Repeats) found in prokaryotic genomes confer small RNA-mediated protection against viruses and other invaders. The CRISPR loci are transcribed to precursor RNAs that are subsequently processed to the individual invader-targeting CRISPR RNA (crRNA). Distinct families of CRISPR-associated Cas proteins function to cleave within the repeat sequence of CRISPR transcripts by vastly different mechanisms. Cas6 represents a family of processing endonucleases that recognize and cleave nonstructured RNA. Crystal structures and complementary biochemical studies of Cas6 bound with a repeat RNA suggest a wrap-around model of processing. This model differs from two other known models of crRNA processing in which the endonuclease either depends on a structured RNA substrate or a guide RNA for processing. The discovery of the families of crRNA processing endonuclease has significantly expanded the repertoire of RNA processing endonucleases.

Keywords: CRISPR RNA, riboendonuclease, protein-RNA complex

MS.71.4

Acta Cryst. (2011) A**67,** C161

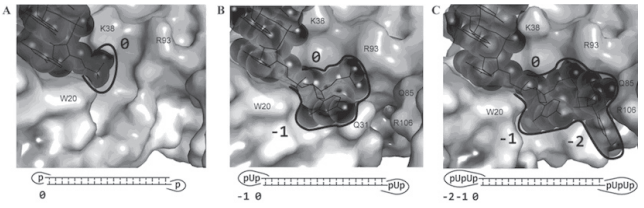
Procrustean bed of RNA silencing suppression

Jevgenia Tamjar,[a] Elizaveta Katorcha,[a] Alexander N. Popov,[b] Lucy Malinina,[a] [a]*Structural Biology Unit, CIC bioGUNE, Technology Park of Bizkaia, Derio, 48160, (Spain).* [b]*European Synchrotron Radiation Facility, Grenoble, 38043, (France).* E-mail: jtamjar@cicbiogune.es

The p19 protein from tombusvirus is a caliper-like machine that recognizes siRNA by length of 19 base pairs [1,2] and blocks RNA silencing through direct binding of siRNA molecules [3]. Here we present four newly solved crystal structures of p19 in complex with (i) 19 bp RNA pGG$(CAG)_5$CC, (ii) 19 bp RNA pUUG$(CUG)_5$CU with 1-nt 5'-overhang, (iii) 19 bp RNA pUUUG$(CUG)_5$CU with 2-nt 5'-overhang, and (iv) 20 bp RNA pG$(CUG)_6$C, that have been studied at 1.73 Å, 1.86 Å, 2.30 Å, and 2.00 Å resolution, respectively.

Comparative analysis of structures (i) to (iii) shows the presence of the 5'-end recognition centre in p19, which can bind 5'P (Fig. A), 1nt 5'-overhang (Fig B) and 2nt 5'-overhang (Fig C). The finding of the recognition center designated for 5'-overhangs of RNA is surprising, as siRNAs are known to contain 3'-overhanging segments rather than 5'-end ones. On the other hand, such a center could facilitate the unwinding of terminal base-pairs in 'longer-than-19bp' RNA fragments. Moreover analysis of 20 bp RNA (iv) revealed the ability of p19 to unwind terminal base pair, hence maintaining the double helical region length of 19 base pairs. Unwound 5'-end nucleotide is stabilized by several direct and water-mediated interactions, whereas unwound 3'end overhang remains invisible in electron density map, in accordance to previously reported data that showed unimportance of 3'overhangs for p19 binding with siRNA [1,2].

As siRNAs are known to be 21-25 nucleotides long the discovered center can be of help for understanding the mechanism of binding of small RNAs of longer length than 19 bp by RNA silencing suppressor p19.



[1] K. Ye, L. Malinina, D.J. Patel, *Nature* **2003***, 426***,** 874-878. [2] J.M. Vargason, G. Szittya, J. Burgyan, T.M. Tanaka Hall, *Cell* **2003**, *115*, 799-811. [3] G.J. Hannon, *Nature* **2002**, *418*, 244-251.

Keywords: protein, RNA, interaction

MS.71.5

Acta Cryst. (2011) A**67,** C161

Biochemical studies of the snRNPs core domain formation in Saccharomyces cerevisiae

Wojciech Galej, Kiyoshi Nagai, *Laboratory of Molecular Biology, Medical Research Council, Cambridge, (UK).* E-mail: wgalej@mrc-lmb.ac.uk

The spliceosome is a dynamic molecular machine, which catalyze excision of non-coding sequences (introns) from precursors of messenger RNAs (pre-mRNAs). It is assembled from 4 canonical subunits – small nuclear ribonucleoprotein particles (U1, U2, U4/U6 and U5 snRNPs) and a number of other, non-snRNP proteins. Each round of splicing requires hierarchical *de novo* assembly of the spliceosomal subunits on pre-mRNA, followed by a series of structural rearrangements to form catalytically competent particle. The catalytic core of the spliceosome consists of a highly structured RNA network formed between U2, U5 and U6 snRNAs and the conserved sequences in the pre-mRNA.

Our goal is to get insight into the structure and the function of splicing machinery by X-ray crystallography combined with electron microscopy and biochemical methods. We are particularly interested in proteins and RNAs involved in U5.U4/U6 tri-snRNP formation in *Saccharomyces cerevisiae*. Our approach is based on *in vitro* reconstitution of spliceosomal subcomplexes from recombinant proteins and *in vitro* transcribed snRNAs. This method has been successfully applied to reconstitute human U1 snRNP[1] and the U4 snRNP core domain[2]. I will present our recent progress in core domain reconstitutions of budding yeast tri-snRNP components and a perspective for future biochemical and structural studies.

[1] D.A. Pomeranz Krummel, C. Oubridge, A.K. Leung, J. Li, K. Nagai *Nature* **2009**, *458(7237)*, 475-80. [2] A.K. Leung, K. Nagai, J. Li *Nature* **2011**, in press.

Keywords: RNA-protein interactions

MS.72.1

Acta Cryst. (2011) A**67,** C161-C162

Graphical tools for structure determination and refinement in PHENIX

Nathaniel Echols,[a] Pavel V. Afonine,[a] Gábor Bunkóczi,[c] Vincent B. Chen,[d] Ian W. Davis,[d] Jeffrey J. Headd,[a] Li-Wei Hung,[e] Gary J. Kapral,[d] Ralf W. Grosse-Kunstleve,[a] Airlie J. McCoy,[c] Nigel W. Moriarty,[a] Robert Oeffner,[c] Randy J. Read,[c] David C. Richardson,[d] Jane S. Richardson,[d] Thomas C. Terwilliger,[e] Peter H. Zwart,[a] Paul D. Adams,[a,b] [a]*Lawrence Berkeley National Laboratory, Berkeley, (USA).* [b]*Department of Bioengineering, UC Berkeley, (USA).* [c]*Department of Haematology, University of Cambridge, Cambridge Institute for Medical Research, (England).* [d]*Department of Biochemistry, Duke University Medical Center, (USA).* [e]*Los Alamos National Laboratory, (USA).* E-mail: NEchols@lbl.gov

PHENIX [1] is a software package for phasing and refinement of macromolecular crystal structures. Although the individual programs

are primarily command-line driven, an emphasis has been placed on ease-of-use and automation. We have developed a graphical interface for the major components of PHENIX, which currently includes phenix.refine, phenix.xtriage, comprehensive validation tools based on the Molprobity web server, Phaser, and the AutoSol, AutoBuild, AutoMR, and LigandFit automation "wizards". The Python-based framework allows new GUIs to be generated semi-automatically while preserving all of the flexibility of the command-line programs, and supports both Macintosh and Linux. Python extensions for Coot and PyMOL facilitate real-time visualization of refinement and automated model-building, and convenient viewing of results. Transitions between separate modules within PHENIX are simplified or eliminated in the GUI, reducing the amount of manual input required and avoiding the use of command-line tools. Further automation is possible with definition of standard parameter sets and input files for individual programs. Future improvements will include greater use of multiprocessing and clusters, tools for handling multiple structures in parallel, and new automation pipelines.

[1] P.D. Adams, et al. *Acta Cryst.* **2010,** *D66*, 213-21.

Keywords: PHENIX, automation, software

MS.72.2

Acta Cryst. (2011) A**67,** C162

More speed, more data, more automation, less work!

Alun W. Ashton, Graeme Winter, *Scientific Software Team, Diamond Light Source Ltd., Diamond House, Harwell Science and Innovation Campus, Didcot, Oxfordshire (UK).* E-mail: alun.ashton@diamond.ac.uk

With samples exchanged, centered and ready for data collection to start in under 60 seconds, 424 x 435 mm area detectors capable of collecting 25 images a second, goniometers rotating at over 10 degrees/second in shutterless data collections, it's not too surprising that many synchrotron beamline users now find themselves deluged with over 7000 images and up to 15 complete datasets or 40 crystal screenings an hour. At Diamond Light Source (UK) the potential Armageddon of more speed, more data and more automation of 5 MX beamlines is only circumvented with a substantial investment in integrated and automated data reduction and structure solution pipelines backed up by high performance data storage and computing clusters.

The implementation so far at Diamond will be outlined along with not only the usage for high throughput experiments and industrial usage but also how the technology is benefiting experiments with more challenging samples.

Keywords: automation, data, reduction

MS.72.3

Acta Cryst. (2011) A**67,** C162

New phasing methods for high throughput crystallography

Carmelo Giacovazzo,[a,b] Maria Cristina Burla,[c] Rocco Caliandro,[a] Benedetta Carrozzini,[a] Giovanni Luca Cascarano,[a] Giuliana Comunale,[d] Annamaria Mazzone,[a] Giampiero Polidori,[c] Dritan Siliqi,[a] [a]*Istituto di Cristallografia, CNR, Bari.* [b]*Dipartimento Geomineralogico, Università di Bari.* [c]*Department of Earth Sciences, University of Perugia.* [d]*Dipartimento di Chimica, Università della Basilicata, Potenza.* E-mail: carmelo.giacovazzo@ic.cnr.it

One year ago a new phasing method was described [1], [2], called *VLD* (*Vive la Difference*), based on the properties of a new type of Difference Fourier Synthesis. A specific feature of the synthesis is that it provides information on the ideal difference map, no matter the quality of the model: in particular, also when the model is uncorrelated with the target structure or is completely random. That allows two types of phasing approaches:

1) ab initio crystal structure solution, even of proteins [3], starting from random phases, or equivalently, from random models;
2) non-ab initio phasing: a starting model is found by one of the most popular phasing methods (e.g., Patterson techniques, molecular replacement, SAD-MAD, SIR-MIR) and *VLD* is applied to extend and to refine phases. Owing the feature described at the point 1) it is expected that even poor models may be a good starting point for *VLD* applications.

The *VLD* algorithm is based on a simple cyclic approach. The basic cycle is the following: the difference Fourier synthesis $\rho_q = \rho - \rho_p$ is estimated, modified and Fourier inverted, the corresponding structure factor F_q is vectorially summed to F_p to provide new phase estimates for the target structure.

Such algorithm is particularly suitable for Structural Solution for High Throughput Crystallography: indeed the phases produced by *VLD* may be submitted with success to automated model building programs.

We are implementing *VLD* into the program IL MILIONE [4].

[1] M.C. Burla, R. Caliandro, C. Giacovazzo, G. Polidori, *Acta Cryst.* **2010,** *A66*, 347-361. [2] M.C. Burla, C. Giacovazzo, G. Polidori, *J. Appl. Cryst*. **2010,** *43*, 825-836. [3] M.C. Burla, C. Giacovazzo, G. Polidori, *J. Appl. Cryst*. **2011,** *44*, 193-199. [4] M.C. Burla, R. Caliandro, M. Camalli, B. Carrozzini, G.L. Cascarano, L. De Caro, C. Giacovazzo, G. Polidori, D. Siliqi, R. Spagna, *J. Appl. Cryst.* **2007,** *40,* 609-613.

Keywords: methods, phasing, proteins

MS.72.4

Acta Cryst. (2011) A**67,** C162

On the systematic scaling and merging of multiple datasets in macromolecular crystallography

James Foadi,[a] Yilmaz Alguel,[a] Wes Armour,[b] Danny Axford,[b] Alex Cameron,[a] Robin Owen,[b] David Waterman[b] and Gwyndaf Evans[b] [a]*MPL, Imperial College, London SW7 2AZ, (UK).* [b]*Diamond Light Source Ltd, Harwell Science and Inovation Campus, Didcot, Oxfordshire OX11 0DE, (UK).* E-mail: j.foadi@imperial.ac.uk

BLEND, a computer program to handle systematically the scaling and merging of data collected from multiple wedges or multiple crystals, is described here for the first time. The availability of more intense and collimated synchrotron sources and fast-reading detectors has enabled protein crystallographers to acquire a large number of datasets from a given structure in a relatively short interval of time. A similar step change in the speed of processing and analysis of such multiple datasets must be realised, or data collection efforts go wasted.

BLEND uses multivariate statistics, mainly in the form of cluster analysis, to bring together datasets with better merging likelihood. The program allows researchers to save time both in avoiding the combinatorial explosion implied in the analysis of multiple datasets and in the cumbersome and time-consuming amount of book keeping that goes with it. BLEND has been successfully used in the solution of a novel membrane protein.

Keywords: multiple, datasets, merging