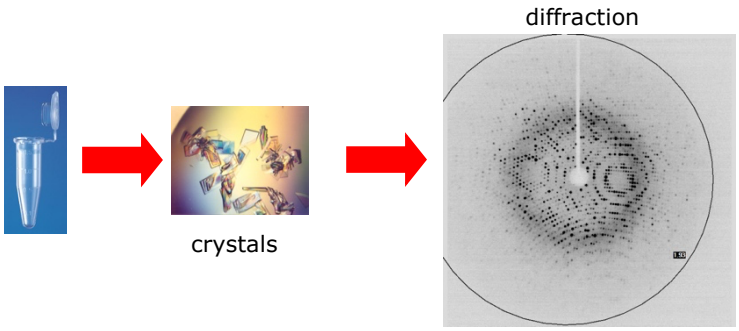
SINGLE CRYSTAL X-RAY DIFFRACTION

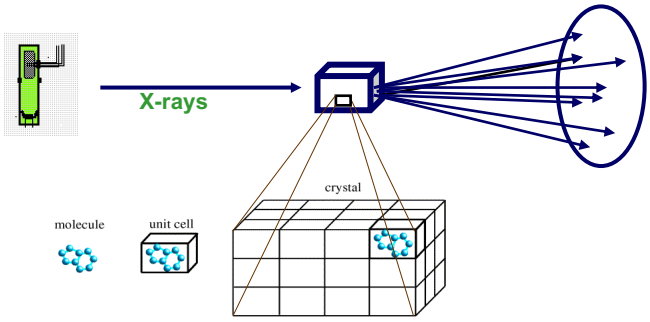
&

DATA PROCESSING

Once a **single crystal** of adequate size is obtained, it is exposed to a **monochromatic beam of X-rays** and the intensity of the diffracted photons is recorded.

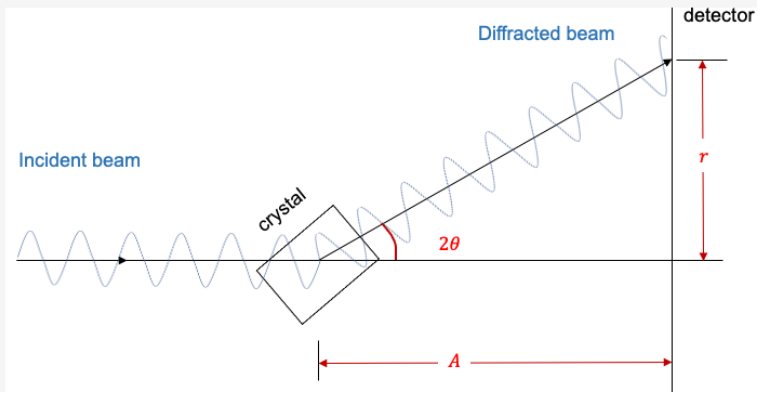
A crystal is built by many **unit cells**, that are described by three cell axes *a,b,c* and by three angles α, β and γ. The specific relationship between these six parameters defines the **crystal's lattice and space group**. Inside the unit cell, the **asymmetric units** (the macromolecules arranged with or without quaternary structure) are distributed according to the symmetry operations defined by the space group. The lattice and the space group will define the exact position of each spot (or ***hkl* reflection**) in the diffraction pattern produced by a crystal. *h, k* and *l* are also known as Miller indices.





h	*k*	*l*	*I*	σ(*I*)
6	0	0	484.38	78.37
6	0	0	502.41	78.91
-6	0	0	443.23	78.59
-6	0	0	488.55	78.36
8	0	0	25.037	5.827
8	0	0	22.527	5.404
-8	0	0	25.589	7.102
-8	0	0	27.919	5.399
-10	0	0	1178.49	186.21
10	0	0	1168.69	186.25
10	0	0	1181.58	186.25
-10	0	0	1030.63	186.66
-12	0	0	459.73	90.88
-12	0	0	557.97	89.00
12	0	0	561.78	89.10
12	0	0	569.92	89.09
-14	0	0	591.40	113.20
14	0	0	705.39	110.52
⋮				
0	-2	-20	-538.91	396.86
0	2	-20	-421.48	216.28
-2	-2	-20	-221.69	382.55
11	3	-20	-253.64	192.06
9	-3	-20	-635.92	565.37
-9	3	20	-50.69	87.92
7	-3	-20	-671.88	460.42
5	-3	-20	-743.76	507.86
3	-3	-20	-121.55	386.71
3	3	-20	-255.39	161.72
-3	3	20	-21.73	106.06
1	-3	-20	-300.02	358.61
1	3	-20	-312.28	152.83
4	4	-20	21.11	128.28
0	0	0	0.00	0.00

A very important mathematical concept in crystallography, **Bragg's Law**, defines that the production of a spot in the diffraction pattern only occurs when **constructive interference** of the scattered radiation is attained. In practice, the diffraction experiment produces a list of intensities (I_{hkl}) and associated errors for all reflections recorded. The intensity of each reflection results from the **sum of photons that constitute the scattered wave**, with an **amplitude**, **phase** and **frequency** (the same as the incident radiation).



A – distance between detector and crystal.
r – distance between a diffracted spot and the primary beam position.

$\tan 2\theta = r/A$

Maximum resolution: Bragg's Law $n\lambda = 2 d \sin\theta \Rightarrow d_{min} = \lambda / (2 \sin\theta_{max})$

($\lambda_{Cu K\alpha}$ = 1.5418 Å d_{min} = 0.77 Å)

The first part of this lesson is dedicated to harvest, cryo-protect and test the diffraction power of a single crystal of **a protein**. The experiment will be carried out in the in-house X-ray diffractometer.

Once a full set of diffraction images is collected, several software packages for processing image data, such as HKL-2000 (Otwinowski & Minor, 1997), XDS (Kabsch, 2010), MOSFLM (Leslie, 2006), d*TREK (Pflugrath, 1999) and DIALS (Beilsten-Edmands *et al.*, 2020), have been developed that differ in the details of their implemented algorithms.

In this tutorial, you will use **MOSFLM** (Battye *et al*, 2011) software to index, integrate and analyze X-ray diffraction data (in the form of diffraction images) obtained on a source of synchrotron radiation.

A. X-ray diffraction data collection

1- Under the microscope, observe the crystallization setups that were previously prepared and register which crystallization conditions have originated protein crystals.

2- Measure the crystals' maximum dimensions.

3- Test the cryo-protectant solutions in the diffractometer, checking for ice formation.

4- Choose a monocrystal adequate for the diffraction experiment and incubate it in the cryo-protectant solution.

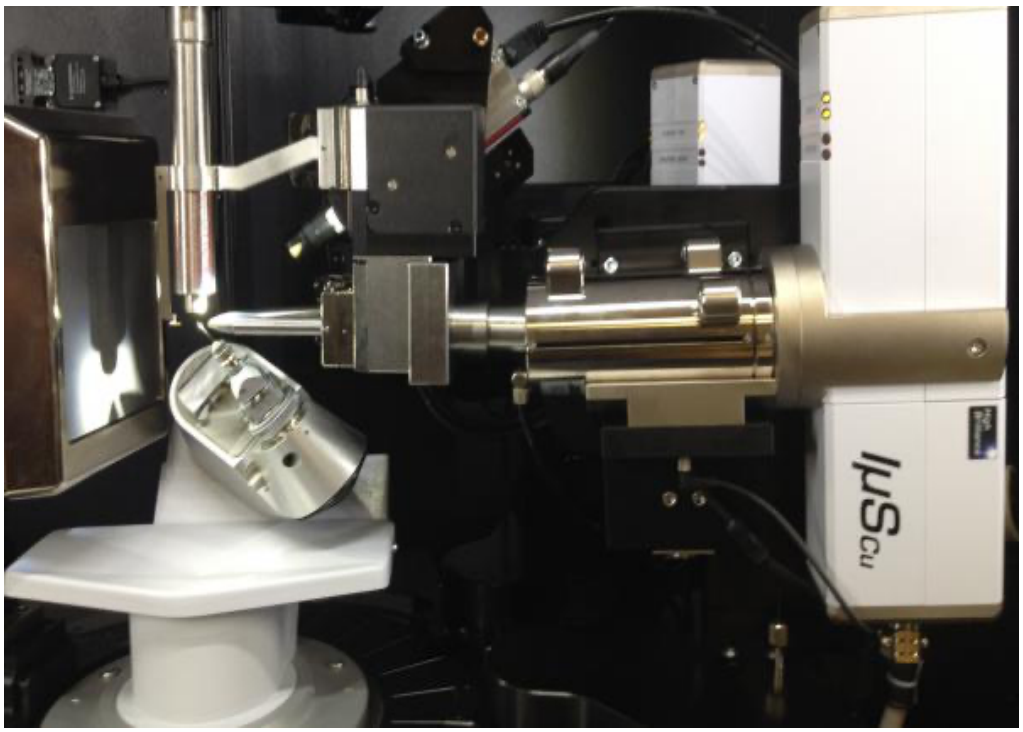
5- Mount the crystal in a loop and place it on the goniometer head, under the cooled nitrogen stream.

6- Centre the crystal.

7- Test the crystal's diffraction quality, by collecting two frames 90º apart. Evaluate the highest resolution that the crystal will produce.

8- Index this data to determine the space group and the cell constants, if possible.

9- Collect data to completeness, as efficiently and as fast as possible (try to measure 100% of all possible reflections).

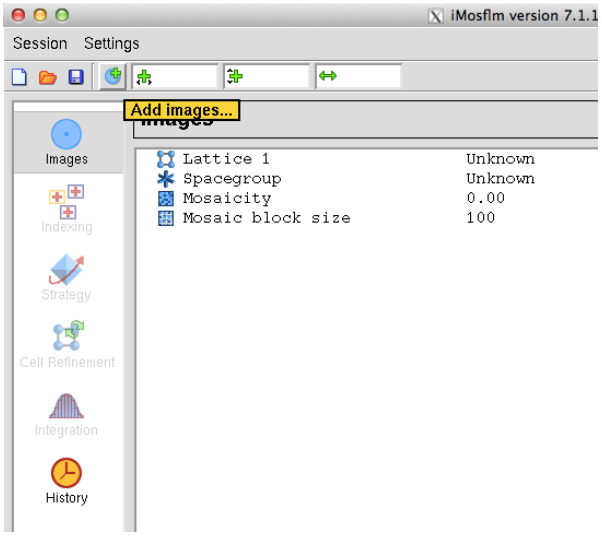


D8 Venture X-ray diffractometer (with KAPPA four-circle goniometer) at FCT-NOVA, Caparica, Portugal.

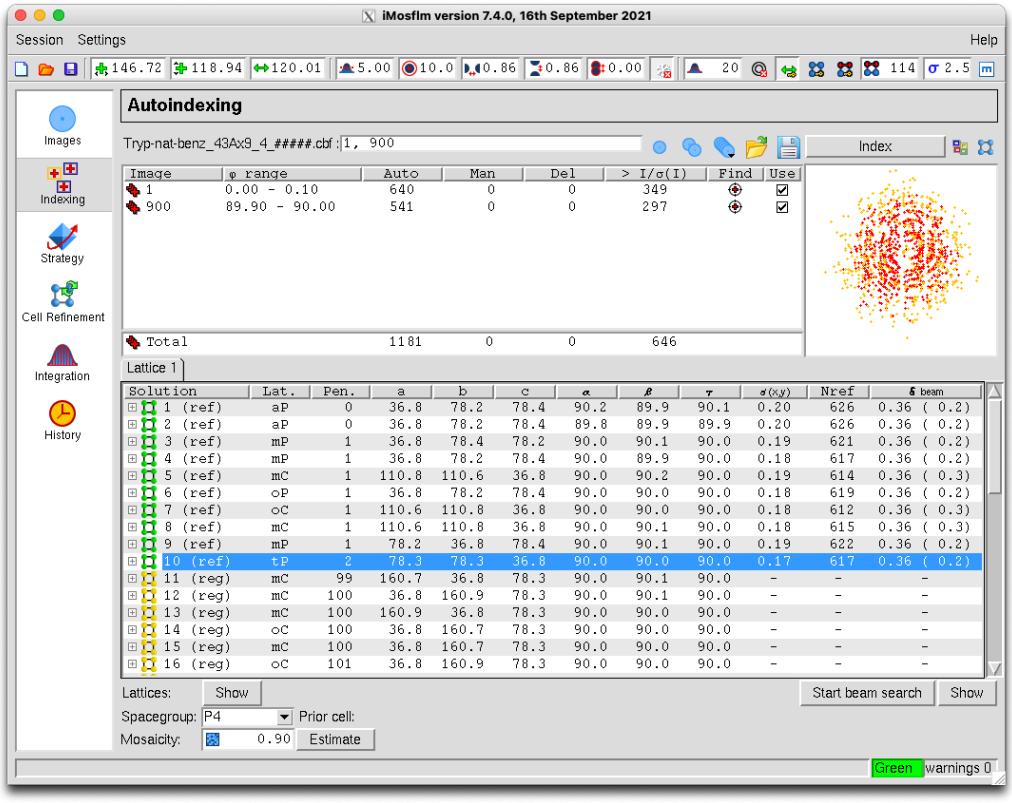
B. Data processing and analysis

1- In your working computer, find the 900 image files provided (0.1º per image; 90º of diffraction data). If necessary, uncompress the folder.

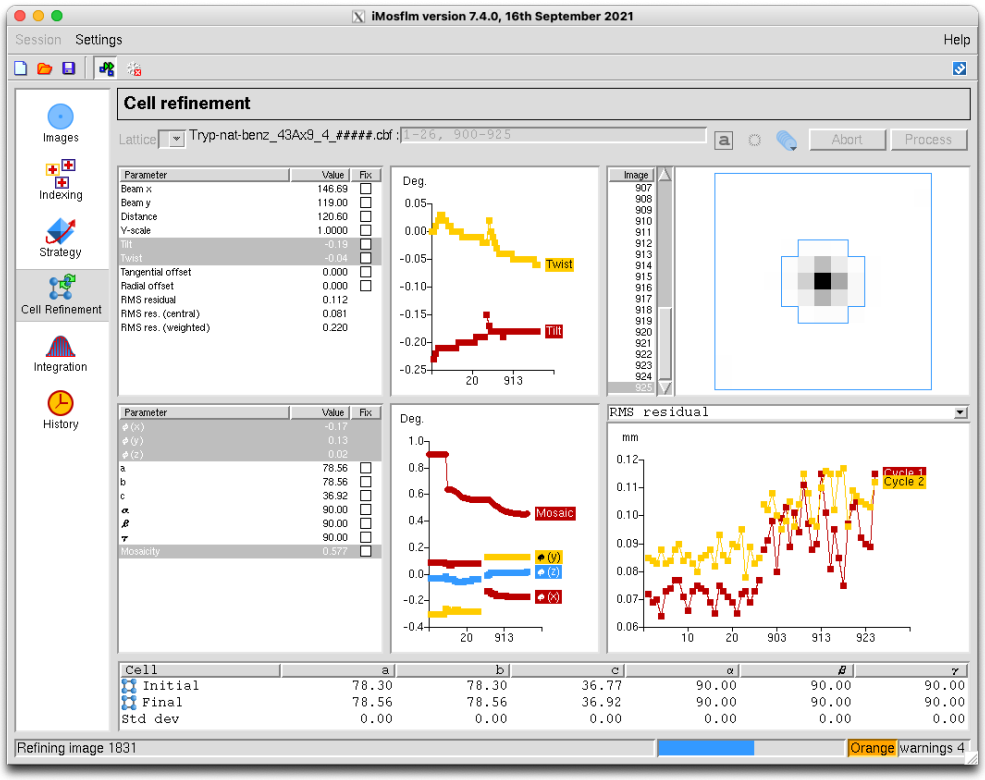
2- Follow the tutors' instructions to start program MOSFLM and read in the images from the data set provided.



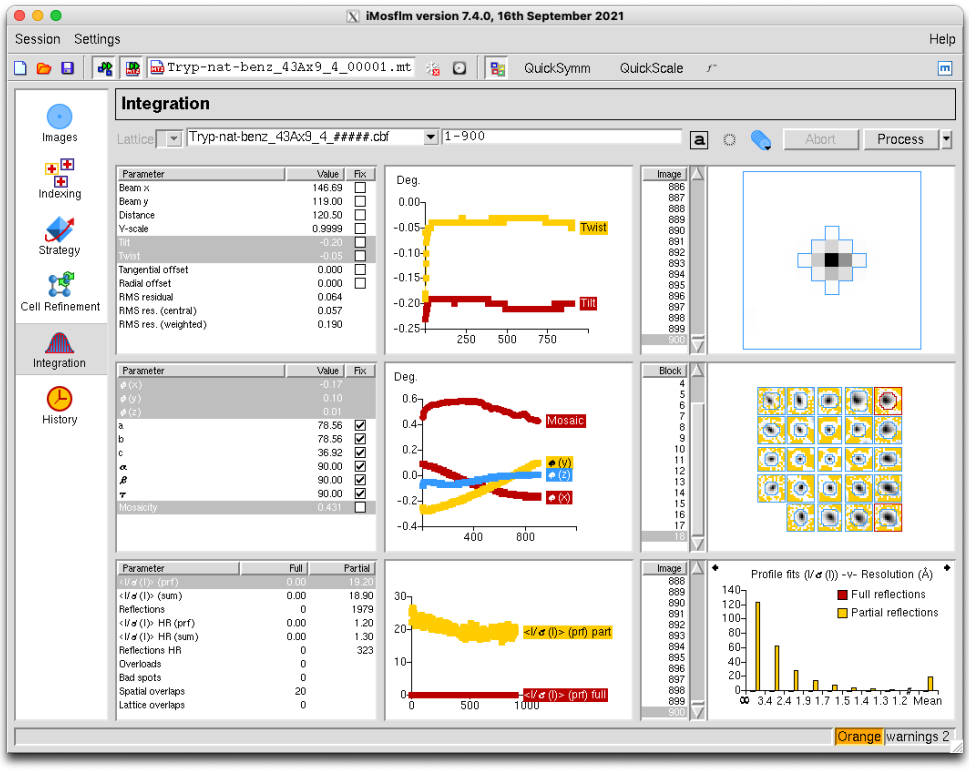
3- Index data (pick 2 images, 90º apart, by default) to estimate cell constants, space group and crystal mosaicity.



4- Refine the cell constants (use at least 2 wedges of data, approximately 90º apart).



5- Check that the cell constants are fixed, choose a name for the file of the integrated images (press *enter*) and process at least 20º of data (200 images). If the program runs fast, process the full 90º of data (900 images).



6- Run QuickSymm option to detect higher symmetry (e.g. screw axis), if present, and check for possible twinning (**Note 1**).

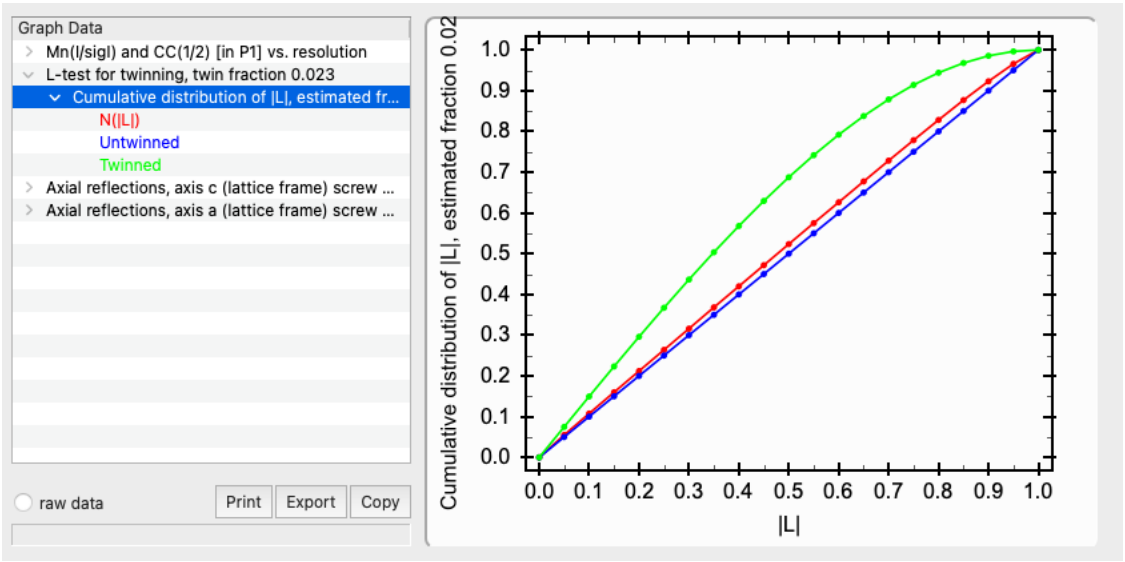
Result

```
Best Solution:    space group P 41 21 2

   Reindex operator:                  [h,k,l]
   Laue group probability:            0.999
   Systematic absence probability:    0.930
   Total probability:                 0.929
   Space group confidence:            0.913
   Laue group confidence              0.999

WARNING: You will have to resolve the enantiomorphic ambiguity later

   Unit cell:    78.56   78.56   36.92      90.00   90.00   90.00

   36.92 to   1.19   - Resolution range used for Laue group search

   36.92 to   1.19   - Resolution range in file, used for systematic absence check

   Number of batches in file:    900

The data do not appear to be twinned, from the L-test
```

Please cite:

P.R.Evans, 'Scaling and assessment of data quality' Acta Cryst. D62, 72-82 (2006).

PDF
P.R.Evans, 'An introduction to data reduction: space-group determination, scaling and intensity statistics' Acta Cryst. D67, 282-292 (2011)



7- Run Quick Scale to scale and analyse diffraction data quality (see **Note 2**). Evaluate true limits of resolution and completeness of data (this will also run program Truncate to convert I_{obs} to $|F_{obs}|$, *ie* measured intensities to observed structure factors; see **Note 3**). Look at the summary files below and compare the statistics for 20º, 90º and the full set of 3600 images (360º).

Summary for 20º of data:

Result

```
Summary data for        Project: New Crystal: New Dataset: New

                                          Overall  InnerShell  OuterShell
Low resolution limit                      36.92     36.92      1.21
High resolution limit                      1.19      6.53      1.19

Rmerge  (within I+/I-)                    0.023     0.012     0.172
Rmerge  (all I+ and I-)                   0.026     0.023     0.179
Rmeas (within I+/I-)                      0.030     0.015     0.233
Rmeas (all I+ & I-)                       0.032     0.030     0.220
Rpim (within I+/I-)                       0.019     0.009     0.155
Rpim (all I+ & I-)                        0.018     0.019     0.125
Rmerge in top intensity bin               0.015       -         -
Total number of observations              48864      320      2304
Total number unique                       21185      129       990
Mean((I)/sd(I))                            15.6      51.2       2.7
Mn(I) half-set correlation CC(1/2)        0.999     0.984     0.953
Completeness                               57.7      50.9      55.4
Multiplicity                                2.3       2.5       2.3
Mean(Chi^2)                                0.68      0.81      0.31

Anomalous completeness                     33.3      36.1      33.2
Anomalous multiplicity                      0.9       1.6       1.6
DelAnom correlation between half-sets     -0.029     0.495    -0.104
Mid-Slope of Anom Normal Probability      0.657       -         -

Anomalous flag switched OFF in input, anomalous signal is weak

Estimates of resolution limits: overall
   from half-dataset correlation CC(1/2) >  0.30: limit =  1.19A  == maximum resolution
   from Mn(I/sd) >  1.50:                        limit =  1.19A  == maximum resolution
   from Mn(I/sd) >  2.00:                        limit =  1.19A  == maximum resolution

Estimates of resolution limits in reciprocal lattice directions:
  Along h k plane
   from half-dataset correlation CC(1/2) >  0.30: limit =  1.19A  == maximum resolution
   from Mn(I/sd) >  1.50:                        limit =  1.21A
  Along l axis
   from half-dataset correlation CC(1/2) >  0.30: limit =  1.19A  == maximum resolution
   from Mn(I/sd) >  1.50:                        limit =  1.19A  == maximum resolution

Anisotropic deltaB (i.e. range of principal components), A^2:  0.20

Average unit cell:    78.56   78.56   36.92   90.00   90.00   90.00
Space group: P 4 2 2
Average mosaicity:   0.55

Minimum and maximum SD correction factors: Fulls   0.00   0.00 Partials   0.53  34.79
```

Summary for 90º of data:

Result

```
Summary data for        Project: New Crystal: New Dataset: New

                                          Overall  InnerShell  OuterShell
Low resolution limit                      36.92     36.92      1.21
High resolution limit                      1.19      6.53      1.19

Rmerge  (within I+/I-)                    0.034     0.021     0.193
Rmerge  (all I+ and I-)                   0.036     0.024     0.208
Rmeas (within I+/I-)                      0.040     0.025     0.231
Rmeas (all I+ & I-)                       0.040     0.027     0.228
Rpim (within I+/I-)                       0.021     0.013     0.125
Rpim (all I+ & I-)                        0.016     0.012     0.091
Rmerge in top intensity bin               0.023       -         -
Total number of observations             225958      1465     10603
Total number unique                       37300       275      1803
Mean((I)/sd(I))                            19.6      48.8       4.2
Mn(I) half-set correlation CC(1/2)        0.999     0.996     0.974
Completeness                               99.9      96.9     100.0
Multiplicity                                6.1       5.3       5.9
Mean(Chi^2)                                0.68      0.70      0.33

Anomalous completeness                     99.1      96.6      99.3
Anomalous multiplicity                      3.1       3.8       3.0
DelAnom correlation between half-sets     0.008     0.403    -0.056
Mid-Slope of Anom Normal Probability      0.663       -         -

Anomalous flag switched OFF in input, anomalous signal is weak

Estimates of resolution limits: overall
   from half-dataset correlation CC(1/2) >  0.30: limit =  1.19A  == maximum resolution
   from Mn(I/sd) >  1.50:                        limit =  1.19A  == maximum resolution
   from Mn(I/sd) >  2.00:                        limit =  1.19A  == maximum resolution

Estimates of resolution limits in reciprocal lattice directions:
  Along h k plane
   from half-dataset correlation CC(1/2) >  0.30: limit =  1.19A  == maximum resolution
   from Mn(I/sd) >  1.50:                        limit =  1.19A  == maximum resolution
  Along l axis
   from half-dataset correlation CC(1/2) >  0.30: limit =  1.19A  == maximum resolution
   from Mn(I/sd) >  1.50:                        limit =  1.19A  == maximum resolution

Anisotropic deltaB (i.e. range of principal components), A^2:  0.81

Average unit cell:    78.56   78.56   36.92   90.00   90.00   90.00
Space group: P 41 21 2
Average mosaicity:   0.53

Minimum and maximum SD correction factors: Fulls   0.00   0.00 Partials   0.43 107.66
```

Please cite:

P.R.Evans and G.N.Murshudov, 'How good are my data and what is the resolution?' Acta Cryst. D69, 1204–1214 (2013).

Summary for 360º of data:

Result

```
Summary data for        Project: New Crystal: New Dataset: New

                                       Overall  InnerShell  OuterShell
Low resolution limit                   36.92     36.92      1.21
High resolution limit                   1.19      6.53      1.19

Rmerge  (within I+/I-)                 0.053     0.038     0.259
Rmerge  (all I+ and I-)                0.054     0.040     0.263
Rmeas (within I+/I-)                   0.055     0.040     0.270
Rmeas (all I+ & I-)                    0.055     0.041     0.269
Rpim (within I+/I-)                    0.015     0.013     0.076
Rpim (all I+ & I-)                     0.011     0.010     0.055
Rmerge in top intensity bin            0.034       -         -
Total number of observations          907807      5892     42650
Total number unique                    37370       290      1803
Mean((I)/sd(I))                         27.9      65.1       5.9
Mn(I) half-set correlation CC(1/2)     1.000     0.997     0.991
Completeness                           100.0      99.7     100.0
Multiplicity                            24.3      20.3      23.7
Mean(Chi^2)                             0.65      1.03      0.20

Anomalous completeness                 100.0      99.3     100.0
Anomalous multiplicity                  12.8      14.2      12.1
DelAnom correlation between half-sets  -0.050     0.302    -0.058
Mid-Slope of Anom Normal Probability    0.584       -         -

Anomalous flag switched OFF in input, anomalous signal is weak

Estimates of resolution limits: overall
   from half-dataset correlation CC(1/2) >  0.30: limit =  1.19A  == maximum resolution
   from Mn(I/sd) >  1.50:                       limit =  1.19A  == maximum resolution
   from Mn(I/sd) >  2.00:                       limit =  1.19A  == maximum resolution

Estimates of resolution limits in reciprocal lattice directions:
  Along h k plane
   from half-dataset correlation CC(1/2) >  0.30: limit =  1.19A  == maximum resolution
   from Mn(I/sd) >  1.50:                       limit =  1.19A  == maximum resolution
  Along l axis
   from half-dataset correlation CC(1/2) >  0.30: limit =  1.19A  == maximum resolution
   from Mn(I/sd) >  1.50:                       limit =  1.19A  == maximum resolution

Anisotropic deltaB (i.e. range of principal components), A^2:  1.24

Average unit cell:    78.56   78.56   36.92   90.00   90.00   90.00
Space group: P 41 21 2
Average mosaicity:   0.50

Minimum and maximum SD correction factors: Fulls   0.00   0.00 Partials   0.61 163.63
```

Please cite:

P.R.Evans and G.N.Murshudov, 'How good are my data and what is the resolution?' Acta Cryst. D69, 1204-1214 (2013).

8- The contents of output MTZ files can be accessed with program VIEWHKL (**Note 4**)

Note 1: *Pointless checks sets of reflections which may be systematically absent to suggest a possible spacegroup. There is also a check for lattice centering, ie a check for whole classes of reflections having essentially zero intensity, including a check for obverse/inverse twinning in rhombohedral systems. (Ref: http://www.ccp4.ac.uk/html/pointless.html)*

Note 2: *Aimless scales together multiple observations of reflections, and merges multiple observations into an average intensity: it is a successor program to SCALA (Ref: http://www.ccp4.ac.uk/html/aimless.html)*

Note 3: *The standard use of the Truncate program is to read a file of averaged intensities (output from SCALA, SCALEPACK2MTZ or DTREK2MTZ) and write a file containing mean amplitudes and the original intensities. If anomalous data is present then F(+), F(-), with the anomalous difference, plus I(+) and I(-) are also written out. The amplitudes are put on an approximate absolute scale using the scale factor taken from a Wilson plot. (Ref: http://www.ccp4.ac.uk/html/truncate.html)*

Note 4: *Interactive graphical viewer and browser for reflection Data (Ref: http://legacy.ccp4.ac.uk/newsletters/newsletter48/articles/ViewHKL/viewhkl.html)*

REFERENCES:

Otwinowski, Z. & Minor, W. (1997). Methods Enzymol. 276, 307-326.

Kabsch, W. (2010a). Acta Cryst. D66, 125-132.

Leslie, A. G. W. (2006). Acta Cryst. D62, 48-57.

Pflugrath, J. W. (1999). Acta Cryst. D55, 1718-1725.

Beilsten-Edmands, J., Winter, G., Gildea, R., Parkhurst, J., Waterman, D. & Evans, G. (2020). Acta Cryst. D76, 385-399.

Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). Acta Cryst. D67, 271-281.

For further reading: Carvalho, Ana Luísa, Teresa Santos-Silva, Maria João Romão, Eurico J. Cabrita, and Filipa Marcelo. 2018. “Structural Elucidation of Macromolecules.” Essential Techniques for Medical and Life Scientists, September, 30–91. https://doi.org/10.2174/9781681087092118010005.