



# THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### Use of intensity quotients and differences in absolute structure refinement

**Citation for published version:**

Parsons, S, Flack, H & Trixie, W 2013, 'Use of intensity quotients and differences in absolute structure refinement' Acta Crystallographica Section B Structural Science Crystal Engineering and Materials, vol 69, no. Part 3, pp. 249–259. DOI: 10.1107/S2052519213010014

**Digital Object Identifier (DOI):**

[10.1107/S2052519213010014](https://doi.org/10.1107/S2052519213010014)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Publisher's PDF, also known as Version of record

**Published In:**

Acta Crystallographica Section B Structural Science Crystal Engineering and Materials

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



Simon Parsons,<sup>a\*</sup> Howard D.  
 Flack<sup>b</sup> and Trixie Wagner<sup>c</sup>
<sup>a</sup>EaStCHEM School of Chemistry and Centre for  
 Science at Extreme Conditions, The University of  
 Edinburgh, King's Buildings, West Mains Road,  
 Edinburgh EH9 3JJ, Scotland, <sup>b</sup>Chimie minérale,  
 analytique et appliquée, University of Geneva,  
 Geneva, Switzerland, and <sup>c</sup>Novartis Institutes for  
 BioMedical Research, 4002 Basel, Switzerland

Correspondence e-mail: s.parsons@ed.ac.uk

# Use of intensity quotients and differences in absolute structure refinement

 Received 5 February 2013  
 Accepted 11 April 2013

Several methods for absolute structure refinement were tested using single-crystal X-ray diffraction data collected using Cu  $K\alpha$  radiation for 23 crystals with no element heavier than oxygen: conventional refinement using an inversion twin model, estimation using intensity quotients in *SHELXL2012*, estimation using Bayesian methods in *PLATON*, estimation using restraints consisting of numerical intensity differences in *CRYSTALS* and estimation using differences and quotients in *TOPAS-Academic* where both quantities were coded in terms of other structural parameters and implemented as restraints. The conventional refinement approach yielded accurate values of the Flack parameter, but with standard uncertainties ranging from 0.15 to 0.77. The other methods also yielded accurate values of the Flack parameter, but with much higher precision. Absolute structure was established in all cases, even for a hydrocarbon. The procedures in which restraints are coded explicitly in terms of other structural parameters enable the Flack parameter to correlate with these other parameters, so that it is determined along with those parameters during refinement.

## 1. Introduction

When applied in crystallography the term *absolute structure* refers to the spatial arrangement of the atoms of a physically identified non-centrosymmetric crystal and its description by way of unit-cell dimensions, space group and representative coordinates of all atoms (Flack & Bernardinelli, 1999, 2008*b*). Since inverted images of a non-centrosymmetric crystal structure are different, the question of absolute structure arises during analysis of any non-centrosymmetric crystal structure. The most important practical application of absolute structure refinement is, however, in the crystallographic determination of the absolute configuration of chiral molecules.

The fact that absolute structure can be obtained at all in a crystal structure determination is the result of resonant scattering, also known as anomalous scattering or anomalous dispersion, which introduces small differences in intensity between reflections  $\mathbf{h}$  and  $\bar{\mathbf{h}}$  which carry the information on absolute structure. Methods for absolute structure determination most commonly used today are based on a formulation first described by Flack (1983), in which the crystal under investigation is considered to be an inversion twin in which the reference domain has the absolute structure of the current refinement model, and the other domain is inverted. Measured intensities are then modelled according to

$$I_{\text{model}}(\mathbf{h}) = (1 - x)|F_{\text{single}}(\mathbf{h})|^2 + x|F_{\text{single}}(\bar{\mathbf{h}})|^2, \quad (1)$$

where  $|F_{\text{single}}(\mathbf{h})|^2$  and  $|F_{\text{single}}(\bar{\mathbf{h}})|^2$  are model quantities based on a single crystal comprised of the reference domain. The two alternative absolute structures can be refined competitively against one another by refining the twin scale factor,  $x$ , which in this application is referred to as the Flack parameter.

The Flack parameter has a physically meaningful value in the range of 0–1, and represents the fraction of the inverted structure present in the crystal. A value of  $x = 0$  implies that none of the crystal is in the inverted form and the model has the correct absolute structure; if  $x = 1$  then all of the crystal is in the inverted form. Intermediate values of  $x$  point to inversion twinning.

It is important to interpret the value of the Flack parameter in the context of its standard uncertainty. From a statistical point of view, a value of 0.2 (8) has such a large standard uncertainty (0.8) that one neither knows whether the crystal is twinned by inversion or not, nor whether it is inverted or not. Further analysis shows that before any conclusions regarding absolute structure can be made, the standard uncertainty of the Flack parameter should be less than 0.1, even if a material is known to be enantiopure (Flack & Bernardinelli, 2000).

The ability to achieve a low standard uncertainty for the Flack parameter depends in part on the resonant scattering effects having sufficient magnitude to lead to measurably different intensities for Friedel pairs of reflections with indices  $\mathbf{h}$  and  $\bar{\mathbf{h}}$ . This depends on the chemical elements present in the crystal and the wavelength of the X-rays used to collect the diffraction data. The magnitude of resonant scattering effects in a given experiment can be conveniently quantified by the Friedif<sub>stat</sub> parameter (Flack & Shmueli, 2007).

If Friedif<sub>stat</sub> has a value of  $\sim 80$  or more, absolute structure determination presents little problem (Flack & Bernardinelli, 2008a). However, resonant scattering effects for elements such as C, N and O are small for commonly available X-ray energies making it difficult to determine the Flack parameter with sufficient precision to establish absolute structure for many organic compounds. For example, the value of Friedif<sub>stat</sub> for the amino acid L-alanine with Cu  $K\alpha$  radiation is only 34. Accordingly, the value of the Flack parameter obtained from a conventional refinement of L-alanine was  $-0.04$  (27). The data-set was of excellent quality, yielding low merging and refinement residuals, yet the precision of the Flack parameter is too low to enable a definitive statement to be made regarding the absolute structure (Flack & Bernardinelli, 2000).

The ability to determine absolute structure precisely also depends on low levels of random and systematic errors in intensity measurements. Analysis of non-centrosymmetric crystal structures published in *Acta Cryst. C* in 2007, 2011 and 2012 has shown that even the intensity data of structures with large Friedif<sub>stat</sub> values may be entirely dominated by random uncertainties and systematic errors (Flack, 2012; Flack *et al.*, 2011).

There is a long-standing interest in finding ways to improve the precision of the Flack parameter in light-atom structures. A post-refinement Bayesian statistical procedure has been described by Hooft *et al.* (2008, 2010), which can be used

either to define a probability that a refined absolute structure is correct or to obtain an estimate of the Flack parameter.<sup>1</sup> Methods in which refinement weights are modified for data in proportion to their sensitivity to the Flack parameter have also been described (Bernardinelli & Flack, 1985; Parsons, Wagner *et al.*, 2012). It has further been shown that precision may be improved by the use of aspherical scattering factors (Dittrich *et al.*, 2006).

While each of the methods described has been shown to yield lower standard uncertainties on Flack parameters than conventional refinement, all are open to potential criticisms. The Hooft method, being a post-refinement method, does not formally allow  $x$  to correlate with other parameters during refinement, and this, theoretically at least, may compromise values of the standard uncertainty obtained (Hooft *et al.*, 2008). Reweighting methods, involving the selection of data for up-weighting, can magnify errors in the intensity measurements, so that values of  $x$  can be precise but inaccurate. Parsons, Wagner *et al.* (2012), for example, refer to one structure where the elimination of just two poorly measured data points shifted the Flack parameter from 0.35 (12) to 0.02 (14). Use of aspherical scattering factors yields improvements in precision, but in most of the examples tested the change was too small to enable sufficiently precise absolute structure determination for light-atom compounds (see Table 3 in Dittrich *et al.*, 2006). This said, the use of invariom models in combination with other methods, described above and herein, merits further investigation.

In this paper we will describe methods based on intensity differences and quotients that enable  $x$  to be refined along with all other parameters. The purpose of this paper is to demonstrate that this leads to more precise estimates of  $x$  than conventional refinement methods while avoiding the potential criticisms discussed above.

## 2. Definitions of intensity differences and quotients

### 2.1. Differences

Differences between the observed intensities of Friedel pairs of reflections

$$D_{\text{obs}}(\mathbf{h}) = I_{\text{obs}}(\mathbf{h}) - I_{\text{obs}}(\bar{\mathbf{h}}) \quad (2)$$

can be modelled following equation (1) with

$$\begin{aligned} D_{\text{model}}(\mathbf{h}) &= I_{\text{model}}(\mathbf{h}) - I_{\text{model}}(\bar{\mathbf{h}}) \\ &= (1 - 2x)[|F_{\text{single}}(\mathbf{h})|^2 - |F_{\text{single}}(\bar{\mathbf{h}})|^2] \\ &= (1 - 2x)D_{\text{single}}(\mathbf{h}), \end{aligned} \quad (3)$$

where  $D_{\text{single}}(\mathbf{h}) = [|F_{\text{single}}(\mathbf{h})|^2 - |F_{\text{single}}(\bar{\mathbf{h}})|^2]$ . The quantities  $D$ , which are also referred to as Friedel or Bijvoet differences, have been used in strategies for absolute structure determination described by Hooft *et al.* (2008, 2010) and Le Page *et al.* (1990), in a procedure available in the *DIRDIF* suite of

<sup>1</sup> The Flack parameter estimated in this way is referred to as the Hooft parameter and given the symbol  $y$ .

**Table 1**

Results of absolute structure refinements using the methods outlined in the text.

$R_1[|F| > 4\sigma(|F|)]$  is the unweighted  $R$  factor based on  $|F|$  for the intensity data only. The  $x$  (twin) column contains values of the Flack parameter ( $x$ ) calculated in the 'conventional' manner against intensity data merged in the relevant non-centrosymmetric point group (*SHELXL2012*); the output of these refinements was used to calculate  $y$  (Hoof) using Bayesian methods (*PLATON*) and the value of  $R_1[|F| > 4\sigma(|F|)]$ . In the  $x$  (quotient) column the top figure refers to post-refinement calculation of  $x$  as described in §4.2 (*SHELXL2012*) and §4.1; the bottom figure refers to the refinement performed as in §4.3 with the intensity quotients applied as restraints during refinement (*TOPAS*). The top figure in the  $x$  (difference) column was obtained from the numerical restraints method implemented in *CRYSTALS* and the bottom figure to the fully coded difference restraints (§4.4) in *TOPAS*. The superscripts A and B in the 'Structure code' column refer to data collected with Agilent or Bruker instruments, respectively. Samples TWA16a and TWA16b are polymorphs; TWA17c is a recollection of TWA17a with a different sample;  $R$ - and  $S$ -CYCLO, FYO11 and FYO12, and TWA17 and TWA20 are enantiomers.

Structure code	Chemical formula	Friedif <sub>stat</sub>	Space group	Redundancy	$R_1[ F  > 4\sigma( F )]$	$x$ (twin)	$y$ (Hoof)	$x$ (quotient)	$x$ (difference)
$R$ -Mandelic acid <sup>A</sup>	C <sub>8</sub> H <sub>8</sub> O <sub>3</sub>	36	$P2_1$	11	0.0511	0.12 (46)	0.03 (6)	0.00 (11)	−0.04 (5)
$L$ -Alanine <sup>B</sup>	C <sub>3</sub> H <sub>7</sub> NO <sub>2</sub>	34	$P2_12_12_1$	25	0.0219	−0.04 (27)	0.01 (4)	0.01 (4)	0.04 (3)
$L$ -Alanine <sup>A</sup>	C <sub>3</sub> H <sub>7</sub> NO <sub>2</sub>	34	$P2_12_12_1$	15	0.0181	0.06 (26)	0.06 (5)	0.01 (3)	0.01 (3)
Glutamine <sup>B</sup>	C <sub>5</sub> H <sub>8</sub> N <sub>2</sub> O <sub>2</sub>	33	$P2_12_12_1$	28	0.0248	0.09 (25)	0.07 (3)	0.05 (4)	0.04 (3)
GKO02 <sup>B</sup>	C <sub>25</sub> H <sub>31</sub> NO <sub>5</sub>	32	$P2_12_12_1$	15	0.0247	0.01 (15)	0.03 (3)	0.08 (4)	0.07 (4)
A0030a <sup>B</sup>	C <sub>21</sub> H <sub>26</sub> N <sub>2</sub> O <sub>3</sub>	29	$P2_12_12_1$	11	0.0263	−0.10 (21)	−0.07 (5)	0.07 (3)	0.04 (2)
A0034a <sup>B</sup>	C <sub>18</sub> H <sub>25</sub> O <sub>2.5</sub>	29	$P2_12_12_1$	11	0.0274	0.00 (21)	0.02 (3)	0.02 (3)	0.02 (2)
A0034b <sup>B</sup>	C <sub>18</sub> H <sub>25</sub> O <sub>2.5</sub>	29	$P2_12_12_1$	35	0.0268	−0.01 (22)	−0.02 (3)	0.03 (3)	0.02 (3)
LRE01a <sup>B</sup>	C <sub>14</sub> H <sub>19</sub> NO	24	$P2_12_12_1$	22	0.0278	−0.01 (33)	−0.03 (5)	−0.07 (5)	−0.05 (3)
TWA18a <sup>B</sup>	C <sub>16</sub> H <sub>20</sub> N <sub>2</sub> O	23	$P2_12_12_1$	17	0.0253	0.04 (26)	0.04 (3)	−0.02 (3)	−0.07 (5)
$R$ -CYCLO <sup>A</sup>	C <sub>19</sub> H <sub>26</sub> N <sub>6</sub> O	21	$P2_12_12_1$	14	0.0425	−0.02 (27)	−0.02 (4)	0.02 (3)	−0.07 (5)
$S$ -CYCLO <sup>A</sup>	C <sub>19</sub> H <sub>26</sub> N <sub>6</sub> O	21	$P2_12_12_1$	16	0.0409	−0.03 (20)	−0.04 (3)	0.02 (3)	−0.07 (5)
TWA21a <sup>B</sup>	C <sub>21</sub> H <sub>29</sub> N <sub>3</sub>	14	$P2_12_12_1$	17	0.0248	0.00 (40)	0.00 (4)	0.02 (2)	0.06 (2)
TWA20c <sup>B</sup>	C <sub>19</sub> H <sub>19</sub> N <sub>2</sub>	12	$P3_2$	26	0.0231	−0.02 (46)	−0.01 (5)	0.02 (2)	0.02 (2)
TWA16a <sup>B</sup>	C <sub>16</sub> H <sub>18</sub> N <sub>2</sub>	13	$P3_2$	13	0.0283	0.00 (69)	0.02 (7)	−0.02 (3)	−0.01 (3)
TWA16b <sup>B</sup>	C <sub>16</sub> H <sub>18</sub> N <sub>2</sub>	13	$P2_1$	8	0.0286	0.02 (37)	0.05 (6)	−0.02 (2)	−0.02 (2)
TWA17a <sup>B</sup>	C <sub>19</sub> H <sub>18</sub> N <sub>2</sub>	12	$P3_1$	9	0.0300	0.00 (60)	0.06 (8)	−0.02 (2)	−0.02 (2)
TWA17c <sup>B</sup>	C <sub>19</sub> H <sub>18</sub> N <sub>2</sub>	12	$P3_1$	15	0.0319	0.00 (63)	0.04 (5)	−0.04 (6)	−0.04 (3)
TWA22a <sup>B</sup>	C <sub>21</sub> H <sub>22</sub> N <sub>2</sub>	12	$P2_12_12_1$	11	0.0262	0.01 (41)	0.04 (7)	−0.02 (5)	−0.04 (3)
FYO12d <sup>B</sup>	C <sub>21</sub> H <sub>22</sub> N <sub>2</sub>	12	$P2_12_12_1$	35	0.0246	0.07 (53)	0.04 (9)	0.04 (3)	0.01 (3)
FYO12e <sup>B</sup>	C <sub>21</sub> H <sub>22</sub> N <sub>2</sub>	12	$P2_12_12_1$	35	0.0252	0.17 (54)	0.01 (8)	−0.04 (2)	−0.02 (4)
FYO11d <sup>B</sup>	C <sub>21</sub> H <sub>22</sub> N <sub>2</sub>	12	$P2_12_12_1$	18	0.0257	0.08 (53)	0.03 (6)	0.00 (3)	0.00 (3)
Cholestane <sup>B</sup>	C <sub>27</sub> H <sub>48</sub>	9	$P2_1$	18	0.0288	−0.01 (77)	−0.04 (9)	0.01 (4)	0.01 (4)
Reduced $\chi^2$						0.03	0.83	0.04 (8)	0.04 (8)
								0.05 (7)	−0.12 (11)
								0.10 (7)	0.02 (7)
								0.06 (6)	0.00 (6)
								0.05 (6)	0.05 (6)
								0.04 (8)	0.04 (8)
								0.04 (9)	0.10 (6)
								0.01 (8)	0.02 (8)
								0.03 (5)	0.04 (5)
								0.03 (5)	0.02 (5)
								0.00 (11)	−0.02 (8)
								1.22	−0.03 (11)
								1.47	0.86

programs (Beurskens *et al.*, 1996) and in the procedure described by Thompson & Watkin (2011).

## 2.2. Quotients

In principle, on a four-circle diffractometer equipped with a point detector it is possible to measure the intensities of

reflections  $\mathbf{h}$  and  $\bar{\mathbf{h}}$  at setting angles ( $2\theta$ ,  $\omega$ ,  $\chi$  and  $\varphi$ ) and ( $-2\theta$ ,  $-\omega$ ,  $\chi$  and  $\varphi$ ) (Le Page *et al.*, 1990). The first set of setting angles is appropriate for reflection  $hkl$  and the second set for  $\bar{h}\bar{k}\bar{l}$ . In the second set, both the incident and reflected beam directions are reversals of those of the first set. If a crystal has a centrosymmetric habit then the beam paths through the crystal of these two measurements are identical. Conse-

quently, their absorption and extinction corrections are identical, and the quotient  $I_{\text{obs}}(\mathbf{h})/I_{\text{obs}}(\bar{\mathbf{h}})$  is absorption and extinction (and scale) free.

The corresponding model quotient can be written in terms of  $|F_{\text{single}}(\mathbf{h})|^2$  and  $|F_{\text{single}}(\bar{\mathbf{h}})|^2$

$$\frac{I_{\text{model}}(\mathbf{h})}{I_{\text{model}}(\bar{\mathbf{h}})} = \frac{(1-x)|F_{\text{single}}(\mathbf{h})|^2 + x|F_{\text{single}}(\bar{\mathbf{h}})|^2}{(1-x)|F_{\text{single}}(\bar{\mathbf{h}})|^2 + x|F_{\text{single}}(\mathbf{h})|^2} \quad (4)$$

While this type of formulation has been used in the structure refinement of the kinase inhibitor roscovitine (Wang *et al.*, 2001), it is non-linear in  $x$ , and the standard uncertainties of quotients defined for  $\mathbf{h}$  and  $\bar{\mathbf{h}}$  are not the same. These problems are removed, and the independence from absorption, extinction and scale maintained, by reformulating the quotients in terms of sums and differences of Friedel-pair intensities, so that the observed quotients

$$Q_{\text{obs}}(\mathbf{h}) = \frac{I_{\text{obs}}(\mathbf{h}) - I_{\text{obs}}(\bar{\mathbf{h}})}{I_{\text{obs}}(\mathbf{h}) + I_{\text{obs}}(\bar{\mathbf{h}})} = \frac{D_{\text{obs}}(\mathbf{h})}{2A_{\text{obs}}(\mathbf{h})} \quad (5)$$

are modelled with

$$\begin{aligned} Q_{\text{model}}(\mathbf{h}) &= \frac{I_{\text{model}}(\mathbf{h}) - I_{\text{model}}(\bar{\mathbf{h}})}{I_{\text{model}}(\mathbf{h}) + I_{\text{model}}(\bar{\mathbf{h}})} = \frac{D_{\text{model}}(\mathbf{h})}{2A_{\text{model}}(\mathbf{h})} \\ &= (1-2x)Q_{\text{single}}(\mathbf{h}) \end{aligned} \quad (6)$$

where

$$Q_{\text{single}}(\mathbf{h}) = \frac{|F_{\text{single}}(\mathbf{h})|^2 - |F_{\text{single}}(\bar{\mathbf{h}})|^2}{|F_{\text{single}}(\mathbf{h})|^2 + |F_{\text{single}}(\bar{\mathbf{h}})|^2},$$

$$2A_{\text{obs}}(\mathbf{h}) = I_{\text{obs}}(\mathbf{h}) + I_{\text{obs}}(\bar{\mathbf{h}}) \quad \text{and}$$

$$2A_{\text{model}}(\mathbf{h}) = I_{\text{model}}(\mathbf{h}) + I_{\text{model}}(\bar{\mathbf{h}}).$$

All of the test data-sets used in this study were collected using modern diffractometers equipped with area detectors. These do not in general perform reversed beam-path measurements, and so our initial contention that quotients can be measured in such a way that errors cancel does not hold.<sup>2</sup> For such data we follow Parsons, Pattison & Flack (2012) in writing

$$I_{\text{obs}}(\mathbf{h}) = [s(\mathbf{h}) + \Delta s(\mathbf{h})]I_{\text{model}}(\mathbf{h}) \quad (7)$$

$$I_{\text{obs}}(\bar{\mathbf{h}}) = [s(\mathbf{h}) - \Delta s(\mathbf{h})]I_{\text{model}}(\bar{\mathbf{h}}), \quad (8)$$

where  $s(\mathbf{h})$  and  $\Delta s(\mathbf{h})$  are the average and half-difference of the systematic errors in reflections  $I_{\text{obs}}(\mathbf{h})$  and  $I_{\text{obs}}(\bar{\mathbf{h}})$  remaining after the application of a multi-scan correction. If the term  $Q_{\text{model}}(\mathbf{h})[\Delta s(\mathbf{h})/s(\mathbf{h})]$  is small enough that  $(1 - Q_{\text{model}}(\mathbf{h})[\Delta s(\mathbf{h})/s(\mathbf{h})])$  is a good approximation for  $(1 + Q_{\text{model}}(\mathbf{h})[\Delta s(\mathbf{h})/s(\mathbf{h})])^{-1}$ , this leads to

<sup>2</sup> A referee to this paper pointed out that reverse-beam measurements can be made for low-angle reflections on a four-circle diffractometer with an area detector by measuring 360°  $\varphi$ -scans at  $\chi = +90^\circ$  and  $-90^\circ$ . In this case Friedel pairs fall on the same pixels of the detector eliminating another source of possible systematic error. The use of data collected in this way for measuring  $Q_{\text{obs}}(\mathbf{h})$  is a very interesting avenue for further work.

$$\begin{aligned} Q_{\text{obs}}(\mathbf{h}) &\simeq Q_{\text{model}}(\mathbf{h}) + [1 - Q_{\text{model}}^2(\mathbf{h})] \frac{\Delta s(\mathbf{h})}{s(\mathbf{h})} \\ &\quad - Q_{\text{model}}(\mathbf{h}) \left( \frac{\Delta s(\mathbf{h})}{s(\mathbf{h})} \right)^2. \end{aligned} \quad (9)$$

In practice  $|Q_{\text{model}}(\mathbf{h})|$  is usually considerably less than 1, and so provided the difference in systematic errors in  $I_{\text{obs}}(\mathbf{h})$  and  $I_{\text{obs}}(\bar{\mathbf{h}})$  is small relative to the overall systematic error, the assumption that  $Q_{\text{model}}(\mathbf{h}) = Q_{\text{obs}}(\mathbf{h})$  should still hold approximately.

Equation (6) can also be interpreted as being equivalent to equation (3) with an additional weighting term,  $1/2A_{\text{model}}(\mathbf{h})$ . The incorporation of this term can be justified on the basis of the leverage analysis presented in Parsons, Wagner *et al.* (2012), in which it was shown that the data with the greatest influence on the precision of the Flack parameter were those with weak to moderate intensities. The factor  $1/2A_{\text{model}}(\mathbf{h})$  up-weights these data. Very weak data, which have little leverage on the Flack parameter but high values of  $Q_{\text{obs}}(\mathbf{h})$  on account of a small value of  $A_{\text{obs}}(\mathbf{h})$ , should be omitted from an analysis based on quotients.

### 3. Experimental

#### 3.1. Data-sets

A series of test data-sets was used in this study; selected crystal data are given in Table 1. The compounds selected for study contain no element heavier than oxygen, and all have Friedel<sub>stat</sub> of 36 or less for Cu  $K\alpha$  radiation. *R*-Mandelic acid and L-alanine were obtained from Sigma–Aldrich and were used as received; for other samples solution-phase optical rotation measurements or chiral separation established enantiomeric excesses of > 98%. Resonant scattering factors  $f'$  and  $f''$  are independent of resolution, and the contribution of resonant relative to non-resonant scattering is therefore greatest at high values of  $\sin \theta/\lambda$ . For this reason the test data-sets were collected at low temperature. Data-sets were highly redundant, with average multiplicities of observations between 8 and 35.

Data-sets for the samples listed in Table 1 carrying the superscript A were collected with Cu  $K\alpha$  radiation on an Agilent Technologies SuperNova incorporating a microsource generator. The temperature of data collection was 150 K except for *R*-mandelic acid. This material crystallizes as plates, and cooling to 150 K was found to cause strain-broadening, and so data were collected at 220 K. Processing, including integration and a multi-scan absorption correction (Blessing, 1995), was accomplished with *CrysAlis PRO* (Oxford Diffraction Ltd, 2010).

Data-sets carrying the superscript B were collected using Cu  $K\alpha$  radiation at 100 K using a Bruker Microstar fine-focus rotating anode generator with a SMART 6000 CCD detector or a Bruker D8 microsource, also equipped with a SMART 6000 detector. Data were processed with *SAINT* (Bruker–Nonius, 2006) and corrected for absorption and other

systematic errors using the multi-scan procedure *SADABS* (Bruker-Nonius, 2006; Sheldrick, 2008a).

Data were merged using *SORTAV* with unit weights and robust-resistant down-weighting of outliers (Blessing, 1997).<sup>3</sup>

### 3.2. Absolute structure refinement in *SHELXL2012*

Structures were solved using direct methods (*SHELXS*; Sheldrick, 2008b) or charge flipping (*SUPERFLIP*; Palatinus & Chapuis, 2007) and refined against  $|F|^2$  in *SHELXL2012* (beta test version 2012/9) using all data (Sheldrick, 2012). Data were merged in point groups 2, 222 or 3 for the monoclinic, orthorhombic and trigonal structures, respectively. Isopropyl groups in structures *R*- and *S*-CYCLO are disordered over two orientations (Wang *et al.*, 2001). The disorder components were restrained to have similar bond distances and angles. The water of crystallization in structure A0034a is disordered about a crystallographic twofold axis, which also induces disorder in two hydroxyl H-atom positions. All full-weight non-H atoms were refined with anisotropic displacement parameters. Full weight H atoms were refined freely; those part of disordered groups were either refined with restraints (A0034a) or placed in ideal positions (*R*- and *S*-CYCLO). The Flack parameter was refined either by full-matrix least squares (*i.e.* using the *TWIN/BASF* commands in the *SHELXL* .ins file) or post-refinement based on quotients defined in §2.2. The results of the first of these refinements were used to calculate the Hooft parameter *via* the *BIJVOET* routine in *PLATON*. In all cases a Gaussian prior was used (Hooft *et al.*, 2008; Spek, 2003).

### 3.3. Absolute structure refinement in *CRYSTALS*

The same models as described above were refined in *CRYSTALS* (Version 14.40; Betteridge *et al.*, 2003) and the Flack parameter estimated *via* the procedure described by Thompson & Watkin (2011).

### 3.4. Absolute structure refinement in *TOPAS-Academic*

*TOPAS-Academic*, Version 5 (Coelho, 2012), allows user-equations to be written in the form of a function, in a similar way to the definition of a function or subroutine in a programming language such as Fortran or C++ (Coelho *et al.*, 2011). These equations can then be used in restraints or to define other parameters. This option enables equation (6) to be incorporated into the refinement in the form of a set of restraints where the equation is coded in terms of atomic positional, displacement and occupancy parameters for each quotient.

The quotients  $Q_{\text{obs}}(\mathbf{h})$  were calculated from the integrated data-sets using reflections for which both  $I_{\text{obs}}(\mathbf{h})$  and  $I_{\text{obs}}(\bar{\mathbf{h}})$  were greater than three times their respective uncertainties.

This cut-off condition eliminates quotients which are large because the term  $A_{\text{obs}}(\mathbf{h})$  is small (see §2.2).  $Q_{\text{obs}}(\mathbf{h})$  may also be calculated using *XPREP* (Sheldrick, 2001). Outlier data for which  $|D_{\text{obs}}(\mathbf{h})|$  was greater than twice the maximum value of  $|D_{\text{single}}|$  were also excluded. This condition mirrors the procedure used for outlier detection in the Bijvoet routine in *PLATON* (Spek, 2003). For surviving data the standard uncertainty  $u[Q_{\text{obs}}(\mathbf{h})]$  was propagated from the values of  $u[I_{\text{obs}}(\mathbf{h})]$  and  $u[I_{\text{obs}}(\bar{\mathbf{h}})]$  obtained from merging

$$u[Q_{\text{obs}}(\mathbf{h})] = \frac{2}{[I_{\text{obs}}(\mathbf{h}) + I_{\text{obs}}(\bar{\mathbf{h}})]^2} \{I_{\text{obs}}^2(\mathbf{h})u^2[I_{\text{obs}}(\bar{\mathbf{h}})] + I_{\text{obs}}^2(\bar{\mathbf{h}})u^2[I_{\text{obs}}(\mathbf{h})]\}^{1/2}.$$

The observations used in *TOPAS* for refinement took the form of Friedel-averaged intensity data,  $A_{\text{obs}}(\mathbf{h})$ , and quotients,  $Q_{\text{obs}}(\mathbf{h})$ , the latter in the form of restraints.

The values of  $A_{\text{obs}}(\mathbf{h})$  and their uncertainties were obtained by merging the centric and paired acentric data in the relevant Laue group (*e.g.* *mmm* for an orthorhombic structure). Any unpaired acentric data were omitted. The structure was first refined in *CRYSTALS* against  $A_{\text{obs}}(\mathbf{h})$  only, and the weighting scheme optimized. This weighting scheme was then held fixed in subsequent refinement in *TOPAS*. The  $A_{\text{obs}}(\mathbf{h})$  values were modelled both in *CRYSTALS* and *TOPAS* using a Flack parameter equal to 0.5 in order to correctly account for the averaging of Friedel-pair intensities.

The quotient data,  $Q_{\text{obs}}(\mathbf{h})$ , were modelled with equation (6) using a second Flack parameter; it is this parameter which characterizes the absolute structure. The quotient restraints were initially given a weight,  $w_{\text{restraint}}(\mathbf{h}) = 1/u^2[Q_{\text{obs}}(\mathbf{h})]$ . After initial cycles of refinement the values of the deviates  $w_{\text{restraint}}^{1/2}(\mathbf{h})[Q_{\text{obs}}(\mathbf{h}) - Q_{\text{model}}(\mathbf{h})]$  were used to calculate a reduced  $\chi^2$  statistic. The structure was then re-refined with the quotient restraint weights scaled by  $1/\chi^2$ . A normal probability plot (Abrahams & Keve, 1971) of deviates was also inspected to detect further outliers and for validation purposes.

The same procedure was applied to the refinement using difference restraints based on equation (3). No cut-off condition based on values of  $I_{\text{obs}}(\mathbf{h})/u(I_{\text{obs}}(\mathbf{h}))$  was applied, but outliers were detected as described above. The initial restraint weights were  $w_{\text{restraint}}(\mathbf{h}) = 1/u^2[D_{\text{obs}}(\mathbf{h})] = 1/(u^2[I_{\text{obs}}(\mathbf{h})] + u^2[I_{\text{obs}}(\bar{\mathbf{h}})])$ , but these were rescaled after initial cycles of refinement, also as described above.

Clearly it is only possible to calculate  $Q_{\text{obs}}(\mathbf{h})$  and  $D_{\text{obs}}(\mathbf{h})$  for acentric data where both  $I_{\text{obs}}(\mathbf{h})$  and  $I_{\text{obs}}(\bar{\mathbf{h}})$  have been measured; coverage statistics along with numbers of data omitted as outliers are given in Table 2. Input files for quotient refinement of *L*-alanine<sup>B</sup> in *TOPAS-Academic* are available in the supplementary material.

Programs written to calculate  $Q_{\text{obs}}(\mathbf{h})$  and  $D_{\text{obs}}(\mathbf{h})$ , detect outliers and write files of symbolic restraints made use of subroutines available in the CrysFML Fortran library (Rodríguez-Carvajal & González Platas, 2009).

<sup>3</sup> Supplementary data for this paper are available from the IUCr electronic archives (Reference: GP5062). Services for accessing these data are described at the back of the journal.

**Table 2**

Data-set statistics.

The columns contain the compound identifier, the total number of data measured, the number of centric data, the number of paired acentric data used for generating difference and quotient restraints, the number of unpaired acentric data, the number of outlying pairs for which  $|D_{\text{obs}}(\mathbf{h})| > 2|D_{\text{single, max}}|$ , and the number of difference and quotient restraints used. The figures in brackets in the last two columns are the number of outliers omitted after normal probability plot analysis.

Code	Measured	Centric	Paired acentric	Unpaired acentric	Outlying pairs	Differences	Quotients
R-Mandelic acid <sup>A</sup>	2858	298	2546	8	3	1273	1140
L-Alanine <sup>B</sup>	776	189	570	5	6	285	284
L-Alanine <sup>A</sup>	740	184	540	0	8	270	270
Glutamine <sup>B</sup>	1177	265	898	2	6	449	449
GKO <sup>B</sup>	4073	588	3450	1	17	1725	1685
A0030a <sup>B</sup>	3333	518	2770	3	21	1385	1331
A0034a <sup>B</sup>	2757	470	2256	3	14	1128	1127
A0034b <sup>B</sup>	2673	463	2202	0	4	1101	1099
LRE01a <sup>B</sup>	2261	418	1800	3	20	900	894
TWA18a <sup>B</sup>	2488	407	2046	1	17	1023	1021
R-CYCLO <sup>A</sup>	7983	921	7008	12	21	3504	3454
S-CYCLO <sup>A</sup>	8240	939	7242	9	25	3620 [1]	3606 [1]
TWA21a <sup>B</sup>	3071	471	2520	6	37	1260	1258
TWA20c <sup>B</sup>	2545	0	2476	1	34	1238	1238
TWA16a <sup>B</sup>	2167	0	2082	73	6	1041	1036 [5]
TWA16b <sup>B</sup>	4711	220	4034	29	214	2017	1984
TWA17a <sup>B</sup>	2657	0	2580	3	37	1290	1286
TWA17c <sup>B</sup>	2657	0	2572	1	42	1286	1283 [3]
TWA22a <sup>B</sup>	2920	508	2344	4	32	1172	1168
FYO12d <sup>B</sup>	2901	463	2372	0	33	1186	1158
FYO12e <sup>B</sup>	2894	458	2380	4	26	1190	1167
FYO11d <sup>B</sup>	2887	453	2366	2	33	1183	1173
Cholestane <sup>B</sup>	8266	478	7120	2	333	3560	3546

## 4. Results and discussion

### 4.1. Conventional refinement of the Flack parameter

The results of ‘conventional’ refinement of  $x$ , that is as a twin factor in full-matrix least squares, are listed in Table 1 in the column  $x(\text{twin})$ . They are completely consistent with the results of Flack & Bernardinelli (2008*a*), which indicates that compounds with  $\text{Friedif}_{\text{stat}}$  in the range 9–36 would be expected to yield standard uncertainties of the Flack parameter of between 0.8 and 0.2. None of the conventional refinements yields a Flack parameter with sufficient precision to enable a conclusion to be made about the absolute structures of the crystals being studied.

It is, however, remarkable that the values of the Flack parameter in Table 1 cluster around zero much more tightly than would be anticipated on the basis of their high standard uncertainties. The reduced  $\chi^2$  calculated from the data in the  $x(\text{twin})$  column (assuming the true value of  $x$  is zero in each case) is only 0.031, suggesting that the uncertainties are overestimated by a factor of 5.5, meaning that more information about absolute structure is present in the data than implied by the uncertainties calculated by least squares. While this finding runs counter to the general *underestimation* of standard uncertainties in crystallographic least squares (Hamilton & Abrahams, 1970), it is consistent with Thompson & Watkin’s (2011) histogram of Flack parameters obtained from refinements of 150 structures with  $\text{Friedif}_{\text{stat}}$  in the range 3.4–10.8. The histogram was centred at zero; had the data been

devoid of absolute structure information, the histogram should have been centred about 0.5.

It appears that conventional refinement of  $x$  yields rather pessimistic uncertainty estimates. Although the data quoted in Table 1 were derived using *SHELXL2012*, similar results were obtained with *CRYSTALS* and *TOPAS*. The methods published elsewhere by Hoofstede *et al.* (2008, 2010), Thompson & Watkin (2011) and Parsons, Wagner *et al.* (2012) and those described below attempt to obtain more realistic uncertainty estimates.

### 4.2. Post-refinement estimation of the Flack parameter using quotients

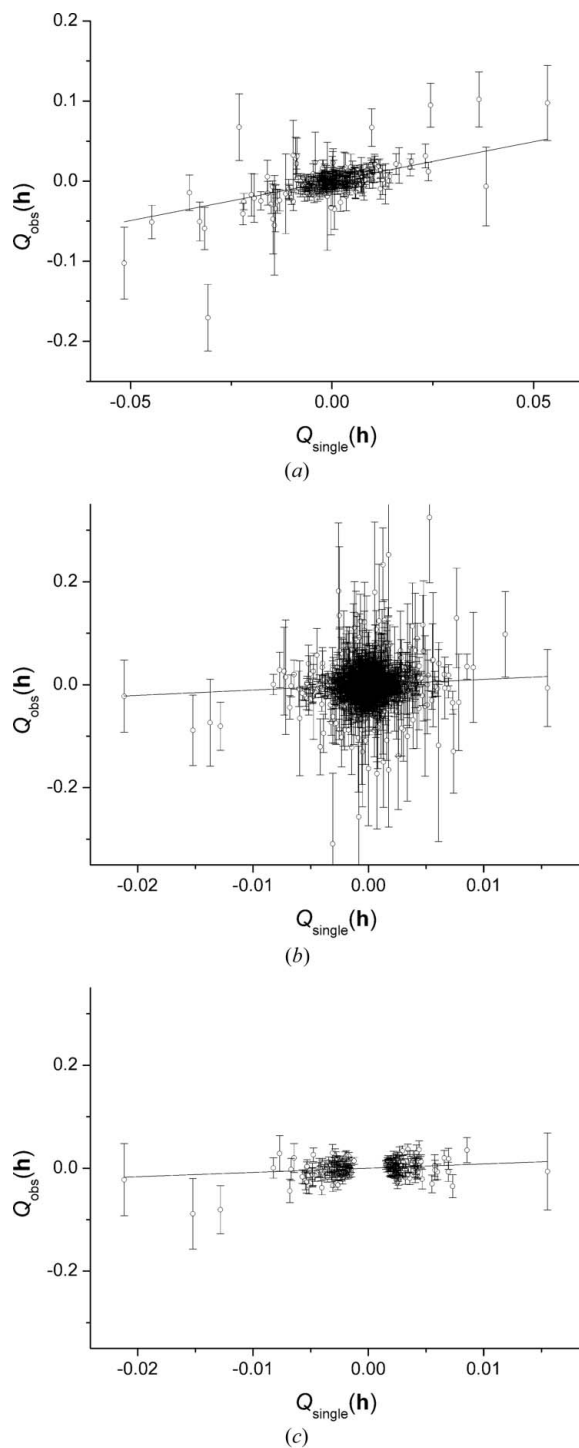
Values of  $Q_{\text{obs}}(\mathbf{h})$  [equation (5)] can be calculated using the observed intensity data, while values of  $Q_{\text{single}}(\mathbf{h})$  [equation (6)] can be calculated from the refinement model. A ‘ $Q$  plot’ of  $Q_{\text{obs}}(\mathbf{h})$  against  $Q_{\text{single}}(\mathbf{h})$  should be a straight line passing through the origin with a gradient of  $(1 - 2x)$ . Similarly, a ‘ $D$  plot’ of  $D_{\text{obs}}(\mathbf{h})$  [equation (2)]

against  $D_{\text{single}}(\mathbf{h})$  [equation (3)] should also be linear with a gradient of  $(1 - 2x)$ .

The method can be illustrated with reference to the data collected for L-alanine. The gradient of the weighted least-squares best straight line in the  $Q$  plot shown in Fig. 1(*a*) is 0.984 (68), which yields a value of the Flack parameter of 0.01 (3). This estimate is very much more precise than that obtained with the same data in a conventional refinement. Some of the error bars in Fig. 1 are huge, but these points have little influence on the (weighted) fit. Much more important is the very clear unit slope in the bulk of the data points.

As the effects of resonant scattering become smaller, the trend in  $Q$ -plots becomes harder to discern by eye, as illustrated in the plot for cholestane ( $\text{Friedif}_{\text{stat}} = 9$ ) in Fig. 1(*b*). The least-squares fit for the data shown nevertheless yields a Flack parameter of  $-0.02$  (12). Part of the problem in Fig. 1(*b*) is that the majority of points have little influence on the fit. Fig. 1(*c*) shows the 200 most influential points in Fig. 1(*b*) (as measured by their leverages; Parsons, Wagner *et al.*, 2012; Prince, 2004; Merli *et al.*, 2000). The noticeable gap in the middle of the plot occurs because for a one-parameter linear fit the leverages are proportional to  $Q_{\text{single}}^2(\mathbf{h})$ ; other missing points in Fig. 1(*c*) have a low leverage on account of their high error bars. Although not exactly obvious, the trend in Fig. 1(*c*) is a little clearer than it is in Fig. 1(*b*). More significantly the value of  $x$  obtained for these points is 0.09 (15): even though we are only working with  $\sim 5\%$  of the data, the precision is only marginally affected.

This linear  $Q$ -fitting method has been incorporated into *SHELXL2012*, and the results for the other compounds studied are the top values listed in the column labelled ' $x$ (quotient)' in Table 1. Also included in Table 1 [in column  $y$ (Hooft)] are the estimates of Hooft's Bayesian method. The results of the two procedures are in excellent agreement, both



**Figure 1**  
Plot of  $Q_{\text{obs}}(\mathbf{h})$  against  $Q_{\text{single}}(\mathbf{h})$  [as defined in equations (5) and (6)] for (a) L-alanine and (b) cholestane. (c) as (b), but only the top 200 most influential points are shown; the axes are chosen to be the same as (b).

showing consistently higher precision than conventional methods [ $x$ (twin)].

#### 4.3. Estimation of the Flack parameter using quotient data

Although the method described in §4.2 yields precise and accurate values of the Flack parameter, like other post-refinement methods, it has the disadvantage that  $x$  is not allowed to correlate with other parameters. A way around this difficulty is to incorporate equation (6) into the structure refinement (Murphy *et al.*, 2010; Parsons & Flack, 2004; Wang *et al.*, 2001). Values of  $Q_{\text{obs}}(\mathbf{h})$  defined in equation (5) are calculated from the intensity data, as before, but facilities available in *TOPAS* enable  $Q_{\text{single}}(\mathbf{h})$  to be written in terms of the parameters of the refinement model (atomic coordinates, displacement parameters and occupancies) and built into the refinement as an equation of restraint; an example is provided in the supplementary material. The Flack parameter can now correlate with other parameters during refinement.

A quotient can be defined for each Friedel pair of intensities in the data-set so that the observations in the refinement now take the form of Friedel-averaged intensity measurements,  $A_{\text{obs}}(\mathbf{h})$ , and the quotients,  $Q_{\text{obs}}(\mathbf{h})$ . The values of  $A_{\text{obs}}(\mathbf{h})$  are obtained by merging the centric and paired acentric data in the centrosymmetric Laue group, and all absolute structure information is contained in the quotients. The number of quotients used in the test examples lay between a few hundred to several thousand depending on the size of the structure (details are in Table 2).

The bottom values listed in the  $x$ (quotient) column of Table 1 show the results. They are very similar to those obtained with the post-refinement procedure described in §4.2, and much more precise than those obtained with the conventional refinement technique. Acceptable precision has been obtained for data-sets with  $\text{Friedif}_{\text{stat}}$  as low as 12.

The Flack parameter is sensitive to outliers in the data, and it is important that these are detected and eliminated. The sensitivity to outliers can be illustrated using the data-set TWA16a. A refinement performed with no outlier elimination at all yielded a Flack parameter equal to 0.18 (8). A normal probability plot calculated for this refinement (Fig. 2) was linear, but with one point (at the bottom left in Fig. 2) deviating substantially from the straight line. Elimination of this one observation changed the value of  $x$  to 0.08 (8).

Normal probability plots are a powerful means for detecting outliers, although in this work we follow Spek's procedure in *PLATON* in eliminating Friedel pairs with  $|D_{\text{obs}}(\mathbf{h})|$  more than twice the maximum calculated absolute difference for the entire data-set. This is a more objective procedure, although it means that more data are omitted the lower the value of  $\text{Friedif}_{\text{stat}}$ . The numbers of outliers omitted are given in Table 2.

#### 4.4. Estimation of the Flack parameter using difference data

An alternative procedure is to base the restraints on differences rather than quotients. The target value for each restraint is  $D_{\text{obs}}(\mathbf{h})$  as defined in equation (2). The model value



**Table 3**

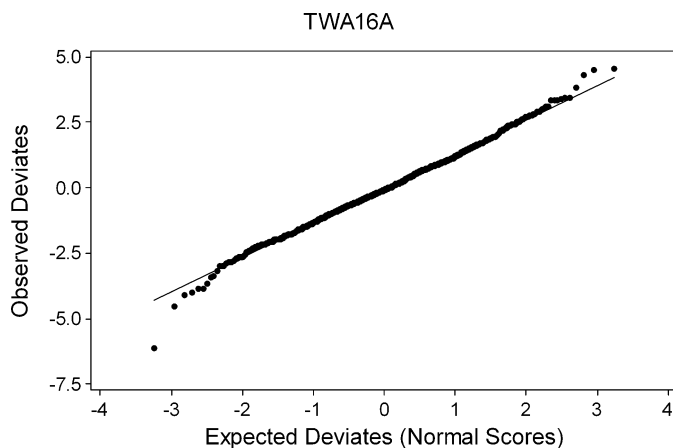
Fitting statistics for the refinements based on quotients and differences.

In the ‘Quotient restraints’ columns, the intercept, gradient and Pearson  $r^2$  refer to normal probability plots calculated using weighted residuals for the restraints,  $w_{\text{restraint}}^{1/2}(\mathbf{h})[Q_{\text{obs}}(\mathbf{h}) - Q_{\text{model}}(\mathbf{h})]$ . Equivalent formulae were used for the ‘Difference restraints’ columns.  $R(Q)$  and  $R(D)$  are defined in equation (10).

Code	Quotient restraints			$R(Q)$	Difference restraints			$R(D)$
	Intercept	Gradient	$r^2$		Intercept	Gradient	$r^2$	
R-Mandelic acid <sup>A</sup>	0.0461	0.993	0.986	1.003	0.0501	0.992	0.985	0.982
L-Alanine <sup>B</sup>	-0.0127	0.995	0.982	0.776	-0.0118	0.994	0.980	0.825
L-Alanine <sup>A</sup>	-0.0503	1.002	0.997	0.892	-0.0471	1.002	0.997	0.907
Glutamine <sup>B</sup>	-0.0158	1.002	0.998	0.913	-0.0219	1.002	0.998	0.904
GKO <sup>B</sup>	0.0019	1.000	0.998	0.925	0.0072	1.000	0.998	0.951
A0030a <sup>B</sup>	-0.0517	0.998	0.997	0.965	-0.0525	0.998	0.997	0.964
A0034a <sup>B</sup>	0.0036	1.000	0.997	0.873	0.0070	1.000	0.997	0.923
A0034b <sup>B</sup>	0.0547	0.999	0.998	0.832	0.0555	0.999	0.998	0.878
LRE01a <sup>B</sup>	0.1018	0.995	0.996	0.959	0.1057	0.994	0.996	0.931
TWA18a <sup>B</sup>	0.0175	0.999	0.996	0.886	0.0166	0.999	0.996	0.921
R-CYCLO <sup>A</sup>	-0.0300	0.999	0.997	0.979	-0.0269	0.999	0.998	0.981
S-CYCLO <sup>A</sup>	-0.0017	1.000	0.999	0.944	-0.0019	1.000	0.999	0.967
TWA21a <sup>B</sup>	0.0015	0.999	0.996	0.916	0.0002	1.000	0.996	0.962
TWA20c <sup>B</sup>	0.0138	1.000	0.998	0.961	0.0153	1.000	0.998	0.990
TWA16a <sup>B</sup>	0.0016	1.000	0.996	1.001	0.0030	1.000	0.993	0.998
TWA16b <sup>B</sup>	-0.0106	0.999	0.997	0.991	-0.0027	0.998	0.994	0.987
TWA17a <sup>B</sup>	-0.0078	0.999	0.996	0.981	-0.0079	0.999	0.996	0.990
TWA17c <sup>B</sup>	0.0159	1.000	0.998	0.990	0.0132	1.000	0.997	0.984
TWA22a <sup>B</sup>	-0.0456	0.997	0.994	0.986	-0.0440	0.997	0.994	0.984
FYO12d <sup>B</sup>	-0.0697	0.997	0.995	0.998	-0.0704	0.996	0.995	0.997
FYO12c <sup>B</sup>	-0.0042	0.998	0.993	0.994	0.0011	0.997	0.993	0.998
FYO11d <sup>B</sup>	-0.0196	0.998	0.994	0.964	-0.0241	0.998	0.994	0.988
Cholestane <sup>B</sup>	-0.0046	1.000	0.998	0.996	-0.0048	1.000	0.998	0.999

$D_{\text{model}}(\mathbf{h})$  is defined in equation (3) and written in terms of the refinement parameters. The procedure was otherwise identical to that described in §4.3.

A similar procedure is available in *CRYSTALS*. Here refinement is carried out against  $|F_{\text{obs}}(\mathbf{h})|^2$  data merged in the appropriate crystal (as opposed to Laue) point group and  $D_{\text{model}}(\mathbf{h})$  is a numerical value calculated from the current model (see Fig. 2 in Thompson & Watkin, 2011).



**Figure 2**

Outlier detection. Normal probability plot calculated using observed quotient restraint deviates  $w_{\text{restraint}}^{1/2}(\mathbf{h})[Q_{\text{obs}}(\mathbf{h}) - Q_{\text{model}}(\mathbf{h})]$  for TWA16a. The extreme point at the bottom left was omitted as an outlier. This plot was calculated prior to rescaling of weights.

The results of the two procedures are listed in the  $x(\text{difference})$  column of Table 1; the top value is obtained from the *CRYSTALS* procedure, the bottom *via* explicit restraints. The results are in agreement with each other and those obtained for quotients.

**4.5. The extent of error cancellation on taking quotients**

The differences defined in equations (2) and (3) are used in the Bayesian and numerical restraint methods for estimation of the Hooft or Flack parameters available in *PLATON* and *CRYSTALS*. Our original idea of basing absolute structure analysis on quotients was conceived about a decade ago, when four-circle instruments with point detectors were still in common use. The cancellation of absorption and extinction errors which occurs on taking intensity quotients using reverse-beam measurements of Friedel pair intensities cannot in

general be achieved with modern area-detector instruments (but see footnote 2, §2.2). The analysis presented in §2 suggests that some approximate cancellation of errors occurs provided the difference in systematic errors in  $I_{\text{obs}}(\mathbf{h})$  and  $I_{\text{obs}}(\mathbf{h})$  is small relative to the overall systematic error.

One means for assessing whether error cancellation is achieved in practice with area-detector data is to compare  $R$  factors based on the observed and model values of quotients or differences [equation (10)]. If errors are really cancelled  $R(Q)$  should be systematically lower than  $R(D)$ .

$$R(Q) = \frac{\sum |Q_{\text{obs}}(\mathbf{h}) - Q_{\text{model}}(\mathbf{h})|}{\sum |Q_{\text{obs}}(\mathbf{h})|}$$

$$R(D) = \frac{\sum |D_{\text{obs}}(\mathbf{h}) - D_{\text{model}}(\mathbf{h})|}{\sum |D_{\text{obs}}(\mathbf{h})|} \tag{10}$$

Values of  $R(Q)$  and  $R(D)$  are listed for each data-set in Table 3. The differences are mostly quite marginal, a finding consistent with a similar listing in Table 10 of Parsons, Pattison & Flack (2012).

A second method for assessing the presence of systematic errors is to examine normal probability plots based on the weighted residuals  $w_{\text{restraint}}^{1/2}(\mathbf{h})[Q_{\text{obs}}(\mathbf{h}) - Q_{\text{model}}(\mathbf{h})]$  and  $w_{\text{restraint}}^{1/2}(\mathbf{h})[D_{\text{obs}}(\mathbf{h}) - D_{\text{model}}(\mathbf{h})]$  for the quotient and difference refinements, respectively. Systematic errors shift the intercept of the plot away from the origin and cause the plot itself to deviate from linearity. The intercepts, gradients and Pearson correlation coefficients ( $r^2$ ) of normal probability plots for quotient and difference refinements are also listed in Table 3. There is generally rather little systematic difference between

the intercepts calculated for quotients or differences, all falling very close to the origin. There is no systematic difference between the correlation coefficients for the quotient and difference plots. The gradients are all near the ideal value of unity, but this is a consequence of the weight-scaling described in §3.4.

The *R*-factor and normal probability calculations indicate that cancellation of systematic errors can occur on taking quotients of intensities collected with area detectors, but the improvement, if present, is usually small and the results of using either method essentially the same. This is possibly because the assumptions about relative systematic errors referred to above are not met or because absorption and extinction are not the principal systematic errors present.

The linearity and small intercepts of the normal probability plots listed in Table 3 indicate that the weights applied to the quotients and differences reflect the uncertainties in the data. The values of reduced  $\chi^2$  listed in Table 1 are near unity for both quotient and difference-based methods, indicating that the magnitudes of the standard uncertainties are realistic. Taken with the accuracy of the values of the Flack parameters listed in Table 1 this shows that quotient and difference methods are both appropriate for absolute structure determination.

#### 4.6. Leverage analysis

Some insight into why the methods presented increase the precision in *x* can be gained by considering the relative influences of the observations  $A_{\text{obs}}(\mathbf{h})$  and  $Q_{\text{obs}}(\mathbf{h})$  or  $D_{\text{obs}}(\mathbf{h})$  on the structural and Flack parameters. Leverage analysis was carried out on the refinement of L-alanine<sup>B</sup> using *CRYSTALS* (Parsons, Wagner *et al.*, 2012).

Fig. 3(a) is a histogram of leverages for the  $|F_{\text{obs}}(\mathbf{h})|^2$  (orange) and the  $D_{\text{obs}}(\mathbf{h})$  (green) data. The  $D_{\text{obs}}(\mathbf{h})$  leverages cluster near zero showing that they have rather little effect on the overall data-fitting. The insensitivity of the structural parameters to the Friedel difference intensities reflects a similar finding described in §4.1 of Flack *et al.* (2011).

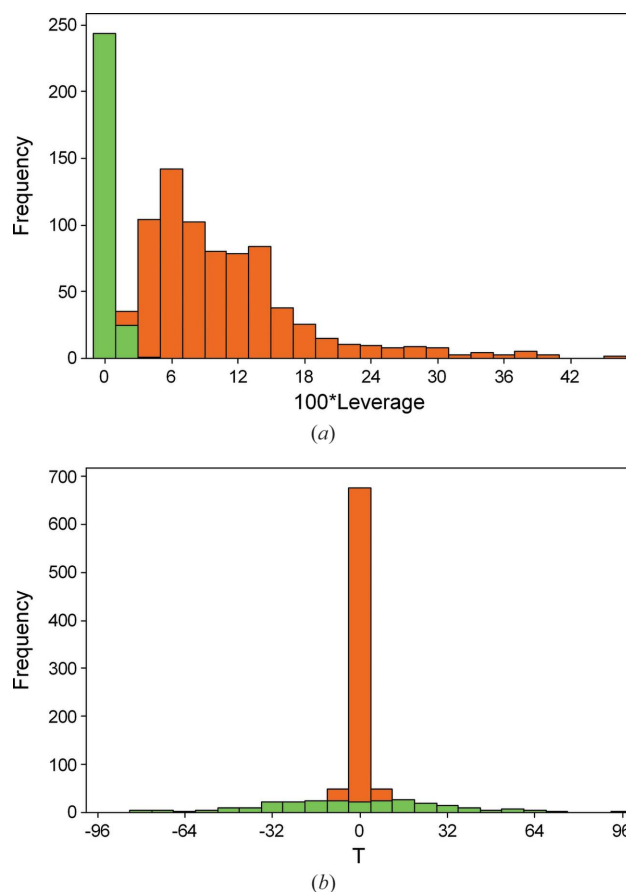
Fig. 3(b) shows a histogram of the quantity *T*, which measures the influence of observations on the precision of the Flack parameter (David, 2004; David *et al.*, 1993; Parsons, Wagner *et al.*, 2012; Prince, 2004). Here the situation seen in the leverage plot is reversed, the orange  $|F_{\text{obs}}(\mathbf{h})|^2$  data cluster about zero, whereas the green  $D_{\text{obs}}(\mathbf{h})$  data span the range  $\pm 100$ .

The improvement in the precision of the Flack parameter which is gained by using differences or quotients is the result of transforming the observations into one set  $[A_{\text{obs}}(\mathbf{h})]$  which is sensitive to the structure but independent of the Flack parameter, and another  $[D_{\text{obs}}(\mathbf{h})$  or  $Q_{\text{obs}}(\mathbf{h})]$  which is sensitive to the Flack parameter, but very insensitive to the atomic parameters. The transformation means that correlation between the Flack parameter and the other refined parameters is essentially absent, and this explains why the results of the post-refinement methods are so similar to those obtained with the method outlined above.

## 5. Concluding remarks

Absolute structure refinements have been carried out for a series of 23 light-atom crystal structures with Friedf<sub>stat</sub> values for Cu *K* $\alpha$  radiation of between 9 and 36. Accurate values of the Flack parameter and its standard uncertainty were obtained, but with a precision higher and more realistic than conventional refinement. The results of the methods implemented in *SHELXL2012*, *CRYSTALS* and *PLATON* are essentially the same as those obtained when quotients or differences are explicitly coded into the refinement in *TOPAS*. We conclude that the potential problems discussed in §1 associated with a lack of correlation between the Flack parameter and the other structural parameters are not significant in absolute structure determinations of light-atom compounds. This justifies the use of post-refinement algorithms for absolute structure determination provided a complete set of Friedel pair intensity measurements is available.

On the basis of the results presented in the current paper, and those cited herein, it is possible to provide an outline of the treatment of diffraction data which leads to a reliable value of the Flack parameter with as low and realistic standard



**Figure 3** Leverage analysis for L-alanine<sup>B</sup>. (a) Leverages of  $|F_{\text{obs}}(\mathbf{h})|^2$  and difference data,  $D_{\text{obs}}(\mathbf{h})$ . (b) Relative influences of  $|F_{\text{obs}}(\mathbf{h})|^2$  and  $D_{\text{obs}}(\mathbf{h})$  on the precision of the Flack parameter (as expressed by the quantity *T*). In each case  $|F_{\text{obs}}(\mathbf{h})|^2$  data are shown in orange and  $D_{\text{obs}}(\mathbf{h})$  data in green.

uncertainty as possible. The steps are described below in terms of intensity differences ( $D$ ), although they could equally well be applied to quotients ( $Q$ )

(i) The reflection data-set is separated into three disjoint classes consisting of centric reflections (class  $c$ ), pairs of Friedel opposites of acentric reflections (class  $ap$ ) and unpaired acentric reflections (class  $au$ ). If the data-set contains an unacceptably large proportion of unpaired acentric reflections, the diffraction data should be recollected with a revised collection strategy. The data in class  $au$  are omitted from further analysis.

(ii) The intensities of the pairs of acentric reflections (class  $ap$ ), are transformed into averages,  $A_{\text{obs}}(\mathbf{h})$ , and differences,  $D_{\text{obs}}(\mathbf{h})$ .

(iii) Least-squares refinement of the structural parameters is undertaken using as data the centric reflections (class  $c$ ) and the average intensities of the paired Friedel opposites [ $A_{\text{obs}}(\mathbf{h})$ ]. This refinement corresponds to a crystal twinned by inversion in a proportion of 50:50, so a value of the Flack parameter fixed at 0.5 should be used. The resulting atomic parameters are unbiased by the effects of resonant scattering and inversion twinning.

(iv) Using the atomic parameters obtained from stage (iii) and a Flack parameter of 0.0, structure-factor amplitudes are calculated for the pairs of Friedel opposites in class  $ap$ . This leads to model values  $D_{\text{single}}(\mathbf{h})$  corresponding to a single crystal untwinned by inversion.

(v) From a plot of  $D_{\text{obs}}(\mathbf{h})$  against  $D_{\text{single}}(\mathbf{h})$  one has a powerful tool for validating the absolute-structure determination and obtaining a value of the Flack parameter from a least-squares fit to a straight line passing through the origin. Critical to the success of this fit is the choice of a suitable weighting scheme and the elimination of outliers. The texts of the current paper, and those cited, contain the corresponding information.

The transformation of the data described in (ii) yields one set of observations [the centric data and  $A_{\text{obs}}(\mathbf{h})$ ] which is sensitive to the structure but independent of the Flack parameter, and another [ $D_{\text{obs}}(\mathbf{h})$  or  $Q_{\text{obs}}(\mathbf{h})$ ] which is sensitive to the Flack parameter, but highly insensitive to the atomic parameters. The agreement between  $A_{\text{obs}}(\mathbf{h})$  and  $A_{\text{model}}(\mathbf{h})$  is usually much better than between  $D_{\text{obs}}(\mathbf{h})$  and  $D_{\text{model}}(\mathbf{h})$  or their equivalents based on quotients (e.g. Flack *et al.*, 2011; Parsons, Pattison *et al.*, 2012); one advantage of the transformation of data into  $A$  and  $D$  (or  $Q$ ) is that different schemes for the selection of outliers and weights can be applied to each. A conventional refinement is compatible only with a single 'one-size-fits-all' weighting scheme. Use of the transformed data also removes correlation between the Flack parameter and the other structural parameters.

All data in this work were collected with Cu  $K\alpha$  radiation, although there is no reason why the methods described could not be applied to data from more than one source, for example, Cu  $K\alpha$  radiation for the Friedel-averaged intensity data  $A_{\text{obs}}(\mathbf{h})$  and Cr  $K\alpha$  radiation for the difference or quotient data  $D_{\text{obs}}(\mathbf{h})$  or  $Q_{\text{obs}}(\mathbf{h})$ . This will be investigated in due course.

We thank Dr Alan Coelho for making the substantial enhancements to *TOPAS-Academic* which made the work reported in this paper possible, and Mr Philippe Piechon, Mr Tom Heathcote, Mr Paul McGovern and Dr Oliver Presly for the collection of data used in this work. We would also like to acknowledge the continued interest and advice given by Dr Richard Cooper, Professor George Sheldrick, Dr Devinder Sivia and Dr David Watkin. We thank Dr Robert Westwood (Cyclacel Ltd) for samples of *R*- and *S*-CYCLO, and the anonymous referees for their comments on the manuscript of this paper.

## References

- Abrahams, S. C. & Keve, E. T. (1971). *Acta Cryst.* **A27**, 157–165.
- Bernardinelli, G. & Flack, H. D. (1985). *Acta Cryst.* **A41**, 500–511.
- Betteridge, P. W., Carruthers, J. R., Cooper, R. I., Prout, K. & Watkin, D. J. (2003). *J. Appl. Cryst.* **36**, 1487.
- Beurskens, P. T., Beurskens, G., Bosman, W. P., de Gelder, R., Garcia-Granda, S., Gould, R. O., Israel, R. & Smits, J. M. M. (1996). *DIRDIF96*. University of Nijmegen, The Netherlands.
- Blessing, R. H. (1995). *Acta Cryst.* **A51**, 33–38.
- Blessing, R. H. (1997). *J. Appl. Cryst.* **30**, 421–426.
- Bruker–Nonius (2006). *SAINT*, Version 7. Bruker-AXS, Madison, Wisconsin, USA.
- Coelho, A. A. (2012). *TOPAS-Academic*, Version 5. Bruker AXS, Karlsruhe, Germany.
- Coelho, A. A., Evans, J., Evans, I., Kern, A. & Parsons, S. (2011). *Powder Diffr.* **26**, s22–S25.
- David, W. (2004). *J. Res. Natl. Inst. Stand. Technol.* **109**, 107.
- David, W. I. F., Ibberson, R. M. & Matsuo, T. (1993). *Proc. R. Soc. London Ser. A*, **442**, 129–146.
- Dittrich, B., Strumpel, M., Schäfer, M., Spackman, M. A. & Koritsánszky, T. (2006). *Acta Cryst.* **A62**, 217–223.
- Flack, H. D. (1983). *Acta Cryst.* **A39**, 876–881.
- Flack, H. D. (2012). *Acta Cryst.* **C68**, e13–e14.
- Flack, H. D. & Bernardinelli, G. (1999). *Acta Cryst.* **A55**, 908–915.
- Flack, H. D. & Bernardinelli, G. (2000). *J. Appl. Cryst.* **33**, 1143–1148.
- Flack, H. D. & Bernardinelli, G. (2008a). *Acta Cryst.* **A64**, 484–493.
- Flack, H. D. & Bernardinelli, G. (2008b). *Chirality*, **20**, 681–690.
- Flack, H. D., Sadki, M., Thompson, A. L. & Watkin, D. J. (2011). *Acta Cryst.* **A67**, 21–34.
- Flack, H. D. & Shmueli, U. (2007). *Acta Cryst.* **A63**, 257–265.
- Hamilton, W. C. & Abrahams, S. C. (1970). *Acta Cryst.* **A26**, 18–24.
- 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.
- Le Page, Y., Gabe, E. J. & Gainsford, G. J. (1990). *J. Appl. Cryst.* **23**, 406–411.
- Merli, M., Ungaretti, L. & Oberti, R. (2000). *Am. Mineral.* **85**, 532–542.
- Murphy, P. J., Thomas, D. A., Sievert, N., Caulket, P. W. R. & Parsons, S. (2010). *Lett. Org. Chem.* **7**, 508–510.
- Oxford Diffraction (2010). *CrysAlis PRO*, Version 1.171.33.55. Oxford Diffraction Ltd, Yarnton, Oxfordshire, England.
- Palatinus, L. & Chapuis, G. (2007). *J. Appl. Cryst.* **40**, 786–790.
- Parsons, S. & Flack, H. (2004). *Acta Cryst.* **A60**, s61.
- 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. (2004). *Mathematical Techniques in Crystallography and Materials Science*, 2nd ed. Berlin: Springer.

- Rodríguez-Carvajal, J. & González Platas, J. (2009). *CrysFML*. Institut Laue Langevin, Grenoble, France, Universidad de La Laguna, La Laguna, Spain.
- Sheldrick, G. M. (2001). *XPREP*. University of Göttingen, Germany.
- Sheldrick, G. M. (2008*a*). *SADABS*, Version 2008–1. University of Göttingen, Germany.
- Sheldrick, G. M. (2008*b*). *Acta Cryst.* **A64**, 112–122.
- Sheldrick, G. M. (2012). *SHELXL2012*. University of Göttingen, Germany.
- Spek, A. L. (2003). *J. Appl. Cryst.* **36**, 7–13.
- Thompson, A. L. & Watkin, D. J. (2011). *J. Appl. Cryst.* **44**, 1017–1022.
- Wang, S., McClue, S. J., Ferguson, J. R., Hull, J. D., Stokes, S., Parsons, S., Westwood, R. & Fischer, P. M. (2001). *Tetrahedron Asymmetry*, **12**, 2891–2894.