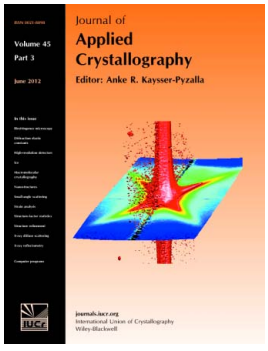
Journal of
Applied Crystallography
ISSN 0021-8898
Editor: **Anke R. Kaysser-Pyzalla**

Applications of leverage analysis in structure refinement

Simon Parsons, Trixie Wagner, Oliver Presly, Peter A. Wood and Richard I. Cooper

J. Appl. Cryst. (2012). **45**, 417–429



Journal of
Applied Crystallography
ISSN 0021-8898

Received 6 October 2011
Accepted 6 April 2012

Applications of leverage analysis in structure refinement

Simon Parsons,[a]* Trixie Wagner,[b] Oliver Presly,[c] Peter A. Wood[a]‡ and Richard I. Cooper[d]

[a]School of Chemistry and Centre for Science at Extreme Conditions, The University of Edinburgh, King's Buildings, West Mains Road, Edinburgh EH9 3JJ, Scotland, [b]Novartis Institutes for BioMedical Research, 4002 Basel, Switzerland, [c]Agilent Technologies, Unit 10, Mead Road, Yarnton, Oxfordshire OX5 1QU, England, and [d]Chemistry Research Laboratory, University of Oxford, 12 Mansfield Road, Oxford OX1 3TA, England. Correspondence e-mail: s.parsons@ed.ac.uk

Leverages measure the influence that observations (intensity data and restraints) have on the fit obtained in crystal structure refinement. Further analysis enables the influence that observations have on specific parameters to be measured. The results of leverage analyses are discussed in the context of the amino acid alanine and an incomplete high-pressure data set of the complex bis(salicylaldoximato)copper(II). Leverage analysis can reveal situations where weak data are influential and allows an assessment of the influence of restraints. Analysis of the high-pressure refinement of the copper complex shows that the influence of the highest-leverage intensity observations increases when completeness is reduced, but low leverages stay low. The influence of restraints, notably those applying the Hirshfeld rigid-bond criterion, also increases dramatically. In alanine the precision of the Flack parameter is determined by medium-resolution data with moderate intensities. The results of a leverage analysis can be incorporated into a weighting scheme designed to optimize the precision of a selected parameter. This was applied to absolute structure refinement of light-atom crystal structures. The standard uncertainty of the Flack parameter could be reduced to around 0.1 even for a hydrocarbon.

1. Introduction

Observations (reflection intensities and restraints) do not contribute equally to data fitting during crystal structure refinement. Some observations are extremely influential, while others have hardly any influence at all. The quantity that measures the influence that an observation has on the fit obtained in a refinement is called the leverage, and it can be calculated from the matrix that is used to describe the model in least squares. The leverage tells us how the value of a data point calculated by the model changes in response to a change in the observed value.

The aim of the present paper is to discuss how information on leverages can be used during structure analysis and interpretation. We will show that leverages provide valuable information on factors such as the importance of weak data in modelling and the efficacy of restraints; we will further show that they can be used to address one of the most pressing issues in chemical crystallography, the precise determination of absolute structure for organic compounds that contain no element heavier than oxygen.

An understanding of the kind of information that leverages convey can be obtained by consideration of a simple one-parameter straight-line fit to $y = mx$. The data in Fig. 1 were constructed to give a best fit line of $y = 0.0x$, and illustrate different ways in which points can contribute to the fit. The figure in parentheses next to each of the points in Fig. 1 is the leverage of that point. Point A, at $x = -5$, has a leverage of 0.11, *i.e.* if the observed value of A changed from $y = 4$ to $y = 5$

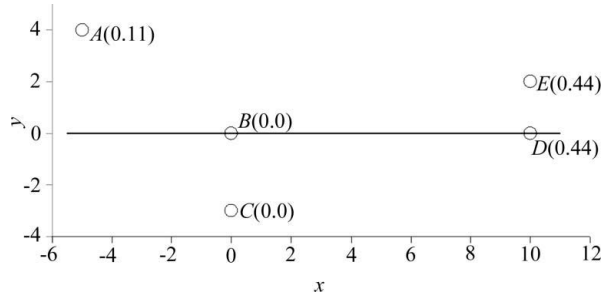


Figure 1
Leverages calculated in the simple linear least-squares fit of the data points A (−5, 4), B (0, 0), C (0, −3), D (10, 0) and E (10, 2) with the function $y = mx$. The figures in parentheses next to each point are the leverages.

‡ Present address: Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, England.

the model would alter such that the calculated value of y at point A would change from zero to 0.11. Leverages can thus be interpreted as the effect that an observation has on its own calculated value (see below). This idea is illustrated further by the points at $x = 0$. The fit to $y = mx$ requires the solution to intercept the y axis at $y = 0$, and the calculated values of y at points B and C will always be zero no matter what the measured value of y is. Both points therefore have zero leverage, and no matter how large their deviation from the model, these points exert no influence on the fit and therefore on their own calculated values. The most extreme points (D and E), at $x = 10$, have the highest leverages (0.44) and therefore the most influence on the model. Point D has zero error and a large leverage, while E has a large error and large leverage. D and E have exactly the same leverage values, despite having different deviations from the model, because the leverage is derived from the model and not the observed values (more detail is given below). Note also that the sum of the five leverages for points A–E is equal to 1, the number of parameters being fitted.

The calculation of leverages in crystallographic least squares has been discussed by Prince and co-workers (Prince, 2004; Prince & Nicholson, 1985; Prince & Spiegelman, 2004*a*,*b*); a discussion of the topic is also available in standard statistics texts such as Rawlings *et al.* (1998). The mathematics is given in full detail in the articles and book by Prince and co-workers, and only a summary is given here. The analysis is based on the projection matrix **P**, which relates the observed (**y**) and calculated ($\hat{\mathbf{y}}$) values of the observations: $\mathbf{P}\mathbf{y} = \hat{\mathbf{y}}$. It is derived as follows: a set of linear equations relates a set of undetermined parameters **x** to a set of observations **y**, so that $\mathbf{y} = \mathbf{A}\mathbf{x}$, where **A** is the design matrix. The parameters $\hat{\mathbf{x}}$, which minimize the squared residual between the observations, **y**, and their calculated values, $\hat{\mathbf{y}}$, are found by solving the normal equations $\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{y} = \mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}\hat{\mathbf{x}}$, where **W** is a weight matrix. Pre-multiplying both sides by the inverse of $\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}$ gives the solution $(\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{y} = \hat{\mathbf{x}}$. Pre-multiplying both sides of this equation by **A** gives $\hat{\mathbf{y}}$: $\mathbf{A}(\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{y} = \mathbf{P}\mathbf{y} = \hat{\mathbf{y}}$. Note that the calculation of **P** is based on the design and weight matrices; the observations are not used.

It is computationally convenient to define a matrix $\mathbf{P}'$ which is related to **P** by pre-multiplying both sides of $\mathbf{y} = \mathbf{A}\mathbf{x}$ by **U**, the upper-right Cholesky factor of the weight matrix, **W**, to give $\mathbf{P}'\mathbf{y}' = \hat{\mathbf{y}}'$, where $\mathbf{y}' = \mathbf{U}\mathbf{y}$. For a diagonal weight matrix, $\mathbf{P}'$ has the same diagonal as **P**, but it is now symmetric and may be constructed using only a single matrix $\mathbf{Z} = \mathbf{U}\mathbf{A}$: $\mathbf{P}' = \mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}$. **P** and $\mathbf{P}'$ are square matrices of dimensions $N_{\mathrm{obs}} \times N_{\mathrm{obs}}$, where N_{obs} is the number of observations used in the refinement.

In other branches of statistics **P** is sometimes referred to as the hat matrix because it relates **y** to $\hat{\mathbf{y}}$. The relationship $\mathbf{P}\mathbf{y} = \hat{\mathbf{y}}$ enables each calculated $\hat{y}_i$ to be written as a linear combination of the observations contained in **y**. This means that an element along the leading diagonal of **P** (P_{ii}) measures the contribution that an observation y_i makes to its own calculated value, something that was illustrated in the simple straight-line-fit example above. The values of P_{ii} are the leverages. They have a maximum value of 1 and a minimum value of 0, and they measure how much influence an observation has on its calculated value. A value of 1.0 means that the observation entirely determines its own calculated value but has no influence on any other observation. The average leverage for a refinement is equal to $N_{\mathrm{parameters}}/N_{\mathrm{observations}}$.

Prince extended his analysis by considering which observations are most important for determining the precision of a particular parameter. The analysis enables us to state the amount by which re-measurement of the ith data point will reduce the variance of the estimate of the jth parameter. The dot product of the ith row of **Z** and the jth column of the inverse normal matrix, $(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}$, yields the value of a quantity designated t_{ij}. The value of $t_{ij}^2/(1 + P'_{ii})$ measures the influence of the ith observation on the variance of the jth parameter; we shall refer to this quantity as T_{ij}^2. It should be noted that the product of **Z** and the inverse normal matrix is related by a matrix transpose to the matrix that is used to solve the normal equations for **x**. The significance of this matrix is that it reveals the magnitude and sense of the contribution that each observed value makes to each model parameter; this feature is discussed in more detail in §3.4.

A high value of T_{ij}^2 implies that the ith observation is very important for determination of the jth parameter. Information of this type was used by David *et al.* (1993) to analyse the influence of different regions of the neutron powder diffraction pattern of C_{60} on parameters used to track disorder that develops as temperature is increased. The procedure was also used by Hazen & Finger (1989) to optimize the precision of the oxygen positional parameters in pyrope by collecting reflections that were most sensitive to these parameters. The most recent work on leverage analysis has been published by Merli *et al.* (2001, 2000, 2002), who have applied it to refinements of mineral structures. Their approach has been applied particularly to understanding the role of different classes of data in determining occupancies on mixed metal sites in minerals. The same group has used leverages and other statistical tools such as Cook's distances to identify outliers in refinement, applying this information to improve the robustness of crystallographic least squares (Merli, 2005; Merli & Sciascia, 2011; Merli *et al.*, 2010).

2. Experimental

2.1. Calculation of leverages and T^2 values

One factor that has hindered wider application of leverage analysis is that the matrices required for the necessary calculations are not available as output from commonly used refinement packages. The program *CRYSTALS* (Betteridge *et al.*, 2003) has been modified to output the matrix **Z**, and the normal matrix and its inverse. (In *CRYSTALS*, the command sequence

```
#SFLS
REFINE PUNCH = MATLAB
END
```

outputs files containing the matrix **Z** and the normal matrix and its inverse, which are used as input to a program called *HATTIE*.)

HATTIE has been written to calculate and output leverages and T and T^2 values for observations to a file suitable for input into a spreadsheet program. Also written to the file are the Miller indices, Y_o, $\sigma(Y_o)$, Y_c, $\sin\theta/\lambda$ and $Y_o/\sigma(Y_o)$ for each reflection, where Y may represent $|F|$ or $|F|^2$, and the subscripts o and c refer to observed and calculated quantities. The calculations apply both to intensity data and to any restraints applied during refinement. The code makes use of several subroutines available in the CrysFML Fortran library (Rodríguez-Carvajal & Platas, 2009). Leverages, which have a maximum value of 1.0, are multiplied by 100, and T values, which are numerically very small, are scaled so that $|T_{max}| = 100$.

Leverage analysis was carried out using both simulated and experimental data on two crystal structures: the amino acid L-alanine and the metal complex bis(salicylaldoximato)-copper(II) [which is abbreviated to $Cu(sal)_2$]. All leverage analyses were performed at refinement minima.

2.2. L-Alanine

L-Alanine is the simplest chiral amino acid (see Fig. S1*a* in the supplementary material[1]). It is zwitterionic in the solid state with formula $^+H_3NCH(Me)CO_2^-$. The crystal structure is orthorhombic, forming in space group $P2_12_12_1$. Experimental intensity data were collected at 100 K on an Agilent Technologies SuperNova diffractometer using a Cu $K\alpha$ microsource. Data were collected to a resolution of 0.84 Å with an average redundancy of 14.9. A multiscan correction for systematic errors was applied, and data were merged (in point group 222) in *SORTAV* (Blessing, 1997). The structure of alanine was refined in *CRYSTALS* against $|F|^2$ using all data. Weights equal to $1/\sigma^2(|F|^2)$ were applied, with a robust-resistant modifier (Prince & Nicholson, 1983) which zero weighted 14 out of 740 reflections as outliers; all such outliers were omitted from further analysis. All non-H atoms were refined with anisotropic displacement parameters. H-atom positions and isotropic displacement parameters were subject to typical bond distance and angle restraints, with $U_{iso}(H)$ restrained to 1.2 or 1.5 times U_{equiv} of the parent C or N atom. The program defaults were used for standard deviations applied to the restraints: 0.02 Å, 2° and 0.002 Å^2 for the distances, angles and displacement parameters, respectively. The extinction coefficient refined to 4.92 (11) and the Flack (1983) parameter refined to 0.00 (13). The final conventional R factor (unweighted, calculated on $|F_o|$ using data with $|F_o| > 4\sigma(|F_o|)$) was 1.59%. The goodness of fit was 2.715, but the normal probability plot was linear, with an intercept of 0.04 and a correlation coefficient of 0.996.

A simulated data set was calculated using *XPREP* (Sheldrick, 2001) to a resolution of 0.4 Å. Uncertainties were estimated according to $\sigma(|F|^2) = 0.02|F|^2 + \langle|F|^2\rangle/1000$. Gaussian random errors were added to the simulated intensities [subroutine *GASDEV* from Press *et al.* (1992)].

2.3. Bis(salicylaldoximato)copper(II) [$(Cu(sal)_2$]

The complex consists of two salicylaldoximate ligands bound to Cu in a square planar arrangement (Fig. S1*b*). The data used for the present calculations were collected as part of a wider investigation into the effects of high pressure on complexes of salicylaldoximate ligands; the full results of this study (Byrne *et al.*, 2011) will be reported later. The crystal structure is monoclinic, forming in $P2_1/c$ with the Cu atoms located on inversion centres. Data were collected with synchrotron radiation on beamline I19 at Diamond Light Source with $\lambda = 0.4959$ Å at a pressure of 0.55 GPa; the crystal was held in a modified Merrill–Bassett diamond anvil cell with a half-opening angle of 40° (Moggach *et al.*, 2008; Merrill & Bassett, 1974). The average redundancy was 6.1. The diffractometer on I19 consists of a Crystal Logic four-circle κ-goniometer with a Rigaku Saturn CCD detector. The data collection images were converted to Bruker `.sfrm` format using the program *ECLIPSE* (Parsons, 2004) and processed using *SAINT* (Version 7; Bruker–Nonius, 2006). Shading of the detector by the pressure cell was taken into account using integration masks, also generated by *ECLIPSE*. A multiscan correction was applied using *SADABS* (Sheldrick, 2008*b*), and data were merged with *SORTAV*. The completeness of the final data set was 51.2% to a resolution of 0.85 Å.

The crystal structure was refined in *CRYSTALS* as described above for L-alanine. A robust-resistant modifier was applied to the $1/\sigma^2(|F|^2)$ weighting scheme, leading to zero weighting of 40 out of 567 reflections, mostly having diffracted beams very close to the opening angle limits of the cell. High-pressure data sets are usually incomplete and it is common practice to apply restraints to help stabilize refinements. The bond distances and angles of the salicylaldoximate ligand were restrained to the values determined from a complete data set measured at ambient pressure. Rigid-bond and rigid-body similarity restraints were applied to the anisotropic displacement parameters of the C, N and O atoms. The H atoms attached to sp^2 carbon atoms were restrained to be coplanar with the ligand. The standard deviations applied to the restraints were 0.01 Å, 1°, 0.01 Å, and 0.005 and 0.04 Å^2 for the distances, angles, planarity, and rigid-bond and rigid-body restraints. Restraints were applied to H atoms as described above for L-alanine (also using the same standard deviations as for L-alanine). The final conventional R factor was 2.87%. The goodness of fit was 1.080, and the normal probability plot had an intercept of −0.07 and a correlation coefficient of 0.999.

For the purposes of comparison a complete data set was collected under ambient conditions using a Bruker APEXII diffractometer and Mo $K\alpha$ radiation. Integration was carried out using *SAINT* and an absorption correction applied using *SADABS*. The structure was refined using the same procedure outlined above for the high-pressure data set.

[1] Fig. S1 is available from the IUCr electronic archives (Reference: HE5536). Services for accessing this material are described at the back of the journal.

2.4. Test data for absolute structure refinements

§3.6 describes a method where leverage analysis is used to improve the precision of the Flack parameter in some absolute structure refinements. Seventeen data sets were used to test the method.

Data sets 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, a Bruker D8 microsource, also equipped with a SMART 6000 detector, or an Agilent Technologies SuperNova, also incorporating a microsource generator. For data collections with the Bruker instruments a typical data collection comprised 16 ω scans at varying φ angles (four scans at $2\theta = 46^\circ$ and 12 scans at $2\theta = 94^\circ$), yielding complete data up to 0.84 Å. The redundancy for orthorhombic crystals is around 11; for monoclinic crystals it is almost 6. The exposure times for the high- and low-resolution scans differed by a factor of 3–4 to ensure sufficient signal-to-noise ratios in the high-resolution shells. Data were processed with *SAINT* and corrected for absorption and systematic errors using *SADABS*. For the data collections using the Agilent system a strategy was calculated to a defined redundancy. Processing, including integration and a multiscan absorption correction, was accomplished with *CrysAlis Pro* (Oxford Diffraction, 2010).

Data were merged using the program *SORTAV* using unit weights and robust-resistant down-weighting of outliers. The standard deviations output by *SORTAV* are estimates of the standard uncertainty of the population rather than of the sample-estimated mean. This quantity should converge to an approximately constant value as redundancy increases. Its use in merging data has been justified by Blessing (1997).

Structures were refined against $|F|^2$ in *CRYSTALS* using all data. All non-H atoms were refined with anisotropic displacement parameters. H-atom positions and isotropic displacement parameters were refined subject to restraints. Flack and extinction parameters were also refined. The weights were equal to $1/\sigma^2(|F|^2)$ multiplied by a robust-resistant modifier as described by Prince & Nicholson (1983). Reflections given zero weight in this procedure were omitted. Goodness-of-fits, S, were in the region of 2, and the weights were rescaled using a facility available in *CRYSTALS* to give $S \simeq 1$. These weights were output along with other files needed for leverage analysis and used for the modified weight calculations described in §3.6.

3. Results and discussion

Figs. 2–4 illustrate the results of the leverage analyses described below. The value of $|F_o|$ (scaled to $|F_{o,max}| = 100$) is used to represent intensity even though refinements were carried out on $|F|^2$; this is to be consistent with existing literature and also aids comparisons and provides clearer dispersion of points for low-intensity data. Leverages were normalized by dividing them by $N_{parameters}/N_{observations}$, that is by the mean leverage value. Observations take the form of intensity data and any restraints applied during refinement.

3.1. Leverages in alanine

Figs. 2(*a*)–2(*c*) show plots of leverage against $|F_o|$, $|F_o|/\sigma(|F_o|)$ and $\sin\theta/\lambda$ for the $|F|^2$ refinement of aniline against all data with $1/\sigma^2$ weights. From Fig. 2(*a*) it can be seen that the most influential data are those with moderately weak intensities, the leverage falling off towards very low or very high intensity; a similar effect is apparent when leverages are plotted against $|F_o|/\sigma(|F_o|)$ (Fig. 2*b*). Fig. 2(*c*) reveals the importance of the high-resolution data, with leverages showing an increasing trend with $\sin\theta/\lambda$.

Although weak data do not appear to be especially influential in alanine the same is not necessarily true of all structures. Weak data may be very important in pseudosymmetric structures, for example in distinguishing between centrosymmetric and noncentrosymmetric models (Dunitz, 1995; Kassner *et al.*, 1993; Marsh, 1981). The organic compound 4-cyano-4′-[(4*R*)-4,5-epoxypentyloxy]biphenyl, which has one asymmetric carbon centre, crystallizes in $P2_1$ with two molecules in the asymmetric unit (Clegg *et al.*, 1998). With the exception of the asymmetric carbon atom these two molecules are related by a pseudo-inversion centre so that the space group is almost $P2_1/n$. The leverages, calculated using the intensity data available as supplementary material to the article by Clegg and co-workers, are plotted against $|F_o|/\sigma(|F_o|)$ in Fig. 3; this should be compared with Fig. 2(*b*), which shows the same data for alanine. There are more high-leverage points amongst the weak data in the former, attesting to the importance of weak data in this structure.

3.2. Leverage analysis of restraints in alanine

Restraints are incorporated into refinement in the least-squares design matrix, and the calculations described above yield leverage values for restraints as well as intensity data. Restraints were applied in the alanine refinement, and the column of points at the far left of the plots in Figs. 2(*a*)–2(*c*) corresponds to their leverages; they are clearest in Fig. 2(*c*). The normalized leverages are generally above average (*i.e.* greater than 1), showing that the restraints have an important influence on the refinement.

The highest leverage values correspond to restraints applied to the isotropic displacement parameters of the H atoms, assigning target values equal to some multiple (1.2 or 1.5) of the equivalent isotropic displacement parameter of their parent atoms. These points have normalized leverages of above 4 and absolute leverage values in the region of 0.5. This means that, though the restraints are important, the values of the H-atom displacement parameters are significantly influenced by the intensity data. Had the absolute leverages been closer to 1 this would have implied that the displacement parameters were simply fitting the restraint applied with little or no influence from the intensity data. The next block of points at the far left of Fig. 2(*c*), with normalized leverages of between 1 and 2, corresponds to restraints applied to N—H and C—H distances, while the lowest points with normalized leverages of less than 1 correspond to the H—N—H and H—C—H angle restraints.

Leverage analysis is useful in the interpretation of the results of a restrained refinement because it shows which restraints are significantly influencing the fit and to what extent they define the final value of a parameter. A leverage value close to 0 implies that the data point in question has little influence. A restraint with a very low leverage might as well be deleted, or, if it is thought to be important, it should have its uncertainty decreased, though not beyond a realistic estimate of the spread of values that the restrained parameter might adopt. Conversely, if a restraint has an absolute leverage near 1.0 this indicates a forced fit: the refinement has converged on whatever value was typed into the restraint list of the refinement program.

3.3. The effects of incomplete data: leverage analysis of $Cu(sal)_2$

The data set for $Cu(sal)_2$ was collected at high pressure, and the completeness is low as a result of shading of reciprocal space by the pressure cell. The plots shown in Figs. 4(*a*)–4(*c*) show leverage *versus* $|F_o|$, $|F_o|/\sigma(|F_o|)$ and $\sin\theta/\lambda$ plots for the refinement of $Cu(sal)_2$. Here the trends are seen to be different from those described above for alanine, with a larger spread of leverage values. There is a broad distribution of points spreading from low to moderately high values of $|F_o|$ in Fig. 4(*a*), and the sharp peak in the $|F_o|$ *versus* leverage plot present in Fig. 2(*a*) is absent. The standard deviations of the normalized leverage values are 0.75 for alanine and 1.07 for $Cu(sal)_2$.

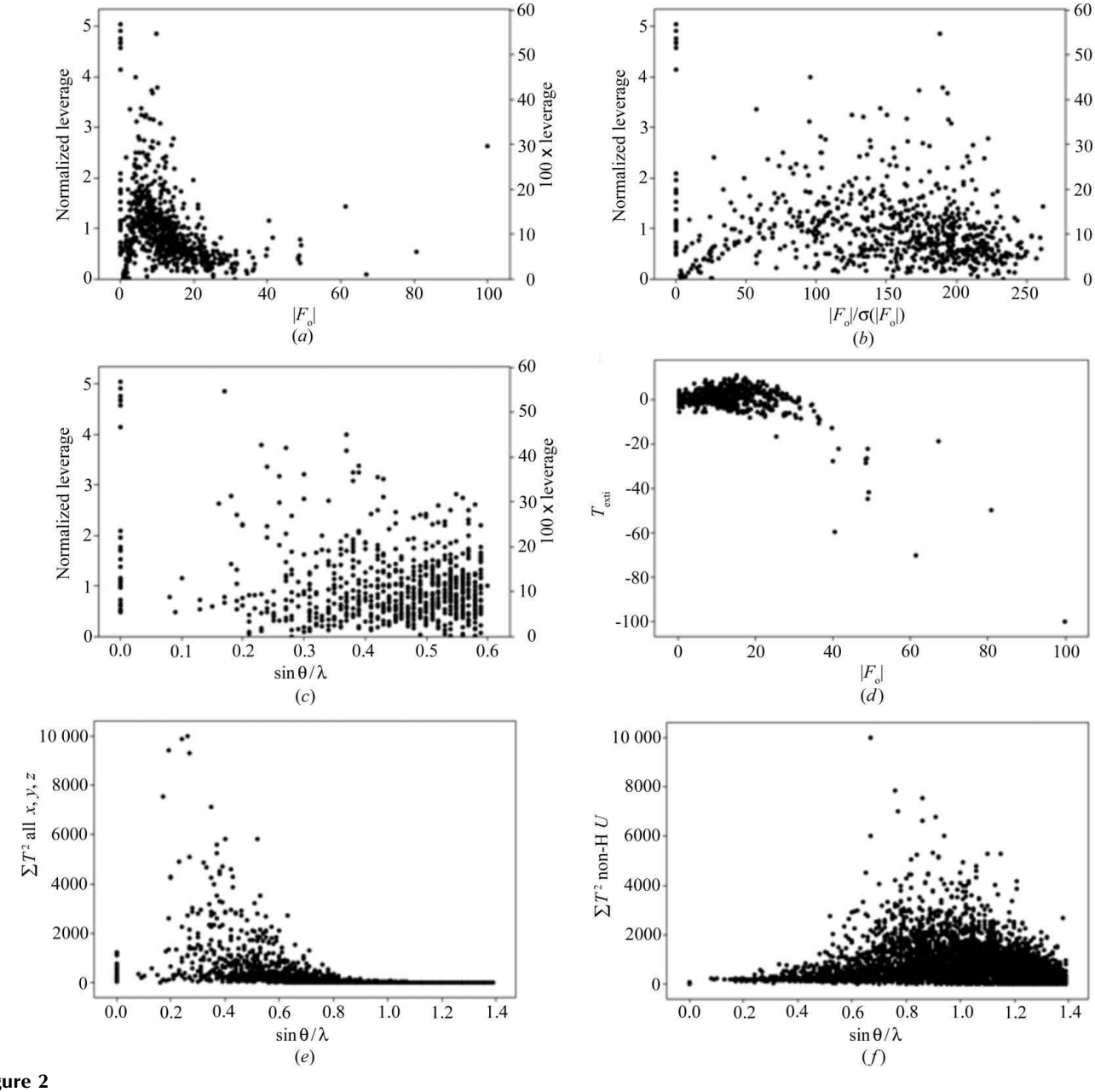


Figure 2
(*a*)–(*c*) Leverage analysis for alanine as a function of $|F_o|$, $|F_o|/\sigma(|F_o|)$ and $\sin\theta/\lambda$, respectively. (*d*) Values of signed T values for the extinction parameter plotted against $|F_o|$. (*e*), (*f*) Sums of T^2 values plotted against $\sin\theta/\lambda$ for, respectively, fractional coordinates and non-H-atom ADPs for simulated data. The columns of points on the far left of the plots correspond to the restraints.

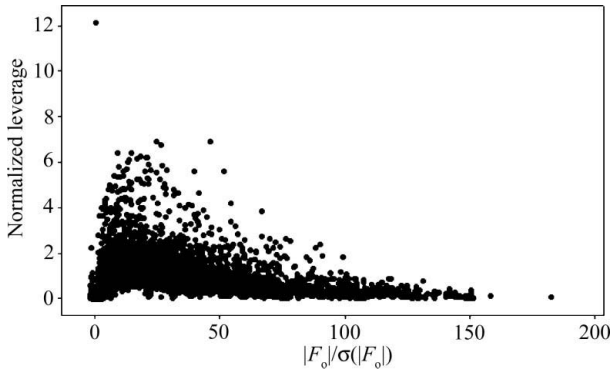


Figure 3
Leverage analysis based on $|F_o|/\sigma(|F_o|)$ for the pseudosymmetric structure referred to in the text. Notice that there are more high-leverage points amongst the weak data than in Fig. 2(*b*).

A number of the restraints have normalized leverages of >5 and absolute leverage values of 0.8 or more; these occur at the top of the column of points at the left of Figs. 4(*a*)–4(*c*). Some of these correspond to restraints applied to H-atom displacement parameters and to planarity restraints involving H atoms. The C—H and N—H distance restraints have absolute values of 0.5–0.7, substantially higher than in alanine. The high leverage values for restraints involving H-atom parameters are quite reasonable for a heavy-atom compound.

Also found amongst the highest leverage values are rigid-bond restraints applied to the anisotropic displacement parameters (ADPs) of atoms forming the ligand; these are known as 'DELU' restraints to *SHELX* (Sheldrick, 2008*a*) users, and apply the Hirshfeld rigid-bond criterion as a restraint. The smallest leverages, with values close to 0, relate to rigid-body ('SIMU') restraints, which restrain the U_{ij} values of neighbouring atoms to be equal. Refinement of ADPs against incomplete high-pressure data sets usually leads to elongation along the direction where data are missing, and it is therefore not unexpected that restraints applied to ADPs should have high leverage values. However, the rigid-bond restraints are much more influential than the rigid-body restraints. Although rigid-bond restraints are usually applied with higher weight than rigid-body restraints, the complete lack of any leverage for the latter was surprising, and the analysis shows that in view of the acceptable ADPs obtained in the refinement (Fig. S1*b*) the rigid-body restraints might as well be deleted.

A possible procedure for assessing the effect of completeness on leverages might be to compare leverages from a refinement using the high-pressure data set just discussed with another using a complete data set collected at ambient pressure. The problem with this procedure is that the experimental values of $\sigma(|F_o|^2)$ would differ between the two data sets and so any comparison would be complicated by the effect of

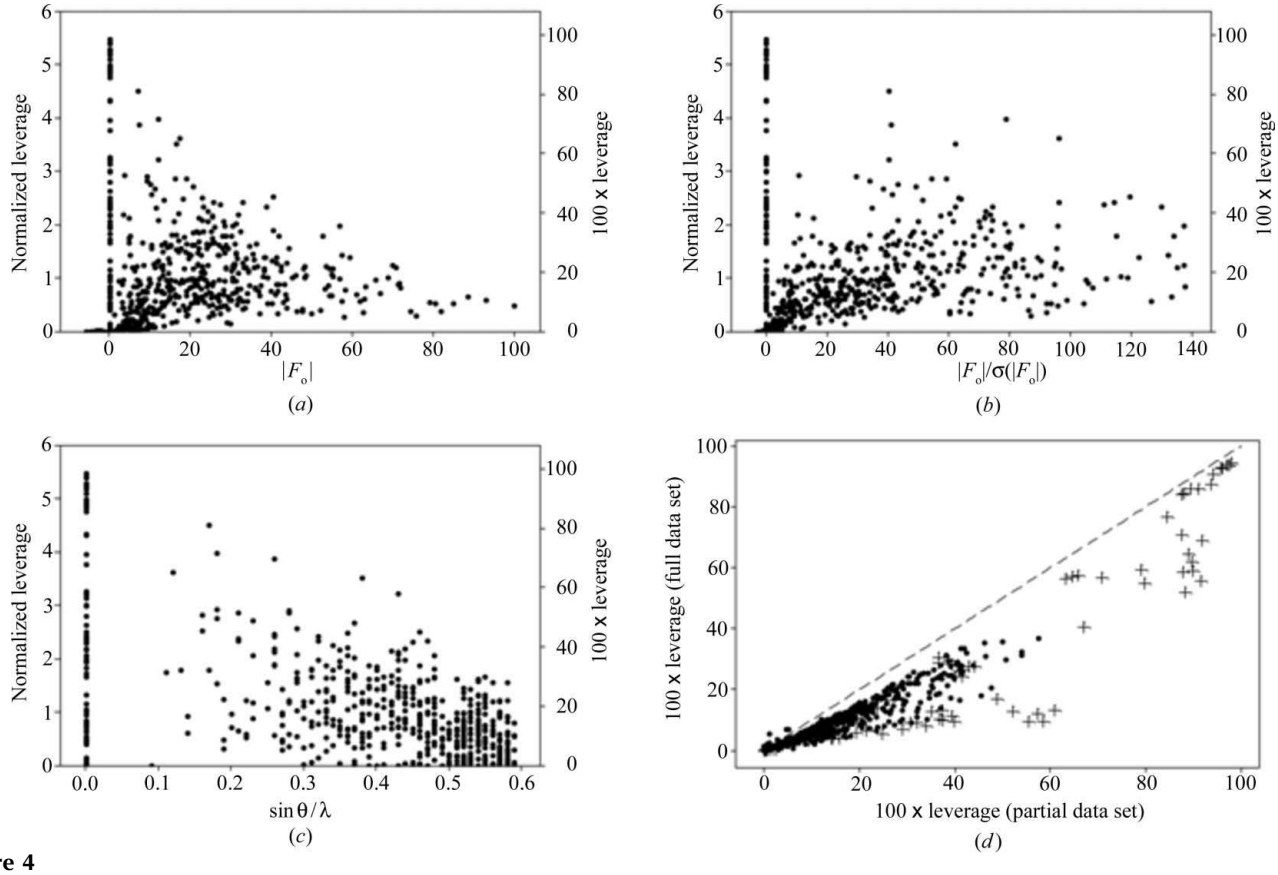


Figure 4
(*a*)–(*c*) Leverage analysis for $Cu(sal)_2$ using data with a completeness of around 50% as a function of $|F_o|$, $|F_o|/\sigma(|F_o|)$ and $\sin\theta/\lambda$, respectively. (*d*) Comparison of leverages from refinements against complete and partial data sets; the points shown as plus signs (+) refer to the restraints, and the dashed line traces the path of $y = x$.

different refinement weights. Instead a complete data set was collected under ambient conditions and a partial data set generated from this by taking only those data which had been measured in the high-pressure data set. The weights [$= 1/\sigma^2(|F_o|^2)$] for equivalent reflections in refinements using the complete and partial data sets are then the same. The same set of restraints (see *Experimental*) was applied in both refinements.

A plot of leverage values for equivalent reflections in the two refinements is shown in Fig. 4(*d*), in which intensity data are shown as dots and restraints as plus signs. The average leverage ($N_{\rm parameters}/N_{\rm observations}$) must be larger in the incomplete data set, and essentially all points in the graph are to the right of the line $y = x$. There is a tendency for intensity data that are already influential when the data are complete to become more influential when the data are incomplete. Low-leverage reflections tend to stay low. Lack of completeness also has a significant effect on some of the restraint leverages. There is a horizontal spread of plus signs in Fig. 4(*d*) near the x axis, corresponding to a marked increase in the influence of rigid-bond restraints applied to the anisotropic displacement parameters of the ligand. The highest restraint leverages, which apply to H-atom isotropic displacement parameters, are the same for both data sets.

3.4. Interpretation of T^2 and T values

While leverages measure the overall influence that a data point has on a refinement, it may be of more interest to ask which data points influence a specific parameter. This information is contained in the T^2 values that can be generated in a leverage analysis. A high T^2 value indicates an influential observation.

David and co-workers (David *et al.*, 1993; David, 2004) have recommended analysis of signed T values [$= t_{ij}/(1 + P'_{ii})^{1/2}$] as they show whether a data point makes a parameter more positive or more negative. These authors illustrated this idea using displacement parameter T values in a Rietveld refinement. Short-*d*-spacing data all had negative T values because a relative increase in the intensities of these data would make the displacement parameter smaller. Conversely, long-*d*-spacing data all had positive T values. Fig. 2(*d*) shows the variation of T values for the extinction parameter in alanine. The numerically largest values of T occur for the strong data, as expected, and they are all negative: increasing the intensities of strong data will reduce the value of the extinction parameter.

Rather than analysing the influence of data on a single parameter it may be of more interest, or simply less time consuming, to study groups of parameters. If only one parameter is being refined the leverage and T^2 values for the parameter in question amount to the same thing; this implies that one method for analysing a group of parameters is to study leverages from a refinement in which only those parameters are allowed to vary. This technique was used by Merli and co-workers in their work on minerals (*e.g.* Merli *et al.*, 2000). An alternative approach, which avoids the need to carry out multiple refinements, is to sum the T^2 values for groups of parameters. Fig. 2(*e*), which shows sums of T^2 values for the fractional coordinates in a refinement of alanine against simulated data, displays a marked drop-off in values above $\sin\theta/\lambda = 0.6$ Å^{-1}. This result can be contrasted with that described in Merli *et al.*'s (2000) leverage analysis of the silicate mineral pyrope. Here, high-resolution data were found to be important in determining the precision of oxygen positional parameters. This result was reflected in the importance of high-resolution data that had been noted anecdotally in Merli's laboratory in systematic work with garnets (Merli *et al.*, 2000).

In alanine, data above $\sin\theta/\lambda = 0.6$ Å^{-1} are most influential for the ADPs (Fig. 2*f*).

3.5. T^2 analysis of the Flack parameter in alanine

The Flack parameter is refined for noncentrosymmetric crystal structures in order to establish the absolute structure (Flack, 1983). The most important practical application of absolute structure refinement is in the determination of the absolute configuration of chiral compounds. The ability to distinguish one absolute structure from its inverted analogue depends on the resonant (or anomalous) scattering effects having sufficient magnitude to lead to measurably different intensities for Friedel pairs, something that depends on the elements present in the crystal and the wavelength of the X-rays used to collect intensity data.

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). 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 organic compounds such as alanine. The likely success of an absolute structure determination can be gauged using the Friedif parameter (Flack & Bernardinelli, 2008; Flack & Shmueli, 2007). If Friedif has a value of about 80, absolute structure determination should present little problem. The value of Friedif for alanine is only 33.9. Accordingly, the value of the Flack parameter obtained from the refinement of alanine was 0.00 (13). The data set was of excellent quality, yet the precision of the Flack parameter is (just) too large to enable a definitive statement to be made regarding the absolute structure (Flack & Bernardinelli, 2000).

Fig. 5 shows the results of a T^2 analysis for the Flack parameter in alanine. T^2 values for reflections that form Bijvoet pairs are strongly correlated, as expected (Fig. 5*a*). Values of $|T|$ are also closely correlated with

$$\frac{||F_c(\mathbf{h})|^2 - |F_c(\bar{\mathbf{h}})|^2|}{\left\{\sigma^2\left[|F_o(\mathbf{h})|^2\right] + \sigma^2\left[|F_o(\bar{\mathbf{h}})|^2\right]\right\}^{1/2}},$$

the calculated Bijvoet difference divided by its uncertainty as derived from those of the experimental observations (Fig. 5*b*). The most influential reflections are those with weak-to-

moderate intensities, 10–15% of $|F|_{max}$ (Fig. 5*c*). It is also notable that there are only a few (about 15) data that strongly affect the precision of the Flack parameter: most data have rather little effect.

Fig. 5(*d*) shows the distribution of T^2 values as a function of $\sin\theta/\lambda$. The most influential data lie at $\sin\theta/\lambda \simeq 0.4$–$0.5$ Å^{-1}, but the trend seems to drop off towards higher resolution. Similar features are seen for the other light-atom structures. Nonresonant X-ray scattering factors decrease with $\sin\theta/\lambda$, whereas the resonant corrections (f' and f'') are constant, and so the relative contribution of resonant scattering effects increases with resolution. Influential observations are expected to lie amongst the high-resolution data.

The increasing contribution of the resonant scattering factors at high resolution has led to the suggestion that collecting very high resolution data should enable precise absolute structure determination even for light-atom structures. However, in order to obtain such data it is necessary to use short-wavelength radiation for which resonant scattering effects are very small. Data for alanine were simulated to a resolution of 0.4 Å using scattering factors for Mo $K\alpha$ radiation. The structure of alanine was refined (along with the Flack

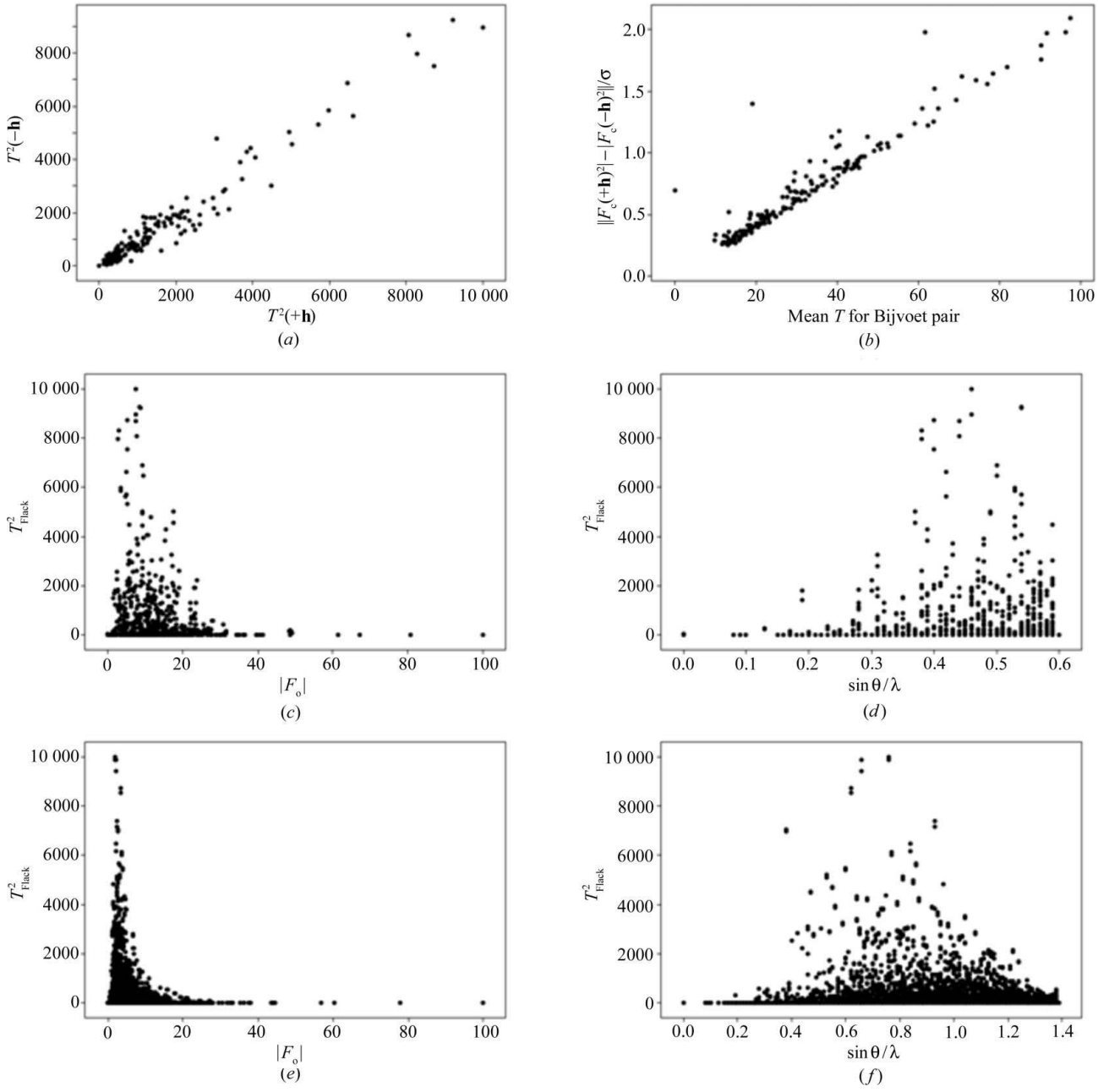


Figure 5
T analysis for the Flack parameter in alanine; (*a*)–(*d*) were generated using experimental Cu $K\alpha$ data, and (*e*) and (*f*) were generated from simulated data calculated to very high resolution for Mo $K\alpha$ radiation. (*a*) The relationship between T^2 values for Friedel pairs. (*b*) The variation of $|T|$ with the calculated Bijvoet ratio/σ [this quantity was calculated in *PLATON* (Spek, 2003)]. (*c*), (*d*) T^2 as a function of $|F_o|$ and $\sin\theta/\lambda$, respectively, for the experimental data sets; (*e*)–(*f*) the same quantities plotted for the simulated data.

parameter) against this data set. The T^2 *versus* $|F_o|$ plots for the experimental data (Fig. 5*c*) and the simulated data (Fig. 5*e*) show the same trend for moderate values of $|F_o|$ being the most influential, though the distribution in Fig. 5(*e*) is sharper. Fig. 5(*f*) shows the values of T^2 for the Flack parameter in this refinement plotted as a function of $\sin\theta/\lambda$. While there is a general increase in the T^2 values with $\sin\theta/\lambda$, the distribution is peaked in the middle of the resolution range, indicating that very high resolution data do not dominate the precision of the Flack parameter.

The reasons for expecting high-resolution data to be influential in determining the precision of the Flack parameter were outlined above, and it is perhaps surprising that there is a fall-off in T^2 values at the highest resolution in Figs. 5(*d*) and 5(*f*). However, Fig. 5(*b*) shows that an important factor in determining the influence that a particular Bijvoet pair has on the Flack parameter is how high the intensity difference is relative to its measurement standard uncertainty. It seems that the influence of reflections on the Flack parameter is the result of a balance between the increased contribution of the resonant scattering factors and the overall reduction in the signal-to-noise ratio of the intensities, which both occur as $\sin\theta/\lambda$ increases. At high resolution data will be weak and the Bijvoet ratios small relative to the measurement uncertainties, leading to a reduced influence on the Flack parameter. The fall-off can also be associated with the trends shown in Figs. 2(*e*) and 2(*f*), which show, respectively, the sums of T^2 values for the positional parameters and the non-H ADPs. The low-angle data most strongly influence the positional parameters, while the highest T^2 values for the ADPs are seen for the high-angle data. The largest Flack parameter T^2 values are seen between these two regions. The leverage of very high resolution data is 'spent' on defining the displacement parameters rather than the Flack parameter.

3.6. Use of *T* values in a weighting scheme

There is a long-standing interest in finding ways to improve the precision of the Flack parameter in light-atom structures. In the past, when four-circle instruments with point detectors were in use, a selected set of data with the highest Bijvoet ratios could be measured to a desired precision and statistical tests performed on the intensities to assess absolute structure (Le Page *et al.*, 1990). More recently, a post-refinement statistical procedure has been described by Hooft *et al.* (2010, 2008), while a method that can be used during refinement, based on combining Bijvoet intensity measurements and applying them as restraints, has been described by Parsons *et al.* (2010). It has also been shown that precision may be improved by the use of aspherical scattering factors (Dittrich *et al.*, 2006).

A method explored by Bernardinelli & Flack (1985) showed that precision can also be improved by modifying the refinement weights, up-weighting reflections calculated to be sensitive to the value of the Flack parameter. By this procedure the standard uncertainty of the Flack parameter could be reduced to an arbitrarily small value, but at the cost of causing the value of the parameter itself to deviate from its true value.

Information on the sensitivity of parameters to specific data is, of course, available from a leverage analysis in the form of the T and T^2 values, and the potential for improving the precision of the Flack parameter by incorporating these into the refinement weights was explored.

After some experimentation the following procedure for reweighting was used. The value of $\tau = 0.5\{\max[a|T(\mathbf{h})|^b, c] + \max[a|T(-\mathbf{h})|^b, c]\}$ was evaluated for each reflection with $a = 0.1$, $b = 1.0$ and $c = 1.0$. The overall mean τ, $\langle\tau\rangle$, was also determined. The reflection weights (w) were then modified (w') according to $w' = [\tau/(\langle\tau\rangle S)]^2 w$, where S is the goodness of fit obtained in the refinement with the original weights w. Larger values of a and b correspond to stronger up-weighting of sensitive data, though the placing of T values on a relative scale with $T_{max} = 100$ also implies a greater up-weighting in cases where resonant effects are weak.

For the alanine data set a Flack parameter of 0.00 (13) was obtained using F^2 refinement with weights equal to $1/\sigma^2(|F_o|^2)$ multiplied by a robust modifier as described by Prince & Nicholson (1983). The value of the Flack parameter obtained on reweighting with $a = 0.1$, $b = 1.0$ and $c = 1.0$ was -0.02 (5). Reweighting using the parameters $a = 0.5$, $b = 1.0$ and $c = 0.5$ yielded $x = -0.02$ (7).

Reweighting modestly increased the value of the unweighted R factor based on $|F|$ and all data by 0.2%. A normal probability plot based on $w^{1/2}(|F_o|^2 - |F_c|^2)$ had a gradient and correlation coefficient near unity and an intercept near 0; analyses of variance based on resolution or intensity were flat.

Hooft *et al.* (2010, 2008) have emphasized the value of normal probability plots (Abrahams & Keve, 1971) based on weighted Bijvoet differences in absolute structure refinement, and these proved to be a much more sensitive procedure for validating the weighting scheme. While the central region of the plot showed the expected behaviour, there was deviation from linearity at the extremes (Fig. 6*a*), suggesting that some data had been over-weighted. Over-weighting could be corrected using a second program, *REWEIGHT*, which fits a straight line to the central region of the normal probability plot and uses the equation of this line to define a factor to down-weight the deviating data points (Fig. 6*b*). The normal probability plot based on $w^{1/2}(|F_o|^2 - |F_c|^2)$ was still linear after this procedure (Fig. 6*c*). The value of the Flack parameter was -0.02 (6).

The procedure described above was tested on a number of other absolute structure refinements, and the results are listed in Table 1. All data sets were collected with high redundancy using Cu $K\alpha$ radiation at 100 K. All are 'difficult cases' for absolute structure refinement, all except one having Friedif parameters of 34 or less. One conclusion to be drawn from Table 1 is that robust-resistant $1/\sigma^2$ weights can be very effective for absolute structure refinements. However, precision was improved by application of the T-scaled weighting scheme, which yielded Flack parameters in most cases with standard uncertainties of around 0.1 or less. In the majority of

cases the Flack parameter itself moved closer to zero, with a value within one standard deviation of zero. In all cases the normal probability plots based on $w^{1/2}(|F_o|^2 - |F_c|^2)$ or Bijvoet differences were linear, while analyses of variance based on intensity, resolution and parity group were flat.

A particularly encouraging result was obtained for entry 17 in Table 1. These data refer to cholestane, a hydrocarbon with a Friedif parameter of only 9. Refinement of the Flack parameter using unmodified weights yielded a value of 0.36 (45), clearly an uninterpretable result. Reweighting yielded a Flack parameter of 0.10 (14); increasing the influence of sensitive data still further using $a = 0.2$ (and $b = c = 1.0$ as before) yielded a value of 0.10 (11).

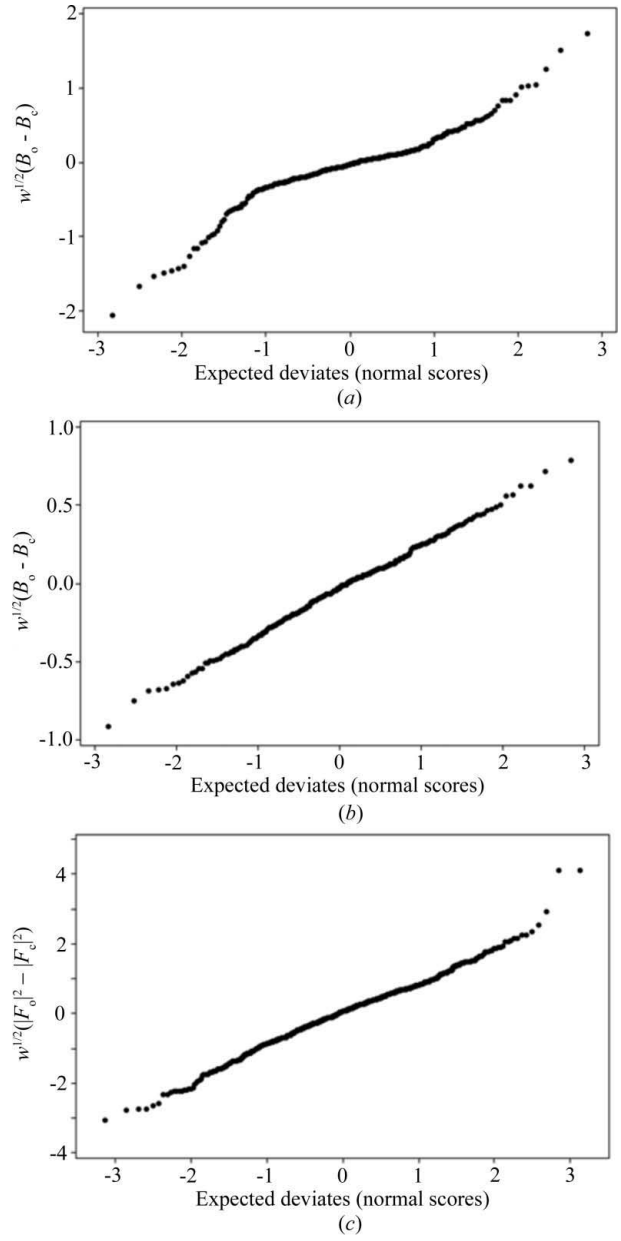


Figure 6
Normal probability plots relating to absolute structure refinement for alanine using experimental data. (*a*), (*b*) Before and after linearization of the *T*-weighted normal probability plot based on observed and calculated Bijvoet differences ($B_o - B_c$). (*c*) Normal probability plot based on $w^{1/2}(|F_o|^2 - |F_c|^2)$.

One disadvantage of the reweighting procedure is that it can amplify noise in the data, and Bijvoet normal probability plots were useful for detecting outliers. Outliers can cause the Flack parameter to deviate from its true value: in example 15, deletion of just two outliers changed x from 0.35 (12) to 0.02 (14). In cases such as this one we recommend, in preference to selective deletion of data, that the whole experiment be repeated.

The down-weighting procedure based on linearization of the weighted Bijvoet difference normal probability plot to some extent reduces the sensitivity of the results to the values of the parameters *a*, *b* and *c* defined above. We note in passing that in all cases the weighted Bijvoet difference normal probability plots had gradients much less than unity, spanning the range 0.28–0.75. Hooft *et al.* (2010) have also noted this feature, pointing out that it implies that the values of the Bijvoet difference uncertainties used to calculate the plots are overestimated. The variances of Bijvoet differences are calculated as $\{\sigma^2[|F_o(\mathbf{h})|^2] + \sigma^2[|F_o(-\mathbf{h})|^2]\}^{1/2}$, but this neglects a further covariance term equal to $-2\mathrm{cov}[|F_o(\mathbf{h})|^2, |F_o(-\mathbf{h})|^2]$. The small numerical values of the probability plot gradients suggests that the errors in $|F_o(\mathbf{h})|^2$ and $|F_o(-\mathbf{h})|^2$ are positively correlated. The correlation between errors suggests that it may be appropriate to include off-diagonal weights in absolute structure refinements. However, we are grateful to Professor Howard Flack for pointing out that the 'AD refinement' method of Flack *et al.* (2011) is equivalent to inclusion of these off-diagonal weighting terms, and when tested, this did not lead to substantial changes in either the Flack parameter or its standard deviation

The procedure described here alters the relative weights of observations in such a way as to improve the precision of a selected parameter. In an absolute structure determination the aim of the experiment is to obtain a precise value of the Flack parameter; our weighting scheme effectively refocuses the information present in the data in line with the aim of the experiment. The precision of other parameters may be decreased in a similar way. As an illustrative example, data sensitive to the *x*, *y* and *z* fractional coordinates of one of the ammonium H atoms in alanine were up-weighted (using $a = b = c = 1.0$). Prior to reweighting the coordinates were 0.4601 (16), 0.4089 (13) and 0.6476 (7); after reweighting they were 0.4608 (10), 0.4090 (8) and 0.6475 (5). The N—H bond distance changed from 0.907 (9) to 0.912 (6) Å.

In principle the precision of other parameters should decrease as a result of reweighting. The effect is small in our absolute structure tests because the number of data being up-weighted is also quite small (there are only a few really sensitive data). For the structures in Table 1 the maximum change in position was 0.004 Å and the maximum change in U_{ij} was 0.002 Å^2, these values being similar to the standard uncertainties in C—C bond distances and U_{ij} values in the structures concerned. In another test (using the data set collected for alanine) data sensitive to the scale factor were up-weighted using parameters $a = b = c = 1.0$. The scale factor

Table 1
The effect of incorporating T into refinement weights in absolute structure refinements of some light-atom structures.

Listed are values of unweighted R factors calculated on F and all data, the gradient, intercept and correlation coefficient of normal probability plots, and the value of the Flack parameter. The first and second lines refer to the refinements without and with T weighting; for the normal probability plot data the values before and after the '/' refer to plots based on $w^{1/2}(F_o^2 - F_c^2)$ and $w^{1/2}(B_o^2 - B_c^2)$, where B is the Bijvoet difference. Structures 2, 3, 5, 8, 11, 12 and 17 are monoclinic ($P2_1$); the remainder are orthorhombic ($P2_12_12_1$). For entries 5 and 15, three and two Bijvoet pairs were omitted, respectively.

					Normal probability plot			
No.	Formula	Friedif	Redundancy	$R1$(all data) (%)	Gradient	Intercept	Correlation coefficient	Flack parameter
1	$C_3H_7NO_2$	34	14.9	1.61	0.922	0.02	0.996	0.00 (13)
				1.71	0.936/0.339	0.01/−0.05	0.994/0.999	−0.02 (6)
2	$C_9H_{15}F_2NO_2$	53	5.7	2.18	0.934	0.01	0.998	0.01 (7)
				2.27	0.942/0.386	0.03/0.02	0.999/0.998	0.00 (4)
3	$C_{13}H_{17}NO_5$	35	5.7	2.55	0.941	0.04	0.998	−0.06 (10)
				2.67	0.945/0.493	0.05/0.00	0.999/0.999	0.00 (5)
4	$C_5H_8N_2O_2$	33	28.5	1.83	0.911	0.04	0.989	0.01 (10)
				1.89	0.938/0.281	0.02/0.02	0.998/0.997	0.01 (5)
5	$C_{13}H_{19}N_3O_4$	32	7.8	2.25	0.915	0.00	0.996	0.10 (9)
				2.33	0.944/0.428	−0.02/−0.03	0.999/0.998	0.06 (4)
6	$C_{25}H_{31}NO_5$	32	11.5	2.30	0.942	0.05	0.997	0.02 (8)
				2.41	0.946/0.405	0.05/0.01	1.000/0.999	0.01 (4)
7	$C_{35}H_{30}N_2O_5$	29	10.2	4.40	0.942	0.11	0.996	−0.04 (12)
				4.53	0.944/0.587	0.10/0.02	0.999/0.999	−0.01 (5)
8	$C_{29}H_{38}N_3O_4$	28	5.6	2.84	0.945	0.05	0.998	−0.05 (6)
				2.95	0.946/0.457	0.05/−0.01	0.998/0.999	0.00 (3)
9	$C_{60}H_{78}N_6O_8$	28	5.8	3.28	0.960	0.05	0.996	0.08 (8)
				3.25	0.947/0.579	0.05/−0.03	1.000/0.999	0.06 (4)
10	$C_{20}H_{21}NO_2$	26	11.5	2.09	0.925	0.04	0.993	−0.04 (8)
				2.16	0.944/0.281	0.03/0.00	0.999/0.999	−0.01 (3)
11	$C_{20}H_{21}NO_2$	26	11.4	2.15	0.934	0.05	0.994	−0.03 (8)
				2.21	0.944/0.229	0.03/0.00	0.999/0.999	−0.01 (4)
12	$C_{45}H_{60}O_3$	23	5.9	3.06	0.936	−0.01	0.996	−0.10 (11)
				3.14	0.942/0.748	−0.01/0.48	0.999/0.989	−0.08 (5)
13	$C_{20}H_{21}N_4O$	21	11.7	2.05	0.933	0.02	0.994	-0.01(11)
				2.14	0.945/0.331	0.03/0.02	0.999/0.998	0.02(5)
14	$C_{21}H_{22}N_2$ #1	12	11.5	2.79	0.941	0.00	0.998	−0.08 (31)
				2.85	0.946/0.490	−0.01/−0.12	0.999/0.998	0.08 (12)
15	$C_{21}H_{22}N_2$ #2	12	10.9	2.91	0.941	0.03	0.997	0.01 (31)
				3.08	0.946/0.447	0.03/−0.04	1.000/0.998	0.02 (14)
16	$C_{21}H_{22}N_2$ #3	12	11.7	2.05	0.912	0.04	0.994	0.00 (19)
				2.11	0.943/0.239	0.04/0.00	0.999/0.996	−0.04 (8)
17	$C_{27}H_{48}$	9	19.8	4.12	0.941	0.04	0.997	0.36 (45)
				4.23	0.949/0.474	0.04/−0.13	0.999/0.994	0.10 (14)

changed from 4.91 (11) to 4.93 (6). The precision of the extinction parameter also improved [20 (4) to 23.0 (9)], reflecting the fact that strong low-resolution data are important for both parameters. The precision of the displacement parameters, which are most sensitive to high-resolution data (see above), decreased slightly, with the average standard uncertainty changing from 0.0028 to 0.0031 Å^2.

4. Conclusions

Leverage analysis can be based either on the values of the leverages themselves, which give information on overall data fitting, or on T values, which enable the influence of observations with respect to specific parameters or groups of parameters to be investigated. Use of leverage analysis in crystallography is still quite rare, and the aim of this paper was to describe how it might prove useful in routine structure analysis.

Application of leverage analysis to outlier detection has been described previously by Merli (2005). Merli and co-workers have also shown that it can be used to rationalize the sensitivities of different mineral structures to the quality of high-resolution data, and to inform or justify refinement strategies of mixed site occupancies (Merli *et al.*, 2000). The role of different classes of data in a refinement has been described by David *et al.* (1993). The identification of refinements where weak data are important was described here.

A further application of the technique is in determining the effectiveness of restraints: a restraint with almost zero leverage might as well be removed or up-weighted. Equally, leverages are useful in deciding whether a parameter is determined solely by the restraints that have been applied or whether the intensity data retain some influence.

These ideas were illustrated using restrained refinements of alanine and Cu(sal)$_2$. In alanine the restraints were applied to H-atom positional and displacement parameters. Restraints placed on C—H and N—H bond distances were found to be more important than restraints placed on the angles involving H atoms. The leverages of the distance restraints were nevertheless only a little higher than average, and the intensity data were still important. The contrary was true in the Cu(sal)$_2$ refinement. In this case the H-atom parameters were effec-

tively determined by the restraints that had been applied. Of the restraints applied to the C-, N- and O-atom ADPs the rigid-bond restraints were very influential, but the rigid-body restraints had hardly any effect at all.

Another application was illustrated in T^2 analysis applied to the Flack parameter in alanine. It has been suggested that a strategy for precise absolute structure determination for light-atom crystal structures is to collect very high resolution data with Mo $K\alpha$ radiation. However, leverage analysis shows that the influence on the Flack parameter peaks at around $\sin\theta/\lambda = 0.6$ Å^{-1} and begins to decline at higher resolution. It was suggested that this trend is related to the observability of statistically significant Bijvoet intensity differences amongst weak high-resolution data.

The final application of leverages described here was in using T values as refinement weight modifiers to increase the precision of a parameter of interest. The parameter chosen was the Flack parameter in light-atom absolute structure refinements. The results obtained using T weighting are promising: not only are values of the Flack parameter more precise, they are also more accurate than values obtained in conventionally weighted refinements, clustering more closely around zero.

The method could, in principle, be applied to any parameter without the need to develop a physical model for identifying the most sensitive data, though we have not investigated this in detail, and careful testing would be required. In this work, it proved very important to examine refinement statistics critically, particularly so when resonant scattering effects are weak as the results are determined by up-weighting of a small number of data. Nevertheless, it does seem that given data of sufficient quality and high redundancy, reweighting based on leverage analysis might be employed to improve the precision of light-atom absolute structure determinations.

5. Programs

Windows executables for the programs *HATTIE* and *REWEIGHT* can be downloaded from the web site http://www.crystal.chem.ed.ac.uk/resource/. The programs are intended to be used in conjunction with *CRYSTALS*, which is available from http://www.xtl.ox.ac.uk/category/crystals.html.

We are grateful to Dr Martin Lutz (University of Utrecht) and Professor Howard Flack (University of Geneva) for their comments on the manuscript. We also thank Professor William David (ISIS and University of Oxford) for insightful comments made following a presentation of the results described in this paper, and an anonymous referee who read the manuscript with great care and diligence. We also thank Diamond Light Source for access to beamline I19 (proposal No. MT1200) and EPSRC (grant No. EP/G015333/1) for funding that contributed to the results on $Cu(sal)_2$ presented here.

References

Abrahams, S. C. & Keve, E. T. (1971). *Acta Cryst.* A**27**, 157–165.
Bernardinelli, G. & Flack, H. D. (1985). *Acta Cryst.* A**41**, 500–511.
Betteridge, P. W., Carruthers, J. R., Cooper, R. I., Prout, K. & Watkin, D. J. (2003). *J. Appl. Cryst.* **36**, 1487.
Bruker–Nonius (2006). *SAINT*. Bruker AXS Inc., Madison, Wisconsin, USA.
Blessing, R. H. (1997). *J. Appl. Cryst.* **30**, 421–426.
Byrne, P. J., Chang, J., Allan, D. R., Tasker, P. A. & Parsons, S. (2011). Unpublished results.
Clegg, W., Coles, S. J., Fallis, I. A., Griffiths, P. M. & Teat, S. J. (1998). *Acta Cryst.* C**54**, 882–885.
David, W. I. F. (2004). *J. Res. Natl Inst. Stand. Technol.* **109**, 107–123.
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.* A**62**, 217–223.
Dunitz, J. D. (1995). *X-ray Analysis and Structure of Organic Molecules*, 2nd ed. New York: VCH Publishers.
Flack, H. D. (1983). *Acta Cryst.* A**39**, 876–881.
Flack, H. D. & Bernardinelli, G. (2000). *J. Appl. Cryst.* **33**, 1143–1148.
Flack, H. D. & Bernardinelli, G. (2008). *Acta Cryst.* A**64**, 484–493.
Flack, H. D., Sadki, M., Thompson, A. L. & Watkin, D. J. (2011). *Acta Cryst.* A**67**, 21–34.
Flack, H. D. & Shmueli, U. (2007). *Acta Cryst.* A**63**, 257–265.
Hazen, R. M. & Finger, L. W. (1989). *Am. Mineral.* **74**, 352–359.
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.
Kassner, D., Baur, W. H., Joswig, W., Eichhorn, K., Wendschuh-Josties, M. & Kupčik, V. (1993). *Acta Cryst.* B**49**, 646–654.
Le Page, Y., Gabe, E. J. & Gainsford, G. J. (1990). *J. Appl. Cryst.* **23**, 406–411.
Marsh, R. E. (1981). *Acta Cryst.* B**37**, 1985–1988.
Merli, M. (2005). *Acta Cryst.* A**61**, 471–477.
Merli, M., Camara, F., Domeneghetti, C. & Tazzoli, V. (2002). *Eur. J. Mineral.* **14**, 773–784.
Merli, M., Oberti, R., Caucia, F. & Ungaretti, L. (2001). *Am. Mineral.* **86**, 55–65.
Merli, M. & Sciascia, L. (2011). *Acta Cryst.* A**67**, 456–468.
Merli, M., Sciascia, L. & Turco Liveri, M. L. (2010). *Int. J. Chem. Kinet.* **42**, 587–607.
Merli, M., Ungaretti, L. & Oberti, R. (2000). *Am. Mineral.* **85**, 532–542.
Merrill, L. & Bassett, W. A. (1974). *Rev. Sci. Instrum.* **45**, 290–294.
Moggach, S. A., Allan, D. R., Parsons, S. & Warren, J. E. (2008). *J. Appl. Cryst.* **41**, 249–251.
Oxford Diffraction (2010). *CrysAlis Pro*. Version 1.171.33.55. Oxford Diffraction Ltd, Abingdon, Oxfordshire, UK.
Parsons, S. (2004). *ECLIPSE*. The University of Edinburgh, UK.
Parsons, S., Flack, H. D., Presly, O. & Wagner, T. (2010). American Crysallographic Association Conference, 24–29 July 2010, Chicago, USA.
Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (1992). *Numerical Recipes in Fortran*, 2nd ed. Cambridge University Press.
Prince, E. (2004). *Mathematical Techniques in Crystallography and Materials Science*, 2nd ed. Berlin: Springer.
Prince, E. & Nicholson, W. L. (1983). *Acta Cryst.* A**39**, 407–410.
Prince, E. & Nicholson, W. L. (1985). *Struct. Stat. Crystallogr. Proc. Symp.* pp. 183–195.
Prince, E. & Spiegelman, C. H. (2004*a*). *International Tables for Crystallography*, Vol. C, pp. 702–706, edited by E. Prince. Dordrecht: Kluwer Academic Publishers.
Prince, E. & Spiegelman, C. H. (2004*b*). *International Tables for Crystallography*, Vol. C, pp. 707–709, edited by E. Prince. Dordrecht: Kluwer Academic Publishers.
Rawlings, J. O., Pantula, S. G. & Dickey, D. A. (1998). *Applied Regression Analysis: A Research Tool*, 2nd ed. New York: Springer.

Rodríguez-Carvajal, J. & González Platas, J. (2009). *CrysFML*. Institut Laue Langevin, Grenoble, France, and Universidad de La Laguna, La Launa, Spain.
Sheldrick, G. M. (2001). *XPREP*. University of Göttingen, Germany, and Bruker AXS Inc., Madison, Wisconsin, USA.
Sheldrick, G. M. (2008*a*). *Acta Cryst.* A**64**, 112–122.
Sheldrick, G. M. (2008*b*). *SADABS*. Version 2008-1. University of Göttingen, Germany, and Bruker AXS Inc., Madison, Wisconsin, USA.
Spek, A. L. (2003). *J. Appl. Cryst.* **36**, 7–13.