MS4-O2 New bulk-solvent models improves model-to-data fit and facilitates map interpretation

Pavel V. Afonine[1], Paul D. Adams[1]

1. Lawrence Berkeley National Laboratory, Berkeley CA, USA

email: pafonine@lbl.gov

Bio-macromolecular crystals contain between 10 and 90% of solvent. This solvent is mostly disordered so it cannot be interpreted in terms of an atomic model. Owing to its simplicity and yet relatively good modeling power, flat bulk-solvent model is the most commonly used model to account for disordered solvent in modern crystallographic software packages such as CNS, CCP4 or Phenix. This model assumes electron density is constant anywhere in the unit cell where there is no atomic model placed. While this may be a reasonable approximation for some crystal structures or at initial stages of structure determination, it may be less accurate at final stages. Major deviations from the assumption of a flat model include: 1) local concentration of solvent component of specific types, such as lipid belts in membrane proteins, 2) unmodeled ligands, 3) partial occupancy of solvent in small isolated regions (between or inside macromolecules), and 4) lack of solvent in certain regions (hydrophobic cores).

These deviations manifest themselves as elevated R factors in the lowest resolution shells as well as residual features in difference maps: positive in cases when the flat solvent model is inadequate in accounting for distinct features, or negative when solvent model is used in regions with no solvent.

To overcome the limitations of the existing bulk solvent model we have proposed a non-uniform bulk-solvent model that allows for solvent variation across the unit cell volume. The new model splits initially binary (0/1) solvent masks into several masks by applying connectivity analysis. These masks are then split further into more masks based on analysis of difference maps. This final set of solvent masks is used to compute the individual bulk solvent contributions to the total model structure factor.

Tests on all deposited structures in PDB that have diffraction data and cross-validation flags available indicate systematic improvement of model-to-data fit with no signs of over-fitting as judged by the R_{free} factor. Tests on selected models demonstrate notable improvements in map quality especially for weak features such as ligands or solvent molecules.

All described tools will be available in *Phenix*.

Keywords: Refinement, bulk-solvent, maps

MS4-O3 Fragon - rapid fragment-based molecular replacement

Huw T. Jenkins[1], Alfred A. Antson[1]

1. York Structural Biology Laboratory, Department of Chemistry, University of York, York, YO10 5DD

email: huw.jenkins@york.ac.uk

Calculating phases from small geometrically-ideal protein fragments presents an attractive method for structure determination when no suitable template structure for molecular replacement is available. Over 15 years ago ACORN could solve the structure of lysozyme by locating a 10-residue poly-alanine α-helix (3.4% of the contents of the unit cell) and improving phases from this fragment through density modification [1]. More recently several approaches have been developed which use Phaser [2] to place either ideal α-helices [3] or fragments from ab initio modeling [4,5] followed by multiple iterations of density modification and auto-building for each fragment to identify correctly placed fragments. The main drawback of this method is obvious – the vast majority of CPU time is spent attempting model building with phases from incorrectly placed fragments.

We have implemented a novel approach – Fragon – that also uses Phaser to place fragments (ideal α-helices, β-strands or even single atoms) but tests solutions by artificial extension of the data to 1.0 Å resolution followed by density modification with ACORN. In many cases this can rapidly identify correctly placed fragments. Subsequent chain tracing with SHELXE into the density-modified maps provides a robust measure of success and enables comparison with previous approaches [3-6].

We generated a challenging test set (103 structures) of mixed α/β folds at resolutions between 1.0 and 1.7 Å. Fragon solved 62% of the test cases searching for one or two ideal α-helices of 7-14 residues. One further structure was solved from a single S atom to bring the final success rate to 63%. Fragon was also benchmarked against the test set used in previous approaches [4-6]. Fragon's success rate of 61% compares favourably with that of previous approaches of 54% [5] and 49% [6].

We will also describe application of Fragon to solve a previously unknown structure starting from ideal β-strands and a 54.5 kDa structure solved from a single P atom.

1. Foadi *et al.* (2000) Acta Cryst. D56:1137-47
2. McCoy *et al.* (2007) J. Appl. Crystallogr. 40:658-74
3. Rodríguez *et al.* (2009) Nature Methods 6:651-3
4. Bibby *et al.* (2012) Acta Cryst. D68:1622-31
5. Keegan *et al.* (2015). Acta Cryst. D71:338-343
6. Sammito *et al.* (2015) Acta Cryst. D71:1921-30

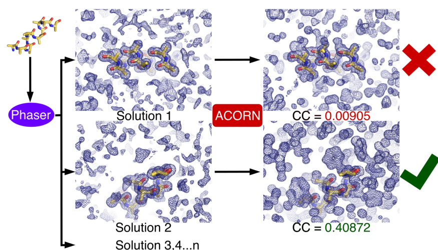

Figure 1. The Fragon process. The correct solution can be selected based on the correlation coefficient (CC) after density modification.

Keywords: Molecular replacement, Phasing, Fragments, Fragon

MS4-O4 BORGES_MATRIX: a tool to generate models for ab initio phasing and for structure interpretation.

Massimo D. Sammito[1,2], Claudia L. Millán-Nebot[2], Rafael J. Borges[2,3], George M. Sheldrick[1], Isabel Usón[2,4]

1. Georg August University of Göttingen, Dept. of Structural Chemistry, 37077 Göttingen, Germany
2. Crystallographic Methods, Institute of Molecular Biology of Barcelona (IBMB), Spanish National Research Council (CSIC), 08028, Barcelona, Spain
3. Dept. Physics and Biophysics, Biosciences Institute (IBB), São Paulo State University (UNESP), 18618-970, Botucatu, São Paulo, Brazil
4. Structural Biology, Catalan Institution for Research and Advanced Studies (ICREA), 08028, Barcelona, Spain

email: massimo.sammito@ibmb.csic.es

ARCIMBOLDO_LITE [1] is a pipeline that combines the search of small fragments, like alpha helices, with PHASER [2], and density modification and autotracing with SHELXE [3]. Even though the model constitutes only a small percentage of the total scattering, the method has proven to be successful for high resolution cases (better than 2.1 Å). In order to correctly locate and extend the input search model it is required that its main chain matches very accurately the one of the final structure. This assumption is generally correct for helices but it does not hold true for composite secondary structure elements. Thus, exploiting the idea that there are common building blocks (such as three beta strands in a sheet, or two small parallel helices) common to unrelated protein structures, we developed BORGES [4], a tool to extract and use libraries of small local folds for phasing. Other bioinformatic tools are also available to search for similar structural occurrences of a fold [5,6], but they tend to retrieve continuous domains, and give a relatively general view of the fold. Our new program BORGES_MATRIX implements a detailed description, based on discrete distributions of characteristic vectors to entail the local conformation of the main chain and to geometrically compare extracted models with a search template. Our method also extracts folds formed through crystallographic and non crystallographic symmetry, and does not require sequence information to retrieve similar occurrences. Recently, a library of three antiparallel strands was used to solve the structure of a viral all beta structure of 130 aa diffracting to 1.55 Å presenting a novel fold [7]. Beyond phasing, the program contributed to the understanding of the structural environment of the binding site by extracting and comparing similar occurrences of the local geometrical conformation.

[1] Millán, C. *et al.* (2015). IUCrJ 2, 95-105.

[2] McCoy, A. J. *et al.* (2007) J. Appl. Cryst. 40, 658-674

[3] Sheldrick, G. M. (2010) Acta Cryst. D66, 479-485.

[4] Sammito, M. *et al.* (2013) Nature Methods 10, 1099-101.

[5] Holm L, *et al.* (2010) Nucl. Acids Res. 38, W545-549.

[6] E. Krissinel *et al.* (2004). Acta Cryst. D60, 2256-268

[7] Fedosyuk *et al.* (2016). PNAS submitted

Keywords: ab initio phasing, fragment libraries, structural bioinformatics