

# Why direct and post-refinement determinations of absolute structure may give different results

David John Watkin\* and Richard Ian Cooper

Chemical Crystallography Laboratory, Department of Chemistry, University of Oxford, 12 Mansfield Road, Oxford, Oxfordshire OX1 3TA, England. \*Correspondence e-mail: david.watkin@chem.ox.ac.uk

Received 13 May 2016  
Accepted 9 August 2016

Edited by S. Parsons, University of Edinburgh, Scotland

**Keywords:** absolute structure determination; Flack parameter; refinement; software; problem structures.

**Supporting information:** this article has supporting information at journals.iucr.org/b

Direct determination of the Flack parameter as part of the structure refinement procedure usually gives different, though similar, values to post-refinement methods. The source of this discrepancy has been probed by analysing a range of data sets taken from the recent literature. Most significantly, it was observed that the directly refined Flack ( $x$ ) parameter and its standard uncertainty are usually not much influenced by changes in the refinement weighting schemes, but if they are then there are probably problems with the data or model. Post-refinement analyses give Flack parameters strongly influenced by the choice of weights. Weights derived from those used in the main least squares lead to post-refinement estimates of the Flack parameters and their standard uncertainties very similar to those obtained by direct refinement. Weights derived from the variances of the observed structure amplitudes are more appropriate and often yield post-refinement Flack parameters similar to those from direct refinement, but always with lower standard uncertainties. Substantial disagreement between direct and post-refinement determinations are strongly indicative of problems with the data, which may be difficult to identify. Examples drawn from 28 structure determinations are provided showing a range of different underlying problems. It seems likely that post-refinement methods taking into account the slope of the normal probability plot are currently the most robust estimators of absolute structure and should be reported along with the directly refined values.

## 1. Introduction

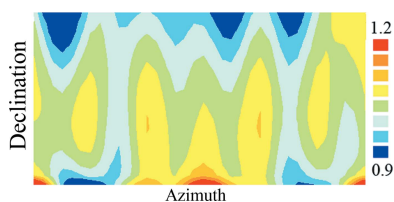
The introduction by Rogers (1981) of a new parameter,  $\eta$ , as a refineable multiplier onto  $f''$  in the least-squares optimization of a crystal structure [equation (1)] was the first attempt to directly determine absolute structures as part of the refinement process (hereafter called *direct* determination).

$$F_h = \sum_j (f_j^o + f_j' + i\eta f_j'') \times (\cos(2\pi h x_j) + i \sin(2\pi h x_j)) \quad (1)$$

Flack (1983) recognized that the  $\eta$  parameter had no physical significance except for values of  $\pm 1$ , and introduced a new formulation of the problem. He proposed that a given sample be regarded as a twin by inversion, and that refining the twin fraction would reveal the absolute structure. Representing  $|F|_h^2$  by  $I^+$  and  $|F|_h^2$  by  $I^-$

$$I_o^+ \simeq I_c^+ = (1-x)I_s^+ + xI_s^-, \quad (2)$$

where the subscript 's' indicates a quantity computed from the atomic model with the Flack parameter  $x$  set to zero (*i.e.* a non-twinned single crystal), 'c' a quantity computed from a twinned model (*i.e.* Flack parameter not necessarily zero) and 'o' an observed quantity. Like the Rogers method, this proposal refined the absolute structure parameter using all the reflection data as part of the normal structure optimization, but had the advantage that the parameter had a real physical significance throughout the whole range from zero to one. This



innovation increased awareness of the existence of twinning by inversion and fears that samples may not have been enantiopure. For convenience we will use the term *Flack parameter* to imply  $x$  determined by an unspecified method, and Flack ( $x$ ) to imply its determination as part of the main structure refinement.

The 1993 release of *SHELXL* included a post-refinement method for determining the Flack parameter, by a method which came to be known as ‘hole-in-one’. Equation (2) can be rearranged to give the Flack parameter directly from observed structure factors and structure factors computed from the atomic model and its inverse (Sheldrick, 2014).<sup>1</sup>

$$I_o^+ - I_s^+ \simeq x(I_s^- - I_s^+) \quad (3)$$

In spite of fears that post-refinement determinations of absolute structure might be compromised because of the neglect of potential covariance with the other refineable parameters, Hooft *et al.* (2008) devised a method based on a Bayesian analysis of Friedel differences (see Müller, 1988, for an interpretation of Friedel pairs). These authors recast equation (3) to treat Friedel pairs of reflections simultaneously.

$$D_o \simeq D_c = (1 - 2x)D_s, \quad (4)$$

where  $D_s = (I_s^+ - I_s^-)$  and similar for  $D_o$  and  $D_c$ .<sup>2</sup> For convenience later, we have called  $x$  computed from equation (4) the Bijvoet ( $d$ ) parameter. The advantage of (4) over (3) is that by taking differences, the significance of the real part of the structure factor is reduced, making the computation less dependent on details of the model structure.

Their process, which used weights derived from the variances of the observed intensities modified by information obtained from the normal probability plot (n.p.p.) of the Friedel residuals (Abrahams & Keve, 1971), yielded values of the absolute structure parameter, Hooft ( $y$ ), not unlike those from the Flack ( $x$ ) method. The underlying assumption, as in Dyadkin *et al.* (2016), was that the error distribution was Gaussian. Hooft *et al.* (2010) show that this distribution is adequate for good data, but that for poor data dramatically improved results are obtained by the use of the student  $t$ -distribution. The method further enabled one to estimate the probabilities of the correctness of absolute structure assignments for enantiopure or 50:50 racemically twinned samples.

Parsons *et al.* (2013) examined the use of equation (4) and its quotient form, equation (5), which we will call Parsons( $q$ ), both for post-refinement determination of the Flack parameter, and as restraints during the direct refinement of Flack ( $x$ ).

$$Q_o \simeq Q_c = (1 - 2x)Q_s, \quad (5)$$

where  $Q_s = (I_s^+ - I_s^-)/2A_s$  etc. and  $A_s = (I_s^+ + I_s^-)/2$ .

These authors showed that the Hooft ( $y$ ), Parsons( $q$ ) and Bijvoet ( $d$ ) estimates of the absolute structure parameter were

<sup>1</sup> In *SHELXL* 2014/7 the ‘hole-in-one’ fit has been renamed ‘classical fit’. This should not be confused with the much older direct refinement, as found, for example, in *X-RAY76* (Flack, 1983) or *CRYLSQ* (Olthof-Hazekamp, 1990).

<sup>2</sup> This equation first appears in this form in Thompson & Watkin (2011).

usually similar to direct refinement of the Flack ( $x$ ), but with significantly lower standard uncertainties. They also observed that using equations (4) or (5) as restraints on the least-squares refinement gave values of the Flack ( $x$ ) in close agreement with post-refinement estimates of absolute structure.

No explanation was given for the observation that direct refinement of the absolute structure consistently gave larger standard uncertainties than any of the post-refinement methods other than to note that the direct refinement was based on all the reflections used in the refinement, while the post-refinement analyses used selected subsets of the full data set. In order to investigate the source of the differences between direct and post-refinements estimations of absolute structure, several different approaches were implemented in the *CRYSTALS* program. Data sets taken from the literature including Escudero-Adán *et al.* (2014), hereafter EBB, Parsons *et al.* (2013), hereafter PFW, and Flack (2013), hereafter HDF, were re-examined using these tools.

## 2. Background

During the period before the common availability of area detector diffractometers, it was generally regarded as too expensive to collect a highly redundant set of all Friedel pairs of reflections. Some of the need for redundancy could be reduced by making measurements in geometries which minimized the differences in the experimental errors between Friedel pairs (Le Page *et al.*, 1990). Even so, full sets of Friedel pairs were generally not measured, and after a structure was resolved and refined from an asymmetric unit of data in the corresponding Laue group, selected Friedel pairs were remeasured and used for absolute structure determination (see, for example, Ealick *et al.*, 1975). The introduction of the Flack parameter has led to attempts to use X-ray crystallography both to determine absolute structure and to determine enantiopurity, *i.e.* whether the sample used for the measurements was twinned by inversion.

### 2.1. Probability methods

Prior to the introduction of Flack’s parameter, structure analysts had simply tried to ascertain the probability of the absolute structure of the crystal being the same as that of the model, so that an absolute structure was chosen to give a best match between selected observed and calculated structure factors. The Hamilton (1965)  $R$ -factor ratio method used all the observed reflections, but was difficult to apply convincingly due to uncertainty about a valid definition of the number of degrees of freedom involved in swapping from one model to its inverted image.

Other methods used reflections carefully selected from the existing data sets, or carefully remeasured. Engel (1972) favoured the ‘Bijvoet Method’, in which a selected set of reflections, the *sensitive reflections*, were remeasured more carefully. Engel used  $B_h = (Q_h - 1)/\frac{1}{2}(Q_h + 1)$  as a measure of the Bijvoet sensitivity, with  $Q_h = |F_h|/|F_{-h}|$ . A comparison of

the signs of the measured and calculated  $B$ s from a selected set of reflections yields the absolute structure. If the intensities of Friedel pairs of reflections, preferably with a  $B$  of the opposite sign, could be found and measured in a neighbouring part of reciprocal space, for which absorption and other errors will be similar, then a 'double quotient' can be estimated which has the effect (as in the Parsons quotient) of reducing the influence of geometry-related experimental errors. Le Page *et al.* (1990), recognizing that Rogers'  $\eta$  should be  $\pm 1$  for an enantiopure sample, computed the probability that the chirality of the model and that of the sample were the same on the basis of a remeasured set of selected reflections. Probability methods have been revisited again by Hooft *et al.* (2008), and using a  $t$ -distribution (Hooft *et al.*, 2010). They constructed tests on the basis that the material is enantiopure, the  $P(2)$  test giving the probability that the model and the material have the same chirality, or possibly twinned, the  $P(3)$  test distinguishing between the correct assignment, a 50:50 inversion twin, or an inverted assignment. The appeal of probability methods is that, under strict assumptions, they appear to give a clear-cut result.

## 2.2. Direct refinement of the Flack ( $x$ ) parameter

Direct refinement of the Flack ( $x$ ) parameter simultaneously with the other structural parameters is now commonplace. Flack *et al.* (2006) recommend that a full set of Friedel pairs be measured on an area detector instrument, preferably with high redundancy in order to optimize empirical intensity scaling, and that refinement be started with the Flack parameter set to 0.5 to minimize the risk of refinement to a false minimum. This is particularly important in the case of space groups with floating origins, in which the structure may distort to accommodate an incorrectly assigned absolute structure – the polar dispersion error (Cruickshank & McDonald, 1967). It has been widely observed that although the Flack ( $x$ ) is rarely in conflict with a known absolute structure (Thompson & Watkin, 2011), it can refine to a value away from the ideal value for an enantiopure material. There is also evidence that the standard uncertainty computed from the full variance–covariance matrix is often over-estimated. Parsons, Wagner *et al.* (2012) have proposed using leverage analysis to identify reflections which are particularly influential in the determination of the Flack parameter, and which could be re-measured and used as supplementary observations (restraints) in the refinement. An alternative approach (Thompson & Watkin, 2011) re-uses Friedel pairs selected from the existing data set to construct supplementary observations.

## 2.3. Post-refinement determination of the Flack parameter

The relation between the absolute structure of a crystalline material and the measured Friedel pairs is given in equation (2). The worryingly high standard uncertainty of the Flack ( $x$ ) parameter determined for many materials of known enantiopurity and absolute structure has led to a search for methods to determine the Flack parameter more robustly than simply

including it in the main least-squares refinement, especially in cases where the resonant signal is likely to be weak. Not infrequently, these methods involve the use of selected subsets of the original or new data.

Given a reasonably well refined model, the absolute structure can be estimated by solving equations (2), (3), (4) or (5) for  $x$  by conventional least squares. The disagreement sometimes seen between the hole-in-one method [equation (3)] and Bijvoet difference method [equation (4)] might, in part, be due to the additional information introduced by pairing up reflections for the differences, with the possibility that certain kinds of errors in the model or in the data might be correlated and tend to cancel out.

The denominators in the Parsons( $q$ ) expression (5) were based (Parsons *et al.*, 2013) on an extension of the earlier recognition that on a serial four-circle diffractometer setting angles could be chosen so that the absorption effect for reflections  $\mathbf{h}$  and  $\bar{\mathbf{h}}$  would be similar (Le Page *et al.*, 1990). On an area detector diffractometer these conditions are rarely satisfied, and in any case the final intensity of each reflection is usually the average of several measurements made with quite different setting angles.

Equation (5) can be rewritten as

$$D_o/2A_o \simeq D_c/2A_c = (1 - 2x)D_s/2A_s. \quad (6)$$

Here  $A_o$  and  $A_s$  seem to be scale factors down-weighting the contribution of strong reflections to the absolute structure parameter. However, when each reflection pair is weighted by the inverse of the variance of the observed quantities, this down-weighting disappears.

If equation (6) is rewritten as

$$D_o \simeq (1 - 2x)D_s(A_o/A_s) \quad (7)$$

we can see that if  $A_o$  can be regarded as  $A_s \pm \text{error}$ , the ratio  $A_o/A_s$  could take large values when the calculated structure amplitude is very small – such reflections must be excluded from any quotient calculation. In fact, if  $A_o$  is not very similar to  $A_s$  then there is a reasonable probability that there is something wrong with the model, the data or both. We can also see that the  $A_o/A_s$  terms act as per-reflection scale factors and should be counted as independent variables.

Just as plots of  $F_o$  versus  $F_c$  can be of diagnostic value in a normal structure refinement, so plots of  $D_o$  versus  $D_s$  and  $2A_o$  versus  $2A_s$  can give insight into absolute structure determination (Parsons, Pattison & Flack, 2012). The  $2A_o - 2A_s$  plot should have a unit gradient and might identify outliers in which the quotient in equation (7) lies far from unity. For enantiopure materials correctly assigned, the  $D_o$  versus  $D_s$  scatterplot should also have a unit gradient, and for materials with a large Friedif (Flack & Bernardinelli, 2008) this is usually clearly evident. For materials with a Friedif less than 100 the linear relationship is always less clear (Cooper *et al.*, 2016).

Fig. 1(a) shows a scatterplot of  $D_o$  versus  $D_s$  and  $2A_o$  versus  $2A_s$  for structure SL-6418 (Friedif = 498; Smith & Lamb, 2012). The best line through  $D_o$  and  $D_s$  (green points) has a gradient of 1.063 (5) and an intercept of  $-0.002$  (21), the correlation

coefficient is 0.960, and the coefficient of determination is 0.929. The value of  $(1 - 2x)$  is reliably determined. Fig. 1(b) is a similar plot for structure EBB-5001 (Friedif = 6.5). The best  $D_o - D_s$  line appears to be independent of the scatter of the observations, yet a least-squares fit gives a gradient of 0.92 (17) [corresponding to a Bijvoet ( $d$ ) of 0.04 (9)], correlation coefficient 0.116 and coefficient of determination of 0.014.

Except when the data points all lie on an exactly vertical line, it is always possible to fit a regression line. However, if the spread of the observations along the dependent axis is much greater than that along the independent axis, the line will have little or no physical significance. The correlation coefficient is independent of the number of observations, but the standard uncertainty is proportional to  $1/\sqrt{(n - 2)}$  so that the standard uncertainty can be reduced by including more ‘vanilla’ data – the Emperor of China Syndrome (Parrish, 1960). Rogers (1981) had been worried that in the Hamilton method, some of the resonant differences would be below the observable threshold, so that ‘*Many of the reflections are mere passengers in the calculations of the  $\Delta F$* ’ yet contribute to the degrees of freedom and falsely improve the apparent reliability of the analysis. Ealick *et al.* (1975) chose to work with reflections for which the ‘sensitivity factor’,  $SF = |D_o - D_s|/A_o$ , was the largest [note that, ignoring the effect of scale factors and the Lp correction *etc.* for Poisson statistics,  $I_o$  is proportional to  $\sigma^2(I_o)$  so that  $SF$  is a measure of signal-to-noise]. Rabinovich & Hope (1980) introduced the idea of ‘observability’,  $D = (D_s A_o)/(A_s \sigma(D_o))$  similar to Ealick’s sensitivity factor. The ratio  $A_o/A_s$  in this expression means that it is strongly related to the Parsons quotient.

The importance of a given datum on its own fitted value is measured by its leverage (Prince, 2004). Since the mean values of  $D_o$  and  $D_s$  (and the corresponding quotients) are close to zero, fitting a straight line can be regarded as a one-parameter model, so that the leverage of each data point is given by

$$P_{ii} = \frac{w_i d_i^2}{\sum w_i d_i^2}, \tag{8}$$

where  $d_i$  are the values of either  $D_s$  or  $Q_s$ . The data with the greatest leverage are those with large absolute values of  $D_s$  or  $Q_s$ . Remember that although  $D_s$  does not depend directly on  $A_s$ , large  $D_s$  can only be possible for large  $A_s$ . If each observation in the post-refinement determination of the Flack parameter is weighted by the inverse of its variance,  $P_{ii}$  is proportional to the square of the signal:noise. To a first approximation,  $\sigma^2(I) \propto I$  (Evans, 2006; but see also §5.2.6) so that the resonant difference originating from strong reflections will have large standard uncertainties, and be down-weighted. The most useful reflections are likely to be those of intermediate intensity and with a large resonant difference. This is in agreement with the leverage analysis for the Flack ( $x$ ) parameter in the least-squares refinement of all structural parameters (Parsons, Wagner *et al.*, 2012).

Equation (4) can be made to yield values of the Flack parameter on a per Friedel pair basis

$$x = (D_s - D_o)/2D_s. \tag{9}$$

Plotting  $x$  from equation (9) against  $D_s$  (Fig. 2) should give a horizontal line at the value of the Flack parameter. If  $|D_s|$  is very small compared to  $|D_o|$ , the value of  $x$  can take extreme values. For a structure with low resonant scattering, individual  $x$  can be ill-determined, and even for good data many extreme values can be seen. The massive vertical distribution near the centre of the plot (which includes both positive and negative estimates of  $x$ ) corresponds to small values of the denominator in equations (4) and (9), and it is only the data lying distant from  $|D_s| = 0$  which contain useful information.

### 3. Data quality

#### 3.1. Friedel completeness

The introduction of the Flack analysis meant that an indication of the absolute structure could be obtained without re-

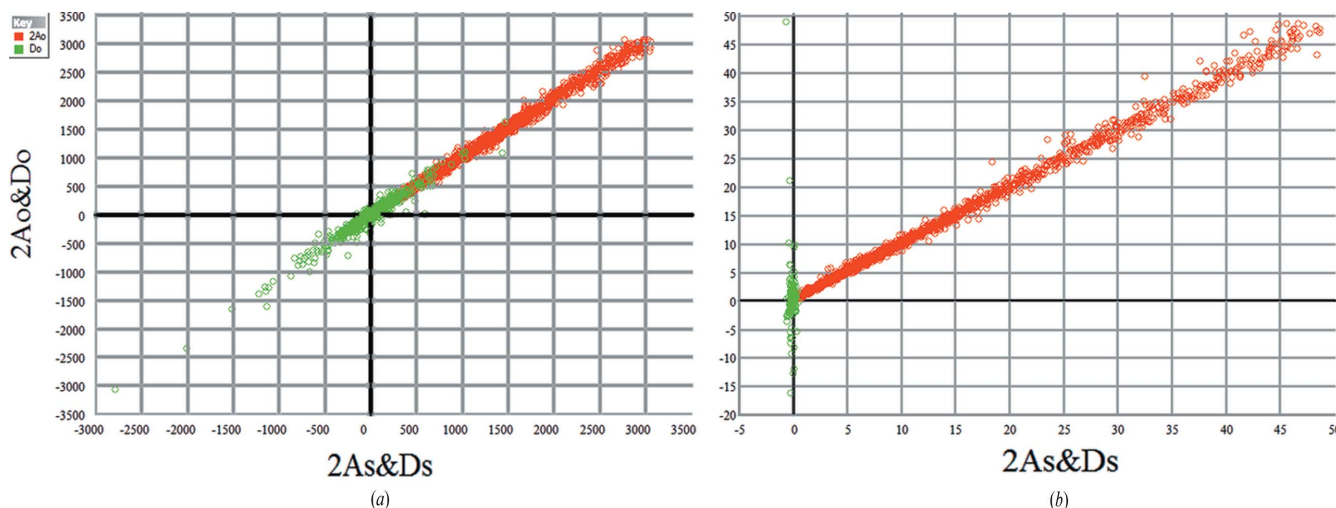


Figure 1  $D_o$  versus  $D_s$  and  $2A_o$  versus  $2A_s$  plots for (a) structure SL-6418 (Friedif 498) and (b) structure EBB-5001 (Friedif 7).

**Table 1**  
Effect of Friedel completeness on the direct refinement of the Flack ( $x$ ) parameter.

$h$ range	$k$ range	$l$ range	No. of reflections	Initial $R$	Final $R$	Final $R_w$	Minimization function	Flack ( $x$ )	$\sigma(x)$
-10	9	0 21	5134	0.0813	0.0431	0.0435	6639	0.002	0.009
-1	9	0 21	3296	0.0815	0.0441	0.0431	4509	-0.008	0.013
0	9	0 21	2927	0.0838	0.0460	0.0437	3038	0.019	0.016

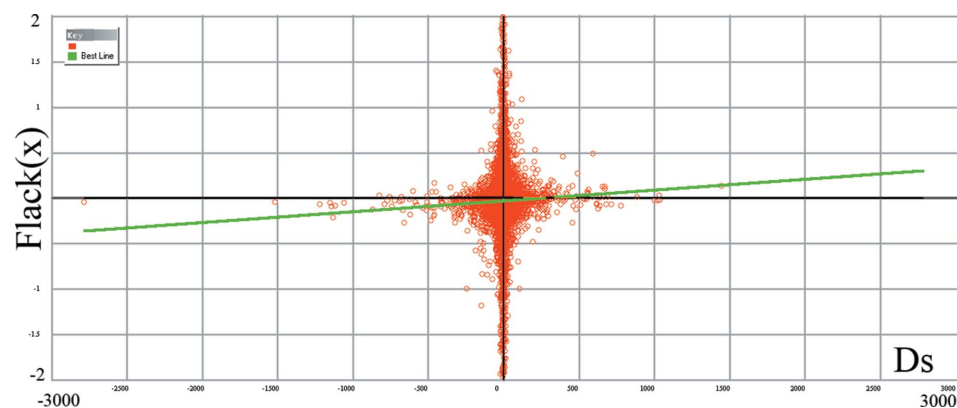
For HDF-gg3255, with Friedif of 600, the Flack ( $x$ ) parameter and its s.u. are determined reasonably reproducibly with few or even no Friedel pairs. Initial  $R$  is for the perturbed structure, final  $R$  factors are at convergence.

measuring any data. Bernardinelli & Flack (1987) showed that, strictly speaking, it does not even require the measurement of any Friedel pairs, but simply that any Friedel pairs that are measured are not merged together. Flack *et al.* (2006) discuss at length the need for extensive Friedel coverage in the case where a structure is pseudo-centrosymmetric. Trials in the 1980s using the Enraf-Nonius CAD-4 serial diffractometer showed that in some cases (for example, an organometallic spontaneously resolving in  $P2_1$ ) a good indication of the absolute structure could be obtained without measuring all Friedel pairs. These results were never published, but an example can be simulated using area-detector data. The model for HDF-gg3255 (Abud *et al.*, 2011), in  $P2_12_12_1$ , Friedif 600 (Flack & Bernardinelli, 2008), was refined using a full data set (only 114 unpaired acentric reflections), the all-positive quadrant of data plus the  $h = -1$  layer, and just the all-positive quadrant. The same model, with Flack( $x$ ) set to 0.5, all atomic coordinates slightly perturbed,  $|F|^2$  observations and the weighting scheme optimized for the full data set, was used to start all three refinements (see Table 1).

This simulation is only indicative since Friedel pairs were measured in the original experiment and used to obtain frame scale factors and absorption corrections, but it casts some light on the robustness of the Flack analysis (see also <http://www.ccp14.ac.uk/ccp/web-mirrors/hugorietveld/stxnews/stx/discuss/dis-fals.htm>).

### 3.2. Outliers and data quality

Merli & Sciascia (2011) and many others, *e.g.* PFW-2013 and Le Page *et al.* (1990), recognized that outliers in the data would degrade the analysis. Hooft *et al.* (2008) provided a filter to try to ensure that only reliable data were used in the determination of the Hooft ( $y$ ) parameter. Parsons *et al.* (2013) give an example in which exclusion of a single reflection changed the Flack parameter from 0.18 (8) to 0.08 (8). The detection of outliers is a vexing problem. Reflections with large  $|F_o^2 - F_c^2|$  residuals can be due to errors in the observed or modelled values, or both quantities. When a model is fully parameterized (all atoms have been found, disorder resolved, twinning dealt with), then there is a good chance that an individual  $I_s$  is more likely to be 'correct' than the corresponding  $I_o$ , because each computed structure factor is, in effect, a complexly weighted average of all the observed structure factors. Under these conditions, a large residual is usually attributed to error in the observation, and these reflections – the outliers – may be filtered out. In structural refinement an outlier can be identified by comparing the residual with the experimentally determined standard uncertainty. If the fully developed model will not refine so that this residual is reduced, it is usually assumed that the discrepancy is a fault in the observation. Robust/resistant weighting schemes are designed to reduce the influence of these suspect reflections in a smoothly continuous way rather than simply



**Figure 2**  
Plot of the Flack parameters *versus*  $D_s$  for all Friedel pairs in compound SL-6418. The small resonant differences give Flack ( $x$ ) in the range  $\pm 2$ , the larger differences give results in the range  $\pm 0.5$ .

very small value for the Friedel difference can still occur even when the two contributing reflections are strong, and are accepted by the ‘three sigma’ criterion. In the PFW-2013 implementation, reflections for which either or both  $I_o^+$  and  $I_o^-$  were less than three standard uncertainties were also eliminated, as were reflections with significant deviations from the  $(D_o - D_s)$  n.p.p. best-line. Whereas in the conventional least-squares refinement of crystal structures some practitioners insist on using all reflections, it is now established practice to filter out some reflections for the post refinement analysis of absolute structure. Filters are provided in *CRYSTALS* to exclude reflections which may either introduce instability into the calculations (very small denominators) or are suspected of being in serious error.

### 3.3. Iterative reweighting

The Le Page algorithm (Le Page *et al.*, 1990) in effect assigns a value of  $\pm 1$  to the Rogers’  $\eta$  value of the selected reflections on a one-by-one basis as opposed to direct refinement of  $\eta$  from all the reflections in the main least-squares calculations. It tacitly assumes that the material is enantiopure. Equation (9) enables us to also evaluate the Flack parameter on a reflection-by-reflection basis – the data used in creating Fig. 2. We could in principle evaluate the Flack parameter from each pair of carefully selected and remeasured reflections – or even from just one very carefully selected and very carefully measured pair. Because  $x$  is a continuously meaningful parameter in the range 0–1, it is not necessary to assign it an integer value. Now, rather than remeasuring selected reflection pairs to estimate  $x$ , we can use all the pairs measured in the original data collection to give individual estimates of  $x$ . With the exception of unknown correlations introduced during the measurement process, these estimates of  $x$  will be experimentally independent (or at least as independent as the measurements of the original data were). As was seen in Fig. 2, the values of  $x$  can take values wildly outside of the 0–1 range – these are physically impossible and correspond to outliers

originating either from large experimental errors, or are artefacts of a small denominator in equation (9). Following the arguments of Blessing & Langs (1987) for the merging of equivalent reflections, we can merge these individual  $x$ -values, and since each  $x$ -value has an associated experimental variance, we can compute both the external variance

$$x' = \frac{\sum w_i x_i}{\sum w_i} \tag{10}$$

$$\sigma_{\text{ext}}^2 = \frac{\sum w_i \sigma^2(x_i)}{\sum w_i} \tag{11}$$

and the internal variance (Appendix A)

$$\sigma_{\text{int}}^2 = \frac{n \sum w_i (x_i - x')^2}{(n - 1) \sum w_i}. \tag{12}$$

The probability of an individual  $x_i$  can be estimated from

$$p_i = \exp\left[\frac{-(x_i - x')^2}{2\sigma_{\text{ext}}^2}\right]. \tag{13}$$

Friedel pairs yielding a value of  $x$  differing from the average value of  $x$  by several variances have a low probability. This probability can be used as a modifier for the weight ( $w'_i = w_i p_i$ ) used to compute a new weighted average value of  $x$ , and the process repeated (Blessing & Langs, 1987). Since the distribution of the computed Flack parameters may be dispersed, skewed or long-tailed, the process is started using the median value of  $x_i$  as an initial estimate of  $x$ . Thus, rather slack values can be set for the various initial filter thresholds used in selecting reflections, and a smoothly varying function can be used to down-weight suspect data. Friedel pairs with a probability  $p_i$  greater than a user-adjustable threshold (typically 0.001) are counted to provide an indication of the number of ‘useful’ reflections in the data. The process is terminated when the number of ‘useful’ reflections is the same for two successive iterations, or until ten iterations are completed. In this latter case, the process is regarded as being

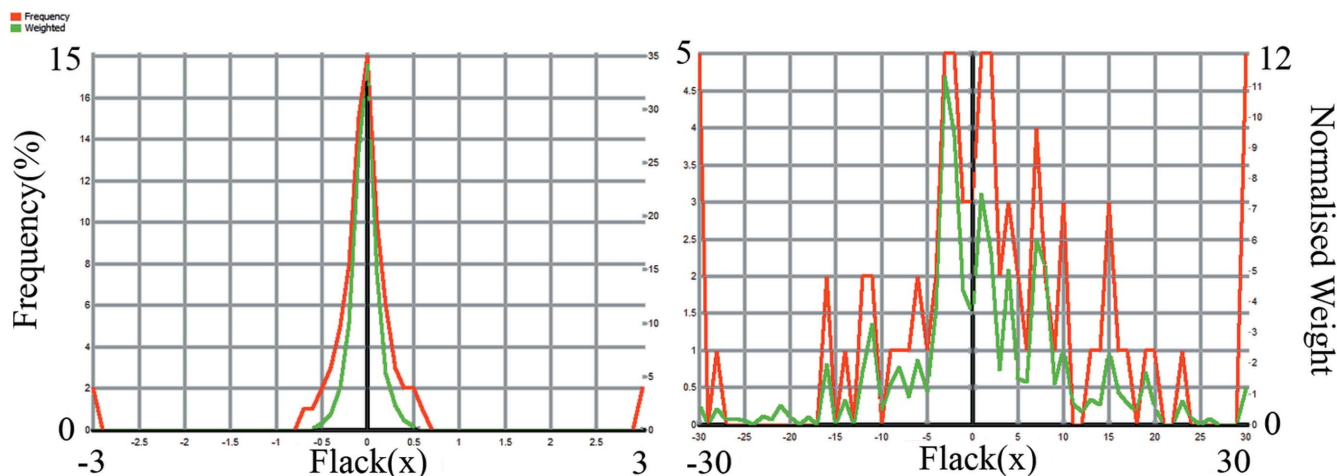


Figure 3 Flack  $x$  distribution plots for SL-6418 (left) and Motherwell (right). The horizontal scale is  $\pm 3$  for SL-6418 and  $\pm 30$  for Motherwell, with a significant number of items in the extreme bins. The red line is the simple percentage frequency, the green line is weighted with the Blessing *et al.* function.

unconverged and unsuccessful. This situation seems to arise when the resonant scattering is small compared with the errors in the intensity measurements. The standard uncertainty on the final value of  $x'$  is estimated from the weighted external variance

$$\sigma_{\text{mean}}^2 = \frac{\sum w'_i \sigma^2(x_i)}{n \sum w'_i} \quad (14)$$

Iterative reweighting (Prince, 1994) using the Tukey biweight algorithm (Tukey, 1976) gave essentially the same results as the Blessing method.

In order to provide the user with a visual representation of the data, a histogram of the frequency distribution of  $x$  can be plotted (Fig. 3). The normalized sum of the weights of the reflections in each bin is also plotted. The number of pairs containing 'useful' information and the number of pairs yielding an  $x$  value falling in the range  $-0.5 < x < 1.5$  is also output. For convenience, we will denote the value of  $x'$  and its s.u. determined by this histogram method as the Histogram ( $h$ ) parameter, and  $\sigma(h)$  its s.u. The standard uncertainty  $\sigma(h)$  can be further scaled by the gradient of the Friedel residual n.p.p. Note that the weights  $w'_i$  could also be used for the computation of a Bijvoet ( $d$ ) or Parsons ( $q$ ) parameter.

The expected and actual information content of the data can be visualized (Fig. 4) by plotting histograms of  $D_s/\sigma(D_o)$  and  $D_o/\sigma(D_o)$  (Bernardinelli & Flack, 1987). A distribution of  $D_s/\sigma(D_o)$  which is very narrow and centred on zero indicates that there is little information in the data. When this is accompanied by a broad  $D_o/\sigma(D_o)$  distribution we have an indication that the data is very noisy.

### 3.4. Ratios of averages and averages of ratios

Letting  $(1 - 2x)$  in equation (4) be represented by  $c$ , then for each Friedel pair we have

$$c_i = D_{oi}/D_{si} \quad (15)$$

An average value of  $c_i$  can be computed as a least-squares estimate (see Appendix A)

$$c = \frac{\sum D_{oi}D_{si}}{\sum D_{si}D_{si}} \quad (16)$$

or as a simple mean

$$c' = \frac{\sum c_i}{n} \quad (17)$$

leading to  $\langle x \rangle$  and  $x'$ . Equation (16) is a ratio of averages (there is a  $1/n$  term in both the numerator and the denominator), equation (17) is the average of the individual ratios,  $c_i$ . In general, if all the summations are made over the same number of data points and there are no wildly eccentric outliers, the values of  $\langle x \rangle$  and  $x'$  are similar. An indication of the presence of outliers can be obtained by computing these coefficients using all the measured Friedel pairs. If they are substantially different, the distribution of the errors in  $D_o$  may be skewed, there may be outliers, the errors may swamp any signal or there may be contributors to (15), where the  $D_{si}$  are tiny. Weighted versions of equations (16) and (17) can be recomputed during the Blessing & Langs (1987) process, where outliers are progressively down-weighted. If convergence is achieved before the maximum number of cycles is reached,  $\langle x \rangle$  and  $x'$  are usually very similar. Both values are output by *CRYSTALS*.

## 4. Experimental considerations

### 4.1. Restraints

The result of using selected reflections as restraints either in the Parsons *et al.* (2013) method or the Thompson & Watkin (2011) method seems at first to be reassuring, but a similar result can also be achieved by computing the value and standard uncertainty of the Flack parameter from the data which would otherwise have been used as restraints and simply using this as one idealized restraint. Using HDF-gg3255 (Friedel = 600) as an example again gave the following results for an unrestrained refinement, and restrained refinements using various target values of the Flack parameter and a requested standard uncertainty of 0.005 (Table 2). The *SHELX*-type weights were optimized for each refinement.

The only impact of imposing the restraint that the Flack parameter should be zero is to reduce the refined value of the parameter from 0.0018 to 0.0004. There is no appreciable change in the *R*-factors or the other estimates of  $x$ . Setting a target of 0.5 with a standard uncertainty of 0.005 leads to a

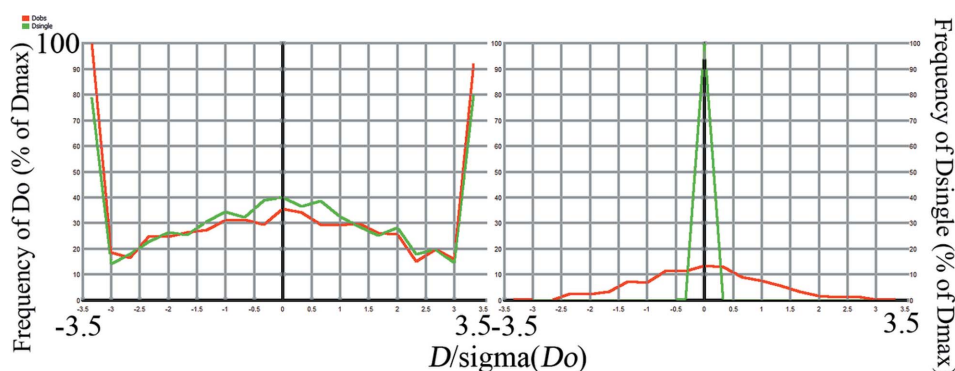


Figure 4

Distribution of the Friedel differences as a function of the standard uncertainty of the corresponding observed difference,  $\sigma(D_o)$ . The left plot is for SL-6418, right for Motherwell. The green line is the percentage frequency of  $D_s/\sigma(D_o)$ , i.e. an estimate of the theoretical signal:noise distribution which could be expected from the sample. The red line is the frequency of  $D_o/\sigma(D_o)$ , i.e. the actual signal:noise. The broader the green distribution, the better the potential usefulness of the data. For SL-6418 there are many observations in the end bins ( $D > 3\sigma(D_o)$ ). For Motherwell the green curve only spans  $\pm 1$ . The red curve has a much broader spread, indicating that random noise is a major contributor to  $D_o$ .

Table 2

Estimates of the Flack parameter derived from refinements with  $x$  restrained to progressively incorrect values for the structure HDF-gg3255.

Post-refinement methods are insensitive to the restraints because the structural model scarcely changes.

Target Flack $x$	$R$	$R_w$	<i>SHELX</i> $a$ parameter	$D_o - D_m$ $R$ factor	Flack ( $x$ )	Hooft ( $y$ )	Histogram ( $h$ )
Free	0.0416	0.0558	0.031	57.29	0.002 (7)	0.009 (5)	0.006 (4)
0.0	0.0415	0.0556	0.031	57.30	0.000 (4)	0.009 (5)	0.006 (4)
0.25	0.0423	0.0594	0.033	67.40	0.202 (5)	0.009 (6)	0.006 (5)
0.5	0.0448	0.0692	0.038	90.62	0.423 (5)	0.008 (6)	0.005 (4)
1.0	0.0541	0.0999	0.042	155.65	0.916 (5)	0.001 (3)	0.000 (4)
†	0.0419	0.0624	–				

†  $R$ -factors for this model but with Flack ( $x$ ) reset to zero.

Table 3

Influence of perturbations of the model on direct and post-refinement determinations of the Flack parameter for the structure HDF-gg3255.

The row labelled 'Original' corresponds to the unperturbed structure. Post-refinement methods are least influenced by the perturbations. The perturbations were drawn from a normal distribution with zero mean and 0.1 Å standard deviation.

Run	$R$	$R_w$	$D_o - D_c$ $R$ -factor	Flack ( $x$ )	Hooft ( $y$ )	Histogram ( $h$ )
1	0.1304	0.3137	67.1	0.08 (48)	0.031 (8)	0.027 (6)
2	0.1366	0.3714	66.5	0.050 (53)	0.032 (8)	0.035 (6)
3	0.1380	0.3750	66.2	0.033 (52)	0.049 (8)	0.051 (5)
4	0.1357	0.3236	67.3	0.074 (51)	0.035 (8)	0.035 (6)
5	0.1300	0.3182	66.6	0.083 (50)	0.040 (7)	0.039 (5)
Mean	0.1341	0.3404	66.7	0.063 (51)	0.037 (8)	0.037 (6)
Original	0.0416	0.0553	57.3	0.002 (10)	0.008 (7)	0.006 (5)

refined Flack parameter close to the target, and causes a small increase in the  $R$ -factors. The Hooft and Histogram estimates of  $x$  decrease a little, and since these are computed from the refined structural model, indicate that the model has relaxed in some way. Raising the target to 1.0 causes a very significant change in the  $R$ -factors, but the refined value of the Flack ( $x$ ) almost satisfies the restraint. The automatically adjusted *SHELX*-type weighting parameter  $a$  increased as the Flack restraint was increased, progressively down-weighting strong reflections in order to try to achieve a flat analysis of residuals, emphasizing the dangers of modifying the weights until the model is finalized. The n.p.p. for the main refinement became progressively more S-shaped as progressively invalid Flack values were imposed. The resonant difference n.p.p.s, using pure statistical weights, remained fairly straight throughout. Preserving the atomic coordinates and weights from this last refinement and resetting the Flack parameter to zero gave the  $R$ -factors in the row labelled with an asterisk. Refinement with a target Flack of unity can be achieved simply by causing a small distortion of the model which has minimal impact on the conventional  $R$ -factor but increases the reweighted  $R$ -factor. For HDF-gg3255 the median bond length distortion with the inverse restraint was 0.01 Å and the maximum 0.03 Å, *i.e.* similar to Müller's (1988) refinement of structures and their inverses. The median change in the arithmetic  $U_{\text{equiv}}$  was 0.001 Å<sup>2</sup> and the maximum 0.004 Å<sup>2</sup>. These results can be

interpreted (for a reasonable data set) as showing that small changes can be forced on the value of the Flack parameter without having an appreciable change on the atomic model, and hence on estimates of the absolute structure based on that model. They also show that while an incorrect assignment of absolute structure will affect fine details of the molecular geometry, small errors in the structural model only have a small effect on the post-refinement determination of the absolute structure.

#### 4.2. Correlation between Flack and other parameters

In order to demonstrate that the absolute structure parameters are only weakly correlated with the atomic structure, the  $x$ ,  $y$  and  $z$  coordinates of the non-H atoms in the fully refined unrestrained structure of HDF-gg3255 (called 'original' in the table) were randomly perturbed from their refined positions with a mean displacement of 0.0 and a standard uncertainty of 0.1 Å. Just the

overall scale and Flack ( $x$ ) parameters were then refined for five different perturbations of the structure, each of which had a conventional  $R$ -factor of  $\sim 14\%$  (Table 3). Although the directly refined Flack ( $x$ ) parameter was less well defined, the table shows why it may be possible to assign a reasonably reliable estimate of the absolute structure quite early on in a structure analysis by the post-refinement methods (Sheldrick, 2015).

#### 4.3. Influence of weighting schemes

In the discussion so far it has been assumed that the weights for the post-refinement analyses have been derived from the observed variances of the original diffraction data *via* equations. However, it has long been established practice to use more complex weighting schemes in the main structure refinement. These weights are computed from empirical formulae with coefficients selected to give a flat distribution of weighted residuals. This process is intended to allow for unidentified errors in the data and shortcomings in the model (Cruickshank, 1961). Weights computed in this way have an influence on the Flack ( $x$ ) parameter and its s.u. as determined during the main refinement (Bernardinelli & Flack, 1987). In order to see the influence of these weights on the post-refinement determination of absolute structure, they can be converted to observational pseudo-variances by



Table 4

PFW-fyo12e: effect of weights on absolute structure analysis.

Post-refinement absolute structure analysis using structure amplitudes computed from a model refined (a) with pure statistical weights, (b) with rescaled weights and (c) with weights computed from a *SHELX*-type formula. Each post-refinement absolute structure determination was performed with simple statistical weights (STAT), and with weights derived from the main least squares (LSQ). The gradients of the n.p.p. were computed using data between the first and ninth deciles, thus reducing the influence of ragged tails.

Main ref.		wR2	Flack ( <i>x</i> )	Post-ref.	Hole-in-one	Hooft ( <i>y</i> )	Bijvoet ( <i>d</i> )	Histogram ( <i>h</i> )	Scaled s.u.	Bijvoet n.p.p.	Main n.p.p.	GoF
Statistical weights	(a)	0.0555	0.02 (38)	STAT LSQ	0.05 (08) 0.02 (08)	−0.00 (08) −0.01 (08)	0.01 (08) 0.01 (08)	0.02 (08) 0.02 (08)	0.08 0.08	1.01 1.01	4.60	4.67
Rescaled weights	(b)	0.0555	0.02 (38)	STAT LSQ	0.05 (08) 0.01 (38)	−0.00 (08) −0.01 (08)	0.01 (08) −0.00 (38)	0.02 (08) 0.02 (36)	0.08 0.08	1.01 0.22	0.98	0.99
<i>SHELXL</i> weights	(c)	0.0662	0.17 (38)	STAT LSQ	0.06 (08) 0.14 (38)	−0.00 (08) 0.13 (08)	0.01 (08) 0.13 (37)	0.03 (07) 0.10 (36)	0.08 0.08	1.01 0.22	1.00	1.01

$$\sigma^2(I_{\text{lsq}}) = 1/\text{weight}_{\text{lsq}}, \quad (18)$$

where  $\text{weight}_{\text{lsq}}$  is the weight assigned to the reflection during refinement.

PFW-fyo12e (Parsons *et al.*, 2013) contains only carbon, nitrogen and hydrogen, and Friedel is 11.8. This data set, specifically collected with a view to exploring the differences between direct and post-refinement evaluations of absolute structure, has Flack (*x*) = 0.17 (38) and Bijvoet (*d*) = 0.01 (08) but contains no evident source for the discrepancy between the two methods. The data set has an average multiplicity of observation of  $\sim 36$ .

The structural model, including the Flack (*x*), was refined under three regimes: (a) using pure statistical weights  $1/\sigma^2(I)$ , (b) in which the weights were rescaled by a common factor to give a goodness-of-fit (GoF) of 1.0, and (c) using optimized *SHELX*-type weights, which involves adding terms to  $\sigma^2(I)$ . For each regime, post-refinement analyses were computed with pure statistical weights, and with ones derived from the least-squares weights. The results are summarized in Table 4.

For this data set we see that the choice of weighting scheme has little influence on the s.u. of the Flack (*x*) parameter determined in the main least squares, although it does have an influence on the value of the parameter itself [column headed Flack (*x*)]. In regime (a), post-refinement analysis gives the

same results whether weighted by simple statistical weights, or weights derived from the LSQ weights (since these were also simple statistical). However, all of the post-refinement methods gave standard uncertainties reduced to  $\sim 20\%$  of those from the direct refinement. The n.p.p. for the weighted Friedel differences was substantially linear with a unit gradient, although the gradient for the n.p.p. of structure factor residuals was 4.5 (Fig. 5a). The histogram of the weighted structure-factor residual  $w(F_o^2 - F_c^2)^2$  as a function of intensity (Fig. 6a) shows an unacceptable upward trend as a function of intensity.

The gradient of the n.p.p. can be made unity simply by rescaling all of the reflection variances. This rescaling has no effect on the refined parameter values, and because of the way parameter standard uncertainties are conventionally computed (Cruickshank & Robertson, 1953), it has no effect on their standard uncertainties. Because structural parameters are unchanged by this scaling, the calculated Friedel differences are unchanged, so that the row (b)STAT in Table 4 is identical to the rows (a) with the exceptions of the GoF and n.p.p. for the main refinement, which are both now close to unity (Fig. 5b). Row (b)LSQ in Table 4 contains some interesting features. Although the n.p.p. for the main refinement now has a unit gradient, the n.p.p. for the Friedel differences has a gradient of 0.2, the inverse of that for the original main

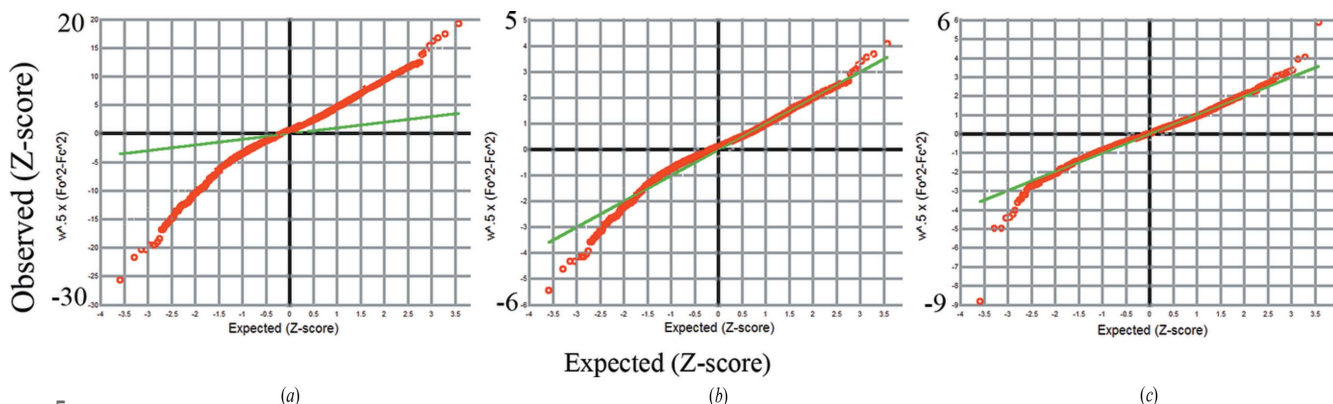
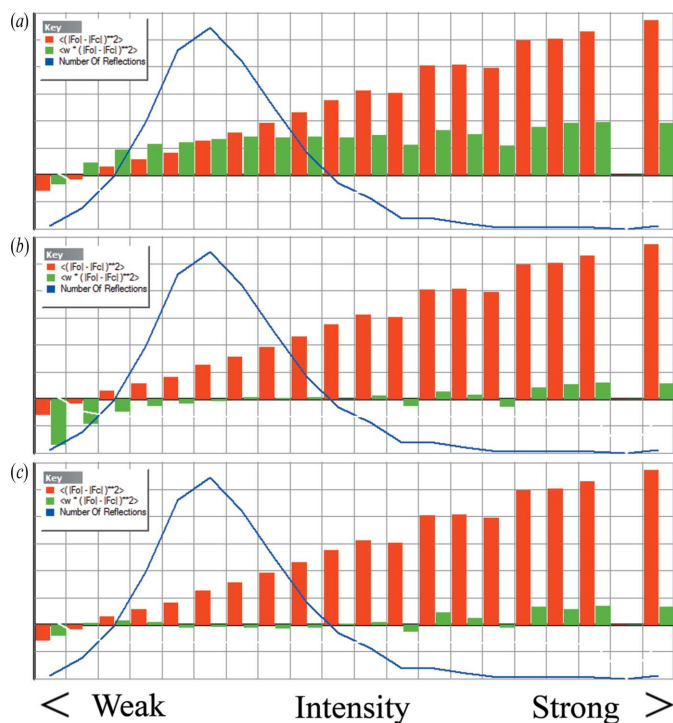


Figure 5

PFW-fyo12e. Normal probability plots for the structure-factor residuals under regimes (a), (b) and (c). The vertical axis is the observed (*z*) score, the horizontal axis the expected (*z*) score. The green lines are unit gradients.



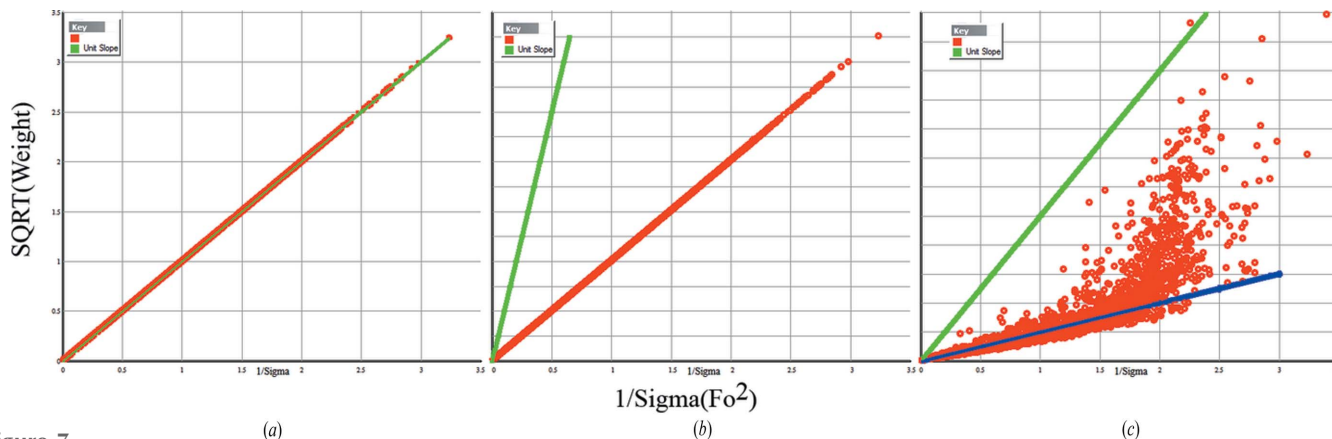
**Figure 6**  
 PFW-fyo12e. Analysis of average residuals for refinement regimes (a), (b) and (c). The red bars are the unweighted residuals, the green bars the weighted residuals on logarithmic scales. The green bars should all be small, centred on the horizontal axis. The dark blue line represents the number of reflections in each interval.

n.p.p. As a consequence, the standard uncertainties in almost all the post-refinement analyses rose to values not dissimilar to those obtained by direct refinement of Flack ( $x$ ). The exceptions to this increase in the s.u. of the absolute structure parameters are those computed by the Hooft method and the histogram method rescaled by the gradient of the n.p.p. Simply rescaling the refinement weights to produce a GoF of unity is, however, not a useful procedure because it fails to produce a

uniform distribution of weighted residuals as a function of intensity (Fig. 6b). For well behaved weights, the average  $\langle w(F_o^2 - F_c^2)^2 \rangle$  should be approximately unity for all intervals across the intensity range. It is now generally accepted that a good strategy for obtaining a uniform distribution of weighted residual is not to scale the observed variances, but to augment them with terms depending upon the magnitude of the observed and/or calculated structure factors (see, for example, the *SHELX76* instruction manual). The structure was re-refined using *SHELX*-type weights giving rows (c) in Table 4. With these weights, the gradient of the refinement n.p.p. was close to unity and the analysis of variance roughly flat. The s.u. of the directly refined Flack ( $x$ ) parameter hardly changed with the new weights, but the parameter itself increased by one-half an s.u. The shifts in the structural parameters had no visible effect on the computed Friedel differences, so that the row (c)STAT is the same as the other purely statistically weighted post-refinement analyses. The standard uncertainties for post-refinement analyses in (c)LSQ are similar to those in (b)LSQ, but the Flack parameter itself has increased. Fig. 7 shows the relationships between the weights and the standard uncertainties of the observations under the three regimes.

We can see that for the strong reflections (to the left of the plots) the *SHELX*-type weighting scheme down-weights the observations in much the same way as a simple scale factor, but that the down-weighting becomes progressively less for the weak data.

Similar results are seen for most of the materials reported in Table S1 of the supporting information. The weighting scheme for the main refinement usually must be more complex than simple statistical weighting in order to achieve a flat distribution of residuals. The effect of these weights is to increase the s.u. of the directly refined Flack ( $x$ ) (Bernardinelli & Flack, 1987). The same effect is seen if the augmented weights are used in the post-refinement determination of the Flack parameter. The n.p.p. computed for Friedel pairs using intensity statistic weights tends to have a unit gradient, suggesting that the error estimates for the differences are valid. The n.p.p. for



**Figure 7**  
 PFW-fyo12e. Square root of weights versus  $1/\sigma(F_o^2)$  plots for refinement regimes (a), (b) and (c). The green lines are unit gradients. (a) corresponds to simple statistical weights. In (b) the weights have been rescaled to give a GoF of unity. The gradient of the observations is 0.2. In (c) the weights are determined from  $1/w \approx \sigma^2(F_o^2) + 0.29F_o^2 + 0.002F_o^4$  (*SHELX*-type weights). The blue line in (c) is drawn at the same gradient (0.2) as the data in the rescaled weights.

Table 5

Sample data for structures EBB-5204 and EBB-5206 extracted from Table S1.

Bold values are referenced in the text. The pairs of lines labelled Scheme refer to refinements using a *SHELX*-type weighting scheme, the pairs labelled Counting refer to refinements using simple counting statistic weights. Lines labelled SIG refer to post-refinement analyses using counting statistic weights, lines labelled LSQ use weights based on the main refinement weights. The second s.u. in the Histogram column has been adjusted for by the gradient of the n.p.p.

Refinement weights			$wR2$	Flack ( $x$ )	Hole-in-one	Bijvoet ( $d$ )	Hooft ( $y$ )	Histogram ( $h$ )	Bijvoet n.p.p.	LSQ n.p.p.	GoF
5204											
C	Scheme	SIG	0.0666	0.05 ( <b>18</b> )	0.01 ( <b>06</b> )	0.02 ( <b>07</b> )	0.02 ( <b>06</b> )	0.04 ( <b>06</b> )( <b>06</b> )	0.94	0.96	1.05
D	Scheme	LSQ		0.02 ( <b>18</b> )	0.02 ( <b>18</b> )	0.04 ( <b>18</b> )	0.03 ( <b>06</b> )	0.03 ( <b>17</b> )( <b>06</b> )	0.35		
E	Counting	SIG	0.0559	0.02 (17)	0.01 (06)	0.02 (07)	0.02 (06)	0.04 (06)(06)	0.94	2.50	2.70
F	Counting	LSQ			0.01 (06)	0.03 (07)	0.02 (06)	0.05 (06)(06)	0.93		
5206											
C	Scheme	SIG	0.0678	<b>0.29</b> (24)	− <b>0.01</b> (08)	0.20 (08)	<b>0.14</b> (08)	<b>0.19</b> (08)(07)	1.00	0.85	0.92
D	Scheme	LSQ			<b>0.26</b> (27)	0.36 (27)	<b>0.26</b> (09)	<b>0.26</b> (26)(09)	0.34		
E	Counting	SIG	0.0677	0.02 (34)	−0.03 (08)	0.20 (08)	0.14 (08)	0.20 (07)(07)	1.00	3.62	4.34
F	Counting	LSQ			0.01 (08)	0.20 (08)	0.14 (08)	0.19 (07)(07)	0.99		

weights based on the LSQ sometimes have a distinctly non-unit gradient, with pronounced curved tails. This seems to suggest that the modifiers added to  $\sigma(I)$  in the refinement weighting scheme to achieve a constant unit  $\chi^2$  may be reflecting deficiencies in the model as much as in the data. Note that the hole-in-one method usually gives similar results to other post-refinement methods when simple statistical weights are used.

## 5. Results and examples

### 5.1. Overview

The above computations were performed on a selection of structures from data collected locally or taken from the literature. The examples were chosen to cover a range of values for Friedif, the Flack parameter, its standard uncertainty or had attracted comments in the body of the paper. When the deposited data included the *SHELX* format .res and .hklf data, this was used in preference to the .cif and .fcf format data. This was especially useful when  $I$  or  $\sigma(I)$  for weak data in the .fcf file had only one significant figure. Each structure was re-refined in *CRYSTALS* and the parameters for a *SHELX*-type weighting scheme optimized. The atomic parameters were first refined in a single matrix together with the overall scale and Flack ( $x$ ) parameters. Additional refinements were then performed from this atomic model on just the overall scale and the Flack ( $x$ ) parameter, first using the optimized weights, and then with weights derived directly from the counting statistics.

Table S1 in the supporting information contains the results of the absolute structure analysis of 28 data sets. In every case the results from the full matrix (rows A&B) were almost identical to those from the small-block (rows C&D), indicating that for a fully refined structure there is little correlation between the structural parameters and the absolute structure parameters (Fig. 8).

Rows E & F give the results of refining Flack ( $x$ ) and scale using simple statistical weights. Refining the whole structure with weights derived from unmodified intensity variances would have led to shifts in the atomic parameters.

Table 5 contains sample data for two materials from Table S1. In each case rows E are almost identical to rows F, showing that direct refinement of the Flack ( $x$ ) parameter using simple intensity statistical weights gives the same results as post-refinement analysis.

The most significant differences are between rows C and D – the *SHELX*-weighted main refinement and post-refinement analyses with either counting or refinement based weights. They show that direct refinement of the Flack ( $x$ ) is more or less unchanged when using either simple statistical or modified (*SHELX*-type) weights, providing the atomic model is not allowed to adjust. However, the post-refinement determination of absolute structure is sensitive to the weights used. Post-refinement analysis using weights derived from those used in the main least squares yields results very similar to those found by direct refinement of Flack ( $x$ ). However, using simple statistical weights almost always leads to significantly lower standard uncertainties (Fig. 9). The influence on the absolute structure parameter itself is more variable (Fig. 10). We find that the discrepancy often seen between direct and

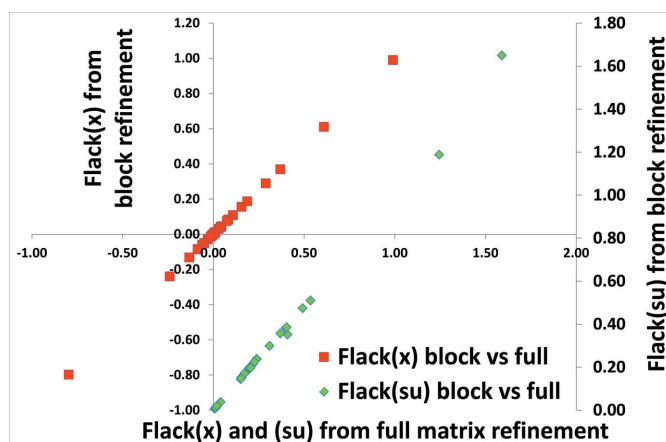


Figure 8

Comparison of the Flack parameter (red) and its s.u. (green) for 28 data sets, determined as part of a full matrix refinement with all the other structural parameters *versus* the values obtained by refinement of a small block containing only the overall scale and the Flack parameters. The points lie on lines of unit gradient, indicating the full and block matrix refinements are essentially equivalent.

post-refinement values of the Flack parameter is linked to the weights used in the refinement.

In Table S1 we see that the slope of the n.p.p. for the statistically weighted Friedel differences is generally close to unity, but the slope with weights from the main refinement is almost always less than unity (Fig. 11).

Because both the Hooft ( $y$ ) and scaled Histogram ( $h$ ) methods take into account the slope of the n.p.p., they give very similar values for both the absolute structure parameter and its s.u. independently of the weighting scheme used. It would seem, for general work at least, that the Hooft ( $y$ ) parameter as implemented in *PLATON* (Hooft *et al.*, 2008) is a widely available suitably robust estimator of absolute structure.

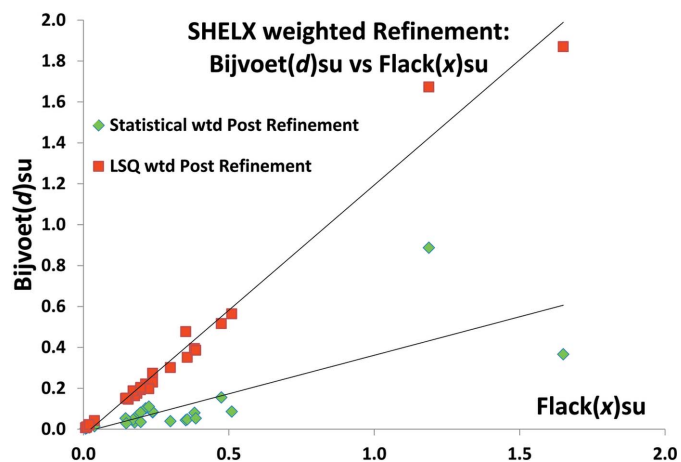
Fig. 12 shows standard uncertainties computed from the main least squares, and by the hole-in-one, Hooft and Bijvoet

difference methods *versus* the histogram method. The main refinement was done with *SHELX*-type weights, the post-refinement analyses with simple statistical weights.

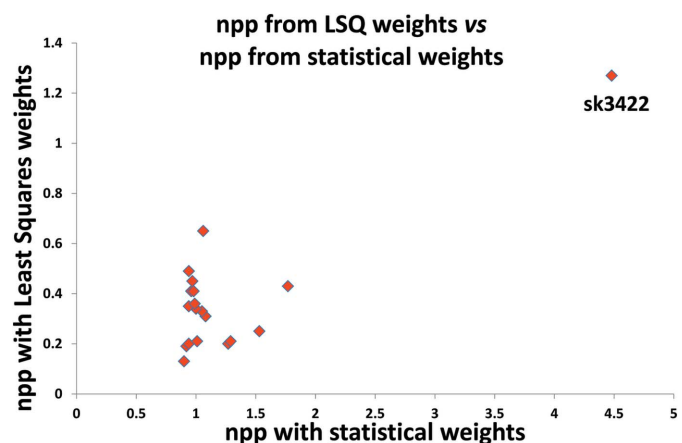
### 5.2. Examples

The various estimators of absolute structure are summarized in Table S2. The refinements for the structures which gave a s.u. for the Flack ( $x$ ) substantially larger than the s.u. determined by other methods (the clear outliers in Fig. 12) were examined in detail to try to understand the source of the discrepancies.

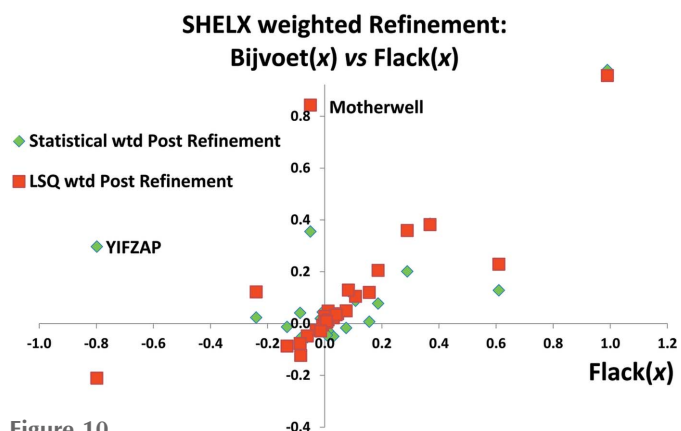
**5.2.1. Motherwell (Watkin, unpublished).** The data for 2-methyl-4-nitroaniline (previously published by Howard *et al.*, 1992; Ferguson *et al.*, 2001), Friedif = 5.94, in *Cc*, was remeasured without the intention of determining the Flack para-



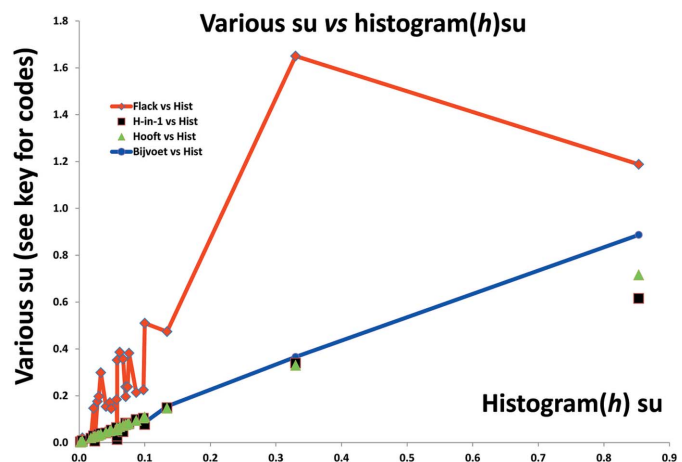
**Figure 9**  
Standard uncertainties of the Bijvoet difference determination of absolute structure *versus* direct determination for materials in Table S1. The count weighted post-refinement analysis always gives smaller standard uncertainties than direct refinement.



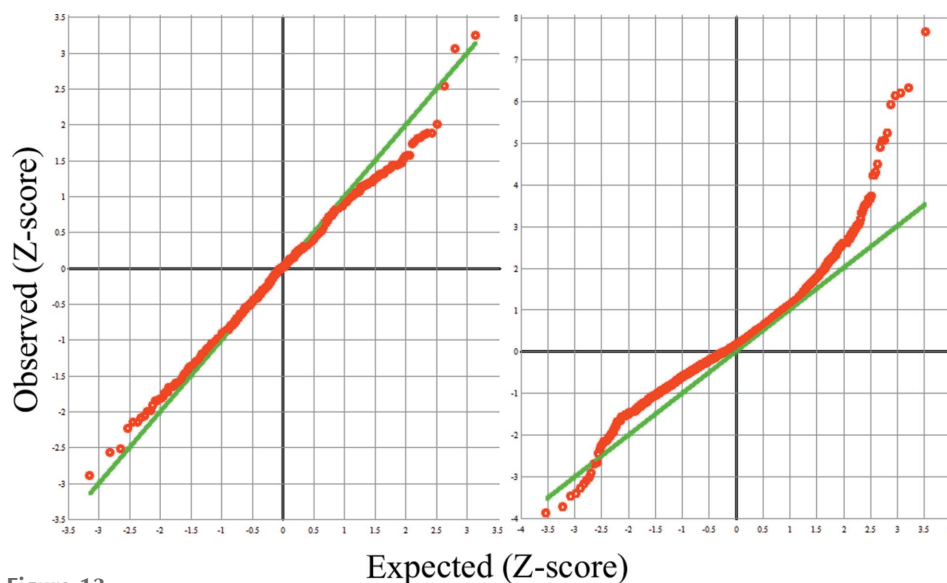
**Figure 11**  
Gradient of the normal probability plot for the Friedel residuals using least-squares weights *versus* the gradient with counting statistics weights. The gradient based on counting weights is usually close to unity. Data set sk3422III (Fäbry *et al.*, 2012) gives very anomalous n.p.p.s.



**Figure 10**  
Bijvoet difference determination of absolute structure *versus* direct determination for materials in Table S1. Using the same *SHELX*-type weights for the post-refinement analysis as in the direct refinement generally results in similar values (red datum) for Flack ( $x$ ). The count weighted post-refinement analysis often gives smaller values (green datum). Largest differences are for YIFZAP and Motherwell, discussed below.



**Figure 12**  
Standard uncertainties computed by direct Flack refinement, hole-in-one, Hooft and Bijvoet difference methods compared with those from the Histogram method for the structures in Table S1. The direct-refinement method almost always gives larger, and in some cases much larger, standard uncertainties. Note that all the post-refinement methods generally give essentially the same results.



**Figure 13**  
HDF-tp3005: (a) N.p.p. for the residuals from the post-refinement absolute structure determination. The plot is acceptably linear. (b) N.p.p. for the residuals in the main structural refinement. There is something wrong with the model, the data or the error estimates.

meter, using Mo radiation from a conventional source. The data collection strategy yielded data containing little or no resonant signal. From equation (4) one would expect the absolute structure parameter to be 0.5 with an s.u. simply reflecting the noise in the data. This result is more or less achieved by all except the Hooft (*y*) post-refinement methods. For other materials, with a larger value for Friedif, one would expect larger values for  $D_s$  and thus a smaller s.u. on the absolute structure parameter, enabling twinning by inversion to be detected.

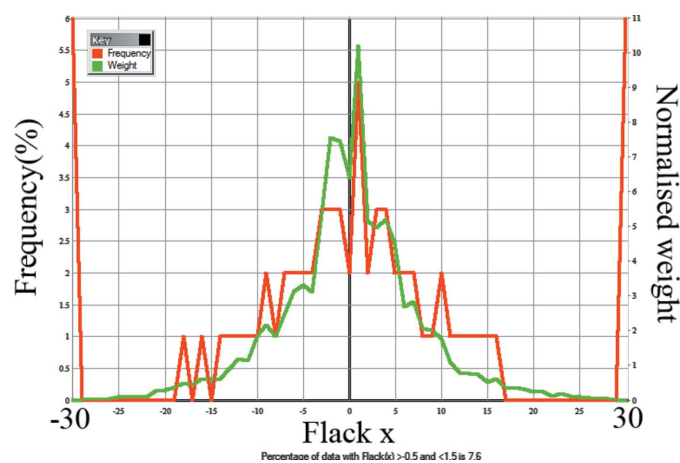
**5.2.2. HDF-tp3005W (Zhang *et al.*, 2012).** The absolute structure of this material was known from the starting materials. Both the refined Flack (*x*) parameter and its standard uncertainty are larger than values obtained by post-refinement analyses. The n.p.p. (Fig. 13a) for the residuals from the post-refinement absolute structure determination is acceptably linear, but the plot for the residuals in the main refinement shows serious deviations from linearity (Fig. 13b). The weights for the plot illustrated were computed from a *SHELXL*-type scheme. Weights derived from three, four- or five-parameter Chebychev polynomials (Carruthers & Watkin, 1979) fared no better. The relatively large value for the second parameter in the *SHELX*-type weighting scheme is often taken as a sign of twinning, but none could be identified using *ROTAX* (Cooper *et al.*, 2002). The original authors reported positional disorder in one of the residues, but this was well modelled. They also reported that the crystals were very small and the data collection was difficult, requiring the use of synchrotron radiation. Since the refinement of a conventional model produces unweighted residuals whose distribution cannot be matched by conventional weighting schemes, it seems likely that the model is deficient or the error distribution in this experiment is unusual.

**5.2.3. HDF-sf3166 (Seela *et al.*, 2012).** This material, of known absolute structure, Friedif of 5.4 and with two molecules in the asymmetric unit, gave a refined Flack (*x*) parameter of  $-0.24$  (49) and a Histogram (*h*) parameter of 0.00 (14). The n.p.p. for the residuals from both the post-refinement absolute structure determination and the main refinement were good straight lines with a unit gradient. However, of the 2630 Friedel pairs having  $D_s > 0.01\sigma(D_o)$ , only 7.6% of the Friedel pairs give a Flack parameter in the range  $-0.5$  to  $1.5$  during the histogram post-refinement analysis (Fig. 14). There are no Friedel differences having a theoretical magnitude of more than  $0.5\sigma(D_o)$ .

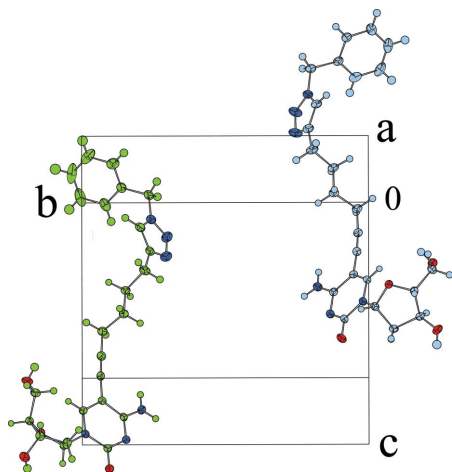
It is not uncommon to find pseudo-symmetry between the independent molecules in structures with  $Z > 1$ . The *CRYSTALS MATCH* procedure identified a pseudo-glide plane parallel to *c*, Fig. 15. If the terminal 2-(hydroxymethyl)tetrahydrofuran-3-ol is excluded from the matching procedure, the remaining 45 atoms conform to the pseudo-glide ( $x, 0.97 - y, z - 0.52$ ) with an r.m.s. deviation in equivalent torsion angles of  $16^\circ$ .

**5.2.4. PFW cholestane (Parsons *et al.*, 2013).** Cholestane contains only carbon and hydrogen, and two molecules in the asymmetric unit, Friedif is 9.0. The effect of including Friedel pairs with progressively smaller resonant differences is shown in Table 6.

Filtering out those reflections with  $|D_s| < 0.1\sigma(D_o)$  gives Bijvoet (*d*), Hooft (*y*) and Histogram (*h*) parameters close to



**Figure 14**  
HDF-sf3166. Frequency distribution of values of the Flack parameter for individual reflections. The horizontal axis (Flack (*x*)) runs from  $-30$  to  $+30$ . Note the large number of reflections in the extreme bins. Percentage of data with Flack *x* values in the interval  $-0.5$  to  $1.5$  is only 7.6.



**Figure 15**  
(100) projection of the two independent molecules in sf3166 showing the pseudo *c*-glide broken only by the 2-(hydroxymethyl)tetrahydrofuran-3-ol groups. Most of the a.d.p.s also obey the pseudo-glide.

**Table 6**

PFW Cholestane: various absolute structure estimators as a function of the ratio of expected signal-to-observed standard uncertainty.

Reflections (1) is the number of Friedel pairs used in the following calculations, Reflections (2) is the number of pairs with a significant robust-resistant weight. %Ok is the percentage of these pairs yielding an individual estimate of Flack(*x*) in the interval  $-0.5$  to  $1.5$ .

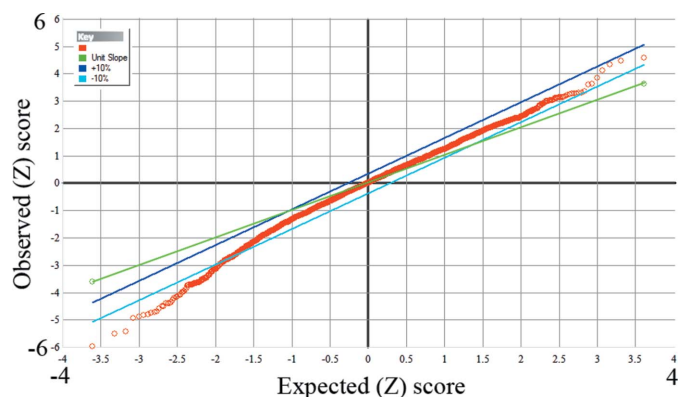
	$ D_s  > 0.1\sigma$		$ D_s  > 0.01\sigma$	
	Number of reflections	Parameters S.u.	Number of reflections	Parameters S.u.
Flack ( <i>x</i> )		-0.061 0.538		
Reflections (1)	737		3274	
Bijvoet ( <i>d</i> )		0.045 0.095		-0.063 0.086
Hooft ( <i>y</i> )		0.020 0.121		-0.036 0.110
Reflections (2)	694		2405	
%Ok	23.6		9.8	
Histogram ( <i>h</i> )		0.058 0.121		0.001 0.101

the ideal value of zero for this enantiopure material. Reducing the threshold to include reflections with  $|D_s| < 0.01\sigma(D_o)$  increases the number of reflections used from 737 to 3274, but the Bijvoet (*d*) and Hooft (*y*) parameters go slightly negative. When the weaker resonant differences are included, the histogram filtering reduces the percentage of reflections used from 94 to 73%. 24% (163) of these reflections have an individual Flack parameter in the range  $-0.5$  to  $1.5$  when  $|D_s| < 0.1\sigma(D_o)$ , falling to only 10% (236) when the threshold is reduced to 0.01. The n.p.p. for the residuals from the main structural refinement lie on a good straight line with a unit gradient, but unlike the case of tp3005 (and most well determined structures), the n.p.p. for the resonant differences has a gradient of 1.30 and a distinct downwards tail (Fig. 16).

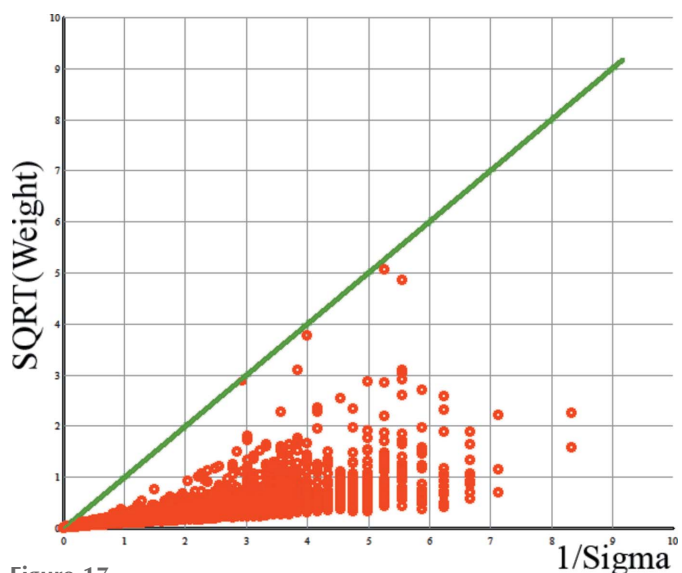
The deviations could be due to errors in  $D_o$ ,  $D_s$  or the weights. As demonstrated earlier,  $D_s$  is not strongly influenced by fine details of the structure, so one is left suspecting the problem is with the intensities or their standard uncertainties. Since the structure refined to a conventional *R* of 0.029, it

seems that the s.u. of the observations may have been underestimated. The weights used in the main refinement are based on the reported intensity standard uncertainties modified to ensure a uniform analysis of variance. Fig. 17 is a plot of  $\text{SQRT}(\text{weight})$  versus  $1/\sigma(F^2)$ .

5.2.5. EBB-threonine (Escudero-Adán *et al.*, 2014). The paper EBB 2014 is a rich mine of useful data sets collected under a variety of conditions with Mo *K* $\alpha$  radiation. Five of the D-threonine data sets were re-refined in *CRYSTALS*, yielding essentially the same results as obtained by the original authors. Those authors drew attention to data set EBB-5206, which had an anomalously large value for the directly refined Flack (*x*) parameter (EBB Fig. 3). The Flack parameters determined by post-refinement methods were also anom-



**Figure 16**  
PFW Cholestane. N.p.p. of the residuals for the post-refinement analysis of resonant differences using weights computed from the standard deviations of the observed intensities. The green line is unit gradient, the blue lines enclose the data used to compute the best line.



**Figure 17**  
PFW Cholestane. Square root of the weight used in the least-squares refinement versus  $1/\sigma(F^2)$ . The weights determined by a *SHELX*-type procedure are almost all much less than purely statistical estimates. The green line would correspond to simple statistical weights. The modification terms augment the standard uncertainties, *i.e.* reduce the weights.

alously high, yet all methods gave standard uncertainties not unlike those from the other threonine data sets. EBB attribute these anomalous results to the reduced number of reflections (6401, redundancy 3.2) compared with other analyses (*e.g.* 8324 for EBB-5204, which had a redundancy of 11.6). We were not convinced by this argument because EBB-5205 had a similar number of reflections and redundancy (7710, 3.7), but yielded a quite normal refined Flack parameter. Fortunately these authors had deposited complete reflection data sets (.hklf files) so we were able to examine them in detail.

**Data completeness:** Data collection EBB-5206 was terminated prematurely to try to reduce the redundancy. As Fig. 18(a) shows, this strategy also had the unfortunate effect of reducing the completeness of the data in the region between the Bragg angles of 40 and 45°, even when Friedel pairs were merged. Most serious is the systematic pattern to some of the missing reflections, including, for example, the row lines ( $h00$ ) where  $h$  is even, ( $h10$ ) where  $h$  is odd and some patches of reciprocal space ( $hk0$ ) where  $h$  is 9–12 *etc.* There was a small dip in completeness for data set 5204 at about 45° (Fig. 13b).

**Signal to noise:** Fig. 19 shows some measures of the quality of the data as a function of resolution. This suggests that for

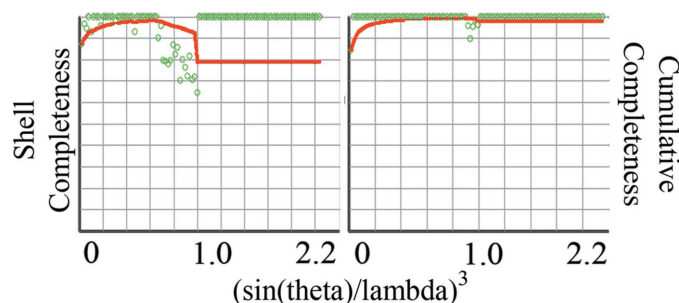


Figure 18

Shell completeness, % (green rings), and cumulative completeness, % (red line), as a function of  $(\sin \theta/\lambda)^2$ . The shells contain approximately the same number of reciprocal lattice points. Left (18a) EBB-5206, right (18b) EBB-5204. Up to the experimental resolution limit for each data set, set EBB-5206 has more than 1000 reflections missing, while set EBB-5206 only has 100 missing.

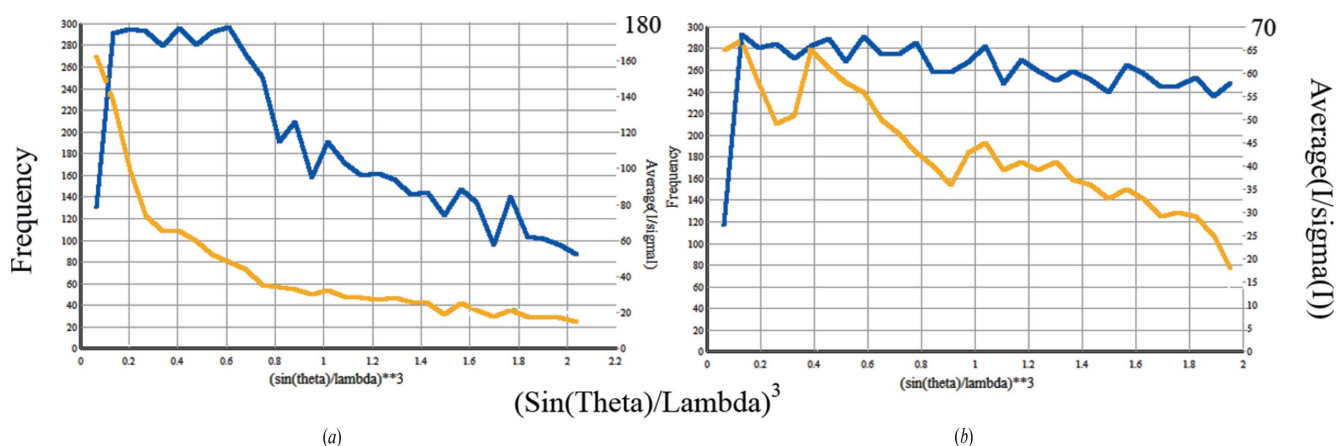


Figure 19

Frequency of occurrence of reflections with  $I > 10\sigma(I)$  (blue) and average  $I/\sigma(I)$  (orange) as a function of resolution. In the data for EBB-5206 (Fig. 19a, left) the average intensity (orange line) falls steadily with resolution, as does the number of strong reflections. For EBB-5204 (Fig. 19b, right) the average intensity starts with a similar trend to EBB-5206, but then at about 30° it rises again before falling steadily away.

EBB-5204 the data collection strategy was not homogeneous, and that the frame exposure time was increased for the high-angle data. There is a hint of a further increase in exposure time at about 45°, a feature more clearly seen in data sets EBB-5213 and EBB-5215. The number of reflections with  $I > 10\sigma(I)$  remains high right across the data set.

**Analysis of refinement residuals:** Both data sets seem to refine well, with *SHELX*-type weighting schemes achieving a goodness-of-fit sufficiently close to unity, Fig. 20.

Some insight into the deviations comes from examination of the weighted and unweighted residuals  $(F_o^2 - F_c^2)^2$  as a function of intensity and of resolution (Fig. 21). The very large number of medium intensity reflections (blue curve) dominates the determination of the parameters for the weighting scheme, which leads to over-weighting of the strong reflections (green bars in the top illustrations). The distribution of residuals as a function of resolution is not good for either data set, with the low-angle data (strong) being over weighted, and the high-angle underweighted. The role of the weighting scheme is to make the binned average value of the weighted residual approximately unity. For conventional data sets it is usually assumed that the principal contributors to  $(F_o^2 - F_c^2)^2$  are errors in  $F_o$ , but for these extended data sets it is possible that the usual independent spherical atom model emphasizes errors into  $F_c$ . A further complication may be that a single weighting scheme may not be appropriate when the data collections are not made under constant conditions.

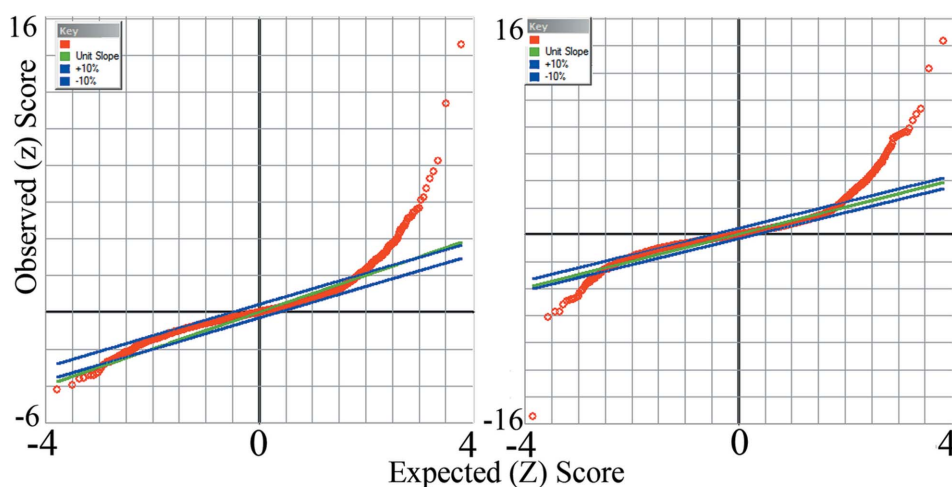
**Analysis of Friedel Residuals:** In spite of the unusual distribution of the structure amplitude residuals, the n.p.p. for the Friedel residuals were very linear with gradients close to unity (Fig. 22). Based on these, one would expect to obtain similar outcomes from post-refinement determination of the absolute structure of both EBB-5206 and EBB-5204.

In Table 7 we can see that for data set EBB-5204 the value for the Flack parameter determined directly or by post-refinement is not strongly affected by the weighting scheme. Refinement of EBB-5206 with simple statistical weights has a goodness-of-fit of 4.3, but gives a directly refined Flack ( $x$ ) of

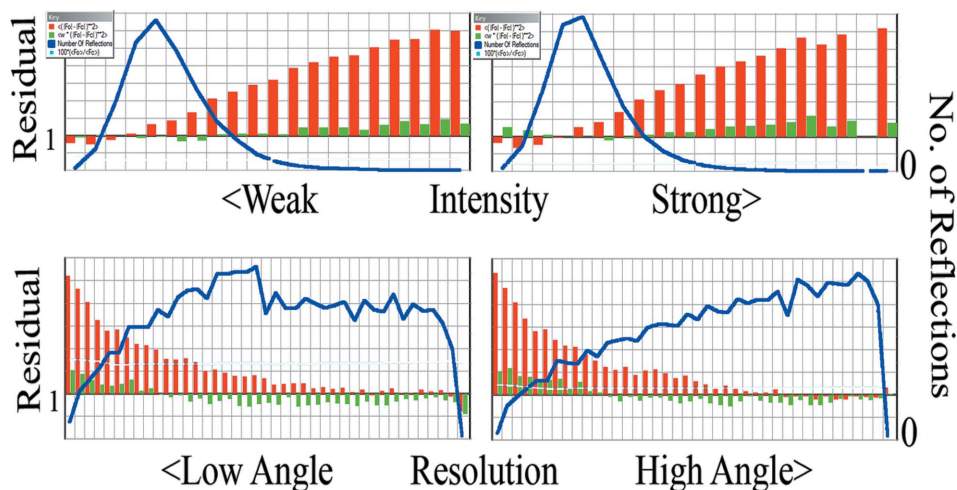
**Table 7**  
Absolute structure determinations for EBB-5206 and EBB-5204.

The Flack ( $x$ ) parameter was refined by conventional least squares with either *SHELX*-type or simple statistical weights. The resulting structure factors were post-refinement analysed with each weighting scheme.

Cpd	Main ref.	Post ref.	$wR2$	Flack ( $x$ )	Hole-in-one	Bijvoet ( $d$ )	Hooft ( $y$ )	Histogram ( $h$ )	Scaled	n.p.p.	Main_npp	GoF
5204	<i>SHELX</i> weights	STAT	0.0666	0.05 (18)	0.01 (06)	0.02 (07)	0.02 (06)	0.04 (06)	0.06	0.94	0.96	1.05
		LSQ			0.02 (18)	0.04 (18)	0.03 (06)	0.03 (17)	0.06	0.35		
	Statistical weights	STAT	0.0559	0.02 (17)	0.01 (06)	0.02 (07)	0.02 (06)	0.04 (06)	0.06	0.94	2.50	2.70
		LSQ			0.01 (06)	0.03 (07)	0.02 (06)	0.05 (06)	0.06	0.93		
5206	<i>SHELX</i> Weights	STAT	0.0678	0.29 (24)	-0.01 (08)	0.20 (08)	0.14 (08)	0.19 (08)	0.07	1.00	0.85	0.92
		LSQ			0.26 (27)	0.36 (27)	0.26 (09)	0.26 (26)	0.09	0.34		
	Statistical weights	STAT	0.0677	0.02 (34)	-0.03 (08)	0.20 (08)	0.14 (08)	0.20 (07)	0.07	1.00	3.62	4.34
		LSQ			0.01 (08)	0.20 (08)	0.14 (08)	0.19 (07)	0.07	0.99		



**Figure 20**  
Normal probability plots for EBB-5206 and EBB-5204. The central regions of the n.p.p. for the intensity residuals are reasonably linear with unit gradient, but both data sets show highly populated, strongly deviating, tails. Reflections lying outside of the band bordered by the blue lines are flagged as outliers.



**Figure 21**  
Analysis of residuals. Data for EBB-5206 on the left, for EBB-5204 on the right. The weighted (green bars) and unweighted (red bars) residuals are on a logarithmic scale. The blue curve represents the number of reflections in each interval. The resolution scale is with equal  $(\sin \theta/\lambda)^2$  bins. For good weights the green bars should form a narrow band across the centre of the plots. The top row shows that the fit is very good for the large number of relatively weak reflections, but poor for the strong intensities.

0.02 (34). Except for the hole-in-one method, other post-refinement procedures lead to larger values of the Flack parameter, but with smaller standard uncertainties. Refinement of EBB-5206 with *SHELX*-type weights give Flack ( $x$ ) of 0.29 (24), in agreement with the original authors. Post-refinement determinations, also using the *SHELX*-type weights, give much the same value for the Flack parameter, but with the Hooft and scaled histogram methods (which involve the gradient of the n.p.p.) giving reduced standard uncertainties. Except for hole-in-one, post-refinement methods using statistical weights yield slightly smaller Flack parameters and much reduced standard uncertainties. These results suggest that the anomalous reported value for the directly refined Flack ( $x$ ) parameter is a consequence of the weights used for the main refinement. The algorithm used to determine the coefficients in the weighting expression was unchanged for all the data sets we examined. We are led to suspect that the failure of this algorithm for data set EBB-5206 is due to the large number of missing reflections in the narrow band between 40 and 45°.

**Leverage analysis:** EBB tried to show that the reliability of an absolute structure analysis increases as the resolution of the data included in the analysis



Table 8

Post-refinement determination of the absolute structure parameter for EBB-5215 using data in the given  $(\sin(\theta)/\lambda)^2$  ranges and using weights from the experimental variances.

The column headed  $n$ -pair lists the number of Friedel pairs lying in each interval,  $n$ -used are the number of pairs after filtering, and  $H$ -used the number remaining after the robust filtering.

$(\sin \theta/\lambda)^2$ range	Bijvoet n.p.p.	Hole-in-one	Hooft ( $y$ )	Bijvoet ( $d$ )	Histogram ( $h$ ) (scaled s.u.)	$n$ -pair	$n$ -used	$H$ -used
0.00–1.75	1.07	0.04 (05)	−0.04 (05)	−0.03 (05)	−0.06 (05)(05)	4011	3973	3151
0.00–0.84	0.96	0.23 (07)	−0.02 (07)	−0.01 (07)	−0.04 (07)(06)	1443	1434	1143
0.84–1.33	1.15	−0.09 (07)	−0.06 (08)	0.01 (08)	0.01 (07)(08)	1296	1286	936
1.33–1.75	1.12	−0.19 (12)	−0.06 (14)	0.15 (13)	0.15 (12)(14)	1272	1253	929

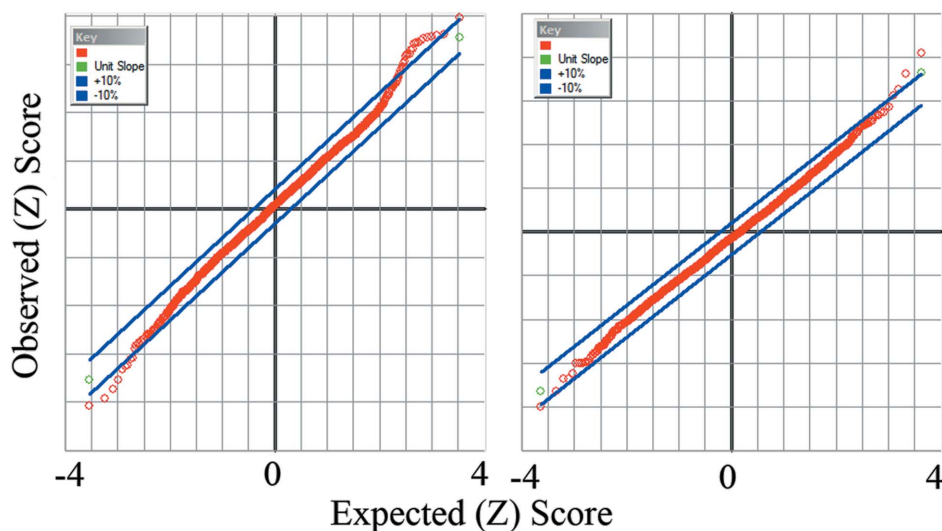


Figure 22

Normal probability plots of the Friedel pair residuals for EBB-5206 and EBB-5204. The plots are linear, with unit gradient and few outliers for both analyses.

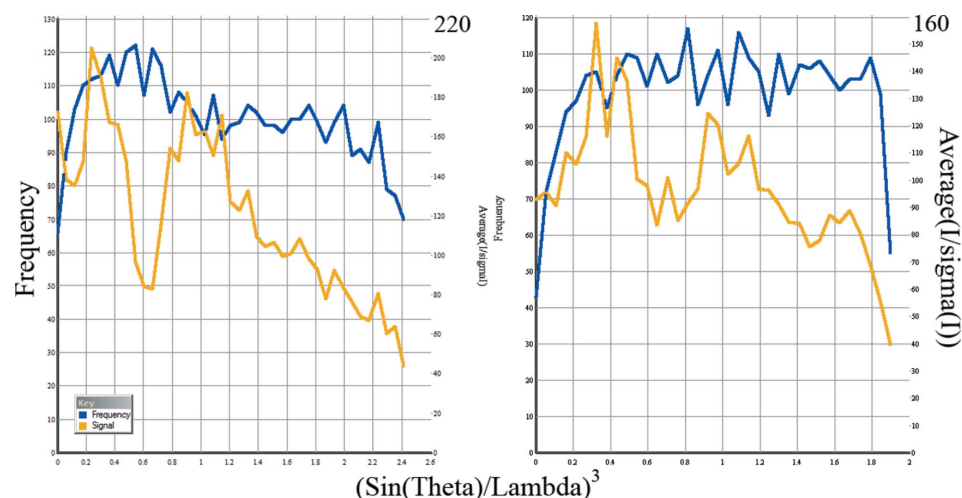


Figure 23

Signal-to-noise distributions. The orange curve is the scaled mean signal:noise ratio for each resolution bin, the blue curve is the number of reflections in each bin. The  $(\sin \theta/\lambda)^3$  bins contain approximately the same number of reciprocal lattice points, but some of the reflections have been eliminated by the reliability filters. Structures EBB-5215 (left) and EBB-5204 (right). The signal-to-noise curves are probably not unimodal because of increased exposure times at higher theta values.

increases. We used their data in a slightly different way, with different conclusions. The leverage of an individual reflection in a post-refinement absolute structure determination is proportional to the square of the signal-to-noise. A histogram of the mean signal-to-noise as a function of resolution should indicate where in the data set the most influential information lies. Fig. 23 is such a plot for EBB-5215 (redundancy = 8.2) and EBB-5204 (redundancy = 11.6).

The atomic and Flack ( $x$ ) parameters of EBB-5215 were refined using all the data and a *SHELX*-type weighting scheme [ $R = 0.023$ ,  $wR_2 = 0.066$ , Flack ( $x$ ) = 0.01 (17)]. The post-refinement Flack parameter was then determined using firstly all reflections, and then only the reflections in three non-overlapping resolution ranges, chosen to contain approximately the same number of reciprocal lattice points. Table 8 shows that for EBB-5215 all methods produce a steady increase in the standard uncertainty as the resolution band increases even though frame exposure times seem to have been increased.

5.2.6. PFW-fyo12e (PFW-2013). This material has previously been referred to in §4.3. Since the data were collected carefully, it was worth further exploring the cause for the difference between the standard uncertainties of the Flack parameter determined by direct and post-refinement methods. The steep gradient of the n.p.p. for the main refinement (4.64) with purely statistical weights suggests that the

standard uncertainties of the observations are severely underestimated. A plot of the internal *versus* the external sample standard uncertainties for the merged data is a rather dispersed straight line (gradient 1.4) showing that the manufacturer's estimates of individual uncertainties reflect reasonably well the dispersion between equivalent measurements (Fig. 24).

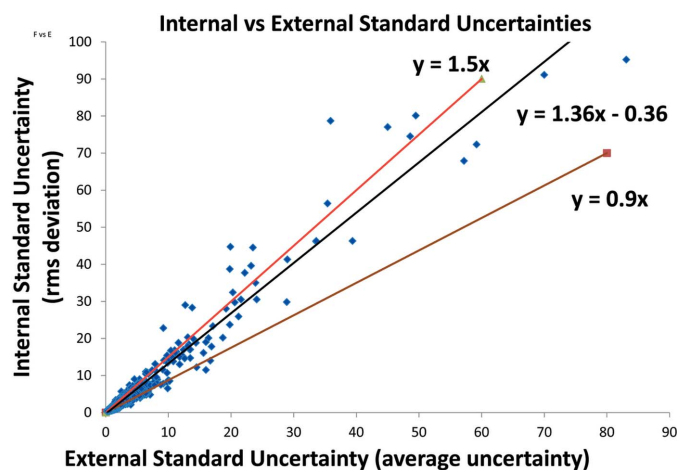
It was expected that a plot of the standard uncertainty of the mean  $\sigma(I)$  against  $I$  [equation (14)] for the merged data would approximate to a Poisson distribution (Evans, 2006). Instead, it was found to be a rather good straight line (Fig. 25) with a gradient of 0.045.

For Poisson statistics the signal:noise  $[I/\sigma(I)]$  can be increased by accumulating more photons. Diederichs (2010) had observed that  $[I/\sigma(I)]$  tended to a limiting value for synchrotron data. Plots of  $I/\sigma_{\text{mean}}(I)$  for PFW-fyo12e show a

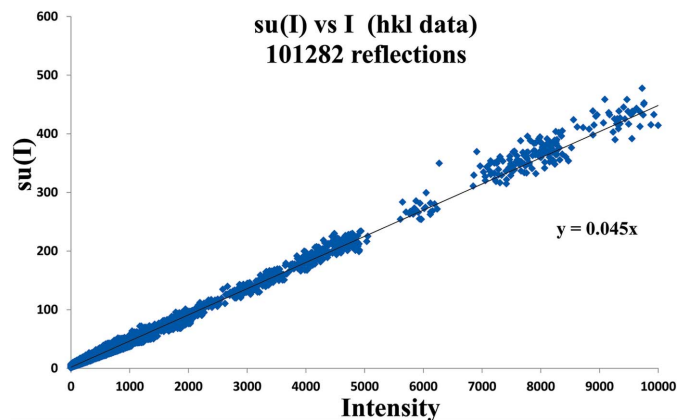
similar tendency, except that there appears to be two (or possibly three) limiting values (Fig. 26).

This raises the possibility that the data is not homogeneous, in the sense that it is derived from more than one experimental regime. A histogram of the frequency of distribution of redundancy is at least bimodal (Fig. 27), with a long low-order tail.

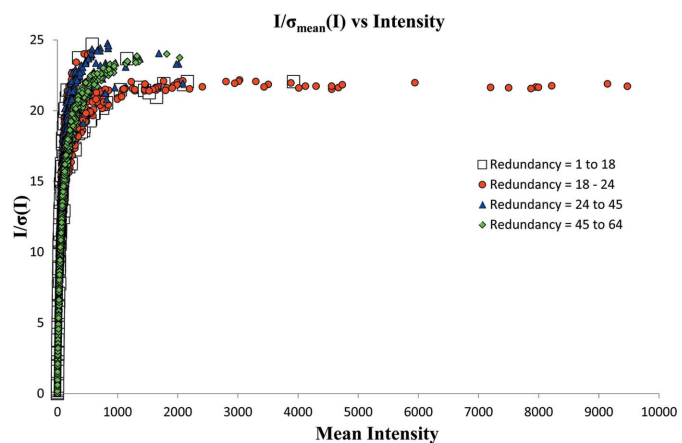
The intensities used during the least squares are usually the (weighted) means of a set of equivalent reflections. The variance of this mean is related to the sample variance [equation (11) or (12)] by  $1/(\text{redundancy})$ . Reflections measured 25 times have a variance almost twice as large as a reflection measured 45 times. It is possible that this variability of redundancy leads to the various asymptotic limits seen in Fig. 26. Whatever the individual variabilities of the standard uncertainties of the means, on average  $\sigma(F^2) \simeq 0.05F^2$  (from



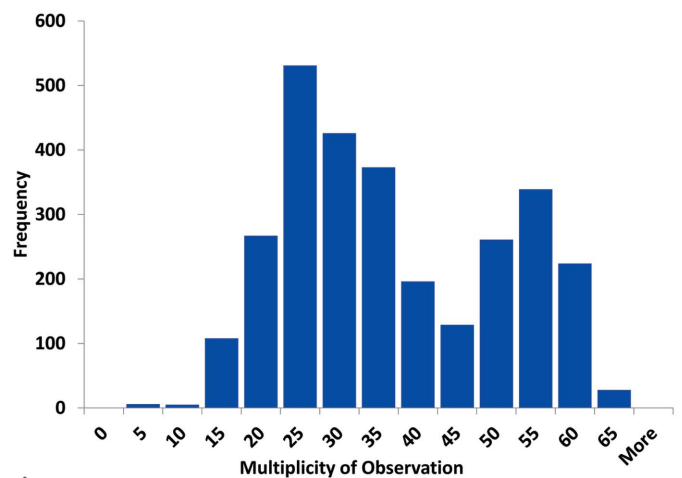
**Figure 24**  
PFW-fyo12e. Plot of the internal standard uncertainty (r.m.s. deviation from group means) against the external uncertainty (mean of reported standard uncertainties). The dispersion around the overall trend line (black) is bound by two trends, one (mainly with small average uncertainties) with a gradient of about 0.9, and the other with a gradient of 1.5.



**Figure 25**  
PFW-fyo12e. Standard uncertainty *versus* mean intensity from the deposited *hkl* file.  $\sigma(I)$  is almost proportional to  $(I)$  rather than proportional to  $\sqrt{(I)}$ .



**Figure 26**  
PFW-fyo12e.  $I_{\text{mean}}/\sigma(I_{\text{mean}})$  *versus* mean intensity. The asymptotic limits are approximately 22 for the data with redundancy 18–24 (red), 24 for redundancy 24–45 (green) and 25 for redundancy 45–64 (blue). The low redundancy data do not fit any pattern.



**Figure 27**  
PFW-fyo12e. Frequency distribution of the redundancy for fyo12e. The most frequent values for redundancy are 25 and 45. One reflection, (9 0 1), was measured only once, two reflections (2 2 10 and 2 4 8) were each measured 64 times.

Fig. 25). The  $a$  and  $b$  terms in the *SHELX*-type weighting scheme are 0.044 and 0.291, so that it is these which dominate the weighting during refinement. Weighting the co-refinement of data measured under different regimes (for example, widely differing redundancies) may warrant further investigation, a situation alluded to in Bernardinelli & Flack (1987).

5.2.7. PFW *R*-mandelic acid (PFW-2013). PFW-2013 report that this material (Friedif = 35) crystallized as plates which on cooling to 150 K showed evidence of strain broadening, so that the actual data collection was performed at 220 K. 32 194 reflections were measured, yielding 2860 unique reflections, an average redundancy of 11.3.  $R_{\text{int}}$  was 0.04, indicating a fair level of self-consistency amongst the data. The final  $R$  factor, 0.0549, was higher than one would expect for this type of material, but might be explained by the strain broadening. There are two molecules in the asymmetric unit, differing by a small rotation about the single bond to the phenyl group, and no evidence for disorder. The  $\sigma(\text{Flack}(x)) = 0.37$  greatly exceeds the  $\sigma(\text{Flack}(h)) = 0.05$ , in spite of the significant value of Friedif. The n.p.p. for the main refinement residuals (Fig. 28) is far from ideal, suggesting a problem with the data, the weights or the model itself.

Alternative weighting schemes to a *SHELX*-type formula did not significantly improve the n.p.p. Although the program *DIFABS* (Walker & Stuart, 1983), once used as a method for estimating empirical absorption corrections, has long been replaced by the use of multi-scan methods, it still provides a useful diagnostic tool. The program fits a smoothly varying function of azimuth and declination of the incident and emergent beams to the residual between  $|F_o|$  and  $|F_c|$ , the so-called *absorption surface*. For merged area detector data there are no 'incident' or 'emergent' beams, but these can be replaced by the scattering vector to generate a visualization of the residual. If the multi-scan procedure has adequately modelled absorption and illuminated volume effects, variations of this surface from unity will indicate that there are

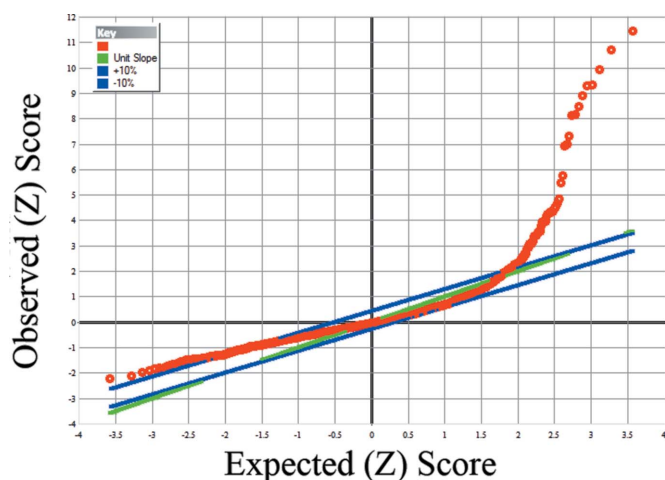


Figure 28  
*R*-Mandelic acid normal probability plot for the published model using *SHELX*-type weights. The gradient is 0.85, but there is a long tail of outliers.

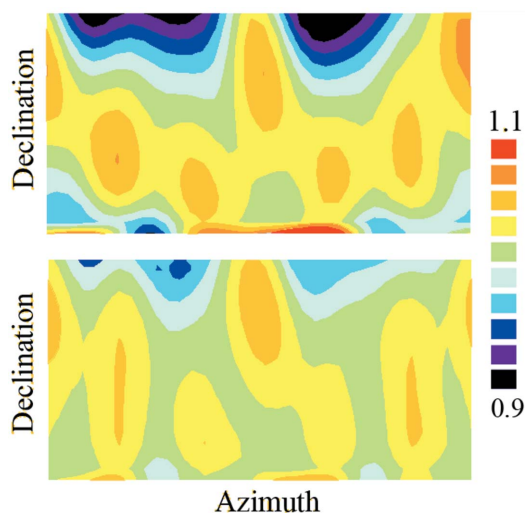


Figure 29  
*R*-Mandelic acid. Variation of  $F_o$  and  $F_c$  as a function of orientation of the diffraction vector. 29a (top) is for the published structure. 29b (bottom) shows the surface much flattened by the inclusion of twinning in the model.

problems with the model, or undetected errors in the data. Fig. 29(a) is the plot for *R*-mandelic acid. It shows variations between 0.9 and 1.1 with some very sharp gradients, indicating that there is a problem with the analysis.

The  $F_o - F_c$  plot (Fig. 30) is a fair straight line with unit gradient, and without any very outstanding outliers. However, although the distribution is bounded by a reasonably well defined lower edge, the upper edge is distinctly ragged. This condition, together with the *DIFABS* plot, is often symptomatic of twinning.

ROTAX analysis (Cooper *et al.*, 2002) indicated the possibility of twinning by the law  $[1,0,0; 0,-1,0; -0.8,0,-1]$ . Refinement including this twinning reduced the  $R$ -factor to 0.0496 and greatly improved the  $F_o$  versus  $F_c$  plot and *DIFABS* surface (Fig. 29b), but did little to improve the n.p.p. If the components of the non-inversion twin are labelled  $A$  and  $B$ , and corresponding inversion twin components by  $a$  and  $b$ , then refinement was continued with the constraint that  $A + B + a + b = 1.0$  and the restraint  $0.000(1) = b - (aB/A)$  on the assumption that the inversion twinning ratios ( $a/A$  and  $b/B$ ) are the same. The refined twin element scale factors are  $A = 0.8$  (2),  $B = 0.13$  (4),  $a = 0.0$  (2),  $b = 0.01$  (4).<sup>3</sup> Inclusion of inversion twinning had no effect on the  $R$ -factor. For the moment the post-refinement absolute structure analysis in *CRYSTALS* will not handle non-inversion twinning.

5.2.8. FSTW-YIFZAP (Gowda *et al.*, 2007). This material, falling at the bottom of Table 1 in FSTW (Flack *et al.*, 2011), caught the interest of those authors because of the small variation of  $R_D (= \Sigma|D_{\text{obs}} - D_{\text{model}}|/\Sigma|D_{\text{obs}}|)$ , summed over the Friedel pairs) as a function of imposed values for the Flack ( $x$ ) parameter. They came to the conclusion that the reported uncertainty in the Flack parameter was *very grossly under-*

<sup>3</sup> The twin scale factors sum to unity if quoted to three decimal places [0.811 (218), 0.133 (036), 0.048 (218), 0.008 (035)].

Table 9

Comparison of published results for FSTW-YIFZAP with new results obtained from *SHELXL-2013/2* and *CRYSTALS*.

$N_{\text{ref}}$  is the number of reflections with  $I > 2\sigma(I)$ ,  $N_{\text{ref}}(\text{all})$  is for all data, and  $N_{\text{ref}}(\text{post})$  in the number of reflections used in the post-refinement analysis.

Program	R1	$N_{\text{ref}}$	R1all	$N_{\text{ref}}(\text{all})$	wR2	Inversion twin	Hole-in-one	$N_{\text{ref}}(\text{post})$	Parsons or Bijvoet	Histogram
Published	0.074	621	0.104	1003	0.260	−0.1 (3)				
<i>SHELXL13</i>	0.088	639	0.117	1054	0.279	−0.94 (49)				
<i>SHELXL13</i>							−0.67 (43)	173	0.35 (42)	
<i>CRYSTALS</i>	0.087	632	0.116	1052	0.276	−0.95 (38)		281	0.35 (04)	0.55 (04)

estimated. Using the CIF and .fcf files recovered from the IUCr we were unable to reproduce with *CRYSTALS* some of the results recorded elsewhere in the CIF. *PLATON* was used to convert the CIF file to a *SHELXL* ins (data) file, and the refinements repeated with *SHELXL-2013/2*, using the *TWIN/*

*BASF* commands or using the hole-in-one and Quotient methods for estimating the Flack ( $x$ ) parameter (Table 9).

The *SHELXL* and *CRYSTALS* analyses are reasonably compatible, but in poor agreement with the published values. In the absence of evidence to the contrary, we attribute this conflict to the fact that the original authors were able to use the full precision of the reflection data stored in an .hklf file, but for the re-calculations we had to use the limited precision of the .fcf file. Whatever the source of the discrepancy, it remains clear that while direct refinement of Flack ( $x$ ) in *CRYSTALS* and the inversion twin in *SHELXL-2013/2* lead to very similar results, these are quite different values from the post-refinement methods. The usual diagnostic tools were used to try to locate the source of the discrepancy. The gradient of the n.p.p. (Fig. 31a) was 0.91, but with substantial displacement from the origin of the graph, which usually implies a feature in the data which cannot be matched by the model. The n.p.p. for the post-refinement analysis of the Friedel differences (Fig. 31b) had a least-squares gradient of 4.7. Examination of the plot showed that very many of the reflections in the central region lay on a line of unit gradient, but there were substantial numbers of outliers at the extremes of the plot. From other analyses, we have seen that the calculated Friedel differences are only weakly correlated with the atomic parameters, so we must assume that the non-linearity of the n.p.p. is due either to errors on the observed Friedel differences or in their standard uncertainties.

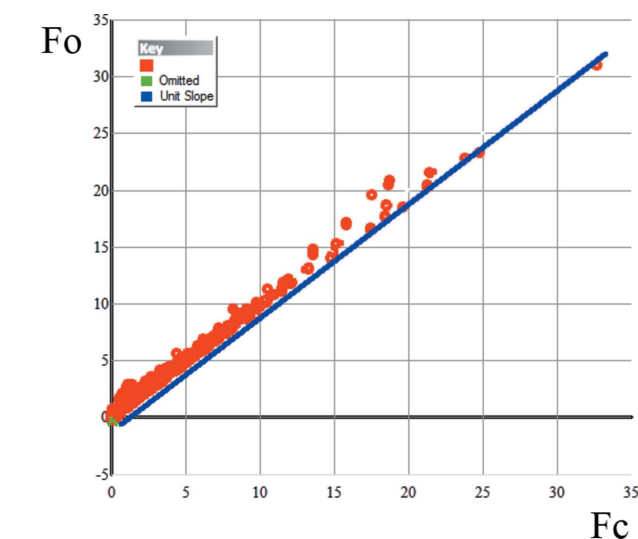


Figure 30 *R*-Mandelic acid. Plot of  $F_o$  versus  $F_c$ . The blue unit gradient line helps show that the lower edge of the distribution is well defined when compared with the upper boundary.

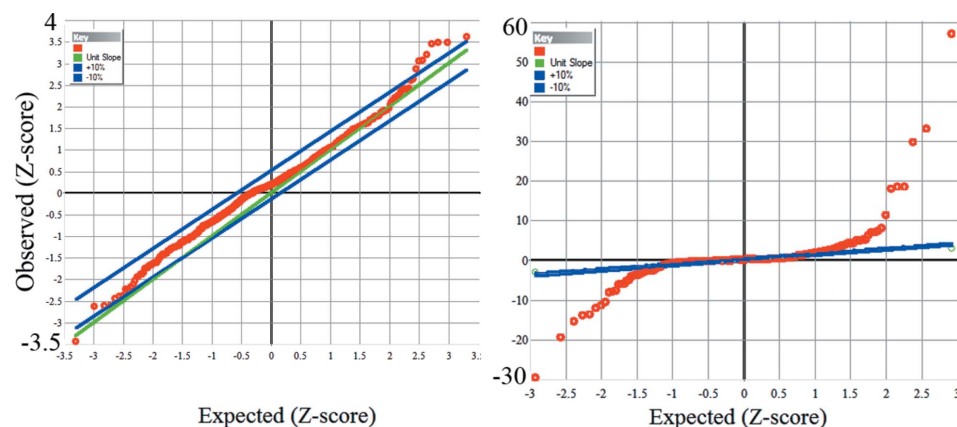


Figure 31 FSTW-YIFZAP. Left: Normal probability plot for the main refinement. There is a distinct step near the origin together with some outliers at the extremes. Right: Normal probability plot for the post-refinement analysis of the Friedel differences. The central region has a unit gradient, but the very large deviations in both tails lead to a meaningless best straight line. Either the observed Friedel differences or the weights computed from their standard uncertainties may be in error.

Examination of the *DIFABS* map, Fig. 32, showed deep hollows and high peaks with a maximum ratio of 1:1.77. This could be indicative of uncorrected absorption. The authors give the crystal size as  $0.52 \times 0.46 \times 0.09$  mm – a thin plate – and used an analytical correction by the method of Clark & Reid (1995) giving minimum and maximum corrections of 0.86 and 1.16, a ratio of 1:1.35.

The  $F_o$  versus  $F_c$  plot was only weakly indicative of twinning, and *ROTAX* suggested an unconvincing twin operation 1,0,0.734; 0,−1,0; 0,0,1. Refinement with this twin law gave a major component of 0.88 (3). The text of the article made no mention of twinning but

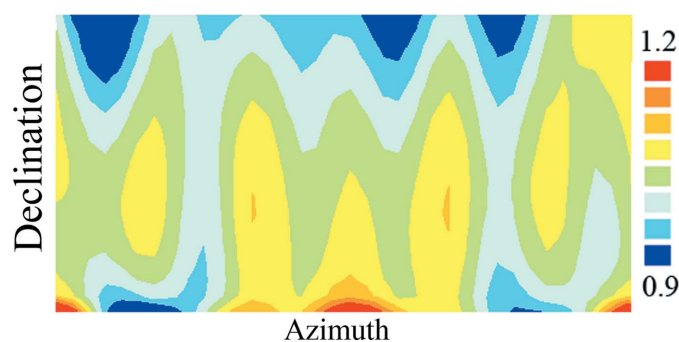
the CIF file contained an entry for the Flack parameter and its standard uncertainty. Because of this, *PLATON* had added the necessary TWIN/BASF instructions to the *SHELXL* instruction file. Attempts to refine the non-merohedral twinning and inversion twinning in *CRYSTALS* failed, the normal matrix becoming singular in spite of the application of appropriate restraints and constraints.

At this point we retrieved the supporting information. From this it was clear that the original authors had detected the same twin law as ROTAX, and had refined this model to a minor element of 0.15 using an HKLF5 reflection file. Strangely, in spite of the Flack entry  $[-0.1(3)]$  in the deposited CIF, the supporting information states ‘*Owing to the poor quality of the data, the absolute structure couldn’t be reliably defined and any references to the Flack parameter have been omitted*’.

Our analysis of the data was repeated using the twinned model, but showed no great improvement in the n.p.p. nor the *DIFABS* surface. The data had been collected with an area detector, standard source and graphite monochromator, so that unless the authors had used a very fine collimator, one might expect the crystal to have been more-or-less fully bathed in the direct beam. 2613 reflections were measured, merging down to 1003 independent observations ( $R_{\text{int}} = 0.086$ ) of which 621 had  $I > 2\sigma(I)$ . Seeing that over 30% of the data could be classed as very weak, the observed and calculated Wilson plots were examined, Fig. 33. The up-turn in the plot of the observed data at about  $\rho = 0.3$  is often characteristic of data being measured to a resolution at which there is little or no signal amongst the noise.

## 6. Conclusions

X-ray crystallography is unique in that it provides both an estimate of the enantiopurity of a sample, and a standard uncertainty for that estimate without special user action. Chiroptical spectroscopies look at a total signed signal and thus require a reference spectrum to compare against in order to judge the proportions of each enantiomer. When this is available then typically the resolution is *ca* 1%. NMR with shift reagents can give separate signals for each enantiomer, but there are substantial complications about the binding of

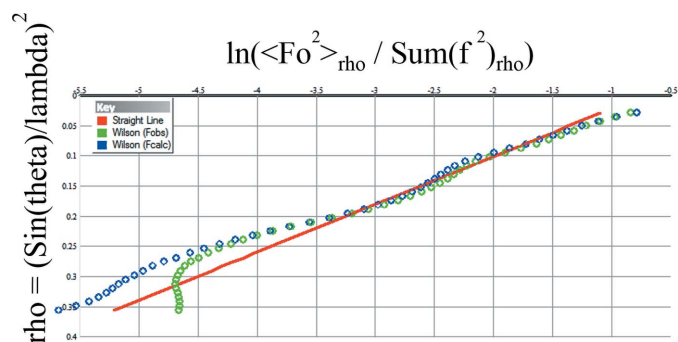


**Figure 32**  
*DIFABS* surface for FSTW-YIFZAP. The maximum excursions occur at similar, low, declinations.

the shift reagent, equilibria *etc.* Chiral HPLC has the advantage of actually separating the enantiomers as individual signals that can be directly ratioed and so can be very deterministic. In many cases one should be able to detect a 1 m.a.u. (a.u. = atomic unit) signal from an enantiomeric impurity alongside a signal of 1 a.u. of the main peak, *i.e.* 0.1%. These techniques are degraded in the presence of impurities. Except for the case of twinning by inversion, crystallography largely avoids the impurity problem, but suffers in that one crystal is taken as representative of the bulk sample. However, for materials known to be enantiopure or to have a large enantiomeric excess, it can be a robust way for assigning the absolute structure of the major (or only) component.

The results of Thompson & Watkin (2011) showed that even in apparently unsuitable cases, there was usually some resonant signal amongst the random noise and systematic errors. Flack used the  $2A/D$  plots to try to visualize the signal. The plots in this paper of  $D_o$  and  $D_s$  versus  $\sigma(D_o)$  provide a clear indication of the best possible signal in the data, and the actual signal in the observed data. We know that the value of the Flack parameter must lie in the interval 0–1 and in favourable cases histograms of the Flack  $x$  peak in this interval. The broader the spread about this interval, the less reliable the estimate of  $\sigma(x)$ . The ratio  $D_s/\sigma(D_o)$  is a measure of the information content of a reflection. Measuring data to high resolution increases  $D_s/I_s$ , but only improves the leverage if care is taken to minimize  $\sigma(D_o)$ .

Direct refinement of the Flack ( $x$ ) parameter usually results in a value with a larger standard uncertainty than that obtained by post-refinement methods using weights derived from the observed variances, making these latter methods more attractive for publication. However, the value of the Flack ( $x$ ) parameter and its standard uncertainty obtained by free refinement in the main least squares should be compared with the values obtained by a post-refinement method. Substantial differences indicate that there may be a problem with the data or with the proposed model, although other techniques will have to be used to identify the problem.



**Figure 33**  
Wilson plot for observed and calculated data from FSTW-YIFZAP. Note that the calculated data (blue) show a steady trend from right to left, while there is a sudden up-turn in the observed data (green) at about  $\rho = 0.3$ . This up-turn is often characteristic of an attempt to measure data to too high a resolution for the given crystal size, X-ray source and exposure time. The high-angle data is essentially just noise (Weiss, 2001).

In the absence of widespread availability of software able to refine structures using both averages and differences of structure amplitudes as observations, the low correlation between the structural parameter values and the Flack  $x$  suggests that a post-refinement estimate of the Flack parameter once the model is fully parameterized can be used to guide the final refinement. The Bijvoet difference method is a good diagnostic for problems with the data or model since it contains the minimum number of assumptions: the Hooft and Parsons methods both allow for some problems with the data or model and so may be most suitable for routine work. If the Parsons quotient and the Bijvoet difference methods give substantially different results, this may be indicative of absorption or other problems with the main refinement. If there is doubt about the enantiopurity of the material, the Flack parameter must be included as part of the model. It can either be refined freely, treated as a constant (constraint) with the value taken from the post-refinement analysis, or a single equation of restraint on the Flack ( $x$ ) parameter can be introduced using the post-refinement estimate of its value and standard uncertainty as target values.

## APPENDIX A

### Ratios of averages and averages of ratios

For an individual Friedel pair we can write equation (4) as

$$x = 0.5 - D_o/(2D_s). \quad (19)$$

Defining

$$c_i = D_{oi}/D_{si} \quad (20)$$

we obtain

$$x_i = 0.5 - 0.5c_i. \quad (21)$$

#### A1. Ratio of averages

Writing (20) as

$$[D_s](c) = [D_o], \quad (22)$$

where the terms in square brackets are column vectors of the model and observed Friedel differences, the least-squares estimate of  $c$  from a set of Friedel pairs is

$$[D_s]^T [D_s](c) = [D_s]^T [D_o] \quad (23)$$

from which a weighted value for  $\langle c \rangle$  can be obtained as

$$\langle c \rangle = \frac{\sum w_i D_{oi} D_{si}}{\sum w_i D_{si} D_{si}} \quad (24)$$

with

$$w_i = 1/\sigma^2 D_{oi}. \quad (25)$$

Equation (24) can be rewritten as a ratio of averages

$$\langle c \rangle = \frac{\frac{1}{n} \sum w_i D_{oi} D_{si}}{\frac{1}{n} \sum w_i D_{si} D_{si}}. \quad (26)$$

Letting

$$\Delta_i = D_{oi} - \langle c \rangle D_{si} \quad (27)$$

gives

$$\sigma^2(c) = \frac{\sum w_i \Delta_i^2}{n \sum w_i D_{si}^2} \quad (28)$$

from which

$$\langle x \rangle = (1 - \langle c \rangle)/2 \quad (29)$$

and

$$\sigma(\langle x \rangle) = \sigma(\langle c \rangle)/2. \quad (30)$$

#### A2. Average of ratios

Alternatively, we can evaluate individual  $c_i$  from equation (20) and  $x_i$  from (21), and form the (weighted) average of these ratios

$$x_i = (D_{si} - D_{oi})/2D_{si} \quad (31)$$

$$x' = \sum w_i x_i / \sum w_i \quad (32)$$

$$\sqrt{w_i} = 1/\sigma(x_i) = 2D_{si}/\sigma(D_{oi}). \quad (33)$$

Following Blessing & Langs (1987) we can form the internal and external estimates of the variance of the sample, and hence the variance of the average

$$\sigma_{\text{int}}^2(x) = \frac{n}{n-1} \frac{\sum w_i (x_i - x')^2}{\sum w_i} \quad (34)$$

$$\sigma_{\text{ext}}^2(x) = \frac{\sum w_i \sigma^2(x_i)}{\sum w_i} \quad (35)$$

$$\sigma^2(x') = \frac{\sigma_{\text{int}}^2}{n} \text{ or } \frac{\sigma_{\text{ext}}^2}{n}. \quad (36)$$

For a list of paired observations, the ratio of averages and the average of ratios will be the same if there is a linear relationship between the observations and the error distributions are similar. A difference between these two statistics indicates a problem that should be investigated.

## APPENDIX B

### Standard uncertainties of differences and quotients

The relation between the absolute structure of a crystalline material and the measured Friedel pairs was given in equation (2)

$$I_o \simeq I_c = (1-x)I_s^+ + xI_s^-. \quad (37)$$

Writing

$$\begin{aligned}D_o &= I_o^+ - I_o^- \\D_s &= I_s^+ - I_s^- \\A_o &= 0.5(I_o^+ + I_o^-) \\A_s &= 0.5(I_s^+ + I_s^-) \\Q_o &= D_o/2A_o \\Q_s &= D_s/2A_s\end{aligned}$$

we obtain the difference expression

$$D_o \simeq D_c = (1 - 2x)D_s \quad (38)$$

and the quotient expression

$$\begin{aligned}D_o/2A_o &\simeq (1 - 2x)D_s/2A_s \\Q_o &\simeq Q_c = (1 - 2x)Q_s\end{aligned} \quad (39)$$

from which

$$\sigma^2(D_o) = \sigma^2(I_o^+) + \sigma^2(I_o^-) \quad (40)$$

$$\sigma^2(A_o) = 0.25\sigma^2(D_o) \quad (41)$$

$$\sigma^2(Q_o) = \left[ \frac{2}{(I_o^+ + I_o^-)^2} \right]^2 [I_o^{-2}\sigma^2(I_o^+) + I_o^{+2}\sigma^2(I_o^-)]. \quad (42)$$

If  $I_o^+ \simeq I_o^- = I_o = A_o$  and  $\sigma^2(I_o^+) \simeq \sigma^2(I_o^-) = \sigma^2(I_o)$ , we get

$$\sigma^2(Q_o) = \sigma^2(I_o)/(2I_o^2)$$

and hence

$$Q_o/\sigma(Q_o) = D_o/\sigma(D_o)$$

and

$$Q_o/\sigma^2(Q_o) = 2I_oD_o/\sigma^2(D_o).$$

## Acknowledgements

The authors wish to thank Howard Flack for critical advice, to Simon Parsons for software which enabled us to verify some calculations, to Ton Spek for code from *PLATON*, to George Tranter (Chiralabs Ltd) for advice on non-X-ray techniques, and many colleagues for suggesting additions to the manuscript. Figs. 8–12 and 24–27 were created using Microsoft Excel 2010. All others are lightly retouched screen-dumps from *CRYSTALS*. *SHELXL* calculations were made with *SHELXL*-2013/2, *CRYSTALS* calculations with the executable dated 17/12/2015 08:36.

## References

- Abrahams, S. C. & Keve, E. T. (1971). *Acta Cryst.* **A27**, 157–165.  
 Abud, J. E., Sartoris, R. P., Calvo, R. & Baggio, R. (2011). *Acta Cryst.* **C67**, m130–m133.  
 Bernardinelli, G. & Flack, H. D. (1987). *Acta Cryst.* **A43**, 75–78.  
 Blessing, R. H. & Langs, D. A. (1987). *J. Appl. Cryst.* **20**, 427–428.  
 Carruthers, J. R. & Watkin, D. J. (1979). *Acta Cryst.* **A35**, 698–699.  
 Clark, R. C. & Reid, J. S. (1995). *Acta Cryst.* **A51**, 887–897.  
 Cooper, R. I., Gould, R. O., Parsons, S. & Watkin, D. J. (2002). *J. Appl. Cryst.* **35**, 168–174.  
 Cooper, R. I., Watkin, D. J. & Flack, H. D. (2016). *Acta Cryst.* **C72**, 261–267.

- Cruickshank, D. W. J. (1961). *Computing Methods and the Phase Problem*, edited by R. Pepinsky, J. M. Robertson & J. C. Speakman, paper 6. Oxford: Pergamon Press.  
 Cruickshank, D. W. J. & McDonald, W. S. (1967). *Acta Cryst.* **23**, 9–11.  
 Cruickshank, D. W. J. & Robertson, A. P. (1953). *Acta Cryst.* **6**, 698–705.  
 Diederichs, K. (2010). *Acta Cryst.* **D66**, 733–740.  
 Dyadkin, V., Wright, J., Pattison, P. & Chernyshov, D. (2016). *J. Appl. Cryst.* **49**, 918–922.  
 Ealick, S. E., Van der Helm, D. & Weinheimer, A. J. (1975). *Acta Cryst.* **B31**, 1618–1626.  
 Engel, D. W. (1972). *Acta Cryst.* **B28**, 1496–1509.  
 Escudero-Adán, E. C., Benet-Buchholz, J. & Ballester, P. (2014). *Acta Cryst.* **B70**, 660–668.  
 Evans, P. (2006). *Acta Cryst.* **D62**, 72–82.  
 Fábry, J., Fridrichová, M., Dušek, M., Fejfarová, K. & Krupková, R. (2012). *Acta Cryst.* **C68**, o76–o83.  
 Ferguson, G., Glidewell, C., Low, J. N., Skakle, J. M. S. & Wardell, J. L. (2001). *Acta Cryst.* **C57**, 315–316.  
 Flack, H. D. (1983). *Acta Cryst.* **A39**, 876–881.  
 Flack, H. D. (2013). *Acta Cryst.* **C69**, 803–807.  
 Flack, H. D. & Bernardinelli, G. (2008). *Acta Cryst.* **A64**, 484–493.  
 Flack, H. D., Bernardinelli, G., Clemente, D. A., Linden, A. & Spek, A. L. (2006). *Acta Cryst.* **B62**, 695–701.  
 Flack, H. D., Sadki, M., Thompson, A. L. & Watkin, D. J. (2011). *Acta Cryst.* **A67**, 21–34.  
 Gowda, B. T., Nayak, R., Kožíšek, J., Tokarčík, M. & Fuess, H. (2007). *Acta Cryst.* **E63**, o2967.  
 Hamilton, W. C. (1965). *Acta Cryst.* **18**, 502–510.  
 Hooft, R. W. W., Straver, L. H. & Spek, A. L. (2008). *J. Appl. Cryst.* **41**, 96–103.  
 Hooft, R. W. W., Straver, L. H. & Spek, A. L. (2010). *J. Appl. Cryst.* **43**, 665–668.  
 Howard, S. T., Hursthouse, M. B., Lehmann, C. W., Mallinson, P. R. & Frampton, C. S. (1992). *J. Chem. Phys.* **97**, 5616–5630.  
 Le Page, Y., Gabe, E. J. & Gainsford, G. J. (1990). *J. Appl. Cryst.* **23**, 406–411.  
 Merli, M. & Sciascia, L. (2011). *Acta Cryst.* **A67**, 456–468.  
 Müller, G. (1988). *Acta Cryst.* **B44**, 315–318.  
 Olthof-Hazekamp, R. (1990). *Xtal 3.0 Reference Manual*, edited by R. S. Hall & J. M. Stewart. University of Western Australia, Perth.  
 Parrish, W. (1960). *Acta Cryst.* **13**, 838–850.  
 Parsons, S., Flack, H. D. & Wagner, T. (2013). *Acta Cryst.* **B69**, 249–259.  
 Parsons, S., Pattison, P. & Flack, H. D. (2012). *Acta Cryst.* **A68**, 736–749.  
 Parsons, S., Wagner, T., Presly, O., Wood, P. A. & Cooper, R. I. (2012). *J. Appl. Cryst.* **45**, 417–429.  
 Prince, E. (1994). *Mathematical Techniques in Crystallography and Material Science*, pp. 80–82. Berlin: Springer-Verlag.  
 Prince, E. (2004). *Mathematical Techniques in Crystallography and Material Science*, 3rd ed., p. 121. Berlin: Springer-Verlag.  
 Rabinovich, D. & Hope, H. (1980). *Acta Cryst.* **A36**, 670–678.  
 Rogers, D. (1981). *Acta Cryst.* **A37**, 734–741.  
 Seela, F., Xiong, H., Budow, S., Eickmeier, H. & Reuter, H. (2012). *Acta Cryst.* **C68**, o174–o178.  
 Sheldrick, G. M. (2014). Personal communication.  
 Sheldrick, G. M. (2015). *Acta Cryst.* **A71**, 3–8.  
 Smith, M. & Lamb, A. (2012). Personal communication. Oxford Archive No. 6418, C21, H23 Br N2 O4.  
 Thompson, A. L. & Watkin, D. J. (2011). *J. Appl. Cryst.* **44**, 1017–1022.  
 Tukey, P. J. W. (1976). *Proceedings of the First ERDA Statistical Symposium*, edited by W. L. Nicholson & J. L. Harris. Ohio: Battelle, Pacific Northwest Laboratories.  
 Walker, N. & Stuart, D. (1983). *Acta Cryst.* **A39**, 158–166.  
 Weiss, M. S. (2001). *J. Appl. Cryst.* **34**, 130–135.  
 Zhang, W., Oliver, A. G. & Serianni, A. S. (2012). *Acta Cryst.* **C68**, o7–o11.