electronic reprint

Journal of
Applied Crystallography
ISSN 0021-8898
Editor: **Gernot Kostorz**

TLSMD web server for the generation of multi-group TLS models

Jay Painter and Ethan A. Merritt

Journal of
Applied Crystallography
ISSN 0021-8898

Received 14 September 2005
Accepted 23 November 2005

TLSMD web server for the generation of multi-group TLS models

Jay Painter and Ethan A. Merritt*

Biomolecular Structure Center, Department of Biochemistry, University of Washington, Seattle, WA 98195-7742, USA. Correspondence e-mail: merritt@u.washington.edu

The *TLSMD* web server extracts information about dynamic properties of a protein based on information derived from a single-crystal structure. It does so by analyzing the spatial distribution of individual atomic thermal parameters present in an input structural model. The server partitions the protein structure into multiple, contiguous chain segments, each segment corresponding to one group in a multi-group description of the protein's overall dynamic motion. For each polypeptide chain of the input protein, the analysis generates the optimal partition into two segments, three segments, ... up to 20 segments. Each such partition is optimal in the sense that it is the best approximation of the overall spatial distribution of input thermal parameters in terms of N chain segments, each acting as a rigid group undergoing TLS (translation/libration/screw) motion. This multi-group TLS model may be used as a starting point for further crystallographic refinement, or as the basis for analyzing inter-domain and other large-scale motions implied by the crystal structure.

1. Introduction

The TLS (translation/libration/screw) formalism can be used to describe bulk motion of an arbitrarily large set of atoms acting as a rigid body (Schomaker & Trueblood, 1968). Even if this group of atoms does not, in fact, truly behave as a rigid body, the TLS description may nevertheless provide a very useful approximation. This is in particular true when the total amplitude of motion is small, as is the case for atoms in a well ordered protein in a crystal lattice at 100 K.

The bulk vibrational motion of a protein within the crystal lattice may be approximated by assigning the entire protein molecule to a single TLS group. The magnitude and specific TLS parameter values for such a bulk motion are obviously specific to the particular crystal lattice packing and symmetry. While such a single TLS-group model provides little insight into protein structure *per se*, it can significantly improve the crystallographic model by yielding better values of $F_{\rm calc}$, and hence lower crystallographic residuals R and $R_{\rm free}$. This in turn may lead to improved electron density maps and ultimately to a better structural model. Such single-group TLS models are easily generated and refined by the *CCP4* program *REFMAC5* (Winn *et al.*, 2001; Collaborative Computational Project, Number 4, 1994).

Partitioning the protein into more than one TLS group will often yield significant additional improvement in the crystallographic residuals R and $R_{\rm free}$, although in general the incremental improvement is less dramatic than the change to a bulk TLS model from no TLS model at all.

Notwithstanding this diminishing sensitivity of the overall crystallographic R factor to the introduction of larger numbers of TLS groups, the correct identification of these groups can be of substantial biological significance. It allows the inference of dynamic behavior, *e.g.* inter-domain hinge motions, directly from a single-crystal structure (Wilson & Brunger, 2000; Papiz *et al.*, 2003; Chaudhry *et al.*, 2004; Bernett *et al.*, 2004).

2. The *TLSMD* web server

2.1. Input models

The *TLSMD* web server automates the optimal partitioning of an existing protein structural model into multiple, contiguous TLS groups. It is implemented as a set of Python scripts, making extensive use of the *mmLib* programming toolkit (Painter & Merritt, 2004). Some routines have been re-coded in C in order to reduce the required computational time substantially. Detailed discussion of the algorithms used and of the physical significance of the results will be presented elsewhere. In brief, the server calculates a best-fit TLS model for every possible subsegment of a protein chain, and then compares the distribution of thermal parameters predicted by that TLS description with the actual thermal parameters present in the input model. Each chain fragment is then assigned a residual that describes the goodness of fit between the predicted and observed thermal parameters. The optimal N-group model is then chosen by selecting $N-1$ break points in the protein chain such that the residual sum of the N-component TLS groups is lower than for any other selection of break points.

The web server is used by uploading a protein structural model from a client machine's web browser *via* a submission form. The submitted model should ideally have been refined with individual atomic displacement parameters (ADPs), either isotropic or anisotropic. If a pre-existing TLS model is found in the REMARK statements of the input PDB file, its contributions are automatically added to the individual ADPs of the atoms in the structure. In this case, it is important that the temperature factors contained in the ATOM records are residual magnitudes from TLS + $B_{\rm iso}$ refinement,

and do not already contain the contribution from the TLS model. This is the current default behavior of *REFMAC5*.

2.2. Returned results

The *TLSMD* analysis is computationally intensive, sometimes requiring several hours of CPU time. The time required grows with the square of the maximum chain length. Therefore, submitted protein structures are queued for execution, and e-mail notification is sent when the run has been completed. The progression of a submitted job through the queue may be monitored *via* the web interface.

The result of TLSMD analysis is presented as a series of web pages, and as dynamically generated files for use with client-side applications that perform further visualization or refinement. Additional web pages provide tabular and statistical data related to the properties of individual TLS groups. Future versions of the *TLSMD* server will attempt to offer automated assistance in interpreting this rich pool of information.

2.2.1. Goodness of fit. For each polypeptide chain in the submitted structure, the server returns a plot of the extent to which an N-group TLS model adequately describes the spatial distribution of ADPs in the submitted structure. Increasing the allowed number of groups from N to $N + 1$ does not always improve the overall residual (Fig. 1), but the residual should be monotonic, non-increasing with N.

2.2.2. Partition of chains into multiple contiguous segments. For each polypeptide chain, the server returns a simple diagram of the optimal partition into $1, 2, \ldots N$ contiguous chain segments. Two such diagrams are shown in Fig. 1. Each row of the diagram embeds a hyperlink to a more detailed presentation of that specific chain partition (Fig. 2).

If multiple copies of the chain are present, additional plots are generated to compare the corresponding chain partitions that have been independently generated for each copy of the chain.

2.2.3. Graphics and visualization. For each partition of each chain into multiple TLS groups, the server generates three visualization aids. A static image of the multi-group TLS model is shown on the web page (Fig. 2). A link is provided to an animation of the corresponding TLS group motion interpreted as three independent orthogonal screw displacements, which may be viewed in the client browser *via* the *Jmol* java applet (JMol Team, 2002). A set of files may be downloaded for more comprehensive interactive visualization using the program *TLSView* (Painter & Merritt, 2005).

2.2.4. Refinement. The user may select for each chain a preferred number of TLS groups. This set of choices, *e.g.* two groups for chain A and four groups for chain B, is then used by the server to generate a corresponding modified PDB file and a matching *TLSIN* file that may be input to the program *REFMAC5* for further crystallographic refinement. The atomic coordinates in this PDB file are identical to those in the original input file, but the $B_{\rm iso}$ or U^{ij} parameters are modified to describe an incremental difference from the underlying TLS model, rather than the original absolute parameter value. Without this modification, subsequent TLS refinement in *REFMAC5* may not behave well numerically; hence it is important to download the modified PDB file rather than recycling the input file into further refinement.

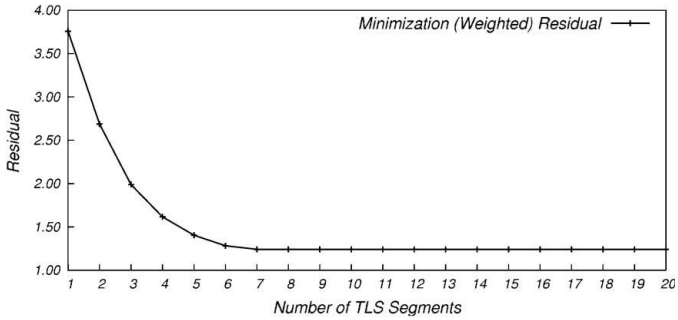
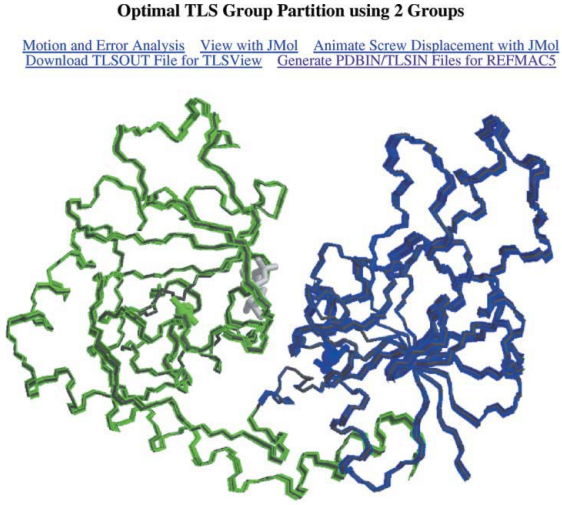


Figure 1
The least-squares residual resulting from optimal partitioning of the protein chain into an increasing number of segments. Each segment s contributes a partial residual $R_s = \sum w_i \left(U_i^{\rm obs} - U_i^{\rm tls}\right)^2$ where the values of $U^{\rm obs}$ are the ADPs for all atoms i in group s as read from the uploaded structure file, and the values of $U^{\rm tls}$ are the corresponding ADPs predicted by the TLS model fit to this segment. For a given partition into N segments, each with its own associated TLS fit, the total residual plotted is the sum of the N components, $R = \sum_{s=1}^{N} R_s$. The shape of this curve depends on the individual protein structure. In general, the shape of the curve parallels the expected decrease in crystallographic residuals R and $R_{\rm free}$ when the corresponding N-group model is adopted for crystallographic refinement. The curve shown here is for PDB entry 3HVP. The residual R drops sharply as the number of TLS groups is increased from 1 to 5, but then becomes relatively flat if additional groups are added. This is an imperfect indication that using the five-group TLS model for further refinement in *REFMAC5* will yield improved R factors, but little further improvement will result from partitioning into six or more groups.

3. Availability

The analysis code used by *TLSMD* is hosted by SourceForge, http://pymmlib.sourceforge.net/, as is the underlying crystallographic toolkit, *mmLib*. The source code is currently available under the Artistic License, but other licensing arrangements are possible.



Analysis with 2 TLS Groups											
Input Structure						TLS Predictions					
Color	Segment	Residues	Atoms	<B>	<Aniso>	LSQR	LSQR/Res	eval(T^t) *B*	eval(L) *DEG*^{2}	<B>	<Aniso>
	1-166	166	1352	22.5	1.00	8.8475	0.0533	20.0 18.5 17.5	0.96 0.45 0.29	21.5	0.80
	167-351	185	1517	19.8	1.00	8.7767	0.0474	17.5 16.6 15.8	0.84 0.47 0.33	19.0	0.81

Figure 2
The top-level summary of information returned by the *TLSMD* web server, describing a specific partition of a single polypeptide chain into multiple TLS groups. The example shown describes a two-group partition of DNA β-glucosyltransferase (PDB accession code 1JG6). The libration axes of the two rigid groups identified by this analysis of a single-crystal structure may be compared with that deduced by morphing multiple conformations of the same protein as observed in multiple crystal structures. Many such multi-crystal analyses are available in the Database of Macromolecular Movements compiled by Gerstein and coworkers (Gerstein *et al.*, 1999).

This work was supported by NIH award GM62617.

References

Bernett, M. J., Somasundaram, T. & Blaber, M. (2004). *Proteins Struct. Func. Bioinf.* **57**, 626–634.
Chaudhry, C., Horwich, A. L., Brunger, A. T. & Adams, P. D. (2004). *J. Mol. Biol.* **342**, 229–245.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Gerstein, M. B., Jansen, R., Johnson, T., Park, B. & Krebs, W. (1999). *Studying Macromolecular Motions in a Database Framework: From Structure to Sequence*, pp. 401–442. Norwell, MA: Kluwer Academic/Plenum Press.
JMol Team (2002). *Jmol molecular viewer*, http://www.jmol.org/.
Painter, J. & Merritt, E. A. (2004). *J. Appl. Cryst.* **37**, 174–178.
Painter, J. & Merritt, E. A. (2005). *Acta Cryst.* D**61**, 465–471.
Papiz, M. Z., Prince, S. M., Howard, T., Cogdell, R. J. & Isaacs, N. W. (2003). *J. Mol. Biol.* **341**, 1523–1538.
Schomaker, V. & Trueblood, K. N. (1968). *Acta Cryst.* B**24**, 63–76.
Wilson, M. A. & Brunger, A. T. (2000). *J. Mol. Biol.* **301**, 1237–1256.
Winn, M., Isupov, M. & Murshudov, G. (2001). *Acta Cryst.* D**57**, 122–133.