

# データ処理結果についてのお知らせ

2020/7/14

SPring-8 PX-BLにおいて自動測定で得られたデータは、プログラムKAMO (Yamashita et al, 2018)によりXDSで自動処理された結果、クラスタリングとマージ処理の結果も併せて送付しています。

## 1. 送付したデータの構造

HDDに保管されたデータは以下の階層で保存されています。

```
"Date_BL45XU_username"  
├ "PuckID"-“PinID” : 回折データ  
│   ├── ##.ppm, raster.jpg : センタリング画像  
│   ├── scan** : スキャンデータ (含むSHIKAの結果)  
│   └── data** : 測定した回折データ (helical, single, multi)  
│  
├ “_kamoproc” : 標準での処理結果  
│   ├── “PuckID”-“PinID” : XDS処理結果  
│   ├── report.html, correct.html (contents) : XDS処理結果のレポート  
│   └── “merge_**_**_SampleName” : マージ処理結果  
│  
├ “_kamo_30deg” : 30°ぶった切り処理結果  
│   ├── “PuckID”-“PinID” : XDS処理結果  
│   ├── report.html, contents : XDS処理結果のレポート  
│   └── “merge_**_**_SampleName” : マージ処理結果  
│  
├ “Date_BL45XU_username”_00.html : ZOOのレポートファイル  
├ “Date_BL45XU_username”.tgz : レポートファイル  
└ *csv, *db, *xlsxなど : 自動測定で使用した設定ファイル
```

## 2. 回折データ

測定データは、自動測定サンプルシートの“PuckID”-“PinID”名で保存されています。

```
├ "PuckID"-“PinID” : 回折データ  
│   ├── ##.ppm, raster.jpg : センタリング画像  
│   ├── scan** : スキャンデータ (含むSHIKAの結果)  
│   └── data** : 測定した回折データ (helical, single, multi)
```

### 2.1. センタリング画像

センタリング画像は、測定時に選択したオプションによって構成が変わります。ご注意ください。

1. before\_warmup.ppm : warmup前の画像
2. after\_warmup.ppm : 設定したwarmup時間後の画像
3. raster.jpg : センタリング後の2Dスキャン範囲を保存  
(“Hor. scan length [μm]”の設定が適切であったか確認することができます)
4. before.ppm : raster.jpgのスキャン範囲なしの画像
5. loop\_zoom.ppm : センタリング後の拡大画像
6. loop\_zoom\_In2.ppm : 霜除去後の拡大画像 (“LN2 Splash”選択の場合のみ)

## 2.2. データ収集ディレクトリの構成

以下はデータ収集ディレクトリの構造を示したもの（ディレクトリツリー）です。バックアップのディレクトリの直下にピンごとのディレクトリがあると考えてください。EIGERのHdf5フォーマット.h5の場合は、100枚ごとに\*\*\*\_000001.h5のようなデータファイルが作成されます。PILATUSのcbfフォーマット.cbfは1枚ごとにファイルが出力されます。今回の説明では、データセットは0.1° oscillation の測定としています。

Multi, Helical, Singleのそれぞれのmodeで測定した構成を説明します。

- MUL0001-01 : Multi modeの回折データ (2.2.1.参照)
- HEL0002-002 : Helical modeの回折データ (2.2.2.参照)
- SIN003-003 : Single mode回折データ (2.2.3.参照)

### 2.2.1. Multi modeの場合

Multi modeでのデータ収集を行ったデータのディレクトリ構造を示します。

```
└ MUL0001-01 : 試料ディレクトリ
├ └ ##.ppm, raster.jpg : センタリング画像 (2.1.参照)
├ └ scan** : スキャンデータを格納 (含むSHIKAの結果)
│   └ 2d : スキャンタイプ (2d = 二次元)
│       └ spotfinder : SHIKA解析結果が格納
│           └ thumb_2d : SHIKA回折像のサムネイルが格納
├ └ data00 : 測定した回折データを格納 (data01だと 2 回めの測定)
│   └ MUL0001-01-multi_###_*****.cbf / h5 : 回折データ
└ (###:"# of crystals"で設定した個数内で回折点検出数の閾値を超えた結晶分を測定)
```

### 2.2.2. helical modeの場合

ヘリカルデータ収集を行ったデータのディレクトリ構造を示します。

```
└ HEL0002-002 : 試料ディレクトリ
├ └ ##.ppm, raster.jpg : センタリング画像 (2.1.参照)
├ └ scan** : スキャンデータを格納 (含むSHIKAの結果)
│   └ 2d : スキャンタイプ (2d = 二次元)
│       └ spotfinder : SHIKA解析結果が格納
│           └ thumb_2d : SHIKA回折像のサムネイルが格納
│   └ lv00_00 : ラスター種類 (lv = 結晶の左端縦スキャン)
│       └ spotfinder : SHIKA解析結果が格納
│           └ thumb_2d : SHIKA回折像のサムネイルが格納
│   └ rv00_00 : ラスター種類 (rv = 結晶の右端縦スキャン)
│       └ spotfinder : SHIKA解析結果が格納
│           └ thumb_2d : SHIKA回折像のサムネイルが格納
│   └ rv00_01 : ラスターやりなおし (index部分が00 -> 01へ)
│       └ spotfinder : SHIKA解析結果が格納
│           └ thumb_2d : SHIKA回折像のサムネイルが格納
├ └ data00 : 測定した回折データを格納 (data01だと 2 回めの測定)
│   └ cry###_*****.cbf / h5 : 回折データ
└ (###:"# of crystals"で設定した個数内で回折点検出数の閾値を超えた結晶分を測定)
```

### 2.2.3. single modeの場合

シングルデータ収集を行ったデータのディレクトリ構造を示します。

```
└ SIN003-003 : 試料ディレクトリ
|   └ ##.ppm, raster.jpg : センタリング画像 (2.1.参照)
|   └ scan** : スキャンデータを格納 (含むSHIKAの結果)
|     └ 2d : スキャンタイプ (2d = 二次元)
|       └ spotfinder : SHIKA解析結果が格納
|         └ thumb_2d : SHIKA回折像のサムネイルが格納
|           └ lv00_00 : ラスター種類 (lv = 結晶の左端縦スキャン)
|             └ spotfinder : SHIKA解析結果が格納
|               └ thumb_2d : SHIKA回折像のサムネイルが格納
|   └ data00 : 測定した回折データを格納 (data01だと2回目の測定)
|     └ single_001_*****.cbf / h5 : 回折データ
|
```

以下、注意点

- lv / rv01\_00などがある  
→結晶を二個以上見つけた場合のそれぞれについてスキャンを行っている
  - lv / rv00\_01などがある (helical modeの場合)  
→端のセンタリングに失敗し、やり直している。  
→ラスタースキャンから90°回転した方向の回折能が弱い (異方性がある) もしくは、何らかの理由により結晶からビームが外れている。
- 端の縦スキャンで見つからなければ、結晶体積が大きいと思われる側に10 μm併進し再度縦スキャンを行う。最大5回繰り返す。結晶が見つからない場合にはギブアップし、データ測定を行わない  
(2020/04/01現在)

### 3. XDSの処理結果

測定したデータ (helical, single, multi) を標準設定で処理した結果は“\_kamoproc”に保存されています。処理内容については XDS.INP や CORRECT.LP で確認してください。

```
└ _kamoproc : KAMOでの標準設定での処理ディレクトリ
|   └ MUL0001-01 : Multi mode の処理結果
|       └ data00 :
|           └ MUL0001-01-multi_001_1-100 : XDS処理結果が入っている
|           └ MUL0001-01-multi_002_1-100
|           └ MUL0001-01-multi_003_1-100
|           └ MUL0001-01-multi_0##_1-100
|       |
|       └ HEL0002-002 : Helical mode の処理結果
|           └ data00 :
|               └ cry00_1-3600 : XDS処理結果が入っている
|                   └ ccp4 : mtz化されたファイルが入っている
|           |
|           └ SIN003-003 : Single mode の処理結果
|               └ data00 :
|                   └ single_1-3600 : XDS処理結果が入っている
|                       └ ccp4 : mtz化されたファイルが入っている
|           |
|           └ report.html : 測定条件、処理の結果レポート
|       |
|       └ correct.html : helical, single modeデータの処理の結果レポート
|   |
```

helicalデータ収集では、不均一な結晶のデータから有効なデータを得るために、30°毎にデータを分断した処理（通称“ぶった切り処理”）およびそれらのマージも自動で行い結果をお返ししています。その結果は“\_kamo\_30deg” に保存されています。ぶった切り処理は、全てのデータを対象に処理しています。これらの処理結果は、自動測定サンプルシートの“PuckID”-“PinID”名で保存されています。

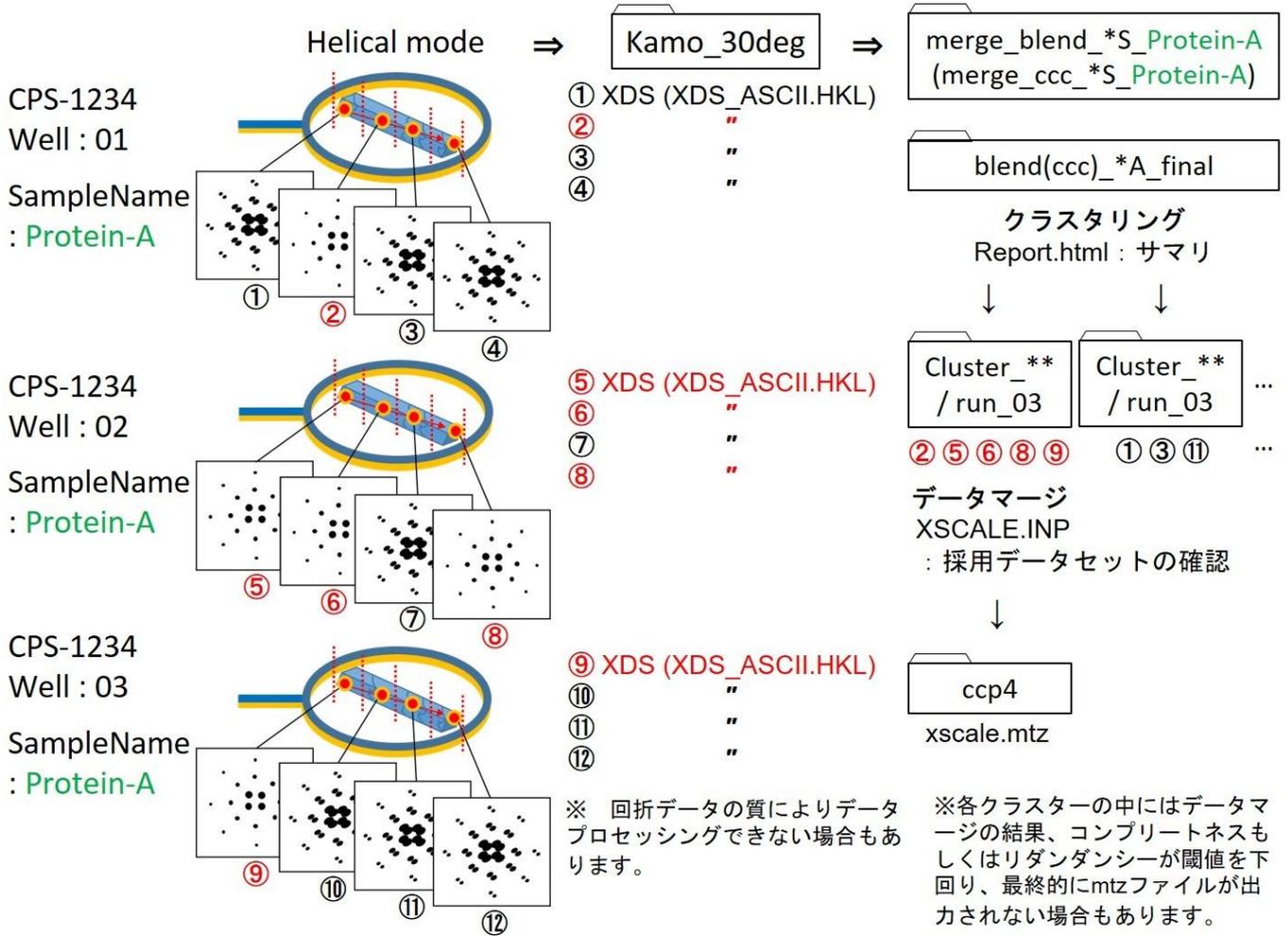
```
└ _kamo_30deg : 30°毎に処理したディレクトリ
|   └ MUL0001-01 : (multi modeの10°のデータは _kamoprocと同じ)
|       └ data00 :
|           └ MUL0001-01-multi_001_1-100 : XDS処理結果が入っている
|           └ MUL0001-01-multi_002_1-100
|           └ MUL0001-01-multi_003_1-100
|           └ MUL0001-01-multi_0##_1-100
|       |
|       └ HEL0002-002 : Helical mode の処理結果
|           └ data00 :
|               └ cry00_1-300 : 0-30°分の処理結果
|               └ cry00_301-600 : 30-60°分の処理結果
|               └ . . . 以下、最後まで処理される
|               └ cry00_3301-3600 : 330-360°分の処理結果
|           |
|           |
```

```

|   └ SIN003-003 : Single mode の処理結果
|   |   └ data00 :
|   |       └ single_1-300 : 0-30°分の処理結果
|   |       └ single_301-600 : 30-60°分の処理結果
|   |       └ . . . 以下、最後まで処理される
|   |       └ single_3301-3600 : 330-360°分の処理結果
|   |
|   └ report.html : 30°で分割した測定条件、処理の結果レポート
|

```

”ぶった切り処理”のイメージ



#### 4. レポートファイルXDSの処理結果

ルートディレクトリに保存されている"Date\_BL45XU\_username".tgzには、以下のレポートファイルが格納されています。

"Date\_BL45XU\_username"

└"Date\_BL45XU\_username"\_00.html : ZOOのレポートファイル

| (複数個含まれる場合があります)

|

└"\_kamoproc" : 標準での処理結果

| └report.html : 測定条件、処理の結果レポート

| └merge\_\*\*\* : マージ処理結果

| | └/\_kamoproc/merge\_"クラスタリングの種類"\_ "スタートの分解能"

| | S\_"SampleName"/blend\_\*/report.html

| |

| └correct.html : helical, single modeデータの処理の結果レポート

|

└"\_kamo\_30deg" : 30°ぶった切り処理結果

└merge\_\*\*\* : マージ処理結果

└/\_kamo\_30deg/merge\_"クラスタリングの種類"\_

"スタートの分解能"\_ "SampleName"/blend\_\*/report.html



## 4.2. <correct.html : helical, single modeデータの処理の結果レポートの見方>

### XDS Report

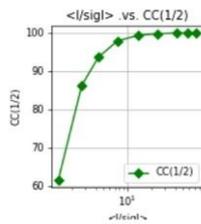
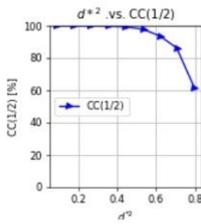
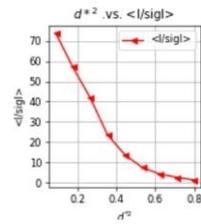
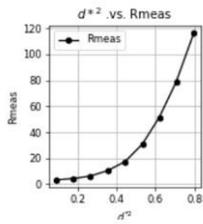
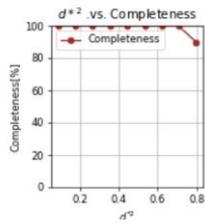
\_kamoproc/correct.html

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

SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE $\geq -3.0$ AS FUNCTION OF RESOLUTION													
RESOLUTION LIMIT	NUMBER OF REFLECTIONS OBSERVED	NUMBER OF REFLECTIONS UNIQUE	SIGNAL/NOISE 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

•  
•  
•

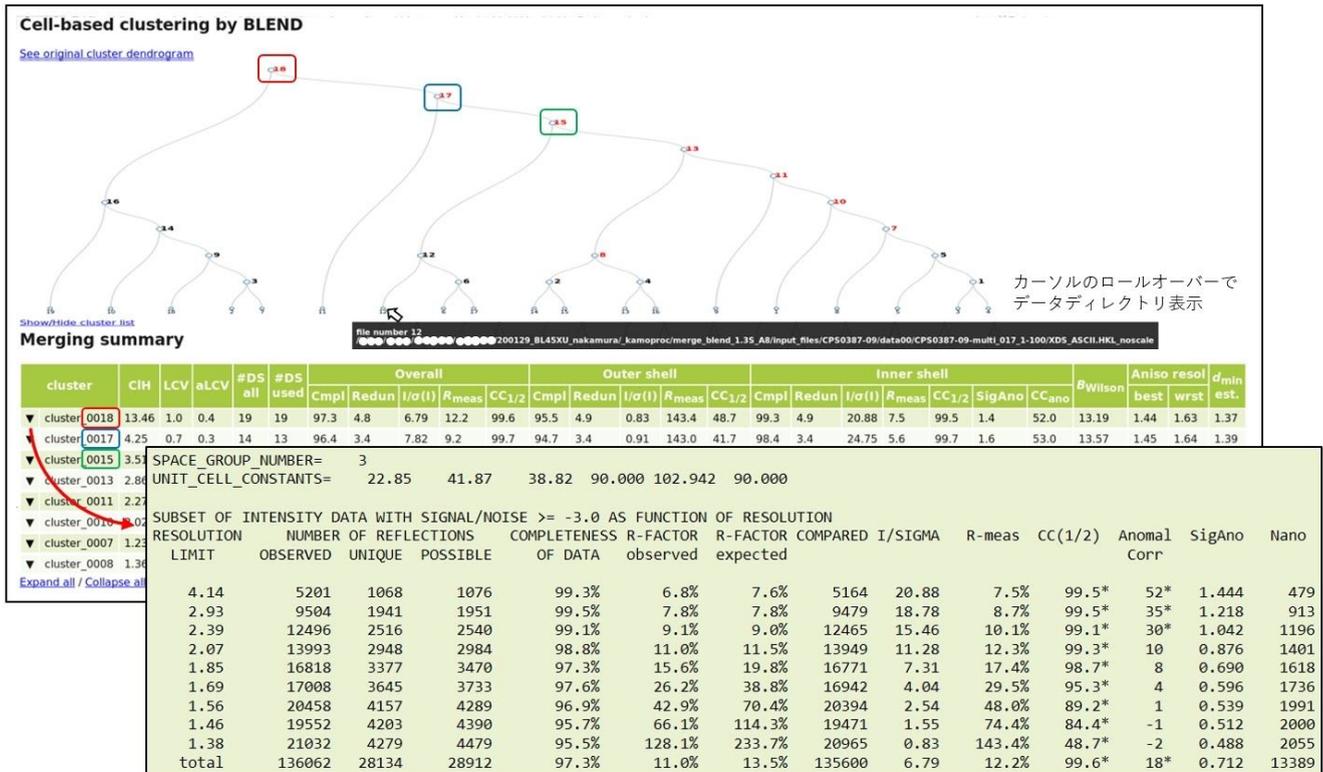
データセットごとのXDS処理結果の統計値 (COLLECT.LP より) を一覧することができます。

※multi 測定については記載されません。mult測定個々のデータの統計値はKAMO report (\_kamoproc 直下の report.html) をご参照ください。

### 4.3. <report.html : マージデータ処理結果のレポートの見方>

#### Merge Report

\_kamoproc/merge\_blend(or cc)\_"Start Resolution"S\_"Sample name"\*/report.html  
 各マージディレクトリ毎にreport.html が格納されています。



cluster一覧の□,□,□内cluster番号は、デンドログラム中の□,□,□内の番号とそれぞれ対応しています。デンドログラム中のcluster番号にカーソルをロールオーバーするとデータの保存先ディレクトリが表示されます。左端の▼印をクリックすると統計値の一覧が表示されます。CC(1/2)が、およそ50%以上になるように分解能がCut offされていることが分かります。

なお、速報として送付されるデータには最上位のクラスターのrun\_03のみが含まれます。その他のクラスターについてはHDDが到着後に、ご確認ください。

## 5. クラスタリングとマージ処理の実行

クラスタリングとマージ処理は、自動測定サンプルシートの"SampleName"で指定した名前毎でデータをクラスタリング[Blend（格子定数）とCC（回折強度の相関）の2種類]します。クラスタリングとマージ処理を実行したディレクトリは、以下のルールで作成され、設定した分解能から処理を行います。マージ可能なデータのしきい値は最小リダンダンシー：2.0、最低コンプリートネス：10%です。

### 詳細

クラスタリングの計算の際、開始分解能は初期設定分解能（ユーザ指定）をエッジとしたカメラ長に基づいて、いわゆるhalf-corner(ハーフコーナー)の分解能を設定します。このため著しく設定分解能よりも低い回折データの場合、有効な反射数も著しく減り、結果として完全度が低くなり、閾値を超えずにマージがされないことがあります。運用上のポリシーとして「マージされた結果を何らかお返しする」ため閾値を非常に低い値に設定し、これを回避しています。KAMOは分解能をCC(1/2)に基づいて決定するシーケンスを自動で行うため単に分解能が低い場合には、低い分解能で完全なデータとしてマージが成功するケースもあります。

一方で、本当にコンプリートしていないデータもマージの結果として残る可能性がありますので、必ずマージの結果についてはコンプリートネスを確認してください。

- \_kamoproc/もしくは\_kamo\_30deg/ 以下に
- ディレクトリ名
  - merge\_ "**クラスタリング種類**" "**スタート分解能**" S\_ "**SampleName**"
    - 例：merge\_blend\_2.5S\_sample1
  - **スタート分解能の設定**
    - half-cornerの分解能から処理を開始
  - **クラスタリング種類**
    - blend
      - 格子定数を利用したクラスタリングの結果
    - cc
      - 回折強度の相関を利用したクラスタリングの結果
  - **SampleName**
    - 測定シートに記載した"SampleName"
    - 同じSampleNameが複数ある場合、は一つのデータにマージされる
- ディレクトリの内容（後述ディレクトリツリー参照）
  - BLENDとccで内容やディレクトリ名が若干異なりますので適宜読み替えてください。
  - 最終データは\_finalにあると考えてください
    - 自動マージでは既定のリダンダンシーとコンプリートネスを満たすために分解能は自動調整される
    - **自動マージは分解能の自動決定をCC(1/2)~50%で行う**
    - スタート分解能は実際の分解能（この段階では不明）よりも高めに入力しておく必要がある→half-cornerとしてある
    - さらに下の階層
      - run\_01, run\_02, run\_03についてはデータの研磨を行った結果
      - 詳しくはKAMOの公式ページを御覧ください

- <https://github.com/keitaroyam/yamtbx/blob/master/doc/kamo-ja.md>
- **マージの結果閲覧**
  - マージディレクトリ（例としてblend\_3.5A\_final）
  - **直下にある”report.html”がマージ全体のサマリとなっている**
    - htmlをウェブブラウザで閲覧して最終データを決定
    - すでに同梱されているMTZファイルを利用して解析を開始します
      - 例) 0634のクラスタのrun\_03を選定した場合
        - [blend\\_3.5A\\_final/cluster\\_0634/run\\_03/ccp4/](#) へ
      - 例) 0321のクラスタのrun\_02を選定した場合
        - [blend\\_3.5A\\_final/cluster\\_0634/run\\_02/ccp4/](#) へ
  - 上記はすべて自動マージを行った第一弾の処理結果であると認識し、空間群の選定や同型な結晶のクラスタリング、データの棄却、分解能決定についても最善であるとは考えていません。
  - **report.htmlの内容、好みのクラスタの“XSCALE.INP”、“XSCALE.LP”を参照し、最終データ処理については各自で行うことを推奨しています。**

<BLENDによるクラスタリングとマージの結果：ファイル情報>  
CCによるクラスタリングも同様の構造になっていますので適切に読み替えてください。

```
└ merge_blend_2.2S_complex1 (BLENDクラスタリングの一つのディレクトリ)
  │ │ └ cluster_summary.dat サマリ
  │ │ └ js
  │ │ │ └ LICENSE
  │ │ │ └ d3.js
  │ │ │ └ d3.min.js
  │ │ │ └ index.js
  │ │ └ multi_merge.log
  │ │ └ report.html このデータマージに関するレポートHTML
  │ └ blend_3.03A_final 最終的な結果が格納されている (_final)
  │ │ └ blend BLENDクラスタリングの結果ファイル
  │ │ │ └ 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 BLENDクラスタリングの結果：デンドログラム
  │ │ │ └ tree.ps
  │ │ │ └ xds_lookup_table.txt
  │ │ └ cluster_0079 あるクラスタのスケールの結果(0079 = clusterID)
  │ │ │ └ args.pkl
  │ │ │ └ files 反射ファイルが格納 (リンクもある)
  │ │ │ │ └ CPS0836-07
  │ │ │ │ │ └ ...
  │ │ │ │ │ └ ...
  │ │ │ │ │ └ ...
  │ │ │ │ └ merge.log マージに関するログ
  │ │ │ │ └ merge_cluster_0079.sh 計算を行ったスクリプト
  │ │ │ │ └ -merge_cluster_0079.sh.o1468055 計算クラスタ上での出力
  │ │ │ │ └ merge_cluster_0079.sh.po1468055 計算クラスタ上での出力
  │ │ │ │ └ result.pkl 結果をピクルしたもの
  │ │ │ │ └ run_01 (run_01, run_02, run_03についてはマニュアル参照)
  │ │ │ │ │ └ XSCALE.INP
  │ │ │ │ │ └ XSCALE.INP.1
  │ │ │ │ │ └ XSCALE.LP
  │ │ │ │ │ └ XSCALE.LP.1
  │ │ │ │ │ └ aniso.log 異方性結晶の分解能評価 (AIMLESSと同様)
  │ │ │ │ └ cc_files.dat
```

