

# Data from automatic data collection at SPring-8

September 28, 2020

Data collected at SPring-8 PX-BLs with ZOO are processed with XDS through automated data processing pipeline KAMO. The processed data are transferred to users with merged data.

## <Reference of automatic data collection and automatic data processing>

If you publish your structures from data collected using ZOO and analyzed with KAMO, please cite the following papers as references. The citations are exceedingly important for us to continue developing more advanced techniques at synchrotron radiation technology.

Thank you for your cooperation.

All staff

1. K. Hirata, K. Yamashita, G. Ueno, Y. Kawano, K. Hasegawa, T. Kumasaka, and M. Yamamoto  
ZOO: an automatic data-collection system for high-throughput structure analysis in protein microcrystallography  
*Acta Cryst. D75*, 1–13, (2019).  
[doi: 10.1107 / S2059798318017795]
2. K. Yamashita, K. Hirata, and M. Yamamoto  
KAMO: towards automated data processing for microcrystals  
*Acta Cryst. D74*, 441-449, (2018).  
[doi: 10.1107 / S2059798318004576]

## 1. Structure of diffraction dataset directories

Files and directories copied on the media.

```
~/YYMMDD_ blname_username/ <- Root directory
├ "PuckID"- "PinID"/ <- Data directory Containing data
│   ├── ##.ppm, raster.jpg <- On gonio loop image after loop centering
│   ├── scan**/ <- Directory for raster scan (including SHIKA analysis results)
│   └── data**/ <- Directory for dataset (helical, single, multi)
├ _kamoproc/ <- KAMO working directory, processing results
│   ├── "PuckID"- "PinID"/ <- Processing results of XDS
│   ├── report.html, correct.html, contents/ <- Reports of XDS processing
│   └── merge_**_**_"SampleName"/ <- Merge results for small wedge data
│       (e.g. multi mode data collection)
├ _kamo_30deg/ <- 30 deg divided data processing for helical mode data
│   ├── "PuckID"- "PinID"/ <- Processing results of XDS
│   ├── report.html, contents/ <- Reports of XDS processing
│   └── merge_**_**_SampleName/ <- Merge results for the divided data
├ YYMMDD_BL45XU_username_00.html <- Report file of ZOO
├ YYMMDD_BL45XU_username.tgz <- Archived report file containing processed reflection
│   files
└ *csv, *db, *xlsx etc. <- Configuration files for automatic data collection
```

## 2. Diffraction data

The directory name for diffraction data corresponds to "PuckID"- "PinID" of the sample sheet.

```
~/YYMMDD_ blname_username/
├ "PuckID"- "PinID"/
│   ├── ##.ppm, raster.jpg <- On gonio loop image after loop centering
│   ├── scan**/ <- Directory for raster scan (including SHIKA analysis results)
│   └── data**/ <- Directory for dataset (helical, single, multi)
```

## 2.1. Loop centering images

Image files for loop centering stored in the directory are altered upon selected options (e.g. zoom capture, LN2 flash)

1. before\_warmup.ppm <- Before warming up, just after mounting
2. after\_warmup.ppm <- After warming up before precise centering
3. raster.jpg  
<- After loop centering, defined 2D scan area is shown with green rectangle
4. before.ppm <- Same as raster.jpg without 2D scan area indication
5. loop\_zoom.ppm <- zoom up image after loop centering
6. loop\_zoom\_ln2.ppm  
<- zoom up image after loop centering and defrosting with liquid N2

## 2.2. Directory structure

The directory structure (directory tree) for the data-collection scheme is described. There are directories named "PuckID"-**"PinID"**, i.e., "sample directory", directly under the root directory. The sample directories contain raw diffraction data from raster scan and data collection and some images produced with loop centering.

EIGER detector produces hdf5 files. In HDF5 format, every 100 frames are archived such as **\*\*\*\_000001.h5**. (BL32XU, BL41XU). PILATUS detector produces a cbf file from a frame. (BL45XU)

### 2.2.1. multi mode

~/YYMMDD\_blname\_username/

```
├ MUL0001-01/ <- Sample directory
├ ##.ppm, raster.jpg <- Images produced with loop centering (ref. 2.1.)
├ scan*/ <- Containing raster scan data (including SHIKA analysis results)
│   └ 2d/ <- Scan type (2d means 2 dimension)
│       └ spotfinder/ <- Containing SHIKA analysis results
│           └ thumb_2d/ <- Thumbnails of SHIKA diffraction images
└ data00/ <- Containing diffraction images from data collection
    │ (If data collection has been done twice, the number is incremented)
    └ MUL0001-01-multi_###_*****.cbf
        (###: number of crystals that satisfy a criteria. It is limited by "# of
        crystals".)
```

### 2.2.2. helical mode

~/YYMMDD\_blname\_username/

```
├ HEL0002-002/ <- sample directory
├ ##.ppm, raster.jpg <- Images produced with loop centering (ref. 2.1.)
├ scan*/ <- Containing raster scan data (including SHIKA analysis results)
│   └ 2d/ <- Scan type (2d means 2 dimension)
│       └ spotfinder/ <- Containing SHIKA analysis results
│           └ thumb_2d/ <- Thumbnails of SHIKA diffraction images
├ lv00_00/ <- Containing data of vertical scan at left edge
│   └ spotfinder/ <- Containing SHIKA analysis results
│       └ thumb_2d/ <- Thumbnails of SHIKA diffraction images
├ rv00_00/ <- Containing data of vertical scan at right edge
│   └ spotfinder/ <- Containing SHIKA analysis results
│       └ thumb_2d/ <- Thumbnails of SHIKA diffraction images
├ rv00_01/ <- Containing data of second attempt of vertical scan at right
│   │ edge (index 00->01)
│   └ spotfinder/ <- Containing SHIKA analysis results
│       └ thumb_2d/ <- Thumbnails of SHIKA diffraction images
└ data00/ <- Containing diffraction images from data collection
    │ (If data collection has been done twice, the number is incremented)
    └ cry##_*****.cbf/h5: Raw data
        (###: number of crystals that satisfy a criteria. It is limited by "# of crystals".)
```

### 2.2.3. single mode

~/YYMMDD\_ blname\_username/

```
└ SIN003-003/ <- sample directory
├ ##.ppm, raster.jpg <- Images produced with loop centering (ref. 2.1.)
├ scan*/ <- Containing raster scan data (including SHIKA analysis results)
├ 2d/ <- Scan type (2d means 2 dimension)
├   └ spotfinder/ <- Containing SHIKA analysis results
├   └   └ thumb_2d/ <- Thumbnails of SHIKA diffraction images
├   └ lv00_00/ <- Containing data of vertical scan at left edge
├   └   └ spotfinder/ <- Containing SHIKA analysis results
├   └   └ thumb_2d/ <- Thumbnails of SHIKA diffraction images
├ data00/ <- Containing diffraction images from data collection
├   (If data collection has been done twice, the number is incremented)
├   └ single_001_*****.cbf/h5/ <- Raw data
```

- `lv/rv01_00` means a vertical scan of the second crystal in the case of multiple crystals in helical mode.
- In some cases, ZOO repeats vertical scan for centering the crystal at the left and right edges due to weak diffraction from this direction or if the X-ray beam does not hit the crystal. If ZOO cannot detect the edge of the crystal with vertical scan, it shifts the exposure position 10 um inside the crystal and executes vertical scan again (repeat up to 5 times). If ZOO fails vertical scan 5 times, it gives up data collection. Results of the attempts are stored in `lv/rv00_01/` (last digit can be incremented)

### 3. Results of automated data processing pipeline KAMO

The data are processed with the automated data processing pipeline KAMO under the `_kamoproc/` directory. Refer to `XDS.INP` and `CORRECT.LP` to evaluate the processing results.

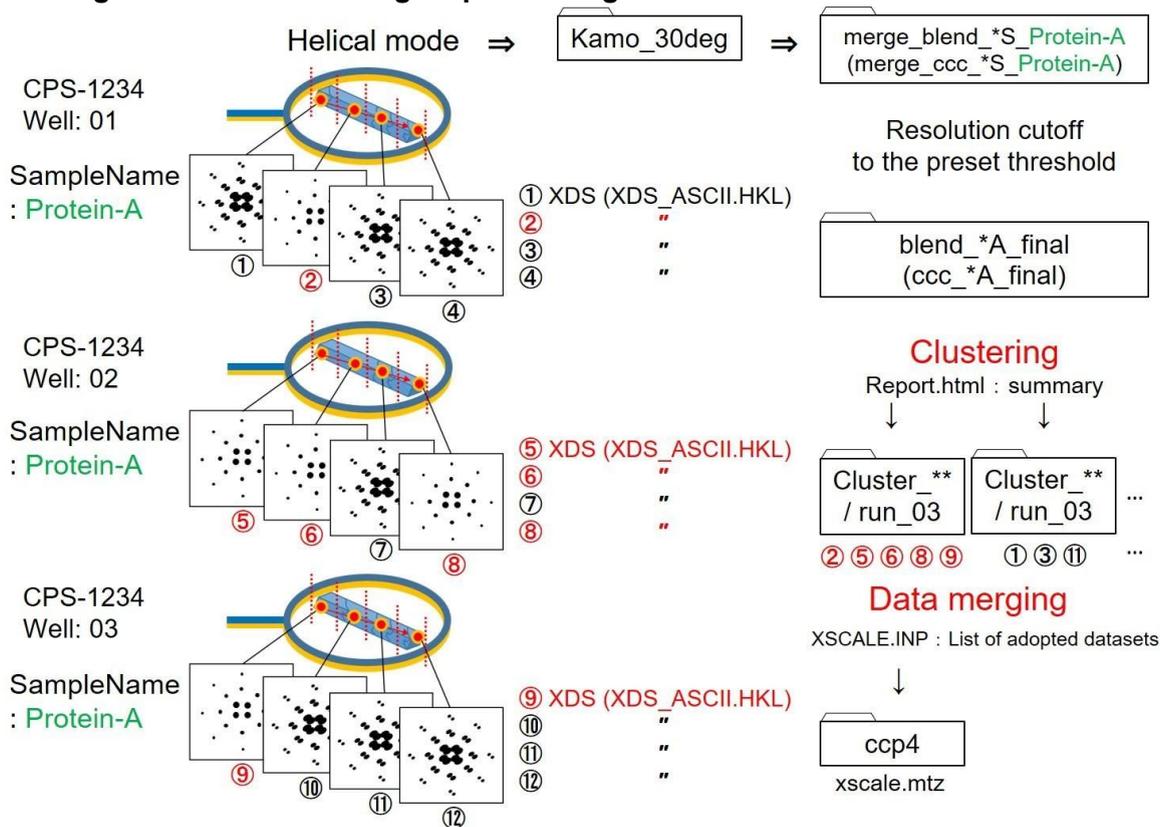
~/YYMMDD\_BLNAME\_USER/

```
└ _kamoproc/ <- KAMO data processing directory
├ MUL0001-01/ <- in multi mode
├   └ data00/ <- same structure as data directory
├   └   └ MUL0001-01-multi_001_1-100/ <- working directory for data processing program
├   └   └ MUL0001-01-multi_002_1-100/
├   └   └ MUL0001-01-multi_003_1-100/
├   └   └ MUL0001-01-multi_0##_1-100/
├ HEL0002-002/ <- in helical mode
├   └ data00/
├   └   └ cry00_1-3600/ <- working directory for data processing program
├ SIN003-003/ <- in single mode
├   └ data00/
├   └   └ single_1-3600/ <- working directory for data processing program
├ report.html <- summary of data processing
├ correct.html <- summary of data statistics of helical/single mode data collection
```

#### Processing helical data as divided small wedge data (KAMO “buttagiri” processing)

For the data collected in helical mode, considering the heterogeneity of the crystals, KAMO processes the data divided into multiple data sets within a 30-deg. oscillation range (e.g. The data that have 720 deg. of total oscillation range is divided into 24 wedges with 30 deg. of oscillation range). The divided data are processed independently followed by merged based on “sample\_name” designated by users. This process is run under `_kamo_30deg/`. The reduced reflection data are found in working directories.

## Conceptual diagram of KAMO “buttagiri” processing



\* Data processing may not be possible depending on the quality of diffraction data

\* In some clusters, completeness or redundancy falls below the threshold, and the mtz file is not output in the end.

~/YYMMDD\_BNAME\_USERNAME/

```

└ _kamo_30deg/ <- KAMO“buttagiri” data processing directory
  └ MUL0001-01/ <- in multi mode
    └ data00/ <- same structure as data directory
      └ MUL0001-01-multi_001_1-100/ <- working directory for data processing program
      └ MUL0001-01-multi_002_1-100/
      └ MUL0001-01-multi_003_1-100/
      └ MUL0001-01-multi_00#_1-100/

  └ HEL0002-002/ <- in helical mode (total 360 deg., 0.1 deg./frame)
    └ data00/ <- same structure as data directory
      └ cry00_1-300/ <- working directory for 0 - 30 deg.
      └ cry00_301-600/ <- working directory for 30 - 60 deg.
      └ ... continue...
      └ cry00_3301-3600/ <- working directory for 330 - 600 deg.

  └ SIN003-003/ <- in helical mode (total 360 deg., 0.1 deg./frame)
    └ data00/
      └ single_1-300/ <- working directory for 0 - 30 deg.
      └ single_301-600/ <- working directory for 30 - 60 deg.
      └ ...continue...
      └ single_3301-3600/ <- working directory for 330 - 600 deg.

  └ report.html <- summary of data processing with “buttagiri”
  
```

#### 4. Report from SPring-8

The archive saved in the root directory as `YYMMDD_BL45XU_USERNAME.tgz` contains report files further described below

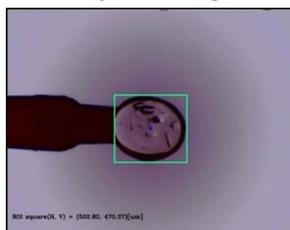
```
~/YYMMDD_BL45XU_USERNAME/
├ report_YYMMDD_BL45XU_USERNAME_00.html <- summary of data collection (ref. 4.1)
├ _kamoproc/ <- outputs from automated data processing pipeline KAMO (ref. 3)
│   ├── correct.html <- summary of data statistics of helical and single mode(ref. 4.2)
│   ├── contents/ <- contains files referenced by correct.html
│   ├── merge_***/ <- results of auto_multi_merge (ref. 5)
│   │   └── blend(/cc)_##A_final/
│   │       ├── report.html (ref. 4.3)
│   │       └── cluster_##/ <- Only the top clusters are included in the report
│   │           └── run_03/
│   │               ├── XSCALE.INP
│   │               ├── xscale.hkl
│   │               ├── XSCALE.LP
│   │               ├── aniso.log
│   │               └── pointless.log
│   ├── "PuckID"- "PinID"/ <- in helical mode
│   │   └── data00/
│   │       ├── cry00_1-3600/ <- working directory for data processing program
│   │       │   ├── CORRECT.LP
│   │       │   └── XDS_ASCII.HKL
│   │   └── "PuckID"- "PinID"/ <- in single mode
│   │       └── data00/
│   │           ├── single_1-###/ <- working directory for data processing program
│   │           │   ├── CORRECT.LP
│   │           │   └── XDS_ASCII.HKL
│   │           └── ... continue...
├ _kamo_30deg/ <- outputs from KAMO buttagiri
│   ├── merge_***/ <- results of auto_multi_merge (ref. 5)
│   ├── blend(/cc)_##A_final
│   │   ├── report.html (ref. 4.3)
│   │   └── cluster_##: Only the top clusters are included in the report
│   │       └── run_03
│   │           ├── XSCALE.INP
│   │           ├── xscale.hkl
│   │           ├── XSCALE.LP
│   │           ├── aniso.log
│   │           └── pointless.log
```

#### 4.1. report\_YMMMDD\_BL45XU\_USERNAME\_00.html, Summary of data collection

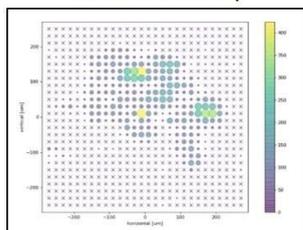
pinid	sample_name	mode	wavelength	total phi	osc width	raster beam	raster area(grid)	#DS	log comment	meas time[min]	loop	2D scan	SHIKA
CPS1012 1	Lys_multi	multi	1.0000	10.0	0.10	20*20 (um)	470*503 (23*25 grids)	30	Normal termination	7.43	O	O	MAP
CPS1012 2	Lys_multi	multi	1.0000	10.0	0.10	20*20 (um)	464*503 (23*25 grids)	24	Normal termination	6.18	O	O	MAP
CPS1012 3	Lys_multi	multi	1.0000	10.0	0.10	20*20 (um)	464*503 (23*25 grids)	25	Normal termination	6.27	O	O	MAP
CPS1012 4	Lys_multi	multi	1.0000	10.0	0.10	20*20 (um)	470*503 (23*25 grids)	30	Normal termination	6.83	O	O	MAP
CPS1012 5	Lys_multi	multi	1.0000	10.0	0.10	20*20 (um)	461*503 (23*25 grids)	30	Normal termination	6.77	O	O	MAP
CPS1012 6	Lys_06	single	1.0000	360.0	0.10	20*20 (um)	260*302 (12*15 grids)	1	Normal termination	5.18	O	O	MAP
CPS1012 7	Lys_07	single	1.0000	360.0	0.10	20*20 (um)	260*302 (12*15 grids)	1	Normal termination	5.17	O	O	MAP
CPS1012 8	Lys_08	single	1.0000	360.0	0.10	20*20 (um)	263*302 (13*15 grids)	1	Normal termination	5.15	O	O	MAP
CPS1012 9	Lys_09	single	1.0000	360.0	0.10	20*20 (um)	256*302 (12*15 grids)	1	Normal termination	5.12	O	O	MAP
CPS1012 10	Lys_10	single	1.0000	360.0	0.10	20*20 (um)	256*302 (12*15 grids)	1	Normal termination	5.07	O	O	MAP
CPS1012 11	Lys_11	helical	1.0000	360.0	0.10	20*20 (um)	256*302 (12*15 grids)	1	Normal termination	6.18	O	O	MAP
CPS1012 12	Lys_12	helical	1.0000	360.0	0.10	20*20 (um)	263*302 (13*15 grids)	1	Normal termination				
CPS1012 13	Lys_13	helical	1.0000	360.0	0.10	20*20 (um)	266*302 (13*15 grids)	1	Normal termination				
CPS1012 14	Lys_14	helical	1.0000	360.0	0.10	20*20 (um)	444*503 (22*25 grids)	1	Normal termination				
CPS1012 15	Lys_15	helical	1.0000	360.0	0.10	20*20 (um)	464*503 (23*25 grids)	1	Normal termination				
CPS1012 16	Lys_16	helical	1.0000	360.0	0.10	20*20 (um)	464*503 (23*25 grids)	1	Normal termination				

#DS: number of datasets collected from the pin.  
 "log comment": Log comment from ZOO (see check).  
 loop: a link to the picture of a loop before raster scan.  
 2D scan: a link to the picture of 2D raster beammap if  
 'X' appears on "loop" or "2D scan", there would not be crystals on the loop.

Crystal image



Scan heat map



SHIKA report

SHIKA report, The diffraction images of any grid of the heat map can be checked.

\* Pop-up images cannot be viewed with this html file alone. Please check after the HDD arrives.

"#DS", number of data sets from the pin

"log comment", status of data collection

1. Normal termination  
Data collection has been done normally.
2. No crystals found in a raster scan  
As a result of raster scan, there was no diffraction exceeding the threshold value; thus data collection has not been done (helical/multi mode).
3. Crystals were not found in a raster scan  
As a result of raster scan, there was no diffraction exceeding the threshold value; thus data collection has not been done (single mode).
4. X-ray centering failed  
As a result of vertical scan (performed after raster scan), there was no diffraction exceeding the threshold value; thus data collection has not been done.
5. Loop centering failed (INOCC)  
Loop centering failed for some reasons; thus data collection has not been done.

"meas time [min]", measurement time for each pin

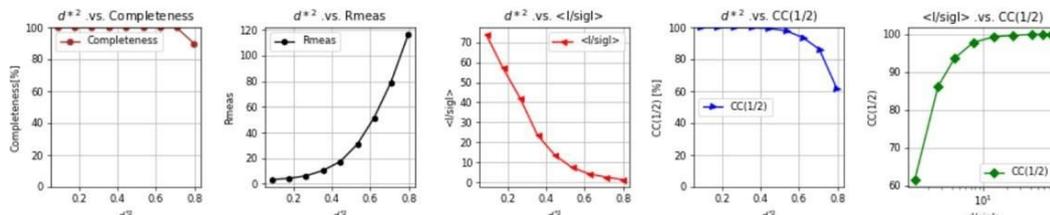
## 4.2. correct.html, summary of data statistics of helical/single modes \_kamoproc/correct.html

CPS1012-06/data00/CPS1012-06-single\_001\_1-3600/CORRECT.LP

Space group(XDS) = P41212, cell: 78.550 78.550 37.110 90.00 90.00 90.00

ISa(XDS) = 23.7

RESOLUTION LIMIT	NUMBER OF OBSERVED REFLECTIONS	NUMBER OF UNIQUE REFLECTIONS	POSSIBLE	COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR expected	COMPARED I/SIGMA	R-meas	CC(1/2)	Anomal Corr	SigAno	Nano
3.35	42144	3179	3180	100.0%	3.3%	3.4%	42144 73.85	3.4%	99.9*	65*	1.374	1319
2.38	74296	5763	5763	100.0%	4.3%	3.9%	74296 57.10	4.5%	99.9*	43*	1.280	2607
1.94	100920	7445	7445	100.0%	6.2%	5.3%	100920 41.99	6.4%	99.9*	33*	1.197	3445
1.68	116965	8818	8818	100.0%	10.4%	9.6%	116965 23.77	10.8%	99.7*	4	0.893	4133
1.50	136570	10008	10008	100.0%	17.4%	17.9%	136570 13.73	18.1%	99.3*	-2	0.779	4726
1.37	145581	11034	11034	100.0%	30.8%	33.5%	145581 7.54	32.0%	97.8*	-3	0.727	5243
1.27	157244	12013	12013	100.0%	51.5%	57.2%	157244 4.37	53.6%	93.6*	-3	0.706	5727
1.19	169027	12907	12907	100.0%	78.6%	91.6%	169027 2.70	81.8%	86.2*	-4	0.667	6176
1.12	110458	12359	13741	89.9%	116.3%	143.0%	110238 1.37	123.4%	61.5*	-2	0.621	5814
total	1053205	83526	84909	98.4%	6.8%	6.9%	1052985 16.89	7.1%	100.0*	7	0.822	39190



CPS1012-07/data00/CPS1012-07-single\_001\_1-3600/CORRECT.LP

Space group(XDS) = P41212, cell: 78.000 78.000 37.050 90.00 90.00 90.00

ISa(XDS) = 20.2

... continue...

The table and graphs are extracted from CORRECT.LP of XDS output.

This report is not applied to the data collected in multi mode; please refer to report.html directly under \_kamoproc/.

## 4.3. report.html; summary of data statistics of multi mode or helical mode "buttಾಗಿರಿ"

\_kamoproc/merge\_blend(or cc)\_"Start Resolution"S\_"Sample name"/\*/report.html

Cell-based clustering by BLEND

See original cluster dendrogram

Merging summary

cluster	CIH	LCV	aLCV	#DS all	#DS used	Overall				Outer shell				Inner shell				Wilson	Aniso best	resol wrst	d_min est.						
						Cmpl	Redun	I/σ(I)	R <sub>meas</sub>	CC <sub>1/2</sub>	Cmpl	Redun	I/σ(I)	R <sub>meas</sub>	CC <sub>1/2</sub>	Cmpl	Redun					I/σ(I)	R <sub>meas</sub>	CC <sub>1/2</sub>	SigAno	CC <sub>ano</sub>	
cluster 0018	13.46	1.0	0.4	19	19	97.3	4.8	6.79	12.2	99.6	95.5	4.9	0.83	143.4	48.7	99.3	4.9	20.88	7.5	99.5	1.4	52.0	13.19	1.44	1.63	1.37	
cluster 0017	4.25	0.7	0.3	14	13	96.4	3.4	7.82	9.2	99.7	94.7	3.4	0.91	143.0	41.7	98.4	3.4	24.75	5.6	99.7	1.6	53.0	13.57	1.45	1.64	1.39	
cluster 0015	3.51																										
cluster 0013	2.84																										
cluster 0011	2.22																										
cluster 0010	2.02																										
cluster 0007	1.23																										
cluster 0008	1.36																										
total						136062	28134	28912		97.3%	11.0%	13.5%	135600	6.79	12.2%	99.6*	18*							0.712	13389		

SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE >= -3.0 AS FUNCTION OF RESOLUTION

RESOLUTION LIMIT	NUMBER OF OBSERVED REFLECTIONS	NUMBER OF UNIQUE REFLECTIONS	POSSIBLE	COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR expected	COMPARED I/SIGMA	R-meas	CC(1/2)	Anomal Corr	SigAno	Nano
4.14	5201	1068	1076	99.3%	6.8%	7.6%	5164 20.88	7.5%	99.5*	52*	1.444	479
2.93	9504	1941	1951	99.5%	7.8%	7.8%	9479 18.78	8.7%	99.5*	35*	1.218	913
2.39	12496	2516	2540	99.1%	9.1%	9.0%	12465 15.46	10.1%	99.1*	30*	1.042	1196
2.07	13993	2948	2984	98.8%	11.0%	11.5%	13949 11.28	12.3%	99.3*	10	0.876	1401
1.85	16818	3377	3470	97.3%	15.6%	19.8%	16771 7.31	17.4%	98.7*	8	0.690	1618
1.69	17008	3645	3733	97.6%	26.2%	38.8%	16942 4.04	29.5%	95.3*	4	0.596	1736
1.56	20458	4157	4289	96.9%	42.9%	70.4%	20394 2.54	48.0%	89.2*	1	0.539	1991
1.46	19552	4203	4390	95.7%	66.1%	114.3%	19471 1.55	74.4%	84.4*	-1	0.512	2000
1.38	21032	4279	4479	95.5%	128.1%	233.7%	20965 0.83	143.4%	48.7*	-2	0.488	2055
total	136062	28134	28912	97.3%	11.0%	13.5%	135600 6.79	12.2%	99.6*	18*	0.712	13389

The cluster numbers (□,□,□) in the list correspond to the numbers in the dendrogram. Roll over the cluster number in the dendrogram to display the data-storage directory. Click the ▼ mark at the left end to display a list of statistical values. The resolution is cut off so that CC(1/2) is about 50% or more.

The resulting bulletins include only the top-level clusters. For other clusters, please check after the HDD arrives.

## 5. Clustering and merging

The clustering and merging process is executed based on the "Sample Name" designated in the automatic measurement sample sheet.

The data are clustered by BLEND (lattice parameter) and CC (diffraction intensity correlation). The directory in which the clustering and merging process was executed is created using the following rules. The process is executed from half corner resolution. The threshold for mergeable data is set as a minimum redundancy of 2.0 and minimum completeness of 10%

- Under `_kamoproc/` or `_kamo_30deg/`
- Directory name
  - `merge_”clustering method based on”_”Starting resolution”S_”SampleName”`
    - e.g. `merge_blend_2.5S_sample1`
  - Starting resolution is defined as “half corner resolution” at a particular wavelength, camera distance, and detector.
  - BLEND based on the hierarchical clustering analysis by unit cell.
  - CC based on the correlation coefficients of diffraction data.
  - SampleName
    - Designated by users as “SampleName” in a sample sheet
    - Data that have the same sample name are merged.

When automerger is started, the data are merged with the reflection up to the starting resolution. After a round of the automerger process, KAMO `auto_multi_merge` estimates the resolution limit and attempts the automerger process until the CC(1/2) of the highest resolution bin reaches around 50%. You can find the `blend/cc_’RESO’A_final` directory under the merge directory.

KAMO `auto_multi_merge` executes hierarchical clustering based on unit cell or pairwise CC.

All data sets are scaled and merged with XSCALE (`run_01`)

Bad frames are removed (`run_02`)

Bad data sets are removed (`run_03`)

These directories are contained under the `blend/cc_’RESO’A` and `blend/cc_’RESO’A_final` directories.

Please refer to the official KAMO documentation.

<https://github.com/keitaroyam/yamtbx/blob/master/doc/kamo-en.md>

### Reviewing the results of auto merge

- **“report.html” under the merge directory is a summary of auto merge.**
  - reflection file can be found in `run_*` directory as `xscale.hkl`.
  - mtz file is also available in the `run_*/ccp4/` directory as `xscale.mtz`.
- All processes through indexing to space group determination, resolution determination, and hierarchical clustering are automated. The results should be reviewed by users. We do not take responsibility for any inconvenience
- We recommend that users should review the content of `“report.html”` and assess the `XSCALE.INP` and `XSCALE.LP` of particular clusters for further analysis.

## e.g.) Results of auto\_multi\_merge with BLEND

\_kamoproc/

```
└─ merge_blend_2.2S_complex1 <- working director for "complex1"
  └─ blend_3.03A_final <- containing final data
    └─ blend <- results of BLEND clustering
      └─ BLEND0.RData
      └─ BLEND_KEYWORDS.dat
      └─ CLUSTERS.txt
      └─ FINAL_list_of_files.dat
      └─ blend0.log
      └─ blend_cluster_summary.dat
      └─ dendro.json
      └─ forR_macropar.dat.bak
      └─ hctojson.R
      └─ tree.png <- dendrogram of BLEND clustering
      └─ tree.ps
      └─ xds_lookup_table.txt

    └─ cluster_0079 <- scaling results of a cluster (0079 = clusterID)
      └─ args.pkl
      └─ files <- containing reflection files (with link)
        └─ CPS0836-07
          └─ ....
          └─ ....
          └─ ....

      └─ merge.log <- log for merge
      └─ merge_cluster_0079.sh <- script for calculation
      └─ merge_cluster_0079.sh.o1468055 <- output from computer cluster
      └─ merge_cluster_0079.sh.po1468055 <- output from computer cluster
      └─ result.pkl <- pickled result
      └─ run_0* <- (refer to section 5 for run_01, run_02, run_0)
        └─ XSCALE.INP
        └─ XSCALE.INP.1
        └─ XSCALE.LP
        └─ XSCALE.LP.1
        └─ aniso.log <- anisotropic resolution analysis
        └─ cc_files.dat
        └─ cc_frames.dat
        └─ ccp4
          └─ ctruncate.log
          └─ logfile.log
          └─ xds2mtz.log
            └─ xscale.mtz <- reflection file for structural analysis

        └─ corfac_absorp.dat
        └─ corfac_decay.dat
        └─ corfac_modpix.dat
        └─ pointless.log
        └─ xscale.hkl <- reflection file from XSCALE

    └─ js
      └─ LICENSE
      └─ d3.js
      └─ d3.min.js
      └─ index.js

    └─ cluster_summary.dat
    └─ multi_merge.log merge log
    └─ report.html <- report of cluster_0079 merging process
```

```
└ cells.dat <- list of cell parameters
└ excluded.lst <- list of datasets excluded by clustering
└ formerge.lst <- list of data used to merge
└ input_files <- reflection files used to merge (with link)
  │
  │ └ CPS0836-07
  │ └ CPS0836-08
  │ └ ...
  │
└ kwards.pkl
└ multi_merge.log <- log of merge
└ multimerge.sh <- shell script for merge (for staff)
└ multimerge.sh.01468049
└ prep_merge.log <- log for preparation of merge
```