

TLS and all that

Ethan A Merritt

CCP4 Summer School 2011 (Argonne, IL)

Abstract

We can never know the position of every atom in a crystal structure perfectly. Each atom has an associated positional uncertainty. When we say that an atom has coordinates $[x,y,z]$, what we really mean is that averaged over all the copies of that atom in the crystal, and averaged over the time of measurement, $[x,y,z]$ is the mean position of the center of the atom. But at some times, or in some copies of the unit cell, the atom is displaced from this mean position. Factors contributing to this displacement include contributions from vibration of the atom itself, from vibrational motions of the molecule it belongs to, and from the static disorder within the crystal lattice that results when different unit cells trap different microconformers of the molecule as it exists in solution. The averaged displacement of each atom is seen crystallographically as non-spherical electron density at the atomic positions. The description of this uncertainty in atomic position due to displacement is called an Atomic Displacement Parameter (ADP), or more colloquially a “B factor”.

TLS (Translation/Libration/Screw) is a mathematical model that describes the local positional displacement of the atoms in a molecule using an approximation that each atom is part of a ideal rigid body that is displaced normally about a mean position. The TLS formalism was originally developed in the late 1960s to help interpret the ‘thermal ellipsoids’ used in small molecule crystallography at high resolution [1, 2]. In that form it was a method of analysing a completed crystal structural model rather than of refining a model in progress. Its success in identifying rigid groups by examination of the individual atomic displacement parameters in small molecule structures makes it attractive for use also in macromolecular crystallography, where it can be used to find domain and loop flexibility in proteins.

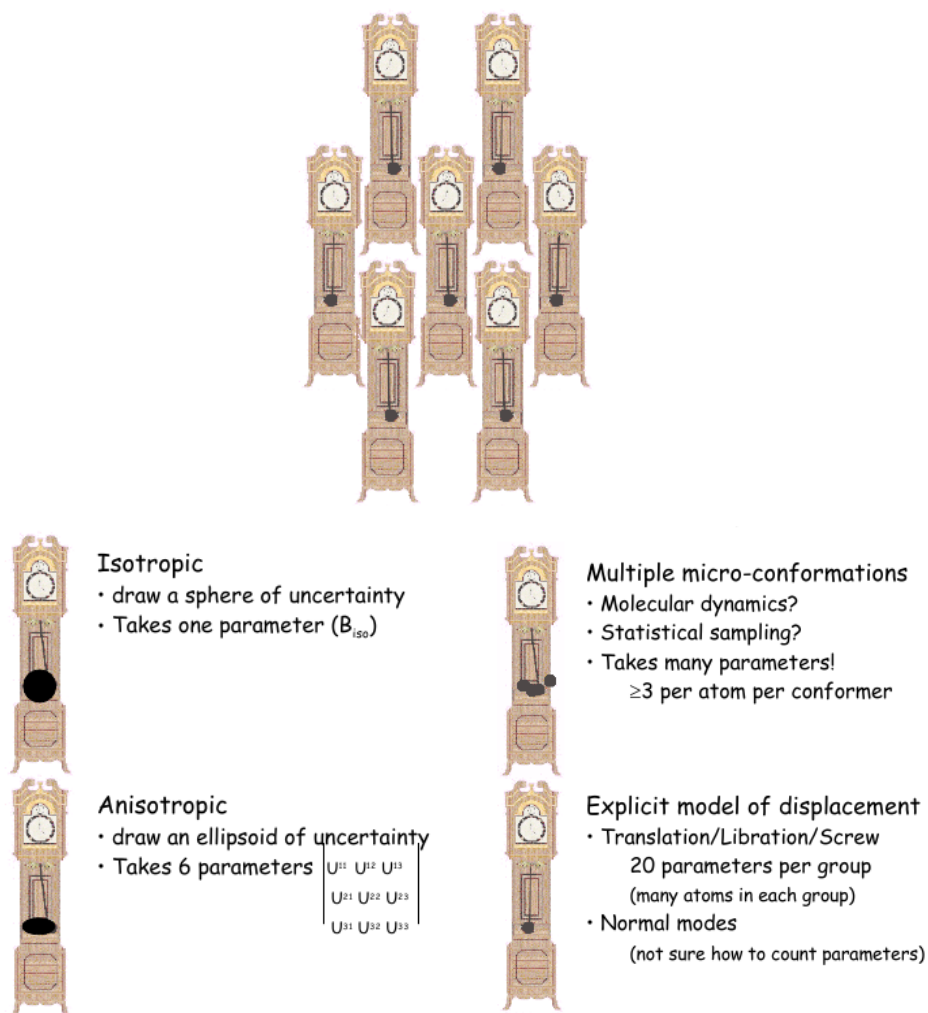
Moreover, the equations that derive a single 20-parameter TLS description from the observed thermal ellipsoids of many atoms can be turned around to great benefit. In this reverse direction, the TLS description is used to predict the individual atomic displacements of many atoms. This turns out to be a great way to describe and refine protein structures. Protein crystals rarely yield the very high-resolution data ($\leq 1.2\text{\AA}$) that would be needed to model anisotropy (thermal ellipsoids) for every atom separately. But even at low resolution ($>3\text{\AA}$) they yield sufficient data to refine TLS models, and these TLS models describe aspects of the crystalline protein that would otherwise not be described at all: dynamic and static disorder due to flexibility¹.

¹A popular corollary to this is that better descriptions of the crystal content allow improved values of F_{calc} , and therefore TLS models generally allow better R factors. This makes crystallographers happy.

A choice of ways to describe the displacement of individual atoms

Consider a crystal of the hypothetical protein Grandfather Clockase. The crystal lattice is established by tight packing of the exterior “Case” domain. The separate “Pendulum” domain is relatively unconstrained by the lattice, and hence is present in a mixture of conformations. Crystallographically we will observe an average of these multiple conformations, which will lead to slight blurring at the hinge-point of this domain but very substantial blurring of the large atom at the bottom of the domain. How best to model this?

Figure 1: Modeling displacement due to an obvious vibrational mode



If Clockase crystals diffract to atomic resolution, we may have the option of modeling the obviously anisotropic displacement of the pendulum tip as an ellipsoid. At lower resolution we don't have that choice. But even at such high resolution it may be worth using an explicit model of the vibrational mode that causes this anisotropy. We will consider only TLS models, although other descriptions are possible.

Simple models vs. complex models - how many TLS groups?

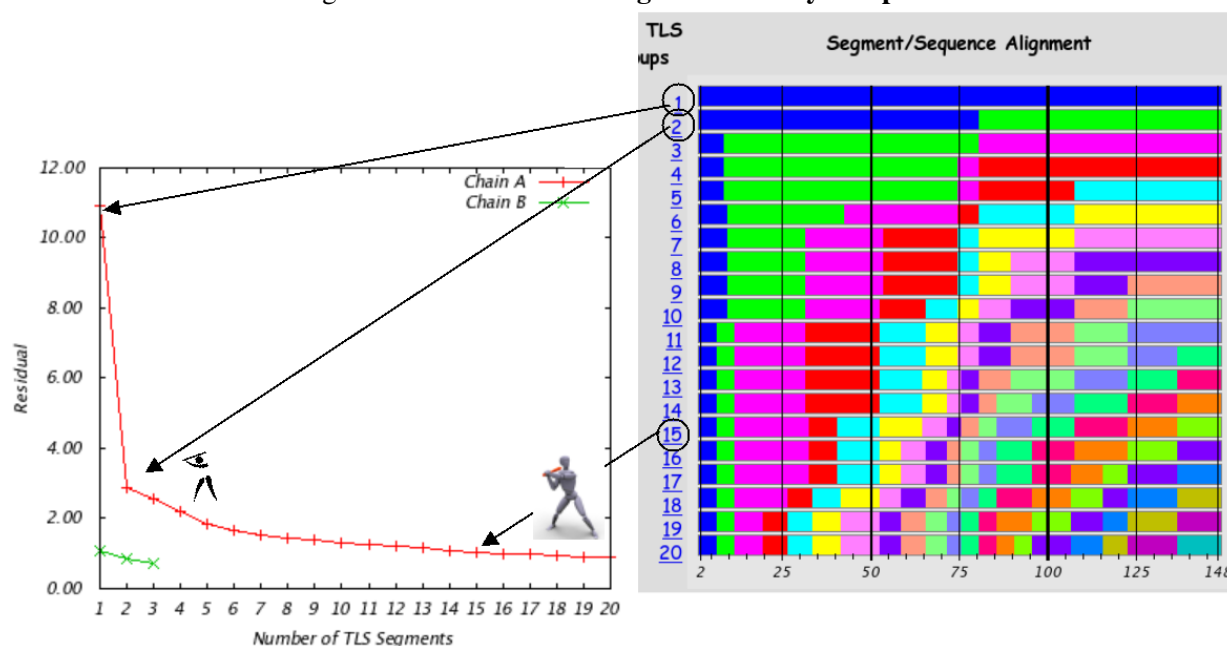
In the Clockase example it was obvious that the structure can be divided into two parts, only one of which is moving/vibrating. It was also obvious which part is moving, and that it really is acting as a rigid body. In the general case we do not know how many flexible parts there are, or how finely we would have to slice up the structure to reach a point where the motion of each individual part is well approximated as a rigid body.

Figure 2: Modeling the swing of a bat



In choosing how many separate pieces should be modeled, it also helps to have a question in mind. To ask “what part of the baseball player may contact the ball?”, we only need a simple model. Perhaps it is sufficient to consider arms, bat, and body. On the other hand, if we want to ask “are the ankles important to the swing?”, a more complex model is needed.

Figure 3: TLSMD: Modeling the flexibility of a protein



Equivalent considerations apply to choosing how many separate TLS groups you will use to model the

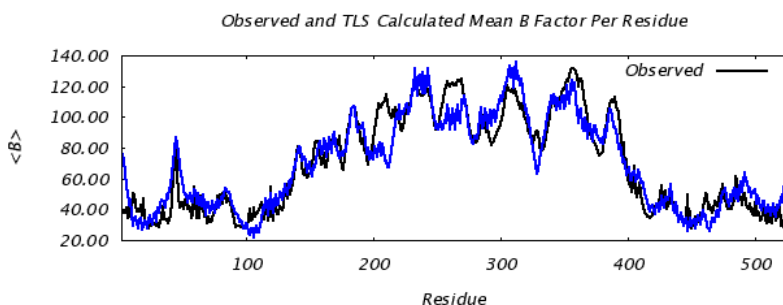
behaviour of a protein. One group only - which essentially says that the entire protein acts approximately as a single rigid body? Two groups? 20 groups? One group may suffice to describe the overall vibration of the protein in the crystal lattice allowed by loose packing. But if you want to look for inter-domain hinge motions, or flexibility in the helices or loops surrounding an active site, then you will need to consider breaking the protein into smaller pieces that can move relative to each other.

Overview of TLS refinement guided by TLSMD

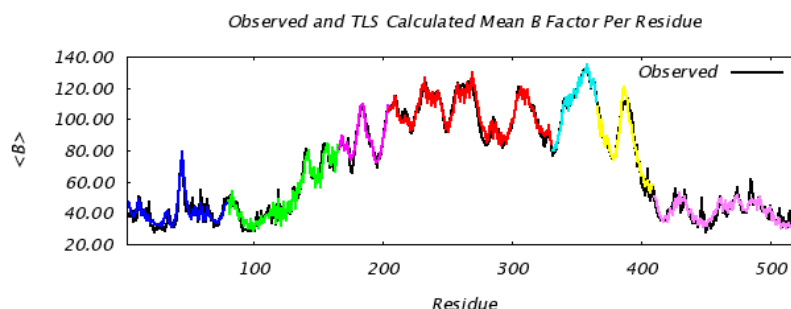
1. [*incredibly important!*] start with a refined set of B values
 - usual case: conventional B_{iso} refinement
 - high resolution: anisotropic U^{ij} refinement
 - low resolution: individual B_{iso} or group B per-residue
 - DON'T use the original B s from a molecular replacement solution
 - DON'T use fixed B . Models with constant B are useless; newly-built residues often have $B=15$.
2. [*the easy part*] submit your starting model to the TLSMD server
<http://skuld.bmsc.washington.edu/~tlsmd>
3. [*the server does the hardest part*] generate optimal multi-segment TLS models that best explain the observed B values
4. Choose the TLS model you prefer; i.e., how many segments per chain?
 - The eventual R value is expected to track the change in residual reported by the TLSMD server. If there isn't much drop in residual as you add TLS groups, probably TLS is not going to help your R factors much. My rule of thumb is that if the residual doesn't eventually drop by a factor of 2, a multi-group TLS model is only worth pursuing if you care about the hinge points or domain boundaries for their own sake.
 - Look for a dog-leg in the plot. For example, Figure 3 shows a huge dog-leg at 2 segments and a less dramatic one at 5 segments.
 - Look down the list of partitions for stable locations of implied hinge-points.
 - The fit of observed to predicted B should be equally good for all segments (Figure 4). But if it isn't, this may indicate a problem in the original refinement rather than a problem with TLS analysis per se.
 - Does the model make sense? The animations may help you decide.
5. [*Optional*] simplify the model further (Figure 5).
6. [*refmac or phenix does this part*] refine the TLS model against the crystallographic data

- You must choose whether to reset the starting B_{iso} values to a constant. If in doubt, try it both ways and see which works better (Figure 6).
7. [... *and this part*] refine the TLS model jointly with coordinates and maybe residual contribution B_{iso}
 - High to medium resolution: refine residual B_{iso} for each atom (Figure 6a).
 - Medium to low resolution: refine only an overall B , *not* individual B_{iso} for each atom. This gives you a pure TLS model (Figure 6b).
 8. repeat steps 2-7. Remember, the quality of the multi-group TLS model suggested by TLSMD is limited by the quality of the B values in the PDB file you give it for analysis. As your model improves, the TLS description may improve also (Figure).
 9. Validate your model. Check not only that the stereochemistry makes sense (bond lengths, angles, rotamers, etc), but also that the B factors and/or TLS groups make sense.
 10. Figure out how to tell the PDB what you have done. Unfortunately, there is still a great deal of confusion about what exactly is in a PDB file that has been refined with TLS. Many programs ignore it. The PDB now requests that before depositing your model, you expand the TLS model you refined to generate individual ANISOU records for each atom. I think this is a bad idea for several reasons, but nevertheless that is the current standard. For ccp4 refinements you can do this either by selecting the `refmac` option “`TLSO ADDU`” or by post-processing using the programs `TLSEXTRACT` and `TLSANL`.

Figure 4: A good choice of TLS segmentation will give roughly equal residuals for each segment. The curves below show the average per-residue agreement of a 1-group TLS model (top) and a 7-group TLS model (bottom) for the same protein structure. It is clear that some portions of the chain are not well described by the single TLS group model. However, after partitioning the chain into 7 separate TLS groups, the fit is of roughly equal quality everywhere.



Input Structure						TLS Predictions				
Color	Segment	Residues	Atoms		<Aniso>	RMSD B	$T^r B$	eval(L) DEG^2		<Aniso>
■	2-526	525	3855	71.1	1.00	11.90	-4.5	4.40, 1.01, 0.00	71.1	0.61



Input Structure						TLS Predictions				
Color	Segment	Residues	Atoms		<Aniso>	RMSD B	$T^r B$	eval(L) DEG^2		<Aniso>
■	2-78	77	569	41.0	1.00	5.73	32.9	10.58, 0.05, 0.00	41.0	0.80
■	79-164	86	597	51.9	1.00	5.25	28.7	12.46, 0.00, 4.03	51.9	0.66
■	165-203	39	297	87.9	1.00	4.35	72.4	8.18, 3.41, 0.00	87.9	0.86
■	204-329	126	942	105.0	1.00	5.05	52.3	5.63, 1.34, 0.00	105.0	0.87
■	330-364	35	272	113.6	1.00	3.18	110.2	4.75, 0.00, 0.00	113.6	0.86
■	365-408	44	324	84.7	1.00	6.07	70.3	9.42, 0.01, 0.00	84.7	0.77
■	409-526	118	854	42.6	1.00	5.71	30.0	5.36, 2.26, 0.00	42.6	0.82

Figure 5: Optional Step: Merge discontinuous chain segments into a single TLS group. Domains are often formed from segments of the polypeptide chain that are not adjacent in the sequence. A simple example of this is shown below in which a 5-strand beta sheet is formed by the N- and C- termini. In a case like this we expect that the whole domain may act as a single approximately rigid body, and be well described as a single TLS group. The automated analysis is not currently smart enough to recognize this by itself, although it does print out tables of how well the individual TLS descriptions for each segment agree with each other. The diagonal terms in the table are the residuals showing how well the TLS model for each segment predicts the B values reported for that segment. The off-diagonal terms are the residuals you would get after combining two groups into a single group. If a pair of off-diagonal terms are approximately the same magnitude as the on-diagonal terms for those same segments, this is a hint that they may actually act as a single group.

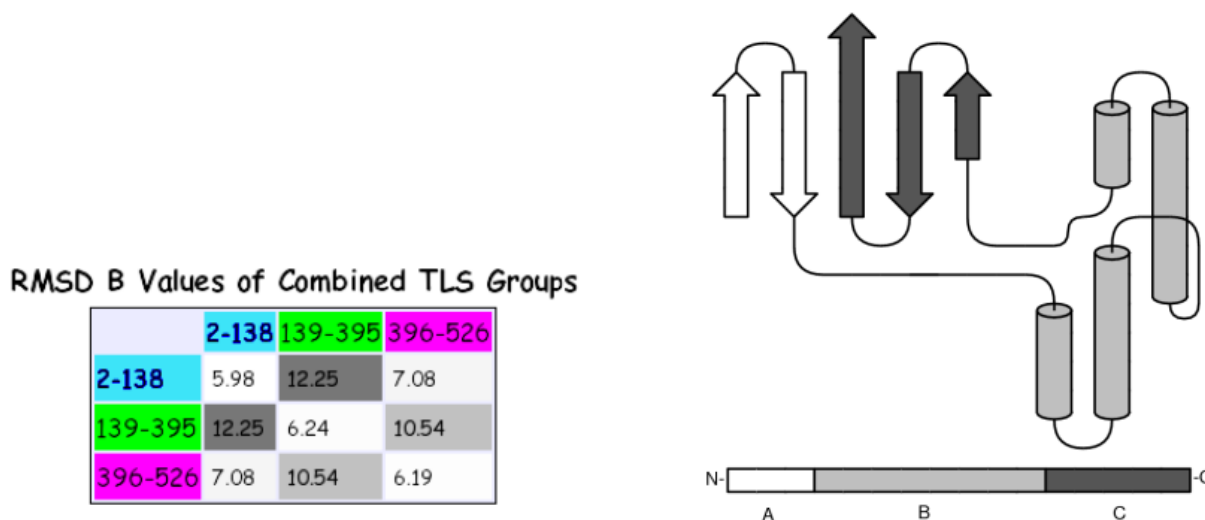
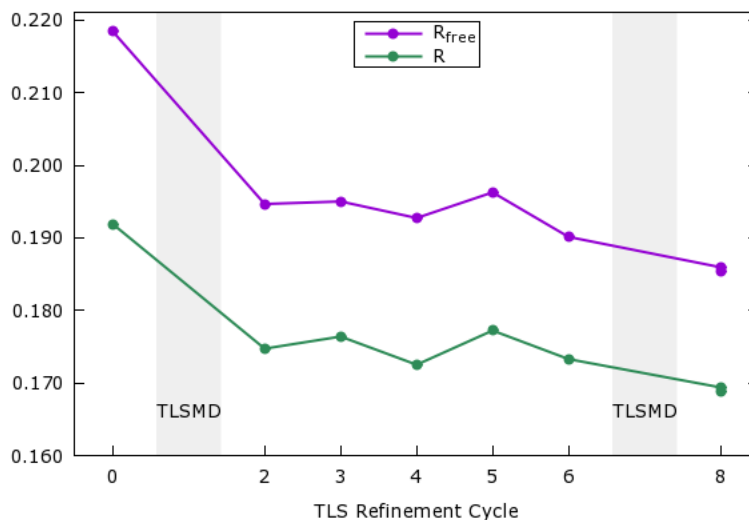


Figure 6: Telling reffmac whether to refine a residual B_{iso} for each atom (a), or a pure TLS model (b)

The figure shows two side-by-side screenshots of the reffmac software interface. Both windows show the 'Data Harvesting' and 'Refinement Parameters' sections. In the left window (a), the 'Set initial B factors to 20.0' checkbox is checked, and the 'Refine' dropdown is set to 'isotropic'. In the right window (b), the 'Set initial B factors to 20.0' checkbox is checked, and the 'Refine' dropdown is set to 'overall'. Red circles highlight these specific settings in both windows.

Figure 7: **Revisiting the TLSMD model after refinement.** You should probably re-evaluate your TLS model as your refinement nears completion. TLSMD constructs a model that optimally predicts the distribution of B values in the 3-dimensional space occupied by your structure. After you have refined the structure you hopefully have better B values and 3D coordinates, so TLSMD analysis may be able to construct a better TLS description.



Validation

How can we test if a TLS model is reasonable? For that matter, how can we test if the B factors are reasonable even aside from whether they came from a TLS model? The same way we can test that the bond angles or bond lengths are reasonable - we compare the distribution of B factors in our new model to the distribution that has previously been observed in very well-refined (usually very high resolution) structures. Of course, “reasonable” doesn’t necessarily mean “correct”. But “unreasonable” usually *does* mean “incorrect”. As an extreme case, a value of $B_{iso} < 0$ is not physically possible for a real atom. If our new model has negative B factors, something has gone wrong. Refinement programs protect you against this extreme case when you are refining individual B factors for each atom, but this protection can fail if your B factors are derived indirectly through a TLS model. Here are some specific validation checks that can be made for B factors and TLS models:

1. Are any atoms non-positive definite, or nearly so?
2. Is the overall distribution of anisotropy within the range seen for well-refined structures?
3. If the model is split into multiple TLS groups, do they make sense?

A number of validation tests are available via the Parvati server:

<http://skuld.bmsc.washington.edu/parvati>

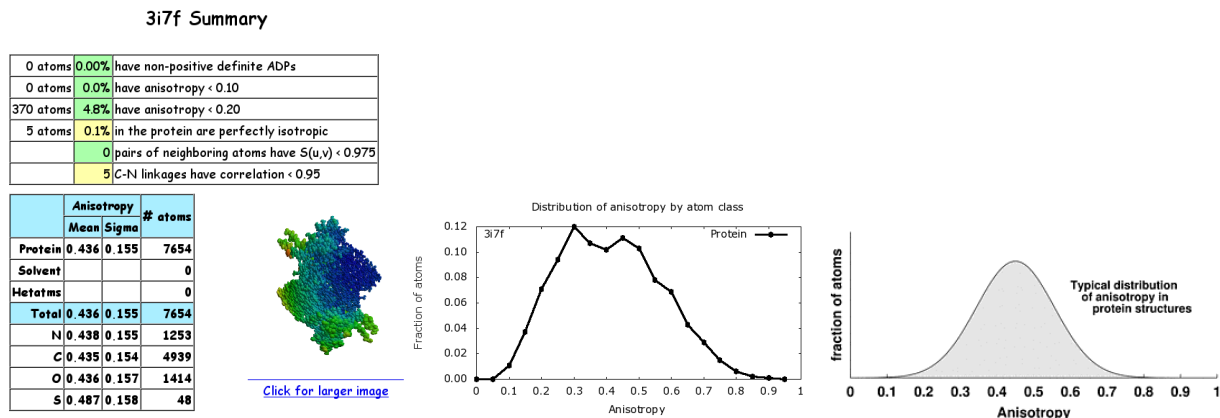


Figure 8: Output from the Parvati validation server

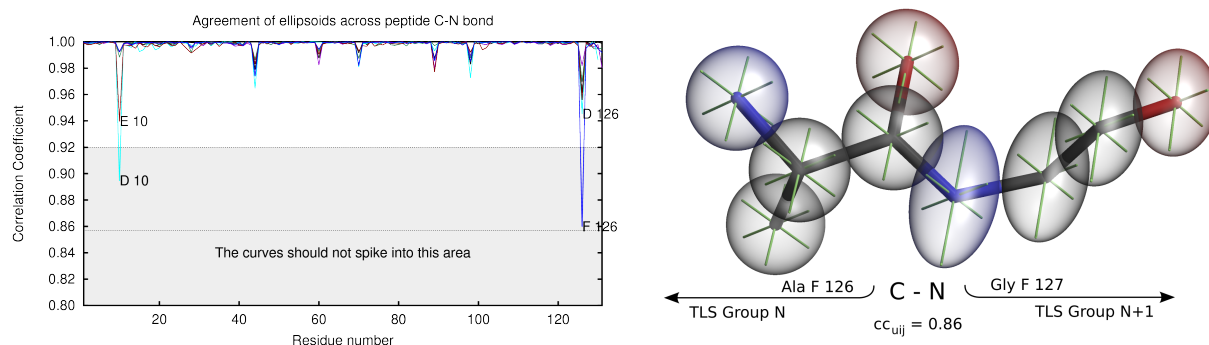
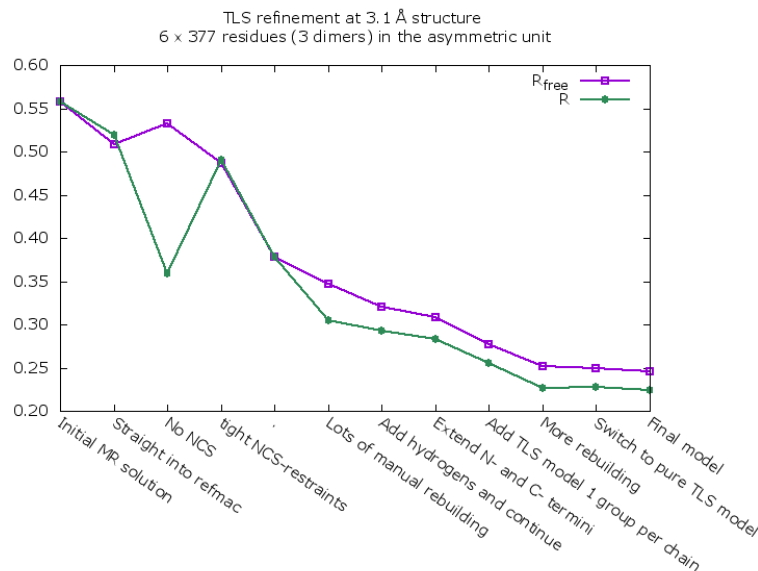


Figure 9: Using the CC_{uij} residual to detect bad junctions between adjacent TLS groups. The atoms of the two residues on either side of a junction between TLS groups are depicted here as thermal ellipsoids drawn at the 33% probability level. The TLS model for the group containing residue 126 describes a relatively isotropic displacement for atoms in this region of space. The TLS model for the group containing residue 127 describes a more anisotropic displacement for atoms in this same region. This discrepancy results in incompatible models for the vibrational motion of the two bonded atoms C-N that bridge the two TLS groups. One measure of this discrepancy is the quantity cc_{uij} . A small value of cc_{uij} may indicate a poorly chosen boundary between the two groups. Alternatively it may indicate that the description refined for one or both of the TLS groups is dominated by inclusion of other residues whose true displacements are different from those of atoms in either of the residues shown here, and thus would better be split off into a TLS group of their own. Both scenarios suggest that the assignment of TLS group boundaries within the protein chain should be reconsidered. The figure is from Zucker *et al* (2010).

Case study: TLS Refinement at 3 Å resolution

Figure 10: Refinement at 3 Å resolution, from initial MR to final model



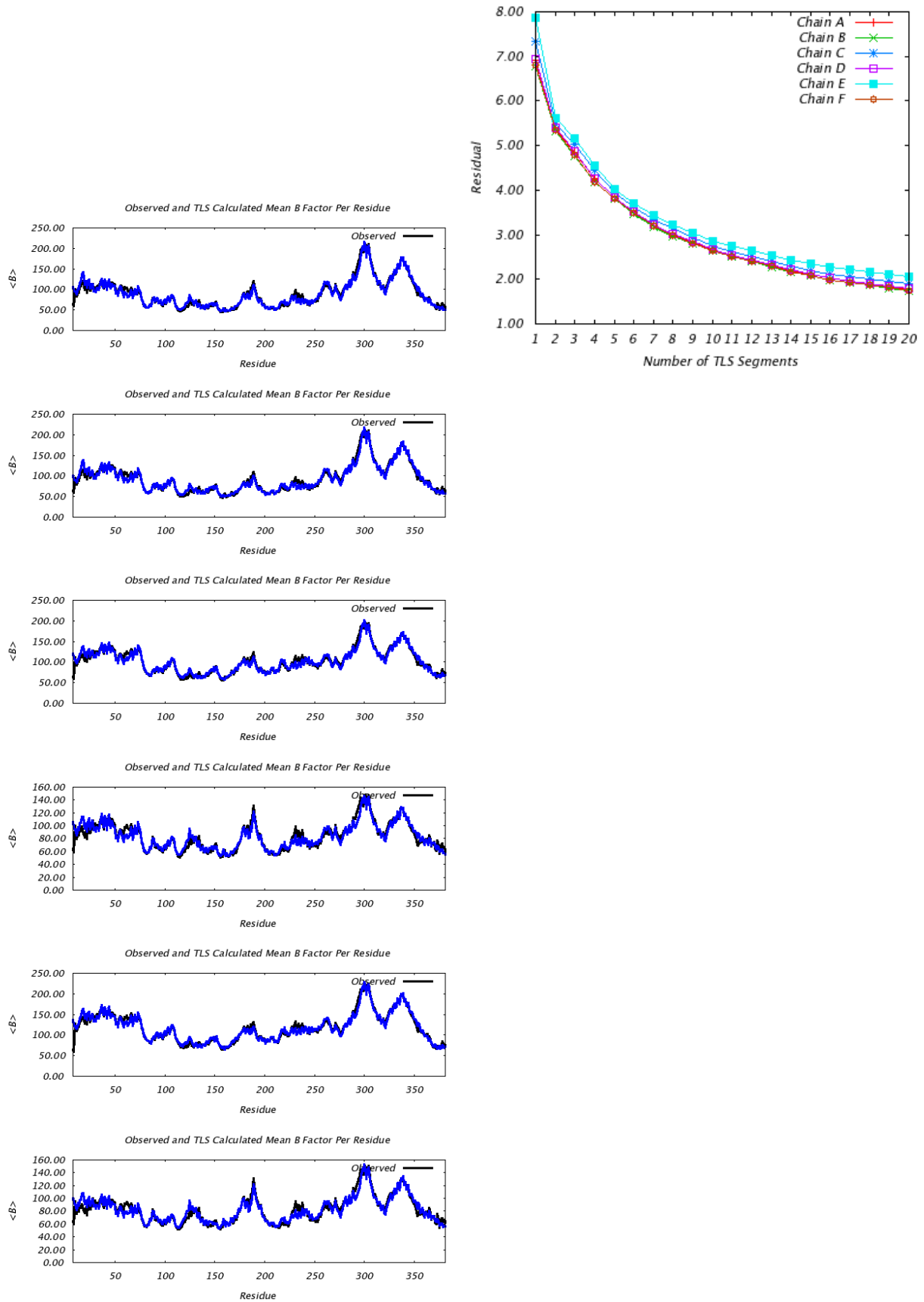
Here is the course of refinement of a real protein at 3 Å resolution. Phaser was able to place 3 copies of a homologous dimer in the asymmetric unit. Conventional refinement of the initial model (individual B_{iso} , no special restraints) was not well-behaved. Adding tight NCS restraints helped a lot. The usual iteration of manual rebuilding with refinement of an NCS-restrained model using individual (NCS-restrained) B_{iso} stalled out at $R = 0.28/R_{free} = 0.31$. At this point I submitted the model to TLSMD. The analysis is shown in Figure 11. Several things are apparent. Partitioning the chain into many segments yields an overall drop in the residual by about a factor of 4. That is pretty good.

However, it is interesting that even a single group TLS model fits the refined B_{iso} very well (Figure 11). This suggested that it was worth refining a very simple, one group per chain, TLS model first. Indeed, adding this simple TLS model to the refinement lowered the crystallographic residuals to $R = 0.23/R_{free} = 0.25$. That is pretty dramatic, and hints that the primary source of displacement in the crystal is the rocking of entire molecules within the lattice (remember that this TLS in this case is treated an entire monomer as a rigid body). This hint is borne out when the refinement is switched over to using a pure TLS model, still treating each monomer being as a rigid body, with no individual B_{iso} terms. This even simpler model yields slightly better R and R_{free} than the model which included individual B terms.

For whatever reason, in this case partitioning each chain into many segments does not significantly drop R or R_{free} beyond this point. Refining a TLS model with 12 groups per chain yielded $R = 0.225/R_{free} = 0.252$. This seems contrary to what I would normally predict based on the curve in Figure 11, but it is consistent with the observation that a 1 group per chain model already fits very well everywhere (Figure 11). Compare this, for instance, to the more typical single group curve in Figure 4. So for this particular crystalline protein, TLS refinement was extremely effective even without splitting the chains into multiple TLS groups.

Figure 11: TLSMD analysis of the conventionally refined model in case study #1.

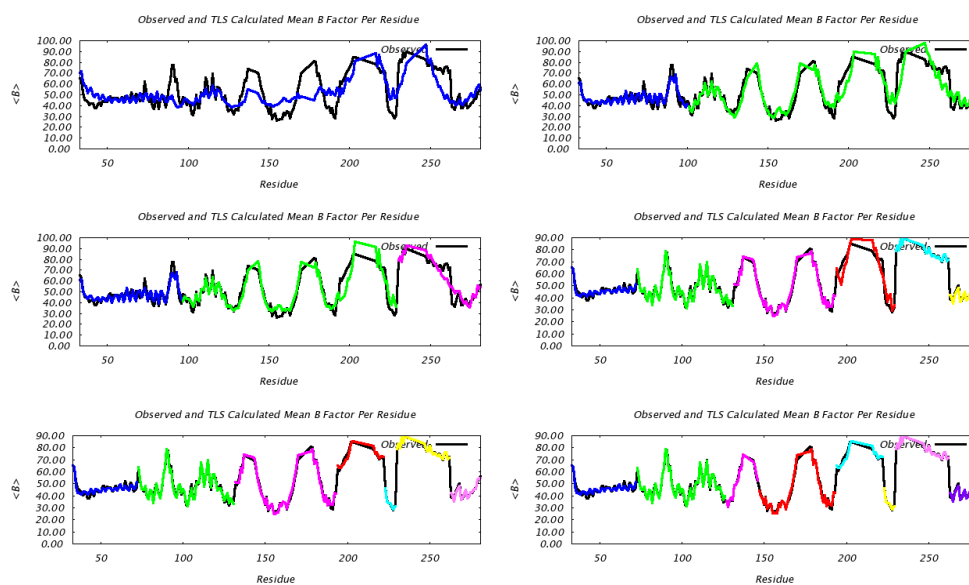
(a) 1-group TLS fits for each chain A B C D E F (b) Overall residual as a function of number of TLS groups



Case study #2: TLS Refinement at 2.8 Å resolution

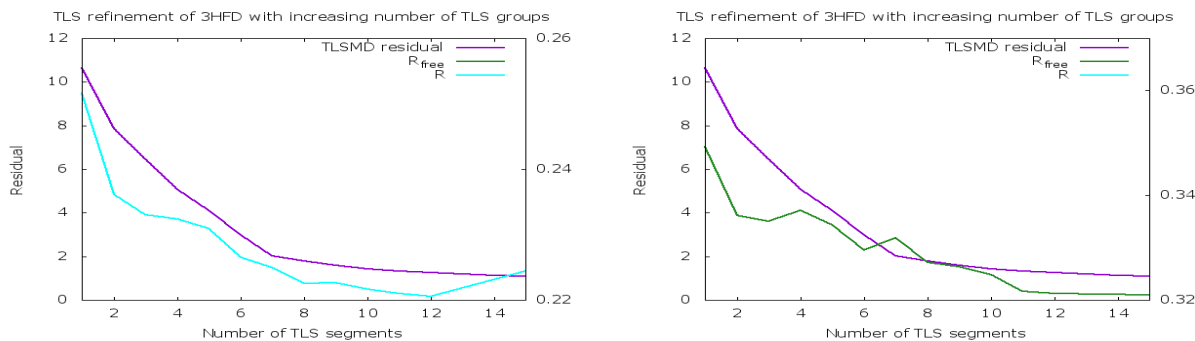
Here is another example of refinement at moderately low resolution, one that works out very differently. In this case the protein is a dimer, but the asymmetric unit contains a single monomer of 270 residues. Again the initial structure solution came from molecular replacement. Manual rebuilding and conventional B_{iso} refinement brought yielded $R = 0.249/R_{free} = 0.341$.² TLSMD analysis of the model indicated a distinct dog-leg in the residual plot at 7 groups. You can see essentially the same thing by noticing that the per-residue fit of predicted to observed B factors is very uneven for partitions with fewer segments (Figure 12).

Figure 12: TLSMD analysis after conventional refinement of case study #2.



In this case these features of the TLSMD analysis do indeed correlate with the improvement in R and R_{free} obtained from refinement of TLS models with an increasing number of TLS groups (Figure 13).

Figure 13: Pure TLS refinement (no B_{iso} terms) after TLSMD analysis of case study #2.



² R and R_{free} are rather far apart, which probably indicates model bias. Unfortunately in this case we had no NCS or experimental phases to help overcome this bias.

References

Theory of TLS

- [1] V. Schomaker, K. N. Trueblood, On the rigid-body motion of molecules in crystals, *Acta Cryst.* B24 (1968) 63–76.
- [2] J. Dunitz, V. Schomaker, K. N. Trueblood, Interpretation of anisotropic displacement parameters from diffraction studies of crystals, *J. Phys. Chem.* 92 (1988) 856–867.
- [3] R. E. Rosenfield, K. N. Trueblood, J. D. Dunitz, Test for rigid-body vibrations, based on a generalization of hirshfeld rigid-bond postulate, *Acta Cryst.* A34 (1978) 828–829.
- [4] B. T. M. Willis, A. W. Pryor, *Thermal Vibrations in Crystallography*, Cambridge University Press, 1975.

TLSMD server

- [5] J. Painter, E. A. Merritt, Optimal description of a protein structure in terms of multiple groups undergoing TLS motion, *Acta Cryst.* D62 (2006) 439–450.
- [6] J. Painter, E. A. Merritt, TLSMD web server for the generation of multi-group TLS models, *J. Appl. Cryst.* 39 (2006) 109–111.

Validation

- [7] F. Zucker, P. C. Champ, E. A. Merritt, Validation of crystallographic models containing TLS or other descriptions of anisotropy, *Acta Cryst.* D66 (Part 8) (2010) 889–900.

Acknowledgments

Source for complex model baseball swing images: <http://blogs.tamu.edu/ariel> [Ariel Chisholm, Texas A&M University Visualization Sciences Coursework]