Acta Cryst. (1982). A**38**, 577–590

Normalized Structure Factors. I. Choice of Scaling Function

BY V. SUBRAMANIAN† AND S. R. HALL

Crystallography Centre, University of Western Australia, Nedlands 6009, *Australia*

(*Received* 6 *May* 1981; *accepted* 20 *January* 1982)

Abstract

The estimation of normalized structure factors depends both on the method used to scale measured intensities and on the expectation values for those intensities. In this study different scaling functions are examined with respect to the 'true' normalized structure factors calculated for a range of refined crystal structures. The exponential scale $\mathbf{k}\exp(Bs^2)$, which includes an overall rescale factor, combined with the random-atom expectation value, is shown to provide the most reliable triplet and quartet structure-invariant relationships.

Introduction

In the application of direct methods the reliability of structure-invariant phase relationships is assumed to be directly related to the magnitude of the normalized structure factor $|E_\mathbf{h}|$. However, the relative importance of the *precision* of normalized structure factors to the success of structure-invariant phasing procedures is not well understood. The conditional probabilities of structure-invariant relationships are largely based on the magnitude of the quasi-normalized structure factor

$$|\mathscr{E}'_\mathbf{h}| = \frac{\sum^N f_j \exp(2\pi\mathbf{h}.\mathbf{x}_j)}{\left[\sum^N f_j^2\right]^{1/2}}, \qquad (1)$$

while the normalized structure factor used in phasing procedures is estimated with the expression

$$|E_\mathbf{h}| = \frac{|F_\mathbf{h}^m|\ \mathbf{K}(\mathbf{h})}{\langle|F_\mathbf{h}^2|\rangle^{1/2}}, \qquad (2)$$

where $|F_\mathbf{h}^m|$ is the measured structure factor, $\mathbf{K}(\mathbf{h})$ is a structure-factor scaling function and $\langle|F_\mathbf{h}^2|\rangle$ is the atoms-at-rest expectation value of $|F_\mathbf{h}^2|$. The importance of $|\mathscr{E}'_\mathbf{h}|$ to the theoretical distribution of the structure invariants suggests that an essential criterion for the 'correctness' of an estimated $|E_\mathbf{h}|$ value is its similarity to the 'true' normalized structure factor

$$|\mathscr{E}_\mathbf{h}| = |\mathscr{E}'_\mathbf{h}|/\varepsilon^{1/2}.$$

This assumption is the basis for an earlier study of normalization procedures by Ladd (1978). The most critical measure of an estimated normalized structure factor is the reliability of the structure-invariant relationships they form with each other. This is the principal test applied to normalized structure factors estimated in this study.

The normalized structure factor squared $|E_\mathbf{h}^2|$ is essentially the ratio of scaled measured intensity $|I_\mathbf{h}^m|$ to the intensity expected for a random distribution of N atoms at rest. If a particular measured intensity is significantly different from its expected random-atom value, then the value of $|E_\mathbf{h}^2|$ will depart significantly from 1·0. The magnitude of this departure indicates the sensitivity of that reflection to non-random aspects of the structure, or, in other words, to the 'presence of structure'. Reflections with $|E_\mathbf{h}^2|$ values much greater than 1·0 are of particular importance in direct methods because of their dominance as Fourier transform coefficients. Structure-factor magnitudes close to zero also contain significant structural information but because their phases are not as well defined, and their contribution to the Fourier transform is small, they are not generally used in phasing procedures.

The different approaches currently in use for evaluating and applying the $\mathbf{K}(\mathbf{h})$ and $\langle|F_\mathbf{h}^2|\rangle$ components in the normalized structure-factor equation (2) are examined in this study. There are at least three ways of formulating the scaling function $\mathbf{K}(\mathbf{h})$. The first is the *K*-curve approach of Karle & Hauptman (1953) which evaluates $\mathbf{K}(\mathbf{h})$ as the monotonically decreasing function best fitting the variation of $|F_\mathbf{h}^m|/\langle|F_\mathbf{h}^2|\rangle^{0.5}$ with s^2. A second method, which is closely allied to the first is to represent $\mathbf{K}(\mathbf{h})$ as the exponential scale, $\mathbf{k}\exp(Bs^2)$; and a third approach allows for non-monotonic variations of $\mathbf{K}(\mathbf{h})$ with s^2. Supplementary to these basic approaches, the scaling function $\mathbf{K}(\mathbf{h})$ is usually adjusted to ensure that the mean $|E_\mathbf{h}^2|$ is 1·0 for either all data, or for different reflection classes (Dewar, 1970).

† Deceased 27 December 1981.

0567-7394/82/050577-14$01.00

Limited tests on several scaling functions have already been reported by Ladd (1978). He concludes that a *minimum-K*-curve approach produces the best $|E_h|$ values; followed closely by the exponential scale $\mathbf{k} \exp(Bs^2)$. These findings were based on the correspondence between the estimated $|E_h|$ and $|\mathscr{E}_h|$ values. The validity of this conclusion makes the fundamental assumption that $|\mathscr{E}_h|$ values provide the most reliable phase relationships. This is not unreasonable since $|\mathscr{E}'_h|$ is the basis in the probabilistic methods of structure-invariant theory. Nevertheless, this theory has been derived largely for $P1$ and $P\bar{1}$ atomic distributions and it remains to be shown that $|\mathscr{E}_h|$ values necessarily provide the *most* reliable phase relationships for higher symmetries, and for a range of structure types. This study examines the reliability of $|\mathscr{E}_h|$ in terms of the precision of the structure-invariant relationships, for a variety of test structures.

Also encompassed in this study is an analysis of alternative approaches to calculation of the atoms-at-rest expectation value $\langle |F_h^2| \rangle$. In the absence of any structural knowledge this expectation value is assumed to be the random-atom approximation,

$$\langle |F_h^2| \rangle = \varepsilon \sum_j^N f_j^2, \tag{3}$$

where N is the number of atoms in the cell.

If conformational information on a fragment of the structure is available then the expectation value may be expressed as a Debye scattering equation (Debye, 1915)

$$\langle |F_h^2| \rangle = \varepsilon \sum_j^M \sum_k^M f_j f_k \frac{\sin 4\pi s d_{jk}}{4\pi s d_{jk}}, \tag{4}$$

where M is the number of atoms in the fragment.

When the orientation of the molecular fragment is known, the expectation value becomes (Main, 1976)

$$\langle |F_h^2| \rangle = \left(\sum_j^M \sum_k^M \sum_p^P f_j f_k \exp 2\pi i \mathbf{h}.(\mathbf{x}_{jp} - \mathbf{x}_{kp}) \right), \tag{5}$$

where P is the number of point groups in the cell.

When actual atomic coordinate information is available the expectation value is the atoms-at-rest structure factor squared given by

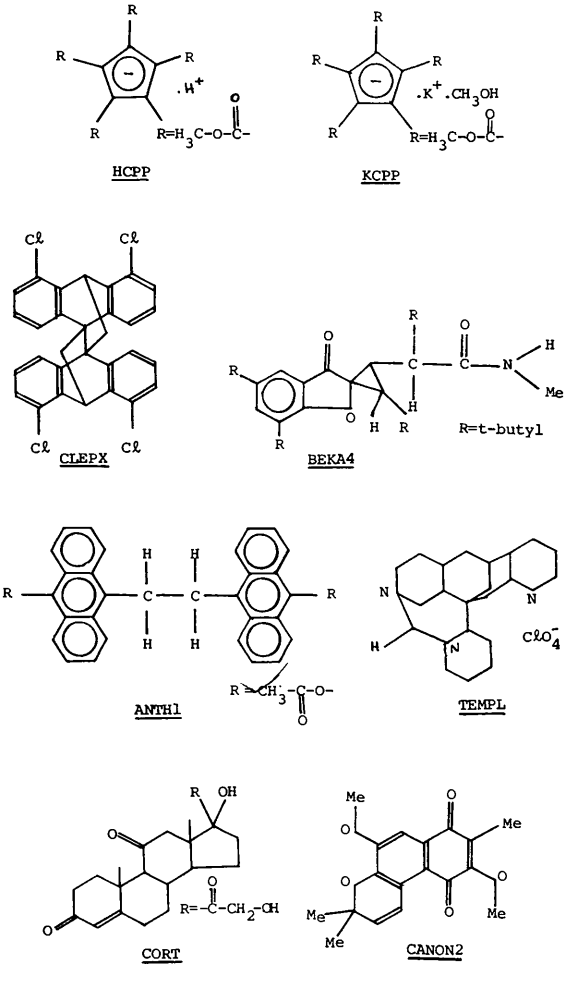
$$\langle |F_h^2| \rangle = \left(\sum_j^M \sum_q^S f_j \exp 2\pi i \mathbf{h}.\mathbf{x}_{jq} \right)^2, \tag{6}$$

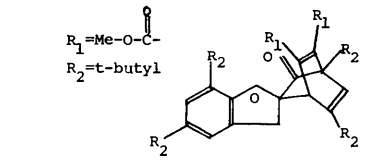
where S is the number of equivalent positions.

It would seem to follow from (3)–(6) that using the expectation value with highest information content in the normalization process provides improved estimates of normalized structure factors. Such a rationale presumes that knowledge of the structure provides more precise estimates of the 'true' normalized structure factor and more reliable structure-invariant relationships. There is little direct evidence for this and Ladd (1978) has shown that the Debye expectation value reduces rather than enhances the precision of the estimated $|E_h|$. An explanation for this is that when the random-atom expectation value is used, the $|E_h^2|$ values differing significantly from 1·0 will contain more (non-random) structural information. On the other hand, if the Debye expectation value (4) is used, structural contributions due to short-range translational symmetry tend to be suppressed in the resulting $|E_h|$ values. In this study we examine these conclusions in terms of the precision of the structure-invariant relationships.

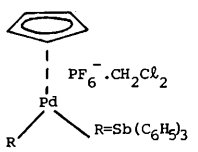
Test structures

The refined structures of eleven organic and inorganic compounds are used in this study to test the different scaling functions. Details of these structures are given in Table 1 and their molecular configurations are shown below.

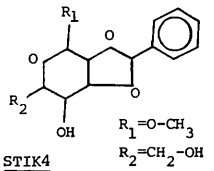




K22BR



PDCPS



STIK4

The test structures were selected to provide a range of space groups, centricity, and chemical composition. Their data sets range from extensive ($s^2_{max} = 0{\cdot}42$) to limited ($s^2_{max} = 0{\cdot}22$), and represent data collected from both strongly and weakly diffracting crystals. All structures were refined by conventional least-squares techniques to R values given in Table 1.

HCPP is a penta-substituted cyclopentadiene derivative, and KCPP is its potassium salt. The former crystallizes in $P\bar{1}$ and the latter in *Pcab*. CLEPX and BEKA4 crystallize in $P\bar{1}$. The former is a small molecule with a chlorine atom and the latter has two molecules per asymmetric unit. STIK4 is a light-atom structure in the polar space group $P2_1$. PDCPS, CANON2, and ANTH1 belong to $P2_1/c$. PDCPS is a heavy-atom structure with an extensive data set. CANON2 is a highly ordered structure with a predominant hexagonal motif and has a limited data set. It also has a history of direct-method failures (Hall, Raston & White, 1978). ANTH1 is hypercentric. CORT, a steroid, and TEMPL, a natural product, both belong to $P2_12_12_1$. The CORT data are supplied as the standard light-atom test deck for the *MULTAN*80 system (Main *et al.*, 1980). TEMPL contains a heavy atom. K22BR is a light-atom structure in space group *Iba*2 and has a very limited data set.

The data of test structures were processed with the normalization routine *EVAL*, the structure-invariant generation routine *SINVAR* and the phase-analysis routine *PHISIN* (Hall, 1978). Another routine *ESCAN* (Hall & Subramanian, 1980) was used to analyse the dependency of estimated $|E_h|$ values and the calculated $|\mathscr{E}_h|$ on a variety of parameters.

Comparison of scaling functions

Two types of scaling functions are used widely in normalization procedures. One is the exponential scale $\mathbf{k}\exp(Bs^2)$, often referred to as the 'linear' scale because of its linear relationship in a Wilson plot (Wilson, 1942), and the other is the 'profile scale'.

The profile scale describes the variation of the mean values of $|F_h^2|/\langle|F_h^2|\rangle$ with s^2. If this scale takes the form of a monotonically decreasing function, it is equivalent to the K curve of Karle & Hauptman (1953). In this form the profile scale is often equivalent to the linear scale of the Wilson plot (Karle, 1976). If, however, the profile scale is to account for short-range variations with s^2 due to radial scattering effects it may be represented by a polynomial-type function, as in the case of the *MULTAN K* curve (Main *et al.*, 1980), or as a smooth curve that follows the significant variations in $\ln(|F_h^2|/\langle|F^2|\rangle)$.

In this study, both the exponential and profile scaling functions are examined. The minimum K-curve method of Ladd (1978) was not studied, principally because of its close similarity to the exponential scale but also because there seems less theoretical justification for its use. The radial effects of short-range translational symmetry on the scattering process result in decreases as well as increases of the mean radial intensities defined by a random-atom structure. Since the random-atom structure is the basis for calculating the quasi-normalized structure factor $|\mathscr{E}'_h|$, it is logical that the function $\mathbf{K}(\mathbf{h})$ should attempt to scale values to the

Table 1. *Test structures*

$R = \sum||F_o| - |F_c||/\sum|F_o|$. $\bar{B}$ is the overall temperature factor.

	Formula	Space group	R value	$\bar{B}$ (Å^2)	s^2_{max}	Reference
HCPP	$C_{15}H_{16}O_{10}$	$P\bar{1}$	0·037	4·6	0·29	(*a*)
CLEPX	$C_{30}H_{18}Cl_4$	$P\bar{1}$	0·037	4·0	0·25	(*a*)
BEKA4	$C_{58}H_{90}N_2O_6$	$P\bar{1}$	0·055	4·5	0·24	(*a*)
STIK4	$C_{14}H_{18}O_6$	$P2_1$	0·050	4·8	0·36	(*a*)
PDCPS	$C_{42}H_{37}Cl_2F_6PPdSb_2$	$P2_1/c$	0·051	3·6	0·42	(*a*)
CANON2	$C_{18}H_{18}O_5$	$P2_1/n$	0·058	3·8	0·24	(*b*)
ANTH1	$C_{34}H_{26}O_4$	$P2_1/c$	0·034	4·5	0·36	(*a*)
TEMPL	$C_{21}H_{34}ClN_3O_4$	$P2_12_12_1$	0·056	4·4	0·29	(*a*)
CORT	$C_{21}H_{28}O_5$	$P2_12_12_1$	0·058	3·3	0·32	(*c*)
K22BR	$C_{35}H_{48}O_6$	*Iba*2	0·049	4·8	0·22	(*a*)
KCPP	$C_{16}H_{19}KO_{11}$	*Pcab*	0·042	3·2	0·36	(*a*)

References: (*a*) Skelton & White (1981); (*b*) Hall *et al.* (1978); (*c*) Declercq, Germain & Van Meerssche (1972).

median line between these positive and negative fluctuations; or, when it is required that the Debye scattering effects be reduced in the estimated $|E_h|$ values, the profile form of **K(h)** should be encouraged to follow these fluctuations. The concept of a *minimum K* curve is therefore difficult to justify. Certainly, the *K*-curve method of Karle & Hauptman (1953) uses a mean, not a minimum, fit of the data points and this is equivalent to the exponential scale $\mathbf{k}\exp(Bs^2)$.

The above methods for evaluating **K(h)** do not require that the overall mean value of $|E_h^2|$ be necessarily 1·0. This is, however, a normalization requirement and is easily achieved by rescaling the initial estimates of $|E_h^2|$ calculated with **K(h)** with the factor $\sum m/\sum m|E_h^2|$. When this factor is applied to all data the process is referred to as *overall rescaling*. Alternatively, if it is applied to individual classes of reflections it is referred to as *index rescaling* (Dewar, 1970). Index rescaling ensures that each reflection group has a similar spectrum of large and small $|E_h|$ values and therefore each group has an equivalent influence on the phasing process.

It is well recognized that the application of index rescaling can be of benefit in the analysis of structures with a predominant subunit. The presence of a substructure exaggerates the intensities in certain reflection classes and makes the elucidation of superstructure difficult. However, index rescaling does have its drawbacks. In particular, its use without careful consideration of scaling effects removes from the estimated $|E_h|$ values structural information that is then inaccessible to the phasing process. The validity of routine application of index rescaling and its effect on the reliability of structure-invariant relationships are also considered in this study.

To compare the various scaling options *four different* estimates of $|E_h|$ were used. These estimates were calculated with the program *EVAL* (Hall, 1978).

Estimate $|E_h|_1$ is calculated with the exponential scale function $\mathbf{k}\exp(Bs^2)$ and the random-atom

Table 2. *Comparison of* $|\mathscr{E}(\mathbf{h})|$ *and* $|E_n(\mathbf{h})|$

All values of $|E_n(\mathbf{h})|$ are estimated using the random-atom expectation value except for the values in brackets where the Debye value was employed. ΔE_n^2 and $\Delta E_n^2/\mathscr{E}^2$ are the mean difference and the mean fractional difference between the estimated $|E_n(\mathbf{h})|$ values and the calculated $|\mathscr{E}(\mathbf{h})|$ value. *n*ref is the number of reflections used to calculate the means.

(*a*) For $|\mathscr{E}(\mathbf{h})| > 0$ (all data)

	*n*ref	ΔE_1^2	$\Delta E_1^2/\mathscr{E}^2$	ΔE_2^2	$\Delta E_2^2/\mathscr{E}^2$	ΔE_3^2	$\Delta E_3^2/\mathscr{E}^2$	ΔE_4^2	$\Delta E_4^2/\mathscr{E}^2$
HCPP	2167	0·33	0·54	0·39	0·57	0·36	0·56	0·40	0·59
CLEPX	2138	0·26	0·42	0·30	0·45	0·27	0·43	0·31	0·46
BEKA4	5413	0·39	0·67	0·44	0·69	0·40	0·68	0·45	0·70
STIK4	1099	0·31	0·44	0·36	0·48	0·32	0·45	0·36	0·49
PDCPS	7493	0·32	0·63	0·33	0·64	0·35	0·67	0·35	0·67
CANON2	1652	0·33	0·42	0·43	0·52	0·38	0·50	0·47	0·60
		(0·33)	(0·45)	(0·41)	(0·51)	(0·40)	(0·52)	(0·46)	(0·59)
ANTH1	2196	0·31	0·57	0·38	0·65	0·31	0·57	0·39	0·65
TEMPL	1874	0·35	0·56	0·39	0·61	0·36	0·57	0·40	0·61
CORT	1603	0·29	0·41	0·35	0·49	0·31	0·45	0·37	0·52
		(0·31)	(0·46)	(0·35)	(0·49)	(0·33)	(0·47)	(0·37)	(0·52)
K22BR	1832	0·43	0·60	0·48	0·65	0·48	0·60	0·49	0·66
KCCP	3943	0·36	0·80	0·38	0·81	0·40	0·87	0·41	0·87
Mean		0·34	0·55	0·39	0·60	0·36	0·58	0·40	0·62

(*b*) For $|\mathscr{E}(\mathbf{h})| > 1·0$

	*n*ref	ΔE_1^2	$\Delta E_1^2/\mathscr{E}^2$	ΔE_2^2	$\Delta E_2^2/\mathscr{E}^2$	ΔE_3^2	$\Delta E_3^2/\mathscr{E}^2$	ΔE_4^2	$\Delta E_4^2/\mathscr{E}^2$
HCPP	702	0·69	0·30	0·83	0·34	0·75	0·31	0·86	0·35
CLEPX	700	0·49	0·22	0·61	0·27	0·52	0·23	0·63	0·27
BEKA4	1699	0·80	0·32	0·94	0·37	0·83	0·33	0·97	0·38
STIK4	404	0·52	0·25	0·62	0·29	0·54	0·25	0·63	0·30
PDCPS	2180	0·68	0·28	0·69	0·29	0·74	0·31	0·75	0·32
CANON2	446	0·87	0·31	1·15	0·42	1·01	0·36	1·24	0·44
		(0·85)	(0·34)	(1·08)	(0·40)	(1·05)	(0·37)	(1·20)	(0·43)
ANTH1	574	0·77	0·29	0·99	0·36	0·77	0·30	0·99	0·37
TEMPL	685	0·58	0·28	0·65	0·32	0·58	0·29	0·66	0·32
CORT	549	0·55	0·27	0·65	0·31	0·57	0·28	0·70	0·33
		(0·57)	(0·28)	(0·66)	(0·32)	(0·60)	(0·29)	(0·71)	(0·34)
K22BR	678	0·76	0·36	0·86	0·40	0·76	0·36	0·86	0·40
KCPP	1085	0·68	0·30	0·76	0·33	0·78	0·34	0·83	0·36
Mean		0·67	0·29	0·80	0·34	0·71	0·31	0·83	0·35

expectation value. The values of B (see Table 1) used in this function were obtained from a Wilson-plot fit based on positioned-atom expectation values [see (6)] calculated from the refined atomic coordinates. The scale **k** is evaluated by overall rescaling.

Estimate $|E_{\mathbf{h}}|_2$ is calculated from a profile scale and the random-atom expectation value $\langle|F_{\mathbf{h}}^2|\rangle$. The profile scale takes the form of 41 values of $\mathbf{K}(\mathbf{h})_i$ ($i = 1$ to 41),

$$\mathbf{K}(\mathbf{h})_i = \left[\frac{\sum \varepsilon \sum^{N} f^2}{\sum |F_{\mathbf{h}}^2|}\right]_i, \tag{7}$$

averaged for shells of reciprocal space bounded by equal intervals of s^2 between 0·0 and $s^2_{\max}$. These shells were symmetrically overlapped to minimize the statistical fluctuations and to provide a relatively smooth variation over the 41 ranges. The degree of overlap ensures at least 200 contributing reflections to the mid-range ($i = 20$). The profile scale was applied by a simple linear interpolation of the 41 values of $\mathbf{K}(\mathbf{h})$, and was followed by overall rescaling.

Estimates of $|E_{\mathbf{h}}|_3$ *and* $|E_{\mathbf{h}}|_4$ are calculated using identical procedures to $|E_{\mathbf{h}}|_1$ and $|E_{\mathbf{h}}|_2$, respectively, except that overall rescaling is replaced with index rescaling based on the eight separate *hkl*-parity groups (*eee*, *eeo*, . . . , *ooo*).

Methods for comparing $|E_{\mathbf{h}}|$ *and* $|\mathscr{E}_{\mathbf{h}}|$

The rationale for using the calculated structure factor $|\mathscr{E}_{\mathbf{h}}|$ as a *posterior* measure for the correct normalized structure factor has already been discussed. In this study, each of the four estimates of $|E_{\mathbf{h}}|$ are compared with $|\mathscr{E}_{\mathbf{h}}|$ using the program *ESCAN* (Hall & Subramanian, 1980). The definitions of mean differences ΔE_n and ΔE_n^2 are

$$\Delta E_n = |(|E_{\mathbf{h}}|)_n - |\mathscr{E}_{\mathbf{h}}|| \tag{8}$$

and

$$\Delta E_n^2 = |(|E_{\mathbf{h}}^2|)_n - |\mathscr{E}_{\mathbf{h}}^2||. \tag{9}$$

The values of $\Delta E_{\mathbf{h}}^2$ and $\Delta E_{\mathbf{h}}^2/|\mathscr{E}_{\mathbf{h}}^2|$ are shown in Table 2 for $|\mathscr{E}_{\mathbf{h}}| > 0{\cdot}0$ and $|\mathscr{E}_{\mathbf{h}}| > 1{\cdot}0$. The differences are calculated as the mean of the sum individual differences {*i.e.* $[\sum m\Delta E_{\mathbf{h}}^2]/[\sum m]$ and $[\sum m\Delta E_{\mathbf{h}}^2/|\mathscr{E}_{\mathbf{h}}^2|]/[\sum m]$} for each of the 41 shells of reciprocal space. These show that the mean difference, ΔE_n^2 and $\Delta E_n^2/|\mathscr{E}_{\mathbf{h}}^2|$, between $|\mathscr{E}_{\mathbf{h}}^2|$ and $|\mathscr{E}_{\mathbf{h}}^2|_n$ are smallest for $|E_{\mathbf{h}}|_1$, the estimate based on the exponential scaling function, and an overall rescale. The average mean of the differences for all test structures (Table 2) indicates that the above combination produces $|E_{\mathbf{h}}|$ values closest to $|\mathscr{E}_{\mathbf{h}}|$, whereas the combined profile scale and index rescale produces the worst agreement.

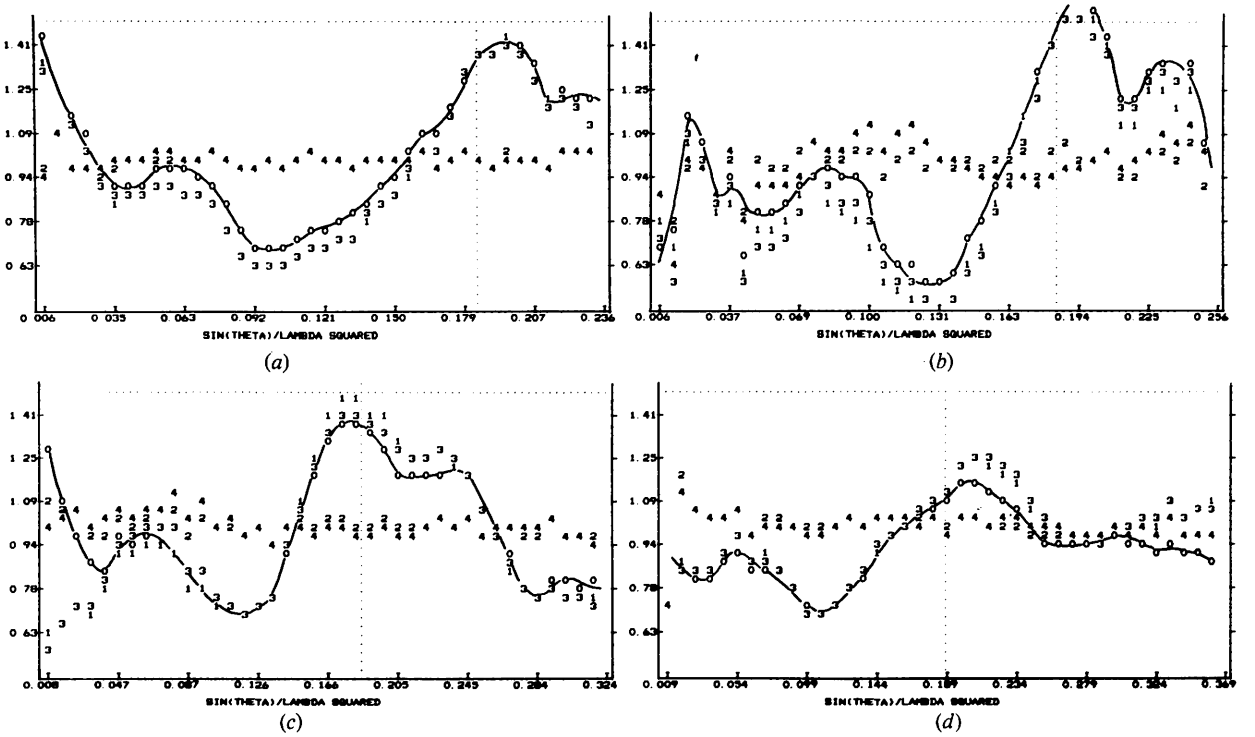
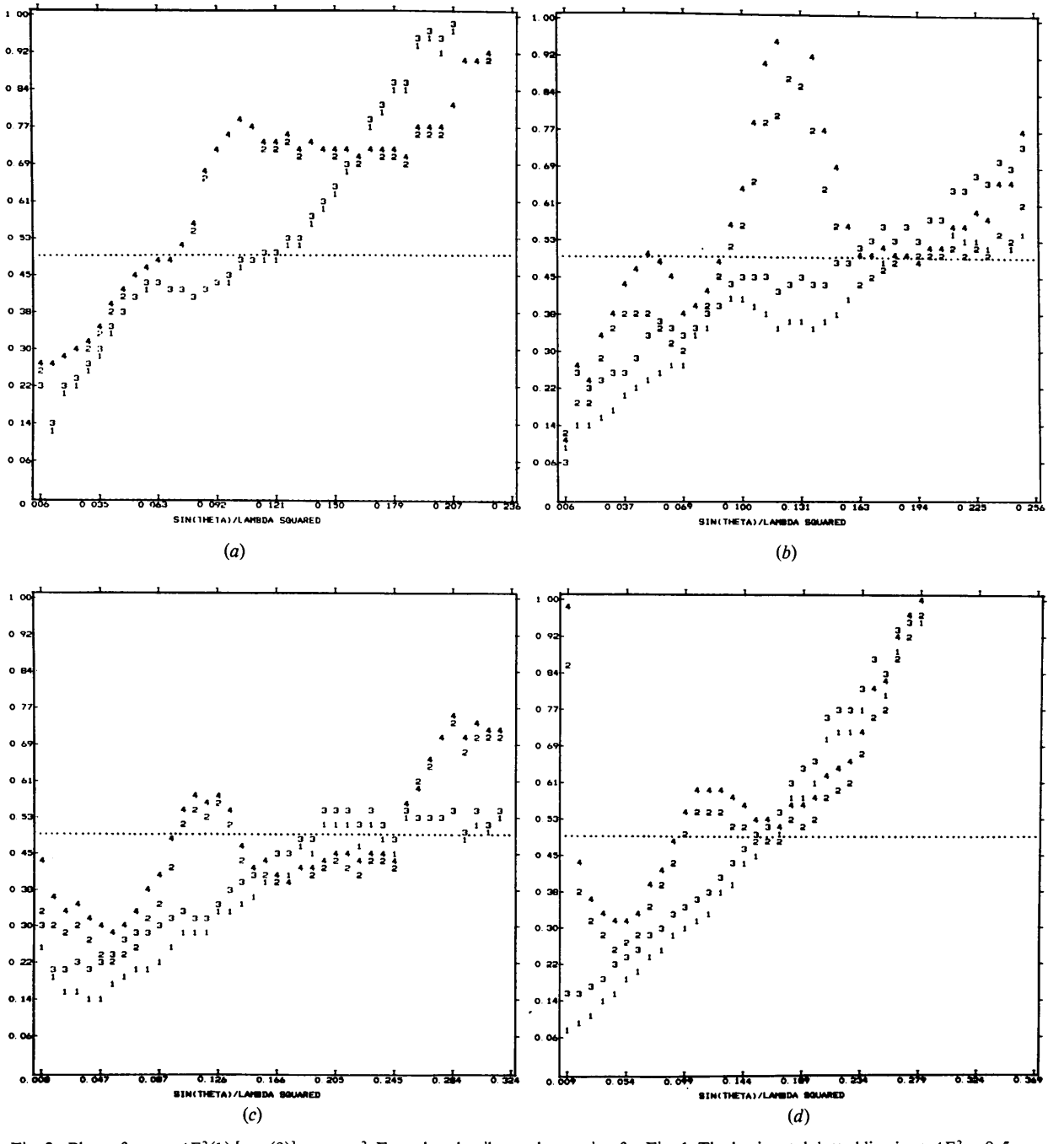


Fig. 1. Plots of $|E_n^2(\mathbf{h})|$ and $|\mathscr{E}^2(\mathbf{h})|$ *versus* s^2 for the structures (*a*) BEKA4, (*b*) CANON2, (*c*) CORT and (*d*) KCPP. The expectation value used in estimates of $|E_n^2(\mathbf{h})|$ is based on the random-atoms-at-rest approximation. The numbers 1–4 correspond to the subscripts of the four different estimates of $|E_n^2(\mathbf{h})|$ (see text). The solid lines connecting the points denoted by '0' represent the mean values of the 'true' normalized structure-factor squared. If data points overlap, the largest number has precedence. The vertical dotted line is at $s^2 = 0{\cdot}18$. The horizontal dotted line is at $|E^2| = 1{\cdot}5$.

The difference between $|\mathscr{E}_{\mathbf{h}}^2|$ and $|E_{\mathbf{h}}^2|_n$ as a function of s^2 is illustrated in Figs. 1, 2, and 3 for the test structures BEKA4, CANON2, CORT, and KCPP. These four test structures were selected as being representative for a range of structural types. Fig. 1 shows the variation of $|\mathscr{E}_{\mathbf{h}}^2|$ with s^2 as a solid line (connecting '0' data points) with different $|E_{\mathbf{h}}^2|$ estimates as data points numbered 1, 2, 3, and 4. In all plots the highest numbered data points have precedence and this means that a 3, for instance, will obscure a 2, 1, or 0 in the same position.

The relative variation of $|\mathscr{E}_{\mathbf{h}}|$ and $|E_{\mathbf{h}}|$ is consistent for all of the eleven test structures. In all cases there is a relatively close agreement between $|\mathscr{E}_{\mathbf{h}}|$, $|E_{\mathbf{h}}|_1$ and



(*a*) (*b*) (*c*) (*d*)

Fig. 2. Plots of mean $\Delta E_n^2(\mathbf{h})$ [see (9)] *versus* s^2. For other details see the caption for Fig. 1. The horizontal dotted line is at $\Delta E^2 = 0{\cdot}5$.

Table 3. *Mean systematic deviation of* $E(\mathbf{h})$ *from* $\mathscr{E}(\mathbf{h})$

All values of $|E_n(\mathbf{h})|$ are estimated using the random atom expectation value.

$$\Delta\bar{E}_n(\mathbf{h}) = \left\{\sum_{l=1}^{41} [|\bar{E}_n(\mathbf{h})| - |\bar{\mathscr{E}}(\mathbf{h})|]_l\right\}/41 \qquad \Delta\bar{E}_n^2(\mathbf{h}) = \left\{\sum_{l=1}^{41} [\bar{E}_n^2(\mathbf{h}) - \bar{\mathscr{E}}^2(\mathbf{h})]_l\right\}/41.$$

(*a*) For $|\mathscr{E}(\mathbf{h})| > 0{\cdot}0$ (all data)*

	$\Delta\bar{E}_1$	$\Delta\bar{E}_1^2$	$\Delta\bar{E}_2$	$\Delta\bar{E}_2^2$	$\Delta\bar{E}_3$	$\Delta\bar{E}_3^2$	$\Delta\bar{E}_4$	$\Delta\bar{E}_4^2$
HCPP	0·25	0·46	0·52	1·36	0·25	0·48	0·52	1·38
CLEPX	0·18	0·34	0·39	0·91	0·17	0·32	0·39	0·92
BEKA4	0·23	0·52	0·77	1·97	0·22	0·52	0·78	2·01
STIK4	0·29	0·45	0·64	1·45	0·27	0·45	0·64	1·46
PDCPS	0·15	0·29	0·21	0·49	0·19	0·61	0·20	0·56
CANON2	0·61	0·87	0·91	2·24	0·47	0·89	0·92	2·29
ANTH1	0·12	0·28	0·67	1·78	0·12	0·30	0·67	1·78
TEMPL	0·27	0·67	0·73	1·66	0·27	0·66	0·73	1·66
CORT	0·25	0·74	0·77	1·72	0·25	0·72	0·82	1·92
K22BR	0·43	0·83	0·90	2·05	0·43	0·84	0·90	2·02
KCPP	0·17	0·53	0·50	1·30	0·20	0·54	0·51	1·25

(*b*) For $\mathscr{E}(\mathbf{h}) > 1{\cdot}0$*

	$\Delta\bar{E}_1$	$\Delta\bar{E}_1^2$	$\Delta\bar{E}_2$	$\Delta\bar{E}_2^2$	$\Delta\bar{E}_3$	$\Delta\bar{E}_3^2$	$\Delta\bar{E}_4$	$\Delta\bar{E}_4^2$
HCPP	0·67	1·76	1·22	3·89	0·71	1·88	1·23	4·00
CLEPX	0·49	1·16	0·85	2·49	0·51	1·17	0·85	2·49
BEKA4	0·99	2·45	1·62	5·15	1·00	2·51	1·64	5·25
STIK4	0·61	1·47	1·22	3·52	0·61	1·52	1·25	3·62
PDCPS	0·36	0·66	0·54	1·42	0·59	0·87	0·51	1·52
CANON2	1·25	2·63	1·96	6·71	1·29	3·38	1·92	6·53
ANTH1	0·38	0·88	1·42	5·32	0·39	1·00	1·41	5·25
TEMPL	0·70	1·72	1·34	3·68	0·71	1·71	1·33	3·66
CORT	0·76	2·19	1·09	3·42	0·72	2·03	1·23	3·93
K22BR	1·13	2·75	1·45	4·19	1·15	2·76	1·46	4·14
KCPP	0·37	0·75	1·40	3·99	0·41	0·88	1·33	3·81

* Numbers of reflections used to calculate the means are as given in Table 2.

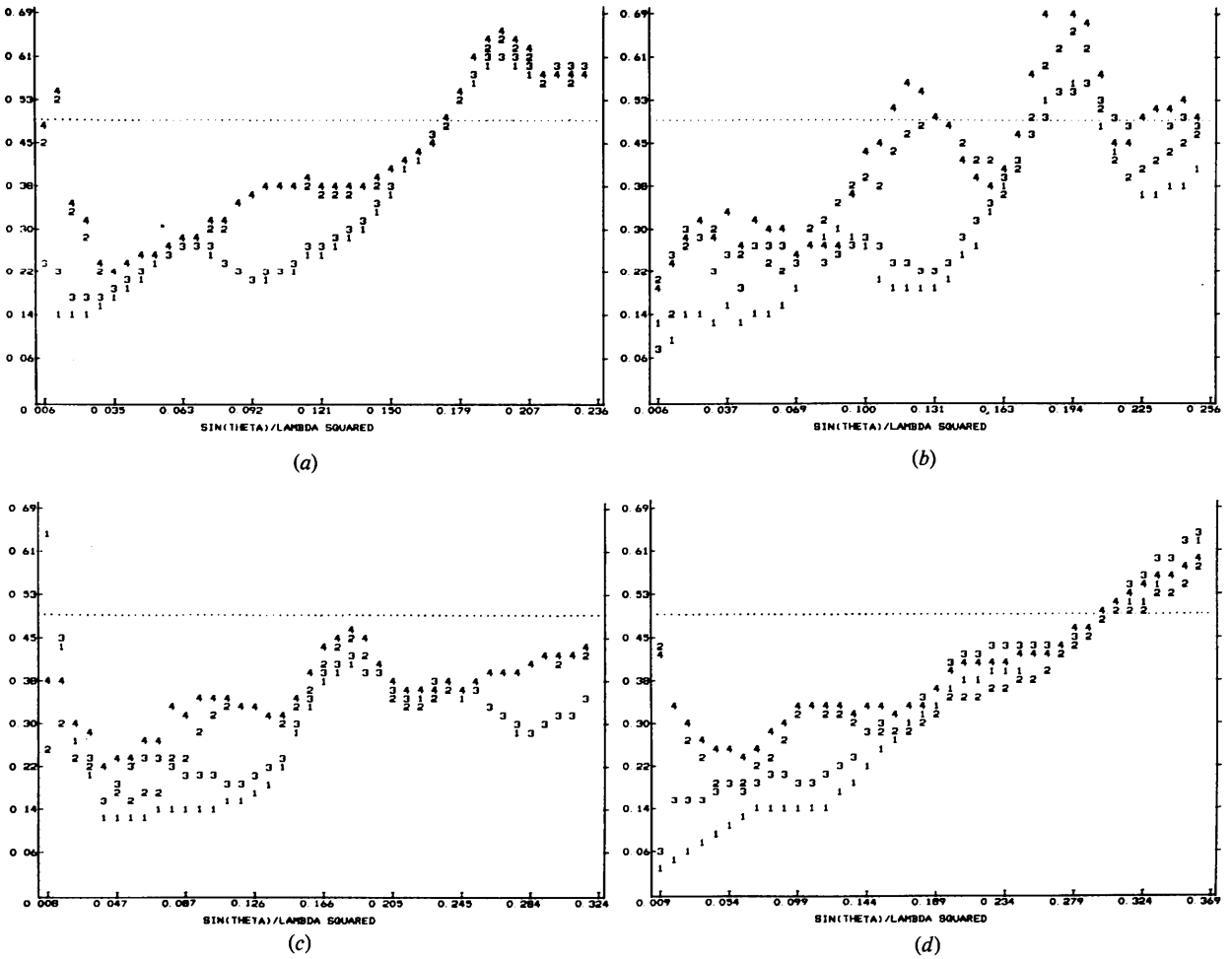


Fig. 3. Plots of the mean ratio $\Delta E_n^2(\mathbf{h})/|\mathscr{E}^2|$ *versus* s. For other details see caption for Fig. 1. The horizontal dotted line is at 0·5.

$|E_h|_3$. This correspondence is also evident in the plots of ΔE_n^2 and $\Delta E_n^2/|\mathscr{E}_h^2|$ *versus* s^2 (see Figs. 2 and 3). Both these plots show the larger differences of ΔE_2^2 and ΔE_4^2 arising from the profile scale. The overall increase in these differences as a function of s^2 is the subject of another study (Hall & Subramanian, 1982*b*).

It should be stressed that the difference measurements shown in Figs. 2 and 3, and listed in Table 2, contain both the random deviations associated with the individual estimates of $|\mathscr{E}_h|$ and $|E_h|$, and the systematic deviations of the mean $|E_h|$ from the mean $|\mathscr{E}_h|$. These measurements constitute the *total deviation* – random plus systematic – associated with the estimates of $|E_h|$. The systematic component is present because the mean value of $|E_h|_n$ in each range often departs significantly from the mean $|\mathscr{E}_h|$ for that range. That is,

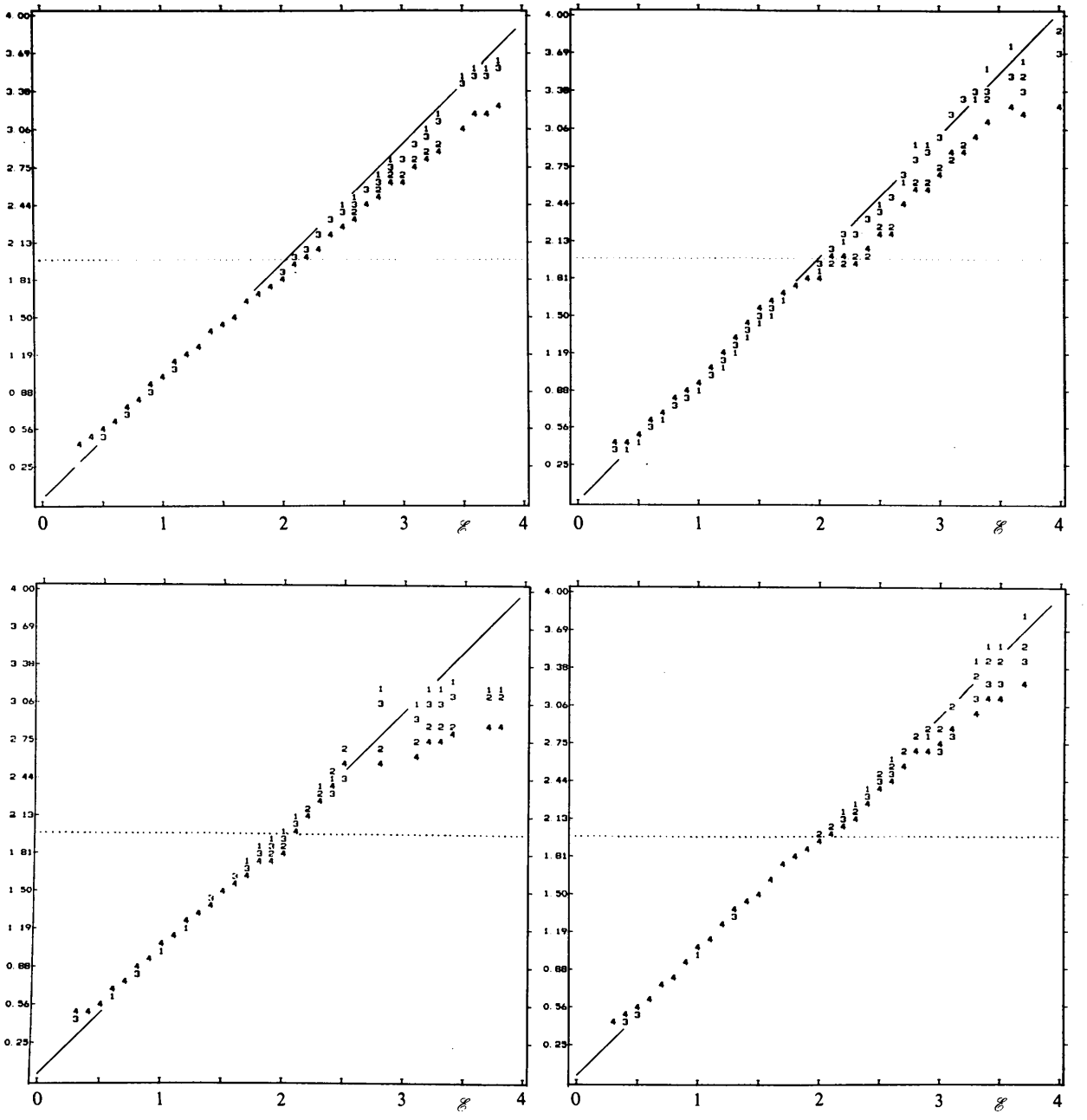


Fig. 4. Plots of $|E_n(\mathbf{h})|$ *versus* $|\mathscr{E}(\mathbf{h})|$. The solid line represents the condition $|E_n(\mathbf{h})|/|\mathscr{E}(\mathbf{h})| = 1{\cdot}0$. For other details see the caption for Fig. 1.

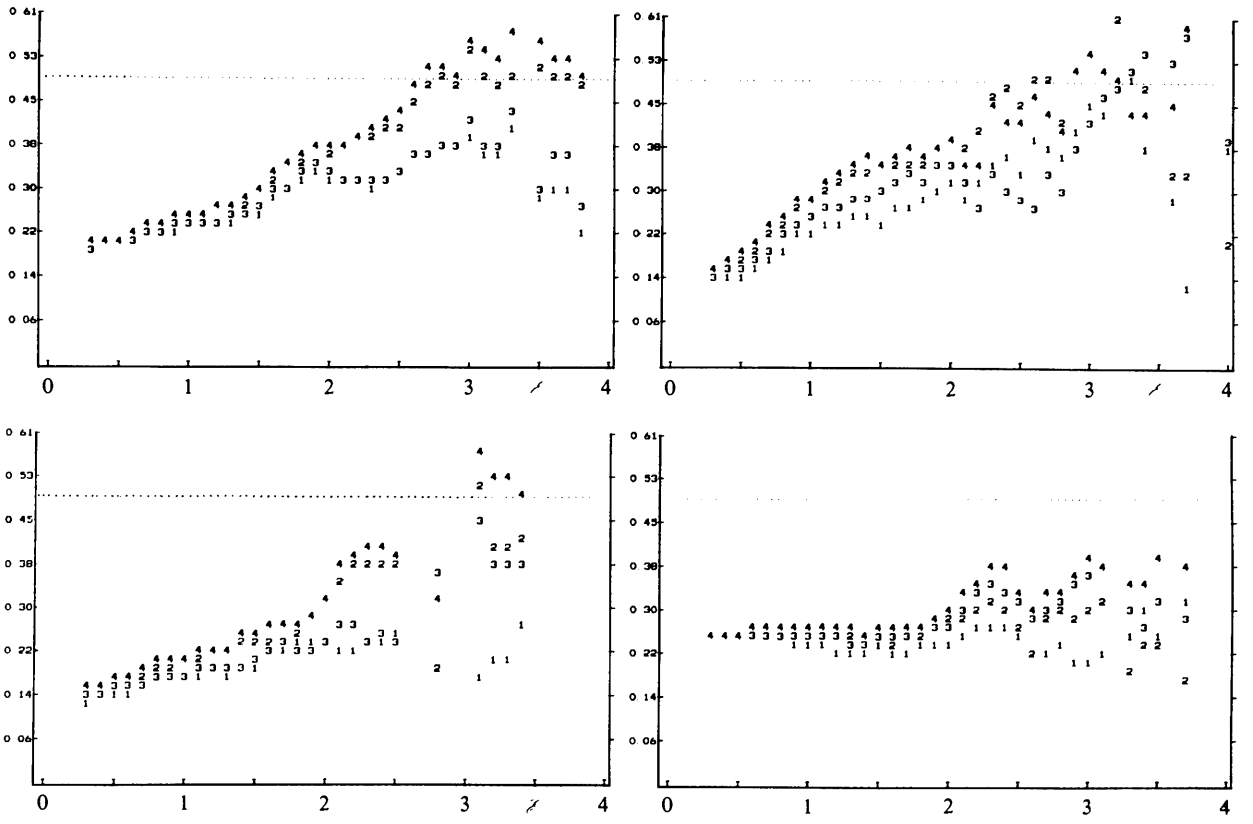


Fig. 5. Plots of $|\Delta E(\mathbf{h})|$ [see (8)] *versus* $|\mathscr{E}(\mathbf{h})|$. For other details see the caption for Fig. 1.

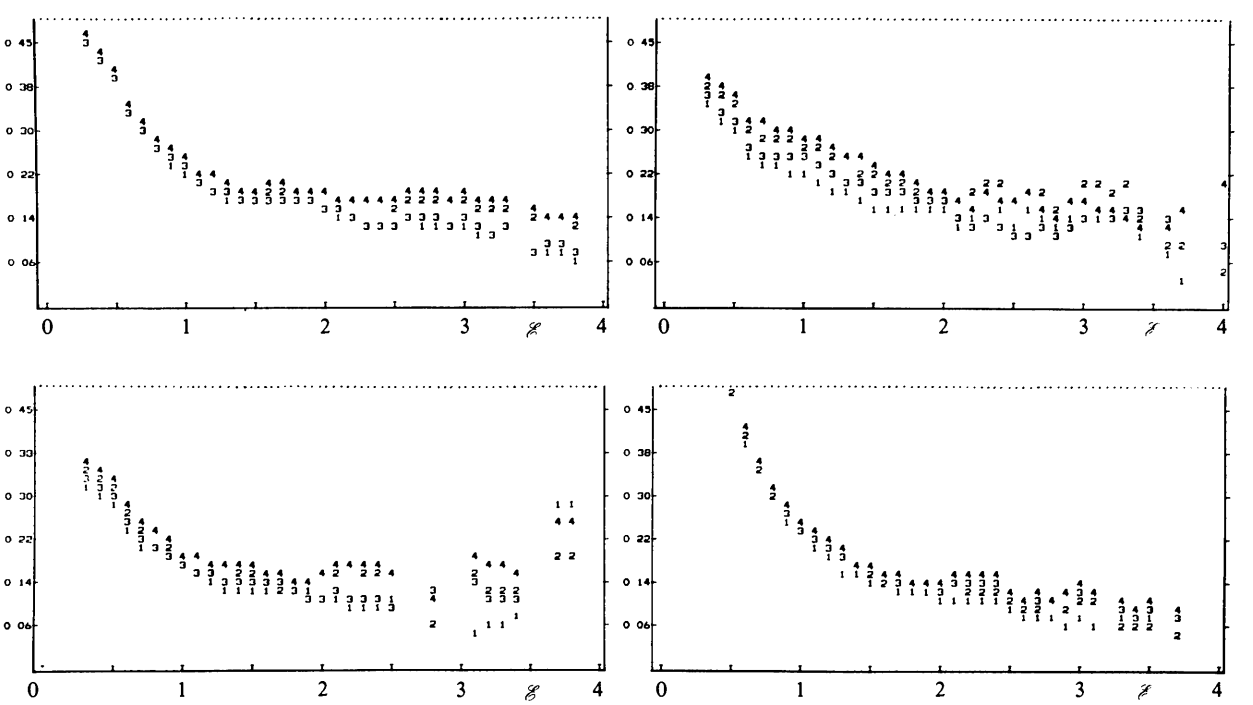


Fig. 6. Plots of the mean ratio $\Delta E_n/|\mathscr{E}_\mathbf{h}|$ *versus* $|\mathscr{E}(\mathbf{h})|$. For other details see the caption for Fig. 1.

the first moment of the differences, $\Delta\bar{E}_h = |E_h| - |\bar{E}_h|$, representing the systematic bias of the measurement, is not zero. This systematic bias component is evident in Fig. 1 and is summarized in Table 3 as the overall mean difference, $\Delta\bar{E}_h$, for all ranges. This table shows clearly the extent of the systematic differences between calculated $|\mathscr{E}_h|$ and the various estimates of $|E_h|$.

The agreement between $|E_h|$ and $|\mathscr{E}_h|$ was also tested by plotting the difference parameters against $|F_h|$, $|\mathscr{E}_h|$, and $\mathscr{E}_h^2$. A modified version of the program *ESCAN* (Hall & Subramanian, 1980) was used to plot $|E_h|_n$, ΔE_n and $\Delta E_n/|\mathscr{E}_h|$ *versus* $|F_h|$, and *versus* $|\mathscr{E}_h|$; $\Delta E_n/|\mathscr{E}_h|$ *versus* $\sigma F_h/|F_h|$; and $\Delta E_n^2/|\mathscr{E}_h^2|$ *versus* $\sigma^2|F_h^2|/|F_h^2|$. A selection of these plots ($|E_h|$, ΔE_n, and $\Delta E_n/|\mathscr{E}_h|$ *versus* $|\mathscr{E}_h|$) is shown in Figs. 4, 5, and 6, respectively, for the structures BEKA4, CANON2, CORT, and KCPP.

In Fig. 4, 1:1 agreement between $|E_h|$ and $|\mathscr{E}_h|$ is shown as a solid line. The agreement for the mean values $|E_h|$ has little systematic bias for $|E_h| < 2{\cdot}0$. However, above this value the $|E_h|$ values tend to be lower than corresponding $|\mathscr{E}_h|$ values, with the large differences occurring for the estimates $|E_h|_2$ and $|E_h|_4$. The magnitudes of these differences with respect to $|\mathscr{E}_h|$ are illustrated in Figs. 5 and 6.

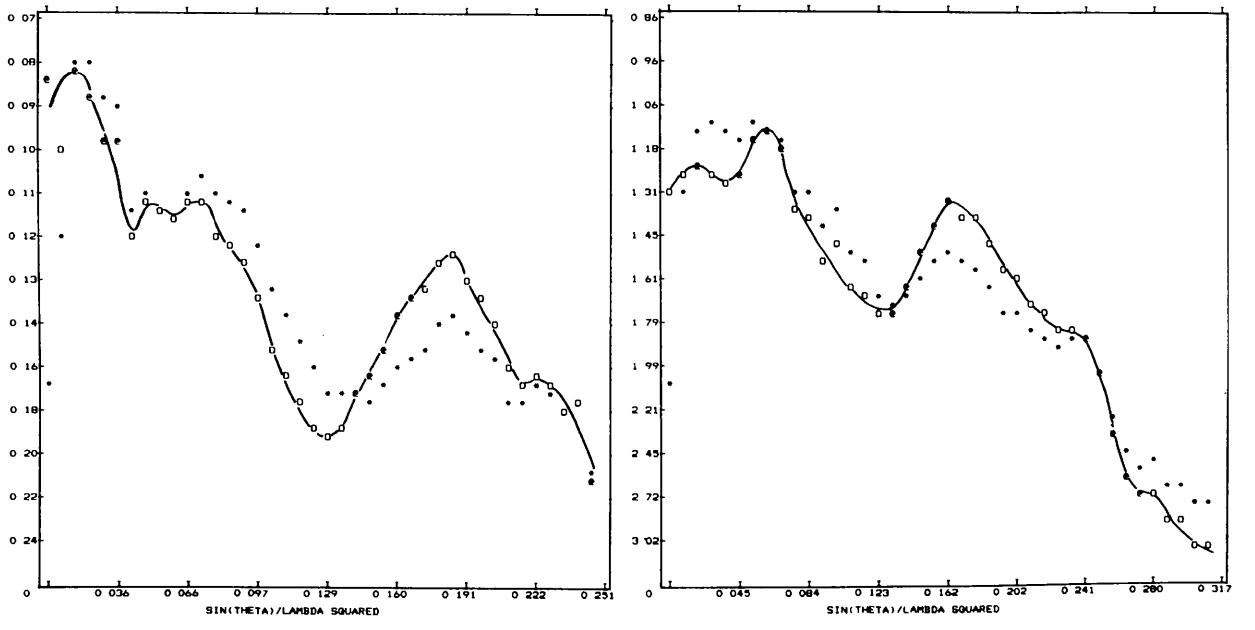
The conclusions to be drawn from these comparisons are inescapable. Firstly, the exponential scale function results in $|E_h|$ estimates that are closer to $|\mathscr{E}_h|$ than those estimated from the profile scale. Secondly, the overall rescaling provides better correlation between $|\mathscr{E}_h|$ and $|E_h|$ than the index rescaling. The tests for all eleven structures are unanimous in this verdict.

The choice of expectation value $\langle|F_h^2|\rangle$

The random-atom expectation value $\varepsilon\sum f_i^2$ was used in all of the above scaling function tests. The question of whether this is necessarily the best practice, or provides the most reliable $|E_h|$'s for a phasing procedure, must now be answered.

It is important to emphasize that phase information calculated from partial structure knowledge and used with structure-invariant relationships is not in question here. This has been shown to provide substantially improved reliability in the application of these relationships (Main, 1976; Hall, 1978). What is at issue is the practice of estimating $|E_h|$'s with other than random-atom expectation values and thus modifying the structural information passed on to the phasing process.

Only the Debye expectation value [see (4)] was tested in this study. There were several reasons for this. The molecular conformation is usually the only available structural information. It is also the procedure recommended in some programs to improve normalization. Furthermore, the phase information extractable from the Debye expression for use in the structure-determination step is negligibly small and need not be considered in this assessment.



(*a*) (*b*)

Fig. 7. Plots of the ratio $\ln|I'^2|/\langle|F^2|\rangle|$ *versus* s^2 for (*a*) CANON2 and (*b*) CORT. The solid line through data points O and @ is the ratio calculated for the random-atom expectation value. The @ are sample points for inflexion-point least squares (Hall & Subramanian, 1982*a*). The asterisk data points (*) correspond to the ratio evaluated using the Debye expectation value.

Some of the rationale opposed to the use of the Debye expectation value has already been discussed. Previous tests (Ladd, 1978) suggest that indiscriminate use of structural information in the estimation of $|E_{\mathbf{h}}|$ values can be detrimental. The results of the scaling tests described above also lead, albeit indirectly, to a similar conclusion. This is because the application of a profile scale is equivalent to the use of the Debye expectation value (see Fig. 7). The profile scale provided consistently less correlation between $|E_{\mathbf{h}}|$ and $|\mathscr{E}_{\mathbf{h}}|$ than alternative approaches (*vide supra*).

The CANON2 and CORT test data were used for the comparison of the Debye and random-atom expectation values. These are typical of structures for which reliable conformational information is often available. Only the rigidly bound non-hydrogen atoms were included in the calculation of the Debye expectation value. Fig. 7 shows the Wilson plots for both the random-atom and Debye expectation values. In Table 2, values in brackets correspond to the mean values of ΔE_n^2 and $\Delta E_n^2/|\mathscr{E}_{\mathbf{h}}^2|$ for the estimates based on the Debye expectation value. Comparison of these values for the four $|E_{\mathbf{h}}|_n$ estimates, and the two distinct expectation values, indicates that the use of random-atom expression (3) provides $|E_{\mathbf{h}}|$ values closer to $|\mathscr{E}_{\mathbf{h}}|$ than the Debye expression (4). It also confirms that the combination of exponential scale function and overall rescale produces consistently better correlation between $|E_{\mathbf{h}}|$ and $|\mathscr{E}_{\mathbf{h}}|$. These conclusions are reinforced by the plots of $|E_{\mathbf{h}}^2|$ and $|\mathscr{E}_{\mathbf{h}}^2|$ *versus* s^2 (see Fig. 8) and by a comparison with the equivalent plots of Fig. 1.

In summary these tests support the conclusions of Ladd (1978) that the *routine* application of the Debye expectation value is unwarranted in normalization procedures.

The reliability of phase relationships

In the preceding two sections it was observed that 'best' $|E_{\mathbf{h}}|$ values are obtained by applying the combined exponential scale, random-atom expectation value and overall rescale factor. In this part of the study it will be shown that the 'true' normalized structure factor $|\mathscr{E}_{\mathbf{h}}^2|$, and the $|E_{\mathbf{h}}|$ estimate closest to it, provide the most reliable structure-invariant relationships.

The triplet and quartet structure-invariant phase relationships

$$\psi_3 = \varphi(\mathbf{h}_1) + \varphi(\mathbf{h}_2) + \varphi(\mathbf{h}_3); \quad \sum^{3} \mathbf{h}_n = 0 \qquad (10)$$

$$\psi_4 = \varphi(\mathbf{h}_1) + \varphi(\mathbf{h}_2) + \varphi(\mathbf{h}_3) + \varphi(\mathbf{h}_4); \quad \sum^{4} \mathbf{h}_n = 0 \qquad (11)$$

were generated for each structure, and for each estimate of $|E_{\mathbf{h}}|_n$, using the program *SINVAR* (Hall, 1978).

The probability thresholds,

$$A_3 = 2N^{-1/2}|E_{\mathbf{h}_1} E_{\mathbf{h}_2} E_{\mathbf{h}_3}| \qquad (12)$$

$$B_4 = 2N^{-1}|E_{\mathbf{h}_1} E_{\mathbf{h}_2} E_{\mathbf{h}_3} E_{\mathbf{h}_4}|, \qquad (13)$$

used in each case and the number of triplets and quartets generated are shown in Table 4. For quartets, only those cross-vector sums

$$X = |E_{\mathbf{h}_1+\mathbf{h}_2} + E_{\mathbf{h}_1+\mathbf{h}_3} + E_{\mathbf{h}_1+\mathbf{h}_4}| \qquad (14)$$

(Hauptman, 1975) outside a preset 'window' defined by values X_T and X_B were analysed (see Table 4).

The structure-factor phases calculated with the refined coordinates were substituted into (10) and (11) to obtain ψ_3 and ψ_4 (base modulo 2π). The ψ_4 values were analysed in two categories those with $X > X_T$ (referred to as 'positive quartets' PQ) and those with $X < X_B$ (referred to as 'negative quartets' NQ). For large B_4 values the expected value of ψ_4 for PQ's, $\langle\psi_4\rangle$, is zero, and for NQ's the value of $\langle\psi_4\rangle$ is π (Hauptman, 1975). Similarly, for large A_3-valued triplets the value of $\langle\psi_3\rangle$ is assumed to be zero.

One measure of the suitability of $|E_{\mathbf{h}}|$'s in phasing procedures may be gauged from the percentage of structure-invariant violations. A violation occurs if the

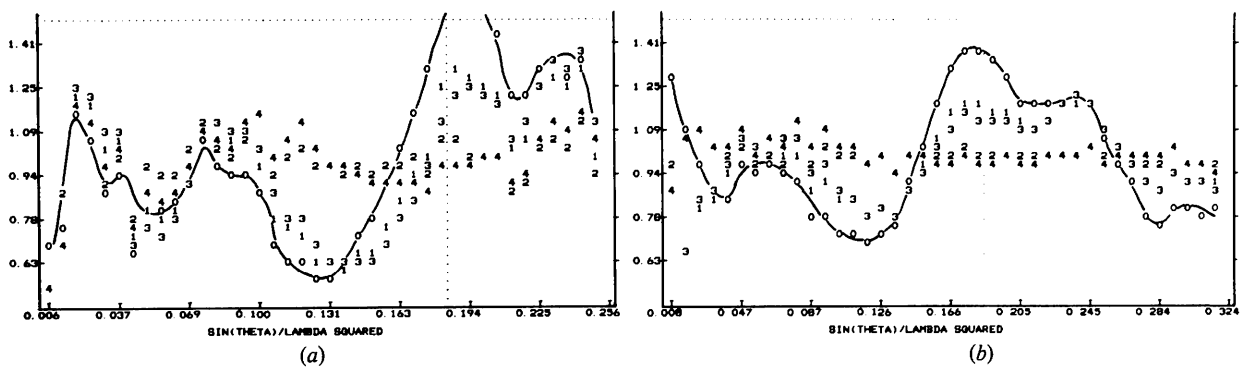


Fig. 8. Plots of $|E_n^2(\mathbf{h})|$ estimated by using the Debye expectation value and $|\mathscr{E}^2(\mathbf{h})|$ *versus* s^2 for (*a*) CANON2 and (*b*) CORT. For other details see caption to Fig. 1.

Table 4. *Summary of structure-invariant information*

All values of $|E_n(\mathbf{h})|$ are estimated using the random-atom expectation value except for values in brackets where the Debye value was employed. nE is the number of E values used as structure-invariant generators; $A_{\min}$ and $B_{\min}$ are the A_3 and B_4 thresholds [see (11) and (12)]; X_T is the X vector sum limit for positive quartets. nT is the number of unique triplets and nPQ is the number of unique quartets with $X \geq X_T$ and nNQ is the number of unique quartets with $X \leq X_B$, $X_B = X_T/2$.

Test	nE	$A_{\min}$	$B_{\min}$	X_T	$\|\mathscr{E}\|$			$\|E\|_1$			$\|E_2\|$			$\|E_3\|$			$\|E_4\|$		
					nT	nPQ	nNQ	nT	nPQ	nNQ	nT	nPQ	nNQ	nT	nPQ	nNQ	nT	nPQ	nNQ
HCPP	375	1·7	1·5	5·0	2094	567	107	2019	568	74	2056	501	111	2074	548	106	2112	441	126
CLEPX	481	2·3	1·7	4·0	2103	512	54	2046	363	71	2058	772	46	2026	385	81	2061	805	45
BEKA4	783	1·8	0·8	6·0	2153	85	339	2079	173	192	2169	194	280	2034	138	170	2121	178	293
STIK4	219	1·2	0·8	4·0	1573	409	57	1579	397	98	1549	428	127	1523	340	94	1576	377	117
PDCPS	483	4·9	4·9	5·0	9301	2644	4	12308	5737	4	12162	4788	5	10902	1717	0	10621	1679	0
CANON2	247	1·5	1·9	7·0	2025	705	30	2029	676	40	2037	689	23	2019	697	35	2031	724	33
	(247)	(1·6)	(1·9)	(7·0)				(2001)	(747)	(34)	(2040)	(673)	(24)	(2005)	(600)	(72)	(2007)	(701)	(42)
ANTH1	210	2·5	1·9	6·0	1585	611	64	1758	674	65	1600	450	102	1776	687	66	1657	467	96
TEMPL	240	1·2	0·8	4·0	2168	69	11	2280	44	14	2378	33	38	2342	49	17	2375	37	54
CORT	193	1·1	0·8	5·0	1285	136	41	1492	89	6	1164	51	14	1188	25	10	1260	12	17
	(193)	(1·1)	(0·8)	(5·0)				(1365)	(25)	(8)	(1207)	(24)	(14)	(1175)	(20)	(5)	(1268)	(30)	(7)
K22BR	278	1·0	0·8	—	524	—	—	310	—	—	311	—	—	146	—	—	147	—	—
KCPP	205	1·8	0·8	5·0	1315	109	84	1353	211	242	1324	163	397	1090	52	119	1184	48	238

ψ value calculated from the refined phases of the known structure exceeds its corresponding expectation value $\langle\psi\rangle$ by more than $\pi/2$. Table 5 lists this information for the triplets and quartets generated for $|\mathscr{E}_\mathbf{h}|$ and the four different estimates of $|E_\mathbf{h}|$. With few exceptions, the relationships resulting from the $|\mathscr{E}_\mathbf{h}|$, $|E_\mathbf{h}|_1$, or $|E_\mathbf{h}|_3$ values produce fewer violations than $|E_\mathbf{h}|_2$ or $|E_\mathbf{h}|_4$. This is borne out by the mean percentages listed at the bottom of Table 5 which indicate that for these test structures $|E_\mathbf{h}|_1$ and $|\mathscr{E}_\mathbf{h}|$ are of similar reliability. The use of Debye expectation values to evaluate $|E_\mathbf{h}|_n$ results in some decreases in the number of violations for CANON2 and CORT (see bracketed values in Table 5) but on the whole their reliability is less. In assessing these results, it is important to remember that percentage violations may not be well suited to testing non-centrosymmetric structures with unrestricted phases. In addition, they do not take into account the expected reliability of individual invariants in the phasing process. For instance, a triplet with $A = 5{\cdot}0$ contributes to the violation statistics identically to one with $A = 1{\cdot}0$. This can be misleading because the former triplet is usually more heavily weighted in the phasing process, particularly in the critical early stages.

A better measure of the reliability of phase relationships is the weighted root-mean-square difference between the calculated ψ and the $\langle\psi\rangle$.

$$\text{r.m.s.d. } \psi_3 = [\sum A_3(\psi_3 - \langle\psi_3\rangle)^2/\sum A_3]^{1/2} \quad (15)$$

$$\text{r.m.s.d. } \psi_4 = [\sum B_4(\psi_4 - \langle\psi_4\rangle)^2/\sum B_4]^{1/2}. \quad (16)$$

In this test the probability values A_3 and B_4 [see (12) and (13)] which play such an important role in most phasing processes are effectively taken into account. With the r.m.s.d. ψ values, errors in invariant relationships of high expected reliability (based on A_3 or B_4) are given highest weight. These results (see Table 6) are therefore a more realistic guide to the effectiveness of the scaling functions or expectation values in providing more reliable structure invariants. The mean values given at the bottom of Table 6 show clearly that the 'true' normalized structure factor produces the most reliable structure-invariant relationships and that the estimated value closest to it, $|E_\mathbf{h}|_1$, is superior to other estimates.

Conclusions

The principal aim of this study was to identify the form of a scaling function that provides the most reliable $|E_\mathbf{h}|$ values for use in phasing procedures. A related objective was to examine the form of the expectation value $\langle|F_\mathbf{h}^2|\rangle$ to be used in the Wilson plot. To achieve these goals, various scaling functions and expectation values were tested against eleven refined structures in two separate stages. In the first stage the correlation

Table 5. *Percentage of invariants violated*

All values of $|E_n(\mathbf{h})|$ are estimated using the random-atom expection value, except for values in brackets where the Debye value was employed. A structure invariant is designated as violated if $|\psi - \langle\psi\rangle| > \pi/2$. *T* triplets; *PQ* positive quartets; *NQ* negative quartets.

	$\|\mathscr{E}\|$			$\|E_1\|$			$\|E_2\|$			$\|E_3\|$			$\|E_4\|$		
Test	*T*	*PQ*	*NQ*	*T*	*PQ*	*NQ*	*T*	*PQ*	*NQ*	*T*	*PQ*	*NQ*	*T*	*PQ*	*NQ*
HCPP	3·9	0·0	1·9	6·0	0·1	2·7	6·3	0·2	5·4	5·5	0·0	6·4	6·5	0·0	10·3
CLEPX	0·8	0·0	12·3	2·4	0·6	12·7	2·4	0·5	10·9	2·5	0·7	22·2	2·1	0·8	8·9
BEKA4	6·9	1·2	39·7	4·8	0·6	35·3	7·6	2·7	37·2	5·2	0·7	41·3	7·4	3·6	38·3
STIK4	7·4	1·6	10·3	7·9	3·9	75·0	9·2	7·1	25·0	7·8	3·7	19·6	10·1	5·8	24·4
PDCPS	0·0	0·0	—	0·0	0·0	2·6	0·0	0·0	—	0·0	0·0	—	0·0	0·0	—
CANON2	10·5	0·8	0·0	9·8	0·4	1·6	11·5	2·7	15·9	10·9	0·8	16·8	12·2	2·2	31·6
				(9·8)	(1·1)	(11·2)	(11·5)	(1·7)	(6·6)	(10·9)	(1·0)	(37·2)	(12·6)	(2·0)	(37·3)
ANTH1	7·8	0·0	4·9	6·0	0·0	1·6	7·4	0·2	18·1	6·3	0·0	6·2	7·7	0·4	17·2
TEMPL	6·6	4·5	26·2	6·3	2·4	35·2	8·8	3·2	51·8	6·7	2·1	41·0	8·8	5·6	56·2
CORT	11·3	5·3	28·2	10·6	4·5	52·4	12·8	0·0	49·0	11·0	4·0	71·0	14·1	0·0	53·1
				(11·6)	(0·0)	(50·0)	(13·3)	(0·0)	(48·9)	(13·3)	(3·8)	(63·1)	(14·2)	(12·9)	(30·8)
K22BR	12·8	—	—	7·4	—	—	12·3	—	—	7·5	—	—	12·2	—	—
KCPP	3·0	2·5	30·6	4·1	5·7	41·9	3·7	20·9	28·0	8·3	4·0	41·9	8·1	12·0	37·1
Mean	6·5	1·6	15·4	5·9	1·8	20·4	7·5	3·8	23·2	6·5	1·6	26·6	8·1	3·0	27·7

Table 6. *Weighted r.m.s.d. ψ of invariants*

All values of $|E_n(\mathbf{h})|$ are estimated using the random-atom expectation value, except for those in brackets where the Debye expectation value was employed. The r.m.s.d. value [see (15) and (16)] for triplets (*T*) and quartets (*Q*) are in degrees.

	$\|\mathscr{E}\|$		$\|E_1\|$		$\|E_2\|$		$\|E_3\|$		$\|E_4\|$	
Test	*T*	*Q*	*T*	*Q*	*T*	*Q*	*T*	*Q*	*T*	*Q*
HCPP	21·	1·	29·	2·	31·	4·	28·	4·	32·	7·
CLEPX	5·	6·	13·	10·	12·	5·	13·	11·	11·	6·
BEKA4	27·	71·	19·	48·	29·	57·	14·	28·	28·	63·
STIK4	47·	39·	48·	48·	52·	49·	47·	52·	52·	50·
PDCPS	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0
CANON2	42·	4·	37·	3·	42·	16·	43·	8·	48·	16·
			(38·)	(8·)	(43·)	(11·)	(42·)	(23·)	(51·)	(17·)
ANTH1	28·	2·	35·	1·	39·	17·	34·	3·	40·	15·
TEMPL	46·	36·	47·	33·	52·	75·	47·	43·	43·	85·
CORT	54·	47·	55·	45·	58·	48·	56·	66·	66·	73·
			(57·)	(46·)	(59·)	(50·)	(58·)	(65·)	(61·)	(63·)
K22BR	51·	—	44·	—	48·	—	44·	—	50·	—
KCPP	20·	37·	24·	71·	21·	81·	38·	76·	76·	86·
Mean	31·	24·	32·	26·	35·	35·	34·	32·	41·	40·

between the 'true' normalized structure factor, $|\mathscr{E}_\mathbf{h}|$, and each of the $|E_\mathbf{h}|$ values estimated from the combinations of either the exponential scale or profile scale, and overall or index rescale procedure, was studied. In the second stage, the reliability of triplet and quartet structure-invariant phase relationships generated both from the calculated $|\mathscr{E}_\mathbf{h}|$ and the estimated values of $|E_\mathbf{h}|$ was investigated.

For all test structures the $|E_\mathbf{h}|$ values estimated using the exponential scale function were consistently closer to $|\mathscr{E}_\mathbf{h}|$ than those estimated with the profile scale. It is noted that the discrepancies between $|\mathscr{E}_\mathbf{h}|$ and the different estimates of $|E_\mathbf{h}|$ usually increase with $|\mathscr{E}_\mathbf{h}|$, and it is the largest $|E_\mathbf{h}|$ values that are often used in the phasing procedures! The combination of exponential scaling function, random-atom expectation value, and overall rescale was found to provide, in that order of importance, $|E_\mathbf{h}|$ values that had consistently better agreement with the calculated values of $|\mathscr{E}_\mathbf{h}|$. The analysis of triplet and quartet structure invariants confirmed that $|\mathscr{E}_\mathbf{h}|$ and $|E_\mathbf{h}|_1$ are, indeed, most reliable for methods based on these relationships and, by implication, for methods that also use structure-seminvariant relationships.

It should be emphasized, however, that a fundamental assumption in this conclusion is that the structure-invariant relationships with the highest overall reliability will provide the best chance for success in a typical direct-methods procedure. This should not be interpreted to mean that the most reliable set of invariants will *always* provide the 'best' solution, or even any solution, for a given structure and a given

phasing procedure. The uncertainty in individual invariant relationships and the relative instability of existing phasing algorithms makes it possible for less-reliable invariants to succeed occasionally where more precise sets have failed. Nevertheless, it must remain true that the most precise invariants have a statistically better chance of providing a solution independent of the methods used to apply these invariants.

In summary, this study has shown that, for the eleven structures examined, normalized structure factors, estimated from a Wilson plot using an exponential scaling function, the overall rescale and the random-atom expectation value are best suited for use in direct methods.

The authors wish to acknowledge the assistance of the Australian Research Grants Committee (Grant: C7915302) during the tenure of this study.

References

DEBYE, P. (1915). *Ann. Phys.* (*Leipzig*), **46**, 809.
DECLERCQ, J. P., GERMAIN, G. & VAN MEERSSCHE, M. (1972). *Cryst. Struct. Commun.* **1**, 13–15.
DEWAR, R. B. K. (1970). In *Crystallographic Computing Techniques*, edited by F. R. AHMED, pp. 63–65. Copenhagen: Munksgaard.
HALL, S. R. (1978). *Acta Cryst.* A**34**, S348.
HALL, S. R., RASTON, C. L. & WHITE, A. H. (1978). *Aust. J. Chem.* **31**, 685–688.
HALL, S. R. & SUBRAMANIAN, V. (1980). *ESCAN*. Program for the comparison of E values. XRAY76 System. Univ. of Western Australia.
HALL, S. R. & SUBRAMANIAN, V. (1982*a*). *Acta Cryst.* A**38**, 590–598.
HALL, S. R. & SUBRAMANIAN, V. (1982*b*). *Acta Cryst.* A**38**, 598–608.
HAUPTMAN, H. (1975). *Acta Cryst.* A**31**, 680–687.
KARLE, I. L. (1976). In *Crystallographic Computing Techniques*, edited by F. R. AHMED, pp. 27–70. Copenhagen: Munksgaard.
KARLE, J. & HAUPTMAN, H. (1953). *Acta Cryst.* **6**, 473–476.
LADD, M. F. C. (1978). *Z. Kristallogr.* **147**, 279–296.
MAIN, P. (1976). In *Crystallographic Computing Techniques*, edited by F. R. AHMED, pp. 97–105. Copenhagen: Munksgaard.
MAIN, P., FISKE, S. J., HULL, S. E., LESSINGER, L., GERMAIN, G., DECLERCQ, J. P. & WOOLFSON, M. M. (1980). *MULTAN*80. Univ. of York.
SKELTON, B. W. & WHITE, A. H. (1981). Personal Communication.
WILSON, A. J. C. (1942). *Nature* (*London*), **150**, 151.

Acta Cryst. (1982). A**38**, 590–598

Normalized Structure Factors. II. Estimating a Reliable Value of B

BY S. R. HALL AND V. SUBRAMANIAN†

Crystallography Centre, University of Western Australia, Nedlands 6009, *Australia*

(*Received* 6 *May* 1981; *accepted* 20 *January* 1982)

Abstract

A reliable estimate of the overall temperature factor B is shown to be important to the calculation of normalized structure factors, and to the application of structure-invariant phasing methods. Methods for obtaining improved estimates of B from the Wilson plot procedure are examined. The use of Bayesian statistics, the inclusion of missing data, the application of least-squares weights and the compensation for Debye scattering effects in the Wilson plot are considered. Estimates of B are compared for fourteen refined structures, including three proteins.

† Deceased 27 December 1981.

0567-7394/82/050590-09$01.00

Introduction

The standard method for estimating the overall temperature factor B and the structure-factor scale **k** from measured intensity data is by a linear least-squares fit to data in a Wilson plot (Wilson, 1942). In this plot of $\ln[|F_h^2|/\langle|F_h^2|\rangle]$ *versus* s^2 the slope of the fitted line is $-2B$ and the intercept at $s^2 = 0$ is $-2\ln(\mathbf{k})$. Because the Wilson-plot method is simple and computationally convenient, it is widely used in many crystallographic laboratories for scaling data. It is therefore surprising that the computer programs applying this technique often produce quite different estimates of B and **k** from the same data. In fact, it is not uncommon for estimates to differ by as much as a