Acta Cryst. (1993). D**49**, 468–477

DNA Helix Structure and Refinement Algorithm: Comparison of Models for d(CCAGGCm^{5}CTGG) Derived from *NUCLSQ*, *TNT* and *X-PLOR*

BY MICHAEL HAHN AND UDO HEINEMANN*

Institut für Kristallographie, Freie Universität Berlin, Takustrasse 6, D-1000 Berlin 33, Germany

(*Received* 24 *February* 1993; *accepted* 14 *May* 1993)

Abstract

In an earlier study [Heinemann & Hahn (1992). *J. Biol. Chem.* **267**, 7332–7341], the crystal structure of the double-stranded B-DNA decamer d(CCAGGCm^{5}CTGG) was refined with *NUCLSQ* to $R = 17.4\%$ against 3799 2σ structure amplitudes in the resolution range 8–1.7 Å. This structure has now been re-refined against the same diffraction data using either *TNT* or *X-PLOR* in order to determine to what extent the resulting DNA conformations would differ and to examine the suitability of these programs for the refinement of oligonucleotide structures. The *R* value from the *NUCLSQ* refinement could not be reached with either *TNT* or *X-PLOR*, although both programs yield reasonably refined DNA models showing root-mean-square deviations against the *NUCLSQ* model of the decamer duplex of 0.25 and 0.32 Å, respectively. Some derived local structure parameters differ depending on the refinement procedure used. This holds true for several exocyclic torsion angles of the sugar-phosphate backbone, whereas sugar puckers as well as helical and base-pair stacking parameters are only weakly influenced. A subset of 15 solvent sites with low temperature factors is conserved in all three models.

Introduction

In the resolution range normally accessible, the crystal structure of a biological macromolecule is not determined by the measured X-ray diffraction data alone. Instead, structure refinement needs to be supplemented with stereochemical constraints or restraints (Jensen, 1985). Refinement packages use different approaches to ensure acceptable geometry, hence the results obtained with them may differ too. This is a problem especially with nucleic acid structures, where the global structure is usually predefined in one of the helical families and sequence-determined fine structural details are of interest.

In earlier studies, the structure of the B-DNA dodecamer d(CGCGAATTCGCG), first published by Dickerson and colleagues (Wing *et al.*, 1980; Dickerson & Drew, 1981), was re-refined by different approaches. The original structure had been refined with the program of Jack & Levitt (1978), and the re-refinements used a version of *CORELS* (Sussman, Holbrook, Church & Kim, 1977) and *NUCLIN*/*NUCLSQ* (Westhof, Dumas & Moras, 1985). Refinement with *CORELS* using anisotropic temperature factors for rigid groups provided an improved fit of the model to the experimental data as judged from the *R* value and from the difference electron density (Holbrook, Dickerson & Kim, 1985). Refinement of the DNA model with *NUCLSQ*, on the other hand, showed that the overall structure was insensitive to the refinement algorithm applied whereas the backbone torsion angles showed some differences (Westhof, 1987).

NUCLSQ is well established for the refinement of nucleic acid crystal structures (Shakked & Rabinovich, 1986; Kennard & Hunter, 1989) but is not easily applicable to the structure determination of DNA in protein–DNA complexes. Although Jordan & Pabo (1988) and Suck, Lahm & Oefner (1988) have modified the Hendrickson–Konnert conjugate-gradient program (Hendrickson, 1985) to accept restraints for nucleic acids based on a version of *NUCLIN*, usually programs like *TNT* (Tronrud, Ten Eyck & Matthews, 1987) or *X-PLOR* (Brünger, Kuriyan & Karplus, 1987) have been used for the refinement of protein–DNA complexes (Schultz, Shields & Steitz, 1991; Somers & Philipps, 1992; Pavletich & Pabo, 1991; Beamer & Pabo, 1992). The aim of the present study is to investigate the suitability of *TNT* and *X-PLOR* for DNA structure refinement and to test their effect on the derived stereochemical parameters. Both programs provide their own standard geometry files for nucleic acids. *X-PLOR* has been used before in the refinement of DNA crystal structures (Yoon, Privé, Goodsell & Dickerson, 1988; Courseille *et al.*, 1990).

The double-stranded DNA decamer d(CCAGGCm^{5}CTGG) has been refined with *NUCLIN*/*NUCLSQ* to an *R* value of 17.4% against 1.7 Å X-ray diffraction data (Heinemann & Hahn, 1992). Here we describe the re-refinement of this structure

* Author for correspondence.

ISSN 0907-4449

against the same experimental data using either *TNT* or *X-PLOR*. d(CCAGGCm^{5}CTGG) seemed especially suitable for this purpose, since the resolution and quality of the diffraction data are fairly typical for DNA structures, and the duplex crystallizes in space group *P*6 without internal dyad symmetry leaving the maximum number of independent structural parameters.

Methods

In all cases, the starting model for the refinement was the structure of d(CCAGGCm^{5}CTGG) refined to 2.25 Å (Heinemann & Alings, 1991). All solvent molecules were deleted from the model, and temperature factors were reset to 19.1 Å^2, the overall *B* value obtained from Wilson (1942) statistics. The structure refined with *NUCLSQ* against the present 1.7 Å data set consisting of 3799 2σ structure amplitudes was published elsewhere (Heinemann & Hahn, 1992) and is not described here therefore. It should be mentioned that bond lengths, bond angles, planes, chiral centers and non-bonded contacts were restrained, whereas torsion angles remained unrestrained in this refinement.

The restraints applied to stereochemical parameters in *TNT* and their weights are shown in Tables 1 and 2 for a number of cycles. The target distances for non-bonded contacts were those from *NUCLIN*, since the values provided by *TNT* led to contacts that were judged unacceptably close. Torsion angles were not restrained, thermal parameters were restrained with the *BCORREL* module of *TNT*. In total, 17 rounds of *TNT* refinement were performed with 20 least-squares cycles each. Water molecules were added to the set after round 5, and the 2σ cutoff was applied on the structure amplitudes after round 10. Finally, 37 water O atoms, all with full occupancy and reasonable temperature factors, were located, and the refinement converged at $R = 20.6\%$.

X-PLOR uses a simulated-annealing (SA) procedure (Kirkpatrick, Gelatt & Vecchi, 1983) to explore the conformational space of the molecule by molecular dynamics. It provides a large radius of convergence compared to conventional least-squares refinement and therefore reduces the need for manual corrections (Brünger, Kuriyan & Karplus, 1987). The coordinates of d(CCAGGCm^{5}CTGG) were prepared for *X-PLOR* by addition of polar H atoms, based on parameter and topology files originally from *CHARMM* (Brooks *et al.*, 1983). The weight for the X-ray pseudo-energy term was determined before each run with the *CHECK* module of *X-PLOR*. 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 to 300 K. The resulting model was then refined by conjugate-gradient methods, with 120 steps for coordinates and 20 steps for temperature factors in each cycle. This yielded an *R* value of 27.9% with 2σ structure amplitudes from 20 to 1.7 Å. In the next cycles, solvent molecules were added, and the structure was further refined by the conjugate-gradient method after setting the low-resolution cutoff to 8 Å. After 18 cycles and addition of 80 water O atoms the refinement converged with $R = 19.1\%$. The course of the *X-PLOR* refinement is summarized in Table 3.

Table 1. *Course of TNT refinement*

R.m.s. Δ, root-mean-square deviation. The *R* value is $\Sigma|F_o - F_c|/\Sigma F_o$, where F_o and F_c are the observed and calculated structure amplitudes, respectively.

	Cycle					
R.m.s. Δ from target	1	5	8	11	14	Final
Sugar/base bonds and angles						
Bond distances (Å)	0.017	0.026	0.028	0.032	0.031	0.031
Bond-angle distances (Å)	0.036	0.065	0.074	0.078	0.081	0.081
Phosphate bonds and angles						
Bond distances (Å)	0.035	0.063	0.083	0.093	0.103	0.104
Bond-angle distances (Å)	0.054	0.089	0.098	0.101	0.106	0.109
Planar groups (Å)	0.030	0.033	0.045	0.046	0.048	0.050
Chiral volumes (Å^3)	0.133	0.180	0.222	0.231	0.225	0.220
Non-bonded contacts						
Single-torsion contacts (Å)	0.113	0.176	0.170	0.169	0.179	0.178
Multiple-torsion contacts (Å)	0.181	0.147	0.115	0.159	0.178	0.189
$\langle F_o - F_c \rangle$	1.51	1.13	0.85	0.75	0.58	0.56
R value (%)	36.7	29.4	24.7	22.7	20.8	20.6
No. of water molecules	0	0	19	30	41	37

Table 2. *Weights used in TNT refinement*

	Cycle					
Weight	1	5	8	11	14	Final
Bond distances	5	0.6	0.5	0.4	0.4	0.4
Bond angles	4	0.5	0.4	0.4	0.4	0.4
Planar groups	10	3.5	3.0	3.0	2.5	2.5
Pseudorotation angles*	3	0.8	0.6	0.6	0.6	0.6
Non-bonded contacts	5	1.5	1.0	0.5	0.5	0.5
Thermal parameters	0.1	0.1	0.1	0.09	0.07	0.07
Diffraction data	0.005	0.005	0.005	0.005	0.005	0.005

* The target value of the pseudorotation angle (Altona & Sundaralingam, 1972) was 158°.

Table 3. *Course of X-PLOR refinement*

R.m.s. Δ, root-mean-square deviation. The *R* value is $\Sigma|F_o - F_c|/\Sigma F_o$, where F_o and F_c are the observed and calculated structure amplitudes, respectively.

	Cycle					
R.m.s. Δ from target	1	5	8	11	14	Final
Sugar/base bonds and angles						
Bond distances (Å)	0.023	0.022	0.025	0.022	0.021	0.028
Bond-angle distances (Å)	0.056	0.057	0.052	0.051	0.050	0.053
Phosphate bonds and angles						
Bond distances (Å)	0.125	0.117	0.126	0.113	0.114	0.111
Bond-angle distances (Å)	0.071	0.067	0.068	0.067	0.067	0.068
Planar groups (Å)	0.081	0.076	0.082	0.076	0.082	0.083
Chiral volumes (Å^3)	0.162	0.174	0.190	0.164	0.158	0.172
Non-bonded contacts						
Single-torsion contacts (Å)	0.103	0.116	0.102	0.084	0.086	0.082
Multiple-torsion contacts (Å)	0.204	0.166	0.153	0.142	0.125	0.136
$\langle F_o - F_c \rangle$	0.32	0.38	0.28	0.24	0.23	0.21
R value (%)	28.7	22.8	21.0	20.0	19.5	19.1
No. of water molecules	0	35	56	68	73	80

In all three refinement schemes, electron-density maps were displayed with the graphics program *FRODO* (Jones, 1985). The peaks of difference Fourier maps were accepted as water O atoms if they had hydrogen-bonding distances to polar DNA atoms in the range of 2.4–3.5 Å. A small number of water molecules that violated the upper threshold were added to the coordinate sets to satisfy persistent difference electron density and retained if they refined to acceptable B values of less than 50 Å^2.

The structure-factor amplitudes of d(CCAGGCm^{5}CTGG) and the atomic coordinates of the *NUCLSQ*-derived model are available from the Brookhaven Protein Data Bank as entries R2D25SF and 2D25, respectively. The atomic coordinates of the *TNT* and *X-PLOR* models have also been deposited.*

Results

Refinement results

The DNA stereochemistry and the agreement of the atomic model with the X-ray observations after structure refinement with *NUCLSQ*, *TNT* and *X-PLOR* are summarized in Table 4. The root-mean-square (r.m.s.) deviations for various parameters are shown with respect to the target values of the *NUCLIN* stereochemical library. Thus, they are likely to be larger for the *TNT*- and *X-PLOR*-derived models, since the standard geometries used in these programs are slightly different. However, since Table 4 is based on the same reference geometry, it permits a direct comparison to be made. Analyses based on the *TNT* and *X-PLOR* libraries (not shown) give smaller r.m.s. differences for the models resulting from these programs, but the trends apparent from Table 4 are observed as well.

The standard geometries defined in the *NUCLIN*, *TNT* and *X-PLOR* (PARAM11.DNA) libraries are appreciably different. Ideal bond lengths in sugars and bases, for example, show r.m.s. differences of 0.014 Å for the *TNT* library and of 0.015 Å for PARAM11.DNA when compared to the *NUCLIN* standard geometry. This accounts for the differences in the corresponding values of Table 4. Engh & Huber (1991) after noticing similar discrepancies in bond-length and bond-angle values for proteins have proposed a set of target distances and angles for refinement with *X-PLOR*.

Table 4. *Comparison of NUCLSQ, TNT and X-PLOR refinements*

R.m.s. Δ, root-mean-square deviation. The R value is $\Sigma|F_o - F_c|/\Sigma F_o$, where F_o and F_c are the observed and calculated structure amplitudes, respectively. The correlation coefficient is $\Sigma[(F_o - \langle F_o\rangle)(F_c - \langle F_c\rangle)]/[\Sigma(F_o - \langle F_o\rangle)^2\Sigma(F_c - \langle F_c\rangle)^2]^{1/2}$.

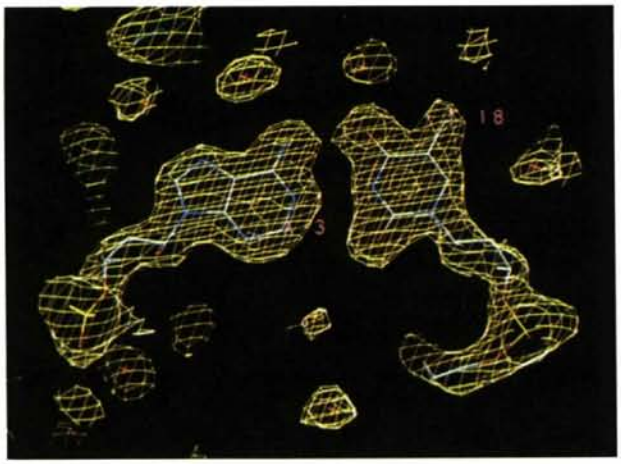
	NUCLSQ		*TNT*		*X-PLOR*	
	Mean	R.m.s. Δ	Mean	R.m.s. Δ	Mean	R.m.s. Δ
Sugar/base bonds and angles						
Bond distances (Å)	1.411	0.016	1.404	0.031	1.409	0.028
Bond-angle distances (Å)	2.392	0.040	2.376	0.081	2.394	0.053
Phosphate bonds and angles						
Bond distances (Å)	2.222	0.045	2.203	0.104	2.174	0.111
Bond-angle distances (Å)	3.007	0.064	2.992	0.109	3.016	0.068
Planar groups (Å)		0.025		0.050		0.083
Chiral volumes (Å^3)		0.087		0.220		0.172
Non-bonded contacts						
Single-torsion contacts (Å)		0.143		0.178		0.082
Multiple-torsion contacts (Å)		0.200		0.189		0.136
Isotropic thermal factors						
Sugar/base bond (Å^2)		2.815		3.385		3.353
Sugar/base angle (Å^2)		4.280		5.406		5.197
Phosphate bond (Å^2)		2.982		3.905		3.919
Phosphate angle (Å^2)		4.048		6.025		5.146
$\langle\|F_o - F_c\|\rangle$	0.17		0.56		0.21	
R value (%)	17.4		20.6		19.1	
$\Sigma F_o/\Sigma F_c$	1.021		1.068		1.026	
Correlation coefficient	0.972		0.955		0.966	
Final difference-density map						
No. of peaks >0.25 e Å^{-3}	2		10		2	
Highest peak (e Å^{-3})	0.30		0.30		0.28	
No. of water molecules	85 + 1 Mg^{2+}		37		80	
R (%) without solvent	27.1		28.2		26.8	

Whereas *X-PLOR* shows the largest deviations from base planarity in the models for d(CCAGGCm^{5}CTGG), it gives the smallest violations of non-bonded contact distances. For single-torsion distances, this is partly as a result of torsion-angle restraints which were not used with *NUCLSQ* or *TNT*. All three refinement schemes produce similar variations of thermal parameters.

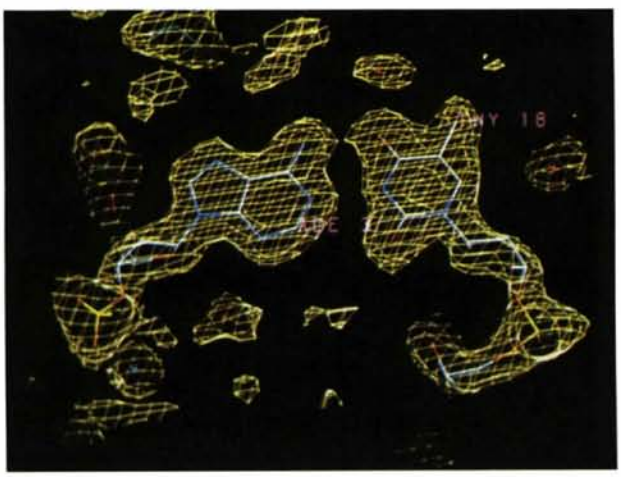
NUCLSQ gives the best model fit to the diffraction data as judged from the R value, the mean $|F_o - F_c|$, the correlation coefficient and the final difference-density map. The higher R value and the noisier difference density given by the *TNT*-refined model is mainly as a result of the presence of only 37 water O atoms in the coordinate set which, in turn, is a consequence of the refinement of B values with *TNT* (see below). The R values of the final DNA models, omitting the solvent, are identical to within 1.4% after refinement with *NUCLSQ*, *TNT* or *X-PLOR*. Thus, judged by this criterion and by the close similarity of the resulting structures (see below), all three programs can provide reasonable DNA structures. Fig. 1 shows electron densities around the base pair A(3)·T(18) for a visual comparison of the model fit to the X-ray data.

The simulated-annealing procedure of *X-PLOR* reduces the R value of the input model rapidly (see Table 3), but subsequent conventional least-squares refinement with this program does not give as good a fit to the data as *NUCLSQ*. Therefore, as a test the

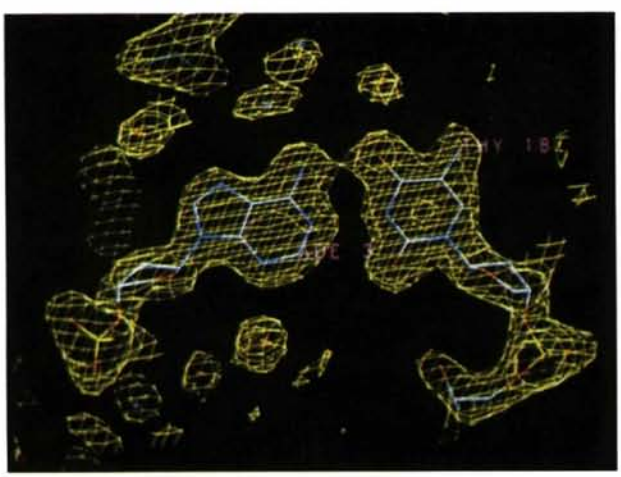
* Atomic coordinates of the *TNT* and *X-PLOR* models have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 122D, 123D). Free copies may be obtained through The Technical Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Supplementary Publication No. SUP 37086). A list of deposited data is given at the end of this issue.



(*a*)



(*b*)



(*c*)

Fig. 1. Portions of the $2F_o - F_c$ difference electron-density maps around base pair A(3)·T(18) of d(CCAGGCm^{5}CTGG). The maps are contoured at 1.5 times their r.m.s. density. Models and maps are derived from (*a*) *NUCLSQ*, (*b*) *TNT* and (*c*) *X-PLOR*.

Table 5. *R.m.s. deviations between models*

	NUCLSQ					Single strands	
Model	10 bp	8 bp	1BD1*	1D25†	*TNT*	*X-PLOR*	*NUCLSQ*
TNT‡	0.25	0.20	1.06	0.29	0.50	—	—
X-PLOR	0.32	0.23	1.08	0.42	—	0.57	—
NUCLSQ	0	0	1.05	0.31	—	—	0.52

* Protein Data Bank entry for d(CCAGGCCTGG) (Heinemann & Alings, 1989).

† Protein Data Bank entry (obsolete) for the 2.25 Å model of d(CCAGGCm^{5}CTGG) (Heinemann & Alings, 1991), the starting coordinates for refinement.

‡ The tabulated values are r.m.s. distances (Å) between equivalent atoms.

X-PLOR-derived model was further refined with 135 cycles of *NUCLSQ*. After removal of three water molecules which adopted high *B* values from the coordinate set, the *R* value converged at 17.7%, the difference density was clean with a maximum of 0.26 e Å^{-3} and the *NUCLSQ* model geometry was very similar to that shown in Table 4.

Intermolecular contacts

In both *TNT* and *X-PLOR*, intermolecular non-bonded contacts are restrained, whereas in *NUCLSQ* they are not. Nevertheless, lattice contacts between duplex DNA molecules, either direct or *via* a bridging water molecule, are at least partially conserved. Four direct hydrogen-bonded contacts are observed in both the *NUCLSQ*- and *TNT*-derived models of which three are identical. Only one of these contacts is also present in the *X-PLOR* model. Seven bridging water molecules making hydrogen bonds to two duplex molecules in the crystal are conserved in all three models, out of a total of 15 from *NUCLSQ*. Thus, the conservation of bridging water molecules is significantly better than for solvent sites in general (see below). Restraints on intermolecular non-bonded contacts do not have a strong influence on the lattice contacts.

Global DNA structures

Least-squares superposition (Fig. 2) shows the global structures of d(CCAGGCm^{5}CTGG) resulting from refinement with *NUCLSQ*, *TNT* or *X-PLOR* to be very similar. The r.m.s. distances between equivalent atoms (Table 5) are close to the average coordinate error calculated to be 0.2–0.3 Å by the method of Luzzati (1952). If the superposition is limited to the inner eight base pairs of the duplexes, the r.m.s. deviations are further reduced. Therefore, the global structures of the DNA decamer are identical within the accuracy of the structure determination no matter which program is used for refinement. The r.m.s. deviations between the final models and the starting structure are between 0.42 Å (*X-PLOR*) and 0.29 Å (*TNT*).

All three models show the same degree of asymmetry which is characterized by r.m.s. deviations of 0.5–0.6 Å between the superimposed atoms of the chemically identical single strands. This asymmetry is clearly above the experimental error and must be attributed to crystal packing. A superposition of the models of d(CCAGGCm^{5}CTGG) with the structure of the unmethylated variant, d(CCAGGCCTGG) (Heinemann & Alings, 1989), gives r.m.s. deviations around 1.05 Å. Despite this difference, the decamer helices in general show similar structural features.

Helical parameters

After having established the close general similarity of the DNA model resulting from the different refinement schemes it has to be discussed to what extent the remaining small structural differences influence derived conformational parameters. The most important helical parameters for the three models of d(CCAGGCm^{5}CTGG) have been calculated with *NEWHEL*91 (Dickerson, 1991) and are displayed in Fig. 3. It is apparent from the plots and the linear correlation coefficients between parameters, which are 0.74 or above, that the helical parameters are only slightly affected by the refinement procedure employed. The parameters twist, rise and slide are remarkably similar between the different DNA models (correlation coefficients 0.94 or above), whereas the agreement is poorest for roll between the *NUCLSQ* and *TNT* models.

A significant difference between the *NUCLSQ* structures on one side and the *TNT* or *X-PLOR* models on the other might be expected for the propeller, since, in the former, the geometry of the Watson–Crick base pairs is explicitly restrained whereas in the latter it is not. It appears, however, that this does not play a major role, since both the mean values and the sequence-dependent variations of propeller are similar in all three models.

Torsion angles and sugar pucker

Torsion angles are determined by only four atoms, whereas the helical parameters discussed above are defined by larger subsets of atoms. We may expect, therefore, a more notable influence of the small coordinate differences between the helical models of d(CCAGGCm^{5}CTGG) on the torsion angles. Fig. 4 shows that this is indeed the case. In all three models, the mean values of the backbone torsion angles fall in the expected ranges of, for example, (−)*gauche* for α, *trans* for β and (+)*gauche* for γ. The distributions of torsion angles and pseudorotation phase angles, however, are quite different between the *NUCLSQ*, *TNT* and *X-PLOR* refined structures (inner, middle and outer ring, respectively). *NUCLSQ* tends to yield the narrowest angular range, although in refinement with this program explicit torsion-angle restraints were not applied. Apparently, the chiral-volume restraints for sugar atoms employed in *NUCLSQ* are very efficient in maintaining reasonable torsion-angle values. The well known correlations between torsion angles in the sugar–phosphate backbone, *e.g.* α/γ, δ/χ, ε/ζ, (Dickerson *et al.*, 1982; Dickerson, Kopka & Drew, 1983) are well preserved in all three models (not shown).

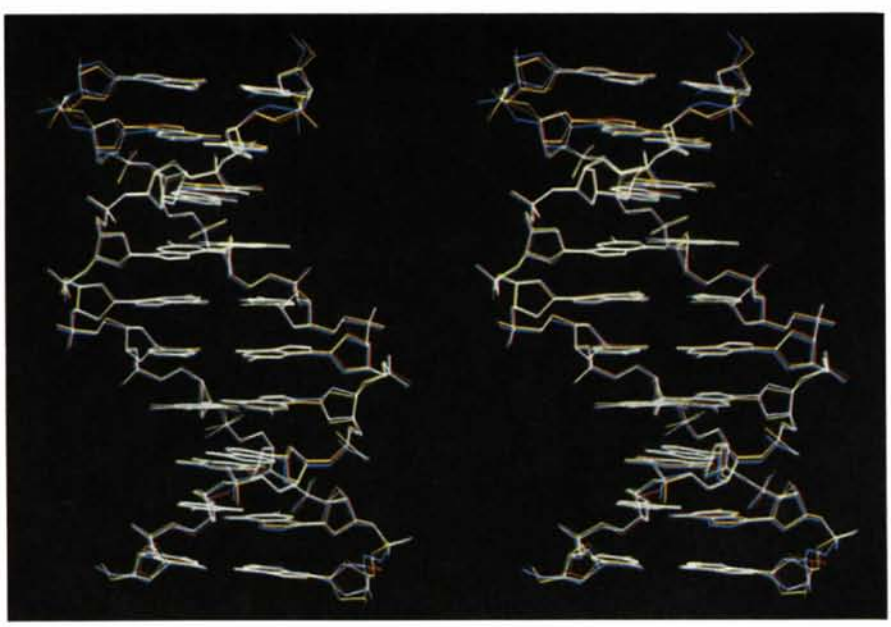


Fig. 2. Stereo drawing of a least-squares superposition of the models of d(CCAGGCm^5-CTGG) from *TNT* (red) and *X-PLOR* (blue) refinement with the *NUCLSQ* model (yellow).

The values for $\varepsilon - \zeta$ usually show a bimodal distribution which defines the B_I ($\varepsilon - \zeta < 0°$) *versus* B_{II} ($\varepsilon - \zeta > 0°$) backbone types (Privé *et al.*, 1987). In the three DNA models discussed here, a clear bimodal distribution is seen exclusively for ζ but not for ε. Hence, the less pronounced bimodal distribution of $\varepsilon - \zeta$ is entirely because of the ζ torsion angle which should, therefore, be used to define the backbone type instead of $\varepsilon - \zeta$.

The sequence-dependent variation of torsion angles along the backbones in the three DNA models is best compared by calculating linear correlation coefficients (Table 6). It is obvious that sugar pucker and the glycosidic torsion χ are fairly well preserved in the three models, whereas the exocyclic torsion angles α, β and γ are less well defined. Thus, even at a resolution of 1.7 Å, the variation of these torsion angles within their expected ranges remains mostly an artefact of structure refinement. The situation is different with the remaining exocyclic torsion angles ε and ζ which vary over a wider range along the strands. Their sequence-dependent change is very

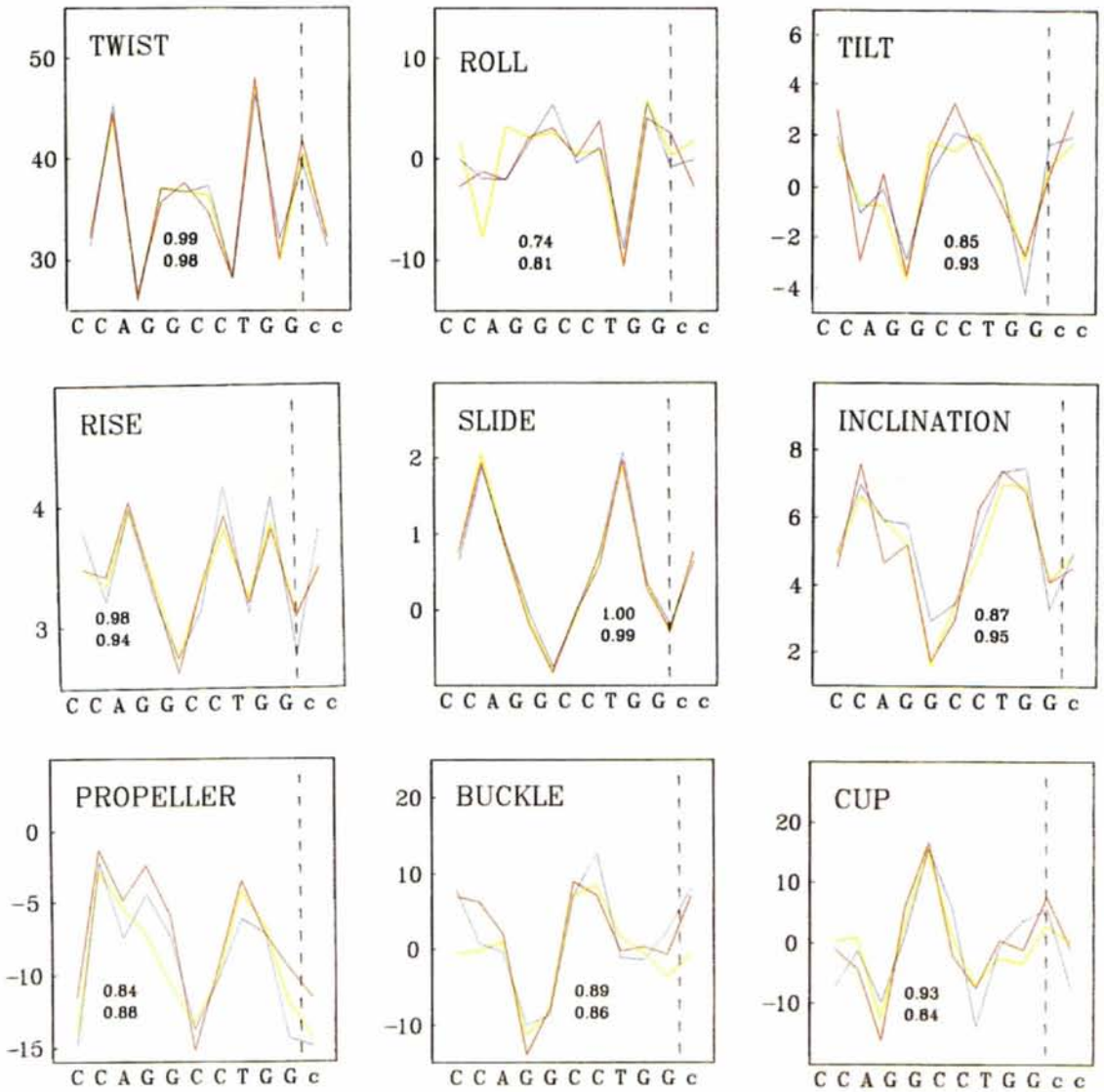


Fig. 3. Helical parameters for the models of d(CCAGGCm⁵CTGG) resulting from refinement with *NUCLSQ* (yellow), *TNT* (red) and *X-PLOR* (blue). The values for the rotational parameters twist, tilt, roll, propeller, buckle, cup and inclination are given in °, those for the translation parameters rise and slide in Å. Parameters are defined according to the EMBO Cambridge Workshop (1989) convention. The dashed vertical line marks the end-to-end stacking of decamer duplexes which creates continuous helices traversing the crystals; lower case letters indicate base pairs in the translationally related DNA molecule. Numbers inside the diagrams give the linear correlation coefficients between the helical parameters in the *TNT* model (top) or the *X-PLOR* model (bottom) and those in the *NUCLSQ* reference structure.

well preserved with correlation coefficients of 0.95 and 0.75, respectively. Thus, the B_I/B_{II} backbone types are well defined at the present resolution and independent of the refinement protocol.

Thermal parameters

In all three refinement schemes, thermal parameters were restrained to prevent unreasonably large *B*-factor differences in atoms linked by one or two covalent bonds. The mean *B* values for the atoms in bases, sugars and phosphate groups of d(CCAGGCm^{5}CTGG) after refinement with *NUCLSQ*, *TNT* and *X-PLOR* are shown in Fig. 5. There is a significant increase of temperature factors in the *TNT*-derived model as compared to the two other models. The average *B* value expected from Wilson (1942) statistics of 19.1 Å^2 is close to the average *B*'s of 22.2 and 21.4 Å^2 of the DNA atoms in the *NUCLSQ* and *X-PLOR* models, respectively. The elevated level of temperature factor encountered in the *TNT* refinement ($\langle B \rangle = 28.0$ Å^2 for DNA atoms) poses a problem in the location of solvent molecules many of which refine to very large *B* values (≥ 60 Å^2) and, therefore, have to be excluded from the model, leading to an incomplete solvent model and causing a noisy difference density and a less than satisfactory *R* value (see above). The increase in thermal parameters usually observed

Table 6. *Backbone torsion angles and pseudorotation angles* (°)

	α	β	γ	δ	ε	ζ	χ	τ	P
NUCLSQ									
Mean	−65	163	40	142	−138	−143	−96	45	159
S.d.*	18	14	17	14	40	49	19	6	23
Corr.†	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
TNT									
Mean	−61	164	45	142	−145	−134	−100	46	151
S.d.	17	16	26	15	27	41	20	8	25
Corr.	0.26	0.80	0.25	0.80	0.90	0.95	0.92	0.71	0.88
X-PLOR									
Mean	−66	162	42	140	−143	−135	−98	41	153
S.d.	29	24	35	17	39	57	20	6	26
Corr.	0.65	0.49	0.61	0.72	0.70	0.75	0.87	0.52	0.88

* Standard deviation of the sample.

† Linear correlation coefficient between the parameter values of the model and those of the *NUCLSQ*-derived reference model. Torsion angles α of residue G(10) and γ of C(1) and G(10) are excluded from calculations of mean, standard deviation and correlation.

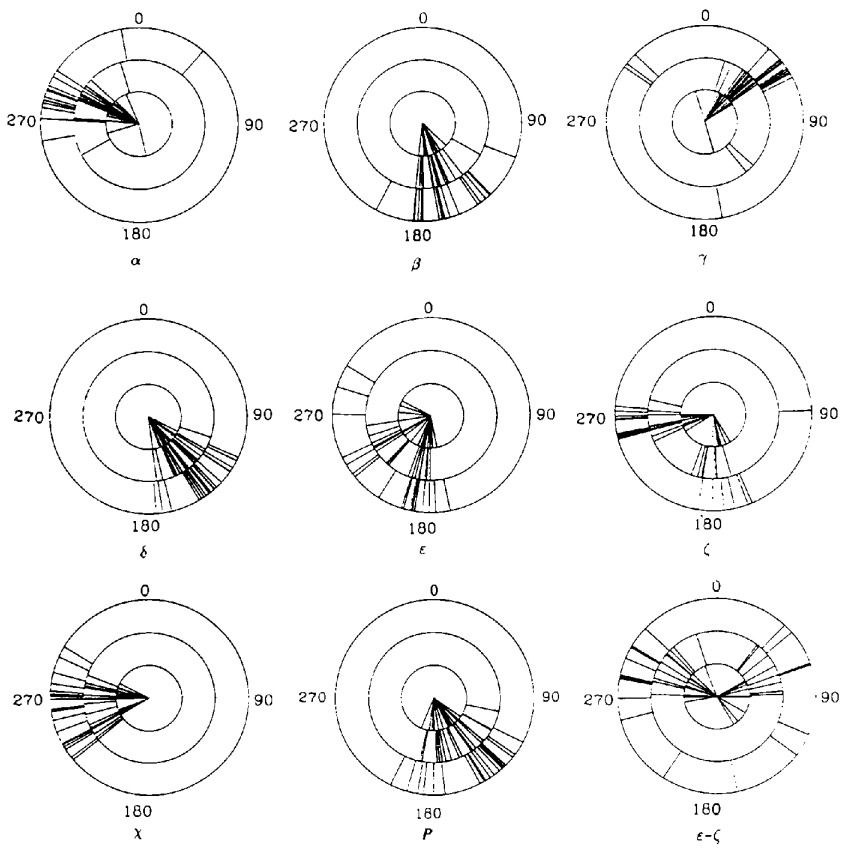


Fig. 4. Distribution of torsion angles and pseudorotation phase angles in the models for d(CCAGGCm^{5}CTGG) derived from NUCLSQ (inner), TNT (middle) and *X-PLOR* (outer ring) refinement.

from bases to sugars to phosphates is seen in all three refinements as well as the increase in *B* values from the center to the ends of the decamer duplex as described previously (Heinemann & Hahn, 1992).

Hydration

Since the water activity determines the global DNA-helix structure, hydration patterns around crystalline DNA molecules have received much attention (Saenger, Hunter & Kennard, 1986; Westhof, 1988; Kennard *et al.*, 1986; Kopka, Fratini, Drew & Dickerson, 1983). For analyses it is mandatory to know the reliability with which solvent sites are determined in DNA crystal structures. By comparing the solvent structure in the three decamer models arising from refinement with *NUCLSQ*, *TNT* and *X-PLOR* we have tried to define those solvent sites that are conserved, *i.e.* not influenced by refinement protocol or operator bias. Solvent sites (water O atoms) were considered conserved when they appeared within a specified cut-off distance from one another in both the *NUCLSQ* model and either the *TNT* or the *X-PLOR* models and shared at least one hydrogen-bond partner. From histograms of the frequency of occurrence in pairwise comparisons *versus* the water–water distance (not shown) the cut-off distance was chosen to be 0.8 Å. Following this definition, 23 solvent sites are conserved between the *NUCLSQ* and *TNT* models and 26 in the *NUCLSQ*/*X-PLOR* pair.

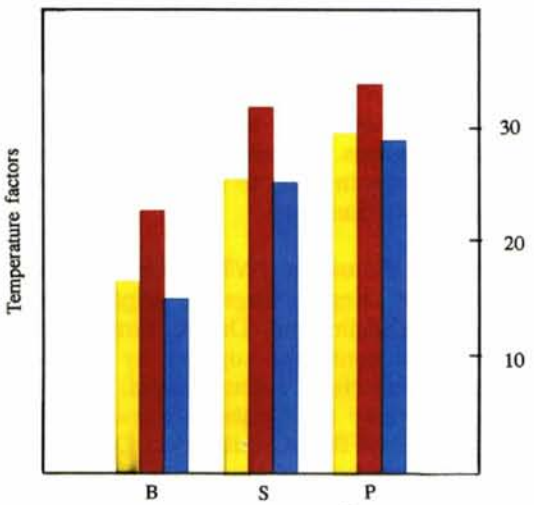


Fig. 5. Average temperature factors (Å^2) for the atoms in bases (*B*), sugars (*S*) and phosphate groups (*P*) of d(CCAGGCm^{5}CTGG) after refinement with *NUCLSQ* (yellow), *TNT* (red) and *X-PLOR* (blue).

Looking at all three models, we find only 15 conserved solvent sites (Fig. 6). This is a surprisingly small number in comparison with the total number of more than 80 sites resulting from refinement with both *NUCLSQ* and *X-PLOR*. The conclusion has to be that most of the remaining solvent sites are determined with questionable reliability. Of the 15 conserved water molecules, three are in the minor groove, five in the major groove, six are hydrogen bonded to phosphate O atoms and one belongs to the secondary hydration shell. Regular patterns of

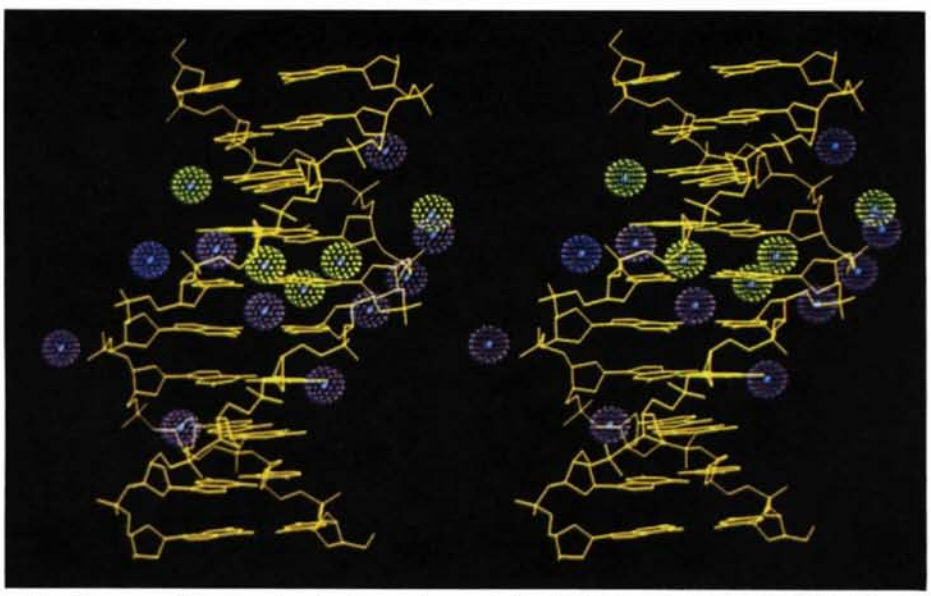


Fig. 6. Stereoview of the 15 conserved water molecules (see text) around the *NUCLSQ* model of d(CCAGGCm^{5}CTGG). Water O atoms are shown with a van der Waals surface and are color coded according to their temperature factors: $B \leq 20$ Å^2 (blue), $20 \leq B \leq 30$ Å^2 (purple), $B \geq 30$ Å^2 (yellow).

hydration are hardly discernible in the small set of 15 conserved sites. As expected, the average temperature factors of the conserved water molecules are smaller, with 27, 33 and 29 Å^2 in the *NUCLSQ*, *TNT* and *X-PLOR* models, than those for the entire solvent spheres which are between 38 and 40 Å^2 in all three models. The absence of conserved water molecules bound to the outer four base pairs may be explained by the elevated *B* values of DNA atoms near the ends of the duplex.

Discussion

Using X-ray diffraction data at 1.7 Å resolution, the crystal structure of the B-DNA decamer d(CCAGGCm^{5}CTGG) has been refined with three different protocols. The DNA conformations obtained after refinement with *NUCLSQ*, *TNT* and *X-PLOR* are nearly identical within the limits of the attainable accuracy. Nevertheless, the small differences in DNA coordinates affect, in different ways, derived conformational parameters routinely used to describe DNA structures.

Global DNA structure, groove dimensions and helical parameters describing the geometry of base pairs, their orientation and position relative to the helix axis, and the geometry of base-pair stacking are preserved very well in the models obtained by the different refinement schemes. Thus, given diffraction data of reasonable quality, X-ray crystallography can provide a reliable description of DNA conformation in these terms. The situation is slightly different regarding the conformation of the sugar–phosphate backbone. Here, the deoxyribose pucker, the glycosidic torsion angles and the B_I *versus* B_{II} backbone type from nucleotide to nucleotide are determined reproducibly whereas the seemingly sequence-dependent variation of the exocyclic torsion angles α, β and γ within their standard ranges appears not to be meaningful at the resolution of this analysis.

The solvent sphere around the DNA molecule is surprisingly poorly determined. Only 15 water molecules were located at equivalent sites in the three models of d(CCAGGCm^{5}CTGG). Out of these 15 only three are hydrogen bonded to minor-groove atoms. Since the minor groove of B-DNA is expected to display an ordered hydration pattern (Privé *et al.*, 1987; Chuprina *et al.*, 1991), we have to ask why it is so ill defined in the present analysis. The reason may rest in an unusual combination of sequence and groove dimensions. For the most part, the minor groove is too narrow to permit a double string of water molecules running down its walls, whereas a spine of hydration at the bottom of the groove is disturbed by the presence of the guanine amino functions in eight out of the ten base pairs of the decamer duplex. At the 1.7 Å resolution of the analysis, cations could not be located unambiguously. A partly hydrated Mg^{2+} was assigned in the *NUCLSQ* model (Heinemann & Hahn, 1992) which was not confirmed in the *TNT* and *X-PLOR* refinements where solvent spheres could be modelled with water O atoms exclusively.

We have shown that DNA models after *TNT* or *X-PLOR* refinement are not systematically different from those resulting from *NUCLSQ*. For this reason, the conformation of DNA in complexes with proteins, usually refined with the former programs, may be directly compared with oligonucleotide crystal structures, usually refined with the latter. Protein–nucleic acid crystal structures will normally be refined with *TNT* or *X-PLOR* alone, whereas for structure analyses of free nucleic acid fragments is it still advantageous to use *NUCLSQ* in the final refinement cycles in order to find the optimum model fit to the diffraction data.

Advice on the use of *TNT* by Dr D. E. Tronrud (University of Oregon, Eugene), helpful discussions with H. Schindelin and Drs J. Granzin and W. Hinrichs, and continued support by Professor W. Saenger are gratefully acknowledged. This research was supported by the Deutsche Forschungsgemeinschaft through SFB 344, Teilprojekt D/3, and by the Fonds der Chemischen Industrie.

References

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

BEAMER, L. J. & PABO, C. O. (1992). *J. Mol. Biol.* **227**, 177–196.

BROOKS, B. R., BRUCCELORI, R. E., OLAFSON, B. D., STATES, D. J., SWAMINATHAN, S. & KARPLUS, M. (1983). *J. Comput. Chem.* **4**, 187–217.

BRÜNGER, A. T., KRUKOWSKI, A. & ERICKSON, J. W. (1990). *Acta Cryst.* A**46**, 585–593.

BRÜNGER, A. T., KURIYAN, J. & KARPLUS, M. (1987). *Science*, **235**, 458–460.

CHUPRINA, V. P., HEINEMANN, U., NURISLAMOV, A. A., ZIELENKIEWICZ, P., DICKERSON, R. E. & SAENGER, W. (1991). *Proc. Natl Acad. Sci. USA*, **88**, 593–597.

COURSEILLE, C., DAUTANT, A., HOSPITAL, M., LANGLOIS D'ESTAINTOT, B., PRÉCIGOUX, G., MOLKO, D. & TÉOULE, R. (1990). *Acta Cryst.* A**46**, FC9–12.

DICKERSON, R. E. (1991). Personal communication.

DICKERSON, R. E. & DREW, H. R. (1981). *J. Mol. Biol.* **149**, 761–786.

DICKERSON, R. E., DREW, H. R., CONNER, B. N., WING, R. M., FRATINI, A. V. & KOPKA, M. L. (1982). *Science*, **216**, 475–486.

DICKERSON, R. E., KOPKA, M. L. & DREW, H. R. (1983). *Conformation in Biology*, edited by R. SRINIVASAN & R. H. SARMA, pp. 227–257. Guilderland, NY: Adenine Press.

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

ENGH, R. A. & HUBER, R. (1991). *Acta Cryst.* A**46**, 392–400.

HEINEMANN, U. & ALINGS, C. (1989). *J. Mol. Biol.* **210**, 369–381.

HEINEMANN, U. & ALINGS, C. (1991). *EMBO J.* **10**, 35–43.

HEINEMANN, U. & HAHN, M. (1992). *J. Biol. Chem.* **267**, 7332–7341.

HENDRICKSON, W. A. (1985). *Methods Enzymol.* **115**, 252–270.
HOLBROOK, S. R., DICKERSON, R. E. & KIM, S.-H. (1985). *Acta Cryst.* **B41**, 255–262.
JACK, A. & LEVITT, M. (1978). *Acta Cryst.* **A34**, 931–935.
JENSEN, L. H. (1985). *Methods Enzymol.* **115**, 227–234.
JONES, T. A. (1985). *Methods Enzymol.* **115**, 157–171.
JORDAN, S. R. & PABO, C. O. (1988). *Science*, **242**, 893–899.
KENNARD, O., CRUSE, W. B. T., NACHMAN, J., PRANGE, T., SHAKKED, Z. & RABINOVICH, D. (1986). *J. Biomol. Struct. Dyn.* **3**, 623–647.
KENNARD, O. & HUNTER, W. N. (1989). *Q. Rev. Biophys.* **22**, 327–379.
KIRKPATRICK, S., GELATT, C. D. JR & VECCHI, M. P. (1983). *Science*, **220**, 671–680.
KOPKA, M. L., FRATINI, A. V., DREW, H. R. & DICKERSON, R. E. (1983). *J. Mol. Biol.* **163**, 129–146.
LUZZATI, V. (1952). *Acta Cryst.* **5**, 802–810.
PAVLETICH, N. P. & PABO, C. O. (1991). *Science*, **252**, 809–817.
PRIVÉ, G. G., HEINEMANN, U., CHANDRASEGARAN, S., KAN, L.-S., KOPKA, M. L. & DICKERSON, R. E. (1987). *Science*, **238**, 498–504.
SAENGER, W., HUNTER, W. N. & KENNARD, O. (1986). *Nature (London)*, **324**, 385–388.
SCHULTZ, S. C., SHIELDS, G. C. & STEITZ, T. A. (1991). *Science*, **253**, 1001–1007.
SHAKKED, Z. & RABINOVICH, D. (1986). *Prog. Biophys. Mol. Biol.* **47**, 159–195.
SOMERS, W. S. & PHILIPPS, S. E. V. (1992). *Nature (London)*, **359**, 387–393.
SUCK, D., LAHM, A. & OEFNER, C. (1988). *Nature (London)*, **332**, 464–468.
SUSSMAN, J. L., HOLBROOK, S. R., CHURCH, G. M. & KIM, S. H. (1977). *Acta Cryst.* **A33**, 800–804.
TRONRUD, D. E., TEN EYCK, L. F. & MATTHEWS, B. W. (1987). *Acta Cryst.* **A43**, 489–500.
WESTHOF, E. (1987). *J. Biomol. Struct. Dyn.* **5**, 560–581.
WESTHOF, E. (1988). *Annu. Rev. Biophys. Biophys. Chem.* **17**, 125–144.
WESTHOF, E., DUMAS, P. & MORAS, D. (1985). *J. Mol. Biol.* **184**, 119–145.
WILSON, A. J. C. (1942). *Nature (London)*, **150**, 151–152.
WING, R., DREW, H., TAKANO, T., BROKA, C., TANAKA, S., ITAKURA, K. & DICKERSON, R. E. (1980). *Nature (London)*, **287**, 755–758.
YOON, C., PRIVÉ, G. G., GOODSELL, D. S. & DICKERSON, R. E. (1988). *Proc. Natl Acad. Sci. USA*, **85**, 6332–6336.