Introduction to single crystal X-ray analysis VI. About CIFs—Alerts and how to handle them—

Akihito Yamano*

1. Introduction

CIF is an abbreviation for Crystallographic Information File, and these files record all of the information pertaining to crystal structure analysis. A CIF is written as a text file, and thus its content can be checked and edited using ordinary text editing software. CIFs are written with a special-purpose syntax, but they have spread rapidly due to their adoption by the IUCr (International Union of Crystallography), and are indispensible in X-ray structure analysis and related fields. In particular, when submitting papers whose focus is reporting the structure of a molecule, authors are frequently asked to submit a CIF, and there are likely many researchers who have struggled to understand and handle the alerts which appear during checking with checkCIF/PLATON[2] on the IUCr website[1].

This paper discusses the purpose of and background behind the adoption of CIFs, alerts which frequently appear when checking using checkCIF/PLATON and how to handle them, details on judgment criteria, and vrf's (validation reply/response forms). To deal with these alerts, one must first be familiar with checkCIF/PLATON, and the author will be thrilled if this paper serves as a opportunity for readers to improve their understanding.

2. CIF significance and background

CIFs were originally devised as a way to transfer crystal structure information in the form of an electronic file. Before CIFs appeared, structure data was described in formats specific to various software packages. In some cases data had to be exchanged in a free format when no common format could be found between different types of software. This method was extremely inconvenient because information such as lattice constants and space groups had to be input separately by hand.

In the late 70s, IUCr began to encourage transfer of data in a fixed format to facilitate tasks such as processing of submitted papers. In the beginning, they encouraged use of a format called SCFS (Standard Crystallographic File Structure). General-purpose files need to be easy to use with various types of software, and, more specifically, it must be easy to program the reading/writing part. However, SCFS did not have general versatility sufficient for responding to these sort of needs.

In 1987, at the 14th Congress of the IUCr held in Perth, Australia, electronic paper submissions were encouraged by the crystallographic journal *Acta Crystallographica*. At the European Crystallographic Meeting held later that same year, it was decided to describe structure information in the STAR format (Self-defining Text Archive and Retrieval) developed for CIF by Mr. Syd Hall. In 1990, CIF was announced at the 15th Congress of the IUCr held in Bordeaux, France. Later, Syd Hall *et al.* published a paper on CIF in *Acta Crystallographica Section A* in 1991[3], and CIF subsequently came into use for transferring crystal structure information. At present, the original purpose of enabling transfer of structure information has been fully achieved, and there are journals such as *Acta Crystallographica Section C* and *E* to which papers are submitted in CIF itself. In addition, software for structure analysis, and almost all software using crystal structure as basic information, can now import CIF data. It can also be imported by modeling software for analysis of protein structure, and is used for analysis of the cocrystal structure of proteins and small molecule compounds.

3. checkCIF/PLATON

When a CIF is created, its content must be checked before submission. Checking is done using checkCIF/PLATON[2], available at the IUCr website. checkCIF/PLATON combines both checkCIF produced by IUCr, and the CIF check function of the PLATON[4] crystallographic software. Some items are duplicated by checkCIF and PLATON, but basically they each have their own validation criteria. A deeper understanding of checkCIF/PLATON can be achieved by understanding the nature of these validation criteria.

When CIF check is executed, the first step is checking of the CIF syntax. When syntax checking is finished, the page changes and crystallographic data are displayed. Alerts are displayed under that (Fig. 1).

On the left side of an alert line, the name of the test involved in the alert is indicated. Here the test name is PLAT230. Next is the type of alert. There are four types of alert: 1–4. In this example, the type is 2. Type 2 is an alert relating to a mistake or defect in the molecule model. To the right of the type, the alert level is indicated. There are four alert levels, and in this line the level is B. This is the second most severe alert level. Finally, there is the alert message. In this example, the content of the message is that there is variation

* Application Laboratories, Rigaku Corporation.

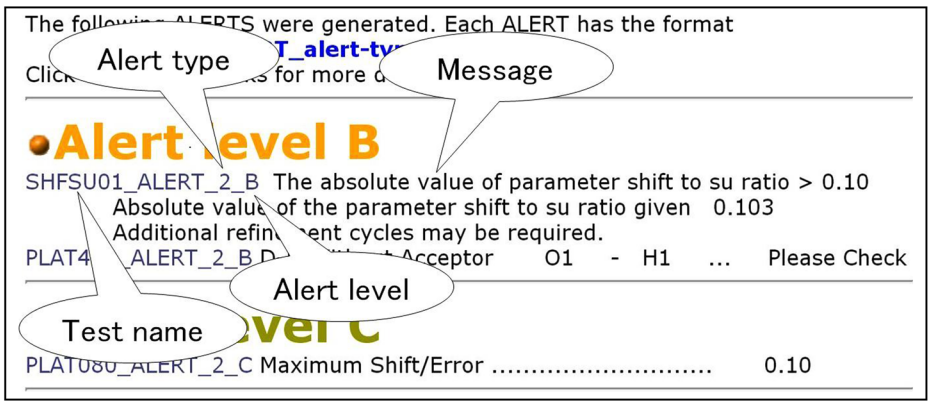

Fig. 1. An alert section of output when checkCIF/PLATON was executed.

in the bonding direction of the bonded atoms O8 and O9. Under that are a summary of the alert, and a vrf (validation replay formula) template. The next section discusses alerts which occur frequently, and how to handle them.

4. Frequently occurring alerts and how to handle them

With the CrystalStructure structure analysis program package, if the Check Acta option is designated, major items are checked at the end of refinement. The following description covers alerts which often appear with checkCIF/PLATON even when all of the Check Acta items of CrystalStructure are satisfied.

Alerts which frequently occur include those relating to: (1) Residual electron density, (2) General temperature factors, (3) Magnitude of temperature factors, (4) Extension direction of anisotropic temperature factors, and (5) Shape of anisotropic temperature factors.

4.1. Alerts relating to residual electron density

The maximum difference density is > 0.1*ZMAX*1.00 _refine_diff_density_max given= 1.020 Test value=0.800

Large Reported Max. (Positive) Residual Density 1.02 eA^{-3}

Here, ZMAX is the atomic number of the heaviest atom. Possible causes of this alert include improper absorption correction and overlooking twins. It also sometimes occurs when crystallinity is poor and data precision is low. The simplest cause may be mistaken atom assignment. In particular, large residual electron density may appear due to an excess or insufficiency of electrons around heavy atoms. In cases where there is a large residual electron density in the solvent region, and this cannot be addressed with a feasible solvent molecule model, the issue can be addressed by smoothing the electron density in the solvent region using SQUEEZE (one of the functions of PLATON).

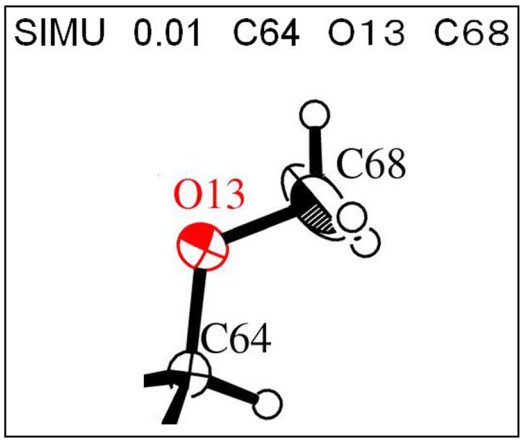

Fig. 2. Causes and handling of an alert relating to the size of temperature factors "Check High (Low) Ueq as Compared to Neighbors for O13".

4.2. Alerts relating to general temperature factors

Large Non-Solvent C Ueq(max)/Ueq(min) ... 6.03 Ratio

In structures at non-solvent parts, this alert indicates that there is a large range of sizes of temperature factors for the same atomic species, in this case carbon. The cause may be mistaken atom assignment. To address the issue, first check that there are no mistakes in assignment of atoms. If there are no mistakes, try to avoid the alert using the DELU instruction of SHELXL.

4.3. Alerts relating to the size of temperature factors

Check High (Low) Ueq as Compared to Neighbors for O13

This alert indicates that, comparing the temperature factors of the bonded atoms, the temperature factor of the parent atom (in this case O13) is too large or too small (Fig. 2). The cause may be mistaken atom assignment. This alert tends to occur with tert-butyl groups and similar structures. If there is thought to be no problem with the structures, try to avoid the alert using such as the SIMU instruction of SHELXL.

4.4. Alerts relating to the extension direction of anisotropic temperature factors

Hirshfeld Test Diff for O20 -- C79 ... 8.57 su
Large Hirshfeld Difference O50 -- C174 ... 0.24 Ang.

This alert originated from the Hirshfeld test. Hirshfeld is a person's name. He was active in the field of precision structure analysis, took notice of temperature factors, and proposed a criterion for determining the validity of temperature factors. The alert given here indicates a difference in displacement in the bonding axis direction of two bonded atoms. This is because the two atoms with the covalent bond can be viewed as approximately rigid. Possible causes of this alert include: occurrence of site disorder where different atomic species are present at the same site, non-crystallographic symmetry in the molecule itself, and overlapping of the entire molecule with a rotated or inverted structure. It is also necessary to check for mistaken atom placement. Temperature factors are sensitive to effects such as measurement error and correction, and thus one should check whether proper absorption correction has been employed. This alert occurs frequently even in structures based on fairly high precision data, and thus some questioned whether the Hirshfeld test is too sensitive, but the basic approach is first to examine the structure while taking into account the above points. If the cause cannot be specified, the alert can be avoided by using DELU instruction of SHELXL.

4.5. Alerts relating to the form of anisotropic temperature factors

Atom C95 has ADP max/min Ratio ... 3.40 prola

This alert indicates unusual elongation of the anisotropic temperature factor in a specific direction. In this example, the temperature factor of C95 in the end methyl group is elongated (Fig. 3). This is a classic sign of an disordered structure. It is handled by introducing a disorder model. The modeling method for disordered structure in SHELXL is given here as an example. C94 to C95 are the atoms in the original structure. C294 to C295 are newly introduced atoms. They are linked with FVAR so that the sum of the occupancies becomes 1.0. In addition, the temperature factors are restrained by inserting DELU and SIMU instructions.

```
SIMU C94 O28 C95 C294 O228 C295
DELU C94 O28 C95 C294 O228 C295
FVAR 0.9956 0.5
PART 1
C94··· 21.0
O28··· 21.0
C95··· 21.0
PART 2
SAME C94 > C95
C294 ··· -21.0
O228 ··· -21.0
C295 ··· -21.0
PART 0
```

Fig. 3. Causes and handling of an alert relating to the form of anisotropic temperature factors "Atom C95 has ADP max/min Ratio ... 3.40 prola".

5. Method of writing in vrf

When it is difficult to eliminate an Alert level A, and if there is a legitimate reason, the alert can be avoided by writing reasons into the CIF as a vrf. There are two ways of inserting a vrf.

With the first method, the vrf is directly written in after a line including data_ in the CIF. When checkCIF/PLATON is executed, a vrf template for Alert level A is displayed at the end. This can be copied, and inserted after the data_ line. The reason should be described in the RESPONSE section.

In the second method, the vrf is written in two stages. Just like the above method, the vrf template shown in checkCIF/PLATON is inserted after the data_ line. Next, in this case, "see publ_section_exptl_refinement" is written in the RESPONSE section, and the reason is written in 'publ_section_exptl_refinement'.

Various examples of CIFs in which vrf's have been written are available on the IUCr website.

6. Let's get familiar with checkCIF/PLATON

Thus far, this article has introduced frequently occurring alerts and how to handle them. Since the number of possible alerts is about 400 for check.def and about 50 for data validation procedure, the total is about 450. It is impossible to memorize methods of handling each of these alerts. Thus the aim of the remaining portion of this article is to improve familiarity with the criteria and mechanisms by which CIF alerts occur, and thereby enable readers to handle a broader range of alerts themselves.

All the check items and thresholds of checkCIF/PLATON can be examined in the 'Details of checkCIF/PLATON tests' located at the bottom of the checkCIF/PLATON page[(2)]. If you click on 'Details of checkCIF/PLATON tests', a table listing full check items will appear (Fig. 4). If you look closely at the test names, you can see that they are divided into two types: those that start with PLAT, and those that do not. Threshold values of tests which start with PLAT are given in a PLATON file called check.def. Threshold values for test names other than PLAT are given in the data validation procedure. checkCIF/PLATON combines the content of these two types of checks.

These two tests are generally complementary, but there are some items included in both. One example is the Flack parameter used to determine the absolute structure (Table 1). From this table it is evident that, depending on the situation, there may be cases where two or more messages are issued simultaneously in connection with the Flack parameter alone.

Here we will ascertain these by actually changing

Test name	Type	Purpose
ABSMU01	1	To check that _exptl_absorpt_coefficient_mu value is consistent with the cell contents.
ABSTY01	1	To check that _exptl_absorpt_correction_type is one of the recognised keywords.
PLAT030	1	Check _diffrn_reflns_number > reflns_number_total
PLAT031	4	Check need for Extinction Correction Parameter
PLAT032	4	Check su Flack Parameter
PLAT033	4	Check Flack Parameter value
PLAT034	1	Check for Flack parameter value specified Z>Si, non-centro
PLAT035	1	Check for _chemical_absolute_configuration

Fig. 4. Table listing full checkCIF/PLATON check items.

Table 1. Overview of judgment criteria and alert levels relating to the Flack parameter.

	Data validation procesures (checkCIF)				check. def (PLATON)			
Test name	SRTVA01				PLAT033		PLAT032	
Parameter	Flack			S.U.	Flack		S.U.	
Thresholds	− 0.2	0.3	0.7	0.5	0.3	10.0	0.2	0.10
Alert type	2	2	2	4	2	2	2	4
Alert level	C	C	C	C	C	A	C	A
Message	Flack parameter is too small.	Flack test results are ambiguous.	Chirality of atom sites is inverted?	Flack test results are meaningless.	Flack Parameter Value Deviates from Zero.		Std. Uncertainty in Flack Parameter too High.	

the value of the CIF Flack parameter. First, we edit the value of refine_ls_abs_structure_Flack of CIF, and try entering the value 0.40.

STRVA01 is a test stipulated in the data validation procedure. The alert level is C, and the alert type is 4 (Fig. 5). The message indicates that the absolute structure is not determined by the Flack parameter. In addition, PLAT033 is a test name stipulated in check.def, and the alert indicates that the Flack parameter deviates significantly from zero. The alert level is C, and the alert type is 4. In this way, multiple alerts are sometimes issued by two tests for one parameter.

Here will try to make the checking system issue more warnings by entering 10.1 for the Flack parameter, and 0.6 for the reliability 'su'. An alert level A should occur. This is an actual alert (Fig. 6).

Just as before, STRVA01 is a test item stipulated in the data validation procedure. PLAT033 is a test result stipulated in check.def. What is different from before is that the alert level has changed to A. PLAT024 indicates that there is no anomalous dispersion signal. In addition, PLAT032 is an alert indicating that the standard deviation 'su' is large. PLAT024 and PLAT032 are also stipulated in check.def.

Alerts from checkCIF/PLATON are stipulated in the data validation procedure and check.def, and thus if you keep these two files on hand, it will likely be extremely useful for dealing with alerts.

In the data validation procedure, if you click on a test name other than PLAT at the page on the details of the checkCIF/PLATON test, the window for the corresponding test will open. At the bottom of this window is an item "Full list of validation algorithms." If you click here, all of the test items can be seen at once. check.def can be obtained at the PLATON website[5]. check.def is frequently revised.

7. Conclusion

Today, CIFs are indispensable for transferring crystal structure information. Crystallographic journals now receive submissions in CIF form, and checking using checkCIF/PLATON is mandatory at a level appropriate to the journal being submitted to. Even in cases where CIF checking is unnecessary, it is extremely useful as a tool for checking whether there are any errors in analysis results.

Checking with checkCIF/PLATON is done using the items and threshold values listed in check.def and the data validation procedure. If you print out each of those files and keep them on hand, it will deepen your

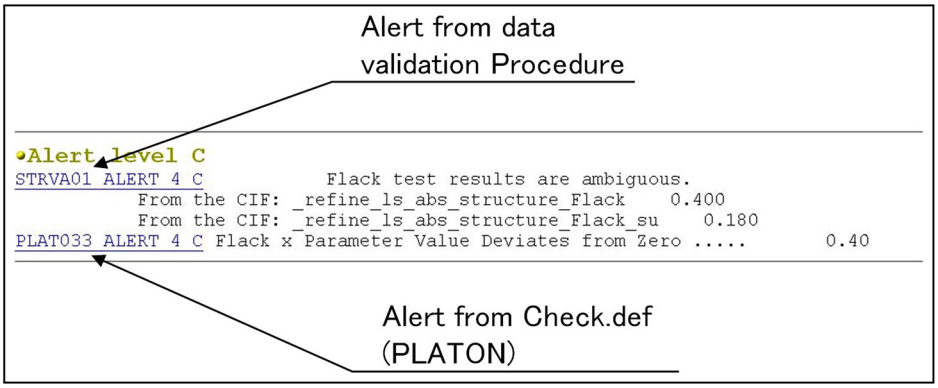

Fig. 5. Output when the value of _refine_ls_abs_structure_Flack was set to 0.40.

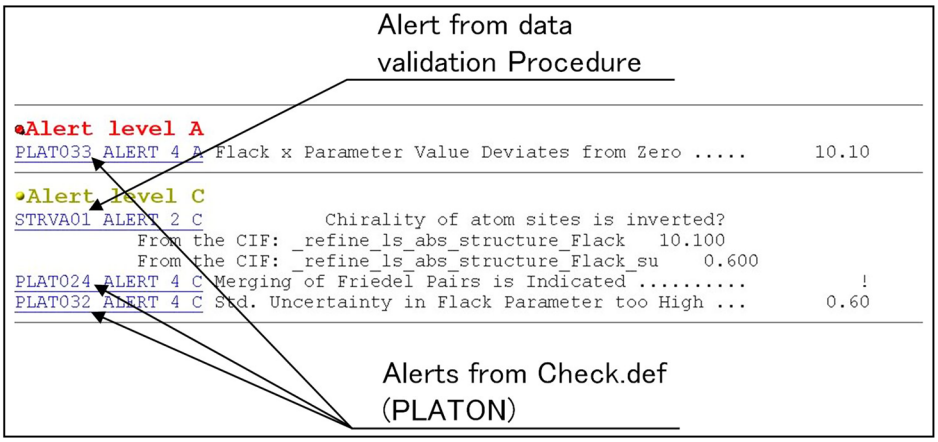

Fig. 6. Output when the value of _refine_ls_abs_structure_Flack was set to 10.1 and reliability was set to 0.6.

understanding and make it easier to respond to alerts. Numerous hints for responding to alerts are given on the checkCIF/PLATON[2] page and the check results page.

References

(1) http://www.iucr.org/
(2) http://journals.iucr.org/services/cif/checking/checkfull.html
(3) S. R. Hall, F. H. Allen and I. D. Brown: *Acta Cryst.* **A 103** (1991), 655–685.
(4) A. L. Spek: *Acta Cryst.* **D65** (2009), 148–155.
(5) http://www.cryst.chem.uu.nl/spek/platon/