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

Swanand Gore, Tjèvår S. G. Olsson, Marina Zhuravleva, EMBL, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton (UK); Cambridge Crystallographic Data Centre, Cambridge (UK); 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

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 posteriori validation by an independent repository such as the PDB certainly serves this need (but does not require that the depositor actually corrects the model), a more proactive process is concurrent (but does not require that the depositor actually corrects the model) bias often provides an overwhelming desire to find what one expects. In the model building program Coot [2], where real space geometry and electron density both are used for model improvement, an attempt of the use of similar procedure at lower resolution failed, while it worked similarly well at higher resolution [2]. 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 1.2 Å a formal procedure of unrestrained refinement gives a useful 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 assignments of alternative conformations originally present in PDB files. It was found that usually residues possessing of abnormal atomic shifts after unrestrained 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 similarly 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.

Keywords: Model validation, Bayes’ theorem, Bias

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 1.2 Å a formal procedure of unrestrained refinement gives a useful 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 assignments of alternative conformations originally present in PDB files. It was found that usually residues possessing of abnormal atomic shifts after unrestrained 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 similarly 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.