Structure Comparison, Analysis and Validation

Ton Spek
National Single Crystal Facility
Utrecht University

Overview

This lecture lists and discusses the various tools and descriptors that are available for the analysis and validation of a single crystal study as implemented in the PLATON program.

Structure Analysis

- Analysis of the Intra-molecular Geometry
- Analysis of the Inter-molecular Geometry
- Analysis of the Coordination Geometry
- Bond Valence Model (Brown et al.)
- 'CALC ALL' - LISTING

Intra-molecular Geometry

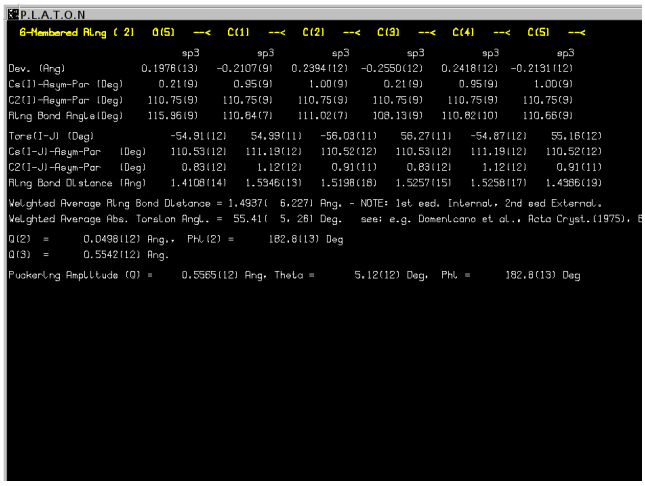
- Generation of the symmetry expended Connected Set on the basis covalent radii plus a tolerance..
- Special tolerances are applied for certain types of X-Y bonds/contacts, either to include or avoid them.
- Residues grow from a starting atom by recursive spherical expansion.

Intra-molecular Geometry

- Detection of residues and derivation of the Moiety formula, Z and Z'.
- Bond distances, Bond Angles, Torsion Angles.
- Automatic ring search,automatic seach of planar parts in the structure

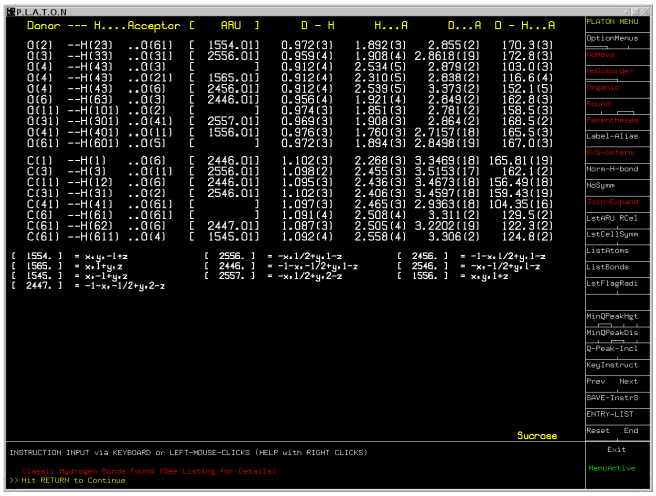
Intra-Molecular (Continued)

- Determination of the hybridization, R/S assignments and ‘topology numbers’.
- Listing of the plane-plane and bond-plane angles.
- Ring puckering analysis (Cremer & Pople)



Inter-Molecular

- Hydrogen Bonds (linear, bi- and trifurcated)
- Automatic analysis in terms of 1, 2 and 3-D networks (aggregates or cooperative)
- Search for pi-pi and C-H..pi interactions

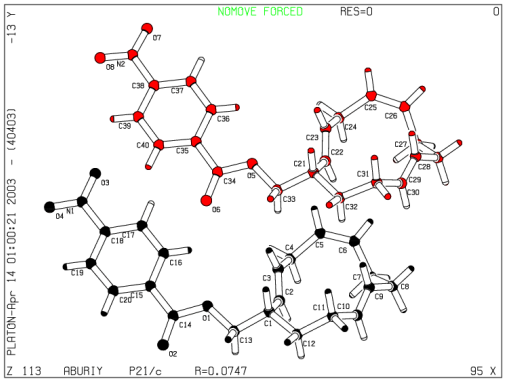


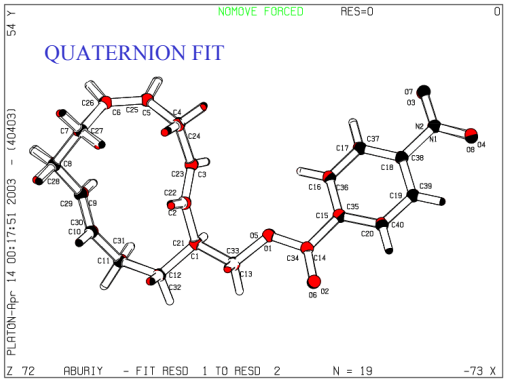
Structure Comparison

- Quaternion Fitting
  - Modified version of A.L.Mackay (1984), A40,165-166 (Note: 180 degree singularity)
  - Alternative: S.K.Kearsley (1989), A45, 208-210.
- Comparison of Simulated Powder Patterns
- StructureTidy (Inorganics)

QUATERNION FIT

- In many cases, an automatic molecule fit can be performed
- A) Identical atom numbering
- B) Sufficient number of Unique Atoms
- C) By manual picking of a few atom pairs



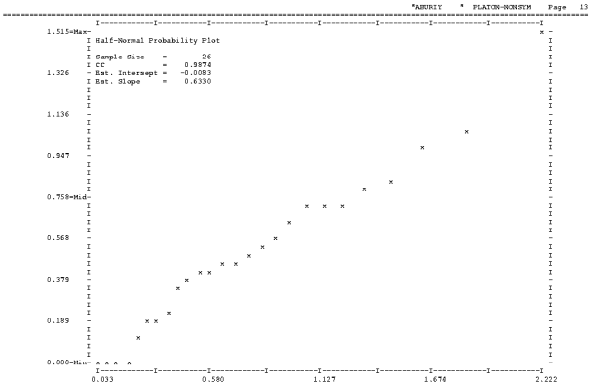



```
==============================================================================
Molfit with Quaternion Method (A.L. Mackay, Acta Cryst.(1984), A40, 165-166)
==============================================================================

Fit Rotation angle about (Pseudo)axis [l,m,n] =        3.29 Degree
Direction Cosines with Orthogonal Cell l,m,n  =    0.436719  0.830383  0.346553
Components in crystal system                       0.394049  1.000000  0.083576

Transf. Orthogonal Coord. Mol1     Orth. Coord. Mol2 with Resp. to C.G. Dist (A)
------------------------------------------------------------------------------
 O(1)      1.224  -0.879  -1.850   O(5)      1.198  -0.889  -1.868       0.033
 O(2)      3.035  -1.529  -2.978   O(6)      3.008  -1.572  -2.964       0.052
 N(1)     -0.456   1.596  -7.411   N(2)     -0.369   1.657  -7.412       0.106
 C(1)      0.535  -1.424   0.352   C(21)     0.498  -1.430   0.337       0.041
 C(2)      0.325   0.021   0.707   C(22)     0.253   0.000   0.695       0.076
 C(3)     -0.645   0.793   0.325   C(23)    -0.743   0.741   0.291       0.116
 C(4)     -0.844   2.220   0.754   C(24)    -0.972   2.175   0.700       0.146
 C(5)     -2.124   2.389   1.499   C(25)    -2.203   2.340   1.516       0.095
 C(6)     -2.269   2.324   2.800   C(26)    -2.291   2.305   2.823       0.037
 C(7)     -1.239   2.048   3.830   C(27)    -1.178   2.089   3.790       0.084
 C(8)     -1.452   0.723   4.578   C(28)    -1.303   0.782   4.602       0.162
 C(9)     -1.384  -0.460   3.664   C(29)    -1.310  -0.424   3.725       0.102

                              ... etc ...

:: Weighted and Unit Weight RMS-Fit =     0.08729   0.08194 Angstrom
```

Cg1 0.946 0.234 0.592

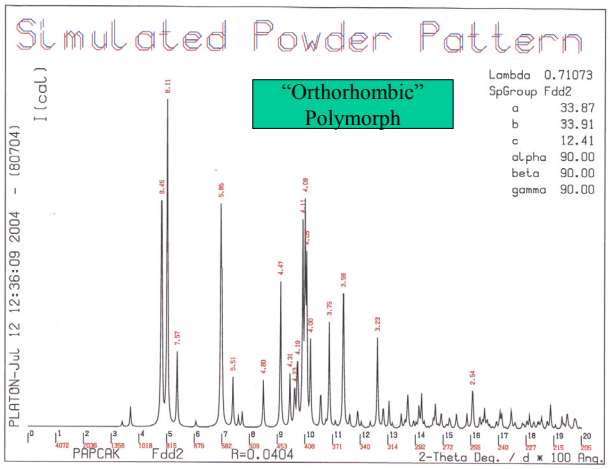
Cg2 0.441 0.253 0.581

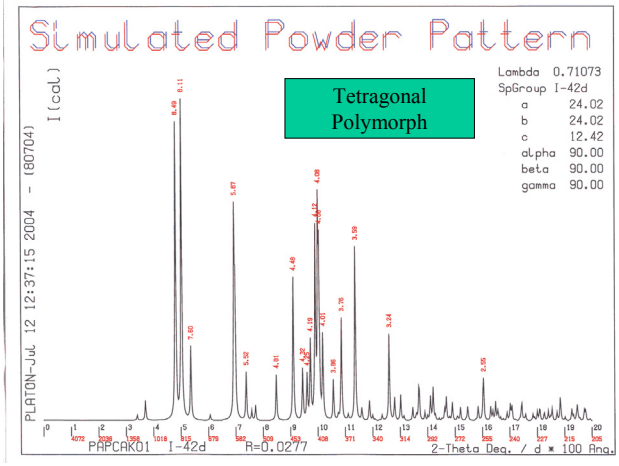
```
                                                                        "ABURI
==============================================================================
Comparison of the Bonds of the Fitted Residues
==========================================

Resd#1              Resd#2                Dist#1        Dist#2      Diff  Diff/Sig
------------------------------------------------------------------------------
O(1)    -C(13)      O(5)    -C(33)      1.449(6)      1.440(6)    0.0090    1.0607
O(1)    -C(14)      O(5)    -C(34)      1.326(6)      1.320(6)    0.0060    0.7071
O(2)    -C(14)      O(6)    -C(34)      1.198(7)      1.198(6)         0         0
O(3)    -N(1)       O(7)    -N(2)       1.225(7)      1.210(7)    0.0150    1.5152
O(4)    -N(1)       O(8)    -N(2)       1.217(7)      1.210(7)    0.0070    0.7071
N(1)    -C(18)      N(2)    -C(38)      1.477(8)      1.475(7)    0.0020    0.1881
C(1)    -C(2)       C(21)   -C(22)      1.502(7)      1.494(7)    0.0080    0.8081
C(1)    -C(12)      C(21)   -C(32)      1.523(7)      1.530(7)   -0.0070   -0.7071
C(1)    -C(13)      C(21)   -C(33)      1.504(7)      1.507(6)   -0.0030   -0.3254
C(2)    -C(3)       C(22)   -C(23)      1.297(7)      1.306(6)   -0.0090   -0.9762
C(3)    -C(4)       C(23)   -C(24)      1.503(7)      1.508(7)   -0.0050   -0.5051
C(4)    -C(5)       C(24)   -C(25)      1.491(8)      1.487(7)    0.0040    0.3763
C(5)    -C(6)       C(25)   -C(26)      1.311(9)      1.311(8)         0         0
C(6)    -C(7)       C(26)   -C(27)      1.483(8)      1.489(8)   -0.0060   -0.5303
C(7)    -C(8)       C(27)   -C(28)      1.537(8)      1.544(7)   -0.0070   -0.6585
C(8)    -C(9)       C(28)   -C(29)      1.496(8)      1.491(8)    0.0050    0.4419
C(9)    -C(10)      C(29)   -C(30)      1.284(8)      1.284(8)         0         0
C(10)   -C(11)      C(30)   -C(31)      1.496(8)      1.501(7)   -0.0050   -0.4704
C(11)   -C(12)      C(31)   -C(32)      1.520(8)      1.526(7)   -0.0060   -0.5644
C(14)   -C(15)      C(34)   -C(35)      1.483(7)      1.483(7)         0         0
C(15)   -C(16)      C(35)   -C(36)      1.393(7)      1.395(6)   -0.0020   -0.2169
C(15)   -C(20)      C(35)   -C(40)      1.387(7)      1.383(7)    0.0040    0.4041
C(16)   -C(17)      C(36)   -C(37)      1.375(7)      1.374(7)    0.0010    0.1010
C(17)   -C(18)      C(37)   -C(38)      1.383(8)      1.374(7)    0.0090    0.8467
C(18)   -C(19)      C(38)   -C(39)      1.374(7)      1.372(7)    0.0020    0.2020
C(19)   -C(20)      C(39)   -C(40)      1.378(7)      1.374(7)    0.0040    0.4041

:: RMS Bond Fit =      0.0060 Ang.
```

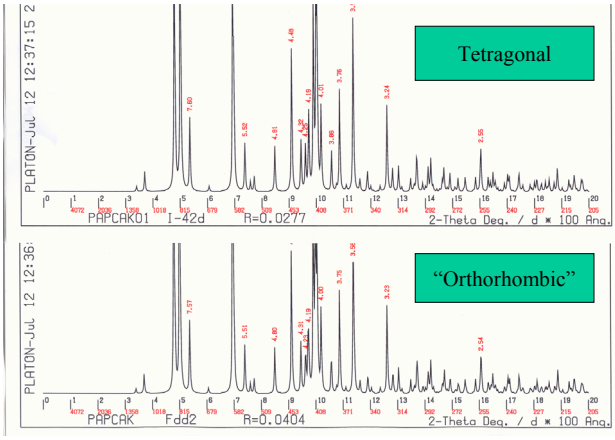


Simulated Powder Patterns

- It is not always apparent that two crystal structures are identical. The assigned unit cell or space group can differ.
- Comparison of the associated calculated powder patterns should solve the issue.
- Example for the CSD:

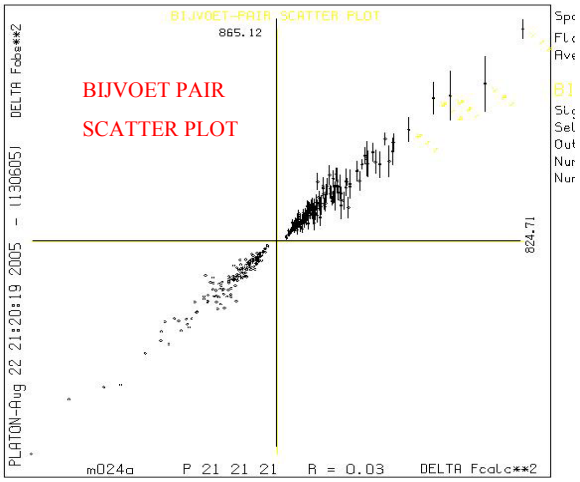






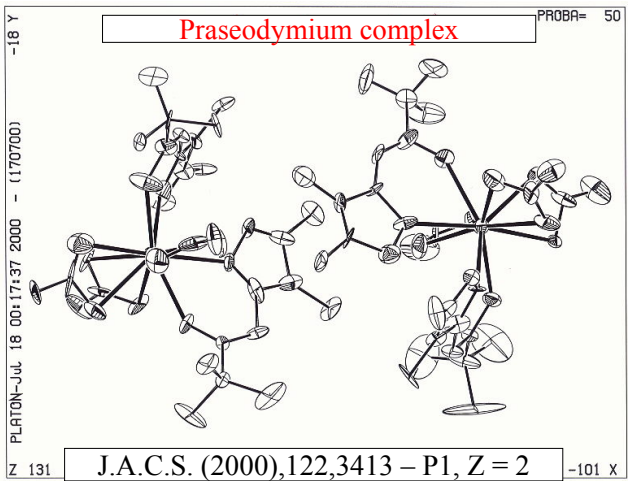
Absolute Structure

- The absolute structure of a compound is normally determined with the refinement of the Flack parameter.
- The value of the Flack parameter can be inconclusive in view of a high su.
- A detailed scatter-plot may be more conclusive.

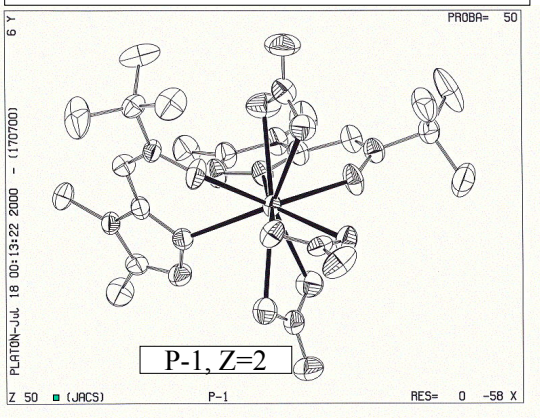


Validation

- ORTEP
- IUCr – CHECKCIF Structure Validation
- FCF- Validation (Completeness & Twinning)



CORRECTLY REFINED STRUCTURE



STRUCTURE VALIDATION

Single crystal structure validation addresses three important questions:

1 – Is the reported information complete?
2 – What is the quality of the analysis?
3 – Is the Structure Correct?

IUCR-CHECKCIF

IUCR-TESTS:
- MISSING DATA, PROPER PROCEDURE, QUALITY

PLATON TESTS:
- SYMMETRY, GEOMETRY, DISPLACEMENT PARAMETERS

ALERT LEVELS:
- ALERT A - SERIOUS PROBLEM
- ALERT B - POTENTIALLY SERIOUS PROBLEM
- ALERT C - CHECK & EXPLAIN

Problems Addressed by PLATON

- Missed Higher Space Group Symmetry
- Solvent Accessible Voids in the Structure
- Unusual Displacement Parameters
- Hirshfeld Rigid Bond test
- Miss-assigned Atom Type
- Population/Occupancy Parameters
- Mono Coordinated/Bonded Metals
- Isolated Atoms

Problems Addressed by PLATON

- Too Many Hydrogen Atoms on an Atom
- Missing Hydrogen Atoms
- Valence & Hybridization
- Short Intra/Inter-Molecular Contacts
- O-H without Acceptor
- Unusual Bond Length/Angle
- CH3 Moiety Geometry

Validation with PLATON

- Details: www.cryst.chem.uu.nl/platon
- Driven by the file **CHECK.DEF** with criteria, ALERT messages and advice.
- Use: **platon –u structure.cif**
- Result on file: **structure.chk**
- Applicable on CIF's and CCDC-FDAT
- FCF-Valid: **platon –V structure.cif**

BAMYEU

Dalton Trans
2003,134-140

Cc

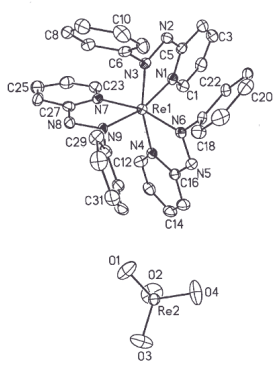
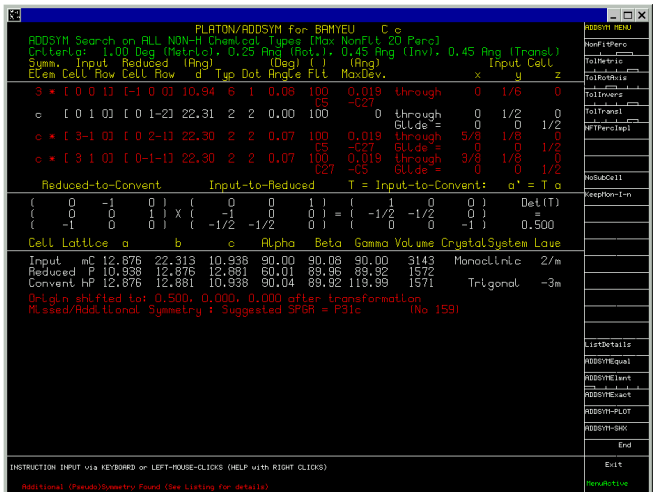
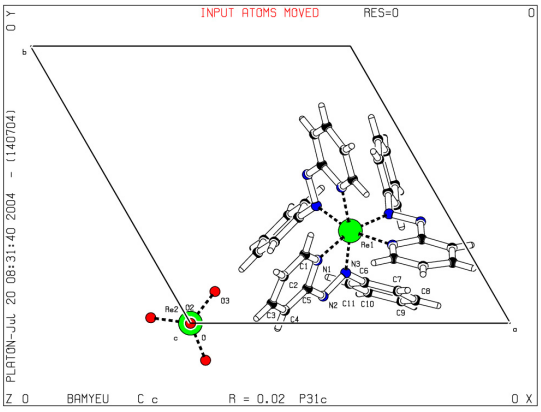


Fig. 4 A perspective view of $[Re(L^3)_2]ReO_4$ **4a**. The atoms are represented by their 30% thermal probability ellipsoids.

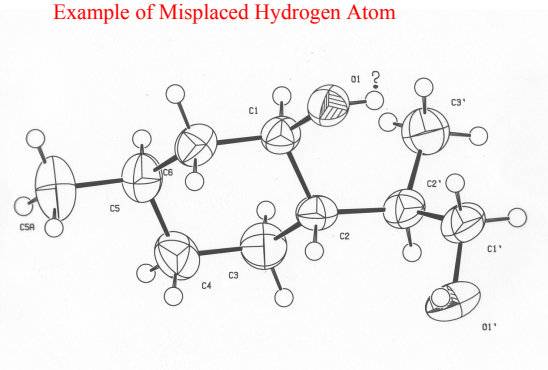




Misoriented O-H

- The O-H moiety is generally, with very few exceptions, part of a D-H..A system.
- An investigation of structures in the CSD brings up many 'exceptions'.
- Closer analysis shows that misplacement of the O-H hydrogen atom is in general the cause.

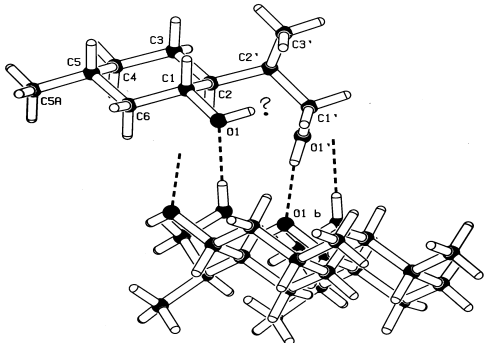
Example of Misplaced Hydrogen Atom

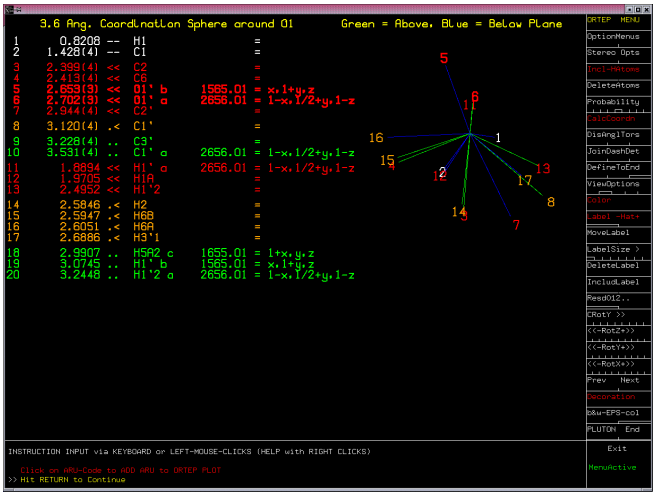


Two ALERTS related to the misplaced Hydrogen Atom

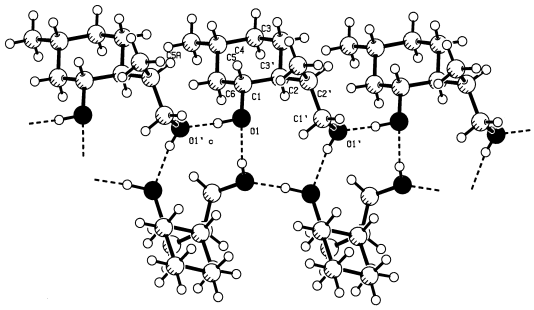
```
#==============================================================================
# PLATON/CHECK run versus check.def version of 21/06/00       for entry:I
# Data From: publ.cif - Data Type: CIF

# CELL   0.71069    8.571    6.466    9.850   90.00  106.78   90.00    522.69
# SpaceGroup P21                  Rep: P 21
# MoietyFormula C10 H20 O2
#       Reported ?
# SumFormula C10 H20 O2    Rep: C10 H20 O2
# Mr          =     172.26[Calc],     172.26[Rep]
# Dx,gcm-3    =       1.094[Calc],      1.094[Rep]
# Z           =           2[Calc],          2[Rep]
# Mu (mm-1)   =       0.074[Calc],      0.074[Rep]
# Calculated T limits: Tmin=0.993 Tmin'=0.978  Tmax=0.996
# Reported   Hmax= 10, Kmax=  7, Lmax= 11, Nref= 1935       , Th(max)= 25.76
# Calculated Hmax= 10, Kmax=  7, Lmax= 12, Nref= 1090( 1991), Ratio= 1.78( 0.97)
# R= 0.0500( 1299), wR2= 0.1870( 1935), S = 0.721, Npar= 112
#==============================================================================

>>> The Following ALERTS were generated <<<

420_ALERT B D-H without acceptor       O(1)   -    H(1)              ?         <<
#==============================================================================
048_ALERT C MoietyFormula not Given ......................          ?
084_ALERT C High R2 value ................................       0.19
086_ALERT C Unsatisfactory S value (Too Low or not given)        0.72
142_ALERT C su on b - axis small or missing (x 100000) ...         30 Ang.
145_ALERT C su on beta      small or missing  (x 10000) ...        30 Deg.
414_ALERT C Short intra D-H..H-X      H(1)    ..  H(12') =       1.93 Ang.   <<
708_ALERT C D-H..A  Calc    170.5(5), Rep    170.00, Dev.       1.05 Sigma
         O1'   -H1'   -O1      1.555   1.555   2.646

  1 ALERT Level B = Potentially Serious Problem
  7 ALERT Level C = Check & Explain
#==============================================================================
```

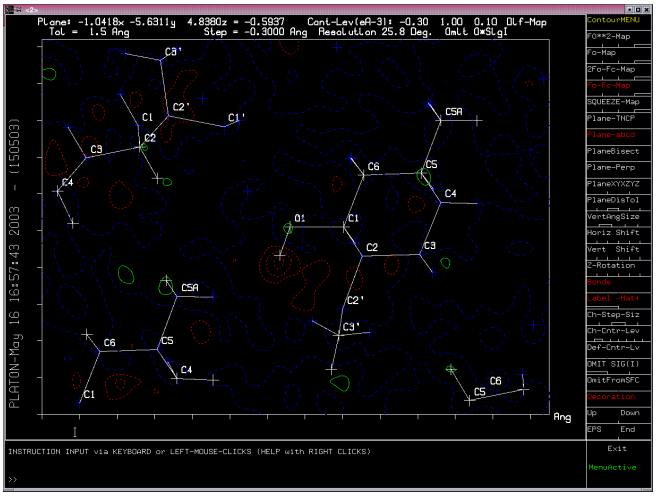
Unsatisfactory Hydrogen Bond Network

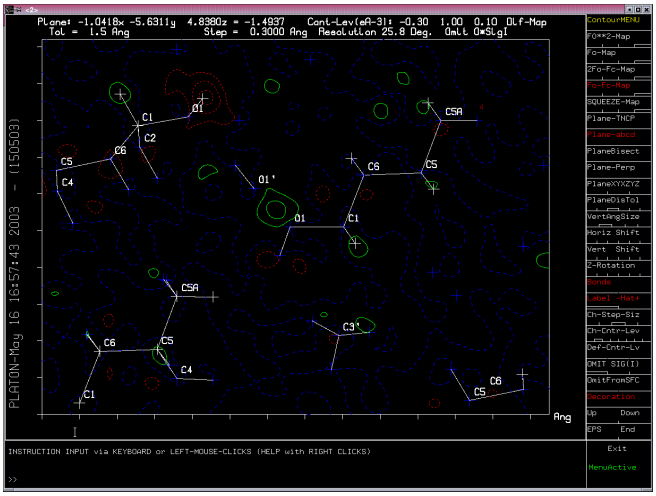




Satisfactory Hydrogen Bond Network with new H-position



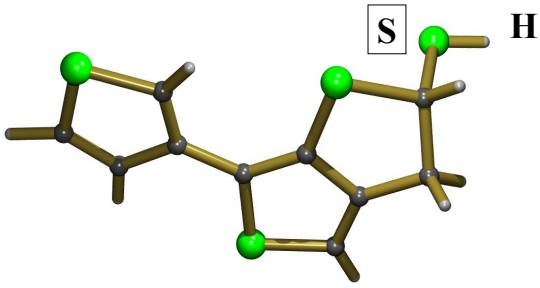




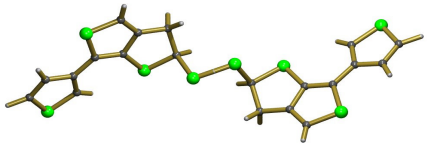
Consult the CSD

- It is a good idea to always consult the CSD for previous reports on structures related to the one at hand.
- The statistics provided by VISTA (CCDC) can be very helpful for this.
- However, such an analysis often shows outliers. Many of those appear to be errors.

Entry from the CAD



But with Space Group Symmetry



=> Different structure with S-S Bond !

Concluding Remarks

- Automatic Validation both ALERTS for potential errors and for interesting features in a structure to be discussed.
- Detailed analysis of intermolecular interactions appears often to be ignored in a service setting.