

The Relative Scaling of Multiple-Wavelength Anomalous Dispersion Data

BY J. KARLE

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

(Received 9 May 1983; accepted 27 July 1983)

Abstract

A formula is presented for the calculation of the values of the $|F_h^n|$, the structure-factor magnitudes in the absence of anomalous dispersion, from the observed structure-factor magnitudes, $|F_{\lambda,h}|$ and $|F_{\lambda,\bar{h}}|$, that are affected by anomalous dispersion and hence vary with wavelength. Since the values of the $|F_h^n|$ are independent of wavelength, the calculation can provide the basis for relatively scaling data that are collected at different wavelengths. Test calculations of the formula with use of accurate data indicate that the formula has the potential to scale relatively experimental data in multiple-wavelength anomalous-dispersion experiments involving heavy-atom-substituted macromolecules to accuracies of 1 or 2% or better.

Introduction

In the collection of anomalous dispersion data at a variety of wavelengths, the relative scale of the data can be variously affected, for example, by absorption and changes in the incident-beam intensity, to such an extent that the data would not be usable for theoretical formulas that require the simultaneous use of multiwavelength data. Examples of such formulas are the simultaneous equations that arise from the exact algebraic analysis of multiwavelength anomalous-dispersion experiments (Karle, 1980) and methods for evaluating triplet phase invariants from anomalous dispersion data (Karle, 1984). The availability of a method for facilitating the appropriate rescaling could play a crucial role in making practical the new theoretical developments.

A formula will be presented that provides the basis for the required rescalings. It is used to calculate $|F_h^n|$, the magnitude of the structure factor in the absence of anomalous dispersion, from experimental data. Although $|F_h^n|$ is independent of wavelength, it should be noted that the experimental data from which this quantity is calculated may be quite dependent upon the wavelength because of anomalous-dispersion effects.

A procedure for performing the relative scaling of multiwavelength anomalous dispersion data can be based on the independence of wavelength of the $|F_h^n|$.

It will be seen from test calculations that 1 to 2% accuracy or better is potentially achievable.

Calculation of $|F_h^n|$ and tests

Computation of $|F_h^n|$

A formula for the computation of $|F_h^n|$ from experimental structure-factor magnitudes, $|F_{\lambda,ph}|$ and $|F_{\lambda,p\bar{h}}|$, is presented in a heuristic fashion,

$$|F_h^n| \approx 0.5 W_{\lambda,ph} (|F_{\lambda,ph}| + |F_{\lambda,p\bar{h}}|), \quad (1)$$

where λ_p represents any particular wavelength,

$$W_{\lambda,ph} = \left\{ \frac{\sum_{j=1}^{N_{\text{non}}} f_{jh}^2 + \sum_{j=1}^{N_{\text{ano}}} (f_{jh}^{\prime\prime})^2}{\sum_{j=1}^{N_{\text{non}}} f_{jh}^2 + \sum_{j=1}^{N_{\text{ano}}} [(f_{jh}^{\prime\prime} + f_j^{\prime})^2 + f_j^{\prime\prime 2}]} \right\}^{1/2}, \quad (2)$$

$f_{jh}^{\prime\prime}$ is the normal part of the atomic scattering factor for the j th atom and f_j^{\prime} and $f_j^{\prime\prime}$ are the real and imaginary corrections, respectively. The quantity N_{non} is the number of nonanomalously scattering atoms and N_{ano} is the number of anomalously scattering atoms in a unit cell. The implication of (1) is therefore that $|F_h^n|$ is approximately the average of $|F_{\lambda,ph}| + |F_{\lambda,p\bar{h}}|$ corrected for the additional contributions of the f_j^{\prime} and $f_j^{\prime\prime}$ by taking the ratio of the average value for $|F_h^n|$ (Wilson, 1949) and the average value of the average of $|F_{\lambda,ph}| + |F_{\lambda,p\bar{h}}|$.

Numerical tests

Numerical tests have been performed on model substances for which the relative contribution of the anomalous scattering ranges over amounts that correspond roughly to those that are normally encountered in heavy-atom derivatives of macromolecules. The two substances and modifications thereof on which the test calculations were based are magnesium hexapeptide, $C_{48}H_{68}N_{12}O_{12}Mg.CIO_4.4CD_3CN$ (Karle & Karle, 1981a) and quinidine sulfate, $(C_{20}H_{25}N_2O_2)_2SO_4.2H_2O$, (Karle & Karle, 1981b). These model substances were further modified by the theoretical substitution of different atoms for the predominant anomalously scattering atoms.

The results of the test computations are shown in Tables 1, 2 and 3. Modifications of the chemical composition of the test substances in individual tests are indicated. The average percentage error, ϵ , was obtained from comparing $|F_h^n|_{\text{calc}}$, computed from (1), with $|F_h^n|_{\text{corr}}$, the correct value computed from the known structure, by

$$\epsilon = \nu^{-1} \sum_h \frac{||F_h^n|_{\text{corr}} - |F_h^n|_{\text{calc}}|}{|F_h^n|_{\text{corr}}} \times 100, \quad (3)$$

where ν is the number of reflections included in the summation.

For test calculations shown in Table 1, the average errors were computed for various numbers of data ranging from the 200 to 2000 largest values of $|F_{\lambda h}|$. For each $|F_{\lambda h}|$, the corresponding $|F_{\lambda \bar{h}}|$ was obtained and $|F_h^n|_{\text{calc}}$ was then determined by use of (1). It is seen in Table 1 that the average percentage error is greatest with Cr $K\alpha$ radiation, reaching 1.32% for quinidine sulfate and 3.13% when an I atom replaces the S atom, as the number of $|F_h^n|$ computed reaches 2000. The full data set was comprised of 2869 $|F_{\lambda h}|$. This number represents the number of data that would be present in an independent set in the absence of anomalous dispersion. When anomalous dispersion is significant, there are twice as many independent data since there is an independent $|F_{\lambda \bar{h}}|$ for each $|F_{\lambda h}|$. Throughout this paper, 'number of data in the data set' refers to the number of independent $|F_{\lambda h}|$ and ignores the corresponding $|F_{\lambda \bar{h}}|$. It therefore also refers to the maximum number of $|F_h^n|$ that could be computed from the data set by use of (1). For quinidine sulfate, the two thousandth $|F_{\lambda h}|$ was approximately 200 times smaller than the largest value and, for the iodine substitution, it was about 100 times smaller.

The calculations in Table 2 were performed for those values of h that are associated with the largest differences $||F_{\lambda h}| - |F_{\lambda \bar{h}}||$ rather than the largest $|F_{\lambda h}|$ as in Table 1. It is seen that the number of data in the data set and the number of $|F_h^n|$ that were computed were varied. The errors listed in Table 2 are a fraction of one percent except in the case of the iodine substitutions at the Cr $K\alpha$ wavelength. There is a decrease in the percentage errors as the number of $|F_h^n|$ computed is decreased. An increase in the percentage error is observed as the number of data in the full data set is decreased. There is also the indication, in comparing the results of Table 2 with those of Table 1, that a decrease in average error would accrue from setting a lower limit on the values of $|F_{\lambda h}|$ and $|F_{\lambda \bar{h}}|$ used for computing $|F_h^n|$ from (1), when basing the selection criterion on the largest values for the differences, $||F_{\lambda h}| - |F_{\lambda \bar{h}}||$. For example, the calculations in rows 5 and 6 of Table 2 were repeated with a lower limit for $|F_{\lambda h}|$ and $|F_{\lambda \bar{h}}|$ of 4000. In this case, instead of average errors of 1.83 and

Table 1. Average percentage error at various wavelengths in the calculation of $|F_h^n|$ by use of (1) for test substances formed from quinidine sulfate

Selection of the $|F_h^n|$ computed was based on the largest values for $|F_{\lambda h}|$. There were 2869 $|F_{\lambda h}|$ in the data set.

Number of $ F_h^n _{\text{calc}}$ computed	Source of anomalous dispersion	Average percentage error		
		Cr $K\alpha$	Cu $K\alpha$	Ag $K\alpha$
200	1 S atom	0.53	0.50	0.42
400	1 S atom	0.56	0.53	0.46
600	1 S atom	0.66	0.63	0.54
800	1 S atom	0.74	0.71	0.60
1000	1 S atom	0.82	0.78	0.66
1500	1 S atom	1.02	0.97	0.82
2000	1 S atom	1.32	1.26	1.06
200	1 I atom*	1.54	0.45	0.50
400	1 I atom*	1.75	0.44	0.49
600	1 I atom*	1.85	0.42	0.48
800	1 I atom*	1.93	0.40	0.54
1000	1 I atom*	2.01	0.41	0.60
1500	1 I atom*	2.28	0.49	0.77
2000	1 I atom*	3.13	0.69	1.04

* I replaces S in quinidine sulfate.

Table 2. Average percentage error at various wavelengths in the calculation of $|F_h^n|$ by use of (1) for test substances formed from quinidine sulfate (Q_2SO_4) and magnesium hexapeptide (MgHex)

Selection of the $|F_h^n|$ computed was based on the largest differences $||F_{\lambda h}| - |F_{\lambda \bar{h}}||$.

Substance	Number of data in data set	Number of $ F_h^n _{\text{calc}}$ computed	Source of anomalous dispersion	Average percentage error		
				Cr $K\alpha$ (%)	Cu $K\alpha$ (%)	Ag $K\alpha$ (%)
Q_2SO_4	2869	200	1 S atom	0.59	0.56	0.47
Q_2SO_4	2869	400	1 S atom	0.65	0.62	0.53
Q_2SO_4	550	100	1 S atom	0.56	0.54	0.46
Q_2SO_4	550	200	1 S atom	0.80	0.78	0.67
Q_2SO_4	2869	200	1 I atom*	1.83	0.41	0.58
Q_2SO_4	2869	400	1 I atom*	2.07	0.44	0.63
Q_2SO_4	550	100	1 I atom*	2.09	0.28	0.59
Q_2SO_4	550	200	1 I atom*	3.27	0.46	0.83
Q_2SO_4	2869	200	2 I atoms†	1.22	0.37	0.57
Q_2SO_4	2869	200	2 Mg atoms‡	0.66	0.40	0.06
MgHex	1000	200	1 Cl atom‡	0.79	0.75	0.18
MgHex	4612	200	1 I atom‡	2.27	0.40	0.54
MgHex	2869	200	1 I atom‡	2.35	0.33	0.54
MgHex	1000	200	1 I atom‡	3.26	0.69	0.81

* I replaces S in the structure.

† One atom replaces S and the second one replaces O(12) in the structure.

‡ The source atom replaces Mg in the structure.

2.07, the average errors were found to be 1.72 and 1.77, respectively.

A calculation was performed to illustrate the effectiveness of the factor W in (1) and the results are shown in Table 3. The calculations were applied to a set of 2869 data for quinidine sulfate and the modification of it in which an iodine atom replaces a sulfur atom. The results show some dramatic improvements with the use of W in the case of the iodine substitution. There is even some modest improvement when sulfur is the source of anomalous dispersion despite the fact that in this case the values of W differ from unity by only a few thousandths. It is also observed that the percentage error is very low for the iodine

Table 3. Comparison calculation of the average percentage error in the calculation of $|F_h^n|$ by use of (1) for test substances formed from quinidine sulfate

The results obtained when W is included in (1) are compared with those obtained when W is omitted from (1). As in Table 2, selection of the $|F_h^n|$ computed was based on the largest differences $\| |F_{\lambda h}| - |F_{\lambda \bar{h}}| \|$.

Number of $ F_h^n _{\text{calc}}$ computed	Source of anomalous dispersion	W in (1)	Average percentage error		
			Cr $K\alpha$	Cu $K\alpha$	Ag $K\alpha$
400	1 S atom	Present	0.65	0.62	0.53
400	1 S atom	Absent	0.95	0.92	0.78
400	1 I atom*	Present	2.07	0.44	0.63
400	1 I atom*	Absent	7.94	0.31	2.54

* I replaces S in the structure.

substitution with Cu $K\alpha$ radiation and, in fact, the average error is slightly smaller when W is omitted. This occurs because at the wavelength of Cu $K\alpha$ radiation internal compensations make the value of W very close to unity.

Procedure for rescaling

It is assumed that, within a data set measured at any particular wavelength, $|F_{\lambda p h}|$ and $|F_{\lambda p \bar{h}}|$ are properly relatively scaled and that the data for, at least, one wavelength, perhaps the smallest wavelength, have been subjected to the usual corrections. Thus, it is assumed that the data at the selected wavelength are corrected for absorption, Lorentz and polarization effects and, in addition, thermal motion and, perhaps, scale. In the case of crystals in capillaries, it would be necessary to make initial corrections for, at least, the absorption of the capillary and solvent when present at all wavelengths in order to ensure that $|F_{\lambda p h}|$ and $|F_{\lambda p \bar{h}}|$ would be properly relatively scaled within a data set. Such corrections have been considered by Wells (1960) and have been the subject of a recent study by DeTitta (1983).

We now modify (1) to incorporate the possibility that the $|F_{\lambda p h}|$ that derive from experiment, $|F_{\lambda p h}|_{\text{exp}}$, may not have the desired scaling,

$$|F_h^n| = 0.5 S_p(\mathbf{h}) W_{\lambda p h} (|F_{\lambda p h}|_{\text{exp}} + |F_{\lambda p \bar{h}}|_{\text{exp}}), \quad (4)$$

where $S_p(\mathbf{h})$ is a scale factor associated with data obtained at wavelength λ_p . We may use the data set at $p = 1$ as the set to which all the remaining data are to be relatively scaled. By noting then that $S_1 = 1$ and that $|F_h^n|$ is independent of λ_p , we have, from (4),

$$S_p(\mathbf{h}) = \frac{W_{\lambda_1 h} (|F_{\lambda_1 h}|_{\text{exp}} + |F_{\lambda_1 \bar{h}}|_{\text{exp}})}{W_{\lambda p h} (|F_{\lambda p h}|_{\text{exp}} + |F_{\lambda p \bar{h}}|_{\text{exp}})}. \quad (5)$$

The desired result, i.e. to have all data properly relatively scaled, is obtained by taking the product $S_p(\mathbf{h}) |F_{\lambda p h}|_{\text{exp}}$.

For the case of one type of predominant anomalous scatterer, an alternative method for relatively scaling

intensity data measured at different wavelengths can be based on an exact relationship that is easily derived from equation (15) of Karle (1980),

$$\frac{|F_{\lambda_1 h}|^2 - |F_{\lambda_1 \bar{h}}|^2}{|F_{\lambda_2 h}|^2 - |F_{\lambda_2 \bar{h}}|^2} = \frac{f''_{\lambda_1 2}}{f''_{\lambda_2 2}}, \quad (6)$$

where $f''_{\lambda_p 2}$ is the imaginary correction to the atomic scattering factor of the anomalous scatterer at some particular wavelength, λ_p , and the 2 in the subscript labels the type of atom, in this case a particular type of anomalous scatterer. If the intensities are scaled properly at λ_1 , (6) permits proper scaling of the intensities at λ_2 . For certain cases involving more than one type of predominant anomalous scatterer, it is conceivable that $f''_{\lambda_1 2}$ and $f''_{\lambda_2 2}$ could be replaced in (6) by weighted averages of the approximate f'' . This matter has many facets and is left for future study.

Concluding remarks

The test substances, quinidine sulfate with a sulfur atom or an iodine atom substituting for the sulfur atom as sources for anomalous scattering, can serve as approximate models over the range of anomalous scattering that occurs in heavy-atom-substituted macromolecules. The iodine-substituted structure is somewhat extreme. Thus, the relative amount of anomalous scattering in heavy-atom-substituted macromolecules is usually in the middle of the range, perhaps closer to that of the sulfur compound. It follows, therefore, from the test calculations summarized in Tables 1 and 2 that, for macromolecules with heavy-atom substituents, formula (1) has the potential to scale relatively experimental data in multiple-wavelength anomalous-dispersion experiments to accuracies of 1 or 2% or better. The ultimate possible accuracy depends upon the relative amount of anomalous scattering that occurs in the experiments as well as the accuracy of the data.

In applying the procedure for relatively scaling multiple-wavelength anomalous-dispersion data, it is important that the data collected at the selected reference wavelength be of high quality and properly corrected, i.e. for Lorentz and polarization effects and absorption. For macromolecules, it is generally necessary to enclose them in capillaries in the environment of mother liquor. This adds difficulties to the making of absorption corrections. As noted above, however, this matter is currently receiving attention. It is also important that at all wavelengths $|F_{\lambda h}|$ and $|F_{\lambda \bar{h}}|$ be properly relatively scaled.

In order to compute W in (1), it is necessary to know the chemical composition and site occupancy of the substance of interest. Even if this information is only approximately known, it should be helpful to use it, as implied by Table 3.

Further evaluation of the usefulness of the rescaling procedure in multiple-wavelength anomalous-disper-

sion experiments awaits the undertaking of such experiments.

The programs for the computations performed in this paper were written by Mr Stephen Brenner. I am very grateful for his cooperation.

This research was supported in part by USPHS grant GM30902.

Acta Cryst. (1984). A40, 4–11

Rules for Evaluating Triplet Phase Invariants by Use of Anomalous Dispersion Data

JEROME KARLE

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

(Received 12 April 1983; accepted 5 August 1983)

Abstract

On the basis of some mathematical and physical characteristics of anomalous dispersion experiments, it has been possible to derive simple rules, $R_{\text{ano},1}$, $R_{\text{ano},2}$ and $R_{\text{ano},3}$, that permit the selection of triplet phase invariants, in a single-wavelength experiment, that have values close to $\pi/2$, $-\pi/2$, 0 and other values in the range from $-\pi$ to π . The rules, $R_{\text{ano},1}$, $R_{\text{ano},2}$ and $R_{\text{ano},3}$, apply to the case of a single type of predominant anomalous scatterer. The simple generalization to more than one type of predominant anomalous scatterer is also described. Test examples show that large numbers of invariants may be evaluated by these means with reliabilities that are potentially high, but depend, of course, on the reliability of the experimental data. The only information required besides the measurements of the diffraction intensities is the chemical composition of the anomalously scattering atoms. In some cases, even this information is not required if two alternative sets of estimates of the values of the triplet phase invariants are considered.

Introduction

Anomalous dispersion experiments afford the opportunity to evaluate triplet phase invariants in single-wavelength experiments. Formulas for accomplishing this have already been described by Heinerman, Krabbendam, Kroon & Spek (1978) and by Hauptman (1982). The formulas of Heinerman *et al.* (1978) were derived from probabilistic arguments and give the sines of the invariants. The twofold ambiguity inherent in the sine of an invariant requires auxiliary

References

- DETTITA, G. T. (1983). *Am. Crystallogr. Assoc. Meet.*, 14–18 March, Columbia, Missouri. Abstract PA8.
 KARLE, I. L. & KARLE, J. (1981a). *Proc. Natl Acad. Sci. USA*, **78**, 681–685.
 KARLE, I. L. & KARLE, J. (1981b). *Proc. Natl Acad. Sci. USA*, **78**, 5938–5941.
 KARLE, J. (1980). *Int. J. Quantum Chem. Symp.* **7**, 357–367.
 KARLE, J. (1984). *Acta Cryst.* **A40**, 4–11.
 WELLS, M. (1960). *Acta Cryst.* **13**, 722–726.
 WILSON, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.

information for its resolution. The condition used in test calculations was to select the value closest to zero. The formulas of Hauptman (1982) were derived from use of the conditional joint probability distribution and can yield unique estimates for the invariants in the whole interval $-\pi$ to $+\pi$. The method of derivation and type of result to be described in this article are quite different from those in the previous studies. They are based on an analysis of some particular mathematical and physical characteristics of the data from anomalous dispersion experiments, in a fashion that is quite comparable to an analysis recently carried out for isomorphous replacement experiments (Karle, 1983).

The characteristics of interest concern observations related to the differences of the magnitudes of selected types of structure factors and also the expected values of triplet phase invariants associated with the structure of the anomalous scatterers. The only information required of the predominant anomalous scatterers in this approach is their chemical nature. It is not necessary to know the number of positions occupied or the occupancy. The results to be obtained here are simple rules, based on several types of structure-factor magnitudes, for selecting triplet phase invariants whose values are near some particular values.

Test examples performed by Heinerman *et al.* (1978) and Hauptman (1982) show that their formulas have the potential for reliable results. Similarly, it will be seen from the results of test calculations in this article that the rules to be developed here also have the potential for yielding reliable evaluations of very large numbers of triplet phase invariants.