MS12-O3

Exploiting residue-residue contact information beyond structure predictions in Molecular Replacement

Felix Simkovic[1], Saulo de Oliveira[2], Charlotte Deane[2], Daniel J Rigden[1]

1. Department of Biochemistry, University of Liverpool, Liverpool, United Kingdom
2. Department of Statistics, University of Oxford, Oxford, United Kingdom

email: hlfsimko@liv.ac.uk

Interactions between amino acids are crucial for protein folding and interactions. Natural selection can therefore limit changes to conservative substitutions or simultaneous changes at two positions, leaving a covariance signal between them. Successful detection of these positions, through the analysis of evolutionary covariance, is now reliably possible [1]. Beyond the immediate impact this data has already made in the field of Structural Biology, many applications specifically in X-ray crystallography are yet to be tested or determined [2].

Ab initio protein structure prediction is typically done by assembling protein fragments selected using the target sequence. Here we describe the use of contact information to select protein fragments for direct use as search models. The selection of fragments is supported by matching the target-specific contact profile to a fragment, thus extracting fragments more likely to be similar to the target fold. We show how such fragments can be directly used as Molecular Replacement search models to successfully elucidate the structure of small globular protein targets. The findings are particularly relevant for targets for which conventional Molecular Replacement fails and idealised helices are not sufficient.

Another example of the use of contact information to aid X-ray crystallography is the applicability of residue-residue contacts to select accurate *ab initio* structure predictions. The use of such predictions to generate ensemble search models in the unconventional MR pipeline AMPLE has been described [3]. Here we consider how it may be possible to select and process single favoured structures as an alternative route to structure solution in difficult cases.

References:

[1] de Oliveira, S. & Deane, C. (2017). F1000Research, 6, 1224.

[2] Simkovic, F., Ovchinnikov, S., Baker D. & Rigden D.J. (2017). IUCrJ, 4, 291-300.

[3] Simkovic, F., Thomas J.M.H., Keegan R.M., Winn M.D., Mayans O. & Rigden D.J. (2016). IUCrJ, 3, 259-270.

Keywords: ab initio modelling, evolutionary covariation, molecular replacement

MS12-O4

ALEPH: a network-oriented approach to structure mapping and comparisons

Massimo Domenico Sammito[1], Ana Medina[2], Claudia Millán[2], Rafael J. Borges[3], Isabel Usón[4]

1. Department of Haematology, Cambridge Institute for Medical Research, University of Cambridge, Cambridge, United Kingdom
2. Crystallographic Methods, Institute of Molecular Biology of Barcelona (IBMB), Spanish National Research Council (CSIC), Barcelona, Spain
3. Dept. Physics and Biophysics, Biosciences Institute (IBB), São Paulo State University (UNESP), Botucatu, Brazil
4. Structural Biology, Catalan Institution for Research and Advanced Studies (ICREA), Institute of Molecular Biology of Barcelona (IBMB), Spanish National Research Council (CSIC), Barcelona, Spain

email: massimo.sammito@gmail.com

The folding of an amino acid sequence into stable secondary structure fragments represents one of the most energetically favorable and hence initial steps for the formation of tertiary structure in almost all protein families. Therefore, a description based only on these basic structural elements is a powerful abstraction that allows structure comparison and classification. In our group, we define as **L**ocal **F**olds[1](LF) any characterization of nonspecific, small, discontinuous composites of secondary structure elements. The retrieval of libraries of ubiquitous LF has been possible through specialised software. Such libraries have been used to phase (from crystallographic data) larger and even all-beta structures following an ab-initio approach[1]. In our group, LF characterization is performed through a new software called ALEPH; it is based only on 3d-geometrical descriptors named **C**haracteristic **V**ectors[1] whose relations are described with a network. Interpretation of this network results in a specific signature for every structure. Structure comparisons through signatures do not require sequence alignments and allow to abstract folds, domain, and even LF without performing superpositions or coordinate-based algorithms. This generalization can be used also for the generation of structural ensembles for Molecular Replacement (MR) or to automatically determine compact secondary structure subgroups to apply new Phaser Gyre/Gimble[2] functions for solution optimization as recently reported in ARCIMBOLDO_SHREDDER[3]. The nature of the problem requires data mining over more than 130,000 structures deposited in the Protein Data Bank and the design of specialized algorithms to build, inspect and compare CV networks.

References:

[1] Sammito, M., et al. (2013) Exploiting tertiary structure through local folds for crystallographic phasing. Nat Methods. 10(11), 1099-101.

[2] McCoy, A. et al., (2018) Gyre and gimble: a maximum-likelihood replacement for Patterson correlation refinement. Acta Crystallogr. D Biol. Crystallogr., 74(Pt 4), 279–289.

[3] Millán, C. & Sammito, M., et al. (2018) Exploiting distant homologues for phasing through the generation of compact fragments, local fold refinement and partial solution combination. Acta Crystallogr. D Biol. Crystallogr., 74(Pt 4), 290-304

Keywords: Structure comparison, Local Folds, Network analysis