Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Optimal description of a protein structure in terms of multiple groups undergoing TLS motion

Jay Painter and Ethan A. Merritt

Biomolecular Structure Center, Department of Biochemistry, University of Washington, Seattle, WA 98195-7742, USA

Received 14 December 2005
Accepted 14 February 2006

A single protein crystal structure contains information about dynamic properties of the protein as well as providing a static view of one three-dimensional conformation. This additional information is to be found in the distribution of observed electron density about the mean position of each atom. It is general practice to account for this by refining a separate atomic displacement parameter (ADP) for each atomic center. However, these same displacements are often described well by simpler models based on TLS (translation/libration/screw) rigid-body motion of large groups of atoms, for example interdomain hinge motion. A procedure, *TLSMD*, has been developed that analyzes the distribution of ADPs in a previously refined protein crystal structure in order to generate optimal multi-group TLS descriptions of the constituent protein chains. *TLSMD* is applicable to crystal structures at any resolution. The models generated by *TLSMD* analysis can significantly improve the standard crystallographic residuals R and $R_{\rm free}$ and can reveal intrinsic dynamic properties of the protein.

1. Introduction

1.1. Dynamic information from a static structure

Although crystallography is generally thought of as providing only a static snapshot of the molecules in the crystal, in fact it is also possible to extract a significant amount of information about dynamic properties. A crystal structure may be viewed as a blurred snapshot in which the blurring highlights portions of the molecule that are in motion and indicates which way they are moving. The underlying physical basis of this blurring is a combination of at least two effects (Bürgi, 2000). Dynamic disorder arises from residual thermal vibration and at low temperature is primarily relevant to the blurred view of individual atoms or small groups of atoms such as a single amino-acid side chain. Static disorder is relevant to larger-scale motions and arises when individual molecules within the crystal lattice are frozen into different microconformations corresponding to states sampled along a trajectory of motion. The relative importance of these contributions to diffraction by real crystals differs greatly between small-molecule crystallography and macromolecular crystallography. Small molecules pack tightly into the crystal lattice, leaving little if any room for large-scale vibrational modes in the crystalline state. By contrast, protein crystals contain anywhere from 10 to 90% solvent, with 40–60% being typical. This means that the protein molecules within the crystal lattice are relatively loosely packed. They are correspondingly less constrained by lattice packing than crystalline small molecules. Even in the crystalline form, protein domains or subunits may be free to flex relative to each other,

doi:10.1107/S0907444906005270

secondary-structural elements may exhibit local displacements, and individual side chains may exhibit torsional flexibility.

These various hierarchies of vibrational freedom can be modeled by describing the protein as consisting of multiple approximately rigid groups, each undergoing TLS (translation/libration/screw) displacement (Schomaker & Trueblood, 1968). We show here that such multi-group TLS models of protein structure can yield improved residuals R and $R_{\rm free}$ compared with conventional crystallographic isotropic or anisotropic refinement. Moreover, such models may reveal intrinsic biologically relevant properties of the protein such as the deformability and local flexibility of a ligand-binding site.

1.2. Crystallographic models of flexibility and vibrational motion

Conventional crystallographic models account for the effects of both dynamic and static disorder on X-ray diffraction by introducing additional parameters describing the probabilistic displacement of each atom about its mean position in the lattice (Willis & Pryor, 1975; Trueblood *et al.*, 1996). These atomic displacement parameters (ADPs) can be either isotropic or anisotropic. The isotropic form is familiarly encountered as a correction to the atomic scattering factor $f^2 = f_0^2 \exp(-B_{\rm iso} \sin^2 \theta/\lambda^2)$, where f_0 is the scattering factor for an atom at rest and f is the corrected form that accounts for displacement. The isotropic ADP describes a spherical Gaussian and in the present context it is useful to realise that $B_{\rm iso} = 8\pi^2\langle u^2\rangle$, where $\langle u^2\rangle$ is the mean-square positional displacement. Similarly, the anisotropic correction is a 3 × 3 symmetric tensor U^{ij} describing a trivariate Gaussian probability density function for the location of the atomic center. $B_{\rm iso}$ is commonly called a 'thermal parameter' and the real-space representation of U^{ij} is commonly called a 'thermal ellipsoid'. This is misleading, however, as the ADP describes both dynamic (thermal) and static contributions to the total displacement (Trueblood *et al.*, 1996; Bürgi, 2000).

If modeled separately for each atom, the isotropic form requires a total of four model parameters per atom $(x, y, z, B_{\rm iso})$, while the anisotropic form requires a total of nine model parameters per atom $(x, y, z, U^{11}, U^{22}, U^{33}, U^{12}, U^{13}, U^{23})$. This is an important point, because only very high resolution diffraction experiments provide enough observations to support refinement of a model containing nine parameters per atom. ADPs, either isotropic or anisotropic, are by definition a property of individual atoms. They are therefore a better model for very localized vibrational motion, *e.g.* along-bond vibration, than for large-scale motions involving many atoms acting in concert. The use and interpretation of anisotropic ADPs associated with individual atoms is well established in small-molecule crystallography, where local vibrational modes dominate. The atomic U^{ij} terms can be used, for instance, to correct for the effect of vibrational modes on apparent bond lengths (Dunitz *et al.*, 1988).

TLS (translation/libration/screw) models constitute an alternative way of describing vibrational modes. Just as for individual ADPs, TLS parameterization of atomic displacement in a crystal structure does not strictly speaking describe actual motion; the description encompasses both dynamic displacement (thermal motion) and static displacement (trapped microconformers). For practical purposes, both of these contributions yield the same probabilistic positional variance about the refined mean coordinates.

In the TLS formalism, rigid-body displacement of an arbitrary set of atoms is described by a set of 20 parameters (Schomaker & Trueblood, 1968). These parameters constitute three 3 × 3 tensors: **T**, **L** and **S**. **T** is a symmetric tensor with elements given in units of Å^2; it describes the anisotropic translational displacement common to all atoms in the rigid-body group. **L** is also a symmetric tensor with elements in units of rad^2; it describes the rotational component (libration) of the rigid-body displacement. The **S** tensor is not usually symmetric; it describes the correlation between the rotation and translation of a rigid body undergoing rotation about three orthogonal axes that do not intersect at a common point.

For small amplitudes of vibration, the locus of points visited by any given atom in the vibrating group can be approximated as a three-dimensional Gaussian, *i.e.* a 'thermal ellipsoid' corresponding to some set of U^{ij} for that atom. Let this approximation be called $U^{ij}_{\rm TLS}$. For an atom located at point (x, y, z) with respect to the TLS tensor origin, the six unique components $U^{ij}_{\rm TLS}$ for that atom may be calculated using equations equivalent to those of Schomaker & Trueblood (1968),

$$\begin{aligned}
U^{11}_{\rm TLS} &= L^{22}z^2 + L^{33}y^2 - 2L^{23}yz + 2S^{21}z - 2S^{31}y + T^{11},\\
U^{22}_{\rm TLS} &= L^{11}z^2 + L^{33}x^2 - 2L^{13}xz - 2S^{12}z + 2S^{32}x + T^{22},\\
U^{33}_{\rm TLS} &= L^{11}y^2 + L^{22}x^2 - 2L^{12}xy - 2S^{23}x + 2S^{13}y + T^{33},\\
U^{12}_{\rm TLS} &= -L^{33}xy + L^{23}xz + L^{13}yz - L^{12}z^2\\
&\quad + (S^{22} - S^{11})z + S^{31}x - S^{32}y + T^{12},\\
U^{13}_{\rm TLS} &= -L^{22}xz + L^{23}xy - L^{13}y^2 + L^{12}yz\\
&\quad + (S^{11} - S^{33})y + S^{23}z - S^{21}x + T^{13},\\
U^{23}_{\rm TLS} &= -L^{11}yz - L^{23}x^2 + L^{31}xy + L^{12}xz\\
&\quad + (S^{33} - S^{22})x + S^{12}y - S^{13}z + T^{23}.
\end{aligned} \tag{1}$$

Thus, a single set of 20 TLS parameters describes the individual displacements of all atoms within the vibratory group. For any group larger than three atoms, the TLS description is more parsimonious than describing each atom separately by six ADP parameters U^{ij}.

1.3. *Post hoc* generation of TLS models from existing refinements

The use of TLS models in crystallography has arisen in two distinct contexts. In small-molecule crystallography, TLS models are generated *via post hoc* analysis of a refined structure in order to explore whether the individual refined atomic displacements may be explained as concerted vibration of a larger group of atoms (Schomaker & Trueblood, 1968; He & Craven, 1993). For example, it may be that the individual thermal ellipsoids refined for the atoms in a six-membered

Table 1
Notation used in this work.

ADP	Atomic displacement parameter.
$U_{\rm iso}$	Isotropic ADP as it would result from direct refinement against diffraction data by conventional isotropic refinement. $B_{\rm iso} = 8\pi^2 U_{\rm iso}$.
$U_{\rm TLS}$	Estimated component of an isotropic ADP arising from TLS motion. The work reported in this paper used $U_{\rm TLS}$ = (1/3)trace($U^{ij}_{\rm TLS}$), but other estimates are possible.
$U_{\rm obs}$	Net isotropic ADP from refinement against diffraction data. For conventional isotropic refinement, $U_{\rm obs} = U_{\rm iso}$. In the presence of explicit TLS models, $U_{\rm obs} = U_{\rm iso} + U_{\rm TLS}$.
$U^{ij}_{\rm obs}$	Anisotropic ADP refined directly against diffraction data using conventional anisotropic refinement.
$U^{ij}_{\rm TLS}$	Anisotropic ADP derived by applying a particular set of TLS parameters to a particular atom (1).
$U_{\rm TLSiso}$	Isotropic ADP derived by applying a restricted set of TLS parameters to a particular atom (3).
$U_{\rm eq}$	Isotropic approximation to an anisotropic ADP U^{ij}, defined as $U_{\rm eq}$ = (1/3)trace(U^{ij}) (Trueblood *et al.*, 1996).

ring are adequately explained by a rocking motion of the entire ring about some axis. In this context, the TLS model serves as a simplified description approximating a much more complicated underlying model that has already been refined against atomic resolution data. The same sort of *post hoc* analysis may be applied to protein structures in order to identify portions of the model that act approximately as rigid bodies. Sternberg and coworkers first showed that a single-group TLS model can be used to approximate the distribution of isotropic *B* values observed in a previously refined structure of hen egg-white lysozyme (Sternberg *et al.*, 1979). This approach was later extended by Kuriyan and Weis, who fitted a separate TLS group to each monomer of trimeric hemagglutinin at 3 Å resolution and fitted separate TLS groups to two previously identified subdomains of glutathione reductase at 1.5 Å (Kuriyan & Weis, 1991). These studies demonstrated that the observed distribution of *B* values within a structure can be partially explained as arising from bulk displacement of the protein, subunits or subdomains. However, they did not address the question of how one might in general identify such approximately rigid groups directly from a refined structure.

1.4. Use of TLS models in crystallographic refinement

The direct refinement of a TLS model against crystallographic data, as opposed to constructing it *post hoc* from a pre-refined model, is a distinct case. Perhaps surprisingly, this is an easier computational task. Refinement of TLS models against protein crystallographic data was first introduced in the programs *RESTRAIN* and *TLSANL* (Driessen *et al.*, 1989; Howlin *et al.*, 1993). These programs were used to model domain motion in protein structures refined at modest resolution (2.5 Å) (Moss *et al.*, 1996; Papiz & Prince, 1996). Howlin and coworkers refined a model of ribonuclease A at 1.45 Å consisting of one TLS group per side chain (Howlin *et al.*, 1989). This ability to refine TLS parameters was later incorporated into the primary *CCP*4 refinement program *REFMAC*5 (Murshudov *et al.*, 1999; Winn *et al.*, 2001). Because *REFMAC*5 is in widespread use, the use of simple TLS models increased greatly at that point. Roughly 1000 structures deposited with the PDB since 2000 have incorporated TLS models refined in *REFMAC*5. Most of these refinements have introduced only a single TLS group into the model or, in the case of oligomers, one TLS group per polypeptide chain. This can lead to a significant improvement in the crystallographic residuals R and $R_{\rm free}$. However, we will show here that appropriate choice of a larger number of TLS groups can yield substantial additional improvement in the crystallographic residuals at all resolutions. The obviously missing piece is a procedure for identifying which parts of a protein may usefully be modeled as belonging to any one TLS group.

1.5. Identification of TLS groups within a refined structure

When the starting point for analysis is a high-resolution crystal structure that has been refined with anisotropic ADPs, then one possible approach to identifying approximately rigid groups is to use Rosenfield's rigid-body postulate (Rosenfield *et al.*, 1978). This postulate states that if the ADPs for any two atoms are entirely a consequence of their joint membership of a larger rigid-body group, then clearly the two ADPs should have the same integrated displacement probability along the vector joining them. So in principle one can search for groups of atoms in a refined structure whose ADPs all agree with each other when considered pairwise. This test was first implemented for small molecules in the programs *THMB* and *THMI* (Trueblood, 1978). More recently, it was revisited by Winn for use with protein structures in the program *ANISOANL* (Winn *et al.*, 2001). Several applications to specific proteins have been reported (Yousef *et al.*, 2002; Wilson & Brunger, 2000; Bernett *et al.*, 2004), but in general it is very difficult to delineate multiple self-consistent almost rigid groups above the background noise of non-rigid contributions to the ADPs and of imperfect refinement. The observed displacements of individual protein atoms will always contain significant components which deviate from the rigid body ideal owing to the fact that vibrations and conformational changes in proteins are at best only approximately rigid-body displacements.

Furthermore, most macromolecular crystals do not diffract well enough to permit refinement of individual atomic anisotropic thermal parameters in the first place. Therefore, a different approach is needed which can handle structures refined with isotropic ADPs. An additional concern is that in most cases we have little prior knowledge of how well any TLS model will describe the protein structure at hand. This is highly dependent on the quality of the crystal lattice and on the nature of conformational flexibility uniquely characteristic of this protein. Therefore, it is reasonable to formulate the search for possible component groups of a multi-group TLS model as an optimization problem without any preconceived requirements on the quality of the fit. With this in mind, we have developed an optimization algorithm which, for a given protein chain in a crystal structure, finds the optimal partition of the chain into any desired number of contiguous TLS groups along its amino-acid sequence.

2. Algorithmic methods

The notation used in this work is given in Table 1.

2.1. Fitting a TLS model to an existing set of anisotropic ADPs

The 20 parameters of a TLS model may be fitted to the refined anisotropic ADPs of a group of atoms. By treating atomic positions as fixed, the quadratic terms in (1) become constants and the expressions relating TLS parameters and ADPs become linear. In matrix form, solving for the TLS parameters is accomplished by solving the matrix equation $Ax = b$, where A is a matrix containing six rows for each atom and 20 columns corresponding to each TLS parameter, x is a column vector of the 20 unknown TLS parameters and b is a column vector containing the six terms $U^{ij}_{\rm obs}$ of the observed anisotropic ADP. The elements of A are a function of each atom's position (x, y, z) from an arbitrary origin.

To uniquely determine the 20 TLS parameters, there must be U^{ij} terms for at least four atoms in each group. However, these terms may contain non-TLS contributions and will inevitably contain noise. Therefore, it is desirable to include enough atoms in the group so that A is overdetermined, allowing any noise and non-rigid contributions to the ADPs to average out. Even so, the matrix is occasionally ill-conditioned. When this is the case, some methods of solving for A^{-1} such as LU decomposition will fail and the most reliable method of solving for x is by singular value decomposition. Any column degeneracies which exist in A are detected as small singular values which can be filtered out. *TLSMD* uses double precision variables for A and filters all singular values which are smaller than 1×10^{-12} of the largest singular value. The *LAPACK* subroutine *DGESDD* is used to perform the singular value decomposition (Anderson *et al.*, 1999).

Since the ADPs of refined crystal structures rarely include experimental standard deviations, it is not possible to weight the contribution of individual ADPs to the minimized residual using the standard weighting $w = 1/\sigma^2$. Therefore, *TLSMD* defaults to using unit weights for all atoms. However, there is a strong correlation between error and overall ADP magnitude, so we have also implemented an empirical weighting $w = U_{\rm eq}^{-1}$ for each atom. This has the effect of down-weighting the contribution of side chains that are poorly ordered and of side chains with substantial displacement contributions from vibration about internal torsion angles. In either case, the weight calculated for each atom is further multiplied by the occupancy of that atom.

The anisotropic ADPs calculated from the resulting TLS parameters may then be compared with the target ADPs and the accuracy of the calculated ADPs assessed. For anisotropic ADPs, the weighted least-squares residual for a group of k atoms is given by

$$R_{\rm TLSaniso} = \frac{\sum w_k (U^{ij}_{k,\rm obs} - U^{ij}_{k,\rm TLS})^2}{\sum w_k}. \qquad (2)$$

2.2. Fitting a TLS model to an existing set of isotropic ADPs

Fitting TLS parameters to isotropic ADPs requires a modified form of (1) that predicts isotropic ADPs instead of anisotropic ADPs. The modified form contains only ten of the 20 parameters of the anisotropic TLS description. The first three equations in (1) can be combined with the definition of $U_{\rm eq}$ to yield (Sternberg *et al.*, 1979)

$$\begin{aligned} U_{\rm TLSiso} = {} & T_{\rm iso} + \tfrac{1}{3}[L^{11}(y^2 + z^2) + L^{22}(x^2 + z^2) + L^{33}(x^2 + y^2) \\ & - 2L^{12}xy - 2L^{13}xz - 2L^{23}yz + 2S^1 z + 2S^2 y + 2S^3 x]. \end{aligned} \qquad (3)$$

In this equation, the off-diagonal elements of the $\mathbf{S}$ tensor appear only as the differences $S^1 = S^{21} - S^{12}$, $S^2 = S^{13} - S^{31}$ and $S^3 = S^{32} - S^{23}$. The anisotropic description of TLS translation described by the $\mathbf{T}$ tensor also collapses to a single isotropic parameter, $T_{\rm iso}$. The design matrix A for the isotropic TLS model contains one row per atom and ten TLS parameter columns. The isotropic TLS model can correctly calculate the value of $U_{\rm iso}$ for rigid bodies and therefore the least-squares residual

$$R_{\rm TLSiso} = \frac{\sum w_k (U_{k,\rm obs} - U_{k,\rm TLSiso})^2}{\sum w_k} \qquad (4)$$

yields the correct quality-of-fit estimate for a group of k atoms as a rigid body. However, parameters which are required for interpreting the TLS parameters as rigid-body screw and translational displacements are missing from the isotropic TLS model. This is an intrinsic problem with *post hoc* fitting of a TLS description to purely isotropic ADPs, but it can be remedied by the introduction of external constraints. It can also be remedied by taking the TLS model back into crystallographic refinement, where the full anisotropic set of TLS parameters is refined (Winn *et al.*, 2001).

2.3. Constrained refinement of TLS parameters

If the set of equations (1) is used to solve for TLS parameters without additional constraint, the resulting solution may lie anywhere in a 20-dimensional space. However, only a portion of this space corresponds to true rigid-body motion. Therefore, an unconstrained fit of TLS parameters to noisy data often yields a solution that does not describe a physically plausible motion. This is exemplified by considering the eigenvalues of the $\mathbf{L}$ tensor. These represent the group's mean-square rotational displacement about three orthogonal axes and should always be ≥ 0. If an eigenvalue of $\mathbf{L}$ is negative, this has the physical interpretation that atoms further from the libration axis move less than atoms close to the libration axis, which clearly violates the assumption of rigid-body motion used to derive the TLS description. This is not a major problem if the TLS model is used purely in the context of structure refinement, since the B factors and positions of the individual atoms are still adequately restrained (Winn *et al.*, 2001). However, the validity of the refined TLS model parameters as a description of component vibrational modes within the protein is called into question.

Therefore, when searching for optimal multi-group TLS descriptions of a previously refined protein, we originally discarded any descriptions which contain groups with negative **L** eigenvalues. This was unsatisfactory, however, because retrospective analysis showed that the assignment of residues to potential TLS groups was often reasonable even though the principal components of **L** had been poorly determined. We have therefore reparameterized the original Schomaker–Trueblood equations to express all components of **L** as arising from the square of three orthogonal principal components (u, v, w) and a rotation matrix $\mathbf{R}_{\alpha\beta\gamma}$ defined by Euler angles (α, β, γ),

$$\mathbf{L} = \mathbf{R}_{\alpha\beta\gamma}\begin{pmatrix} u^2 & 0 & 0 \\ 0 & v^2 & 0 \\ 0 & 0 & w^2 \end{pmatrix}\mathbf{R}^T_{\alpha\beta\gamma}. \tag{5}$$

The refinement of a TLS model can be reformulated in terms of these new parameters rather than directly refining the components of **L**. This strictly constrains **L** to describe a true rigid-body librational displacement. The change in variables also transforms minimization into a non-linear problem, but minimization is still straightforward using the Levenberg–Marquardt algorithm. An alternative constraint method applied directly to refinement of the off-diagonal terms of **L** was suggested by Winn *et al.* (2001). Ideally, similar constraints should be developed for the **S** and **T** tensors as well.

2.4. Optimal partitioning as a graph-minimization problem

The *TLSMD* algorithm finds the optimal one-group, two-group, three-group, ..., p-group partition of a single protein chain into multiple TLS groups. The partitioning is performed by breaking the chain into non-overlapping contiguous segments that together span the entire protein chain. For this partitioning scheme, the problem can be restated as follows: given a set S, the set of all possible segments of a protein chain, select p contiguous non-overlapping segments from S which span the chain such that the sum of the TLS fit residuals for these segments is lower than the sum for any other choice of segments. This selection requires that the set S be constructed and that each segment in S be individually fitted with TLS parameters to yield a TLS fit residual. It is also desirable to impose a lower limit m on the number of residues in any given segment to ensure there are enough data points from ADPs to completely determine the TLS parameters. With this constraint, the number of segments $s(n, m)$ in the set S may be calculated by $s(n, m) = [n(n+1)/2] - \sum_{i=2}^{m}(n+1-i)$, where n is the number of residues in the chain and m is the minimum number of residues in a segment. Once the set S is constructed, an exhaustive search may be used to calculate the TLS residual sum of all unique p segment partitions, or configurations, of the chain and select the configuration with the lowest residual.

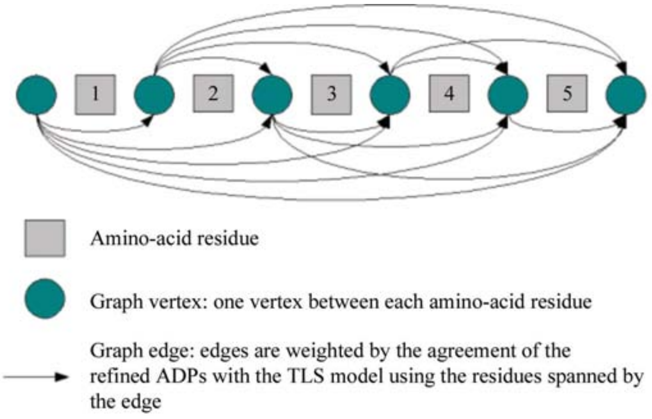
1
2
3
4
5
Amino-acid residue
Graph vertex: one vertex between each amino-acid residue
Graph edge: edges are weighted by the agreement of the refined ADPs with the TLS model using the residues spanned by the edge

Figure 1
The graph constructed in stage 2 of the *TLSMD* algorithm for finding the optimal segmentation of a peptide chain into multiple contiguous TLS groups.

TLSMD is implemented as a two-stage process. The first stage performs a least-squares fit of the TLS parameters to all possible segments in a protein chain (the set S) and stores the results in a database. The second stage searches for a selection of segments that together describe the entire chain and have the minimum summed residual. This is computationally expensive, because the number of configurations grows rapidly with increasing n and p.

2.4.1. Constructing a graph over a protein chain using TLS segments. Remarkably, by posing the search for the configuration with minimum residual as a shortest path problem over a weighted directed graph, the solution may be found in order $O(np)$ time using a variant of the Bellman–Ford algorithm (Bellman, 1958; Ford & Fulkerson, 1962). A graph is constructed to represent a protein chain of n residues as follows: a source vertex $v_s = v_1$ is constructed preceding the first residue, the vertex v_{i+1} is constructed between residues i and $i + 1$ and the destination vertex $v_d = v_{n+1}$ is constructed following the last residue. Thus, the vertices of the graph lie between adjacent amino-acid residues and edges connecting them may be viewed as spanning contiguous segments of the protein chain. In the context of this optimization algorithm the edges represent TLS segments (Fig. 1) and any given set of edges forming a path from the source vertex v_s to destination vertex v_d is equivalent to a set of adjacent protein segments which span the chain. For a given segment containing residues i to j, a corresponding edge is added to the graph connecting vertex v_i to vertex v_{j+1} with weight equal to the number of residues times the weighted TLS residual (4) for that segment: cost = $(j - i + 1)R_{\rm TLS}$. This graph has two important features which simplify the algorithm required to find the shortest path. There are no cycles in the graph and there are no negative edge weights.

A modified version of the Bellman–Ford shortest path algorithm (provided as supplementary material[1]) is used to

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: SX5049). Details for accessing these data are given at the back of the journal.

find the least-cost path from vertex v_1 to vertex v_{n+1}. Each iteration solves for the shortest path containing one additional edge. Thus, the first iteration trivially yields the shortest path using a single edge, *i.e.* the entire chain. The second iteration yields the shortest path using at most two edges and so on. Since graph edges correspond to TLS groups, iteration p yields the optimal p-group TLS partition of the protein chain into segments.

3. Experimental methods

3.1. Refinement protocols

For each test case presented here, the starting point for *TLSMD* analysis was a model generated by conventional refinement of the protein using *REFMAC*5 with one isotropic ADP, $B_{\rm obs}$, for each atom. In the case of structures drawn from the PDB (1kp8, 1y7p), this starting model was generated by subjecting the PDB entry to ten cycles of conventional isotropic refinement in *REFMAC*5.

The starting model was submitted to *TLSMD* analysis using unit weights for each atom as described above. 20 separate models for each structure were generated for subsequent crystallographic refinement. These corresponded to selection of the optimal one-group, two-group, ..., 20-group TLS partition of each peptide chain in the structure as found by *TLSMD*. Each model is described by a pair of files: a `tlsin` file containing the parameters fitted for each TLS group in the model during *TLSMD* analysis and a PDB-format `xyzin` file containing the atomic coordinates from the starting model paired with modified $B_{\rm iso}$ values.

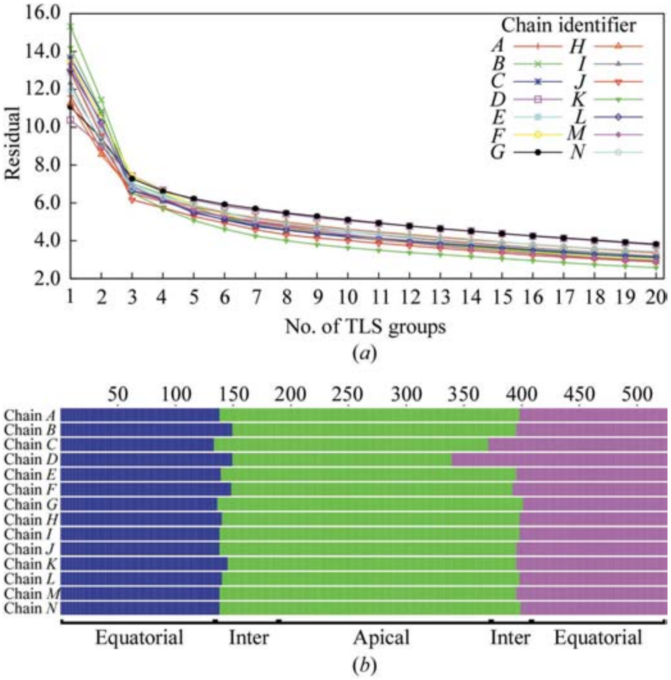

Figure 2
(*a*) The least-squares residual $R_{\rm TLSiso}$ resulting from fitting one-group, two-group, ..., 20-group TLS models to the $B_{\rm iso}$ values obtained by conventional isotropic refinement of *E. coli* GroEL (PDB code 1kp8). Residuals are shown here for each of the 14 crystallographically independent chains in the unit cell. The shape of these curves may be compared with the crystallographic *R* factors obtained from subsequent refinement of the corresponding multi-group TLS models (Fig. 7). (*b*) The optimal three-group partition for each chain based on the analysis shown in (*a*). Residue numbers are shown at the top; domain assignments proposed for this structure by Chaudhry *et al.* (2004) are shown underneath. Figs. 2(*b*) and 3 were prepared using $T_EXshade$ (Beitz, 2000).

Refinement of TLS parameters against crystallographic data employed a locally modified version of *REFMAC*5 that uses singular value decomposition rather than eigenvalue filtering to solve for TLS parameter shifts.

3.1.1. Pure TLS refinement. In this refinement protocol, the ADP for every protein in the structure is described entirely by the TLS parameters for the TLS group to which it belongs. The modified `xyzin` file provided to *REFMAC*5 contains $B_{\rm iso} = 0$ for all protein atoms and $B_{\rm iso} = B_{\rm obs}$ for all non-protein atoms. The refinement consists of five cycles of TLS refinement during which only the TLS parameters are refined against the crystallographic data $F_{\rm obs}$. All other model parameters remain fixed; *i.e.* there is no coordinate refinement or modification of the non-protein thermal parameters.

3.1.2. TLS + $B_{\rm iso}$ refinement. In this refinement protocol, a supplemental isotropic contribution $B_{\rm iso}$ for each protein atom is refined in addition to the TLS parameters. After *TLSMD* analysis, we have an estimated isotropic ADP component $U_{\rm TLS} = (1/3){\rm trace}(U^{ij}_{\rm TLS})$ for each protein atom arising from the TLS model constructed for its particular TLS group. This component can be subtracted from the original input ADP, $U_{\rm obs}$, to yield an estimate for the non-TLS component of the ADP, $U_{\rm iso} = U_{\rm obs} - U_{\rm TLS}$. The atoms of each TLS group are then inspected to find the smallest value of $U_{\rm iso}$ for that group. If necessary, the **T** tensor is modified to shift some of the displacement amplitude out of **T** and into the individual $U_{\rm iso}$ values for atoms in that group, guaranteeing that $U_{\rm iso}$ is always positive and greater than some minimum value. The $U_{\rm iso}$ values are converted to $B_{\rm iso}$ and used to create an `xyzin` file for subsequent refinement. The TLS parameters, including the modified **T** tensor, are used to create a corresponding `tlsin` file. Non-protein atoms retain their original $B_{\rm iso}$. The subsequent *REFMAC*5 refinement against $F_{\rm obs}$ consists of five cycles of TLS-parameter refinement followed by ten cycles of joint coordinate and ADP refinement for both protein and non-protein atoms.

3.1.3. TLS-restrained anisotropic refinement. This refinement protocol is based on local modifications to *REFMAC*5 that allow an input TLS model to be used as a restraint during full anisotropic refinement of U^{ij} terms for each protein atom. The restraint is applied as an additional term in the overall residual being minimized, equivalent to (1). Preparation of `tlsin` and `xyzin` proceeds as in the TLS + $B_{\rm iso}$ protocol. After five cycles of TLS refinement in *REFMAC*5, each protein atom has an associated ADP that consists of an anisotropic description $U^{ij}_{\rm TLS}$ from the TLS model and an isotropic contribution $B_{\rm iso}$ that was present in the `xyzin` file. The modified program combines these by adding the isotropic component into the diagonal terms of $U^{ij}_{\rm TLS}$ and writes our a PDB-format file containing `anisou` records corresponding to an anisotropic ADP for each protein atom. Two copies of this file are then provided as input to a second run of the modified *REFMAC*5. One copy is presented as a standard xyzin file describing a starting model for anisotropic refinement. The second copy is presented as a file of restraint targets. All other ADP restraints to protein atoms are disabled. Non-protein atoms are treated as purely isotropic. For the refinement of *Leishmania braziliensis* initiation factor 5A at 1.6 Å, ten cycles of joint coordinate and restrained anisotropic ADP refinement were then carried out. This refinement was run several times in parallel, so that the weight of the TLS restraint relative to other terms in the minimized residual could be adjusted to yield a final mean anisotropy of 0.45 for the protein atoms (Merritt, 1999).

4. Results

4.1. Output from *TLSMD* analysis

Fig. 2(*a*) shows the net TLS residual $R_{\rm TLSiso}$ resulting from use of an increasing number of TLS groups to model the observed *B* factors from previous conventional refinement of *Escherichia coli* GroEL (PDB code 1kp8). The shape of this curve is dependent on the individual protein, but is expected to correlate with the improvement in crystallographic residuals *R* and $R_{\rm free}$ that would be achieved by taking the corresponding *p*-group TLS model into further crystallographic refinement. In the case of GroEL, a three-group model is dramatically better at explaining the observed distribution of $B_{\rm iso}$ values in the deposited structure than either a one- or two-group model.

TLSMD analysis is performed on all chains present in the unit cell. This offers the opportunity to compare the result of analyzing chains related by non-crystallographic symmetry. Since NCS-related chains will in general have different lattice contacts, we do not necessarily expect the overall magnitude of their net vibrational mode within the lattice to be the same, nor do we necessarily expect the magnitude of the TLS residual as shown in Fig. 2(*a*) to be the same. However, to the extent that *TLSMD* succeeds in identifying groups whose relative flexibility is an intrinsic property of the protein, we expect the shape of the residual curve to be the same for all copies. Furthermore, and more importantly, the optimal partition of the chain into *p* groups should identify similar segment boundaries in each crystallographically independent copy (Fig. 2*b*).

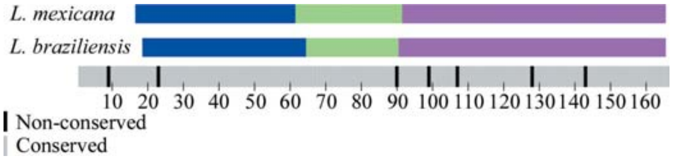

Figure 3
Optimal partition into three TLS groups of two close homologs. Analysis of the 2.7 Å *L. mexicana* structure agrees well with that of the 1.6 Å resolution *L. braziliensis* structure. The two initiation factor 5A homologs differ in sequence at seven residues.

4.2. Crystallographic refinement of multi-group TLS models

We have used structures determined as part of the SGPP structural genomics consortium as test cases for refinement of multi-group TLS models generated by *TLSMD* analysis. Multi-group models were of substantial benefit in many but not all cases. For eight of these structures (PDB entries 2a0k, 2a0u, 2ar1, 1x6o, 1xo7, 1xq9, 1xtd and 1zso) a multi-group TLS model was evaluated as significantly better than either conventional isotropic *B* refinement or refinement of a single-group TLS model supplemented by individual $B_{\rm iso}$ terms. Two of these are structures of eukaryotic initiation factor 5A from the related species *Leishmania mexicana* and *L. braziliensis*. Both homologs contain 166 residues and share 96% sequence identity. Crystals of the *L. mexicana* protein diffract poorly compared with those of the *L. braziliensis* protein but are essentially isomorphous, so these constitute a controlled pair of data sets for trial application of *TLSMD* at different resolutions.

4.2.1. *L. mexicana* eukaryotic initiation factor 5A at 2.7 Å resolution. The plot of $R_{\rm TLSiso}$ *versus* number of TLS groups for this structure suggested that inclusion of at least three TLS groups was warranted, although not as dramatically as the plot in Fig. 2. The optimal three-group partitions found by *TLSMD* for the two IFSA homologs are shown in Fig. 3. The two analyses agree well despite the large difference in resolution of the two input structures.

Conventional refinement of the *L. mexicana* structure yielded $R = 0.198$, $R_{\rm free} = 0.250$, which is somewhat better than expected at this resolution, probably owing to the fact that the original structure solution used the 1.6 Å structure of the *L. braziliensis* homolog as a starting point. In this case, refinement of a pure TLS model is not an improvement. In fact $R_{\rm free}$ begins to rise if more than 15 TLS groups are chosen,

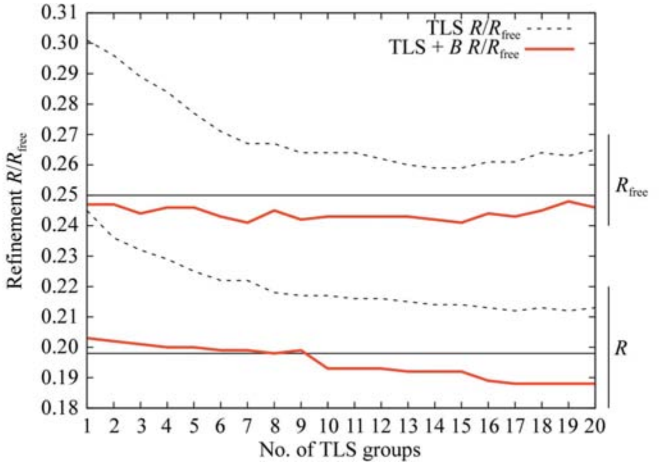

Figure 4
Crystallographic refinement of multi-group TLS models for *L. mexicana* initiation factor 5A. This 2.7 Å data set was collected from a SeMet form of the protein. The *R* and $R_{\rm free}$ residuals from conventional refinement with individual isotopic *B* factors are shown as thin horizontal lines. The dotted lines track the residuals from pure TLS refinement (individual atoms are described only the TLS model, with no individual $B_{\rm iso}$ component). The heavy solid lines correspond to refinement of TLS models supplemented by individual $B_{\rm iso}$ components.

indicating that addition of these additional parameters to a pure TLS model is not justified. On the other hand, refinement of the TLS + $B_{\rm iso}$ model yields better $R_{\rm free}$ values than conventional refinement for any number of TLS groups and better R values than conventional refinement for nine or more TLS groups (Fig. 4).

4.2.2. *L. braziliensis* eukaryotic initiation factor 5A at 1.6 Å resolution. This example is particularly striking in that even a pure TLS model performs well compared with conventional refinement of individual $B_{\rm iso}$ values. The crystallographic model contains 1070 protein atoms. Thus, conventional isotropic refinement requires 1070 thermal parameters, while a 20-group pure TLS model requires only 20 × 20 = 400 thermal parameters. Conventional isotropic refinement yielded R = 0.177, $R_{\rm free}$ = 0.210. A 20-group pure TLS model refines to R = 0.178, $R_{\rm free}$ = 0.200. Thus, R is equivalent to that from conventional refinement, while $R_{\rm free}$ is significantly better. A similar preference for multi-group pure TLS refinement over conventional refinement as judged by $R_{\rm free}$ was previously reported by Winn *et al.* (2001) for a 2.05 Å refinement of GAPDH. In the present case, TLS + $B_{\rm iso}$ refinement protocol yields better residuals R and $R_{\rm free}$ for any number of TLS groups (Fig. 5).

Although 1.6 Å resolution is insufficient for reliable anisotropic refinement using conventional protocols, we were interested in evaluating the utility of multi-group TLS models as restraints for fully anisotropic refinement. Conventional anisotropic refinement in *REFMAC*5 yielded R = 0.149, $R_{\rm free}$ = 0.190. The standard ADP restraints (`rbon bfac sphe`) were adjusted to minimize $R_{\rm free}$, which resulted in a mean anisotropy of 0.59 (0.12), somewhat more spherical than typical near-atomic resolution structures (Merritt, 1999). Using the 20-group TLS model as a restraint rather than using the standard ADP restraints yielded R = 0.150, $R_{\rm free}$ = 0.186. The resulting mean anisotropy was 0.50 (0.16), somewhat closer to the typical value of 0.45. Thus, the residuals from restrained anisotropic refinement are only marginally better than those of the 20-group TLS + $B_{\rm iso}$ model itself, while the total number of thermal parameters increases drastically from 1070 × 1 + 20 × 20 = 1470 for the TLS + $B_{\rm iso}$ to 1070 × 6 = 6420 for full anisotropic treatment. It is apparent that the TLS + $B_{\rm iso}$ model is statistically preferred over restrained anisotropic refinement.

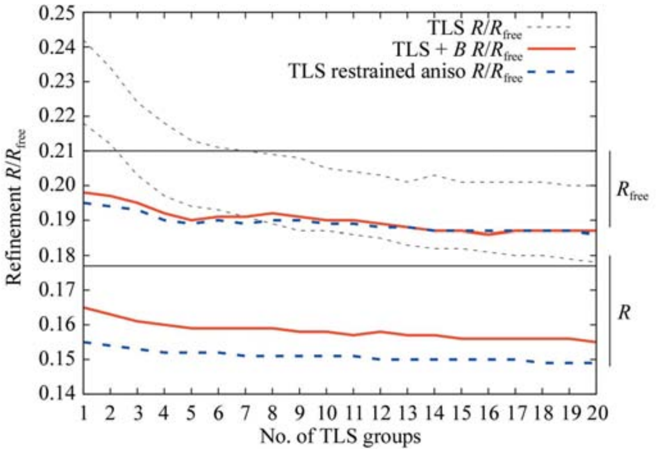

Figure 5
Crystallographic refinement of multi-group TLS models for *L. braziliensis* initiation factor 5A. This 1.6 Å data set was collected from a native (SMet) form of the protein. The R and $R_{\rm free}$ residuals from conventional refinement with individual isotopic B factors are shown as thin horizontal lines. The dotted lines track the residuals from pure TLS refinement (individual atoms are described only for the TLS model, with no individual $B_{\rm iso}$ component). The heavier solid lines correspond to refinement of TLS models supplemented by individual $B_{\rm iso}$ components. The heavy dashed lines correspond to TLS + $B_{\rm iso}$ refinement followed by restrained anisotropic refinement in which individual U^{ij} terms are restrained to agree with the $U^{ij}_{\rm TLS}$ values predicted by the TLS model.

4.2.3. Hypothetical protein AF1403 at 1.9 Å resolution. PDB entry 1y7p is a hypothetical protein from *Archeoglobus fulgidus* selected as structural genomics target AF1403 (R. Zhang, T. Skarina, A. Savchenko, A. Edwards & A. Joachimiak, unpublished work). This is a 223-residue protein of unknown function, although it is structurally homologous to several sugar-binding proteins and the crystal structure contains one bound ribose molecule per protein chain. The structure is a dimer, with one and a half dimers in the asymmetric unit. The original refinement deposited with the PDB

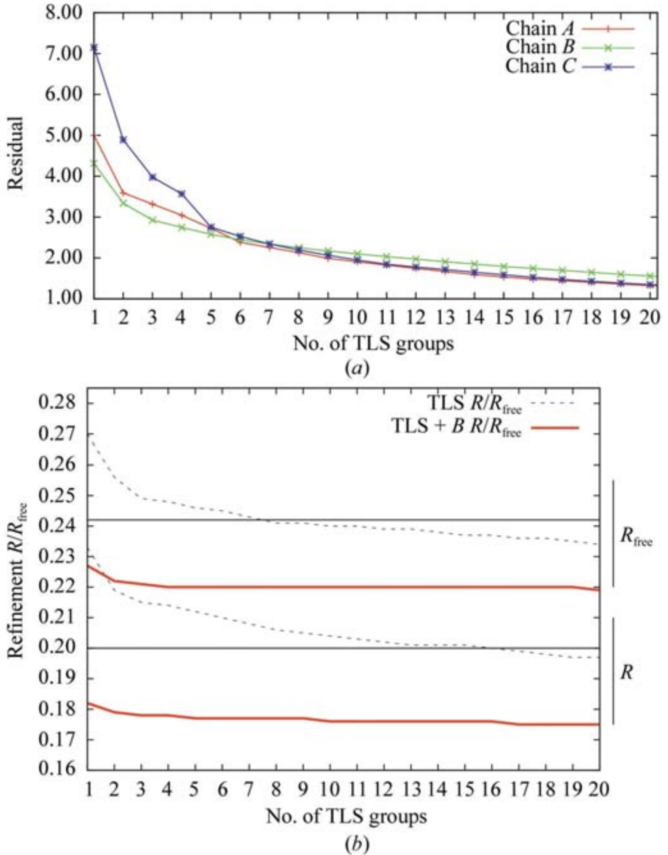

Figure 6
(a) The least-squares residual $R_{\rm TLSiso}$ resulting from fitting multi-group TLS models to the $B_{\rm iso}$ values obtained by conventional isotropic refinement of the three crystallographically independent chains of AF1403 (PDB code 1y7p). (b) Crystallographic refinement of multi-group TLS models for AF1403. The R and $R_{\rm free}$ residuals from conventional refinement with individual isotopic B factors are shown as thin horizontal lines. The dotted lines track the residuals from pure TLS refinement, while the heavier solid lines correspond to refinement of TLS models supplemented by individual $B_{\rm iso}$ components.

contained no TLS model and had crystallographic residuals $R = 0.211$, $R_{\rm free} = 0.255$.

Fig. 6(*a*) shows the result of *TLSMD* analysis applied to the conventionally refined structure and Fig. 6(*b*) the result of subsequent multi-group TLS refinement. Following our standard TLS + $B_{\rm iso}$ protocol, inclusion of a single TLS group per chain lowered R from 0.200 to 0.182 and lowered $R_{\rm free}$ from 0.242 to 0.227 compared with conventional isotropic B-factor refinement. Inclusion of a second TLS group per chain yielded another slight drop in the residuals, but inclusion of three or more TLS groups per chain using the TLS + $B_{\rm iso}$ protocol did not yield substantial additional improvement. We interpret this as indicating significant anisotropic displacement of each atom in the structure arising from overall vibration of the protein within the crystal lattice, anisotropy which is well described by TLS motion of the entire molecule. *TLSMD* analyses of the three crystallographically independent chains agreed identically that the incremental improvement seen from partitioning each chain into two TLS groups corresponds to a split into an N-terminal domain and a C-terminal domain, with a hinge point at residues 80–81. The remaining net displacement of atoms in the structure is small enough that inclusion of conventional $B_{\rm iso}$ parameters for each atom easily accounts for the total magnitude, which is why the residuals for TLS + $B_{\rm iso}$ refinement are flat beyond two TLS groups in Fig. 6(*b*). On the other hand, it is striking that pure TLS models, with no individual $B_{\rm iso}$ terms, continue to yield significantly improved crystallographic residuals as groups are added. The residuals for a pure TLS model with 12 groups per chain are better than for conventional $B_{\rm iso}$ refinement and are still improving at least to the point of 20 groups per chain despite the reduction in the total number of protein thermal parameters from 4898 (one $B_{\rm iso}$ per protein atom) in the conventional $B_{\rm iso}$ model to 1200 (20 × 20 for each of three chains) for a pure TLS model with 20 groups per chain. The continuing drop in crystallographic residual as more groups are added to the pure TLS model supports the idea that these groups describe real, if small, vibrational modes in the protein structure.

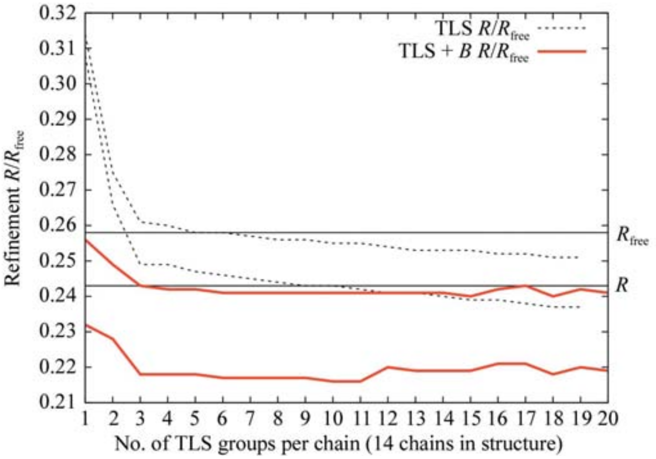

Figure 7
Crystallographic refinement of multi-group TLS models for *E. coli* GroEL. The R and $R_{\rm free}$ residuals from conventional refinement with individual isotopic B factors are shown as thin horizontal lines. The dotted lines track the residuals from pure TLS refinement (individual atoms are described only for the TLS model, with no individual $B_{\rm iso}$ component). The heavier solid lines correspond to refinement of TLS models supplemented by individual $B_{\rm iso}$ components.

4.2.4. GroEL at 2.0 Å resolution

We next show results from refinement of a large structure at medium resolution. The *E. coli* chaperonin GroEL forms a 14-mer with 547 residues per chain and has been studied crystallographically by a number of groups at resolutions from 3.6 to 2.0 Å. Chaudhry *et al.* (2004) reported an analysis of several GroEL complexes based on TLS refinement; each chain was split into three TLS groups by manual assignment of domain boundaries and comparison to Rosenfield analysis performed by the program *ANISOANL*. We chose the 2.0 Å GroEL structure (PDB code 1kp8) reported by Wang & Boisvert (2003) as the starting point for *TLSMD* analysis to avoid any possible bias by previous TLS treatment of the crystallographic model.

This is a large structural model, consisting of residues 2–526 in each of 14 chains. Each chain contains three distinct domains: an apical domain consisting of residues 191–374, an intermediate domain consisting of residues 136–190 and 375–409 and an equatorial domain consisting of residues 2–135 and 410–525 (Chaudhry *et al.*, 2004). It is striking that the three-group TLS models generated by *TLSMD* refine dramatically better than the one- or two-group models (Fig. 7). However, the N-terminal and C-terminal TLS groups in the three-group model should properly both be recognized as belonging to the same (equatorial) domain. Because the two-stage *TLSMD* algorithm assigns only contiguous residues to any one TLS group, it does not notice that these two groups may be merged into a single larger group. The same limitation arises with regard to the intermediate domain, which also contains non-contiguous chain segments. Thus, in its current state *TLSMD* would need a five-group model to describe the three actual domains. This limitation may be overcome by adding a third stage to the *TLSMD* algorithm, in which previously identified groups may be merged (Fig. 8*b*).

Pure TLS models yield better R and $R_{\rm free}$ than conventional isotropic refinement if the chain is broken into nine or more contiguous groups. This is despite a 20-fold reduction in the total number of thermal parameters from 53 970 in the conventional isotropic model to 2520 in the nine group per chain pure TLS model. The TLS + $B_{\rm iso}$ models have better R and $R_{\rm free}$ than conventional isotropic refinement for any number of TLS groups, although $R_{\rm free}$ decreases only marginally with the partition of each chain into more than three groups.

5. Discussion

5.1. TLS models of nested vibrational modes

It may seem inconsistent to model a nested hierarchy of dynamic groups using a linear sequence of rigid-body segments. That is, if a particular flexible loop lies within a larger domain undergoing hinge motion, its net vibrational motion contains contributions both from the local flexibility and from the bulk motion of the domain in which it is embedded. It also contains contributions from the overall vibration of the entire protein within the crystal lattice. Yet if *TLSMD* assigns this flexible loop to its own chain segment, it will be described by only one set of TLS parameters. Fortunately, it is straightforward to show that the cumulative effect of nested TLS motions can always be described by a single set of TLS parameters.

Consider some group of atoms that is described by two nested TLS models A and B. We know that the origin of the TLS model tensors is arbitrary and may be shifted to any choice of origin by appropriate transformation of the **T** and **S** tensors (Schomaker & Trueblood, 1968). By shifting the TLS descriptions of groups A and B to the same origin, it is easy to see they combine to form an equivalent single TLS model C. This is illustrated below by showing the equation for calculating one component of the ADP for an atom in C by combining the contributions of A and B,

$$\begin{aligned}U_C^{11} &= L_A^{22}z^2 + L_A^{33}y^2 - 2L_A^{23}yz + 2S_A^{21}z - 2S_A^{31}y + T_A^{11}\\ &\quad + L_B^{22}z^2 + L_B^{22}y^2 + L_B^{23}yz + 2S_B^{21}z - 2S_B^{31}y + T_B^{11}\\ &= (L_A^{22} + L_B^{22})z^2 + (L_A^{33} + L_B^{33})y^2 - 2(L_A^{23} + L_B^{23})yz\\ &\quad + 2(S_A^{21} + S_B^{21})z - 2(S_A^{31} + S_B^{31})y + (T_A^{11} + T_B^{11})\\ &= L_C^{22}z^2 + L_C^{33}y^2 - 2L_C^{23}yz + 2S_C^{21}z - 2S_C^{31}y + T_C^{11}.\end{aligned}$$

Similar equations are obtained for the other terms U_C^{ij} of the combined ADP.

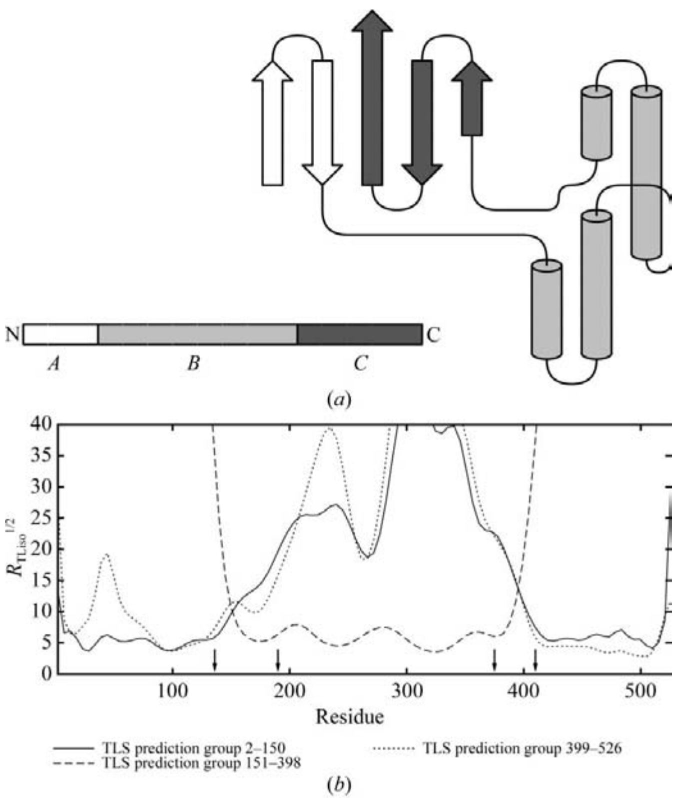

Figure 8
(*a*) Schematic topology of a two-domain protein in which one domain is made up of chain segments from both the N- and C-termini. The *TLSMD* algorithm will partition the chain into three groups *A*, *B* and *C* without considering that *A* and *C* may properly belong to the same TLS group. (*b*) One possible metric for evaluating whether two TLS groups should be merged into a single group. After *TLSMD* analysis, each group has its own set of TLS parameters which approximately predict the initial ADPs observed for atoms belonging to that group. If the TLS parameters fitted for one group are also found to predict also the ADPs observed for a second group, this is an indication that the two groups are in fact non-contiguous segments of a single larger group. The figure shows a three-group partition for one chain in the 2 Å GroEL structure 1kp8, for which the TLS parameters fitted for residues 2–150 also generate a low residual when applied to residues in the range 399–526 and *vice versa*. Neither of these two sets of TLS parameters yield a good residual when applied to residues in the intervening portion of the chain. The arrows indicate domain boundaries identified by the analysis of Chaudhry *et al.* (2004).

This result assures us that we can describe the effect of nested TLS groups without an explicit model for the nesting. Nevertheless, physical interpretation of the multi-group TLS model is complicated by this collapse of terms describing several nested levels of motion into a single combined set of parameters.

5.2. Limitations

5.2.1. Non-contiguous TLS groups. An obvious limitation of the current *TLSMD* algorithm is that it only considers contiguous runs of residues within a chain when searching for TLS groups. It fails to identify the common case of a compact protein domain made up of residues from several portions of the chain. A common example is a domain containing residues from both the N- and C-termini (Fig. 8*a*). We have considered several possible ways of implementing a separate post-processing step that would merge two or more groups identified by the existing *TLSMD* algorithm into a single non-contiguous TLS group. One promising approach is to assign a score to potential mergers of two groups based on their ability to cross-predict each other's thermal parameters. Fig. 8(*b*) shows a three-group partition of the 2 Å GroEL structure, in which the TLS parameters fitted for residues 2–150 also generate a low residual when applied to residues in the range 399–526 and *vice versa*. Neither of these two sets of TLS parameters yields a good residual when applied to residues in the intervening portion of the chain. Therefore, it would be reasonable to combine both the N-terminal and C-terminal segments into a single TLS group. This is in reasonably good agreement with the literature characterization of the GroEL equatorial domain as consisting of residues 2–136 and 410–525 (Chaudhry *et al.*, 2004).

5.2.2. TLS-group boundaries are not identical to hinge points or domain boundaries. The optimized boundaries between adjacent TLS groups do not match up precisely to specific hinge-point residues in the structure. Two portions of the protein that are well described by TLS may be separated by a short run of residues that are not so well described and which constitute a hinge region rather than a hinge point. The *TLSMD* analysis will place a break point somewhere within this region, but since these residues are not well described by the TLS parameters of either neighboring TLS group, the precise placement of the break within the hinge region is sensitive to small perturbations or noise in the input model for these residues. Thus, in general the break points assigned for multiple NCS-related chains should not be expected to match up identically. Furthermore, if the protein chain is partitioned into many TLS groups, then the insertion of additional break points into the middle of a larger group may obscure the nature and extent of the larger group. For example, the presence of a flexible loop in the middle of a large domain may be correctly found and modeled, but this will split the original single TLS group describing the domain into three TLS groups. The bulk of the domain is still well described by the original single set of TLS parameters, but the residues making up this larger bulk are no longer contiguous. This is essentially the same case illustrated in Fig. 8 and may be addressed in the same manner by adding a post-processing step to identify and merge compatible TLS groups.

5.2.3. Chain segments that are not well described by any TLS model. Although the derivation of the TLS formalism assumes a perfectly rigid group, even non-rigid groups may be described well by TLS models for small amplitudes of motion. Nevertheless, the observed displacements for some groups of atoms may be poorly predicted by any TLS model, even for small amplitudes of motion. *TLSMD* will not assign residues in such a region to a group of their own precisely because they are not described well by any TLS model and hence putting them in a separate group will not usually improve the overall TLS residual. However, for the purpose of inferring protein dynamic properties, it would be useful to identify and label such regions separately rather than including them in a neighboring TLS group.

5.2.4. Relevance to protein–protein docking and other applications. We have yet to systematically evaluate the accuracy and utility of *TLSMD*-generated models for applications outside of crystallographic refinement. Nevertheless, some case-by-case comparison is possible using individual proteins that have been characterized previously by other techniques. For example, Akif and coworkers have shown good agreement between multi-group TLS models and normal-mode analysis based on an elastic network model, applying both techniques to a 3 Å structure of *M. tuberculosis* thioredoxin reductase (Akif *et al.*, 2005). A second opportunity for comparison is the availability of a large number of protein backbone motions collected in the Database of Macromolecular Movements (MolMovDB; Echols *et al.*, 2003). These are generated by interpolation ('morphing') between distinct conformations found in various crystal structures. *TLSMD* analysis can be applied to one or more of the constituent structures contributing to the MolMovDB interpolation, followed by qualitative comparison of animation or other visualizations from MolMovDB morphing (Krebs & Gerstein, 2000) and from *TLSMD* analysis (Painter & Merritt, 2005). We have collected several examples of such paired animations on the *TLSMD* website (Painter & Merritt, 2006).

Some potential applications of *TLSMD* have well defined quantitative criteria for success. For example, the difficulty of modeling backbone flexibility is a major limitation in current methods for predicting protein–protein interactions, even when the structures of both target proteins are already known individually (Janin, 2005). Use of *TLSMD* analysis to identify probable modes of backbone deformation may therefore provide a powerful extension to the current generation of docking programs such as *RosettaDock* (Gray *et al.*, 2003; Schueler-Furman *et al.*, 2005; Wang *et al.*, 2005) and can be evaluated against the protein-docking targets comprising the *CAPRI* challenge set (Janin *et al.*, 2003; Janin, 2005).

5.2.5. Availability. *TLSMD* is a Open Source code base released under the Artistic License. It is hosted on SourceForge as a subproject of the Python Macromolecular Library (Painter & Merritt, 2004). We have also created a web server http://skuld.bmsc.washington.edu/~tlsmd to perform *TLSMD* analysis of structures submitted by external groups (Painter & Merritt, 2006). Because the analysis is CPU-intensive, jobs are queued for later execution and the submitter is notified by e-mail when the analysis has been completed.

5.2.6. Summary. We have shown here that the addition of multiple TLS groups into conventional isotropic refinement typically lowers both the resulting R and $R_{\rm free}$ residuals. It is particularly striking that in some cases even a pure TLS model with no refinement of individual B factors outperforms conventional crystallographic refinement. In three of the test cases presented here, pure TLS refinement of a model containing 10–20 TLS groups yielded better R and $R_{\rm free}$ than conventional isotropic refinement despite reduction in the number of thermal parameters in the model by a factor of 2.6 (1x6o) to 20 (1kp8). This reflects the fact that multi-group TLS models constitute a parsimonious approximation to the true underlying behaviour, both static and dynamic, of atoms in a protein crystal structure.

The *TLSMD* algorithm described here allows the construction of optimal multi-group TLS models from a previously refined structure or as part of the refinement of a new structure. Moreover, the ability to identify and model specific modes of backbone flexibility on the basis of single-crystal structures is of great interest even apart from any improvement to crystallographic refinement.

We are grateful to Mark Robien and Jürgen Bosch for providing the *L. mexicana* and *L. braziliensis* eukaryotic initiation factor 5A data sets for trials of *TLSMD* analysis and to Ron Stenkamp for stimulating discussions. This work was supported by NIH awards GM62617 and GM64655.

References

Akif, M., Suhre, K., Verma, C. & Mande, S. C. (2005). *Acta Cryst.* D**61**, 1603–1611.

Anderson, E., Bai, Z., Bischof, C., Blackford, S., Demmel, J., Dongarra, J., Du Croz, J., Greenbaum, A., Hammarling, S., McKenney, A. & Sorensen, D. (1999). *LAPACK Users' Guide*, 3rd ed. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Beitz, E. (2000). *Bioinformatics*, **16**, 135–139.

Bellman, R. (1958). *Q. Appl. Math.* **16**, 87–90.

Bernett, M. J., Somasundaram, T. & Blaber, M. (2004). *Proteins*, **57**, 626–634.

Bürgi, H. B. (2000). *Annu. Rev. Phys. Chem.* **51**, 275–296.

Chaudhry, C., Horwich, A. L., Brunger, A. T. & Adams, P. D. (2004). *J. Mol. Biol.* **342**, 229–245.

Driessen, H., Haneef, M. I. J., Harris, G. W., Howlin, B., Khan, G. & Moss, D. S. (1989). *J. Appl. Cryst.* **22**, 510–516.

Dunitz, J., Schomaker, V. & Trueblood, K. N. (1988). *J. Phys. Chem.* **92**, 856–867.

Echols, N., Milburn, D. & Gerstein, M. (2003). *Nucleic Acids Res.* **31**, 478–482.

Ford, L. R. Jr & Fulkerson, D. R. (1962). *Flows in Networks.* Princeton University Press.

Gray, J. J., Moughon, S., Wang, C., Schueler-Furman, O., Kuhlman, B., Rohl, C. A. & Baker, D. (2003). *J. Mol. Biol.* **331**, 281–299.

He, X.-M. & Craven, B. (1993). *Acta Cryst.* A**49**, 10–22.

Howlin, B., Butler, S. A., Moss, D. S., Harris, G. W. & Driessen, H. P. C. (1993). *J. Appl. Cryst.* **26**, 622–624.

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

Janin, J. (2005). *Protein Sci.* **14**, 278–283.

Janin, J., Henrick, K., Moult, J., Ten Eyck, L., Sternberg, M. J. E., Vajda, S., Vakser, I. & Wodak, S. J. (2003). *Proteins*, **52**, 2–9.

Krebs, W. G. & Gerstein, M. (2000). *Nucleic Acids Res.* **28**, 1665–1675.

Kuriyan, J. & Weis, W. I. (1991). *Proc. Natl Acad. Sci. USA*, **88**, 2773–2777.

Merritt, E. A. (1999). *Acta Cryst.* D**55**, 1109–1117.

Moss, D. S., Tickle, I. J., Theis, O. & Wostrack, A. (1996). *Proceedings of the CCP4 Study Weekend. Macromolecular Refinement*, edited by E. Dodson, M. Moore, A. Ralph & S. Bailey, pp. 105–113. Warrington: Daresbury Laboratory.

Murshudov, G. N., Lebedev, A., Vagin, A. A., Wilson, K. S. & Dodson, E. J. (1999). *Acta Cryst.* D**55**, 247–255.

Painter, J. & Merritt, E. A. (2004). *J. Appl. Cryst.* **37**, 174–178.

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

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

Papiz, M. Z. & Prince, S. M. (1996). *Proceedings of the CCP4 Study Weekend. Macromolecular Refinement*, edited by E. Dodson, M. Moore, A. Ralph & S. Bailey, pp. 115–123. Warrington: Daresbury Laboratory.

Rosenfield, R. E., Trueblood, K. N. & Dunitz, J. D. (1978). *Acta Cryst.* A**34**, 828–829.

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

Schueler-Furman, O., Wang, C. & Baker, D. (2005). *Proteins*, **60**, 187–194.

Sternberg, M. J., Grace, D. E. & Phillips, D. C. (1979). *J. Mol. Biol.* **130**, 231–252.

Trueblood, K. N. (1978). *Acta Cryst.* A**34**, 950–954.

Trueblood, K. N., Bürgi, H.-B., Burzlaff, H., Dunitz, J. D., Gramaccioli, C. M., Schulz, H. H., Shmueli, U. & Abrahams, S. C. (1996). *Acta Cryst.* A**52**, 770–781.

Wang, C., Schueler-Furman, O. & Baker, D. (2005). *Protein Sci.* **14**, 1328–1339.

Wang, J. & Boisvert, D. C. (2003). *J. Mol. Biol.* **327**, 843–855.

Willis, B. T. M. & Pryor, A. W. (1975). *Thermal Vibrations in Crystallography.* Cambridge University Press.

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.