Acta Cryst. (1995). D**51**, 979–989

On the Application of Anomalous Scattering in Oligonucleotide Crystallography

By S. R. Hubbard, R. J. Greenall and M. M. Woolfson

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

(*Received* 1 *February* 1995; *accepted* 11 *May* 1995)

Abstract

Simulated anomalous-scattering differences, at wavelengths between 1.5 and 5.5 Å, were used with *MULTAN* to locate P atoms in an oligonucleotide hexamer. The success of the method depended heavily on the level of errors in the data. With error-free data most or all P atoms were located at all wavelengths. With noisy data, the best results were obtained by refining the phases associated with the largest values of $|\Delta F|/\sigma(|\Delta F|)$ rather than with the largest values of $|\Delta F|$. In this case a few of the P-atom positions could be located, with the best results occurring at wavelengths between 3.0 and 4.0 Å. Further improvements were gained by reducing the values of the thermal parameters of the P atoms. *MULTAN* figures of merit had limited success in indicating the best phase sets, but a small improvement was gained by modifying the procedure for selecting those reflections used in the calculation of PSIZERO.

Introduction

We have investigated the use of anomalous scattering with direct methods in the solution of oligonucleotide crystals. Anomalous scattering results in a breakdown of Friedel's Law, that is $|F^+(\mathbf{h})| \neq |F^-(\mathbf{h})|$, where $F^+(\mathbf{h})$ and $F^-(\mathbf{h})$ are the structure factors of the reflection and its Friedel mate, **h** and **-h**, respectively (Fig. 1). The structure factor $F^+(\mathbf{h})$ and the complex conjugate of its inverse, $F^-(\mathbf{h})^*$, can be expressed as follows,

$$F^+(\mathbf{h}) = F_N(\mathbf{h}) + F'_A(\mathbf{h}) + F''_A(\mathbf{h}) = F'(\mathbf{h}) + F''_A(\mathbf{h})$$
$$F^-(\mathbf{h})^* = F_N(\mathbf{h}) + F'_A(\mathbf{h}) - F''_A(\mathbf{h}) = F'(\mathbf{h}) - F''_A(\mathbf{h}), \quad (1)$$

where $F_N(\mathbf{h})$ is the structure-factor term due to the normal scatterers, $F'_A(\mathbf{h})$ is the term due to the real part of the scattering by the anomalous scatterers, and $F''_A(\mathbf{h})$ is the term due to the imaginary part of the scattering factor, $\Delta f''$, of the anomalous scatterers. If there is only one type of anomalous scatterer present in the structure, which is generally the case for native oligonucleotides, then $F''_A(\mathbf{h})$ is perpendicular to $F'_A(\mathbf{h})$. The anomalous-scattering difference, $\Delta F(\mathbf{h})$, is related to $F''_A(\mathbf{h})$ by the following approximation (Hendrickson, Smith & Sheriff, 1985),

$$\Delta F(\mathbf{h}) = |F^+(\mathbf{h})| - |F^-(\mathbf{h})| \simeq 2|F''_A(\mathbf{h})| \cos(\Delta\varphi), \quad (2)$$

where $\Delta\varphi = \psi + \nu - \varphi$ (Fig. 1).

In principle, the Friedel differences, $|\Delta F|$, can be used to locate the positions of the anomalous scatterers in the structure. The complete structure might then be solved by employing one of a number to techniques which are available (Fan, Woolfson & Yao, 1993). At the wavelengths commonly used in X-ray diffraction experiments, the anomalous-dispersion effects are usually only significant if the structure contains a heavy-metal atom. Nevertheless, it has been shown by Hendrickson & Teeter (1981) that the anomalous differences were sufficiently large, with Cu $K\alpha$ radiation, to solve the sulfur-containing protein crambin. Theses authors also suggested that the same technique could be used to solve short

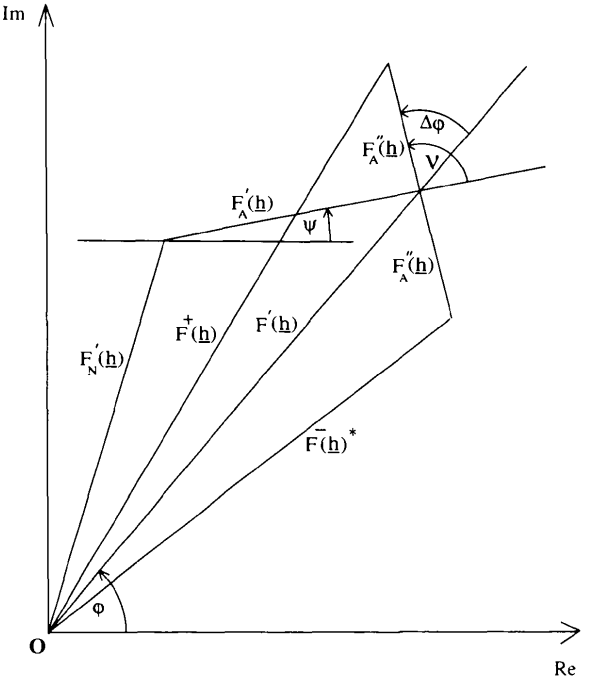

Fig. 1. Argand diagram of the structure factors of the reflection and its Friedel mate, $F^+(\mathbf{h})$ and $F^-(\mathbf{h})$, respectively. The structure factor of the inverse reflection is shown reflected across the real axis and is indicated as $F^-(\mathbf{h})^*$. In the case considered in this work, all the anomalous scatterers are identical, and so $\nu = \pi/2$.

oligonucleotide sequences such as d(CGCGCG). The availability of tunable radiation from synchrotron sources allows the collection of X-ray diffraction data at wavelengths extending beyond 3.0 Å, where the anomalous scattering from the atoms of intermediate atomic number in a structure becomes more significant . For example, Lehmann, Müller & Stuhrmann (1993) and Stuhrmann, Hütsch, Trame, Thomas & Stuhrmann, (1995) have made measurements for lysozyme crystals and ribosome crystals using X-rays with a wavelength of approximately 5.0 Å. Therefore, the possibility exists of solving some protein and DNA structures by measuring the anomalous scattering from S or P atoms at wavelengths near their absorption edges of 5.018 and 5.787 Å, respectively. However, at such long wavelengths, absorption is considerable , although it may be reduced through the use of small crystals. It is thus a question of balancing the benefit of collecting data at a longer wavelength, in order to give an increased anomalous-scattering signal, against the problems caused by greater absorption.

One approach for finding the anomalous scatterers is to calculate a Patterson map, using $|\Delta F|^2$ as the coefficients (Rossmann, 1961), in which the peaks correspond to vectors between the anomalous scattering sites. Strictly $|F_A''|^2$ should be used in the calculation, but both types of map have similar major features. If there are many sites in the structure then the interpretation of the map becomes rather difficult, and, in such cases, direct methods, utilizing observed $|\Delta F|$ or $|\Delta E|$ values, provide an alternative approach. From (2), large values of $|\Delta F|$ are related to large values of $|F_A''|$, although the converse is not always true; therefore, the largest $|\Delta F|$ values form a subset of the largest $|F_A''|$ values. Since direct methods would make use of only the largest $|\Delta E|$ values, they are particularly suitable for locating the anomalous scatterers. The observed differences, $|\Delta F|_{obs}$, can be converted to normalized $|\Delta E|$ values, and, for the largest differences, multiple sets of different random phases can be generated and then refined using the tangent formula. E maps calculated from the best sets of phases should reveal the anomalous scatterer's positions. Such an approach was tested by Mukherjee, Helliwell & Main (1989), who successfully used observed anomalous-scattering differences with *MULTAN* (Main *et al.*, 1988) to determine the positions of the anomalous scatterers in three metalloproteins and in a small molecule. However, in their concluding remarks, the authors stated that further work was required to discover whether this technique would be effective in solving a structure containing many anomalous scatterers.

We present here the results of using *MULTAN* with anomalous differences to locate the P atoms in oligonucleotide crystal structures over a range of wavelengths. The effect of observational errors on the efficacy of the methods has been investigated, and we suggest a method for choosing the differences which should be used. We also considered the effect of varying the thermal B factors and the number of differences used in the refinement. The method is limited in that the conventional *MULTAN* figures of merit are not very successful in discriminating the best phase sets; therefore, we have considered the possibility of using alternative figures of merit.

Methods

Data simulation

Various error-free and noisy X-ray data sets were simulated for the Z-DNA structure d(CGCG*X*G) (Van Meervelt, Moore, Lin, Brown & Kennard, 1990), where X is the modified base MeO^{4}C. The asymmetric unit contains 116 C, 48 N, 148 O and ten P atoms plus 78 water molecules. The space group is $P2_12_12_1$.

The program *CROSSEC* (Cromer, 1983) was used to calculate the real parts, $\Delta f'$, and the imaginary parts, $\Delta f''$, of the anomalous contributions to the atomic scattering factors for all the atoms in the structure (Table 1). The anomalous scattering from phosphorus increases with wavelength, λ, up to the K absorption edge of phosphorus, which occurs at 5.787 Å, so the data sets were simulated, using the program *SAPI* (Yao *et al.*, 1985), at wavelengths of 1.50, 1.80, 2.10, 2.40, 2.70, 3.00, 3.30, 3.50, 3.70, 4.00, 5.00 and 5.50 Å. The resolution of the observed data for the structure d(CGCG*X*G) was 1.7 Å, so the error-free data sets were also simulated to this resolution at the seven shortest wavelengths. At the five longest wavelengths, the resolutions of the data sets were restricted to 1.75, 1.85, 2.00, 2.50 and 2.75 Å, respectively, which in each case corresponded to the maximum allowed value of $\lambda/2$.

The P atoms are found on the extremities of the DNA duplexes, and so they are subject to the greatest static and thermal disorder. As a result, the Debye factors, B_P, of the P atoms are generally higher than those, B_O, of the other atoms in the DNA molecule. The water molecules tend to be the most disordered, and hence their Debye factors, B_W, are generally higher than B_P and B_O. Although there is usually a distribution of observed B_P, B_O and B_W values, for simplicity we have assumed that all the P atoms have the same Debye factors, and similarly for the other atoms in the DNA and the water molecules. The static disorder in such a crystal structure could be reduced by producing more perfect crystals, whereas the thermal disorder may be reduced by collecting the X-ray data at a lower temperature. In this study, we considered the effect of reducing the B values on the success of the anomalous-scattering methods. For each of the 12 wavelengths listed above, three data sets were simulated, with the atoms having the following B factors: for set 1, $B_P = 30.0$, $B_O = 10.0$, and $B_W = 40.0$ Å^2; for set 2, $B_P = 15.0$, $B_O = 5.0$ and

Table 1. *Real and imaginary parts of the anomalous scattering factors at various wavelengths for the atoms in the structure d*(CGCGXG)

Wavelength	C		N		O		P	
(Å)	$\Delta f'$	$\Delta f''$	$\Delta f'$	$\Delta f''$	$\Delta f'$	$\Delta f''$	$\Delta f'$	$\Delta f''$
1.50	0.016	0.009	0.028	0.017	0.044	0.030	0.276	0.412
1.80	0.023	0.013	0.039	0.025	0.061	0.045	0.328	0.580
2.10	0.030	0.018	0.051	0.035	0.078	0.061	0.365	0.769
2.40	0.038	0.024	0.063	0.046	0.097	0.080	0.382	1.051
2.70	0.046	0.030	0.077	0.058	0.115	0.102	0.365	1.283
3.00	0.055	0.038	0.090	0.072	0.134	0.126	0.314	1.529
3.30	0.064	0.046	0.104	0.088	0.153	0.152	0.233	1.787
3.50	0.070	0.052	0.113	0.099	0.166	0.171	0.156	1.966
3.70	0.077	0.058	0.123	0.111	0.178	0.190	0.112	2.057
4.00	0.086	0.068	0.137	0.129	0.197	0.221	−0.056	2.337
5.00	0.119	0.106	0.183	0.200	0.254	0.339	−1.176	3.342
5.50	0.136	0.128	0.206	0.240	0.279	0.405	−2.669	3.891

$B_W = 20.0\,\text{Å}^2$; for set 3, $B_P = 10.0$, $B_O = 3.0$ and $B_W = 15.0\,\text{Å}^2$. Set 1 corresponds to values that are typically found when oligonucleotide crystals are solved and refined using data obtained at room temperature. For brevity we shall henceforth refer to these sets by the B_P values alone.

Since the anomalous scattering differences are often rather small, the accuracy of the X-ray measurements is obviously an important consideration. Therefore, to investigate the sensitivity of our methods to the accuracy of the data, random errors having Gaussian distributions were added to the error-free data sets. This was carried out for all the data sets using a similar procedure to that employed previously (Hubbard, Greenall & Woolfson, 1994). In this earlier work, we were attempting to solve the structures of oligonucleotide crystals using simulated structure factors in the direct-methods program *SAYTAN* (Debaerdemaeker, Tate & Woolfson, 1985). The standard deviation $\sigma(I)$ of each reflection having an intensity I was given a value of $I/10$, which assumed that the background intensity was negligible. This approximation is reasonable for the strong reflections, but it is much less so for the weak reflections. However, since direct methods involve the estimation of triple-phase relationships between the strong reflections, the procedure used in the earlier work was justified. In the present paper, we are considering large anomalous-scattering differences $|\Delta F|$ and, in principle, these may be found for either moderately weak or strong reflections. Therefore, in this case we adopted a more sophisticated approach for simulating the noisy data which takes into account the presence of a background intensity. For each reflection of intensity I, the standard deviation $\sigma(I)$ was calculated from a quadratic function of the form,

$$\sigma(I) = aI^2 + bI + c, \qquad (3)$$

where the coefficients a, b and c were determined from a least-squares fit of the function to the observed distribution of $\sigma(I)$ values, which were found in the experimental data set for the structure d(CGCGXG) (Van Meervelt *et al.*, 1990). The following values for the coefficients a, b and c were found: $a = 2.85 \times 10^{-8}$, $b = 4.15 \times 10^{-3}$ and $c = 6.90 \times 10^2$. The root-mean-square difference between the observed and calculated values of $\sigma(I)$ was 24%. The experimental values of $\sigma(I)$ and the result of the least-squares fit are shown in Fig. 2. A random number generator based on Gaussian distributions with mean values of zero and standard deviations $\sigma(I)$ was then used to produce corrections ΔI to the intensities I of the reflections. In this way, uncorrelated random errors were added to the intensity of the reflection, $I(hkl)$, and to its Friedel mate, $I(\bar{h}\bar{k}\bar{l})$, for all the pairs of reflections in each data set.

Phase refinement

Initially, we considered the ideal, error-free data simulated at the various wavelengths with different thermal parameters. For these data sets, we assumed that the large values of $|\Delta F|$ represented a subset of the large anomalous-scattering contributions from the phosphorus substructure, but corrupted by an unknown factor $\cos(\Delta\varphi)$ [see (2)]. Therefore, we attempted to solve the substructure by using the largest values of $|\Delta F|$ as structure amplitudes. For each of the error-free data sets, the true phases, derived from the phosphorus substructure, were refined using the tangent formula. In each case the 250 largest values of $|\Delta F|$ were used in the refinement. Centric reflections were excluded from the analysis, since their anomalous scattering differences are zero. The procedure was then repeated starting from 800 sets of random phase, and the mean phase error of each refined phase set was calculated.

E maps were calculated for the phase set refined from the true phases and for that which had the lowest mean phase error of the sets refined from random phases. These calculated maps were compared with the true map,

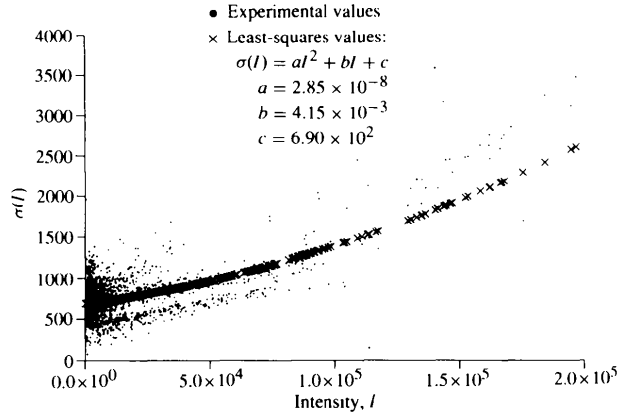

Fig. 2. A least-squares fit of a quadratic function to the distribution of $\sigma(I)$ values in the observed data set for the structure d(CGCGXG).

Table 2. *Anomalous scattering signal,* $\langle|\Delta F|^2\rangle^{1/2}/\langle|F'|^2\rangle^{1/2}$, *at various wavelengths for the atoms in the structure* d(CGCG*X*G)

The numbers of atoms of each kind in the asymmetric unit are shown in parentheses.

	$\langle\|\Delta F\|^2\rangle^{1/2}/\langle\|F'\|^2\rangle^{1/2}$			
Wavelength	P	O	N	C
(Å)	(10)	(148)	(48)	(116)
1.50	1.41	0.39	0.13	0.10
1.80	1.98	0.59	0.19	0.15
2.10	2.63	0.80	0.26	0.21
2.40	3.59	1.05	0.34	0.28
2.70	4.38	1.34	0.43	0.35
3.00	5.22	1.66	0.54	0.44
3.30	6.10	2.00	0.66	0.53
3.50	6.71	2.25	0.74	0.60
3.70	7.02	2.50	0.83	0.67
4.00	7.98	2.90	0.96	0.79
5.00	11.41	4.45	1.50	1.38
5.50	13.28	5.32	1.80	1.58

in order to try to match the positions of the highest peaks with the true P-atom positions.

Next, we considered the effect of adding noise to the error-free data sets. The method adopted here was slightly different from that outlined above for the error-free data sets. Instead of refining the phases associated with the largest values of $|\Delta E|$, the 250 phases used were those associated with the largest values of $|\Delta F|/\sigma(|\Delta F|) = |\Delta E|/\sigma(|\Delta E|)$. The justification of this is that it tends to select differences that are large and whose errors are small. We surmised that using differences with these two properties would improve the chance of solving the structure since the anomalous-scattering signal due to the P atoms in d(CGCG*X*G) is rather weak, particularly at the shorter wavelengths (Table 2) and, therefore, error prone. Most of the results presented here were obtained using this method, which we will call procedure 1. To test whether this was indeed the best way in which to select the differences, we also tested two alternatives. In procedure 2 the phases associated with the 250 largest values of $|\Delta E|/\sigma(|\Delta F|)$ were refined in the tangent formula. Procedure 3 was that described by Mujherjee, Helliwell & Main (1989), who used the phases associated with the largest values of $|\Delta E|$, but who excluded the error-prone weak reflections with $F_{obs} < 4\sigma(F_{obs})$ and also any reflections for which $|\Delta E| > 5 \times$ r.m.s.$(|\Delta E|)$.

In each case described above, we have refined 250 phases. This number was selected rather arbitrarily and, therefore, we have also investigated the effect of varying the number of phases used in the refinement between 150 and 400.

The quality of the refined phase sets was judged according to the standard *MULTAN* figures of merit ABSFOM, PSIZERO and RESID, and also from a combined figure of merit, CFOM, plus the experimental figure of merit, TFOM (which is defined by Hubbard *et al.*, 1994). Unless otherwise stated, the weights associated with ABSFOM, PSIZERO and RESID used to calculate CFOM were given their default values of 0.60, 1.20 and 1.20, respectively.

Results

Strength of the anomalous signal

It is useful to calculate the expected r.m.s. value of the anomalous-scattering difference, $\langle|\Delta F|^2\rangle^{1/2}$, expressed as a fraction of the expected value of the total scattering, $\langle|F'|^2\rangle^{1/2}$, at zero scattering angle, which is given by (Hendrickson, Smith & Sheriff, 1985),

$$\left(\langle|\Delta F|^2\rangle^{1/2}\right)/\left(\langle|F'|^2\rangle^{1/2}\right) = 2^{1/2}(N_A/N)^{1/2}(\Delta f_A''/Z_{\text{eff}}), \quad (4)$$

where N_A is the number of anomalous scatterers per molecule, N is the number of non-H atoms per molecule, and Z_{eff} is the effective average atomic number [which is approximately 6.7 for proteins and 7.3 for oligonucleotides (Hendrickson & Teeter, 1981)]. The results (Table 2), show that the anomalous scattering due to the P atoms is the dominant effect, accounting for about 60–70% of the total anomalous scattering at each wavelength. However, other atoms in the structure, in particular the O atoms, also produce a significant anomalous-scattering signal, especially at the longer wavelengths.

Refinement with error-free data

The results of the phase refinements are summarized in Tables 3, 4 and 5. They show that, except at the two longest wavelengths, a good set of phases can be found by refining multiple sets of random phases using the error-free data, in the sense that the E map calculated from these phases reveals most if not all of the P-atom positions in the structure (for example, see Table 6). Therefore, in principle, a *MULTAN* run starting from multiple sets of random phases can produce a substantially correct solution.

The mean phase errors are smaller at the shorter than at the longer wavelengths, although this general trend is blurred by statistical fluctuations. In particular, at the two longest wavelengths, 5.00 and 5.50 Å, the results obtained from both refining the true phases and refining random phases are markedly worse. This is because, at the shorter wavelengths, the 250 phases used in the refinement were chosen from a larger set of data extending to a resolution of 1.7 Å, whereas the resolutions of the data sets at the longer wavelengths were necessarily more restricted. Therefore, at the shorter wavelengths, the normalized $|\Delta E|$ values were in general larger and so the phase relationships were more reliable. Consequently, the phases associated with the largest values of $|\Delta F|$ refined to give lower mean phase errors.

Hubbard, Greenall & Woolfson (1994) showed that attempts to solve the structure (dCGCG*X*G) by refining

Table 3. *Results obtained using MULTAN with anomalous scattering data sets for the structure d*(CGCGXG)

These data sets were simulated with the P atoms having thermal factors: $B_P = 30\,Å^2$. For each data set of wavelength λ, NAR gives the number of pairs of acentric reflections and RES specifies the resolution. For the error-free data sets, the phases associated with the 250 largest values of $|\Delta E|$ were refined by the tangent formula. For the noisy data sets, the phase refinement used the 250 largest values of $|\Delta F|/\sigma(|\Delta F|)$. For each data set, the upper figure gives the mean phase error (°), calculated from the ten P-atom positions in the asymmetric unit, after the true phases were refined. The lower figure gives the lowest mean phase error (°) after phase refinement using the tangent formula, starting from 800 sets of different random phases. For each of the resulting E maps, the number of peaks at distances < 1.0 Å from true P-atom positions, together with the number of the lowest successfully matched peak, are given in parentheses. For the noisy data sets, the average signal-to-noise ratio, $\overline{|\Delta F|}/\overline{\sigma(|\Delta F|)}$, is given for all the acentric reflections (SNR_{All}) and for the 250 acentric reflections with the largest values of $|\Delta F|/\sigma(|\Delta F|)$ which were used in the phase refinement (SNR_{250}).

Wavelength λ (Å)	NAR	RES (Å)	Error-free data	Noisy data	SNR_{All}	SNR_{250}
1.50	2233	1.70	17 (10/10)	64 (4/10)	0.61	2.10
			21 (10/10)	75 (1/1)		
1.80	2233	1.70	19 (10/10)	65 (4/10)	0.62	2.13
			28 (10/10)	73 (2/5)		
2.10	2233	1.70	21 (10/10)	69 (2/5)	0.63	2.17
			22 (10/10)	75 (1/2)		
2.40	2233	1.70	16 (10/10)	61 (5/13)	0.64	2.25
			41 (10/13)	70 (2/20)		
2.70	2233	1.70	19 (10/10)	54 (7/15)	0.65	2.34
			31 (10/10)	72 (1/1)		
3.00	2233	1.70	19 (10/10)	63 (3/4)	0.65	2.45
			42 (9/11)	69 (1/1)		
3.30	2233	1.70	18 (10/10)	64 (3/8)	0.65	2.59
			36 (10/11)	69 (1/1)		
3.50	1937	1.75	22 (10/10)	63 (2/17)	0.72	2.80
			36 (10/12)	58 (4/10)		
3.70	1709	1.85	20 (10/10)	63 (4/16)	0.79	2.79
			30 (10/10)	66 (1/1)		
4.00	1331	2.00	18 (10/10)	57 (4/11)	0.89	2.81
			52 (7/18)	66 (2/4)		
5.00	988	2.50	36 (10/10)	69 (3/17)	1.44	3.66
			61 (6/19)	65 (3/13)		
5.50	758	2.75	65 (3/7)	41 (7/10)	1.91	3.92
			64 (5/7)	65 (4/19)		

multiple sets of random phases in the tangent formula were unsuccessful, even when using the exact simulated $|F|$ values, if the resolution of the data was worse than about 1.0 Å. However, the results given here show that the P atoms can be located from random starting phases using the error-free simulated $|\Delta F|$ values, when the resolution of the data is in the range 1.7–2.0 Å. Therefore, success in using direct methods with anomalous-scattering differences is less critically dependent upon having high-resolution data. The reason for this is that the P atoms are separated by about 6.0–8.0 Å, so that they can be resolved using lower resolution data.

The main conclusion from this is that, in principle, one can determine the positions of the P atoms in an oligonucleotide structure using this approach.

Refinement with noisy data

The results of refinements using the noisy data sets are also given in Tables 3, 4 and 5. The results clearly

Table 4. *Results obtained using MULTAN with anomalous scattering data sets for the structure d*(CGCGXG).

As for Table 3, except that these data sets were simulated with the P atoms having thermal factors $B_P = 15\,Å^2$.

Wavelength λ (Å)	NAR	RES (Å)	Error-free data	Noisy data	SNR_{All}	SNR_{250}
1.50	2233	1.70	15 (10/10)	63 (5/11)	0.73	2.20
			15 (10/10)	76 (1/16)		
1.80	2233	1.70	15 (10/10)	64 (3/8)	0.74	2.29
			32 (10/11)	68 (1/1)		
2.10	2233	1.70	14 (10/10)	59 (5/11)	0.77	2.42
			17 (10/10)	72 (1/1)		
2.40	2233	1.70	15 (10/10)	42 (8/14)	0.81	2.64
			17 (10/10)	69 (3/5)		
2.70	2233	1.70	14 (10/10)	42 (9/9)	0.84	2.90
			19 (10/10)	61 (2/3)		
3.00	2233	1.70	14 (10/10)	36 (9/10)	0.88	3.19
			19 (10/10)	59 (2/3)		
3.30	2233	1.70	14 (10/10)	30 (10/13)	0.91	3.54
			27 (10/10)	59 (3/6)		
3.50	1937	1.75	15 (10/10)	34 (9/15)	0.98	3.70
			15 (10/10)	59 (3/6)		
3.70	1709	1.85	18 (10/10)	39 (10/14)	1.08	3.77
			15 (10/10)	62 (3/10)		
4.00	1331	2.00	19 (10/10)	27 (10/11)	1.22	4.21
			37 (10/12)	62 (2/16)		
5.00	988	2.50	62 (5/11)	65 (4/19)	2.27	5.32
			58 (5/16)	64 (3/15)		
5.50	758	2.75	50 (6/9)	41 (9/18)	3.12	5.74
			60 (3/5)	61 (4/6)		

Table 5. *Results obtained using MULTAN with anomalous scattering data sets for the structure d*(CGCGXG)

As for Table 3, except that these data sets were simulated with the P atoms having thermal factors $B_P = 10\,Å^2$.

Wavelength λ (Å)	NAR	RES (Å)	Error-free data	Noisy data	SNR_{All}	SNR_{250}
1.50	2233	1.70	16 (10/10)	66 (2/4)	0.78	2.25
			19 (10/10)	75 (1/1)		
1.80	2233	1.70	16 (10/10)	52 (5/6)	0.81	2.39
			16 (10/10)	72 (1/1)		
2.10	2233	1.70	16 (10/10)	54 (5/15)	0.85	2.59
			17 (10/10)	64 (2/10)		
2.40	2233	1.70	16 (10/10)	37 (9/16)	0.93	2.98
			23 (10/10)	65 (3/11)		
2.70	2233	1.70	16 (10/10)	31 (10/13)	1.00	3.33
			43 (8/13)	57 (2/10)		
3.00	2233	1.70	16 (10/10)	25 (10/11)	1.07	3.77
			16 (10/10)	56 (4/18)		
3.30	2233	1.70	17 (10/10)	25 (10/12)	1.15	4.20
			19 (10/10)	54 (4/14)		
3.50	1937	1.75	18 (10/10)	23 (10/10)	1.19	4.42
			26 (10/10)	56 (6/15)		
3.70	1709	1.85	20 (10/10)	33 (10/16)	1.29	4.55
			20 (10/10)	48 (6/8)		
4.00	1331	2.00	23 (10/10)	22 (10/10)	1.54	5.07
			53 (8/11)	44 (9/18)		
5.00	988	2.50	61 (5/12)	65 (4/16)	2.86	6.01
			62 (4/9)	63 (5/18)		
5.50	758	2.75	51 (5/9)	65 (4/6)	3.64	6.43
			62 (4/9)	63 (4/10)		

demonstrate the extreme sensitivity of the method to the accuracy of the anomalous-scattering differences. However, this is only to be expected, given the relatively small size of the anomalous-scattering signal, in particular at the shorter wavelengths (see Table 2). With

Table 6. *List of the* 12 *highest peak positions in the E map calculated from the best set of refined phases, having a mean phase error of* 19°, *after refinement of multiple sets of random phases using the error-free data simulated at* $\lambda = 1.5$ Å *with* $B_P = 10$ Å^2

The true coordinates of the P-atom positions (P1–P10), which were matched with the peak positions, are given below the peak coordinates. The corresponding distances between the true P-atom positions and the peak positions are also shown.

Peak No.	Peak height	Fractional coordinates x/a	y/b	z/c	Distance (Å)
1	1907	−0.175	0.270	−0.481	
	(P1)	−0.183	−0.729	0.522	0.214
2	1883	−0.676	0.270	−0.017	
	(P2)	0.328	0.270	−0.016	0.084
3	1738	−0.300	0.315	−0.342	
	(P3)	−0.305	−0.686	0.657	0.087
4	1357	−0.830	0.401	0.356	
	(P4)	0.165	−0.597	0.354	0.126
5	1226	−0.517	0.423	−0.187	
	(P5)	0.482	0.425	0.813	0.061
6	1162	−0.984	0.414	0.489	
	(P6)	0.008	−0.590	0.490	0.200
7	1143	−0.157	0.313	−0.119	
	(P7)	0.849	0.308	0.879	0.220
8	1093	−0.944	0.494	−0.296	
	(P8)	0.054	0.491	0.703	0.119
9	994	−0.885	0.394	−0.175	
	(P9)	0.120	0.401	0.826	0.226
10	900	−0.420	0.346	0.050	
	(P10)	0.592	0.345	0.052	0.241
11	715	−0.171	0.387	0.193	
	—		Spurious peak		—
12	604	−0.480	0.489	−0.225	
	—		Spurious peak		—

Table 7. *List of the* 12 *highest peak positions in the E map calculated from the best set of refined phases, having a mean phase error of* 75°, *after refinement of multiple sets of random phases using the noisy data simulated at* $\lambda = 1.5$ Å *with* $B_P = 10$ Å^2

The true coordinates of the P-atom positions (P1–P10), which were matched with the peak positions, are given below the peak coordinates. The corresponding distances between the true P-atom positions and the peak positions are also shown.

Peak No.	Peak height	Fractional coordinates x/a	y/b	z/c	Distance (Å)
1	2830	0.603	0.165	0.062	
	(P1)	0.592	−0.845	0.068	0.568
2	765	0.134	0.168	0.206	
	—		Spurious peak		—
3	733	0.594	0.201	0.562	—
	—		Spurious peak		—
4	691	0.607	0.234	0.328	
	—		Spurious peak		—
5	678	0.550	0.090	0.070	
	—		Spurious peak		—
6	651	0.537	0.058	0.185	
	—		Spurious peak		—
7	650	0.840	0.042	0.806	
	—		Spurious peak		—
8	648	0.905	0.068	0.467	
	—		Spurious peak		—
9	636	0.893	0.202	0.502	
	—		Spurious peak		—
10	635	0.599	0.208	0.814	
	—		Spurious peak		—
11	619	0.536	0.175	0.613	
	—		Spurious peak		—
12	608	0.390	0.164	0.086	
	—		Spurious peak		—

the noisy data sets, even refinement of the true phases generally fails to provide good solutions. In several cases, especially when $B_P = 30$ Å^2, the true phases refine to give mean phase errors that are greater than 60°, and the resulting E maps revealed only a few of the P-atom positions. The true phases tend to refine better at the longer wavelengths, although there are statistical fluctuations in this general trend. The explanation for this is that the anomalous-scattering signal is stronger at the longer wavelengths, giving higher average signal-to-noise ratios, which means that the phase refinements at these wavelengths are less susceptible to noise in the data.

When multiple sets of initially random phases are refined using the noisy data, the best phase sets usually give partial solutions, with only one or a few P-atom positions being located in the resulting E maps (for example, see Table 7). Again the better solutions tend to be found at the longer wavelengths where the anomalous-scattering signal is greater.

The effect of varying the B factors

Comparison of the results in Tables 3, 4 and 5 illustrates the effect of reducing the values of the thermal parameters of the P atoms in the structure. For the error-free data sets, the results in the lower B-factor cases show a general improvement on those in the higher B-factor cases. When the true phases were refined, the final mean phase errors were, in general, slightly smaller for the lower B-factor data sets. Further, when multiple sets of random phases were refined, the smallest mean phase errors for the lower B-factor data sets corresponded to good solutions and were mostly smaller than for the higher B-factor data sets.

The results for the noisy data sets simulated with lower B factors were generally superior to those for the corresponding noisy data sets simulated with higher B factors. The phase refinements of both the true phases and of multiple sets of random phases were more effective in the lower B-factor cases. However, the accuracy of the anomalous-scattering differences was still the most critical factor affecting the success of the method in these cases.

Selecting the best differences

The three procedures for phase refinement using noisy data, described above, were tested using the data sets simulated with $B_P = 10$ Å^2 over a range of wavelengths. A comparison of the results obtained using each of the three procedures is given in Table 8. The first two procedures give generally similar results after refinement of both the true and the random phases. However, the

Table 8. *Comparison of three different procedures for determining* P*-atom positions using noisy data sets*

As for the noisy data sets in Table 5 with $B_P = 10\,Å^2$, except that each of the following three procedures was tested. Procedure 1, tangent phase refinement of the 250 reflections having the largest values of $|\Delta F|/\sigma(|\Delta F|)$. Procedure 2: tangent phase refinement of the 250 reflections having the largest values of $|\Delta E|/\sigma(|\Delta F|)$. Procedure 3: tangent phase refinement of the 250 reflections having the largest values of $|\Delta E|$, but excluding any reflections with $F < 4\sigma(F)$ or with $|\Delta E| > 5|\Delta E|_{r.m.s.}$.

Wavelength		RES	Noisy data		
λ (Å)	NAR	(Å)	Procedure 1	Procedure 2	Procedure 3
1.50	2233	1.70	66 (2/4)	52 (6/17)	61 (4/17)
			75 (1/1)	65 (1/1)	74 (1/1)
1.80	2233	1.70	52 (5/6)	66 (2/4)	70 (2/3)
			72 (1/1)	68 (1/1)	71 (1/1)
2.10	2233	1.70	54 (5/15)	47 (6/8)	63 (2/5)
			64 (2/10)	64 (2/15)	67 (1/1)
2.40	2233	1.70	37 (9/16)	39 (8/9)	68 (1/1)
			65 (3/11)	57 (2/5)	66 (2/12)
2.70	2233	1.70	31 (10/13)	31 (10/18)	41 (9/18)
			57 (2/10)	57 (3/4)	65 (2/12)
3.00	2233	1.70	25 (10/11)	24 (10/10)	40 (9/11)
			56 (4/18)	56 (3/3)	61 (3/4)
3.30	2233	1.70	25 (10/12)	23 (10/10)	26 (10/11)
			54 (4/14)	55 (4/11)	55 (6/17)
3.50	1937	1.75	23 (10/10)	26 (10/10)	41 (7/14)
			56 (6/15)	55 (4/10)	63 (2/2)
3.70	1709	1.85	33 (10/16)	40 (9/15)	30 (9/14)
			48 (6/8)	41 (8/10)	56 (5/19)
4.00	1331	2.00	22 (10/10)	23 (10/10)	28 (10/15)
			44 (9/18)	50 (6/10)	62 (1/1)
5.00	988	2.50	65 (4/16)	65 (4/12)	64 (4/19)
			63 (5/18)	64 (2/4)	64 (3/18)
5.50	758	2.75	65 (4/6)	52 (7/11)	65 (4/6)
			63 (4/10)	61 (4/19)	63 (4/10)

refined sets of phases produced using the third procedure are in several cases markedly worse than, and at best similar to, the corresponding sets of phases obtained using the first two procedures. Therefore, it would seem that refining the phases associated with the largest values of $|\Delta F|/\sigma(|\Delta F|)$ or $|\Delta E|/\sigma(|\Delta F|)$ is more effective than refining the phases associated with the largest values of $|\Delta E|$ as in procedure 3.

Varying the number of differences used

For the noisy data sets, after refinement of initially random phases associated with the 250 largest values of $|\Delta F|/\sigma(|\Delta F|)$, the best results shown previously were those obtained using the data simulated at the longer wavelengths with $B_P = 10\,Å^2$. Therefore, we considered the effect of varying the number of large $|\Delta F|/\sigma(|\Delta F|)$ values used in the phase refinement for the noisy data sets with $B_P = 10\,Å^2$ at wavelengths of 3.7, 4.0, 5.0 and 5.5 Å. The results are shown in Table 9. Allowing for statistical variations, the phase refinement is most effective when using between 150 and 250 of the largest $|\Delta F|/\sigma(|\Delta F|)$ values. If we further increase the number of phases included in the refinement, we start to include reflections with smaller values of $|\Delta F|/\sigma(|\Delta F|)$, due to the limited resolutions of the data at these longer

Table 9. *The effect of varying the number of largest* $|\Delta F|/\sigma(|\Delta F|)$ *values used in the phase refinement*

The noisy data sets simulated at the wavelengths $\lambda = 3.7$, 4.0, 5.0 and 5.5 Å with $B_P = 10\,Å^2$ were tested as in Table 5, but using different numbers (NREF) of the largest $|\Delta F|/\sigma(|\Delta F|)$ values. The number of pairs of acentric reflections (NAR) is also given for each data set.

	Noisy data Wavelength λ (Å)			
	3.7	4.0	5.0	5.5
	NAR			
NREF	1709	1331	988	758
150	27 (10/16)	30 (10/18)	43 (7/10)	57 (5/16)
	51 (7/7)	60 (4/19)	51 (4/6)	57 (6/20)
200	33 (9/15)	27 (10/15)	61 (4/16)	64 (4/20)
	49 (6/20)	55 (4/9)	55 (5/9)	64 (4/17)
250	33 (10/16)	22 (10/10)	65 (4/16)	65 (4/6)
	48 (6/8)	44 (9/18)	63 (5/18)	63 (4/10)
300	33 (10/11)	20 (10/10)	65 (5/18)	41 (9/15)
	55 (6/8)	63 (3/16)	61 (5/15)	65 (4/11)
350	31 (10/10)	21 (10/10)	67 (3/12)	33 (10/12)
	60 (5/9)	63 (5/14)	66 (3/11)	65 (4/11)
400	29 (10/10)	24 (10/10)	67 (3/16)	37 (10/12)
	62 (5/14)	61 (4/13)	66 (3/13)	66 (4/9)

wavelengths. Although we thus increase the number of three-phase relationships employed in the tangent formula, the additional relationships are in general not so reliable and so the phase refinement is less effective. In Table 9, the best solution found after refining random phases had a mean phase error of 44° and was obtained using the 250 largest values of $|\Delta F|/\sigma(|\Delta F|)$ in the noisy data set simulated at $\lambda = 4.0$ Å.

Efficacy of the MULTAN figures of merit

An important consideration is whether the phase sets having the lowest mean phase errors were also indicated as good solutions by the conventional figures of merit employed in *MULTAN*. ABSFOM measures the extent to which the triplet phase relationships hold: it is zero for random phases and unity for correct phases. RESID measures the discrepancy between the actual and the estimated values of the phases: it should be small for correct phases. The expected value of PSIZERO is unity for correct phases. The combined figure of merit CFOM, when calculated with the default weighting scheme, takes values in the range 0–3, with the best phase sets having the highest values. TFOM is a measure of the difference between the cosines of the structure invariants and their theoretical expectation values; therefore, low values of TFOM indicate good phase sets.

An examination of our results, which have been presented in detail by Hubbard (1994), showed that these figures of merit were not always effective in determining the sets of refined phases corresponding to the best solutions. For example, with error-free data at a wavelength of 1.5 Å, the best phase set (*i.e.* that with the lowest mean phase error) was indicated by the maximum CFOM and the minimum RESID and TFOM, but the set with the maximum ABSFOM was not a

solution. At a wavelength of 2.7 Å the best set had only the sixth highest CFOM; the set with the maximum CFOM and the minimum TFOM had a low mean phase error, but it was not the best; and the set with the minimum RESID and the maximum ABSFOM had essentially random phases. When the wavelength was increased further to 4.0 Å the best set was not indicated by any figure of merit, and the sets with the best figures of merit had random phases. The results were worse when noise was added to the data: at the short wavelength the best set was indicated by the maximum ABSFOM but by no other figures of merit; at the intermediate and long wavelength the best set was not indicated by any figure of merit, and the sets with the optimum figures of merit were not even partial solutions of the structure. PSIZERO was unsuccessful at indicating good phase sets in all cases. The conclusion to be reached from these results is that the *MULTAN* figures of merit are only poorly discriminating for phase sets generated from limited resolution data. This corroborates the work of Mukherjee & Woolfson (1993), who applied the direct-methods program *SAYTAN* to data sets at various restricted resolutions for the protein structure aPP, and who found that the conventional figures of merit were not very effective in selecting the best phase sets. Therefore, we decided to explore ways in which the figures of merit used in *MULTAN* could be made more effective in selecting the best refined phase sets associated with anomalous-scattering partial structure.

First, since PSIZERO had proved to be particularly unreliable, we considered the effect of excluding it from the calculation of CFOM. The refinement of multiple sets of random phases was repeated for the error-free data sets simulated at wavelengths of 1.5 and 2.7 Å with $B_P = 10\,\text{Å}^2$, and CFOM was calculated only from the values of ABSFOM and RESID, both with weights of 1.50. However this modified weighting scheme gave no improvement in the discrimination of the best phase sets using CFOM.

PSIZERO is likely to be unreliable in selecting good phase sets with anomalous data since small values of $|\Delta E|$ are involved in its calculation, in addition to the large $|\Delta E|$ values employed in the phase determination and, even in the case of error-free data, small anomalous differences do not necessarily correspond to small values of $|F_A''|$. Therefore, we considered trying to improve PSIZERO by selecting the 50 smallest values of $|\Delta E|$, used in its calculation, from amongst only the weak reflections. This was done by, first, finding the 100 smallest values of $|F^+| + |F^-|$, since it is more probable that the values of $|F_A''|$ are small for weak reflections. Then, from these differences, we selected the 50 which had the smallest values of $|\Delta F|$. With the error-free data set simulated at 1.5 Å wavelength, the mean value of $|F_A''|$ for the 50 smallest $|\Delta F|$ values was 0.911, whereas for the 50 small $|\Delta F|$ values selected from amongst the weakest reflections, the mean value of $|F_A''|$ was 0.835. At $\lambda = 2.7$ Å, the mean value of $|F_A''|$ for the 50 smallest $|\Delta F|$ values was 2.831, whereas for the 50 small $|\Delta F|$ values selected from the weak reflections, the mean value of $|F_A''|$ was 2.608. These $|\Delta F|$ values were all set equal to 0.001, in order to ensure that *MULTAN* used them to calculate PSIZERO. The refinement of both the true phases and multiple sets of initially random phases was then repeated using the low B factor, error-free data at the wavelengths of 1.5 and 2.7 Å.

When the true phases were refined against the error-free data at a wavelength of 1.5 Å PSIZERO was 1.511 if the new selection procedure was used compared to the value of 2.000 attained earlier. After refining initially random phases the best solution with a mean phase error of 19°, which previously had a PSIZERO of 2.031 and the maximum CFOM of 2.215, now had a PSIZERO of 1.549 and still had the maximum CFOM of 2.460. Other good solutions found in this case similarly had lower values of PSIZERO and so higher values of CFOM than observed previously. Therefore, the effectiveness of PSIZERO seemed to have been improved by the new selection procedure. However, it was still not a reliable figure of merit when used on its own, since some poor phase sets had low values of PSIZERO. For example, one refined set of phases was obtained having a PSIZERO of 1.387 and the ninth highest CFOM of 2.023, but the mean phase error for this set was 83°.

A similar pattern was found using the error-free data at $\lambda = 2.7$ Å. Refining the true phases gave a set of phases having a PSIZERO of 1.526, which was less than the value of 1.792 obtained previously. After refining initially random phases it was again found that the best solutions had lower values of PSIZERO and so higher values of CFOM than those obtained previously. Therefore, this selection procedure for the small $|\Delta E|$ values again provided an improvement in the figures of merit.

Finally, we considered the use of alternative figures of merit for selecting the best phase sets generated from lower resolution data. The modified figures of merit proposed by Mukherjee & Woolfson (1993) were applied to the refinement of multiple sets of initially random phases using the low B factor, error-free data set simulated at $\lambda = 4.0$ Å at a resolution of 2.0 Å. The *MULTAN* figures of merit – ABSFOM, PSIZERO and RESID – were replaced by the modified figures of merit ABSM, PSIM and RESM, respectively. A combined figure of merit, CFOM, which was a weighted sum of the other three figures of merit, was also calculated. The weights associated with ABSM, PSIM and RESM in the calculation of CFOM were 0.60, 1.20 and 1.20, respectively.

For the error-free data at $\lambda = 4.0$ Å, after again refining multiple sets of random phases, the best solution having a mean phase error of 53° was not distinguished by the modified figures of merit. The value of CFOM (2.227) for this set of phases was relatively small and all the sets of phases with the highest values of CFOM had

mean phase errors greater than 70°. Therefore, the application of the modified figures of merit suggested by Mukherjee & Woolfson (1993) did not give any improved discrimination of the best phase sets in this case.

Discussion

The results presented above clearly show that *MULTAN* is successful in locating the positions of the anomalous scatterers when the data are error-free, but it is much less effective when relatively small amounts of random noise are added to the intensities. The reason for this sensitivity to errors in the data can be illustrated by plotting graphs of the largest values of $|\Delta F|$ *versus* the corresponding values of $|F_A''|$ due to the P atoms, which are calculated using,

$$|F_A''| = \left| \Delta f'' \exp(-B \sin^2\theta/\lambda^2) \times \sum_{j=1}^{N} \exp i2\pi(hx_j + ky_j + lz_j) \right|, \quad (5)$$

where the summation is over all the $N = 40$ P atoms in the unit cell. Such graphs were plotted for both the error-free and noisy data sets simulated at the wavelengths of 1.5, 2.7 and 4.0 Å, with $B_P = 10.0$ Å^2, showing the 250 largest values of $|\Delta F|$ in each case (Fig. 3).

For the error-free data sets at each of the three wavelengths, it is apparent that the very largest values of $|\Delta F|$ correspond to large values of $|F_A''|$, which are the required Fourier coefficients. However, because these two quantities are related by a trigonometric factor [(2)], the very largest values of $|F_A''|$ do not necessarily correspond to the largest values of $|\Delta F|$. Nevertheless, since the 250 largest values of $|\Delta F|$ include no very small values of $|F_A''|$, these $|\Delta F|$ values can be successfully used in *MULTAN* to find a solution by refining multiple sets of random phases.

For the noisy data sets at each of the three wavelengths, there is no longer any correlation between the largest values of $|\Delta F|$ and the large values of $|F_A''|$. In each of these cases, the 250 largest values of $|\Delta F|$ include both large and small values of $|F_A''|$, so that the direct-methods procedure is less effective in refining the phases associated with the phosphorus substructure.

Concluding remarks

We have shown that, for error-free data sets, good solutions can usually be found starting from initially random phases, with the resulting E maps revealing most if not all of the P atoms in the structure. However, the success of this method is very sensitive to relatively small errors in the values of $|\Delta F|$, so that for the noisy data sets the phase refinement is much less effective, particularly at the shorter wavelengths. We have also established that of the three procedures for phase refinement in the tangent formula using noisy data, procedures 1 and 2 are more effective than procedure 3. The phase refinement was found to be most effective when using between 150 and 250 of the largest $|\Delta F|/\sigma(|\Delta F|)$ values. The lower B-factor data sets generally give better results than do those simulated with higher B factors, although the accuracy of the $|\Delta F|$ values is still the most critical factor affecting the success of the phase refinement in these cases. The figures of merit used in *MULTAN* have difficulty in identifying the best phase sets, although the procedure we have devised for selecting the reflections to be used in the calculation of PSIZERO does result in a small improvement.

The successful determination of the structure of the protein crambin was based upon the measurement of a very small anomalous-scattering signal due to the S atoms, which allowed their positions to be found. This work provided encouraging evidence that a similar approach could be successfully applied to solve DNA crystal structures. The sensitivity of our method to the presence of small errors in the $|\Delta F|$ values makes the solution of the structure of crambin appear all the more remarkable. However, it was pointed out by Hendrickson *et al.* (1985) that, by employing suitable data-collection strategies, the values of $|\Delta F|$ can often be measured more accurately than those of $|F|$. In order to achieve this, firstly the $|\Delta F|$ values should be measured from the same crystal, so that major scaling errors can be eliminated. Secondly, if the pairs of Friedel mates are measured close together in time, then errors due to radiation damage or variations in the beam intensity are minimized. Finally, one should measure the Friedel pairs such that the paths traced by the diffracting rays through the crystal are similar for both of the reflections. In this case the absorption corrections for the two measurements will be nearly equivalent for favourable crystal morphologies. Numerical techniques can be applied to remove many of the residual experimental errors in the anomalous-scattering difference measurements. Systematic errors due to uncorrected absorption differences can be effectively minimized by using local scaling factors. Such procedures were very important in the case of crambin, where factors in the range 0.97–1.01 removed errors which otherwise completely masked the anomalous-scattering signal.

The method which we have described here is not dependent upon having high-resolution data, in contrast to the case in which attempts were made to solve oligonucleotide crystals from the $|F|$ values using direct methods. Since many oligonucleotide crystals have been solved from rather limited resolution (> 2.0 Å) data sets, in particular the B-form crystals, the anomalous-scattering method could be quite generally applicable to solving these structures. In the case of crambin, it was seen that

anomalous-scattering techniques could be successful even at the Cu $K\alpha$ wavelength. The anomalous signal does however increase markedly as the phosphorus absorption edge is approached, and we have shown that the most favourable wavelength at which to record the data is between 3.0 and 4.0 Å if the anomalous scatterers are to be located by direct methods. The availability of synchrotron-radiation facilitates collection of data at such wavelengths, and, as stated previously, the absorption might be reduced by using very small crystals.

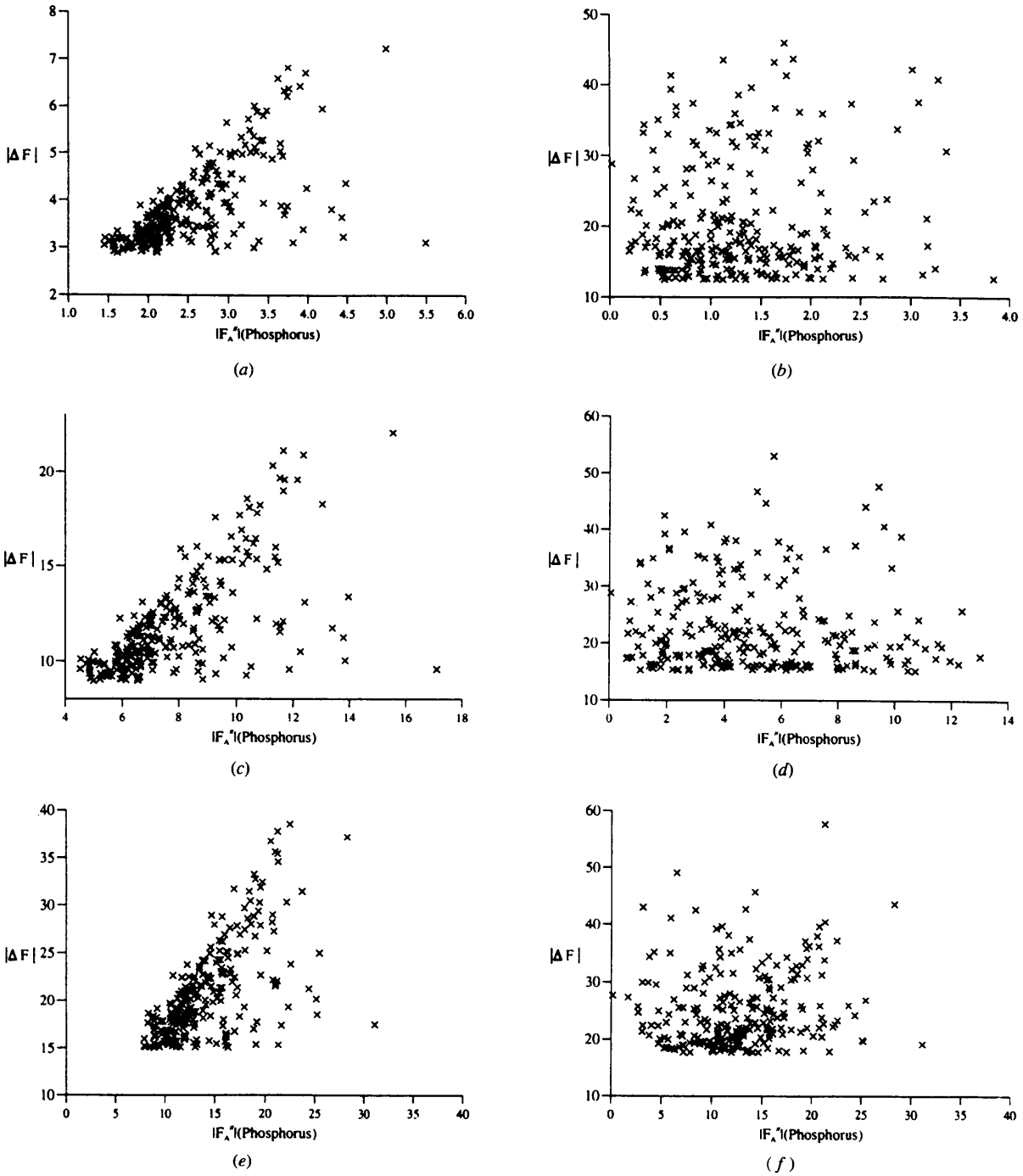

Fig. 3. Graphs of the 250 largest values of $|\Delta F|$ *versus* the corresponding values of $|F_A''|$ due to the P atoms in the structure d(CGCG*X*G), calculated for both the error-free and the noisy data sets with B_P values of 10.0 Å^2. Graphs on the left are for error-free data, and those on the right are for noisy data. (*a*) and (*b*) $\lambda = 1.5$ Å; (*c*) and (*d*) $\lambda = 2.7$ Å; (*e*) and (*f*) $\lambda = 4.0$ Å.

We wish to express our gratitude to Drs Jia-Xing Yao, L. Refaat, C. Tate & Hao Quan for their help and advice. We are also grateful to the Science and Engineering Research Council who provided a research studentship for SRH.

References

CROMER, D. T. (1983). *J. Appl. Cryst.* **16**, 437.

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

FAN, H.-F., WOOLFSON, M. M. & YAO, J.-X. (1993). *Proc. R. Soc. London Ser. A*, **442**, 13–32.

HENDRICKSON, W. A. & TEETER, M. M. (1981). *Nature (London)*, **290**, 107–113.

HENDRICKSON, W. A., SMITH, J. L. & SHERIFF, S. (1985). *Methods Enzymol.* **115**, 41–55.

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

HUBBARD, S. R., GREENALL, R. J. & WOOLFSON, M. M. (1994). *Acta Cryst.* D**50**, 833–841.

LEHMANN, M. S., MÜLLER, H.-H. & STUHRMANN, H. B. (1993). *Acta Cryst.* D**49**, 308–310.

MAIN, P., DEBAERDEMAEKER, T., GERMAIN, G., REFAAT, L. S., TATE, C. & WOOLFSON, M. M. (1988). *MULTAN*88. *Computer Programs for the Automatic Solution of Crystal Structures from X-ray Diffraction Data*. Univ. of York, England, and Louvain, Belgium.

MUKHERJEE, A. K., HELLIWELL, J. R. & MAIN, P. (1989). *Acta Cryst.* A**45**, 715–718.

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

ROSSMANN, M. G. (1961). *Acta Cryst.* **14**, 383–388.

STUHRMANN, S., HÜTSCH, M., TRAME, C., THOMAS, J. & STUHRMANN, H. B. (1995). *J. Synchrotron Rad.* **2**, 83–86.

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

YAO, J.-X., ZHENG, C.-D. QIAN, J.-Z., HAN, F.-S., GU, Y.-X. & FAN, H.-F. (1985). *SAPI*85. *A Computer Program for Automatic Solution of Crystal Structures from X-ray Diffraction Data*. Institute of Physics, Chinese Academy of Sciences, Beijing, China.