Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Description of ordered solvent molecules in a platinated decanucleotide duplex refined at 1.6 Å resolution against experimental MAD phases

Franck Coste,[a] William Shepard[b]† and Charles Zelwer[a]*

[a]Centre de Biophysique Moléculaire, Centre National de la Recherche Scientifique, affiliated to the University of Orléans and to the Institut National de la Santé et de la Recherche Médicale, Rue Charles Sadron, 45071 Orléans CEDEX 2, France, and [b]Laboratoire pour l'Utilisation du Rayonnement Electromagnétique, BP 34, Bâtiment 209d, Centre Universitaire Paris-Sud, 91898 Orsay CEDEX, France

† Current address: ESRF, BP 220, 38043 Grenoble CEDEX, France.

Correspondence e-mail: zelwer@cnrs-orleans.fr

Received 30 July 2001
Accepted 18 December 2001

NDB Reference: platinated decanucleotide duplex, DDJ075.

Accurate experimental phases derived from a MAD experiment may be useful to enable the identification of solvent molecules during the course of an atomic parameter refinement. The structure of a double-stranded DNA decanucleotide bearing a cisplatin interstrand cross-link at 1.6 Å resolution, whose phases were first determined experimentally using the L_{III} edge of the Pt atom, was refined by various methods. The previously published structure resulted from a least-squares refinement using the structure-factor magnitudes and stereochemical restraints (program *SHELX*). In this paper, these previous results are compared with a model obtained by the likelihood-maximization method (program *REFMAC*) which allows the combination of the observed magnitudes with experimental MAD phases. This solution corresponded to a lower R_{free} (18.8 compared with 20.3%), a lower R factor and accounted for 135 water molecules and one spermine molecule collected by the program *wARP* during refinement. The previously published *SHELX* solution exhibited no spermine molecule and accounted for 92 water molecules, only 74 of which are also present in the model obtained with the MAD phases. In order to verify that these improvements were actually related to the use of the MAD phases, the same type of procedure without the MAD phases was applied starting from the initial model. The resulting solution had a higher R_{free} (20.3%), which could be related to the loss of 22 water molecules and the addition of 20 new ones. MAD phases therefore seem especially helpful in preventing the model bias which may affect the solvent molecules. All models have in common a hydration cage of nine water molecules which surround the platinum residue. In addition to the spermine molecule, the model obtained with the MAD phases allows description of the water-molecule organization, with reproducible motifs related to the base pairs and to the phosphodiester backbone.

1. Introduction

In macromolecular crystallography, the number of measured intensities is often insufficient to refine the individual atomic parameters reliably and/or accurately when the diffraction data does not reach atomic resolution. Usually, the experimental diffraction data must be supplemented with chemical information in the form of stereochemical constraints or the minimization of the internal energy of the molecule in order to improve the balance between the number of observations and the number of unknown parameters to refine. However, accurate stereochemical information is lacking for the ordered solvent molecules which are associated with the macromolecular structures. To date, there is little means available to

enable distinction between features caused by ordered solvent molecules and those reflecting errors of experimental origin or arising from model bias. In 1996, Burling and coworkers included in a least-squares refinement the minimization of the differences between the computed structure-factor phases and those estimated in a MAD experiment (Burling *et al.*, 1996). This was performed for the structure determination of a mannose-binding protein at 1.8 Å resolution. The resulting model exhibited better statistics and lower thermal parameters than in the refinement without the experimental phases. This new refinement allowed the identification of many ordered solvent molecules as well as partially ordered solvent shells. Enhancing the power of the model refinement with accurate experimental phases may help to assign solvent molecules reliably in the electron-density maps and to improve the accuracy of the macromolecular model, especially when the scattering limit of the crystals does not approach atomic resolution. To our knowledge, this refinement method based on experimentally derived phases does not seem to be used regularly in the case of structures derived from MAD experiments and in particular has never been applied to the case of oligonucleotides.

Hydration of nucleic acids has frequently been discussed at the structural level. In DNA, the recognition of particular sequences by nuclear proteins may be mediated by a primary recognition of water molecules as in the case of the tryptophan repressor–DNA complex (Schwabe, 1997). In addition, DNA is a flexible molecule in which the different conformations exhibit different patterns of hydration. Although in some cases the hydration patterns have been extensively studied (the Drew–Dickerson DNA dodecamer; Kopka *et al.*, 1983; Chiu *et al.*, 1999), there is a need for systematic investigations relying on high-resolution crystal structures. A recent compilation of nucleic acid crystal structures pointed out that about 30% of the water molecules may be incorrectly positioned (Das *et al.*, 2001). Anomalous diffraction data provide the ability to experimentally determine the structure-factor phases and therefore to increase the number of structure-factor observations at a given resolution. Since the anomalous scattering contribution to the structure factors is effectively constant with resolution, the effect of experimentally derived phases on the refinement accuracy is relatively strong at higher resolution. However, even if diffraction data are not available at atomic resolution, valuable information useful to correct the bias owing to the model can be obtained if MAD phases are included in the model refinement.

The structure of a double-stranded DNA containing a cisplatin [*cis*-diamminedichloroplatinum (II)] interstrand cross-link was solved at 1.6 Å resolution in previous work (Coste *et al.*, 1999). The anomalous scattering of the Pt atom was the unique source of phase information. The experimental electron-density map was of high quality, allowing each atom from the cross-linked DNA molecule to be assigned unambiguously (Fig. 1). The model was refined conventionally (*i.e.* against scattered magnitudes only) with *SHELX*97 (Sheldrick, 1997) and 92 water molecules could be identified from σ_A-weighted difference ($mF_o - DF_c$) maps using the program *ARP* (Lamzin & Wilson, 1993). Nine of these water molecules constitute a cage around the platinum residue: two of them are located at 3.6 Å from the platinum in the axial positions of the local fourfold axis of the platinum coordination (Coste *et al.*, 1999). This could explain the instability of this particular adduct under physiological conditions (Perez *et al.*, 1997). In this paper, we describe the results of a new refinement of the model, using the program *REFMAC* (version 4.0; Collaborative Computational Project, Number 4, 1994) with and without the special option which allows refinement against experimental or partial model phases. We discuss the stereochemical information drawn from the water molecules determined by the new refinement with MAD phases and included in the refined model.

2. Phase determination and model refinement

2.1. Crystallization conditions and data collection

The sequence of the double-stranded DNA decamer containing an interstrand cross-link is d($C_1C_2T_3C_4G^*_5C_6$-$T_7C_8T_9C_{10}$)·d($G_{11}A_{12}G_{13}A_{14}G^*_{15}C_{16}G_{17}A_{18}G_{19}G_{20}$), where G^*_5 and G^*_{15} are guanine residues cross-linked by *cis*-diammine-dichloroplatinum (II) (cisplatin) at their N7 position. The crystals have been obtained by vapour diffusion at 277 K from a drop containing the cross-linked DNA molecule (82.5 µ*M*), 20 m*M* sodium cacodylate pH 6.0, 5 m*M* NaCl, 30 m*M* KCl, 2.5 m*M* spermine, 2%(*v*/*v*) 2-methyl-2,4-pentanediol (MPD) and equilibrated against a buffer containing 2.5% MPD. The crystals belong to space group *C*2 with one molecule in the asymmetric unit and diffracted to 1.6 Å resolution on the DW21b beamline at LURE. The unit-cell parameters are $a = 42.89$, $b = 29.98$, $c = 46.57$ Å, $\beta = 95.98^\circ$. For data collection at 100 K, the crystals were transferred into a buffer containing 40%(*v*/*v*) MPD with the other concentrations unchanged. One crystal mounted with its twofold axis close to the rotation axis of the spindle was used for data collection at four wavelengths about the platinum L_{III} edge (1.091, 1.072, 1.071 and 1.030 Å). Other details are given in Coste *et al.* (1999).

Figure 1
Initial electron-density map (1σ contour) and refined model of a G·C base pair. Model and density contour are from *TURBO-FRODO* (Roussel & Cambillau, 1991). C atoms are green, P atoms are orange, N atoms are blue and O atoms are red.

2.2. Structure determination

Determination of the structure-factor phases was performed with the program *SHARP* (de La Fortelle & Bricogne, 1997), taking λ_4 magnitudes (high-energy remote) as the native data set (pseudo-MIR method). After convergence, high figures of merit (overall FOM = 0.84 and FOM = 0.77 in the 1.74–1.63 Å resolution shell) and phasing-power values (overall, 5.1 for the high-energy remote amplitudes and 8.6 for the peak) were obtained. The MAD electron-density map is unambiguously interpretable at the atomic level. Solvent flattening (*SOLOMON*; Abrahams & Leslie, 1996), however, yielded a map with a higher contrast, which was used to build the initial model. The O atoms of the phosphate groups have distinct density maxima and all the bond angles are clearly defined for all non-H atoms of the macromolecule. An example of the quality of this initial map superimposed with the refined atomic model is given in Fig. 1 for one of the G·C base pairs.

2.3. Model refinements

The initial model built with *TURBO-FRODO* v.5.5 (Roussel & Cambillau, 1991) gave an *R* factor of 32%. A first model refinement based on structure-factor amplitudes only was performed with *SHELX*97. Alternate sets of least-squares refinement cycles and analysis of σ_A-type weighted difference maps ($2mF_o - DF_c$ and $mF_o - DF_c$) using *ARP* led to the localization of 92 water molecules. These water molecules had densities above 1σ and were checked for hydrogen-bond stereochemistry consistency. For the last cycles of refinement, anisotropic temperature factors were refined for both the backbone P and Pt atoms (2091 individual parameters, 407 restrained atoms, 92 unrestrained atoms). In this refinement, the final *R* and $R_{\rm free}$ factors were 16.92 and 20.32%, respectively. Other statistics are given in Table 1.

To take advantage of the high quality of the MAD phases, a second refinement was undertaken using *REFMAC*. In this we used the MAD phases derived from the program *SHARP* with their corresponding probabilities given by the Hendrickson–Lattman coefficients. Furthermore, the solvent flattening is able to introduce a bias whose effects would be opposite to the collection of ordered solvent molecules. Indeed, we estimated that the initial phases from *SHARP* have a more straightforward physical meaning than the phases derived from *SOLOMON*, which depend on the definition of a mask, itself subjected to possible errors. This program has the option of including in the refinement not only the observed moduli and the constraints of the bond distances and angles, but also the structure-factor phases of any origin provided that they are associated with individual probabilities. The minimization method used was based upon gradient minimization associated with maximum-likelihood ranking. The use of the MAD phases is equivalent to augmenting the number of structure-factor observations with an equal number of experimental values which are generally less accurately known than the magnitudes (we suppose that the probability functions associated with the MAD phases are well represented by the Hendrickson–Lattman coefficients). The initial unrefined model of the cross-linked DNA molecule was taken as a starting point for this new refinement and did not include any solvent molecules. Stereochemical restraints were applied to all non-H atoms from the DNA molecule and isotropic thermal parameters were refined for all atoms. When the *R* factor reached the value of 22%, we began to add solvent molecules located as before in σ_A-type weighted difference maps from the program *wARP* ($2mF_o - DF_c$ and $mF_o - DF_c$). After several steps of refinement and introduction of water molecules in the model, we identified a continuous stretch of density in which we could build 12 of the 14 atoms of a spermine molecule. A total of 135 putative water molecule O atoms have been included in the last refinement cycles as well as the partially built spermine molecule. The results of this refinement using experimentally derived phases (*REFMAC*/*MLKMAD*) and its comparison with the more conventional refinement (*SHELX*) are summarized in Table 1 and Fig. 2.

Table 1
Comparison of the refinement statistics obtained without (*SHELX*, *REFMAC*/*MLK*) and with (*REFMAC*/*MLKMAD*) the MAD phases.

	SHELX	*REFMAC*/*MLKMAD*	*REFMAC*/*MLK*
R factor (%)	16.92	15.85	15.63
$R_{\rm free}$ (%)	20.32	18.83	20.32
Total No. of atoms	499	554	552
No. of solvent atoms			
Water	92	135	133
Spermine	0	12	12
Mean *B* factors (Å^2)			
DNA	18.07	15.04	16.36
Solvent	30.25	29.88	27.83
R.m.s. on bond distances and angles (Å)			
Distances	0.007	0.010	0.010
Angles	0.016	0.026	0.026

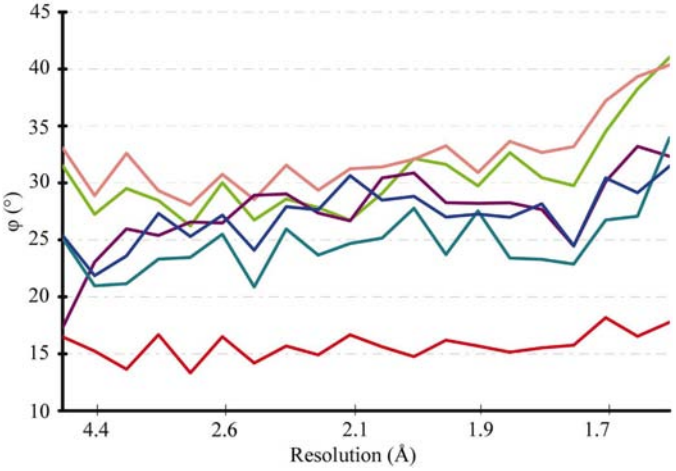


Figure 2
Statistics of phase differences between MAD (exp), *SOLOMON*, *SHELX* and *REFMAC* phases. Resolution shells contain an approximately equal number of reflections (350). Light green, exp/*REFMAC*; orange, exp/*SHELX*; purple, exp/*SOLOMON*; red, *REFMAC*/*SHELX*; dark green, *REFMAC*/*SOLOMON*; blue, *SHELX*/*SOLOMON*.

Although anisotropic temperature factors were not applied to the P and Pt atoms in the *REFMAC* refinement, the R and R_{free} factors are lower than those resulting from the *SHELX* refinement (R = 16.0% compared with 17.0%; R_{free} = 18.6 compared with 20.3%). In addition, the average of the individual B factors decreases from 18 to 15 Å^2 for the DNA molecule. The B factors of the two refinements can be compared in Fig. 3. This difference in B factors has an effect on the apparent resolution of the maps. With the experimental MAD phases and with the *SOLOMON* phases, the purine rings have shapes where each valence angle can be identified (Fig. 1), whereas surprisingly with the *SHELX* phases, the corresponding rings exhibit holes of a smaller size for the hexagonal moieties and have a smoother appearance. As in the first refinement, all occupancy ratios for the solvent molecules were set to 100%. The atomic coordinates for the cross-linked DNA had an r.m.s. displacement of 0.11 Å with respect to the *SHELX* model. The strongest shifts, which occur for the phosphate O atoms, are partly a consequence of the fact that anisotropic thermal parameters were not used for the P atoms in this case.

Fig. 2 gives a comparative analysis of discrepancies between the different sets of phases (for simplicity, we denote the conventional *SHELX*97 refinement as the *SHELX* model, the *REFMAC* refinement with phases as the *REFMAC* model, the initial phases from *SHARP* before solvent flattening as MAD phases and those resulting from solvent flattening as the *SOLOMON* phases). The phase differences between the two refined models vary from 13 to 18° (average 16°). For the *REFMAC* refinement the differences from the experimental phases vary from 26° (2.7 Å resolution) to 40° (1.6 Å resolution) and the average discrepancy is 30.5°. For the *SHELX* refinement, the extrema are 28 and 40° and the average is 32°. We notice that the (*REFMAC* − experimental) phase differences curve is systematically below the (*SHELX* − experimental) curve, except for the 1.66 Å shell. Furthermore, the *REFMAC* phases are closer to those derived from *SOLOMON* than the *SHELX* phases by an average of 2°. This is interesting, as the *SOLOMON* phases correspond to a map corrected for errors in the cell volume occupied by the bulk solvent and have not contributed to the refinement. However, we may notice that beyond 1.75 Å all the slopes of the difference curves increase significantly and this may mean that experimental MAD and *SOLOMON* phases become less accurate. This is not surprising as the resolution limits are 1.72 Å for the low-energy remote data, 1.69 Å for the peak and inflexion data, and 1.63 Å for the high-energy remote and the corresponding R_{sym} values are higher (Coste *et al.*, 1999).

2.4. Identification of water molecules

The water molecules were identified automatically by the program *ARP* from ($mF_o - DF_c$) and ($2mF_o - DF_c$) maps with phases computed from the atomic model after each range of refinement cycles. Afterwards, the density of each new solvent molecule was inspected visually and then included definitively into the atomic model only if its hydrogen-bonding stereochemistry is sensible and its B factors refined below 50 Å^2. A total of 18 water molecules from the *SHELX* model are absent from the *MLKMAD* list and 74 water molecules (identified after the visual inspection of the list of atoms sorted according to one of the atomic parameters) are common to both models. Since it was not possible to relate the water molecules from the *SHELX* model to characteristic patterns (except for the water molecules in contact with the platinum residue), it is difficult to relate these 18 water molecules to specific interactions. On the contrary, in the case of the *MLKMAD* model it has been possible to relate the ordered water molecules to reproducible interactions along the DNA molecule. We may therefore suspect that the 18 solvent molecules removed from the *SHELX* model were a consequence of model bias and that the addition of the 61 new water molecules helped to propose logical descriptions for the hydration patternslloyd

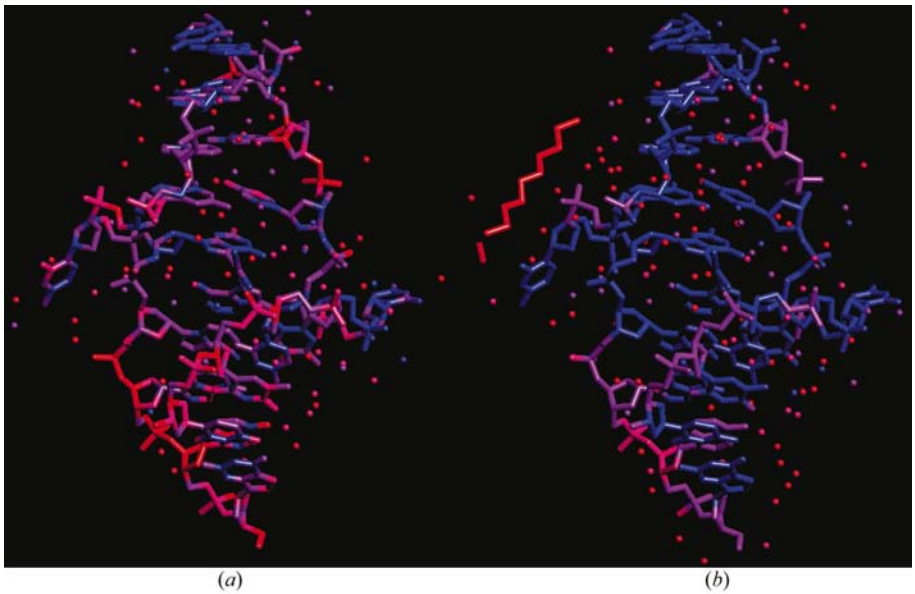


Figure 3
Colour representation (*TURBO-FRODO*) of the thermal B factors for both refined models. (*a*) Without MAD phases (*SHELX*), (*b*) with MAD phases (*REFMAC*). Dark blue, 12 Å^2; red, 30 Å^2.

2.5. Identification of a spermine molecule during the refinement process

In the course of the refinement process, a continuous stretch of density appeared in an area close to the extrahelical cytosine C_{16}. The program *ARP* tentatively interpreted this electron density as a cluster of water molecules. However, owing to the close contacts the density could not be properly modelled as a cluster of water molecules. We verified that the map could convincingly accommodate 12 non-H

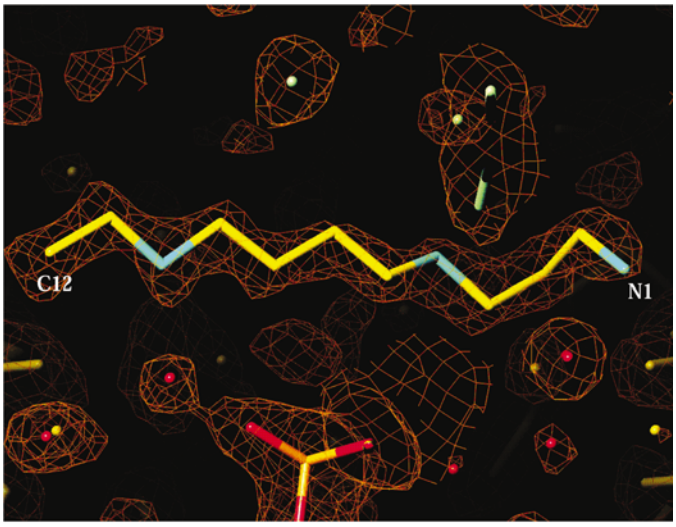
atoms from a spermine molecule over 14 (from N1 to C12). The atoms from C6 to C12 were not stable when further refined again against the MAD experimental phases. Assigning the spermine molecule an occupancy level of 0.8 helped the refinement procedure to reach a stable solution. The electron density around the spermine molecule is shown in Fig. 4.

2.6. *REFMAC* refinements based on structure-factor magnitudes only and/or without stereochemical restraints

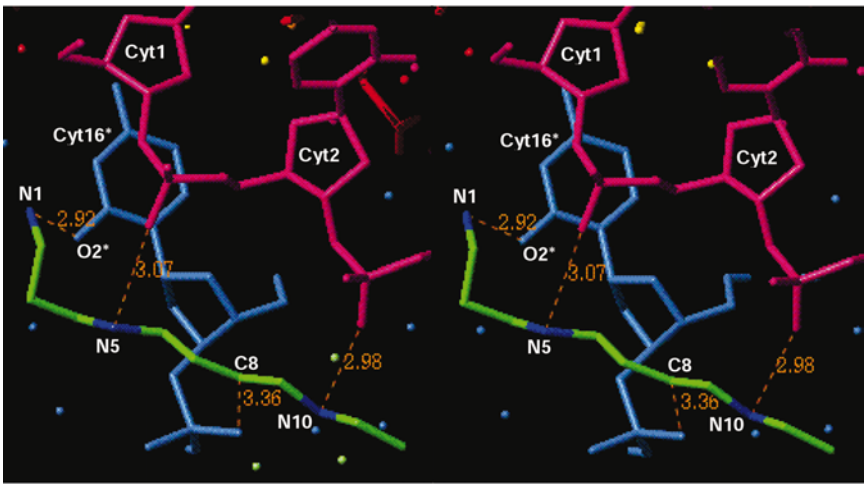
Since *SHELX* and *REFMAC* are different programs using different minimization methods to refine the atomic models, the question of the origin of the improvements obtained with *REFMAC* must be addressed. More precisely, the use of likelihood ranking might improve the model by itself according to our criteria. Another fact which may complicate the interpretation of the results is that the refinements performed with *SHELX* and *REFMAC* did not involve the same number of atomic parameters for the DNA molecule. In *SHELX* the possibility has been implemented of refining the anisotropic thermal parameters for specified atoms only, which is not the case with *REFMAC*. The question of the number of solvent molecules which may artificially improve the refinements could also be raised. As we pointed out in the introduction, we are aware of two main criteria which are not completely independent in checking the quality of the structure refinement: (i) being able to assign a maximum number of solvent molecules on the basis of an automatic analysis of the electron density in the course of the refinement and (ii) avoiding the possibility that this electron density is biased by the model. The program *REFMAC* allowed us to compare two refinements performed in parallel with and without the MAD phases by maximum-likelihood ranking. The corresponding refinement statistics are given in Table 1. The $R_{\rm free}$ factor is higher by 1.5% when MAD phases are omitted, whereas the R factors have comparable values. The mean thermal parameters have intermediate values between the *SHELX* and *MLKMAD* refinements. The examination of the set of solvent molecules shows that the two solutions have 115 water molecules and the partially occupied spermine molecule in common. Conversely, 22 water molecules from the *MLKMAD* model are absent from the *MLK* model and 20 other water molecules present in the *MLK* model are absent in the *MLKMAD* model.

Now, if we start from the *MLKMAD* model and refine it without the MAD phases, it remains stable and the R factors are even better than with the *MLKMAD* refinement (data not shown). This means that the *MLKMAD* solution is also compatible with the structure-factor magnitudes alone when refined with the *MLK* conditions. Conversely, if the *MLK* procedure is replaced by the classical least-squares option, 15 water molecules disappear from a $(2F_o - F_c)$ map and the spermine molecule is altered. Editing these solvent molecules would probably converge towards the *SHELX* solution.

Another means to check the relative contributions of likelihood ranking *versus* use of MAD phases, lies in the comparison of the DNA atomic models corresponding to the different refinement procedures (MLKMAD, *MLK*, *LSQ*) in the absence of stereochemical restraints. This comparison (Table 2) shows that *MLKMAD* shifts are the smallest and *LSQ* ones the largest. R.m.s., mean and maximum shifts for the *MLK* procedure are approximately halfway between *MLKMAD* and *LSQ*. This is a confirmation that both MAD phases and likelihood ranking contribute to refinement results and subsequent model improvements.



(*a*)



(*b*)

Figure 4
(*a*) Electron-density map (0.8σ contour) and refined partial model of the spermine molecule given by *TURBO-FRODO*. C and P atoms are yellow, N atoms are blue. (*b*) Stereo pair showing the hydrogen-bond linkage of the spermine to two neighbouring oligonucleotide molecules. C atoms of spermine are green and N atoms are blue; one DNA molecule is magenta and a symmetry-related molecule is blue. Figure generated by *MOLSCRIPT* (Kraulis, 1991) and *Raster3D* (Merritt & Murphy, 1994).

3. Structural and biological importance of the ordered solvent molecules

At the end of the refinement process based on the experimental values of the structure-factor moduli and phases, 135 solvent molecules and one partial spermine molecule involved in intermolecular contacts have been identified. The solvent molecules were assigned in the first instance as water molecules. The chemical nature of the solvent molecules and ions present in the crystals are given in the crystallization conditions. The K^+ ions (30 m*M*) and the Cl^- ions (35 m*M*) cannot in principle be confused with water molecules because of their putative high electron density. Only Na^+ ions (25 m*M*) could be confused for water molecules as they have the same number of electrons. The question of the possible attribution of solvent peaks to Na^+ ions has been discussed recently in the case of the structure of the Drew–Dickerson dodecamer (Shui *et al.*, 1998; Chiu *et al.*, 1999). We applied the equation of Brown & Wu (1976) to identify possible cation sites (*WASP* program; Nayal & Di Cera, 1996), but we did not obtain any clear results for the presence of Na^+ ions (data not shown). For the substitution of water molecules by K^+ ions, six sites have valence strengths larger than 1. However, the corresponding atoms refined as oxygen have large *B* values and the corresponding densities were not convincing for fully occupied K^+ sites. We therefore consider all the well resolved single solvent peaks to represent O atoms from water molecules.

With respect to the *SHELX* refinement, the ratio of water molecules per nucleotide increases from 4.5, which is the average number found for 15 B-DNA molecules in the PDB by Schneider *et al.* (1998), to 6.9, in spite of the suppression of 18 water molecules from the *SHELX* model. Of the 74 water molecules occupying similar positions, only those surrounding the platinum residue seemed to belong to a well defined motif. Conversely, the water molecules of the *REFMAC* model exhibit four different regions of interaction: with the platinum residue in the minor groove, with the Watson–Crick base pairs, with the phosphodiester backbone and between neighbouring symmetry-related molecules. The interactions with the platinum residue have been already well described (Coste *et al.*, 1999) and are conserved in the *REFMAC* model. One water molecule bound to an N atom of an ammine group

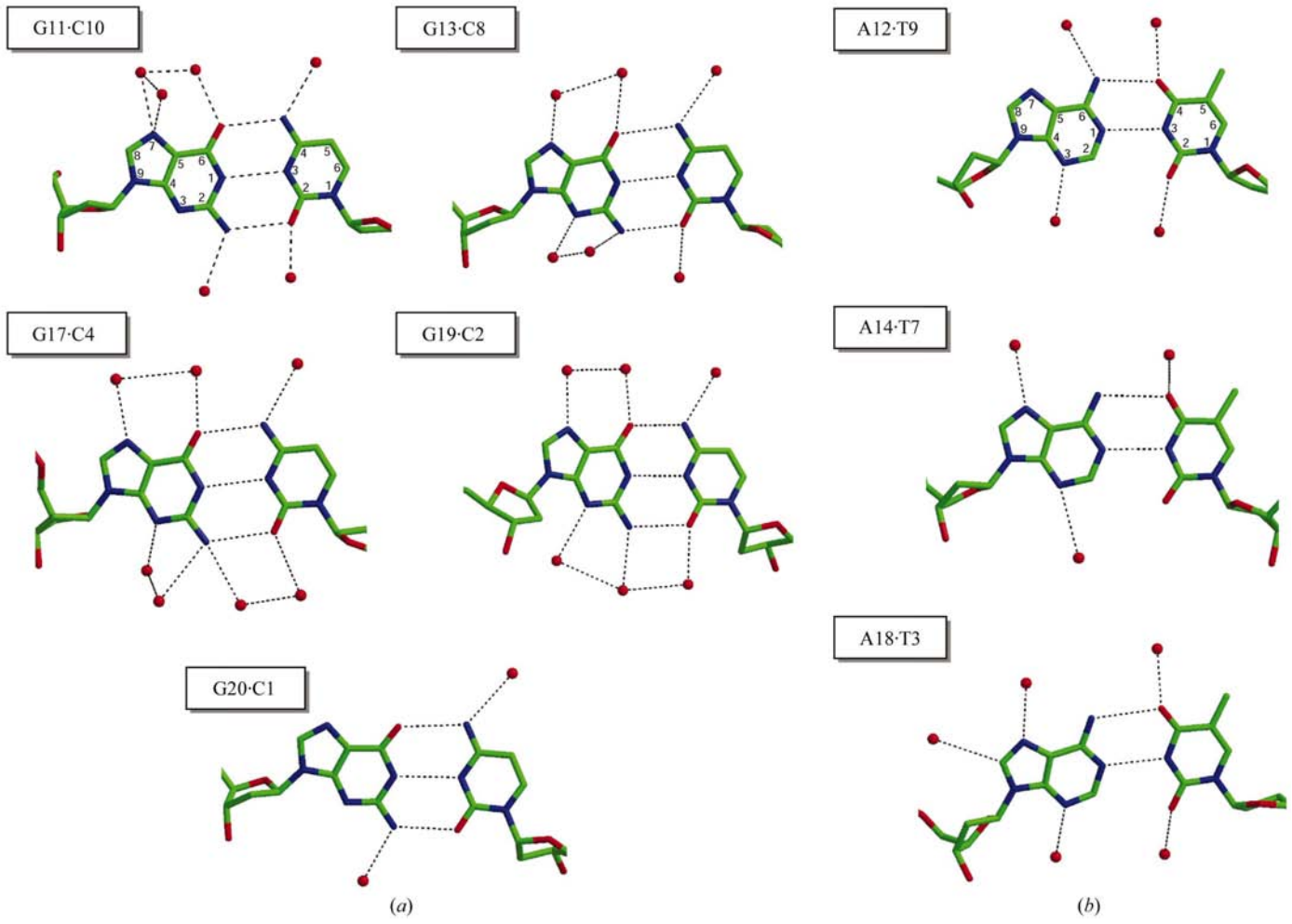


Figure 5
The eight Watson–Crick base pairs and the surrounding water molecules (red). (*a*) G·C pairs, (*b*) A·T pairs. Figure generated by *MOLSCRIPT* and *Raster3D*.

undergoes a displacement of about 1 Å, which gives a more symmetrical shape to the cage surrounding the platinum coordination sphere with respect to the platinum coordination plane. These results from the new refinement using experimentally derived phases now allow an exhaustive description of the interactions of other solvent molecules.

3.1. Hydration of the Watson–Crick base pairs.

On the basis of structural models deposited in the NDB, Schneider *et al.* (1992, 1993), Berman (1994, 1997) and Auffinger & Westhof (2000) proposed some general rules and gave averaged patterns of hydration for the two different types of standard Watson–Crick base pairs. In the case described in this paper, some parts of the DNA molecule have different conformations not representative of the classical B-DNA structure, such as the cytosines C_5 and C_{16} that extend into the solvent and the complementary guanines cross-linked to the platinum moiety. These large distortions modify the arrangement of the water molecules in the vicinity of the lesion only. However, the other base pairs shown in Fig. 5 display strong similarities to the standard patterns given in the reviews cited above. If we exclude the terminal base pairs $C_1{\cdot}G_{20}$ (which makes a triplet with C_6 from another molecule) and $C_{10}{\cdot}G_{11}$, the three other G·C pairs are quite similar, especially in the major groove which does not contain the platinum residue. For less disturbed parts of the DNA molecule, the atoms N7, O6 and O2 of the guanines are always bound to a water molecule. The N3 positions of the purines as well as the 2 and 4 positions of the pyrimidines are also hydrated (except for the terminal G·C pairs). All these O atoms from water molecules lie close to the mean plane of the Watson–Crick base pairs and follow the propeller twist. In the three A·T base pairs, the N6 of A_{12} is hydrated as well as the N7 of A_{14} and A_{18}. The heterocyclic O4 from the three T nucleotides (T_3, T_7 and T_9) are hydrated. The pair $T_7{\cdot}A_{14}$ is close to the lesion and O2 of T7 makes a hydrogen bond with one of the ammine groups of the platinum residue instead of a water molecule. This specificity is related to a shift of the water molecule bound to N3 of A_{14}.

3.2. Hydration of the phosphodiester backbone

While the hydration of DNA base pairs is well documented, the hydration of the phosphodiester backbone is less well known. Nonetheless, the hydration of the phosphate groups is of particular importance in the overall stabilization and in the transitions of nucleic acids from one polymorph to another (Saenger *et al.*, 1986). In B-DNA, no regular hydration patterns at neighbouring phosphate groups have been observed while water molecules bridging two O(II) atoms are frequently seen in A-DNA crystal structures (Schneider *et al.*, 1998; Wahl & Sundaralingam, 1997). Our *REFMAC* refinement reveals a regular pattern of water molecules along the two strands of the phosphodiester backbone in the areas where the double-helical structure is conserved and in the absence of intermolecular contacts (Fig. 6). Indeed, the O(II) atom at the 5′ side of a nucleotide is linked to the O(I) atom of the phosphate at the 3′ side by two consecutive water molecules. When the nucleotides are not in standard B conformation, there may be a linkage between two successive O(II) atoms or by a single water molecule. The linkage by two water molecules could be a characteristic of B-DNA. Egli *et al.* (1996) describes an RNA with consecutive O(I) phosphates bridged by a single water molecule. In a recent molecular-dynamics study (Auffinger & Westhof, 2000), several adjacent phosphate groups of a RNA duplex were bridged by one water molecule. However, it is interesting to note that in the same study no long-lived water bridge was found between the major-groove phosphate groups of a B-DNA duplex. It seems that the phosphate–phosphate distance and the

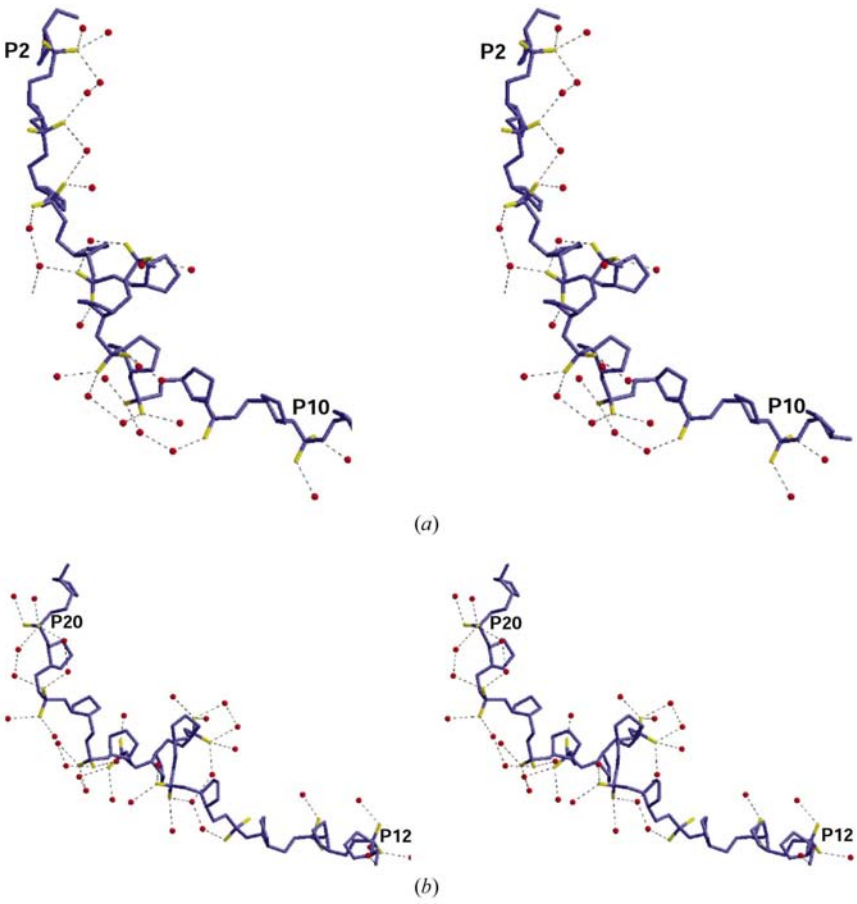


Figure 6
Stereo pairs for (*a*) the upper and (*b*) the lower phosphodiester backbone strands (blue) and hydrogen-bond linked water molecules (red). Figure generated by *MOLSCRIPT* and *Raster3D*.

hydration of phosphate groups are directly related (Saenger *et al.*, 1986). The type of nesting between adjacent phosphates found in this platinated DNA molecule seems to exclude the presence of monovalent cations close to the phosphate groups. More particularly, the coordination of these solvent peaks is not characteristic of a monovalent cation and these sites were not identified by the equation of Brown & Wu (1976). The cations are probably present in the bulk solvent and the negative charge of the phosphate groups may be dislocated between the different O atoms including those of the water molecule bridge. This network of water molecules bound to the phosphate moieties of the phosphodiester backbone shields the deoxyribose moieties and especially protects the 3′ and 5′ ester functions against hydrolysis.

3.3. Intermolecular contacts

In the case of crystals of oligonucleotide crystals, the stacking of terminal base pairs mimics the continuity of the double helix, ensuring the crystal cohesion in one direction. In our case, a twofold axis is close to both terminal base pairs, so each terminal base pair ($C_{10}\cdot G_{11}$ and $C_1\cdot G_{20}$) stacks with the same base pair from a neighbouring molecule (Coste *et al.*, 1999). In the case of the $C_1\cdot G_{20}$ pair, a triplet is also formed with the extra-helical cytosine C_6 from another molecule, ensuring the cohesion in a direction perpendicular to the mean double-helix axis. The extra-helical cytosine C_{16} interacts at the side opposite to C_6 with a C_1 phosphate group from another molecule.

Some putative water molecules are also involved in close contacts in the area of the extra-helical cytosine C_6. Interestingly, Wa3 and Wa4 form a bridge between two phosphate groups: the O(II) from nucleotide 5 from one molecule and the O(I) from the nucleotide 4 of an adjacent molecule (Fig. 7). This interaction seems to reproduce the type of nesting which has been found between two successive phosphate groups of the same strand. Wa3 and Wa4 have low B values for water molecules (about 12 Å^2), but their coordination number is too small for a metal ion. Wa3 is also involved with Wa60 in an intramolecular bridge between two successive phosphate groups.

The contacts of the spermine molecule are shown in Fig. 4(*b*). The distance between the two secondary amine groups (N5 and N10) is close to that of two successive phosphodiester groups (C2 and T3) and they form hydrogen bonds with the O(I) of residues C2 and T3. The spermine molecule shields the phosphodiester backbone in a way similar to the nesting of the water molecules which form intramolecular bridges. The N5 secondary amine group is also hydrogen bonded to Wa90, which is itself linked to the O(II) of residue G17 from another molecule. Here we again have a succession of two hydrogen-bond linkers which connect two phosphate groups from different molecules. The terminal primary amine from the spermine molecule N14 is hydrogen bonded to the O2 oxygen moiety from the extra-helical cytosine C16 which precedes the G17 residue involved in the interaction with Wa90. The N1 primary amine moiety of the spermine molecule is not visible in the electron-density maps and is probably disordered owing to the absence of any strong interactions.

4. Discussion

4.1. Refinement technique

A MAD experiment at 1.6 Å resolution around the Pt L_{III} absorption edge of a crystallized double-stranded decanucleotide bearing a cisplatin interstrand cross-link allowed us to determine accurate phases related to the prior determination of the Pt-atom position. The refinement of the atomic model by minimization of the mean-square error between the calculated structure factors and the phased observed amplitudes led to atomic parameters for the DNA molecule close to those resulting from a refinement performed with the observed moduli only. However, the atomic thermal parameters are significantly lower (3 Å^2) for the *SHELX* refinement when the MAD phases are used and this resulted in an increased contribution of the high-resolution reflections to the computation of the maps and a more accurate model. The main interest in such a procedure to refine macromolecular crystal structures lies in the identification of solvent molecules, because the difference maps from which such molecules are collected are less subjected to model bias. In fact, it would be much better to compute a map resulting from a combination of the MAD phases with those computed from the model, with a proper weighting of each source of information. Such a weighting would require a rigorous statistical treatment which is beyond the scope of this paper. In our work,

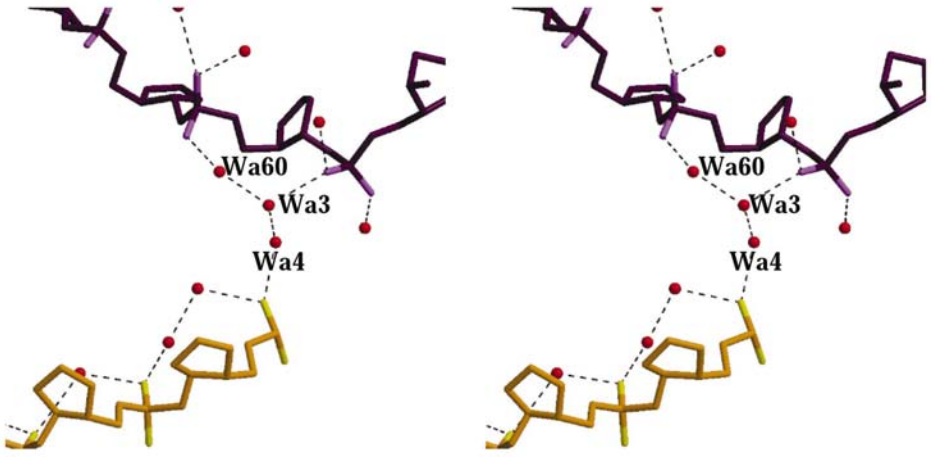


Figure 7
Stereo pair showing the water linkage between two neighbouring DNA molecules. Figure generated by *MOLSCRIPT* and *Raster3D*.

we showed that even at 1.6 Å resolution and with the use of an improved model only to compute the phases, the collection of solvent molecules surrounding the macromolecule yields results strikingly different than in the case of the traditional refinement method. In our case, 18 water molecules (of the 92 found after the refinement without the MAD phases) were removed, whereas 61 new water molecules have been added to the list. In addition, partially occupied atomic positions have been found for a spermine molecule. Since these solvent molecules may be important in the discussion of such topics as the recognition mechanism between macromolecules, enzymatic mechanisms, folding mechanism of proteins, oligomeric proteins assembly and crystallization processes, the reliability of their identification and location is of biological importance. Until now, although some researchers have scanned the databases to point out some interesting features related to solvent molecules, little work has been performed to check the reliability of this part of the data. Since the number of structures determined with MAD techniques, with Se labelling and with data from a unique crystal is increasing, it would be rewarding to generalize the use of MAD phases in refinement whatever the resolution. Of course, when atomic resolution is reached by the diffraction data, MAD phases are probably not mandatory. However, they regardless provide a means to verify the reliability of the solvent-molecule assignments by checking their stability during the refinement process.

The use of maximum-likelihood ranking without the MAD phases is able to improve the model significantly. Indeed, the *MLK* refinement (see Tables 1 and 2) and the examination of the corresponding difference maps indicate that the quality of the refinement is intermediate between *SHELX* and *MLKMAD*. The water molecules which disappear are peripheral ones, particularly those related to the phosphodiester backbone. It has not been possible to check with the program *REFMAC* a combination of the use of the MAD phases and the classical least-squares method as this possibility was not operational. Anyway, in the absence of stereochemical restraints, the use of MAD phases combined with likelihood ranking is able to maintain the atoms of the DNA molecules at relevant positions, better than likelihood ranking alone. This fact suggests that accurate MAD phases have played an important role in terms of the overdetermination of the parameters of the atomic model and would be able to help solving the problem of model bias at non-atomic resolution.

The use of experimental phases was first proposed by Burling *et al.* (1996) in a 1.8 Å resolution study of a protein. We used a similar approach to study the hydration of a DNA molecule at a higher resolution. In our case, conspicuous biochemical information has been derived from the identification of solvent molecules included in the DNA structure.

4.2. The physicochemical importance of ordered water molecules in the structure of DNA crystals

Macromolecular crystallography now allows detailed description of the interaction between proteins or nucleic acids and the ordered solvent molecules. However, three limitations have to be considered: (i) the crystallization mother liquor generally includes many components in addition to water, which may complicate the interpretation of the map features attributed to solvent or (ii) modify the structure of the hydration shell with respect to what occurs in the cellular medium and (iii) the intermolecular contacts which are specific to the crystal lattice are also able to interfere with the primary hydration layer. Nevertheless, the crystallographic analysis of the solvent–DNA interaction provides a unique opportunity to understand some aspects of the crystallization process at the atomic level and relate it to thermodynamic considerations.

Table 2
Comparison of shifts between the *MLKMAD* atomic positions from the DNA molecule obtained with stereochemical restraints and those obtained without such restraints after 120 cycles of *REFMAC* refinement under *MLKMAD*, *MLK* and *LSQ* conditions.

Unrestrained	*MLKMAD*	*MLK*	*LSQ*
R.m.s. shifts (Å)	0.128	0.153	0.166
Mean shifts (Å)	0.104	0.128	0.141
Max shift (Å)	0.528	0.677	0.742
R factor (%)	0.145	0.132	0.136
R_{free} (%)	0.206	0.208	0.204

In this study, two aspects of the hydration patterns are related to the specificity of the DNA molecule: the interactions with the DNA bases and the interactions with the platinum moiety. G·C and A·T base pairs display different hydration patterns. These patterns are modulated by the sequence context and/or the conformation of the DNA molecule. In our case, standard patterns are less affected by the distortions induced by the cross-link in the major groove than in the minor groove, where the adduct protrudes. This can be verified for the pairs $G_{13}\cdot C_8$, $G_{17}\cdot C_4$ and $G_{19}\cdot C_2$, for which the hydration pattern displays three different arrangements at the minor-groove side and only one at the major-groove side. The positions N7, O6, N2 of guanine residues and the positions N4 and O2 of cytosine are always hydrated in these three pairs. The pairs $G_{11}\cdot C_{10}$ and $G_{20}\cdot C_1$ are involved in intermolecular contacts, including the formation of a triplet ($G_{20}\cdot C_1$), and cannot be considered further. In A·T pairs, the N7 position of adenine is hydrated in $A_{18}\cdot T_3$ and $A_{14}\cdot T_7$, whereas for $A_{12}\cdot T_9$ the corresponding water molecule is shifted toward N6. For the three A·T pairs, the N4 from thymine is always hydrated. In the minor groove, the N3 from adenine is always hydrated, whereas the O2 interacts with a water molecule in T_9 and T_3 and in T_7 it is hydrogen bonded to an ammine group from the cisplatin adduct.

4.3. The role of solvent molecules in the cohesion of the crystals

Spermine molecules have been difficult to identify in crystals of nucleic acids, although they are frequently required for crystallization. To date, a partial spermine molecule (seven of the 14 atoms) has been located in only one B-DNA crystal structure (Shui *et al.*, 1998). A possible explanation of this difficulty in seeing the spermine is that polyamines play the

same role as cations in the non-specific shielding of the charged phosphate groups. This type of ionic interaction is less stable than hydrogen bonding and is subject to disorder. According to the results of a recent molecular-dynamics simulation, the highly flexible spermine molecules do not form structurally stable complexes with B-DNA (Korolev *et al.*, 2001). In our case, the spermine molecule has an extended conformation, enabling two of its amino groups to interact with two consecutive phosphates. The spermine interacts also with the extra-helical cytosine C_{16} in a non-ionic fashion and with the G_{17} phosphate group from another molecule. It seems that for the first time in a nucleic acid crystal a clear connection has been found between the requirement for spermine in crystallization and its role in crystal cohesion. Interestingly, the required concentration of spermine in the mother liquor is not high compared with other oligonucleotide crystals (2.5 m*M*); neither is that of the precipitating agent (2.5% MPD). In these crystals, the extra-helical cytosine residues are essential to link together the molecules in a direction perpendicular to the double-helix axis. Consistently, the spermine molecule helps to establish this interaction with the cytosine C_{16} which is not involved in a triplet formation like the cytosine C_6.

5. Conclusions

The use of MAD phases to enhance the strength of a structure refinement at 1.6 Å, combined with likelihood maximization, allowed determination of structural details at the boundary of the molecule, which are difficult to model even at this resolution. In nucleic acid biochemistry, knowledge of the structure of the ordered solvent molecules is essential to understand the interactions of these molecules with drugs and proteins. Experimental MAD phases could also be used to investigate the relationship between a given nucleic acid sequence and the structure of the ordered solvent shell systematically. Such experiments would require further progress in nucleic acid synthesis to incorporate strong anomalous scatterers at chosen positions within the DNA molecule and with minimal disturbance of the solvent shell.

References

Abrahams, J. P. & Leslie, A. G. W. (1996). *Acta Cryst.* D**52**, 30–42.
Auffinger, P. & Westhof, E. (2000). *J. Mol. Biol.* **300**, 1113–1131.
Berman, H. M. (1994). *Curr. Opin. Struct. Biol.* **4**, 345–350.
Berman, H. M. (1997). *Biopolymers*, **44**, 23–44.
Brown, I. D. & Wu, K. K. (1976). *Acta Cryst.* B**32**, 1957–1959.
Burling, F. T., Weis, W. I., Flaherty, K. M. & Brünger, A. T. (1996). *Science*, **271**, 72–77.
Chiu, T. K., Kaczor-Grzeskowiak, M. & Dickerson, R. E. (1999). *J. Mol. Biol.* **292**, 589–608.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Coste, F., Malinge, J. M., Serre, L., Shepard, W., Roth, M., Leng, M. & Zelwer, C. (1999). *Nucleic Acids Res.* **27**, 1837–1846.
Das, U., Chen, S., Fuxreiter, M., Vaguine, A. A., Richelle, J., Berman, H. M. & Wodack, S. J. (2001). *Acta Cryst.* D**57**, 813–828.
Egli, M., Portmann, S. & Usman, N. (1996). *Biochemistry*, **35**, 8489–8494.
Kopka, M. L., Fratini, A. V., Drew, H. R. & Dickerson, R. E. (1983). *J. Mol. Biol.* **163**, 129–146.
Korolev, N., Lyubartsev, A. P., Nordenskiold, L. & Laaksonen, A. (2001). *J. Mol. Biol.* **308**, 907–917.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
Lamzin, V. S. & Wilson, K. S. (1993). *Acta Cryst.* D**49**, 129–147.
Merritt, E. A. & Murphy, M. E. P. (1994). *Acta Cryst.* D**50**, 869–873.
Nayal, M. & Di Cera, E. (1996). *J. Mol. Biol.* **256**, 228–234.
Perez, C., Leng, M. & Malinge, J. M. (1997). *Nucleic Acids Res.* **25**, 896–903.
Roussel, A. & Cambillau, C. (1991). *Silicon Graphics Geometry Partners Directory*, pp. 77–78. Mountain View, CA, USA: Silicon Graphics.
Saenger, W., Hunter, W. N. & Kennard, O. (1986). *Nature (London)*, **324**, 385–388.
Schneider, B., Cohen, D. M., Schleifer, L., Srinivasan, A. R., Olson, W. K. & Berman, H. M. (1993). *Biophys. J.* **65**, 2291–2303.
Schneider, B., Ginell, S. L., Jones, R., Gaffney, B. & Berman, H. M. (1992). *Biochemistry*, **31**, 9622–9628.
Schneider, B., Patel, K. & Berman, H. M. (1998). *Biophys. J.* **75**, 2422–2434.
Sheldrick, G. M. (1997). *SHELX97. Program for Refinement of Crystal Structures.* University of Göttingen, Germany.
Shui, X., McFail-Isom, L., Hu, G. G. & Williams, L. D. (1998). *Biochemistry*, **37**, 8341–8355.
Schwabe, J. (1997). *Curr. Opin. Struct. Biol.* **7**, 126–134.
Wahl, M. C. & Sundaralingam, M. (1997). *Biopolymers*, **44**, 45–63.

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

addenda and errata

Description of ordered solvent molecules in a platinated decanucleotide duplex refined at 1.6 Å resolution against experimental MAD phases. Addendum

Franck Coste,[a] William Shepard[b] and Charles Zelwer[a]*

[a]Centre de Biophysique Moléculaire, Centre National de la Recherche Scientifique, affiliated to the University of Orléans and to the Institut National de la Santé et de la Recherche Médicale, Rue Charles Sadron, 45071 Orléans CEDEX 2, France, and [b]Laboratoire pour l'Utilisation du Rayonnement Electromagnétique, BP 34, Bâtiment 209d, Centre Universitaire Paris-Sud, 91898 Orsay CEDEX, France

An acknowledgment is published for the paper by Coste *et al.* [*Acta Cryst.* (2002) D**58**, 431–440]. The acknowledgement is as follows: This work was supported by grants from l'Association pour la Recherche sur le Cancer, and la Ligue Contre le Cancer du Loiret; FC was the recipient of a PhD fellowship from the Région Centre.

References

Coste, F., Shepard, W. & Zelwer, C. (2002). *Acta Cryst.* D**58**, 431–440.