A global Ramachandran score identifies protein structures with unlikely stereochemistry

O. V. Sobolev1, P. V. Afonine1, N. W. Moriarty1, M. L. Hekkelman2,3, R. P. Joosten2,3, A. Perrakis2,3, P. D. Adams1,4

Molecular Biosciences and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA, Division of Biochemistry, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, the Netherlands, Oncode Institute, Amsterdam, the Netherlands, Department of Bioengineering, University of California, Berkeley, CA 94720, USA osobolev@lbl.gov

Keywords: Ramachandran plot, validation, Rama-Z.

Ramachandran plots report the distribution of the (φ, ψ) torsion angles of the protein backbone and are one of the best quality metrics of experimental structure models. Typically, validation software reports the number of residues belonging to “outlier”, “allowed” and “favored” regions. While “zero outliers” is considered the “gold standard”, it can be misleading if deviations from expected distributions, even within the favored region, are not considered.

At low resolution it is often necessary to use additional information such as internal molecular symmetry, homologous structure models, information about secondary structure and rotameric states of protein amino-acid side chains. Clearly, the well-defined distribution of protein main-chain φ and ψ angles in Ramachandran space is yet another source of information that can guide model building and refinement. Ramachandran restraints can prevent deterioration of backbone conformation during low-resolution refinement, thereby helping to obtain chemically meaningful models. Many software packages provide an option to use Ramachandran restraints.

While helpful for refinement, actively using the Ramachandran plot as a source of restraints reduces its merit as an independent validation metric. The refined model may then appear to have desirable Ramachandran statistics in terms of expected fractions of residues belonging to favored/allowed/outlier regions, while the distribution of (φ, ψ) itself is improbable: this may not be obvious to an untrained eye.

We therefore turned our attention to the Ramachandran Z-score (Rama-Z), a quality metric introduced three decades ago, but underutilized [1].

We re-implemented Rama-Z in CCTBX using modern high quality models from top8000 database used to derive Ramachandran plot contours in MolProbity. Increased number of high quality models allowed us to use smaller grid size. We advocate that Rama-Z is well suited to assess backbone geometry and highlight unusual distributions of (φ, ψ) angles on Ramachandran plot that otherwise can be unnoticed. The metric does not depend on experimental data (X-ray or Cryo-EM) used to derive atomic model.

One of the limitations of the Rama-Z score is that it is not very suited for small structures with few residues. This is mostly caused by the nature of the metric that relies on normalization against a control set of structure models. Normalization is not well suited to small sample sizes, i.e. few available residues. To estimate the reliability of the Rama-Z score for a particular model we use the Jackknife method to estimate RMSD.

The method is implemented and available in open-source CCTBX (mmtbx.rama_z) library as well as in Phenix as command line tool phenix.rama_z and also in various validation reports generated by Phenix. The tortoise implementation is available in PDB-REDO and will become available in the CCP4 and CCP-EM suites in the near future.

We suggest new cutoffs for Rama-Z validation being |Rama-Z| > 3 is for improbable backbone geometry, 2 < |Rama-Z| < 3 for unlikely yet possible, |Rama-Z| < 2 is for normal backbone geometry. Particular Rama-Z scores should be reported along with its reliability.

We argue for a greater acceptance of this metric by the community. PDB-REDO has been reporting Rama-Z score since its inception. More discussion of this metric is published recently [2].


This work was supported by the NIH (Project P01 GM063210), the Phenix Industrial Consortium and by the Netherlands Organization for Scientific Research (NWO; Vidi grant 723.013.003).