STRUCTURAL BIOLOGY
ISSN 2059-7983

From deep TLS validation to ensembles of atomic models built from elemental motions. II. Analysis of TLS refinement results by explicit interpretation

Pavel V. Afonine,[a,b]* Paul D. Adams[a,c] and Alexandre Urzhumtsev[d,e]*

[a]Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA, [b]Department of Physics and International Centre for Quantum and Molecular Structures, Shanghai University, Shanghai 200444, People's Republic of China, [c]Department of Bioengineering, University of California Berkeley, Berkeley, California, USA, [d]Centre for Integrative Biology, Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS–INSERM–UdS, 1 Rue Laurent Fries, BP 10142, 67404 Illkirch, France, and [e]Faculté des Sciences et Technologies, Université de Lorraine, BP 239, 54506 Vandoeuvre-les-Nancy, France. *Correspondence e-mail: pafonine@lbl.gov, sacha@igbmc.fr

Received 1 August 2017
Accepted 12 April 2018

Edited by R. J. Read, University of Cambridge, England

Keywords: TLS model; TLS refinement; atomic displacement parameters; rigid-body motion; ensemble of atomic models; atomic model validation; PDB.

TLS modelling was developed by Schomaker and Trueblood to describe atomic displacement parameters through concerted (rigid-body) harmonic motions of an atomic group [Schomaker & Trueblood (1968), *Acta Cryst.* B**24**, 63–76]. The results of a TLS refinement are **T**, **L** and **S** matrices that provide individual anisotropic atomic displacement parameters (ADPs) for all atoms belonging to the group. These ADPs can be calculated analytically using a formula that relates the elements of the TLS matrices to atomic parameters. Alternatively, ADPs can be obtained numerically from the parameters of concerted atomic motions corresponding to the TLS matrices. Both procedures are expected to produce the same ADP values and therefore can be used to assess the results of TLS refinement. Here, the implementation of this approach in *PHENIX* is described and several illustrations, including the use of all models from the PDB that have been subjected to TLS refinement, are provided.

1. Introduction

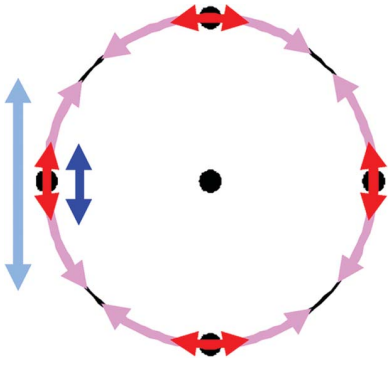
1.1. Atomic positions in crystal structures

Describing atomic positions in crystal structures by Cartesian coordinates is a mathematical abstraction. Atomic positions are averages over the diffraction data-collection time and over all of the unit cells in the crystal. The variation of positions may range from large, representing discrete conformations, to small, reflecting atomic motion around a central position.

If a motion is harmonic (in particular, this means that the motion amplitude is small), the probability of a shift of an atom n by a vector $\mathbf{r}_n = \Delta x_n\mathbf{i} + \Delta y_n\mathbf{j} + \Delta z_n\mathbf{k}$ is defined by individual isotropic (B_n) or anisotropic ($\mathbf{U}_n$) atomic displacement parameters (ADPs):

$$\mathbf{U}_n = \begin{pmatrix} \langle \Delta x_n^2 \rangle & \langle \Delta x_n \Delta y_n \rangle & \langle \Delta x_n \Delta z_n \rangle \\ \langle \Delta x_n \Delta y_n \rangle & \langle \Delta y_n^2 \rangle & \langle \Delta y_n \Delta z_n \rangle \\ \langle \Delta x_n \Delta z_n \rangle & \langle \Delta y_n \Delta z_n \rangle & \langle \Delta z_n^2 \rangle \end{pmatrix}. \quad (1)$$

These characteristics of atomic mobility are part of the structural information that is associated with models of crystal structures. As discussed in the literature (see, for example, Dunitz & White, 1973; Murshudov *et al.*, 1999; Winn *et al.*, 2001), the atomic displacement is a superposition of various motions that arise from different sources. These include displacement of atoms as part of a group and individual vibrations. A group motion itself can have several sources such



OPEN ACCESS

as motion of the whole molecule, motion of its domains, side-chain libration *etc*. Typically, modern refinement programs treat these motions using three separate components: motion of the whole crystal (modelled as an overall anisotropic scale factor), motion of non-overlapping groups that are considered to be rigid, and individual atomic motions.

1.2. Rigid-group motion

Atomic displacements arising from rigid-body motions can be accounted for using the TLS model (Schomaker & Trueblood, 1968). Such a model is based on simple geometric considerations allowing the description of elemental harmonic motions of atomic groups in terms of three matrices **T**, **L** and **S** (for a review, see Urzhumtsev *et al.*, 2013). This provides a convenient mathematical way to present these motions in terms of an individual anisotropic ADP,

$$\mathbf{U}_{\mathrm{TLS},n} = \mathbf{T} + \mathbf{A}_n\mathbf{L}\mathbf{A}_n^{\tau} + \mathbf{A}_n\mathbf{S} + \mathbf{S}^{\tau}\mathbf{A}_n^{\tau}, \qquad (2)$$

where

$$\mathbf{A}_n = \begin{pmatrix} 0 & z_n & -y_n \\ -z_n & 0 & x_n \\ y_n & -x_n & 0 \end{pmatrix} \qquad (3)$$

is an antisymmetric matrix expressed using the Cartesian coordinates (x_n, y_n, z_n) of atom n with respect to the origin of the TLS group. The symbol τ denotes the matrix transpose. The TLS approach may be seen as a statistical model for the analytical averaging of atomic positions that vary according to the given elemental motion parameters. The simplest example of a common motion is an isotropic vibration of a group that is equivalent to the assignment of the same B value to all atoms of the group. In the TLS model, the symmetric matrix **L** corresponds to the libration of a group, the symmetric matrix **T** corresponds to its common vibrations[1] (also including a correction for the position of the libration axes) and the matrix **S** reflects correlations between the motions as well as the position of the axes.

A set of TLS matrices is defined by 21 parameters (six for **T**, six for **L** and nine for **S**). There is a linear constraint on the diagonal elements of the **S** matrix (Schomaker & Trueblood, 1968), resulting in 20 independent parameters. If individual atomic displacements can be ignored and the assumption that atomic motions are purely rigid can be accepted (at low resolution, for example), then modelling atomic displacements using TLS can significantly reduce the overall number of fitting parameters. In the following, we refer to the parameterization of an atomic group motion using parameters of elemental rigid-body motions as direct parameterization and that using elements of the **T**, **L** and **S** matrices, such as in (2), as indirect parameterization.

Indirect parameterization is mathematically and computationally more straightforward compared with direct parameterization. This is because of the simple relationship between the refinable elements of the **T**, **L** and **S** matrices and the atomic displacement parameters **U** using (2). In contrast, direct parameterization requires a nontrivial number of mathematical steps that link the parameters of atomic motions (such as the amplitudes of vibration and libration *etc.*) to the elements of TLS matrices (see, for example, Urzhumtsev *et al.*, 2015). It is thus unsurprising that model-refinement programs such as *phenix.refine* (Afonine *et al.*, 2012) and *REFMAC* (Murshudov *et al.*, 1997) use indirect parameterization for TLS owing to its simplicity; that is, they refine the elements of TLS matrices and not the actual parameters of atomic motions. This approach is inherently problematic because unconstrained or unrestrained refinement of TLS matrices does not guarantee that the derived parameters of atomic motions are physically realistic or comply with TLS theory (see, for example, Zucker *et al.*, 2010; Merritt, 2012; Urzhumtsev *et al.*, 2015). This is very similar to unrestrained refinement of atomic coordinates at typical 'macromolecular resolutions' (*e.g.* 2–3 Å): factually, such refinement would almost certainly result in distorted stereochemistry.

1.3. Two possible interpretations of TLS models

One may think of at least two possible ways to interpret the results of TLS refinement. One interpretation considers TLS modelling to be successful if it leads to an improvement in the R factors and if the atomic displacement parameters $\mathbf{U}_{\mathrm{TLS},n}$ derived from the refined TLS matrices using (2) are realistic (for example, they vary smoothly between neighbouring atoms). A more conservative approach considers TLS modelling to be successful if, in addition to meaningful ADPs and improved model-to-data fit, the TLS parameters comply with the basic assumptions of the corresponding theory set out

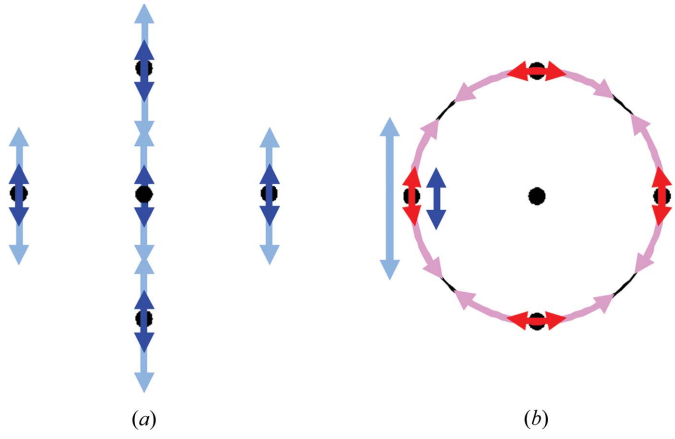


Figure 1
(*a*) A schematic representation of the atomic displacement for pure vibrations along the vertical axis (light and dark blue arrows) and (*b*) for libration around the axis perpendicular to the view (light and dark red arrows) shown for a five-atom dummy model (black dots). Lighter coloured arrows correspond to displacements with larger amplitudes. The displacements for vibration and libration are similar for small amplitudes and different for large amplitudes (*b*). The curvature of libration displacements with large amplitudes (*b*) makes them anharmonic.

[1] Traditionally, the term *translation* is used to denote the pure shift of the group without rotation. We prefer to use the term *vibration*, which is more coherent when using the term *libration*, and not *rotation*. This allows different components of the **T** matrix to be distinguished: those owing to the proper vibrations and those owing to apparent shifts.

by Schomaker & Trueblood (1968). This additional requirement is important when atomic motions modelled using TLS parameters are used to describe molecular motions (Trueblood, 1978; Trueblood & Dunitz, 1983; and references therein) or diffuse X-ray scattering data (Van Benschoten *et al.*, 2015), or are analysed for biological significance (Kuriyan & Weis, 1991; Harris *et al.*, 1992; Šali *et al.*, 1992; Wilson & Brunger, 2000; Raaijmakers *et al.*, 2001; Yousef *et al.*, 2002; Papiz *et al.*, 2003; Chaudhry *et al.*, 2004); see also discussion in Merritt (1999).

1.4. Analytical and numerical calculations of ADPs from TLS models

An interpretation of TLS refinement results in terms of elemental motions (see, for example, Howlin *et al.*, 1993; Urzhumtsev *et al.*, 2015, 2016) provides an opportunity to verify whether the corresponding motion parameters agree with TLS theory. This can be performed in two steps as follows. Firstly, the parameters of elemental group motions extracted from refined TLS matrices can be used to obtain an ensemble of models that samples these motions. In turn, (1) can be used to convert the ensemble back to a single model with the uncertainties in atomic positions described using the corresponding ADP values, $\mathbf{U}_{\rm ensemble,}_n$. Secondly, (2) can be used to calculate the uncertainties $\mathbf{U}_{\rm TLS,}_n$ in atomic positions directly from the TLS matrices. It is intuitive to expect that $\mathbf{U}_{\rm TLS,}_n$ and $\mathbf{U}_{\rm ensemble,}_n$ will match within some tolerance. The tolerance is needed to account for rounding errors and the finite number of models in the ensemble. A difference between $\mathbf{U}_{\rm TLS,}_n$ and $\mathbf{U}_{\rm ensemble,}_n$ beyond this tolerance may be indicative of various problems with the corresponding TLS set.

Since currently used refinement programs utilize an indirect TLS parameterization that does not use restraints or constraints, it may be the case that extracting motion parameters from refined TLS matrices is mathematically impossible (Urzhumtsev *et al.*, 2015, 2016). The simplest example is the **T** or **L** matrices being non-positive definite. A more subtle example is when the parameters of elemental motions can be extracted from the TLS matrices but may not satisfy the basic assumptions about the TLS model (for example, libration amplitudes being too large, resulting in atomic motions that are anharmonic; see Fig. 1 and §2).

When motion parameters can be extracted from TLS matrices, comparison of $\mathbf{U}_{\rm ensemble,}_n$ and $\mathbf{U}_{\rm TLS,}_n$ requires a measure of and a threshold for the tolerance mentioned above (discussed in §2.1). Since $\mathbf{U}_{\rm ensemble,}_n$ depends on the number of models in the ensemble, we use a simple test system to estimate how many models are required to sample the group motion accurately and also to estimate a possible threshold value for the similarity of respective matrices (§2.2). The results are then validated using a more realistic protein model (§2.3). These tests highlighted reasons for differences between $\mathbf{U}_{\rm TLS,}_n$ and $\mathbf{U}_{\rm ensemble,}_n$ matrices and prompted further improvements for TLS analysis (§2.4). In §3 we discuss the results of the application of our procedures to all models in the PDB (Bernstein *et al.*, 1977; Berman *et al.*, 2000) that contain TLS information. Respective tools have been added to the *PHENIX* suite (Adams *et al.*, 2010).

2. Anisotropic displacement matrices and the corresponding model ensembles

2.1. Metrics for matrix comparison

To evaluate the similarity of the two sets of anisotropic displacement matrices, $\mathbf{U}_{\rm TLS,}_n$ and $\mathbf{U}_{\rm ensemble,}_n$, for a group composed of N atoms, $n = 1, 2, \ldots, N$, we arbitrarily choose to use a simple R-factor-type metric,

$$R_{\mathbf{U}}(\mathbf{U}_1, \mathbf{U}_2) = 2\frac{\sum\limits_n \sum\limits_{i,j=1}^{3} \left|\mathbf{U}_{1,n}^{(i,j)} - \mathbf{U}_{2,n}^{(i,j)}\right|}{\sum\limits_n \sum\limits_{i,j=1}^{3} \left[\left|\mathbf{U}_{1,n}^{(i,j)}\right| + \left|\mathbf{U}_{2,n}^{(i,j)}\right|\right]}, \qquad (4)$$

where $\mathbf{U}_{1,n} = \mathbf{U}_{\rm TLS,}_n$ and $\mathbf{U}_{2,n} = \mathbf{U}_{\rm ensemble,}_n$. Here, the sums are calculated over all elements of the matrices and over all atoms of the group. Other metrics can also be used (Dunitz & White, 1973; Zucker *et al.*, 2010). Specifically, Kullback–Liebler (KL) divergence (Kullback & Leibler, 1951; Murshudov *et al.*, 2011; Merritt, 2011, 2012) and the correlation coefficient ($\mathrm{CC}_{\mathbf{UV}}$; Merritt, 1999) seem to be most prominent, with the caveat that they require matrix inversion, which is not always possible in numeric tests where only one single motion can be considered.

The calculation of (4) depends on the randomly generated ensemble models that are used to obtain $\mathbf{U}_{\rm ensemble,}_n$. This is a stochastic procedure that depends on random seed values and on the number of models in the ensemble. Below, we analyze how these parameters affect the estimate of $\mathbf{U}_{\rm ensemble}$. Also, we check whether using $\mathrm{KL}_{\mathbf{UV}}$ or $\mathrm{CC}_{\mathbf{UV}}$ leads to conclusions that differ from those obtained using $R_{\mathbf{U}}$.

2.2. Illustrations using a one-atom model

For simplicity, in this section we drop the subscript n from $\mathbf{U}_{\rm ensemble}$ and $\mathbf{U}_{\rm TLS}$ because only a single atom is considered.

2.2.1. Effect of vibration. In this test, we consider a model composed of a single atom vibrating along the **Ox** axis. For each trial root-mean-square deviation (r.m.s.d., which we call the vibration amplitude t), we generated M random copies of this atom and then took all of these copies to calculate $\mathbf{U}_{\rm ensemble}$ using (1). We then used (4) to compare $\mathbf{U}_{\rm ensemble}$ with the corresponding $\mathbf{U}_{\rm TLS} = \mathbf{T}$ calculated analytically using (2) with $\mathbf{L} = \mathbf{S} = 0$. For each trial t we repeated these calculations 100 times, each time with a different random seed. Obviously, for different trials $\mathbf{U}_{\rm TLS}$ remains the same while $\mathbf{U}_{\rm ensemble}$ varies. Fig. 2(*a*) shows the average (over 100 trials) $R_{\mathbf{U}}$ for different trial values of t and M. The results are essentially independent of t. This is expected since $\langle \Delta x^2 \rangle$ in (1) is proportional to $T_{xx} = t_x^2$ in (2) for a sufficiently large number of models. We observe that $R_{\mathbf{U}}$ becomes close to 0.01 once the size of the ensemble reaches about 10 000 models.

2.2.2. Effect of libration. Here, we used the same single-atom model as above and the same calculation workflow,

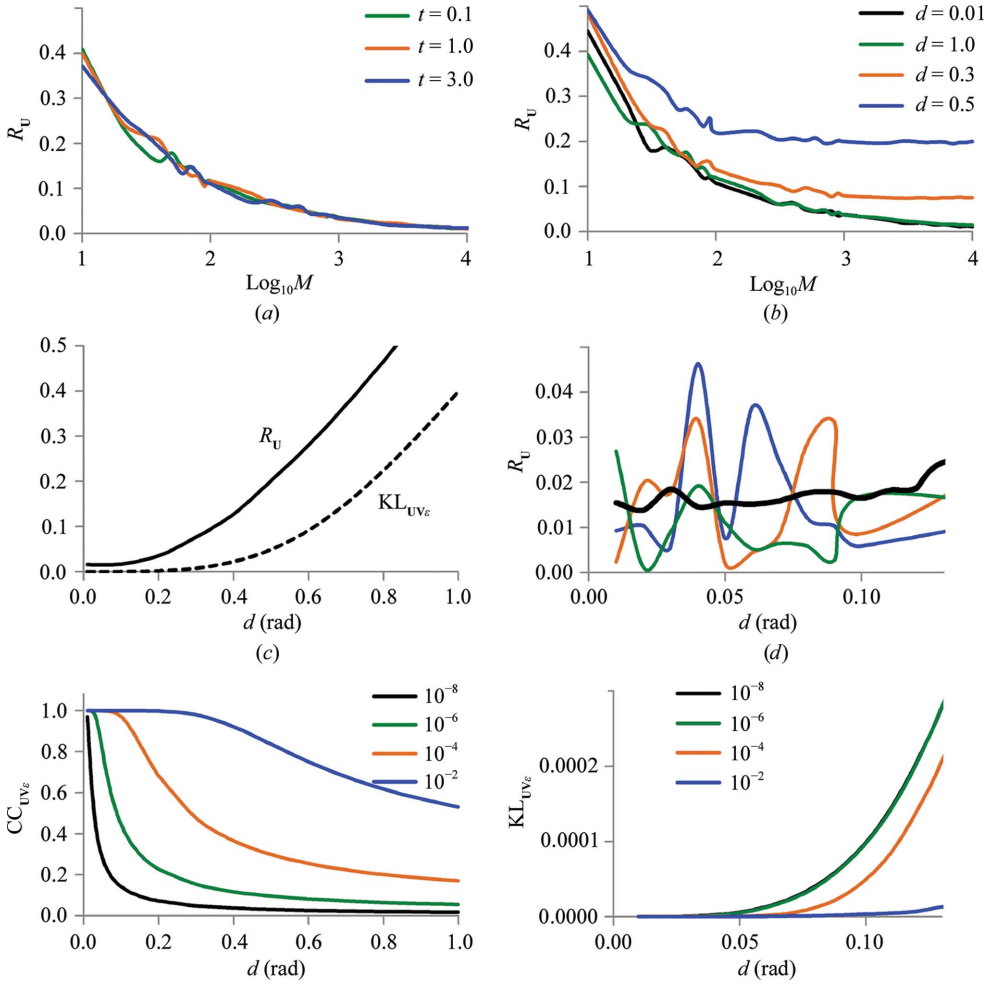


Figure 2
Agreement between the $\mathbf{U}_{\rm ensemble}$ and $\mathbf{U}_{\rm TLS}$ matrices calculated for a single-atom model. $R_{\mathbf{U}}$ (averaged over 100 random runs) is shown as a function of the logarithm of the number M of models for different (*a*) vibration and (*b*) libration r.m.s.d. values. (*c*) $R_{\mathbf{U}}$ (solid line) and $\mathrm{KL}_{\mathbf{UV}\varepsilon}$ (dashed line) with $\varepsilon = 10^{-6}$ as a function of the vibration r.m.s.d. value d for ensembles composed of 5000 generated models. (*d*) $R_{\mathbf{U}}$ as a function of the vibration r.m.s.d. value d zoomed on the d = 0.0–0.1 rad range and shown for the average (black curve) as well as for three individual runs (in maroon, blue and green) selected from the 100 runs used for averaging. (*e*) $\mathrm{CC}_{\mathbf{UV}\varepsilon}$ calculated for several ε values (10^{-2}, 10^{-4}, 10^{-6} and 10^{-8}). (*f*) $\mathrm{KL}_{\mathrm{UV}\varepsilon}$ calculated for the same ε values and for small d values; the curves for ε values of 10^{-6} and 10^{-8} are indistinguishable. See the text for details.

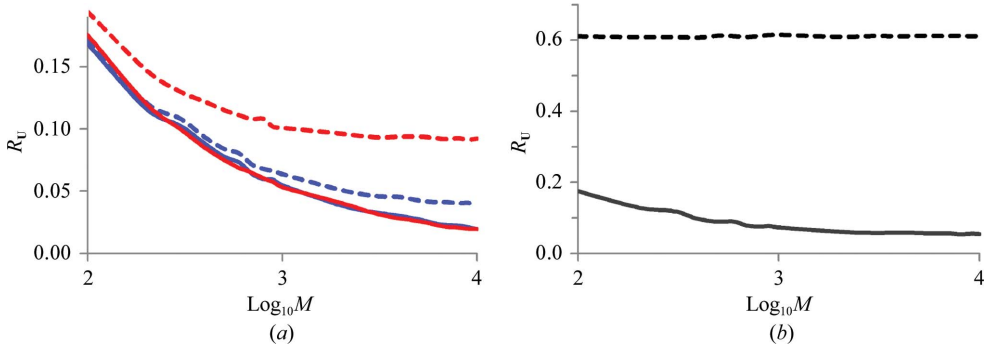


Figure 3
Agreement of $R_{\mathbf{U}}$ between $\mathbf{U}_{\rm ensemble}$ and $\mathbf{U}_{\rm TLS}$ matrices as a function of the number of generated models calculated for protein data. (*a*) Results for 2igd models composed of all main-chain atoms (red) and C^{α} atoms only (blue) using different approaches to extract the elemental motions: dashed lines for (10) and full lines for (11). (*b*) Results for the 4muy model using (10) shown as a dashed line and (11) shown as a full line.

except that we varied the libration r.m.s.d. value d. Similarly to the previous example, $R_{\mathbf{U}}$ as a function of ensemble size reaches a plateau for about 5000–10 000 models in the ensemble (Fig. 2*b*); however, the plateau level depends on the value of d. Then we sampled a broad range of d values keeping the ensemble size fixed at 5000 models (Figs. 2*c* and 2*d*). We observe that $R_{\mathbf{U}}$ remains approximately constant (∼0.02) up to a d_0 of ∼0.15 rad and then starts increasing monotonically. The d_0 value obtained in this numerical experiment corresponds to the limit of a linear approximation to the small rotations discussed in Urzhumtsev *et al.* (2013) and other works cited therein, for example Cruickshank (1956). Owing to rounding errors and the differences between a rotation motion and a linear motion, the $R_{\mathbf{U}}$ values never reached zero even for very small d and large ensemble sizes (Fig. 2*c*). Also, while the average values over several trials are stable, they may vary between individual trials (Fig. 2*d*). These results allowed us to draw two conclusions. Firstly, generating about 5000–10 000 models is sufficient to estimate $\mathbf{U}_{\rm ensemble}$ reliably (in §2.3 we show that this is still the case for realistic macromolecular models). Secondly, we may consider that $\mathbf{U}_{\rm ensemble}$ agrees with $\mathbf{U}_{\rm TLS}$ for a particular TLS set if $R_{\mathbf{U}}$ is approximately 0.05 or less.

2.2.3. Checking other metrics. To illustrate that the results obtained in previous tests are independent of matrix-comparison metrics, we repeated the test described in §2.2.2 using other metrics such as $\mathrm{KL}_{\mathbf{UV}}$ and $\mathrm{CC}_{\mathbf{UV}}$. Since these metrics require a matrix inversion, we had to use a minor modification consisting of adding a small value to all of the diagonal elements of the respective matrices $\mathbf{U}_{\rm ensemble}$ and the corresponding $\mathbf{U}_{\rm TLS}$, which is a

Table 1
Analysis of the discrepancy between $\mathbf{U}_{\mathrm{ensemble},n}$ and $\mathbf{U}_{\mathrm{TLS},n}$ using $R_{\mathbf{U}}$.

For PDB entry 2igd, the two TLS sets, referred to as $\mathrm{TLS_{CA}}$ and $\mathrm{TLS_{MC}}$, are derived from anisotropic ADPs of C^{α} atoms only or of main-chain atoms, respectively. For each of the sets the parameters of the elemental motions were determined using either (10) or (11) with the constraints described in Urzhumtsev *et al.* (2015). For both TLS sets the same model composed of C^{α} atoms only was used to generate $\mathbf{U}_{\mathrm{ensemble},n}$ and compare it with the respective $\mathbf{U}_{\mathrm{TLS},n}$. For the 4muy model all atoms are used both to determine the TLS matrices and to generate $\mathbf{U}_{\mathrm{ensemble},n}$; the elemental motions were determined using either (10) or (11). The $R_{\mathbf{U}}$(all) column shows the results of comparison when the whole set of motions (librations and vibrations) were used (5). The $R_{\mathbf{U}}$(no V) column indicates the case when only three librations were used while vibration components were excluded (6). The next three columns [$R_{\mathbf{U}}(d_x, s_x)$, $R_{\mathbf{U}}(d_y, s_y)$ and $R_{\mathbf{U}}(d_z, s_z)$] show the results for cases when only one single libration and a corresponding screw were used (7). The last three columns [$R_{\mathbf{U}}(d_x)$, $R_{\mathbf{U}}(d_y)$ and $R_{\mathbf{U}}(d_z)$] represent the pure librations (8).

TLS	Method	$R_{\mathbf{U}}$(all)	$R_{\mathbf{U}}$(no V)	$R_{\mathbf{U}}(d_x, s_x)$	$R_{\mathbf{U}}(d_y, s_y)$	$R_{\mathbf{U}}(d_z, s_z)$	$R_{\mathbf{U}}(d_x)$	$R_{\mathbf{U}}(d_y)$	$R_{\mathbf{U}}(d_z)$
PDB entry 2igd									
$\mathrm{TLS_{CA}}$	(10)	0.04	0.07	0.14	0.05	0.03	0.00	0.02	0.01
$\mathrm{TLS_{MC}}$	(10)	0.09	0.15	0.28	0.01	0.03	0.00	0.02	0.01
$\mathrm{TLS_{CA}}$	(11)	0.01	0.02	0.01	0.02	0.04	0.00	0.02	0.01
$\mathrm{TLS_{MC}}$	(11)	0.01	0.02	0.03	0.04	0.04	0.00	0.02	0.01
PDB entry 4muy									
$\mathrm{TLS_{all}}$	(10)	0.61	0.85	0.89	0.25	0.27	0.01	0.01	0.01
$\mathrm{TLS_{all}}$	(11)	0.05	0.11	0.02	0.27	0.42	0.01	0.02	0.00

convolution with an isotropic vibration. For example, the modified $\mathrm{KL}_{\mathbf{UV}}$ metric is $\mathrm{KL}_{\mathbf{UV}\varepsilon} = \varepsilon\,\mathrm{tr}(\mathbf{U}_{\varepsilon}\mathbf{V}_{\varepsilon}^{-1} + \mathbf{V}_{\varepsilon}\mathbf{U}_{\varepsilon}^{-1} - 2\mathbf{I})$ with $\mathbf{U}_{\varepsilon} = \mathbf{U} + \varepsilon\mathbf{I}$ and $\mathbf{V}_{\varepsilon} = \mathbf{V} + \varepsilon\mathbf{I}$. The scale factor ε before 'tr' is used to put the results on a similar scale (to facilitate comparisons). By trial and error, we found that a value of ε in the range 10^{-8}–10^{-6} allows the calculation of $\mathrm{KL}_{\mathbf{UV}\varepsilon}$ and $\mathrm{CC}_{\mathbf{UV}\varepsilon}$ but does not significantly affect the results. Overall, $\mathrm{KL}_{\mathbf{UV}\varepsilon}$, $\mathrm{CC}_{\mathbf{UV}\varepsilon}$ and $R_{\mathbf{U}}$ do not contradict each other (Fig. 2*c*), with $\mathrm{CC}_{\mathrm{UV}\varepsilon}$ showing a much stronger dependence on ε (Fig. 2*e*) and both $\mathrm{KL}_{\mathbf{UV}\varepsilon}$ and $\mathrm{CC}_{\mathbf{UV}\varepsilon}$ showing a less prominent drop (Figs. 2*e* and 2*f*) at a d_0 of ∼0.10–0.15 rad (see Cruickshank, 1956) compared with R_{U}. In the following we use $R_{\mathbf{U}}$ because the original matrices can be used without modification and it has a predictable range of values, unlike $\mathrm{KL}_{\mathbf{UV}\varepsilon}$.

2.3. Illustrations using a protein model

As a more realistic example, we selected the model of IgG-binding domain III (PDB entry 2igd; S. Butterworth, V. L. Lamzin, D. B. Wigley, J. P. Derrick & K. S. Wilson, unpublished work) refined at 1.1 Å resolution using individual anisotropic ADP values. We chose the core of this model as a single TLS group containing residues 6–61 (leaving out the flexible N-terminus).

We considered two models derived from these data. One model contained C^{α} atoms only (56 in total) and the other model included all main-chain atoms (C^{α}, O, C and N). Each of the two models was treated as a single TLS group. For each model we fitted TLS matrices to individual anisotropic $\mathbf{U}_n$ values (ANISOU records from the PDB file) using the *phenix.tls* tool; we refer to these matrices as $\mathrm{TLS_{CA}}$ and $\mathrm{TLS_{MC}}$, respectively. Then, using each of the two TLS sets ($\mathrm{TLS_{CA}}$ and $\mathrm{TLS_{MC}}$) we calculated $\mathbf{U}_{\mathrm{TLS},n}$ using (2) and generated $\mathbf{U}_{\mathrm{ensemble},n}$ as described in Urzhumtsev *et al.* (2015). Similarly to as described in §2.2, we sampled a range of different numbers of models per ensemble.

The blue dashed curve in Fig. 3(*a*) shows that for the C^{α}-only model ($\mathrm{TLS_{CA}}$) $R_{\mathbf{U}}$ becomes smaller than 0.05 when the ensemble contains about 5000 models; using a larger ensemble does not change $R_{\mathbf{U}}$ significantly. This agrees with the conclusions derived in §2.2. For the main-chain model $R_{\mathbf{U}}$ reaches a plateau for ensembles containing the same number of models, but the value of $R_{\mathbf{U}}$ does not decrease below 0.09 (red dashed curve in Fig. 3*a*). To investigate the source of such a significant difference in $R_{\mathbf{U}}$ we performed the following tests.

Firstly, we note that the only difference between the two models is their composition and TLS matrices (Fig. 4). To determine which of the two, the composition or TLS matrices, contributes to the large $R_{\mathbf{U}}$ value, we repeated the calculations above using the C^{α}-only model with $\mathrm{TLS_{MC}}$ matrices and the main-chain model with $\mathrm{TLS_{CA}}$ matrices. In the first case $R_{\mathbf{U}}$ was 0.09 and in the second case it was 0.05. This shows that the difference in $R_{\mathbf{U}}$ is owing to the TLS matrices and is not owing to the model composition. To find out which features of $\mathrm{TLS_{MC}}$ are responsible for the increased $R_{\mathbf{U}}$, we performed a further analysis.

The elemental motions encoded by TLS matrices are three screw librations (around three mutually orthogonal axes $\mathbf{l}_x$, $\mathbf{l}_y$, $\mathbf{l}_z$) coupled with three vibrations, also about three mutually orthogonal axes $\mathbf{v}_x$, $\mathbf{v}_y$, $\mathbf{v}_z$. In the following, d_x, d_y, d_z stand for libration amplitudes, t_x, t_y, t_z stand for vibration amplitudes, s_x, s_y, s_z stand for the corresponding screw parameters and $\mathbf{w}_x$, $\mathbf{w}_y$, $\mathbf{w}_z$ stand for the points that belong to the respective libration axes (for formal definitions, see Urzhumtsev *et al.*, 2013). As discussed in §2.2, the condition $\mathbf{U}_{\mathrm{ensemble},n} \simeq \mathbf{U}_{\mathrm{TLS},n}$ (for a sufficiently large number of models in the ensemble) may

$$\mathbf{T}_{\mathrm{CA}} = \begin{pmatrix} 0.0691 & 0.0048 & -0.0039 \\ 0.0048 & 0.0854 & -0.0005 \\ -0.0039 & -0.0005 & 0.1047 \end{pmatrix} \quad \mathbf{T}_{\mathrm{MC}} = \begin{pmatrix} 0.0695 & 0.0039 & -0.0041 \\ 0.0039 & 0.0842 & 0.0014 \\ -0.0041 & 0.0014 & 0.0992 \end{pmatrix}$$

$$\mathbf{L}_{\mathrm{CA}} = \begin{pmatrix} 2.1972 & 0.5440 & 0.0026 \\ 0.5440 & 0.5110 & 0.2089 \\ 0.0026 & 0.2089 & 1.3269 \end{pmatrix} \quad \mathbf{L}_{\mathrm{MC}} = \begin{pmatrix} 2.1725 & 0.4879 & -0.0881 \\ 0.4879 & 0.5881 & 0.2404 \\ -0.0881 & 0.2404 & 1.31562 \end{pmatrix}$$

$$\mathbf{S}_{\mathrm{CA}} = \begin{pmatrix} 0.0612 & 0.0383 & 0.0986 \\ -0.0977 & -0.0764 & -0.0418 \\ -0.0211 & -0.0272 & 0.0152 \end{pmatrix} \quad \mathbf{S}_{\mathrm{MC}} = \begin{pmatrix} 0.0512 & 0.0543 & 0.0992 \\ -0.0892 & -0.0659 & -0.0441 \\ -0.0042 & -0.0302 & 0.0147 \end{pmatrix}$$

Figure 4
The TLS matrices calculated for the 2igd model for all main-chain atoms (right) and for C^{α} atoms only (left). The matrices are given according to the PDB conventions: **T** is in $\mathrm{Å}^2$, **L** is in deg^2 and **S** is in Å deg.

Table 2
Components of the elemental motions.

The four upper blocks correspond to the TLS matrices for PDB entry 2igd calculated for C^{α} atoms only (TLS_{CA}) and for the main-chain atoms (TLS_{MC}). The TLS matrices were decomposed with (10) or (11) using the constraints described in Urzhumtsev *et al.* (2015). The two bottom blocks correspond to the model for PDB entry 4muy. The vectors $\mathbf{v}_x$, $\mathbf{v}_y$, $\mathbf{v}_z$ and $\mathbf{l}_x$, $\mathbf{l}_y$, $\mathbf{l}_z$ of the vibration and libration bases, respectively, are given in Cartesian coordinates in the principal basis [**M**] with the origin at the group centre of mass and with the axes parallel to the crystal axes. The points $\mathbf{w}_x$, $\mathbf{w}_y$, $\mathbf{w}_z$ (in Å) are given in the orthonormal basis [**L**] composed of the principal libration axes $\mathbf{l}_x$, $\mathbf{l}_y$, $\mathbf{l}_z$ and describe the shift of these axes from the origin. The libration amplitudes d_x, d_y, d_z are given in radians and the vibration amplitudes t_x, t_y, t_z and the screw components s_x, s_y, s_z are in Å. For details of the definitions, see Urzhumtsev *et al.* (2013).

TLS	t_x, t_y, t_z	$\mathbf{v}_x$, $\mathbf{v}_y$, $\mathbf{v}_z$	d_x, d_y, d_z	$\mathbf{l}_x$, $\mathbf{l}_y$, $\mathbf{l}_z$	$\mathbf{w}_x$, $\mathbf{w}_y$, $\mathbf{w}_z$	s_x, s_y, s_z
PDB entry 2igd						
TLS_{CA} (10)	0.163	(−0.085, 0.437, 0.896)	0.011	(−0.262, 0.915, −0.308)	(−12.67, −0.39, 16.71)	−2.07
	0.278	(0.905, 0.410, −0.114)	0.019	(−0.067, 0.301, 0.951)	(1.65, 0.97, 8.55)	−0.88
	0.304	(−0.417, 0.801, −0.430)	0.027	(0.963, 0.270, −0.017)	(−4.67, −3.47, 0.76)	0.80
TLS_{MC} (10)	0.089	(−0.082, 0.334, 0.939)	0.010	(−0.272, 0.943, −0.193)	(−14.16, −1.74, 22.42)	−5.70
	0.277	(0.948, 0.316, −0.030)	0.020	(−0.113, 0.168, 0.979)	(0.49, 0.11, 11.77)	−0.24
	0.314	(−0.306, 0.888, −0.343)	0.027	(0.956, 0.288, 0.061)	(−4.92, −3.54, −0.25)	0.89
TLS_{CA} (11)	0.163	(−0.085, 0.433, 0.897)	0.011	(−0.262, 0.915, −0.308)	(−12.67, −0.39, 16.71)	−0.09
	0.279	(0.902, 0.417, −0.116)	0.019	(−0.067, 0.301, 0.951)	(1.65, 0.97, 8.55)	−0.30
	0.305	(−0.424, 0.799, −0.426)	0.027	(0.963, 0.270, −0.017)	(−4.67, −3.47, 0.76)	1.12
TLS_{MC} (11)	0.083	(−0.078, 0.332, 0.940)	0.010	(−0.272, 0.943, −0.193)	(−14.16, −1.74, 22.42)	−0.43
	0.282	(0.931, 0.362, −0.051)	0.020	(−0.113, 0.168, 0.979)	(0.49, 0.11, 11.77)	0.97
	0.314	(−0.357, 0.871, −0.337)	0.027	(0.956, 0.288, 0.061)	(−4.92, −3.54, −0.25)	1.58
PDB entry 4muy						
TLS_{all} (10)	0.0	(0.951, 0.286, −0.117)	0.001	(0.649, 0.500, −0.573)	(−219.91, −11.67, −256.03)	303.63
	0.257	(−0.220, 0.893, 0.393)	0.008	(−0.633, 0.773, −0.042)	(−49.29, 57.65, −1.63)	2.90
	0.363	(0.216, −0.348, 0.912)	0.014	(0.421, 0.390, 0.819)	(−72.36, −52.48, −124.89)	−3.11
TLS_{all} (11)	0.241	(0.227, 0.947, 0,225)	0.001	(0.649, 0.500, −0.573)	(−219.91, −11.67, −256.03)	0.11
	0.321	(−0.582, −0.053, 0.811)	0.008	(−0.633, 0.773, −0.042)	(−49.29, 57.65, −1.63)	−3.38
	0.396	(0.780, −0.316, 0.540)	0.014	(0.421, 0.390, 0.819)	(−72.36, −52.48, −124.89)	−5.14

break down owing to inadequate librations and not owing to vibrations. To separate the contribution of vibrations and librations, we first derived the set of parameters of elemental motions

$$\mathrm{MP}_{\mathrm{all}} = (d_x, d_y, d_z; s_x, s_y, s_z; t_x, t_y, t_z) \quad (5)$$

from the corresponding TLS matrices as described in Urzhumtsev *et al.* (2015). Here, and in the following, to simplify the text we drop the parameters $\mathbf{l}_x$, $\mathbf{l}_y$, $\mathbf{l}_z$; $\mathbf{w}_x$, $\mathbf{w}_y$, $\mathbf{w}_z$; $\mathbf{v}_x$, $\mathbf{v}_y$, $\mathbf{v}_z$ from the list in (5) since they are invariant within these tests. Then, using the parameters in (5) (Table 2) and the C^{α}-only model, we calculated $\mathbf{U}_{\mathrm{ensemble},n}$ and $\mathbf{U}_{\mathrm{TLS},n}$ for the following different scenarios.

Firstly, we considered a scenario where all three librations are used together, including their screw components, while vibrations are excluded:

$$\mathrm{MP}_{\mathrm{no\ V}} = (d_x, d_y, d_z; s_x, s_y, s_z; 0, 0, 0). \quad (6)$$

Excluding vibrations led to an increase in $R_{\mathbf{U}}$ for both TLS_{MC} and TLS_{CA}; the values in the $R_{\mathbf{U}}$(no V) column in Table 1 are ∼1.5 times larger[2] than those for $R_{\mathbf{U}}$(all).

Secondly, we calculated $R_{\mathbf{U}}$ separately for each individual libration, including its corresponding screw component (Table 1, columns 5–7),

$$\begin{aligned}\mathrm{MP}_{d_x,s_x} &= (d_x, 0, 0; s_x, 0, 0; 0, 0, 0),\\ \mathrm{MP}_{d_y,s_y} &= (0, d_y, 0; 0, s_y, 0; 0, 0, 0),\\ \mathrm{MP}_{d_z,s_z} &= (0, 0, d_z; 0, 0, s_z; 0, 0, 0),\end{aligned} \quad (7)$$

and without it,

$$\begin{aligned}\mathrm{MP}_{d_x} &= (d_x, 0, 0; 0, 0, 0; 0, 0, 0),\\ \mathrm{MP}_{d_y} &= (0, d_y, 0; 0, 0, 0; 0, 0, 0),\\ \mathrm{MP}_{d_z} &= (0, 0, d_z; 0, 0, 0; 0, 0, 0).\end{aligned} \quad (8)$$

(Table 1, columns 8–10). For the screw librations, $R_{\mathbf{U}}$ is large for the rotation around $\mathbf{l}_x$ (Table 1), which is likely to be owing to a large magnitude of the screw component s_x. This component is two and a half times greater for TLS_{MC} compared with TLS_{CA} (−5.70 *versus* −2.07; Table 2), resulting in an about twofold larger $R_{\mathbf{U}}$ value. Removing all screw components results in an $R_{\mathbf{U}}$ of the order of 0.01 for all librations [$R_{\mathbf{U}}(d_x)$, $R_{\mathbf{U}}(d_y)$ and $R_{\mathbf{U}}(d_z)$ columns in Table 1].

These tests let us draw two conclusions. Firstly, the ensemble size required for reliable calculation of $\mathbf{U}_{\mathrm{ensemble},n}$ does not depend on the model size and, similarly to the one-atom case (§2.2), 5000–10 000 models are sufficient. Secondly, large values of the screw components are responsible for the disagreement between $\mathbf{U}_{\mathrm{ensemble},n}$ and $\mathbf{U}_{\mathrm{TLS},n}$ and the large reesulting $R_{\mathbf{U}}$. This conclusion prompted us to revisit the TLS

[2] $\mathbf{U}_{\mathrm{TLS,vibr}}$ and $\mathbf{U}_{\mathrm{ensemble,vibr}}$ from vibration cancel each other in the numerator of (4) since pure vibrations in the TLS model are always harmonic. Therefore, the denominator of (4) is larger for the full **U** matrices than for the matrices for librations only (this is easier to see in the coordinate system where $\mathbf{U}_{\mathrm{TLS,vibr}}$ is diagonal; the diagonal elements of all **U** matrices are always non-negative). This means that $R_{\mathbf{U}}$ for the overall motion can increase after excluding the vibration component.

decomposition algorithm described in Urzhumtsev *et al.* (2015).

2.4. Improvement of the TLS decomposition

In the decomposition of the TLS matrices into elemental motions, some parameters, including libration amplitudes and axes, are defined unambiguously. However, the screw parameters are derived using the **S** matrix, which is not unique but is defined with an arbitrary constant σ that can be added to or subtracted from its diagonal elements (Schomaker & Trueblood, 1968). This freedom in the definition of **S** does not change the ADP and provides the possibility for alternative (and possibly better) decompositions of the TLS matrices. In Urzhumtsev *et al.* (2013) we discussed a possible argument for the traditional choice of σ from the condition

$$\mathrm{tr}(S) = (S_{xx} - \sigma) + (S_{yy} - \sigma) + (S_{zz} - \sigma) = 0. \qquad (9)$$

Here S_{xx}, S_{yy}, S_{zz} are the diagonal elements of the matrix **S** expressed in the basis [**L**] of the principal libration directions; these directions are eigenvectors of the matrix **L**. In Urzhumtsev *et al.* (2015) we showed that (9) may result in TLS matrices that do not correspond to elemental motions, and to address this issue we suggested a better choice for the t value,

$$|\mathrm{tr}(S)| = |(S_{xx} - \sigma) + (S_{yy} - \sigma) + (S_{zz} - \sigma)| \rightarrow \min_{\sigma} \qquad (10)$$

under some additional constraints on σ that are discussed in that paper.

As shown in the previous paragraph, excessively large screw parameters lead to significant discrepancies between $\mathbf{U}_{\mathrm{ensemble},n}$ and $\mathbf{U}_{\mathrm{TLS},n}$. This suggests that a better alternative to (9) and (10) might be to choose σ such that it minimizes the norm of the screw vector $|\mathbf{s}|$. The new condition is then

$$|\mathbf{s}|^2 = s_x^2 + s_y^2 + s_z^2 = L_{xx}^{-2}(S_{xx} - \sigma)^2 + L_{yy}^{-2}(S_{yy} - \sigma)^2 + L_{zz}^{-2}(S_{zz} - \sigma)^2 \rightarrow \min_{\sigma}. \qquad (11)$$

Here, according to equations (5) and (8) in Urzhumtsev *et al.* (2015), $S_{xx} - \sigma = s_x\langle d_x^2\rangle$ and $L_{xx} = \langle d_x^2\rangle$ (and similar expressions for the four other terms) are the diagonal elements of the matrices **S** and **L** given in the basis [**L**].

In order to test the new approach for adjusting the **S** matrix, we used the same models and sets of TLS matrices as described in §2.3. For each set of matrices, $\mathrm{TLS_{MC}}$ or $\mathrm{TLS_{CA}}$, we extracted elemental motions using (11), generated $\mathbf{U}_{\mathrm{ensemble},n}$ using corresponding models and then computed $R_{\mathbf{U}}$ values using the previously obtained $\mathbf{U}_{\mathrm{TLS},n}$. Table 1 shows that the updated $R_{\mathbf{U}}$ values calculated using (11) to adjust **S** are acceptably low not only for the total motion but for each of the individual components, both for $\mathrm{TLS_{CA}}$ and for $\mathrm{TLS_{MC}}$. Fig. 3(*a*) shows $R_{\mathbf{U}}$ plots as a function of the number of generated models. The curves are nearly identical for both models, showing even lower values for $R_{\mathbf{U}}$ than the original curve for the C^{α}-only model.

A more striking result is obtained when applying the new correction method to a real-life example: PDB entry 4muy (Span *et al.*, 2014). In all tests TLS groups were used as defined in the PDB file. The 4muy model is composed of 40 TLS groups, and we focus on group No. 6 (residues 65–77 in chain *A*). The decomposition of the reported TLS matrices into motion parameters using the approach described previously (Urzhumtsev *et al.*, 2015) suggests removing $\sigma = 10^{-5}$ from the diagonal elements of the **S** matrix (expressed in Å rad). The $R_{\mathbf{U}}$ corresponding to these matrices is very high at 0.61, indicating a large disagreement between $\mathbf{U}_{\mathrm{ensemble},n}$ and $\mathbf{U}_{\mathrm{TLS},n}$ (Figs. 5*a* and 5*b* and the dashed curve in Fig. 3*b*). We suspected that this disagreement was owing to a very large value of the screw parameter s_x of 303.6 for the screw rotation around the

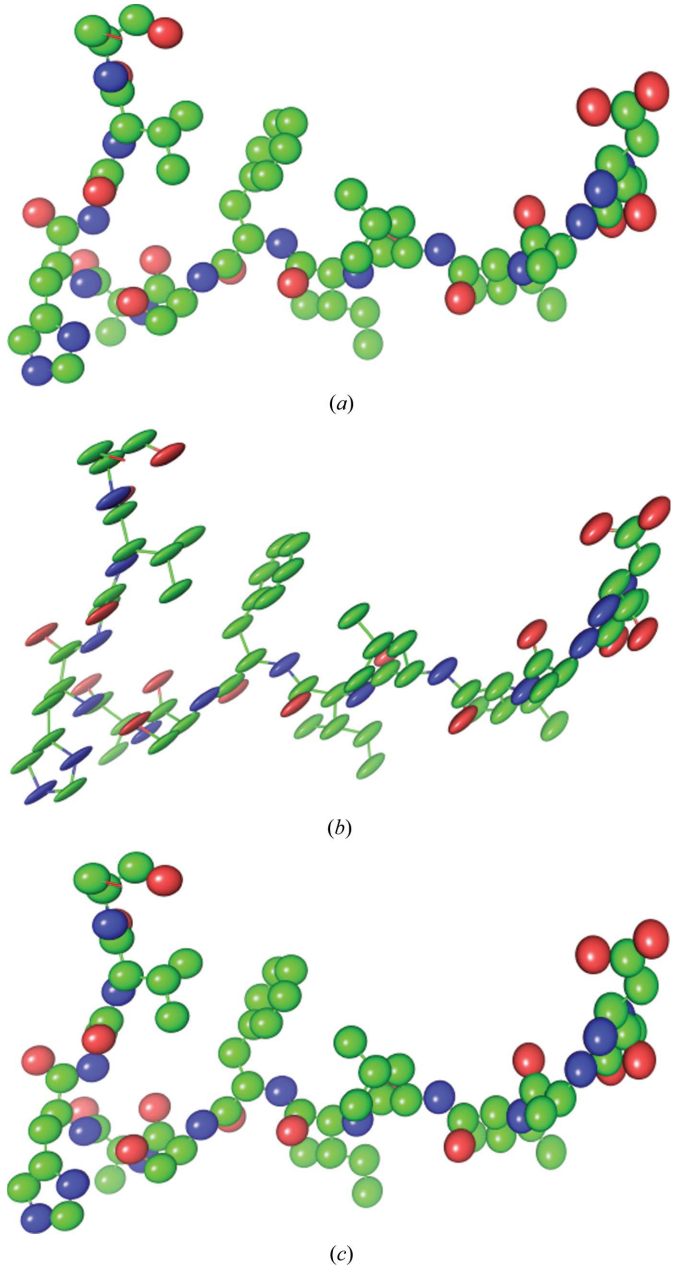


Figure 5
The **U** ellipsoids shown with *PyMOL* (DeLano, 2002) for the atoms of the sixth TLS group of the 4muy model. (*a*) $\mathbf{U}_{\mathrm{TLS}}$ matrices. (*b*) $\mathbf{U}_{\mathrm{ensemble}}$ matrices calculated with the elemental motions obtained using (10). (*c*) $\mathbf{U}_{\mathrm{ensemble}}$ matrices calculated with the elemental motions obtained using (11).

Table 3
Number of TLS groups with ADP matrices that are reproducible by explicit group motions ($R_{\mathbf{U}} \leq 0.05$).

PDB content (November 2016): 32 162 entries containing TLS records, 260 353 TLS groups in total. For 263 TLS groups all three matrices were zero and these groups were excluded from further work. Decomposition of TLS matrices into parameters of elemental motions was performed using (9), (10) and (11). The 'Extracted groups' column shows the total number of TLS groups for which parameter extraction was possible and 'Extracted entries' shows the number of PDB entries for which this was possible for all of the groups. 'Wrong content' shows the number of groups for which random-model generation was impossible for technical reasons and 'Libration undefined' shows the number of groups for which all libration matrices were zero. Other columns: overall motion (5), overall libration (6), conditions verified for each of the three librations of the group including their screw components (7) and conditions verified for each of the three pure librations of the group (8).

Method	Extracted entries	Extracted groups	Wrong content	Overall motion	Libration undefined	Overall libration	Individual screw	Individual libration
Equation (9)								
Total	4290	88434	314	88120	167	87953	87953	87953
$R_{\mathbf{U}} \leq 0.05$				45093		23042	6107	87908
Equation (10)								
Total	4826	95152	332	94820	167	94653	94653	94653
$R_{\mathbf{U}} \leq 0.05$				46627		24163	7478	94596
Equation (11)								
Total	4826	95150	332	94818	167	94651	94651	94651
$R_{\mathbf{U}} \leq 0.05$				57463		31395	11238	94590

axis $\mathbf{l}_x$ (Table 1). Applying (11) to adjust the **S** matrix resulted in σ increasing to 42×10^{-5}, which in turn reduced s_x to 0.1 and also reduced the respective $R_{\mathbf{U}}$ from 0.89 to 0.02. Fig. 5(*c*) shows the thermal ellipsoids obtained using the screw libration parameters extracted with (11): clearly, $\mathbf{U}_{\mathrm{ensemble},n}$ and $\mathbf{U}_{\mathrm{TLS},n}$ are much more similar (compare with Fig. 5*a*). Fig. 6 shows the variation of $R_{\mathbf{U}}$ and of $|\mathbf{s}|$ as a function of the σ value; indeed, the minimum of $R_{\mathbf{U}}$ is observed for σ obtained using (11). The $R_{\mathbf{U}}$ for the overall motion decreased to an acceptable value of 0.05, and for the libration alone it decreased from 0.85 to 0.11. The latter value is still high, possibly because by reducing s_x the procedure increased the magnitudes of s_y and s_z (from 2.90 to −3.38 and from −3.11 to −5.14, respectively; Tables 1 and 2). This test shows both the advantage of the new approach (11) compared with (9) and (10) and also its limitations. In this test using other norms, in particular $\max\{|s_x|, |s_y|, |s_z|\}$, in (11) did not improve the result. In general, there is no guarantee that (11) always results in the best screw parameters and further improvements may be needed, for example by using a local search around the σ value obtained with (11).

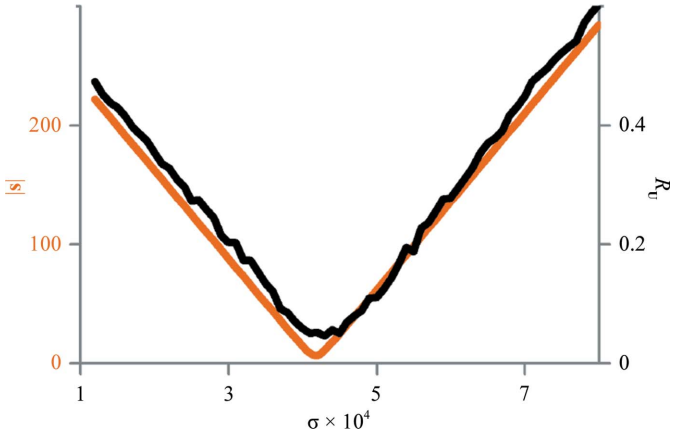


Figure 6
Variation of the vector norm $|\mathbf{s}|$ (11) (maroon) and of the $R_{\mathbf{U}}$ value (black) as a function of the parameter σ that is subtracted simultaneously from all diagonal elements of the **S** matrix during the decomposition of TLS matrices into parameters of elemental motions (4muy data; see §2.4). Small oscillations in $R_{\mathbf{U}}$ illustrate its stochastic nature.

Fig. 3(*b*) shows $R_{\mathbf{U}}$ as a function of the ensemble size for the 4muy model generated using parameters obtained with (10) and (11). It shows the significant difference between the results of the two approaches for correcting the **S** matrix and also confirms the previous observation that 5000–10 000 models are sufficient.

3. PDB analysis and improvement of the TLS decomposition

3.1. Model selection and analysis setup

The PDB (as of 14 November 2016) contains 123 954 entries, of which 32 162 contain TLS records. Since each PDB entry may contain more than one TLS record, a total of 260 353 TLS groups are available in the PDB. For each of these groups we tried to determine the corresponding elemental motions. This was performed using *phenix.tls_as_xyz* as described in Urzhumtsev *et al.* (2015).

88 697 groups could be interpreted in terms of elemental motions. In 263 of these cases all three matrices were composed of zeros. Some further 314 groups were excluded because the deposited TLS information was corrupted in a number of different ways (missing TLS group origin, non-interpretable atomic or TLS records *etc.*).

The remaining 88 120 TLS sets were subjected to three independent rounds of decomposition into elemental motions, each applying corrections to the **S** matrix using (9), (10) and (11), respectively. When using (10) and (11) the constraints on σ described in Urzhumtsev *et al.* (2015) were applied. For each set of extracted parameter values we analyzed the following motions.

(i) A combination of three screw rotations and the vibration component, *i.e.* the overall motion (5).

(ii) A combination of the screw rotations with no vibration components (6).

(iii) Each of the three screw rotations individually (7).

(iv) Each of the three pure rotations (8).

For each of these motions we calculated $\mathbf{U}_{\mathrm{TLS},n}$. We then generated an ensemble of 5000 models using *phenix.tls_as_xyz* and we used this ensemble to calculate $\mathbf{U}_{\mathrm{ensemble},n}$. Finally, we compared $\mathbf{U}_{\mathrm{ensemble},n}$ with $\mathbf{U}_{\mathrm{TLS},n}$ using $R_{\mathbf{U}}$. Details of this analysis are given in Table 3 and are commented on below.

3.2. Analysis of the elemental motions using (9)

Table 3 shows the overall statistics and the number of TLS groups with $R_{\mathbf{U}} \leq 0.05$. For the overall motion combining all elemental components together (5), about half of the TLS groups for which we could extract the motion parameters pass the test condition $R_{\mathbf{U}} \leq 0.05$ (this is approximately 17% of the total number of deposited TLS groups). The same condition applied when considering libration components only (6) reduced the number of acceptable groups roughly by half. In the case of considering librations individually (equations 7 and 8) the criterion $R_{\mathbf{U}} \leq 0.05$ selects only 2.3% of the TLS groups (6107 groups).

There are 45 sets where using pure rotations gives $R_{\mathbf{U}} > 0.05$, all of which correspond to large libration amplitudes. Thus, the main source of the discrepancies between $\mathbf{U}_{\mathrm{ensemble},n}$ and $\mathbf{U}_{\mathrm{TLS},n}$ are the screw components.

We checked (Fig. 7*a*) the distribution of the TLS groups as a function of $R_{\mathbf{U}}$ calculated for the overall motion (5), for the motion excluding vibrations (6) and separately for the screw components (7). The first distribution (maroon full rectangles) has a peak in the interval 0.02–0.05 which corresponds to the TLS matrices that comply with the underlying study. Nevertheless, for a significant number of sets this value is above 0.05. Major problems come from screw components, for which many TLS groups have an $R_{\mathbf{U}}$ above 0.10 or even above 0.20 (blue full rectangles).

The largest value of $|\mathbf{s}|$ observed across all TLS groups is greater than 1000 Å. Such a large value means that for a rotation of 0.01 rad, *i.e.* approximately 0.6°, the rotated atoms would move by 10 Å in the direction of the rotation axis, which is clearly physically unrealistic. Fig. 7(*b*) shows that there are many groups with large values of $|\mathbf{s}|$. The larger the screw parameter $|\mathbf{s}|$, the larger the $R_{\mathbf{U}}$ values (Fig. 7*c*). However, since a particular screw motion also depends on the libration amplitude and on the positions of the axes, this does not allow anharmonic rotations to be discriminated unambiguously using this value alone (Fig. 7*c*).

3.3. Analysis of the elemental motions using (10) and (11)

Using the approach in (10) allows motion parameters to be extracted for 6700 more TLS sets compared with (9). Table 3 and Fig. 7(*b*) show that the distributions of $R_{\mathbf{U}}$ values and the screw parameters $|\mathbf{s}|$ are similar to those using (9).

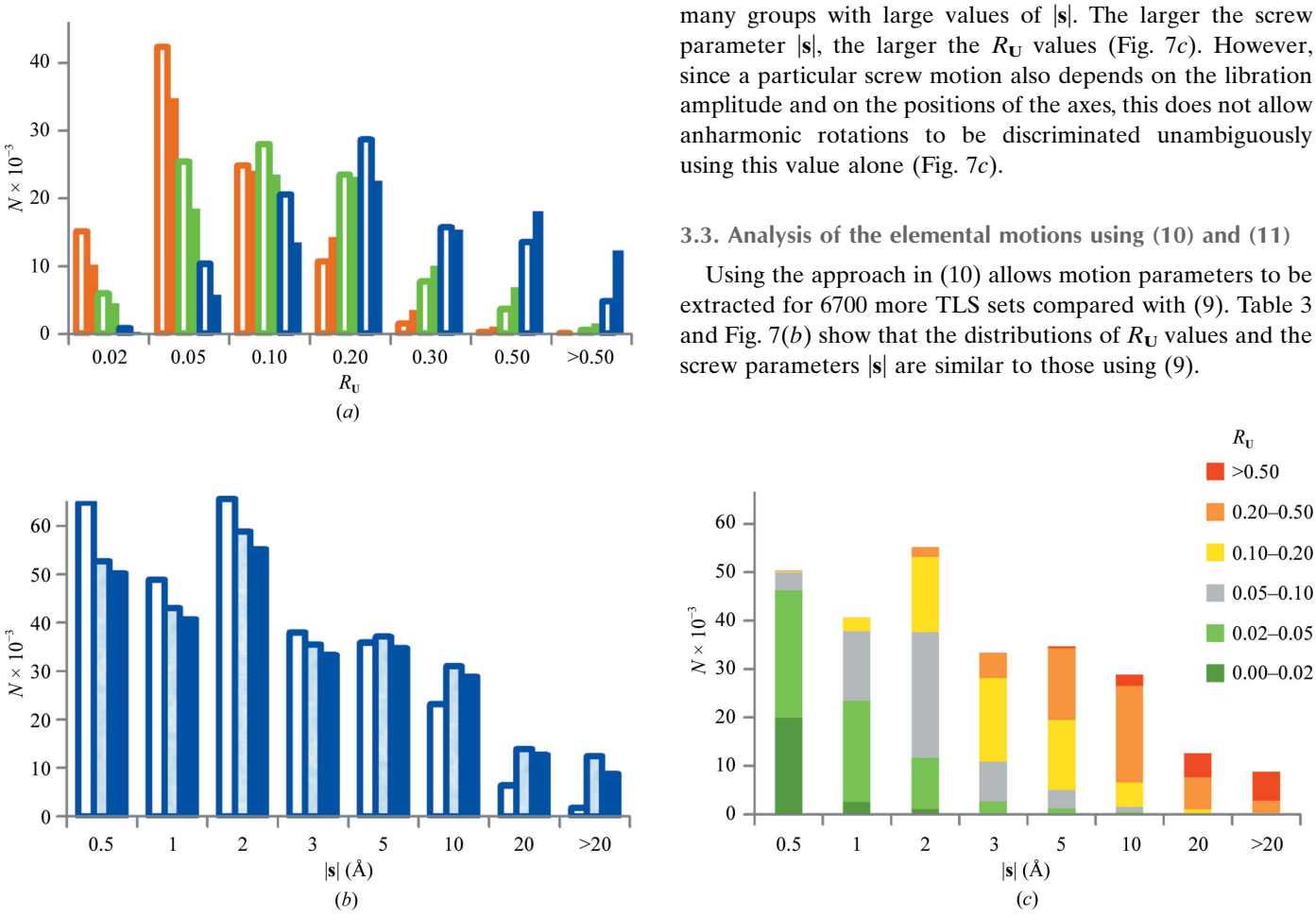


Figure 7
Distribution of TLS groups in the PDB. (*a*) Number of TLS groups with $R_{\mathbf{U}}$ values in the given intervals; distributions are shown for the total motions (maroon), for the total motions excluding vibration components (green) and for the individual screw rotations (blue). The histograms are shown when using (9) (full rectangles) and (11) (open rectangles). (*b*) Number of screw rotations as a function of the screw parameter $|\mathbf{s}|$; the histograms are shown when using (9) (blue rectangles), (10) (light blue rectangles) and (11) (open rectangles). $R_{\mathbf{U}}$ values are calculated for all independent screw librations (7). (*c*) Number of TLS groups with different $R_{\mathbf{U}}$ values for the given interval of $|\mathbf{s}|$. The screw parameters were extracted by the procedure using (9); $R_{\mathbf{U}}$ values are calculated as in (*b*). See §3 for details.

Repeating the same calculations using (11) shows a significantly greater difference compared with using (10) (Table 3). Considering all motions together, the number of groups for which $R_{\mathbf{U}} \leq 0.05$ increased results in more than 12 000 groups compared with using (9). Considering only screw librations, the number of groups satisfying the condition $R_{\mathbf{U}} \leq 0.05$ is doubled compared with using (9) ('Individual screw' column in Table 3). Fig. 7(*b*) shows that the number of rotations with a large value of the screw parameter $|\mathbf{s}|$ is significantly reduced. The largest value of $|\mathbf{s}|$ fell to below 700 Å (which is still overly large).

Fig. 7(*a*) shows that using the approach in (11) instead of that in (9) significantly shifts all three distributions to the left (compare the open rectangles with the full rectangles in Fig. 7*a*), *i.e.* it improves the similarity between $\mathbf{U}_{\mathrm{ensemble},n}$ and $\mathbf{U}_{\mathrm{TLS},n}$. In particular, $R_{\mathbf{U}} \leq 0.10$ for the majority of TLS sets when analyzing only the matrices for the total motion (5). However, considering vibrations alone, $R_{\mathbf{U}} > 0.10$ for more than a third of the models even when using the improved decomposition into elemental motions (11).

4. Discussion

Validation of atomic models is now routine in macromolecular crystallography and is an integral part of structure submission to the Protein Data Bank (Read *et al.*, 2011; Gore *et al.*, 2017). It requires nomenclature compliance and fit to experimental data. Atomic coordinates are subjected to validation that includes analysis of stereochemistry and molecular packing. Atomic displacement parameters (ADPs) are also subjected to validation. For isotropic ADPs the existing validation criteria are rather simple: their values must be positive, not excessively large and not vary too much between neighbouring atoms. For anisotropic ADP values the criteria are somewhat more complex (Hirshfeld, 1976; Schneider, 1996). Similarly to atomic coordinates and displacement parameters, TLS matrices are model parameters and therefore should be subjected to some form of validation. Depending on the accepted paradigm (§1.3) the scope of TLS validation may refer to two questions: (i) how well does the the TLS approximation explain the experimental data and how well does it describe the atomic displacement parameters (see, for example, Merritt, 2011, 2012) and (ii) are the particular descriptors of the TLS model also consistent with the TLS formalism in addition to (i). Addressing the first question does not require analysis of the TLS matrices themselves but only of the derived ADP values. This includes making sure that the ADPs are positive definite and vary smoothly between adjacent atoms and TLS groups (Winn *et al.*, 2001; Zucker *et al.*, 2010; Merritt, 2011, 2012). The current work addresses the second question, which focuses exclusively on the analysis of TLS matrices and the parameters of group motion that they encode. Since modern atomic model refinement packages use an indirect TLS parameterization (§1.3), *i.e.* they refine the elements of the TLS matrices and not the parameters of group motions, it is unsurprising to find that some TLS matrices do not comply with the assumption of harmonic motion that the TLS modelling theory is built upon. The number of such cases may vary based upon the different measures or thresholds that are used. For example, using the criteria discussed above we find that only 2.3% of the TLS groups reported in the PDB can be interpreted in terms of elemental harmonic motions. We envisage two reasons for this. Firstly, the validation of TLS refinement results, focusing on TLS matrices and corresponding group motions, has never been enforced. Secondly, the implementation of TLS refinement in modern refinement packages does not allow control of the parameters of group motion by means of restraints or constraints (see a discussion in Painter & Merritt, 2006) because these parameters are refined indirectly. Unsurprisingly, such unrestrained refinement provides no guarantee of TLS matrices that are interpretable in terms of harmonic elemental motions.

In this work, we have developed methods and a software implementation in the *PHENIX* suite to analyze the results of TLS refinements. These methods are based on comparison of individual atomic displacement parameters calculated analytically from the TLS matrices with ADPs derived numerically using parameters of elemental motions extracted from the TLS matrices. We theorise that large differences between these matrices indicate problematic TLS parameters. In particular, this may indicate a suboptimal choice of TLS groups or refinement protocol. We show that a post-refinement correction of the deposited TLS matrices makes it possible to curate some but not all of the problematic TLS groups.

The analyses presented in this work rely on the choice of particular criteria (metrics and thresholds). These criteria may be optimized further, which is a nontrivial project and may help to diagnose the problem while still not addressing it. A possibly better investment of effort would be to improve TLS refinement protocols so that they operate in terms of elemental parameters of motions, which has been proposed previously (Tickle & Moss, 1999). This would make it possible to control the refinable parameters directly during refinement and therefore keep them physically realistic without the need for post-refinement corrections. This is a major undertaking both mathematically and algorithmically, which may be considered as a future improvement to the *PHENIX* refinement software.

The methods and tools discussed in this manuscript have been implemented and are available in *PHENIX* 1.12 and later. Data and scripts that can be used to reproduce the figures and tables are available at http://phenix-online.org/phenix_data/afonine/tls2/.

Acknowledgements

We thank Dr N. W. Moriarty for critical reading of the manuscript and useful discussions.

Funding information

This research was supported by the NIH (grant GM063210) and the Phenix Industrial Consortium. This work was partially supported by the US Department of Energy under Contract

DE-AC02-05CH11231. AU acknowledges the support and the use of the resources of the French Infrastructure for Integrated Structural Biology (FRISBI; ANR-10-INBS-05) and of Instruct-ERIC.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M. W., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Cryst.* D**68**, 352–367.
Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.
Chaudhry, C., Horwich, A. L., Brunger, A. T. & Adams, P. D. (2004). *J. Mol. Biol.* **342**, 229–245.
Cruickshank, D. W. J. (1956). *Acta Cryst.* **9**, 757–758.
DeLano, W. L. (2002). *PyMOL*. http://www.pymol.org.
Dunitz, J. D. & White, D. N. J. (1973). *Acta Cryst.* A**29**, 93–94.
Gore, S. *et al.* (2017). *Structure*, **25**, 1916–1927.
Harris, G. W., Pickersgill, R. W., Howlin, B. & Moss, D. S. (1992). *Acta Cryst.* B**48**, 67–75.
Hirshfeld, F. L. (1976). *Acta Cryst.* A**32**, 239–244.
Howlin, B., Butler, S. A., Moss, D. S., Harris, G. W. & Driessen, H. P. C. (1993). *J. Appl. Cryst.* **26**, 622–624.
Kullback, S. & Leibler, R. A. (1951). *Ann. Math. Stat.* **22**, 79–86.
Kuriyan, J. & Weis, W. I. (1991). *Proc. Natl Acad. Sci. USA*, **88**, 2773–2777.
Merritt, E. A. (1999). *Acta Cryst.* D**55**, 1997–2004.
Merritt, E. A. (2011). *Acta Cryst.* A**67**, 512–516.
Merritt, E. A. (2012). *Acta Cryst.* D**68**, 468–477.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. (1999). *Acta Cryst.* D**55**, 247–255.
Painter, J. & Merritt, E. A. (2006). *Acta Cryst.* D**62**, 439–450.
Papiz, M. Z., Prince, S. M., Howard, T., Cogdell, R. J. & Isaacs, N. W. (2003). *J. Mol. Biol.* **326**, 1523–1538.
Raaijmakers, H., Törö, I., Birkenbihl, R., Kemper, B. & Suck, D. (2001). *J. Mol. Biol.* **308**, 311–323.
Read, R. J. *et al.* (2011). *Structure*, **19**, 1395–1412.
Šali, A., Veerapandian, B., Cooper, J. B., Moss, D. S., Hofmann, T. & Blundell, T. L. (1992). *Proteins*, **12**, 158–170.
Schneider, T. R. (1996). *Proceedings of the CCP4 Study Weekend. Macromolecular Refinement*, edited by E. Dodson, M. Moore, A. Ralph & S. Bailey, pp. 133–144. Warrington: Daresbury Laboratory. http://purl.org/net/epubs/work/35264.
Schomaker, V. & Trueblood, K. N. (1968). *Acta Cryst.* B**24**, 63–76.
Span, I., Wang, K., Eisenreich, W., Bacher, A., Zhang, Y., Oldfield, E. & Groll, M. (2014). *J. Am. Chem. Soc.* **136**, 7926–7932.
Tickle, I. & Moss, D. S. (1999). *Modelling Rigid-body Thermal Motion In Macromolecular Crystal Structure Refinement.* http://people.cryst.bbk.ac.uk/~tickle/iucr99/iucrcs99.html.
Trueblood, K. N. (1978). *Acta Cryst.* A**34**, 950–954.
Trueblood, K. N. & Dunitz, J. D. (1983). *Acta Cryst.* B**39**, 120–133.
Urzhumtsev, A., Afonine, P. V. & Adams, P. D. (2013). *Crystallogr. Rev.* **19**, 230–270.
Urzhumtsev, A., Afonine, P. V., Van Benschoten, A. H., Fraser, J. S. & Adams, P. D. (2015). *Acta Cryst.* D**71**, 1668–1683.
Urzhumtsev, A., Afonine, P. V., Van Benschoten, A. H., Fraser, J. S. & Adams, P. D. (2016). *Acta Cryst.* D**72**, 1073–1075.
Van Benschoten, A. H., Afonine, P. V., Terwilliger, T. C., Wall, M. E., Jackson, C. J., Sauter, N. K., Adams, P. D., Urzhumtsev, A. & Fraser, J. S. (2015). *Acta Cryst.* D**71**, 1657–1667.
Wilson, M. A. & Brunger, A. T. (2000). *J. Mol. Biol.* **301**, 1237–1256.
Winn, M. D., Isupov, M. N. & Murshudov, G. N. (2001). *Acta Cryst.* D**57**, 122–133.
Yousef, M. S., Fabiola, F., Gattis, J. L., Somasundaram, T. & Chapman, M. S. (2002). *Acta Cryst.* D**58**, 2009–2017.
Zucker, F., Champ, P. C. & Merritt, E. A. (2010). *Acta Cryst.* D**66**, 889–900.