AUTOMATED DETECTION OF POOR OR INCORRECT SINGLE CRYSTAL STRUCTURES

A.L. Spek

Utrecht University Chemistry Padualaan 8 UTRECHT 3584 CH NETHERLANDS

The need for automated validation of reported crystal structures is obvious. The volume of published structures is increasing rapidly with relevant crystallographic details often lacking or buried in inaccessible supplementary material. Validation is becoming even more urgent in view of the current practice where non-professionals carry out structure determinations using 'user-friendly-firmware' and referees with limited expertise. The PLATON/CHECK option (http://www.cryst.chem.uu.nl/platon), developed as part of the IUCr CHECKCIF validation project, was designed to automatically generate an objective list of unusual issues (ALERTS) for a file containing the results of a single crystal study in CIF-format. ALERTS will in general point either to unresolved problems of the analysis or to potentially interesting non-standard observations. Detailed discussion and analysis in the latter case should be obligatory since many unusual observations are related to pitfalls (e.g. bond-stretch isomerism).

Major issues addressed by the software are missed symmetry problems, unusual intra-molecular geometry, (short) inter-molecular contacts and solvent accessible voids in the structure. Symmetry problems can be either of the 'Marsh-missed-symmetry' type or blunt errors, sometimes only detectible with an analysis of short inter-molecular contacts. Several of the last variety made it undetected into the CSD with the side effect of increasing the 'noise-level' in statistical studies. Hydrogen atoms can be either misplaced, too many or missing. False atom type assignments usually show up with various ALERTS including those related to displacement parameters. The validation software is available free-of-charge or through the IUCr-WEB server.

Keywords: STRUCTURE VALIDATION , CHECKCIF, PLATON

ADVENTURES OF A DATA VALIDATION EDITOR

A. Linden

Institute of Organic Chemistry, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

CHECKCIF is the data validation tool used by authors of structural papers destined for Acta Crystallographica, although it can (should?) be used to check any crystal structure. The system is now familiar to most people. Naturally, the CHECKCIF and submission procedures need flexibility that software alone cannot provide, so a human must be part of the system. For Section C of Acta Cryst., a Data Validation Editor fills that role. What should be done when CHECKCIF complains? First, try to resolve every validation alert, even the minor 'C' type. Alerts concerning syntax, CIF completeness or cosmetic aspects (e.g. molecular center of gravity outside unit cell) can normally be cured readily. Alerts that indicate unresolved problems with the structure analysis or non-standard observations may need more careful consideration and additional corrective action. If the crystallography has been performed thoroughly and conscientiously, yet validation alerts still persist and it is believed there is a sound scientific explanation for the observation, an explanatory comment must be included in the CIF before it is submitted to the journal. These explanations are then considered by the Data Validation Editor. When it is clear that an alert is inapplicable (e.g. pseudo-symmetry), or that the situation has been addressed appropriately, the manuscript will be processed normally. Otherwise, the Data Validation Editor may request additional action, but will try to be as helpful as possible and wherever feasible suggest a means of addressing the situation. Sometimes an author has overlooked an alternative strategy or a trivial point.

Keywords: DATA VALIDATION STRUCTURE CHECKING CHECKCIF

VALIDATION DURING STRUCTURE REFINEMENT USING 'CRYSTALS'

R. I. Cooper D. Watkin

[1]Chemical Crystallography Laboratory, 9 Parks Road, Oxford OX1 3PD, UK.

Crystal structures of small- and medium- sized molecules are increasingly being determined by users of crystallographic software, who have little or no formal crystallographic training. In general, crystal structure analysis software does not incorporate chemical knowledge or specialist guidance, and the task of weeding out poor results is left to the journals.

The CRYSTALS software package for structure refinement and analysis contains a complete system for guiding users through a routine crystal structure analysis from diffractometer data to final CIF files. Recent developments have concentrated on improving the variety and quality of validation checks carried out while the analysis is underway, so that the user, and the software that is guiding them, can spot problems early on. In particular, a link to MOGUL (Cambridge Crystallographic Data Centre, to be published), a new software tool in the Cambridge Structural Database System has been developed. MOGUL is a library derived from the main CSD and designed to allow very fast retrieval of molecular geometries. These data can be compared with the crystallographic model under refinement as a useful validation measure, which will immediately reveal errors or 'interesting' features.

The current release of CRYSTALS runs under Microsoft Windows 95/98/Me/NT/XP. Recommended minimum specifications are 100MHz Pentium processor, 32Mb of RAM and 10Mb spare disk space. The program uses OpenGL to display the molecular diagram. The CRYSTALS program and manuals can be downloaded from http://www.xtl.ox.ac.uk/crystals.html. The program is free for academic and not-for-profit use. New releases are generally provided every one to two months.

Keywords: VALIDATION REFINEMENT ANALYSIS

A CODE OF STANDARDS: VALIDATION OF PROTEIN STRUCTURES IN THE ERA OF STRUCTURAL GENOMICS

J. Hendle J. Badger

[1]Structural GenomiX (SGX), 10505 Roselle Street, San Diego, CA 92121, USA

SGX has established a high-throughput structure determination process based on a large-scale genomics approach to increase the efficiency and effectiveness of the structure-based drug discovery process.

The SGX protein structure quality-control process uses an automated and effective structure validation system including a feedback loop that enables the crystallographer to refine a protein structure in the shortest time possible without compromising the quality of the structure. This automated structure deposition and validation system (1) captures data processing statistics and reflection lists, (2) standardizes the input coordinates, (3) calculates R-factors, real-space density correlation and other quantities relating the model to the diffraction data, (4) calculates coefficients for difference maps and for experimentally phased maps, (5) calculates stereochemical measures for the model quality and (6) compares the amino acid sequence with the contents of the coordinate file.

All data processed by this system is complete, calculated by standardized methods and represented by a completely uniform annotation scheme. Outputs from this system include a set of crystallographic information files (CIFs) containing key measures of structure quality and data processing statistics plus a list of amino acids that should be examined for errors (Badger and Hendle, Acta Cryst D**58**, 284-291, 2002).

Our standard operating procedures require that a well-defined set of numerical quality measures are met before a structure is uploaded into our Oracle database and made available for drug discovery activities. These quality measures were initially based on analysis of the first 26 structures solved at SGX and compared favorably with statistics obtained for recent structures taken from the PDB. We have now verified and refined these standards using a set of over 70 new structures.

Keywords: STRUCTURE VALIDATION STRUCTURAL GENOMICS COMPUTER PROGRAM