



COMPUTATIONAL CRYSTALLOGRAPHY INITIATIVE

Automated TLS group determination in Phenix

Pavel Afonine

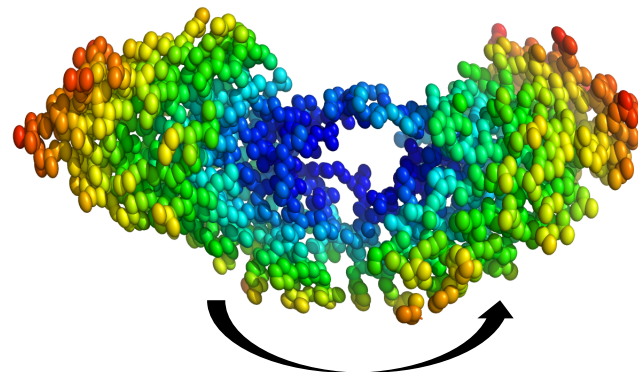
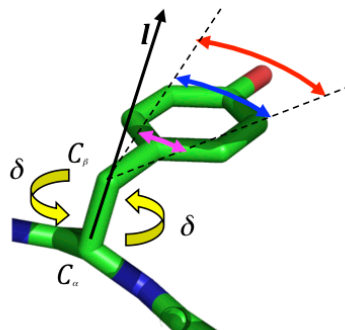
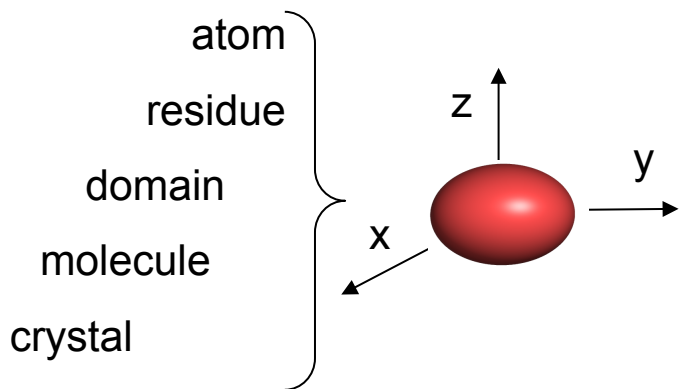
Computation Crystallography Initiative
Physical Biosciences Division
Lawrence Berkeley National Laboratory, Berkeley CA, USA

October, 2010

PHYSICAL BIOSCIENCES DIVISION

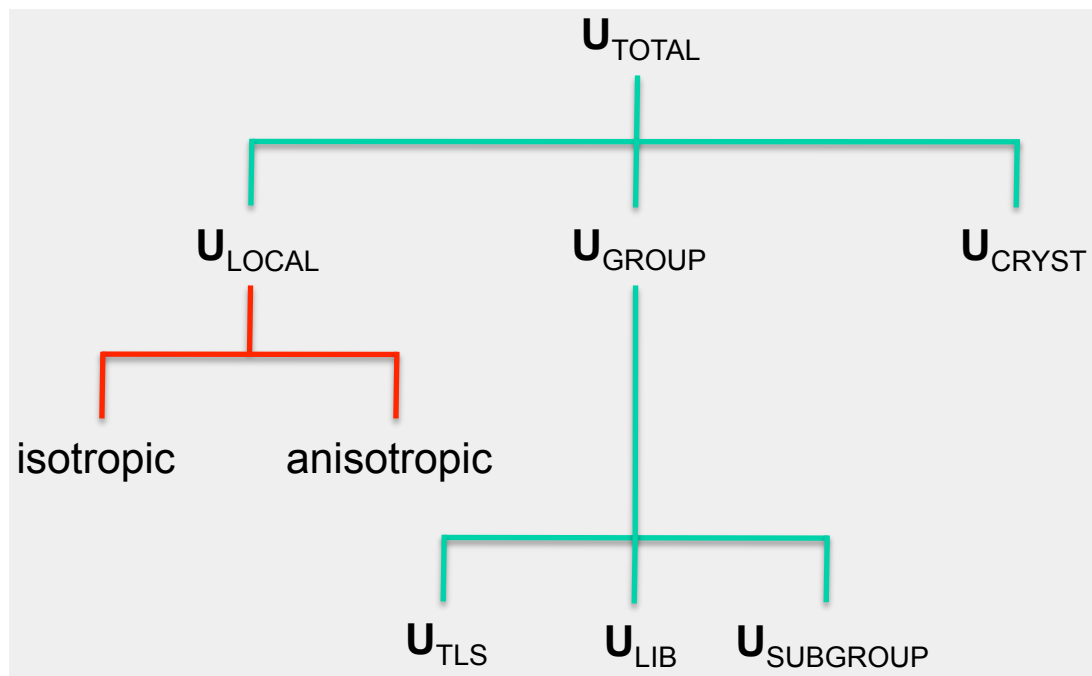
Atomic Displacement Parameters (ADP or “B-factors”)

- ADPs model relatively small atomic motions (within harmonic approximation)
- Hierarchy and anisotropy of atomic displacements



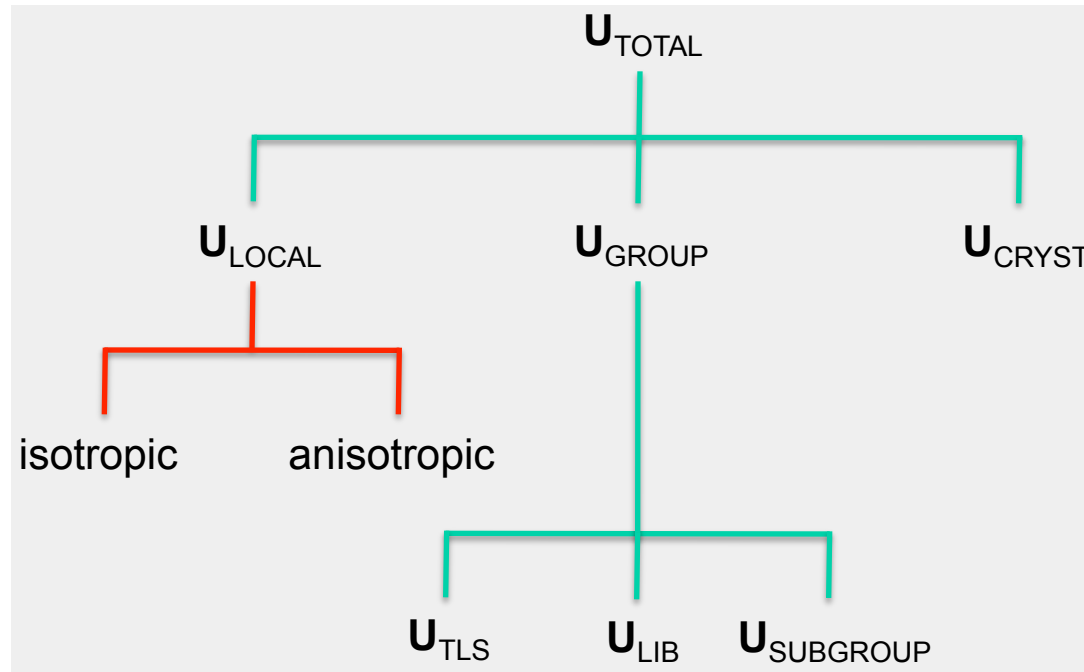
Total ADP:

$$\mathbf{U}_{\text{TOTAL}} = \mathbf{U}_{\text{CRYST}} + \mathbf{U}_{\text{GROUP}} + \mathbf{U}_{\text{LOCAL}}$$



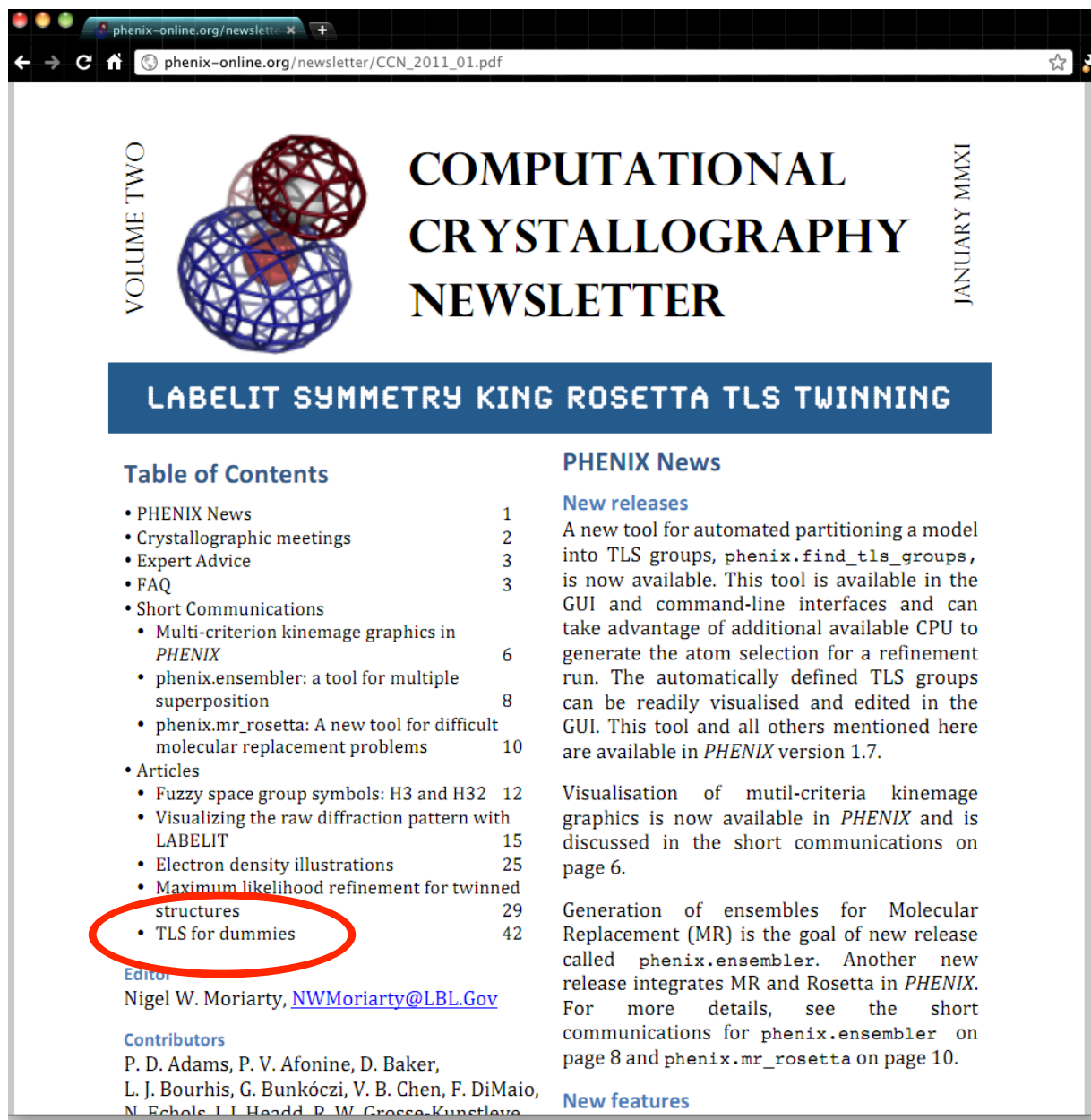
Atomic Displacement Parameters (ADP or “B-factors”)

▪ Total ADP $\mathbf{U}_{\text{TOTAL}} = \mathbf{U}_{\text{CRYST}} + \mathbf{U}_{\text{GROUP}} + \mathbf{U}_{\text{LOCAL}}$



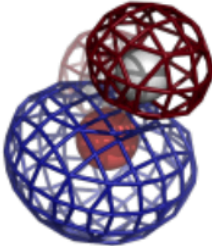
- $\mathbf{U}_{\text{CRYST}}$ – overall anisotropic scale.
- \mathbf{U}_{TLS} – rigid body displacements of molecules, domains, secondary structure elements.
- $\mathbf{U}_{\text{LOCAL}}$ – local vibration of individual atoms.
- \mathbf{U}_{LIB} – librational motion of side chain around bond vector.

TLS and ADP: comprehensive overview (~50 pages)



phenix-online.org/newslett...
phenix-online.org/newsletter/CCN_2011_01.pdf

VOLUME TWO



COMPUTATIONAL
CRYSTALLOGRAPHY
NEWSLETTER

JANUARY MMXI

LABELIT SYMMETRY KING ROSETTA TLS TWINNING

Table of Contents

• PHENIX News	1
• Crystallographic meetings	2
• Expert Advice	3
• FAQ	3
• Short Communications	
• Multi-criterion kinemage graphics in <i>PHENIX</i>	6
• phenix.enssembler: a tool for multiple superposition	8
• phenix.mr_rosetta: A new tool for difficult molecular replacement problems	10
• Articles	
• Fuzzy space group symbols: H3 and H32	12
• Visualizing the raw diffraction pattern with LABELIT	15
• Electron density illustrations	25
• Maximum likelihood refinement for twinned structures	29
• TLS for dummies	42

Editor
Nigel W. Moriarty, NWMoriarty@LBL.Gov

Contributors
P. D. Adams, P. V. Afonine, D. Baker,
L. J. Bourhis, G. Bunkóczi, V. B. Chen, F. DiMaio,
N. Echols, J. L. Headd, P. W. Grosse-Kunstleve

PHENIX News

New releases

A new tool for automated partitioning a model into TLS groups, `phenix.find_tls_groups`, is now available. This tool is available in the GUI and command-line interfaces and can take advantage of additional available CPU to generate the atom selection for a refinement run. The automatically defined TLS groups can be readily visualised and edited in the GUI. This tool and all others mentioned here are available in *PHENIX* version 1.7.

Visualisation of mutil-criteria kinemage graphics is now available in *PHENIX* and is discussed in the short communications on page 6.

Generation of ensembles for Molecular Replacement (MR) is the goal of new release called `phenix.enssembler`. Another new release integrates MR and Rosetta in *PHENIX*. For more details, see the short communications for `phenix.enssembler` on page 8 and `phenix.mr_rosetta` on page 10.

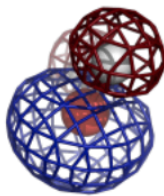
New features

Parameterization and refinement of ADP in Phenix

Afonine *et al* (2010). On atomic displacement parameters and their parameterization in Phenix

phenix-online.org/newslett...
phenix-online.org/newsletter/CCN_2010_07.pdf

VOLUME ONE



COMPUTATIONAL
CRYSTALLOGRAPHY
NEWSLETTER

JULY MMX

IOTBX.PDB SECONDARY-STRUCTURE SPOTFINDER

Table of Contents

- PHENIX News 1
- Crystallographic meetings 2
- Expert Advice 3
- FAQ 3
- Articles
 - cctbx PDB handling tools 4
 - Secondary structure restraints in `phenix.refine` 12
 - cctbx Spotfinder: a faster software pipeline for crystal positioning 18
 - On atomic displacement parameters (ADP) and their parameterization in PHENIX 24
- Short communications
 - Non-periodic torsion angle targets in PHENIX 32
 - Model building updates & new features 34

Editor
Nigel W. Moriarty, NWMoriarty@LBL.Gov

Contributors
P. D. Adams, P. V. Afonine, V. B. Chen, N. Echols, J. J. Headd, R. W. Grosse-Kunstleve, N. W. Moriarty, D. C. Richardson, J. S. Richardson, N. K. Sauter, T. C. Terwilliger, A. Urzhumtsev

PHENIX News

New releases

A new feature in the PHENIX GUI is designed to compare the refined structures of similar proteins. Structure comparison uses the protein sequences

to display the differences between each protein chain loaded in both a table format and graphically in COOT. Several features of the protein structure can be compared including rotamers and secondary structure. NCS chains can be overlaid in COOT, edited and saved in the original orientation.

The Java kinemage viewer KING (Protein Science 2009, 18:2403-2409) has been incorporated into PHENIX, so that the KING jar files and supporting scripts will be part of the distribution package, while relying on the Java virtual machine that is standard on the user's Mac or Linux platform. Without any other setup, the KING program in PHENIX would allow viewing of macromolecular structures directly from PDB files, and viewing of kinemages (such as multi-criterion kinemages from MolProbity). We are currently working to incorporate validation kinemage creation directly within the PHENIX GUI, allowing even more seamless and complete user evaluation of model quality during the refinement process. Later developments might also provide kinemage displays to provide help in evaluating other aspects of refinement and model building progress.

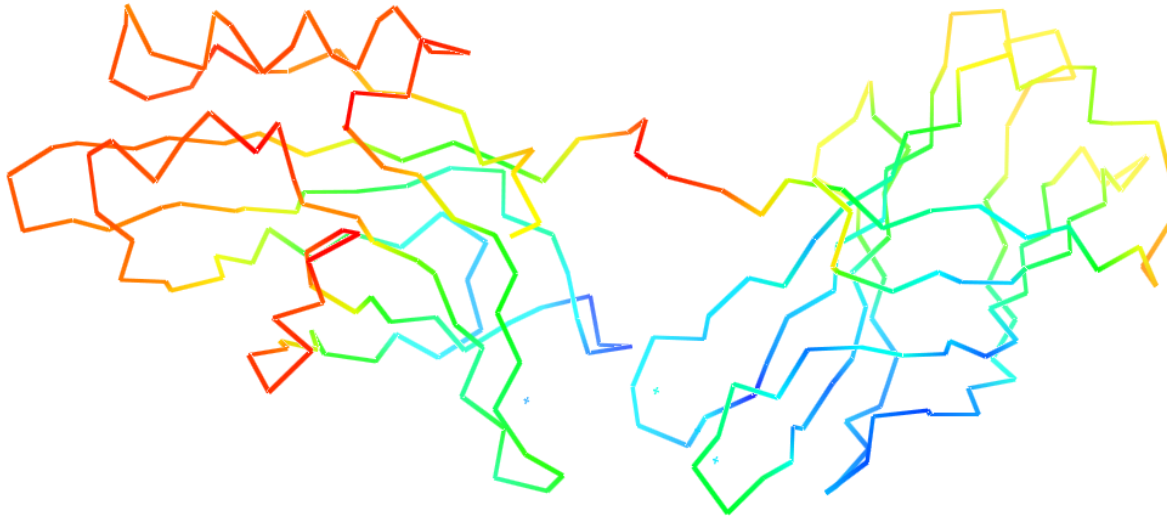
In a similar vein to `phenix.superpose_pdbs`, a new program has been added to PHENIX that is specifically designed for superposing ligands. Superposing protein models is generally done using the C_{α} positions. Ligands require different algorithms to create atomic correspondences, some of which has been implemented in `eLBOW`

The Computational Crystallography Newsletter (CCN) is a regularly distributed electronically via email and the PHENIX website, www.phenix-online.org/newsletter. Feature articles, meeting announcements and reports, information on research or other items of interest to computational crystallographers or crystallographic software users should be submitted to the editor at any time for consideration. Submission of text by email or word-processing files using the CCN templates is requested.

Computational Crystallography Newsletter (2010). Volume 1, Part 1. 1

TLS

- Using TLS in refinement requires partitioning a model into TLS groups. This is typically done by
 - visual model inspection and deciding which domains may be considered as rigid
 - using TLSMD method
 - Painter & Merritt. (2006). Acta Cryst. D62, 439-450
 - Painter & Merritt. (2006). J. Appl. Cryst. 39, 109-111

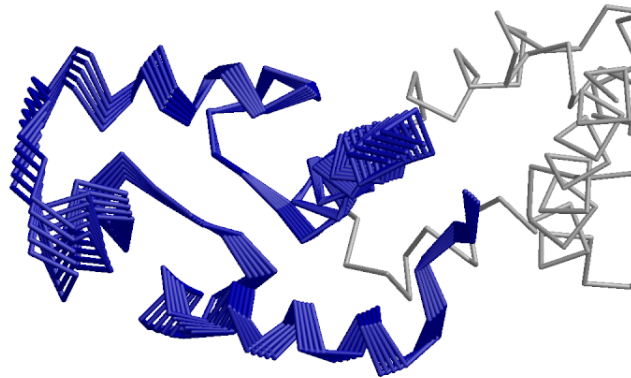


Methodology

TLS contribution of an individual atom participating in a TLS group can be computed from T, L and S matrices:

$$\mathbf{U}_{\text{TLS}} = \mathbf{T} + \mathbf{A}\mathbf{L}\mathbf{A}^t + \mathbf{A}\mathbf{S} + \mathbf{S}^t\mathbf{A}^t$$

(20 TLS parameters per TLS group)

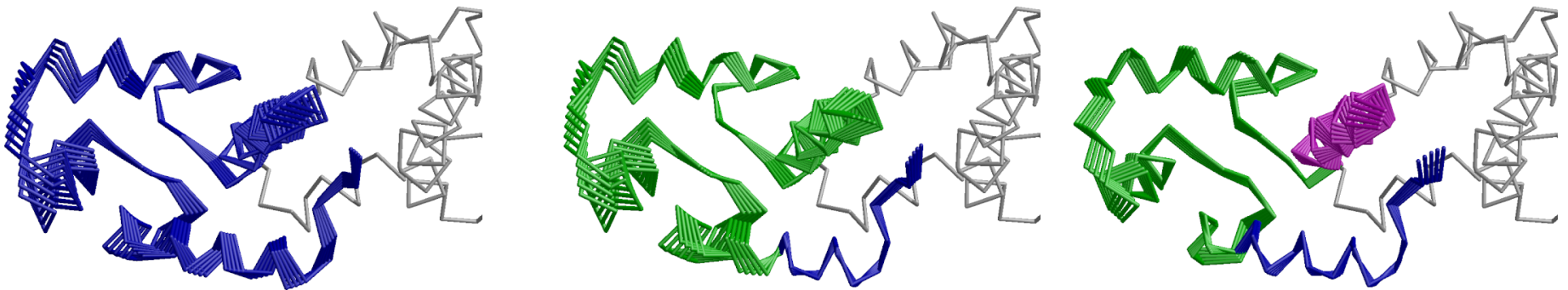


Methodology (TLSMD)

- Split a model into 1, 2, 3, ..., N contiguous segments.
 - If n is the number of residues in the chain, m is minimum number of residues in a segment, then the number of segments is

$$s(n, m) = n(n + 1)/2 - \sum_{i=2}^m (n + 1 - i)$$

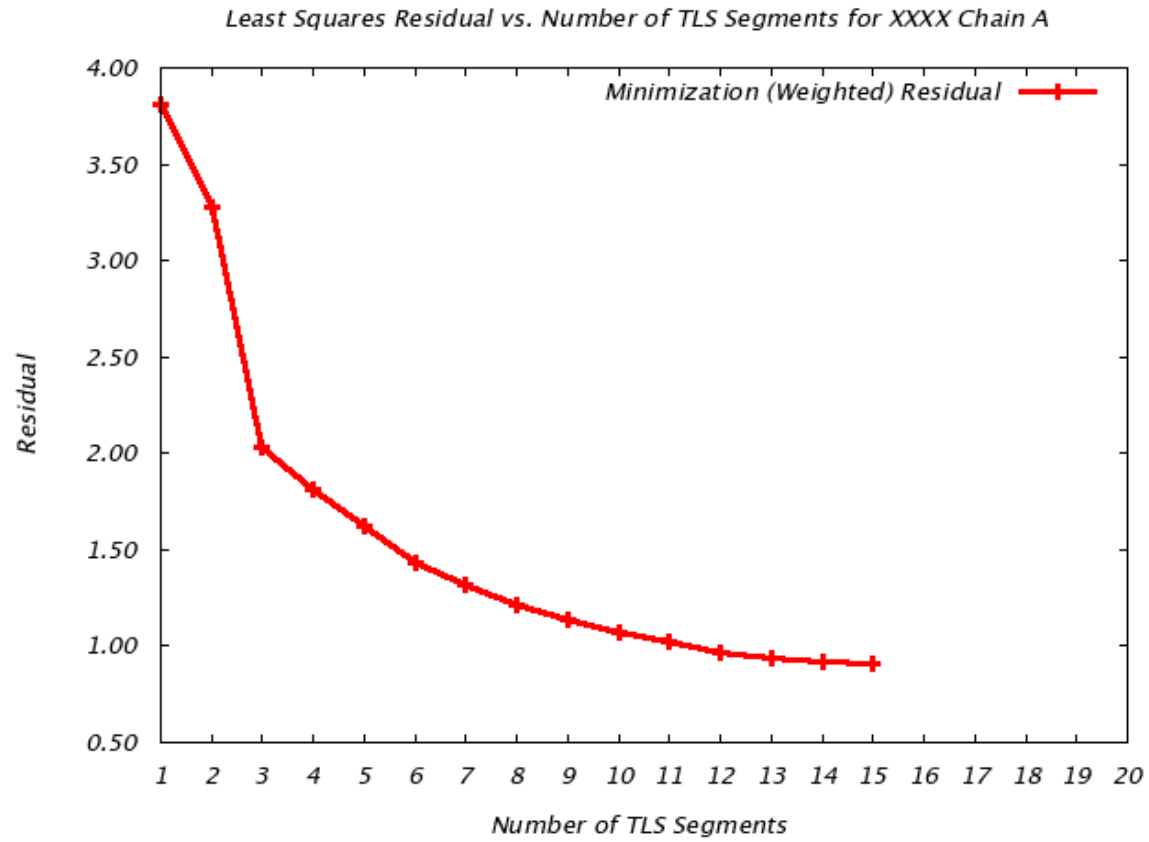
- For example if $n=100$ and $m=3$ (minimal number of residues so there is more “observations” than parameters), then we get 4853 possible partitions.



- Compute residual for all segment partitions:

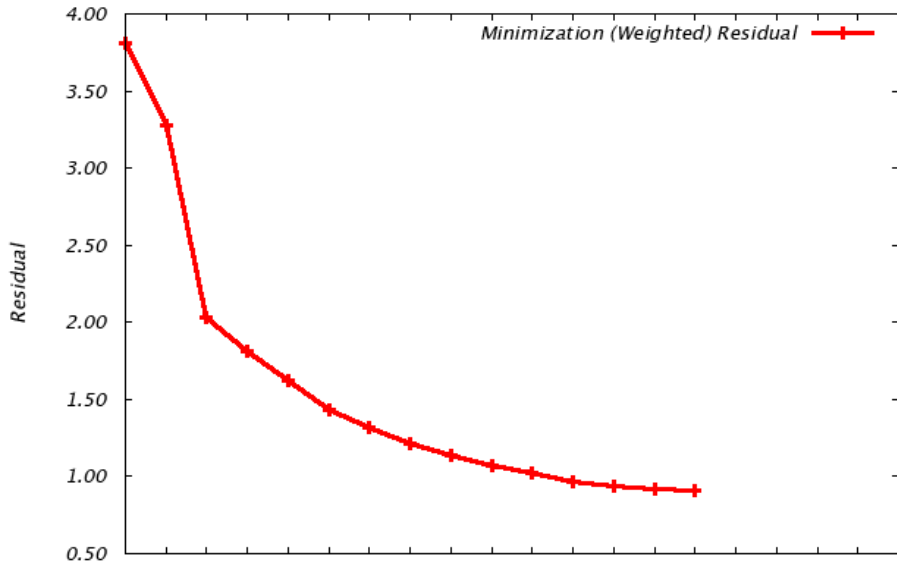
$$R = w \sum_{\text{atoms in segment}} W \left(\sum_{i=1}^6 (U_{ATOM}^i - U_{TLS}^i)^2 \right)$$

Methodology

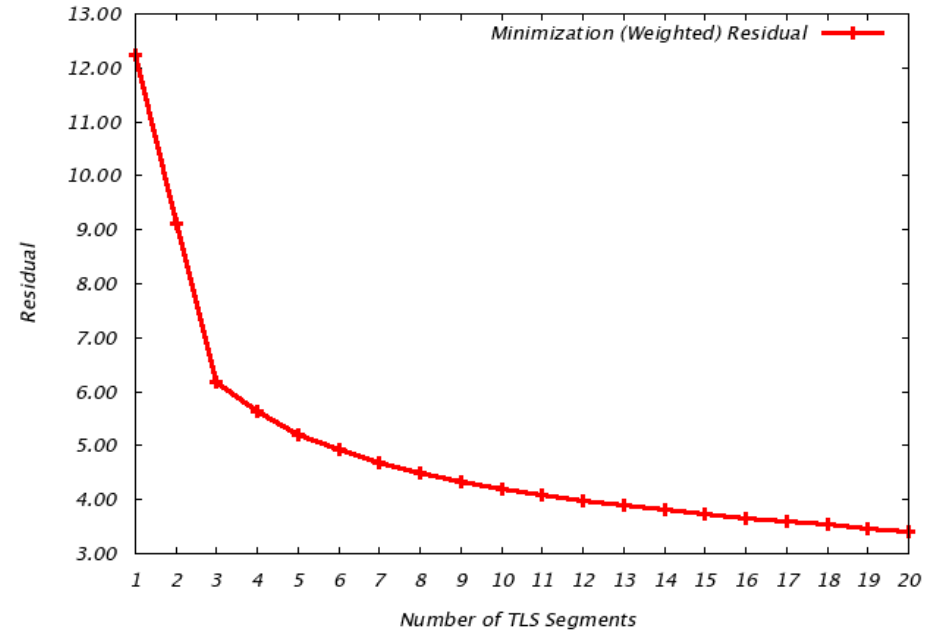


Methodology

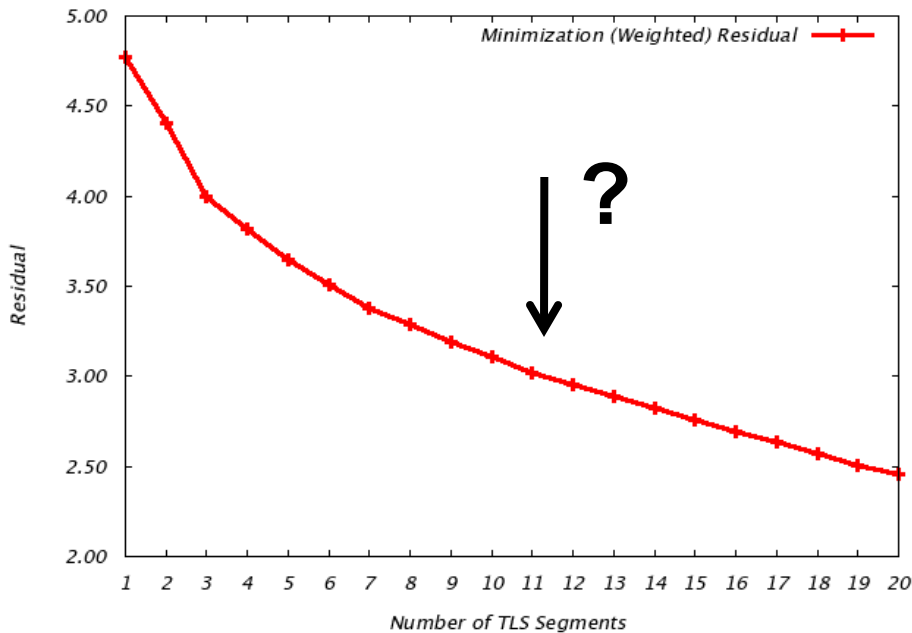
Least Squares Residual vs. Number of TLS Segments for XXXX Chain A



Least Squares Residual vs. Number of TLS Segments for XXXX Chain A



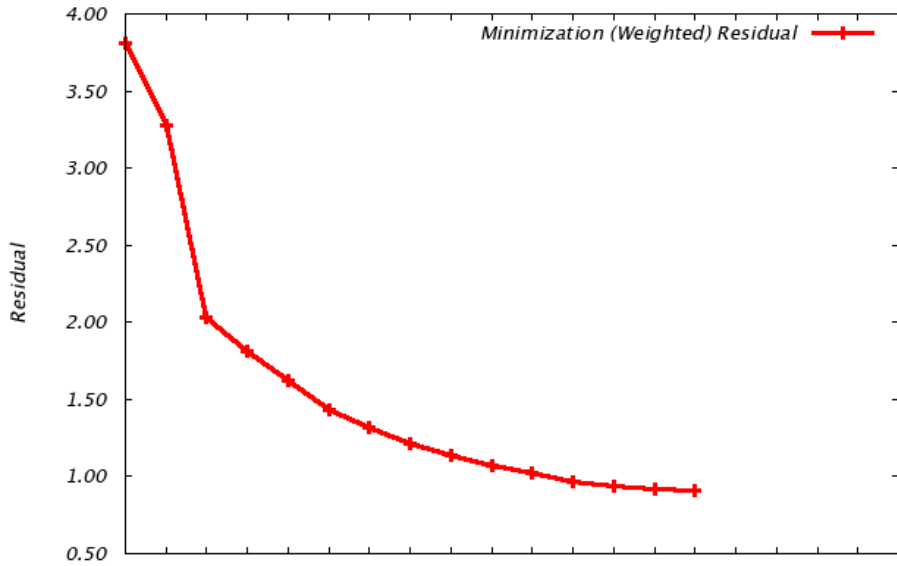
Least Squares Residual vs. Number of TLS Segments for XXXX Chain A



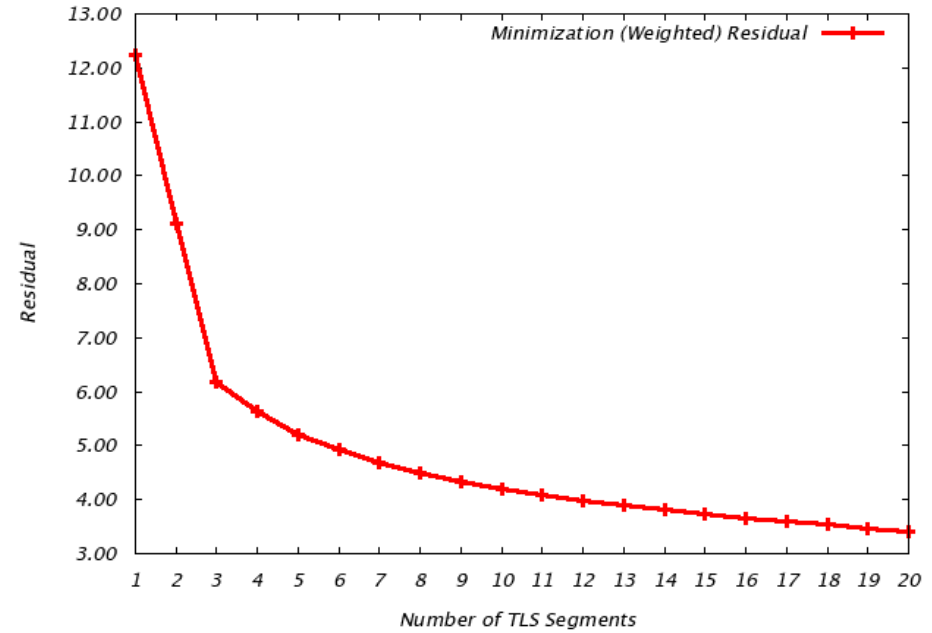
How to pick up the right number of TLS groups?

Methodology

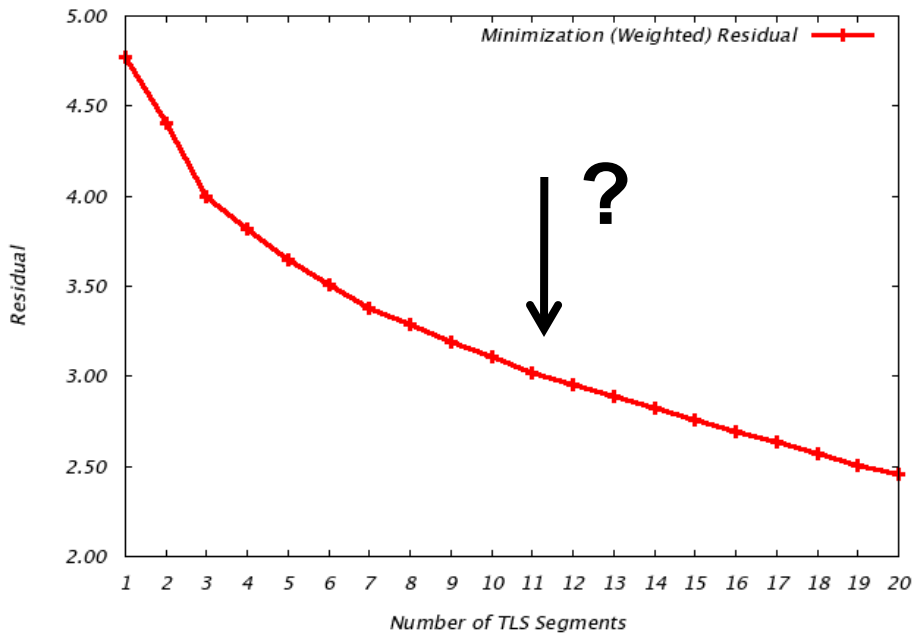
Least Squares Residual vs. Number of TLS Segments for XXXX Chain A



Least Squares Residual vs. Number of TLS Segments for XXXX Chain A



Least Squares Residual vs. Number of TLS Segments for XXXX Chain A



How to pick up the right number of TLS groups?

One can try all 20 or choose an arbitrary break point

PHENIX approach to finding TLS groups

- **Design Goals:**

- Have it as integrated part of PHENIX system:
 - No need to run external software or use web servers (=send your data somewhere, which your policy may even not allow you to do).
 - Use it interactively as part of refinement (update TLS group assignment as model improves during refinement).
 - Make it fast
- Eliminate subjective decisions (procedure should give THE UNIQUE answer and not an array of possible choices leaving the room for subjective decisions).

PHENIX approach to finding TLS groups

Step 1: For each chain find all secondary structure and unstructured elements

- Number of elements defines maximum possible number of TLS groups
- A secondary structure element can't be split into multiple TLS groups. Large unstructured elements, can be split into smaller pieces.

Chain



S – Secondary structure element

U – Unstructured stretch of residues (loop)

Step 2: Find all possible contiguous combinations



3 groups



2 groups

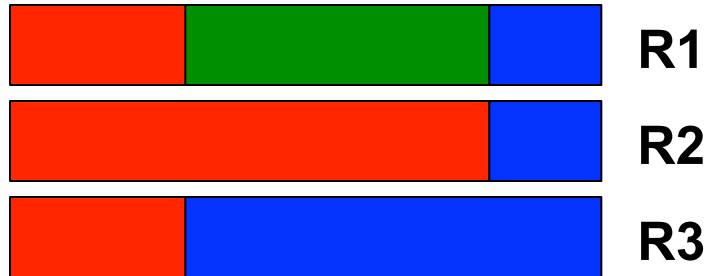


2 groups

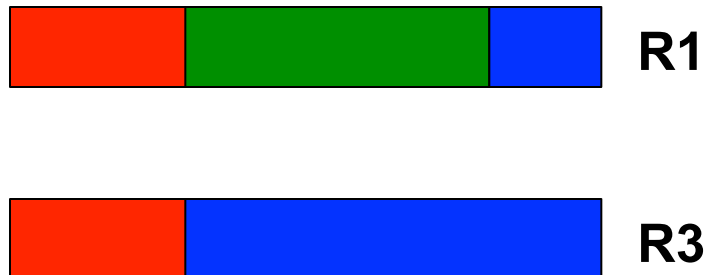
$N_{\text{ELEMENTS}} : N_{\text{POSSIBLE PARTITIONS}}$ 3:3, 4:7, 5:15, 6:31, ..., 10:511, ...

PHENIX approach to finding TLS groups

Step 3: For each partition fit TLS groups and compute the residual



Step 4: Find the best fit among the groups of equal number of partitions. In this example, if $R3 < R2$:



Step 5: Find the best partition...

- Challenge: we can't directly compare **R1** and **R3** because they are computed using different number of TLS groups (different number of parameters)

PHENIX approach to finding TLS groups

Step 5 (continued): Find the best partition...

- Randomly generate a pool of partitions for each candidate, fit TLS matrices compute, average residuals, and compute score:



R1



R11



R12

... many (20-50)

Average residuals: $R1_{\text{AVERAGE}}$

$$\text{Score} = (R1_{\text{AVERAGE}} - R1) / (R1_{\text{AVERAGE}} + R1) * 100$$

Do the same for the next candidate:



R3

The final solution is the one that has the highest score.

PHENIX approach to finding TLS groups: examples

GroEI structure (one chain):

No. of groups	Targets			
	best	rand.pick	diff.	score
2	680.7	869.2	188.5	12.2
3	297.1	665.7	368.6	38.3
4	260.4	448.3	187.8	26.5
5	206.2	342.1	135.8	24.8
6	188.4	264.7	76.3	16.8
7	182.3	251.2	68.9	15.9
8	176.9	229.5	52.5	12.9
9	173.1	207.3	34.1	9.0
10	170.2	196.8	26.6	7.2
11	167.8	183.1	15.2	4.3
12	165.6	179.0	13.4	3.9
13	163.9	170.8	6.9	2.1

PHENIX approach to finding TLS groups: usage

Command line:

```
phenix.find_tls_groups model.pdb nproc=N
```

where nproc is the number of available CPUs

TLS groups for refinement automatically

phenix.refine

PHENIX Preferences Help Run Abort Save Graphics ReadySet NCS TLS Xtrigae

Configure

Input data Refinement settings Output

Input files

File path	Format	Data type
/Users/afonine/PHENIX-dev-625/phenix_001.pdb	PDB	Input model

Space group: P 62 2 2

X-ray data and experimental phases

Data labels:

High resolution: Å

Options...

Neutron data

Data labels:

High resolution: Å

Idle

phenix.refine

Find TLS groups TLSMD server Load file Save file Revert Cancel Update and exit Graphics

TLS refinement groups

You can run phenix.find_tls_groups automatically to identify suitable TLS groups using a server* at the University of Washington. If you have selected the TLS refinement groups, phenix.refine will treat the group, but this is not recommended.

TLS groups:

Atom selection
chain 'A' and (resseq 295:394)
chain 'A' and (resseq 395:434)
chain 'A' and (resseq 435:569)

+ Add - Delete Update item View/pick

Edit selected: chain 'A' and (resseq 295:394)

(* <http://skuld.bmsc.washington.edu/~tlsmd>)

Number of processors for phenix.find_tls_groups: 2

phenix.refine

Close window Models Maps Settings Help Select atoms Zoom selection Clear selection Revert

Object: phenix_001.pdb Selection: chain 'A' and (resseq 295:394)

Mouse: Rotate view 827 atoms selected

Fast Automatic TLS

Examples:

- GroEL structure (3668 residues, 26957 atoms, 7 chains):

PHENIX: 135 seconds

TLSMD : 3630 seconds

(plus lots of clicking to upload/download the files
and making arbitrary decisions)

- Lysozyme structure:
 - 9.5 seconds with one CPU
 - 2.5 seconds using 10 CPUs

Automatic TLS

Why it is faster:

- Use isotropic TLS model,
- Solve optimization problem analytically (no minimizer used)
- A secondary structure element cannot belong to more than one TLS group

$$U_{TLS} = T_{iso} + \frac{1}{3} [L_{11}(y^2+z^2) + L_{22}(x^2+z^2) + L_{33}(x^2+y^2) - 2L_{12}xy - 2L_{13}xz - 2L_{23}yz + 2S_1z + 2S_2y + 2S_3x] \quad \Delta$$

$$U_{TLS} = T_{iso} + L_{11} \frac{y^2+z^2}{3} + L_{22} \frac{x^2+z^2}{3} + L_{33} \frac{x^2+y^2}{3} - \frac{2}{3}xyL_{12} - \frac{2}{3}xzL_{13} - \frac{2}{3}yzL_{23} + \frac{2}{3}zS_1 + \frac{2}{3}yS_2 + \frac{2}{3}xS_3$$

$$X = \frac{y^2+z^2}{3}; Y = \frac{x^2+z^2}{3}; Z = \frac{x^2+y^2}{3}; W = -\frac{2}{3}xy; V = -\frac{2}{3}xz; T = -\frac{2}{3}yz; S = \frac{2}{3}z; R = \frac{2}{3}y; Q = \frac{2}{3}x$$

$$U_{TLS} = T_{iso} + XL_{11} + YL_{22} + ZL_{33} + WL_{12} + VL_{13} + TL_{23} + S \cdot S_1 + R \cdot S_2 + Q \cdot S_3$$

$$LS = \sum_{ATOMS} (U_{TLS} - U_{iso})^2; LS = \sum_{ATOMS} (U_{TLS}^2 - 2U_{TLS}U_{iso} + U_{iso}^2)$$

$$LS = \sum_{ATOMS} (U_{TLS}^2 - 2U_{TLS}U_{iso}) + \sum_{ATOMS} U_{iso}^2$$

$$\frac{\partial LS}{\partial PAR} = \sum_{ATOMS} (2U_{TLS} \cdot \frac{\partial U_{TLS}}{\partial PAR} - 2U_{iso} \cdot \frac{\partial U_{TLS}}{\partial PAR})$$

$$\frac{\partial LS}{\partial PAR} = 2 \sum_{ATOMS} (U_{TLS} \frac{\partial U_{TLS}}{\partial PAR} - U_{iso} \frac{\partial U_{TLS}}{\partial PAR}) \quad \textcircled{1}$$

~~$$\frac{\partial LS}{\partial PAR} = 2 \sum_{ATOMS} (U_{TLS} \frac{\partial U_{TLS}}{\partial PAR} - U_{iso} \frac{\partial U_{TLS}}{\partial PAR})$$~~

$$\frac{\partial LS}{\partial PAR} = 2 \sum_{ATOMS} ([U_{TLS} - U_{iso}] \cdot \frac{\partial U_{TLS}}{\partial PAR})$$

$$\frac{\partial LS}{\partial L_{11}} = 2 \sum (U_{TLS} - U_{iso}) \cdot X = 2 \sum (U_{TLS} \cdot X - U_{iso} \cdot X) = \Delta$$

$$= 2 \sum (U_{TLS} \cdot X) - \frac{2 \sum U_{iso} \cdot X}{P_x} = \frac{2 \sum (U_{TLS} \cdot X) + P_x}{P_x}$$

$$\frac{\partial LS}{\partial L_{22}} = 2 \sum (U_{TLS} \cdot Y) + P_y \quad \frac{\partial LS}{\partial L_{23}} = \sum (U_{TLS} \cdot T) + P_T$$

$$\frac{\partial LS}{\partial L_{33}} = 2 \sum (U_{TLS} \cdot Z) + P_z \quad \frac{\partial LS}{\partial S_1} = \sum (U_{TLS} \cdot S) + P_S$$

$$\frac{\partial LS}{\partial L_{12}} = 2 \sum (U_{TLS} \cdot W) + P_w \quad \frac{\partial LS}{\partial S_2} = \sum (U_{TLS} \cdot R) + P_R$$

$$\frac{\partial LS}{\partial L_{13}} = 2 \sum (U_{TLS} \cdot V) + P_v \quad \frac{\partial LS}{\partial S_3} = \sum (U_{TLS} \cdot Q) + P_Q$$

$$\frac{\partial LS}{\partial T_{iso}} = 2 \sum_{ATOMS} ([U_{TLS} - U_{iso}] \cdot 1) = 2 \sum_{ATOMS} (U_{TLS} - U_{iso})$$

$$U_{TLS} \cdot X = XT_{iso} + X^2L_{11} + XYL_{22} + XZL_{33} + XWL_{12} + XVL_{13} + XT_{23} + XS \cdot S_1 + XR \cdot S_2 + XQ \cdot S_3$$

$$U_{TLS} \cdot Y = YT_{iso} + YXL_{11} + Y^2L_{22} + YZL_{33} + YWL_{12} + YVL_{13} + YT_{23} + YS \cdot S_1 + YR \cdot S_2 + YQ \cdot S_3$$

$$U_{TLS} \cdot Z = ZT_{iso} + ZXL_{11} + ZYL_{22} + Z^2L_{33} + ZWL_{12} + ZVL_{13} + ZT_{23} + ZS \cdot S_1 + ZR \cdot S_2 + ZQ \cdot S_3$$

$$U_{TLS} \cdot W = WT_{iso} + WXL_{11} + WYL_{22} + WZL_{33} + W^2L_{12} + WVL_{13} + WT_{23} + WS \cdot S_1 + WR \cdot S_2 + WQ \cdot S_3$$

$$U_{TLS} \cdot V = VT_{iso} + VXL_{11} + VYL_{22} + VZL_{33} + VWL_{12} + V^2L_{13} + VT_{23} + VS \cdot S_1 + VR \cdot S_2 + VQ \cdot S_3$$

$$U_{TLS} \cdot T = TT_{iso} + TXL_{11} + TYL_{22} + TZL_{33} + TWL_{12} + TVL_{13} + T^2L_{23} + TS \cdot S_1 + TR \cdot S_2 + TQ \cdot S_3$$

$$U_{TLS} \cdot S = ST_{iso} + SXL_{11} + SYL_{22} + SZL_{33} + SWL_{12} + SVL_{13} + STL_{23} + S^2 \cdot S_1 + SR \cdot S_2 + SQ \cdot S_3$$

7 more pages ...

ADP refinement: what goes into PDB

phenix.refine outputs TOTAL B-factor (iso- and anisotropic):

$$U_{\text{TOTAL}} = U_{\text{ATOM}} + U_{\text{TLS}} + \dots$$

Isotropic equivalent

ATOM	1	CA	ALA	1	37.211	30.126	28.127	1.00	26.82	C	
ANISOU	1	CA	ALA	1	3397	3397	3397	2634	2634	2634	C

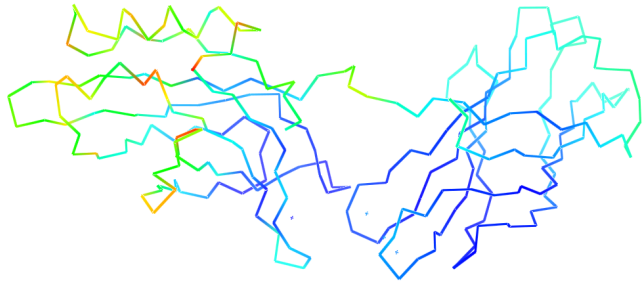
$$U_{\text{TOTAL}} = U_{\text{ATOM}} + U_{\text{TLS}} + \dots$$

Atom records are self-consistent:

- ✓ Straightforward visualization (color by B-factors, or anisotropic ellipsoids)
- ✓ Straightforward computation of other statistics (R-factors, etc.) – no need to use external helper programs for any conversions.

ADP refinement: example

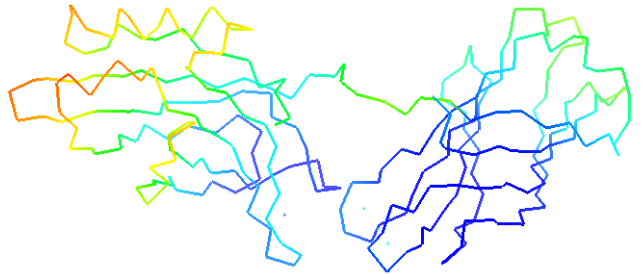
Synaptotagmin refinement at 3.2 Å



Original refinement (PDB code: 1DQV)

R-free = **34** %

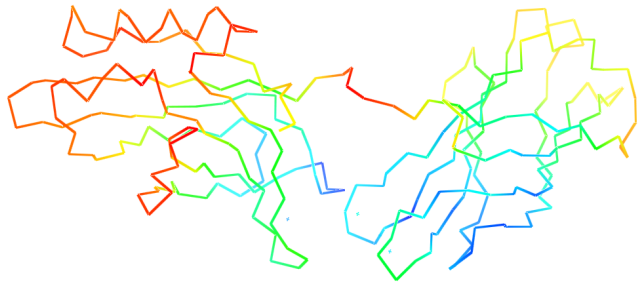
R = **29** %



PHENIX – Isotropic restrained ADP

R-free = **28** %

R = **23** %



PHENIX – TLS + Isotropic ADP

R-free = **25** %

R = **20** %

9% improvement in both Rwork and Rfree !

TLS groups determined automatically...