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_\text{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
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. Sammito1,2, Claudia L. Millán-Nebot2, Rafael J. Borges2,3, George M. Sheldrick1, Isabel Usón2,4

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 autotraging 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.


Keywords: ab initio phasing, fragment libraries, structural bioinformatics