Published online 11 November 2014 *Nucleic Acids Research, 2015, Vol. 43, Database issue* ***D369–D375***
doi: 10.1093/nar/gku1118

ValidatorDB: database of up-to-date validation results for ligands and non-standard residues from the Protein Data Bank

David Sehnal[1,2,3,†], Radka Svobodová Vařeková[1,2,†], Lukáš Pravda[1,2], Crina-Maria Ionescu[1], Stanislav Geidl[1,2], Vladimír Horský[3], Deepti Jaiswal[1], Michaela Wimmerová[1,2] and Jaroslav Koča[1,2,*]

[1]CEITEC—Central European Institute of Technology, Masaryk University Brno, Kamenice 5, 625 00 Brno, Czech Republic, [2]National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Kotlářská 2, 611 37 Brno, Czech Republic and [3]Faculty of Informatics, Masaryk University Brno, Botanická 68a, 602 00 Brno, Czech Republic

Received August 29, 2014; Revised October 24, 2014; Accepted October 24, 2014

ABSTRACT

Following the discovery of serious errors in the structure of biomacromolecules, structure validation has become a key topic of research, especially for ligands and non-standard residues. ValidatorDB (freely available at http://ncbr.muni.cz/ValidatorDB) offers a new step in this direction, in the form of a database of validation results for all ligands and non-standard residues from the Protein Data Bank (all molecules with seven or more heavy atoms). Model molecules from the wwPDB Chemical Component Dictionary are used as reference during validation. ValidatorDB covers the main aspects of validation of annotation, and additionally introduces several useful validation analyses. The most significant is the classification of chirality errors, allowing the user to distinguish between serious issues and minor inconsistencies. Other such analyses are able to report, for example, completely erroneous ligands, alternate conformations or complete identity with the model molecules. All results are systematically classified into categories, and statistical evaluations are performed. In addition to detailed validation reports for each molecule, ValidatorDB provides summaries of the validation results for the entire PDB, for sets of molecules sharing the same annotation (three-letter code) or the same PDB entry, and for user-defined selections of annotations or PDB entries.

INTRODUCTION

Validation of biomacromolecular structures has become a very important topic, because some published structures have been found to contain serious errors (1–4). The first step in the validation of biomacromolecules and their complexes is checking the standard building blocks, namely, standard amino acids and nucleotides. The usual procedure is to evaluate specific properties of each residue (e.g. electron density, atom clashes, bond lengths, bond angles, torsion angles, etc.). Various software tools have been developed to perform such analyses, e.g. WHAT_CHECK (5), PROCHECK (6), MolProbity (7) and OOPS (8).

The next key step is the validation of ligands and non-standard residues in biomacromolecular structures, which can be performed in a similar manner as for standard residues (focus on electron density, atom clashes, etc.). An example of software specialized on this type of validation is ValLigURL (9). This approach was also added to several software tools focused on the validation of standard residues (Mogul (10), Coot (11), PHENIX (12)).

A different ligand validation approach, which can be denoted as validation of annotation, was developed later. The goal of this approach is to evaluate if the ligand or non-standard residue is annotated correctly (i.e. if its structure corresponds to the three-letter code it was assigned in the Protein Data Bank (PDB) file format). Specifically, the topology and stereochemistry of the validated molecule are compared to those of a reference molecule (model), and any differences found are reported. The first software tool implementing this methodology has been pdb-care (13), a tool specialized on carbohydrates. The next step has been MotiveValidator (14), which allows validation of all ligands and residues, performs basic validation analyses and reports

*To whom correspondence should be addressed. Tel: +420 54949 4947; Fax: +420 54949 2556; Email: Jaroslav.Koca@ceitec.muni.cz
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.