Acta Cryst. (1994). D**50**, 833–841

On the Application of Direct Methods to Oligonucleotide Crystallography

BY S. R. HUBBARD, R. J. GREENALL, M. M. WOOLFSON

Department of Physics, University of York, Heslington, York YO1 5DD, England

(*Received* 4 *March* 1994; *accepted* 20 *May* 1994)

Abstract

The direct methods program *SAYTAN* was applied to simulated data at various resolutions from three oligonucleotides. Success in solving the structures was found to depend more upon the resolution of the data than upon errors in the data or the complexity of the structure. Collecting the data at a reduced temperature has little effect, unless it alters the mosaicity of the crystal or changes the resolution of the data. The presence of a heavy atom dramatically improved the phase refinement, particularly at low resolution.

Introduction

A large number of oligonucleotide structures have been crystallized, ranging from the tetramer d(CGCG) (Drew, Dickerson & Itakura, 1978; Drew, Takano, Tanaka, Itakura & Dickerson, 1980; Drew & Dickerson, 1981) to the dodecamer d(CGCGAATTCGCG) (Wing *et al.*, 1980; Drew *et al.*, 1981) and beyond. A number of crystals contain mismatched base pairs such as in the sequence d(TCGCGCG) (Kennard, 1985) and some contain unpaired bases such as the fourth base in the sequence d(CGCAGAATTCGCG) (Joshua-Tor *et al.*, 1988). Several oligonucleotide–drug complexes have also been crystallized. There has been particular interest in the binding of anti-cancer drugs such as daunomycin to specific DNA sequences (Quigley *et al.*, 1980; Moore, Hunter, d'Estaintot & Kennard, 1989).

The presence of a large number of water molecules and counterions in these crystals results in disorder which limits the resolution of their diffraction patterns. The right-handed A-DNA and B-DNA crystals diffract mostly to resolutions in the range 1.7–2.5 Å, whereas the left-handed Z-DNA hexamers produce the most perfect crystals, which diffract to resolutions of 1.0 Å or higher. The counterions are generally disordered and are difficult to identify except in the Z-hexamers, in which the spermine molecules which aid crystallization can also be detected. Many of the water molecules are disordered and usually only about one third per duplex, around 80 to 100, can be identified.

Many of the successful determinations of oligonucleotide crystal structures have employed multi-dimensional direct-space search or molecular-replacement techniques using models based on the well defined coordinates of the *A*, *B* and *Z* helices (Rabinovich & Shakked, 1984). The multiple isomorphous replacement (MIR) method has also been used (Wang *et al.*, 1979). In one case, crystals of the 15-base-pair oligomer d(CGCGAAATTTACGCG) diffracted to only 3 Å resolution, and the structure could not be solved by MIR methods. Conventional direct methods were used to find an initial set of phases, which were then expanded using Sayre's equation. A map calculated from these starting phases revealed large regions of density, but it could not be interpreted. The starting phases were then refined by searching for the minimum-cross-entropy solution; a map calculated from the refined phases clearly revealed the location and orientation of the DNA molecule (Harrison, 1989).

Conventional direct methods, using either the tangent formula or the Sayre-tangent formula incorporated in *MULTAN* and *SAYTAN* (Main *et al.*, 1988; Debaerdemaeker, Tate & Woolfson, 1985) have been very successful in solving many small molecular structures. They have also been successfully applied to solve a previously known small protein structure, avian pancreatic polypeptide, aPP, in which there were 36 amino-acid residues with a heavy Zn atom and also 80 water molecules, giving a total of 382 non-H atoms in the asymmetric unit. The aPP crystal diffracted to 0.98 Å resolution. *SAYTAN*, solved the structure starting from a large number of different sets of random phases, the best solution giving a mean phase error of less than 40° (Woolfson & Yao, 1990). Further testing with *SAYTAN*, using aPP data at various restricted resolutions, showed that useful sets of phases could be obtained even at 3.0 Å resolution (Mukherjee & Woolfson, 1993). However, the aPP structure is a rather special case, since the presence of a Zn heavy atom certainly helps in the solution.

Oligonucleotide structures are intermediate in size between the largest of the 'small' molecular structures and macromolecules. If direct methods could be used to solve these structures they would be

ISSN 0907-4449

Table 1. *Data sets used in the investigation*

Set No.	Sequence	No. of atoms in asymmetric unit	Space group	Resolution of observed data (Å)	Comment
1	CG	76 DNA + 15 solvents/ions	$P2_12_12_1$	0.85	Data simulated using *B* values reported by Ramakrishnan & Viswamitra (1988).
2	CGCG	158 DNA + 84 solvents	$B22_12$	1.6	Data simulated using *B* values reported by Drew, Dickerson & Itakura (1978); Drew *et al.* (1980); Drew & Dickerson (1981).
3	CGCG*X*G *X* = meO⁴C	244 DNA + 78 solvents	$P2_12_12_1$	1.7	Data simulated using *B* values reported by Van Meervelt *et al.* (1990).
4	CGCG*X*G *X* = meO⁴C	The conditions of the data simulation were the same as for set 3 except that the *B* factors were given the following fixed values: *B*(phosphorous) = 10 Å², *B*(other DNA atoms) = 5 Å², *B*(water) = 20 Å².			
5	CGCG*X*G *X* = meO⁴C	244 DNA No water	The same conditions were used as for set 3 except that no water molecules were included and all the *B* factors were set equal to zero.		
6	CGCG^{Br}CG 1 Br	243 DNA + 1 Br + 78 solvents	Conditions as for set 3 except that the methyl group of the modified cytosine on one chain was substituted by a Br atom with *B* = 10 Å².		
7	CGCG^{Br}CG 2 Br	242 DNA + 2Br + 78 solvents	Conditions as for set 3 except that the methyl groups of the modified cytosines on both chains were substituted by Br atoms with *B* = 10 Å².		

preferable to MIR in that no derivatives would be needed. Direct methods might also be used instead of molecular replacement in those cases where no suitable model structure is available. We report here an investigation into the applicability of direct methods to oligonucleotide crystallography. In particular, we have used, in the main, simulated data from known structures to investigate the effects of oligonucleotide size and the resolution of the data on the efficacy of the methods. We have assessed the value of heavy atoms since oligonucleotides may contain brominated base pairs, for example. We have also considered whether it would be advantageous to collect oligonucleotide data at low temperatures.

Data simulation

Several oligonucleotide structures of varying complexity were selected (see Table 1) and noise-free data sets were simulated to 1.0 Å resolution using the program *SAPI* (Yao *et al.*, 1985). Other data sets were then produced, as required, for each of these structures, having restricted resolutions of 1.5 and 2.0 Å and in some cases intermediate resolutions of 1.25 and 1.75 Å. Experimentally observed data were also available for the Z-form hexamer d(CGCG*X*G) (Van Meervelt, Moore, Lin, Brown & Kennard, 1990).

If direct methods were successful in solving a structure using the error-free data, then 10% random errors were added to the intensities, to investigate the sensitivity of the method to noise. The standard deviation $\sigma(I)$ for each reflection was assigned a value of $I/10$. A random-number generator, based on Gaussian distributions having mean values of zero and standard deviations $\sigma(I)$, was used to produce a correction ΔI to the exact intensity I of each reflection.

Phase refinement using *SAYTAN*

The simulated data sets for the various structures at different resolutions were tested using *SAYTAN*, by refining first the true phases. The program assesses the quality of the phase sets according to the conventional *SAYTAN* figures of merit, that is ABSFOM, PSIZERO and RESID, and from these a combined figure of merit, CFOM. In addition, a further figure of merit, TFOM, is evaluated as follows,

$$\text{TFOM} = \sum_h \sum_k \kappa(\mathbf{h},\mathbf{k})\{\cos[\varphi_3(\mathbf{h},\mathbf{k})] - I_1[\kappa(\mathbf{h},\mathbf{k})]/I_0[\kappa(\mathbf{h},\mathbf{k})]\}^2, \quad (1)$$

where all the reflections involved in the summations over h and k are those with strong E's; the quantity $\varphi_3(\mathbf{h},\mathbf{k}) = \varphi(\mathbf{h}) - \varphi(\mathbf{k}) - \varphi(\mathbf{h}-\mathbf{k})$ is a structure invariant; $\kappa(\mathbf{h},\mathbf{k}) = 2N^{-1/2}|E(\mathbf{h})E(\mathbf{k})E(\mathbf{h}-\mathbf{k})|$ in which N is the number of atoms in the unit cell; and $I_1[\kappa(\mathbf{h},\mathbf{k})]$ and $I_0[\kappa(\mathbf{h},\mathbf{k})]$ are the modified Bessel functions of the first and zero orders, respectively. Since $\langle\cos[\varphi_3(\mathbf{h},\mathbf{k})]\rangle = I_1[\kappa(\mathbf{h},\mathbf{k})]/I_0[\kappa(\mathbf{h},\mathbf{k})]$, TFOM is a measure of the difference between the values of the cosines of the structure invariants and their theoretical expectation values. The function on the right-hand side of (1) was proposed by Debaerdemaeker & Woolfson (1983) as a means of phase determination and a very similar normalized function was also suggested for the same purpose by Hauptman & Han (1993). In our approach, no scale factor was applied to TFOM, although a scale factor of $1/\sum_h\sum_k\kappa(\mathbf{h},\mathbf{k})$ would give a normalized result.

The program also calculates the mean phase error (MPE) of the refined phases, which provides a measure of the effectiveness of the phase refinement. Random phases have an MPE of 90.0°, and if the MPE is less than say 70.0°, then an E map calculated using these phases should contain some structural information. Further refinement of these phases may

Table 2. *Selected results from phase refinements*

	Data set (*a*)	Initial phases (*b*)	Errors (*c*)	N_{REF}/N_{TOT} (*d*)	T.R. (*e*)	Resolution (Å)	Phase set	ABSFOM	PSIZERO	RESID	CFOM	TFOM	MPE1 (°) (*f*)	MPE2 (°) (*f*)
A	1	*T*	*N*	364/3685	7453	1.0	1	1.207	1.287	22.33	3.000	1.791	78	16
B	1	*R*	*N*	364/3685	7453	1.0	133	1.201	1.075	23.53	2.569	1.983	78	21
							729	1.226	1.112	25.55	2.408	1.977	79	22
							73	1.185	1.356	26.89	1.992	1.686	77	24
C	1	*T*	*Y*	264/3685	7680	1.0	1	1.088	1.553	22.41	3.000	1.780	80	18
D	1	*R*	*Y*	364/3685	7680	2.0	93	1.166	0.982	25.16	2.747	2.132	77	25
E	1	*T*	*N*	273/1162	8349	1.5	1	1.849	1.417	51.02	3.000	5.652	87	58
F	1	*R*	*N*	273/1162	8349	1.5	714	1.793	1.431	49.98	1.146	5.876	88	64
							213	0.891	0.872	31.40	2.491	5.782	83	83
G	2	*T*	*N*	909/22196	21399	1.0	1	1.094	1.167	19.71	3.000	1.341	69	19
H	2	*R*	*N*	909/22196	21399	1.0	73	1.075	1.135	21.44	2.376	1.379	69	21
I	2	*T*	*Y*	909/22196	21142	1.0	1	1.034	1.266	19.27	3.000	1.313	71	19
J	2	*R*	*Y*	909/22196	21142	1.0	82	1.077	1.077	20.23	2.673	1.353	71	21
K	2	*T*	*N*	545/6806	12289	1.5	1	0.880	0.980	21.31	3.000	2.158	71	22
L	2	*R*	*N*	545/6806	12289	1.5	468	0.747	1.140	25.79	2.201	2.313	82	74
M	3	*T*	Obs.	600/2944	25692	1.7	1	0.803	1.138	39.49	3.000	9.506	86	69
N	3	*R*	Obs.	600/2944	25692	1.7	781	1.404	1.779	39.21	1.431	4.986	79	69
O	3	*T*	*N*	909/13740	31916	1.0	1	1.031	1.044	21.47	3.000	2.010	69	28
P	3	*R*	*N*	909/13740	31916	1.0	508	1.030	1.072	21.88	2.659	2.019	82	29
Q	3	*T*	*Y*	909/13740	32122	1.0	1	1.002	1.034	21.00	3.000	1.981	82	29
R	3	*R*	*Y*	909/13740	32122	1.0	535	0.998	1.026	21.14	2.746	1.986	83	30
S	3	*T*	*N*	636/4246	31078	1.5	1	0.994	1.125	27.65	3.000	4.886	81	40
T	3	*R*	*N*	636/4246	31078	1.5	360	1.784	2.360	49.62	1.340	4.977	81	74
U	4	*T*	*N*	909/13740	30247	1.0	1	0.811	0.837	20.42	3.000	2.546	78	16
V	4	*R*	*N*	909/13740	30247	1.0	999	0.823	1.372	29.39	2.133	2.608	80	67
W	5	*T*	*N*	909/13740	31701	1.0	1	0.980	1.098	23.09	3.000	2.500	79	13
X	5	*R*	*N*	909/13740	31701	1.0	749	1.107	2.462	38.07	1.250	2.541	80	66
Y	6	*T*	*N*	909 13740	35390	1.0	1	1.528	2.679	31.34	3.000	1.189	53	25
Z	6	*R*	*N*	909/13740	35390	1.0	340	1.560	2.837	33.87	1.800	1.267	52	26
							714	1.530	3.264	36.50	1.587	1.375	56	37
							832	1.217	2.908	24.17	1.823	1.202	53	26
							935	1.114	0.986	31.40	2.392	2.439	82	59
							824	1.529	2.727	31.20	1.875	1.193	53	25
							5	1.533	2.749	31.62	1.862	1.200	53	25
AA	6	*T*	*N*	727/7190	30772	1.25	1	2.096	3.552	62.54	3.000	3.203	47	36
BB	6	*R*	*N*	727/7290	30772	1.25	43	2.096	3.597	62.23	0.666	3.180	48	40
							47	2.089	3.572	62.71	0.658	3.196	46	37
CC	6	*T*	*N*	545/4246	20661	1.5	1	1.982	2.462	61.78	3.000	3.839	50	38
DD	6	*R*	*N*	545/4246	20661	1.5	57	2.024	2.657	63.20	0.666	3.820	50	48
							222	2.018	2.660	62.76	0.675	3.800	50	49
							549	1.967	2.471	62.07	0.804	3.870	49	41
EE	6	*T*	*N*	545/2731	29896	1.75	1	2.923	2.485	92.91	3.000	8.129	52	47
FF	6	*R*	*N*	545/2731	29896	1.75	323	2.950	2.567	92.44	0.637	7.968	54	50
							109	2.940	2.594	92.90	0.598	8.010	54	50
							117	2.919	2.466	92.60	0.716	8.174	51	48
GG	6	*T*	*N*	364/1858	12998	2.0	1	2.821	1.804	77.61	3.000	7.551	55	53
HH	6	*R*	*N*	364/1858	12998	2.0	849	2.497	1.547	73.39	1.200	8.229	54	50
JJ	6	*T*	*N*	364/1328	17311	2.25	1	3.881	1.850	101.84	1.800	12.613	83	64
KK	6	*R*	*N*	364/1328	17311	2.25	79	3.980	1.961	102.02	1.111	12.715	84	66
LL	6	*T*	*N*	364/988	20433	2.5	1	4.625	1.480	119.56	1.800	15.967	84	69
MM	6	*R*	*N*	364/988	20433	2.5	161	4.640	1.505	120.50	1.357	15.941	85	69

Notes: (*a*) See Table 1; (*b*) *T* = true, *R* = random; (*c*) *N* = no errors added, *Y* = 10% random errors added to the intensities, Obs. = observed data; (*d*) N_{REF} is the number of reflections refined from the full set of N_{TOT} reflections; (*e*) number of triplet relationships used; (*f*) MPE1 and MPE2 are the respective mean phase errors for the two enantiomers.

lead to a complete structure determination. If the true phases were not greatly changed by the tangent-formula refinement, then the phase-determination process was carried out using 1000 sets of random starting phases, and the *SAYTAN* figures of merit including TFOM and a final MPE were calculated for each set of phases. *E* maps were determined for the sets of refined phases adjudged to be the best according to their figures of merit and weighted Fourier recycling techniques were used to improve these maps.

Initially, we investigated the 'small' molecular structure d(CG) (Table 1, data set 1) which in principle should be easily solved by direct methods. Refinement of the phases, with 1.0 Å data, starting from either the true phases (Table 2, line A) or 1000 sets of random phases (Table 2, line B) gave solutions with small mean phase errors. In these cases, the *SAYTAN* figures of merit clearly discriminated the best phase sets: set 133, which had the lowest MPE, had the maximum CFOM and the minimum RESID; set 729 had the maximum ABSFOM, and

set 73 had the minimum TFOM. Calculation of an E map using the refined phases of set 133 revealed 72 atoms amongst the top 100 peaks at distances less than 0.30 Å from their true positions with the first two very large peaks corresponding to the two P atoms. Starting from the two P-atom positions plus the 48 next highest peaks, three cycles of weighted Fourier recycling revealed all the atomic positions in the structure, including the water molecules and the NH_4^+ ions. The addition of 10% errors to the data had little effect on the mean phase errors of the refined phases starting from either the true phases (Table 2, line C) or random phases (Table 2, line D). In the latter case, set 93 had the maximum ABSFOM and CFOM and the minimum RESID. There were 61 peaks within 0.30 Å of true atomic sites in the E map, and the structure was fully solved in two cycles of Fourier synthesis.

When the data were restricted to 1.5 Å resolution, even the refinement of the true phases yielded a final set with a 58° MPE (Table 2, line E). Starting from 1000 sets of random phases (Table 2, line F), the best solution, having an MPE of 64°, was poorly discriminated by the figures of merit and it was not distinguishable from the other phase sets as a partial solution; the set with the maximum CFOM, set 213, had an MPE of 83°.

For the larger molecule d(CGCG) (Table 1, data set 2), the true phases of 1.0 Å data refined to an MPE of 19° (Table 2, line G), whereas the best set refined from random phases (Table 2, line H) had an MPE of 21° and was readily identified since it had the maximum ABSFOM and CFOM and the minimum RESID and TFOM. An E map calculated from these phases revealed 108 atomic sites amongst the 200 largest peaks, with the top three, the 21st and the 104th peaks corresponding to five of the six P atoms. Several cycles of weighted Fourier recycling revealed 137 of the 158 DNA atoms, including all the P-atom positions, and 64 of the 84 solvent molecules. The DNA atoms that were difficult to identify in the maps were mostly sugar atoms with large thermal parameters. The addition of 10% errors to the intensities again had little effect on the success of the method (Table 2, lines I and J). When the resolution of the error-free data was limited to 1.5 Å the refinement of the true phases gave an MPE of only 22° (Table 2, line K). However, when random phases were used the best MPE was 74° and this phase set was not clearly indicated by the figures of merit (Table 2, line L).

The largest structure we considered was d(CGCGXG) (Table 1, data set 3). Refinement of either the true or random phases using the observed 1.7 Å data both gave an MPE of 69° (Table 2, lines M and N), and in the latter case the best phase set was not distinguished from the others by the figures of merit. Indeed the value of CFOM was much lower than the maximum of 2.471 attained with another set. Refinement against error-free simulated data at 1.0 Å resolution gave an MPE of 28° with the true phases (Table 2, line O) and 29° with random phases (Table 2, line P) and the best set was clearly indicated by the minimum RESID and TFOM and the maximum CFOM. An E map calculated with these phases revealed 155 peaks amongst the top 400 that were within 0.30 Å of true atomic sites. The highest peaks corresponded to base atoms and the lowest corresponded to sugar and phosphate atoms. The addition of 10% errors to the intensities did not lead to deterioration of the MPE's (Table 2, lines Q and R). At 1.5 Å resolution phase refinement was unsuccessful (Table 2, lines S and T) and, when starting from random phases, the best set was not indicated by the figures of merit.

The general pattern emerging from these results is that the success of direct methods is critically limited by the resolution of the data, but is less dependent upon the complexity of the structure or the presence of errors in the data. It is of some interest to investigate the radius of convergence of the tangent refinement. Scatter plots were produced of the refined mean phase error as a function of the initial mean phase error. This was done by refining amalgamated phases in the Sayre-tangent formula. The amalgamated phases, φ_f, were obtained by combining the true phases, φ_t, with random phases, φ_r, using the following expression,

$$\tan \varphi_f = [w \sin \varphi_t + (1 - w) \sin \varphi_r] / [w \cos \varphi_t + (1 - w) \cos \varphi_r]. \quad (2)$$

The weight w assumes a value in the range $0.0 \leq w \leq 1.0$, with $w = 1.0$ producing true phases and $w = 0.0$ giving random phases. Thus, by varying the value of w, phases were generated having a mean phase error lying in the range $0.0 \leq \text{MPE} \leq 90.0°$.

Amalgamated phases were produced for both the observed and simulated data sets for the structure d(CGCG*X*G). The value of the weight w was increased from 0.0 to 1.0, in increments of 0.01, and so in each case a set of amalgamated phases was generated, having an initial MPE which decreased overall from 90.0 to 0.0°. These phases were input to *SAYTAN* for refinement and a final MPE was calculated for each set of phases; the results were then plotted in scatter diagrams. Error-free simulated data sets at resolutions of 1.0, 1.25, 1.5, 1.75 and 2.0 Å, and also the observed data set at 1.7 Å resolution were used; the numbers of large E-value reflections refined with *SAYTAN* in each case were 1000, 800, 600, 600, 400 and 600, respectively. The water atoms were included in the calculations of all the mean phase errors. The resulting scatter plots are shown in

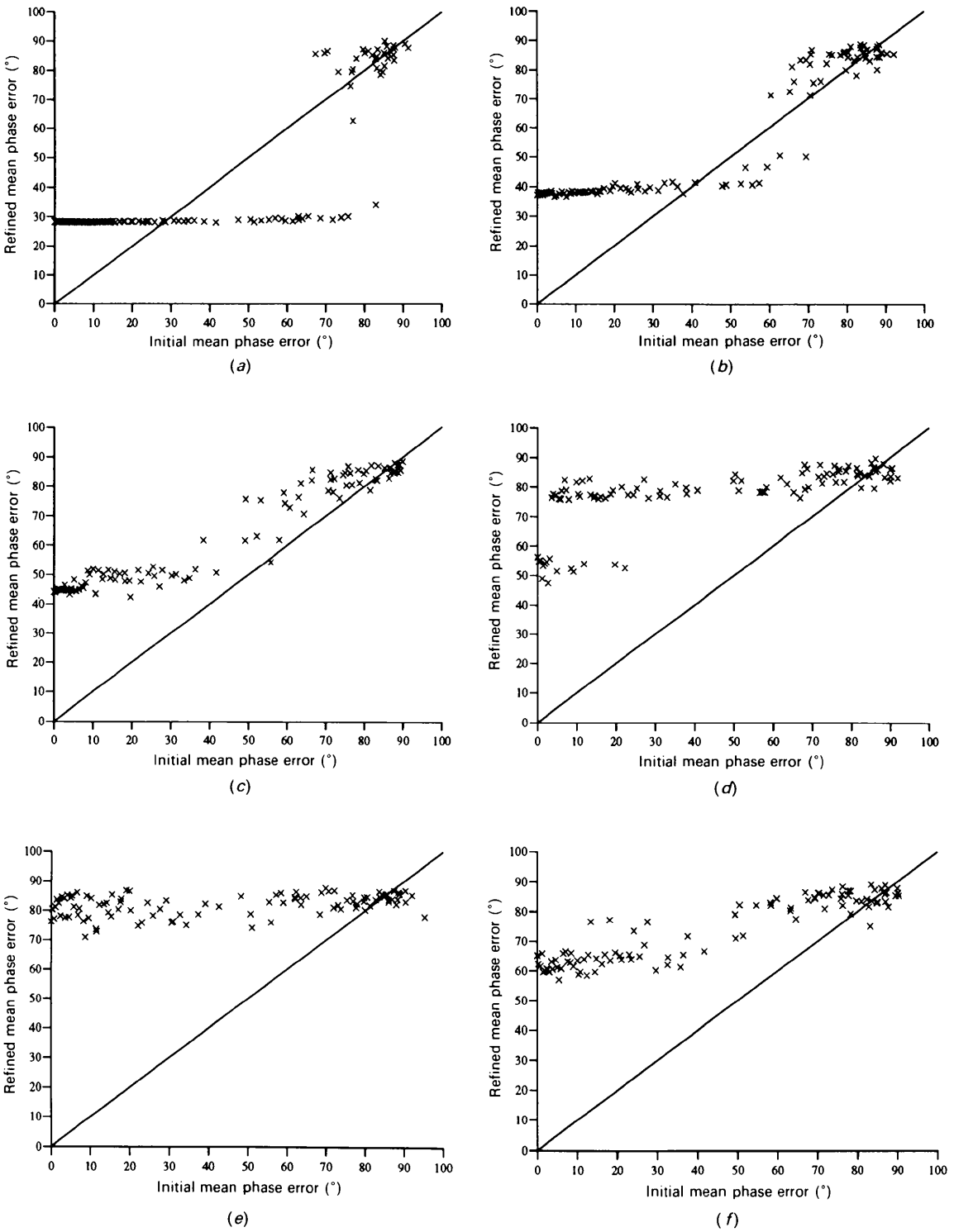

Fig. 1. Refinement of amalgamated phases by *SAYTAN*. Structure: d(CpGpCpGp*X*pG) *X* = meO⁴C. Observed *B* factors used in data simulation. Simulated data are at resolutions: (*a*) 1.0, (*b*) 1.25, (*c*) 1.5, (*d*) 1.75, (*e*) 2.0 Å; (*f*) observed data at resolution 1.7 Å.

Fig. 1. The line corresponding to the initial and final MPE's being equal is drawn in all the diagrams. Clearly, only scatter points which fall below this line represent a positive phase refinement, that is, the final MPE is less than the initial MPE.

The scatter plots show that useful phase refinement by *SAYTAN* is only achieved in the case of the 1.0 Å resolution data set, for which the structure can be solved starting from random phases. Interestingly, the plots show a two-state pattern, with nearly all the refined phases having values either close to the true phases or else having random values. For the 1.25 Å resolution data set, there are only a few points lying below the drawn line; positive phase refinement is only achieved when the initial mean phase error is in the intermediate range of 40–60°. At resolutions ≥ 1.5 Å, there is no effective phase refinement and, at 2.0 Å resolution, even refinement of the true phases does not give a good solution. The phase refinement of the observed data at 1.7 Å resolution confirms this pattern, with no marked improvement in the phases for all initial mean phase errors in the range 0–90°.

The diffraction data from oligonucleotide crystals are affected by thermal disorder. For example, the *B* factors of phosphate group atoms are typically about 15–30 Å^2. It might be supposed that this could affect the ability of direct methods to solve oligonucleotide structures. Therefore, we simulated data from d(CGCG*X*G) with reduced *B* factors (Table 1, data set 4) that would be consistent with a low-temperature diffraction pattern. However, phase refinement using these data (Table 2, lines U and V) was less successful than in the other cases described above. Even setting all the *B* factors to zero and reducing the complexity of the structure by removing all the solvent molecules (Table 1, data set 5) led to no improvement (Table 2, lines W and X). Scatter plots for both these cases (not shown) were very similar to those in Fig. 1 (Hubbard, 1994). Thus, reducing the values of the atomic thermal parameters appears not to improve the effectiveness of direct methods at a given resolution. However, collecting the diffraction data at a low temperature will not only reduce the thermal parameters, but will also change some other important factors. For example, at low temperature it is probable that the mosaicity of the crystal will be altered and that the crystal will be less subject to radiation damage, which will affect both the quality and resolution of the data obtained.

Phase refinement with heavy-atom data

The protein aPP is similar in size to the hexamer d(CGCG*X*G), and it has been solved using direct methods, even with low-resolution data (Woolfson & Yao, 1990; Mukherjee & Woolfson, 1993). However, aPP contains a Zn atom which helps in the solution. We have, therefore, considered whether a DNA molecule containing a heavy atom could be solved by direct methods. We simulated data for the DNA hexamer containing both one (Table 1, data set 6) and two (Table 1, data set 7) Br atoms, and we then attempted phase refinement with *SAYTAN* at various resolutions as described above. The results with DNA + 1Br are summarized in lines Y to MM of Table 2. Clearly, the presence of the heavy atom enables *SAYTAN* to produce good phase sets, from which the structure could be solved, even at 2.5 Å resolution. However, the conventional *SAYTAN* figures of merit were not always clearly indicative of the best solution. For example, with the 1.0 Å data (Table 2, line Z) set 340 had the maximum value of ABSFOM, set 714 had the maximum PSIZERO, set 832 had the minimum RESID, set 935 had the maximum CFOM and set 824 had the minimum TFOM. The sets with the highest values of CFOM did not correspond to the best solutions; the low values of PSIZERO for these sets gave the high values of CFOM. At the lower resolutions there were similar problems, which were particularly acute at resolutions of 2.0 Å or less, where the best solutions were indicated by none of the conventional figures of merit (Hubbard, 1994). Therefore, while one can obtain a direct methods solution in this case using data at 2.5 Å resolution, the effective resolution limit, for being able to identify correctly a good solution according to the *SAYTAN* figures of merit, is about 2.0 Å or higher. These findings corrobate those of Mukherjee & Woolfson (1993), who reported that useful phase sets could be obtained for the aPP structure at resolutions as low as 3.0 Å, but that the conventional figures of merit were not very discriminating for the phase sets developed. They proposed modified figures of merit which were found to be capable of selecting the better phase sets, at least for those produced from data having resolutions of 2.0 Å or higher. These modified figures of merit have not been tested in this study, but they would conceivably offer improved discrimination of the best phase sets. Hence, they would lower the resolution necessary for the practical solution of such heavy-atom oligonucleotide structures using direct methods.

Scatter plots of the refined MPE *versus* initial MPE are shown (Fig. 2) at resolutions of 1.0, 1.25, 1.5, 1.75, 2.0 and 2.5 Å. The numbers of strong reflections refined in each case were 1000, 800, 600, 600, 400 and 400, respectively. Comparison of Fig. 1 and 2 shows that the presence of the heavy atom results in a considerable improvement in the phase refinement, particularly with regard to the low-resolution refinements. Further improvements in the final MPE are gained when two Br atoms are added

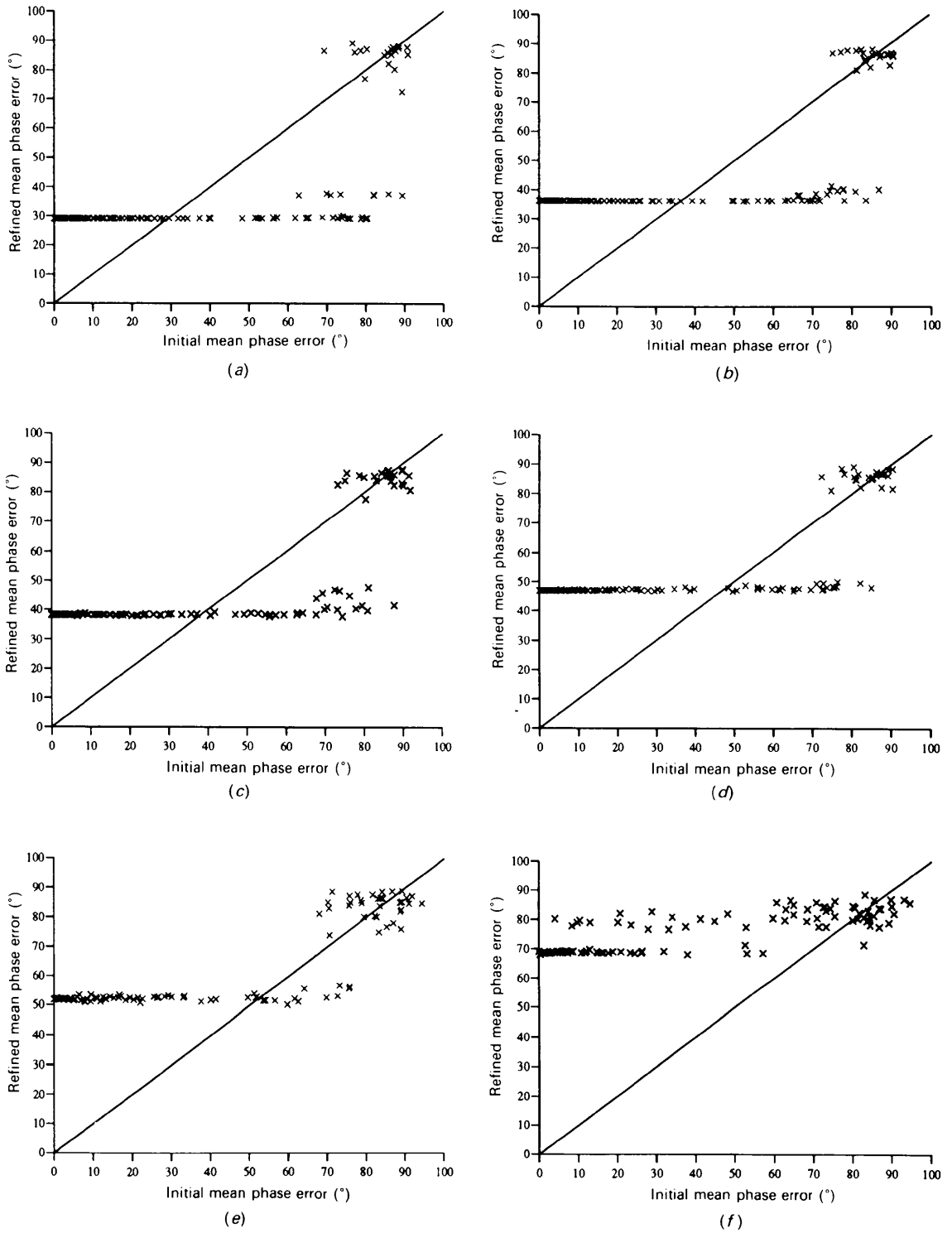

Fig. 2. Refinement of amalgamated phases by *SAYTAN*. Structure: d(CpGpCpGp^{Br}CpG) – 1 Br. Observed B factors used in data simulation; Br atom has $B = 10$ Å^2. Simulated data are at resolutions: (*a*) 1.0, (*b*) 1.25, (*c*) 1.5, (*d*) 1.75, (*e*) 2.0, (*f*) 2.5 Å.

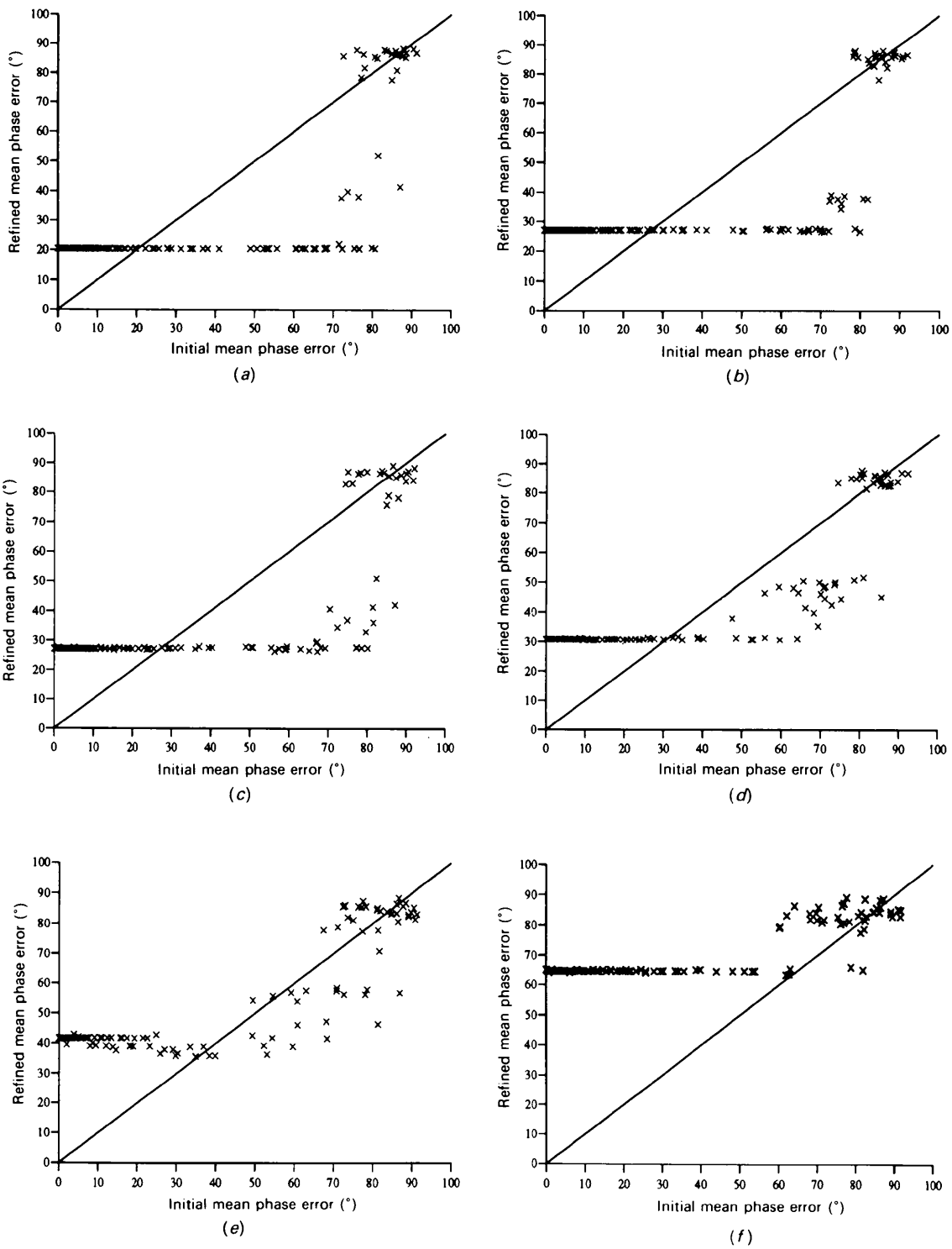

Fig. 3. Refinement of amalgamated phases by *SAYTAN*. Structure: d(CpGpCpGp^{Br}C$_P$G) – 2 Br. Observed B factors used in data simulation; Br atoms have $B = 10$ Å^2. Simulated data are at resolutions: (*a*) 1.0, (*b*) 1.25, (*c*) 1.5, (*d*) 1.75, (*e*) 2.0, (*f*) 2.5 Å.

to the DNA (Fig. 3). In this case, solutions could be found starting from multiple sets of random phases at all resolutions.

Concluding remarks

Direct methods should be able to solve the structure of oligonucleotides even at relatively low resolution, provided that heavy atoms (*e.g.* Br) are present. However, in the absence of heavy atoms, success is dependent upon the availability of high-resolution ($\sim$1.0 Å) data. With data of high enough resolution the quality of the data is not critical, although large errors would obviously have deleterious effects.

We wish to thank Drs Yao Jia-xing, L. S. Refaat and C. Tate for their helpful discussions of this work. We are also grateful to Dr M H Moore who provided the Z-DNA hexamer data. SRH was supported by a Science and Engineering Reseach Council studentship.

References

DEBAERDEMAEKER, T., TATE, C. & WOOLFSON, M. M. (1985). *Acta Cryst.* A**41**, 286–290.

DEBAERDEMAEKER, T. & WOOLFSON, M. M. (1983). *Acta Cryst.* A**39**, 193–196.

DREW, H. R. & DICKERSON, R. E. (1981). *J. Mol. Biol.* **152**, 723–736.

DREW, H. R., DICKERSON, R. E. & ITAKURA, K. (1978). *J. Mol. Biol.* **125**, 535–543.

DREW, H. R., TAKANO, T., TANAKA, S., ITAKURA, K. & DICKERSON, R. E. (1980). *Nature (London)*, **286**, 567–573.

DREW, J. R., WING, R. M., TAKANO, T., BROKA, C., TANAKA, S., ITAKURA, K. & DICKERSON, R. E. (1981). *Proc. Natl Acad. Sci. USA*, **78**, 2179–2183.

HARRISON, R. W. (1989). *Acta Cryst.* A**45**, 4–10.

HAUPTMAN, H. & HAN, F. (1993). *Acta Cryst.* D**49**, 3–8.

HUBBARD, S. R. (1994). DPhil thesis, Univ. of York, England. Submitted.

JOSHUA-TOR, L., RABINOVICH, D., HOPE, H., FROLOW, F., APPELLA, E. & SUSSMAN, J. L. (1988). *Nature (London)*, **334**, 82–84.

KENNARD, O. (1985). *J. Biomol. Struct. Dynam.* **3**, 205–225.

MAIN, P., DEBAERDEMAEKER, T., GERMAIN, G., REFAAT, L. S., TATE, C. & WOOLFSON, M. M. (1988). *MULTAN88. Computer Programs for the Automatic Solution of Crystal Structures from X-ray Diffraction Data.* Univs. of York, England, and Louvain, Belgium.

MOORE, M. H., HUNTER, W. N., LANGLOIS D'ESTAINTOT, B. & KENNARD, O. (1989). *J. Mol. Biol.* **206**, 693–705.

MUKHERJEE, M. & WOOLFSON, M. M. (1993). *Acta Cryst.* D**49**, 9–12.

QUIGLEY, G. J., WANG, A. H.-J., UGHETTO, G., VAN DER MAREL, G. A., VAN BOOM, J. H. & RICH, A. (1980). *Proc. Natl Acad. Sci. USA*, **77**, 7204–7208.

RABINOVICH, D. & SHAKKED, Z. (1984). *Acta Cryst.* A**40**, 195–200.

RAMAKRISHNAN, B. & VISWAMITRA, M. A. (1988). *J. Biomol. Struct. Dynam.* **6**, 511–523.

VAN MEERVELT, L., MOORE, M. H., LIN, P. K. T., BROWN, D. M. & KENNARD, O. (1990). *J. Mol. Biol.* **216**, 773–781.

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

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

WOOLFSON, M. M. & YAO, J.-X. (1990). *Acta Cryst.* A**46**, 409–413.

YAO, J.-X., ZHENG, C.-D., QIAN, J.-Z., HAN, F.-S., GU, Y.-X. & FAN, H.-F. (1985). *SAPI85. A Computer Program for Automatic Solution of Crystal Structures from X-ray Diffraction Data.* Institute of Physics, Chinese Academy of Sciences, Beijing, China.