Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Use of TLS parameters to model anisotropic displacements in macromolecular refinement

M. D. Winn,[a]* M. N. Isupov[b] and G. N. Murshudov[a,c]

[a]Daresbury Laboratory, Daresbury, Warrington WA4 4AD, England, [b]Department of Chemistry and Biological Sciences, University of Exeter, Exeter EX4 4QD, England, and [c]Chemistry Department, University of York, Heslington, York YO1 5DD, England

Correspondence e-mail: m.d.winn@dl.ac.uk

Received 30 May 2000
Accepted 19 October 2000

An essential step in macromolecular refinement is the selection of model parameters which give as good a description of the experimental data as possible while retaining a realistic data-to-parameter ratio. This is particularly true of the choice of atomic displacement parameters, where the move from individual isotropic to individual anisotropic refinement involves a sixfold increase in the number of required displacement parameters. The number of refinement parameters can be reduced by using collective variables rather than independent atomic variables and one of the simplest examples of this is the TLS parameterization for describing the translation, libration and screw-rotation displacements of a pseudo-rigid body. This article describes the implementation of the TLS parameterization in the macromolecular refinement program *REFMAC*. Derivatives of the residual with respect to the TLS parameters are expanded in terms of the derivatives with respect to individual anisotropic U values, which in turn are calculated using a fast Fourier transform technique. TLS refinement is therefore fast and can be used routinely. Examples of TLS refinement are given for glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and a transcription activator GerE, for both of which there is data to only 2.0 Å, so that individual anisotropic refinement is not feasible. GAPDH has been refined with between one and four TLS groups in the asymmetric unit and GerE with six TLS groups. In both cases, inclusion of TLS parameters gives improved refinement statistics and in particular an improvement in R and free R values of several percent. Furthermore, GAPDH and GerE have two and six molecules in the asymmetric unit, respectively, and in each case the displacement parameters differ significantly between molecules. These differences are well accounted for by the TLS parameterization, leaving residual local displacements which are very similar between molecules and to which NCS restraints can be applied.

1. Introduction

An essential part of the model for a macromolecule is the set of atomic displacement parameters, which give the mean-square deviation of each atom from its average position within the usual Gaussian approximation (Willis & Pryor, 1975; Trueblood *et al.*, 1996). For a model derived from X-ray diffraction data, these displacement parameters may describe static disorder, with equivalent atomic coordinates differing between unit cells, dynamic disorder, where the diffraction data represent a time average over atoms in motion, or they may simply account for errors in the model. In any case, an accurate model for displacement parameters is required to provide a good fit to the diffraction data and hence to make reliable predictions of the average positions.

In general, a structure model will contain parameters describing the displacements of each atom and parameters describing the correlations between them. However, interatomic correlations cannot be determined directly from Bragg intensities [but see Kidera & Go (1992) for a discussion of the indirect inference of interatomic correlations], and one is left with atomic displacements only, each represented by a one-parameter isotropic model or a six-parameter anisotropic model. These atomic displacement parameters are usually individually determined in structure refinement. However, more sophisticated descriptions are possible in which the atomic displacement parameters are derived from a more general model for the displacements of the molecule or the crystal in which it resides.

For the purposes of discussion, it is convenient to consider four separate (and in general anisotropic) contributions to the total atomic displacement parameter,

$$U = U_{\rm crystal} + U_{\rm TLS} + U_{\rm internal} + U_{\rm atom}. \quad (1)$$

$U_{\rm crystal}$ represents the overall anisotropy of the crystal and is a single anisotropic displacement parameter applied to the entire contents of the unit cell; as such it obeys the symmetry of the crystal space group when refined against merged data. Inclusion of such anisotropic scaling is known to give improvements in crystallographic R and free R factors of up to several percent and improved behaviour of refinement (Sheriff & Hendrickson, 1987; Murshudov *et al.*, 1998). $U_{\rm TLS}$ represents translations and librations of pseudo-rigid bodies within the asymmetric unit of the crystal. These bodies may be whole molecules or identifiable molecular subunits. Next, $U_{\rm internal}$ includes various kinds of intramolecular collective motions, such as libration about particular torsion angles or internal normal modes of a molecule. Finally, $U_{\rm atom}$ represents displacements of individual atoms and ideally includes local displacements only.

In small-molecule crystallography, refinement of atomic anisotropic displacement parameters (atomic ADPs) is routine. Identification of the contributions to the total ADP, for example $U_{\rm TLS}$, is generally performed after refinement. For macromolecules, because of weaker diffraction, the six parameters per atom needed to model atomic ADPs is usually prohibitive and isotropic displacement parameters are determined instead, although with improvements in data collection individual anisotropic refinement is becoming more common (Merritt, 1999). In this respect, models involving the anisotropic collective motions represented by $U_{\rm TLS}$ and $U_{\rm internal}$ are of interest to macromolecular studies since they generally involve far fewer parameters than the corresponding individual ADPs. Furthermore, these terms incorporate in a natural way the fact that the displacements of neighbouring atoms are likely to be highly correlated. When individual ADPs are refined, restraints are applied to the terms in $U_{\rm atom}$. These restraints are primarily a means to stabilize the refinement, but may also reflect the same correlations between displacements, for example the rigid-bond restraint (Rollett, 1970; Hirshfeld, 1974; Sheldrick & Schneider, 1997).

The earliest analysis of rigid-body displacements of molecules in crystals was that of Cruickshank (1956), who interpreted refined individual ADPs in terms of two tensors for translation and libration. A more general method including the average quadratic correlation between translation and libration, the so-called TLS method, was introduced by Schomaker & Trueblood (1968). In these approaches, the tensors describing the rigid-body motion are fitted to ADPs which have been previously refined against X-ray data. Pawley (1966) introduced an alternative approach (using the two tensors of Cruickshank) in which the rigid-body tensors are refined directly against the X-ray data. This bypasses the need for obtaining individual ADPs and is the method we are concerned with here.

The first application of the refinement of TLS parameters against X-ray data for a macromolecule was that of Holbrook and co-workers (Holbrook & Kim, 1984; Holbrook *et al.*, 1985), who used a modified version of the *CORELS* program to refine TLS parameters for each phosphate, ribose and base of a duplex DNA dodecamer. Subsequently, Moss and co-workers extended the program *RESTRAIN* (Driessen *et al.*, 1989; Collaborative Computational Project, Number 4, 1994) to allow refinement of TLS parameters and applied this to bovine ribonuclease A (Howlin *et al.*, 1989) and papain (Harris *et al.*, 1992). Šali *et al.* (1992) used *RESTRAIN* to refine TLS parameters for the two domains of an endothiapepsin complex. Papiz & Prince (1996) used *RESTRAIN* to refine two TLS models of light-harvesting complex II. However, these programs are not widely used (*RESTRAIN* does not use fast Fourier transforms and is slow to run) and TLS refinement is not yet a common technique.

With the increasing number of atomic resolution protein structures in recent years, the comparison and fitting of TLS parameters to refined individual ADPs is possible, as is more common for small molecules. Stec *et al.* (1995) refined crambin against atomic resolution data with the protein molecule divided into one, two or three TLS groups and compared the predicted anisotropic U values with directly refined values. They further compared the six sets of TLS tensors with each other, extracted the largest common part and hypothesized that this represented the true external motion of the molecule with contributions from internal motions removed. The external motion was found to represent more than 60% of the total mobility. Very recently, Wilson & Brunger (2000) have compared a full anisotropic refinement of calmodulin with a TLS refinement in order to identify domain displacements. Harata and co-workers (Harata *et al.*, 1998, 1999) have refined individual ADPs for turkey egg-white and human lysozymes and human α-lactalbumin and subsequently determined TLS tensors by a least-squares fit.

Groups associated with rigid-body tensors are rarely completely rigid and a variety of additional motions internal to the group can be identified, as represented by the third term on the right-hand side of (1). For small molecules, the approach introduced by Dunitz & White (1973) is usually used, whereby one also determines the mean-square amplitudes of a number of specifically identified internal torsions.

This approach has been generalized to include correlations between the rigid-body motion and the individual torsions and has been implemented in the program *THMA* (Schomaker & Trueblood, 1998).

For macromolecules, early studies of possible internal motion were performed by Sternberg *et al.* (1979) and Artymiuk *et al.* (1979), who compared several models for the refined isotropic displacement parameters of hen egg-white and human lysozymes. They concluded that these displacements must have an intramolecular component superimposed on a rigid-body motion or breathing motion. Diamond (1990) and Kidera & Go (1990, 1992) refined the amplitudes of normal modes describing the internal motion of the protein, together with TLS parameters for the overall motion of each molecule. These studies showed that while account of the internal motion was necessary for a complete description of the protein and is perhaps the most biologically informative aspect, the TLS parameters for the overall motion make the largest contribution to atomic displacements and are a necessary precursor for separating out the internal motion.

The aim of the present study is to implement a refinement protocol using a model for the terms $U_{\rm crystal}$, $U_{\rm TLS}$ and $U_{\rm atom}$. The term $U_{\rm TLS}$, representing rigid-body displacements of groups of atoms, is modelled by the TLS formalism (Schomaker & Trueblood, 1968). The next section gives the TLS approach in more detail and describes its implementation in the macromolecular refinement program *REFMAC* (Murshudov *et al.*, 1997). *REFMAC* uses a fast Fourier transform method which means that TLS refinement is fast and convenient. Examples of its application are given in the following section. The collective displacements represented by $U_{\rm internal}$ will be the subject of future work.

In the context of (1), large-scale contributions are removed successively from the total atomic ADP by $U_{\rm crystal}$ and $U_{\rm TLS}$, leaving $U_{\rm atom}$ (or more usually its isotropic equivalent) to describe local displacements. This is in contrast to earlier versions of *REFMAC*, in which $U_{\rm atom}$ (either anisotropic or isotropic) accounts for all contributions to the total atomic displacement apart from the overall anisotropic scaling $U_{\rm crystal}$. In addition, use of $U_{\rm TLS}$ allows a degree of anisotropic refinement in cases where the data-to-parameter ratio does not justify refinement of individual ADPs.

One application of this division between $U_{\rm TLS}$ and $U_{\rm atom}$ is in applying restraints on ADPs between atoms related by non-crystallographic symmetry. While such atoms may have locally similar environments, the molecules to which they belong are likely to have different overall displacement parameters. Representing the latter by TLS parameters allows the restraints to be applied to the residual terms in $U_{\rm atom}$, which should in that case be more similar.

2. TLS parameterization

2.1. Definition of TLS parameters

The theory behind the TLS parameterization has been presented in detail by Schomaker & Trueblood (1968), with useful summaries in Howlin *et al.* (1989) and Schomaker & Trueblood (1998). We therefore restrict ourselves here to a short summary.

Any displacement of a rigid body can be described as a rotation about an axis passing through a fixed point together with a translation of that fixed point. The corresponding displacement of a point at $\mathbf{r}$ relative to the fixed point is given by

$$\mathbf{u} = \mathbf{t} + \mathbf{D} \cdot \mathbf{r}, \qquad (2)$$

where $\mathbf{t}$ is a column vector for the translation and $\mathbf{D}$ is the rotation matrix. For small displacements, the last term in (2) can be linearized with respect to the amplitude of the rotation to give

$$\mathbf{u} \simeq \mathbf{t} + \boldsymbol{\lambda} \times \mathbf{r}, \qquad (3)$$

where $\boldsymbol{\lambda}$ is a vector along the rotation axis with a magnitude equal to the angle of rotation and $\times$ denotes a cross product (higher order expansions have been considered by, for example, Sygusch, 1976). The corresponding dyad product is then

$$\mathbf{u}\mathbf{u}^T = \mathbf{t}\mathbf{t}^T + \mathbf{t}\boldsymbol{\lambda}^T \times \mathbf{r}^T - \mathbf{r} \times \boldsymbol{\lambda}\mathbf{t}^T - \mathbf{r} \times \boldsymbol{\lambda}\boldsymbol{\lambda}^T \times \mathbf{r}^T, \qquad (4)$$

where superscript T denotes the row vector. Finally, performing a time and spatial average over all displacements yields

$$\mathbf{U} \equiv \langle \mathbf{u}\mathbf{u}^T \rangle = \mathbf{T} + \mathbf{S}^T \times \mathbf{r}^T - \mathbf{r} \times \mathbf{S} - \mathbf{r} \times \mathbf{L} \times \mathbf{r}^T, \qquad (5)$$

where $\mathbf{T} \equiv \langle \mathbf{t}\mathbf{t}^T \rangle$, $\mathbf{L} \equiv \langle \boldsymbol{\lambda}\boldsymbol{\lambda}^T \rangle$ and $\mathbf{S} \equiv \langle \boldsymbol{\lambda}\mathbf{t}^T \rangle$. In this context, the cross product is used as follows: $\mathbf{L} \times \mathbf{r}^T$ yields a matrix whose ith row is the cross product of the ith row of $\mathbf{L}$ and $\mathbf{r}^T$.

Equation (5) gives the mean-square displacement of a point $\mathbf{r}$ in a rigid body in terms of three tensors $\mathbf{T}$, $\mathbf{L}$ and $\mathbf{S}$. Considering in particular the set of points $\{\mathbf{r}\}$ corresponding to the rest positions of atoms in a single rigid body, $\mathbf{U}$ is the mean-square displacement of each such atom and can be identified as the anisotropic displacement parameter that occurs in the Debye–Waller factor in the expression for the structure factor. The linearization used to obtain (3) is equivalent to retaining only quadratic terms in the expression for the ADP. Note that one could also construct expressions for terms $\langle \mathbf{u}_i \mathbf{u}_j^T \rangle$ describing the correlation between the displacements of atoms i and j in the same rigid body. Such cross terms cannot be determined directly from Bragg intensities, but are implicitly included in models for X-ray diffuse scattering.

Given a set of refined ADPs, (5) can be used to make a least-squares fit of TLS parameters. Alternatively, and the approach we use here, (5) can be used to derive ADPs and hence calculated structure factors from TLS refinement parameters. $\mathbf{T}$ and $\mathbf{L}$ are symmetric tensors, while $\mathbf{S}$ is in general asymmetric. Expanding (5) out fully shows that the trace of $\mathbf{S}$ is not fixed by $\mathbf{U}$. Hence, there are a total of 20 refinable parameters (six from $\mathbf{T}$, six from $\mathbf{L}$ and eight from $\mathbf{S}$).

Equation (2) expanded to quadratic terms in $\boldsymbol{\lambda}$ and averaged gives an expression for the mean position of each atom (Howlin *et al.*, 1989),

$$\mathbf{x} = \mathbf{r} + \tfrac{1}{2}[\mathbf{L} - \mathrm{tr}(\mathbf{L})\mathbf{I}]\mathbf{r}. \quad (6)$$

The correction relative to the rest position $\mathbf{r}$ is $O(\mathbf{L})$ and can therefore be neglected in (5) for $\mathbf{U}$, *i.e.* no distinction is made between rest and mean positions. However, this distinction needs to be taken into account when applying distance restraints, which apply to distances between rest positions $\mathbf{r}$ rather than between the observed mean positions $\mathbf{x}$ (the former being a better measure of the mean distance). Given the observed distance d_0, the distance between rest positions d can be estimated as (Howlin *et al.*, 1989)

$$d = d_0\{1 + \tfrac{1}{2}[\mathrm{tr}(\mathbf{L}) - \hat{\mathbf{n}}^T\mathbf{L}\hat{\mathbf{n}}]\}, \quad (7)$$

where $\hat{\mathbf{n}}$ is a unit vector along the bond in question. For small TLS groups such as amino-acid side chains, d can be greater than d_0 by 0.01 Å or more (Howlin *et al.*, 1989) and therefore can have a significant effect on the agreement between ideal and observed distances. For larger TLS groups, such as we consider later, values of $\mathbf{L}$ and hence the distance correction tend to be an order of magnitude smaller and the correction is probably less important.

2.2. Choice of rigid groups

An important component of the TLS model of anisotropic displacements is the choice of rigid groups. The simplest method of choosing the make-up of these groups is to use chemical knowledge of the rigidity of certain groups of atoms. For example, phenyl groups such as those in phenylalanine and tyrosine side chains are known to be quite rigid and were used as TLS groups along with other conjugated side chains in the studies of Howlin *et al.* (1989) and Harris *et al.* (1992).

For lower resolutions, we wish to consider TLS groups for larger groups of atoms. Groups larger than a single side chain clearly are not entirely rigid. However, there may be a significant component of the atomic displacements which can be attributed to a rigid-like motion, with non-rigid motions superimposed on this. Thus, one may expect that secondary-structure elements such as α-helices, or larger groupings such as domains, can be treated as TLS groups.

More robust definitions of TLS groups require additional information. If more than one crystal form of a particular protein is available, then changes of conformation between the different forms can be used to identify so-called dynamic domains, *i.e.* domains that move as quasi-rigid bodies. This has been implemented in the computer program *DYNDOM* (Hayward & Berendsen, 1998). In conflating these dynamic domains with TLS groups, the assumption is that relative displacements between different crystal forms reflect likely displacements within a single crystal form. Dynamic domains can also be identified from molecular-dynamic simulations by superimposing a series of instantaneous configurations, each of which may be rotated relative to the others. Regions which superimpose well under some set of rotations can be considered as potential TLS groups. This method has been used in a recent TLS refinement of light-harvesting complex II (Papiz, 2000).

Another handle on likely quasi-rigid-body displacements can be obtained from previously refined individual ADPs. Rigid-body components can be identified by optimizing the fit of TLS parameters to the refined U values. In itself, this is only useful as a means of interpreting the refined U values and there is no need to repeat the refinement with TLS parameters. However, if there is a series of similar structures, then knowledge of suitable TLS groups can be transferred to those structures for which U values cannot be reliably determined. Such an approach was adopted by Holbrook and coworkers (Holbrook & Kim, 1984; Holbrook *et al.*, 1985), who compared seven different rigid-body models of deoxycytidine 5′-phosphate and used the best (as measured by two indices for the agreement between refined and derived Us) in subsequent TLS refinements of other nucleic acids.

Another approach using refined individual ADPs is to use the rigid-body criterion (Rosenfield *et al.*, 1978) in which a $\mathbf{\Delta}$ matrix is built up between all pairs of atoms, with elements equal to the difference in the projected U values along the interatomic vector. Pairs of atoms belonging to the same quasi-rigid group should have a $\mathbf{\Delta}$ value close to zero. Brock *et al.* (1985) used this approach in their analysis of triphenylphosphine oxide and similar ideas were used by Schneider (1996) for the protein SP445.

2.3. Implementation in *REFMAC*

Refinement of TLS parameters has been implemented in the macromolecular refinement program *REFMAC* (Murshudov *et al.*, 1997). Maximization of the likelihood function with respect to TLS parameters requires the calculation of derivatives of the likelihood function with respect to the elements of the tensors $\mathbf{T}$, $\mathbf{L}$ and $\mathbf{S}$ for each TLS group defined. These derivatives are obtained by the chain rule from the derivatives with respect to individual anisotropic displacement parameters. The latter can be calculated efficiently using fast Fourier transforms as described previously (Murshudov *et al.*, 1999). One consequence of (5) is that $\mathbf{U}$ is now a function of the atomic coordinates. Hence, derivatives of the residual with respect to atomic coordinates also include terms involving the derivative of the residual with respect to the elements of $\mathbf{U}$, reflecting the fact that shifts in the atomic positions change the contribution of the TLS parameters to the calculated structure factor.

Equation (5) relating the individual ADPs to the TLS parameters can be rewritten in terms of a 6×6 variance–covariance matrix $\mathbf{\Sigma}$ for the translation and libration (Kidera & Go, 1992). Although this matrix should be positive definite, this is not guaranteed by the minimization procedure. This can be enforced by a Cholesky decomposition of the TLS groups in terms of a lower triangular matrix $\boldsymbol{\sigma}$ ($\boldsymbol{\sigma}_{mn} = 0$, $m < n$),

$$\mathbf{\Sigma} = \boldsymbol{\sigma}\boldsymbol{\sigma}^T. \quad (8)$$

Derivatives with respect to these new parameters can be obtained from those with respect to the TLS parameters by another application of the chain rule. We note that this decomposition introduces an extra minimum at $\boldsymbol{\sigma} = 0$ which

must be avoided. At the expense of added computations, this re-parameterization ensures correct behaviour of the minimization.

In tests on a single example (not included here), we found that this method worked reasonably well, correcting for slight deviations from non-positive definiteness. In practice, however, such deviations are minor and are acceptable provided that the total ADPs (which include also the individual isotropic B factors) are positive definite, which is in fact found to be the case. We therefore do not use this procedure routinely; in particular, it has not been used for the examples considered in this article.

Each TLS group contributes 20 refinement parameters. There will also typically be an isotropic or anisotropic overall scale factor and one isotropic displacement parameter per atom. In fact, there is some redundancy here, since one can for example increase all atomic B values in a TLS group while decreasing the mean of the **T** tensor by the same amount. There is no numerical instability in the current implementation since each set of parameters is refined independently, but such correlations would have to be accounted for in a full-matrix refinement. For the large TLS groups that we consider here, the use of TLS represents a large reduction in the number of parameters compared with the use of individual anisotropic displacement parameters. (We note that TLS parameterization can also be used for smaller groups; in this case, certain singularities must be avoided; Schomaker & Trueblood, 1968.)

Experience so far shows that it is best to refine the overall scale factor and TLS parameters first, while holding all atomic B factors fixed at a constant value. When the TLS parameters have converged reasonably well, the B factors can then be released to give a more detailed picture of atomic displacements.

2.4. Interpretation of results

The results of TLS refinement are 20 TLS parameters per chosen group. Besides improving the fit of the model to the observed data, it may be possible to extract useful information from the TLS values obtained. It must be stressed, however, that a fit of the TLS model to observed structure-factor amplitudes implies nothing about the relative phases of atomic displacements within the group. The model used assumes that all atoms within the group move in phase, but another model with some other phase relationship, keeping the same amplitudes, would give an equally good fit (although we note that the anisotropy makes the fit more stringent than models which fit TLS groups to isotropic displacement parameters). Furthermore, displacement parameters are well known to mop up errors as well as a variety of kinds of displacements and this must also be true of TLS parameters. With these caveats in mind, it may nevertheless be useful to analyse the TLS parameters. In the end, one must be guided by what is physically reasonable.

The 20 TLS parameters are output as three matrices referred to orthogonal axes and the chosen origin. Different representations can be obtained by rotating and shifting the coordinate frame: the *CCP*4 program *TLSANL* (Howlin *et al.*, 1993; Collaborative Computational Project, Number 4, 1994) is useful for doing this. Principal axes can be calculated for the translation and libration tensors and used to define a coordinate frame for the other tensors. These axes may yield a physical picture of the overall displacements represented by the TLS model, but the correlation between the translations and librations implied by the screw-rotation tensor can affect this picture. By shifting the origin to the so-called centre of reaction, the **S** tensor is made symmetric. Like any origin shift, the **T** tensor but not the **L** tensor is also changed by this shift.

A different picture of the TLS model can be obtained by abandoning the use of intersecting axes. Overall, the TLS parameters can be represented as three translations together with three screw displacements along three mutually perpendicular non-intersecting axes which lie parallel to the principal axes of the **L** tensor (Schomaker & Trueblood, 1968). The translations correspond to a reduced translation tensor $'\mathbf{T}$. The model parameters are then three translation amplitudes, three parameters describing the reduced translation principal axes, three parameters describing the screw axes orientations, six axes shifts, three rotation amplitudes and three screw pitches, giving a total of 21 parameters, one of which is indeterminate. Scheringer (1973) has argued from a description of the lattice dynamics that these six displacements are not statistically independent and that derived independent displacements necessarily mix translation and libration in a complex fashion. Nevertheless, this reduction is probably the simplest way of visualizing the possible displacements of the TLS group.

As well as studying the TLS parameters directly, one can of course derive individual atomic ADPs according to (5). Adding in the refined atomic B factors [or more generally the term $U_{\rm atom}$ in (1)] gives the total atomic ADP for atoms contained in a TLS group. Some atoms (for example, solvent atoms) may not be included in any TLS group, in which case the total atomic ADP is simply the refined atomic B factor. ADPs can be displayed graphically as thermal ellipsoids. For the purposes of discussion, it is often convenient to characterize the ADP by two scalar quantities, namely the equivalent isotropic displacement parameter $U_{\rm equiv} = B_{\rm equiv}/8\pi^2$ and the anisotropy A, defined as the ratio of the smallest to the largest eigenvalue of its ADP (Trueblood *et al.*, 1996). Values of the anisotropy A range from $A = 1$ for a completely isotropic atom to $A = 0$ for a prolate or oblate ellipsoid in the limit that the smallest dimension becomes zero.

3. Examples of application

3.1. Glyceraldehyde-3-phosphate dehydrogenase

The crystal structure of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) from the hyperthermophilic archaeon *Sulfolobus solfataricus* has recently been determined at 2.05 Å resolution (Isupov *et al.*, 1999). The enzyme crystallizes in space group $P4_12_12$, with two molecules per asymmetric unit, referred to in the model as chains *O* and *Q*. The biological unit

Table 1
Anisotropic refinement of GAPDH: refinement of TLS parameters only.

Atomic coordinates are fixed at the previously refined values and individual isotropic B factors are fixed either at the previously refined value (models 1 to 5) or at 20 Å^2 (models 6 to 10). For models 1, 2, 6 and 7 only the scale factors are refined. The TLS models are described in the text.

Model	Scaling	TLS model	Individual B factors (Å^2)	R factor (%)	R_{free} (%)
1	Isotropic	None	Previously refined	23.6	30.3
2	Anisotropic	None	Previously refined	22.9	29.5
3	Anisotropic	1 group	Previously refined	21.3	26.8
4	Anisotropic	2 groups	Previously refined	21.2	26.6
5	Anisotropic	4 groups	Previously refined	21.1	26.5
6	Isotropic	None	20	30.0	35.7
7	Anisotropic	None	20	29.5	35.2
8	Anisotropic	1 group	20	25.1	29.4
9	Anisotropic	2 groups	20	24.7	29.1
10	Anisotropic	4 groups	20	24.4	28.8

Table 2
Anisotropic refinement of GAPDH: complete refinement of coordinates, scale factors, TLS parameters and individual isotropic B factors.

All models use overall anisotropic scaling and start with individual isotropic B factors set to 20 Å^2. The TLS models are as in Table 1. Also indicated is whether or not NCS restraints are applied between chains O and Q.

Model	TLS model	NCS	R factor	R_{free}
11	None	No	23.8	30.4
12	1	No	21.4	25.9
13	2	No	21.2	26.2
14	4	No	21.1	25.8
15	None	Yes	25.0	30.3
16	4	Yes	22.0	25.7

is a tetramer, with chains P and R related to chains O and Q by a crystallographic twofold rotation.

A model of the enzyme, together with 12 sulfate ions and 525 waters, had previously been refined to an R factor of 22.9% and an R_{free} of 29.5% (Isupov *et al.*, 1999). No NCS restraints were imposed. Overall anisotropic scaling was used and individual atomic displacements were modelled by isotropic temperature factors. The average isotropic temperature factors for chains O and Q are 33.7 and 40.5 Å^2, respectively, thus indicating significantly different overall displacements for NCS-related molecules. It was suggested that the relatively high R factors were a consequence of unmodelled anisotropic displacements (Isupov *et al.*, 1999). We therefore attempted to model such anisotropic displacements with TLS refinement.

As a first test of TLS refinement, we investigated a variety of TLS models keeping all atomic coordinates and individual B factors fixed. The atomic coordinates used were those from the earlier refinement (Isupov *et al.*, 1999). Individual B factors were either fixed at the previously refined values or set to a constant value. Three TLS models were studied, as well as a model with no TLS parameters for comparison. The first treated both molecules in the asymmetric unit together as a single TLS group. The second treated each molecule individually as a TLS group. Finally, the third had one TLS group for the NAD-binding domain (residues 1–137 and 303–340) and one for the catalytic domain (residues 138–302) for each molecule, giving four TLS groups in total for the asymmetric unit. For all TLS groups used, all protein atoms were included in the group (including only main-chain atoms was found to give poorer results). Sulfate ions and water molecules were not included in any TLS group. There is no natural division of the protein molecule into more than two TLS groups until the level of secondary-structure elements is reached. One could, for example, model α-helices as TLS groups, which in this case would involve 12 groups per molecule. Such models have not been investigated.

The results are shown in Table 1. Models 1 to 5 correspond to a starting model with previously refined B factors (model 2 is in fact the result reported in Isupov *et al.*, 1999). The most obvious result is that including TLS gives a drop in R of up to 1.8% and in R_{free} of up to 3.0%, in comparison with an improvement in R and R_{free} of only 0.7 and 0.8%, respectively, upon including overall anisotropic scaling. However, while a TLS model gives a clear improvement over a simple scaling function, the difference between the three TLS models employed is relatively small.

Models 6 to 10 in Table 1 show the results when all individual B factors are set to a constant value before refining TLS parameters and overall scale parameters. While the values of R and R_{free} are larger owing to the lack of any individual displacement parameters, the effect of TLS refinement is even more marked in this case, with a drop in R of up to 5.1% and in R_{free} of up to 6.4% upon including TLS groups. The difference between the three TLS models (models 8 to 10) is now greater than before (models 3 to 5), although it is still relatively small compared with the improvement upon including TLS in the first place. It is noteworthy that the model with four TLS groups and constant individual B factors (model 10) gives a significantly better R_{free} factor than the model with refined individual B factors and no TLS parameters (model 2). Thus, the relatively simple model of anisotropic displacements with $4 \times 20 = 80$ parameters gives a better fit to the observed diffraction data than an isotropic model with 5268 parameters.

We conclude from these results that to get the most out of TLS refinement it is probably best to refine TLS parameters first, with individual B factors set to a constant value. Once the TLS refinement has converged, the TLS parameters are held fixed and atomic coordinates and individual B factors are refined. We have adopted this refinement strategy in all subsequent studies. The results for GAPDH are shown in Table 2 as models 11 to 14. Comparing models 12 to 14 with models 3 to 5 in Table 1, in which the order of TLS and B-factor refinements are reversed, we see there is little change in the R factor, but a slight improvement in R_{free}, providing *a posteriori* justification for the refinement strategy. Comparing the three TLS models (models 12 to 14) with each other, there is little difference and in fact R_{free} for model 13 is higher than for models 12 and 14. We now describe the results of models 12 to 14 in more detail.

The results of the full TLS refinement are atomic coordinates, **T**, **L** and **S** tensors for each TLS group chosen and

Table 3
Refined TLS parameters from model 12 of GAPDH.

(*a*) Elements of the **T**, **L** and **S** tensors in the orthogonal coordinate system, as output by the program.

T (Å^2)	0.1668	0.1773	0.0517	0.0579	−0.0816	−0.0715		
L (°^{2})	1.3785	0.4506	0.7920	0.0448	0.0830	0.3809		
S (Å°)	0.0008	0.2331	−0.0370	−0.1823	−0.1398	0.0888	0.0643	0.1396

(*b*) Eigenvectors and eigenvalues of the reduced translation tensor in the orthogonal coordinate system.

Group	**T** axis	Direction cosines			Mean-square **t** (Å^2)
1	1	0.639	0.656	−0.401	0.267
	2	−0.696	0.715	0.061	0.112
	3	0.327	0.240	0.914	−0.005

(*c*) Orientation and position of the non-intersecting screw axes, together with the rotation and pitch of the motion about these axes.

Group	Screw axis	Direction cosines			Position (Å)			Mean-square **l** (°^{2})	Pitch (Å)
1	1	0.968	0.131	0.213	103.968	10.739	57.700	1.403	0.888
	2	0.250	−0.528	−0.811	104.113	8.550	58.220	1.014	1.472
	3	0.006	0.839	−0.544	105.649	6.805	53.289	0.204	1.214

individual isotropic *B* factors for each atom. Table 3 gives the output TLS tensors for model 12. Also given are the values associated with the six independent translation and screw motions which form an equivalent description of the tensors (see §2.4). The latter are also given in Table 4 for model 14. For both models, the translational and screw motions are clearly anisotropic, although no single motion dominates. The mean-square magnitudes of these motions are of a similar magnitude to those seen in previous studies (see, for example, whole-molecule TLS refinement of ribonuclease A in Howlin *et al.*, 1989). Note that such apparently small angular displacements can lead to large linear displacements when the chosen TLS group is large, as it is here.

The full ADP for each atom is the sum of that derived from the TLS parameters according to (5) (if the atom belongs to a TLS group) and the individually refined *B* factor (hereafter termed the 'residual' *B* factor). The relative sizes of these two contributions can be gauged by comparing the equivalent isotropic displacement parameter of the TLS contribution $B_{\rm TLS}$ with the residual *B* factor $B_{\rm res}$. This is shown in Fig. 1 for model 14, with values averaged over main-chain atoms in each residue. Also shown is the sum of these two contributions and for comparison the *B* factors from the original refinement (Isupov *et al.*, 1999). It is immediately obvious that the present model mimics the original refinement very closely, with a correlation coefficient between the latter two curves of 0.944 (with similar values for models 12 and 13; see Table 5). This plot of course hides the anisotropic component of the present model. Secondly, the TLS contribution is clearly the more significant and accounts for the major variations of the total *B* factors. The residual *B* factors are largely constant over the protein chain. Localized peaks can nevertheless be seen which reflect local displacements not accounted for by the TLS parameterization. In general, such displacements may indicate model errors (Kuriyan & Weis, 1991) or unmodelled multiple conformations (Stec *et al.*, 1995), but the relatively small magnitude of these peaks in the present case suggest that this is unlikely.

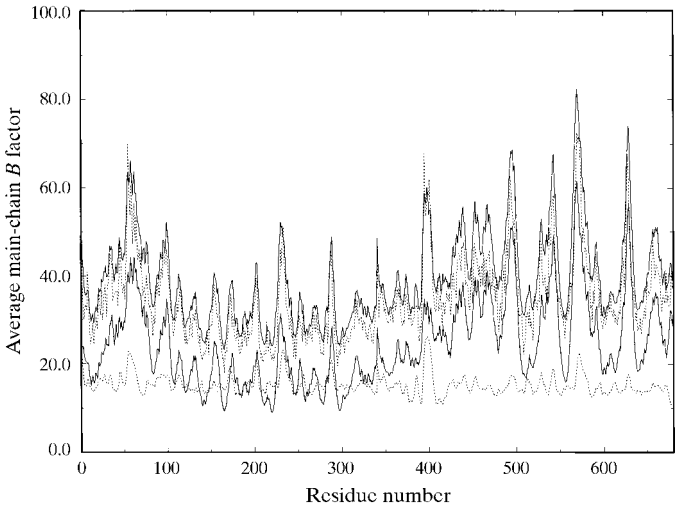

Figure 1
Contributions to the equivalent isotropic *B* factor for model 14 of GAPDH. The residual *B* factors $B_{\rm res}$ (lower dotted line), the contribution from the refined TLS parameters $B_{\rm TLS}$ (lower full line) and their sum (upper full line) are shown, together with the deposited *B* factors (upper dotted line) from PDB entry 1b7g. For each residue, the *B* factors are averaged over the main-chain atoms. Chain *O* consists of residues 1–340 and chain *Q* consists of residues 341–680.

Fig. 2 shows the residual *B* factors for the different TLS models. Close inspection shows that the inclusion of more TLS groups accounts for more of the variation in *B*, but the difference is relatively small and even the simplest TLS model gives a good description of the overall variation in *B*.

One aim of the TLS approach is to account for differences in displacement parameters between NCS-related molecules. Fig. 3 shows the refined *B* factors obtained in the original refinement for chains *O* and *Q* superimposed. The *B* factors for chain *Q* are in general higher than those for chain *O* and although the overall variation is similar, there are some obvious differences, for example the peaks around residue 120. Also shown in Fig. 3 are the residual *B* factors from the current approach using model 14 with four TLS groups; the contribution to the total *B* factor from the TLS parameters is excluded. The residual *B* factors for chains *O* and *Q* are clearly very similar to each other, in contrast to the overall *B* values. As noted above, the residual *B* factors vary little along the chains, but there are some features such as the peaks around residues 60 and 235. We suppose that these features, which

Table 4
Refined TLS parameters for the four groups of model 14 of GAPDH.

Quantities given are as in Table 3(*b*) and 3(*c*).

Group	**T** axis	Direction cosines			Mean-square **t** (Å^2)
1	1	0.415	0.871	−0.262	0.280
	2	−0.901	0.434	0.013	0.144
	3	0.125	0.231	0.965	0.003
2	1	0.656	0.737	−0.161	0.232
	2	−0.743	0.668	0.028	0.113
	3	0.128	0.101	0.987	−0.013
3	1	0.386	0.802	−0.456	0.389
	2	−0.841	0.509	0.184	0.165
	3	0.379	0.312	0.871	0.068
4	1	0.487	0.695	−0.529	0.368
	2	−0.789	0.610	0.074	0.140
	3	0.374	0.381	0.845	0.079

Group	Screw axis	Direction cosines			Position (Å)			Mean-square **l** (°^{2})	Pitch (Å)
1	1	−0.601	0.574	0.556	92.776	12.609	65.970	1.889	0.761
	2	0.476	−0.302	0.826	96.843	18.404	63.179	1.450	0.529
	3	0.642	0.761	−0.092	91.629	17.194	61.530	1.151	0.582
2	1	0.646	0.435	0.627	93.054	14.075	52.091	2.468	−1.457
	2	0.763	−0.356	−0.539	93.070	10.256	53.479	1.625	1.517
	3	−0.011	0.827	−0.562	96.318	12.569	52.590	0.717	1.577
3	1	0.611	0.778	0.143	120.121	13.130	56.557	2.746	4.436
	2	0.791	−0.603	−0.102	123.691	16.542	58.445	1.123	0.896
	3	0.007	0.176	−0.984	118.460	7.635	55.927	1.066	12.366
4	1	0.382	0.355	0.853	109.857	2.976	51.327	2.633	−4.339
	2	0.901	−0.350	−0.258	109.986	1.168	53.996	1.815	2.223
	3	0.207	0.867	−0.454	108.431	0.582	48.580	0.817	9.046

occur for both chains, are intrinsic properties of the molecule and are not dependent on its location in the asymmetric unit. (In fact, the two peaks around residues 60 and 235 correspond to exposed α-helical regions.) Similar results are found with the models 12 and 13 with one and two TLS groups. The correlation coefficient between the average residual *B* factors for chains *O* and *Q* is 0.752, 0.812 and 0.821 for one, two and four TLS groups, respectively, compared with 0.537 without any modelling of TLS (see Table 5).

The refinement with four TLS groups was repeated with NCS restraints imposed (model 16), as well as the refinement with no TLS groups defined (model 15). In both cases, there is an increase in the *R* factor and a very small decrease in $R_{\rm free}$ (see Table 2) compared with the equivalent refinement with no NCS restraints. In fact, since the reflections used to calculate $R_{\rm free}$ are not completely independent of the model by virtue of the NCS, the improvement of the $R_{\rm free}$ factor is probably more significant than it appears. Hence, removal of NCS restraints is possibly not justified. For models 15 and 16, the correlation coefficients between *B* factors for chains *O* and *Q* are, by construct, high (see Table 5). While the use of NCS restraints is justified by $R_{\rm free}$ for both models, the highly correlated *B* factors that are produced by *B*-factor restraints are only physically reasonable for model 16.

So far, we have concentrated on the predicted equivalent isotropic displacement parameters, but the reason the TLS models give better agreement with observed structure factors

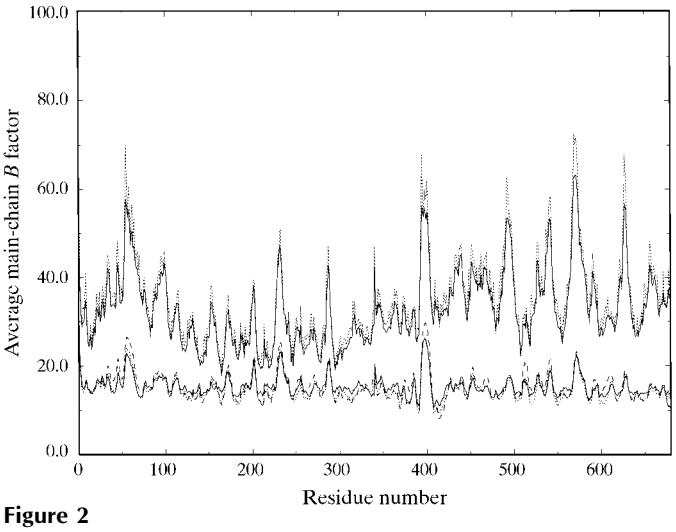

Figure 2
$B_{\rm res}$ from models 11 (upper full line), 12 (dashed line), 13 (lower dotted line) and 14 (lower full line) of GAPDH, together with the deposited *B* factors (upper dotted line) from PDB entry 1b7g for comparison. For each residue, the *B* factors are averaged over the main-chain atoms. Chain *O* consists of residues 1–340 and chain *Q* consists of residues 341–680.

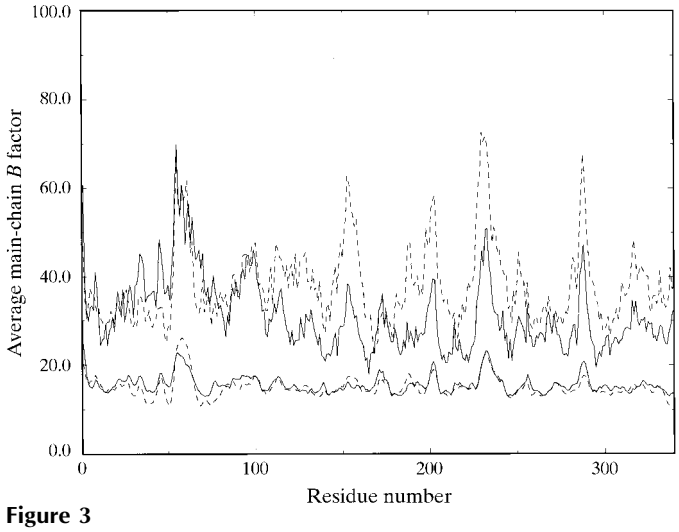

Figure 3
The top two curves are the deposited *B* factors of GAPDH from PDB entry 1b7g for chains *O* (full line) and *Q* (dashed line) superimposed. The bottom two curves are $B_{\rm res}$ from model 14 for chains *O* (full line) and *Q* (dashed line). For each residue, the *B* factors are averaged over the main-chain atoms. The contributions from the TLS parameters are not included.

than simple *B*-factor refinement is of course the inclusion of an anisotropic component. The anisotropy *A* (see §2.4) calculated from the ADPs, averaged over main-chain atoms, is plotted in Fig. 4 along with the equivalent isotropic *U* value. There is a clear anti-correlation between the two curves, *i.e.* large anisotropy (small *A*) correlates with large $U_{\rm equiv}$. This is a consequence of the libration modelled by the TLS and is particularly clear for chain *Q*, where the TLS contribution is most significant (see Fig. 1). The correlation is poorer in regions where the residual *B* factors make a significant contribution, for example around residue 60 of chain *Q*. Most *A* values lie in the range 0.3–0.6, which is similar to the range of values seen in models with individually refined ADPs (Merritt, 1999). The anisotropy implied by the TLS model can therefore be considered reasonable.

Turning to the refined TLS parameters themselves, Fig. 5 shows the screw axes for model 12. The axes lie parallel to the libration axes of the TLS group. The screw axes do not intersect (although unclear on the scale of the figure) and the starting point of each arrow is shifted from the origin of the TLS group according to equation (17) of Schomaker & Trueblood (1968). The length of each axis is proportional to the mean-square libration. The results for models 13 and 14 can be displayed similarly. One can also generate TLS parameters for symmetry mates and in particular for chains *P* and *R* (related to *O* and *Q* by a crystallographic twofold rotation) which complete the biological tetramer. Direct viewing of the axes representing the TLS parameters does not in this case lead to any simple physical interpretation, so in Fig. 6 we instead show the equivalent isotropic *B* factor derived from the TLS parameters for the biological tetramer. There is a region of low *B* values around the interface between chains *O* and *P* and a general increase away from this region, with the highest values around the periphery of chains *Q* and *R*. There is no indication of separate motions for the eight TLS groups included; rather, it seems that the individual TLS motions combine to give a fairly simple overall motion of the tetramer.

We have above considered three TLS models, with increasing detail included going from one TLS group to four TLS groups. Values of $R_{\rm free}$ suggest that there is only marginal improvement in going to more sophisticated models (though we stress that the improvement over simple anisotropic scaling is very significant). Close examination of the results, such as the TLS-derived values given in Tables 3 and 4 and the distribution of *B* factors shown in Fig. 6, hints that the models with two and four TLS groups may in fact be modelling similar overall motion to the model with one TLS group. Hence, while

Table 5
Correlation coefficients cc_B between different calculations of the equivalent isotropic displacement parameters for GAPDH.

The latter are averaged over main-chain atoms on a residue-by-residue basis. $B_{\rm TLS}$ is the contribution of TLS parameters to the equivalent isotropic displacement parameter. $B_{\rm res}$ is the residual *B* factor.

First calculation	Second calculation	cc_B
Deposited *B* factors	Sum of $B_{\rm TLS}$ and $B_{\rm res}$: model 12	0.938
Deposited *B* factors	Sum of $B_{\rm TLS}$ and $B_{\rm res}$: model 13	0.941
Deposited *B* factors	Sum of $B_{\rm TLS}$ and $B_{\rm res}$: model 14	0.944
Deposited *B* factors: chain *O*	Deposited *B* factors: chain *Q*	0.537
$B_{\rm res}$: model 12, chain *O*	$B_{\rm res}$: model 12, chain *Q*	0.752
$B_{\rm res}$: model 13, chain *O*	$B_{\rm res}$: model 13, chain *Q*	0.812
$B_{\rm res}$: model 14, chain *O*	$B_{\rm res}$: model 14, chain *Q*	0.821
Refined *B* factors: model 15, chain *O*	Refined *B* factors: model 15, chain *Q*	0.989
$B_{\rm res}$: model 16, chain *O*	$B_{\rm res}$: model 16, chain *Q*	0.999

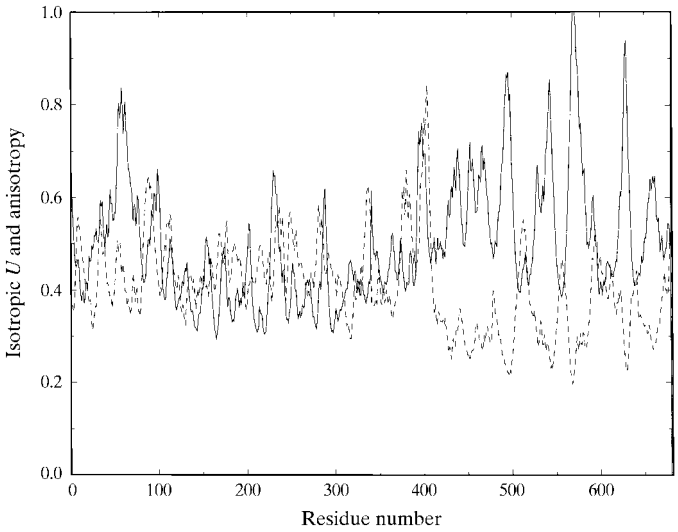

Figure 4
The equivalent isotropic *U* factor $U_{\rm equiv}$ (full line) and the anisotropy *A* (dashed line) for model 14 of GAPDH. $U_{\rm equiv}$ is equal to $B_{\rm equiv}/8\pi^2$. *A* is defined as the ratio of the smallest to the largest eigenvalue of the ADP. For each residue, $U_{\rm equiv}$ and *A* are averaged over the main-chain atoms. Chain *O* consists of residues 1–340 and chain *Q* consists of residues 341–680.

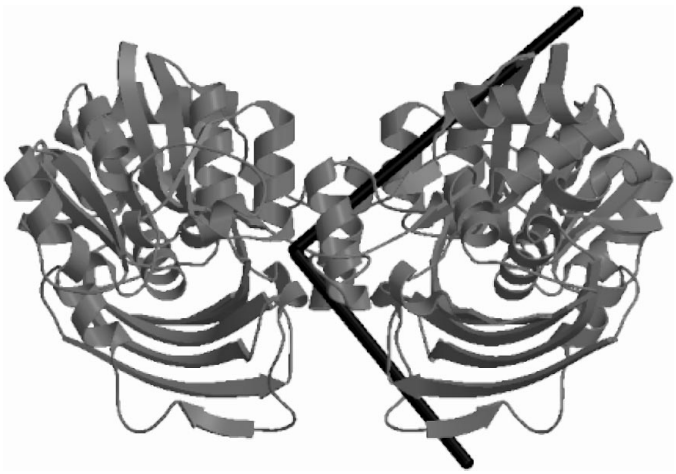

Figure 5
The non-intersecting screw axes for model 12 of GAPDH superimposed on chains *O* (left) and *Q* (right). The non-crystallographic twofold axis relating the two chains is vertical in the plane of the page. For each chain, the lower half (including the clearly visible β-sheet) is the catalytic domain and the upper half is the NAD-binding domain. The screw axes lie parallel to the libration axes of the TLS group and the length of each axis is proportional to the mean-square libration. Two axes are clearly visible in the figure; the third and shortest is approximately perpendicular to the plane of the page and is hidden. This figure was prepared using *Molscript* (Kraulis, 1991) and *Raster3D* (Merritt & Bacon, 1997).

it has been informative to investigate a larger number of TLS groups for this study, in practice there is probably little justification to include more than the single TLS group.

3.2. GerE

The second example is GerE, a transcription activator from *Bacillus subtilis* (Ducros *et al.*, 2001). X-ray data have been collected in space group *C*2 to a resolution of 2.05 Å. The model of the asymmetric unit consists of six copies of the molecule (chains *A* to *F*), together with four sulfate ions, three glycerol molecules and 325 waters. Chains *A*, *C* and *F* are related to chains *B*, *D* and *E* by a non-crystallographic twofold rotation. The monomer contains 74 residues, although the N-terminus is disordered and between four and ten residues (depending on the chain) are omitted from the model. There is also a disordered region between residues 28 and 42 of chain *F* which has not been modelled.

The model has been refined with and without TLS parameters and with or without NCS restraints. The TLS model consists of a single TLS group for each of the six protein chains. Solvent molecules were not included in any TLS group. As for the previous example, all isotropic *B* values are first set to a constant value, TLS parameters are then refined and finally coordinates and *B* values are refined. When used, NCS restraints are applied to residues 12–72 for each protein chain, with medium restraints for main-chain atoms and loose restraints for side-chain atoms. In all cases, overall anisotropic scaling is used.

Table 6
Anisotropic refinement of GerE.

All models use overall anisotropic scaling and start with individual isotropic *B* factors set to 40 Å^2. The correlation coefficient cc_B is the average over all pairs of chains taken from *A* to *E* of the correlation coefficient of the residual *B* factor, calculated from residues 12 to 72.

Model	TLS model	NCS	*R* factor	R_{free}	cc_B
1	None	No	21.9	29.3	0.519
2	None	Yes	22.5	30.0	0.553
3	6	No	21.3	27.1	0.510
4	6	Yes	21.4	27.2	0.816

The *R* and R_{free} factors for each of the four cases are listed in Table 6. Use of TLS parameters gives a reduction of 2.2% in R_{free} when no NCS restraints are used and 2.8% when NCS restraints are used. Fig. 7 shows the relative contributions of the TLS parameters and the individually refined *B* factors for model 3 and, as for GAPDH, we see that the TLS parameters account for most of the large-scale variations. Some quantities derived from the TLS parameters are collected together in Table 7. The magnitudes are similar to those for GAPDH, although there are some particularly large librations. In particular, chains *E* and *F* have the largest **T** and **L** parameters and these chains are indeed the least ordered part of the structure as judged by the quality of the electron density. The TLS parameterization thus gives a rough measure of the degree of order.

Fig. 8 compares the refined *B* factors for model 1 with the residual *B* factors of model 3, with the values for the six chains superimposed. With no TLS refinement, the values vary widely showing that the different chains have widely different

Figure 6
The biological tetramer of GAPDH viewed in the same orientation as Fig. 5. Chains *O* and *Q* are in the top left and top right quadrants of the figure, respectively. Chains *P* and *R* are at bottom left and bottom right respectively, with *R* partly in front of *P*. The multimer is coloured according to the value of B_{TLS} derived from the TLS parameters of model 14, with low values shown in blue and high values in red. This figure was prepared using *Molscript* (Kraulis, 1991) and *Raster3D* (Merritt & Bacon, 1997).

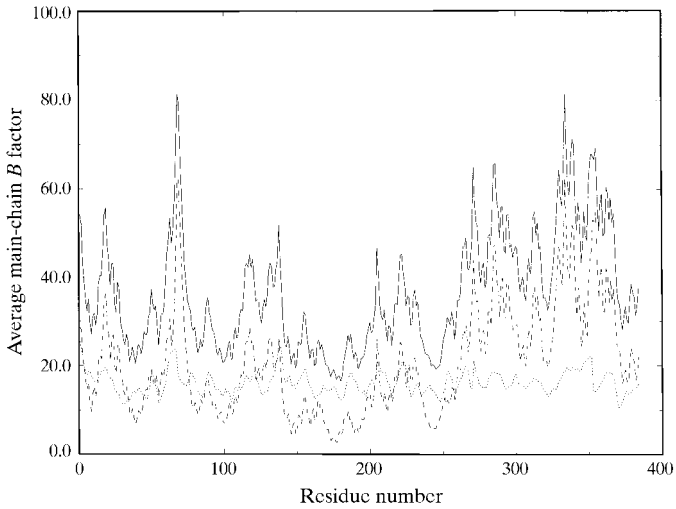

Figure 7
Contributions to the equivalent isotropic *B* factor for GerE refined with six TLS groups and no NCS restraints (model 3). The residual *B* factors B_{res} (dotted line), the contribution from the refined TLS parameters B_{TLS} (dashed line) and their sum (full line) are shown. For each residue, the *B* factors are averaged over the main-chain atoms. Chain *A* consists of residues 1–67, chain *B* 68–137, chain *C* 138–204, chain *D* 205–270, chain *E* 271–334 and chain *F* 335–384.

Table 7
Refined TLS parameters for GerE.

The eigenvalues of the reduced translation tensor are given for each TLS group, followed by the rotation and pitch of the motion about the three non-intersecting screw axes.

Chain	Mean-square **t** (Å^2)			Mean-square **l** ($^\circ$^{2})			Pitch (Å)		
A	0.007	0.163	0.076	1.279	12.445	2.976	0.336	−0.042	0.030
B	0.075	0.041	0.102	8.031	1.059	1.865	0.226	−4.329	1.482
C	0.054	0.022	−0.002	2.867	5.991	1.048	0.075	0.137	0.990
D	0.064	−0.010	0.127	4.027	2.084	5.165	0.731	−1.498	0.034
E	0.391	0.144	0.088	1.943	20.395	2.922	6.666	−1.088	3.163
F	0.035	0.146	0.320	18.639	5.283	7.169	0.348	1.113	−1.726

displacement characteristics. After TLS refinement, the residual *B* factors are much closer in overall magnitude, though some differences persist. Clearly, the application of NCS restraints in the latter case is more reasonable and this would account for the small increase of $R_{\rm free}$ from 27.1 to 27.2, compared with the larger increase from 29.3 to 30.0 on applying NCS restraints to the model with no TLS parameters. Table 7 also lists the average correlation coefficient between residual *B* factors for NCS-related chains. The larger increase in the correlation coefficient on application of NCS restraints when TLS is used again testifies to the use of restraints being more reasonable in this case.

4. Conclusions

In conclusion, we have implemented the refinement of TLS parameters in the macromolecular refinement program *REFMAC*. Test cases show that a large reduction in *R* and $R_{\rm free}$ factors can be obtained by the inclusion of a few large TLS groups. Such an extension of the model requires the addition of just 20 refinement parameters per TLS group, which makes the method amenable at lower resolution than individual anisotropic refinement. For the GAPDH example studied in §3.1, we found that inclusion of one, two or four TLS groups (20, 40 or 80 refinement parameters), keeping individual *B* factors constant, in fact gave a better $R_{\rm free}$ factor than the model with refined individual *B* factors and no TLS parameters (5268 refinement parameters). This is because, in addition to reproducing the main features of the variation in *B* factors, the TLS model includes anisotropy, which can be very important.

The success of the TLS model does not of course prove the validity of the rigid-body assumption. Attempts to model X-ray diffuse scattering by whole-molecule rigid-body displacements have given mixed results with, for example, a good fit for tetragonal lysozyme crystals and a poor fit for orthorhombic lysozyme crystals (Perez *et al.*, 1996; Héry *et al.*, 1998). It is possible that crystal packing restricts whole-molecule displacements (Héry *et al.*, 1998), implying that the rigid-body assumption is more reasonable for loosely packed protein crystals.

The prime role for TLS parameters in the present context is the removal of large-scale domain or molecular motions which otherwise mask local atomic displacements. Removal of this motion leaves residual *B* factors which are chemically more meaningful and which may have NCS restraints applied. This was illustrated by GAPDH, which has two molecules in the asymmetric unit with significantly different overall displacement parameters. After inclusion of TLS parameters, the NCS relationships between residual *B* factors became much more apparent and application of NCS restraints becomes more reasonable. Similar results were found for GerE, where the six molecules in the asymmetric unit again have very different displacement parameters which are accounted for well by the TLS parameterization.

Collective variables are clearly a powerful way of including more sophisticated models of displacements while keeping the number of refinement parameters at a justifiable level. Pseudo-rigid-body variables, described by TLS parameters, are a basic example of this and are a necessary precursor to other methods. Future work will investigate the use of other collective variables, such as normal modes (Diamond, 1990; Kidera & Go, 1990, 1992). In situations where individual anisotropic refinement is not viable, there are likely to be many other levels of modelling which may be considered.

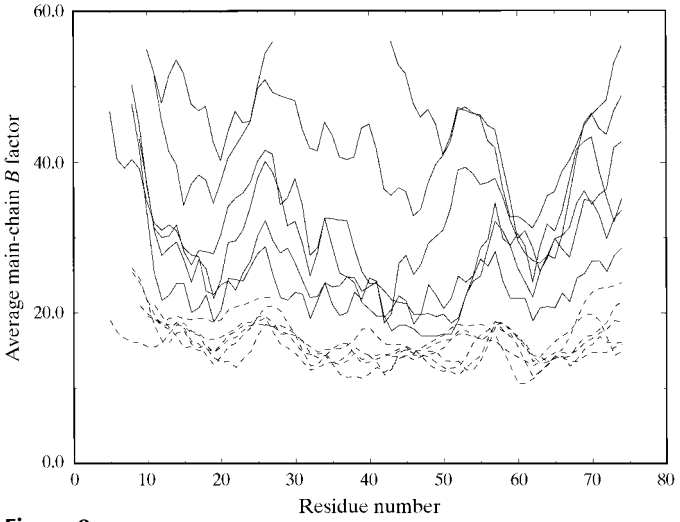

Figure 8
Refined main-chain *B* factors for the six chains of GerE superimposed when refined with no NCS restraints and no TLS (model 1, full lines), compared with $B_{\rm res}$ after TLS refinement (model 3, dashed lines). For each residue, the *B* factors are averaged over the main-chain atoms. The contributions from the TLS parameters are not included. There is a gap in the curves for chain *F* where the protein could not be modelled.

MDW and GNM are supported by the BBSRC through the CCP4 grant (B10200). MDW is grateful to Miroslav Papiz for useful comments.

References

Artymiuk, P. J., Blake, C. C. F., Grace, D. E. P., Oatley, S. J., Phillips, D. C. & Sternberg, M. J. E. (1979). *Nature (London)*, **280**, 563–568.
Brock, C. P., Schweizer, W. B. & Dunitz, J. D. (1985). *J. Am. Chem. Soc.* **107**, 6964–6970.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cruickshank, D. W. J. (1956). *Acta Cryst.* **9**, 754–757.
Diamond, R. (1990). *Acta Cryst.* A**46**, 425–435.
Driessen, H., Haneef, M. I. J., Harris, G. W., Howlin, B., Khan, G. & Moss, D. S. (1989). *J. Appl. Cryst.* **22**, 510–516.
Ducros, V. M.-A., Lewis, R. J., Verma, C. S., Dodson, E. J., Leonard, G., Turkenburg, J. P., Wilkinson, A. J. & Brannigan, J. A. (2001). In preparation.
Dunitz, J. D. & White, D. N. J. (1973). *Acta Cryst.* A**29**, 93–94.
Harata, K., Abe, Y. & Muraki, M. (1998). *Proteins Struct. Funct. Genet.* **30**, 232–243.
Harata, K., Abe, Y. & Muraki, M. (1999). *J. Mol. Biol.* **287**, 347–358.
Harris, G. W., Pickersgill, R. W., Howlin, B. & Moss, D. S. (1992). *Acta Cryst.* B**48**, 67–75.
Hayward, S. & Berendsen, H. J. C. (1998). *Proteins Struct. Funct. Genet.* **30**, 144–154.
Héry, S., Genest, D. & Smith, J. C. (1998). *J. Mol. Biol.* **279**, 303–319.
Hirshfeld, F. (1974). *Acta Cryst.* A**32**, 239–244.
Holbrook, S. R., Dickerson, R. E. & Kim, S. (1985). *Acta Cryst.* B**41**, 255–262.
Holbrook, S. R. & Kim, S. (1984). *J. Mol. Biol.* **173**, 361–388.
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.
Isupov, M. N., Fleming, T. M., Dalby, A. R., Crowhurst, G. S., Bourne, P. C. & Littlechild, J. A. (1999). *J. Mol. Biol.* **291**, 651–660.
Kidera, A. & Go, N. (1990). *Proc. Natl Acad. Sci.* **87**, 3718–3722.
Kidera, A. & Go, N. (1992). *J. Mol. Biol.* **225**, 457–475.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Kuriyan, J. & Weis, W. I. (1991). *Proc. Natl Acad. Sci. USA*, **88**, 2773–2777.
Merritt, E. A. (1999). *Acta Cryst.* D**55**, 1109–1117.
Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.
Murshudov, G. N., Davies, G. J., Isupov, M. N., Krzywda, S. & Dodson, E. J. (1998). *CCP4 Newslett.* **35**, 37–42.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–253.
Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. (1999). *Acta Cryst.* D**55**, 247–255.
Papiz, M. Z. (2000). Personal communication.
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.
Pawley, G. S. (1966). *Acta Cryst.* **20**, 631–638.
Pérez, J., Faure, P. & Benoit, J.-P. (1996). *Acta Cryst.* D**52**, 722–729.
Rollett, J. S. (1970). *Crystallographic Computing*, edited by F. R. Ahmed, S. R. Hall & C. P. Huber, p. 167. Copenhagen: Munksgaard.
Rosenfield, R. E., Trueblood, K. N. & Dunitz, J. D. (1978). *Acta Cryst.* A**34**, 828–829.
Šali, A., Veerapandian, B., Cooper, J. B., Moss, D. S., Hofmann, T. & Blundell, T. L. (1992). *Proteins Struct. Funct. Genet.* **12**, 158–170.
Scheringer, C. (1973). *Acta Cryst.* A**29**, 554–570.
Schneider, T. R. (1996). *Proceedings of the CCP4 Study Weekend. Macromolecular Refinement*, edited by E. Dodson, M. Moore, A. Ralph & S. Bailey, pp. 133–144. Warrington: Daresbury Laboratory.
Schomaker, V. & Trueblood, K. N. (1968). *Acta Cryst.* B**24**, 63–76.
Schomaker, V. & Trueblood, K. N. (1998). *Acta Cryst.* B**54**, 507–514.
Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.
Sheriff, S. & Hendrickson, W. A. (1987). *Acta Cryst.* A**43**, 118–121.
Stec, B., Zhou, R. & Teeter, M. M. (1995). *Acta Cryst.* D**51**, 663–681.
Sternberg, M. J. E., Grace, D. E. P. & Phillips, D. C. (1979). *J. Mol. Biol.* **130**, 231–253.
Sygusch, J. (1976). *Acta Cryst.* B**32**, 3295–3298.
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.
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.