Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

High-throughput structure determination

R. M. Esnouf,[a] D. I. Stuart[a] and K. S. Wilson[b]

[a] The Division of Structural Biology and Oxford Protein Production Facility, The Henry Wellcome Building for Genomic Medicine, Roosevelt Drive, Headington, Oxford OX3 7BN, UK, and [b] York Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York YO10 5DD, UK

A tidal wave of high-throughput technology is sweeping through biology, revolutionizing research methods, and structural biology will not be spared. Crystallographers have traditionally been at home with technology and sophisticated software so it is perhaps not surprising that improvements in software for phasing and refinement and steps towards beamline automation lead the way, but new techniques and ideas for protein production and crystallization are now catching up. It is an exciting time to be a structural biologist, since a result should be that the emphasis of research can shift from the mechanics of structure determination to the understanding of biology.

Since its inception in 1979, the Collaborative Computational Project in Macromolecular Crystallography (CCP4) has developed from primarily curation and standardization of contributed software to be a major force behind software development. At this key moment in the development of structural biology CCP4 is reviewing its own development and so it was natural to use its annual study weekend to review emerging technologies with the whole structural biology community. CCP4 depends on its wide user-base just as much as that user-base depends on CCP4 software, thus it is in all our interests to support CCP4.

The 2002 Study Weekend was unusual in considering the whole process of structural biology from target selection through protein production, crystallization, data collection and structure solution to deposition, rather than directly computational aspects. Such a break with tradition predictably raised eyebrows and provoked discussion, both favourable and unfavourable, but the vigorous debate and record attendance confirmed the relevance of CCP4 and the study weekend in the high-throughput era. In keeping with the ground-breaking nature of the meeting, many presentations are available on the web (http://www.ccp4.ac.uk/talks/stwk2002/main.html).

The weather took an unusually heavy toll on the speakers with the introductory session being particularly badly affected. Dave Stuart expanded his introductory remarks into an impromptu overview of structural genomics from a European perspective. Tom Terwilliger then divided his time between a complementary US perspective, focussing on an NIH-funded structural genomics institution, and developments in his *SOLVE/RESOLVE* package for automated phasing and model building. In the final talk, Gerhard Materlik gave an update on the most important single investment in UK science in the past 25 years, the new Diamond synchrotron.

In the high-throughput regime, bioinformatics and IT pervade the whole process from selection of initial targets to providing a biological context for the final structure. Malcolm Weir discussed how commercial bioinformatics service companies can drive the drug discovery process with genome-wide calculations. Julian Gough presented the SUPERFAMILY database that encapsulates SCOP domain classifications by hidden Markov models. Laboratory information management systems (LIMS), or at least tracking databases, will drive research work and provide the opportunity for methodological improvements through data mining. Joe Peden described the benefits of LIMS from a commercial perspective. Pedro Alzari then described how bioinformatics is being used at the Pasteur Institute to target the research as part of the NIH-funded Mycobacterium tuberculosis consortium. Lastly, a perspective from a commercial structural genomics company was provided by Eric de La Fortelle from SGX where dependence on LIMS and data mining are far more accepted than in academia.

The final Friday session saw the stage given over to presentations on how techniques other than crystallography will be part of the structural genomics revolution. Ian Jones summarized some of the problems associated with high-throughput expression and considered possibilities for streamlining the process. Fergal Hill also discussed aspects of

the same problem and the developments at Avidis including refolding technologies. The area of crystallization trials is particularly amenable to parallelization and Lajos Nyársik gave a tour of the robotized facility being implemented at the Berlin Protein Structure Factory. Iain Campbell bravely stood up in front of 500 crystallographers to discuss the high-throughput potential of NMR and last, but not least, Jan-Pieter Abrahams outlined the possibilities for micro-crystallization.

Saturday began with Glen Spraggon and Julie Wilson giving papers covering somewhat different approaches to the formidable computational problem of detecting crystal growth among the millions of crystallization trials that will be generated by crystallization robots. Filling in for Ashley Deacon, stranded in Atlanta by the weather, Martin Walsh gave an update of the state of beamline automation in the US based on his fact-finding tour on behalf of BM14, the UK MAD beamline at the ESRF. Martin was followed by Andy Thompson who described parallel European developments in Grenoble and Andrew Leslie who described the developments in the area of expert systems for automated data collection control and processing based around *MOSFLM*.

Phasing and structure solution were covered in a dedicated session. Zbigniew Dauter described strategies for real-time structure solution at Brookhaven National Laboratory. Martyn Winn outlined the commitment of CCP4 to enabling high-throughput structure determination, the evolution of software libraries underlying the *CCP*4 suite and project tracking facilities. The methods incorporated in *ACORN*, a recent phasing and phase-refinement addition to the *CCP*4 suite, were described by Yao Jia-Xing along with some examples of successful application. Garib Murshudov then gave a presentation on the new version of the stalwart CCP4 refinement program, *REFMAC*. Finally, Paul Adams described the early stages of development of *PHENIX*, a python-based framework for automating crystallography.

After lunch Dusan Turk gave an introduction to his graphics and model-building package, *MAIN*. Graphics was also the theme for Liz Potterton as she gave an update on progress with the CCP4 molecular graphics initiative. A wider perspective was then provided by the final speaker, Alwyn Jones, with a thought-provoking overview of the problems associated with validation and interpretation of structures produced by high-throughput studies. The meeting closed with a discussion stimulated by single overheads presented by a panel chosen for their strong opinions and lively debating manner.

And so the study weekend was over for another year and all that is left is to do apart from thanking all the speakers once again is to thank David Brown and his replacement, Maeri Howard Eales, for their organizational skill and patience. Thanks are also due to the staff of the University of York conference office for their efforts and to Louise Jones at the IUCr in Chester for bringing the proceedings together in this issue.