MS5-02 **Ensemble refinement of protein crystal structures in PHENIX.** Tom Burnley,[a] Piet Gros,[a] [a]*Crystal and Structural Chemistry, Bijvoet Center for Biomolecular Research, Department of Chemistry, Faculty of Science, Utrecht University, The Netherlands*
E-mail: b.t.burnley@uu.nl

Despite recent advances in data collection and processing in protein crystallography, there remains a significant discrepancy between measurement and model error for macromolecular crystal structures. This has been attributed, at least in part, to the incomplete modelling of atomic disorder. Here we present an alternative refinement method which simultaneously includes anisotropic and anharmonic disorder. This ensemble refinement uses a molecular dynamics approach augmented with time-averaged x-ray restraints [1] to produce a population of Boltzmann-weighted structures that represents the conformational space sampled during a simulation. The resulting ensemble model typically contains 100-250 structures and is shown to significantly improve the model error (as judged by Rfree), in comparison with traditional refinement methods that yield a single structure. This method is suitable for diffraction data with upper resolution limits in the range of 1-3Å d-spacing and does not require excessive computation time and can be run on a standard desktop machine.

Ensemble refinement was developed, and will be available, within the PHENIX software suite [2]. It utilises a maximum-likelihood target function in conjunction with a dual explicit- and bulk-solvent model and can be used with any heterogeneous atom or group. In addition to the improved global statistics, ensemble refinement reveals highly-resolved local disorder features which are demonstrated to reflect important functional details for a number of test cases. Furthermore, the reproducibility and reliability of the sampled atomic distributions will be discussed along with the effect of the upper resolution limit in addition to methods for assisting the viewing and the usage of ensemble models.

[1] P. Gros, W.F. van Gunsteren and W.G.J. Hol. (1990). *Science* ***249***, 1149-1152.
[2] P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger and P. H. Zwart. (2010). *Acta Cryst.* D**66**, 213-221.

Keywords: macromolecular; refinement; disorder

MS5-03 **Use of the Crystallography Open Database as a source of prior knowledge for molecular modelling.** Saulius Grazulis,[a] [a]*Vilnius University Institute of Biotechnology, Lithiania*
E-mail: grazulis@ibt.lt

Crystallography has yielded outstanding amount of important information about structure of matter which is used in many academic and practical applications. The information obtained from crystallographic studies can be instrumental to derive restrains for macromolecular refinement [1], to predict accurate structures of chemical compounds [2] and to investigate their properties [3], as an example. The ubiquitous availability of accessible computing power puts a pressing need to have data available instantly in multiple computing nodes. The Crystallography Open Database (COD)[4,5] is an open-access database that can offer such functionality. Currently, around 200 thousand records have been accumulated in the COD (http://www.crystallography.net/), covering organic, inorganic and metal-organic molecules. The distinctive feature of COD is its free availability – the database itself, or any derived data, may and can be distributed freely along with any software that is used for refinement or other computations. The amount of data available in the COD permits to extract representative set of high resolution structures with high accuracy in many cases. We are thus using the COD to derive restrains for the refinement of biological macromolecules, especially in complexes with various ligands. Since the COD contains primary information about the solved and published crystal structures, full molecule geometry has been derived and tables of bond lengths, bond angles and dihedral angles are calculated. From the current revision of the COD, data about 13 mln. bonds, 25 mln. bond angles and 45 mln. dihedral angles were obtained. Subsequently, using a COD provided atom classification based on the atomic environment connected via chemical bonds, a custom atom classification, or a molecular connectivity graph matches, bond length and angle distributions along with their parameters (averages and standard deviations, higher moments) were extracted and used to predict best molecular geometries and their variabilities, which can then serve as a source of restrains for macromolecular refinement. Thus, COD will serve as a source of reliable restrains based on experimental data that can be used for macromolecular ligand complex refinement. In order to maintain full reproducibility of all computations based on COD, both by COD team and by any external COD users, we maintain version numbers of all derived COD data records, thus ensuring possibility to rerun any computations and to trace provenance of the data.

[1] Engh, R. & Huber, R. (2006). in *Intl. Tables Cryst. vol.* **F**, ed. by Rossmann, M. G. & Arnold, E., 382-392.
[2] Andronico, A.; Randall, A.; Benz, R. W. & Baldi, P. (2011). *J. chem. inf. mod.* **51**, 760-76.
[3] Pepponi, G.; Gražulis, S. & Chateigner, D. (2011). *Nucl. Instr. Meth. Phys. Res. B*, doi:10.1016/j.nimb.2011.08.070
[4] Gražulis, S., Daškevič, A., Merkys, A., Chateigner, D., Lutterotti, L., Quirós, M., Serebryanaya, N. R., Moeck, P., Downs, R. T. & Le Bail, A. (2012). *Nucleic Acids Res.* **40**, D420-D427.
[5] Gražulis, S.; Chateigner, D.; Downs, R. T.; Yokochi, A. F. T.; Quirós, M.; Lutterotti, L.; Manakova, E.; Butkus, J.; Moeck, P. & Le Bail, A. (2009). *J. Appl. Crystallogr.* **42**, 726-729.

Keywords: crystallographic databases; open-access; molecular geometry