D3R Structure Refinement Protocol (April 2017)

Typically, structures contributed to the D3R project are only partially refined, particularly for parts of the protein far from the bound ligand. Full refinement with state-of-the-art methods/standards and validation is completed before release for D3R challenges and deposition in the PDB, in order to provide accurate models validated for the challenges and PDB users. To produce a high quality refined structure, a semi-automatic procedure is used for data preparation, parameter optimization, structure refinement, and protein/ligand validation (Figure 1).

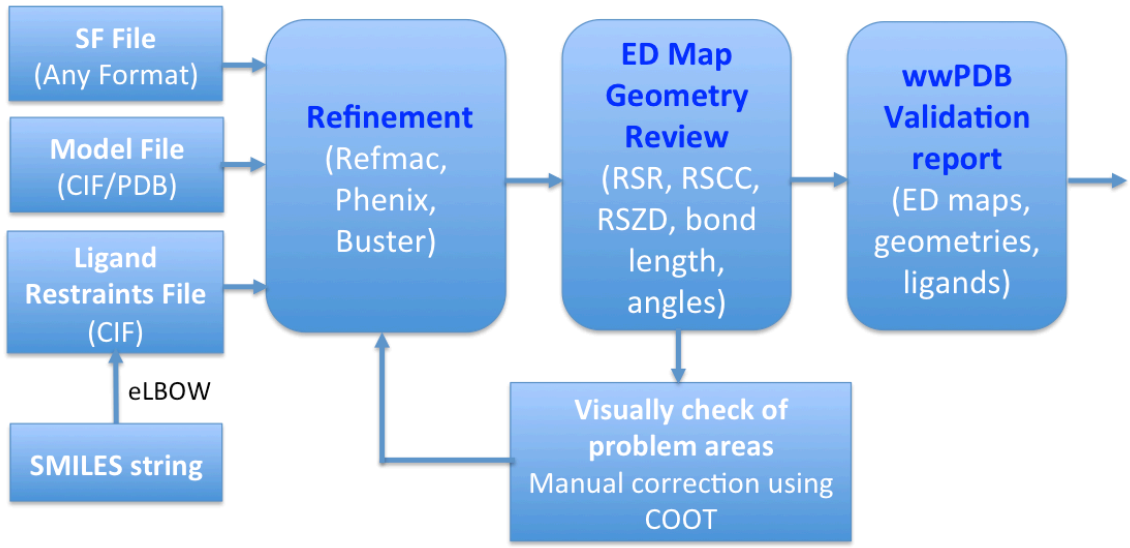

Figure 1. D3R Structure Refinement Workflow. Only residues flagged in the validation report are visually checked.

a) Data preparation

sf_convert is used to extract required diffraction data and convert it to the appropriate refinement program file formats.

eLBOW (version phenix 1.11.1-2575) with the AM1 optimizer (Moriarty et al., 2009) and Mogul (Bruno et al., 2004) reference values from the CSD database was used to generate restraints for co-crystal structure refinement based on contributor-provided SMILES strings, and the results were examined to ensure options “-opt” (use AM1 optimizer) and “-mogul” (use Mogul reference) interoperate appropriately.

b) Structure refinement

A python script named “D3R_refine” was used to automate the entire refinement process, utilizing three different refinement programs including PHENIX (Adams et al., 2010), REFMAC (Murshudov et al., 1997), and BUSTER (Bricogne et al., 2016), with

facilitating CCP4 package utilities (Winn et al., 2011). Results from different programs are assessed and compared to obtain the model of the best quality. Individual anisotropic B-factor refinement is used when the ratio of reflections/atom>30. Individual isotropic B-factor refinement is used when the ratio of reflections/atom<13. Between 30 and 13 reflections/atom, both isotropic and anisotropic B-factor refinement approaches are tested, using automated weighting, and the final B-factor model is selected on the basis of the lowest free R-factor (Rfree) (Brünger, 1992).

Non-Crystallographic Symmetry (NCS) refinement is not used for structures with resolution better than 1.7 Å, and only local NCS restraints are used during refinement of lower resolution structures (>2.0 Å). For resolution limits in the range of 1.7-2.0 Å, refinements with and without NCS are evaluated, and the best model is selected on the basis of refinement outcome statistics. Because NCS can affect the value of Rfree, we use multiple statistics to judge a model quality; e.g.,

	:	original	re-refine
R_value_R_free	:	0.27033	0.2293
R_value_R_work	:	0.19915	0.1607
Rfree-Rwork	:	0.0712	0.0686
dpi_free_R	:	0.3929	0.3472
correlation_overall	:	0.9466	0.9622
real_space_R_overall	:	0.1751	0.1606
Ramachandran_outlier_percent	:	0.96	0.08
Ramachandran_favored_percent	:	92.26	96.49
rotamer_outliers_percent	:	5.58	2.36
cbeta_deviations	:	3	0
all_atom_clashscore	:	13.92	3.71
overall_score	:	2.70	1.67
bond_overall_rms	:	0.0125	0.0077
angle_overall_rms	:	1.4016	0.9970
dihedral_overall_rms	:	17.9772	11.1694
chirality_overall_rms	:	0.0937	0.0583
planarity_overall_rms	:	0.0041	0.0062
non-bonded_rms	:	2.1035	2.1008
bond_ligand_rms	:	0.0486	0.0060
angle_ligand_rms	:	2.0396	1.4432
RSCC_ligand	:	0.938	0.957

Translation-Libration-Screw (TLS) corrections (Painter and Merritt, 2005) are used to improve outcomes in the final stages of refinement. Typically, 15 cycles of TLS refinement are computed. TLS rigid groups were selected for each chain when using REFMAC for refinement. Auto TLS options were invoked for refinement by PHENIX and BUSTER.

Weight matrices were, in most cases, automatically calculated by each of the three refinement programs. Rarely, manual adjustments of the weight matrix were required to obtain better protein geometry.

Although H atoms make little contribution to protein crystal diffraction data, they are included during refinement to help improve the Clashscore quality metric.

Following initial refinement of the protein structure, the best-refined protein model from the three parallel refinements (i.e., lowest R-free values) is selected. Electron density map and polypeptide chain geometry validations are performed as described in section **c,** below. Following any necessary manual corrections to that atomic model, bound ligand structures are built and refined to convergence. Ligand electron density map features and geometry are validated as described in section **d**. In some cases, ligand geometry restraints files are modified to ensure appropriate ligand stereochemistry. In the final round of refinement, water molecules are automatically added to the atomic model, using default PHENIX criteria. Water molecules found in the vicinity of bound ligands are manually checked.

c) Protein structure validation

wwPDB Validation Reports (Young et al. 2017) are generated for each refined structure and reviewed, with particular attention to the value of Rfree, non-bonded atom-atom clashes assessed by a scaled Clashscore (Chen et al., 2010), % Ramachandran Outliers (Ramachandran et al., 1963), % Sidechain Rotamer Outliers (Chen et al., 2010), and % Real Space R-factor Z-Score (RSRZ) Outliers (Kleywegt et al., 2004). Extreme outliers are examined and errors corrected as required.

The Real Space R factor (RSR) (Jones et al., 1991), Real Space Correlation Coefficient (RSCC) (Brändén and Jones, 1990), and Real Space Zscore of Difference map (RSZD) (Tickle, 2012) are calculated using the DCC program for main chain and sidechains at the individual residue level. Difference electron density maps are examined and any problematic residues rebuilt using COOT (Emsley et al., 2010).

d) Ligand validation

Mogul (Bruno et al., 2004) is used to validate bound ligand geometric features, including bond lengths, bond angles, planarity, and chirality using small molecule X-ray crystal structures in the Cambridge Structural Database (Groom et al., 2016). Geometric restraints are modified and the ligand structures re-refined until satisfactory outcomes were obtained, based on visual inspection.

Electron density map features corresponding to bound ligands are evaluated using RSR, RSCC and RSZD (RSZD+ and RSZD-), Occupancy Weighted Average B factor (OWAB) quality metrics, and errors corrected using COOT.

For higher resolution structures, any alternate conformations are detected by visual inspection of electron density maps and are then included in structure refinements.

Finally, intermolecular interactions between protein and bound ligands are visually inspected for plausibility, with particular focus on pseudo-symmetric ligands that could

be modeled in opposite orientations without penalty in terms of electron density map quality measures.

References

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W.*, et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D *66*, 213-221.

Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., Roversi, P., Sharff, A., Smart, O.S., Vonrhein, C.*, et al.* (2016). BUSTER version 2.10.3. (Cambridge, United Kingdom, Global Phasing Ltd.).

Brünger, A., Krukowski, A., and Erickson, J. (1990). Slow-cooling protocols for crystallographic refinement by simulated annealing. Acta Crystallogr A *46*, 585-593.

Brünger, A.T. (1992). Free R -value - a novel statistical quantity for assessing the accuracy of crystal structures. Nature *355*, 472-474.

Bruno, I.J., Cole, J.C., Kessler, M., Luo, J., Motherwell, W.D., Purkis, L.H., Smith, B.R., Taylor, R., Cooper, R.I., Harris, S.E.*, et al.* (2004). Retrieval of crystallographically-derived molecular geometry information. J Chem Inf Comput Sci *44*, 2133-2144.

Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D *66*, 12-21.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr D *66*, 486-501.

Groom, C.R., Bruno, I.J., Lightfoot, M.P., and Ward, S.C. (2016). The Cambridge Structural Database. Acta Crystallogr B *72*, 171-179.

Jones, T.A., Zou, J.-Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr A *47*, 110-119.

Kleywegt, G.J., Harris, M.R., Zou, J.Y., Taylor, T.C., Wahlby, A., and Jones, T.A. (2004). The Uppsala Electron-Density Server. Acta Crystallogr D Biol Crystallogr *60*, 2240-2249.

Moriarty, N.W., Grosse-Kunstleve, R.W., and Adams, P.D. (2009). electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. Acta Crystallogr D Biol Crystallogr *65*, 1074-1080.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D *53*, 240-255.

Painter, J., and Merritt, E.A. (2005). A molecular viewer for the analysis of TLS rigid-body motion in macromolecules. Acta Crystallogr D *61*, 465-471.

Ramachandran, G.N., Ramakrishnan, C., and Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. J. Mol. Biol. *7*, 95-99.
Tickle, I.J. (2012). Statistical quality indicators for electron-density maps. Acta Crystallogr D *68*, 454-467.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A.*, et al.* (2011). Overview of the CCP4 suite and current developments. Acta Crystallogr D *67*, 235-242.

Yang, H., Peisach, E., Westbrook, J.D., Young, J., Berman, H.M., and Burley, S.K. (2016). DCC: a Swiss army knife for structure factor analysis and validation. J Appl Cryst *49*, 1081-1084.

Young, J.Y., Westbrook, J.D., Feng, Z., Sala, R., Peisach, E., Oldfield, T.J., Sen, S., Gutmanas, A., Armstrong, D.R., Berrisford, J.M.*, et al.* (2017). OneDep: Unified wwPDB System for Deposition, Biocuration, and Validation of Macromolecular Structures in the PDB Archive. Structure *25*, 536-545.