From deep TLS validation to ensembles of atomic models built from elemental motions

Alexandre Urzhumtsev[1,2,#], Pavel V. Afonine[3], Andrew H. Van Benschoten[4], James S. Fraser[4] and Paul D. Adams[3,5]

[1]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
[2]Université de Lorraine, Faculté des Sciences et Technologies, BP 239, 54506 Vandoeuvre-les-Nancy, France
[3]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA
[4]Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, California, USA
[5]Department of Bioengineering, University of California Berkeley, Berkeley, CA 94720 USA

[#]sacha@igbmc.fr

Synopsis

Procedures for the decomposition of *TLS* matrices into elementary vibrations and librations indicates possible errors in the definition of these matrices and corrects them when possible. The program outputs the corresponding vibration-libration parameters and generates structural ensembles.

Keywords

TLS model, libration, vibration, correlated motions, validation, ensemble of models

Abstract

The widely used *Translation Libration Screw* (TLS) approximation describes concerted motions of atomic groups in X-ray refinement. TLS refinement often provides a better interpretation of diffraction data and the resulting rigid body motions may subsequently be assigned biochemical significance. In TLS refinement, three matrices (*T, L and S*) describe harmonic vibration, libration and their correlation. Because these matrices describe specific motions, they impose a number of conditions on their elements. Ignoring these conditions while refining the matrix elements may result in matrices that cannot be interpreted in terms of physically realistic motions. We describe a mathematical framework and the computational tools to analyze refined *TLS* matrices through their decomposition into descriptors of underlying motions. This allows for straightforward validation and identification of implausible *TLS* parameters. An algorithm for the generation of structural ensembles that are consistent with given *TLS* parameters, implemented as a part of the *Phenix* project, is also described.

1. **Introduction**

Crystallographic models fit to X-ray, neutron, or electron diffraction data describe each atom by its central position $\mathbf{r_0}$ and additional parameters. Structural disorder, particularly thermal motion, is primarily fit by the so-called Debye-Waller factor (also known as the B-factor or displacement parameter) that reflects the probability of an atom moving from its central position. This factor is accounted for as a multiplier for each structure factor corresponding to integer indices (h,k,l). In a harmonic approximation, this factor can be presented as following (see for example, Grosse-Kunstleve & Adams, 2002, and references therein)

$$\exp\left(-2\pi^2\mathbf{h}^\tau O^{-1}U_{Cart}\left(O^{-1}\right)^\tau\mathbf{h}\right) \qquad . \qquad (1)$$

Here O is the orthogonalization matrix for the given crystal, $\mathbf{h}$ is the column vector with the indices (h,k,l), and τ signifies matrix or vector transposition. (In Grosse-Kunstleve & Adams (2002) the orthogonalization matrix is noted as A; here this letter is used for the matrix in development of U_{Cart}, following Tickle & Moss, 1999 and Urzhumtsev $et\ al.$, 2013). The symmetric matrix U_{Cart} is defined by the average shifts (and their correlations) along the coordinate axes. The matrix U_{Cart} varies between atoms and for isotropic atoms is diagonal with equal elements.

However, purely isotropic motion is rarely observed in macromolecular crystals where the displacement of groups of atoms can be correlated. For example, while each atom of a side chain may oscillate independently around its central position, the side chain itself may oscillate around one (or several) bonds. This concerted motion is of special interest for several reasons. First, it may characterize the mobility of macromolecular domains and thus give insight into molecular mechanism. Second, describing only the common motion may drastically simplify the model and reduce the number of parameters needed to model the data. This may be especially attractive when working at relatively low resolutions.

A description of concerted molecular motions has been introduced by Cruickshank (1956) and Schomaker & Trueblood (1968) and later developed in a number of works (for example Johnson, 1970; Scheringer, 1973; Howlin $et\ al.$, 1989, 1993; Kuriyan & Weis, 1991; Schomaker & Trueblood, 1998; Tickle & Moss, 1999; Winn $et\ al.$, 2001, 2003; Painter & Merritt, 2005, 2006a, 2006b; for a recent comprehensive review see Urzhumtsev $et\ al.$ 2013). Since rigid body displacement is a composition of translation and rotation (see for example Goldstein, 1950), Schomaker & Trueblood presented the matrices $U_{Cart,n}$ for all atoms $n = 1, 2, \ldots N$ within the model as a sum

$$U_{Cart,n} = T + A_n L A_n^\tau + A_n S + S^\tau A_n^\tau \qquad (2)$$

Anti-symmetric matrices A_n are functions of the Cartesian coordinates $\left(x_n, y_n, z_n\right)$ of atom n

$$A_n = \begin{pmatrix} 0 & z_n & -y_n \\ -z_n & 0 & x_n \\ y_n & -x_n & 0 \end{pmatrix} \qquad . \qquad (3)$$

Matrix S and symmetric matrices T and L are common to all atoms of the moving group and are defined by the following elementary motions. L describes libration (oscillating rotations) around three principal rotation axes mutually orthogonal to each other. T describes an apparent translation of the atomic group (a term 'vibration' may be more appropriate for random translations around a central position). S describes the correlation between librations and vibrations (for a detailed definition Urzhumtsev *et al*. 2013). We use the term 'apparent translation' because matrix T may have an additional contribution from librations as discussed in Section 2.

Compared to calculating *TLS* matrices from corresponding libration and vibration parameters, the inverse procedure of decomposing TLS matrices into these parameters is more difficult. As discussed previously (for example, Johnson, 1970; Scheringer, 1973; Tickle & Moss, 1999) the problem itself is poorly posed since the same set of diffraction data (and as a consequence the same set of the *TLS* matrices) may correspond to different common motions of the contributing atoms or atomic groups. Nevertheless, TLS refinement allows one to search for composite motions (librations and vibrations) that correspond to a particular set of matrices. Thus, one may refine either the parameters of the composite motions or the elements of the *TLS* matrices as independent parameters, as is currently implemented in modern refinement programs,.

To describe underlying elementary motions, the matrices must obey certain conditions; otherwise, the physical motions they represent are nonsensical. First, there is the trivial condition that the matrices T and L must be symmetric and positive semidefinite. Second, the matrices must be compatible with each other. For example, since matrix S describes the correlation between libration and vibration, its elements should be consistent with the elements of T and L. Ignoring this condition may result in a non-positive definite matrix T after the apparent contribution of libration is subtracted. Interestingly, refinement of *TLS* matrix elements without taking into account these conditions can still result in physically reasonable individual atomic displacement parameters and an improved model to data fit. Zucker *et al*. (2010) analyzed all PDB entries (Bernstein *et al*., 1977; Berman *et al*., 2000) that employ *TLS* and suggested tools for validating *TLS* parameters to ensure that individual atomic displacement parameters were physically realistic. While programs for the

4

extraction of libration and vibration parameters from *TLS* matrices (Howlin *et al.*, 1993; Painter & Merritt, 2005). A comprehensive mathematical protocol for accomplishing this task could be helpful for determining if decomposition into physically interpretable motions is even possible. This manuscript describes the corresponding algorithm (Fig. 1) and provides practical calculation protocols in a ready-to-program style. In particular, we discuss a list of conditions defining the plausibility of such decomposition. Below, we discuss the decomposition of a single set of *TLS* matrices; however, the procedure is similar when molecular models are composed of several such groups. For a PDB model that contain several sets of *TLS* matrices, each set can be processed independently.

Thus, analyzing the physical reasonableness of the individual atomic displacement parameters and the rigid-body motions that give rise to them require separate validation tools. Indeed, from roughly 200,000 sets of *TLS* matrices in the PDB (25,284 entries from the total 194,633 PDB depositions) about one-third contain T or L matrices that are non positive-semidefinite and another one-third (Table 1) cannot describe vibration-libration correlated motions . Although these numbers can be slightly reduced by removing marginal cases, these results suggest that decomposition of *TLS* matrices into elementary motions can be an important validation step. This analysis can be used to build ensembles of models representing plausible conformations, which will allow for a better description of diffraction data and for the testing of model agreement with diffuse X-ray scattering (Van Benschoten *et al.*, accompanying paper).

2. TLS matrices and different bases

2.1. *Matrix presentations and decomposition difficulties*

Even if a rigid body is involved in several simultaneous motions (supposing that the amplitudes of these motions are relatively small and the motions are harmonic), each motion can be described by a libration around three mutually orthogonal axes and by a vibration following three other mutually orthogonal axes (Urzhumtsev *et al*. 2013, and references therein). For example, if the principal vibration axes coincide with the coordinate axes, and $\langle t_x^2 \rangle, \langle t_y^2 \rangle, \langle t_z^2 \rangle$ are the corresponding squared root-mean-square deviations (*rmsd*) along these axes, the matrix T in equation (2) is equal to

$$T = V = \begin{pmatrix} \langle t_x^2 \rangle & 0 & 0 \\ 0 & \langle t_y^2 \rangle & 0 \\ 0 & 0 & \langle t_z^2 \rangle \end{pmatrix} \tag{4}$$

Similarly, if the libration axes coincide with the coordinate axes, the matrix L in (2) becomes:

$$L = \begin{pmatrix} \langle d_x^2 \rangle & 0 & 0 \\ 0 & \langle d_y^2 \rangle & 0 \\ 0 & 0 & \langle d_z^2 \rangle \end{pmatrix} \tag{5}$$

Here $\langle d_x^2 \rangle, \langle d_y^2 \rangle, \langle d_z^2 \rangle$ are the squared *rmsd*s of the vibration angle expressed in radians; for small deviations they are numerically equal to the squared *rmsd*s of points at a unit distance from the corresponding axes. One difficulty is that in practice the principal vibration and libration axes are not parallel to the coordinate axes, which requires transforming the matrices T and L into a more general form. Another difficulty is that vibration and libration motions may not be independent, as seen in helical rotations. If the correlations between translation and libration are characterized by parameters s_x, s_y, s_z (see Urzhumtsev *et al.* 2013, for a formal definition), the librations generate an additional contribution to translation so that matrix T becomes different from V as defined by equation (4) adding diagonal terms:

$$T = T_C = V + C_S = V + \begin{pmatrix} s_x^2 \langle d_x^2 \rangle & 0 & 0 \\ 0 & s_y^2 \langle d_y^2 \rangle & 0 \\ 0 & 0 & s_z^2 \langle d_z^2 \rangle \end{pmatrix} \tag{6}$$

and also resulting in the non-zero S matrix

$$S = \begin{pmatrix} s_x \langle d_x^2 \rangle & 0 & 0 \\ 0 & s_y \langle d_y^2 \rangle & 0 \\ 0 & 0 & s_z \langle d_z^2 \rangle \end{pmatrix} \tag{7}$$

Yet another difficulty is that the principal libration axes do not need to pass through the origin, or even through a common point. This generates an additional apparent translation component and adds another term

$$D_W = \tag{8}$$

$$
\begin{pmatrix}
\left(w_z^j\right)^2\langle d_y^2\rangle + \left(w_y^k\right)^2\langle d_z^2\rangle & -w_x^k w_y^k\langle d_z^2\rangle & -w_x^j w_z^j\langle d_y^2\rangle \\[2ex]
-w_x^k w_y^k\langle d_z^2\rangle & \left(w_z^i\right)^2\langle d_x^2\rangle + \left(w_x^k\right)^2\langle d_z^2\rangle & -w_y^i w_z^i\langle d_x^2\rangle \\[2ex]
-w_x^j w_z^j\langle d_y^2\rangle & -w_y^i w_z^i\langle d_x^2\rangle & \left(w_y^i\right)^2\langle d_x^2\rangle + \left(w_x^j\right)^2\langle d_y^2\rangle
\end{pmatrix}
$$

to the T matrix (Urzhumtsev *et al*. 2013). Similarly, matrix S must account for

$$
\begin{pmatrix}
0 & w_z^i\langle d_x^2\rangle & -w_y^i\langle d_x^2\rangle \\[2ex]
-w_z^j\langle d_y^2\rangle & 0 & w_x^j\langle d_y^2\rangle \\[2ex]
w_y^k\langle d_z^2\rangle & -w_x^k\langle d_z^2\rangle & 0
\end{pmatrix}
\tag{9}
$$

in addition to (7). Finally, the same actual motion can be presented by different combinations of the elementary motions since one may shift a rotation axis and compensate for this by an additional translation. The same total motion may also be split differently between correlated and independent contributions. These nuances complicate the procedure of extracting the descriptors of elementary motions from the *TLS* matrices.

2.2. *Molecular basis and center of reaction*

Formally, any origin (and, more generally, any basis) can be chosen to describe the composite motions. Clearly, description of the vibration does not depend on the choice of the origin, while the position of the libration axes changes as function of origin choice. Most often, the origin is taken in the center of mass of the atomic group or in between the axes so that the atomic displacements due to the choice of these axes are similar. This point is called the center of diffusion (Brenner, 1967) or the center of reaction (Tickle & Moss, 1999). Shifting from one origin to another does not change L but both T, minimizing its trace, and S, making it symmetric for the center of reaction (Brennen, 1967; Tickle & Moss, 1999; Urzhumtsev *et al*. 2013). Changing the origin does not modify the algorithm of the search for the composite motions. Thus, the *TLS* matrices depend on the choice of the coordinate basis and in particular its origin. In what follows, we suppose that the original *TLS* matrices are defined in a basis chosen with a relation to the molecule, which we will call the M basis (for 'Molecular'). As we will be switching between different bases, we indicate the corresponding basis by index in brackets such as $T_{[M]}$, $L_{[M]}$ and $S_{[M]}$.

3. Basis of libration axes

3.1. Diagonalization of the L matrix (L basis; step A)

We start the procedure from the matrix $L_{[M]}$ that depends only on the libration parameters. Let

$$T = T_{[M]} = \begin{pmatrix} T_{xx} & T_{xy} & T_{xz} \\ T_{yx} & T_{yy} & T_{yz} \\ T_{zx} & T_{zy} & T_{yz} \end{pmatrix}, L = L_{[M]} = \begin{pmatrix} L_{xx} & L_{xy} & L_{xz} \\ L_{yx} & L_{yy} & L_{yz} \\ L_{zx} & L_{zy} & L_{zz} \end{pmatrix}, S = S_{[M]} = \begin{pmatrix} S_{xx} & S_{xy} & S_{xz} \\ S_{yx} & S_{yy} & S_{yz} \\ S_{zx} & S_{zy} & S_{yz} \end{pmatrix}, \qquad (10)$$

with $T_{xy} = T_{yx}$, $T_{xz} = T_{zx}$, $T_{yz} = T_{zy}$ and $L_{xy} = L_{yx}$, $L_{xz} = L_{zx}$, $L_{yz} = L_{zy}$, be the *TLS* matrices defined in the original M basis. Importantly, for further analysis of the T matrices, we remind the reader that if a matrix is symmetric, it remains symmetric for any rotation of the coordinate system.

The principal libration axes correspond to three mutually orthogonal eigenvectors of the symmetric matrix $L_{[M]}$. First we search for the corresponding eigenvalues $0 \le \lambda_1 \le \lambda_2 \le \lambda_3$, which must be non-negative (see equation (5); eigenvalues do not change with the coordinate system). Let l_1, l_2, l_3 be the corresponding normalized eigenvectors from which we construct a new basis L as

$$l_x = \pm l_1, l_y = l_2, l_z = l_3 \qquad (11)$$

The appropriate sign for l_x is chosen so that the vectors in (11) form a right-hand triad; for example one can take $l_x = l_y \times l_z$ such that the condition is guaranteed.

The transition from the M basis into the L basis is a rotation with the corresponding matrix composed from the coordinates of the new basis vectors

$$R_{ML} = \begin{pmatrix} (l_x)_x & (l_y)_x & (l_z)_x \\ (l_x)_y & (l_y)_y & (l_z)_y \\ (l_x)_z & (l_y)_z & (l_z)_z \end{pmatrix} \qquad (12)$$

The *TLS* matrices in the L basis are

8

$$L_{[L]} = R_{ML}^{\tau} L_{[M]} R_{ML}$$

$$T_{[L]} = R_{ML}^{\tau} T_{[M]} R_{ML} \tag{13}$$

$$S_{[L]} = R_{ML}^{\tau} S_{[M]} R_{ML}$$

where matrix $L_{[L]}$ is diagonal with the elements $L_{[L]xx} = \lambda_1$, $L_{[L]yy} = \lambda_2$, $L_{[L]zz} = \lambda_3$ at its diagonal. The atomic coordinates in the L basis are

$$\begin{pmatrix} x_{[L]n} \\ y_{[L]n} \\ z_{[L]n} \end{pmatrix} = R_{ML}^{-1} \begin{pmatrix} x_{[M]n} \\ y_{[M]n} \\ z_{[M]n} \end{pmatrix} = R_{ML}^{\tau} \begin{pmatrix} x_{[M]n} \\ y_{[M]n} \\ z_{[M]n} \end{pmatrix} \tag{14}$$

3.2. *Position of the libration axes in the L basis (step B)*

Given matrix $L_{[L]}$ in the basis of its eigenvectors, we obtain the estimates $\langle d_x^2 \rangle = L_{[L]xx}$, $\langle d_y^2 \rangle = L_{[L]yy}$, $\langle d_z^2 \rangle = L_{[L]zz}$ of the squared libration amplitudes around the three principal libration axes.

In the general case these libration axes do not pass through the origin but through some points $\mathbf{w}^{lx} = \left(w_x^{lx}, w_y^{lx}, w_z^{lx} \right)$, $\mathbf{w}^{ly} = \left(w_x^{ly}, w_y^{ly}, w_z^{ly} \right)$, $\mathbf{w}^{lz} = \left(w_x^{lz}, w_y^{lz}, w_z^{lz} \right)$. In fact, the *x*-component of $\mathbf{w}^{lx}$, *y*-component of $\mathbf{w}^{ly}$ and *z*-component of $\mathbf{w}^{lz}$ can be any values. In other words, the coordinate axes chosen by construction as eigenvectors of $L_{[L]}$ are parallel to the libration axes but do not necessarily coincide with them.

Using equations (5) and (9) we calculate the positions of the rotation axes (Urzhumtsev *et al.* 2013) as

$$w_{[L]y}^{lx} = -\frac{S_{[L]xz}}{L_{[L]xx}}, \quad w_{[L]z}^{lx} = \frac{S_{[L]xy}}{L_{[L]xx}},$$

$$w_{[L]x}^{ly} = \frac{S_{[L]yz}}{L_{[L]yy}}, \quad w_{[L]z}^{ly} = -\frac{S_{[L]yx}}{L_{[L]yy}}, \tag{15}$$

$$w_{[L]x}^{lz} = -\frac{S_{[L]zy}}{L_{[L]zz}}, \quad w_{[L]y}^{lz} = \frac{S_{[L]zx}}{L_{[L]zz}}$$

9

A zero value of any denominator in (15) means that there is no rotation around the corresponding axis; in this case the two corresponding numerator values must be equal to zero too, otherwise the input matrices are incompatible and the procedure must stop. If a libration axis is absent zero values are assigned to the corresponding coordinates in (15). For presentation purposes it might be useful to assign

$$\mathbf{w}_{[L]x}^{lx} = \tfrac{1}{2}\left(\mathbf{w}_{[L]x}^{ly} + \mathbf{w}_{[L]x}^{lz}\right),\ \mathbf{w}_{[L]y}^{ly} = \tfrac{1}{2}\left(\mathbf{w}_{[L]y}^{lx} + \mathbf{w}_{[L]y}^{lz}\right),$$

$$\mathbf{w}_{[L]z}^{lz} = \tfrac{1}{2}\left(\mathbf{w}_{[L]z}^{lx} + \mathbf{w}_{[L]z}^{lz}\right) \tag{16}$$

which gives the points in the middle of axes.

Knowing the position (15) of the libration axes and elements of $L_{[L]}$ we can calculate an apparent translation due to the displacement of the libration axes from the origin

$$D_{W[L]} = \tag{17}$$

$$\begin{pmatrix} \left(w_{[L]z}^{ly}\right)^2 L_{[L]yy} + \left(w_{[L]y}^{lz}\right)^2 L_{[L]zz} & -w_{[L]x}^{lz}w_{[L]y}^{lz}L_{[L]zz} & -w_{[L]x}^{ly}w_{[L]z}^{ly}L_{[L]yy} \\ -w_{[L]x}^{lz}w_{[L]y}^{lz}L_{[L]zz} & \left(w_{[L]z}^{lx}\right)^2 L_{[L]xx} + \left(w_{[L]x}^{lz}\right)^2 L_{[L]zz} & -w_{[L]y}^{lx}w_{[L]z}^{lx}L_{[L]xx} \\ -w_{[L]x}^{ly}w_{[L]z}^{ly}L_{[L]yy} & -w_{[L]y}^{lx}w_{[L]z}^{lx}L_{[L]xx} & \left(w_{[L]y}^{lx}\right)^2 L_{[L]xx} + \left(w_{[L]x}^{ly}\right)^2 L_{[L]yy} \end{pmatrix}$$

Then we can calculate the residual T matrix after removal of the contribution (17):

$$T_{C[L]} = T_{[L]} - D_{W[L]} \tag{18}$$

that must be positive semidefinite. Matrices $S_{[L]}$ and $L_{[L]}$ are not modified at this step.

4. Determination of the screw components (step C)

4.1. Correlation between libration and vibration and usual choice of S diagonal elements

Next we use the matrices $L_{[L]}$ and $S_{[L]}$ to determine the correlation parameters s_x, s_y, s_z, then remove the corresponding contribution from the T_C matrix (equation (6)) and extract the

matrix $V_{[L]}$ for uncorrelated vibrations. However, there is an ambiguity in the definition of $S_{[L]}$ which is apparent from the observation that the matrices $U_{Cart,n}$ of individual atoms will not change if the same number t is removed simultaneously from all three diagonal elements of $S_{[L]}$. Usually t is obtained by minimizing the trace of the resulting matrix S_C (Schomaker & Trueblood, 1968):

$$S_C(t) = S_{[L]} - \begin{pmatrix} t & 0 & 0 \\ 0 & t & 0 \\ 0 & 0 & t \end{pmatrix} \tag{19}$$

which corresponds to the minimal vibration-libration correlation (Urzhumtsev *et al*. 2013). This unconditioned minimization

$$\min_t |S_C(t)| = \min_t \left| \left(S_{[L]xx} + S_{[L]yy} + S_{[L]zz} \right) - 3t \right| , \tag{20}$$

gives the value

$$t_0 = \tfrac{1}{3} \left(S_{[L]xx} + S_{[L]yy} + S_{[L]zz} \right) = \tfrac{1}{3} tr S_{[L]} \tag{21}$$

However, this value may lead to physically unrealistic matrices for which libration-vibration decomposition is impossible. First, it is clear that the S elements cannot be large in comparison with the elements of the T and L matrices; in particular, if there is no rotation around one of the libration axes, the corresponding S elements must be zero. Second, if the elements of matrix S and the corresponding values s_x, s_y, s_z are too large, the matrix V in (6) may not be positive definite. We have not identified any previous discussion of these complications in the crystallographic literature. The following sections describe a procedure that defines the constraints on the diagonal elements of matrix S when using (19).

4.2. *Cauchy-Schwarz constraints*

Since the diagonal elements of the positive definite matrix V are non-negative in any Cartesian basis, according to equations (5)-(7) the diagonal elements of the matrices must satisfy the Cauchy-Schwarz inequality

$$S_{C,xx}^2 \le T_{C[L]xx} L_{[L]xx} ; S_{C,yy}^2 \le T_{C[L]yy} L_{[L]yy} ; S_{C,zz}^2 \le T_{C[L]zz} L_{[L]zz} \tag{22}$$

(This condition is also apparent from equations (8.5) - (8.7) in Urzhumtsev *et al.* 2013 which give $S_{xx} = \langle d_x u_x \rangle$, $T_{xx} = \langle u_x u_x \rangle$, $L_{xx} = \langle d_x d_x \rangle$ and similar conditions for other axes). In turn, this defines the limits as

$$
\begin{aligned}
\left(S_{[L]xx} - t\right)^2 &\leq T_{C[L]xx} L_{[L]xx} \\
\left(S_{[L]yy} - t\right)^2 &\leq T_{C[L]yy} L_{[L]yy} \\
\left(S_{[L]zz} - t\right)^2 &\leq T_{C[L]zz} L_{[L]zz}
\end{aligned}
\tag{23}
$$

or

$$
t_{min,C} \leq t \leq t_{max,C}
\tag{24}
$$

with

$$
t_{min,C} = \max\left\{ S_{[L]xx} - \sqrt{T_{C[L]xx} L_{[L]xx}} \; ; S_{[L]yy} - \sqrt{T_{C[L]yy} L_{[L]yy}} \; ; S_{[L]zz} - \sqrt{T_{C[L]zz} L_{[L]zz}} \right\}
\tag{25}
$$

$$
t_{max,C} = \min\left\{ S_{[L]xx} + \sqrt{T_{C[L]xx} L_{[L]xx}} \; ; S_{[L]yy} + \sqrt{T_{C[L]yy} L_{[L]yy}} \; ; S_{[L]zz} + \sqrt{T_{C[L]zz} L_{[L]zz}} \right\}
$$

4.3. Positive semidefinition of the V matrix

Now we analyze the case of the matrix V being positive semidefinite. First, we suppose that all diagonal elements of the matrix $L_{[L]}$ are non-zero and note that Section 4.4 considers the alternative case. From equations (5) - (7) and (19) we find the screw contribution

$$
C_{[L]}(t) =
\begin{pmatrix}
\dfrac{S_{C,xx}^2}{L_{[L]xx}} & 0 & 0 \\
0 & \dfrac{S_{C,yy}^2}{L_{[L]yy}} & 0 \\
0 & 0 & \dfrac{S_{C,zz}^2}{L_{[L]zz}}
\end{pmatrix}
=
\begin{pmatrix}
\dfrac{\left(S_{[L]xx} - t\right)^2}{L_{[L]xx}} & 0 & 0 \\
0 & \dfrac{\left(S_{[L]yy} - t\right)^2}{L_{[L]yy}} & 0 \\
0 & 0 & \dfrac{\left(S_{[L]zz} - t\right)^2}{L_{[L]zz}}
\end{pmatrix}
\tag{26}
$$

to be subtracted from matrix (18) as

$$
V_{[L]} = T_{C[L]} - C_{[L]}(t) \qquad .
\tag{27}
$$

Matrix $V_{[L]}$ is positive semidefinite along with

$$
V_\Lambda = \Lambda^\tau V_{[L]} \Lambda = \Lambda^\tau T_{C[L]} \Lambda - \Lambda^\tau C_{[L]}(t) \Lambda = T_\Lambda - C_\Lambda(t)
\tag{28}
$$

12

where

$$\Lambda = \Lambda^{\tau} = \begin{pmatrix} \sqrt{L_{[L]xx}} & 0 & 0 \\ 0 & \sqrt{L_{[L]yy}} & 0 \\ 0 & 0 & \sqrt{L_{[L]zz}} \end{pmatrix} \tag{29}$$

$$T_{\Lambda} = \begin{pmatrix} T_{C[L]xx} L_{[L]xx} & T_{C[L]xy} \sqrt{L_{[L]xx} L_{[L]yy}} & T_{C[L]xz} \sqrt{L_{[L]xx} L_{[L]zz}} \\ T_{C[L]yx} \sqrt{L_{[L]xx} L_{[L]yy}} & T_{C[L]yy} L_{[L]yy} & T_{C[L]yz} \sqrt{L_{[L]yy} L_{[L]zz}} \\ T_{C[L]zx} \sqrt{L_{[L]xx} L_{[L]zz}} & T_{C[L]zy} \sqrt{L_{[L]yy} L_{[L]zz}} & T_{C[L]zz} L_{[L]zz} \end{pmatrix} =$$

$$= \begin{pmatrix} t_{xx} & t_{xy} & t_{xz} \\ t_{yx} & t_{yy} & t_{yz} \\ t_{zx} & t_{zy} & t_{zz} \end{pmatrix} \quad \text{with} \quad t_{xy} = t_{yx}, t_{yz} = t_{zy}, t_{zx} = t_{xz} \tag{30}$$

$$C_{\Lambda}(t) = \begin{pmatrix} \left(S_{[L]xx} - t\right)^2 & 0 & 0 \\ 0 & \left(S_{[L]yy} - t\right)^2 & 0 \\ 0 & 0 & \left(S_{[L]zz} - t\right)^2 \end{pmatrix} =$$

$$= \begin{pmatrix} c_{xx}(t) & 0 & 0 \\ 0 & c_{yy}(t) & 0 \\ 0 & 0 & c_{zz}(t) \end{pmatrix} \tag{31}$$

Necessarily, all diagonal terms of (31) cannot be larger than the maximal eigenvalue $\tau_{\max}$ of (30) giving a condition

$$t_{\min,\tau} = \max\left\{S_{[L]xx}; S_{[L]yy}; S_{[L]zz}\right\} - \sqrt{\tau_{\max}} \leq t \leq \min\left\{S_{[L]xx}; S_{[L]yy}; S_{[L]zz}\right\} + \sqrt{\tau_{\max}} = t_{\max,\tau}. \tag{32}$$

Another obvious condition that these terms are not larger than the minimum eigenvalue $\tau_{\min}$ of (30) is sufficient but not necessary. We have not yet found a direct solution to this problem and thus use the numerical approach described below.

Matrix V_{Λ} is positive semidefinite if and only if all three of the real eigenvalues are non-negative (some of them may coincide with each other). They are the roots of the cubic characteristic equation

$$v^3 + a_S v^2 + b_S v + c_S = 0 \tag{33}$$

with the coefficients

$$a_S(t) = -v_{XX} - v_{YY} - v_{ZZ}, \tag{34}$$

$$b_S(t) = \det\begin{pmatrix} v_{XX} & v_{XY} \\ v_{YX} & v_{YY} \end{pmatrix} + \det\begin{pmatrix} v_{YY} & v_{YZ} \\ v_{ZY} & v_{ZZ} \end{pmatrix} + \det\begin{pmatrix} v_{ZZ} & v_{ZX} \\ v_{XZ} & v_{XX} \end{pmatrix} \tag{35}$$

$$c_S(t) = -\det\begin{pmatrix} v_{XX} & v_{XY} & v_{XZ} \\ v_{YX} & v_{YY} & v_{YZ} \\ v_{ZX} & v_{ZY} & v_{ZZ} \end{pmatrix} \tag{36}$$

The positivity of the roots is equivalent to a system of three inequalities:

$$a_S(t) = \left[(t - S_{[L]xx})^2 - t_{xx} \right] + \left[(t - S_{[L]yy})^2 - t_{yy} \right] + \left[(t - S_{[L]zz})^2 - t_{zz} \right] \le 0 \tag{37}$$

$$b_S(t) = \left[(t - S_{[L]xx})^2 - t_{xx} \right]\left[(t - S_{[L]yy})^2 - t_{yy} \right] + \left[(t - S_{[L]yy})^2 - t_{yy} \right]\left[(t - S_{[L]zz})^2 - t_{zz} \right] +$$
$$+ \left[(t - S_{[L]zz})^2 - t_{zz} \right]\left[(t - S_{[L]xx})^2 - t_{xx} \right] - \left[t_{xy}^2 + t_{yz}^2 + t_{zx}^2 \right] \ge 0 \tag{38}$$

$$c_S(t) = \left[(t - S_{[L]xx})^2 - t_{xx} \right]\left[(t - S_{[L]yy})^2 - t_{yy} \right]\left[(t - S_{[L]zz})^2 - t_{zz} \right] - \tag{39}$$
$$- t_{yz}^2\left[(t - S_{[L]xx})^2 - t_{xx} \right] - t_{xz}^2\left[(t - S_{[L]yy})^2 - t_{yy} \right] - t_{xy}^2\left[(t - S_{[L]zz})^2 - t_{zz} \right] - 2 t_{xy} t_{yz} t_{xz} \le 0$$

where the left parts are polynomials of order 2, 4 and 6 of the parameter t, all with the unit highest coefficient. The first condition (37) defines the interval for t values :

$$t_{min,a} = t_0 - t_a \le t \le t_{max,a} = t_0 + t_a \tag{40}$$

with

$$t_a = \sqrt{t_0^2 + \frac{t_{xx} + t_{yy} + t_{zz}}{3} - \frac{S_{[L]xx}^2 + S_{[L]yy}^2 + S_{[L]zz}^2}{3}} \tag{41}$$

We failed to find analytical expressions corresponding to the two other inequalities and the following numerical procedure was implemented to find the best t value that is physically acceptable:

a) Calculate t_0 value (equation (21));

b) Calculate the interval (t_{min}, t_{max}) for allowed t values as intersection of intervals (24), (32)

and (40); $t_{\min} = \max\{t_{\min,C}, t_{\min,\tau}, t_{\min,a}\}$, $t_{\max} = \min\{t_{\max,C}, t_{\max,\tau}, t_{\max,a}\}$; if $t_{\min} > t_{\max}$ the problem has no solution and the procedure stops;

c) If $t_{\min} = t_{\max}$ we check the conditions $b_S(t_{\min}) \geq 0$, $c_S(t_{\min}) \leq 0$, or directly the condition that V_Λ is positive semidefinite; if the conditions are satisfied we assign $t_S = t_{\min}$ otherwise the problem has no solution and the procedure stops;

d) If $t_{\min} < t_{\max}$ we search numerically, in a fine grid, for the point t_S in the interval $(t_{\min}, t_{\max})$ and closest to t_0 such that $b_S(t_S) \geq 0$, $c_S(t_S) \leq 0$; if at any point of this interval at least one of these inequalities is wrong, then the procedure stops;

e) If steps c and d give us the t_S value, we accept it as the solution of the problem otherwise there is no acceptable t value and the procedure of the *TLS* matrices decomposition stops.

4.4. Singular sets of rotation

When one of the $L_{[L]xx}, L_{[L]yy}, L_{[L]zz}$ values is zero (that is, there is no rotation around the corresponding axis), straightforward use of the standard procedure including (26) becomes impossible. However, in this case the t_S value must equal to $S_{C,xx}$, $S_{C,yy}$ or $S_{C,zz}$, corresponding to the axes with no rotation, turning the corresponding inequality in (25) into an equality and making the corresponding diagonal element in (26) equal to zero; for example if $L_{[L]xx} = 0$ then $t_S = S_{L,xx}$ resulting in $S_{C,xx} = 0$. We simply need to check two other conditions in (22) and the condition that the residual matrix is positive semidefinite, for example calculating (37)-(39). If t_S does not satisfy these conditions, the problem has no solution and the procedure of the *TLS* matrices decomposition stops.

4.5. Screw parameters

For the t_S determined above we calculate the matrix $S_C(t_S)$ (19). From this matrix we obtain the estimates $\bar{s}_x = \dfrac{S_{C,xx}}{L_{[L]xx}}$, $\bar{s}_y = \dfrac{S_{C,yy}}{L_{[L]yy}}$, $\bar{s}_z = \dfrac{S_{C,zz}}{L_{[L]zz}}$ for the screw parameters following the rotation axes currently aligned with the coordinate axes of the basis L. We remind the reader that if

one of the $L_{[L]xx}, L_{[L]yy}, L_{[L]zz}$ values is equal to zero, the corresponding diagonal element of S_C must also be equal to zero (otherwise the matrices are inconsistent with each other and the procedure stops) and we assign the corresponding screw parameter, $\bar{s}_x$, $\bar{s}_y$ or $\bar{s}_z$, to be zero.

5. Determination of the vibration components (step D)

5.1. Matrix V and vibration parameters in L basis

When the t_S value is known, matrix $C_{[L]}(t_S)$ and then $V_{[L]}$ are calculated according to equations (26) - (27). The step of obtaining independent vibrations from the $V_{[L]}$ matrix is similar to that for getting the independent librations from $L_{[L]}$. First, we calculate the three eigenvalues $0 \le \mu_1 \le \mu_2 \le \mu_3$ of matrix $V_{[L]}$ (in practice, all of them are strictly positive). Then we identify three corresponding unit eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ that are orthogonal to each other and assign

$$\mathbf{v}_x = \pm\mathbf{v}_1, \mathbf{v}_y = \mathbf{v}_2, \mathbf{v}_z = \mathbf{v}_3 \tag{42}$$

(the sign for $\mathbf{v_x}$ is taken so that the vectors (42) form a right-hand triad). The three eigenvectors give the directions of the uncorrelated translations and the corresponding eigenvalues μ_1, μ_2, μ_3 are squared *rmsd*s along these axes. These axes define a new orthonormal basis V in which matrix $V_{[L]}$ becomes the diagonal matrix $V_{[V]}$ with elements $V_{[V]xx} = \mu_1$, $V_{[V]yy} = \mu_2$, $V_{[V]zz} = \mu_3$; these values are the estimates $\langle \bar{t}_x^2 \rangle = V_{[V]xx}$, $\langle \bar{t}_y^2 \rangle = V_{[V]yy}$, $\langle \bar{t}_z^2 \rangle = V_{[V]zz}$ of the squared vibration amplitudes.

5.2. Vibration and libration axes in M basis

The libration and vibration amplitudes and screw parameters are independent of the choice of the basis, and the direction of the libration axes (11) is known in the principal M basis. However the directions of the uncorrelated translations $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ that were calculated in section 4 and the points $\mathbf{w}_{[L]}^{lx}$, $\mathbf{w}_{[L]}^{ly}$, $\mathbf{w}_{[L]}^{lz}$ belonging to the libration axes (section 3.2) are given in the L basis.

To obtain the coordinates $\mathbf{w}_{[M]}^{lx}$, $\mathbf{w}_{[M]}^{ly}$, $\mathbf{w}_{[M]}^{lz}$ of these points in the M basis we apply the transformation

16

$$
\begin{pmatrix} w^{lx}_{[M]x} \\ w^{lx}_{[M]y} \\ w^{lx}_{[M]z} \end{pmatrix} = R_{ML} \begin{pmatrix} w^{lx}_{[L]x} \\ w^{lx}_{[L]y} \\ w^{lx}_{[L]z} \end{pmatrix} , \quad \begin{pmatrix} w^{ly}_{[M]x} \\ w^{ly}_{[M]y} \\ w^{ly}_{[M]z} \end{pmatrix} = R_{ML} \begin{pmatrix} w^{ly}_{[L]x} \\ w^{ly}_{[L]y} \\ w^{ly}_{[L]z} \end{pmatrix} , \quad \begin{pmatrix} w^{lz}_{[M]x} \\ w^{lz}_{[M]y} \\ w^{lz}_{[M]z} \end{pmatrix} = R_{ML} \begin{pmatrix} w^{lz}_{[L]x} \\ w^{lz}_{[L]y} \\ w^{lz}_{[L]z} \end{pmatrix} \tag{43}
$$

using (12). Similarly, the direction of the axes $\mathbf{v}_x, \mathbf{v}_y, \mathbf{v}_z$ in the M basis and can be obtained as

$$
\mathbf{v}_{[M]x} = R_{ML}\mathbf{v}_{[L]x} \quad , \quad \mathbf{v}_{[M]y} = R_{ML}\mathbf{v}_{[L]y} \quad , \quad \mathbf{v}_{[M]z} = R_{ML}\mathbf{v}_{[L]z} \tag{44}
$$

5.3. Calculations in V basis

Knowledge of $\mathbf{v}_x, \mathbf{v}_y, \mathbf{v}_z$ (42) in the basis M defines the rotation matrix from the M basis to the V basis as

$$
R_{MV} = \begin{pmatrix} (v_x)_x & (v_y)_x & (v_z)_x \\ (v_x)_y & (v_y)_y & (v_z)_y \\ (v_x)_z & (v_y)_z & (v_z)_z \end{pmatrix} . \tag{45}
$$

The coordinates of a vector expressed in the two bases are linked by the equation

$$
\begin{pmatrix} x_{[M]n} \\ y_{[M]n} \\ z_{[M]n} \end{pmatrix} = R_{MV} \begin{pmatrix} x_{[V]n} \\ y_{[V]n} \\ z_{[V]n} \end{pmatrix} \quad , \quad \begin{pmatrix} x_{[V]n} \\ y_{[V]n} \\ z_{[V]n} \end{pmatrix} = R_{MV}^{-1} \begin{pmatrix} x_{[M]n} \\ y_{[M]n} \\ z_{[M]n} \end{pmatrix} \tag{46}
$$

This step finalizes extracting the parameters of the motions that correspond to the given set of *TLS* matrices. Section 6 provides some examples of this procedure's application to models deposited in the PDB.

6. Examples of *TLS* matrix decomposition

6.1. General procedure

To illustrate the algorithm described above we analyzed several entries selected from the PDB. For each structure, we applied a standard *TLS* refinement protocol from *Phenix* (Adams *et al.*,

2010; Afonine *et al.*, 2012) including automatic determination of the *TLS* groups. During refinement, 20 matrix elements (6 for *T*, 6 for *L* and 8 for *S* where the third diagonal element of *S* has been chosen to make the trace of the matrix equal to 0) were refined independently. The procedure described above was then applied to all sets of the *TLS* matrices obtained. By current conventions, the *L* and *S* matrices for the PDB models are expressed in degrees^2 and in Å·degree, respectively, and first we need to convert them into radians^2 and Å·radians.

6.2. *Synaptotagmin*

The crystals of synaptotagmin III (PDB code 1dqv) contain two copies of the molecule in the asymmetric unit. The structure, when re-refined by *phenix.refine*, has R_{work} = 0.200 and R_{work} = 0.231. The *TLS* refinement, with each molecule taken as a single *TLS* group, reduced *R*-factors to R_{work} = 0.177 and R_{work} = 0.211 and shows the significance of this *TLS* modeling. Table 2 shows the two sets of the matrices, and Table 3 shows the corresponding motion parameters extracted using our routine. For the two groups both vibrations and librations are practically isotropic and are of the same order of magnitude for both groups. Fig. 2a shows the principal axes of these motions.

6.3. *Calmodulin*

The structure of calmodulin (PDB code 1exr) was automatically split into 4 *TLS* groups. This case was an example of possible problems that could be solved by a minimal manual intervention. For the first group, one of the eigenvalues of the matrix *L* was marginally negative, equal to -0.000023; we allowed the program to substitute the value 0 thus accepting this group to be singular with one degenerate libration. The second group also had one degenerate libration. After correcting the slightly negative eigenvalue for the first group, all composited motions were extracted without problems.

For the third group, the screw parameters and axes positions were extremely large, leading to the procedure's inability to find a positive semidefinite $V_{[L]}$. This was caused by the fact that all three eigenvalues of the matrix *L* were practically equal to 0 (0.000000, 0.000008 and 0.000035) resulting in high computational instability. We replaced matrix *L*, and respectively *S*, by zero matrices, defining all librations as absent, and obtained significant values for vibration parameters. This group is a helix held at both ends by large domains, which explains the impossibility of librations within the helix.

Finally, for the fourth group one of the diagonal elements of the matrix *T* was marginally negative. Increase all diagonal elements of the matrix *T* by 0.002 makes this matrix positive

definite. This manipulation can be considered a form of 'borrowing' roughly 0.16 Å^2 from either the overall scale factor or from individual atoms (*i.e.* such that the *T* modification and the simultaneous removing of 0.16 Å^2 from all corresponding ADP parameters does not change the structure factors). An accurate separation of total ADP values into contribution from several sources (for example, Winn *et al*., 2001, 2003; Afonine *et al*., 2013) is a separate ongoing project (Afonine & Urzhumtsev, 2007). This group vibrates essentially in a plane (Fig. 2b) and the principal vibration axis of the group 3 (the helix) is parallel to this plane, leading to the plausible hypothesis that these groups at least partially move together.

6.4. Initiation translation factor 2 (IF2)

The structure of IF2 (PDB code 4b3x) has recently been solved in one of our laboratories (Simonetti *et al*., 2013) with R_{work} = 0.180 and R_{free}= 0.219 at a resolution of 1.95 Å. *TLS* refinement was performed with two groups: the first included the N-terminal and the following long helix, and the second included the rest of structure. This refined model had lower R_{work} = 0.176 and R_{free}= 0.203. For this example, the *TLS* matrices from the first group were not directly interpretable because the residual matrix $V_{[L]}$ was not positive semidefinite having one of its eigenvalues equal to -0.05. Similarly to the last group of the calmodulin, we artificially added 0.06 to all diagonal elements of the matrix *T* which corresponds to 'borrowing' of about 5 Å^2 from individual atoms or from the overall scale factor (as above). With such a correction the procedure worked well. It should be noted that for the first *TLS* group one of the rotations was degenerated and that the assignment *trS* = 0 would make this matrix incompatible with *L*. Table 3 shows that vibrations of this group are essentially anisotropic. Fig. 2c also shows that the libration axes for this group pass quite far away from the molecule, which makes the corresponding rotation similar to a translation. Additionally, the large $\bar{s}_z$ value leads us to believe that the matrix *S* is not well defined to be physically significant. The matrices for the second group were routinely interpreted showing isotropic vibrations and librations.

We then tried to apply the same procedure after manually choosing the *TLS* groups as residues 1-50 (N terminal), 51-69 (helix), 70-333 (G domain) and 343-363 (connector to the C domain absent in this structure). As before, the G domain the matrices were physically interpretable. For the other two groups, after a manual adjustment similar to those discussed above (a slight artificial increasing of the diagonal *T* elements with an accompanying decrease in the residual ADP of the individual atoms), we obtained a pure vibration for the helix (as for the calmodulin case) and a

libration around a single axis for the terminal group. In contrast, we failed to find reasonably small corrections for the matrices of the first group that would make them physically interpretable.

7. Interpretation of *TLS* matrices with an ensemble of models

7.1. *Generation of an explicit set of atomic models with a given variability consistent with TLS*

Some structural problems may explicitly require a set of models that describe a given mobility, *e.g.* corresponding to the *TLS* matrices for harmonic motion. We do not discuss here larger-scale anharmonic motions for which other techniques such as molecular dynamic trajectories have traditionally been used, as originated by McCammon *et al.*, 1977. An example of such a problem is described in the accompanying paper by Van Benschoten *et al.* in which X-ray diffuse scattering data were compared with calculated data corresponding to different types of molecular motion.

As soon as vibration and libration parameters are known, we can build a corresponding set of models explicitly. If a model deposited in the PDB contains *TLS* matrices, the procedure of their decomposition described above can be applied. A decomposition of this motion into three independent vibrations and three independent librations provides the corresponding atomic shifts, the sum of which gives the total displacement.

It is generally more convenient to generate each group of atomic shifts in its own basis: shifts $\Delta^V_{[V]}\mathbf{r}_n$ due to vibration in the V basis and shifts $\Delta^L_{[L]}\mathbf{r}_n$ due to libration in the L basis. Here we are working in a linear approximation such that rotation angles are of order of 0.1 radian. For each particular set of generated shifts, they are the transformed into the M basis as $\Delta^V_{[M]}\mathbf{r}_n$ and $\Delta^L_{[M]}\mathbf{r}_n$ and their sum

$$\Delta_{[M]}\mathbf{r}_n = \Delta^L_{[M]}\mathbf{r}_n + \Delta^V_{[M]}\mathbf{r}_n \tag{47}$$

is applied to the corresponding atoms. Details of the generation of such a model are discussed in the next sections. This procedure is repeated independently many times generating randomly moved models distributed accordingly to the *TLS* matrices.

7.2. *Calculation of the model shift due to libration*

Let us suppose that, using the procedure described above, we know the three mutually orthogonal axes $\boldsymbol{l_x}$, $\boldsymbol{l_y}$, $\boldsymbol{l_z}$ for independent libration and the points that belong to these axes. If the coordinates of these points $\mathbf{w}^{lx}_{[M]}$, $\mathbf{w}^{ly}_{[M]}$, $\mathbf{w}^{lz}_{[M]}$ are given in the molecular basis M we can recalculate them into the L basis $\mathbf{w}^{lx}_{[L]}$, $\mathbf{w}^{ly}_{[L]}$, $\mathbf{w}^{lz}_{[L]}$ using (14). We can do the same for the coordinates of all atoms of the group that are subject to the *TLS* motion: $\left(x_{[L]n}, y_{[L]n}, z_{[L]n}\right)$, $n = 1, 2, \ldots, N$. We note that the squared libration amplitudes $\left\langle d_x^2 \right\rangle = L_{[L]xx} = \lambda_1$, $\left\langle d_y^2 \right\rangle = L_{[L]yy} = \lambda_2$, $\left\langle d_z^2 \right\rangle = L_{[L]zz} = \lambda_3$ (Section 3.2) and the screw parameters s_x, s_y, s_z (estimates for which $\bar{s}_x$, $\bar{s}_y$, $\bar{s}_z$ are defined in Section 4.5) are independent of the chosen basis.

For an atom at a distance $R = 1$ Å from the rotation axis, the probability of the shifts d_x, d_y, d_z, which are numerically equal to the rotation angle in radians, are equal to :

$$\text{axis parallel to } \boldsymbol{l_x}: \quad P(d_x) = \sqrt{2\pi\lambda_1}\,\exp\left(-\frac{d_x^2}{2\lambda_1}\right) \tag{48}$$

$$\text{axis parallel to } \boldsymbol{l_y}: \quad P(d_y) = \sqrt{2\pi\lambda_2}\,\exp\left(-\frac{d_y^2}{2\lambda_2}\right) \tag{49}$$

$$\text{axis parallel to } \boldsymbol{l_z}: \quad P(d_z) = \sqrt{2\pi\lambda_3}\,\exp\left(-\frac{d_z^2}{2\lambda_3}\right) \tag{50}$$

If one of the eigenvalues is equal to 0 then the corresponding d is equal to 0 with unit probability. Let us obtain particular values of d_{x0}, d_{y0}, d_{z0} using a random numbers generator with the normal distribution (48-50).

For the rotation around the axis parallel to $\boldsymbol{l_z}$ and crossing the point $\mathbf{w}^{lz}_{[L]} = \left(w^{lz}_{[L]x}, w^{lz}_{[L]y}, w^{lz}_{[L]z}\right)$, for each atom n described by the vector $\mathbf{r}_n$ we recalculate the coordinates of $\mathbf{r}_n - \mathbf{w}^{lz}_{[L]}$ in the L basis

$$x_{[A]n} = x_{[L]n} - w^{lz}_x; \; y_{[A]n} = y_{[L]n} - w^{lz}_y; \; z_{[A]n} = z_{[L]n} \tag{51}$$

If $\mathbf{r}'_n$ stands for the position of the same atom after rotation by angle d_{z0} around this axis, the coordinates of $\mathbf{r}'_n - \mathbf{w}^{lz}_{[L]}$ can be written as

$$\left(x_{[A]n}\cos d_{z0} - y_{[A]n}\sin d_{z0}\right)\boldsymbol{l_x} + \left(x_{[A]n}\sin d_{z0} + y_{[A]n}\cos d_{z0}\right)\boldsymbol{l_y} + \left(z_{[A]n} + \bar{s}_z d_{z0}\right)\boldsymbol{l_z} \;. \tag{52}$$

This gives the corresponding atomic shift $\mathbf{r}'_n - \mathbf{r}_n$

$$
\begin{aligned}
\Delta_{[L]}^{lz}\mathbf{r}_n &= \left(x_{[A]n}\cos d_{z0} - y_{[A]n}\sin d_{z0} - x_{[A]n}\right)\mathbf{l}_x + \\
&+ \left(x_{[A]n}\sin d_{z0} + y_{[A]n}\cos d_{z0} - y_{[A]n}\right)\mathbf{l}_y + \left(z_{[A]n} + \bar{s}_z d_{z0} - z_{[A]n}\right)\mathbf{l}_z = \\
&= \left[\left(x_{[L]n} - w_x^{lz}\right)(\cos d_{z0} - 1) - \left(y_{[L]n} - w_y^{lz}\right)\sin d_{z0}\right]\mathbf{l}_x + \\
&+ \left[\left(x_{[L]n} - w_x^{lz}\right)\sin d_{z0} + \left(y_{[L]n} - w_y^{lz}\right)(\cos d_{z0} - 1)\right]\mathbf{l}_y + \bar{s}_z d_{z0}\mathbf{l}_z
\end{aligned}
\tag{53}
$$

Similarly for two other axes:

$$
\begin{aligned}
\Delta_{[L]}^{lx}\mathbf{r}_n &= \left[\left(y_{[L]n} - w_y^{lx}\right)(\cos d_{x0} - 1) - \left(z_{[L]n} - w_z^{lx}\right)\sin d_{x0}\right]\mathbf{l}_y + \\
&+ \left[\left(y_{[L]n} - w_y^{lx}\right)\sin d_{x0} + \left(z_{[L]n} - w_z^{lx}\right)(\cos d_{x0} - 1)\right]\mathbf{l}_z + \bar{s}_x d_{x0}\mathbf{l}_x
\end{aligned}
\tag{54}
$$

$$
\begin{aligned}
\Delta_{[L]}^{ly}\mathbf{r}_n &= \left[\left(z_{[L]n} - w_z^{ly}\right)(\cos d_{y0} - 1) - \left(x_{[L]n} - w_x^{ly}\right)\sin d_{y0}\right]\mathbf{l}_z + \\
&+ \left[\left(z_{[L]n} - w_z^{ly}\right)\sin d_{y0} + \left(x_{[L]n} - w_x^{ly}\right)(\cos d_{y0} - 1)\right]\mathbf{l}_x + \bar{s}_y d_{y0}\mathbf{l}_y
\end{aligned}
\tag{55}
$$

The overall shift due to libration (around the three axes) is the sum of shifts (53)-(55)

$$
\Delta_{[L]}^{L}\mathbf{r}_n = \Delta_{[L]}^{lx}\mathbf{r}_n + \Delta_{[L]}^{ly}\mathbf{r}_n + \Delta_{[L]}^{lz}\mathbf{r}_n
\tag{56}
$$

It changes from one atom of the group to another and must be calculated for all atoms of the group with the same (d_{x0}, d_{y0}, d_{z0}) values for a particular instance of the three rotations.

The atomic shift (56) is given in the L basis. To transform this shift from the L basis into the initial M basis, we invert equation (14):

$$
\Delta_{[M]}^{L}\mathbf{r}_n = R_{ML}\Delta_{[L]}^{L}\mathbf{r}_n
\tag{57}
$$

7.3. Calculation of the model shift due to vibration

In the harmonic approximation, the independent vibration shifts t_x, t_y, t_z expressed in the V basis are distributed accordingly to the probability laws:

$$
P(t_x) = \sqrt{2\pi V_{[V]xx}}\exp\left(-\frac{t_x^2}{2V_{[V]xx}}\right) = \sqrt{2\pi\mu_1}\exp\left(-\frac{t_x^2}{2\mu_1}\right)
\tag{58}
$$

$$P(t_y) = \sqrt{2\pi V_{[V]yy}}\, \exp\left(-\frac{t_y^2}{2V_{[V]yy}}\right) = \sqrt{2\pi\mu_2}\, \exp\left(-\frac{t_y^2}{2\mu_2}\right) \qquad (59)$$

$$P(t_z) = \sqrt{2\pi V_{[V]zz}}\, \exp\left(-\frac{t_z^2}{2V_{[V]zz}}\right) = \sqrt{2\pi\mu_3}\, \exp\left(-\frac{t_z^2}{2\mu_3}\right) \qquad (60)$$

Using a random numbers generator, for each model we obtain particular values of t_{x0}, t_{y0}, t_{z0} using (58-60). If one of the eigenvalues μ is equal to zero, the zero value is assigned to the corresponding shift. The overall translational shift, common to all points of the rigid group, is equal to

$$\Delta_{[V]}^{V} \mathbf{r}_n = t_{x0}\mathbf{v}_x + t_{y0}\mathbf{v}_y + t_{z0}\mathbf{v}_z. \qquad (61)$$

In order to obtain this shift in the M basis we calculate, similarly to equation (57),

$$\Delta_{[M]}^{V} \mathbf{r}_n = R_{MV}\Delta_{[V]}^{V} \mathbf{r}_n \qquad (62)$$

7.4. Validation and application to GpdQ

We generated the ensembles produced by alternative *TLS* refinements of the glycerophosphodiesterase GpdQ (Jackson *et al.*, 2007). GpdQ is found in *Enterobacter aerogenes* and contributes to the homeostasis of the cell membrane by hydrolyzing the 3'-5' phosphodiester bond in glycerophosphodiesters. Each dimer contains three distinct domains per monomer: an α/β sandwich fold containing the active site, a domain-swapped active site cap and a novel dimerization domain comprised of dual-stranded antiparallel β-sheets connected by a small β-sheet. Due to the high global *B*-factors and presence of diffuse signal, Jackson and colleagues performed three separate *TLS* refinements to model the crystalline disorder (Fig. 3): *Entire molecule*, *Monomer* and *Sub-domain*. All *TLS* refinement attempts improved the R_{free} values when compared to the standard isotropic *B*-factor refinement; however, there was no significant difference among the final R_{free} values from the three *TLS* groupings.

We hypothesized that the diffuse scattering produced by each *TLS* motion would contain significant differences, as diffuse signal is a direct result of correlated motion. The notion that *TLS* refinement produces unique diffuse signal has been suggested previously (Tickle & Moss, 1999). Physical ensembles of the *TLS* motion, rather than an explicit mathematical description, were required to generate 3D diffuse scattering maps from *phenix.diffuse*. Visual inspection confirmed that the ensemble produced by *phenix.tls_as_xyz* replicated the anisotropic motion predicted by *TLS* thermal ellipsoids (Fig. 4). Additionally, we calculated the structure factors predicted by the original *TLS* refinement *Entire molecule* and compared them to F_{model} values (for example as defined in

Afonine *et al*., 2012) produced by various *phenix.tls_as_xyz* ensemble sizes. The structure factors converged to a global correlation value of 0.965, demonstrating that *phenix.tls_as_xyz* ensembles accurately represent the motions predicted by *TLS* refinement. Physical representation of the underlying motion also revealed that, while two of the *TLS* refinements produced motion with small variances (a necessity within *TLS* theory), the *Subdomain TLS* grouping produced fluctuations that are clearly chemically unreasonable (Fig. 3). Thus, viewing *TLS* refinement in the form of a structural ensemble is a valuable check of the validity of the results, as matrix elements that satisfy the previously described conditions may still produce motions that are clearly implausible.

8. Discussion

While our previous review on the subject (Urzhumtsev *et al*. 2013) describes the calculation of *TLS* matrices from a known set of vibration and libration parameters, this work focuses on the opposite problem of extracting vibration and libration parameters from a given set of *TLS* matrices. The problem is not as simple as it may at first seem because identical motions may be represented by different vibration-libration combinations. As a consequence, the matrix S is not uniquely defined. However, there remains a necessary consistency between the S matrix and the T and L matrix elements. Because current structure refinement programs ignore this constraint, some definitions of S may result in *TLS* matrices that cannot be interpreted in terms of physically meaningful motions.

The detailed algorithm for obtaining molecular vibration and libration parameters makes decomposition of the *TLS* matrices into ensemble models possible. The constraints on the matrices can also be used to identify 'non physical' combinations of *TLS* matrices. Beyond the well-known conditions of non-negativity for the eigenvalues of T and L, we also discuss the conditions that relate matrices, which is crucial for ensuring that the *TLS* refinement corresponds to physically meaningful motions. Table 1 sheds more light on how well these conditions are verified for the full collection of PDB structures. For approximiately 4,500 *TLS* sets (roughly 700 PDB entries), correcting the diagonal elements of the S matrix as described previously (instead of the standard requirement making its trace equal to zero) corrects the underlying problems. However, even after this correction is applied, however, there still remain a significant number of PDB entries with unphysical TLS matrices.

We suggest that the problem of obtaining physically meaningless *TLS* matrices and the need for their post-refinement correction may be entirely eliminated if *TLS* refinement were to use

vibration and libration parameters directly as refined variables. The mathematical basis and computation algorithms for such a refinement is under active development and will be discussed elsewhere. However, eliminating physically meaningless rigid body motions may increase the *R*-factors for some structures. The procedures for analysis and validation of *TLS* parameters, as well as the algorithm for generating a set of models from given libration and vibration parameters, are implemented in the *Phenix* suite as *phenix.tls_analysis* and *phenix.tls_as_xyz*, respectively and are available starting with version dev-1890.

Acknowledgements

References

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C. & Zwart, P.H. (2010). *Acta Cryst., D***66**, 213-221.

Afonine, P., Urzhumtsev, A. (2007). *CCP4 Newsletter on Protein Crystallography,* **45**, http://www.ccp4.ac.uk/newsletters/newsletter45/articles/

Afonine, P.V., Echols, N., Grosse-Kunstleve, R.W., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T., Urzhumtsev, A., Zwart, P.H., Adams, P.D. (2012). *Acta Cryst.,* **D68,** 352-367.

Afonine, P.V., Grosse-Kunstleve, R.W., Adams, P.D., Urzhumtsev, A. (2013). *Acta Cryst.,* **D69***, 625-634

Brenner, H. (1967). *J.Colloid Interface Chem.*, **23**, 407-435.

Cruickshank, D.W.J. (1956). *Acta Cryst.* **9**, 754-756.

Goldstein, H. *Classical Mechanics*. **1950**, Cambridge, Massachusetts: Addison-Wesley.

Howlin, B., Moss, D.S. & Harris, G.W. (1989). *Acta Cryst.*, A**45**, 851-861.

Howlin, B., Butler, S.A., Moss, D.S., Harris, G.W. & Driessen, H.P.C. (1993). *J.Appl. Cryst.*, **26**, 622-626.

Jackson, C. J., Carr, P. D., Liu, J. W., Watt, S. J., Beck, J. L. & Ollis, D. L. (2007). *J. Mol. Biol.*, **367**, 1047-1062.

Johnson, C.K. (1970). In *Crystallographic Computing*, ed. F.R.Ahmed, Munksgaard, Copenhagen, **1970**, 220-226.

Kuriyan J. & Weis W.I. (1991) *Proceedings of the National Academy of Sciences*, **88**, 2773-2777.

McCammon, J.A., Gelin, B.R. & Karplus, M. (1977). *Nature*, **267**, 585-590.

Painter, J. & Merritt, E.A. (2005). *Acta Cryst.*, D**61**, 465-471.

Painter, J. & Merritt, E.A. (2006a). *Acta Cryst.*, D**62**, 439-450.

Painter, J. & Merritt, E.A. (2006b). *J.Appl. Cryst.*, **39**, 109-111.

Scheringer, C. (1973). *Acta Cryst.* A**29**, 554-570.

Schomaker, V. & Trueblood, K.N. (1968). *Acta Cryst.* B**24**, 63-76.

Schomaker, V. & Trueblood, K.N. (1998). *Acta Cryst.* B**54**, 507-514.

Tickle, I. & Moss, D.S. (1999). Notes from IUCr Cryst.Computing School, *http://public-1.cryst.bbk.ac.uk/~tickle/iucr99/iucrcs99.htm.*

Urzhumtsev, A., Adams, P.D. & Afonine, P. (2013). *Crystallography reviews,* **19**, 230-270.

Winn, M.D., Isupov, M.N. & Murshudov, G.N. (2001). *Acta Cryst.*, **D**57**,** 122-133.

Winn, M.D., Murshudov, G.N. & Papiz, M.Z. (2003). *Methods in Enzymology.*, **374**, 300-321.

Zucker, F., Champ, P.C. & Merritt, E.A. (2010). *Acta Cryst.*, **D**66, 889-900.

Table 1. Number of PDB entries with at least one unphysical TLS matrix condition. Note: when one of the conditions was found to be broken, the other conditions were not checked (Figure 1). The conditions are (from left to right) : ($T{\geq}0$ & $L{\geq}0$) matrices T and L are positive semidefinite; (s=0 & w=0) for the L matrix with one degenerate libration the corresponding elements of the S matrix are equal to 0; ($S{\leq}TL$) elements of the S matrix are limited by the corresponding elements of the T and L matrices following the Cauchy conditions (23); ($TC{\geq}0$) matrix T is positive semidefinite after the contribution due to the displacement of libration axes is removed; ($V{\geq}0$) residual V matrix is positive semidefinite. Right columns summarize the total number of the TLS sets with one of the condition broken and the number of the PDB entries concerned. The two lines shows the statistics for the original matrices in the PDB with the default condition $trS = 0$ (upper line) and after the optimal choice of the diagonal S elements whenever possible as described in Sections 3-4 (bottom line). The total number of the PDB entries is 104,633.

	total PDB number	total TLS number	conditions			broken		total TLS broken	total PDB broken
			$T{\geq}0$ & $L{\geq}0$	s=0 & w=0	$S{\leq}TL$	$TC{\geq}0$	$V{\geq}0$		
$t_S = 0$	25284	197920	66438	13824	1271	43934	7941	133408	22263
$best\ t_S$	25284	197920	66438	13824	0	43090	5699	129051	21540

Table 2. Examples of the *TLS* matrices.

PDB code	chain, residue number	T (Å^2)			L (degree^2)			S (Å·degree)		
1dqv	A1–A97	0.1777	0.0090	−0.0044	1.4462	−0.0160	−0.2656	0.0467	−0.0523	0.0566
		0.0090	0.1306	0.0019	−0.O160	1.2556	0.4713	0.1010	0.0032	−0.0164
		−0.0044	0.0019	0.1372	−0.2656	0.4713	0.8689	0.0090	0.0188	0.0560
	B1–B97	0.1777	0.0090	−0.0044	1.4462	−0.0160	−0.2656	0.0467	−0.0523	0.0566
		0.0090	0.1306	0.0019	−0.O160	1.2556	0.4713	0.1010	0.0032	−0.0164
		−0.0044	0.0019	0.1372	−0.2656	0.4713	0.8689	0.0090	0.0188	0.0560
4b3x	A1–A65	0.4663	0.0991	−0.0764	0.4738	0.0063	0.2318	0.0391	−0.0307	−0.4316
		0.0991	0.5443	−0.0321	0.O063	0.2120	−0.0584	0.0587	0.1786	−0.2003
		−0.0764	−0.0321	0.5001	0.2318	−0.0584	0.1312	0.3665	0.4293	0.0403
	A66–A363	0.1649	−0.0259	0.0184	0.8808	−0.0912	−0.1736	−0.0345	0.0102	−0.0661
		−0.0259	0.1422	0.0055	−0.O912	0.9522	0.0972	0.1159	−0.0222	0.0999
		0.0184	0.0055	0.2028	−0.1736	0.0972	1.6563	0.0424	−0.1330	−0.0237

Table 3. Examples of parameters of the elementary motions found form the decomposition of the *TLS* matrices.

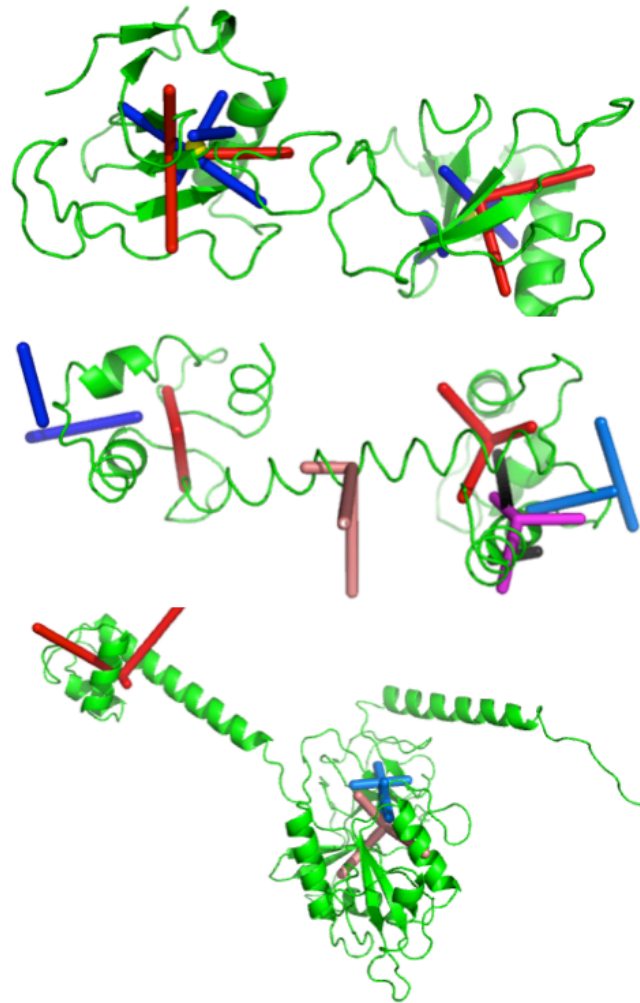| PDB code	chain, residue number	$T: t_x, t_y, t_z$ (Å)			$L: d_x, d_y, d_z$ (rad)			$S: s_x, s_y, s_z$ (Å)			trS
1dqv	A1–A97	.3455	.3671	.4172	.01239	.02044	.02273	1.343	1.137	−1.319	0
	B1–B97	.3634	.3885	.4166	.01608	.01753	.03069	0.679	−1.177	0.200	0
1exr	A2–A30	.1944	.2663	.2870	.00000	.01602	.02182	0.000	2.951	3.408	>0
	A31–A74	.2110	.2939	.3068	.00000	.00860	.01637	0.000	−18.14	−5.028	<0
	A75–A84	.1692	.4906	.6598	.00000	.00000	.00000	0.000	0.000	0.000	0
	A85–A147	.0002	.2270	.3078	.00553	.01418	.02109	20.83	0.800	−1.672	≈0
4b3x	A1–A65	.0994	.6064	.7116	.00000	.00825	.01343	0.000	2.718	−11.05	<0
	A66–A363	.3306	.4102	.4413	.01568	.01720	.02283	3.164	−2.276	−0.197	0

Figure 1. General flowchart of the *TLS* decomposition into libration and vibration composite motions. Yellow ellipses are for conditions to the verified. Green rectangles are for the output parameters of the composite motions. Letters A-E indicate different steps of the procedure as described in the text.

Figure 2. Examples of the vibration-libration ensembles. Red / salmon / magenta sticks indicate the principal vibration axes with the origin in the centre of the group; blue / marine / black sticks are for the libration axes. Yellow spheres for the 1dqv model show the reaction centers. a) 1dqv model. b) 1exr model; note pure vibrations for the group 3 (the helix) and absence of one of libration axes for the groups 1 and 2. c) 4b3x model. Libration axes for the first group are not shown being too far from the molecule.

Figure 3. GpdQ TLS ensembles. The GpdQ *TLS* groups are projected onto the protein structure. The corresponding ensembles produced by *phenix.tls_as_xyz* are shown below. Each *TLS* PDB ensemble is shown as a single asymmetric unit outlined by the unit cell. An increase in overall motion is apparent going from left to right. The 20 member ensemble is shown for visual simplicity.

Figure 4. *phenix.tls_as_xyz* ensembles replicate *TLS* anisotropic motion. a) GpdQ backbone with thermal ellipsoid representation of "entire molecule" *TLS* anisotropic *B*-factors. b) *phenix.tls_as_xyz* ensemble backbones produced from "entire molecule" *TLS* refinement. c) Complete electron density predicted by "entire molecule" *TLS* refinement. d) Global correlation coefficient between experimental structure factor amplitudes F_{obs} of the original GpdQ 'entire motion' refinement and *phenix.tls_as_xyz* ensembles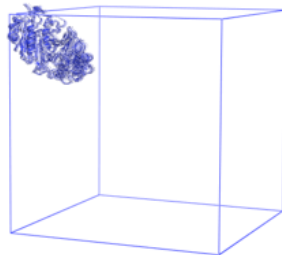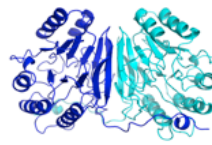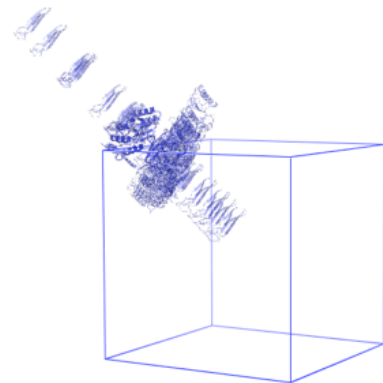 of vario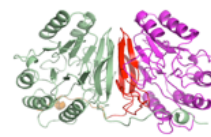us sizes. Convergence values plateau at 0.935.