

Evaluation of phasing models used for molecular replacement structure determination

Kaushik Hatti<sup>1</sup>, N Srinivasan<sup>1</sup>, MRN Murthy<sup>1</sup>

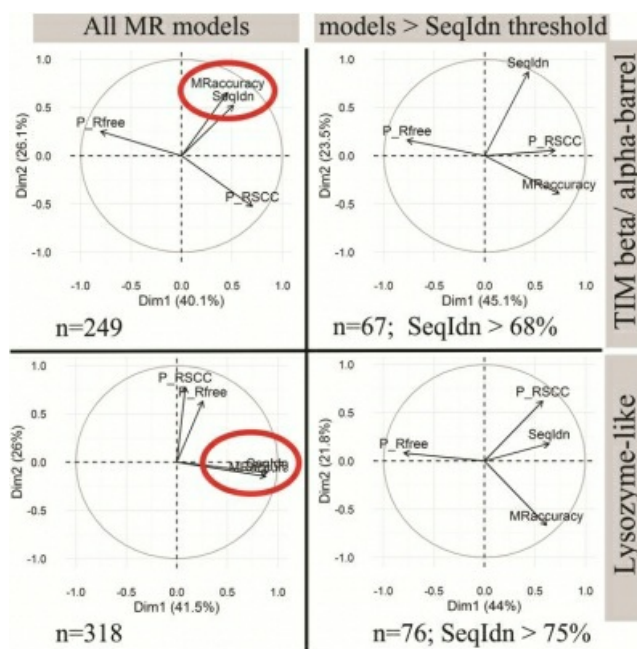
<sup>1</sup>Molecular Biophysics Unit, Indian Institute Of Science, Bangalore, India

E-mail: hskaushik@gmail.com

Nearly 90% of the structures deposited in the Protein Data Bank (PDB) have been determined by X-ray crystallographic methods. Of these structures, 78% of the entries deposited in the last five years have been determined using the Molecular Replacement (MR) technique. MR has several advantages over other crystallographic techniques as it is based on the phase information that could be obtained from the structure of a related protein (phasing model) as a valid approximation to the unknown structure and hence, eliminates the need for more diffraction data to carry out experimental phasing. Since 2010, PDB has grown tremendously by over 50,000 depositions while the growth in the number of unique folds is negligible in comparison [1]. For a given target protein, it is likely that more than one structure is available in the PDB that could be used as a suitable phasing model. In such cases, selecting the best phasing model from among the available pool of structures requires careful examination of the parameters that determine the reliability of the phasing model for MR structure determination. Hence, it is necessary to understand the relationship between properties of phasing model and quality of the structure determined (MR model) to arrive at most reasonable MR model. In this study, we provide quantitative measures of intuitive ideas on the strategies that might be useful in choosing the best phasing model. Towards this goal, re-determination of selected structures from the PDB was carried out using the X-ray intensity data of the selected protein deposited in the PDB and several homologous structures as phasing models. A total of 716 phasing models were considered for MR structure determination of three proteins from Triosephosphate isomerase fold and Lysozyme-like fold. The RMSD of corresponding Ca positions between MR model and the structure of the selected protein re-determined by an identical MR protocol using the deposited coordinates of the protein (positive control) was calculated. A 'Q score' based on the polypeptide length normalized RMSD was considered as a measure of the accuracy of the MR model (MR accuracy). Resolution of the target protein was found to be the most important factor for the success of MR. The success of MR increased with the increase in sequence identity between target protein and phasing model. However, CART modeling [2] indicated that after a defined sequence identity threshold, quality of phasing model (measured by Resolution, Real-space correlation coefficient, Rwork/Rfree) seem to have a greater influence on the MR accuracy than sequence identity. Correlation of phasing model properties and MR accuracy scores obtained by using phasing models with sequence identity above the threshold also supports this observation in both the folds studied. Further, a similar trend is observed in proteins from other folds as well. However, the sequence identity threshold above which the quality of the phasing model assumes importance varies for different folds.

[1] Hatti, K., et. al., (2016). Acta Crystallogr. Sect. D. 72, 1081–1089.

[2] Breiman, L., et. al., (1984). Classification and Regression Trees.



Keywords: [Molecular Replacement](#), [phasing](#), [fold](#)