MS6-01 **Bioinformatics at OPPF-UK.** Jonathan Diprose,[ab] Robert Esnouf,[c] Chris Morris,[d] Raymond Owens,[ab] David Stuart,[ab] [a]*Oxford Protein Production Facility UK, Research Complex at Harwell, Harwell Oxford, UK,* [b]*Division of Structural Biology, Wellcome Trust Centre For Human Genetics, Oxford, UK,* [c]*Wellcome Trust Centre For Human Genetics, Oxford, UK,* [d]*Daresbury Laboratories, Daresbury, UK*
E-mail: jon@strubi.ox.ac.uk

The Oxford Protein Production Facility UK (OPPF-UK) is a National Resource Centre for protein production and crystallization in the UK[1]. OPPF-UK operates as a free-at-the-point-of-use user facility with a peer-reviewed application process for UK academics and as part of the European Union-funded P-Cube[2] and Instruct[3] transnational access programmes. The processes available include construct design, cloning, bacterial and eukaryotic expression and crystallization. OPPF-UK runs a standard set of bioinformatic analyses on target proteins including sequence alignment against various public databases and the prediction of signal sequences, glycosylation sites, transmembrane helices and disordered regions, with results stored in a database (OPTIC[4]). OPPF-UK has also implemented informatics systems based on OPTIC and the Protein Information Management System (PiMS/xtalPiMS[5,6]) to assist in construct design and the management of samples, experiments and results. The presentation will cover how OPPF-UK makes use of these tools, and others, to improve its efficiency and success rate.

[1] OPPF-UK, http://www.oppf.ox.ac.uk/ (2012).
[2] P-Cube, http://www.p-cube.eu/ (2012).
[3] Instruct, http://www.structuralbiology.eu/ (2012).
[4] Albeck, S., Alzari, P., Andreini, C., Banci, L., Berry, I. M., Bertini, I., Cambillau, C., Canard, B., Carter, L., Cohen, S. X., Diprose, J. M., Dym, O., Esnouf, R. M., Felder, C., Ferron, F., Guillemot, F., Hamer, R., Ben Jelloul, M., Laskowski, R. A., Laurent, T., Longhi, S., Lopez, R., Luchinat, C., Malet, H., Mochel, T., Morris, R. J., Moulinier, L., Oinn, T., Pajon, A., Peleg, Y., Perrakis, A., Poch, O., Prilusky, J., Rachedi, A., Ripp, R., Rosato, A., Silman, I., Stuart, D. I., Sussman, J. L., Thierry, J.-C., Thompson, J. D., Thornton, J. M., Unger, T., Vaughan, B., Vranken, W., Watson, J. D., Whamond, G. and Henrick, K. (2006). *Acta Cryst.* D**62**, 1184-1195.
[5] Morris, C., Pajon, A., Griffiths, S. L., Daniel, E., Savitsky, M., Lin, B., Diprose, J. M., Wilter da Silva, A., Pilicheva, K., Troshin, P., van Niekerk, J., Isaacs, N. , Naismith, J., Nave, C., Blake, R., Wilson, K. S., Stuart, D. I., Henrick, K. and Esnouf, R. M. (2011). *Acta Cryst.* D**67**, 249-260.
[6] Daniel, E., Lin, B., Diprose, J. M., Griffiths, S. L., Morris, C., Berry, I. M., Owens, R. J., Blake, R., Wilson, K. S., Stuart, D. I. and Esnouf, R. M. (2011). *J. Struct. Biol.* **175**, 230-235.

Keywords: bioinformatics; informatics; protein production

MS6-02 **Applying bioinformatics to improve models for molecular replacement.** Gábor Bunkóczi, Randy J. Read
Department of Haematology, University of Cambridge, UK
E-mail: gb360@cam.ac.uk

In molecular replacement, the quality of search models can often be improved by capturing and applying information provided by bioinformatics tools. Calculating local sequence similarity based on an alignment between the target and the model sequence can identify regions in the model that are likely to be missing or markedly different in the target and therefore should either be removed or downweighted [1]. In addition, flexible residues identified using an accessible surface area calculation could also be downweighted to improve the model [2]. When several, sufficiently diverse alternative models are available, structural homology and flexibility can be captured using an ensemble model. This requires precise superposition of the structurally conserved segments from each potential model. In the program Ensembler, this is achieved by an iteratively weighted multiple superposition [3, 4] using a robust-resistant weighting scheme. Based on pairwise overall root-mean-square deviations (r.m.s.d.), it is also possible to group the structures into structure families, remove structures from the superposition that are not sufficiently or overly distinct from the others. In addition, the r.m.s.d. of a superposed position has been proved useful in trimming the structures to their conserved cores.

[1] Bunkóczi, G. & Read, R. J (2011). *Acta Cryst.* **D67**, 303-312.
[2] Lebedev, A. A., Vagin, A. A. & Murshudov, G. N. (2008). *Acta Cryst.* **D64**, 33-39.
[3] Diamond, R. (1992). Protein Sci. 1, 1279-1287.
[4] Wang. X. & Snoeyink, J. (2008). *IEEE/ACM Trans. Comput. Biol. Bioinform.* **5**, 525-533.

Keywords: molecular replacement; model improvement; ensemble model