New techniques of applying multi-wavelength anomalous scattering data

By Fan Hai-Fu[1], M. M. Woolfson[2] and Yao Jia-Xing[2]

[1]*Institute of Physics, Chinese Academy of Sciences, Beijing 100080, People's Republic of China*

[2]*Department of Physics, University of York, York YO1 5DD, U.K.*

Several different methods of using multi-wavelength anomalous scattering data are described and illustrated by application to the solution of the known protein structure, core streptavidin, for which data at three wavelengths were available. Three of the methods depend on the calculation of Patterson-like functions for which the Fourier coefficients involve combinations of the anomalous structure amplitudes from either two or three wavelengths. Each of these maps should show either vectors between anomalous scatterers or between anomalous scatterers and non-anomalous scatterers. While they do so when ideal data are used, with real data they give little information; it is concluded that these methods are far too sensitive to errors in the data and to the scaling of the data-sets to each other. Another Patterson-type function, the P_s function, which uses only single-wavelength data can be made more effective by combining the information from several wavelengths. Two analytical methods are described, called AGREE and ROTATE, both of which were very successfully applied to the core streptavidin data. They are both made more effective by preprocessing the data with a procedure called REVISE which brings a measure of mutual consistency to the data from different wavelengths. The best phases obtained from AGREE lead to a map with a conventional correlation coefficient of 0.549 and this should readily be interpreted in terms of a structural model.

1. Introduction

Most applications of anomalous scattering (AS) for structure solution have involved the use of one-wavelength data; obtaining AS data has, until relatively recently, not been straightforward and scaling together data collected at different wavelengths has been an additional hurdle for multi-wavelength work. We have in our laboratories developed several techniques for the use of one-wavelength data (Fan Hai-fu *et al.* 1984, 1990*a*, *b*; Hao Quan & Woolfson 1989; Ralph & Woolfson 1992) and these have been shown to give useful structural information for proteins with up to 100 amino-acid residues. Other methods of using one-wavelength data have also been published (see, for example, Karle 1985, 1989).

With the availability of tunable synchrotron sources the use of multi-wavelength data is increasingly a realistic option; a very good example of the application of such data is the solution of the structure of core streptavidin, a 125–127 residue protein, (Hendrickson *et al.* 1989) where data was used at three wavelengths. The analysis was based on an algebraic approach developed by Karle (1980) in which unknown

Proc. R. Soc. Lond. A (1993) **442**, 13–32

Printed in Great Britain

quantities of interest are connected by linear equations and solved by a least-squares process which also takes into account nonlinear constraints which link the unknown quantities.

We have investigated a number of different ways of getting phase information from multi-wavelength AS data, some based on the ideas inherent in the one-wavelength work we have done previously. These methods vary enormously in their effectiveness, although they are all, on the face of it, equally sound in a theoretical sense. This variation of performance suggests to us that it is sensible to continue to explore new techniques even though some existing ones give reasonable results. In order to compare the different schemes we put forward, with each other and with what has been available hitherto, we use the core streptavidin data previously used by Hendrickson *et al.* (1989). The data for this crystal are: space group $I222$, $a = 95.20$, $b = 105.63$ and $c = 47.41$ Å for one crystal used and $a = 95.27$, $b = 105.41$ and $c = 47.56$ Å for another. The anomalous scatterers for this structure are selenium atoms of which there are two in the asymmetric unit, 16 in the complete cell. Intensity data were measured at four wavelengths at SSRL, 1.1000, 0.9792, 0.9789 and 0.9000 Å and at three wavelengths at the Photon Factor, 0.9809, 0.9795 and 0.9000 Å. We had available only the Photon Factor data for which the real and anomalous components of the AS are:

$$\lambda_1 = 0.9000\ \text{Å}, \quad f_1' = -1.622, \quad f_1'' = 3.285,$$
$$\lambda_2 = 0.9795\ \text{Å}, \quad f_2' = -6.203, \quad f_2'' = 3.663,$$
$$\lambda_3 = 0.9809\ \text{Å}, \quad f_3' = -8.198, \quad f_3'' = 2.058.$$

Before describing detailed methods of using AS data we first give some general principles for selecting the most suitable wavelengths for making measurements which will enhance the chances of a successful structure determination.

2. Optimizing the wavelengths for multi-wavelength data

In figure 1 there is shown the basic diagram for illustrating the Bijvoet difference for one wavelength AS. The contribution of the non-anomalous scattering, including that from the anomalous scatterers, is $\boldsymbol{F}(\boldsymbol{h})$, the real and imaginary parts of the AS are $\boldsymbol{F}'(\boldsymbol{h})$ and $\boldsymbol{F}''(\boldsymbol{h})$ respectively and the structure factors of the Friedel pair are $\boldsymbol{F}^+(\boldsymbol{h})$ and $\boldsymbol{F}^-(\boldsymbol{h})$. The complex conjugate of $\boldsymbol{F}^-(\boldsymbol{h})$ is $\boldsymbol{F}^-(\boldsymbol{h})^*$. From the diagram we find

$$\boldsymbol{F}^+(\boldsymbol{h}) = \boldsymbol{F}(\boldsymbol{h}) + \boldsymbol{F}'(\boldsymbol{h}) + \boldsymbol{F}''(\boldsymbol{h}), \tag{1}$$
$$\boldsymbol{F}^-(\boldsymbol{h})^* = \boldsymbol{F}(\boldsymbol{h}) + \boldsymbol{F}'(\boldsymbol{h}) - \boldsymbol{F}''(\boldsymbol{h}). \tag{2}$$

There are available techniques for finding the positions of the anomalous scatterers using the magnitudes of the Bijvoet differences (Mukherjee *et al.* 1989), in which case both $\boldsymbol{F}'(\boldsymbol{h})$ and $\boldsymbol{F}''(\boldsymbol{h})$ are determinable in both magnitude and phase. Combining (1) and (2) it follows that

$$\boldsymbol{F}^+(\boldsymbol{h}) - \boldsymbol{F}^-(\boldsymbol{h})^* = 2\boldsymbol{F}''(\boldsymbol{h}). \tag{3}$$

The magnitudes of $\boldsymbol{F}^+(\boldsymbol{h})$ and $\boldsymbol{F}^-(\boldsymbol{h})^*$ are found from experiment but there are two possible ways to draw the triangle corresponding to (3). The two triangles are mirrored to each other by the vector $2\boldsymbol{F}''(\boldsymbol{h})$ and lead to two possible solutions for the structure factor, $\boldsymbol{F}(\boldsymbol{h})$. If the true one is as indicated in figure 1 then the false one is shown by the thick dashed line.

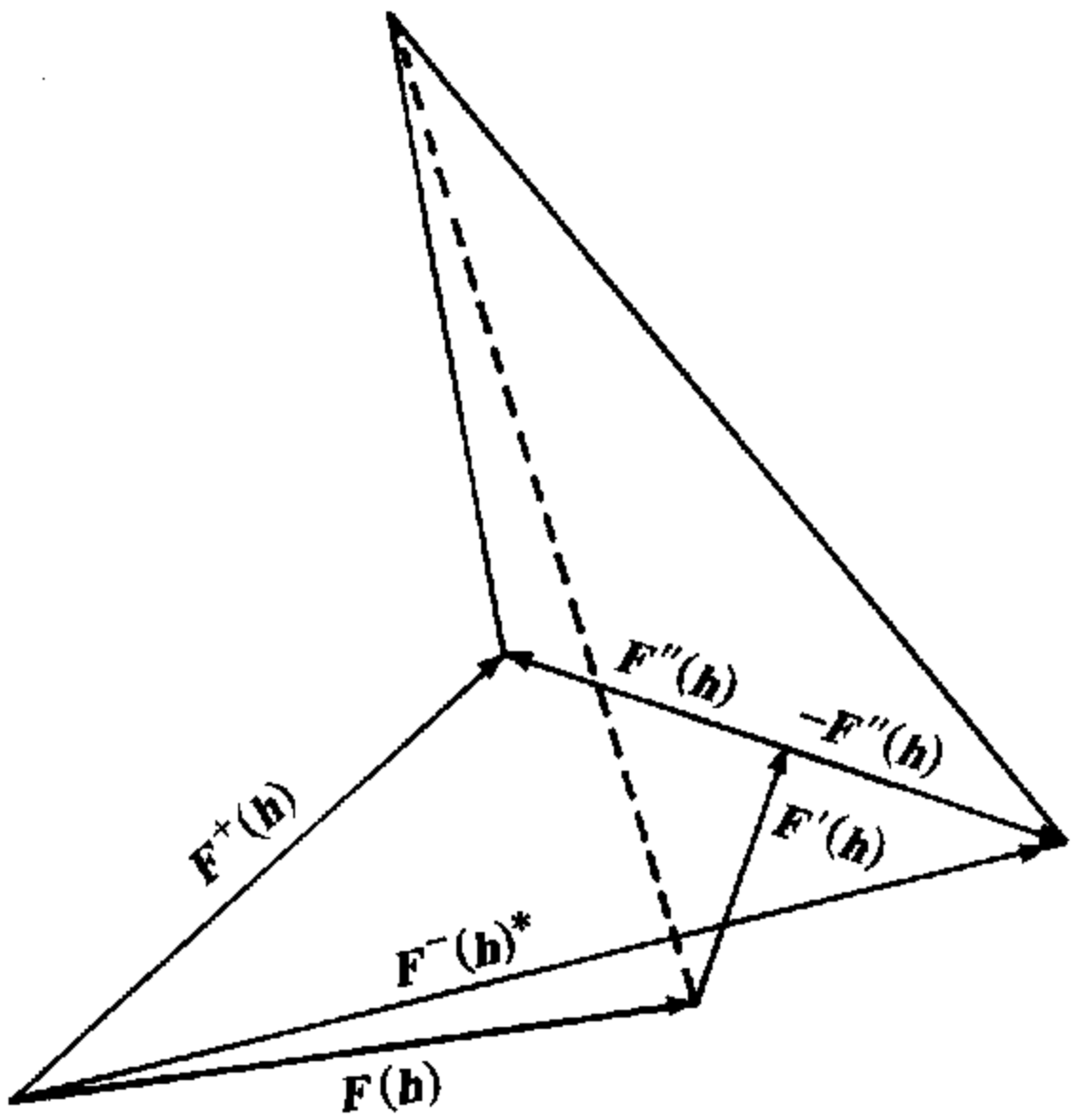

Figure 1. The phase ambiguity for one-wavelength anomalous scattering. $F'(\boldsymbol{h})$ and $F''(\boldsymbol{h})$ are the real and imaginary parts of the anomalous scattering and are assumed to be known. The anomalous structure amplitudes can be explained either by $F(\boldsymbol{h})$, the correct contribution of the non-anomalous scattering, or by a false contribution shown by the dashed line.

To break the phase ambiguity a second wavelength is introduced which will also lead to two possible solutions for $F(\boldsymbol{h})$. However, from the total of four possible solutions the two true solutions, one from each wavelength, will coincide and so be recognized. This coincidence of the correct solutions is shown in figure 2. It should be noticed that the two wavelengths should be chosen so that $F'_1(\boldsymbol{h}) \neq F'_2(\boldsymbol{h})$ otherwise $F''_2(\boldsymbol{h})$ will coincide with $F''_1(\boldsymbol{h})$ giving a coincidence of the two dashed lines thus leaving the ambiguity unresolved.

From a practical point of view a larger f'' will give more accurate phase doubles (figure 1) while a larger difference of f' will provide better discrimination in resolving the phase ambiguity (figure 2). When selecting wavelengths for collecting AS data these considerations should be borne in mind – especially if data is to be collected at only two wavelengths.

3. Effects of data inaccuracy

Bijvoet differences usually have magnitudes comparable with those of the experimental errors. For this reason a considerable number of the observed Bijvoet differences will be incorrect not only in their magnitudes but also in their signs, which can, in some circumstances have a very large effect on the estimate of the phase angle. This is illustrated in figure 3, where the effect of exchanging the values of $|F^+|$ and $|F^{-*}|$ is shown. Tests show that an important factor in getting good results from anomalous scattering data is to get the signs of the Bijvoet differences correct; most methods are fairly tolerant to errors in magnitude alone. We now illustrate this.

For the available data-sets at three wavelengths for core streptavidin there were 4217 reflections at 3.0 Å resolution. Table 1 shows the result of breaking the phase ambiguity for each data set separately by means of a direct-methods approach (Fan

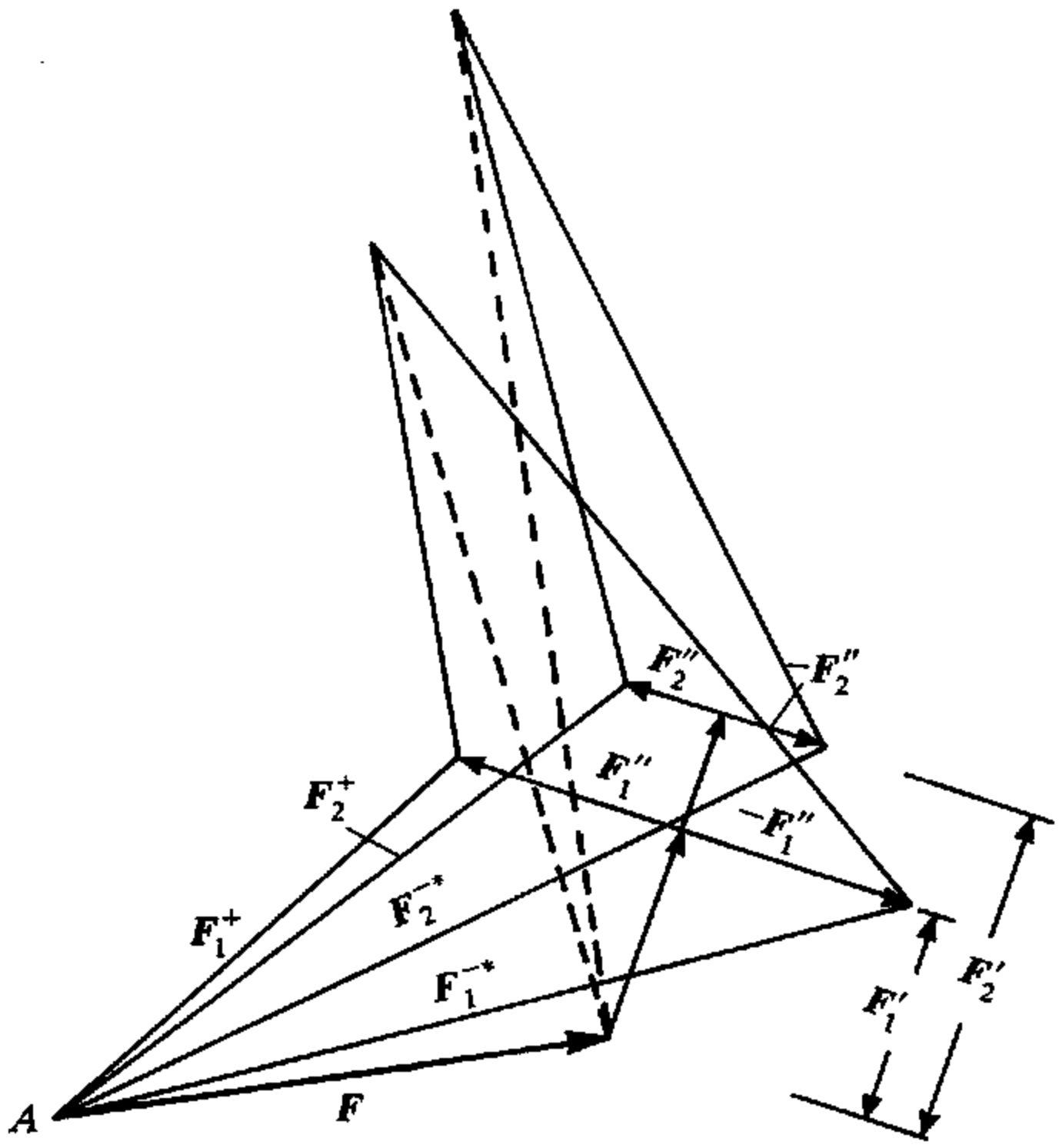

Figure 2. Resolving the ambiguity with two wavelengths. The point A is a common solution giving the same F for the two wavelengths.

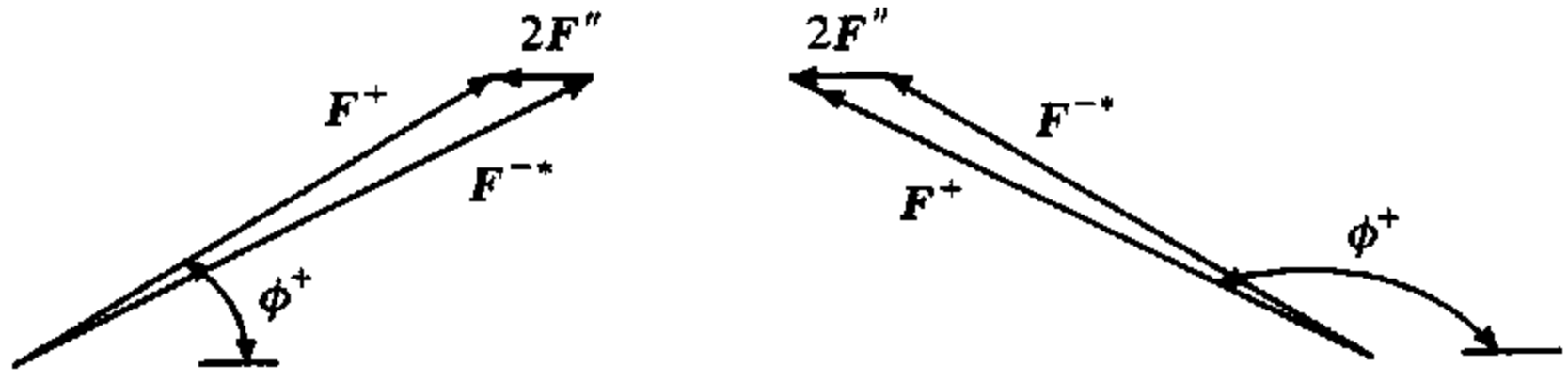

Figure 3. Exchange of the magnitudes of F^+ and F^{-*} results in a great change in the value of ϕ.

Table 1. *The effect of errors in the anomalous-scattering data for core streptavidin on the number of Bijvoet differences estimated with the wrong sign and the corresponding mean phase error with the direct method of resolving the phase ambiguity*

		number of reflections with incorrect signs of Bijvoet differences	weighted average phase error/deg	
λ/Å	f''		for all reflections	for the 1000 largest $\|F\|$s
0.9809	2.058	1360	71	63
0.9000	3.285	1280	68	59
0.9795	3.663	1145	64	53

Hai-fu *et al.* 1984). As will be seen the number of false signs of Bijvoet differences is related to f'' as is the accuracy of the estimated phase angles. A further test was done with the third data set ($\lambda = 0.9795$ Å) in which false signs were corrected for

Table 2. *The effect on the mean phase error using the direct method when the signs of the Bijvoet differences (BD) are corrected for BDs greater than the limit in the first column. Magnitudes are used as estimated*

absolute value of BDs beyond which the correct signs were used in the direct method	weighted average phase error/deg	
	for all reflections	for the 1000 largest $\|F\|$s
0	45	43
10	49	47
15	51	49
27[a]	55	51
∞	64	53

[a] The mean value of the observed magnitudes of the BSs.

Bijvoet differences with observed magnitudes above some limit but with the observed magnitude retained. The results are shown in table 2. If all the signs are corrected (limit of observed Bijvoet magnitude = 0) then the mean phase error of the estimated phases falls from 64° to 45°, a considerable improvement.

4. Refining observed magnitudes with multi-wavelength data

There is a way of looking for consistency of the Bijvoet differences if multiwavelength data is available and if there is only one kind of anomalous scatterer, which is often the case. From figure 4 we find

$$|\boldsymbol{F}^+|^2 = |\boldsymbol{F}|^2 + g^2 + 2|\boldsymbol{F}|\, g\cos(\theta+\delta), \tag{8}$$

$$|\boldsymbol{F}^-|^2 = |\boldsymbol{F}|^2 + g^2 + 2|\boldsymbol{F}|\, g\cos(\theta-\delta). \tag{9}$$

Subtracting

$$|\boldsymbol{F}^-|^2 - |\boldsymbol{F}^+|^2 = 4|\boldsymbol{F}|\, g\sin\theta\sin\delta. \tag{10}$$

The values of $|\boldsymbol{F}|$ and θ are independent of wavelength and

$$g\sin\delta = |\boldsymbol{F}''| \propto f''. \tag{11}$$

From this we find

$$C = (|\boldsymbol{F}^-|^2 - |\boldsymbol{F}^+|^2)/f'' \tag{12}$$

is independent of wavelength. The consistency of the determined values of C for different wavelengths gives an indication of the quality of the data. The result (12) was given by Karle (1984).

The data which is obtained from an AS experiment usually consists of the magnitudes of Friedel pairs, $|\boldsymbol{F}^+|$ and $|\boldsymbol{F}^-|$ together with standard deviations σ^+ and σ^-. We shall find quantities x^+ and x^-, regarded as corrections to the observed structure amplitudes, such that

$$C_i = \{(|\boldsymbol{F}_i^+| + x_i^+)^2 - (|\boldsymbol{F}_i^-| + x_i^-)^2\}/f'' \tag{13}$$

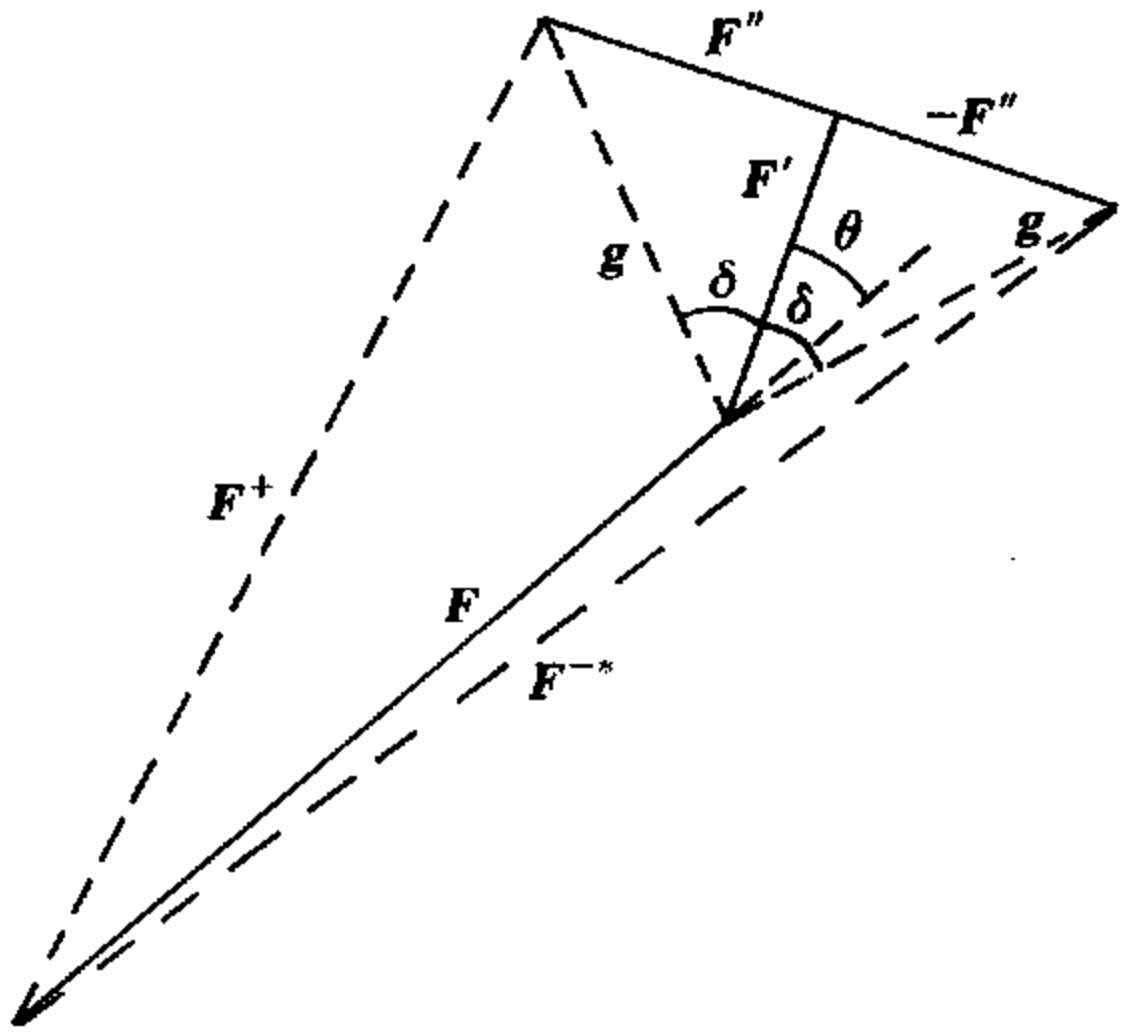

Figure 4. Diagram illustrating the relationship between F, F' and F'' which contributes to the difference of the magnitudes of F^+ and F^-. The quantity g is the magnitude of the anomalous contribution.

has the same value for each wavelength, λ_i, under the condition

$$G = \sum_{i=1}^{n} \left\{ \left(\frac{x_i^+}{\sigma_i^+} \right)^2 + \left(\frac{x_i^-}{\sigma_i^-} \right)^2 \right\} \tag{14}$$

is a minimum, where there are observations at n different wavelengths. This gives the maximum joint probability of the set of values of x under the condition that the values of C are all the same. If, for example, $n = 3$, then from the consistency condition,

$$\left.\begin{aligned} x_2^- &= [(|\boldsymbol{F}_2^+| + x_2^+)^2 - (f_2''/f_1'')\{(|\boldsymbol{F}_1^+| + x_1^+)^2 - (|\boldsymbol{F}_1^-| + x_1^-)^2\}]^{\frac{1}{2}} - |\boldsymbol{F}_2^-|, \\ x_3^- &= [(|\boldsymbol{F}_3^+| + x_3^+)^2 - (f_3''/f_1'')\{(|\boldsymbol{F}_1^+| + x_1^+)^2 - (|\boldsymbol{F}_1^-| + x_1^-)^2\}]^{\frac{1}{2}} - |\boldsymbol{F}_3^-|. \end{aligned}\right\} \tag{15}$$

Substituting these values in (14) makes the task that of minimizing $G(x_1^+, x_1^-, x_2^+, x_3^+)$ without constraint, for the values of x_2^- and x_3^- defined by (15) automatically satisfy the consistency conditions linking the three data-sets. Writing

$$X_1 = x_1^+, \quad X_2 = x_1^-, \quad X_3 = x_2^+ \quad \text{and} \quad X_4 = x_3^+$$

the minimum for G is found by iterative application of

$$\Delta \boldsymbol{X} = \boldsymbol{A}^{-1}\boldsymbol{b} \tag{16}$$

where the ith element of $\Delta \boldsymbol{X}$, ΔX_i, is the change in X_i, the ith element of $\boldsymbol{b}$ is $-\partial G/\partial X_i$ and $A_{ij} = \partial^2 G/\partial X_i \partial X_j$. In numerical applications of this refinement procedure the partial derivatives have been determined by finite-difference approximations.

In table 3 the effect of applying this process, which we call REVISE, to data at three wavelengths for two reflections is shown. For the first reflection, (8 6 2), the observed signs of the anomalous differences were all the same but this was not so for the second reflection (8 10 2). By changing the magnitudes by, at most, one standard deviation there are obtained modified magnitudes consistent with each other according to the values of C given by (12). Later it will be shown that in some phasing procedures better results are obtained using the modified magnitudes rather than the original ones.

Table 3. *The effect of applying the* REVISE *process to observed structure factors at three wavelengths for the structure core streptavidin*

(*a*) Reflection (8 6 2)

	$\lambda = 0.9000$ Å		$\lambda = 0.9795$ Å		$\lambda = 0.9809$ Å	
	h	$\bar{h}$	h	$\bar{h}$	h	$\bar{h}$
exp. $\|F\|$	917.1	958.3	949.5	989.9	942.6	976.3
σ	8.25	8.37	8.78	8.58	6.66	6.60
value of C	−23544		−38063		−17681	
modified $\|F\|$	918.7	956.6	958.3	981.3	938.9	980.2
revised C	−21653		−21653		−21653	

(*b*) Reflection (8 10 2)

exp. $\|F\|$	229.0	178.6	237.7	243.4	240.2	204.4
σ	20.06	24.48	19.29	17.71	14.84	18.03
value of C	−6254		758		−7742	
modified $\|F\|$	221.8	187.7	251.0	233.0	239.8	204.9
revised C	−4242		−4242		−4242	

5. A Patterson-like function involving two wavelengths

Just as there are Patterson-like functions involving one-wavelength data, the P_s and P_c functions of Okaya *et al.* (1955), so there are functions giving similar information with data at two or three wavelengths. Here we shall consider the two-wavelength case.

When anomalous scattering occurs

$$\boldsymbol{F}(\boldsymbol{h}) = \sum_{i=1}^{N} (f_i + f_i' + \mathrm{i}f_i'') \exp(2\pi\mathrm{i}\boldsymbol{h}\cdot\boldsymbol{r}_i). \tag{17}$$

Multiplying this by its complex conjugate, taking account that for each term involving (i,j) there is another involving (j,i), gives

$$|\boldsymbol{F}(\boldsymbol{h})|^2 = \sum_{i=1}^{N}\sum_{j=1}^{N} ((f_i + f_i')(f_j + f_j') + f_i''f_j'')\cos\{2\pi\mathrm{i}\boldsymbol{h}\cdot(\boldsymbol{r}_i - \boldsymbol{r}_j)\}$$
$$- \sum_{i=1}^{N}\sum_{j=1}^{N} ((f_j + f_j')f_i'' - (f_i + f_i')f_j'')\sin\{2\pi\mathrm{i}\boldsymbol{h}\cdot(\boldsymbol{r}_i - \boldsymbol{r}_j)\}. \tag{18}$$

We define

$$X_s(\boldsymbol{h}) = |\boldsymbol{F}(\boldsymbol{h})|^2 + |\boldsymbol{F}(\bar{\boldsymbol{h}})|^2 \tag{19}$$

for wavelength λ_s. From (18) we find

$$X_s(\boldsymbol{h}) = 2\sum_{i=1}^{N}\sum_{j=1}^{N} ((f_i + f_{i,s}')(f_j + f_{j,s}') + f_{i,s}''f_{j,s}'')\cos\{2\pi\mathrm{i}\boldsymbol{h}\cdot(\boldsymbol{r}_i - \boldsymbol{r}_j)\}. \tag{20}$$

For data at two wavelengths, λ_1 and λ_2, then

$$X_1(\boldsymbol{h}) - X_2(\boldsymbol{h}) = 2\sum_{i=1}^{N}\sum_{j=1}^{N} \eta_{i,j}\cos\{2\pi\mathrm{i}\boldsymbol{h}\cdot(\boldsymbol{r}_i - \boldsymbol{r}_j)\}, \tag{21}$$

where

$$\eta_{i,j} = ((f_i + f_{i,1}')(f_j + f_{j,1}') + f_{i,1}''f_{j,1}'')) - ((f_i + f_{i,2}')(f_j + f_{j,2}') + f_{i,2}''f_{j,2}'')). \tag{22}$$

Table 4. *The coordinates of the two independent selenium atoms and the independent Patterson peaks generated by the space group*

(*a*) Atomic coordinates

	x	y	z
Se1	0.3387	0.1129	−0.0944
Se2	0.2074	0.0510	0.2426

(*b*) Patterson peaks

	u	v	w
1	0.0000	0.2258	0.1888
2	0.6774	0.0000	0.1888
3	0.6774	0.2258	0.0000
4	0.0000	0.1020	0.4852
5	0.4148	0.0000	0.4852
6	0.4148	0.1020	0.0000
7	0.1313	0.0619	0.3370
8	0.1313	0.1639	0.1482
9	0.4539	0.1639	0.3370
10	0.4539	0.0619	0.1482

If a centrosymmetric Fourier synthesis, $P_{x_2}(\boldsymbol{u})$, is calculated with coefficients $X_1(\boldsymbol{h})-X_2(\boldsymbol{h})$ then from the form of (22) it is found that peaks will be present linking the anomalous scatterers (AS), assumed similar, with weights proportional to

$$(2f_{\mathrm{A}}+f'_{\mathrm{A},1}+f'_{\mathrm{A},2})(f'_{\mathrm{A},1}-f'_{\mathrm{A},2})+(f''_{\mathrm{A},1})^2-(f''_{\mathrm{A},2})^2$$

and between anomalous and non-anomalous scatterers (NAS) with weights proportional to

$$Z_{\mathrm{N}}(f'_{\mathrm{A},1}-f'_{\mathrm{A},2}),$$

where Z_{N} is the atomic number of the NAS.

In principle it should be possible to choose wavelengths so as to eliminate one set or other of these peaks. For example, with $f'_{\mathrm{A},1}=f'_{\mathrm{A},2}$ the only peaks would be between the anomalous scatterers but in practice this would not be simple to achieve. With the core streptavidin data for wavelengths λ_2 and λ_3 AS–AS peaks would have weights 116 while those between AS and NAS (assumed carbon) would have weight 12. It seems that it might be possible to pick out the vectors between the AS, which could be useful if other ways of doing so were unsuccessful.

To test the idea under ideal conditions data were calculated from the refined structure of core streptavidin for wavelengths λ_1, λ_2 and λ_3. These data were then used to calculate P_{x_2} maps for all three possible pairs of wavelengths and the peak-search routine of MULTAN was used to find the highest 100 peaks in each of the maps. In table 4 we show the coordinates of the two selenium atoms and the ten interatomic peaks which would be sought in the P_{x_2} map.

The best P_{x_2} maps were those for the wavelength pairs (λ_1, λ_2) and (λ_1, λ_3) which showed eight of the ten interatomic vectors in the top 100 peaks. In the case of (λ_1, λ_2) these peaks were, in ranking order excluding the origin peak, at positions 2, 5, 9, 21, 34, 46, 57 and 97 and they matched the true positions to within about 1 Å, which is all that can be expected in the presence of a large number of other peaks and the 3 Å resolution of the data.

Results with the observed data were much worse: the best map, for (λ_1, λ_3), showed four peaks at positions 4, 8, 22 and 64. We conclude that the maps are extremely sensitive to experimental error and that the P_{x_2} map has no practical applications with data of existing accuracy. Even with ideal data it might not be straightforward to pick out the AS–AS vectors.

6. Patterson-like functions involving three wavelengths

From (20) it can be shown that a centrosymmetric Fourier synthesis, P_{x_3}, with coefficients

$$X_1(\boldsymbol{h})\,(f'_{2,\mathrm{A}}-f'_{3,\mathrm{A}})+X_2(\boldsymbol{h})\,(f'_{3,\mathrm{A}}-f'_{1,\mathrm{A}})+X_3(\boldsymbol{h})\,(f'_{1,\mathrm{A}}-f'_{2,\mathrm{A}}) \tag{23}$$

will show only peaks between anomalous scatterers and these will be of weight proportional to

$$(f'^2_{1,\mathrm{A}}+f''^2_{1,\mathrm{A}})\,(f'_{2,\mathrm{A}}-f'_{3,\mathrm{A}})+(f'^2_{2,\mathrm{A}}+f''^2_{2,\mathrm{A}})\,(f'_{3,\mathrm{A}}-f'_{1,\mathrm{A}})+(f'^2_{3,\mathrm{A}}+f''^2_{3,\mathrm{A}})\,(f'_{1,\mathrm{A}}-f'_{2,\mathrm{A}}). \tag{24}$$

For core streptavidin this quantity, showing the contribution of the three terms in the summation (24) separately, is

$$26.8-341.3+327.3=12.8.$$

Although the form of expression (24) is bound to give some positive and some negative indications the near balance of contributions in this case does not bode well for detection of the peaks in the $P_{x'_3}$ map.

To see whether, in principle, AS–AS interatomic vectors could be found by this process we used the magnitudes of the calculated ideal data for core streptavidin for the wavelengths λ_1, λ_2 and λ_3 to produce a $P_{x'_3}$ map. In table 5 we show the top 20 peaks as given by the peak-search program in MULTAN; we see that after the origin peak the highest ten peaks correspond to correct selenium–selenium vectors and that all false peaks are much smaller than the true ones. Encouraged by this result we then went on to use the observed data and we show the results of this in table 6 where the top 30 peaks are shown. The result was extremely disappointing; in the top 30 peaks only two could be associated with Se–Se vectors.

We concluded that the near numerical balance of the three terms in (24) was responsible for the poor results since it will lead to great sensitivity of the relative magnitudes of the Fourier coefficients (23) to errors in the data. If this conclusion was correct then a better choice of three wavelengths might have yielded a better outcome. If the three contributions contributing to the sum in (24) are c_1, c_2 and c_3, then a sensible figure of merit to decide whether or not the data are likely to give good results is

$$S=\Big|\sum_{i=1}^{3} c_i\Big|^2\Big/\sum_{i=1}^{3}|c_i| = \Big|\sum_{i=1}^{3} c_i\Big|\times\Big|\sum_{i=1}^{3} c_i\Big|\Big/\sum_{i=1}^{3}|c_i|, \tag{25}$$

where, expressed as a product, the first term is proportional to the magnitude of the peaks in the map and the second term expresses the balance between the three terms. A high value of S is favourable. For the three wavelength data we had available $S=0.23$. To see what might be possible we calculated data from the refined structure for wavelength 1.1000 Å, which had been measured at SSRL but which was presumably of lower quality than those of the three wavelengths actually used in the

Table 5. *The positions of the 20 highest peaks in the $P_{x_3'}$ map with calculated data for wavelengths λ_1, λ_2 and λ_3 with the particular* Se–Se *peak from table 4 indicated where appropriate*

peak no.	peak height	x	y	z	Se–Se peak
1	6000	0.0000	0.0037	0.0000	—
2	995	0.0000	0.1017	0.5000	4
3	947	0.4155	0.0041	0.5000	5
4	680	0.1319	0.0630	0.3377	7
5	609	0.6775	0.0037	0.1882	2
6	576	0.1322	0.1644	0.1486	8
7	576	0.4155	0.1016	0.0000	6
8	574	0.0000	0.2255	0.1888	1
9	570	0.6772	0.2255	0.0000	3
10	535	0.4545	0.0623	0.1477	10
11	527	0.4543	0.1641	0.3376	9
12	170	0.4723	0.0786	0.3800	
13	168	0.1137	0.0754	0.1037	
14	166	0.1118	0.0450	0.1846	
15	154	0.4356	0.0454	0.3762	
16	151	0.1490	0.0445	0.1090	
17	150	0.4391	0.1503	0.1062	
18	148	0.4699	0.0452	0.2988	
19	147	0.0113	0.0210	0.0876	
20	145	0.1501	0.1451	0.3786	

structure solution. We then corrupted this data by multiplying each intensity independently by random numbers drawn from a gaussian distribution with a mean of unity and a standard deviation which was adjusted so that the number of Bijvoet differences with incorrect signs was about 1400. From the indications in table 1 this artificial data set was then comparable to the measured ones, although somewhat worse as would be expected from having smaller values of f' and f'' and hence smaller values of Bijvoet differences. We refer to this data-set as corresponding to wavelength λ_4. The values of S for various combinations of available data are:

$$S_{1,2,4} = 32.4, \quad S_{1,3,4} = 47.3 \quad \text{and} \quad S_{2,3,4} = 0.56.$$

A $P_{x_3'}$ map was calculated with data-sets λ_1, λ_2 and λ_4, which seemed to give a quite favourable situation, but the results were actually poorer than with λ_1, λ_2 and λ_3; there was just one peak corresponding to an AS–AS vector in the top 30 peaks.

Another Patterson-like function we have discovered, $P_{X_3''}$, is the centrosymmetric one with coefficients

$$X_1(\boldsymbol{h})\,(Y_2 - Y_3) + X_2(\boldsymbol{h})\,(Y_3 - Y_1) + X_3(\boldsymbol{h})\,(Y_1 - Y_2), \tag{26}$$

where $Y_i = (f_{\mathrm{A}} + f'_{\mathrm{A},i})^2 + f''^2_{\mathrm{A},i}$. An analysis based on (20) shows that this synthesis gives only peaks between anomalous scatterers and non-anomalous scatterers and that the weights of these peaks are proportional to the quantity given in (24). Hence the conditions for getting a result less sensitive to experimental error is the same for $P_{x_3''}$ as for $P_{x_3'}$.

The $P_{x_3''}$ map gives the same vectors as the P_s map (Okaya *et al.* 1955; Hao Quan & Woolfson 1989). This function is

$$P_s(\boldsymbol{u}) = \frac{1}{V}\sum_{\boldsymbol{h}} (|F(\bar{\boldsymbol{h}})| - |F(\boldsymbol{h})|) \sin(2\pi \boldsymbol{h}\cdot\boldsymbol{u}) \tag{27}$$

Table 6. *The positions of the 30 highest peaks in the $P_{x_3'}$ map with observed data for wavelengths λ_1, λ_2 and λ_3 with the particular* Se–Se *peak from table 4 indicated where appropriate*

peak no.	peak height	x	y	z	Se–Se peak
1	758	0.0000	0.0208	0.0000	—
2	331	0.3701	0.0044	0.5000	—
3	295	0.0000	0.0000	0.2083	—
4	273	0.0000	0.0483	0.2628	
5	254	0.4167	0.0035	0.5006	5
6	219	0.0313	0.2255	0.5000	—
7	215	0.0000	0.1573	0.4576	—
8	211	0.5000	0.1081	0.5000	—
9	208	0.2545	0.0040	0.1538	—
10	198	0.0000	0.0765	0.2213	—
11	197	0.0000	0.2062	0.4660	—
12	196	0.0000	0.2116	0.1771	1
13	195	0.0230	0.0225	0.1008	
14	195	0.2809	0.1932	0.0000	
15	194	0.0391	0.1625	0.5000	
16	192	0.0891	0.0336	0.5000	
17	191	0.3223	0.1787	0.1456	
18	186	0.5000	0.1834	0.5000	
19	182	0.3580	0.0037	0.0007	
20	180	0.0000	0.2202	0.0000	
21	177	0.5000	0.1953	0.0000	
22	177	0.5000	0.0679	0.3831	
23	176	0.0000	0.1880	0.0817	
24	176	0.0000	0.0742	0.5000	
25	176	0.0546	0.0507	0.0000	
26	175	0.5000	0.1656	0.1741	
27	175	0.4020	0.1355	0.5000	
28	175	0.1356	0.0043	0.3127	
29	174	0.1615	0.0863	0.5000	
30	174	0.0687	0.0038	0.1874	

and it shows positive peaks corresponding to vectors between anomalous scatterers and non-anomalous scatterers and negative peaks in the opposite directions. There is some cancellation of overlapping positive and negative peaks, and thus some loss of information. Since the $P_{x_3''}$ function gives all positive peaks it might be expected to be superior in its information content. The way in which the P_s function is used is described in the following section; in the light of what has previously been said about the poor performance of the $P_{x_3'}$ function it is enough just to state here that the $P_{x_3''}$ function, treated in a way similar to that used for the P_s function, gives virtually no useful information if observed data are used.

7. Combining one-wavelength maps

The exploration of the various two-wavelength and three-wavelength functions described above gave very disappointing results with real data although they are capable, in principle, of giving useful information as was shown by the use of ideal calculated data. The conclusion is that these particular methods are very sensitive to the combination of errors in the data plus the difficulty of properly bringing the data sets to the same scale. Nevertheless, Hao Quan & Woolfson (1989) showed that the

P_s function, which only uses one-wavelength data, is capable of giving useful information for proteins so it seems illogical that using the information in three sets of data should be less successful. We decided to apply the P_s function with each data set separately and then to see whether by combining the information we could get a result better than that from any one wavelength alone.

Hao & Woolfson found after some experimentation that the best way of using the information in a P_s map is

(i) calculate a sum-function based on the positions of the anomalous scatterers,
(ii) equate all negative density to zero,
(iii) Fourier transform the resulting map to get phase estimates,
(iv) compute a map with observed amplitudes and estimated phases.

Since the final result is a map, which one then tries to interpret, we use as a figure of merit for the outcome of the analysis the conventional map correlation coefficient (MCC) between the density in the final map, ρ', and the density obtained using phases calculated from the refined structure, ρ. This is given by

$$r = (\overline{\rho\rho'} - \bar{\rho}\,\overline{\rho'})/\sigma_\rho \sigma_{\rho'}, \tag{28}$$

where σ_ρ and $\sigma_{\rho'}$ are the standard deviations of the indicated density distributions. The MCCs for the wavelengths λ_1, λ_2 and λ_3 were 0.231, 0.208 and 0.305 respectively. According to the conventional wisdom of judging the quality of maps by the values of MCC even the highest of these values corresponds to a map which would be difficult to interpret.

We next took the maps obtained at stage (ii) for each of the three wavelengths separately, added the maps together, equated negative density to zero and then Fourier transformed the final map to give phase estimates. These phases were then combined with the average observed amplitudes to give a final map. The map correlation coefficient for this is 0.334, better than that given by any individual data-set. Finally we applied the process REVISE to the data-sets. Since the values of C, given by equation (12) are now all the same it is only necessary to calculate one P_s map since, with the revised data, the maps from the three separate wavelengths would just be scaled versions of each other. The map correlation coefficient, 0.342, is better than any other obtained and can be compared with 0.455 which comes from a P_s function with coefficients from calculated data.

Our conclusion is that it is possible to exploit the additional information in several data-sets, as compared with a single set, but for methods involving the calculation of maps this is best done by eliminating the need to scale the data-sets together, as when calculating individual Fourier coefficients incorporating components from each set of observations. However, for core streptavidin even the best available map would be difficult to interpret although it might offer glimpses of the structure here and there.

8. The AGREE method

From equations (8) and (9), by addition and subtraction, we find

$$|F^-|^2 - |F^+|^2 = 4|F|\,g \sin\theta \sin\delta \tag{10}$$

and

$$|F^-|^2 + |F^+|^2 = 2|F|^2 + 2g^2 + 4|F|\,g \cos\theta \cos\delta. \tag{29}$$

Eliminating θ from these equations gives

$$Pg^4 + Qg^2 + R = 0, \tag{30}$$

where $$P = 4\sin^2\delta,$$

$$Q = -4\{|F^-|^2+|F^+|^2-2|F|^2\}\sin^2\delta+|F|^2\sin^2 2\delta$$

and

$$R = |F^+|^4+|F^-|^4+2|F^+|^2|F^-|^2\cos 2\delta-4|F|^2\sin^2\delta[|F^-|^2+|F^+|^2-|F|^2].$$

It should be noted that $\tan\delta = f''/f'$ so that, since f' is negative and f'' positive, $\frac{1}{2}\pi \leqslant \delta \leqslant \pi$. While this is not as it is represented in figure 4 it makes no difference to the analysis that follows and the use of correctly signed values of f' and f'' gives the correct results.

If the value of $|F|$ were known then the ambiguity associated with anomalous scattering would be thrown on to the magnitude of g for, from (30)

$$g^2 = (-Q\pm(Q^2-4PR)^{\frac{1}{2}})/2P. \tag{31}$$

Since

$$g^2 = |F'|^2+|F''|^2,$$

$$F'(\boldsymbol{h}) = f'\sum_{j=1}^{M}\exp(2\pi i\boldsymbol{h}\cdot\boldsymbol{r}_j), \quad F''(\boldsymbol{h}) = f''\sum_{j=1}^{M}\exp(2\pi i\boldsymbol{h}\cdot\boldsymbol{r}_j),$$

where there are M anomalous scatterers at positions $\boldsymbol{r}_j$, then it follows that

$$m^2 = \frac{g^2}{(f')^2+(f'')^2} = \left|\sum_{j=1}^{M}\exp 2\pi i\boldsymbol{h}\cdot\boldsymbol{r}_j\right|^2 \tag{32}$$

is independent of wavelength.

It is clear that, for a wavelength λ_i, there is a maximum value of $|g_i|$,

$$|g_i|_{\max} = M\{(f'_i)^2+(f''_i)^2\}^{\frac{1}{2}}. \tag{33}$$

These maximum values, together with the geometrical constraints implied in figure 4 lead to the following set of inequalities, applicable for any wavelength, λ_i:

$$\left.\begin{aligned} |F| < |F_i^+|+|g_i|_{\max}, \quad |F| < |F_i^-|+|g_i|_{\max},\\ |F| > |F_i^+|-|g_i|_{\max}, \quad |F| > |F_i^-|-|g_i|_{\max}. \end{aligned}\right\} \tag{34}$$

Applying these constraints for all the available wavelengths gives a possible range of values of $|F|$, between $|F|_{\min}$ and $|F|_{\max}$.

The AGREE method explores the possible range of $|F|$ values, usually at 100 equi-interval points over the range, and compares the values of m_i^2 found for each wavelength. These are then checked for consistency by calculating

$$T = \left\{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(m_i^2-m_j^2)^2\right\}^{\frac{1}{2}}, \tag{35a}$$

or a scaled version

$$T' = T/\Sigma m_i^2, \tag{35b}$$

and the minimum value of T, or T', is taken to indicate the most probable values of m, hence g, and also $|F|$. From (10) and (29) it is then possible to find the value of θ, the angle between F and F', from

$$\tan\theta = \frac{(|F^-|^2-|F^+|^2)\cos\delta}{(|F^+|^2+|F^-|^2-2|F|^2-2g^2)\sin\delta}, \tag{36}$$

where the ambiguity in the value of θ is resolved by the signs of the numerator and divisor of the expression.

It is evident that the analysis leading to the values of g and $|F|$ has not required any information about the positions of the anomalous scatterers, but only their total number and type. If the estimated values of g are good enough then it should be fairly straightforward, by the use of a Patterson function with Fourier coefficients g^2 (or m^2) or with MULTAN with structure amplitudes g (or m), to find the positions of the anomalous scatterers. From these positions one can find F' in both magnitude and phase (ψ) and hence the phase of F from

$$\phi = \psi - \theta \quad \text{(see figure 4)}. \tag{37}$$

In table 7 there are shown the results of the AGREE process for two different reflections for core streptavidin. For reflection (8 2 2) there is a fairly sharp minimum of T' and the different values of m^2 agree quite well. On the other hand for the (9 7 2) reflection the minimum of T' is much more poorly defined and this is accompanied by poorer agreement of the values of m^2 at the minimum. Most situations fall between these two extremes; there are also occasional situations where two minima of almost equal depth occur but these are rare enough not to disturb the overall effectiveness of the process.

It might be expected that smaller values of T' would correspond to more reliable estimates of m^2 and this turns out to be true. In table 8 there are shown the values of the residual

$$R = \Sigma |m^2_{\text{est}} - m^2_{\text{true}}| / \Sigma m^2_{\text{true}}, \tag{38}$$

for batches of reflections, chosen according to their ranges of values of T', where m^2_{est} and m^2_{true} are the estimated and true values of m^2 and the sums are over all the reflections in the batch. It is evident that smaller values of T' give more reliable estimates of m although even the best agreement does not look impressive. However, the estimated values of m^2 are actually good enough to give useful information.

An initial step in most anomalous scattering approaches is to find the positions of the anomalous scatterers. If these are heavy atoms, mercury for example, then it is often straightforward to find them from a Patterson function. However, if the anomalous scatterers will not show themselves in this way then the anomalous differences

$$\Delta F = \| F^+| - |F^-| \|$$

are frequently used, either as magnitudes in a direct methods procedure (Mukherjee *et al.* 1989), or in the form ΔF^2, as the coefficients of a Patterson function.

If multi-wavelength data are available, the use of REVISE and AGREE offers an attractive approach. The values of $\overline{m^2}$ can be used as the coefficients of a Patterson map which should show the vectors between the anomalous scatterers or may be fed in as intensities to a direct-methods programme. In table 9 there are shown the top 20 peaks from a map with the 2000 most reliably indicated values of m^2, as indicated by T', used as coefficients. The peaks are found by an automatic peak-search routine from the MULTAN direct-methods program and so require no visual examination of a map. The ten interatomic peaks between the selenium atoms are present (table 4) and, given that the correct peaks are a related set, they can readily be recognized. The number of reflections used in the map is not very critical and maps of similar quality are produced with any number of terms between 1500 and 3000.

With the positions of the anomalous scatterers known, and with values of $|F|$ and

Table 7. *Values of* $|F|$, m_i^2, $\overline{m^2}$ *and* T' *for two reflections for values of* $|F|$ *in the vicinity of the minimum of* T' *together with the values of* $|F^+|$ *and* $|F^-|$ *for the three wavelengths*

(*a*) Reflection (8 2 2)

	$\lvert F^+\rvert$	$\lvert F^-\rvert$
λ_1	985.8	993.2
λ_2	1042.2	1049.5
λ_3	1021.7	1031.3

$\lvert F\rvert$	m_1^2	m_2^2	m_3^2	$\overline{m^2}$	T'
973.99	87.69	78.21	72.06	79.31	0.0813
974.37	83.71	77.42	71.03	77.39	0.0669
974.74	79.82	76.63	70.04	75.50	0.0539
975.12	76.02	75.84	69.06	73.64	0.0440
975.49	72.30	75.06	68.09	71.82	0.0399
975.86	68.68	74.28	67.13	70.03	0.0439
976.24	65.15	73.51	66.17	68.28	0.0546
976.61	61.71	72.74	65.23	66.56	0.0692
976.99	58.36	71.98	64.28	64.87	0.0859
977.36	55.10	71.22	63.35	63.22	0.1041
977.74	51.93	70.46	62.42	61.60	0.1231

(*b*) Reflection (9 7 2)

	$\lvert F^+\rvert$	$\lvert F^-\rvert$
λ_1	564.8	587.0
λ_2	556.7	604.7
λ_3	555.9	579.0

$\lvert F\rvert$	m_1^2	m_2^2	m_3^2	$\overline{m^2}$	T'
591.63	117.31	136.15	24.99	92.82	0.5233
591.84	120.37	136.17	25.26	93.93	0.5215
592.06	123.48	136.19	25.53	95.07	0.5201
592.27	126.64	136.21	25.80	96.22	0.5190
592.48	129.85	136.24	26.08	97.39	0.5184
592.69	133.11	136.27	26.36	98.58	0.5182
592.90	136.43	136.29	26.64	99.79	0.5183
593.11	139.79	136.32	26.93	101.01	0.5188
593.32	143.21	136.35	27.21	102.26	0.5197
593.53	146.67	136.38	27.50	103.52	0.5208
593.74	150.19	136.42	27.79	104.80	0.5224

Table 8. *The residual, as defined in* (38), *for the* N *most reliably indicated values of* m^2 *according to the values of* T^v, *found from* REVISE *and* AGREE

(The final row, for $N = 4018$, gives the overall residual for all available reflections.)

N	residual	N	residual
500	0.744	3000	0.862
1000	0.783	3500	0.880
1500	0.790	4000	0.928
2000	0.811	4018	0.946
2500	0.837		

Table 9. *The top 20 peaks, excluding the origin peak, from a map with the 2000 most reliably indicated values of m^2, as indicated by T' following the use of* REVISE *and* AGREE, *used as coefficients*

(The peaks between selenium atoms are numbered according to the scheme used in table 4.)

peak no.	height	x	y	z	table 4 peak no.
1	530	0.0000	0.0997	0.5000	4
2	480	0.4177	0.0050	0.5000	5
3	402	0.0000	0.0526	0.5000	—
4	390	0.1330	0.0622	0.3396	7
5	363	0.1312	0.1671	0.1472	8
6	343	0.5000	0.0023	0.4333	—
7	342	0.3140	0.0030	0.1693	2
8	335	0.0000	0.0205	0.0798	—
9	335	0.4239	0.1090	0.0000	6
10	329	0.5000	0.1229	0.5000	—
11	322	0.3321	0.2220	0.0000	3
12	311	0.0216	0.0395	0.0000	—
13	305	0.0000	0.2287	0.1900	1
14	293	0.2098	0.0004	0.4991	—
15	280	0.0425	0.0006	0.0485	—
16	273	0.4492	0.0649	0.1526	10
17	262	0.4706	0.1869	0.5000	—
18	254	0.5000	0.0961	0.3304	—
19	253	0.4556	0.1630	0.3389	9
20	252	0.0000	0.1550	0.0404	—

g determined by AGREE, it is possible by the use of (36) and (37) to find estimated phases for all 4217 reflections, including the centric ones. When the estimated values of ϕ, using data unmodified by REVISE, are compared with those calculated from the refined structure the mean phase error is 62.0°. The MCC for these phases is 0.456; the map contains regions which can be interpreted with a model and would probably be a starting point for a complete structure determination.

The phase values deduced from the data modified by REVISE give a substantially better result although only 4018 reflections are available since the magnitudes of some reflections did not seem to refine properly with our REVISE procedure. The mean phase error is 54.3° and the MCC is 0.549. This shows the benefit of subjecting the values of $|F^+|$ and $|F^-|$ to REVISE before using them in any phase-estimating process – and the more data-sets at different wavelengths that are collected the more should be the benefit.

We conclude that with the data collected at wavelengths λ_1, λ_2 and λ_3, modified by REVISE and analysed by AGREE the structure of core streptavidin could almost certainly be solved without too much difficulty.

9. The ROTATE method

In figure 5 we show the diagram which gives the anomalous structure amplitudes rotated so that the non-anomalous contribution, F, is drawn horizontally and the anomalous contributions for λ_i, are CP_i and CQ_i for F^+ and F^- respectively. The angle θ is, as in figure 4, that between F and the real contribution of the anomalous

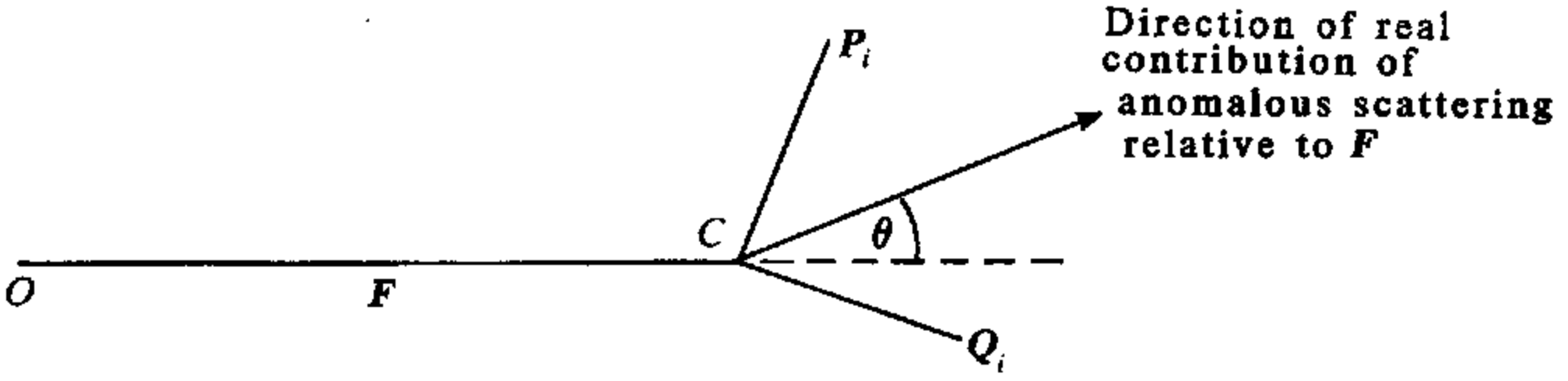

Figure 5. CP_i and CQ_i, the contributions of the anomalous scattering to F^+ and F^{-*}, are known. If $|F|$ and θ are correct then $OP_i = |F^+|$ and $OQ_i = |F^-|$.

scattering, and is wavelength independent. In this analysis we assume that the positions of the anomalous scatterers are known; we have seen that there are various ways by which they can be found. Thus we also know the values of

$$(CP_i)^2 = (CQ_i)^2 = m^2\{(f_i')^2 + (f_i'')^2\}. \tag{39}$$

For particular assumed values of $|F|$ and θ it is simple to calculate the values of OP_i and OQ_i which, in a perfect situation and with $|F|$ and θ at their correct values, would equal $|F_i^+|$ and $|F_i^-|$. For the assumed values of $|F|$ and θ we are actually considering and for the observed data we write

$$\left.\begin{aligned} \Delta|F_i^+| = OP_i - |F_i^+|, \quad \Delta|F_i^-| = OQ_i - |F_i^-|, \\ S(|F|,\theta) = \sum_i \left\{\left(\frac{\Delta|F_i^+|}{\sigma_i^+}\right)^2 + \left(\frac{\Delta|F_i^-|}{\sigma_i^-}\right)^2\right\}, \end{aligned}\right\} \tag{40}$$

where the summation is over all available wavelengths λ_i.

Starting with some values of $|F|$ and θ which are not too far from the correct values we might expect to be able to get improved estimates of their values by minimization of S. This can be done simply, if not most efficiently, by a steepest descents approach where the successive shifts in $|F|$ and θ are given by

$$\delta\theta = -\alpha S \frac{\partial S}{\partial \theta} \Big/ \left\{\left(\frac{\partial S}{\partial \theta}\right)^2 + \left(\frac{\partial S}{\partial |F|}\right)^2\right\}^{\frac{1}{2}} \tag{41a}$$

and

$$\delta|F| = -\alpha S \frac{\partial S}{\partial |F|} \Big/ \left\{\left(\frac{\partial S}{\partial \theta}\right)^2 + \left(\frac{\partial S}{\partial |F|}\right)^2\right\}^{\frac{1}{2}}, \tag{41b}$$

where α is a damping constant which is made small (*ca.* 0.01) to avoid the refinement repeatedly jumping over the minimum point.

In our approach, incorporated in a computer programme ROTATE, we have started with six different values of θ, the sextant values 30°, 90°, 150°, 210°, 270° and 330° and a value of $|F|$ equal to the average of all the available values of $|F^+|$ and $|F^-|$. Some results for the core streptavidin data are shown in table 10. Most of the situations are as seen for (8 2 2), (8 4 2) and (9 5 2) where all starting points refine to the same final values. Occasionally a situation arises, such as for (9 7 2) and (8 24 2) where two different minima are found but normally they are well distinguished in plausibility by the associated values of S. When the ROTATE program was first produced provision was made for a range of starting values of $|F|$ but this was found to be unnecessary and merely added to the computing effort. It should be said at this stage that the program takes little computer time so there has been little incentive to use a more efficient refinement procedure.

Table 10. ***Results from ROTATE for a selection of core streptavidin data***
(The final column gives the number of cycles in the refinement, limited to 500.)

h	k	l	θ_i/deg	θ_f/deg	$\|F\|_i$	$\|F\|_f$	$\overline{\|\Delta\|F^{\pm}\|\|}$	number of cycles
8	2	2	30	157	1020.6	987.4	11.07	500
			90	158		986.9	10.65	500
			150	157		987.3	10.99	500
			210	159		986.8	10.52	500
			270	159		986.5	10.24	500
			330	159		986.9	10.57	500
8	4	2	30	311	721.2	736.4	7.85	171
			90	311		736.3	7.86	162
			150	305		734.6	7.82	251
			210	307		735.1	7.82	192
			270	308		735.5	7.82	166
			330	307		735.3	7.82	182
8	24	2	30	84	1562.2	1564.0	31.22	19
			90	4		1610.9	20.87	242
			150	85		1563.5	31.28	30
			210	3		1610.8	20.86	271
			270	7		1609.9	20.95	214
			330	6		1610.7	20.93	236
9	5	2	30	166	220.5	204.9	11.66	163
			90	166		204.9	11.65	164
			150	166		205.0	11.66	163
			210	165		204.9	11.65	208
			270	163		205.1	11.64	375
			330	167		204.9	11.66	204
9	7	2	30	88	574.7	573.9	12.74	156
			90	87		564.7	12.75	5
			150	89		573.8	12.74	41
			210	99		566.8	12.84	116
			270	95		569.0	12.76	45
			330	276		575.3	36.38	22

The results of using ROTATE have been quite encouraging. Without applying REVISE to the data the values of θ led to a mean phase error of 57.0° with a corresponding MCC of 0.512. Some improvement was obtained by preprocessing the data with REVISE, the phase error reducing to 54.9° with corresponding MCC 0.519. A slightly better result than either of these results was obtained by taking that value of θ from the two approaches, with and without using REVISE, which had the smaller value of S. This gave a mean phase error of 54.4° with an MCC of 0.527. The map corresponding to this last result shows the form of the most of the molecule although there are obviously distortions and deficiencies; a section of this core streptavidin map with part of the structure overlaid is shown in figure 6 both for the map with phases derived from the refined structure and also that with phases from ROTATE. The section was chosen randomly, not as the best available, and the whole map should provide a good basis from which to determine a refined structure.

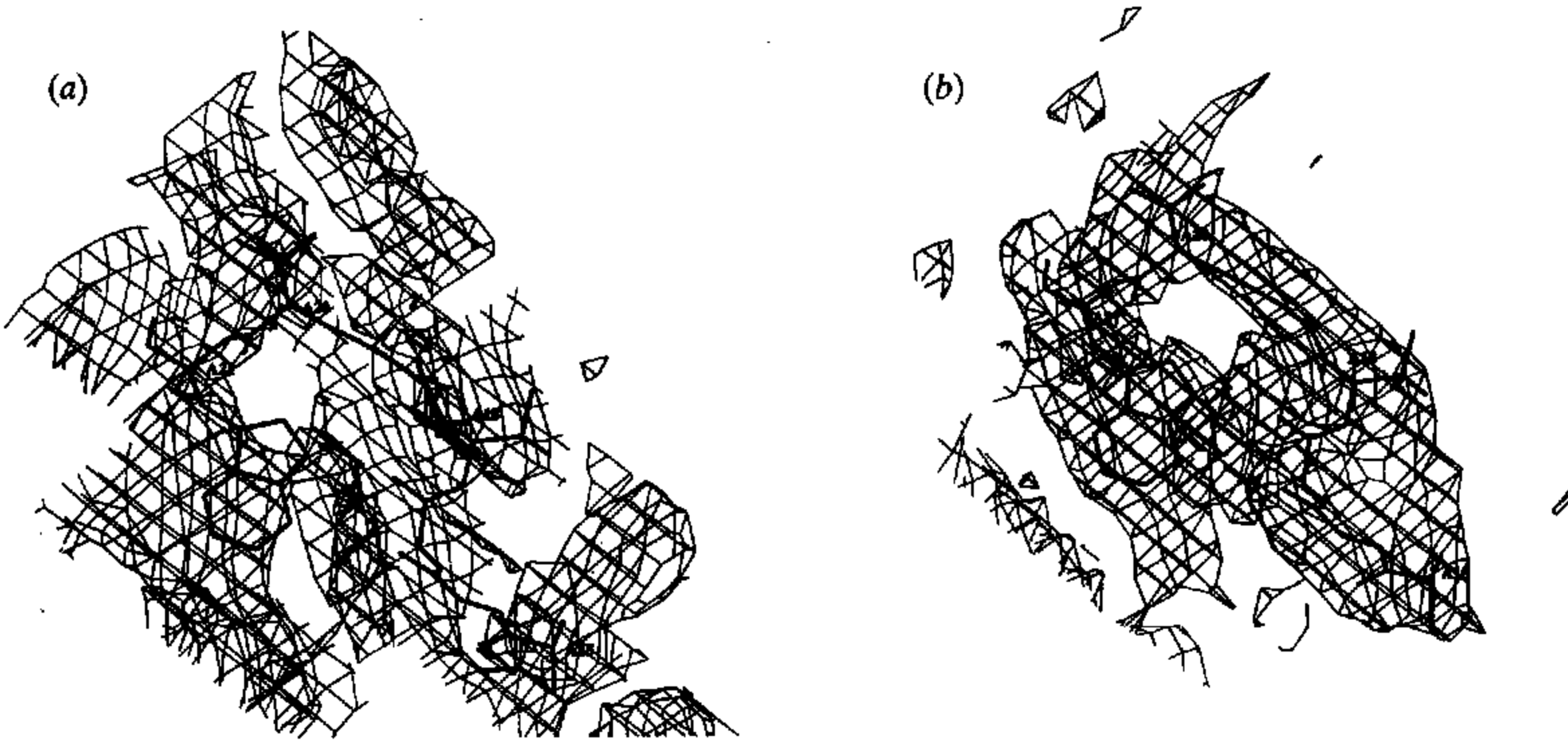

Figure 6. (*a*) A section of a core streptavidin map calculated with phases from a combination of ROTATE and REVISE + ROTATE with part of the molecule superimposed. (*b*) The same section of map for true phases calculated from the final refined structure.

Table 11. *The results from a selection of methods of using anomalous scattering data at three wavelengths for core streptavidin giving the map correlation coefficients* (MCC) *and, where appropriate, the mean phase error*

(For the P_s function only non-centric reflections are used. For comparison the result obtained with ideal data, calculated from the coordinates of the refined structure, is shown.)

method	no. of reflections	mean phase error/deg	MCC
P_s function λ_1	3364	—	0.231
P_s function λ_2	3364	—	0.208
P_s function λ_{31}	3364	—	0.305
P_s function $\lambda_1+\lambda_2+\lambda_3$	3364	—	0.334
REVISE + P_s function	3364	—	0.343
P_s function with ideal data	3364	—	0.455
AGREE	4217	62.0	0.456
REVISE + AGREE	4018	54.3	0.549
ROTATE (A)	4217	57.0	0.512
REVISE + ROTATE (B)	4018	54.9	0.519
best indicated of A and B	4018	54.4	0.527

10. General conclusions

We have considered several methods of using multi-wavelength anomalous scattering data and for ease of comparison we show together in table 11 the results from use of the P_s function, and also AGREE and ROTATE with and without REVISE. The methods P_{x_3}, $P_{x_3'}$ and $P_{x_3''}$ are all clearly unsuccessful because they are too sensitively dependent on the errors in the measured values of $|F^+|$ and $|F^-|$. On the other hand one does get some benefit from map-dependent methods by the use of the P_s function with the individual wavelengths and then using a sensible combination of the separate maps.

Of the two analytical methods AGREE, in conjunction with REVISE, seems very attractive giving the best result with core streptavidin, although it might not be safe to generalize from a single example. We hope that, if we can acquire further multi-wavelength data sets, we can carry out further experiments with REVISE, AGREE and ROTATE with a view to making them even more effective. In the meanwhile we recommend their use to those who collect multi-wavelength anomalous scattering data.

We express our gratitude to Professor Wayne Hendrickson, whose data have enabled us to carry out these studies. We must also sincerely thank the Science and Engineering Research Council and the Wellcome Trust for their support, without which we could do little.

References

Fan Hai-fu, Hao Quan & Woolfson, M. M. 1990*a* The application of one-wavelength anomalous scattering. I. Combining results of different methods. *Acta crystallogr.* A **46**, 656–659.

Fan Hai-fu, Hao Quan & Woolfson, M. M. 1990*b* The application of one-wavelength anomalous scattering. II. An analytical approach for phase determination. *Acta crystallogr.* A **46**, 659–664.

Fan Hai-fu, Han Fu-son, Qian Jin-zi & Yao Jia-xing 1984 Combining direct methods with isomorphous replacement or anomalous scattering data. *Acta crystallogr.* A **40**, 489–495.

Hao Quan & Woolfson, M. M. 1989 Application of the P_s-function method to macromolecular structure determination. *Acta crystallogr.* A **45**, 794–797.

Hendrickson, W. A., Pähler, A., Smith, J. L., Satow, Y., Merritt, E. A. & Phizackerley, R. P. 1989 Crystal structure of core streptavidin determined from multiwavelength anomalous diffraction of synchrotron radiation. *Proc. natn. Acad. Sci. USA* **86**, 2190–2194.

Karle, J. 1980 Some developments in anomalous dispersion for the structural investigation of macromolecular systems in biology. *Int. J. quant. Chem. Symp.* **7**, 357–367.

Karle, J. 1984 The relative scaling of multiple-wavelength anomalous dispersion data. *Acta crystallogr.* A **40**, 1–4.

Karle, J. 1985 Unique or essentially unique results from one wavelength anomalous dispersion data. *Acta crystallogr.* A **41**, 387–394.

Karle, J. 1989 Linear algebraic analyses of structures with one predominant type of anomalous scatterer. *Acta crystallogr.* A **45**, 303–307.

Mukherjee, A. K., Helliwell, J. & Main, P. 1989 The use of MULTAN to locate the positions of anomalous scatterers. *Acta crystallogr.* A **45**, 715–718.

Okaya, Y., Saito, Y. & Pepinsky, R. 1955 New method in X-ray crystal structure determination involving the use of anomalous dispersion. *Phys. Rev.* **98**, 1857–1858.

Ralph, A. C. & Woolfson, M. M. 1991 The application of one-wavelength anomalous scattering. III. The Wilson distribution and MPS methods. *Acta crystallogr.* A **47**, 533–537.