MS45-O2 **A cluster and truncate approach to processing *ab initio* models for Molecular Replacement.** Daniel Rigden,[a] Jaclyn Bibby,[a] Olga Mayans,[a] Ronan Keegan, [b] Martyn Winn,[c] *[a]University of Liverpool, United Kingdom, [b]STFC Daresbury Laboratory, United Kingdom, [c]STFC Rutherford Appleton Laoratory, United Kingdom.*
E-mail: drigden@liv.ac.uk

Molecular Replacement (MR) uses a known search model to solve the unknown crystal structure of a related protein, but is dependent on the availability of a model having sufficient structural similarity. *Ab initio* modelling has developed to the extent that its results can sometimes be used to successfully phase diffraction data. Thus, *ab initio* models can be tried as search models where structural homologues are not available and experimental phasing is difficult [1]. Our method employs *ab initio* polyalanine models (or 'decoys'), produced in large numbers then clustered based on the presence of similar core structures. The largest of these clusters is likely to be closest to the native structure [2]. Such *ab initio* modelling may result in an accurate prediction of the structural core of the target, but with inaccurate loops and termini. We have developed an automated pipeline for the processing of *ab initio* models for use in MR. We show that clustering of models into ensembles combined with targeted truncation can give a successful result where a single search model would fail. A pipeline has been tested on 296 proteins between 40-120 residues long, using Rosetta to produce 1000 decoys for each target. Ensembles were produced from these decoys, and MR carried out using MrBUMP. A success rate of 42% in this fully automated pipeline was achieved. á-helical proteins are particularly successful (79% success), reflecting the greater accuracy of their modelling *ab initio*, while all-â protein (5% success) remain problematic. Modelling of side chains onto the *ab initio* models is sometimes required for success. Successes are achieved at different levels of truncation from <30 to >90% of target sequence in the successful ensemble. Importantly, the likely success or failure of the modelling can be predicted based on characteristics of the protein such as length and secondary structure, and by the convergence of the modelling program to produce a large cluster of models with a similar core structure. Successes at the upper size limit of our test set suggest that larger proteins may sometimes be solved successfully. There are further promising avenues for application of our core methodology to MR cases in which NMR structures or only distant homologues are available. Unlike other computationally intensive methods [3], this method is suitable for modest hardware, allowing for broader adoption. An implementation of the pipeline, AMPLE, has been developed and a beta release version is included in the latest release of the CCP4 software suite.

[1] Das, R., Baker, D. (2009) *Acta Cryst*, **D65**, 169-175.
[2] Rigden, D. J., Keegan, R. M., Winn, M. D. (2008) Acta Cryst. **D64**, 1288-1291.
[3] Qian B., Raman S., Das R., Bradley P., McCoy A.J., Read R.J., Baker D. (2007) Nature, **450**, 259-264.

Keywords: molecular replacement; *ab initio* modeling; computational modeling methods

MS45-O3 **DNA – Protein Complex Structure Prediction ARCIMBOLDO structure-solution with DNA – Protein fragment subsets** Kevin Pröpper,[a] Kathrin Meindl,[b] Dayté Rodríguez,[b] Massimo Sammito,[b] Birger Dittrich,[a] George M. Sheldrick,[a] Ehmke Pohl,[c *] Isabel Usón,[b,d *] *[a]Lehrstuhl für Strukturchemie, Universität Göttingen, Tammannstrasse 4, 37077 Göttingen, Germany, [b]Instituto de Biología Molecular de Barcelona (IBMB–CSIC), Barcelona Science Park, Baldiri Reixach 15, 08028 Barcelona, Spain, [c]Department of Chemistry & School of Biological and Biomedical Sciences, Durham University, Durham DH1 3LE, UK, [d]Institució Catalana de Recerca i Estudis Avançats (ICREA), Spain.*
E-mail: kproepper@shelx.uni-ac.gwdg.de, uson@ibmb.csic.es

DNA – Protein interactions have a major role in all aspects of genetic activity within an organism, such as transcription, packaging, rearrangement, replication and reparation. We have witnessed an increasing determination of high quality DNA - binding proteins structures despite the fact that crystallographic methods for nucleic acids are considerably behind those available for proteins.[1] DNA – Protein interactions has become an important field in state of the art scientific research and structures emphasising insight into the principles of binding and base sequence recognition. Crystallographic structure solution of DNA – Protein complexes remains a challenging area and requires the combination of theoretical and practical methods. We present a study on a newly developed computational method focused on data from DNA – Protein binding motives.[2]. We have assessed the potential of the ab initio structure-solution program ARCIMBOLDO on protein data complexes, based on the combination of locating small subset DNA – Protein motif fragments with density modification with the program SHELXE in a multisolution frame.[3,4] For proteins, mainchain alpha helices provide the ideal, almost ubiquitous, small fragment to start searches. In the case of DNA complexes, the binding motives constitute a suitable search fragment. Our long-term aspiration is to gain a better understanding of sequence-specific DNA recognition, for an ultimately ability to interfere with this process in order to develop new and better antibacterial drugs.

[1] Luscombe, N. M., Augustin, S. E., Berman, H. M., Thornton, J. M., Genome Biology, (2000), 1(1).
[2] Lucarelli, D., Vasil, M. L., 2, Meyer-Klaucke, W., Pohl, E., Int. J. Mol. Sci. (2008), 9, 1548-1560.
[3] Rodríguez, D., Grosse, C., Himmel, González, C., Ilarduya I. M., Becker, S., Sheldrick, G. M., and Usón, I., Nature Methods (2009) 6(9), 651-653.
[4] Rodríguez, D., Sammito, M., Meindl, K., Ilarduya, I. M., Potratz, M., Sheldrick, G. M., Usón, I., Acta Cryst. (2012). D68, 336–343.

Keywords: ab-initio structure determination; DNA-protein complexes & DNA-binding proteins; Macromolecules