

Oral Contributions

[MS44-04] BORGES-ARCIMBOLDO exploiting tertiary folds as fragments libraries for phasing

Massimo Sammito¹, Claudia Millán¹, Iñaki M de Ilarduya¹, Ivan De Marino¹, Kathrin Meindl¹, Isabel Usón^{1,2}

¹Instituto de Biología Molecular de Barcelona, IBMB-CSIC, Barcelona, Spain

²Institució Catalana de Recerca I Estudis Avançats, Barcelona, Spain

Corresponding author: msacri@ibmb.csic.es

Protein structures may share the geometry of local folds in cases of distant homology or even if totally unrelated. BORGES is a software that builds customized libraries of composite secondary structure fragments extracting them from all possible real structures deposited in the PDB. Description of the folds is independent of the primary sequence, but relies on a novel algorithm based on distribution of characteristic vectors computed over the mainchain coordinates atoms. Filtering out models with no significant differences expected in terms of phasing reduces the size of the library without loss of generality. BORGES libraries are used within a novel procedure combining rotation clustering, rotation refinement, and model refinement (BORGES) with the model location (PHASER[1]), density modification and autotracing (SHELXE[4]), extending the method underlying the previous ARCIMBOLDO[2,3,5], which exploited enforcing secondary structure to a new method enforcing local unspecific tertiary structure to constrain phasing. BORGES exploits cloud computing technologies (in particular middleware CONDOR[6] and SGE[7]) to parallelize jobs, building up a dynamic multiresolution frame in which decisions are taken in running time to lead towards the final structure of the protein. Usability is considered providing the user an easy graphical input/output interface, this also includes the possibility to remote access to external supercomputing center for cloud

computing. With this new approach three novel structures were so far solved, where both the classical ARCIMBOLDO and other software had failed, and many test cases of both alpha and beta structures.

[1] Airlie J. McCoy, Ralf W. Grosse-Kunstleve, Paul D. Adams, Martyn D. Winn, Laurent C. Storoni, and Randy J. Read. Phaser crystallographic software. *J Appl Crystallogr.* 2007 August; 40(Pt 4): 658–674.

[2] Dayte D. Rodríguez Martínez, Christian Grosse, Sebastian Himmel, César González, Iñaki M. de Ilarduya, Stefan Becker, George M. Sheldrick and Isabel Usón. ARCIMBOLDO: Crystallographic Ab Initio protein solution far below atomic resolution *Nature Methods* 6(9): 651-653 (2009)

[3] Dayté D. Rodríguez Martínez, Massimo Sammito, Kathrin Meindl, Iñaki Martínez de Ilarduya, Marianus Potratz, George M. Sheldrick, Isabel Usón. Practical structure solution with ARCIMBOLDO. *Acta Crystallogr D Biol Crystallogr.* 2012 Apr; 68(Pt 4): 336-43.

[4] George M. Sheldrick. Macromolecular phasing with SHELXE. *Z. Kristallogr.* 2002, 217, 644-650

[5] Isabel Usón, Claudia Millán, Massimo Sammito, Kathrin Meindl, Iñaki M. de Ilarduya, Ivan De Marino, Dayté D. Rodríguez. Phasing Through Location of Small Fragments and Density Modification with ARCIMBOLDO, In: *Advancing Methods for Biomolecular Crystallography*, R. Read, A. G. Urzhumtsev, V. Y. Lunin (Eds.), Springer, Dordrecht, The Netherlands, 2013, pp. 123-132.

[6] Tannenbaum, T., "Condor - A Distributed Job Scheduler", in Thomas Sterling, editor, *Beowulf Cluster Computing with Linux*, The MIT Press, 2002. ISBN: 0-262-69274-0 (2002).

[7] <https://arc.liv.ac.uk/trac/SGE>

Keywords: Model Libraries, Phasing methods, Bioinformatics