (Å)									0.15	0.13	0.11	0.14
Ave. displacement (Å)		—					—		0.11	0.10	0.10	0.11
Max. displacement (Å)									0.79	0.56	0.25	0.78

3. Results

3.1. *R.m.s. differences between global structures*

The root-mean-square differences between the *NUCLSQ*, *PROLSQ* and *SHELXL*93 models of the non-solvated complex, as calculated in the *CCP*4 program *LSQKAB*, are very small, varying from 0.12 Å (shelxl93-prolsq) to 0.164 Å (shelxl93-nuclsq) (Table 4). This is less than/equal to the upper estimate of coordinate error, 0.1870 Å, as calculated by the σ_A method of Read (1986). The root-mean-square differences between the *X-PLOR* model and the other models are all higher, from 0.203 Å (x-plor-prolsq) to 0.239 Å (x-plor-nuclsq), due to two anomalies. One is the C26 methyl group of NOG1 in the *X-PLOR* model, which shows an orientation differing by 180° with respect to the models from the other refinement programs. This conformation lay in well defined 1.25σ $2F_o-F_c$ density with no positive or negative 3σ F_o-F_c density to indicate that it may have been misplaced. The other anomaly consists of the O1P and O2P atoms of residue T10 which have undergone 30 and 60° rotations, respectively, as reflected in the outlying α and β angles (Fig. 2*b*). A small volume of positive and negative 3σ F_o-F_c density in the vicinity of the two non-esterified phosphate O atoms indicated the mobility of residue T10 phosphate. Both groups are in close proximity to one another and display high B values on refinement using all programs. The uncertainty in their atomic positions, is accentuated in the *X-PLOR* model where a different conformation to that seen in the other three models was attained during simulated annealing. Calculating the difference between such poorly defined positions does not provide relevant information, therefore it was decided to remove the three outlying atoms from all further r.m.s. calculations. With these exclusions, the range in r.m.s. differences between the *X-PLOR* and the other three models falls within that for the three programs quoted initially. The maximum displacements between corresponding atoms of the four complexes vary from 0.558 Å (5O1P shelxl93-prolsq) to 0.815 Å (11O2P prolsq-nuclsq) and are always between phosphate O atoms.

In general, the r.m.s. fit between the NOG1 molecules is always lower than that between the complexes or between the DNA structures, whilst the r.m.s. fit between the NOG2 molecules is always higher. However, this is not always the case in comparisons involving the *X-PLOR* model.

R.m.s. differences between the *NUCLSQ*, *PROLSQ*, *SHELXL*93 and *X-PLOR* models of the conserved solvent structure are also small, the highest being 0.296 Å (xplor-nuclsq).

The four final structures of the complex are identical within the accuracy of the structure determination irrespective of the program used for refinement.

3.2. *Effect of the dictionary on fine structural details*

The r.m.s. differences between the target bond lengths in the nucleic acid dictionaries of each program were negligible, typically 0.012 Å. The dictionaries for nogalamycin used in *NUCLSQ*, *X-PLOR* and *PROLSQ* are based on molecule *A*, molecule *B* and the mean of both molecules, respectively, in the asymmetric unit of the small-molecule structure (Arora, 1983). R.m.s. differences calculated between the dictionaries for nogalamycin used in *NUCLSQ*, *PROLSQ* and *X-PLOR* are around 0.050 Å, indicating that the variations in ideal drug distances between the programs are also insignificant. Similar values are found on comparing these dictionaries to the final model obtained from *SHELXL*93, which uses similar distance restraints to define nogalamycin and DNA backbone geometry.

In Fig. 2(*a*) both dictionary and final bonded distances within the sugar–phosphate backbone are plotted for

each program. As expected the smallest spread in distances is found in the model obtained from *SHELXL93* refinement, since this program restrains the backbone distances to be similar. It is noteworthy that without target restraints *SHELXL93* refines bond lengths to values similar to the dictionary distances from the other refinement programs, indicating the validity of the latter. The largest spread in final bond lengths was found for the *NUCLSQ* refinement. This could be expected, since the weighting of the restraints relative to the data was less than in the other refinements. *X-PLOR* restrains all distances between pre-specified atom types to be similar, for example C1′—O4′ and C4′—O4′ are regarded to be similar, while *NUCLSQ* and *PROLSQ* target slightly different values for both bonds. In this respect, the latter programs appear to be better equipped for DNA refinement, since *SHELXL93* reproduces the same small differences. The *PROLSQ* dictionary takes into account the possible protonation of one phosphate O atom and therefore contains a longer target bond length for the P—O1P than for P—O2P. However, all P—O1P/O2P bonds refine to distances typical of the unprotonated state.

Fig. 2(*b*) illustrates the distribution of backbone torsion angles. Apart from the few outlying angles, all four refined structures show almost an identical overall distribution and fall within the ranges typical of B-DNA. Furthermore, the different programs reflect the same variation of torsion angles within the individual ranges for each angle, which therefore provides information about the fine structural details of DNA.

As expected, the glycosidic torsion angle χ is well preserved throughout all refinement models. All χ angles lie in the range 215–272°, except four which are affected by drug binding. At both intercalation sites the χ angles for the four pyrimidine bases are distorted from normal B-DNA geometry whereas the purine χ angles are unaffected. In all models, a reduced χ (193.8°) for residues T1 and T7 is observed whereas an increased χ (277.8°) is observed for residues C5 and C11. These distortions of the pyrimidine χ angles are indicative of the asymmetrical changes that the DNA backbone undergoes to accommodate the nogalamycin molecule. The endocyclic δ angle is also well preserved.

Few aberrations are seen in the exocyclic torsion angles α, β and γ. Firstly, both of the *NUCLSQ* and *PROLSQ* models show one outlying γ angle for residue T7 due to its position at the end of the DNA duplex. Secondly, the *X-PLOR* model shows substantially deviating α and β angles for residue T10. In all models a bimodal distribution for ζ is seen which is characteristic for B-DNA (Privé *et al.*, 1987); except for residue G2, which falls in between the two groups in the *NUCLSQ* model. The ζ value of only one residue, A9, is not consistent in all models, varying from *–synclinal* to *–antiperiplanar*. On the contrary, for ε no bimodal distribution is seen but a similar variation from *antiperiplanar* to *anticlinal* is demonstrated in all refinements. This lack of bimodal distribution for ε has previously been reported (Hahn & Heinemann, 1993). The inner residues of the DNA helix show *antiperiplanar* ε angles. The ε angle is adjusted from *antiperiplanar* (–169°), characteristic of B-DNA, to *anticlinal* for the outer residues C5, T7, C11 and to a lesser extent C1. These distortions combine to result in an increase in base-pair separation, from 3.4 to 6.6 Å, to accommodate the drug.

Table 5. *Sugar conformations and pseudorotation angles (P) for the NUCLSQ, PROLSQ, SHELXL93 and X-PLOR final models*

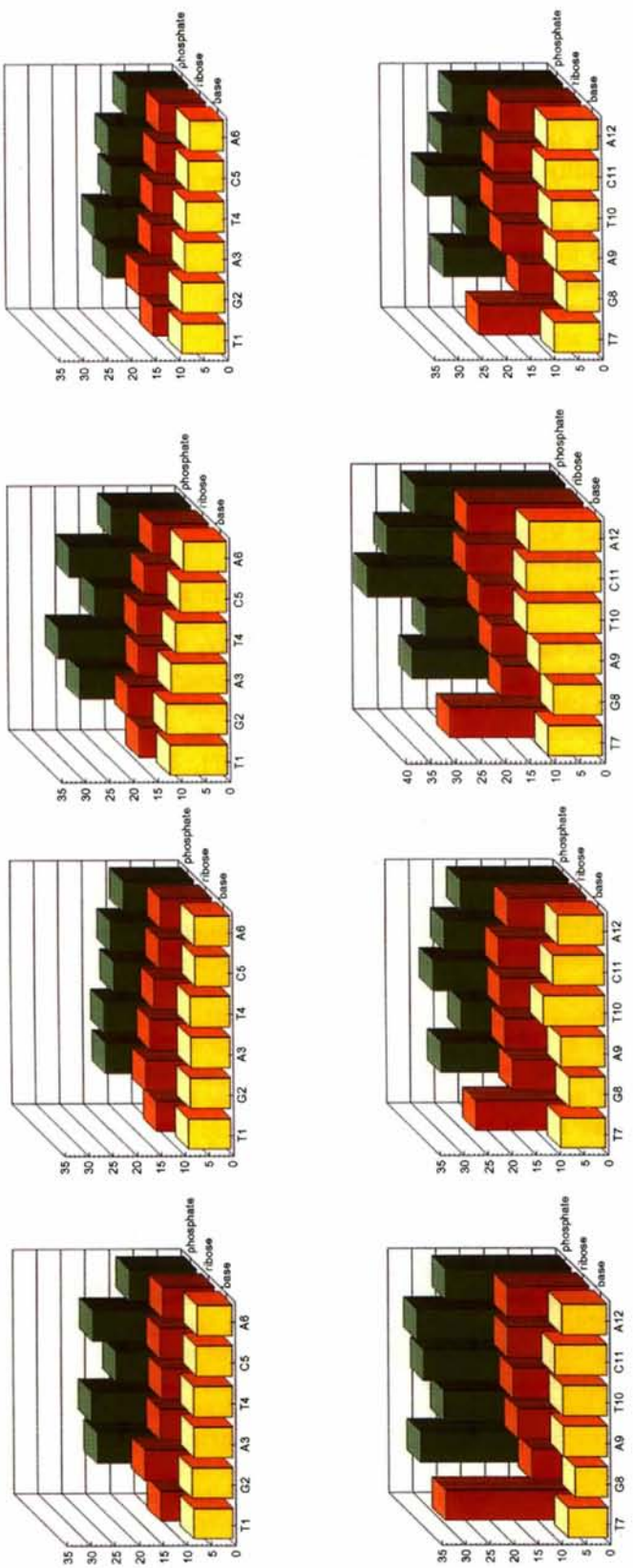
Pseudorotation angles were calculated using the program *NEWHEL93* distributed by R. E. Dickerson.

BASE	Sugar conformation	*NUCLSQ* P	*PROLSQ* P	*SHELXL93* P	*X-PLOR* P
Strand 1					
T1	C4′ *exo*	68.9	68.1	61.5	67.1
G2	C2′ *exo*	115.4	129.2	124.6	133.6
A3	C2′ *endo*	141.5	149.8	157.2	143.5
T4	O4′ *endo*	87.4	88.0	78.9	89.2
C5	C2′ *endo*	158.9	149.5	153.0	154.6
A6	C2′ *endo*	174.1	174.2	172.4	176.9
Strand 2					
T7	C4′ *endo*	230.6	230.2	250.8	234.7
G8	C1′ *exo*	130.9	131.3	139.6	133.9
A9	C1′ *exo*	116.1	146.6	136.8	143.7
T10	O4′ *endo*	105.1	100.9	95.1	90.7
C11	C2′ *endo*	156.4	149.1	151.2	139.4
A12	C2′ *endo*	172.8	164.3	168.3	160.7

Sugar conformations and pseudorotation angles for each residue are well conserved throughout all models (Table 5). The overall similarity in backbone and furanose puckering angles, between programs and to normal B-DNA demonstrates that, at this resolution, it was not necessary to restrain torsion angles. Thus, the values they adopt in the final structures provide independent validations of the final models.

The helical parameters describing the relationship between two base pairs, as calculated with the program *RNA* (Babcock & Olson, 1993), are displayed in Fig. 2(*c*). The plotted parameters are calculated using local Cartesian coordinate frames and are therefore not directly comparable with those calculated with the program *NEWHEL93* (distributed by R. E. Dickerson), which uses one global helix axis. The plots show almost identical helical parameters in all four refinement models. In all models the T·A base pairs above the chromophore of both nogalamycin molecules are buckled by 12°, while the G·C base pairs below the chromophore have a larger buckle of approximately 25° in the opposite direction.

The twist angle of the d(T1pG2) and d(C5pA6) intercalation-site steps is 31.4 and 34.4°, respectively. The other steps have an average twist angle of 30.7°. The overall unwinding angle of the four inner base pairs of the DNA double helix, due to the intercalation of nogalamycin is different for both drug molecules, being



NUCLSQ *PROLSQ* *SHELXL93* *X-PLOR*

Fig. 3. Bar diagram representing the thermal vibrations of (*a*) strand 1 and (*b*) strand 2 of the d(TGATCA)–nogalamycin models from the four refinements. The thermal parameter B is averaged for atoms of the base (yellow), the sugar (excluding O3′ and O5′) (red) and the phosphate group (including O3′ and O5′) (green) of each residue. For the *SHELXL*93 model B_{eq} values are quoted, where $B_{eq} = 8\pi^2 U_{eq}$.

3.7° for NOG1 and 0.7° for NOG2. The terminal base pairs T1·A12 and A6·T7 at each end of the DNA helix are shifted over the short axis of the base pair towards the minor groove.

This analysis of fine structural details demonstrates that the use of different refinement programs and their associated stereochemical dictionaries does not leave a significant imprint on the atomic coordinates of the final fully refined structure.

3.3. *Thermal parameters*

The mean *B* values for the atoms in base, sugar and phosphate groups of each nucleotide of d(TGATCA) after refinement with *NUCLSQ*, *PROLSQ*, *X-PLOR* and *SHELXL*93 are shown in Fig. 3. In general, *B* values from the *SHELXL*93 model are the highest (averaging at 16.8 Å^2 for the non-solvated complex after anisotropic refinement) and those from the *X-PLOR* model the lowest (averaging at 11.3 Å^2 for the non-solvated complex). In particular, in *SHELXL*93 *B* values of both phosphate O atoms of residue T10 (60.4 and 53.0 Å^2) and methyl group C26 of NOG1 (31.4 Å^2) are allowed to expand to higher values compared to the other programs, which restrain them tightly (phosphate O atoms from 30 to 35 Å^2 and C26 from 17 to 20 Å^2).

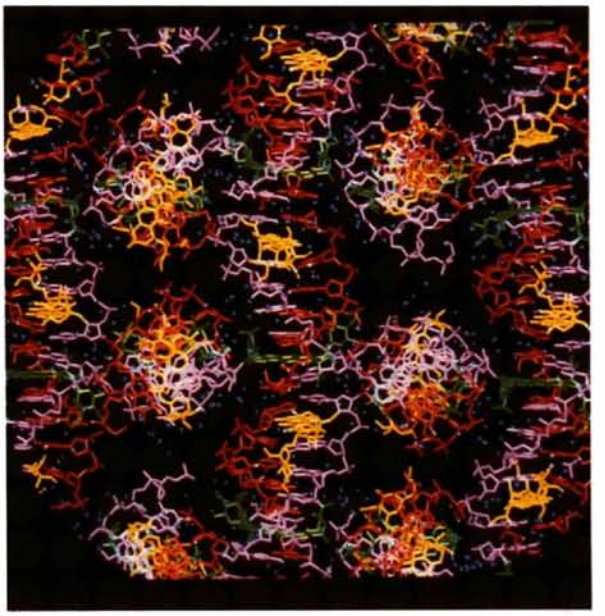
In general, the largest difference in temperature factors is observed not between the models from the different programs, but between the two strands of the DNA helix in each model. The second strand always has significantly higher *B* values (+5 Å^2) than the first strand. The same is seen for the nogalamycin molecules. The aminoglucose and the nogalose moiety of NOG2 are situated closer to the second strand than to the first strand, and hence show higher *B* values than the same groups in NOG1, which is slightly shifted towards the first strand.

The higher *B* values at the second strand site can be explained by the crystal packing of the complexes, as shown in Fig. 4. The sugar of nucleotide T7 shows in all four models an increased mean *B* value compared to the other sugars. This is due to its position at the end of the DNA duplex.

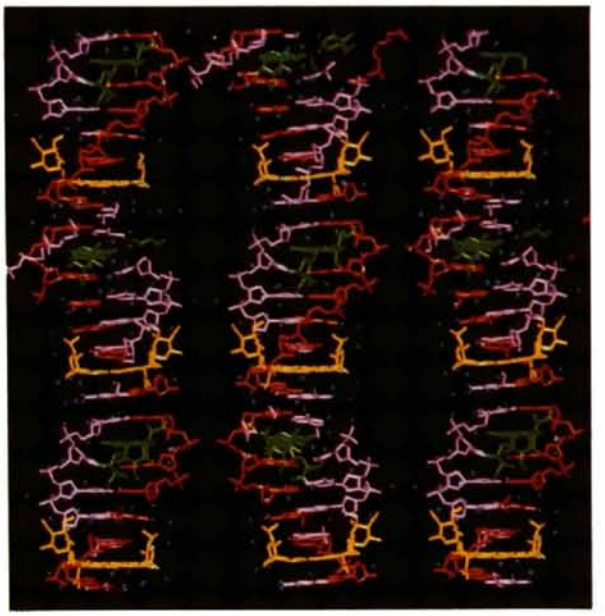
Temperature factors of the waters range from *ca* 9 Å^2 to *ca* 50 Å^2 and average at *ca* 28 Å^2 in all four models.

3.4. *Static disorder*

The distinct double conformation of the methyl ester group at C10 in NOG2, has been modelled in three of the four programs. In *PROLSQ* the occupancies of the disordered groups were fixed at 0.5 and not refined. In *X-PLOR* occupancies of the disordered group were refined, although their sum was not constrained to unity. *SHELXL*93 is especially well equipped for realistic refinement of a statically disordered group, by constraining the sum of the occupancy factors to be unity and restraining chemically equivalent distances to be equal to each other.



(*a*)



(*b*)

Fig. 4. Orthogonal views (*a*) along the *ab* diagonal and (*b*) along the *c* axis, representing the crystal packing of the d(TGATCA)–nogalamycin complex. Segments of the complex are represented in the following colours: strand 1 (T1·A6) (pink), strand 2 (T7·A12) (red), NOG1 (green) and NOG2 (yellow). It can be seen from (*a*) that the central part of strand 2, (residues A9, T10 and C11), is more open to the local widened solvent channel and has few stabilizing interactions with the neighbouring complexes. (*b*) clearly indicates the unfavourable contact made by the phosphate groups of residue G8 of strands 2 of neighbouring complexes. Both factors may account for the higher *B* values of strand 2 and NOG2.

Both, the *SHELXL*93 and to a lesser extent the *X-PLOR* model, revealed the same predominant conformation with respective occupancies of 0.67 and 0.53. This conformation is the one found in all other DNA–nogalamycin structures. The minor conformation refined to occupancies of 0.33 and 0.42 for *SHELXL*93 and *X-PLOR*, respectively.

The presence of double conformations was most clearly revealed in the *SHELXL*93 model. In *NUCLSQ* a double conformation was the least obvious and was not modelled. A double conformation for T7 O5′ was modelled in the *X-PLOR* refinement only. The new conformation refined to an occupancy of 0.56, while surprisingly the conformation in common with that found in all other models refined to a slightly lower occupancy of 0.42. Insertion of this double conformation was also tried in *PROLSQ* and *SHELXL*93 without improvement of the model.

3.5. *Hydration of DNA*

Since solvent is believed to play an important role in determining the DNA conformation and its recognition characteristics, hydration patterns around crystalline DNA molecules have received much attention (Saenger, Hunter & Kennard, 1986; Kopka, Fratini, Drew & Dickerson, 1983). If the solvent structure is that critical for recognition and specificity, at least the same first-shell water molecules should be found, whichever refinement protocol is used.

Comparison of the solvent structure of the final models from *NUCLSQ*, *PROLSQ*, *SHELXL*93 and *X-PLOR* resulted in 53 solvent sites which were conserved in all four models. As a criterion for being conserved a maximum distance of 0.8 Å between two O atoms from different refinements was applied (Hahn & Heinemann, 1993). The total amount of conserved solvent molecules ranges from 82% of the total (62) in *X-PLOR* to 59% of the total (86) in *NUCLSQ*. This is significantly more than was found in a previous 8–1.7 Å study at room temperature (Hahn & Heinemann, 1993). Amongst the 53 conserved solvent sites, 38 are first-shell water molecules. In addition four first-shell waters are partially conserved, *i.e.* in three of the four final models. In two of the four final models, five other first-shell waters are conserved. Only four first-shell waters are uniquely found with one of the refinement programs. Thus, the first-shell hydration sphere around the DNA–drug complex is determined with satisfactory reliability whichever refinement program is used.

Not the position, but rather the number of first-shell waters differs in the different refinement models, in *SHELXL*93 the fewest are located and in *NUCLSQ* the most. Apart from the previously mentioned 38 sites, in the *SHELXL*93 model only three partially conserved first-shell waters could be located, whereas in the *NUCLSQ* model there were 12.

Of the 38 first-shell waters, 12 are located in the minor groove, 18 in the major groove and eight are hydrogen bonded to phosphate groups. Hence, the majority comprises a well ordered hydration pattern in both grooves, while the few remaining interact with the backbone. Only one conserved backbone water is bonded to a second-strand phosphate; in contrast, all except one of the first-strand phosphates are hydrated in all refinements. The second hydration shell contains 14 conserved waters, six in the minor groove and eight in the major groove. This is approximately half of the total amount of second-shell waters found in all models, except the *X-PLOR* model where only one extra second-shell water is found. Even one third-shell water is conserved.

As many as 69 water molecules are found to be conserved between the *NUCLSQ* and *PROLSQ* final models.

3.6. *Hydration of nogalamycin by conserved waters*

In terms of DNA recognition, the nature of the noncovalent interactions between DNA, intercalator and solvent are important. Hence, stable interactions should be conserved in all four programs.

The hydration patterns of both nogalamycin molecules are illustrated in Fig. 5. NOG1 is surrounded by seven conserved first-shell water molecules and NOG2 in a similar way by six. In both nogalamycins a water molecule is bound to atoms N1 and O16 of the aminoglucose moiety and to the O1′ and O3′ atoms of the nogalose sugar. In NOG1 the hydroxyl substituent O9 on ring *A* forms two hydrogen bonds with two conserved water molecules, while in NOG2 there is only one conserved water bound. The hydration of O12 on the chromophore of NOG2 is not observed at O12 of NOG1 where the T7 O5′ is folded back in towards the helix forming a direct hydrogen bond in place of a water molecule.

In addition, weak hydrogen bonds between conserved waters and both O4 (3.44, 3.39 Å) are seen in NOG1 and NOG2, respectively. Further stabilization of nogalamycin by van der Waals contacts is observed [*e.g.* the contact between a conserved water and O14 (3.52, 3.68 Å)]. Distances quoted in this and the following section are averaged over the four final models.

3.7. *Interactions of nogalamycin with DNA*

Considering the DNA–drug interactions (illustrated in Fig. 5), only one strong direct hydrogen bond between the N7 of the intercalation-site guanine and O15 of nogalamycin is observed in the major groove. In addition, in both molecules a direct weak interaction is seen in the minor groove, between the O7 glycosyl linkage of NOG2/NOG1 and N2 of G2 (3.38 Å) and N2 of G8 (3.39 Å), respectively. Both strong and weak interactions are consistent with previously reported nogalamycin–DNA structures (Liaw *et al.*, 1989; Gao,

Liaw, Robinson & Wang, 1990; Williams *et al.*, 1990; Egli, Williams, Frederick & Rich, 1991).

The number of indirect interactions with DNA is larger. In the major groove, both nogalamycins interact with the base pair adjacent to the intercalation site through a conserved water-mediated hydrogen bond from N1 to O4 of thymine (T4/T10). A hydrogen bond, mediated by three conserved waters, to N7 of adenine (A3/A9) provides an extra stabilization of the nogalamycin N1 atoms. NOG2 makes an additional water-mediated hydrogen bond from O4 to N7(A12) of the intercalation site in the major groove, while NOG1 makes an additional one from O15 to a phosphate O atom of a symmetry-related G2 residue. The latter interaction is a consequence of the crystal packing and provides supplementary stabilization of the aminoglucose moiety of NOG1, which is not seen for NOG2. Both nogalamycins form a hydrogen bond mediated by two conserved water molecules from O16 to N4 of cytosine (C5/C11), which is stabilized by a van der Waals interaction between the drug O16 and N4 of C5/C11 (3.58, 3.41 Å).

In the minor groove, a conserved water-mediated hydrogen bond is seen from O1′ of nogalamycin to O3′ of the intercalation-site adenine (A6/A12). The O1′ is further stabilized by a hydrogen bond, mediated by two conserved waters, to N3 of adenine (A6/A12). In NOG1, atom O9 on ring *A* is connected by a hydrogen bond, mediated by three conserved waters, to the N3 atom of A6. As an effect of packing, the same

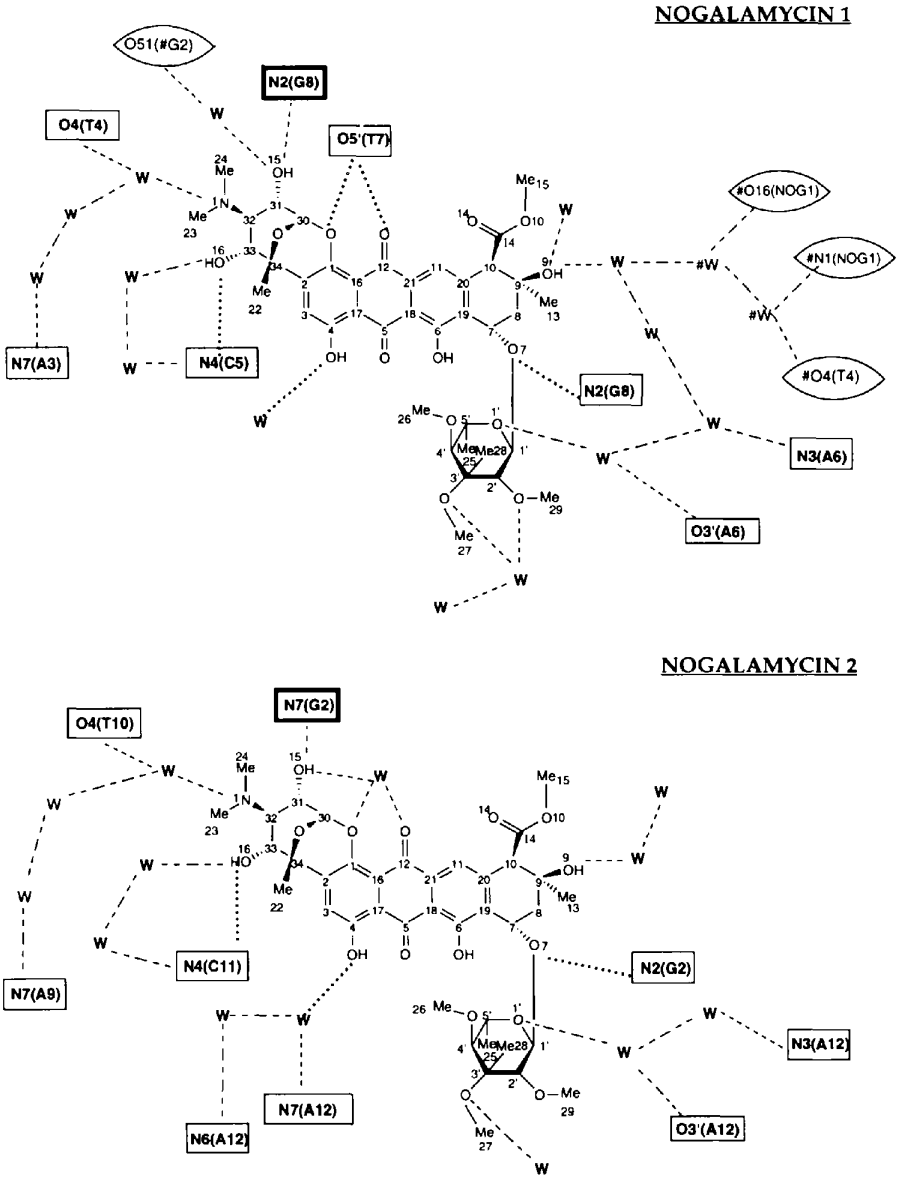


Fig. 5. Two-dimensional schematic diagrams illustrating the hydration patterns and the direct interactions to DNA for both nogalamycin molecules, NOG1 and NOG2, in the d(TGATCA)–nogalamycin complex. The N7 of guanine, which makes a strong interaction to O15 of the drug, is outlined in a bold rectangle. Normal hydrogen bonds (2.6–3.3 Å) are shown in dashed lines (- - -), whilst weak interactions (3.3–3.4 Å) are represented as dotted lines (. . .). The DNA atoms which may be acceptors or donors in the hydrogen-bonded network are outlined in rectangular shapes if they are intramolecular and in oval shapes if they are from a neighbouring molecule.

atom O9 forms other long-distance solvent interactions with both the NOG1 aminoglucose and the DNA of a symmetry-related complex. Again these interactions are not observed for NOG2.

Stacking interactions between the anthracycline chromophore and the intercalation-site base pairs stabilize the complex. Also the parallel orientation of the carbonyl O2 of cytosine (C5/C11) and the carbonyl O5 of the drug chromophore (3.34/3.25 Å) contributes slightly to these π–π interactions. The O2 atom of thymine (T1/T7) affects the orientation of the methyl ester group at C14 by long-range repulsion of the O14 atom in both NOG2 (3.31 Å) and NOG1 (3.58 Å).

Thus, although both drug molecules show similar hydration and DNA–interaction patterns, in this crystal structure NOG1 is more stabilized by solvent-mediated interactions with symmetry-equivalent atoms than NOG2 as a consequence of packing.

4. Discussion

4.1. *Factors influencing the refinement*

This work presents the comparison of the final models obtained using a single 1.4 Å low-temperature data set in four independent refinement programs. The data set is of high quality with respect to completeness, multiplicity, $\%I > 3\sigma(I)$ and $R_{\rm merge}$ statistics (Table 1). Ambiguities and differences between refinements could not then be attributable to any reason pertaining to the data. Therefore, in this comparison the only cause of differences in the final models stem from either the refinement methodologies or strategies.

4.1.1. *Refinement methodologies of the programs.* Differences exist in the ways various programs minimize the difference between observed and calculated structure factors and minimize the deviations from a set of ideal geometric criteria. *NUCLSQ*, *PROLSQ* and *SHELXL*93 all use conjugate-gradient least-squares methods which could fall into local minima. To overcome this problem, *X-PLOR* minimizes an energy function by simulated-annealing conjugate-gradient methods and conformational-search procedures. Monitoring only the conventional R factor in least-squares refinement will not reveal the overfitting of the observations. It was also found that even if the value of $R_{\rm free}$ at the end of a set of cycles was less than that at the end of the previous set but fluctuated within the set, this did not necessarily mean that the model was totally correct. Therefore, the $R_{\rm free}$ value was calculated after every cycle as an independent diagnostic for overfitting of the density in the *PROLSQ* and *X-PLOR* refinements. By comparison in *SHELXL*93 it is only practically possible to monitor $R_{\rm free}$ at the end of a set of cycles as opposed to *PROLSQ* and *X-PLOR*. *NUCLSQ* is not adapted for the sensible inclusion of an $R_{\rm free}$ calculation. It is not surprising that with the latter program, where the $R_{\rm free}$ has not been used as a criterion in refinement, the largest number of waters were located. A better agreement between R and $R_{\rm free}$ from the *X-PLOR* refinement in comparison to the other refinements is attributed to the fact that only in *X-PLOR* are the working and test sets scaled independently.

Convergence was achieved by using a combination of empirically optimized weighting schemes and restraints which are different for each program. It is, therefore, difficult to assess the individual contributions from the various parameters within each refinement. *NUCLSQ*, *PROLSQ* and *X-PLOR* use dictionary values to restrain geometry. *SHELXL*93, the only program using similar distance restraints instead of imposing target values for the sugar–phosphate backbone, gives final bond lengths comparable to those from the other three refinements, indicating the validity of the dictionaries of the latter (Fig. 2*a*). Although the thermal-parameter restraints were tighter for *PROLSQ* than for *NUCLSQ*, the average B values of the two complexes are similar. *SHELXL*93 refinement, where isotropic temperature factors were not restrained, resulted in the highest average B values of the four models. *X-PLOR* refinement resulted in the lowest temperature factors, where the default values restrain B more tightly (Fig. 3).

It appears that using this high-quality 8.0–1.4 Å data set neither the dictionary nor the refinement program used leave an imprint on the final fully refined complex.

4.1.2. *Refinement strategies.* As the initial model was identical, all refinements were subject to the same coordinate bias. In addition, every effort was made to adopt the same strategy for each refinement.

The inclusion of double conformations and solvent molecules are the steps of refinement most open to subjective interpretation. A set of strict water-selection criteria were adhered to at all times and double conformations were only included when a definite improvement in local map quality was obtained on subsequent refinement. After water fitting was complete in all refinements the $F_o - F_c$ density maps contained negligible density which has not been accounted for (*e.g.* maximum 0.45 e Å^{-3} and minimum –0.34 e Å^{-3} in *SHELXL*93).

H atoms were not included in *NUCLSQ* and *PROLSQ* refinement. In *X-PLOR* positions of polar H atoms were generated as their inclusion was necessary to calculate the appropriate empirical energy term. In *SHELXL*93 all H atoms were added to the model at calculated positions before anisotropic temperature-factor refinement was performed. Their positions remained unrefined, but they were allowed to 'ride' with the atoms to which they were attached. However, the coordinates of the non-H atoms in all final models were essentially the same irrespective of the addition of H atoms.

In *NUCLSQ* and *PROLSQ* only isotropic temperature factor refinement was performed. In *SHELXL*93 refinement, restraining anisotropic displacement parameters of bonded or neighbouring atoms to be equal did not appear to be strong enough to prevent numerous splittings and

thermal ellipsoids of a large number of atoms becoming non positive definite. Therefore, the U_{ij} components were restrained to have approximate isotropic behaviour.

4.2. *Structural comparison*

The four final models for the d(TGATCA)–nogalamycin complex, using either *NUCLSQ*, *PROLSQ*, *SHELXL*93 or *X-PLOR*, are identical within the discrepancy of the observed and calculated structure factors, 0.20 Å as calculated by the Luzzati method (Luzzati, 1952) and 0.17 Å by the σ_A method (Read, 1986).

An analysis of backbone torsion angles and sugar puckering shows that the different programs reflect the same variation of angles within their individual ranges (Fig. 2*b* and Table 5). Furthermore, helical parameters describing the geometry of base pairs and their position relative to the neighbouring base pairs, are preserved very well in the different refinement models (Fig. 2*c*). This high degree of reproducibility throughout all refinements allowed us to discuss conformational details derived from the final crystal structures.

Asymmetric variations in backbone geometry and sugar pucker are caused by intercalation of nogalamycin. The bases of residue T1/T7 and C11/C5 adopt altered glycosidic torsion angles to accommodate the nogalamycin molecules, while A6/A12 and G2/G8 remain unaffected. The values of the exocyclic torsion angle ζ show a bimodal distribution typical of B-DNA. This is not seen for ε, which displays values from *antiperiplanar* for the inner residues of the helix to *anticlinal* for the terminal residues. These distortions combine to result in the separation of adjacent bases from 3.4 to 6.5 Å. The intercalation of nogalamycin causes an overall unwinding of 3.7° per NOG1 molecule and 0.7° per NOG2.

As a consequence of crystal packing, throughout all final models a higher B value was observed for the second strand as well as for NOG2 which is slightly shifted towards this strand.

Although different numbers of water molecules are found in the four different refinement models (*NUCLSQ* = 86, *PROLSQ* = 77, *SHELXL* = 66 and *X-PLOR* = 62), the majority of them, 53, are located in the same positions. The first-shell hydration sphere around the DNA–drug complex is relatively well conserved. The number of conserved first-shell waters, 38, represents almost all first-shell waters found in the *SHELXL*93 refinement, but only 3/4 of the total found in the *NUCLSQ* refinement.

4.3. *Comparison with previously reported nogalamycin –DNA structures*

The decrease in thermal disorder on collecting low-temperature data has enabled double conformations to be seen in this complex structure which were previously unseen in the room-temperature refinement.

In general, in both the 1.4 and 1.8 Å (Smith *et al.*, 1995) structures similar conformational parameters and a similar solvation pattern around the nogalamycin were seen. However, the 1.4 Å structure revealed more clearly the long-range DNA–drug interactions. The difference between the temperature factors of the two strands, as observed in the 1.4 Å structure, was less pronounced in the 1.8 Å structure.

A number of DNA–nogalamycin interactions are conserved when comparing this structure with previously reported structures of the drug intercalated in the sequences d[CGT(pS)ACG] and d[^{m5}CGT(pS)A^{m5}CG] (Liaw *et al.*, 1989; Gao *et al.*, 1990; Williams *et al.*, 1990; Egli *et al.*, 1991). The strong hydrogen bond between the O15 atom of the drug and the N7 of guanine observed in this structure, was also seen, not only in the 1.8 Å refinement but also in the d[CGT(pS)ACG] and d[^{m5}CGT(pS)A^{m5}CG]-nogalamycin complexes. Some water mediated interactions present in both 1.4 and 1.8 Å structures are also conserved when the drug is bound to a different sequence. Notably, the O4 atom of T4/T10 in this structure, which makes a water-mediated hydrogen bond to the N1 atom of nogalamycin, is replaced by N6 of A4/A10 in the previous DNA–nogalamycin complexes. The stabilizing van der Waals interaction between O16 of the drug and N4 of C5/C11 is also preserved in all structures.

In the minor groove, the weak interaction between the glycosidic O7 and N2 of the intercalation-site guanine is also present in all DNA–nogalamycin structures.

Certain water-mediated interactions are not conserved, for example, the water-mediated stabilizing hydrogen bonds between the O1′ of the drug and both O3′ and N3 of the intercalation-site adenine (A6/A12).

4.4. *Suitability of the programs*

From the point of view of the suitability of the different programs, *NUCLSQ* and *PROLSQ* are the most closely related and perform, in general, equally well. Each program has its own advantages. In *NUCLSQ*, the most widely used nucleic acid refinement package, different weighting of the phosphate sugar and base bonds is allowed according to the increased flexibility of the DNA backbone with respect to the bases. However, in this refinement, application of looser restraints to the phosphates resulted in bad geometry. The program is less adapted for the incorporation of multiple conformations. In *PROLSQ*, all nucleic acid bonds and angles are weighted equivalently. Multiple conformations are easily incorporated but the occupancy refinement is not trivial. In the *X-PLOR* package atom types are used to define the restraints used. In some cases the atom types may be too general, refining bonds to be the same when chemically they are not identical. *X-PLOR* refinement is better than *PROLSQ* with respect to including double conformations as occupancies are refined but they are not constrained to add up to one. It is the only program

which gives an altered orientation for the C26 methyl group of nogalamycin NOG1. The *SHELXL*93 program includes all of the necessary requirements for a high-resolution nucleic acid refinement. The differing bond types in the DNA backbone and bases are weighted accordingly. Refinement is performed against F^2, with no σ cut-off. In this way more experimental information is incorporated, thus improving the data:parameter ratio. The maximized data:parameter ratio allowed restrained anisotropic temperature-factor refinement and may be the reason why *SHELXL*93 revealed much more clearly than the other refinement programs the presence of a double conformation. The program is well equipped for realistic refinement of a disordered group, by constraining the sum of the occupancy factors to be unity. It also has the advantage of combining different restraints so as to make the most appropriate use of available dictionaries. Where good distance restraints are available, *i.e.* for the bases, dictionary distances were used. For the sugars and the phosphates, where bond distances are not so well defined, it is better to assume that chemically equivalent 1,2 and 1,3 distances are equal without the need to specify target values. Therefore, the use of similar distance restraints for DNA refinement was advantageous.

This study indicates that the protein-refinement packages *PROLSQ* and *X-PLOR*, as well as the small-molecule refinement program *SHELXL*93, result in the same model as found using *NUCLSQ*, although they are not all as easily adaptable for sensible high-resolution DNA refinement. Hence, results of structure refinements of the DNA in DNA–protein complexes may directly be compared with earlier oligonucleotide crystal structures refined using program packages typical for the refinement of nucleic acids.

The authors acknowledge the help of Jim Brannigan, Eleanor Dodson, Anke Gelbin (Nucleic Acid Database), George Sheldrick, Giles Wilson and the SERC- and YCRC-supported Structure Group at York. GSS and LVM are supported by the research council of the Katholieke Universiteit, Leuven, Belgium. LVM is a Senior Research Associate of the National Fund for Scientific Research (Belgium).

References

Allen, F. H., Kennard, O. & Taylor, R. (1983). *Acc. Chem. Res.* **16**, 146–153.

Arora, S. K. (1983). *J. Am. Chem. Soc.* **105**, 1328–1332.

Babcock, M. S. & Olson, W. K. (1993). *Computation of Biomolecular Structures: Achievements, Problems, and Perspectives*, edited by D. M. Soumpasis & T. M. Jovin, pp. 65–85. Heidelberg: Springer-Verlag.

Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Deveny, T., Hsiek, S.-H., Srinivasan, A. R. & Schneider, B. (1992). *Biophys. J.* **63**, 751–759.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.

Brünger, A. T. (1992*a*). *X-PLOR. Version 3.1. A System for X-ray Crystallography and NMR*. Yale University, New Haven, CT, USA.

Brünger, A. T. (1992*b*). *Nature (London)*, **355**, 472–474.

Brünger, A. T. (1993). *Acta Cryst.* D**49**, 24–36.

Brünger, A. T., Krukowski, A. & Erickson, J. W. (1990). *Acta Cryst.* A**46**, 585–593.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Driessen, H., Haneef, M. I. J., Harris, G. W., Howlin, G., Khan, G. & Moss, D. S. (1989). *J. Appl. Cryst.* **22**, 510–516.

Egli, M., Williams, L. D., Frederick, C. A. & Rich, A. (1991). *Biochemistry*, **30**, 1364–1372.

EMBO Cambridge Workshop (1989). *EMBO J.* **8**, 1–4.

Gao, Y.-G., Laiw, Y.-C., Robinson, H. & Wang, A. H.-J. (1990). *Biochemistry*, **29**, 10307–10316.

Gewirth, D. T. & Sigler, P. B. (1995). *Nature Struct. Biol.* **2**, 386–394.

Hahn, M. & Heinemann, U. (1993). *Acta Cryst.* D**49**, 468–477.

Hendrickson, W. A. & Konnert, J. H. (1981). In *Biomolecular Structure, Conformation, Function and Evolution*, edited by R. Srinivasan, Vol. 1, pp. 43–57. Oxford: Pergamon Press.

Jochimiak, A., Haran, T. E. & Sigler, P. B. (1994). *EMBO J.* **13**, 367.

Jones, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.

Jones, T. A. & Cambillau, C. (1989). *Silicon Graphics Geometry Partners Directory*, Vol. 89. Silicon Graphics Inc., Mountain View, CA, USA.

Kleywegt, G. J. (1995). *Jnt CCP4 ESF–EACBM Newslett. Protein Crystallogr.* **31**, 45–50.

Kopka, M. L., Fratini, A V., Drew, H. R. & Dickerson, R. E. (1983). *J. Mol. Biol.* **163**, 129–146.

Langlois d'Estaintot, B., Gallois, B., Brown, T. & Hunter, W. N. (1992). *Nucleic Acids Res.* **20**, 3561–3566.

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**, 38–64.

Liaw, Y.-C., Gao, Y.-G., Robinson, H., van der Marel, G. A., van Boom, J. H. & Wang, A. H.-J. (1989). *Biochemistry*, **28**, 9913–9918.

Luzzati, V. (1952). *Acta Cryst.* **5**, 802–810.

Otwinowski, Z. (1991). *DENZO. A Film Processing Program for Macromolecular Crystallography*. Yale University, New Haven, CT, USA.

Otwinowski, Z., Schevitz., R. W., Zhang, R.-G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F. & Sigler, P. B. (1988). *Nature (London)*, **335**, 321–329.

Privé, G. G., Heinemann, U., Chandrasegaran, S., Kan, L.-S., Kopka, M. L. & Dickerson, R. E. (1987). *Science*, **238**, 498–504.

Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.

Saenger, W., Hunter, W. N. & Kennard, O. (1986). *Nature (London)*, **324**, 385–388.

Sheldrick, G. M. (1993). *SHELXL93. Program for Crystal Structure Refinement*. University of Göttingen, Germany.

Smith, C. K., Davies, G. J., Dodson, E. J. & Moore, M. H. (1995). *Biochemistry*, **34**, 415–425.

Taylor, R. & Kennard, O. (1982). *J. Am. Chem. Soc.* **104**, 3209–3212.

Westhof, E., Dumas, P. & Moras, D. (1985). *J. Mol. Biol.* **184**, 119–145.

Williams, L. D., Egli, M., Gao, Q., Bash, P., van der Marel, G. A., van Boom, J. H., Rich, A. & Frederick, C. A. (1990). *Proc. Natl Acad. Sci. USA*, **87**, 2225–2229.