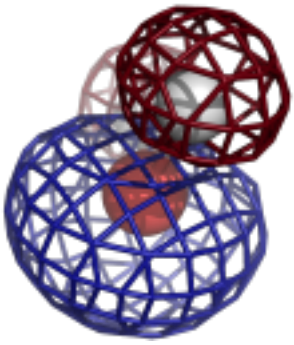
VOLUME EIGHT



COMPUTATIONAL CRYSTALLOGRAPHY NEWSLETTER

JULY MMXVII

1.12, SER & THR IN HELICES

Table of Contents

Editor

Nigel W. Moriarty, NWMoriarty@LBL.Gov

Phenix News

Announcements

Phenix 1.12 release

The Phenix developers are pleased to announce that version 1.12 of Phenix is now available (build 1.12-2829). Binary installers for Linux, Mac OSX, and Windows platforms are available at the download site.[1] Highlights for this version include new tools for secondary structure validation, comparing multiple related structures, cryo-EM model and map validation, and calculation of real space difference maps. Video tutorials are now available on the new Phenix Tutorial YouTube channel.[2] A suite of tools for cryo-EM model building and structure refinement including automated map sharpening, segmentation and model building has also been added. Real space refinement now has an option to use a histogram-equalized map as a refinement target. Phaser version 2.8 has several new features, including unit cell refinement for EM data, more extensive use of eLLG in decision making and analysis, anomalous substructure determination (Phassade), improved tNCS detection and analysis, and improvements to SCEDS.

New programs

Structure Comparison

Tool for parallel validation and analysis of near-identical protein structures. Results include the analysis of ligands, rotamers,

[1] http://phenix-online.org/download/

[2] www.youtube.com/c/phenixtutorials

Ramachandran angles, missing atoms, water locations and B-factors.

phenix.mtriage

Command-line tool to validate cryo-EM models has been addded. GUI is available. Computes various map statistics including resolution estimates, FSC between half-maps, full and model maps. Outputs mask file used to compute FSC. Still under development.

phenix.real_space_diff_map

Command-line tool to compute difference map. This is a real-space analogue of Fo-Fc map mainly designed for cryo-EM maps.[1]

Crystallographic meetings and workshops

24th Congress and General Assembly of the International Union of Crystallography, August 21–28, 2017

Location: Hyderabad, India. A *Phenix* workshop will be held on the 21st of August. See conference website for details.

Expert advice

Fitting Tip #14 – How Ser & Thr behave oddly on helices

Jane Richardson and David Richardson

What is typical behavior on helices?

For most sidechain types, which have a Cγ group, the **p** (near +60°) rotamer of χ1 is completely forbidden on regular α-helix, because that group will seriously clash with the *i*-3 backbone CO of the preceding turn. At top left in figure 1, the clashes (red spike clusters) are shown between Cγ H atoms and the preceding helix turn (and a water), for a Thr misfit with the Cγ at **p** χ1. The no-**p**-on-helix rule was noted very early (McGregor 1987) and tabulated as percent occurrence for α, β, and 'other' categories in the penultimate rotamer library paper (Lovell 2000). In contrast, the **p** rotamer position is actually preferred for Ser and Thr on α-helix.

In general, **p** is by far the least preferred of the three χ1 angle optima (**p**, **m**, & **t**). For 14 of the 17 rotameric amino acids, the percentage of **p** averages only 10.9%, with a range of 6.2 to 17.7%. In contrast, for Ser and Thr the **p** rotamer is the commonest, at 48%. Leu is an exception in the other direction, at 0.8%, because its branched Cγ means that any χ2 value clashes with backbone if χ1 is **p**. This means that even before folding, Leu residues are set up with the right rotamers for α-helix, which is presumably a major reason that Leu has the highest helix propensity.

Serine and Threonine

Of course, the answer to this conundrum is that for Ser or Thr, the Oγ can make a short, favorable H-bond with the *i*-3 CO, as seen at top right in figure 1. When the misfit Thr is rebuilt, every criterion gets better, as listed in figure 1. At this high resolution it is easy to see what happens. In order to fit the reversed O and C branches into their strong density peaks, the Cβ is forced entirely out of its density and has huge geometry distortions. The right answer of an H-bond seems really obvious if you have all-atom contacts to judge by, which was not the case for these depositors in the 1990s. Also, one is used to expecting an OH with solvent exposure to point outward, not inward.

This tucked-in **p** rotamer arrangement results in a bifurcated H-bond for the *i*-3 CO: one branch is from the Thr Oγ and the other is the normal helical H-bond. We do not know of a study addressing this specific issue, but our

[1]See http://phenix-online.org/newsletter/CCN_2017_01.pdf for details.

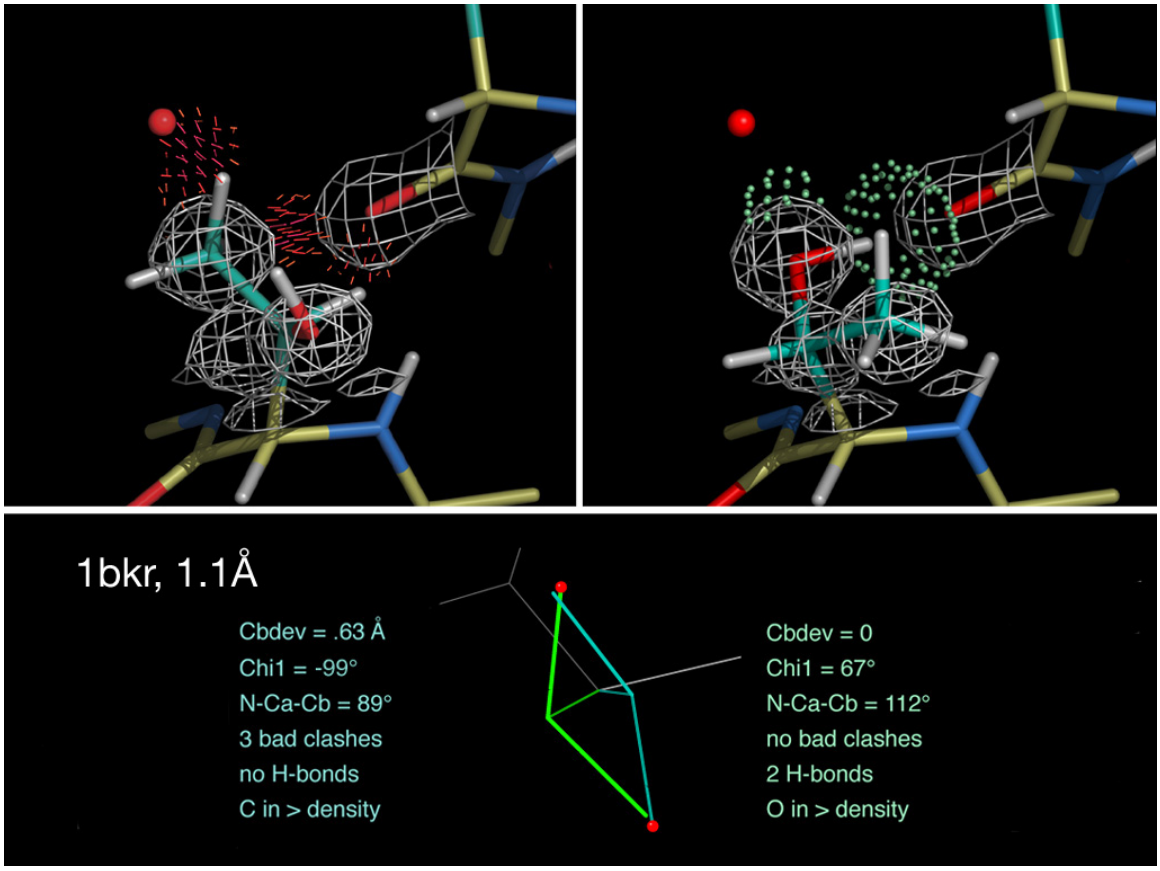


Figure 1: A misfit (left) vs corrected (right) Thr sidechain on an α-helix at 1.1Å resolution. The methyl group clashes with the *i*-3 carbonyl O and also with a nearby water (red ball), while in contrast the OH in that position can make two good H-bonds. Geometry and fit to density improve greatly, as listed in the lower panel. All-atom clashes are shown as clusters of red spikes, and H-bonds as lenses of pale-green dots. Thr 101 in 1bkr (Banuelos 1998).

impression is that such bifurcated H-bonds are in general somewhat weaker than a good single H-bond. That idea is supported by the fact that Ser and Thr have quite low helix propensities. Asn, also disfavored in helix, can make a similar bifurcated H-bond from its sidechain amide to the *i*-4 CO, in an **m-80** rotamer; however, more often the amide lies against the outer surface of the *i*-4 peptide, in rotamer **m-20** (Lovell 1999)

In membrane proteins

The H-bonded **p** conformation of Ser and Thr on helix is fairly common in globular proteins. It is even more favorable on trans-membrane helices, where there is very seldom a polar group to H-bond with an outward-pointing OH. A membrane-buried helical Ser or Thr is therefore very likely to adopt a **p** rotamer, where presumably it gains more from forming an H-bond than the helix loses from the bifurcation destabilizing a backbone H-bond somewhat. Figure 2 illustrates one of the many Ser/Thr H-bonded to their trans-membrane helices in the 5u9w trimeric transporter structure (Hirschi 2017). The left panel shows chain A Ser 110 within the membrane trimer. The right panel is a closeup, showing the inward-pointing

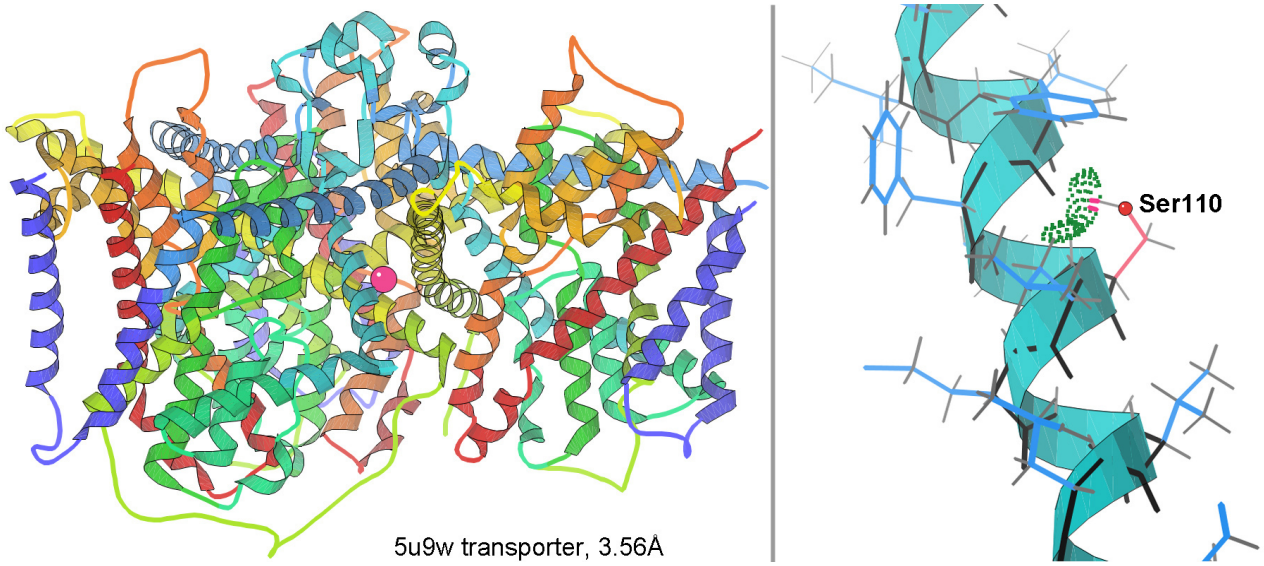


Figure 2: A **p**-rotamer, *i*-3 H-bonded serine on a trans-membrane helix at 3.56Å. Left panel shows ribbons for the trimeric transporter molecule viewed in the plane of the membrane, with Ser 110 as a hot-pink ball. Right panel is a closeup of Ser 110 on its helix, with green dot lenses for the bifurcated H-bond, one H-bond branch from the *i* -3 CO to the Ser sidechain OH and the other branch the normal helical H-bond. From 5u9w (Hirschi 2017)

sidechain, the bifurcated H-bond, and the regular α-helix.

The bottom line

In any protein structure, it is worth considering an i-3 H-bonded **p** rotamer for a Ser or Thr on an α-helix. For a trans-membrane helix, that arrangement is the dominant one, unless the sidechain happens to be next to a polar group or solvent channel. Therefore, if the Ser/Thr orientation is unclear in a membrane protein by low-resolution X-ray or high-resolution cryoEM, choose the default H-bonded **p** rotamer.

References:

Banuelos S, Saraste M, Carugo KD (1998) Structural comparisons of calponin homology domains: impications for actin binding, *Structure* **6**: 1419-1431

Hintze BJ, Lewis SM, Richardson JS, Richardson DC (2016) MolProbity's ultimate rotamer-library distributions for model validation, *Proteins: Struct Func Bioinf* **84**: 1177-1189

Hirschi M, Johnson ZL, Lee S-Y (2017) Visualizing multistep elevator-like transitions of a nucleoside transporter, *Nature* **545**: 66-70

Lovell SC, Word JM, Richardson JS, Richardson DC (1999) Asparagine and glutamine rotamers: B-factor cutoff and correction of amide flips yield distinct clustering, *Proc Natl Acad Sci USA* **96**: 400-405

Lovell SC, Word JM, Richardson JS, Richardson DC (2000) The Penultimate Rotamer Library, *Proteins: Struct Func Genet* **40**: 389-408

McGregor MJ, Islam SA, Sternberg MJE (1987) Analysis of the relationship between side-chain conformation and secondary structure in globular proteins, *J Mol Biol* **198**: 295-310

FAQ

What is in the MTZ output by *phenix.refine*?

The MTZ file from *phenix.refine* contains four kinds of data:

1. Copy of input data. Why? For convenience and consistency. Inputs may not necessarily be in MTZ format and may be spread across multiple files of different format.
2. Data that were actually used in refinement. Why? A user has options to cut resolution from both ends, as well as apply cutting by sigma. Plus, *phenix.refine* may choose not to use a handful of reflections as outliers (Read, 1999). So it may be good to have set of reflections that were used in given refinement run.
3. Model in reciprocal space (F_{model}). Why? This is a reciprocal space representation of what's in model file except that it is richer because contains not only atomic model (F_{calc}) but also solvent contributions (bulk and non-uniform) as well as all scales. F_{model} taken from this array and F_{obs} described in item 2 are expected to reproduce reported R_{factor} exactly.
4. Fourier map coefficients ($2mF_{obs}$-DF_{model}, mF_{obs}-DF_{model}, anomalous map if applicable).

Item 1 and 3 can be optional because with trivial scripting one can obtain F_{model} using data from item 2 and 4.

Item 1 does duplicate data and it's not unreasonable to assume they are still available by the time you finalize your structure. However, space is cheap and it is much easier to have relevant arrays of data centralized in one place (file) rather than scattered across hard drive.

References:

Read, R. J. 1999. "Detecting Outliers in Non-Redundant Diffraction Data." *Acta Crystallographica Section D: Biological Crystallography* 55 (10): 1759–64. doi:10.1107/S0907444999008471.

phenix.mtriage: a tool for analysis and validation of cryo-EM 3D reconstructions

Pavel V. Afonine

Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

phenix.mtriage is a new *Phenix* tool for analysis of cryo-EM derived maps. Given a map it provides a basic statistics about the map, such as box parameters (three-dimensional gridding), map origin, min/max/mean map values, histogram of map values, etc. Also, it estimates the map resolution obtained by map perturbation and that we refer to as d_{99}. In a nutshell, this measure indicates up to what resolution Fourier map coefficients corresponding to the real-space map can be omitted before the map starts changing significantly. This is done by Fourier transforming the map to obtain a full set of possible Fourier map coefficients. Then highest resolution terms are removed gradually and for each chunk of removed terms the correlation between the original map and the map corresponding to the truncated set of Fourier coefficients is computed. Obviously, at the start the correlation is 1 and it decreases with the amount of removed terms. Empirically we find that 0.01 in correlation drop is indicative of a significant change in the map, so we note the resolution cut-off at which this occurs, d_{99}.

Providing half-maps will enable more validation: the FSC plot will be calculated (and available as a graphic picture) and the resolution derived at FSC=0.143 ("gold standard") will be shown (d_{FSC}). Additionally, three map histograms (full and two for each half-map) will be shown. The all three histograms are expected to be similar.

Finally, if atomic model is provided then even more statistics will be generated. It is possible to use an atomic model to estimate map resolution by generating a series of model-calculated maps at different resolutions and comparing them with the experimental map. The resolution of the model-calculated map (d_{model}) that maximizes the correlation is considered as the most representative of the map resolution. Overall B-factor is calculated and reported as part of d_{model} calculation. Also, it is possible to calculate the correlation between model and experimental maps in Fourier space, similarly to FSC for half-maps. It is informative as it gives an estimate up to what resolution (d_{FSC_model}) there is at least some signal (FSC=0). Clearly, these measures are model-dependent and as accurate as accurate the model is.

As will be demonstrated elsewhere, it is informative to consider all four resolution estimates d_{FSC}, d_{99}, d_{FSC_model}, d_{model} and overall B as collectively they may indicate issues ranging from as trivial as typos to particular quirks of the data.

This is an ongoing development and details will be published elsewhere. More functionality is likely to appear in future. *phenix.mtriage* is available as a command line tool and in *Phenix* GUI. Users are encouraged to provide their feedback and requests for options.

More TLS validation with *phenix.tls_analysis*

Pavel V. Afonine

Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

phenix.tls_analysis is a *Phenix* tool for analysis and validation of refined TLS parameters (Computational Crystallography Newsletter (2015). Volume 6, Part 2., page 27). The original implementation validates TLS matrices by interpreting them in terms of parameters of elemental motions. This approach is efficient at catching cases when refined elements of T, L and S matrices cannot describe a motion in turn indicating that these matrices are invalid. The TLS formalism assumes that atomic motions being modeled are harmonic. However, this assumption is not enforced neither in refinement protocols nor in decomposition of TLS matrices into parameters of translation-vibration motion. This makes it possible that refined TLS parameters do describe a motion but the kind of motion they describe is beyond the scope of TLS theory. The proposed way to verify that TLS parameters describe a harmonic motion consists of following steps. First, individual atomic displacement (ADP) parameters are calculated from TLS matrices using the known formula $U_{TLS} = T + ALA^t + AS + S^tA$. Then refined TLS matrices are used to extract parameters of concerted motions that they describe. These parameters include amplitudes of translations and vibrations, positions and orientations of corresponding axes, etc. In turn, these parameters are used to generate an ensemble of explicit atomic models with each model from the ensemble being consistent with TLS matrices. Having the ensemble of models allows converting positional uncertainty of each atom into corresponding displacement parameter, $U_{ensemble}$. Since $U_{ensemble}$ and U_{TLS} originate from the same source they are expected to match. Their mismatch indicates that motions that corresponding TLS matrices parameterize are anharmonic and, consequently, TLS matrices themselves are invalid. This analysis is now part of *phenix.tls_analysis* tool; details will be described elsewhere.