Current status of automatic data collection by ZOO at SPring-8

Kunio Hirata
(RIKEN/SPring-8 Center)
Beamlime BL32XU at SPring-8

SPring-8

BL45XU
Fully automated beamline

Micro-focus beamline BL32XU

Loop centering
Define scan area

Raster scan
(50-100 Hz)

ZOO system

Find crystals

Crystal positions

Dose estimation

Automatic data processing/merging
How to use ZOO system

Just send us UniPucks and data collection sheet.
ZOO solved many structures with merging

---

**Orexin 2 receptor**
2.0 Å

R. Suno et al.

**Angiotensin II receptor**
3.2 Å

H. Asada et al.

**M2 muscarinic receptor**
2.3 Å

R. Suno et al.

**Human ETB receptor**
2.0 Å

W. Shihoya et al.

**5-HT2A Receptor**
2.7 Å

K. Kimura et al.

**Prostaglandin E receptor (EP4)**
3.2 Å

Y. Toyoda et al.

**Prostaglandin E subtype EP3**
3.2 Å

K. Morimoto et al.

**β2AR regulation**
3.2 Å

X. Liu et al.

**Melanocortin-4 receptor**
2.8 Å

J. Yu et al.

---

"Quantitative changes" make "Qualitative change"
Merits

• **Un-attended measurements**
  • Users can sleep
  • Users can do other experiments

• **Fast data collection**
  • ZOO shortens data collection time
  • No human errors

• **No prior knowledges/expertise**
  • ZOO enabled ‘technically difficult experiments’
  • under dose control

• **Automatic filter enhancing data quality**
  • KUMA : controls dose
  • SHIKA : rejects bad crystals via the raster scan
  • KAMO : selects better frames/datasets => merge
ZOO system


**SHIKA**
- fast spot finder
- crystal detection

**KUMA**
- dose estimation
- set exp. conditions

**HEBI**
- prepares ‘helical’ scheme

**HITO**
- prepares ‘mixed’ scheme

**INOC**
- loop centering
- raster area definition

**BSS**
- high-level device controller

**ZOO navigator**
- sequence manager
- make/sends commands

**Microscope**

**Goniometer**

**SPACE**

**Detector**

**Data storage Detector server**

*Program package with ‘animal names’ in Japanese*
Current status of user operation at BL45XU

Auto (remote)  Auto (attended)  Manual  Industrial use

Shutdown

Shutdown

Shutdown

Shutdown
Specifications of MX ID beamlines at SPring-8

- Beam size: 5~30 µm
  (1 µm beam is available for SSROX@BL32XU)

- Photon flux: $10^{12} \sim 10^{13}$ phs./s.

- Wavelength: 0.8 ~ 1.9 Å
  (0.35~0.65 Å available at BL41XU)

ZOO is available at MX ID beamlines at SPring-8
BL32XU/BL41XU/BL45XU.
Available schemes in ZOO

<table>
<thead>
<tr>
<th>Scheme</th>
<th>Rot. (x)tal</th>
<th>10–300 crystals</th>
<th>single or a few crystals</th>
<th>small &amp; large crystals</th>
<th>300–10,000 crystals</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Small wedge&quot; (SWSX)</td>
<td>5–10 deg.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;helical&quot; or &quot;single&quot;</td>
<td>90–360 deg.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;mixed&quot;</td>
<td>5–10 deg.</td>
<td></td>
<td></td>
<td>30–360 deg.</td>
<td></td>
</tr>
<tr>
<td>&quot;SSROX&quot;</td>
<td>&lt; 1 deg.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Users should choose ‘scheme’ for each cryo-pin.
How ZOO works
(automates helical scheme)

HEBI eliminates ‘expertise’ for 3D centering of helical vector.
Keitaro Yamashita

**KAMO**
(Automatic data processing)

- Index/Integrate each dataset
- Grouping datasets (cells)
- Determination of space group
- Clustering (cell/intensity CC)
- Selecting good cluster (Completeness, Redundancy)
- Merging datasets
- Outlier detection/rejection

**Data processing**
XDS (or DIALS)

**Grouping**
crystals with equivalent cells

- **Clustering**
  - BLEND (cell parameters)
  - CC of intensities


Keitaro Yamashita
Data back to users

Small-wedge

5-10º/crystal

Full data/crystal

Ex: 180º

BUTTAGIRI

Partial data

Clustering

Full data

Full datasets

Partial data

Clustering

Full datasets
Data back to users

Small-wedge

5-10º/crystal

Partial data

Clustering

Full datasets
Data back to users

Full data/crystal

Ex: 180°

Full dataset

H, K, L, l, sig(l)

........................

........................

........................
BUTTAGIRI process

divides ‘complete data’ to ‘small wedge data’
(Ex. : Full data $180^\circ = 30^\circ \times 6$ wedges)

→Hierarchical clustering is applied to them
(SELECT better data in ‘merging process’)
Merits in ‘BUTTAGIRI’ process

• extracts/merges better crystal volumes

• is very powerful when it works with ‘helical data collection’
  – especially for ‘inhomogenous’ crystals
Data collection from ‘inhomogeneous crystal’

Is a human superior to ‘automatic data collection’, especially for ‘inhomogenous crystal’?

This case is one of the answers to this question.
Data collection from ‘inhomogeneous crystal’

1. Helical data collection x 11 crystals

2. 720 deg/crystal (0.1deg oscillation)

3. Data processing
   - KAMO: 10° BUTTAGIRI processing
   - \( \frac{720°}{10°} = 72 \) wedges/crystal
   - \( 72 \times 11 \) crystals = 792 datasets
   - KAMO: Hierarchical clustering
   - Merge \( \rightarrow 2.8 \) Å resolution

Succeeded data collection from ‘inhomogeneous’ crystals

The best way: Helical data collection with high redundancy
- Data collection from ‘entire crystal volume’
- 360-720 deg/crystal
- Data collection in different orientation/volume from each crystal volume

BUTTAGIRI process
- To select better volumes in data processing

Does ‘redundant data collection’ deteriorate resolution limit?
- Redundant measurements can recover accumulated signals.
- (25% x 4 times → 100%) in the ideal case

Multiple crystal strategy is effective for data collection if ‘better’ volumes in a crystal are small
- Only 10%/crystal is good → Merging 10 crystals

The protocol would be superior to the manual measurement.
From clustered ‘micro’ crystals

Are they crystals??

40 x 20 µm² beam

Low resolution powder-like diffraction

~8 Å

Consultation:
Could we collect ‘full dataset’ from them?

Problem solving by ZOO : case2

Courtesy of Dr. Numoto
Problem solving by ZOO : case2

From clustered ‘micro’ crystals

Data collection with ZOO

5 µm beam (scan & data collection)

• Oscillation data : 5 deg/point
• ~1,400 wedges from 16 loops
• 670 datasets: equivalent cell
• Merging gave 2.5 Å dataset
• Measurement time: 3 hours

Ex) Crystal 1

Ex) Crystal 2

Scan results

Structural analysis from a crystal that doesn't look good at first glance!

Courtesy of Dr. Numoto
Future perspectives

• **Towards  “Fully” automatic**
  – Minimal prior information from users

• **Feedback**
  – Results of data processing
    • to direct residual measurements

• **Fully automated exchanging pucks**
  – “Endless” automatic data collection
HITO (mixed mode)

Automatically selects ‘data collection scheme’.

合計 1,270 deg / 1mm loop 25 min. → 1.3 Å resolution
(c.f. small wedge 1,200 deg./24 min.)
Future perspectives

• Towards “Fully” automatic
  – Minimal prior information from users

• Feedback
  – Results of data processing
    • to direct residual measurements

• Fully automated exchanging pucks
  – “Endless” automatic data collection
Acknowledgements

RIKEN
Keitaro Yamashita
Go Ueno
Yoshiaki Kawano
Naoki Sakai
Hiroaki Matsuura
Masaki Yamamoto

JASRI
Seiki Baba
Nobuhiro Mizuno
Yuki Nakamura
Kazuya Hasegawa
Takashi Kumasaka

The university of Tokyo
Takafumi Kato
Wataru Shihoya
Kaoru Kumazaki
(Astellas)
Hideaki Kato
Yoshiki Tanaka
Tomohiro Nishizawa
Osamu Nureki

Kyoto University
Ryoji Suno
Hidetsugu Asada
Im Dohyun
Yosuke Toyoda
Takuya Kobayashi
So Iwata

Tokyo Tech University
Satoshi Abe
Takafumi Ueno

• KAMO is open source programs
https://github.com/keitaroyam/yamtbx