Automated TLS group determination in Phenix

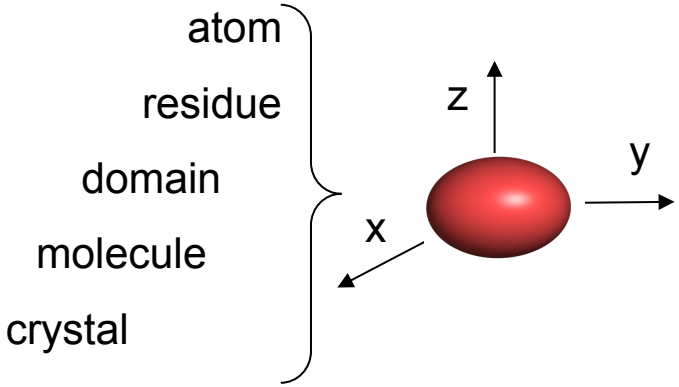
Pavel Afonine

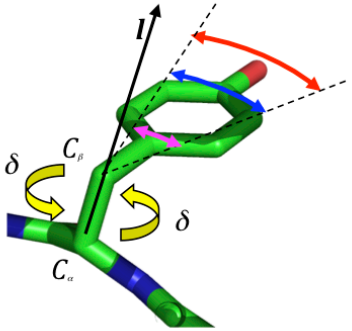
Computation Crystallography Initiative
Physical Biosciences Division
Lawrence Berkeley National Laboratory, Berkeley CA, USA

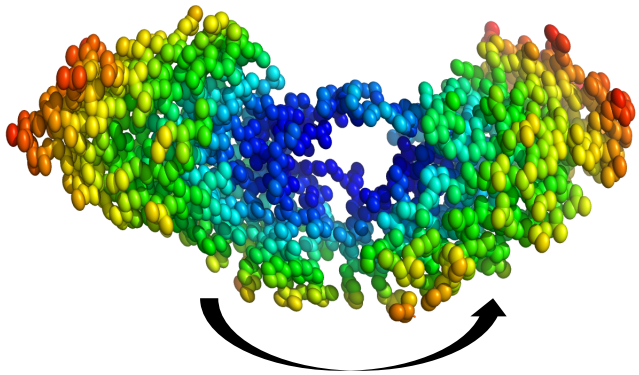
October, 2010

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

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

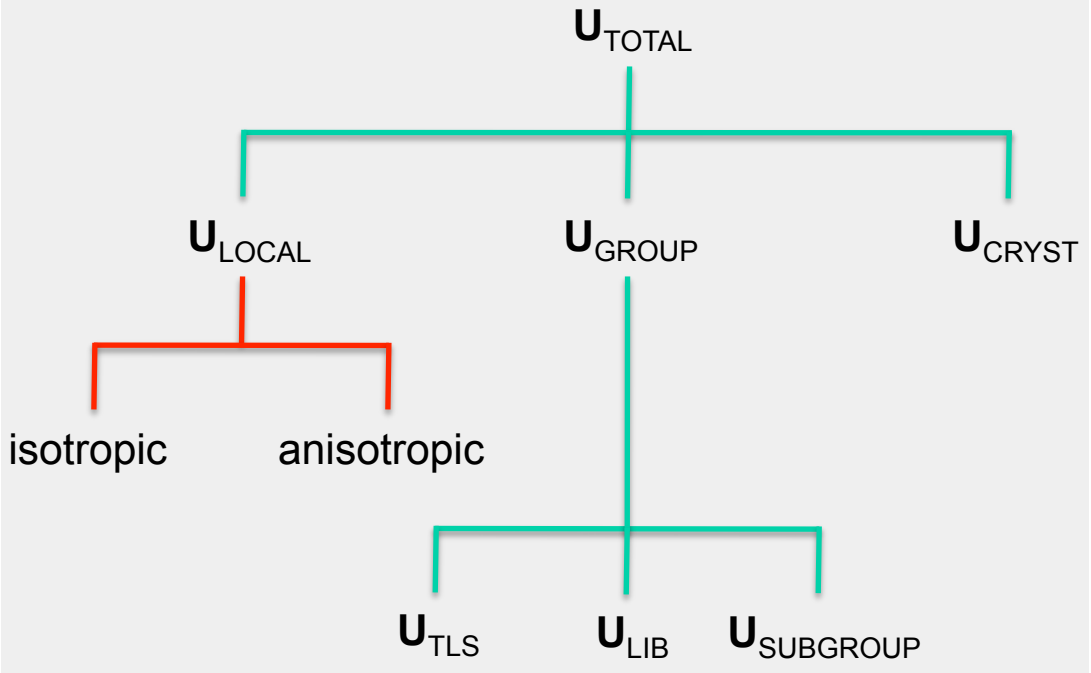






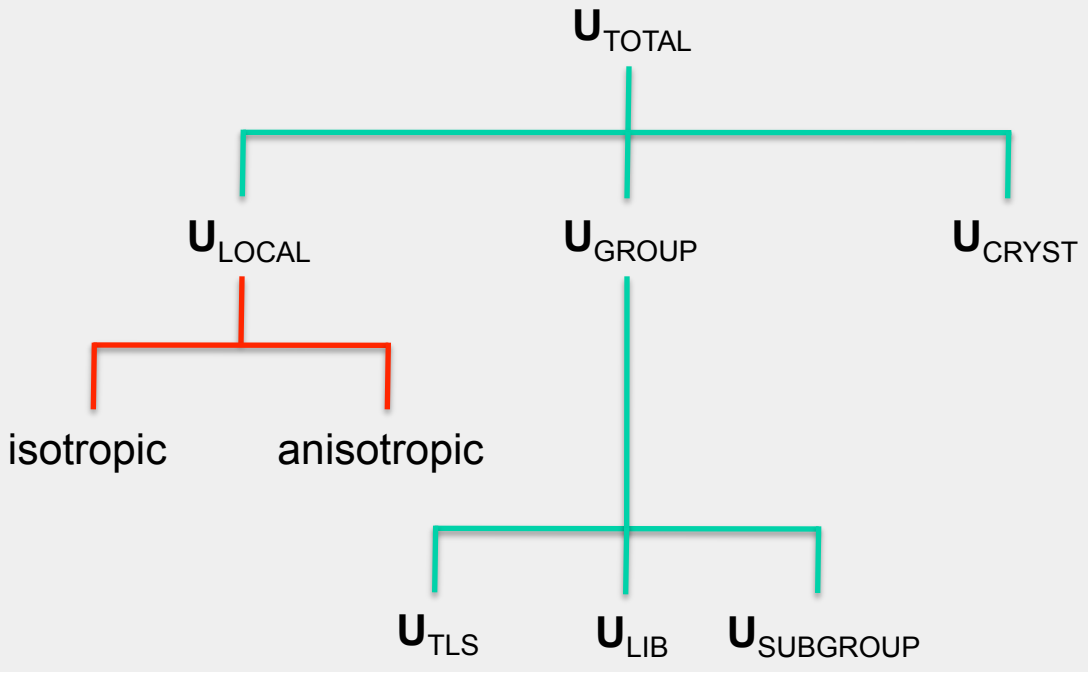
Total ADP:

$U_{TOTAL} = U_{CRYST} + U_{GROUP} + U_{LOCAL}$



Atomic Displacement Parameters (ADP or "B-factors")

- **Total ADP** $U_{TOTAL} = U_{CRYST} + U_{GROUP} + U_{LOCAL}$

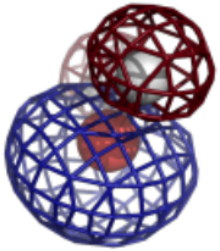


- U_{CRYST} – overall anisotropic scale.
- U_{TLS} – rigid body displacements of molecules, domains, secondary structure elements.
- U_{LOCAL} – local vibration of individual atoms.
- U_{LIB} – librational motion of side chain around bond vector.

TLS and ADP: comprehensive overview (~50 pages)

phenix-online.org/newsletter/CCN_2011_01.pdf

VOLUME TWO



COMPUTATIONAL CRYSTALLOGRAPHY NEWSLETTER

JANUARY MMXI

LABELIT SYMMETRY KING ROSETTA TLS TWINNING

Table of Contents

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. J. Headd, R. 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.

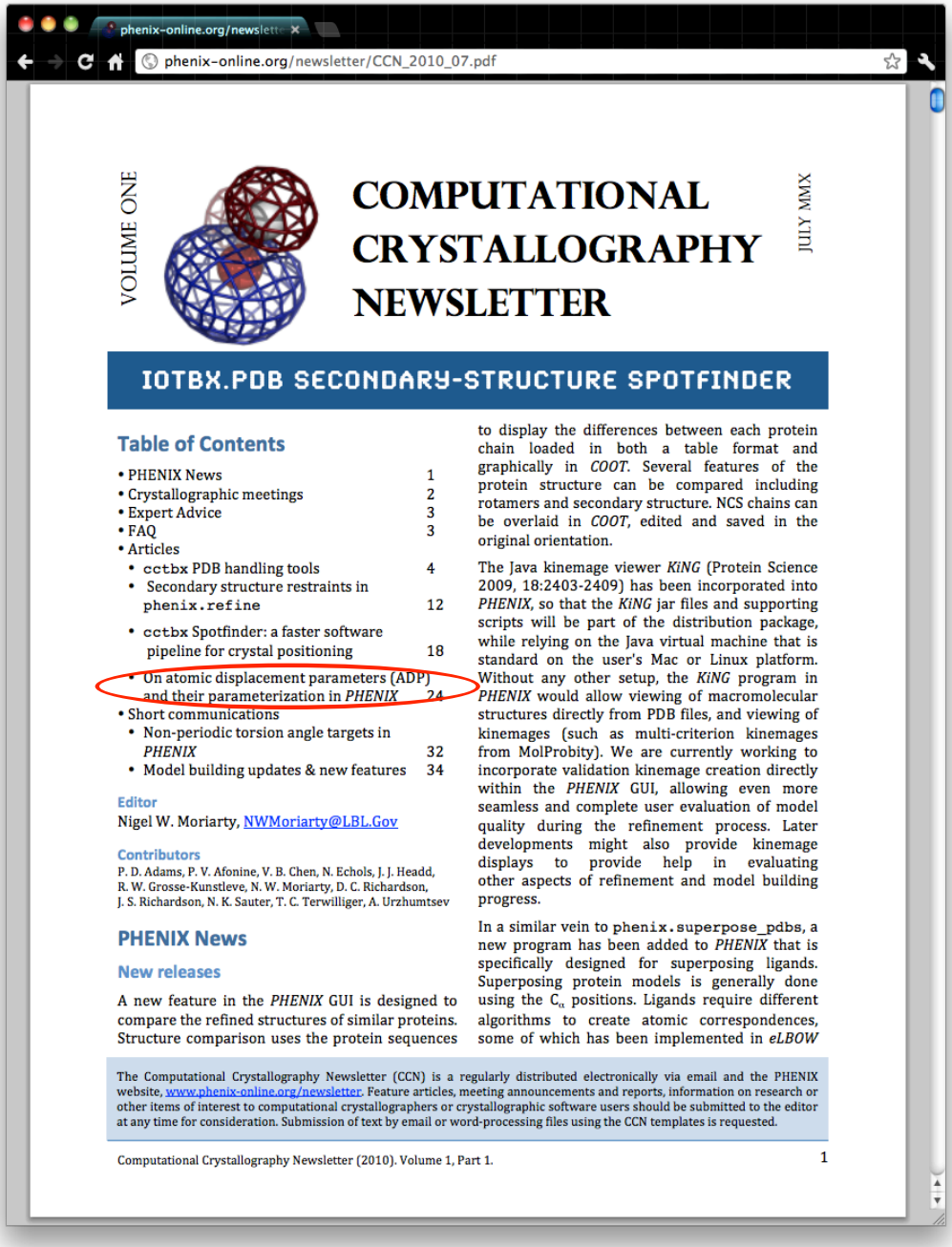
Generation of ensembles for Molecular Replacement (MR) is the goal of new release called `phenix.ensembler`. Another new release integrates MR and Rosetta in *PHENIX*. For more details, see the short communications for `phenix.ensembler` on page 8 and `phenix.mr_rosetta` on page 10.

New features

www.phenix-online.org

Parameterization and refinement of ADP in Phenix

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



phenix-online.org/newsletter/CCN_2010_07.pdf

VOLUME ONE

COMPUTATIONAL CRYSTALLOGRAPHY NEWSLETTER

JULY MMX

IOTBX.PDB SECONDARY-STRUCTURE SPOTFINDER

Table of Contents

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*

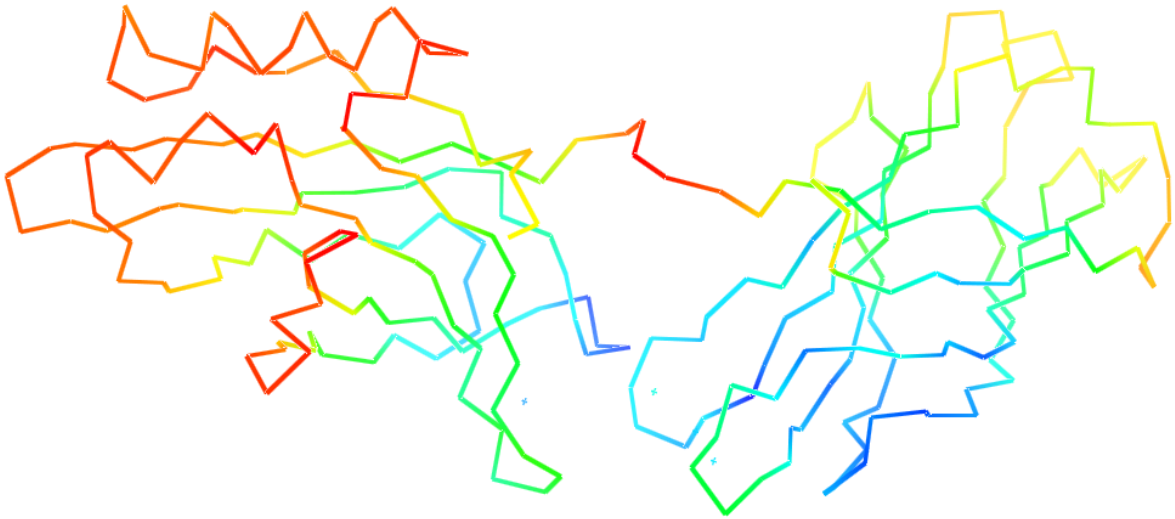
The Computational Crystallography Newsletter (CCN) is a regularly distributed electronically via email and the PHENIX website, www.phenix-online.org/newsletter. Feature articles, meeting announcements and reports, information on research or other items of interest to computational crystallographers or 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.

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

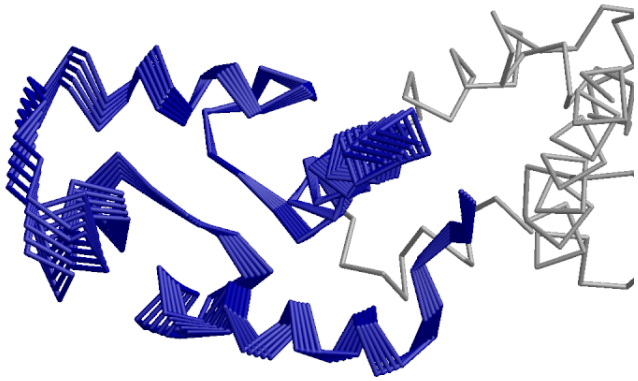


Methodology

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

$$\mathbf{U}_{TLS} = \mathbf{T} + \mathbf{ALA}^t + \mathbf{AS} + \mathbf{S}^t\mathbf{A}^t$$

(20 TLS parameters per TLS group)

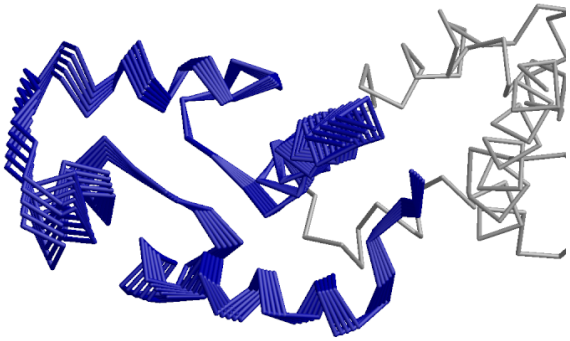
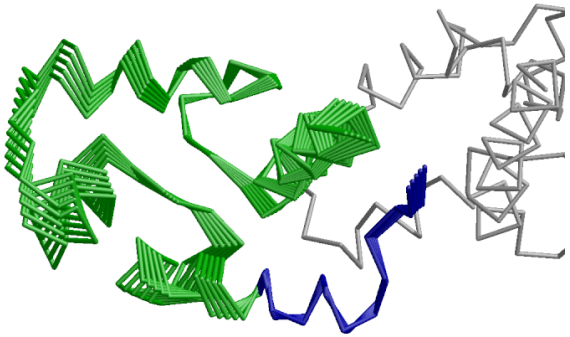
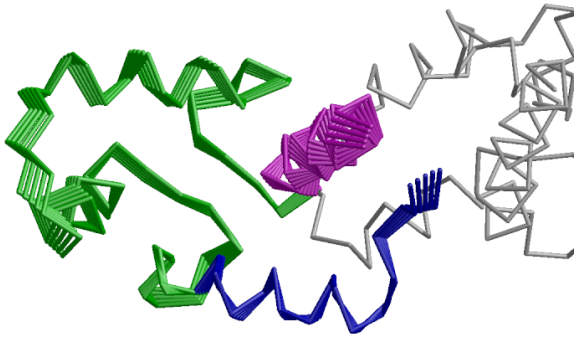


Methodology (TLSMD)

- 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

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

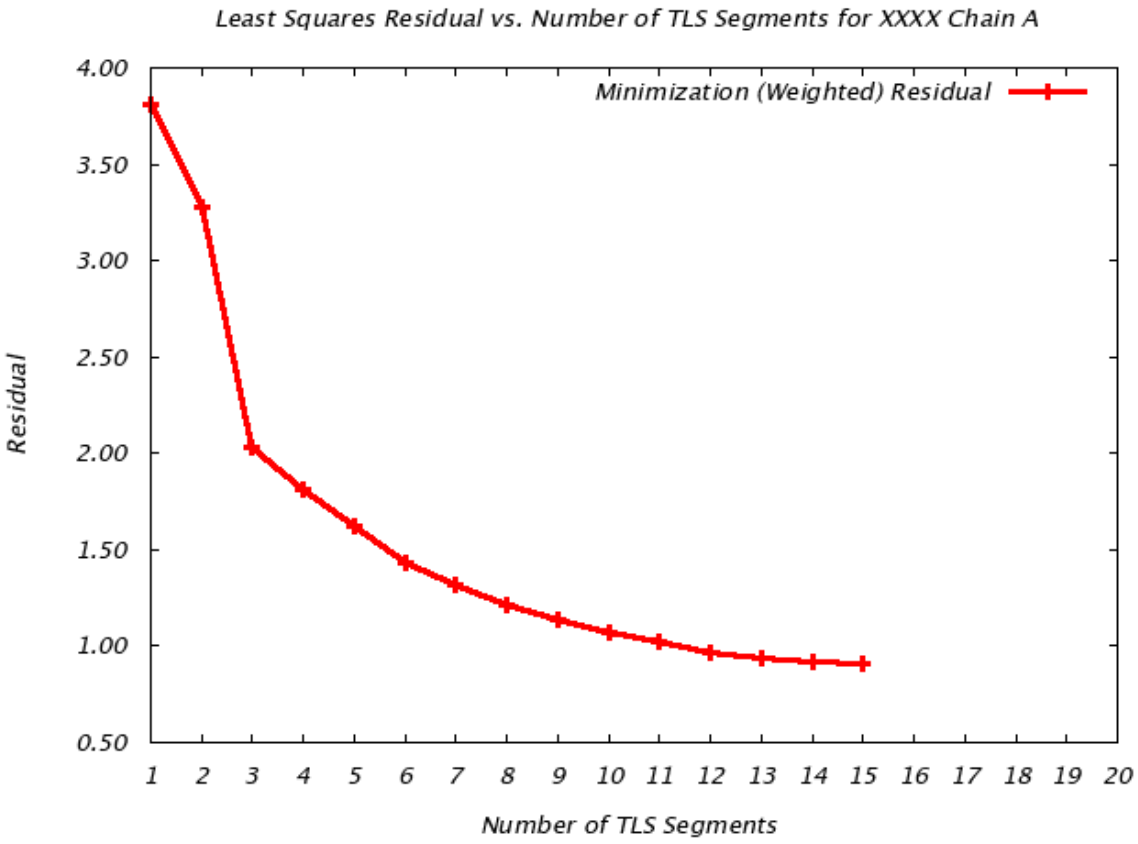
  - For example if n=100 and m=3 (minimal number of residues so there is more “observations” than parameters), then we get 4853 possible partitions.

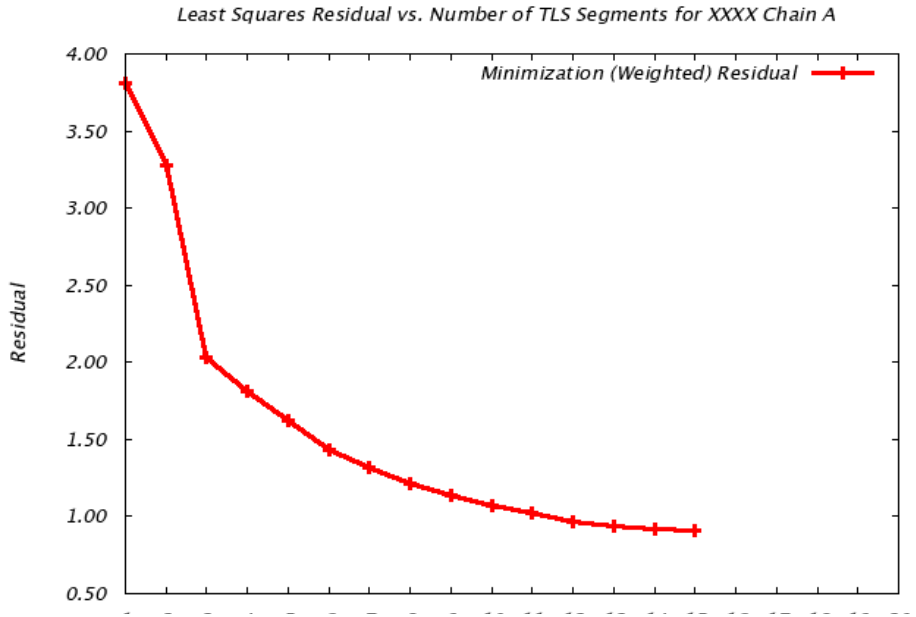
  - Compute residual for all segment partitions:

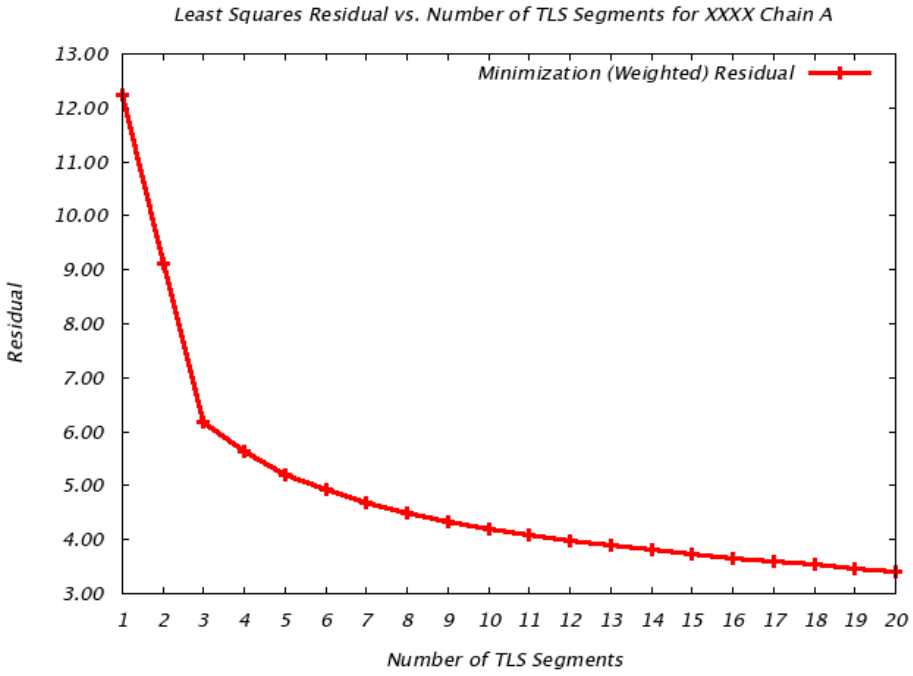
$$R = w \sum_{\text{atoms in segment}} W\left(\sum_{i=1}^{6}\left(U^{i}_{ATOM} - U^{i}_{TLS}\right)^{2}\right)$$

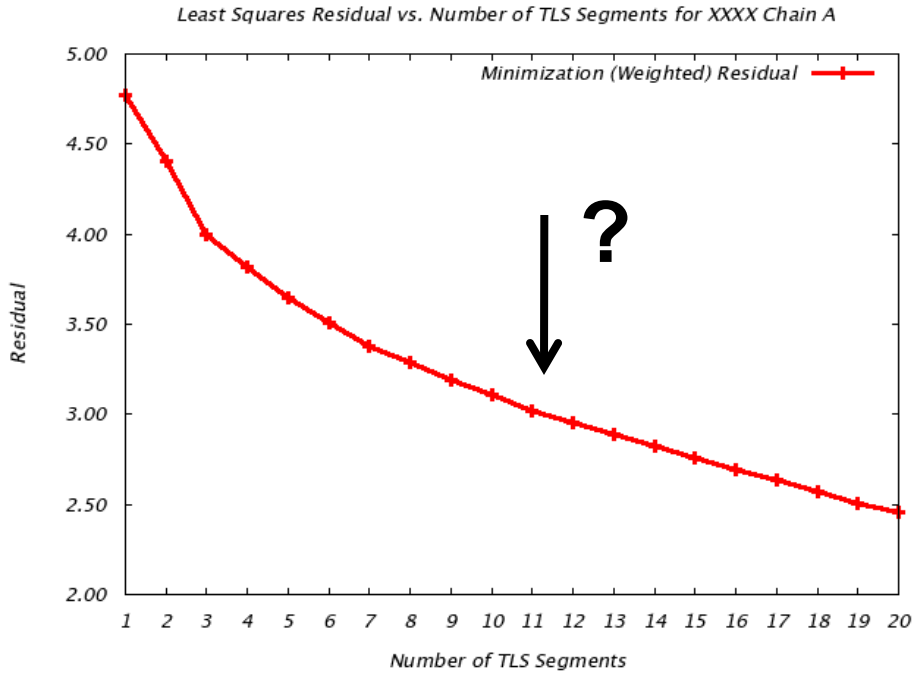
Methodology



Methodology

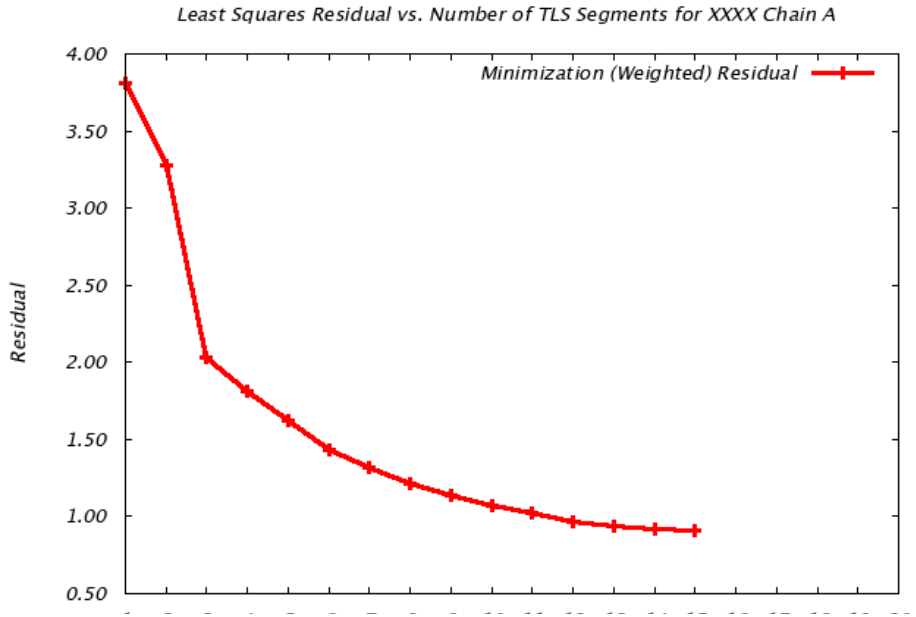


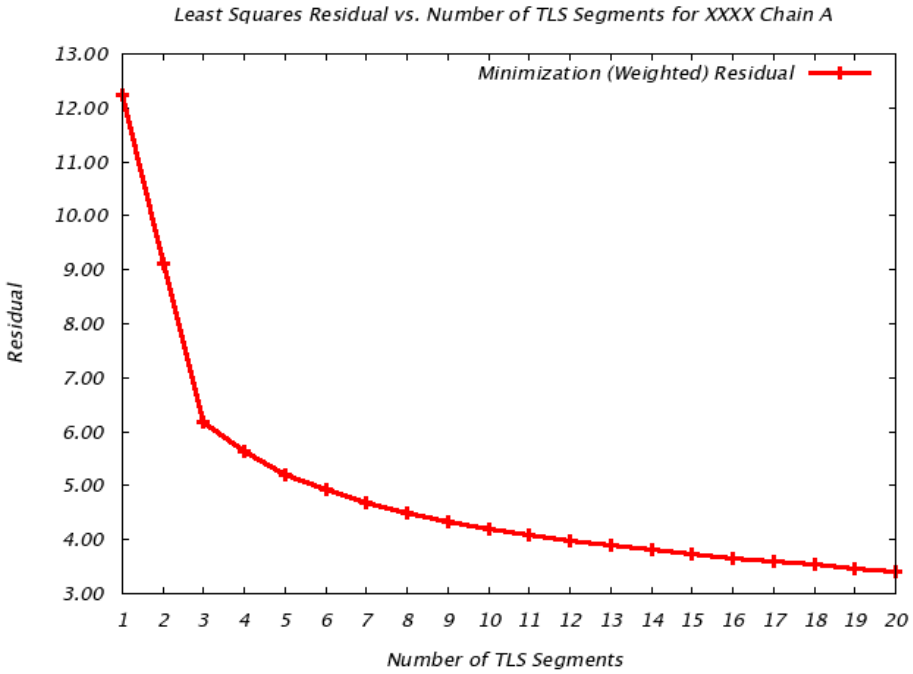


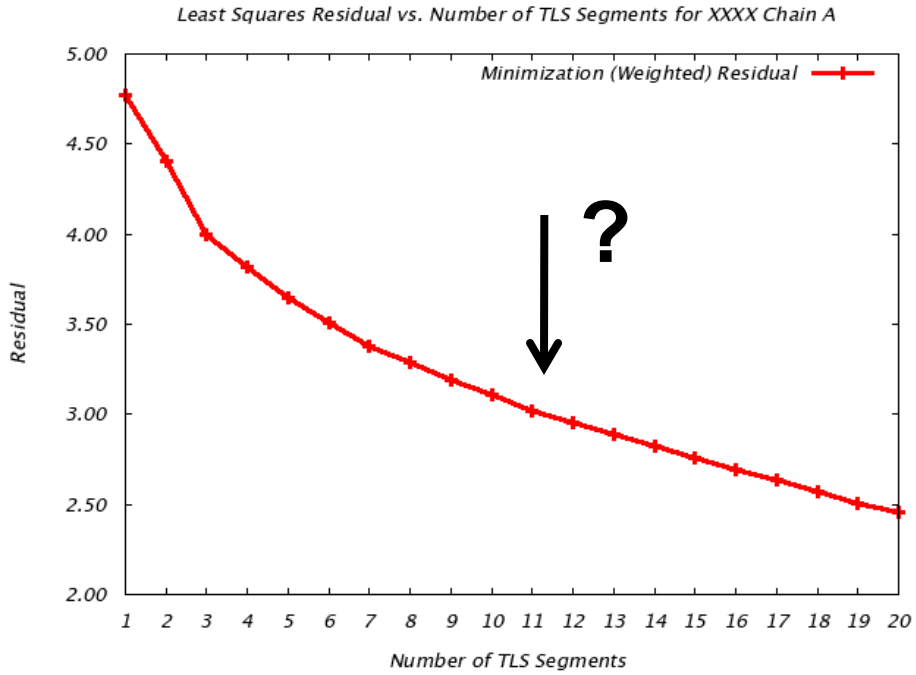


How to pick up the *right* number of TLS groups?

Methodology







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

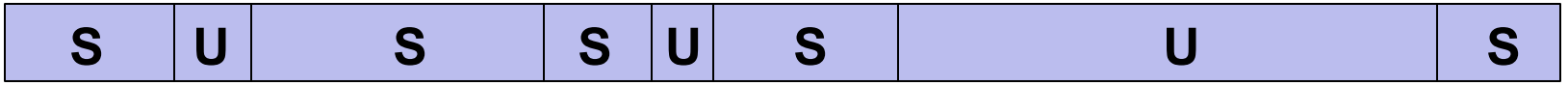
- **Design Goals:**
  - Have it as integrated part of PHENIX system:
    - No need to run external software or use web servers (=send your data somewhere, which your policy may even not allow you to do).
    - Use it interactively as part of refinement (update TLS group assignment as model improves during 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

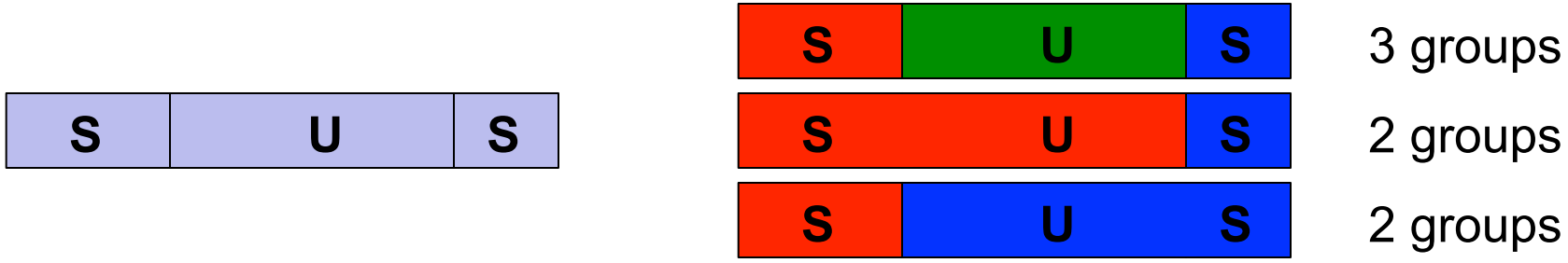
- *Number of elements defines maximum possible number of TLS groups*
- A secondary structure element can't be split into multiple TLS groups. Large unstructured elements, can be split into smaller pieces.

Chain



S – Secondary structure element
U – Unstructured stretch of residues (loop)

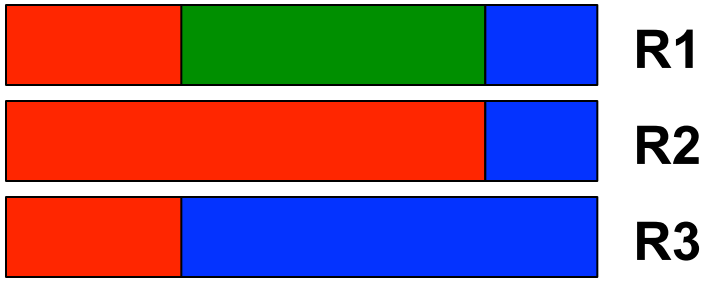
Step 2: Find all possible contiguous combinations



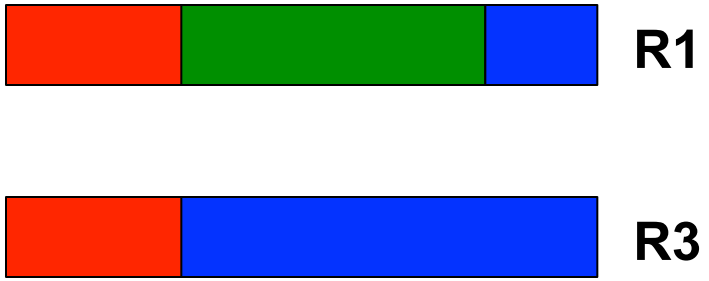
$N_{ELEMENTS} : N_{POSSIBLE\ PARTITIONS}$ 3:3, 4:7, 5:15, 6:31, …, 10:511, …

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: $\mathbf{R1_{AVERAGE}}$

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

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

GroEl 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

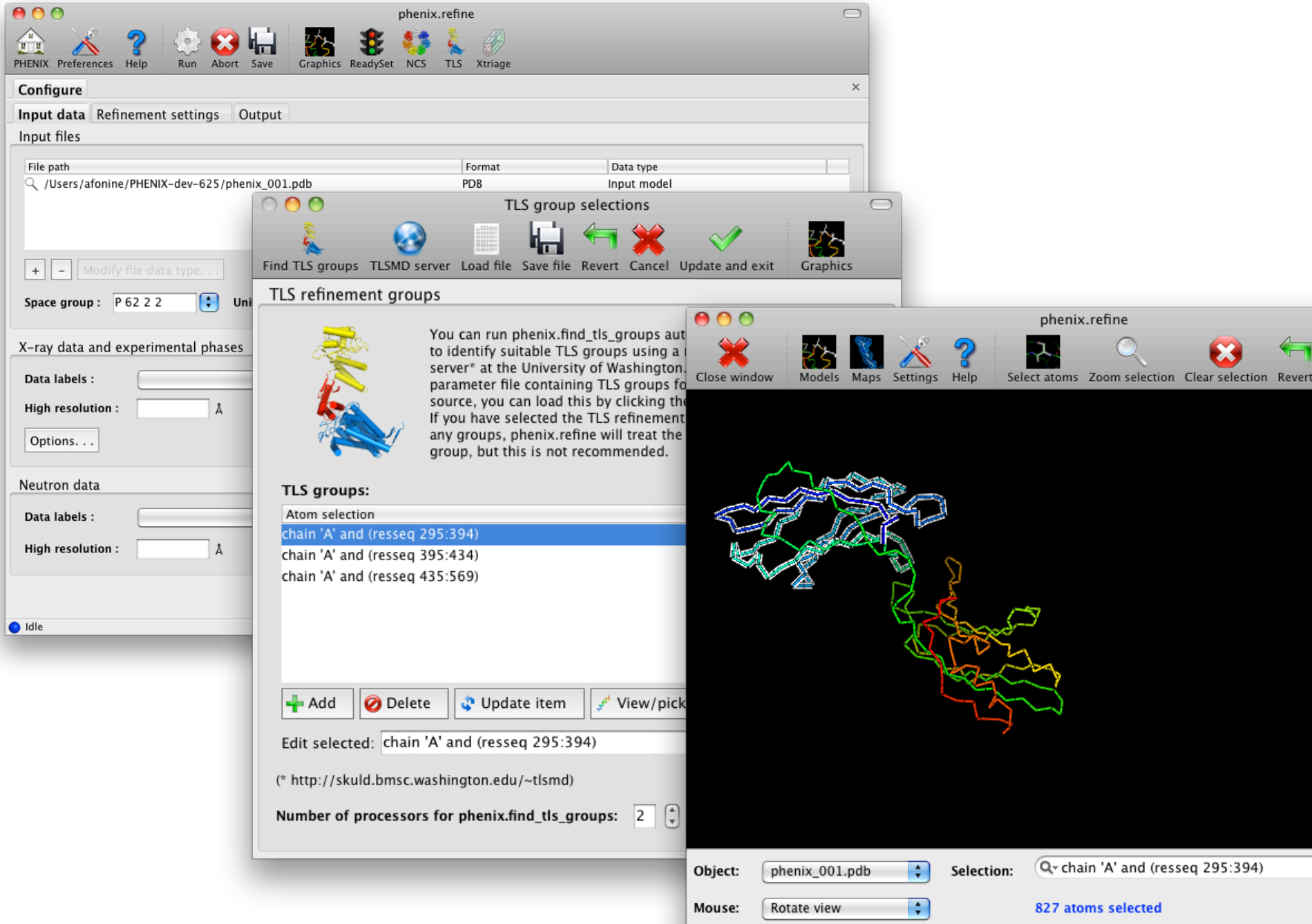
PHENIX approach to finding TLS groups: usage

Command line:

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

where nproc is the number of available CPUs

TLS groups for refinement automatically



Fast Automatic TLS

Examples:

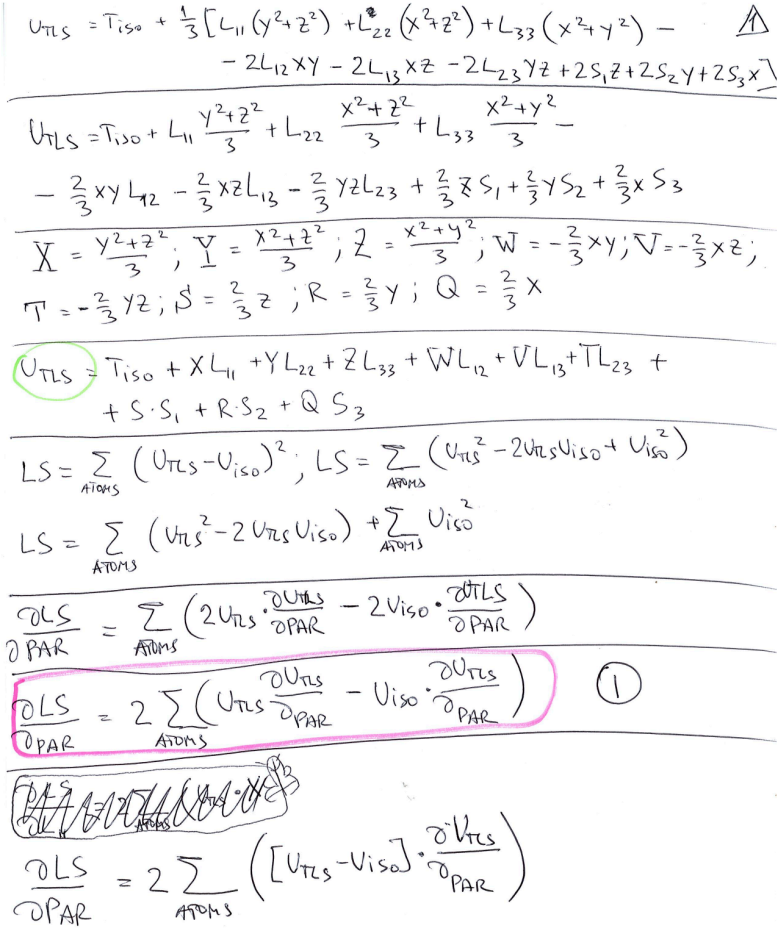
- GroEL structure (3668 residues, 26957 atoms, 7 chains):

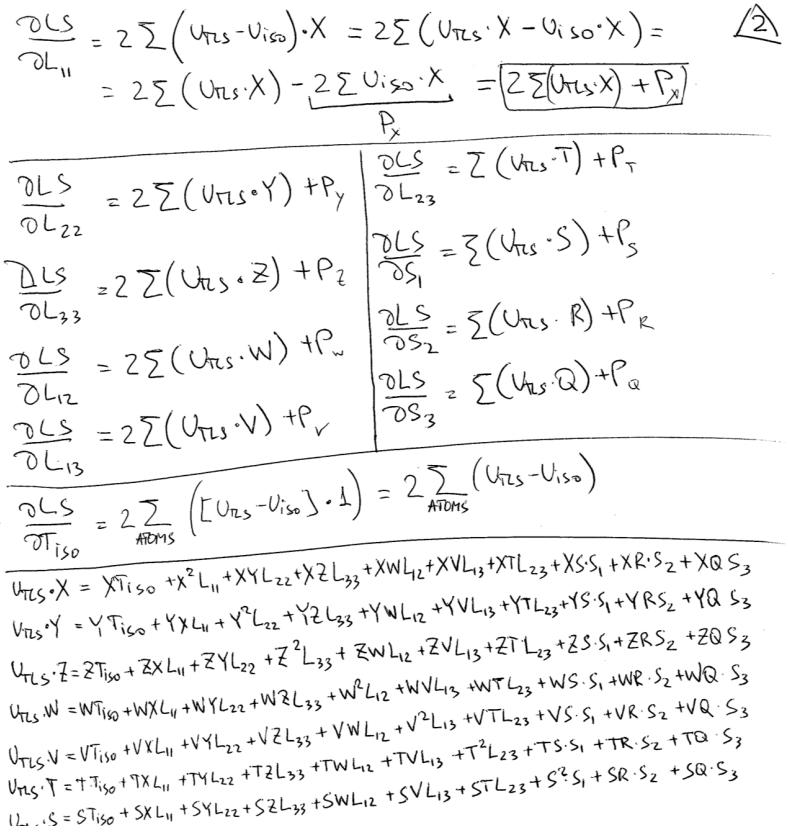
  PHENIX: 135 seconds
  TLSMD : 3630 seconds
  (plus lots of clicking to upload/download the files and making arbitrary decisions)

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

Automatic TLS

- **Why it is faster:**
  a) Use isotropic TLS model,
  b) Solve optimization problem analytically (no minimizer used)
  c) A secondary structure element cannot belong to more than one TLS group

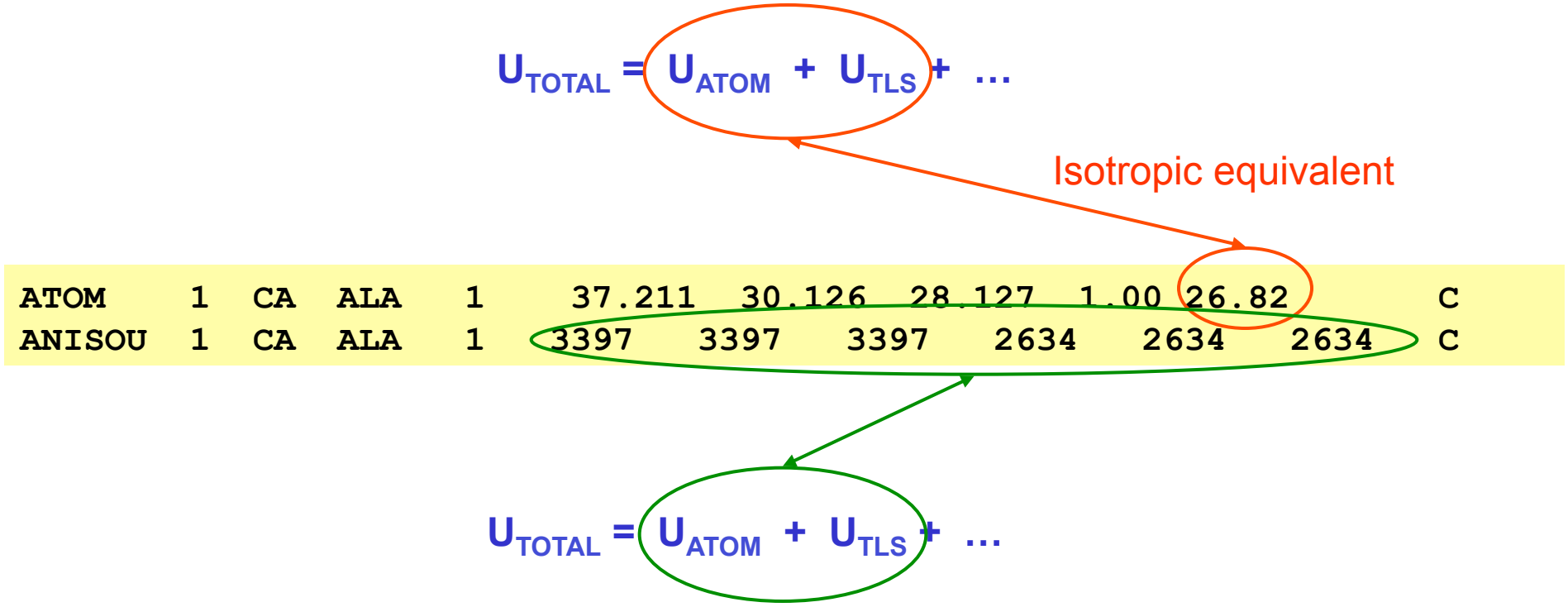




7 more pages …

ADP refinement: what goes into PDB

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

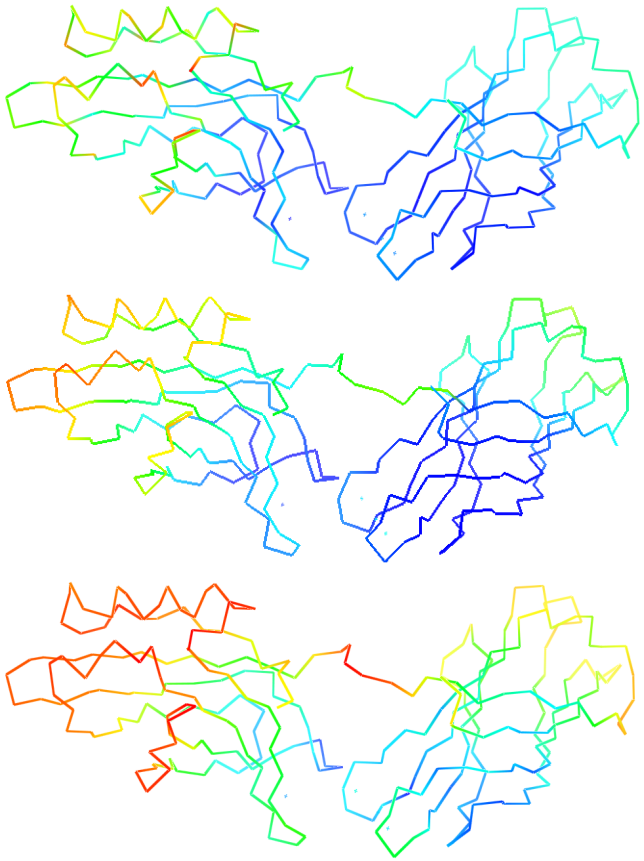


Atom records are self-consistent:

- ✓ 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…