# An overview of PLATON/PLUTON crystal structure validation

**Anthony L Spek**, Utrecht University, Utrecht, The Netherlands

## 1 Introduction

Crystal structure studies are often essential as proof for new reported inorganic and metal-organic chemistry. Nowadays, single crystal X-ray crystallography, the most used analysis technique for that, is often conceived as being routine, highly automated and reliable. For that reason, the main result of such a study is often just summarized in the printed part of a non-crystallographic journal with an ORTEP illustration showing the three-dimensional molecular geometry without presenting the supporting numerical details: *Seeing is Believing*. Fortunately, such a figure gives also, apart from the three-dimensional molecular geometry, an overall impression of the quality of the structure determination and may point experts at possible unresolved problems. Issues with the structure model and its refinement often show up in such a picture as unusual Atomic Displacement Parameter (ADP) ellipsoidal shapes. The full analysis details are expected to be made available as supplementary material to the printed paper, mostly with a reference to the archived CIF in the Cambridge Structural Database (CSD).[1] At least one of the referees of such a paper is expected to also inspect that material for validity and completeness. This applies in particular when unusual structural results are claimed to be sure that those results are not based on a misinterpretation of the experimental data or just refinement artifacts. In addition, future readers should be able to make use of the deposited data to allow them to repeat or improve on the analysis or to use the experimental data for an alternative or more detailed investigation beyond the original purpose of the study. The experimental data might also be unique or not easily obtainable again from scratch. Some interesting metastable polymorphs are good examples for that since they cannot always be reproduced.

Validation of the result of a crystal structure determination is not trivial. Expertise is often required to evaluate whether the available data support the authors analysis and claims. Various types of avoidable pitfalls that may invalidate a claimed result need to be recognized. The available refereeing expertise is limited for the handling of the exploding number of reports of routinely

obtained crystal structures determined by insufficiently trained analysts using mainly GUI driven black box software. Unfortunately, also the relatively rare cases of faked structure reports should be detected.[2]

The time-consuming and proper validation of structure reports became problematic in the early 1990s due to their increasing numbers. New software requiring less user input made structure solution and refinement easier. At that time, the results were still mainly available only in printed form, often incomplete and with typos introduced with the preparation of the manuscript. Using those data for additional calculations and archival in databases such as the CSD required retyping again and was time consuming. Newer data collection hardware started to generate an explosion of new structure reports, making their proper processing soon unmanageable.

One of the problems was that the various refinement program systems had their own I/O formats. Of those, a still popular and surviving exchange format for selected refinement results is the free format RES file of the SHELXL program[3] that is often used for data exchange with molecular graphics programs. The same applies for the fixed format HKL file containing the reflection data on which the refinement model is based and used by other refinement programs. The information in those files is not complete.

Early attempts to standardize electronic data exchange involving a fixed formatted computer readable archival style, based on the 80 column IBM punched card model, were not widely adopted. The solution for the data exchange and archival issue of the crystallographic results was eventually found in the creation of the flexible computer readable Crystallographic Information Framework (CIF) file format.[4]

The free format CIF file is flexible with a data-name keyword and associated data-value structure. This standard was pioneered by the International Union of Crystallography (IUCr) with their Acta Crystallographica section B, C and E journals. Syd Hall, co-author of a at that time popular structure determination package, XTAL,[5] and section editor of Acta Crystallographica, section C, was very influential in pushing CIF as a data exchange and archival standard. He managed to convince George Sheldrick to be one of the first adopters of this standard in his still today widely used SHELXL refinement program, either in its native form or as part of free software packages such as OLEX2,[6] WINGX[7] and commercial software packages that are provided with the diffractometer hardware. CIF style crystal structure data deposition is now adopted as a standard requirement by all major journals. The CSD also requires CIF as deposition standard. Most current crystallographic programs can read and/or write CIF formatted files.

The CIF standard also opened the way to the publication of structure reports where both the manuscript and the relevant data are submitted electronically as a single file. The IUCr journals Acta Cryst. C and E were among the early adopters. It was their answer to the exponential growth of manuscripts received by the journals reporting crystal structure reports due to advances in data collection hardware, structure solution software and computing facilities.

In view of the large increase in the number of received manuscripts, often reporting routine structure reports, it was also clear that the classical refereeing process was inadequate. Automated validation of the CIF data was therefore introduced to facilitate that process. Initially, that involved tests that checked for the completeness of the data and their internal consistency. Subsequently, more detailed content related tests as supplied by the PLATON program[8] were included in the IUCr/checkCIF server (https://checkcif.iucr.org) based report. Currently several hundreds of tests have been implemented and their result collected in the form of a set of ALERT messages with four levels of potential relevance. This validation report is easily available for the structure analyst, authors, referees and reader. Details about IUCr/checkCIF can be found on the IUCr journals WEB-site (https://journals.iucr.org). This chapter will give an overview of the multiple tests done by the program PLATON (http://platonsoft.nl). In addition, several PLATON tools will be discussed that can be used to investigate ALERTed issues in more detail.

## 2　Crystal structure determination

A crystal Structure determination can be divided into several stages: data collection, data reduction, solution of the phase problem, structure refinement, analysis of the 3D structure, graphical presentation of the result and validation. Though very important for the validity of a structure report, the first three are currently only marginally validated. Work is going on for standardized archival of the primary experimental data, i.e., the diffraction images.[9] Archived diffraction images may be useful to resolve problems encountered with a structure determination by looking for details in the diffraction images that are not accounted for as part of the standard block-box image processing and data reduction.

### 2.1　Data collection and data reduction

Today, most 3D structure determinations of organic and metal-organic compounds are based on the collection of diffraction data on a single crystal using MoK$\alpha$, CuK$\alpha$ or synchrotron X-ray radiation and 2D diffraction images produced by 2D detectors. Alternatives are the less routine neutron diffraction and electron diffraction. The diffraction images are processed, based on the diffraction spots in those images, into information about the translation lattice (cell dimensions, space group) and the intensity of the indexed diffraction spots, to be used in the subsequent structure determination. In that data-reduction process, valuable information may be lost such as diffraction spots that do not fit for various reasons in the assumed translation lattice or the diffuse scattering and streaks in between diffraction spots. The latter may provide information about multiple types of structural disorder and twinning. It is important to be able to go back to the primary experimental data, i.e., the diffraction images, to try to resolve unexplained problems with a structure determination such as twinning or poor structure refinement issues.

## 2.2    Solution of the phase problem

Crystal structure determination in essence amounts to obtaining a three-dimensional electron density map of the unit cell content from the experimental set of diffraction spot intensities. Such a map can be analyzed in terms of isolated atomic densities from which the three-dimensional coordinates of their centers can be extracted and used, after refinement, for molecular geometry calculations such as bond distances and angles and for illustrations of the crystal structure.

The electron density map can be calculated with a Fourier synthesis, based on the amplitudes and phases of the reflections as coefficients. The amplitudes are easily derived as proportional to the square root of the intensity of the diffraction spots. Unfortunately, the corresponding phases are lost in the experiment. However, as it turns out, in most cases approximate phases can be recovered from the set of observed intensity data, given that the number of observations is usually much larger than the number of atomic parameter values to be determined. Those preliminary phases are subsequently improved iteratively in the refinement stage. Early approaches to the phase recovery issue relied on the introduction of a heavy atom into the molecule to be studied, when not already present next to the otherwise light atoms, or co-crystallized with a molecule containing a heavy atom. Subsequently, statistical methods were developed (symbolic addition, tangent formula) that no longer needed the introduction of heavy atoms in the compound to be investigated. Those methods are today again mostly superseded by even more powerful black-box techniques such as the charge flipping algorithm[10] or the intelligent brute force procedure as implemented in the program SHELXT.[11]

## 2.3    Structure refinement

Most of the reported structures are today refined into a final set of atomic parameters using the least-squares program SHELXL, either in its native form or as part of a structure solution package [WINGX, OLEX2]. Alternative software packages such as CRYSTALS,[12] JANA2006[13] and OLEX2 include refinement options not available in SHELXL.

Traditional refinement programs such as SHELXL use the AIM (Atom-in-Molecule) model where the electron density map is approximated as a collection of spherical atomic densities with associated anisotropic displacement parameters. Such a model is usually sufficient for the purpose of most studies. However, they ignore the bonding and lone pair effects that will show up as residual density peaks in difference electron density maps, in particular in case of refinement with high quality and high-resolution diffraction data. More involved refinement techniques, e.g., NoSpherA2,[14] and IDEAL,[15] that use aspherical scattering models (involving quantum chemical calculations) will not be discussed here. The same applies to incommensurate structures that can be modeled and refined with JANA2006.

## 2.4    Analysis of the results, illustrations and validation

Multiple programs are available for the calculation of a variety of derived geometry parameters such as bond distances and angles and molecular illustrations. This chapter will concentrate on those available in the program PLATON, some of which are also used as part of the checkCIF structure validation.

## 3    The programPLATON

PLATON[16−18] is a computer program that has been developed and extended over a period of more than 40 years since 1980 and is used in the context of the National Service Facility for single crystal structure determinations in the Netherlands. It grew out of a geometry analysis tool by adding multiple new options that were found to be useful for our service facility or suggested by outside users along with its involvement in the IUCr/checkCIF project. Some of those tools are unique such as TwinRotMat, others such as ADDSYM [MISSYM[19]] and ORTEP[20] are adaptations and extensions of pre-existing programs that we found useful to include. PLATON can be seen as a collection of knowledge and experience assembled over more than 50 years in this field. As a single program, PLATON is designed to be as much as possible to be independent from external libraries. It is available on the three common computer platforms: LINUX, MacOS and MS-WINDOWS. PLATON is developed and updated on the current FORTRAN platform that is also used by other widely used programs such as SHELXL and SHELXT.

The central subject of this overview chapter is the validation tool checkCIF in PLATON. That tool makes use of a selection of the other available tools in PLATON. PLATON/checkCIF creates a validation report in the form of a set of so-called ALERTS, short messages that need to be further investigated and possibly acted upon. Also there, PLATON tools can be helpful for that task.

## 3.1    PLATON tools and functions

Fig. 1 shows the opening window of PLATON when it is invoked with a CIF file, in this example the file yk2161.cif. Alternatively, it can be invoked with a <name>.ins or <name>.res file, being the standard input or output files from the SHELXL refinement program. The program attempts to automatically search for an associated reflection file such as <name>.fcf or <name>.hkl. The center of the PLATON overview window displays a tableau showing the available tools and functions that can be invoked by left-clicking on their respective names. Right-clicking on an item will offer info, downloaded from the internet, on that item in a new
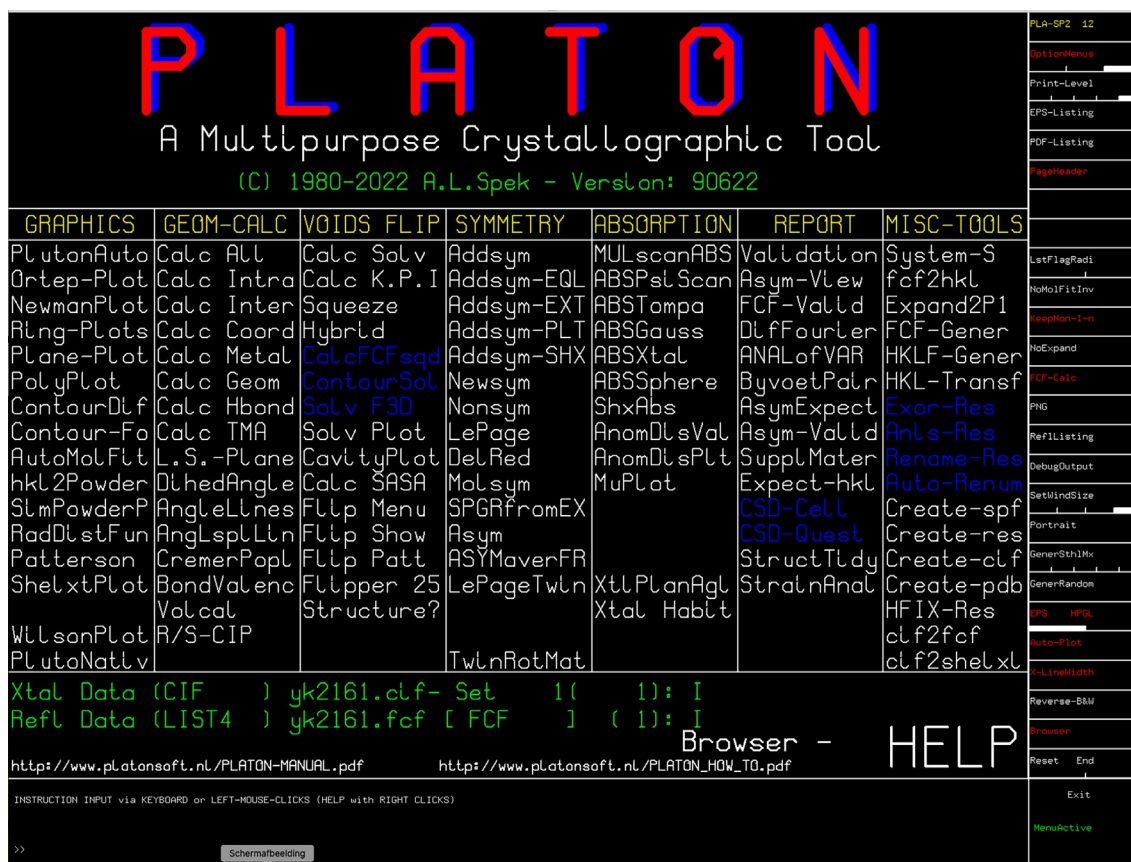
**Fig. 1**  Opening window of the PLATON program with in the central box an overview of the clickable tools (Blue when currently not applicable). The side menu gives access to various program options. The lower section lists the active data files. The box at the bottom provides an option for keyboard input of instructions.

window. As an example, left-clicking on VALIDATION, [1, 6], will produce a validation report for data set <name>.cif and optionally <name>.fcf in the report files <name>.chk and <name>.ckf respectively. Note that the notation [1, 6] indicates the position row 1, column 6 in the tools tableau. The validation tool invokes in the background several of the other tools for its report. Examples are ADDSYM, [1, 4], Calc Solv, [1, 3], TwinRotMat [17, 4] and BijvoetPair, [6, 6]. Those tools can also be used directly to investigate ALERTed for issues in more detail.

On the right of the tool tableau there is a left clickable submenu with options and features such as listing symmetry details, symm, (10,13,2). Note that the latter code indicates its click position in submenu 10, vertical option 13, horizontal box 2. Most tools have one or more sub-menus. The main PLATON tool has three submenus [Fig. 1, (12,1,3)]. Below the tools tableau there is a box where information is displayed about the associated data files and some help pointers to more information. Instructions to the program can be given either by clicking on menu options or by keyboard data entry in the bottom window box after "≫."

### 3.1.1   CALC ALL

A single click on CALC ALL [1, 2] will create an extensive listing file with a large range of detailed geometrical information such a bond distances, bond angles, torsion angles, least-squares planes, dihedral angles, ring puckering parameters, coordination geometry, intermolecular contacts including hydrogen bonds and many other descriptors such as tentative valence and chirality assignments. CALC INTRA [2, 2], CALC INTER [3, 2], CALC COORD [4, 2], CALC METAL [5, 2], CALC METAL [6, 2], HBOND [7, 2] and CALC TMA [8, 2] will create subset listings.

The graphical tool L.S.-PLANE [9, 2] can be used to calculate the least-squares plane though a set of atoms along with their deviations from that plane by clicking on the partaking atoms in the graphical molecular display. Similarly dihedral angles [DihedAngle [10, 2]], the angle between lines (or bonds) [AngleLines [11, 2]] and the angle between a line and a plane [AngLsplLin [12, 2]] can be calculated interactively.

### 3.1.2   PLUTON

PLUTON is a molecular graphical tool that visualizes the 3D connectivity of the atom sites as derived from the electron density map with associated atom types and atom labels. By default, atoms are represented by balls and bonds by sticks in a minimum overlap orientation. Alternatively, space-filling models can be created. There are two clickable options to invoke this tool. PlutoNative [17, 1] to start building the content and style of the molecular display from scratch and PlutonAuto [1, 1] for an automatic preliminary

display that can also be extended with additional interactively supplied instructions. Fig. 2 gives an example of a unit-cell content illustration in space group P-1. PLUTON can be used to investigate polymeric and hydrogen bond networks and as interface to external programs such as POVRAY for ray-traced images or RASMOL for dynamic rotations of the structure.

### 3.1.3   ORTEP

ORTEP is a molecular graphics tool [2, 1] similar to PLUTON with the difference that not only the position and connectivity of the atoms are shown but also the Atomic Displacement Parameters (ADPs). Atoms are represented with an ellipsoidal probability surface drawn at a given probability level (usually 50%). Those ellipsoids do not represent the electron density but the refined atomic thermal motion parameters that can be convoluted by systematic factors such as unaccounted for disorder or unsuitable correction for absorption. Fig. 3 gives an example of a 50% probability ellipsoid plot. Atoms with severely elongated or flattened ellipsoids or with a size that deviates significantly from those of neighboring atoms may indicate model and/or refinement issues to be investigated. The components of the ellipsoids along a bond are expected to have similar values.[21]

### 3.1.4   CONTOUR

Fourier transformation of the square root of the experimental diffraction intensity data, $F_{obs}(hkl)$, together with associated phases obtained with one of the available methods to recover them from the same intensity data results in a 3D electron density map. Such
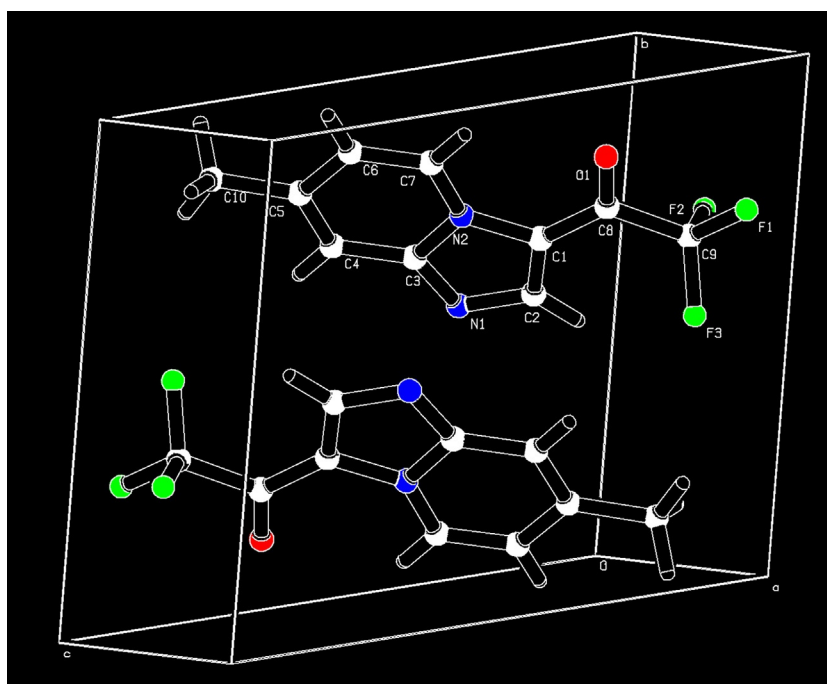


**Fig. 2**   A PLUTON style ball-and-stick packing plot of the content of the triclinic unit-cell.
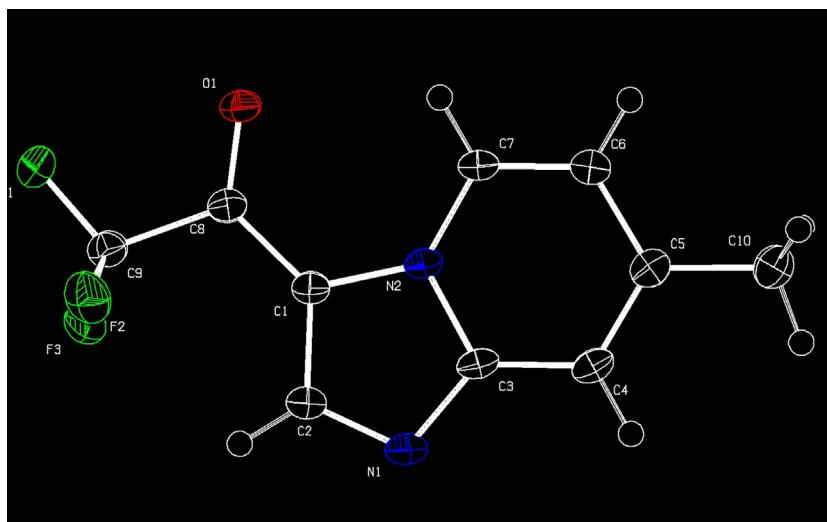


**Fig. 3**   A labeled ORTEP plot with ellipsoids drawn at the 50% probability level.
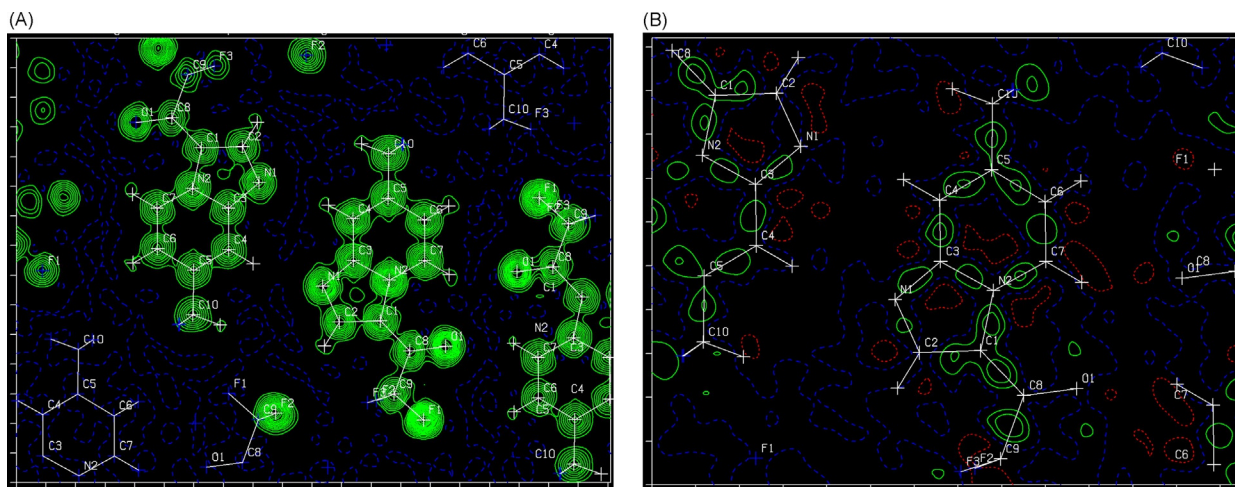
**Fig. 4** (a) A contoured electron density map section with contour levels 1.0 e/Å$^3$ apart. (b) Difference electron density map showing residual densities on bonds that are not accommodated by the AIM model. Contour levels 0.1 e/Å$^3$ apart.

a map represents the electron density averaged over all unit cells in the crystal, i.e., space and time averaged. CONTOUR is a tool for displaying density level contoured sections of that map. The CONTOUR-Fo option [8, 1] displays a section of the $F_{obs}$ map (Fig. 4a) defined by the 6-membered ring atoms in the structure shown in Fig. 3. For validation purposes, the related CONTOUR-DIF option [7, 1] can be used to create a density map that displays the difference electron density between a map calculated with observed data and a map calculated using the derived model parameters (Fig. 4b). Nonzero densities in such a map may indicate several types of problems such as misplaced hydrogen atoms, missing hydrogen atoms, wrong element type assignment, unresolved twinning and inadequate correction for absorption. The residual density on the bonds in Fig. 4b is due to refinement based on the AIM refinement model.

### 3.1.5    Simulated powder pattern
Fig. 5 shows a simulated powder pattern for the compound illustrated in Fig. 3. Such a pattern can be calculated based on the experimental reflection data with hkl2Powder [10, 1] or based on reflection data as calculated with the structure model parameters
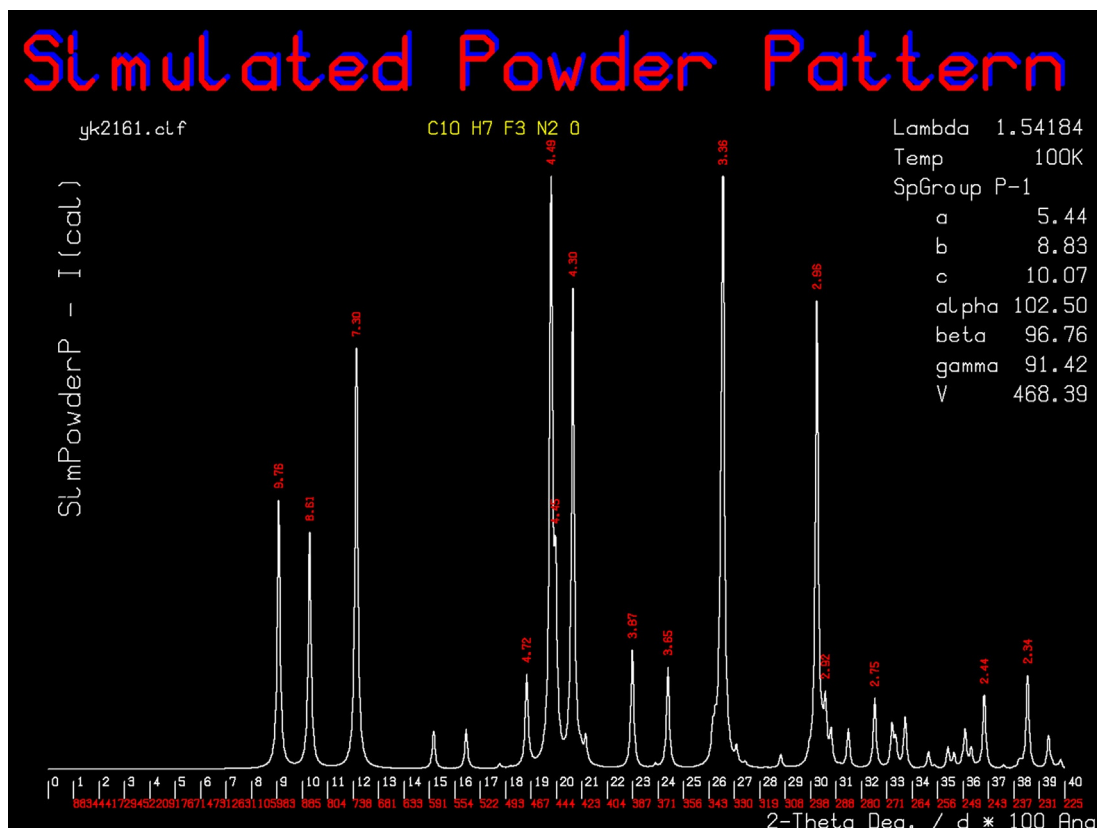


**Fig. 5** A Simulated Powder Pattern based on intensities calculated from the structure model parameters.

with SimPowderP [11, 1]. This tool can be useful for checking whether two structures are identical but described in different settings or space groups or compared with experimental powder patterns.

### 3.1.6   LEPAGE, DELRED and ADDSYM

The assignment of the proper space group to a crystal structure is not always obvious. A preliminary structure may have been obtained (or obtainable only) in a lower symmetry space group then the actual higher symmetry space group. In that case, atoms that are related by symmetry are refined as independent. This might lead to refinement artifacts and distorted geometry, in particular when the missing symmetry elements are an inversion center or a lattice translation.

The LEPAGE [CREDUC[22]] or DELRED [DELOS[23]] tools can be used to investigate the lattice symmetry of a supplied unit cell parameter set for higher lattice symmetry within applied tolerances. The actual content of the unit cell will determine whether this is the actual symmetry or just accidental. Both tools are inspired by the respective tools in square brackets.

ADDSYM is a tool for checking the correctness of the preliminary space group assignment by inspection of the structure model coordinate set for additional symmetry elements. Invoking the ADDSYM option [1, 4] will report on the possible additional symmetry elements along with a proposed revised space group. The related option ADDSYM-PLT [4, 4] will show an updated model when applicable and the ADDSYM-SHX [5, 4] option will create an averaged coordinate set suitable for a test refinement in the proposed new space group. Comparison of the old and new refinement results should determine which one provides the best description of the structure. The reported higher symmetry elements can be either only approximate (pseudo-symmetry) or to be implemented, depending on the applied distance tolerances and experimental error. Poor data quality, disorder, pseudo-symmetry and twinning may complicate the analysis. In those cases, all possibilities should be investigated and the best one reported with an associated discussion.

Example 1: The structure shown in Fig. 6 with CSD code BAMYEU was originally reported in monoclinic non-centrosymmetric space group Cc. Interestingly, the associated publication included an ORTEP illustration in approximately the same orientation as in Fig. 6, clearly suggesting an additional threefold symmetry axis. ADDSYM indeed points out (Fig. 7) that the real space group is P31c in the trigonal system (Fig. 8). A correction was published by Dick Marsh[24] who has also corrected hundreds of similar published cases of missed higher symmetry. Adding a threefold axis does not change the structure into centrosymmetric thus no drastic changes in the molecular geometry are to be expected for this structure other than averaging for systematic errors in the reflection data.

Example 2: The structure with CSD code EKOKOE was published with two independent molecules in the non-centrosymmetric space group P1. Inspection of Fig. 9 and ADDSYM clearly suggest an additional center of inversion. Fitting both molecules with
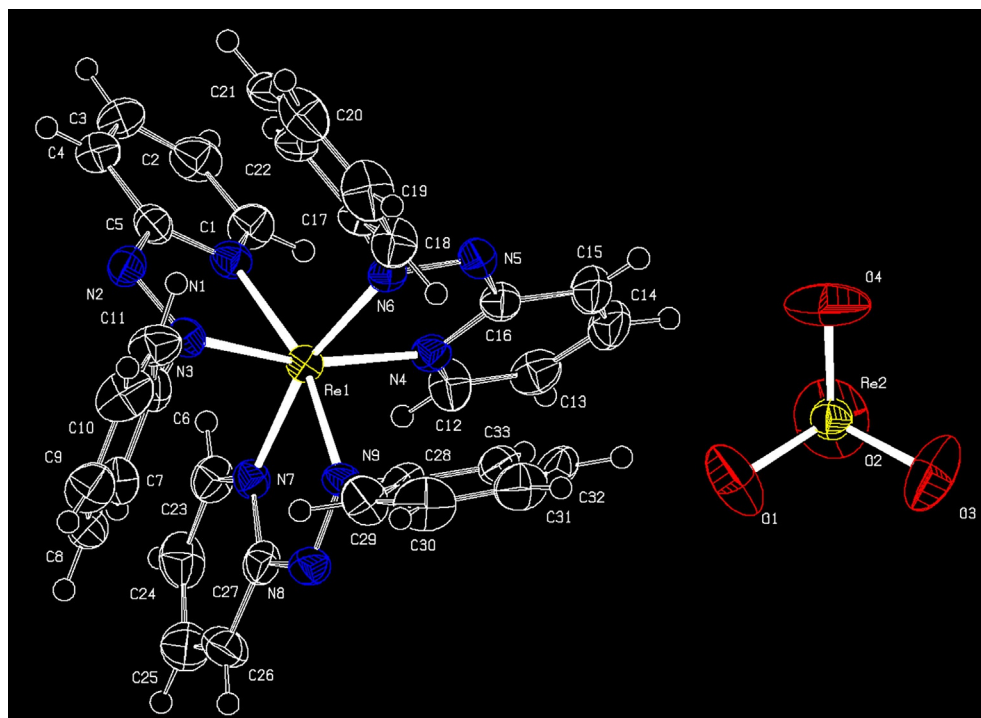


**Fig. 6**   Structure of CSD entry BAMYEU published in the monoclinic space group Cc that suggests threefold axes as additional symmetry elements of both the cation and the anion.

**Fig. 7** ADDSYM suggests the higher trigonal P31c space group symmetry based on the detection of three more symmetry elements. The proposed additional symmetry elements are shown in red. The transformation matrix from the C-centered to P-trigonal lattice is also shown. The new symmetry applies to 100% of the atoms, subject to the distance and angle tolerances shown in green.
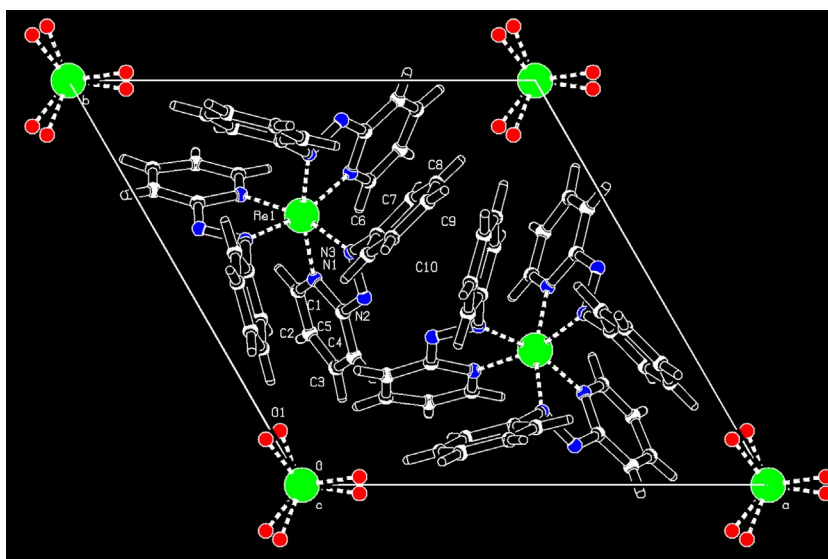


**Fig. 8** Packing diagram of the unit cell content in the trigonal space group P31c with all molecular species on threefold axial sites. The symmetry independent atoms are labeled.

the AUTOMOLFIT [9, 1] routine in PLATON shows a good configurational fit (Fig. 10). A description in the centrosymmetric space group P-1 is indicated[25] but should be investigated with a refinement in that space group for a definite proof. Unfortunately, no reflection data are available for that. Refinement of a centrosymmetric structure in a non-centrosymmetric space group usually leads to refinement artifacts due to numerical instability. This is clearly shown in Fig. 11 where corresponding bond distances are compared. As an example, C10-C11 and C10′-C11′ differ significantly by 12.1 s.u. units. Their average is near the expected value for such a bond. Another signal is the observation that corresponding ADP ellipsoids tend to be perpendicular due to the numerical instability.

Example 3: See Section 4.2. That example shows that additional symmetry proposed by ADDSYM, representing the (approximate) symmetry in the coordinate set of the refined model, should always be checked with the symmetry in the reflection data set (i.e., systematic absences and statistical tests for an inversion center). The latter can be done with the SPGRfromEx [11, 4] tool. In addition, there is the Newsym [6, 4] tool to investigate the reflection data symmetry of reflection data calculated from the refined model parameters, which may differ from the symmetry of the observed reflection set. A detailed statistical analysis for inversion symmetry is available with the WilsonPlot [17, 1] tool.
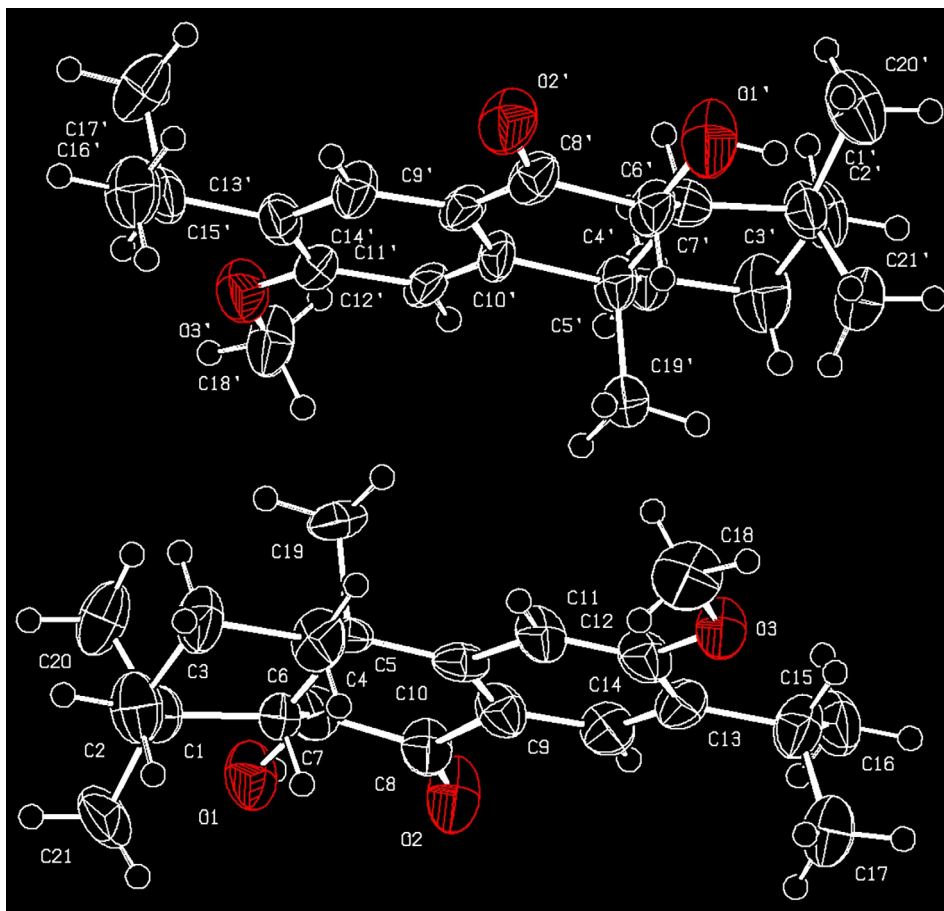
**Fig. 9**   Displacement ellipsoid plot of the two independent molecules of CSD entry EKOKOE published in space group P1. Corresponding ADP ellipsoids tend to be perpendicular (e.g., C10 and C10′) suggest inversion symmetry and space group P-1.
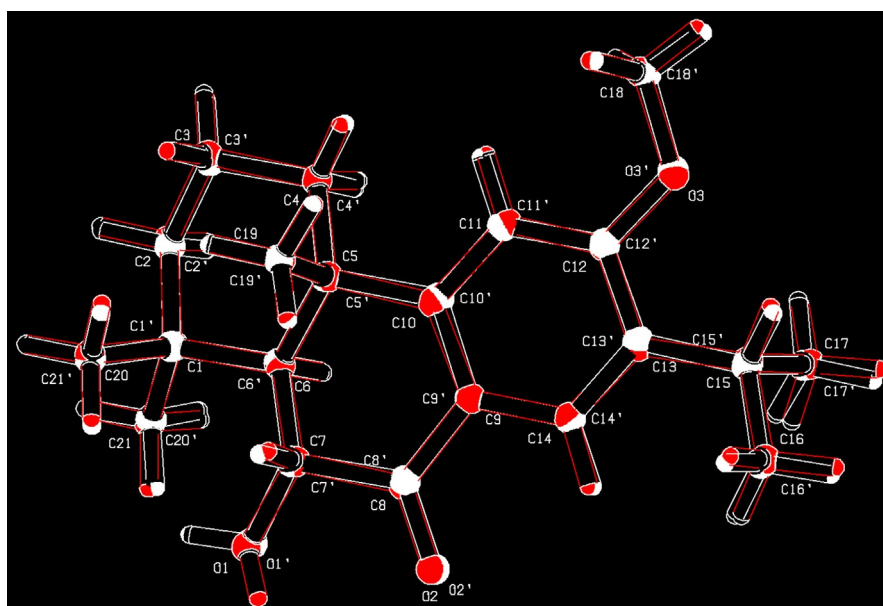


**Fig. 10**   Quaternion based fitting of the two independent molecules of EKOKOE shows a close configurational consistency.

| Resd#1 | Resd#2 | Dist#1 | Dist#2 | Dif | Dif/Sig |
|--------|--------|--------|--------|------|---------|
| O1 –C7 | O1' –C7' | 1.435(10) | 1.384(11) | 0.051 | 3.431 |
| O2 –C8 | O2' –C8' | 1.193(11) | 1.248(10) | -0.055 | -3.700 |
| O3 –C12 | O3' –C12' | 1.357(10) | 1.353(9) | 0.004 | 0.297 |
| O3 –C18 | O3' –C18' | 1.453(12) | 1.352(13) | 0.101 | 5.709 |
| C1 –C2 | C1' –C2' | 1.561(13) | 1.498(14) | 0.063 | 3.298 |
| C1 –C6 | C1' –C6' | 1.582(9) | 1.553(10) | 0.029 | 2.156 |
| C1 –C20 | C1' –C20' | 1.576(11) | 1.518(16) | 0.058 | 2.987 |
| C1 –C21 | C1' –C21' | 1.551(14) | 1.511(12) | 0.040 | 2.169 |
| C2 –C3 | C2' –C3' | 1.552(14) | 1.482(15) | 0.070 | 3.412 |
| C3 –C4 | C3' –C4' | 1.544(11) | 1.507(12) | 0.037 | 2.273 |
| C4 –C5 | C4' –C5' | 1.540(10) | 1.548(11) | -0.008 | -0.538 |
| C5 –C6 | C5' –C6' | 1.578(9) | 1.559(10) | 0.019 | 1.412 |
| C5 –C10 | C5' –C10' | 1.485(8) | 1.567(7) | -0.082 | -7.714 |
| C5 –C19 | C5' –C19' | 1.539(9) | 1.570(10) | -0.031 | -2.304 |
| C6 –C7 | C6' –C7' | 1.525(10) | 1.556(10) | -0.031 | -2.192 |
| C7 –C8 | C7' –C8' | 1.495(9) | 1.560(8) | -0.065 | -5.398 |
| C8 –C9 | C8' –C9' | 1.486(11) | 1.427(9) | 0.059 | 4.151 |
| C9 –C10 | C9' –C10' | 1.394(10) | 1.416(9) | -0.022 | -1.635 |
| C9 –C14 | C9' –C14' | 1.387(11) | 1.407(9) | -0.020 | -1.407 |
| C10–C11 | C10'–C11' | 1.464(9) | 1.318(8) | 0.146 | 12.125 |
| C11–C12 | C11'–C12' | 1.368(10) | 1.426(8) | -0.058 | -4.529 |
| C12–C13 | C12'–C13' | 1.427(11) | 1.389(9) | 0.038 | 2.674 |
| C13–C14 | C13'–C14' | 1.341(11) | 1.402(10) | -0.061 | -4.103 |
| C13–C15 | C13'–C15' | 1.517(10) | 1.524(9) | -0.007 | -0.520 |
| C15–C16 | C15'–C16' | 1.560(14) | 1.497(13) | 0.063 | 3.298 |
| C15–C17 | C15'–C17' | 1.515(14) | 1.550(12) | -0.035 | -1.898 |

**Fig. 11**   Comparison of corresponding bond distances in both molecules in EKOKOE as reported in space group P1. Bond distances such as those of C10-C11 and C10'-C11' differ up to 12 times the estimated error in the bond.

### 3.1.7   CALC SOLV

The crystal structure of a compound of interest often includes solvents of crystallization in voids left by the packing of the main molecules. Those solvent molecules are often disordered when there is limited interaction with the framework of the main molecules or might even have been partly evaporated. Disordered density in those voids is easily missed. Peak search algorithms that are used to identify atoms in an electron density map assume ellipsoidal atomic density shapes. Disordered solvents often do not satisfy that assumption. Missed solvent contributions to the structure model may lead to unsatisfactory refinement results. CALC SOLV reports on solvent accessible voids in a structure model. That information is used in the SQUEEZE tool (Section 3.1.8). CAVITYPLOT [9, 3] creates a crude image of those voids and their location in the structure (Fig. 12).

Even when there are no solvent accessible voids in a structure there is still space in pockets in between molecules. The volume of the molecules in the unit-cell is usually characterized with the value of the Kitaigorodskii[26] packing index. That value can be calculated with the related CALC K.P.I instruction and is based on the volume taken by the molecules in the unit cell with van der Waals radii assigned to the atoms. Typical packing indices are in the order of 0.65.

### 3.1.8   SQUEEZE

It is not always possible to model the content of voids meaningfully with a discrete set of parameters due to severe disorder, in particular when the nature of the solvent is unknown, a mixture of solvents or even an impurity with its origin in earlier synthesis steps. Channels that are incommensurately filled with molecules are difficult to model meaningfully with a disorder model. Ignoring their contribution will result in higher R-values and low-quality structural parameter values of the molecules of interest. The SQUEEZE [3, 3] tool provides an alternative method to handle the scattering contribution of disordered solvents in a crystal structure to the total scattering of the structure model in the least-squares refinement. The geometry of the molecules of interest and R-values generally improve with the application of SQUEEZE.
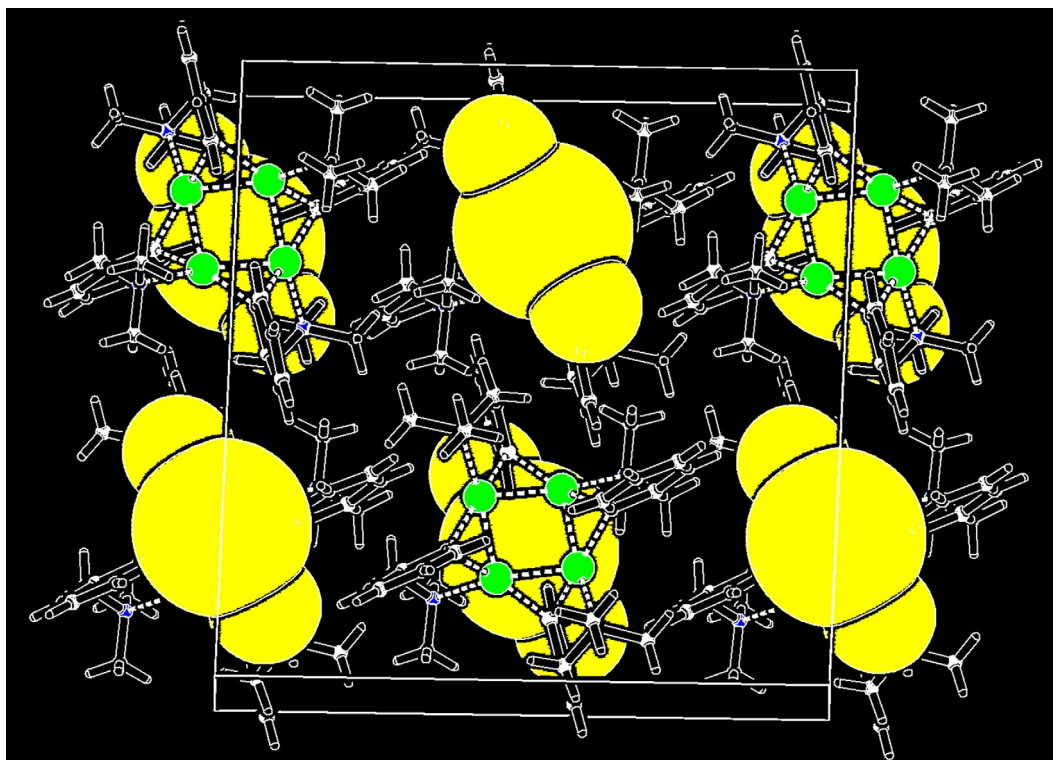
**Fig. 12** Visualization of the solvent accessible voids in the unit-cell. Voids are represented by yellow spheres of variable radius.

Preliminary, the Calc Solv [1, 3] option can be invoked to investigate whether a structure contains solvent accessible volumes. When that is the case, the SQUEEZE tool can be used to find out how much density can be found in those voids. SQUEEZE generates a new set of instruction files suitable for implementing the solvent contribution to the calculated structure factors with the final refinement using SHELXL. More details about SQUEEZE and examples can be found in Ref. 27.

### 3.1.9   TWINROTMAT

Twinned crystals can often be detected at the experimental stage, either by inspection of the crystal under polarized light or by inspection of the diffraction images, but not always. When twinned, the main diffraction lattice is overlayed with one or more similar lattices in a different orientation. The different lattices may be completely overlapping or only partial. In the latter case it is often possible to deconvolute the data for a preliminary structure determination. In many other cases, it is still possible to preliminary solve such a structure but with unsatisfactory high R-values and significant residual density peaks in the difference density map. TwinRotMat [17, 4] is a tool to detect a twinned structure in such a dataset and to provide information about the type of twinning model needed for a proper final refinement. Fig. 13 shows the result of such an analysis. A twofold rotation about the a-axis is detected causing a 100% lattice overlap resulting in a close to 50:50 racemic mixture. Taking the reported rotation matrix
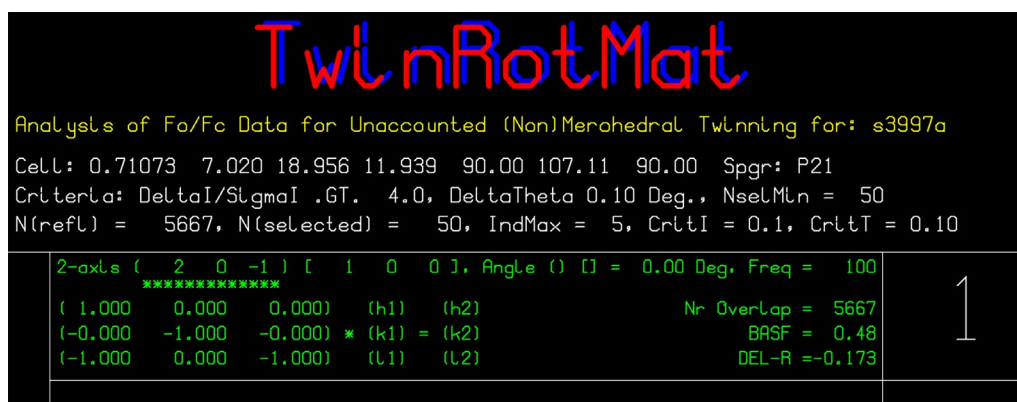


**Fig. 13** Example output of TwinRotMat reporting a twinning operation and its estimated reduction of the R-factor when applied.
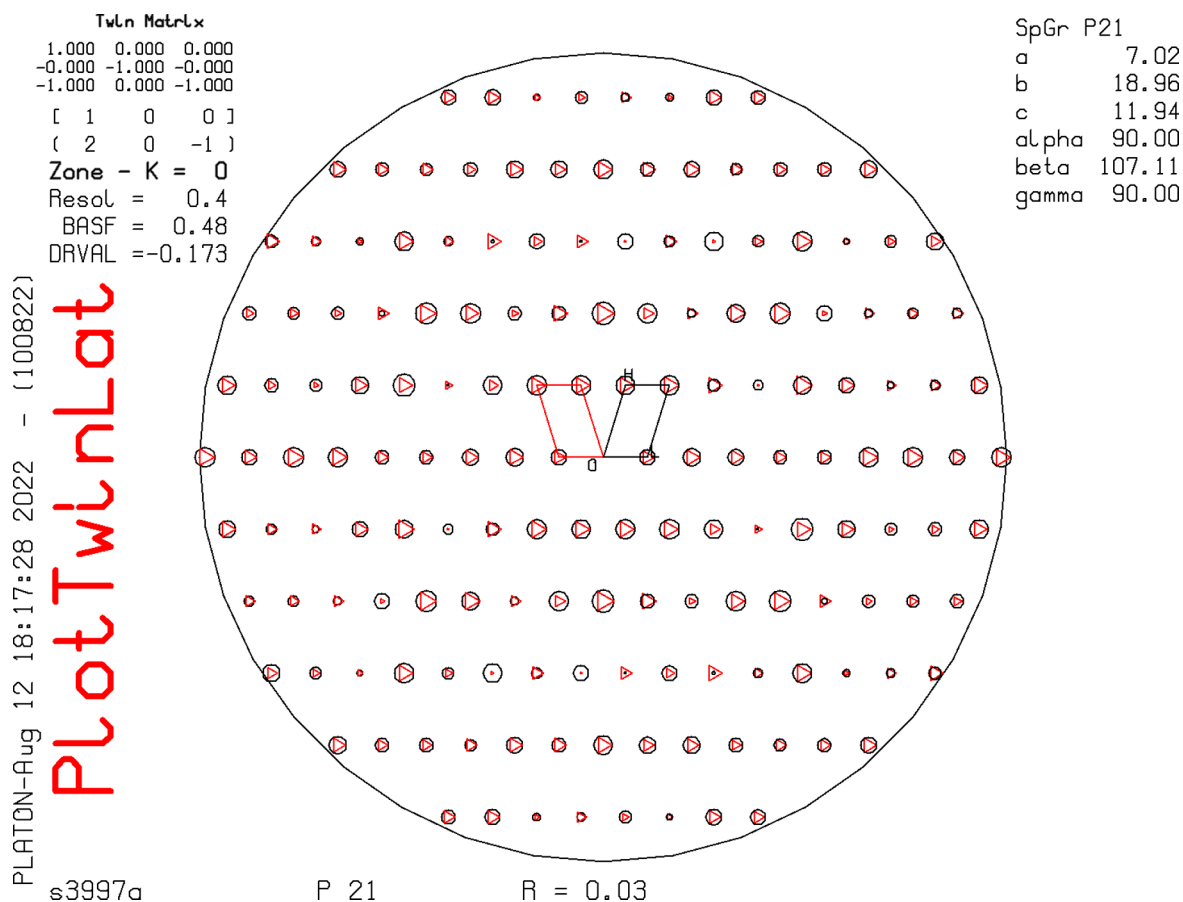
**Fig. 14**    Display of the two overlapping reciprocal lattices in the h0l section. The twin operation is a twofold rotation about the vertical a-axis.

into account in the subsequent least-squares refinement is expected to lower the R-value by in the order of 17%. Fig. 14 illustrates the lattice overlap. See also Fig. 2 in Ref. 28.

### 3.1.10    ASYM-VIEW

ASYM-VIEW [2, 6] is a graphical tool to inspect the reciprocal lattice for missing data and the distribution of weak data. It can be used with either HKL or FCF data. ASYM-VIEW is one of the options of the ASYM [12, 4] tool. It can be used to create an averaged/unique reflection file. The <name>.cfk file (see Section 4.1.2) includes a listing of the merged reflections and excluded systematic absences.

### 3.1.11    BIJVOET-PAIR

X-Ray crystallography allows the determination of the absolute structure of crystal structures missing an inversion center. In such a case, the intensities of Friedel pairs of reflections (hkl and -h-k-l), or more generally Bijvoet pairs, have slightly different intensities, depending on the atomic resonant scattering parameter values f' and f''. This implies, in the absence of a mirror plane, the determination of the chirality of the molecules in the structure or the polarity of the crystal otherwise. The absolute structure can be determined as a special case of twinning with the refinement of a inversion twinning parameter = Flack parameter.[29] Values of 0 and 1 represent enantiopure structures where 1 indicates that the reported structure model needs to be inverted to be consistent with the diffraction data. A value of 0.5 indicates a racemic mixture. The refined Flack parameter comes with an s.u. (standard uncertainty) to be used as a reliability indicator for the absolute structure assignment. Refinement with a complete set of Friedel pairs is advised to avoid correlation effects with the positional parameters.

Alternatively, post-refinement estimates of the Flack parameter value can be determined based on the comparison of the observed differences in the Friedel pair intensities vs the corresponding calculated differences. Those Flack parameter value estimates tend to have smaller s.u.'s than the refined Flack parameter value. The BIJVOET-PAIR tool [6, 6] reports the values of two of those estimates: The PARSONS parameter[30] and the HOOFT parameter.[31–33]

Fig. 15 gives an overview of the result of the various absolute structure determination approaches for a gold containing compound in space group P1. Relevant numerical results are collected in the table on the right. The Hooft and Parsons estimates of the Flack parameter value show indeed lower s.u. values than the refined Flack parameter value. All values are close to zero,
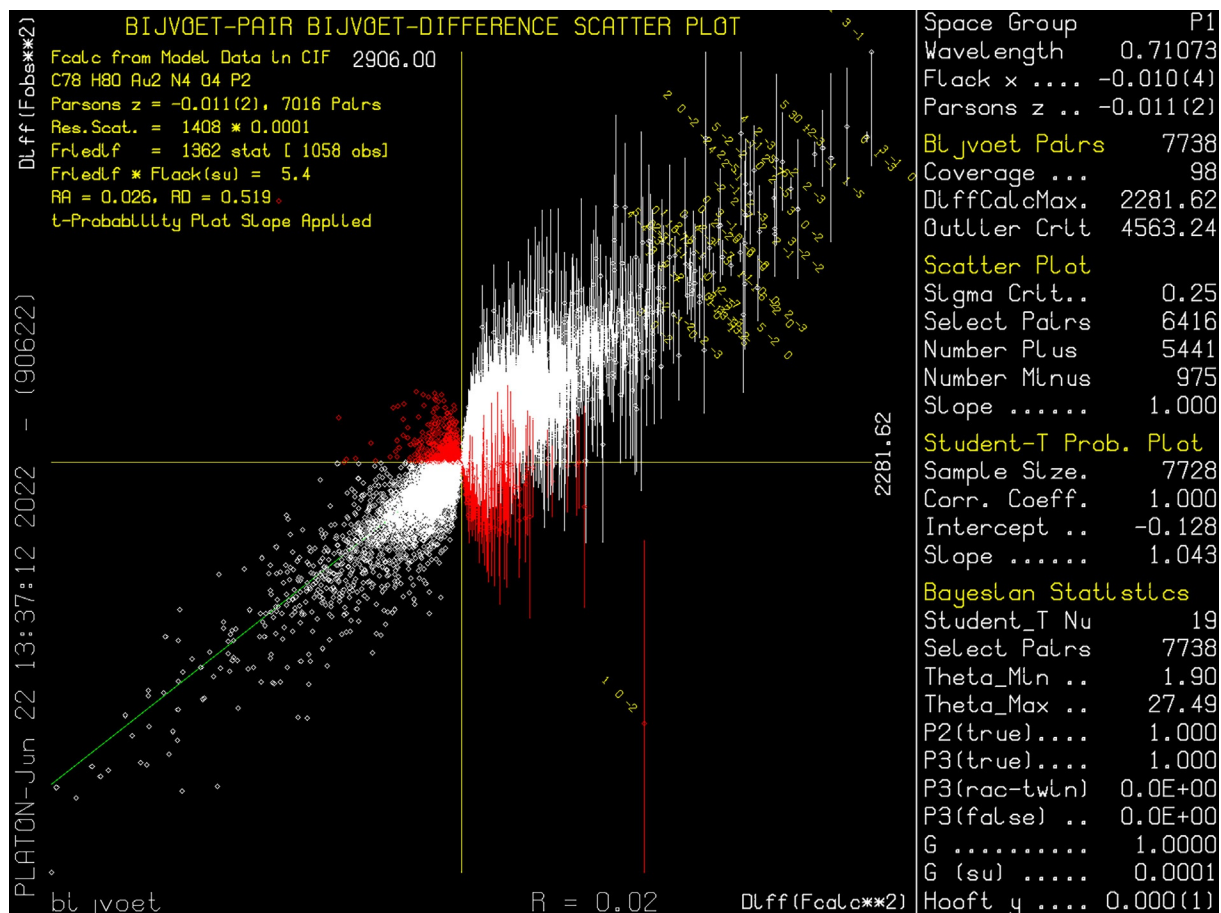
**Fig. 15** Overview of the Bijvoet-pair absolute structure analysis results.

indicating that the structure is essentially enantiopure. The Friedel pair coverage of 0.98 is also high. A scatter plot of observed vs calculated Friedel pair differences is shown on the left. The slope of the least-squares line (green) though the data point is positive, again indicating the correct absolute structure assignment. Individual counter indications are shown in red. The data points on the right are shown with one sigma error bars. The same datapoints, but now without error bars, are shown on the left.

## 4   Crystal structure validation

To quote Dorothy Hodgkin in her 1964 Nobel Lecture: ". . . [The] great advantage of X-ray analysis as a chemical structure analysis is its power to show some totally unexpected and surprising structure with, at the same time, complete certainty . . .." The catch with this statement is the implicit assumption that all procedures have been carried out correctly. That is the task of structure validation to certify.

A key step in a crystal structure determination is the interpretation of the preliminary 3D electron density map of the unit cell content in terms of discrete atomic densities. The next and critical step is the correct assignment of an element type to that atomic density. That is done based on peak height, chemical knowledge and assumed chemistry such as the expected element types. The preliminary assignment of the non-hydrogen atoms is today mostly done by computer software such as SHELXT. In general, this works well with good experimental diffraction data with reasonable resolution. Twinning, when not detected at the experimental stage, may hamper a smooth structure determination. Disorder, either configurational, conformational, partial occupation or substitutional, will need human intervention. The correct assignment of atom types differing by only one electron, such as C, N and O, can be a problem when not known beforehand from the chemistry. The same applies for nearby transition elements such as Cu and Zn and the lanthanides. The assignment of hydrogen atoms can be trivial with good data and involved in other cases. Notorious are problems with hydrogens on water molecules and -O-H moieties, when the hydrogen atoms are disordered over more than one site. Mis-assignment of atom types may lead to different assumed chemistry. Attempts to synthesize such a compound may lead to one with completely different properties than the original compound. A famous example is the complete synthesis of a natural product based on an incorrect crystal structure report.[34]

Wrong symmetry assignment may also lead to unusual geometry claims based on artifacts. Solution of the phase problem may only lead to an interpretable structure in a particular space group such as P1. However, that space group may be a subgroup of the correct higher symmetry one. Refinement in a too low symmetry group may lead to refinement artifacts and "interesting chemistry." See Fig. 2 in Ref. 16 where an example is shown where the authors claim significantly different coordination bond lengths that are chemically expected to be identical. CheckCIF validation includes multiple checks to detect artifacts as described above. Newly discovered issues often lead on a regular basis to the addition of new ALERTS.

### 4.1    The PLATON/checkCIF report

A full PLATON/checkCIF report for a compound <name> consists of two parts. The first one is the file <name>.chk file with an overview of relevant data of the analysis and a list of ALERT messages about the structure analysis, mainly based on the structure model parameters as supplied in the supplied <name>.cif file, being *the authors interpretation of the experimental data*. An example is shown in Section 4.2, Fig. 18. That information is also displayed in a graphical window when validation is invoked from the main PLATON main menu. The second one is the <name>.cfk file with extensive details on the quality of the refinement and the reflection data, as supplied with the <name>.fcf file, on which the structure model is based. This file contains more explicit information on issues reported in the <name>.chk file such as data set resolution, missing reflections, twinning, absolute structure and details of the refinement and reflection merging.

CheckCIF validation reports on issues such as missing information, inconsistencies, quality, potential errors, unusual results, possible improvements or interesting features. Those messages come as ALERT messages but are not necessarily errors. They come with three levels of relative importance, A, B and C, and a mainly informative G level. Level A messages may be serious or simple to address when pointing to missing relevant information. A set of low-level G ALERTS in combination may point to a serious issue after all.

PLATON based checkCIF ALERTS are included in IUCr/checkCIF with an identifier PLATxyz, where xyz is an ALERT number. ALERT messages are compact. More details and suggestions are listed on the terminal window in the background.

#### 4.1.1    CIF-validation

Structure validation aims at providing information on (1) The quality of the data on which the study is based (i.e., based on the best attainable data quality or sufficient for the purpose of the study). (2) The quality of the refined model (i.e., are all issues such as disorder resolved or sufficient to prove the structural features of interest). (3) Details on the experimental, data reduction and handling procedures used (allowing to repeat the analysis or to use the experimental data for follow-up research). (4) Messages about interesting, unusual or erroneous structural features.

#### 4.1.2    FCF-validation

CIF-validation mainly addresses the result of the structure determination. FCF-validation inspects the diffraction data on which that analysis is based and the refinement in more detail. The key to that analysis is the FCF file containing the (merged) observed and calculated diffraction data on which the refinement is based. That file should be provided and archived explicitly along with the CIF unless when that file can be easily recreated automatically from the relevant (embedded RES, HKL, FAB file) information in the CIF as is the case with current SHELXL version-based refinements.

The FCF file is checked for missing reflection data, sufficient resolution, outliers, signs for unresolved twinning, absolute structure, checking of a normal distribution of errors with a Normal Probability Plot (Fig. 16) or of deviating Analysis-of-Variance values (Section 4.3.2, Fig. 23). Relevant ALERTS are added to the <name>.chk file with more details in the <name>.ckf file.

### 4.2    A PLATON/checkCIF report example

Fig. 17 shows the packing diagram of a structure that refined to an R-factor of 3% in the non-centrosymmetric space group Pca2$_1$. This figure suggests an additional center of inversion at the nickel atom. Notice that the whole structure can be moved in this space group freely in the c-axis direction to have the Ni atom close to a center of inversion of the lattice. A PLATON based validation report for this structure is shown in Fig. 18. The checkCIF report starts with an overview of some relevant data for this structure (Fig. 18a). It reports that structure was refined with SHELXL-2018/3 and that both the CIF and FCF are provided. For most displayed items, both the data as reported in the CIF and those as calculated from the content of the CIF are shown. Those values should be (close to) identical. The best choice of the formula unit is not always obvious. The choice made by the algorithm in checkCIF may differ from that of the authors. When a different formula unit has been reported, the values of Z, Z′ and M$_r$ should be consistent with that choice. The reported residual difference map density should be similar to that calculated by checkCIF. Three sets of R, wR2 and S values are reported. The first one reports values calculated from the data in the CIF and FCF. The second set is calculated based on the observed and calculated F$^2$ values in the FCF. The third set are those values as reported in the CIF. All three should be consistent. A large difference may indicate that the FCF is not created in the same job where the CIF was created. The value of the Flack parameter is the one reported in the CIF. Those for Parsons and Hooft are calculated estimates.

Fig. 18b lists the various ALERTS. The first eight ALERTS are easy to address. ALERT_907 can be ignored for this racemic structure. ALERT_111 and ALERT_113 need some more work since they seem to confirm a missing inversion center within the default distance tolerances in ADDSYM. The PLATON tool ADDSYM-SHX can be used to create the input for a test refinement in the suggested new
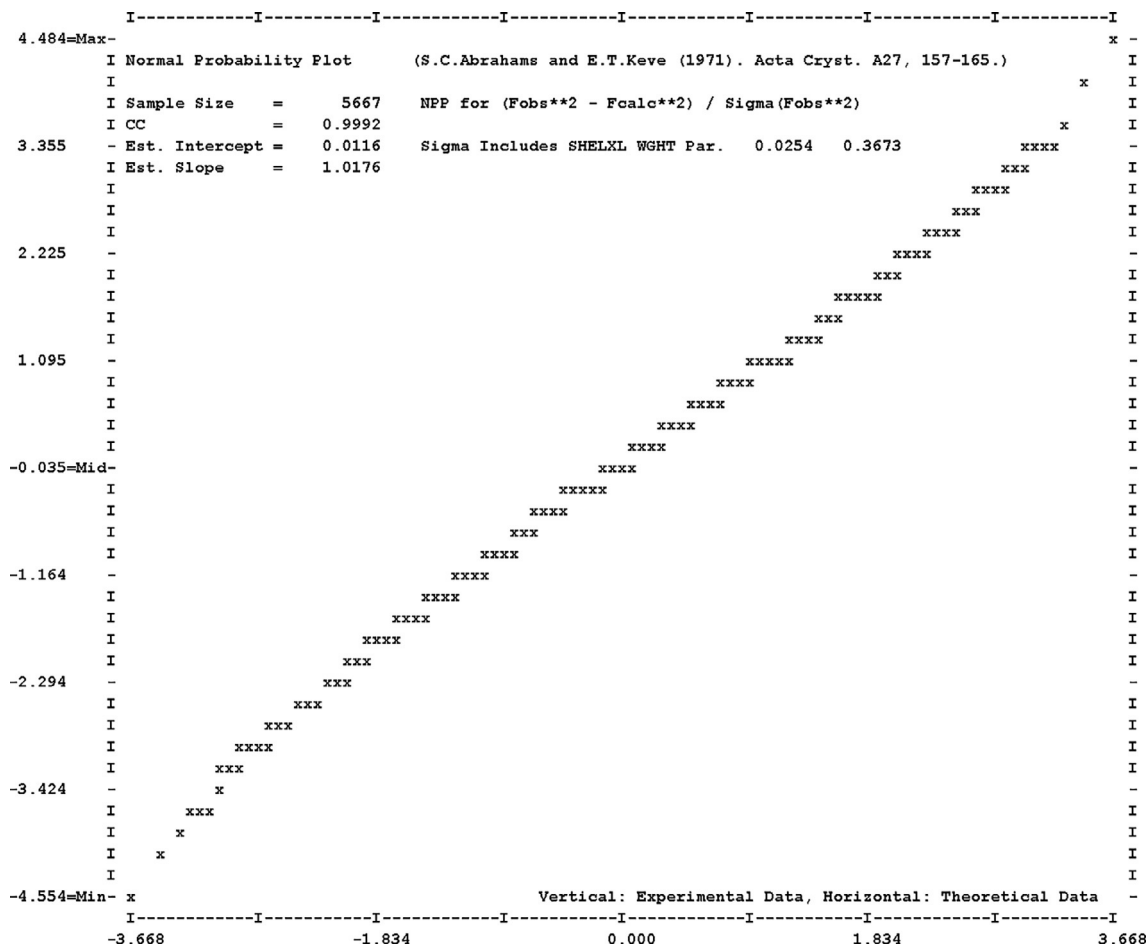
**Fig. 16**  Normal Probability Plot testing the normal error distribution between $F_o^2$ and $F_c^2$ that is expected to be close to linear.
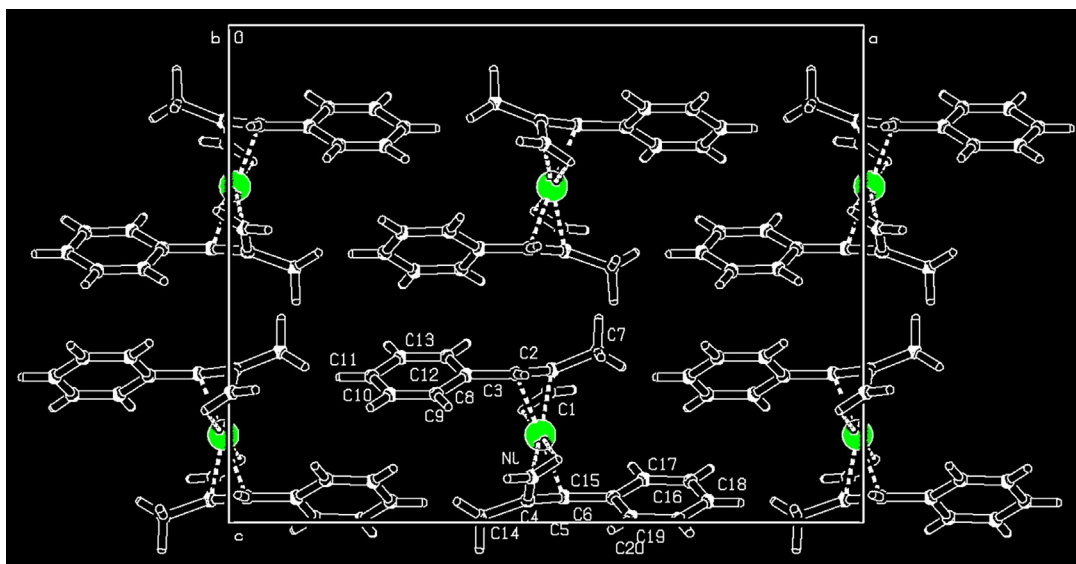


**Fig. 17**  Projection of a structure in the non-centrosymmetric space group Pca21, suggesting an additional center of inversion on Ni.
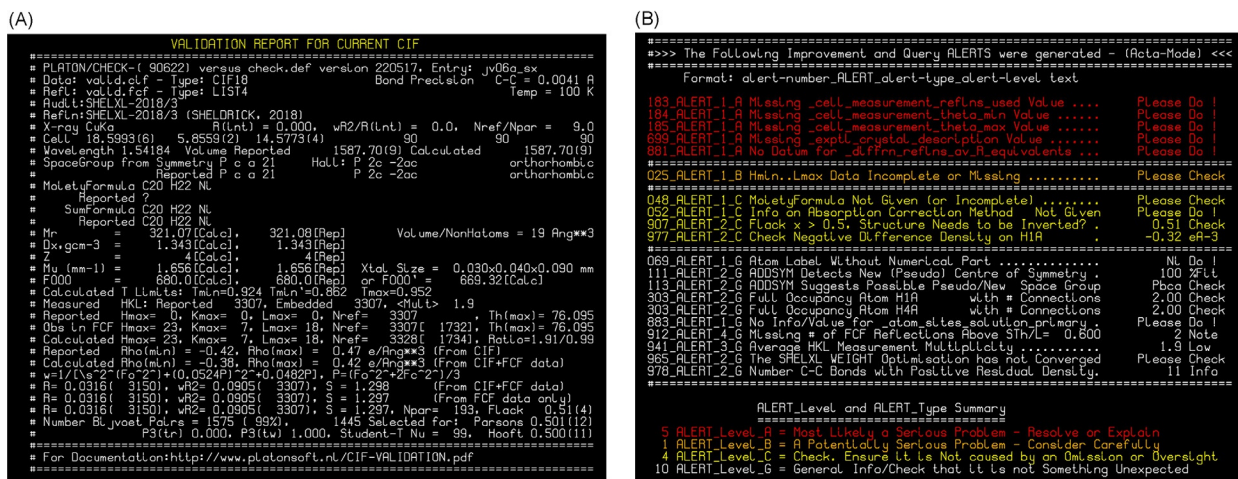
(A)



(B)



**Fig. 18**    Validation output: (a) Overview of selected information about the structure determination. (b) Validation ALERT messages listing.
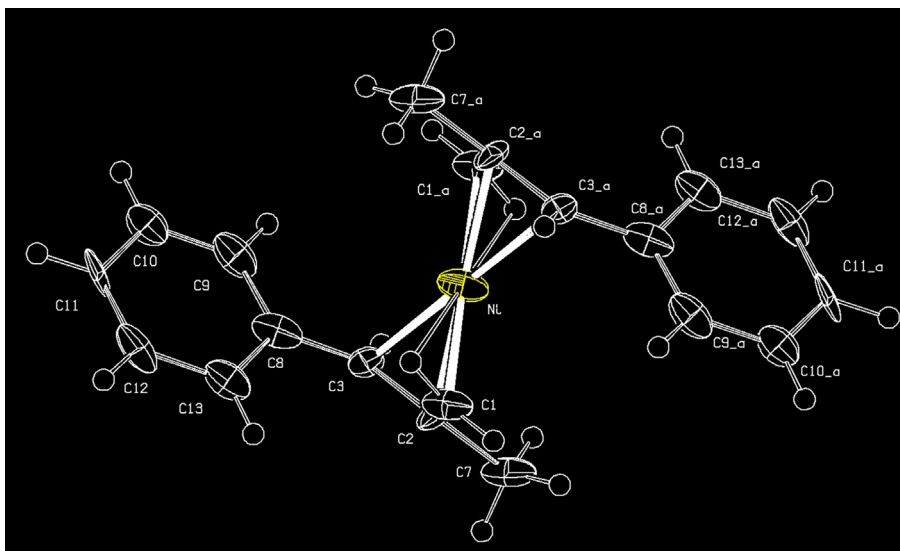


**Fig. 19**    Refinement result when the Pca2$_1$ structure shown in Fig. 17 is refined in space group Pbca, based on the inclusion of a center of inversion, as suggested by ADDSYM.

space group Pbca. That resulted in a refined structure as shown in Fig. 19 with an unsatisfactory R value of 22% and some poor displacement ellipsoids. From this, it can be concluded that this is a case of pseudo-symmetry which could also be verified by re-running ADDSYM with more strict distance tolerances. Further support for description of the structure in space group Pca2$_1$ comes from the NZ test (Fig. 20) that clearly shows a non-centrosymmetric intensity distribution. Finally, space group Pbca is not supported by the required set of systematic absences for that space group. Only two of the three glide-planes are fully supported. This can be investigated with the SPGRfromEX [11, 4] tool. Close inspection of the Pbca structure shows that layers of molecules perpendicular to the c-axis have slightly shifted to an average position, consistent with the direction of the elongation of the ADP ellipsoids as shown in the obviously averaged structure shown in Fig. 19.

ALERT_977 and ALERT_303 are related. Fig. 21 suggests agostic Ni...H interactions what would be chemically interesting when true. Inspection of the difference density map (Fig. 22) shows negative density (red) at the H1A and H4A atom sites, indicating that there is too much density put in the model at that site. The nearby green density maxima indicate the proper location of those hydrogen atoms. Refinement with the hydrogen atoms at the proposed positions removes both ALERTs, thus removing the agostic interaction issue.

ALERT_941 reports about low measurement multiplicity. Sufficient reflection multiplicity, i.e., the same reflection measured in different orientations multiple times, is required for a meaningful application of the multi-scan type correction for absorption with
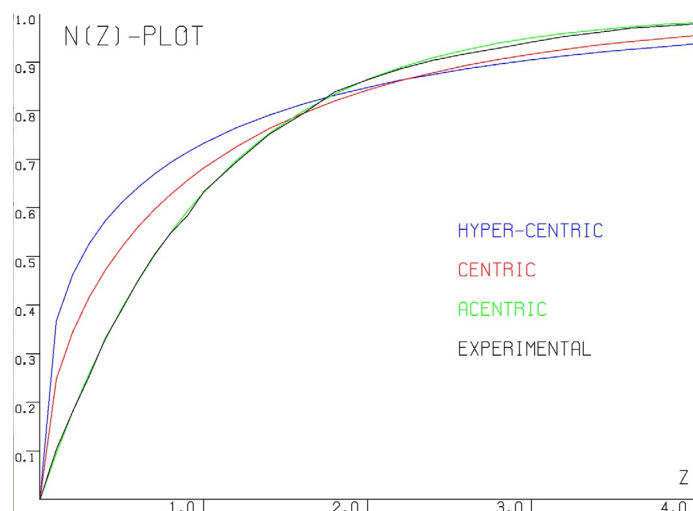
**Fig. 20**   NZ-Plot test showing that the investigated structure is likely non-centrosymmetric. The experimetal data based curve closely coincides the theoretical green curve for a non-centrosymmetric structure.
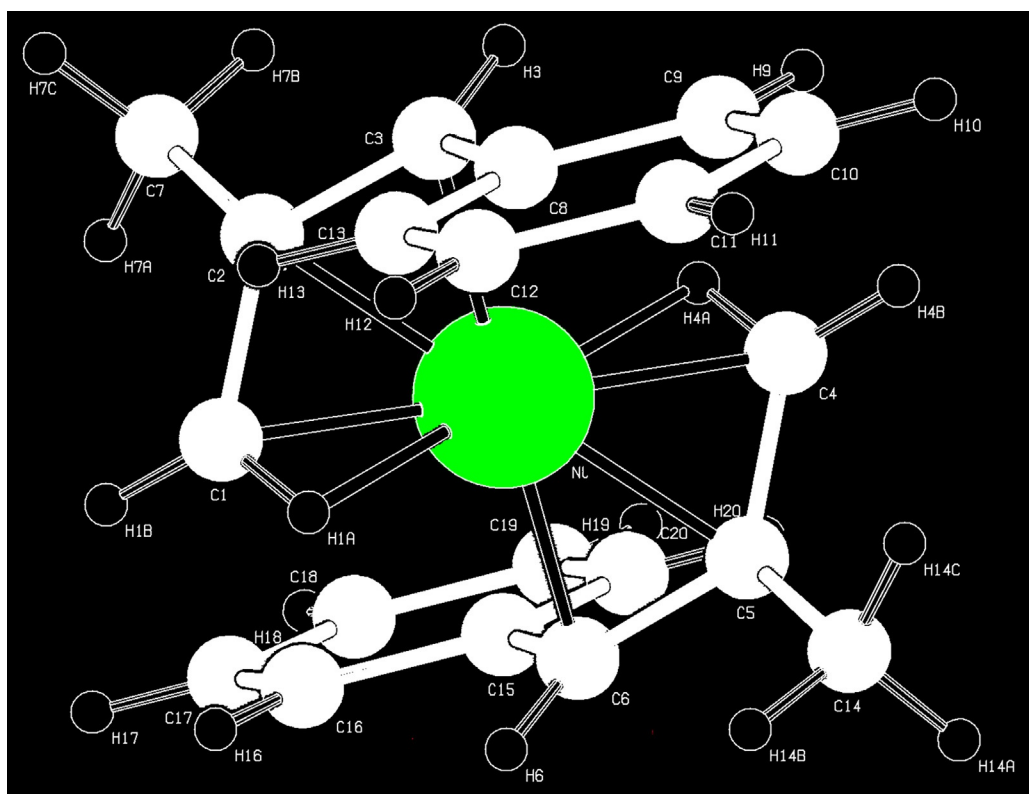


**Fig. 21**   PLUTON illustration showing erroneous short Ni-H contacts.

programs such as SADABS or MULABS [1, 5]. The positive and negative peaks around Ni may indicate inadequate absorption correction for that reason (Fig. 22).

ALERT_978 reports residual density on bonds. This generally indicates good quality data and result. There absence may indicate non-AIM type refinement, poor data or "observed" data "erroneously?" based on calculated structure factors.

ALERT_965 suggests that the reflection weighting scheme should be optimized. The target is a GOOF = S value close to 1, both globally and as a function of intensity or resolution. The ANALofVAR [8, 6] tool provides such an overview (Fig. 23).
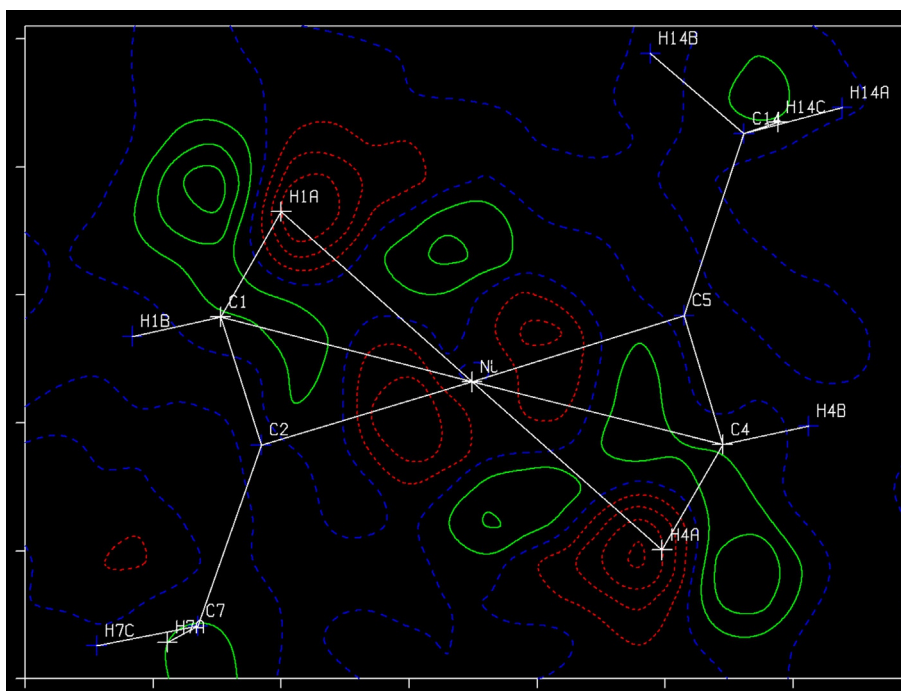
**Fig. 22** Difference density map showing negative density (red) at the location where H1A and H4A are in the refined structure model and nearby positive density (green) for their correct position.



**Fig. 23** Analysis-of-Variance statistics. Values of GooF and K are expected to be close to 1.

## 4.3 Some common validation issues

### 4.3.1 Reflection dataset completeness

Two types of information are expected to be given in the CIF: (a) The completeness of the dataset at theta_max, the maximum theta value of the reflection set, (b) The theta value (theta_full) at which the dataset is essentially complete. A dataset is expected to be complete to at least 25.2° (MoKα) radiation. SHELXL reports by default the completeness at 25.2° (MoKα). A completeness ALERT will be issued when lower than 95%. Proper theta_full values where near completeness is reached can be gleaned from the <name>.ckf listing associated with the checkCIF report (Section 4). A lower than 25.2° theta_full value needs a valid justification such as no significant reflections beyond that value or experimental restriction such as data collected with high-pressure gadgets.

### 4.3.2 Negative or large K values

The K values in Fig. 23 are expected to be close to 1. Strongly deviating values are reported with an ALERT in the checkCIF report. The most common ones are those associated with relatively weak $F_c^2$ values. A value much larger than 1 may indicate unresolved twinning, model errors or integration problems. A negative value may indicate problems with the background handling.

### 4.3.3 Residual density peaks

The difference electron density map should be essentially clean as mentioned in Section 3.1.4. Large positive and negative peaks often appear near heavy atoms. Most of those densities are caused by inadequate correction for absorption. Such peaks on atom sites may indicate wrong atom type assignments or partial site occupation. Other causes may be disorder or twinning. The argument of diffraction ripples does not apply for difference density maps.

### 4.3.4 Hydrogen atoms

Hydrogen atoms are often introduced at calculated positions according to the assumed hybridization of the atom they are attached to. ALERTS will be issued when either negative or positive density is found in the difference density map on those sites. Short inter-molecular contacts due to such a mis-assignment may also result in an ALERT.

## 5 Implementation and availability

PLATON is developed on the UNIX platforms LINUX and MacOS with graphics and GUI based on the X-Windows system. Hardcopy versions of the molecular graphics may be either PostScript or HPGL. The program is provided in source code and is easily compiled on UNIX platforms using free of charge compilers such as gfortran. For the MS-Windows platform there are two options. A compiled version with additional GUI is available from http://chem.gla.ac.uk/~louis/software/platon/. That version does not include, due to implementation issues, the System-S tool. The current WINDOWS11 platform also allows to install a virtual LINUX machine in it. In that way, a full LINUX version of PLATON can be compiled and installed as well on that platform.

Some PLATON tools rely on the availability of readily available external programs such as SHELXL, SHELXT, RASMOL and POVRAY.

Some program packages such as OLEX2, CRYSTALS and SHELXLE[35] use PLATON tools and features in the background. The same applies to IUCr/checkCIF. The latter facility includes most of the PLATON based ALERTS. The PLATON software and more details are available from http://platonsoft.nl.

## 6 Concluding remarks

PLATON/checkCIF validation is designed to validate mainly supposedly routine 3D structure determination reports. Powder diffraction studies are validated mainly for the reported 3D geometry. For a cautionary tale see Ref. 36. Incommensurate structure reports will need a specialist reviewer and are currently not covered by checkCIF. Similarly, the details of reports based on (quantum-mechanical) methods to address the non-spherical atomic density distribution are validated only partly.

CheckCIF is not only useful as part of the refereeing process. It should also not be used only in that late finalization stage but also at various points of the structure determination. Issues with a structure determination might be easier to address at that time than years later when submitted for publication. Not all ALERTS are errors that can be corrected. A good scientific explanation should suffice.

## References

1. Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C. *Acta Crystallogr.* **2016**, *B72*, 171–179.
2. Harrison, W. T. A.; Simpson, J.; Weil, M. *Acta Crystallogr.* **2010**, *E66*, e1–e2.
3. Sheldrick, G. M. *Acta Crystallogr.* **2015**, *C71*, 3–8.
4. Hall, S. R.; Allen, F. H.; Brown, I. D. *Acta Crystallogr.* **1991**, *A47*, 655–685.
5. Stewart, M.; Hall, S. R. *J. Appl. Cryst.* **1985**, *18*, 283.
6. Dolomanov, O. V.; Bourhis, L. J.; Gildea, R. J.; Howard, J. A. K.; Puschmann, H. *J. Appl. Cryst.* **2009**, *42*, 339–341.
7. Farrugia, L. J. *J. Appl. Cryst.* **2012**, *45*, 849–854.
8. Spek, A. L. *J. Appl. Cryst.* **2003**, *36*, 7–13.
9. Kroon-Batenburg, L. M. J.; Helliwell, J. R.; McMahon, B.; Terwilliger, T. C. *IUCrJ* **2017**, *4*, 87–99.
10. Palatinus, L.; Chapuis, G. *J. Appl. Cryst.* **2007**, *A60*, 604–610.
11. Sheldrick, G. M. *Acta Crystallogr.* **2015**, *A71*, 3–8.
12. Betteridge, P. W.; Carruthers, J. R.; Cooper, R. I.; Prout, K.; Watkin, D. J. *J. Appl. Cryst.* **2003**, *36*, 1487.
13. Petriček, V.; Dusek, M.; Palatinus, L. *Z. Kristallogr.* **2014**, *229* (5), 345–352.
14. Kleemiss, F.; Dolomanov, O. V.; Bodensteiner, M.; Peyerimhoff, N.; Midgley, M.; Bourhis, L. J.; Genoni, A.; Malaspina, L. A.; Jayatilaka, D.; Spencer, J. L.; White, F.; Grundkoetter-Stock, B.; Steinhauer, S.; Lentz, D.; Puschmann, H.; Grabowsky, S. *Chem. Sci.* **2021**, *12*, 1675–1692.
15. Lübben, J.; Wandike, C. M.; Hübschle, C. B.; Ruf, M.; Sheldrick, G. M.; Dittrich, B. *Acta Crystallogr.* **2019**, *A75*, 50–62.
16. Spek, A. L. *Acta Crystallogr.* **2009**, *D65*, 148–155.
17. Spek, A. L. *Inorg. Chim. Acta* **2018**, *470*, 232–237.
18. Spek, A. L. *Acta Crystallogr.* **2020**, *E76*, 1–11.
19. Le Page, Y. *J. Appl. Cryst.* **1987**, *20*, 264–269.
20. Johnson, C. K. *ORTEP-II*; Oak Ridge National Laboratory: Oak Ridge, TN, ORNL-3794, 1965.
21. Hirshfeld, F. L. *Acta Crystallogr.* **1976**, *A32*, 239–244.
22. LePage, Y. *J. Appl. Cryst.* **1982**, *15*, 255–259.

23. Zimmermann, H.; Burzlaff, Z. *Z. Kristallogr.* **1985**, *170*, 231–246.
24. Marsh, R. E. *Acta Crystallogr.* **2004**, *B60*, 252–253.
25. Marsh, R. E. *Acta Crystallogr.* **2005**, *B61*, 359.
26. Kitaigorodskii, A. I. *Molecular Crystals and Molecules*; Academic Press: New York, 1973.
27. Spek, A. L. *Acta Crystallogr.* **2015**, *C71*, 9–18.
28. Linden, A. *Acta Crystallogr.* **2020**, *E76*, 765–775.
29. Flack, H. D. *Acta Crystallogr.* **1983**, *A39*, 876–881.
30. Parsons, S.; Flack, H. D.; Wagner, T. *Acta Crystallogr.* **2013**, *B69*, 249–259.
31. Hooft, R. W. W.; Straver, L. H.; Spek, A. L. *J. Appl. Cryst.* **2008**, *41*, 96–103.
32. Hooft, R. W. W.; Straver, L. H.; Spek, A. L. *Acta Crystallogr.* **2009**, *A65*, 319–321.
33. Hooft, R. W. W.; Straver, L. H.; Spek, A. L. *J. Appl. Cryst.* **2010**, *43*, 665–668.
34. Li, J.; Burgett, W. G.; Esser, L.; Amezcua, C.; Harran, P. G. *Angew. Chem.* **2001**, *113*, 4906–4909.
35. Hübschle, C. B.; Sheldrick, G. M.; Dittrich, B. *J. Appl. Cryst.* **2011**, *44*, 1281–1284.
36. Schlesinger, C.; Fitterer, A.; Buchsbaum, C.; Habermehl, S.; Chierotti, M. R.; Nervi, C.; Schmidt, M. U. *IUCrJ* **2022**, *9*, 406–424.