describe here and possibly with the majority of other methods to be found in the literature.

The results are very similar to those for direct methods (Fan *et al.*, 1984), the P_s-function method (Hao & Woolfson, 1989) and the analytical method (Fan, Hao & Woolfson, 1990). This really raises the question whether the errors from the various methods are highly correlated with respect to individual reflections. If the correlation was low and the weighting schemes were sensible then it should be possible to produce a better answer than that from any single method by combining the results from all the methods. Our experiments in this direction show that it is indeed possible to produce a better result from pairs of methods, or even from triplets, but combining results from any more than three methods produces no gain. We suspect that each of the five methods we have explored in our laboratories in Beijing and York are exploiting the information from OAS in approximately equivalent ways and so are giving similar results. A related conclusion is that it would probably not be productive to look for even more methods of similar type for exploiting OAS data, although we do not discount the possibility that more sophisticated approaches may be more successful.

Finally, we point out from the results in Table 3 that it may be rather misleading to assess the practical effectiveness of procedures by using only calculated data sets – even those with added errors if the error simulation is not done realistically. The virtue of using calculated data is that they reveal the intrinsic properties of the method without the complication of errors in the data which will vary from one experimenter to another and with the technique of data collection. However, error-free data do not exist and errors do not occur in convenient ways. Methods like the P_s-function method, which use values of $|F(\mathbf{h})| - |F(\bar{\mathbf{h}})|$, can comfortably tolerate random errors of factors of, say, zero to three in these quantities. However, for real observed data, the error factors can be much higher and even the sign will be wrong for some of the anomalous differences. Whether or not any particular method will work with real data is not easy to predict. However, since there are very few occasions when real observed data cannot be used, we would advocate their use whenever possible. If this is not possible then, as can be seen from Table 4, it is possible to simulate errors in a more complicated way that gives data with characteristics comparable to those of observed data – including giving anomalous differences with the wrong sign.

References

BLOW, D. M. & ROSSMANN, M. G. (1961). *Acta Cryst.* **14**, 1195–1202.
DODSON, G. G., SEVCIK, J., DODSON, E. & ZELINKA, J. (1987). *Metabolism of Nucleic Acids Including Gene Manipulation*, pp. 33–36. Prague: Slovak Academy of Sciences.
FAN, H., HAN, F., QIAN, J. & YAO, J. (1984). *Acta Cryst.* A**40**, 489–495.
FAN, H., HAO, Q. & WOOLFSON, M. M. (1990). *Acta Cryst.* A**46**, 656–659.
GLOVER, I. D., MOSS, D. S., TICKLE, I. J., PITTS, J. E., HANEEF, I., WOOD, S. P. & BLUNDELL, T. L. (1985). *Adv. Biophys.* **20**, 1–12.
HAO, Q. & WOOLFSON, M. M. (1989). *Acta Cryst.* A**45**, 794–797.
KARLE, J. (1985). *Acta Cryst.* A**41**, 387–394.
KARTHA, G. (1961). *Acta Cryst.* **14**, 680–686.
OKAYA, Y., SAITO, Y. & PEPINSKY, R. (1955). *Phys. Rev.* **98**, 1857–1858.
WANG, B. C. (1985). *Diffraction Methods in Biological Macromolecules*, edited by H. W. WYCKOFF, C. H. W. HIRS & S. N. TIMASHEFF, pp. 90–112. New York: Academic Press.

Some Additional Features of One-Wavelength Anomalous Dispersion

BY JEROME KARLE

Laboratory for the Structure of Matter, Naval Research Laboratory, Washington, DC 20375-5000, USA

Abstract

By use of appropriate algebraic formulas, illustrations are given of several characteristics of one-wavelength anomalous-dispersion data, for the case that one predominant type of anomalous-scattering atom is present. It is shown that, when the structure of the anomalous scatterer is known, some simple algebraic formulas may be used to generate initial values of many phases associated with a macromolecular structure. In some cases, there may be enough phases determined to permit further refinement and

extension by use of current techniques for doing so. Another calculation shows the virtue of including isomorphous-replacement data and how readily that is done in an algebraic system. It is also shown that there is an advantage in accuracy through the use of a particular statistical calculation of the magnitude of the structure factors for the anomalous scatterers, even when the magnitudes are known accurately because the structure of the anomalous scatterers is known. This occurs because the statistical values are scaled to adjust to errors in the data and thus avoids a disparity in scale that would occur otherwise in the algebraic system. With highly accurate data, the advantage would disappear. It is also shown, however, that, even with accurate one-wavelength anomalous-dispersion data, the coupling of isomorphous-replacement data leads to greater accuracy in evaluating phase differences than coupling with known magnitudes for the structure factors associated with the anomalous scatterers. Finally, there is a discussion of the use of known and statistically estimated values of the structure-factor magnitudes that correspond to the total structure in which all atoms (including the anomalous scatterers) scatter as if they were all scattering normally.

Introduction

In previous studies (Karle, 1985b, 1989), it has been shown that there are circumstances in which the usual ambiguity associated with the calculation of the values of two-phase invariants can be largely resolved by making use of statistical information. This was illustrated for the case of macromolecules in which most atoms scatter X-rays with negligible anomalous dispersion and only one predominant type of anomalously scattering atoms is present. In this article, several features of one-wavelength anomalous-dispersion data, some of which were alluded to in previous publications, will be discussed further. They concern such matters as the initial evaluation of phases for the macromolecular structure, the resolution of the twofold ambiguity and the relative virtue of using a theoretical calculation of the values of the intensities for the structure of the anomalously scattering atoms instead of the values computed from a known structure.

The subject of this article is mainly one-wavelength anomalous dispersion. There are, of course, other sources of data which could be additionally quite helpful, such as multiple-wavelength anomalous dispersion and multiple isomorphous replacement. With good data, a one-wavelength experiment may prove to be useful and convenient. In any case, studies of the latter serve to characterize the anomalous-dispersion technique.

For other approaches to resolving the twofold ambiguity, there are the use of the heavy-atom posi-

tions as applied, for example, in the investigation of crambin (Hendrickson & Teeter, 1981), the filtering method of Wang (1985) and the sine Patterson function of Okaya, Saito & Pepinsky (1955) discussed further by Hao & Woolfson (1989).

Description of calculations

Initial phase evaluation when the structure of the anomalous scatterers is known

It was previously noted (Karle, 1986) that the following exact equations obtain for the case that one type of anomalous scatterer is present:

$$\sin(-\varphi_{A\bar{h}} - \varphi_{2,h}^n) = [|F_{Ah}|^2 - |F_{A\bar{h}}|^2 - 4(f_{A2}''/f_{2,h}^n)^2|F_{2,h}^n|^2]$$
$$\times [4(f_{A2}''/f_{2,h}^n)|F_{A\bar{h}}||F_{2,h}^n|]^{-1} \quad (1)$$

$$\sin(\varphi_{Ah} - \varphi_{2,h}^n) = [|F_{Ah}|^2 - |F_{A\bar{h}}|^2 + 4(f_{A2}''/f_{2,h}^n)^2|F_{2,h}^n|^2]$$
$$\times [4(f_{A2}''/f_{2,h}^n)|F_{Ah}||F_{2,h}^n|]^{-1} \quad (2)$$

where F_{Ah} is the structure factor for a macromolecule containing one type of anomalous scatterer at some incident wavelength, λ, φ_{Ah} is the corresponding phase, f_{A2}'' is the imaginary correction to the normal atomic scattering factor, $f_{2,h}^n$, of the anomalously scattering atoms, $F_{2,h}^n$ is the structure factor for the anomalously scattering atoms, scattering as if there were no anomalous corrections, and $\varphi_{2,h}^n$ is the corresponding phase. If the structure of the anomalous scatterers is known, it is possible to compute the values of the $F_{2,h}^n$ and therefore the values of the $|F_{2,h}^n|^2$ and $\varphi_{2,h}^n$. Equations (1) and (2) can then be used to obtain initial values of φ_{Ah} by specifying a magnitude somewhat less than unity that the right sides of (1) and (2) must attain. When this value or greater is attained, the sine functions may be set approximately equal to +1 or −1, as appropriate, and then the $-\varphi_{A\bar{h}}$ or φ_{Ah} may be set approximately equal to $\varphi_{2,h}^n + \pi/2$ or $\varphi_{2,h}^n - \pi/2$, corresponding to the values of +1 and −1, respectively. This could give an initial set of values for a number of phases. It appears to be a convenient way to obtain an initial set of phase values having an acceptable error. Evidently, in this calculation, we are dealing with the classical single-wavelength anomalous-scattering phase ambiguity. An average value between two indistinguishable ambiguous alternatives is assigned when the alternatives differ by a tolerable amount from their average value. With errorless data, only when the magnitude of the right sides of (1) and (2) is equal to unity can these equations give a unique exact evaluation.

Calculations based on (1) and (2) are shown in Table 1. They were performed with data for cytochrome c550.PtCl$_4^{2-}$ (Timkovitch & Dickerson, 1976) at 2·5 Å resolution. The calculations were made at three different limiting magnitudes for the right sides of (1) and (2), 0·938, 0·766 and 0·000. The error

Table 1. *The evaluation of individual phases, $\varphi_{\lambda h}$ and $\varphi_{\lambda \bar{h}}$, with anomalous-dispersion data by use of* (1), (2) *and* 2·5 Å *data for cytochrome* c550.$PtCl_4^{2-}$ (Cu Kα *radiation*)

Those values of the phases $\varphi_{\lambda h}$ and $\varphi_{\lambda \bar{h}}$ that were determined to be sufficiently close to $\varphi_{2,h}^n + \pi/2$ or $\varphi_{2,h}^n - \pi/2$ by use of (1) and (2) were set equal to those values. This was done when the magnitudes of the right sides of (1) and (2) were equal to or larger than the magnitudes in column 2, otherwise the phases were not evaluated. The total number of independent noncentric data for cytochrome is 3252 at 2·5 Å resolution. The factor 1·5 means that an error was introduced into the correct difference $|F_{\lambda h}| - |F_{\lambda \bar{h}}|$ by making it 1·5 times larger. The factor 1·0 signifies that no error was introduced into the difference and 0·5 that the correct difference was reduced by a factor of 2. The average magnitudes of the errors for $\varphi_{\lambda h}$ and $\varphi_{\lambda \bar{h}}$ at the various acceptance levels are shown in columns 5 and 7, respectively.

Number of reflections	Limit for |right sides| of (1), (2)	Error factor for $\|\|F_{\lambda h}\| - \|F_{\lambda \bar{h}}\|\|$	Number of $\varphi_{\lambda h}$ accepted	$\langle\|Error\|\rangle$ for $\varphi_{\lambda h}$ (rad)	Number of $\varphi_{\lambda \bar{h}}$ accepted	$\langle\|Error\|\rangle$ for $\varphi_{\lambda \bar{h}}$ (rad)
3252	0·938	1·5	46	0·313	45	0·457
3252	0·766	1·5	715	0·444	699	0·443
3252	0·000	1·5	3252	0·849	3252	0·848
3252	0·938	1·0	645	0·183	618	0·177
3252	0·766	1·0	1266	0·353	1267	0·352
3252	0·000	1·0	3252	0·835	3251	0·836
3252	0·938	0·5	5	0·244	4	0·169
3252	0·766	0·5	629	0·423	622	0·427
3252	0·000	0·5	3252	0·840	3252	0·841

factors are factors that multiplied the correct differences $|F_{\lambda h}| - |F_{\lambda \bar{h}}|$. These errors were distributed equally among the $|F_{\lambda h}|$ and $|F_{\lambda \bar{h}}|$. The factor 1·0 implies that no error was introduced into the data. The results of the calculations are shown in the last four columns. It is seen that numerous phases can be evaluated in this way with sufficient accuracy to be potentially useful, even for relatively large errors in $\|F_{\lambda h}| - |F_{\lambda \bar{h}}\|$. Phases obtained by this calculation are identical to those that would be obtained by classical single-anomalous-scattering methods, *i.e.* the centroid of the bimodal probability distribution. The use of a particular lower limit for the right sides of (1) and (2) is identical to selecting the subset of reflections having the highest conventional single-anomalous-scattering figures of merit. The phase values would be suitable for further refinement and extension techniques, for example, with use of triplet phase invariants or as starting values in the least-squares analysis of two-phase invariants (Karle, 1989).

Use of isomorphous-replacement data

In this section, a system of equations is considered for a structure in which there are atoms having negligible anomalous scattering and, in addition, only one type of anomalously scattering atom. The appropriate equations are (Karle, 1980, 1985b, 1989)

$$|F_{\lambda h}|^2 = \chi_1 + \alpha_h \chi_2 + \beta_h \chi_3 + \gamma_h \chi_4 \quad (3)$$

$$|F_{\lambda \bar{h}}|^2 = \chi_1 + \alpha_h \chi_2 + \beta_h \chi_3 - \gamma_h \chi_4 \quad (4)$$

$$\chi_3^2 + \chi_4^2 = \chi_1 \chi_2 \quad (5)$$

$$\chi_1 = |F_{1,h}^n|^2 \quad (6)$$

$$\chi_2 = |F_{2,h}^n|^2 \quad (7)$$

$$\chi_3 = |F_{1,h}^n||F_{2,h}^n| \cos(\varphi_{1,h}^n - \varphi_{2,h}^n) \quad (8)$$

$$\chi_4 = |F_{1,h}^n||F_{2,h}^n| \sin(\varphi_{1,h}^n - \varphi_{2,h}^n) \quad (9)$$

$$\alpha_h = 1 + 2f'_{\lambda 2}/f_{2,h}^n + (f'^2_{\lambda 2} + f''^2_{\lambda 2})/(f_{2,h}^n)^2 \quad (10)$$

$$\beta_h = 2(1 + f'_{\lambda 2}/f_{2,h}^n) \quad (11)$$

$$\gamma_h = 2f''_{\lambda 2}/f_{2,h}^n. \quad (12)$$

The quantity $f'_{\lambda 2}$ is the real part of the correction to the normal atomic scattering factor, $f_{2,h}^n$, of the anomalously scattering atoms (labeled by the subscript 2), $F_{1,h}^n$ is the structure factor for the non-anomalously scattering atoms and $\varphi_{1,h}^n$ is the corresponding phase. $F_{\lambda h}$ and $F_{2,h}^n$ are structure factors as defined for (1) and (2).

The system of equations (3)–(5) has been used in a one-wavelength context to evaluate $\varphi_{1,h}^n - \varphi_{2,h}^n$ by making use of a statistical estimate of $|F_{2,h}^n|^2$, thus reducing the number of unknown quantities from four to three (Karle, 1985b). A least-squares technique was used to evaluate the unknown quantities from the equations which contain a twofold ambiguity. The calculation of initial values of the unknown quantities for use in the least-squares procedure with (3)–(5) has been described (Karle, 1985a, 1989). It was apparent from the results of the calculation that the method usually selected the ambiguous alternative that was closer to the correct value. A geometric analysis (Karle, 1985a) suggested that the reason for this was the degree of accuracy obtained from the statistical method used to obtain starting values of the $|F_{1,h}^n|^2$ for the least-squares calculation.

The calculation presented here is related to the latter conclusion. Here, instead of estimating the values of the $|F_{2,h}^n|^2$ and holding them fixed, the values of the $|F_{1,h}^n|^2$ are considered known and introduced into the system (3)–(5). These values would be available from an isomorphous-replacement experiment in which the anomalous scatterers were introduced by isomorphous substitution. The values of the $|F_{1,h}^n|^2$ would correspond to the data obtained from the native protein.

Table 2. *Averages of the magnitudes of error for data from cytochrome c550.PtCl$_4^{2-}$ for the phase differences, $\varphi_{1,h}^{n} - \varphi_{2,h}^{n}$, as a function of the type of sample and errors in known values of $|F_{1,h}^{n}|^2$*

The data were treated as exact and calculated as if only the Pt atoms scattered anomalously. The samples are based on **h** associated with an ordered sequence of $\||F_{\lambda h}| - |F_{\lambda \bar{h}}|\|$ in which the largest one is first. Initial values of $|F_{2,h}^{n}|^2$ were obtained from (13) for application of system (3)-(5) in a least-squares fashion. The total number of independent noncentric data is 3252 at 2·5 Å resolution and the radiation is Cu $K\alpha$.

	$\langle|\text{Error}|\rangle$ (rad) for $\varphi_{1,h}^{n} - \varphi_{2,h}^{n}$ given $\langle|\text{Error}|\rangle$ for $|F_{1,h}^{n}|^2$		
Sample	0%	5%	10%
1–100	0·001	0·061	0·135
901–1000	0·007	0·113	0·230
1601–1700	0·000	0·111	0·212
1–1700	0·003	0·098	0·190

Table 2 shows the average magnitude of error for the values of $(\varphi_{1,h}^{n} - \varphi_{2,h}^{n})$ obtained from (3)-(5) when the values of the $|F_{1,h}^{n}|^2$ are introduced into the system with a variety of average errors. The calculations were made for 3252 exact independent noncentric data at 2·5 Å resolution from cytochrome c550.PtCl$_4^{2-}$ (Cu $K\alpha$ radiation). The data were calculated as if only the Pt atoms scattered anomalously. Samples were composed from the **h** associated with an ordered sequence of $\||F_{\lambda h}| - |F_{\lambda \bar{h}}|\|$ in which the largest one is first.

It is seen from the second column of Table 2 that, when exact values of the $|F_{\lambda h}|$ and $|F_{1,h}^{n}|$ are used, essentially exact values of the $\varphi_{1,h}^{n} - \varphi_{2,h}^{n}$ are obtained from (3)-(5). Evidently the correct ambiguous alternative is generally obtained. This is consistent with the earlier geometric interpretation of ambiguity resolution, mentioned above, when statistically estimated values for the $|F_{2,h}^{n}|^2$ were used in a least-squares calculation of (3)-(5) (Karle, 1985b). It is also seen that 5 and 10% random errors in $|F_{1,h}^{n}|^2$ lead to small average errors in the $\varphi_{1,h}^{n} - \varphi_{2,h}^{n}$.

Use of the known structure of the anomalous scatterers

If the structure of the anomalous scatterers is known, it is possible to calculate readily the values of $|F_{2,h}^{n}|^2$ and $\varphi_{2,h}^{n}$. In the previous section, it was seen that knowledge of the exact values of the $|F_{1,h}^{n}|^2$ leads to high accuracy in the application of (3)-(5). The question arises concerning what the comparable accuracy would be if exact values of $|F_{2,h}^{n}|^2$ were used in (3)-(5) with exact values of the $|F_{\lambda h}|^2$ and $|F_{\lambda \bar{h}}|^2$. Such a calculation is shown in Table 3 for cytochrome c550.PtCl$_4^{2-}$ and Hg elastase (Shotton & Watson, 1970; Watson, Shotton, Cox & Muirhead, 1970).

The sampling level is a point at some fraction of the total data set, listed in decreasing values of $\||F_{\lambda h}| - |F_{\lambda \bar{h}}|\|$. The average magnitudes of error in the phase differences, $\varphi_{1,h}^{n} - \varphi_{2,h}^{n}$, are averages of several hun-

Table 3. *Averages of the magnitudes of error for data from cytochrome c550.PtCl$_4^{2-}$ and Hg elastase for the phase differences $\varphi_{1,h}^{n} - \varphi_{2,h}^{n}$ as a function of sample level for exact values of $|F_{2,h}^{n}|^2$*

Sampling level is a point at some fraction of the total data set, listed in order of decreasing values of $\||F_{\lambda h}| - |F_{\lambda \bar{h}}|\|$ and the average magnitude of error in the phase differences, $\varphi_{1,h}^{n} - \varphi_{2,h}^{n}$, is the average of several hundred determinations taken in the vicinity of the sampling level. The total number of independent noncentric data for cytochrome was 3252 and for elastase 6864, both at 2·5 Å resolution. At 2·0 Å resolution for elastase, there were 13 672 data. The results were obtained from the application of system (3)-(5) in a least-squares fashion. Exact data for Cu $K\alpha$ radiation were used.

	$\langle|\text{Error}|\rangle$ (rad) for $\varphi_{1,h}^{n} - \varphi_{2,h}^{n}$		
	Cytochrome c550.PtCl$_4^{2-}$	Hg elastase	
Sampling level	2·5 Å resolution	2·5 Å resolution	2·0 Å resolution
0·10	0·119	0·177	0·159
0·25	0·206	0·290	0·272
0·50	0·319	0·440	0·443
0·75	0·560	0·623	0·593

dred determinations taken in the vicinity of the sampling level. At 2·5 Å resolution the total number of independent noncentric data for cytochrome is 3252 and for Hg elastase 6864. At 2·0 Å resolution for elastase there are 13 672 data. In all cases, Cu $K\alpha$ radiation was used.

It is seen from Table 3 that although the average magnitude of errors of the phase differences, $\varphi_{1,h}^{n} - \varphi_{2,h}^{n}$, are within what would be considered acceptable bounds, they do not compare with the almost perfect results shown in column 2 of Table 2 from use of exact values for the $|F_{1,h}^{n}|^2$.

Use of theoretical and known values of $|F_{2,h}^{n}|^2$ in system (3)-(5) when there are errors in the data

As has been discussed previously, one way to use (3)-(5) in a single-wavelength experiment is to make a theoretical estimate of $|F_{2,h}^{n}|^2$ and hold its value while a least-squares solution to (3)-(5) is found. Alternatively, if the structure of the anomalous scatterers is known, $|F_{2,h}^{n}|^2$ can be calculated. Since the theoretical estimate of $|F_{2,h}^{n}|^2$ is defined, in part, by $\||F_{\lambda h}| - |F_{\lambda \bar{h}}|\|$ (13) and, in a sense, scaled to the error in the difference of magnitudes, it is not obvious that it is better to use the known values of $|F_{2,h}^{n}|^2$ rather than the theoretical estimate in the evaluation of $\varphi_{1,h}^{n} - \varphi_{2,h}^{n}$ from (3)-(5). In this section, calculations are discussed which afford some insight into this matter.

Estimated values for $|F_{2,h}^{n}|^2$ can be obtained from

$$|F_{2,h}^{n}|^2 \simeq S\{\||F_{\lambda h}| - |F_{\lambda \bar{h}}|\|/[2(f''/f_{2,h}^{n})]\}^2, \quad (13)$$

where S is the scale factor that is equal to 1 when the angles $\varphi_{\lambda h}$ and $-\varphi_{\lambda \bar{h}}$ are equal. An estimate of

Table 4. *Averages of the magnitudes of error for data from* Hg *elastase for the phase differences* $\varphi_{1,h}^{n} - \varphi_{2,h}^{n}$ *determined from system* (3)-(5) *as a function of the type of sample, error factor and source of fixed value for* $|F_{2,h}^{n}|^{2}$

The samples were selected in the same fashion as described in Table 2 and the error factor has the same meaning as in Table 1. The theoretical values for $|F_{2,h}^{n}|^{2}$ were obtained from (13). A convergence technique, in which convergence of the least-squares calculation based on (3)-(5) is facilitated by a systematic narrowing of the gap $\||F_{\lambda h}|-|F_{\lambda \bar{h}}|\|$, was applied in column 4 to those calculations that did not otherwise converge. There was no convergence problem for exact values of $|F_{2,h}^{n}|^{2}$ and 0·5 error factor. There were therefore no calculations to report in the lower half of column 4. The calculations were made with data for Hg elastase at 2·0 Å resolution for which the total number of independent noncentric data is 13 672. The data were calculated for Cu $K\alpha$ radiation.

| | | $\langle|\text{Error}|\rangle$ (rad) for $\varphi_{1,h}^{n} - \varphi_{2,h}^{n}$ given values for $|F_{2,h}^{n}|^{2}$ | | |
Sample	Error factor	Exact value	Exact value + convergence	Theoretical value
1–600	1·5	Nonconvergence	0·697	0·210
4701–5300	1·5	0·537	0·784	0·464
9401–10 000	1·5	0·701	0·763	0·762
1–10 000	1·5	0·570*	0·756	0·511
1–600	0·5	0·824		0·481
4701–5300	0·5	0·946		0·716
9401–10 000	0·5	0·918		0·946
1–10 000	0·5	0·930		0·741

* For all that converge and pass acceptance criteria, *i.e.* a calculation is rejected if the calculated value of $|F_{1,h}^{n}|$ is less than 0·33 $(|F_{\lambda h}|+|F_{\lambda \bar{h}}|)$ or less than a preassigned value which, in this case, was set at 100, or both.

the values for S can be based on test examples having the same atomic composition as the substance of interest. For the test examples a list is made of the differences $\||F_{\lambda h}| - |F_{\lambda \bar{h}}|\|$ in order of decreasing magnitude. For successive subsets of these differences, average values of the known functions on both sides of (13) are evaluated, giving an average value of S for each of the subsets. If the number of differences in each of the subsets is sufficiently large to give average values of S for the test examples with small statistical variance, these values can be used for the unknown structure.

Averages of the magnitudes of error for the phase differences, $\varphi_{1,h}^{n} - \varphi_{2,h}^{n}$, determined from (3)-(5), are illustrated in Table 4 as a function of sample type, error factor and the source of fixed values for $|F_{2,h}^{n}|^{2}$. The samples have been selected in the same way as in Table 2 and the error factor has the same meaning as in Table 1. Theoretical values for $|F_{2,h}^{n}|^{2}$ were obtained from (13) and exact values were obtained from the known structure. A convergence technique, in which convergence was not initially obtained in the least-squares calculation based on (3)-(5), was facilitated by a systematic narrowing of the gap, $\||F_{\lambda h}| - |F_{\lambda \bar{h}}|\|$. The results are seen in column 4. There were no convergence problems for exact values of $|F_{2,h}^{n}|^{2}$ and an error factor of 0·5, and therefore there were no calculations to report in the lower half of column 4. The calculations were made for Cu $K\alpha$ radiation on data for Hg elastase at 2·0 Å resolution for which the total number of independent noncentric data is 13 672.

It is seen from Table 4 that, for this calculation, there is an advantage in proceeding with the use of the theoretical value for $|F_{2,h}^{n}|^{2}$ obtained from (13) rather than the exact value. This advantage is likely to decrease as the errors in $\||F_{\lambda h}| - |F_{\lambda \bar{h}}|\|$ decrease.

Unique values for $\psi_{2,h} - \varphi_{2,h}^{n}$ *and* $|F_{2,h}^{n}|^{2}$ *from a one-wavelength experiment*

The mathematical analysis that forms the basis for this section can be found in an earlier publication (Karle, 1985a). Specifically, with the use of (24) and (25), Tables 1 and 2 (with corrections for cases $m = 1$ and 2 by the insertion of omitted magnitude signs in the third column) from the latter reference, we obtain for $m = 1$

$$\sin ({}_{1}\bar{\psi}_{2,h} - \varphi_{2,h}^{n}) \simeq \frac{|F_{\lambda h}| - |F_{\lambda \bar{h}}|}{2(f_{\lambda 2}''/f_{2,h}^{n})|F_{2,h}^{n}|} \qquad (14)$$

since $f_{\lambda 2}''$ is nonnegative and for $m = 2$

$$\cos ({}_{2}\bar{\psi}_{2,h} - \varphi_{2,h}^{n}) \simeq \frac{|F_{\lambda h} + F_{\lambda \bar{h}}^{*}| + |F_{\lambda \bar{h}} + F_{\lambda h}^{*}| - 2|F_{h}^{n}|}{4(|f_{\lambda 2}'|/f_{2,h}^{n})|F_{2,h}^{n}| \cos \delta_{2\lambda 2}} \qquad (15)$$

where F_{h}^{n} is the structure factor for the entire structure in the absence of anomalous scattering, $\cos \delta_{2\lambda 2}$ is equal to $+1$ if $f_{\lambda 2}'$ is positive and equal to -1 if $f_{\lambda 2}'$ is negative and

$$_{m}\bar{\psi}_{2,h} = 0·5({}_{m}\psi_{2,h} - {}_{m}\psi_{2,\bar{h}}). \qquad (16)$$

The phase ${}_{1}\psi_{2,h}$ is the phase associated with $F_{\lambda \bar{h}}^{*}$, as indicated by the subscript 2 (Karle, 1985a, Table 1). It should be noted that it was shown in this reference that, to the approximation represented in (14) and (15), ${}_{m}\psi_{2,h}$ and ${}_{m}\psi_{1,h}$ are interchangeable. The phase ${}_{1}\psi_{1,h}$ is the phase associated with $F_{\lambda h}$, ${}_{2}\psi_{1,h}$ is the phase associated with $F_{\lambda h} + F_{\lambda \bar{h}}^{*}$ and ${}_{2}\psi_{2,h}$ is the phase associated with F_{h}^{n}. Calculations have indicated (Karle, 1985a, Table 3) that these angles differ from one another on the average by only a few degrees.

Table 5. *Average magnitude of error for* $\bar{\psi}_{2,\mathbf{h}} - \varphi_{2,\mathbf{h}}^{n}$ *and* $|F_{2,\mathbf{h}}^{n}|^{2}$ *when quantities are calculated from* (22) *and* (23) *for exact and statistical values of* $|F_{\mathbf{h}}^{n}|$

Statistical values for $|F_{\mathbf{h}}^{n}|$ are obtained from (23). The calculations are based on data calculated from the structure of cytochrome c550.PtCl$_4^{2-}$ for Cu $K\alpha$ radiation in which it is assumed that Pt is the only anomalous scatterer. The data were calculated to 2·5 Å resolution corresponding to 3252 independent acentric data. In formulating the samples, the data were ordered on $\|F_{\lambda\mathbf{h}}| - |F_{\lambda\bar{\mathbf{h}}}\|$ with the first $|F_{\lambda\mathbf{h}}|^{2}$ corresponding to the largest difference. Results based on exact values of $|F_{\mathbf{h}}^{n}|^{2}$ are given in columns 2 and 4 and those based on statistical values are given in columns 3 and 5.

Sample	⟨|Error|⟩(rad) for $\bar{\psi}_{2,\mathbf{h}} - \varphi_{2,\mathbf{h}}^{n}$	⟨|Error|⟩(rad) for $\bar{\psi}_{2,\mathbf{h}} - \varphi_{2,\mathbf{h}}^{n}$	⟨|Error|⟩(%) for $|F_{2,\mathbf{h}}^{n}|^{2}$	⟨|Error|⟩(%) for $|F_{2,\mathbf{h}}^{n}|^{2}$
1-500	0·056	0·397	1·4	21
501-1000	0·036	0·481	3·2	39
1001-1500	0·036	0·605	4·2	52
1501-2000	0·022	0·684	5·4	67
1-2000	0·038	0·542	3·6	45

With the use of

$$|F_{\lambda\mathbf{h}} + F_{\lambda\bar{\mathbf{h}}}^{*}| + |F_{\lambda\bar{\mathbf{h}}} + F_{\lambda\mathbf{h}}^{*}|$$

$$= 2[|F_{\lambda\mathbf{h}}|^{2} + |F_{\lambda\bar{\mathbf{h}}}|^{2} + 2|F_{\lambda\mathbf{h}}||F_{\lambda\bar{\mathbf{h}}}| \cos (\varphi_{\lambda\mathbf{h}} + \varphi_{\lambda\bar{\mathbf{h}}})]^{1/2}$$

$$\simeq 2[|F_{\lambda\mathbf{h}}|^{2} + |F_{\lambda\bar{\mathbf{h}}}|^{2} + |F_{\lambda\mathbf{h}}||F_{\lambda\bar{\mathbf{h}}}|]^{1/2} \qquad (17)$$

and

$$\bar{\psi}_{\mathbf{h}} = 0 \cdot 25 (_{1}\bar{\psi}_{1,\mathbf{h}} + _{1}\bar{\psi}_{2,\mathbf{h}} + _{2}\bar{\psi}_{1,\mathbf{h}} + _{2}\bar{\psi}_{2,\mathbf{h}}) \qquad (18)$$

where

$$_{m}\bar{\psi}_{1,\mathbf{h}} = 0 \cdot 5 (_{m}\psi_{1,\mathbf{h}} - _{m}\psi_{1,\bar{\mathbf{h}}}), \qquad (19)$$

we may write (16) and (17) as

$$\sin (\bar{\psi}_{\mathbf{h}} - \varphi_{2,\mathbf{h}}^{n}) \simeq \frac{|F_{\lambda\mathbf{h}}| - |F_{\lambda\bar{\mathbf{h}}}|}{2(f_{\lambda2}''/f_{2,\mathbf{h}}^{n})|F_{2,\mathbf{h}}^{n}|} \qquad (20)$$

and

$$\cos (\bar{\psi}_{\mathbf{h}} - \varphi_{2,\mathbf{h}}^{n})$$

$$\simeq \frac{[|F_{\lambda\mathbf{h}}|^{2} + |F_{\lambda\bar{\mathbf{h}}}|^{2} + 2|F_{\lambda\mathbf{h}}||F_{\lambda\bar{\mathbf{h}}}|]^{1/2} - |F_{\mathbf{h}}^{n}|}{2(|f_{\lambda2}'|/f_{2,\mathbf{h}}^{n})|F_{2,\mathbf{h}}^{n}| \cos \delta_{2\lambda2}}. \qquad (21)$$

The average angle $\bar{\psi}_{\mathbf{h}}$ may be written with (18) and appropriate definitions (Karle, 1985a, Table 1)

$$\bar{\psi}_{\mathbf{h}} = 0 \cdot 25 \left[\varphi_{\lambda\mathbf{h}} - \varphi_{\lambda\bar{\mathbf{h}}} + \varphi_{\mathbf{h}}^{n} \right.$$

$$\left. + \tan^{-1} \frac{|F_{\lambda\mathbf{h}}| \sin \varphi_{\lambda\mathbf{h}} - |F_{\lambda\bar{\mathbf{h}}}| \sin \varphi_{\lambda\bar{\mathbf{h}}}}{|F_{\lambda\mathbf{h}}| \cos \varphi_{\lambda\mathbf{h}} + |F_{\lambda\bar{\mathbf{h}}}| \cos \varphi_{\lambda\bar{\mathbf{h}}}} \right]. \qquad (22)$$

A statistical formula for the evaluation of $|F_{\mathbf{h}}^{n}|$ for use in the application of (20) and (21) in a one-wavelength context is (Karle, 1984)

$$|F_{\mathbf{h}}^{n}| \simeq 0 \cdot 5 W_{\lambda\mathbf{h}}(|F_{\lambda\mathbf{h}}| + |F_{\lambda\bar{\mathbf{h}}}|), \qquad (23)$$

where

$$W_{\lambda\mathbf{h}} = \left\{ \frac{\sum\limits_{j=1}^{N_{non}} f_{j\mathbf{h}}^{2} + \sum\limits_{j=1}^{N_{ano}} (f_{j\mathbf{h}}^{n})^{2}}{\sum\limits_{j=1}^{N_{non}} f_{j\mathbf{h}}^{2} + \sum\limits_{j=1}^{N_{ano}} [(f_{j\mathbf{h}}^{n} + f_{j}')^{2} + f_{j}''^{2}]} \right\}^{1/2}. \qquad (24)$$

Test calculations were made with (20) and (21) in which the unknown quantities were $\bar{\psi}_{\mathbf{h}} - \varphi_{2,\mathbf{h}}^{n}$ and $|F_{2,\mathbf{h}}^{n}|$

(Table 5). Two types of calculations were made, one with exact values for $|F_{\mathbf{h}}^{n}|$ and a second with statistical values for $|F_{\mathbf{h}}^{n}|$ computed from (23). The test molecule was cytochrome c550.PtCl$_4^{2-}$ in which Pt was regarded as the sole type of anomalous-scattering atom. Exact data were used. Errors were calculated from knowledge of the structure and use of (22) to give optimal values for $\bar{\psi}_{\mathbf{h}}$.

Examination of the results of Table 5, particularly column 3, and comparing them with the results for exact data in Table 1 of an earlier study based on statistical estimates of $|F_{2,\mathbf{h}}^{n}|^{2}$ (Karle, 1985b) shows that no advantage accrues from evaluating phase differences on the basis of statistical estimates of $|F_{\mathbf{h}}^{n}|$ rather than on statistical estimates of $|F_{2,\mathbf{h}}^{n}|^{2}$. This was also observed in a previous study involving an alternative algebraic analysis (Karle, 1989). In this case as well as the earlier study made with statistical estimates of $|F_{2,\mathbf{h}}^{n}|$, the errors in phase differences with the use of actual rather than exact data may be small enough in some experiments to permit the use of one-wavelength data, if it were considered expedient to do so.

Concluding remarks

It has been shown in Table 1 that it may be possible to obtain, from a simple calculation, a large number of initial values for the phases of structure factors relating to a macromolecular structure, simply from knowing the structure of the anomalous scatterers. Various procedures for phase extension and refinement, including the possibility of using triplet phase invariants, may then be applied with use of the initial values.

The value of combining isomorphous-replacement data with one-wavelength anomalous-dispersion data has long been known. The results shown in Table 2 illustrate that the virtue of this combination may be readily achieved by calculations that use the simple system of equations (3)–(5). As Table 3 shows, however, if instead of having isomorphous replacement data it is assumed that the structure of the

anomalous scatterers is known, a somewhat less-accurate result is obtained from use of the simple system of equations despite the fact that the number of unknown quantities remains the same. The reason for this is the need to use a quadratic equation in addition to the linear ones, giving a twofold ambiguity which is accurately resolved by isomorphous-replacement data, but somewhat less accurately so with use of the known structure of the anomalous scatterers.

Another question concerning the system of algebraic equations concerns whether greater accuracy is obtained in the calculation of the phase differences when exact values of the structure factors for the anomalous scatterers are used, as obtained from their known structure, rather than the statistical values defined in (13). As seen in Table 4, it is better to use the statistical values of the structure-factor magnitudes when significant errors are present in the data.

Finally, the question concerning whether (20) and (21) could be used with rather accurate statistical estimates of $|F_h^n|$ from (23) to give more accurate values for phase differences than alternative algebraic calculations is answered in Table 5. The accuracy of the values is about the same as that obtained in alternative systems of equations. It is also obvious from Table 5 that more-accurate values of $|F_h^n|$ would lead to potentially accurate values for phase differences. More-accurate values for $|F_h^n|$ would come from experiments that included shorter wavelengths.

This article has been devoted, for the most part, to one-wavelength anomalous-dispersion data. It has revealed some further characteristics of the information available which imply that, for some systems, a one-wavelength experiment may suffice to yield a solution. This is already known to be true from experience. It would seem, however, that unless it is quite expedient to work with one-wavelength data, the addition of isomorphous-replacement and multiple-wavelength information would be quite beneficial.

References

HAO, Q. & WOOLFSON, M. M. (1989). *Acta Cryst.* A45, 794–797.
HENDRICKSON, W. A. & TEETER, M. M. (1981). *Nature (London)*, 290, 107–113.
KARLE, J. (1980). *Int. J. Quantum Chem.* 7, 357–367.
KARLE, J. (1984). *Acta Cryst.* A40, 1–4.
KARLE, J. (1985a). *Acta Cryst.* A41, 182–189.
KARLE, J. (1985b). *Acta Cryst.* A41, 387–394.
KARLE, J. (1986). *Acta Cryst.* A42, 246–253.
KARLE, J. (1989). *Acta Cryst.* A45, 303–307.
OKAYA, Y., SAITO, Y. & PEPINSKY, R. (1955). *Phys. Rev.* 98, 1857–1858.
SHOTTON, D. M. & WATSON, H. C. (1970). *Nature (London)*, 225, 811–816.
TIMKOVITCH, R. & DICKERSON, R. E. (1976). *J. Biol. Chem.* 251, 4033–4046.
WANG, B. C. (1985). In *Diffraction Methods in Biological Macromolecules*, edited by H. WYCKOFF, pp. 90–112. New York: Academic Press.
WATSON, H. C., SHOTTON, D. M., COX, J. M. & MUIRHEAD, H. (1970). *Nature (London)*, 225, 806–811.

Continuous Diffuse Scattering from Polymethylene Chains – an Electron Diffraction Study of Crystalline Disorder

BY DOUGLAS L. DORSET AND HENGLIANG HU

Electron Diffraction Department, Medical Foundation of Buffalo, Inc., 73 High Street, Buffalo, NY 14203, USA

AND J. JÄGER

Abteilung Elektronenmikroskopie, Fritz-Haber-Institut der Max-Planck-Gesellschaft, Faradayweg 4-6, D-1000 Berlin 33, Germany

Abstract

Continuous diffuse scattering is noted in electron diffraction patterns from polymethylene compounds such as *n*-paraffins and polyethylene. In a projection down the chain axes, experimentally produced by solution crystallization, the diffuse scatter in *hk*0 patterns disappears at low temperature, in accord with a thermal-diffuse-scattering model, which explains the intensity distribution and its temperature dependence. For a projection onto the chain axes, experimentally achieved by epitaxic orientation on benzoic