[MS4-05] A Combined Method for Automatic Structure Solution of Proteins of Unknown Fold.

Pavol Skubak, Navraj S. Pannu.

Department of Biostructural Chemistry, Leiden University, Leiden (the Netherlands).
E-mail: p.skubak@chem.leidenuniv.nl

Traditionally, the process of solving macromolecular crystal structure of unknown fold from X-ray data consists of separate step-wise experimental phasing, density modification and model building with refinement (Fig. 1a).
Here we present a novel "combined" multivariate likelihood target function that directly considers phase information from via the experimentally collected X-ray data and simultaneously combines it with the information from density modification and model building into a single unified process (Fig. 1b). The unified process consists of iterative minimization of the minus log-likelihood of the new combined likelihood target in reciprocal space, followed by current density modification and model building procedures in crystal space. Thus, the structure solution process no longer relies on successive step-wise approximations of the experimental data.
We tested the performance and robustness of the new method for combined structure solution (Fig. 1b) against the current step-wise approach (Fig. 1a) on 147 SAD data sets spanning a wide range of resolutions from 0.94 to 3.88 Angstroms and anomalous scatterers including selenium, sulfur, chloride, iodide, bromide, calcium and zinc. We use the program REFMAC5 [1] for both the new combined function as well as the step-specific substructure phasing, phase combination and model refinement multivariate functions. The automated structure solution package CRANK [2] is used to link these programs for both approaches.
Fig. 1. a) The traditional process of structure solution. The steps of experimental phasing, density modification and refinement are performed separately, using a different minimization function in each step.
For all data sets, the average fraction of model correctly built increases from 60% to 74%. If we exclude the data sets built to at least 85% completeness by the stepwise method and data sets where the heavy atom substructure could not be found, 45 data sets remain with 28% of the model correctly built on average by the stepwise approach and 77% by the combined algorithm. These differences clearly demonstrate the synergistic effect of simultaneously combining the steps and that the current limits of X-ray crystallography can be significantly extended.
The performance of the new method at low resolution can be demonstrated on the 12-Subunit RNA Polymerase II (3950 residues in the asymmetric unit) singlewavelength anomalous diffraction (SAD) data set diffracting anisotropically to 3.88 Angstroms [3] that could not be utilized to build the structure neither automatically nor manually by the original authors. The combined method results in automatic building of a majority of the protein backbone solely from the anomalous signal of eight intrinsic zinc atoms and the single SAD dataset. The quality of the automatically built structure is evident from the low R-free value of 37.6%.

[1] Murshudov, G.N., Skubak P., Lebedev A.A., Pannu N.S., Steiner R.A., Nicholls R.A., Winn M.D., Long F. and Vagin A.A. (2011). *Acta Cryst.* **D67**, 355-367.
[2] Pannu, N.S., Waterreus W.J., Skubak P., Sikharulidze I., Abrahams J.P. and de Graaff RAG. (2011). *Acta Cryst.* **D67**, 331-337.
[3] Martinez-Rucobo, F.W., Sainsbury, S., Cheung, A.C.M., Cramer (2011). P. EMBO J. **30**, 1302-1310.