MS.43.2

Acta Cryst. (2011) A**67,** C104

Validation of small- and macro-molecular X-ray structures: PDB and CCDC collaborations

Swanand Gore,[a] Tjelvar S. G. Olsson,[b] Marina Zhuravleva,[c] *[a]EMBL, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton (UK). [b]Cambridge Crystallographic Data Centre, Cambridge (UK). [c]RCSB Protein Data Bank, Department of Chemistry and Chemical Biology, Rutgers The State University of New Jersey (USA).* E-mail: swanand@ebi.ac.uk

The Protein Data Bank (PDB) and the Cambridge Crystallographic Data Centre (CCDC) have been collecting and curating X-ray structures for almost half a century. Although the databases store different types of structures, macro-molecules and small-molecules respectively, both databases exhibit an exponential growth. This is partly due to experimental techniques becoming more automated and partly due to structure solution software packages becoming more accessible to "non-expert" users.

This raises two issues for both the PDB and the CCDC. First of all we need to design processes and workflows that allow our finite resources to deal with the exponential growth in structures. Secondly, we need mechanisms for flagging honest mistakes made by less experienced (and experienced) users.

In this talk we will discuss issues arising when processing small-molecule and macro-molecule structures. We will also discuss how the PDB and the CCDC have learnt from studying each other's data processing procedures and how we are sharing information and technologies to improve the quality of data in the databases.

Keywords: validation, PDB, CCDC

MS.43.3

Acta Cryst. (2011) A**67,** C104

The most powerful crystallographic validation tool: Common sense

Bernhard Rupp, *k.-k. Hofkristallamt, Livermore, CA 94551 (USA).* E-mail: hofkristallamt@gmail.com.

A recent rush of retractions of protein structures published in high impact journals has led to the call for more rigorous validation of crystallographic protein structure models as well as the underlying structure factor amplitudes [1]. While a posterior validation by an independent repository such as the PDB certainly serves this need (but does not require that the depositor actually corrects the model diligently), a more proactive process is concurrent *a priori* model validation during the building-refinement cycles. This is achieved for example in the model building program Coot [2], where real space geometry and electron density both are used for model improvement.

Interestingly enough, despite the availability of powerful validation tools, the majority of flawed models that passed (or ignored) the tests (and also fooled editors and reviewers) could have been prevented by using the most powerful validation tool at our disposal: the common sense of the person building the model. Basic probabilistic inference models that consider both the crystallographic evidence and the compatibility of the model with all available prior knowledge suffice to intercept almost all major problems - including some rare cases of data fabrication. While such probability analysis provides a solid qualitative measure for model correctness, it still can be fooled by the most insidious problem in biomolecular crystallography: Mental phase (or model) bias often provides an overwhelming desire to find what one seeks. Particularly ligand structures suffer from this inherent temptation, and depend heavily on critical and unbiased plausibility analysis. Again, such modeler-bias-introduced problems need to be preferably addressed *a priori* in crystallographic education and curriculum [3], with a posteriori validation remaining only the final safeguard against errors. With the great power of modern crystallography comes great responsibility - and that responsibility ultimately rests with the model depositor, irrespective of any automated validation.

[1] E.N. Baker, Z. Dauter, H. Einspahr, et al., *Acta Crystallogr*. **2010**, *D66(2)*, 115. [2] P. Emsley, B. Lohkamp, W.G. Scott, et al., *Acta Crystallogr*. **2010**, *D66(4)*, 486-501. [3] B. Rupp, *J. Appl. Crystallogr*. **2010,** *43(5)*, 1242-1249.

Keywords: Model validation, Bayes' Theorem, Bias

MS.43.4

Acta Cryst. (2011) A**67,** C104-C105

Unrestrained reciprocal space refinement can indicate alternative conformations

Oleg V. Sobolev, Vladimir Y. Lunin, *Institute of Mathematical Problems of Biology, Russian Academy of Sciences, Pushchino, (Russia).* E-mail: oleg@impb.ru

The reciprocal space refinement with geometrical restraints turned off becomes a common practice when working at sub-atomic resolution. Nevertheless even at ultra-high resolution the stereochemical restraints are usually kept for the residues found to be in alternative conformations. In other case these residues deteriorate significantly. We suggest that this property can be used in an opposite direction as an indicator that can reveal the necessity of alternative conformations for a given residue when applied at early stages of refinement with all residues present in a single conformation. Our tests demonstrated that for the resolution higher than 1.2Å a formal procedure of unrestrained refinement gives a usefull hint for which residues might be checked thoroughly with electron density maps as possible candidates for the presence of alternative conformations.

To check this suggestion we designed a pipeline to select structures from PDB with desired resolution and R-factor values (using PDB search engine), download and process experimental data and perform unrestrained refinement with the use of PHENIX [1], and calculate atomic shifts. This analysis allowed to estimate "normal" value for coordinate shifts taking into account resolution and atom properties (main or side chain, protein surface or core residues *etc*.).

The most thorough analysis was performed for structures refined in 1.2-1.1 Å resolution range. It included visual analysis of several electron density maps and comparison with asignments of alternative conformations originaly present in PDB files. It was found that usually residues possessing of abnormal atomic shifts after unrestraind refinement either are already present with alternative conformations in PDB file or the electron density map suggests such idea. Some correlation was found as well in magnitudes of atomic shifts and relative occupancies of alternative conformations. The maximal coordinate deviations were obtained for residues that have alternative conformations with equal occupancies.

An attempt of the use of similar procedure at lower resolution had failed, while it worked similary well at higher resolution [2]. The exclusion from the model of the water molecules resulted in significant growth of shift for the most part of the structure.