

## Chapter 9.3

# PLATON

**This is a preliminary Introduction to the program. See the Platon website at**

**<http://www.cryst.chem.uu.nl/platon>  
for up-to-date program information.**

A.L.SPEK,  
BIJVOET CENTRE FOR BIOMOLECULAR RESEARCH,  
VAKGROEP KRISTAL- EN STRUCTUURCHEMIE,  
UNIVERSITY OF UTRECHT,  
PADUALAAN 8, 3584 CH UTRECHT,  
THE NETHERLANDS.

<b>1. INTRODUCTION TO PLATON .....</b>	<b>4</b>
1.1 COORDINATE DATA INPUT STANDARDS .....	4
1.2 REFLECTION DATA FORMAT .....	4
1.3 MENUS AND MOUSE CLICKING .....	4
1.4 INTRODUCTORY TUTORIAL .....	5
1.5 ANALYTICAL ABSORPTION CORRECTION .....	6
1.6 RECIPROCAL LATTICE SYMMETRY .....	6
1.7 PROGRAM PLUTON .....	6
1.8 WINDOW MENU .....	6
1.8.1 PLATON .....	6
1.8.2 PLUTON .....	6
1.9 HARD-COPY GRAPHICS .....	7
1.10 RAY-TRACED MOLECULAR GRAPHICS .....	7
1.11 PRINTING .....	7
1.12 CIF-VALIDATION .....	7
1.13 MISCELLANEOUS .....	7
<b>2. ON THE INTERNAL WORKINGS OF PLATON .....</b>	<b>7</b>
<b>3. SUMMARY OF KEYBOARD COMMANDS FOR PLATON .....</b>	<b>8</b>
3.1 INSTRUCTIONS PRECEDING ALL CALCULATIONS .....	9
3.2 CALCULATIONS .....	10
3.3 PLOTS .....	12
3.4 GENERAL INSTRUCTIONS .....	13
<b>4. OTHER OPTIONS IN PLATON .....</b>	<b>14</b>
4.1 CELL TRANSFORMATION .....	14
4.2 TRANSFORMATION OF HKL AND SHELX STYLE DIRECTION COSINES .....	15
4.3 COLOUR OPTIONS IN PLUTON .....	16
4.4 VOID & SOLV CALCULATIONS .....	16
4.5 ASYM-VIEW .....	17
4.6 LEPAGE - METRICAL SYMMETRY CHECK .....	18
4.7 TECHNIQUES FOR ABSORPTION CORRECTION IN PLATON .....	18
4.8 MULABS - BLESSING'S METHOD FOR ABSORPTION CORRECTION .....	18
4.9 PSI-SCAN BASED ABSORPTION CORRECTION .....	19
4.10 XTAL-PLOT .....	19
4.11 RESIDUE NUMBERS IN PLATON AND PLUTON .....	19
4.12 ATOM RENAMING IN PLUTON (.RES FILES) .....	20
4.13 DEFAULT INSTRUCTIONS FOR PLUTON .....	20
4.14 GRAPHICS (META, POVRAY, RASMOL) .....	20
4.15 OUTPUT LISTINGS .LIS & .LPS (PRINT & VIEW) .....	21
4.16 PLATON MAIN-MENU OPTIONS .....	21
4.17 PLATON/ORTEP (SUB) MENU OPTIONS .....	21
<i>PLATON/ORTEP SUB-MENU</i> .....	23
4.18 PLATON AND CSD-FDAT FILES .....	23
<b>5. TERMS AND NOTIONS .....</b>	<b>23</b>
5.1 THE N.IJK SYMMETRY OPERATION ON INPUT .....	23
5.2 DISORDER .....	24
5.3 MOLECULES AND RESIDUES .....	24
<b>6. PARAMETER FILES .....</b>	<b>24</b>
6.1 THE SPF FORMAT .....	24
<b>7 SPACE GROUP SYMMETRY .....</b>	<b>26</b>
7.1 SPACE GROUP NAMES KNOWN TO THE PROGRAM .....	28

---

7.2 NONSTANDARD SETTINGS WITH CELL TRANSFORMATIONS.....	29
<b>8. ATOMIC RADII USED IN PLATON .....</b>	<b>30</b>
<b>9. INTERNAL PARAMETERS.....</b>	<b>30</b>
<b>10. CIF-VALIDATION DOCUMENTATION (VALIDATION.DOC) .....</b>	<b>30</b>
<b>11. SQUEEZE, AN EFFECTIVE CURE FOR THE DISORDERED SOLVENT SYNDROME IN CRYSTAL STRUCTURE REFINEMENT.....</b>	<b>41</b>
11.1 GENERAL .....	41
11.2 IMPLEMENTATION AND USE .....	41
11.3 NOTES ON SQUEEZE .....	42
11.4 INTERPRETATION OF THE RESULTS .....	42
<b>12 PLATON IN THE <i>WINGX</i> IMPLEMENTATION.....</b>	<b>43</b>

## 1. Introduction to PLATON

Platon is a general crystallographic tool implementing:

- a large variety of standard geometrical calculations, *i.e.* bonds, angles, torsions, planes, rings, inter-molecular contacts (H-Bond analysis), co-ordination *etc.*, either fully automatic or as specified
- tests: *e.g.* ADDSYM (based on Yvon LePage's powerful published MISSYM (C) algorithm), VOIDS *etc.*
- utilities: cell transformation, SHELXL input *etc.*, PDB & CIF output
- graphics: *e.g.* automatic labelled 'ORTEP-lookalike plots, the molecular graphics program PLUTON and NEWMAN plots, inspection for completeness of the dataset with ASYM-VIEW
- several filters: *e.g.* exact analytical/de Meulenaer & Tompa face-indexed correction for absorption, DELABS (the locally modified Walker & Stuart DIFABS technique for empirical absorption correction), psi-scan based on with MULABS using multiple scanned reflections (following the Blessing algorithm) and SQUEEZE for handling disordered solvents.
- validation checks for data supplied in CIF-Format, implementing most published tests done in Chester on papers submitted for publication in Acta Cryst + a large number of 'PLATON-related' tests

Most PLATON features complement those available in the widely distributed public domain SHELX97 package.

**NOTE** Most features are currently available only for non-protein structures.

### 1.1 Coordinate data input standards

Parameter data may be given in various formats including CSD-FDAT, CIF, PDB & SHELX .RES standards. There are some restrictions on atom labelling (A4). A CIF file is recommended when su's (= e.s.d.'s) on the derived geometry parameters are desired.

### 1.2 Reflection data format

Reflection data should be provided in the SHELX format (optionally with SHELX76-style direction cosines. Some program options also accept SHELXL97-style FCF CIF's as reflection input.

### 1.3 Menus and mouse clicking

PLATON automatically opens a Window Menu. In the menu mode, input can be given either from the keyboard (useful for the more complex instructions) or via mouse clicks.

The Window Menu area is divided in four sub-windows.

1. The main graphics area in the upper-left corner (also used for the main PLATON (opening) menu from which various (sub)programs can be started by mouse clicks).
2. The upper-right window carries various sup-menus with clickable options. Currently active options are displayed in RED. Entries in this menu with vertical lines indicate click-

position sensitive options or magnitudes. The yellow top-field may be clicked to refresh the window contents.

3. The lower-Left window is used for short messages and the display of keyboard input.
4. The lower-right window informs about the current input status. It also provides an easy 'Exit' by clicking on it.

More extensive output is produced on the window from which the program was started and on the listing file.

### 1.4 Introductory tutorial

A sample run of the program using the test data in 'sucrose.spf' is executed by the interactive instructions (on the program prompt >>):

```
>> CALC (or click on 'CALC ALL')
```

This will give an exhaustive geometry analysis. The main results are on the listing file to be printed on an ASCII lineprinter. Some summary information is also directed to the display.

Alternatively, after restarting the program, the instruction :

```
>> CALC ADDSYM
```

will provide a test for missed symmetry (*cf.* the LePage MISSYM (C) algorithm), or:

```
>> CALC SOLV
```

will do a calculation to identify missed solvent areas or

```
>> TABL ACC
```

will generate a file as a starting point of a CIF-file suitable for submission to Acta Cryst C.

A default labelled ORTEP is produced with:

```
>> PLOT ADP
```

The SQUEEZE option is executed with:

```
>> CALC SQUEEZE
```

A file named name.HKP contains a new reflection file where the disordered solvent contribution is subtracted from the observed data. Further refinement can be done against this new file. See Section 10 below for further details.

A similar filter may be set-up to do a DIFABS-type of absorption correction.

```
>> CALC DELABS
```

This should be done in a separate PLATON run. You will need SHELXL files *name*.INS and *name*.HKL (HKLF 4 - with direction cosines in the SHELX-76 style - their validity is checked. A DELABS run may be done (when applicable) on the reflection file modified by SQUEEZE.

**NOTE** Be aware of the fact that the use of DELABS/DIFABS as an empirical absorption correction technique is considered to be 'sloppy crystallography' by some crystallographers/referees/journals. This implementation attempts to avoid possible human errors due to (or related to) cell transformation matrices, scaling factors, anisotropic displacement parameter model *etc.*

### 1.5 Analytical absorption correction

For those having crystals that can be face-indexed there is the easy-to-use implementation of the de Meulenaar & Tompa absorption correction method. This filter may be invoked with:

```
PLATON name.ABS
```

again, you will need a SHELX *name*.HKL file including the direction cosines. The file *name*.ABS should contain:

```
TITL text
CELL lambda a b c alpha beta gamma
FACE h k l d
.....
ABST mu
```

The d-values are the distances of an arbitrary point within the crystal to their corresponding faces. d & mu should be given in mm.

### 1.6 Reciprocal lattice symmetry

The symmetry of the reciprocal lattice at hand may be inspected with the PLATON/ASYM utility for a given SHELX .HKL

### 1.7 Program PLUTON

The previously available stand-alone molecular graphics program PLUTON is now available as an option within PLATON. The original PLUTON is now obsolete and no longer supported & maintained. Clicking in the main PLATON-MENU on the button PLUTON will result in a PLATON-run to generate an input file for PLUTON to make both consistent.

### 1.8 Window Menu

#### 1.8.1 PLATON

By default, a menu is shown from which the desired options can be MOUSE-CLICKED. Traditional keyboard entry is possible either through the 'MenuOff' button.

#### 1.8.2 PLUTON

By default, a PLOT instruction will bring up the Window Menu from which mouse-driven options may be run. The more complex instructions may be given via the keyboard. The

menu-feature is switched off either through a menu option or with the instruction MENU OFF (to be given before the first PLOT instruction).

### 1.9 Hard-copy graphics

Hard copies of the current display are conveniently obtained as a (meta) file by clicking on the 'Meta' option. By default, a PostScript file (Landscape Mode), suitable to be sent to a PostScript laser printer will be produced. Alternatively, the instruction 'SET META HPGL' will make the default HPGL (suitable when graphics is to be included in a wordprocessor environment (e.g. WordPerfect)).

### 1.10 Ray-traced molecular graphics

High quality (colour)graphics is easily obtained with the excellent Public Domain package POV-Ray. The required input files for that program are conveniently generated using the 'POV' button in the PLUTON menu. This will generate a *name*.POV file with the current content, style, colour and orientation of the current PLUTON-display.

POV-Ray (information) may be obtained from:

<http://www.povray.org> and <ftp.povray.org>.

### 1.11 Printing

PLATON generates an extensive listing file *name*.LIS that can be viewed with an editor, listed on a lineprinter or sent (via a filter) to a laserprinter. A utility that can be used for this purpose is 'pstext' (available in directory 'util' as 'pstext.shar.z'). An ASCII text is converted into a PostScript file. We use: 'pstext -s 9 -l -i 15 -2 platon.lis | lpr -Ppostscript' to print the file 'platon.lis' on out HP-postScript laser printer.

### 1.12 CIF-validation

PLATON can be run in a special mode to provide some validation checking of CIF's. It is still experimental with criteria not necessarily final. Most warnings can be signs of unresolved problems that should be looked at more closely and at least discussed in the experimental section of the paper when sent for publication. See Section 9 below for further details

### 1.13 Miscellaneous

Further information on available instructions is obtained with the instruction:

HELP

A right mouse click on the menu launches the hypertext help facility with a direct connection to the PLATON home site at: <http://www.cryst.chem.uu.nl/platon/pl000000.html>

## 2. On the internal workings of PLATON

This section provides a framework to understand the effects of the various available instructions.

The input atomic coordinates (X, Y, Z) are with reference to user-defined axes (a, b, c), which will usually be either crystallographic unit cell axes or an arbitrary orthogonal set; these coordinates are input as fractions of the unit cell edges or as Å units (in the latter case they are converted and stored as fractions of dummy cell edges). A second, orthogonal system (A, B, C) with coordinates (XO, YO, ZO) in Å's is set up internally (see J.D. Dunitz, X-Ray analysis and structure of Organic molecules, p236): A is a unit vector along a, B is a unit vector normal to a in the ab-plane, and C is normal to A and B. B will coincide with b in monoclinic cells in the b-setting. If the input axes are orthogonal, the two sets of axes a,b,c and A,B,C are coincident. The third system is the plotting coordinate system in cm: XP across the picture from left to right, YP up the picture from bottom to top and ZP out of the paper. All these axial sets are right-handed and absolute configuration is preserved in all rotations.

As atoms are input to the program, they are stored in the X,Y,Z and XO,YO,ZO axes systems. Each atom also has additional information stored for it such as estimated standard deviations, thermal motion parameters, a name (the embedded element name is used by default to set various radii to be used during the subsequent calculations) and various bit flags such as the inclusion bit. Coordinate data are checked for duplications on input. Atom labels not conforming to the required format are renamed.

A CALC instruction generally initiates a distance search on the basis of the previously set or supplied radii. In the INTRA mode this results in the set-up of an array that stores per atom all connections that are found. This list is used subsequently by a geometry listing routine that generates all unique bond distances, bond angles and torsion angles. Simultaneously with the set-up of the connectivity array all atoms are transformed (when necessary) to obtain a connected set. In addition, in the case that the molecule lies on a special position, the primary coordinate list is expanded with additional symmetry generated atoms in order to handle the geometry of the complete molecule. See also Section 4 Terms & Notions.

### 3. Summary of keyboard commands for PLATON

This Section provides a description of the keyboard commands which were available in PLATON92. Most are still available in current versions (which have a much larger number of options), though there may be alternative ways of doing things, in particular using the Window Menu. The full current manual is available in HTML format from

<http://www.cryst.chem.uu.nl/platon/pl000000.html>

Commands are grouped together as compound specific, calculation, plot, list and general instructions. The logical order of calculations is intra-molecular, inter-molecular and coordination geometry.

In the description of individual instructions below the following applies:

- (Sub)keywords are in uppercase and user data in lower case
- Data in parentheses are optional.
- Choices are separated by a slash.
- Parentheses in atom names (on input) are ignored except for that Ag denotes the atom type and Ag() the individual atom.
- Lower case input is automatically converted to upper case.
- Lines with a blank character in position 1 are ignored.

- Input lines may be continued with data on the next line by placing the symbol = at the end of the line.

### 3.1 Instructions preceding all calculations

These instructions will be necessary only in special situations.

#### **ROUND (ON/OFF)**

This option defines whether primary input data and derived geometrical parameter values will be rounded based on their standard deviations or not. The default setting is ON.

Example: ROUND OFF

#### **PARENTHESES (ON/OFF)**

By default, the numerical part of an atomic label will be enclosed within parentheses.

Example: PARENTHESES OFF

#### **INCLUDE E11 E12 ...**

Only the elements specified in the include list will be included in the calculations.

Example: INCLUDE C N O

#### **EXCLUDE E11 E12 ...**

The elements in the exclude list will be excluded from all calculations.

Example: EXCLUDE H

#### **DOAC E11 E12 ....**

The elements N, O, Cl, S, F and Br are treated as potential donor/acceptor atoms for hydrogen bonding by the program. This list will be replaced by the one specified in the instruction.

Example: DOAC N O

#### **HBOND p1 p2 p3**

Default criteria for hydrogen bonds are: distance between donor and acceptor atom less than the sum of their van der Waals radii + p1 (= 0.5 Å); distance H to acceptor atom less than sum of corresponding van der Waals radii + p2 (= -0.12 Å) and angle D-H...A greater than p3 (= 100 degree). The default values may be changed with the HBOND instruction.

#### **LSPL atom\_name1 atom\_name2 ..**

This instruction specifies the set of atoms for which a least-squares plane should be calculated. In this way it is possible to include special planes in the following calculations that include the generation of least-squares planes for planar parts in the structure.

#### **RING atom\_name1 atom\_name2 ...**

Rings in the structure up to 8 membered are found automatically. This instruction provides a facility to include larger rings (up to 30 membered) in the calculations. The atoms should be specified in bonded order.

#### **FIT At11 At21 At12 At22 ..... (etc)**

PLATON contains a FIT routine based on quaternion rotation (A.L. Mackay, *Acta Cryst.* (1984), **A40**, 165-166).

The general instruction to fit two molecules or residues is as above, where atoms to be fitted are given pairwise. Note: The FIT instruction may be broken up over more than one line. Lines that are to be continued should end with '='. There are two modes of operation:

1. when specified before any CALC instruction, the actual calculation will be done along with the subsequent CALC GEOM or CALC INTRA calculation. Listing of the results will be on the .lis' file only.
2. when specified after a CALC INTRA or CALC GEOM calculations will be done directly. Listing of the results of the calculation are both on the interactive output window and in the listing file.

A special case is the situation where the two molecules to be fitted have similar numbering of the atoms. The automatic sorting feature of PLATON will put the atoms in the same order. In such a case, specification of only one atom from each of the molecules will be sufficient to fit all non-hydrogen atoms in both molecules, e.g.

```
FIT O11 O21
```

### 3.2 Calculations

The full range of molecular geometry calculations will be carried out automatically with a the single keyword instruction **CALC**. This includes all the calculations that may be executed alternatively with the instruction sequence **CALC INTRA**, **CALC INTER**, **CALC COORDN** and **CALC METAL**.

```
CALC INTRA ((E11 r1 E12 r2 ..)/(TOLA p1)) (NOBOND) (NOANG)
(NOTOR) (NOLSPL) (NORING) (NOTMA) (NOBPA) (NOSTD) (WLSPL)
(NOPESD) (NOMOVE) (NOSYMM) (VERIFY) (TOLP t2)
```

The default instruction **CALC INTRA** produces a full calculation and listing of all relevant intra-molecular geometrical parameter options using default covalent radii drawn from internal tables. Atoms with distances less than the sum of their covalent radii plus a tolerance ( $TOLA = 0.4 \text{ \AA}$ ) are considered to be bonded. The default radii values may be modified with their explicit specification (in which case  $TOLA$  is set to zero, unless specified explicitly). Alternatively the parameter  $TOLA$  may be modified. In the automatic radii mode an additional  $0.6 \text{ \AA}$  is added to the tolerance to catch (Earth)alkali to non-metal contacts.

The calculation and listing of bonds, bond angles, torsion angles, least-squares planes, rings, angles between bonds and least-squares planes and thermal motion analysis may be suppressed with the specification of the sub-keywords **NOBOND**, **NOANG**, **NOTOR**, **NOLSPL**, **NORING**, **NOBPA** and/or **NOTMA**. The calculation of standard deviations may be suppressed with **NOSTD**.

The **NOMOVE** sub-keyword has the effect that atoms are left at their input positions in the course of the generation of a connected set.

The sub-keyword **VERIFY** may be used to control the inclusion of a connection in the bond list for plotting.

**TOLP** is an out-of-plane deviation parameter (by default  $0.1 \text{ \AA}$ ) that determines the inclusion of an atom in the process of automatic least-squares plane search.

**NOPESD**, when specified, has the effect that the e.s.d. of the plane parameters is not included in the calculation of the e.s.d. in out-of-plane deviations.

WLSPL invokes mass-weighted least-squares plane calculations as opposed to unit weighted.

NOSYMM limits the search for connections within the input coordinate set without the application of translation or rotation symmetry.

Example: CALC INTRA NOLSPL NORING

**CALC GEOM (SHELX/OMEGA/MOGLI/EUCLID) (VERIFY) (NOMOVE) (EXPAND)**

This instruction executes a short intra calculation, mainly producing a list of bond distances, bond angles and torsion angles, as an alternative for the exhaustive CALC INTRA calculations. The sub-keyword SHELX may be used to generate an ordered coordinate file suitable for SHELX; OMEGA generates a file suitable for the tabulation of primary and derived parameters; MOGLI results in a DGE-file suitable for the program MOGLI and EUCLID gives a new SPF style file.

The NOMOVE sub-keyword has the effect that atoms are left at their input positions in the course of the generation of a connected set.

The EXPAND option may be useful for the generation of a file with the complete molecule as opposed to just the unique part.

Example: CALC GEOM EUCLID EXPAND

#### **CALC TMA**

This invokes the execution of a rigid-body thermal motion analysis and the calculation of derived quantities. It is automatically included in a CALC INTRA calculation. Note: No TMA analysis is done when the residue contains too few atoms or when the R-index of the observed and calculated Uij's is too high.

**CALC INTER (E11 p1 E12 p2 ..)/(TOLR p1)**

Short inter-molecular contacts are listed with this instruction. By default van der Waals radii drawn from internal tables are used in conjunction with a default tolerance (TOLR = 0.2 Å). Hydrogen bonds are automatically found and analyzed.

**CALC HBOND (p1 p2 p3)**

This instruction provides a subset of the information generated with the CALC INTER instruction and may be of use when interest is concentrated on H-bonds.

**CALC COORDN (p1/E11 r1 E12 r2 .. (NOANG) FIVE (TBA))**

This instruction provides for the analysis of co-ordination spheres. Bond distances and bond angles are calculated for atoms within the specified sphere. By default such a calculation is done for all atoms (excluding C and H) and with radius 3.2 Å. This default may be changed with the specification of the desired value. Alternatively a list of selected elements and their corresponding co-ordination radii may be specified for the co-ordination geometry calculations. Bond angles may be excluded from the listings with the NOANG sub-keyword. A Berry pseudo rotation analysis is carried out automatically when an atom is found to be bonded to exactly 5 atoms. Such a calculation may be enforced for the five shortest contacts with the sub-keyword FIVE optionally followed with the value for the trans-basal-angle (default 150 degree).

**CALC COORDN atom\_name p1**

The co-ordination geometry about a single atom may be examined with this instruction.

Example: CALC COORDN O3 3.2

**CALC METAL (p1)**

Distances between metal atoms less than p1 (default 10 Å) are calculated. This option is included in the default CALC calculations.

**CALC VOID (LIST) (GRID p1) (TOLV p2)**

This option may be used to check the structure for voids as possible sites for solvents. The GRID (default value 0.4 Å) and the minimum VOID radius (1.2 + p2 Å) may be changed (default p2 = 0.0). The LIST option gives a map on the lineprinter. Positions with a shortest contact distance to the van der Waals surface of at least 1.2 + p2 Å are indicated with >. Solvent accessible areas are indicated with a dot. Blank areas indicate small voids, all other gridpoints are within the molecular van der Waals volume. Note: This option may also be used to study cases where the unit cell contents are misplaced with respect to the symmetry elements, since this fault will generally result in both areas with short molecular contacts and areas with voids.

The VOID option is more compute intensive than the rest of the instructions. It is advised to run this option in BATCH mode.

**CALC DIST (eltype p1)**

A distance scan is done for all vectors between the specified element and within the specified radius. By default a scan is done for H-atoms.

Example: CALC DIST I 4.0

**DIST atom\_name1 atom\_name2**

With this option a distance between two specified and not necessarily bonded atoms may be calculated between atoms in the atom array.

**ANGL atom\_name1 atom\_name2 atom\_name3**

The angle between the specified and not necessarily bonded atoms is calculated.

**TORS atom\_name1 atom\_name2 atom\_name3 atom\_name4**

The dihedral angle involving the four specified atoms (not necessarily bonded) is calculated.

**LSPL atom\_name1 atom\_name2 atom\_name3 atom\_name4 ...**

The least-squares plane determined by the specified atoms is calculated.

### 3.3 Plots

The program provides graphics options to support the geometry analysis.

**PLOT (LSPL/PLAN/RING/RESD) (ALONG/PERP) (HPD/HPP/CAL)**

Plots of the structure viewed perpendicular to or along the various least-squares planes may be produced for inspection.

**PLOT NEWMAN (DISPLAY/PLOTTER/META) (at1 at2)**

Newman plots are produced, provided that a CALC INTRA instruction was carried out previously in order to prepare a file with the relevant data for all Newman projections. The

Newman plots may be examined sequentially or for an individual one to be selected by specifying the relevant central bond.

**PLOT ADP (*nr*) (*COLOR*) (*VERIFY*) (*DISPLAY/PLOTTER/META*)**

A thermal motion ellipsoid plot is produced for residue number *nr*. The *COLOR* option provides for the distinction of atom types in the plot (oxygen RED, Nitrogen BLUE and halogens GREEN). The three plot angles *xr*, *yr* and *zr* to reconstruct the present orientation are plotted in the lower right corner, upper left corner and lower left corner respectively. The probability level of the ellipsoid surfaces is shown in the upper right corner. When no VIEW instruction was given previously, the program will calculate a minimum overlap view. Example: PLOT ADP 3 COLOR

**PLOT (*STICK/SOLID/CPK*) (*MONO/STEREO*) (*HATCH/CONTOUR*) (*nr*)  
(*SHN/SHH/SHF/SHC*) (*VERIFY*) (*COLOR*) (*DISPLAY/PLOTTER/META*)**

Plots of the molecule in various styles are produced in a orientation that may be modified with VIEW instructions. The STICK style provides a fast plot with labelled atoms. The SOLID sub-keyword results in a ball-and-stick plot and the CPK sub-keyword in a spacefilling plot. The COLOR option colours oxygen atoms RED Nitrogen atoms BLUE and Halogen atoms GREEN. *nr* specifies the number of the residue to be plotted. The three plot angles *xr*, *yr* and *zr* to reconstruct the present orientation are plotted in the lower right, upper left and lower left corners respectively.

**BOX (*ON/OFF[ON]*) (*RATIO ratio[1.333]*)**

By default the drawing will be surrounded with a rectangular box outline. This setting may be changed with the ON and OFF sub-keywords. The three numbers shown in the bottom right, top left and bottom left corner of the box are the rotation angles *xr*, *yr* and *zr* respectively. These numbers may be used to reconstruct this particular orientation directly from the default UNIT orientation via a VIEW XR *xr* YR *yr* ZR *zr* instruction. The default horizontal to vertical size ratio of the box for a ADE plot is 4/3. A ratio of 1 produces a square box. Example: BOX ON RATIO 1.0

**VIEW (*XR xr*) (*YR yr*) (*ZR zr*) ...**

The current orientation of the molecule for plotting may be modified with a VIEW instruction: VIEW XR 45 YR -55 will rotate the molecule first clockwise about the horizontal X-axis, followed by an anti-clockwise rotation by 55 degrees about the vertical Y-axis. VIEW instructions are accumulative. The single keyword instruction VIEW will bring the molecule back in the default orientation.

**SET PROB (*30/40/50*)**

The probability level for the ellipsoid surfaces is set by default to 50%. Example: SET PROB 30

### 3.4 General instructions

**HELP (*SPGR*)**

This instruction provides an on-line HELP facility. The SPGR option lists all space groups known to PLATON.

**LIST BONDS/ATOMS/SYMM/CELL/RADII**

This provides for on-line inspection of BOND and ATOM tables, the current symmetry, CELL dimensions and default radii.

**LIST IPR/PAR (*ival1* (*ival2*))**

Internal parameter values (see Appendix VII) may be inspected with this instruction. A range will be listed when two values are specified and the full range when none is given.

Example:LIST PAR 3 5

**SET PAR *p1 p2***

This instruction is not meant for general use. It provides a facility to modify internal parameter values, in particular those with no equivalent (sub)keyword. *p1* is the parameter number and *p2* the new value.

**SET IPR *p1 p2***

This instruction is not meant for general use. It provides a facility to modify internal parameter values, in particular those with no equivalent (sub)keyword. *p1* is the parameter number and *p2* the new value.

**SAVE**

This instruction causes the saving of subsequent instructions on a file to be executed on all data sets, separated by ENDS cards, on the parameter file.

**END**

This results in a normal end of program when the .SPF file contains only one data set, otherwise the program restarts for the next data set on the file.

**QUIT**

This results in an immediate stop of the program, ignoring possible further datasets on the input file.

**STOP**

This results in an immediate stop of the program, ignoring possible further datasets on the input file.

**4. Other options in PLATON****4.1 Cell transformation**

PLATON can be used to transform CELL, SYMM and coordinate data according to a specified transformation matrix. The general format of the transformation instruction line is:

```
TRMX r11 r12 r13 r21 r22 r23 r31 r32 r33 t1 t2 t3
```

in which 'r11 r12 r13' expresses the new a-axis in terms of the old

e.g.  $a' = b + c$  is encoded as 0 1 1

t1 t2 t3 indicate a shift of origin after the cell transformation.

The TRMX will affect only data following it! Symmetry operations may be protected for transformation by placing [] e.g. SPGR [C2/C]. This may be useful when the target space group is known and the transformation doesn't seem to work otherwise (which should of course never happen). The transformed data may be written out as a SHELX .RES file by

1. click on proper button in the PLATON opening window, or
2. type CALC SHELX

#### 4.2 Transformation of hkl and SHELX style direction cosines

Although it is generally better not to touch the SHELX reflection file with a transformation matrix (to avoid chaos), since SHELX avoids that need via specification of a transformation matrix on the HKLF line to transform reflection data as provided to the 'CELL-standard', PLATON provides such a tool (the HKLT option).

The following example illustrates the procedure: Two files are needed

1. the reflection file : *name*.HKL (SHELX HKLF 4 - STYLE)
2. instruction file : *name*.INS

For the transformation (P21/c to P21/n : (1 0 1/0 1 0/ -1 0 0)

```
TRMX 1 0 1 0 1 0 -1 0 0
CELL 0.71073 10.6597 9.1517 21.8466 90.0 102.25 90.0
SPGR P21/c
HKLF 4 1 1 0 0 0 1 0 0 0 1
HKLT
```

or

```
TITL N1848A
CELL 0.71073 22.1829 9.1517 10.6593 90.0 105.76 90.0
SPGR P21/n
HKLF 4 1 1 0 1 0 1 0 -1 0 0
HKLT
```

The specification of the space group is optional. This example run would produce a new file *name*.HKP and a short screen output.

```
:: TITL N1848A
:: CELL 22.1829 9.1517 10.6597 90.000 105.758 90.000
2082.7
:: SPGR P21/n
```

Direct Cell axes and HKL Transformation Matrix

```
1.0000 0.0000 1.0000
0.0000 1.0000 0.0000
-1.0000 0.0000 0.0000
```

Coordinate Transformation Matrix

```
0.0000 0.0000 1.0000
0.0000 1.0000 0.0000
-1.0000 0.0000 1.0000
```

Transformation Matrix for SHELX Direction Cosines

```
0.0000 0.0000 1.0000
0.0000 1.0000 0.0000
-0.9848 0.0000 0.4805
```

```
:: HKLTRANS hkl on :hkl.t.hkp
```

### 4.3 Colour Options in PLUTON

The assignment of colour to plot-items can be done at four levels

1. global colour
2. per atom-type
3. per residue-type
4. per ARU

Option 1:Instruction:

```
COLOR BLACK/RED/GREEN/BLUE/YELLOW/ORANGE/VIOLET/BROWN
```

Option 2:

Colour assignment is done by default on the basis of element-type. The default setting may be changed with:

```
COLOR TYPE atom-type col (atom-type col ..)
```

Colour is switched on/off with

```
COLOR (on/off)
```

or implicitly with

```
STRAW COLOR
```

This option may be combined with the 'Black-and-White' Patterns:

```
BWC (on/off)
```

Option 3:

Residues (i.e. unconnected species) can be displayed with differing colours with:

```
COLOR RESD
```

Option 4:

ARU's may be given distinguishing colours with instructions such as

```
ARU red 1555.01 1556.01
```

```
ARU green 1565.01
```

ARU-related colours are displayed with:

```
COLOR ARU (on/off)
```

or by clicking the 'col ARU' menu field. This option may be combined with the 'Black-and-White' patterns:

```
BWC (on/off)
```

### 4.4 VOID & SOLV calculations.

PLATON offers two options to detect and analyse solvent accessible voids in a crystal structure. SOLV is a faster version of VOID. VOID is useful when, in addition to the detection of solvent areas, a packing coefficient (Kitaigorodski) is to be calculated. The SOLV option is used as part of a SQUEEZE calculation. Some background information may be obtained from the paper *Acta Cryst* (1990) **A46**, 194-201. The algorithm used to detect solvent accessible areas may be summarised as follows.

1. The unitcell is filled with atoms of the (symmetry expanded) structural model with van der Waals radii assigned to each atom involved.
2. A grid search (with approximately 0.2Å grid steps) is set up to generate a list of all gridpoints in the unitcell which are at a minimum distance of 1.2Å from the nearest van der Waals surface.
3. The list generated under 2 is used to grow lists of gridpoints (possibly supplemented with gridpoints within 1.2 Å around 2-list points) constituting (isolated) solvent accessible areas.
4. For each set of 'connected gridpoints' a number of quantities are calculated.
  - the centre of gravity
  - the volume of the void
  - the second moment of the distribution (The centre of gravity can be seen as a first moment). The corresponding properties of the second moment (ellipsoid) can be calculated via the eigenvalue/ eigenvector algorithm. The shape of the ellipsoid can be guessed from the square-root of the eigenvalues: a sphere will give three equal values.
5. For each void in the structure a list of shortest distances to atoms surrounding the void is calculated. Short contacts to potential H-bond donors/acceptors may point to solvents with donor/acceptor properties.

As a general remark it can be stated that crystal structures do not contain solvent accessible voids larger than in the order of 25Å<sup>3</sup> However it may happen that solvent of crystallisation leaves the lattice without disrupting the structure. This can be the case with strongly H-bonded structures or framework structures such as zeolites. It should also be remarked that structures have a typical packing index of in the order of 65 %. However, the missing space is in small pockets, too small to include isolated atoms.

#### 4.5 ASYM-VIEW

This option may be used to get an overview over the dataset in reciprocal space in terms of resolution, data quality and missing data. The feature requires a *name.RES* or *name.CIF* file and a *name.HKL* or *name.FCF* structured reflection file and is invoked via 'ASYM-VIEW' on the opening window. Data completeness is an important issue for CCD and imageplate derived datasets.

A series of resolution rings is shown [ $\sin(\theta) / \lambda$ ] starting at 0.50 in steps of 0.05. The red ring represents the 'critical' 0.6 (about 25 degrees for MoKa) minimum resolution level required for Acta Cryst papers. Only a hemisphere of data is shown if Friedel related reflections are averaged.

Reflections in the asymmetric section of the hemisphere are represented by 'L' for weak reflections, '\*' for those with intensities > 10 sigma(I) or the number of sigma's. Symmetry related sections show a '+' for reflections with a symmetry related reflection in the asymmetric section. 'Blank' areas either indicate missing reflections or systematic absences, left out on the basis of the symmetry provided in the *name.RES* (or *name.CIF*) file.

#### 4.6 LEPAGE - metrical symmetry check

The metrical symmetry of a lattice may be investigated with the LEPAGE algorithm. The input to the program may be a *name.RES*, *name.CIF* or similar file containing cell parameters and lattice centring information. The feature may be invoked either via the 'METRICSYMM' button on the PLATON opening window or with the keyword 'LEPAGE'

Note: This algorithm only gives the symmetry of the lattice. The actual symmetry may be lower, depending on the content of the unit cell. When the content of the unitcell is known, it is suggested to run the ADDSYM option, based on the MISSYM(C) algorithm by Y. Le Page.

#### 4.7 Techniques for absorption correction in PLATON

PLATON implements a large variety of established techniques for correction for absorption.

1. Numerical Methods (Supposedly close to exact and based on FACE indexing)
  - ABST: Analytical following the Alcock version of "de Meulenaer & Tompa"
  - ABSG: Gaussian Integration (Modified from Coppens)
  - ABSS: Spherical Correction
2. Semi-empirical methods (based on additional experimental data)
  - ABSP: Psi-Scan data based correction (North et al.)
  - MULABS: Based on multiscanned reflection data (based on Blessing)
3. Empirical Methods
  - DELABS: Modified implementation of the DIFABS algorithm (Walker & Stuart)
4. ABSX: Comparison of calculated (i.e. Face-Indexed Alcock) and experimental psi-scans.

#### 4.8 MULABS - Blessing's method for absorption correction

MULABS implements a semi-empirical method for absorption correction using multiple scanned reflections (*i.e.* multiple symmetry or azimuth equivalent reflection data) following the excellent algorithm published by Bob Blessing, *Acta Cryst* (1995), **A51**, 33-38 (also available in his SORTAV program). MULABS as implemented in PLATON requires two files:

1. a reflection file *name.HKL* containing the redundant data set (SHELXL HKLF 4 FORMAT + DIRECTION COSINES)
2. A small pertinent data/instruction file '*name.ABS*'  
*name.ABS* should contain the following (free format) data:
 

```
TITL ..
CELL lambda a b c alpha beta gamma
SPGR name
MULABS mu radius tmin tmax l0max l1max
```

The CELL should correspond to that of the dataset, *i.e.* the one used to collect the set of equivalent reflections. SPGR can be either P21/c or P2/m *etc* LATT & SYMM line if necessary. **mu** should be in mm<sup>-1</sup>, **radius** the equivalent radius (in mm), **tmin** & **tmax** the minimum and maximum crystal dimensions, **l0max** & **l1max**, respectively the even and odd order limits of spherical harmonic expansion. Generally, only **mu** and **radius** are needed on input. An example

```
TITLE test
```

```
CELL 0.71073 10.1 10.2 10.3 90 90 90
LATT P C
SYMM -x, y, -z
MULABS 6.5 0.3
```

MULABS without parameter data gives a correction without the theta dependence (i.e. average value = 1.0). Parameters may be set interactively with MUMM (for mu in mm) and RADIUS (mm) instructions. The MULABS implementation of the Blessing Algorithm is EXPERIMENTAL and may change somewhat on the basis of future experience. MULABS can also be called from the main PLATON menu (MULABS button) In that case leave out the MULABS line from *name.ABS* file. MULABS should be useful in particular for redundant data sets collected with area detectors (e.g. KCCD). However it can be also useful for redundant datasets originating from serial detector diffractometers.

Note: This implementation is independent of the particular brand/type of diffractometer used; all information needed is included in the (SHELX-style) direction cosines (apart from the cell dimensions and Laue symmetry). A new file *name.HKP* is written (SHELX HKLF 4 style) implementing the absorption correction. Some additional information is written to *name.LIS*

#### 4.9 Psi-scan based absorption correction

PLATON/ABSPSI for psi-scan based absorption correction requires two files

1. a reflection file (SHELX HKLF 4 + Direction Cosines) : *name.HKL*
2. an instruction file : *name.ABS*

The content of the instruction file should look like

```
TITL S588A
CELL 0.71073 11.1364 12.0577 15.3782 90 90 90
ABSP
```

The SHELXL .HKL file should contain two sections, separated by a blank line. The psi-scan data should go in the second section. Such a file is produced for CAD4 data using the data-reduction program HELENA. Each psi-scan set should consist of 36 measurements, measured in 10° steps.

#### 4.10 Xtal-Plot

The face indexed absorption correction options contain a feature to draw crystals for checking the correctness of the description that can also be used stand-alone. The essential input file *name.INS* should contain

```
TITL
CELL
FACE h k l dist
.....
FACE h k l dist
```

It may be invoked from the main PLATON menu (ABSTOMPA). Planes may be added, modified or deleted interactively.

#### 4.11 Residue Numbers in PLATON and PLUTON

PLUTON & PLATON use different algorithms to assemble residues, resulting in differing assigned numbers. In order to get consistent residue numbering, PLUTON should be run through PLATON (PLUTONauto button). PLATON will then generate a *name.ELD* file with

the order and numbering of residues as in PLATON. This feature is call the 'PLATON Compatibility mode'

#### 4.12 Atom renaming in PLUTON (.res files).

PLUTON can be used for interactive atom renaming. A new file (.new) including changes is written only under certain conditions:

- PLUTON should have been called directly from the command line, either via an alias 'pluton' or 'platon -p' or via the 'toPLUTON' option in the PLATON opening menu. The PLUTONauto button will not work because the derived *name*.ELD will be read, rather than the *name*.RES file
- The input file must be recognised as 'shelx.res' format type.

There are two options for RENAME

##### 1. KEYBOARD

- REN C100 C1 C150 C2 C5 P1
- This will change C100 into C1 etc.

##### 2. VIA MENU: Click in CONTENT MENU on RENAME ATOM

- Click on ATOM to be renamed and give new name via KEYBOARD

#### 4.13 Default instructions for PLUTON

PLUTON (i.e. the PLUTON path through PLATON) starts to read data from an input file (e.g. *name*.RES ). That file usually, but not necessarily, contains the pertinent data for the structure only. Before switching to interactive input (console or menu) a file named *name*.DEF is read. This file may be used to execute a number of instructions before the first plot.

Example of *name*.DEF

```
STRAW COLOR
LABEL
PLOT
```

Such a file is generated automatically when PLUTON is run in the PLATON-compatibility mode via the PLUTONauto button.

#### 4.14 Graphics (Meta, POVRAY, RASMOL)

PLATON will generate by default a PostScript file of the image displayed on the screen by clicking on the 'EPS' button. The PostScript default may be changed into HPGL or TEK4014 with a button on the PLATON opening window. In the PLUTON mode, there are two additional graphical interfaces.

1. clicking on 'Pov' will generate a file *name*.POV suitable for the raytracing program POVRAY. POVRAY will be executed when implemented.
2. clicking on 'Ras' will generate a 'PDB' tructured *name*.RAS file suitable for the molecular visualisation program RasMOL. RasMOL will be executed when implemented.

The executables 'povray' and 'rasmol' are assumed to be globally accessible. Alternatively, a path may be given in environment variables.

```
e.g. 'setenv POVEXE '/usr/local/bin/povray'
      'setenv RASEXE '/usr/local/bin/rasmol'
```

#### 4.15 Output Listings .lis & .lps (Print & View)

PLATON produces a detailed 132 column listing file (line printer ASCII format) on a file with extension '.lis' in addition to the limited (80 column) listing that is sent to the display window from which the program was invoked. The listing file can be printed or read by a post-processing program to extract relevant data from it. The .lps file contains the same information as the .lis file but now in EPS-format. This file may be either sent to a PostScript printer or viewed with programs such as ghostview. Generation of the '.lps' file can be suppressed by clicking on the 'EPS-Listind' button in the PLATON-Opening menu.

#### 4.16 PLATON Main-Menu Options

The following options set special features: RED when active !

- To Pluton - Direct path to PLUTON function (Native-mode). Symmetry codes may be inconsistent with those in PLATON.
- Nomove - Coordinates are not transformed from input positions. This option is ON by default for CIF-Input data, assuming a connected set. Set 'Nomove' OFF when coordinates do not form a connected set.
- AtomSort - By default, the atom list is sorted on atom type and numerical value in the label.
- Organic - Different Radii are used for Organic/Inorganic Structures
- Round - By default, Coordinates and derived data are rounded following the 1-19 rule.
- Parentheses - By default, the numerical part of the label is enclosed in parentheses.
- Nosymm - When Active, No symmetry is applied.
- R/S-Determ - R/S-Determ is enforced when active.
- SHELXL ATWT - Mol.Weight data 'as-in' SHELXL
- Norm H-Bond - X-H distances are normalised to standard values
- Save-InstrS - Toggle for saving instructions
- Print-Level - (Not completely implemented yet)  
Click on leftmost position results in no listing file and on the rightmost one a full listing on '.lis'.
- EPS-Listing - Toggle to switch off the generation of a PostScript version on '.lps' of the '.lis' file.
- EPS HGL TEK - Default setting of the graphics output (PostScript, HP-GL or Tektronix 4014) on the Meta-file.
- Auto-Plot - New Plot on new instruction (on/off) toggle.
- Menu-off - Keyboard input only. Type 'MENU ON' to return to menu-mode.

#### 4.17 PLATON/ORTEP (Sub) Menu Options

The following options are available from the Ortep Menu (Red = Active)

- ToOptions - Path to a sub-menu
- StereoOpts - RedGreen and BlueGreen Stereo (Alternative = Mono)

- H-Atoms - In/Exclude H-Atoms toggle
- PLUTON-ELD - Direct path to PLUTON with the same connectivity.
- Probability - Displacement ellipsoid probabilities from 10 to 90 % in steps of 10%. The default is 50%.
- StyleB - (not implemented)
- Hetero El.s - Hetero atom octants shaded (3D ellipsoids).
- Envelope El - 2-D ellipsoids
- Octant El.s - Octant 3-D ellipsoids for all atoms.
- ViewOptions - Starting default orientations  
a) All angles zero
- Auto-Plot - Some option changes result in a redrawing of the image. This can be overrules by clicking this feature OFF.
- Label -Hat+ - Global label on/off toggle. Left position for non-hydrogen atoms only. Right click position includes H-atom labels,
- MoveLabel - When active, atom labels may be moved to a new position. Clicking on the lowerleft corner of a label will delete it to appear again at the clickposition of a second click.
- LabelSize - The labelsize can be changed as a function of the click position.
- DeleteLabel - Individual labels can be deleted by clicking on them.
- IncludLabel - Deleted labels (shown in red) can be reinstated by clicking on them.
- Resd012.. - By default, all residues are drawn. This corresponds to the leftmost click position. The other click positions bring up individual residues.
- CRotY - Continuous rotation with speed corresponding to click position.
- <<-RotZ+>> - Stepwise rotation about Z (perpendicular to image). Stepsize and direction based on click position.
- <<-RotY+>> - Stepwise rotation about Y (vertical in image) Stepsize and direction based on click position.
- <<-RotX+>> - Stepwise rotation about X (horizontal in image) Stepsize and direction based on click position.
- Color - Colour on/off toggle.
- Decoration - Include/Exclude Border text.
- EPS - A copy of the current display (excluding menu's) is written in EPS (PostScript) code to a file with extension '.ps'
- MenuOff End - Switch to Keyboard input (from Input Window) or terminate ORTEP menu.  
To start the menu feature again type 'MENU ON'

#### Remarks:

- The three numbers in the corners indicate the angles required to reconstruct the current orientation from the default the default orientation with a 'VIEW UNIT XROT xr YROT yr ZROT zr' instruction.

- The image can be refreshed by either hitting <cr> or clicking on the menu header

#### PLATON/ORTEP SUB-MENU

ToOptions - Back to main PLATON/ORTEP menu  
DisplayText - Display additional text toggle.  
NewText - Input on new text to move to its proper place subsequently.  
MoveText - Move Text item to new position  
Textsize - Change default textsize.  
DeleteText - Delete textitem by clicking on it.  
ChTextSize - Change size of a text item.  
ViewInvert - Invert to opposite absolute configuration.

#### 4.18 PLATON and CSD-FDAT files

PLATON can be run on multiple entry FDAT files. An END instruction loads the next entry from the FDAT file. Certain instructions can be saved in order to run them also for the next entries. This can be done by activating the 'SAVE-InstrS' button. An instruction sequence is terminated by clicking on END (or by typing END). e.g. in order to examine a set of entries:

```
Click 'SAVE-InstrS'  
Click 'ORTEP/ADP'
```

```
Click 'END'  
Click 'END'
```

### 5. Terms and notions

#### 5.1 The n.ijk symmetry operation on input

Atomic coordinates as input will be transformed in general by symmetry operations following certain rules. In the default automatic mode this will result in a connected set with residues properly positioned within the unit cell range. The symmetry operation applied to the input data will be listed under the header trans in the atomic coordinates listing and is encoded as n.ijk. n stands for the number of the symmetry operation as specified on the first page of the output listing and ijk for the unit cell translations in the three directions relative to 555: ijk=564 means 1 positive translation in the b direction, 1 negative translation in the c direction and none in the a direction.

The automatic mode transformation may be overruled for a given atom by preceding the data for that particular atom by a TRNS instruction e.g. TRNS 3.564. This facility may be used to determine the part of the molecule that is to be considered as the asymmetric part of a symmetrical molecule. The transformation to be applied only to the first atom as a starting point of a new residue can be forced with a negative symmetry transformation code e.g. TRNS -5.354. Its position in the input stream determines the atoms to which it will apply. The input stream may contain several of such instructions, each apply to the atoms that follow until overruled by a new one. Their effect will only be on atoms that are chosen to start a new residue.

## 5.2 Disorder

The program attempts to manage the problems that are encountered with several types of disorder. Only two-fold disorder is allowed. Populations higher than 0.5 are understood as major disorder components and those less than 0.5 as minor disorder components. The usual transformations on input coordinates are restricted. In general it will be necessary to supply disordered molecules as connected sets. The calculation of distances and angles etc. will extend only to entities involving the major disorder component or the minor disorder component but not both.

## 5.3 Molecules and Residues

The concepts of molecules and residues are related but not always synonymous within the PLATON realm. A residue is defined as a part of the structure that is connected by intra-molecular bonds only and is associated with a number. A structure may thus contain one or more residues. Residues may be chemically equivalent or chemically distinct. A molecule is defined as an asymmetric part of the structure connected by intra-molecular bonds only. Several molecules may join by crystallographic symmetry into one residue. A particular molecule is designated by a code: [nijk.rr] where n denotes the symmetry operation with respect to the basic molecule, ijk the translation with respect to 555 and rr the residue number. The structure of sucrose thus consists of two molecules (e.g. [1555.01] and [2545.01]) but only one residue.

## 6. Parameter files

The atomic parameters (including unit cell parameters, coordinates and temperature parameters) for a given structure may be input in various ways: the STANDARD PARAMETER FILE Format (SPF), SHELX type .INS/.RES files, simple XYZ Angstrom coordinate format, CIF, PDB, CSD-FDAT formats

The SPF-format is card image oriented. The first four characters on a card specify the nature of the data that follow on that card. Data that are not needed for the current program are simply skipped. All data are in free format.

### 6.1 The SPF format

#### **TITL *text***

This text may be used for various titling purposes. It may be overridden at any time by another TITL instruction.

#### **CELL (*wavelength*) *a b c alpha beta gamma***

Optional wavelength and cell parameters in Å's and degrees respectively. No CELL card is needed for Å data input. The wavelength is used for the calculation of the linear absorption coefficient.

#### **CESD *sig(a) sig(b) sig(c) sig(alpha) sig(beta) sig(gamma)***

This optional card specifies standard deviations in the cell parameters. No CESD card is needed for Angstrom data. The cell e.s.d. is combined with the coordinate e.s.d. for the calculation of the e.s.d. in derived parameters.

**SPGR** *space-group-name*

Space group symbol. See 9.4.7 below for more details.

**LATT** (*P/A/B/C/I/F*) (*A/C*)

First parameter specifies the Bravais lattice type and the second whether the lattice is acentric or centric.

**SYMM** *symmetry-operation*

Symmetry operation. See 9.4.7 below

**ATOM** *atom\_name x y z (pop) (sig(x) sig(y) sig(z)) (spop)*

This specifies the positional parameters, the population and their estimated standard deviations. The atom\_name should conform some rules in order to be acceptable since it is interpreted. The first one or two characters should correspond to an element name known to the program (see Appendix V). The number of characters of the element type and the attached digital number cannot exceed four. ' and " are allowed as part of an atom name. Labels not conforming with the PLATON-rules are modified in a new label including the symbol #. The atom-name may contain parentheses enclosing the numerical part.

**UIJ** *atom\_name U11 U22 U33 U23 U13 U12*

Anisotropic thermal parameters. Note the order of the components that is the same as in SHELX but often different in other systems (such as the XRAY and XTAL systems).

$$TF = \exp[-2*\pi^2(U11*h**2(a*)^2+...+2*U12*h*h*(A*)(B*)+...)]$$

**SUIJ** *atom\_name sig(U11) sig(U22) sig(U33) sig(U23) .. sig(U12)*

Estimated standard uncertainties (e.s.d's) for the anisotropic thermal parameters.

**U** *atom\_name U sig(U)*

Isotropic temperature factor along with its associate standard deviation.

**BIJ** *atom\_name Beta11 Beta22 Beta33 Beta23 Beta13 Beta12*

Anisotropic thermal parameters. Note the order of the components.

$$TF = \exp[-(Beta11*h^2+Beta22*k^2+...+2*Beta12*h*h+...)]$$

$$\text{Definition: } Beta11 = 2*\pi^2 *astar^2$$

$$Beta12 = 2*\pi^2*astar*bstar.$$

**SBIJ** *atom\_name sig(Beta11) .. sig(Beta23) .. sig(Beta12)*

Estimated standard deviations for the anisotropic thermal parameters.

**B** *atom\_name B sig(B)*

Isotropic temperature factor along with its associate standard deviation. Definition:  $B = 8 \cdot \pi^2 \cdot U$

**TRNS** *-n.klm*

Facility to influence the applied symmetry operation for the first atom in a new residue. (see appendix I)

**TRNS** *n.klm*

When placed in front of an ATOM card this instruction will transform the input coordinates on that card by the named symmetry operation: n is the number of the symmetry operation and k,l,m are the translations. (see 4)

**TRNS** *T11 T12 T13 T21 T22 T23 T31 T32 T33 (SH1 SH2 SH3)*

Transformation matrix on cell axis and origin shift to be applied to the data following (CELL parameters, atomic coordinates and thermal parameters).

Example:

```
TITL NI-COMPOUND
CELL NI .123 .544 -.176 1 .001 .002 .001 0.0
UIJ NI .011 .013 .025 -.011 .004 .009
SUIJ NI .001 .001 .002 .002 .002 .001
ATOM C1 .345 .675 -.334 1 .010 .009 .005 0.0
U C1 0.04 0.01
(etc)
```

**7 Space group symmetry**

Space group symmetry is handled in PLATON with a general space group symmetry management routine that permits the specification of the symmetry either explicitly in terms of the general equivalent positions as presented in the International Tables or implicitly in terms of space group generators. The generators for all space groups in their standard setting and some commonly used non-standard settings are also implicitly retrievable by the program from internal tables (see tables below) on the basis of the specified name of the space group (e.g. R-3m)

EXAMPLE: The symmetry for space group nr. 19 ( $P2_12_12_1$ ) may be specified either as:

```
LATT P A
SYMM X, Y, Z
SYMM 1/2 + X, 1/2 - Y, -Z
SYMM -X, 1/2 + Y, 1/2 - Z
SYMM 1/2 - X, - Y, 1/2 + Z
```

or

```
LATT P A
SYMM 1/2 + X, 1/2 - Y, -Z
SYMM -X, 1/2 + Y, 1/2 - Z
```

or

```
SPGR P212121
```

LATT card should precede any SYMM card in order that the symmetry arrays are initialised to either, by default, a primitive non-centrosymmetric lattice or to the specified lattice type: (P/A/B/C/I/F) and (A)Centric type (A/C). The general equivalent positions should be given as

specified in International Tables and should have the centre of symmetry at the origin, in the case that the space group is centrosymmetric. The symmetry operation SYMM X,Y,Z is always implicitly assumed as the first symmetry operation and needs not be given although any redundancy in the symmetry input will be ignored.

Note: Rhombohedral lattice types (in hexagonal setting) should be specified explicitly using an extra symmetry generator. Thus the generators for space group *R3* are:

```
LATT P A
SYMM -Y, X-Y, Z
SYMM 1/3+X, 2/3+Y, 2/3+Z
```

The same space group on rhombohedral axes should be specified as *R3R*.

The translation part may be specified either as a ratio or as a real (e.g. 1/4 or 0.25).

Monoclinic-b is taken as the standard setting for monoclinic space groups. Other settings are to be specified by the full space group name: e.g. *P112* for the monoclinic-c setting of *P2*.

Non-standard orthorhombic settings such as space group *A2aa* may be handled by specifying *Ccc2 -cba* on the SPGR card (see International Tables Vol A). In fact the program automatically modifies the input line accordingly for non-standard settings (see table below). The standard setting symmetry is then transformed accordingly.

Note: Symmetry may also be presented in the SHELX style. However a LATT card should always be supplied since the default symmetry of PLATON is always *P1* whereas SHELX defaults to *P-1*.

The names of the space groups known to the program are given in the following table and are in accordance with the usage in the CAMBRIDGE CRYSTALLOGRAPHIC DATA BASE files.

## 7.1 Space group names known to the program

P1	P-1	P2	P112	P21	P1121	C2	A2
B112	PM	P11M	PC	PA	PN	P11B	CM
AM	B11M	CC	IC	IA	AA	B11B	P2/M
P112/M	P21/M	P1121/M	C2/M	A2/M	B112/M	P2/C	P2/A
P112/B	P21/C	P21/A	P2/N	P1121/B	P21/N	P21/N11	C2/C
A2/A	C2/N	I2/C	I2/N	I2/M	I2/A	B112/B	P222
P2221	P2212	P2122	P21212	P212121	C2221	B2212	C222
F222	I222	I212121	PMM2	PMC21	PCC2	PMA2	PCA21
PNC2	PMN21	PBA2	PNA21	P21NB	PC21N	PN21A	PBN21
P21CN	PNN2	CMM2	CMC21	CCC2	AMM2	ABM2	AMA2
ABA2	FMM2	FDD2	IMM2	IBA2	IMA2	PMMM	PNNN
PCCM	PBAN	PMMA	PNNA	PMNA	PCCA	PBAM	PCCN
PBCM	PNNM	PMMN	PBCN	PBCA	PCAB	PNMA	PBNM
PMCN	PNAM	PMNB	PCMN	CMCM	CMCA	CMMM	CCCM
CMMA	CCCA	FMMM	FD3D	IMMM	IBAM	IBCA	IMMA
P4	P41	P42	P43	I4	I41	P-4	I-4
P4/M	P42/M	P4/N	P42/N	I4/M	I41/A	P422	P4212
P4122	P41212	P4222	P42212	P4322	P43212	I422	I4122
P4MM	P4BM	P42CM	P42NM	P4CC	P4NC	P42MC	P42BC
I4MM	I4CM	I41MD	I41CD	P-42M	P-42C	P-421M	P-421C
P-4M2	P-4C2	P-4B2	P-4N2	I-4M2	I-4C2	I-42M	I-42D
P4/MMM	P4/MCC	P4/NBM	P4/NNC	P4/MBM	P4/MNC	P4/NMM	P4/NCC
P42/MMC	P42/MCM	P42/NBC	P42/NNM	P42/MBC	P42/MNM	P42/NMC	P42/NCM
I4/MMM	I4/MCM	I41/AMD	I41/ACD	P3	P31	P32	R3
R3R	P-3	R-3	R-3R	P312	P321	P3112	P3121
P3212	P3221	R32	R32R	P3M1	P31M	P3C1	P31C
R3M	R3MR	R3C	R3CR	P-31M	P-31C	P-3M1	P-3C1
R-3M	R-3MR	R-3C	R-3CR	P6	P61	P65	P62
P64	P63	P-6	P6/M	P63/M	P622	P6122	P6522
P6222	P6422	P6322	P6MM	P6CC	P63CM	P63MC	P-6M2
P-6C2	P-62M	P-62C	P6/MMM	P6/MCC	P63/MCM	P63/MMC	P23
F23	I23	P213	I213	PM3	PM-3	PN3	PN-3
FM3	FM-3	FD3	FD-3	IM3	IM-3	PA3	PA-3
IA3	IA-3	P432	P4232	F432	F4132	I432	P4332
P4132	I4132	P-43M	F-43M	I43M	P-43N	F-43C	I-43D
PM3M	PM-3M	PN3N	PN-3N	PM3N	PM-3N	PN3M	PN-3M
FM3M	FM-3M	FM3C	FM-3C	FD3M	FD-3M	FD3C	FD-3C
IM3M	IM-3M	IA3D	IA-3D				

## 7.2 Nonstandard settings with cell transformations

ABC	BA-C	CAB	-CBA	BCA	A-CB
P2221	P2221	P2122	P2122	P2212	P2212
P21212	P21212	P22121	P22121	P21221	P21221
C2221	C2221	A2122	A2122	B2212	B2212
C222	C222	A222	A222	B222	B222
PMM2	PMM2	P2MM	P2MM	PM2M	PM2M
PMC21	PCM21	P21MA	P21AM	PB21M	PM21B
PCC2	PCC2	P2AA	P2AA	PBAB	PBAB
PMA2	PBM2	P2MB	P2CM	PC2M	PM2A
PCA21	PBC21	P21AB	P21CA	PC21B	PB21A
PNC2	PCN2	P2NA	P2AN	PB2N	PN2B
PMN21	PNM21	P21MN	P21NM	PN21M	PM21N
PBA2	PBA2	P2CB	P2CB	PC2A	PC2A
PNA21	PBN21	P21NB	P21CN	PC21N	PN21A
PNN2	PNN2	P2NN	P2NN	PN2N	PN2N
CMM2	CMM2	A2MM	A2MM	BM2M	BM2M
CMC21	CCM21	A21MA	A21AM	BB21M	BM21B
CCC2	CCC2	A2AA	A2AA	BB2B	BB2B
AMM2	BMM2	B2MM	C2MM	CM2M	AM2M
ABM2	BMA2	B2CM	C2MB	CM2A	AC2M
AMA2	BBM2	B2MB	C2CM	CC2M	AM2A
ABA2	BBA2	B2CB	C2CB	CC2A	AC2A
FMM2	FMM2	F2MM	F2MM	FM2M	FM2M
FDD2	FDD2	F2DD	F2DD	FD2D	FD2D
IMM2	IMM2	I2MM	I2MM	IM2M	IM2M
IBA2	IBA2	I2CB	I2CB	IC2A	IC2A
IMA2	IBM2	I2MB	I2CM	IC2M	IM2A
PCCM	PCCM	PMAA	PMAA	PBMB	PBMB
PBAN	PBAN	PNCB	PNCB	PCNA	PCNA
PMMA	PMMB	PBMM	PCMM	PMCM	PMAM
PNNA	PNNB	PBNN	PCNN	PNCN	PNAN
PNNA	PNMB	PBMN	PCNM	PNCM	PMAN
PCCA	PCCB	PBAA	PCAA	PBCB	PBAB
PBAM	PBAM	PMCB	PMCB	PCMA	PCMA
PCCN	PCCN	PNAA	PNAA	PBNB	PBNB
PBCM	PCAM	PMCA	PMAB	PBMA	PCMB
PNNM	PNNM	PMNN	PMNN	PNMN	PNMN
PMMN	PMMN	PNMM	PNMM	PMNM	PMNM
PBCN	PCAN	PNCA	PNAB	PBNA	PCNB
PBCA	PCAB	PBCA	PCAB	PBCA	PCAB
PNMA	PMNB	PBNM	PCMN	PMCN	PNAM
CMCM	CCMM	AMMA	AMAM	BBMM	BMMB
CMCA	CCMB	ABMA	ACAM	BBCM	BMAB
CMMM	CMMM	AMMM	AMMM	BMMM	BMMM
CCCM	CCCM	AMAA	AMAA	BBMB	BBMB
CMMA	CMMB	ABMM	ACMM	BMCM	BMAM
CCCA	CCCB	ABAA	ACAA	BBCB	BBAB
IBAM	IBAM	IMCB	IMCB	ICMA	ICMA
IBCA	ICAB	IBCA	ICAB	IBCA	ICAB
IMMA	IMMB	IBMM	ICMM	IMCM	IMAM

**8. Atomic radii used in PLATON**

Ac	1.88	Er	1.73	Na	0.97	Sb	1.46
Ag	1.59	Eu	1.99	Nb	1.48	Sc	1.44
Al	1.35	F	0.64	Nd	1.81	Se	1.22
Am	1.51	Fe	1.34	Ni	1.50	Si	1.20
As	1.21	Ga	1.22	Np	1.55	Sm	1.80
Au	1.50	Gd	1.79	O	0.68	Sn	1.46
B	0.83	Ge	1.17	Os	1.37	Sr	1.12
Ba	1.34	H	0.23	P	1.05	Ta	1.43
Be	0.35	Hf	1.57	Pa	1.61	Tb	1.76
Bi	1.54	Hg	1.70	Pb	1.54	Tc	1.35
Br	1.21	Ho	1.74	Pd	1.50	Te	1.47
C	0.68	I	1.40	Pm	1.80	Th	1.79
Ca	0.99	In	1.63	Po	1.68	Ti	1.47
Cd	1.69	Ir	1.32	Pr	1.82	Tl	1.55
Ce	1.83	K	1.33	Pt	1.50	Tm	1.72
Cl	0.99	La	1.87	Pu	1.53	U	1.58
Co	1.33	Li	0.68	Ra	1.90	V	1.33
Cr	1.35	Lu	1.72	Rr	1.47	W	1.37
Cs	1.67	Mg	1.10	Re	1.35	Y	1.78
Cu	1.52	Mn	1.35	Rh	1.45	Yb	1.94
D	0.23	Mo	1.47	Ru	1.40	Zn	1.45
Dy	1.75	N	0.68	S	1.02	Zr	1.56

Note: OW is equivalent to O and Q1 is equivalent to C1.

Covalent radii are those given in the Cambridge Structural data base manual.

**9. Internal parameters**

The program contains internal integer and real parameter arrays (IPR and PAR respectively). They include default parameter settings and values that may be either explicitly or implicitly manipulated with the (sub)keywords. Below is a list of some of them. Their values may be changed with SET PAR and SET IPR instructions or examined with LIST PAR and LIST IPR instructions. It should be noted that there is no checking for side-effects.

IPR(141) - Nplane parameter in TME

IPR(142) - Lines parameter in TME

PAR(73) - Letter size

**10. CIF-validation documentation (VALIDATION.DOC)**

This document should provide information on the CIF-validation tests as specified in the file 'check.def' and implemented in PLATON.

```
# check.def contains the tests, error thresholds and warning messages
#   issued by the program PLATON when run in the 'VALIDATION' mode
#
#       This file is read when PLATON is called with the '-u' switch
#       (e.g. 'platon -u sk1500.cif')
#
# - Test output will be in the order as defined implicitly below
#
```

```
# - Tests are identified with three-digit numbers
#   _0xx - general
#   _1xx - cell/symmetry
#   _2xx - adp-related
#   _3xx - intra geometry
#   _4xx - inter geometry
#   _5xx - coordination geometry
#   _6xx - void tests
#   _7xx - various test
#
```

**\_000-013 : Pseudo Requests**

This set of pseudo requests generates a listing of some relevant reported and calculated data including transmission range and completeness of the data set.

**\_020 : Check on R(int)**

R(int) should be well below 10% and in the order of magnitude of the reported R-values. R(int) may be relatively meaningless when based on a very limited number of averaged data.

**\_021 : Check ratio of observed/expected number of reflections**

The expected number of reflections corresponds to that in the asymmetric unit of the Laue group.  
Expected ratio: less-or-equal 1 for centro symmetric structures and less than 2 for non-centrosymmetric structures.

Reasons to exceed those numbers can be:

- 1 - Systematic extinction not omitted from the observed data count
- 2 - Refinement with redundant (i.e. not averaged/unique) data set.

**\_022 : Check for completeness**

Test for data completeness. The ratio of the reported number of unique reflections and expected for the resolution given.

**\_023 : Check Theta-max**

Check resolution of the data set. Alerts are issued when  $\sin(\theta)/\lambda < 0.6$

**\_024 : Check for Friedel Averaging Preferred**

Averaging of Friedel pairs is preferred for MoKa datasets containing atoms not heavier than Si. Use MERG 4 for SHELXL97 refinements.

**\_025 : Check h,k,l - range**

Check reported h,k,l - range with calculated range based on reported theta-max.

**\_026 : Check for weak data**

Check whether a sufficient fraction of the unique data is indeed above the 2 sigma level.

\_030 Check on meaningfully refined extinction parameter.

This test checks whether a refined extinction parameter is meaningful i.e. whether its value is significantly larger than its corresponding s.u. If not, this parameter should be removed from the model and the structure refined without this meaningless additional parameter. The current default gives a warning when its value is within 3.33 s.u.

\_032 Check S.U. of the Flack parameter

Check meaningfulness of Flack parameter

\_033 Check deviation of Flack parameter from zero

Check validity of absolute structure determination

\_040 Check for the presence of H-atoms in carbon containing compounds.

Alert for 'no H-atoms' in CIF

\_041 Compare the Reported and Calculated Sum Formula

\_042 Compare the reported and Calculated Moiety Formula

\_043 Compare Reported and Calculated Molecular Weight

Note: atomic weights used in the calculation of the molecular weight are those used in SHELXL97.

\_044 Compare Reported and Calculated density

\_045 Compare Reported and Calculated Z

\_046 Check on Reported Z, MW, D(calc) consistency

D(calc) as calculated from the reported Z and MW is compared for consistency with the reported d(calc).

\_050 Test for mu given

\_051 Test difference of reported and calculated mu

\_052 Test for specification of absorption correction method

\_053 Test for specification xtal\_dimension\_min

\_054 Test for specification xtal\_dimension\_mid

\_055 Test for specification xtal\_dimension\_max

\_056 Test for specification xtal\_radius

\_057 Test for absorption correction needed

\_058 Test for Tmax specified

\_059 Test for Tmin specified

\_060 RR Test

\_061 RR' Test

\_062 Rescale Tmin & Tmax

\_063 Test crystal size

Alert for crystals with at least one dimension probably too large for the homogeneous part of the Xray beam when used for data collection using crystal monochromated radiation. An exception will be datacollection using a beta-filter and a sufficiently large collimator.

\_064 Test for T(max) .GE. T(min)

\_065 Test for applicability of (semi-)empirical abs.corr.

\_070 Test for duplicate atom labels

\_071 Uninterpretable label

This alert relates to unintelligible labels on input and are deleted from the model & checking.

\_080 Test maximum shift/error

\_081 Test for maximum shift/error given

\_082 Test for reasonable R1 value

\_084 Test for reasonable R2 value

R2 will in general have a value twice of that of R1 with refinement on F\*\*2. Significantly larger values usually indicate a poor refinement model.

\_086, \_087 Test for reasonable S value

S should in general be close to 1 at the end of a refinement with a proper weighting scheme. If not, there might be significant unresolved problems with the model.

\_088 Test for reasonable Data / parameter ratio (centro)

\_089 Test for reasonable Data / parameter ratio (non-centrosymmetric)

\_095 Test for residual density maximum given

\_096 Test for residual density maximum given

\_097 Test maximum residual density

\_098 Test for minimum residual density

\_099 Test for minimum residual density less zero

\_11x : ADDSYM Tests for missed symmetry (with an expanded MISSYM (C) clone)

These tests warn for missed or possible higher (pseudo) symmetry in the structural model (i.e. based on the coordinate data). Close examination of the situation at hand is indicated in order to prove/disprove the issue (usually in combination with the reflection data).

The tests are split-up in three classes of problems:

\_110 Reports on potential (pseudo/real) lattice centering or cell halving.

\_111 Reports on potential additional (pseudo/real) inversion centres.

\_112 Reports on potential additional (pseudo/real) rotation axes and mirrors.

\_113 Report New spacegroup suggested by ADDSYM

Note: Disordered atoms are not taken into account in the tests.

\_12x : Tests for consistency and completeness of symmetry data items in CIF.

Symmetry in the CIF should be provided in the CIF both explicitly with a `_symmetry_equiv_pos_as_xyz` loop and implicitly with `_symmetry_space_group_name_H-M`.

The tests done include (with some redundancy to point to the real problem)

`_120` Test for consistency of the `_symmetry_equiv_pos_as_xyz` set with `_symmetry_space_group_name_H-M`.

`_121` Test for valid `_symmetry_space_group_name_H-M` symbol.

`_122` Test for missing (i.e. ?) `_symmetry_space_group_name_H-M` symbol.

`_123` Test for uninterpretable or inconsistent Spacegroup information.

`_124` Test for uninterpretable or absent explicit symmetry records.

`_130 - _140` : Test for symmetry restrictions on cell-parameters.

Symmetry restrictions on cell dimensions are checked.

`_141 - _146` : Test Su's on cell parameters

The presence of su's (where required) and value are checked. Su's as given by the diffractometer software are often much smaller than realistic.

`_150` Check Reported versus Calculated volume

An alert is issued when the reported unit cell volume differs significantly from the volume calculated on the basis of the supplied cell dimensions.

`_16x` : Tests for missing s.u. on coordinates or refined C-H

Positional parameters for all non-hydrogen atoms in general positions are checked for the presence of a non-zero s.u. on them. This includes parameters fixed to fix the origin in polar space groups which is no longer necessary when refinement is done with modern programs (e.g. SHELXL, XTAL).

`_161` Missing or Zero su (esd) on x-coordinate.

`_162` Missing or Zero su (esd) on y-coordinate.

`_163` Missing or Zero su (esd) on z-coordinate.

`_164` Refined C-H H-atoms in heavy-atom structure (warning)

Such H-atoms are better refined at calculated positions riding on the atoms they are attached to.

`_201` Test for isotropic non-hydrogen atoms in main residue (molecule)

This test reports on non-hydrogen atoms that were refined with isotropic displacement parameters only in the main residue. Such a practice is unusual by modern standards and only needed for minor disorder modelling.

`_202` Test for isotropic non-hydrogen atoms in anion/solvent (small moiety)

This test reports on isotropically refined atoms in small moieties (usually anions or solvent)

`_211` Test for NPD ADP's in main residue(s)

This test reports on non-positive definite (i.e. complex and unrealistic) anisotropic displacement parameters in the main residue.

\_212 Test for NPD ADP's in small moieties (solvent/anions)

This test reports on non-positive definite (i.e. complex and unrealistic) anisotropic displacement parameters in small moieties (solvent/anion).

\_213 ADP maximum/minimum ratio test for main residue

The maximum and minimum main axis ADP ratio (Angstrom Units) is tested for the main residue. Large values may indicate unresolved disorder.

\_214 ADP maximum/minimum ratio test for small moiety

The maximum and minimum main axis ADP ratio (Angstrom Units) is tested for the minor residue(s). Large values may indicate unresolved disorder.

\_220 Test Ueq(max)/Ueq(Min) range for non-H in main residue(s)

This test reports on a larger than usual U(eq) range for the non-hydrogen atoms. Too high or too low Ueq's may be an indication for falsely identified atomic species (i.e. Br versus Ag)

\_221 Test Ueq(max)/Ueq(Min) range for non-H atoms in minor residue(s)

This test reports on a larger than usual U(eq) range for the non-hydrogen atoms. Too high or too low Ueq's may be an indication for falsely identified atomic species (i.e. Br versus Ag)

\_222 Test Ueq(max)/Ueq(Min) range for H atoms in main residue(s)

This test reports on large ranges in displacement parameters for hydrogen atoms.

\_223 Test Ueq(max)/Ueq(Min) range for H atoms in minor residues(s)

This test reports on large ranges in displacement parameters for hydrogen atoms.

\_230, \_233 : Hirshfeld rigid bond test

Components of anisotropic displacement parameters along chemical bonds are assumed to be equal in magnitude. Large differences supposedly indicate contamination of these parameters with other (unresolved) effects such as disorder and/or overrefinement. Atom sites assigned the wrong scattering type (e.g. Ag versus Br) should generate 'problem signals' with this test.

Data sets corrected for absorption effects with DELREF techniques (e.g. DIFABS, SHELXA, XABS2) often show large DELU values for bonds involving the heaviest atom.

\_241, \_242 Test for too high/low U(eq) as compared to neighbours

The U(eq) value of an atom is compared with the average U(eq) for non-hydrogen atoms bonded to it. Large differences may indicate that the wrong atom type was assigned (e.g. N instead of O).

\_301, \_302 : Detection of Disorder

Atom sites that are not fully occupied are counted. A large fraction of disordered atoms may be both a signal for serious structure determination problems or less reliable/interesting results. A distinction is made for 'solvent molecules of less interest' (\_302) and the 'main components' (\_301)

**\_305 Test for 'Isolated' Hydrogen atoms**

This test reports on hydrogen atoms that are not on bonding distance to any atom.

**\_306 & \_311 Test for 'Isolated' Oxygen atoms**

This test reports on oxygen atoms that are not within bonding distance to any other atom in the structure. A common reason may be that no hydrogen atoms are given for a water molecule.

Alert \_306 is issued for full weight atoms, \_311 otherwise.

**\_307 Test for 'Isolated' Metal atoms**

This test reports on metal atoms that are not bonded or at coordination distance of other atoms. Isolated ions are very unusual (or non-existent ?)

**\_308 Test for single bonded metal atoms**

This test reports on single bonded (coordinated) metal atoms/ions. This represents a very unusual situation. There are literature examples where such a 'single bonded metal' was shown to be a halogen.

**\_310 Test for 'Too close' (symmetry related) full weight atoms.**

This test identifies (very) short contacts between atoms that only becomes apparent after the application of symmetry on the primary coordinate set.

**\_320, \_321 : Checks on Hybridisation of C atoms**

This test is split up in two sections: the main molecule (\_320) and solvent molecules (\_321).

The test attempts to assign one of three hybridisations to C atoms: sp, sp<sup>2</sup> or sp<sup>3</sup> on the basis of the angles around C. In this way, missing H atoms or too many H-atoms on a carbon atom should be detected.

**\_350 : \_373 : Test for Short/Long Bonds as compared with standard values**

Large deviations from generally accepted values may indicate model problems, overrefinement etc. Default (X-Ray) values are from SHELXL (X-H) and Ladd & Palmer, Structure Determination by Xray Crystallography (1985).

Note: C4-C3 indicates a bond between an atom with 4 bonds and one with 3 bonds. Conjugated systems may cause some 'false alarm' messages.

_350 Short C - H	Reference Xray : 0.96 Ang
_351 Long C - H	0.96
_352 Short N - H	0.87 Ang.
_353 Long N - H	0.87
_354 Short O - H	0.82
_360 Short C4 - C4	1.54

_361 Long C4 - C4	1.54
_362 Short C4 - C3	1.52
_363 Long C4 - C3	1.52
_364 Short C4 - C2	1.46
_365 Long C4 - C2	1.46
_367 Long C? - C?	1.50
_368 Short C3 - C3	1.34
_369 Long C3 - C3	1.34
_370 Short C3 - C2	1.31
_371 Long C3 - C2	1.31
_372 Short C2 - C2	1.25
_373 Long C2 - C2	1.25

**\_380 Check for Methyl Moiety refinement**

This test alerts for possible unrefined CH<sub>3</sub> moieties. (E.g. AFIX 33 instead of AFIX 137 etc. within the SHELXL realm)

**\_390, \_391 Test for unusual CH<sub>3</sub> geometry.**

**\_390 Reports on unusual X-C-H Angles**

**\_391 Reports on unusual H-C-H Angles**

Methyl moieties easily refine to poor geometries when refined in a poorly defined electron density or due to orientational disorder. If so, refinement with relevant con-/restraints should be considered.

**\_410, \_413 : Tests for Short Intra and Intermolecular H...H Contacts**

Short intermolecular H..H contacts may indicate incorrectly determined structures (i.e. wrong symmetry, missed translation symmetry, wrong position with reference to the symmetry elements, hydrogen atoms on atoms where there should not be any etc..). Short intramolecular contacts may arise when H-atoms are in (false) calculated positions. Short intramolecular contacts may also be a sign for a false structure with the molecule sitting on a site with improper site symmetry (e.g. '2' instead of '-1') which may happen when a lattice translation is missed. Short contacts are defined using a van der Waals radius of 1.2 Angstrom. For intermolecular contacts, an alert is generated for contacts less than 2.4 Angstrom. For intramolecular contacts alerts are generated for contacts less than 2.0 Angstrom.

Shorter intramolecular contacts may make sense for crowded configurations, in particular when hydrogen atoms are calculated in idealised positions.

\_412 & \_413 report on short H .. H contact involving CH<sub>3</sub> H-atoms. Those positions are often optimal when in calculated position.

**\_416 : Test for short intra D-H .. H-D contacts**

Such contacts may be related to disordered or misplaced H-atoms.

**\_417 : Test for short inter D-H .. H-D contacts**

Such contacts may be related to disordered or misplaced H-atoms.

**\_420 : Test for D-H without acceptor**

Potential hydrogen bond donors are checked for the presence of suitable acceptors using commonly used (Jeffrey) H-bond criteria. As a general rule there should be an acceptor for each donor. Exceptions are very rare for O-H and more common for -NH and -NH<sub>2</sub>. A common error is an -OH on a calculated position pointing in the wrong direction.

**\_430 : Test for short intermolecular D..A distances**

This test alerts for possibly missed hydrogen bonds as indicated by short (i.e. shorter than sum of the van der Waals radii - 0.2) donor - acceptor distances.

**\_431 : Test for short intermolecular H1...D distances**

This test reports on short intermolecular halogen .. donor/acceptor atom-type distances.

**\_432 : Test for short intermolecular distances**

This test reports on short intermolecular atom-type distances falling in another class than alerts \_430, \_431.

**\_601, \_602 : Solvent Accessible Void(s) Test(s)**

Crystal structures in general do not contain large solvent accessible voids in the lattice. Most structures lose their long-range ordering when solvent molecules leave the crystal. Only when the remaining network is strongly bonded (e.g. zeolites and some hydrogen bonded networks) the crystal structure may survive. Residual voids in a structure may indicate the omission of (disordered) density from the model. Disordered density may go undetected when smeared since peak search programs are not designed to locate maxima on density ridges. The presence or absence of residual density in the void may be verified on a printed/plotted difference Fourier map or with PLATON/SQUEEZE. Voids of 40 Å<sup>3</sup> may accommodate water. Small molecules such as tetrahydrofuran have typical volumes in the 100 to 200 Å<sup>3</sup> range.

\_601 : This test reports the volume of the largest solvent accessible void in the structure.

\_602 : This test reports on a void, too large for the current PLATON version for more detailed analysis.

**\_701, \_702, \_703 Check for consistency of derived geometry and coordinates**

The bond distances, angles and torsion angles given in the CIF are cross-checked with corresponding values calculated from the coordinates. Alerts are set at 1, 2 and 3 sigma deviation levels.

Note: Default s.u.'s are used where no su given (e.g. for C-H)

In general, all differences should be within the associated su. Small differences may arise from rounding. Very large deviation (or zero distance) normally indicate incorrectly specified symmetry operations on the associated atoms, or 'cut-and-pasting' of incompatible CIF's.

**\_704 Check for consistency of contact distance geometry and coordinates**

Intermolecular contacts listed in the CIF are checked against the coordinates in the CIF. Alerts are set at 1,2 and 3 sigma deviation levels.

`_705, _706, _707, _708` Check of H-Bond geometry against coordinates

Hydrogen-Bond data (D-H, H..D, D..A, D-H..A) listed in the CIF are checked. Alerts are set at 1,2 and 3 sigma deviation levels.

`_710` Check for linear torsions

Torsion angles specified in the CIF are checked for the 'linear variety' where one or both of the 1-2-3 and 2-3-4 bond angles are close to 180 Deg. SHELXL97 will generate those 'torsions' for molecules containing linear moieties (E.g. metal-C=O).

`_711 to _718` Inconsistent labels

When labels are found on geometry items (bonds, angles etc.) that are not in the coordinate list, and alert `_71n` is issued, related to alert `_70n`.

`_721, _722, _723, _724` Check for consistency of derived geom. and coordinates

Same as `_701, _702, _703, _704`, but for distances, angles and torsions without su (esd). Difference is tested in terms of Angstroms and Degrees.

`_725, _726, _727, _728` Check of H-Bond geometry against coordinates

Same as `_705, _706, _707, _708`, but for distances and angles without su (esd). Differences are tested in terms of Angstrom and Degrees.

`_731 to _738` Check on 'Comparable' reported and calculated s.u.

Same as `_701 to _708` but for reported and calculated s.u.'s (Ratio)

This range of checks should warn for erroneous rounding:

E.g.           105.5(19) to 105.5(2)  
              105.0(5) to 105(5)

etc.

`_741 to _748` Check for missing s.u.'s on bonds, angles, torsions

An alert in this series generally indicates a missing s.u. on the reported bond, angle, torsion.

`_751 to _758` Check for senseless s.u.'s on Bonds, Angles, Torsions

Alerts are issued for constrained bonds, angles, torsions (i.e. with the R flag in the CIF or of the type X-Y-X' = 180 where Y is on an inversion center).

`_761, _762` Check for presence of X-H bonds and X-Y-H or H-Y-H angles

The CIF should also contain those data.

`_763, _764` Check for too few or too many non X-H Bonds in CIF

`_780` Check for Connected Atom SET

Atoms given in a CIF should form a 'connected set', i.e. no symmetry operations are needed to get atoms in a bonding position. A connected set of atoms is not needed for the refinement (unless hydrogen atoms are to be added at calculated positions). Geometry listings (bonds, angles, torsions & H-bonds) become unwieldy for non-connected atom sets.

_790 Check for Residue Centres to be within the Unitcell bounds
---

Unless for a good reason, molecular species should be transformed (by symmetry and/or translation) so that their centres of gravity are close to or within the unitcell bounds.

This is a strict rule for the main species. Deviations from this general rule are smaller additional species when relevant for intermolecular interactions with the main species.

## 11. SQUEEZE, An effective cure for the disordered solvent syndrome in crystal structure refinement.

### 11.1 General

See P. v.d. Sluis & A.L. Spek. *Acta Cryst.* (1990), **A46**, 194

The current version of SQUEEZE has been designed, dimensioned and tested for small moiety structures containing disordered solvent molecules of the type toluene, CH<sub>2</sub>Cl<sub>2</sub>, tetrahydrofuran, water, methanol etc.. Anions may be treated in the same way. However, take care of the problem of uncertainty of charge balance.

Large voids currently require significant computing in the stage where the size and shape of the solvent accessible void is determined. All calculations are done in the triclinic system (data are expanded automatically when appropriate). Reflection data and FFT-array are stored in memory i.e. large structures (and high symmetry) may require large amounts of memory (change parameter NP21, that defines the available scratch area, globally to a larger value).

### 11.2 Implementation and Use

SQUEEZE as implemented in the program PLATON should be compatible with small-moiety structure refinement usage of the program SHELXL-97. The program is used as a filter. Input files are :

- *name*.HKL SHELX file (HKLF type 4)
- *name*.RES SHELX file (complete set of refined model parameters, including hydrogen atoms but excluding any dummy atoms used to describe the disorder region)

Invoke the program by giving, on the prompt >> (or click on the menu) the instruction:  
CALC SQUEEZE

The result will be in two files:

- *name*.LIS a listing file giving details of the calculations
- *name*.HKP: a modified reflection file against which the ordered structure parameters can be refined (the solvent contribution has been eliminated from the reflection data.)

In order to run SHELXL-97 on the 'solvent-free'  $F_o^2$  data you will need to rename the file *name*.HKP to *name*.HKL In order to get an .FCFstyle file ( $F_o^2 + F_c^2$  (model + solvent)) you will need to run PLATON with the option  
CALC FCF

Final R-values are reported on the basis of the WGHT parameters in the *name*.RES file. There will be a difference in reflection count as compared with the SHELXL-run due to the differing number of surviving 'observed' reflections. The procedure (starting from the original reflection data) can be repeated using the newly refined parameters when desired. This may define a 'refined' void area. However, there will be rarely a need to repeat the procedure. The 'difference-map' improvement potential of this technique has been demonstrated for small molecule structures. The technique should also work for protein data. However, this has not been tested by us as yet with PLATON/SQUEEZE. Current design features may cause problems when tried.

A directory containing test-examples is in (the PLATON ftp site at)  
/pub/unix/platon/TEST/SQUEEZE\_TEST.'

### 11.3 Notes on SQUEEZE

- The record length of the *name*.HKP file has been increased above 80 to accommodate additional data (the exact details are not fixed)
- The SHELXL-TWIN instruction is not available as yet in PLATON
- PLATON internally calculates structure factors for the model given in the shelxl.ins file.
- The solvent contribution to the structure factors is taken as 'error-free'. This means that the 'solvent-free'  $F_o^2$  keep their original su's (esd's)
- The method relies heavily on the quality of the low-order reflections. The dataset should be as complete as possible. Systematic errors may hamper the quality of the results.

Current conditions for applicability are

1. Reasonable data-resolution (say 25 degrees Mo)
2. Structure of the known part completed with H-atoms
3. No more than 26 'voids' in the unit-cell
4. Disorder density should be well outside the van der Waals surface of the known structure
5. The area to be SQUEEZED should not be too large (say less than ~30% of the unit-cell volume).

### 11.4 Interpretation of the results

1. A successful application of SQUEEZE will show the following results:
  - A new hkl-file against which a satisfactory refinement of the discrete model can be done (purpose: good geometry, good R-value)
  - Smooth convergence of the SQUEEZE iteration.
  - Significant improvement of the R-value in low resolution data. (see table at the end of the listing file).
  - The number of electrons reported to be found in a void is close to that expected for the assumed solvent.
  - The difference map peaklist should not contain significant peaks outside the VOID areas.
  - Peaklist is given in the file *name*.SQZ

2. Problems are indicated when:

- Convergence is unstable
- A large number of reflections left out during the iteration process (This may be well indicative for problems with the data, and should be checked for).

3. A report on the use of SQUEEZE should always state for each (significant) independent void:

- Where the void is (i.e. x,y,z)
- Its volume
- The number of electrons recovered.

## 12 PLATON in the *WinGX* implementation

The version of PLATON in *WinGX* is virtually identical to the UNIX workstation version, except that System S is not available (it is too UNIX specific). The major differences concern the input and output files. To maintain the normal *WinGX* file naming protocols, the ASCII listing file is renamed as PLATON.LST while the input files are generally created automatically. In normal operation, a temporary file called *name*.SPF is created by reading, in the following order, the SHELXL.LST (to try and obtain errors on the coordinates), the *name*.RES file or the *name*.INS file. If there is some problem with any/either of these files, the program may abort. The first thing to check is whether the SHELXL.LST file is from a refinement which has terminated normally - if not delete this file. The output *name*.HKP files are also usually renamed depending on the operation carried out. The program is ALWAYS run in locked mode because of this renaming procedures. The hypertext help functions exactly as in the UNIX version. .