Validation of F(obs)/F(calc) Data with PLATON (Version 17-07-2010)

The result of a single crystal structure determination is generally available in the form of two data files. In the SHELXL world they have the extensions **.cif** and **.fcf**. Unfortunately, not all journals require the archival of the reflection data. Thus only the **.cif** file with the model parameters may be available for structure validation. In such a case many issues cannot be pursued such as the reason for an unexpectedly high R-value that might in fact be due to an incomplete analysis. Availability of the reflection data (**.fcf**) solves this issue. Recently, it was also shown that fraud is less likely to go undetected with the availability of both data files.

The current implementation of CheckCIF/PLATON takes both files (**.cif** and **.fcf**) into account when available. The validation report comes as two files with extensions **.chk** and **.ckf**. The CheckCIF report (on **.chk**) is extended with ALERTS generated by the analysis of the reflection data. The **.ckf** file provides a supporting listing to be used for detailed inspection of the variety of analyzes that are done

Currently, full FCF validation is available only for SHELXL LIST 4 style Fo/Fc reflection files. Current versions of JANA follow this style as well.

Following are details about the information that is listed in the **.ckf** listing. The sample output listing sections are in *Italic* and are taken from various structures.

Section 1: General Data

```
==============================================================================
Crystal Data From: sk1758.cif
Fo/Fc   Data From: sk1758.fcf FCF-TYPE=SHELXL
Space Group      : P21/c
Wavelength (Ang) :    0.71073
Unit Cell (CIF)  :    6.8750  25.3741   8.3964   90.000   96.303   90.000
SHELX WGHT Pars. :    0.0763   0.0000
Extinction Par.  :    0.0590
==============================================================================
```

The files on which the analysis is done are listed along with a guestimate on the program that produced the **.fcf** file. Data names should be identical in both **.cif** and **.fcf.** The same applies for the cell dimensions. Also the SHELXL weight parameters are listed as found in the **.cif**. The latter are used to recalculate R, wR2 and S values.

Section 2: Deviating I(obs), I(calc) Data

```
==============================================================================
Section 2: Reflections with abs((I(obs) - I(calc)) / SigW(I) .GT.  3.0
==============================================================================
   Nr   H   K   L  Theta     I(obs)    I(calc) Sigma(I)   Ratio SigW(I)  RatioW
------------------------------------------------------------------------------
    1   0   2   0   1.61    4157.68    3215.74    77.11   12.22  280.14    3.36
    2  -2   1   1   6.25     139.54     184.96     3.45  -13.17   13.41   -3.39
    3   8   1   1  25.03      13.10      21.72     2.42   -3.56    2.82   -3.06
    4   2  12   1  11.80      76.89     100.67     3.17   -7.50    7.75   -3.07
```

5	-1	15	1	12.68	16.48	23.47	1.59	-4.40	2.26	-3.09
6	-1	4	2	6.32	85.91	111.48	3.16	-8.09	8.47	-3.02
7	-1	7	2	7.84	38.01	55.64	2.51	-7.02	4.55	-3.87
8	0	7	2	7.46	131.70	176.27	6.67	-6.68	14.01	-3.18
9	-1	10	2	9.74	8.71	0.21	2.43	3.50	2.44	3.48
10	-1	14	2	12.58	39.46	27.37	2.50	4.84	3.46	3.49
11	0	24	2	20.29	14.20	21.99	2.03	-3.84	2.51	-3.10
12	1	11	3	12.14	28.01	14.77	2.08	6.37	2.54	5.21
13	3	15	3	17.37	48.47	37.23	1.78	6.31	3.60	3.12
14	3	18	3	19.24	18.73	27.76	1.72	-5.25	2.55	-3.54
15	-1	7	4	11.45	71.10	53.43	3.44	5.14	5.69	3.11
16	4	10	4	18.40	27.59	17.94	1.73	5.58	2.37	4.08
17	1	4	7	18.25	27.26	37.22	1.68	-5.93	3.08	-3.23
18	1	4	8	20.77	25.09	17.05	1.89	4.25	2.42	3.33
19	0	6	8	20.54	15.18	7.91	1.99	3.65	2.14	3.40
								--------		--------
							Average =	-0.72		0.00

For I(calc) < 2 Sigma(I): <I(obs)> = 1.90 and <I(calc)> = 1.62

==========

Reflections with a ratio larger than 3.0 of abs[I(obs)-I(calc)] / SigW(I) are listed.
Ratio's are calculated both on the reflection Sigma(I) and sigma's that take the weight parameters into account (SigW(I)). Large values in the last column may indicate reflections that might be effected by systematic error. With a good explanation (e.g. reflection behind the beamstop) those reflections can be taken out of the refinement.

The section ends with some statistics. The ideal 'average' value should be zero, indicating balancing positive and negative outliers. The average observed intensity for weak reflections should not be much different from the average calculated intensity.

Section 3: Missing Reflections

==========

Missing Reflections (Asym. Refl. Unit) below sin(th)/lambda = 0.5

==========

Nr	H	K	L	sin(th)/lambda	Theta	I(calc)	I(calc)/I(max)
1	2	2	0	0.152	6.18	20875.79	2.98716
2	2	3	0	0.158	6.44	16594.51	2.37455
3	-1	1	1	0.091	3.73	20042.01	2.86786
4	1	4	1	0.127	5.18	18037.32	2.58101

** Note: I(max) is the maximum I(obs) encountered in the fcf-file **
Starred Reflections have a Theta below Theta(Min) = 1.61
From CIF: Theta(Min) = 1.61

==========

The Friedel averaged data set is checked for completeness. All reflections that are missing below a resolution of sin(theta) / lambda < 0.5 are listed. Missing reflections below the minimum resolution as reported in the CIF are starred. The expected intensity (as calculated from the model data in the CIF) of a missing reflection is compared with the largest intensity found in the data set. A reflection with a large ratio may have been left out due to an overflow.

Section 4: Resolution and Completeness Statistics

```
==========================================================================================
Resolution & Completeness Statistics (Cumulative and Friedel Pairs Averaged)
==========================================================================================
Theta sin(th)/Lambda Complete  Expected Measured  Missing
------------------------------------------------------------------------------
 20.82    0.500    0.997       1528    1524      4
 23.01    0.550    0.998       2040    2035      5
 25.24    0.600    0.998       2631    2627      4
------------------------------------------------------- ACTA Min. Res. ---
 26.20    0.621    0.997       2936    2926      10

Note: The Reported Completeness refers to the Actual H,K,L Index Range
==========================================================================================
```

This section gives the completeness of the Friedel averaged data set as a function of the resolution. The table entries under 'Expected', 'Measured' and 'Missing' are cumulative. The minimum resolution expected for Acta Cryst. papers is indicated.

Section 5: R-value statistics as a function of the resolution

```
==========================================================================================
R-Value Statistics as a Function of Resolution (in Resolution Shell)
==========================================================================================
Theta sin(Th)/Lambda  #    R1    wR2     S  av(I/SigW)   av(I)    av(SigW)
------------------------------------------------------------------------------
 12.38    0.302   328  0.039  0.117  1.264   10.42     751.81     64.29
 15.68    0.380   341  0.031  0.095  0.969    9.41     198.43     16.93
 18.02    0.435   325  0.039  0.102  0.985    8.58     144.51     12.72
 19.90    0.479   344  0.043  0.112  0.992    7.72     132.20     12.13
 21.51    0.516   336  0.046  0.120  0.949    6.63      69.08      7.29
 22.94    0.548   342  0.043  0.119  0.813    5.55      39.89      5.12
 24.22    0.577   323  0.054  0.146  0.785    4.10      21.36      3.67
 25.40    0.603   345  0.065  0.169  0.852    3.75      19.23      3.63
 26.20    0.621   242  0.072  0.198  0.873    3.16      14.19      3.45
------------------------------------------------------------------------------

From FCF: R1 =  0.042(   2206), wR2 =  0.118(   2926), S = 0.988
From CIF: R1 =  0.042(   2206), wR2 =  0.117(   2926), S = 0.996, Npar =  202
==========================================================================================
```

At the end of this section, R-values as reported in the CIF are listed with the values that are calculated from the F(calc) values in the reflection file. Both lines are expected to differ only due to rounding errors. When they differ, the reason might be that the .cif and .fcf were not produced in the same refinement run and/or with different weight parameters.

Section 6: Summary of Reflection Data

```
================================================================================
Summary of Reflection Data in FCF - Note: Friedel Pairs Averaged
================================================================================
Total # of Reflections in FCF.    2926 (Hmax =  8, Kmax = 31, Lmax = 10) Obs
Number above Rep. Theta(Max) .       1
Actual   Theta(Max) (Deg.) ... 26.202 (Hmax =  8, Kmax = 31, Lmax = 10) Exp
Reported Theta(Max) (Deg.) ... 26.200 (Hmax =  8, Kmax = 31, Lmax = 10) Rep
Actual   Theta(Min) (Deg.) ...  1.605
Reported Theta(Min) (Deg.) ...  1.610

Unique (Expected) .............   2935
Unique (in FCF) ...............   2926

Observed [I .gt. 2 Sig(I)] ...    2206
Less-Thans ...................     720
Negative Intensities .........     129
Negative Intensities < - 2 SIG       0


Missing (Total) ..............      10
Missing Below Th(Min) ........       0
Missing Th(Min) to STh/L=0.600       4
Missing STh/L=0.600 to Th(Max)       6
Missing Very Strong Refl. ....       4
Beamstop Effected Reflections        2


Space Group Extinctions ......      85
================================================================================
```

A summary is given the reflection data in the CIF. Expected, actual and reported (in the CIF) values are listed.

Section 7: Intensity Distribution

```
                                     I                  I                  I

================================================================================
Intensity Distribution [Decay of I/Sigma(I) versus sin(theta)/lambda]
================================================================================
sh  st/l   Ang     #  0.25   1.0   2.0  Percent  Distr. for I .gt. 2.0 * sig(I)
================================================================================
 1 0.301 1.661   322  99.4  96.9  96.3 ************************************.
 2 0.379 1.318   343  97.7  92.7  89.5 ***********************************....
 3 0.434 1.152   320  96.3  90.0  85.3 *********************************......
 4 0.478 1.046   340  95.9  88.5  82.1 ********************************.......
 5 0.515 0.971   337  95.5  86.9  77.2 ******************************.........
 6 0.547 0.914   339  90.9  80.8  72.6 *****************************...........
 7 0.576 0.868   323  90.1  75.2  61.6 ************************.................
 8 0.602 0.830   337  88.7  72.4  58.2 ***********************..................
 9 0.626 0.798   265  83.8  65.7  51.3 ********************...................
                                    I                  I                  I
                     Percent Observed:   0                 50                100

Maximum Percentage of Reflections with I .gt. 2*s(I) in any Resolution Shell 96
================================================================================
```

This section visualizes the percentage of observed data (i.e. I > 2 * Sigma(I)) as a function of sin(theta / lambda). Note: percentages are given for three different Sigma levels. Only the 2 * Sigma version is displayed. The slope of this distribution (i.e. the curve consisting of the starred point just before the period) indicates the intensity decay as a function of sin(theta) / Lambda. A fast decay points to disorder. An indication for missed translation symmetry might be a low level of observed data in the first shell.

Section 8: Search for Unaccounted for Twinning

Two searches for possibly missed twinning are done based on differences between F(obs) and F(calc) values. In the first analysis F(calc) is based on the model in the CIF and in the second analysis on the F(calc) values in the CIF. In case of missed twinning, both analyses result in similar proposals for applicable twin laws. When a proper twin law was included in the refinement model, twin laws should be reported only in the first analysis since the twinning contribution is in that case included in the F(calc) values in the CIF.

Section 9: Absolute Structure analysis

This section is present only for non-centrosymmetric structures.

The Bijvoet pair analysis is done twice: First on the basis of structure factors calculated from the model parameter is the .**cif**, followed by the same analysis of the calculated structure factors in the **.fcf.** The results of both analyses are expected to be the same when the least squares refinement was not done with a BASF/TWIN instruction set.

```
================================================================================
Bijvoet Pair Analysis - F(calc) from CIF
================================================================================

Excluded Outliers with Observed Bijvoet Difference .GT.    456.71
================================================================================
Nr     H    K    L      FOKD      FCKD      SIGD
================================================================================
  1    0    2   -6    629.45    204.28    451.81
  2    2    0   -4   -703.96   -124.76    968.31
  3    0    2   -4   -713.02    -86.88    529.87
  4    2    2   -4  -1137.35     84.25    977.21
  5    0    1   -3   -531.19   -158.70    392.40
  6    2    1   -3  -1457.54   -221.38    576.87

Flack Parameter Value ....    0.12(10)
Number of Bijvoet Pairs ..    1792[  1858]
Friedel Pair Coverage ....      96%
Res.Scat.................  0.0075
Friedif ..................     120
```

For a definition of Res.Scat. see E.Girard et al. (2003) Acta Cryst. D59,1914-1922 and for Friedif see H.Flack & U. Shmueli (2007). Acta Cryst. A63, 257-265.

H1	K1	L1	I1(OBS)	H2	K2	L2	I2(OBS)	DEL OBS	DEL CALC	SIG DEL	DLC SIG	DLO DLC	|I+-I-| I+ + I-
1	3	-8	2961.6	1	3	8	2822.1	139.5	97.9	80.70	1.21	1.42	0.01583
1	1	-7	334.0	1	1	7	374.9	-40.9	-28.3	23.39	1.21	1.45	0.04197
2	2	-9	1190.4	2	2	9	1104.2	86.2	61.1	51.91	1.18	1.41	0.02451
3	1	-6	942.7	3	1	6	994.0	-51.3	-55.5	55.50	1.00	0.92	0.02717
1	4	-5	806.8	1	4	5	808.5	-1.6	-31.9	32.94	0.97	0.05	0.01941
1	2	-5	2949.8	1	2	5	2921.8	28.0	85.7	89.95	0.95	0.33	0.01506
3	4	-5	295.2	3	4	5	364.9	-69.7	-32.3	35.70	0.91	2.16	0.03935
2	1	-5	2249.0	2	1	5	2238.9	10.1	66.8	74.58	0.90	0.15	0.01573
1	3	-6	1926.2	1	3	6	1855.3	70.9	52.2	59.30	0.88	1.36	0.01225
1	8	-11	781.9	1	8	11	807.9	-26.0	-36.1	42.47	0.85	0.72	0.02125
1	5	-4	1310.4	1	5	4	1396.8	-86.4	45.8	54.38	0.84	-1.89	0.01690
1	3	-10	589.0	1	3	10	591.1	-2.2	-25.6	30.83	0.83	0.08	0.02145
3	6	-1	2350.4	3	6	1	2332.3	18.1	-75.8	91.18	0.83	-0.24	0.01528
6	2	-1	1732.5	6	2	1	1705.9	26.6	65.6	79.95	0.82	0.41	0.01949
1	4	-3	378.4	1	4	3	382.0	-3.7	-22.5	29.31	0.77	0.16	0.02895
0	1	-11	545.1	0	1	11	522.3	22.8	-31.9	41.16	0.77	-0.71	0.03137
4	2	-2	991.6	4	2	2	1053.4	-61.8	-40.2	52.21	0.77	1.54	0.01731
1	1	-11	353.5	1	1	11	308.5	45.0	20.6	27.10	0.76	2.18	0.03001
4	3	-5	1671.3	4	3	5	1554.6	116.7	60.5	79.74	0.76	1.93	0.01947

etc.

Hooft y Parameter Value . 0.15(6)

Section 10: Analysis of the Difference Fourier Map

```
================================================================================
Analysis of Difference Map Grid Point Density. - (MIN = -0.18, MAX =   0.19)
================================================================================
  eA-3 Frequency Plot Sqrt(Frequency) - Average =  0.000, sigma =  0.039 eA-3
================================================================================
```

```
 -0.25         0
 -0.20        12 *
 -0.15       256 ****
 -0.10      3676 **************
 -0.05     28808 ************************************
  0.00     64854 ************************************************************
  0.05     29978 ************************************
  0.10      3280 *************
  0.15       188 ***
  0.20        20 *
  0.25         0
```

```
================================================================================
Unique Density Maxima in          Difference Map (CutOff level =     0.10 eA-3)
================================================================================
   #   x     y     z  (e/A^3)     Shortest Contacts within 3.2 Ang. (Excl. H)
================================================================================
   1 0.199 0.961 0.754  0.19      C10  0.11; C5   1.29; C9   1.45; C4   1.53;
   2 0.363 0.448 0.369  0.15      C2   0.36; C3   1.28; O1   1.34; C21  1.71;
   3 0.721 0.992 0.349  0.15      C9   0.18; O1   1.24; C8   1.39; C10  1.55;
   4 0.280 0.042 0.394  0.13      C8   1.48; C9   2.42; C7   2.53; O1   2.64;
   5 0.873 0.938 0.101  0.13      C2   1.42; C3   1.45; O1   2.23; C21  2.30;
   6 0.104 0.908 0.939  0.13      O4   1.23; C4   1.88; C5   2.12; C10  2.25;
   7 0.741 0.277 0.511  0.12      O29  1.32; C25  2.29; C29  2.46; C26  2.56;
   8 0.499 0.156 0.975  0.12      C21  1.33; O29  1.34; C29  1.34; O28  1.35;
```

```
================================================================================
Density Maxima within  1.2 Angstrom from Atoms (CutOff level =     0.10 eA-3)
================================================================================
Atom      # e/A^3   Ang    # e/A^3   Ang    # e/A^3   Ang    # e/A^3   Ang \
================================================================================
C2         2  0.15  0.36:  23  0.10  0.90:
C3        16  0.11  0.38:  23  0.10  0.95:
C5        18  0.11  0.19:
C7        14  0.11  1.15:
C8        14  0.11  0.38:
C9         3  0.15  0.18:
C10        1  0.19  0.11:
C21       19  0.11  0.11:
C27       11  0.12  0.21:
```

```
================================================================================
Unique Density Minima in          Difference Map (CutOff level =    -0.10 eA-3)
================================================================================
   #   x     y     z  (e/A^3)     Shortest Contacts within 3.2 Ang. (Excl. H)
================================================================================
   1 0.408 0.141 0.153 -0.18      O28  0.80; C29  0.97; C30  1.56; O29  2.13;
   2 0.999 0.740 0.936 -0.18      C25  0.82; C24  1.25; C26  2.19; C23  2.56;
   3 0.723 0.727 0.912 -0.17      C24  0.78; C23  1.28; C25  2.15; C22  2.60;
   4 0.877 0.415 0.066 -0.14      C6   0.83; C7   0.90; C5   2.16; C8   2.23;
   5 0.316 0.037 0.626 -0.13      O1   0.79; C9   0.98; C8   1.59; C2   2.13;
```

```
   6 0.376 0.325 0.247 -0.13      C23  0.77; C22  1.18; C21  1.89; C24  2.14;

Density Minima within  1.2 Angstrom from Atoms (CutOff level =    -0.10 eA-3)
================================================================================
Atom      # e/A^3   Ang    # e/A^3   Ang    # e/A^3   Ang    # e/A^3   Ang
================================================================================
O1         5 -0.13  0.79:  19 -0.11  0.82:
O4        13 -0.13  1.06:  25 -0.10  1.11:
O28       10 -0.13  0.78:   1 -0.18  0.80:
O29        7 -0.13  0.80:  14 -0.12  1.01:
C2        19 -0.11  1.05:
C3        11 -0.13  0.72:
C4        13 -0.13  0.72:  25 -0.10  0.95:  20 -0.11  1.01:
```

A histogram is displayed of the gridpoint values. The distribution is expected to be symmetrical and centered around zero eA-3.

Unique maximum difference density peaks are listed along with their distance to the nearest atoms in the model.

Alternatively, for each atom in the model, the nearest density peaks are listed.

Similar tabulations are given unique negative density peaks.