Poster Presentations

[MS04-P11] Improved estimates of coordinate error for molecular replacement Robert D Oeffner[a], Gábor Bunkóczi[a], Airlie J McCoy[a] and Randy J Read[a]

[a]*Cambridge Institute for Medical Research, University of Cambridge, Hills Road, Cambridge, CB2 0XY, United Kingdom.*
E-mail: rdo20@cam.ac.uk

The estimate of the root-mean-square deviation (RMSD) in coordinates between the model and the target is an essential parameter for calibrating likelihood functions for molecular replacement (MR) [1]. Good estimates of the RMSD lead to good estimates of the variance term, , in the likelihood functions [2], which increases signal to noise and hence success rates in the MR search. Phaser [3] has hitherto used the Chothia & Lesk estimate of the RMSD, [4], which depends only on the sequence identity between the model and target and moreover was not optimised for the MR likelihood functions. Variance refinement functionality was added to Phaser to enable determination of the VRMS, the RMSD that optimises the log-likelihood gain (LLG) for a correct MR solution. Variance refinement was subsequently performed on a database of over 21000 MR problems, each with one component in the asymmetric unit, sampling a range of sequence identities, protein sizes and protein fold classes. Success of the MR calculations was monitored with the translation function Z-score (TFZ), where a peak TFZ of 8 and over was found to be a reliable indicator that MR had succeeded. Good estimates of the RMSD are correlated with the sequence identity and the protein size. A VRMS value differs from traditional RMSD values by not being biased by an explicit atom pair assignment and also in not being dominated by outliers. A new estimate, the eVRMS, that uses these two parameters in a function optimized to fit the mean of the VRMS values is implemented in Phaser and improves MR outcomes. The distribution of VRMS values has an approximately Gaussian deviation from the eVRMS, with the size of the error being proportional to the eVRMS. The eVRMS estimates have a small but measurable dependence on SCOP fold class [5]. Perturbing the eVRMS from the mean of the VRMS values in steps of standard deviations further increases MR success rates. Phaser documentation is available at http://www.phaser.cimr.cam.ac.uk.

[1] M. G. Rossmann and D. M. Blow, *Acta Cryst*., vol. 15, pp. 24-31, 1962. [2] R. J. Read, "Pushing the boundaries of molecular replacement with maximum likelihood," *Acta Cryst* D, pp. 57, 1373-1382, 2001. [3] A. J. McCoy, R. W. G. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read, *Acta Cryst* D, vol. 40, pp. 658-674, 2007. [4] C. &. L. A. M. Chothia, "The relation between the divergence of sequence and structure in proteins," The EMBO Journal, vol. 5, no. 4, pp. 823-826, 1986. [5] A. G. Murzin, S. E. Brenner, T. Hubbard and C. Chothia, "SCOP: a structural classification of proteins database for the investigation of sequences and structures.," J. Mol. Biol., vol. 247, pp. 536-540, 1995.

Keywords: Phaser, RMSD, variance refinement