# Data evaluation, integration and analysis II

**L.M.J. Kroon-Batenburg**
Utrecht University
The Netherlands

# Statistics for data quality

$$R_{merge} = \frac{\sum_{hkl}\sum_{i}^{n}|I_i(hkl) - \langle I(hkl)\rangle|}{\sum_{hkl}\sum_{i}^{n} I_i(hkl)}$$

$$R_{meas} = \frac{\sum_{hkl}\sqrt{\frac{n}{n-1}}\sum_{i}^{n}|I_i(hkl) - \langle I(hkl)\rangle|}{\sum_{hkl}\sum_{i}^{n} I_i(hkl)}$$

Rrim

$$R_{pim} = \frac{\sum_{hkl}\sqrt{\frac{1}{n-1}}\sum_{i}^{n}|I_i(hkl) - \langle I(hkl)\rangle|}{\sum_{hkl}\sum_{i}^{n} I_i(hkl)}$$

$$R_{anom} = \frac{\sum_{hkl}\left|I(hkl) - I(\overline{hkl})\right|}{\sum_{hkl}\langle I(hkl)\rangle}$$

Merging and standard uncertainties

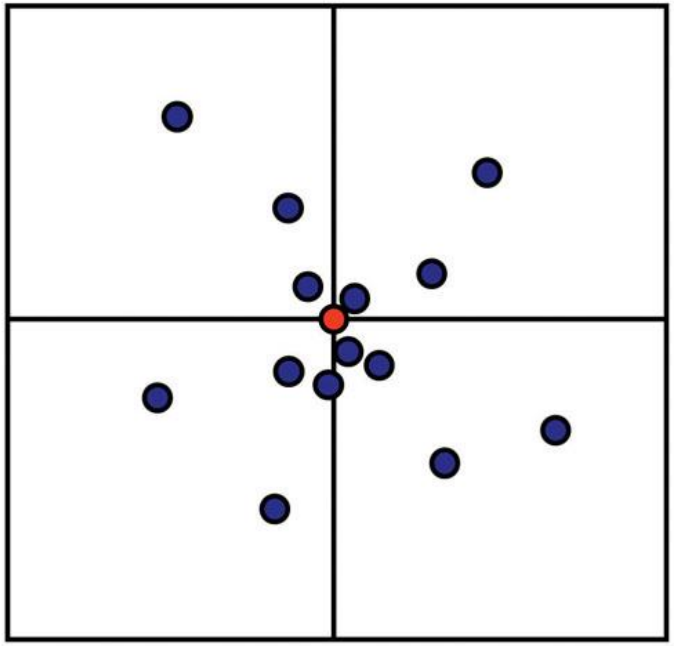
$$\sigma^2 = \frac{\sum_i w_i \sigma_i^2}{n\sum_i w_i} \qquad \sigma \approx \frac{\sigma_i}{\sqrt{n}}$$

Correlation between two half datasets

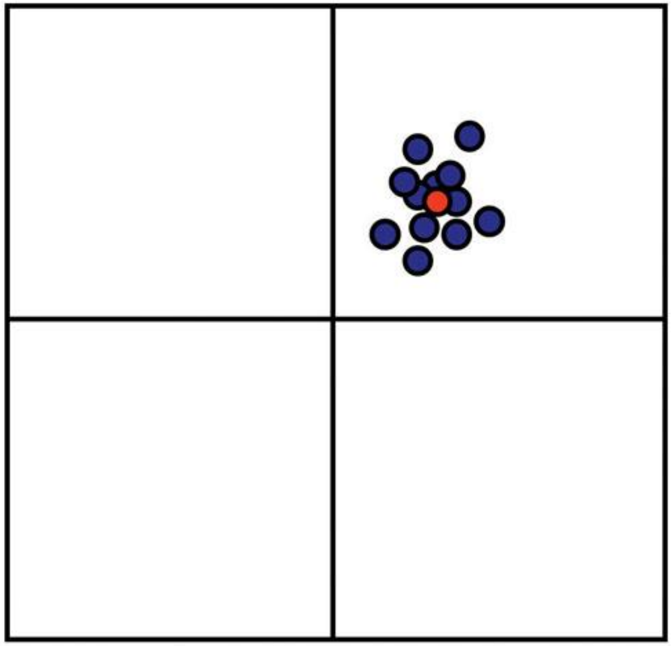
$$CC1/2 = \frac{\sum_i^n((x_i - \langle x\rangle)(y_i - \langle y\rangle))}{\sqrt{\sum_i^n(x_i - \langle x\rangle)^2 \sum_i^n(y_i - \langle y\rangle)^2}}$$

# Accuracy vs precision

Spread of observation around the true value



Internal consistency



# Anomalous signal indicators

Anomalous signal-to-noise ratio: $\dfrac{\Delta F}{\sigma_{\Delta F}}$

Anomalous signal: $\dfrac{\langle \Delta F \rangle}{\langle F \rangle}$ Expected: $\dfrac{\langle \Delta F \rangle}{\langle F \rangle} = \dfrac{\sqrt{2N_A}\,\delta f''}{\sqrt{N_p}\,Z_{eff}}$

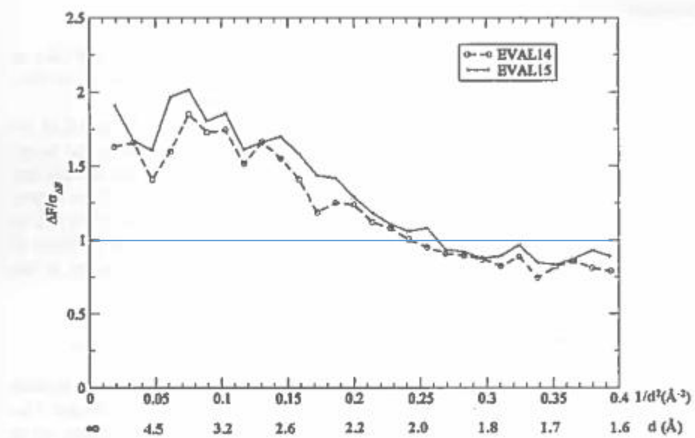



**Fig. 3.14** The anomalous signal-to-noise ratio $\Delta F/\sigma_{\Delta F}$ of the glucose isomerase data versus resolution.

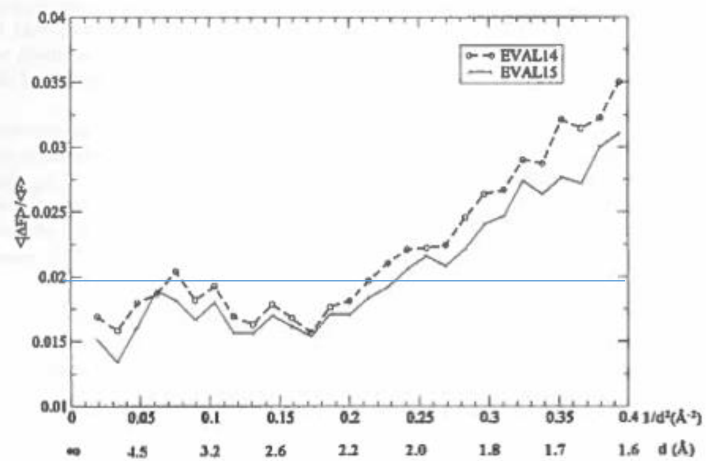
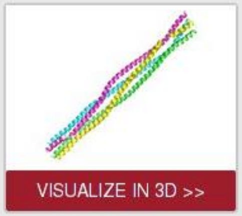


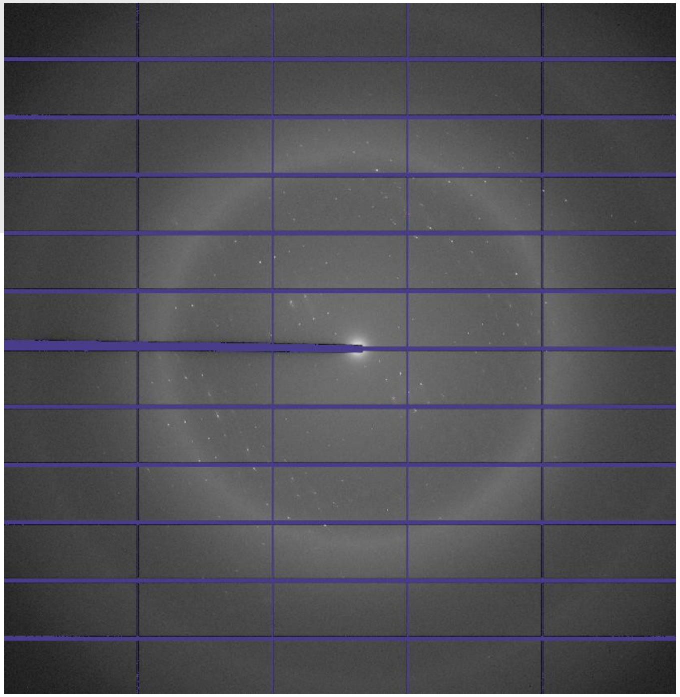

**Fig. 3.15** $\langle \Delta F \rangle / \langle F \rangle$ as a function of resolution.

# More noisy data

X-Ray Diffraction data from M. musculus SYCP3 residues 105-248, source of 6DD8 structure




Data DOI: 10.15785/SBGRID/583 | ID: 583
Publication DOI: 10.7554/eLife.40372
6DD8 Coordinates: Viewer, PDB (RCSB) (PDBe), MMDB
Corbett Laboratory, University of California, San Diego
Release Date: Jan. 25, 2019



Pilatus3 6M detector at APS 24-ID-C
Rotation increment 0.4°

# Statistics

SBgrid 583

Completeness and Rmerge for Shells
Forbid: EDGEVER EDGEROT BADUNIF MAXSHIFT
Allow: GOOD WEAK NEGATIVE
Require: NONE
limit reso 30.79 2.6 inside limit theta 0.91 10.86 inside
theta from 0.0 to 10.855

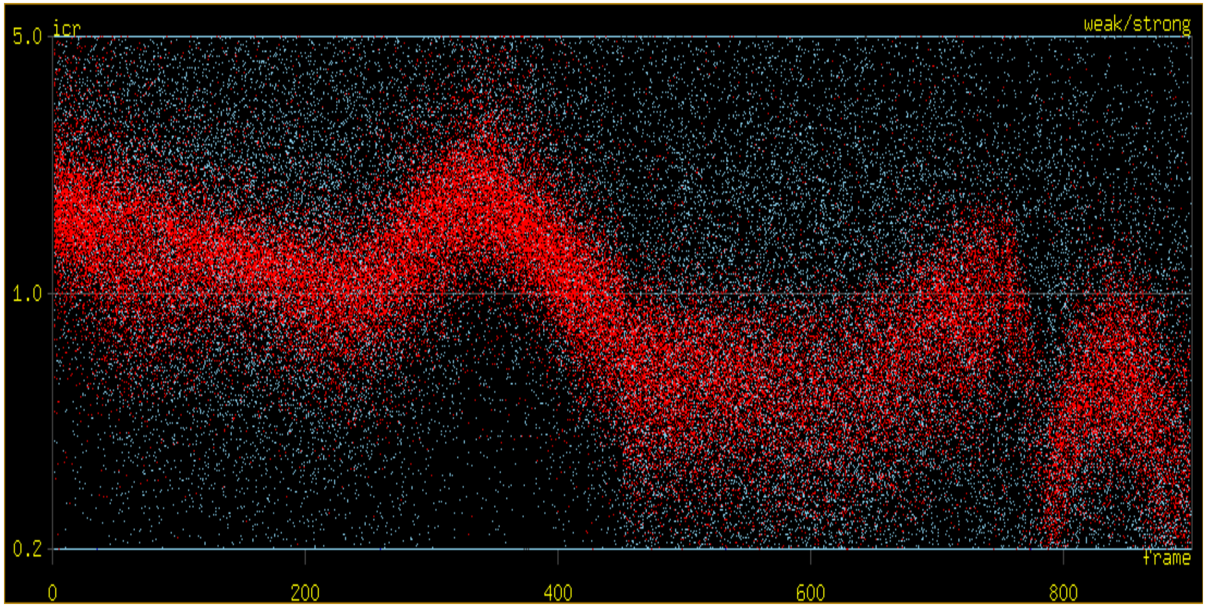
| Sh | Theta | Reso | Meas | Equi | Obs | Mis | Lost | Total | Perc | Cum | Uni1 | Uni2+ | Nrsym | Redun | Rsym | Rmeas | Rpim | Chi2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 5.01 | 5.602 | 6932 | 910 | 7842 | 48 | 170 | 8060 | 97.3 | 97.3 | 18 | 2102 | 13284 | 6.32 | 0.270 | 0.298 | 0.124 | 428223.25 |
| 2 | 6.32 | 4.446 | 7254 | 836 | 8090 | 0 | 52 | 8142 | 99.4 | 98.3 | 14 | 2105 | 13981 | 6.64 | 0.435 | 0.474 | 0.186 | 1985.25 |
| 3 | 7.24 | 3.884 | 6932 | 1028 | 7960 | 0 | 178 | 8138 | 97.8 | 98.2 | 27 | 2038 | 13077 | 6.42 | 0.480 | 0.522 | 0.204 | 641.12 |
| 4 | 7.98 | 3.529 | 6938 | 926 | 7864 | 0 | 60 | 7924 | 99.2 | 98.4 | 23 | 2011 | 13233 | 6.58 | 0.536 | 0.584 | 0.227 | 167.30 |
| 5 | 8.60 | 3.276 | 7346 | 784 | 8130 | 0 | 52 | 8182 | 99.4 | 98.6 | 19 | 2078 | 14144 | 6.81 | 0.629 | 0.682 | 0.260 | 54.97 |
| 6 | 9.14 | 3.083 | 6850 | 1080 | 7930 | 0 | 130 | 8060 | 98.4 | 98.6 | 18 | 2024 | 12863 | 6.36 | 0.670 | 0.730 | 0.286 | 29.18 |
| 7 | 9.63 | 2.928 | 6830 | 1138 | 7968 | 0 | 104 | 8072 | 98.7 | 98.6 | 43 | 2004 | 12647 | 6.31 | 0.681 | 0.741 | 0.288 | 42.58 |
| 8 | 10.07 | 2.801 | 7178 | 980 | 8158 | 0 | 56 | 8214 | 99.3 | 98.7 | 31 | 2067 | 13608 | 6.58 | 0.877 | 0.953 | 0.368 | 13.14 |
| 9 | 10.48 | 2.693 | 7156 | 862 | 8018 | 0 | 22 | 8040 | 99.7 | 98.8 | 13 | 2043 | 13654 | 6.68 | 1.078 | 1.168 | 0.445 | 7.13 |
| 10 | 10.86 | 2.600 | 7116 | 778 | 7894 | 0 | 98 | 7992 | 98.8 | 98.8 | 22 | 2005 | 13612 | 6.79 | 1.297 | 1.407 | 0.538 | 5.14 |
| | 10.86 | 2.600 | 70532 | 9322 | 79854 | 48 | 922 | 80824 | 98.8 | 98.8 | 228 | 20477 | 134103 | 6.55 | 0.410 | 0.448 | 0.178 | 42720.21 |

Intensity distribution for Shells, unmerged and merged
Forbid: EDGEVER EDGEROT BADUNIF MAXSHIFT
Allow: GOOD WEAK NEGATIVE
Require: NONE
limit reso 30.79 2.6 inside limit theta 0.91 10.86 inside

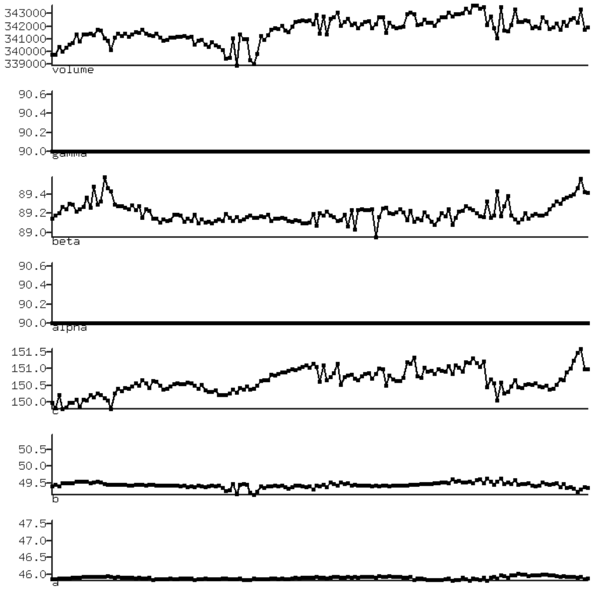
| Sh | Theta | Reso | N | <I> | <s> | <I/s> | Nmerge | <I> | <s> | <I/s> | cc1/2 | cc* | npair |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 5.01 | 5.602 | 13302 | 152.64 | 1.14 | 100.68 | 2120 | 179.26 | 23.03 | 7.71 | 0.918 | 0.979 | 2102 |
| 2 | 6.32 | 4.446 | 13995 | 100.13 | 1.70 | 36.76 | 2119 | 117.96 | 20.32 | 6.44 | 0.905 | 0.975 | 2105 |
| 3 | 7.24 | 3.884 | 13104 | 45.48 | 1.84 | 19.18 | 2065 | 52.27 | 9.69 | 5.39 | 0.872 | 0.965 | 2038 |
| 4 | 7.98 | 3.529 | 13256 | 26.18 | 2.12 | 10.42 | 2034 | 31.09 | 6.35 | 4.59 | 0.841 | 0.956 | 2011 |
| 5 | 8.60 | 3.276 | 14163 | 12.26 | 2.12 | 5.29 | 2097 | 15.00 | 3.38 | 3.77 | 0.820 | 0.949 | 2078 |
| 6 | 9.14 | 3.083 | 12881 | 7.96 | 2.09 | 3.30 | 2042 | 9.35 | 2.47 | 2.69 | 0.835 | 0.954 | 2024 |
| 7 | 9.63 | 2.928 | 12690 | 8.34 | 2.10 | 3.12 | 2047 | 9.37 | 2.54 | 2.00 | 0.877 | 0.967 | 2004 |
| 8 | 10.07 | 2.801 | 13639 | 3.79 | 2.12 | 1.65 | 2098 | 4.44 | 1.58 | 1.55 | 0.781 | 0.936 | 2067 |
| 9 | 10.48 | 2.693 | 13667 | 2.60 | 2.16 | 1.06 | 2056 | 3.01 | 1.33 | 1.16 | 0.827 | 0.952 | 2043 |
| 10 | 10.86 | 2.600 | 13634 | 1.92 | 2.21 | 0.83 | 2027 | 2.32 | 1.24 | 1.05 | 0.795 | 0.941 | 2005 |
| | 10.86 | 2.600 | 134331 | 36.25 | 1.96 | 18.23 | 20705 | 43.04 | 7.28 | 3.66 | 0.918 | 0.978 | 20477 |

# Radiation damage?

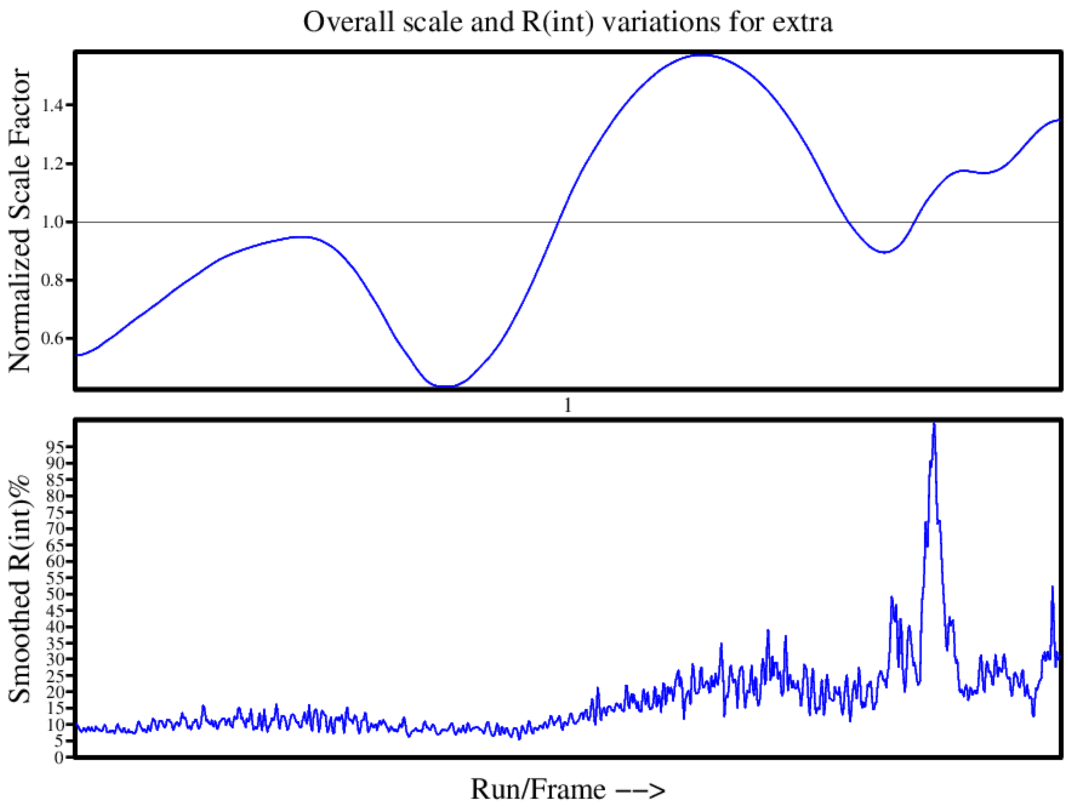
icr = intensity control reflection




Cell dimensions



# Scaling in SADABS




- Increased disorder -> larger B
- Loss in intensity due to structural change: correction to zero dose

| B(start) | B(mid) | B(end) | Rad. damage factors |
|---|---|---|---|
| 0.000 | 17.497 | 34.995 | 0.558 - 4.381 |

# Statistics after scaling

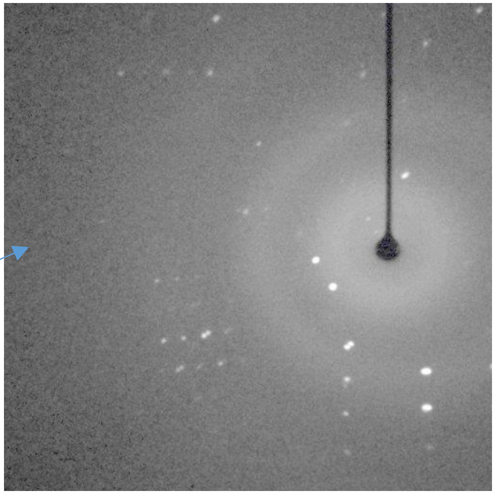
```
Completeness and Rmerge for Shells
Forbid: EDGEVER EDGEROT BADUNIF MAXSHIFT SADABS
Allow: GOOD WEAK NEGATIVE
Require: NONE
limit reso 29.99 2.6 inside limit theta 0.94 10.86 inside
theta from 0.935 to 10.855
Sh Theta  Reso  Meas  Equi   Obs Mis Lost Total Perc  Cum Uni1 Uni2+  Nrsym Redun  Rsym Rmeas  Rpim Chi2
 1  5.01 5.602  6080  1686  7766   0  238  8004 97.0 97.0   77  2015   9710  4.82 0.069 0.079 0.036 0.87
 2  6.32 4.446  6931  1123  8054   0   88  8142 98.9 98.0   24  2083  11545  5.54 0.072 0.080 0.035 0.99
 3  7.24 3.884  6679  1249  7928   0  210  8138 97.4 97.8   41  2016  11295  5.60 0.102 0.113 0.048 1.10
 4  7.98 3.529  6806  1052  7858   0   66  7924 99.2 98.1   30  2002  12039  6.01 0.155 0.170 0.070 1.19
 5  8.60 3.276  7292   838  8130   0   52  8182 99.4 98.4   27  2070  13389  6.47 0.271 0.295 0.116 1.19
 6  9.14 3.083  6819  1111  7930   0  130  8060 98.4 98.4   20  2022  12467  6.17 0.386 0.423 0.169 1.08
 7  9.63 2.928  6794  1172  7966   0  104  8070 98.7 98.4   44  2003  12321  6.15 0.394 0.432 0.173 1.01
 8 10.07 2.801  7163   997  8160   0   56  8216 99.3 98.5   33  2065  13430  6.50 0.777 0.844 0.326 0.93
 9 10.48 2.693  7143   875  8018   0   22  8040 99.7 98.7   15  2041  13497  6.61 1.195 1.296 0.496 0.91
10 10.86 2.600  7105   789  7894   0   98  7992 98.8 98.7   25  2002  13506  6.75 1.608 1.741 0.660 0.91
=========================================================================================================
   10.86 2.600 68812 10892 79704   0 1064 80768 98.7 98.7  336 20319 123199  6.06 0.142 0.157 0.065 1.02
```

```
Intensity distribution for Shells, unmerged and merged
Forbid: EDGEVER EDGEROT BADUNIF MAXSHIFT SADABS
Allow: GOOD WEAK NEGATIVE
Require: NONE
limit reso 29.99 2.6 inside limit theta 0.94 10.86 inside
Sh Theta  Reso      N   <I>  <s> <I/s> Nmerge   <I>  <s> <I/s> cc1/2   cc* npair
 1  5.01 5.602   9787 55.10 8.11  5.72   2092 54.29 4.14 11.96 0.989 0.997  2015
 2  6.32 4.446  11569 32.74 5.01  5.04   2107 38.44 2.82 11.62 0.997 0.999  2083
 3  7.24 3.884  11336 16.62 2.84  4.28   2057 18.14 1.43  9.80 0.994 0.999  2016
 4  7.98 3.529  12069  9.94 2.15  3.42   2032 11.31 1.06  8.07 0.989 0.997  2002
 5  8.60 3.276  13416  4.83 1.66  2.49   2097  5.46 0.71  6.07 0.976 0.994  2070
 6  9.14 3.083  12487  3.08 1.58  1.68   2042  3.49 0.66  3.95 0.971 0.993  2022
 7  9.63 2.928  12361  3.11 1.72  1.29   2047  3.53 0.70  3.06 0.983 0.996  2003
 8 10.07 2.801  13467  1.56 1.69  0.92   2098  1.75 0.61  2.21 0.913 0.977  2065
 9 10.48 2.693  13512  1.05 1.79  0.65   2056  1.22 0.60  1.58 0.886 0.969  2041
10 10.86 2.600  13531  0.84 1.92  0.56   2027  0.98 0.62  1.38 0.777 0.935  2002
=================================================================================
   10.86 2.600 123535 11.45 2.69  2.45  20655 13.98 1.34  5.99 0.990 0.998 20319
```

# Multi-scan data

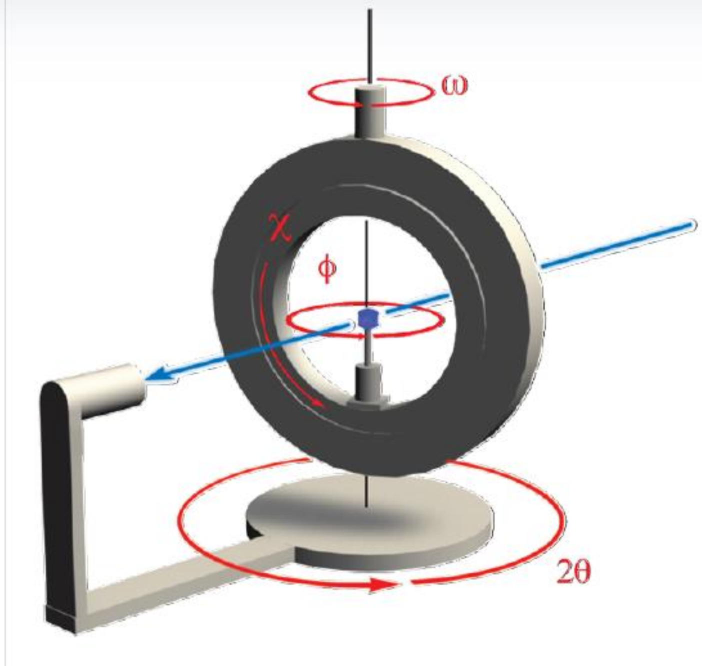
Bruker APEXII CCD

Resolution:
$\theta$max=27.5 d=0.77 Å

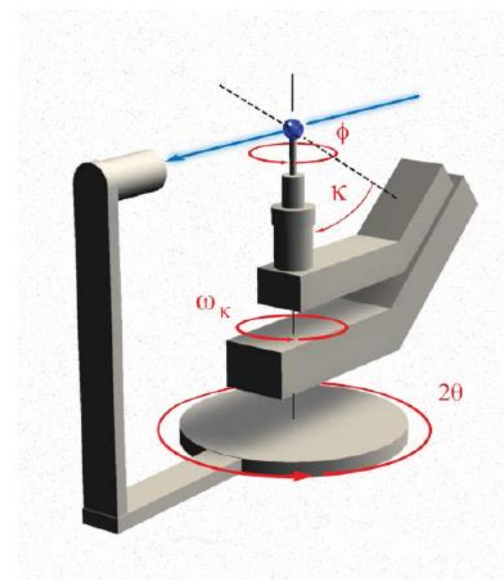


| Nr | Exp | Rotax | dist | swing | inc | frame | frame | Omega | Chi | Phi | rotstart | rotend | sweep |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | s01f | Phi | 45.01 | 22.1 | -0.3 | 1.00 | 1076.00 | 0.00 | 35.00 | 0.00 | 0.00 | 37.45 | -322.80 |
| 2 | s02f | Phi | 45.04 | 22.1 | -0.3 | 1077.00 | 1276.00 | 0.00 | 35.00 | 0.00 | 45.00 | -14.98 | -60.00 |
| 3 | s03f | Omega | 45.01 | 22.1 | 0.3 | 1277.00 | 1689.00 | 0.00 | -54.74 | -156.00 | 13.18 | 137.06 | 123.90 |
| 4 | s04f | Omega | 45.00 | 22.1 | 0.3 | 1690.00 | 1961.00 | 0.00 | 64.88 | 0.00 | -51.68 | 29.87 | 81.60 |
| 5 | s05f | Omega | 44.99 | 22.1 | 0.3 | 1962.00 | 2374.00 | 0.00 | -54.74 | 102.00 | 13.18 | 137.07 | 123.90 |
| 6 | s06f | Omega | 45.01 | 22.1 | 0.3 | 2375.00 | 2787.00 | 0.00 | -54.74 | 0.00 | 13.18 | 137.08 | 123.90 |
| 7 | s07f | Omega | 45.00 | 22.1 | 0.3 | 2788.00 | 3200.00 | 0.00 | -54.74 | 51.00 | 13.18 | 137.06 | 123.90 |
| 8 | s08f | Omega | 44.99 | 22.1 | 0.3 | 3201.00 | 3613.00 | 0.00 | -54.74 | 153.00 | 13.18 | 137.07 | 123.90 |

# Goniometer Geometry




Eulerian Geometry




Kappa Geometry

# Binary image header

- Bruker .sfrm image files have an ascii header
- We normally rely on the data processing software to read metadata from the header and make correct use of it

```
Id=APEX2 Model=D85 [10/02-2786] with KAPPA [50.00000] DetectorType=CCD-LDI-APEX2S
DetectorId=smart10022786 CalibrationId=smart10022786 GoniostatType=x8 GoniostatId=smart10022786
Date=10/03/16 12:28:01 repeats=1 IntegrationTime=15.0 nx=512 ny=512 Binned=no Theta=11.054 Omega=0.0
Chi=35.0 Phi=-0.0 Interval=-0.3 Dx=45.0 RadiationType=XRAY HV=50 MA=30 PixelXsize=120.0
PixelYsize=120.0 Detgain=15.668 DataTypeRead=u8 DataTypeWrite=u8 Target=MO Alpha1=0.7093
Alpha2=0.71359 Alpha ratio=1.99996 Polarisation=PARALLEL CryoTemperature=150.0
CryoActualTemperature=150.0 DetectorTemperature=-57.43 BeamHor=-0.757 BeamVer=-0.042 Format=100
Nunderflow=76 UnderflowSize=1 Noverflow1=37985 Noverflow2=0
Swing=22.108 Dist=45.0 Axis=3 RotAxis=Phi StartPos=0.0 35.0 -0.0 EndPos=0.0 35.0 -0.3
MeanPos=0.0 35.0 -0.15 Goniostat=0.0 35.0 0.0 RotStart=-0.0 RotInc=-0.3 RotEnd=-0.3 RotValue=-0.15
OverflowLevel=960000 RescaleFactor=16.0 RescaleLevel=120000.0 DoSwapHeader=on DoSwapData=on
No AdcZero correction fast=right slow=down
```

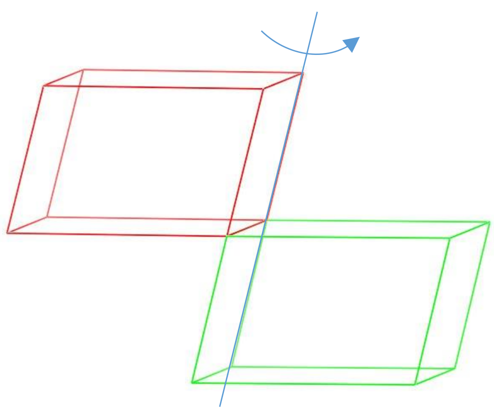
VIEW (EVAL software suite)

# Twin crystal: indexing

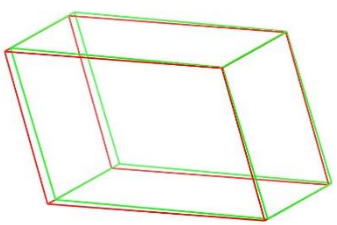
```
1000 c-vectors from file i.drx. 3749 input reflections ignored
Dirax> go
165999834 triplets
30000 triplets used
Randomizing [i,j,k]...
30000 random triplets
29999 triplet vectors
Squishd: 29999 t vectors ==> 29996 t vectors
Sorting 29996 t vectors...
Reducing 29996 t vectors ==> 28185 t vectors
Acl    nH      a       b      c  alpha   beta  gamma Volume S
371     9  4.681   5.223 58.300  90.10  91.53 114.50   1296
362     5  1.789   4.269  8.373  89.99  92.93  96.13     63 ?
356    56  8.351   9.153 13.554 104.49  90.22  90.16   1003
262    21  8.256   9.133 13.538 105.37  91.96  91.23    983
                             .
                             .

250   153  8.340   9.078 13.572 105.34  91.37  90.69    991
240    20  8.390   9.176 13.336 104.95  90.79  91.85    991
239    51  8.333   9.147 13.459 104.54  90.85  90.60    993
236    17  8.408   9.002 13.784 105.90  91.88  90.99   1002
                             .
                             .

 34     7  6.226   6.228  9.377  81.93  82.52  84.71    356
 28     3  3.952   4.107  4.853  90.36 104.70  98.26     75 ?

250   153  8.340   9.078 13.572 105.34  91.37  90.69    991
selected ACL 250
```

# Non-merohedral twin crystal: indexing

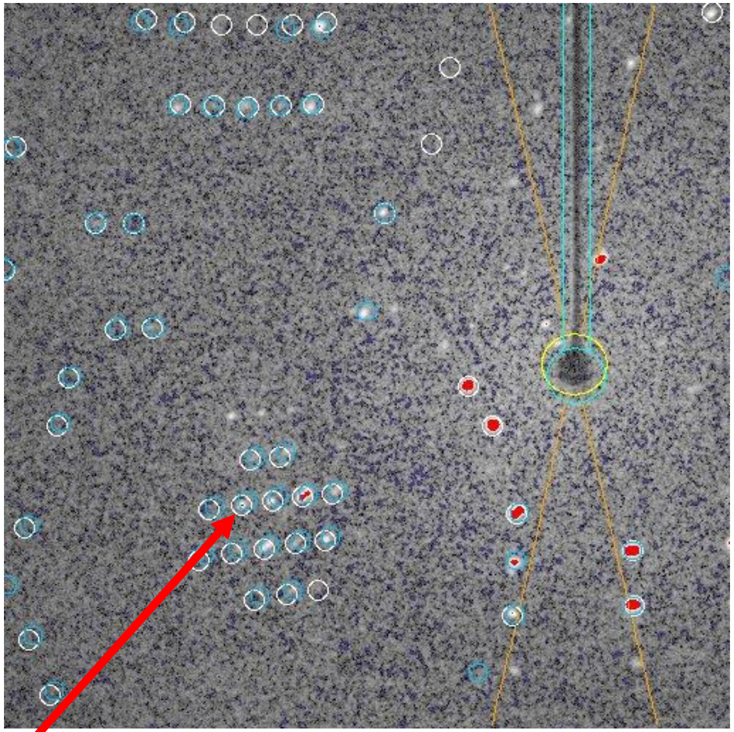
```
Correlation=-0.82
                  a        b        c      alpha     beta    gamma     volume
Save A :    8.340    9.078  13.572  105.34    91.37    90.69      990.5
Save B :    8.367    9.075  13.593  105.22    91.48    91.03      995.2
Volume ratio = 0.995 Trying 64 solutions
Nr Rotangle        Rotvec(xyz)             RotVec(hkl)      ( angle)      RotVec(uvw)      ( angle) Obliq      Fom
 1     1.819 -0.8603  0.5098  0.0042    0.04   7.01 -8.00 (  0.34)   0.06 12.00 -5.02 (  0.22)  0.18    8.860
 2   179.989  0.4740  0.7873  0.3942  -1.00   0.01   0.04 (  1.79)  -1.00   0.00   0.00 (  0.07)  1.75    1.140 U
Selected Solution 2
                  a        b        c      alpha     beta    gamma     volume
Save A :    8.340    9.078  13.572  105.34    91.37    90.69      990.5 <
Save B :    8.367    9.075  13.593  105.22    91.48    91.03      995.2 <
H'= +1.003*H +0.001*K +0.001*L
K'= -0.032*H -1.000*K -0.001*L
L'= -0.079*H -0.002*K -1.002*L
Nr Rotangle        Rotvec(xyz)             RotVec(hkl)      ( angle)      RotVec(uvw)      ( angle) Obliq      Fom
 2   179.989  0.4740  0.7873  0.3942  -1.00   0.01   0.04 (  1.79)  -1.00   0.00   0.00 (  0.07)  1.75    1.140 U<
```

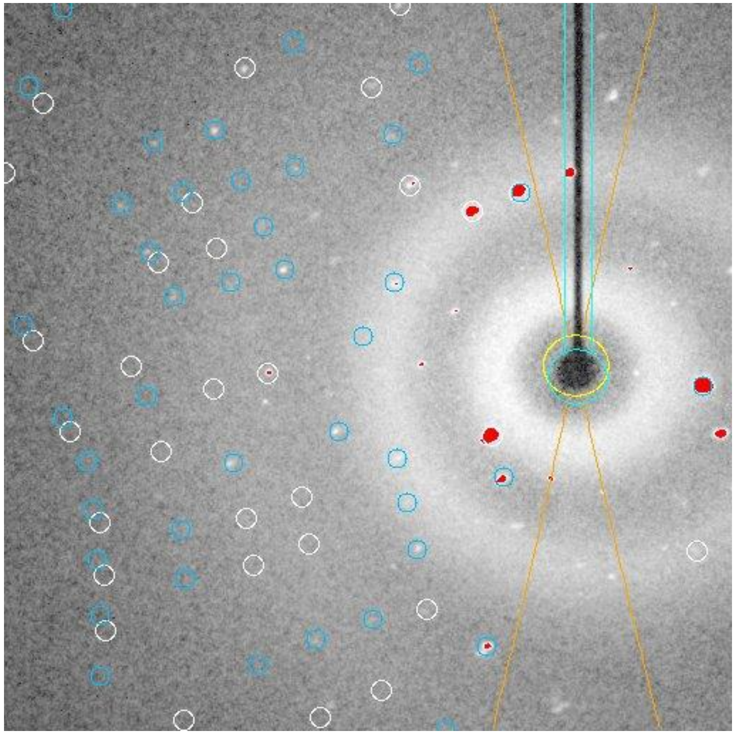


(h,k,l) transforms to ~ (h,-k,-l)



# Twin crystal: indexing





Frame 1

Frame 200

View ~ a*

A: hkl 2 2 9
B: hkl 2 -2 -9

# Validation and open science

Inspect the images:

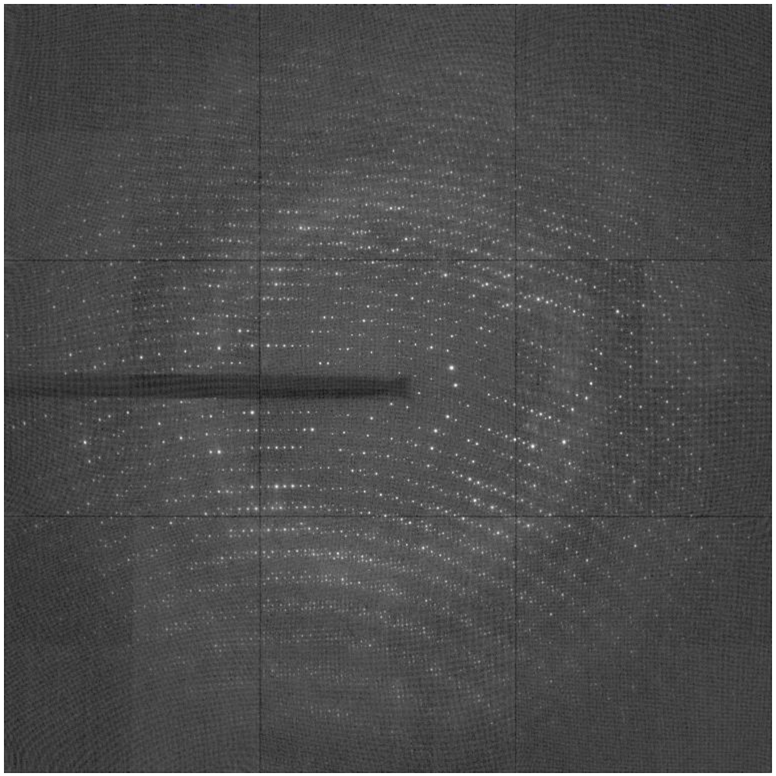
- Where all Bragg spots predicted and integrated?
- Did we understand all features we observed?
- Are we confident the structure solution and refinement is not (strongly) affected by remaining unexplained features
- Data may be useful for other researcher or software developers. Open science is advocated by research funders: please archive you data in a public repository
- Take care the data is FAIR. Correct and sufficient metadata is essential.

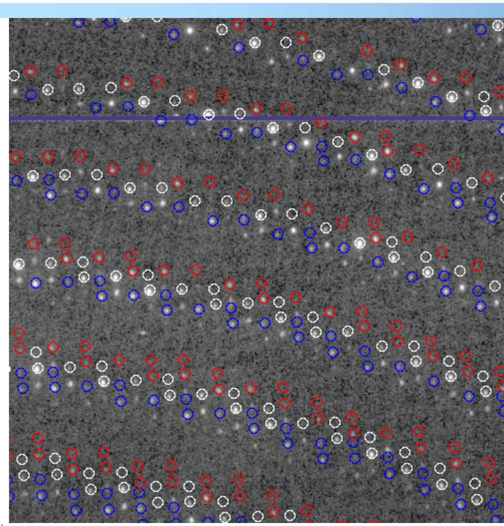
# Core Metadata

| Core Metadata | imgCIF tags |
| --- | --- |
| • Data binary format | _array_structure_byte_order,_array_structure_compression_type |
| • Number of pixels, pixel size (binning mode) | _array_structure_list.index; _array_structure_list.dimensions _array_element_size.size |
| • Beam Center (mm, pixels)<br>• Origin of data frame | _diffrn_detector_element.center[1] _diffrn_detector_element.center[2] |
| • Wavelength<br>• Rotation axis<br>• Rotation range per frame | _diffraction_radiation.wavelength.wavelength _diffrn_scan_axis.axis_id, _diffrn_scan_axis.displacement_start _diffrn_scan_axis.displacement.increment |
| • Axes and offsets<br>• Detector-to-sample distance | _axis.id, _axis.vector[1].., _ _axis.offset[1].. |

# Incommensurate modulation

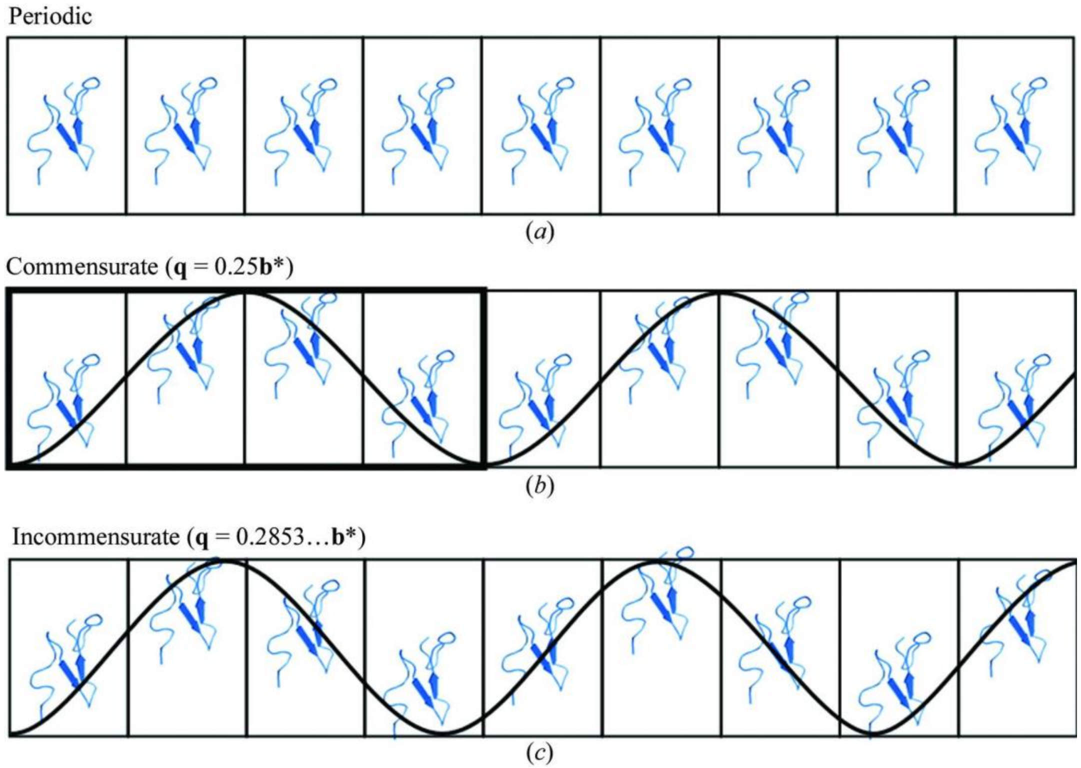
E. coli enzyme N-acetyl-neuraminic lyase





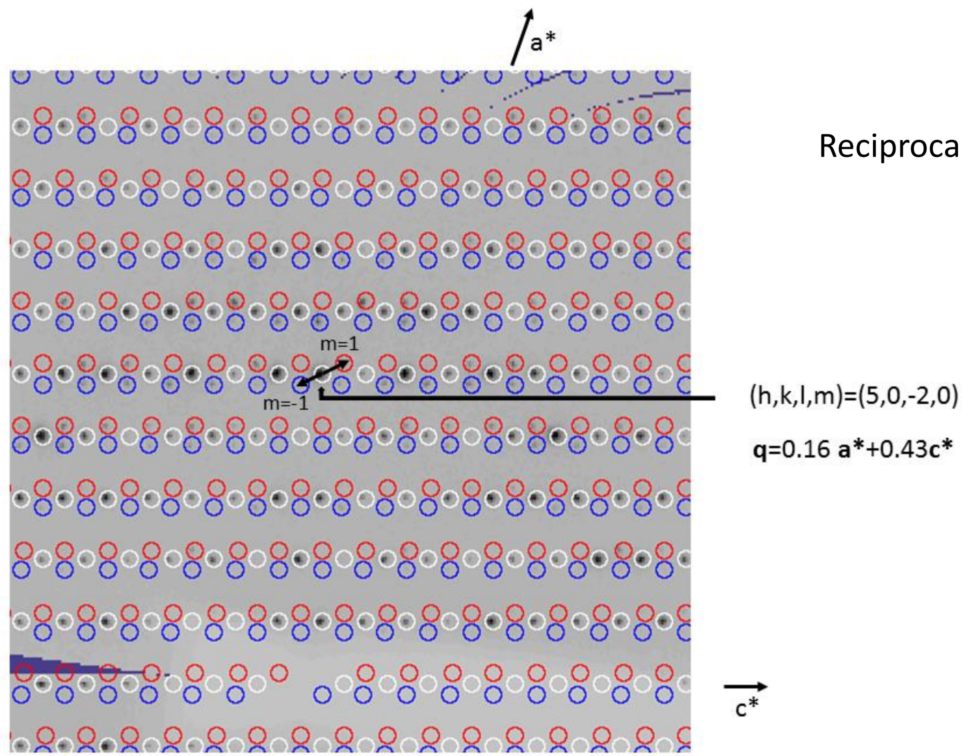
```
rmat from j1r.rmat
checking laue with one qvector
qvec setup for symmetry 2/m
RMAT 1 j1r
                RMAT                                              DMAT
  -0.0172030  -0.0026438  -0.0008334      -50.8232994 -20.1229591    6.3027682
  -0.0075459   0.0064414  -0.0017993      -53.9012222 131.3275757 -15.3532276
  -0.0041504  -0.0007531  -0.0124649       20.1790657  -1.2342228 -81.3962860
Determinant:  0.1606476E-05                             622480.2
cell from rmat: 55.02423 142.78653 83.86938 90.0003 109.1485 90.0002 V= 622480.25
pg constrained: 55.02423 142.78653 83.86938 90.0000 109.1485 90.0000 V= 622480.19
Sigma 0.0266 0.0627 0.0144 0.011 0.019 0.025 Volume 283.49
Bravais=P pg=2/m
Qvectype=2 MODULATED
ActiveQvecRadius=0.007152
QVEC     dh       dk       dl    Order
   1  0.1653  0.0000  0.4316   1
3 generated QV Combinations
```

# Incommensurate modulation: example




Porta et al. (2011). *Acta Cryst.* D**67**, 628–638, 745

# Incommensurate modulation: q-vector




Reciprocal space reconstruction

# Incommensurate modulation

SCIENTIFIC REPORTS

OPEN

Received: 23 January 2018
Accepted: 19 September 2018
Published online: 05 October 2018

## Pathological macromolecular crystallographic data affected by twinning, partial-disorder and exhibiting multiple lattices for testing of data processing and refinement tools

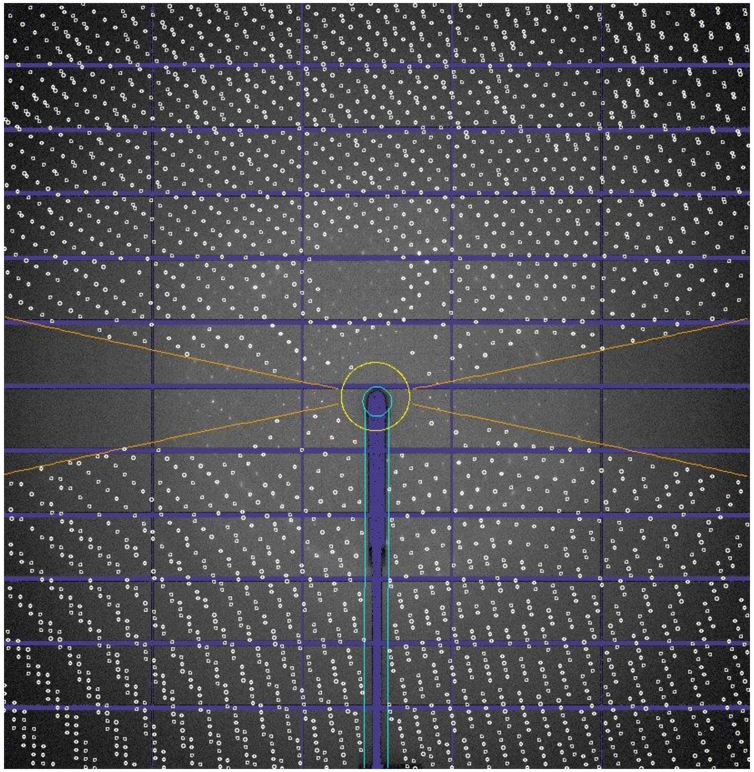
Ivan Campeotto[1,2,3], Andrey Lebedev[4], Antoine M. M. Schreurs[5], Loes M. J. Kroon-Batenburg[5], Edward Lowe[2], Simon E. V. Phillips[1,4], Garib N. Murshudov[6] & Arwen R. Pearson[1,7]

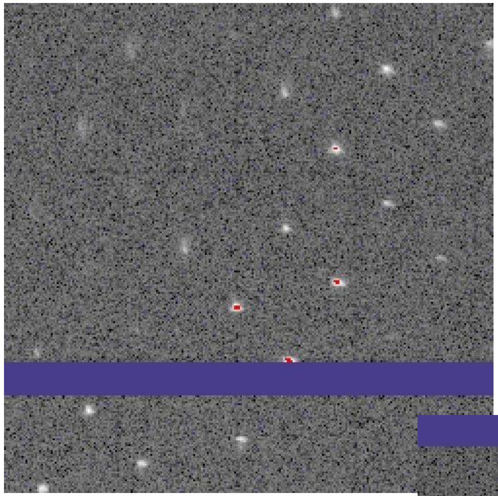
### Data Records

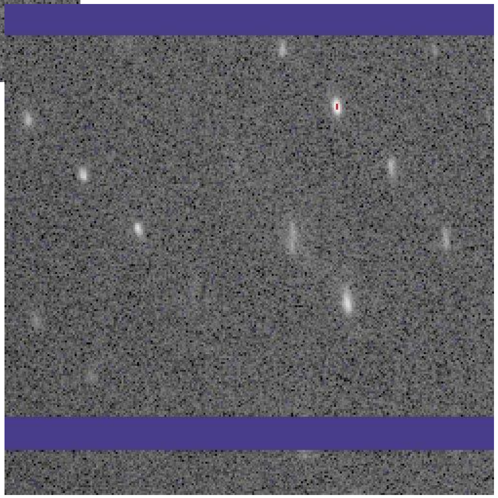
The datasets (raw diffraction images) discussed in this manuscript have been deposited in the publicly available database zenodo at, https://doi.org/ 10.5281/zenodo.54568 and 10.5281/zenodo.1240503. Structural models and processed structure factor data deposited in the PDB are available under the accession codes given in Table 1, with the exception of dataset Y137A, as the R factor indices were not satisfactory for PDB deposition.

# Watch out! Diffuse scattering

Frame 1





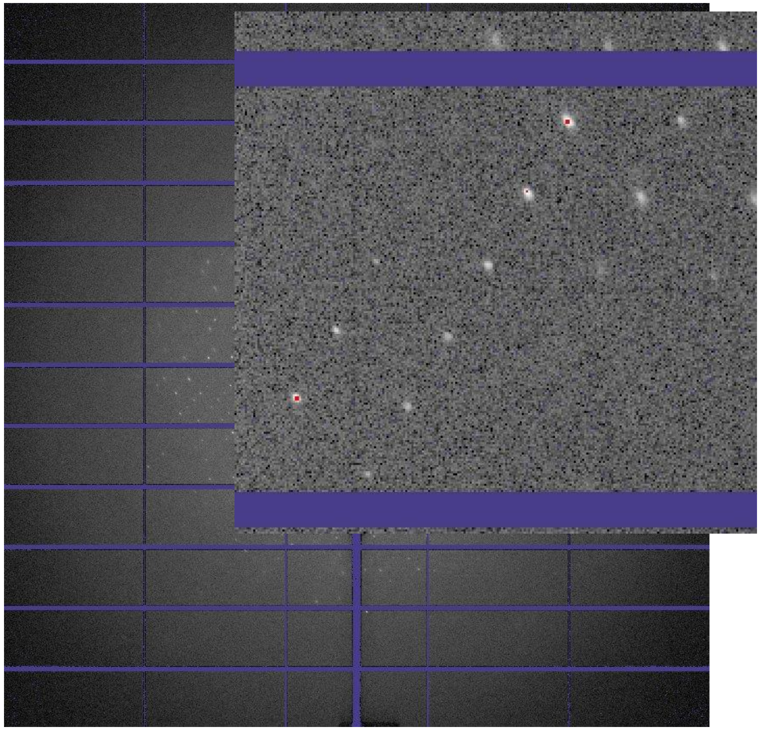


ROR2-CRD protein
R. Van Scherpenzeel
Thesis, Utrecht 2018

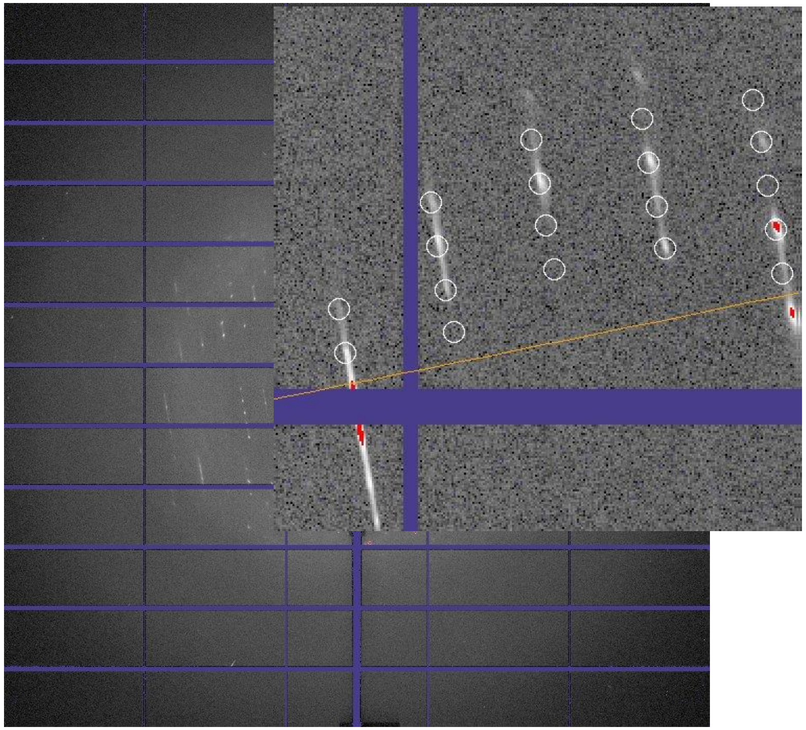
**Pilatus 6M hybrid pixel**

# Watch out! Diffuse scattering

Frame 90



Frame 45

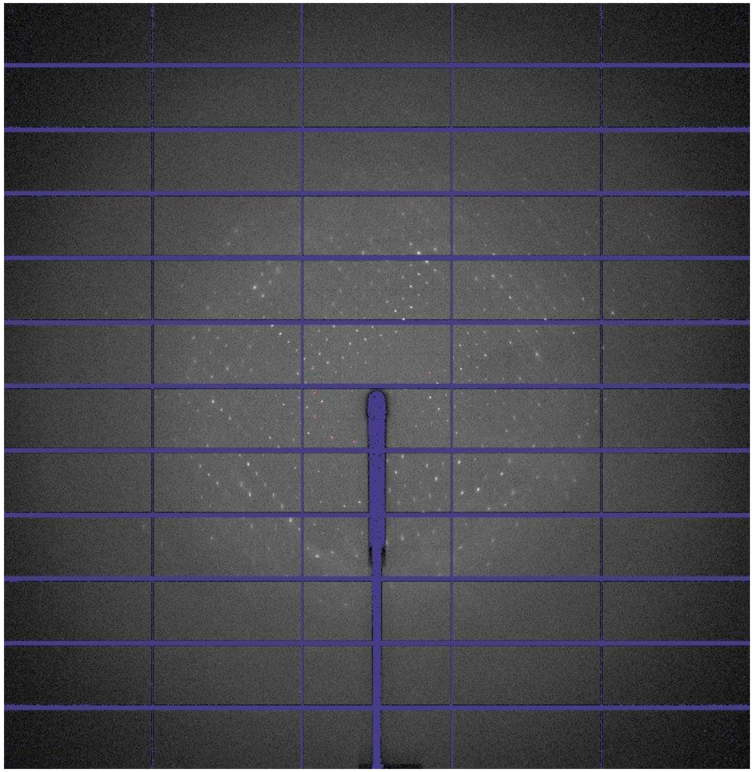


23

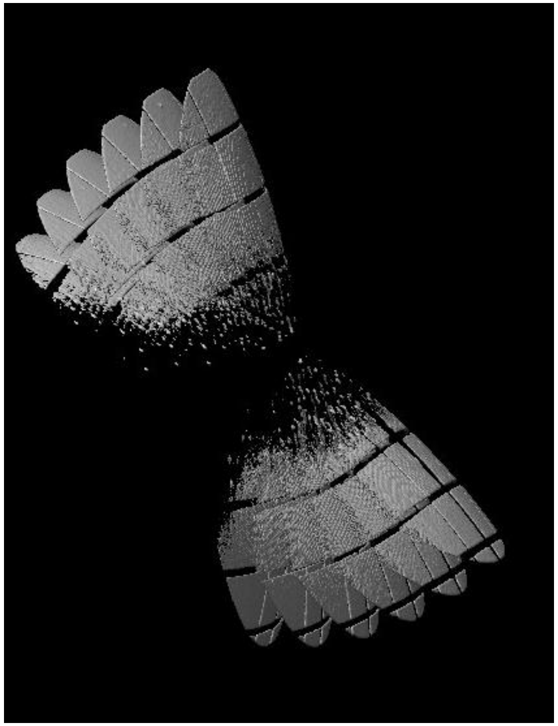
# Mapping detector pixels to reciprocal space

Pixel -> η

Detector panel



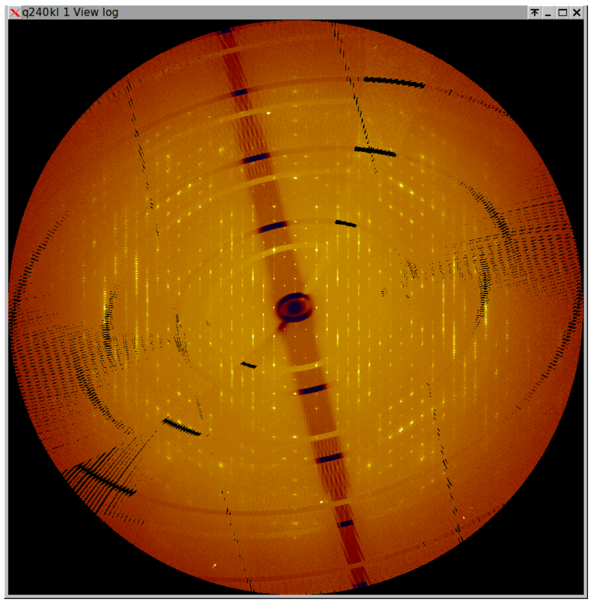
Ewald sphere



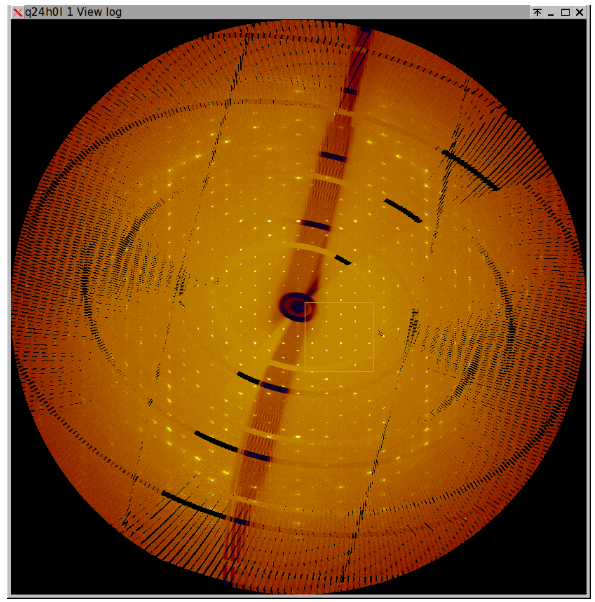
IMG2HKL

# Reciprocal space reconstructions

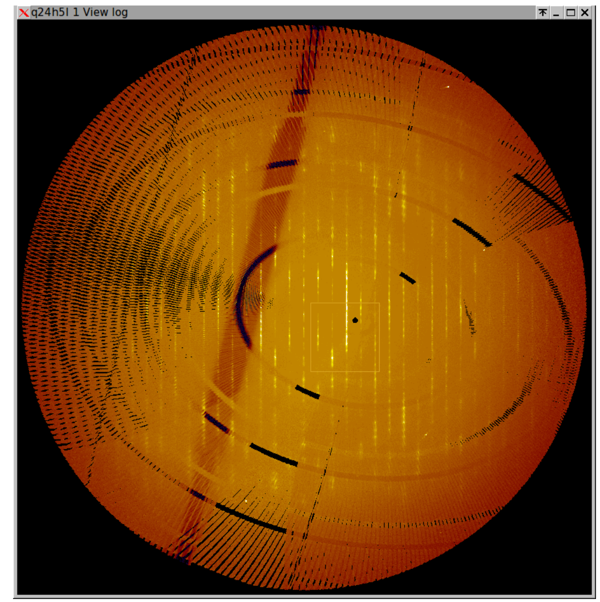
Streaks in b*-direction




hk0

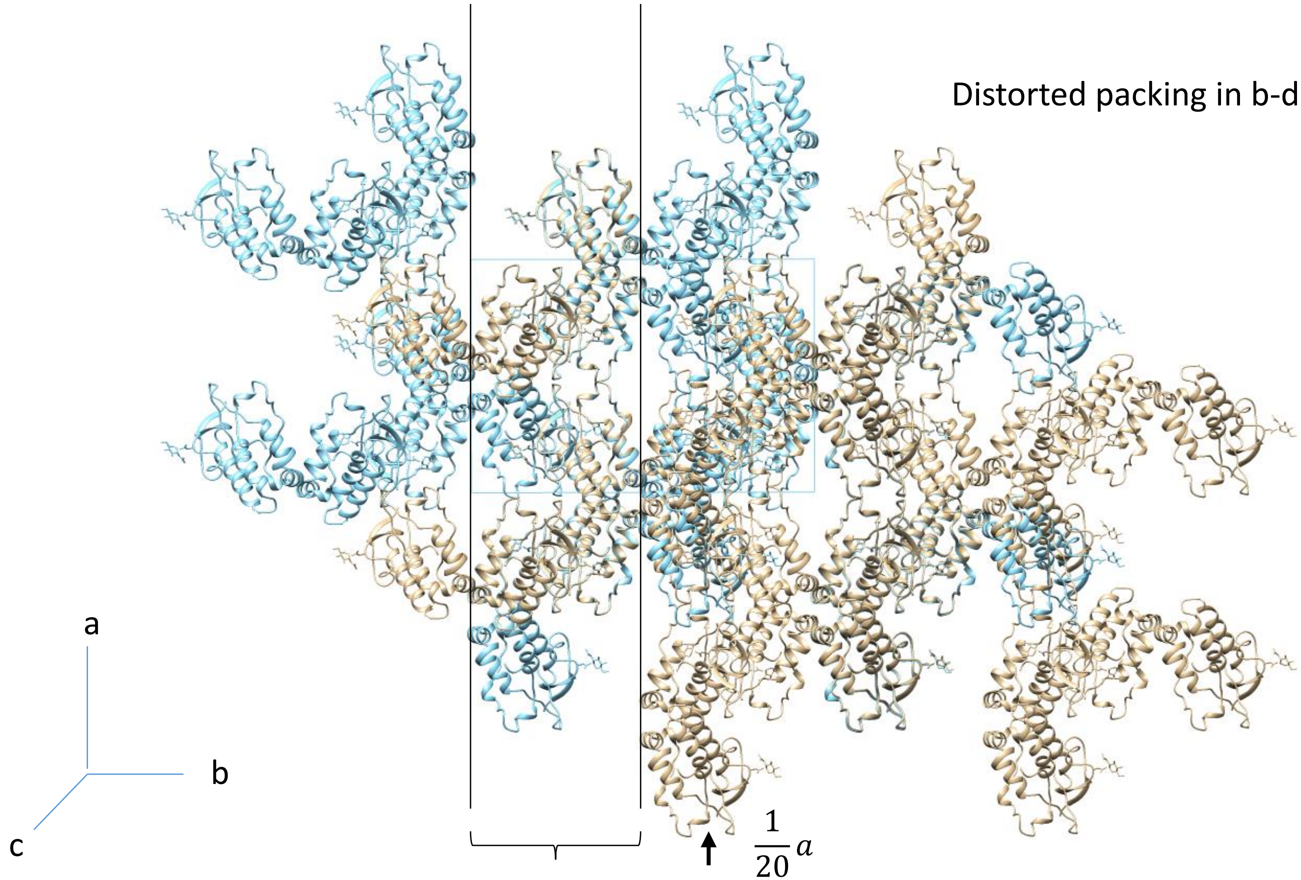



0kl




5kl

Distorted packing in b-direction



# Space group determination

**XPREP**

SBgrid 583

```
Determination of reduced (Niggli) cell

 Transformation from original cell (HKLF-matrix):
    1.0000  0.0000  0.0000     0.0000  1.0000  0.0000     0.0000  0.0000  1.0000

 Unitcell:      45.898   49.431  150.563    90.00    89.26    90.00

 Niggli form:     a.a =    2106.62       b.b =    2443.44       c.c =  22669.28
                  b.c =       0.00       a.c =      89.30       a.b =      0.00

 Search for higher METRIC symmetry
 Identical indices and Friedel opposites combined before calculating R(sym)

 -------------------------------------------------------------------------------
 Option A: FOM = 0.740 deg.   ORTHORHOMBIC P-lattice   R(sym) = 0.732 [  25203]
 Cell:    45.898  49.431 150.563   90.00   89.26   90.00    Volume:     341567.50
 Matrix: 1.0000  0.0000  0.0000  0.0000  1.0000  0.0000  0.0000  0.0000  1.0000
 -------------------------------------------------------------------------------
 Option B: FOM = 0.000 deg.   MONOCLINIC    P-lattice   R(sym) = 0.054 [  15691]
 Cell:    45.898  49.431 150.563   90.00   90.74   90.00    Volume:     341567.50
 Matrix: 1.0000  0.0000  0.0000  0.0000 -1.0000  0.0000  0.0000  0.0000 -1.0000
 -------------------------------------------------------------------------------
 Option C: FOM = 0.740 deg.   MONOCLINIC    P-lattice   R(sym) = 0.751 [  16048]
 Cell:    45.898 150.563  49.431   90.00   90.00   89.26    Volume:     341567.50
 Matrix: 1.0000  0.0000  0.0000  0.0000  0.0000  1.0000  0.0000 -1.0000  0.0000
 -------------------------------------------------------------------------------
 Option D: FOM = 0.740 deg.   MONOCLINIC    P-lattice   R(sym) = 0.777 [  15447]
 Cell:    49.431  45.898 150.563   90.74   90.00   90.00    Volume:     341567.50
 Matrix: 0.0000  1.0000  0.0000  1.0000  0.0000  0.0000  0.0000  0.0000 -1.0000
 -------------------------------------------------------------------------------
 Option E retains original cell
```

```
[A] Triclinic,  [M] Monoclinic,  [O] Orthorhombic,  [T] Tetragonal,
[H] Trigonal/Hexagonal,  [C] Cubic  or  [E] EXIT

Select option [M]:


Lattice exceptions: P      A      B      C      I      F     Obv    Rev    All

N (total) =         0  61703  61680  61679  61633  92531  82442  82433 123535
N (int>3sigma) =    0  22732  22794  22980  22872  34253  30745  30520  45882
Mean intensity =  0.0    2.5    2.6    2.4    2.4    2.5    2.5    2.4    2.5
Mean int/sigma =  0.0    2.6    2.6    2.6    2.6    2.6    2.6    2.6    2.6

Lattice type [P, A, B, C, I, F, O(obv.), R(rev. rhomb. on hex. axes)]

Select option [P]:


Mean |E*E-1| = 1.045 [expected .968 centrosym and .736 non-centrosym]


Systematic absence exceptions:

        -21-   -a-   -c-   -n-

N         35  2395  2383  2396
N I>3s     4  1133  1087  1062
<I>      0.1   4.6   5.3   5.1
<I/s>    1.1   3.3   3.2   3.1


Identical indices and Friedel opposites combined before calculating R(sym)

Option  Space Group  No.  Type  Axes  CSD  R(sym) N(eq)  Syst. Abs.   CFOM

[A] P2          #  3  chiral    1    29  0.054  15691  0.0 /  1.1  22.00
[B] P2/m        # 10  centro    1    21  0.054  15691  0.0 /  1.1  14.25
[C] Pm          #  6  non-cen   1     1  0.054  15691  0.0 /  1.1  68.67
[D] P2(1)       #  4  chiral    1  3543  0.054  15691  1.1 /  2.6  15.13
[E] P2(1)/m     # 11  centro    1   402  0.054  15691  1.1 /  2.6   6.38
```

Chiral -> $P2_1$

# Negative intensities

From Intensities to structure factor amplitudes

$$F = \sqrt{I}$$

$$2\frac{\sigma_F}{F} = \frac{\sigma_I}{I} \quad \rightarrow \quad \sigma_F = \frac{1}{2}\frac{\sigma_I}{\sqrt{I}}$$

Conserving relative errors

# Wilson distribution

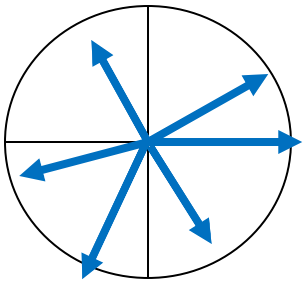
centric:

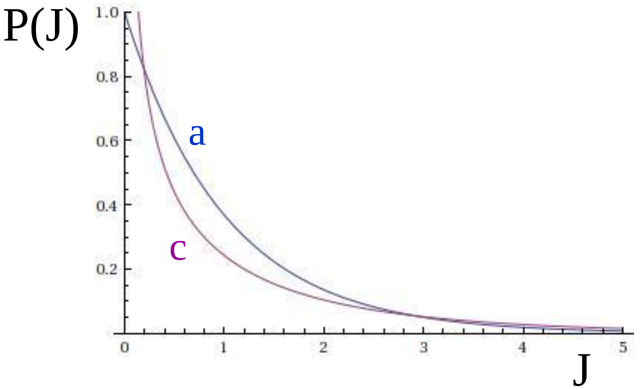
$$P(J) = (2\pi \sum_N J)^{-1/2} \exp(-J/2\sum_N)$$

acentric:

$$P(J) = \sum_N^{-1} \exp(-J/\sum_N)$$

$$\sum_N = \sum_j f_j^2$$

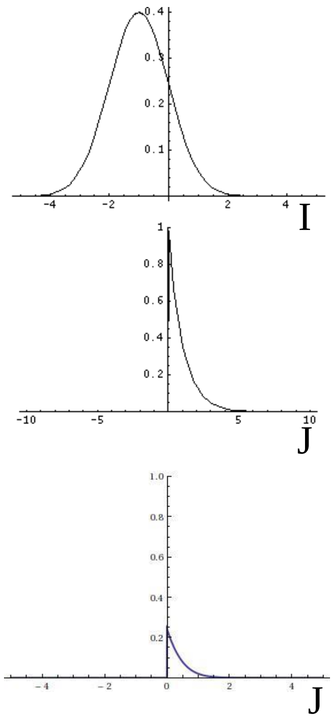


# Unconditional structure factor probability distribution

Inflating negative and weak intensities

$$P(J|I) = P(I|J)P(J)$$




$$P(I|J) = \frac{1}{\sigma\sqrt{2\pi}}\exp(-(I-J)^2/2\sigma^2)$$

I = -1.0
σ= 1.0

J = true value

$$P(J) = \sum\nolimits_N^{-1} \exp(-J/\Sigma_N) \quad \text{if } J\geq 0$$

$$P(J) = 0 \quad \text{if } <0$$

Wilson distribution

$\Sigma_N$ = mean I in resolution shell=20

Estimate: $E(J|I) = \int_0^\infty J\, P(I|J)P(J)dJ$ =0.51

Implemented in TRUNCATE

French & Wilson (1978): Bayesian statistics

# Wilson plot

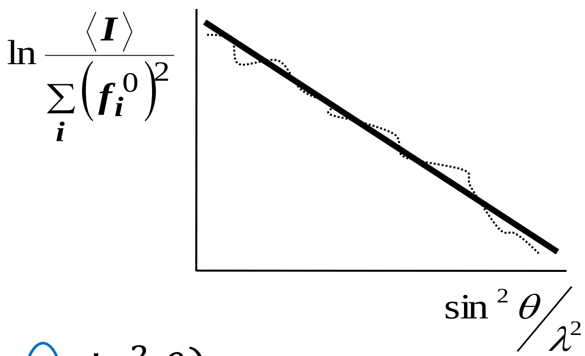
$$P(I) = \sum_N^{-1} \exp(-I / \sum_N)$$

$$I_{\mathbf{abs}}(\boldsymbol{\eta}) = |F(\boldsymbol{\eta})|^2 = F(\boldsymbol{\eta}) \cdot F^*(\boldsymbol{\eta}) = \sum_i \sum_j f_i f_j \exp\{2\pi i(\mathbf{r_i} - \mathbf{r_j}) \cdot \boldsymbol{\eta}\}$$

$$\langle I_{\mathbf{abs}}(\boldsymbol{\eta}) \rangle = \sum_i f_i^2 = \sum_N$$

$$f_i^2 = \left(f_i^0\right)^2 \exp\{-2B \frac{\sin^2\theta}{\lambda^2}\}$$

Average over resolution shells




$$\langle I_{\exp}(\vec{s}) \rangle = k\langle I_{\text{abs}}(\vec{s}) \rangle = k \exp\left\{-2B \frac{\sin^2\theta}{\lambda^2}\right\} \sum_i \left(f_i^0\right)^2$$

Wilson B-factor

# Scale factor and B-factor

After determining the scale factor and B-factor, $F_{calc}$ and $F_{obs}$ can be compared in refinement

Minimize:
$$\sum_{hkl}(|F_{obs}(hkl)| - k\,|F_{calc}(hkl)|)^2$$

$$F_{\mathrm{calc}}(hkl) = \sum_{j=1}^{N} f_j \,\mathrm{exp}\{-B_j(sin\theta/\lambda)^2\}\,\mathrm{exp}\{\,2\pi i(hx_j + ky_j + lz_j)\}$$

# Cherish your data

- ❑ Check that you have understood what you see in the diffraction images
- ❑ Could unprocessed features influence your structure determination/refinement results?
- ❑ Archive your raw data in a FAIR way to:
  - Allow other researchers to conduct further research based your experiments
  - Allow reanalysis at a later date, especially to extract 'new' science as new techniques are developed
  - Provide example materials for teaching and learning.