



JOURNAL OF
APPLIED
CRYSTALLOGRAPHY

Volume 52 (2019)

Supporting information for article:

Aquarium: an automatic data-processing and experiment information management system for biological macromolecular crystallography beamlines

Feng Yu, Qisheng Wang, Minjun Li, Huan Zhou, Ke Liu, Kunhao Zhang, Zhijun Wang, Qin Xu, Chuanyan Xu, Qiangyan Pan and Jianhua He

SSRF BL17U1 Home Data collection WebCams bl17u1

Filter by data name (case insensitive) Auto Refresh Select date range Open BL17U1

| Data name II | Spacegroup II | a II | b II | c II | α II | β II | γ II | OscWidth II | Frames II | Resolution II | Inner Rmeas II | Outer Rmeas II | Rmeas II | Inner AnomCC II | AnomCC II | Anom Limit Res II | Lstat II | Score II | Residues(built) II | Built Ratio II | NCS II | Time II |
|---------------|---------------|--------|--------|--------|-------|-------|-------|-------------|-----------|---------------|----------------|----------------|----------|-----------------|-----------|-------------------|----------|-----------|--------------------|----------------|--------|------------------------|
| pus9_16 | P 21 21 2 | 121.30 | 214.07 | 33.93 | 90.00 | 90.00 | 90.00 | 0.20 | 1800 | 2.10 | 0.092 | 1.833 | 0.205 | -0.101 | 0.236 | | 0.504 | 199.859 | | | | 2018-05-03 17:07 GMT+8 |
| pus9_15 | P 21 21 2 | 121.10 | 213.64 | 33.85 | 90.00 | 90.00 | 90.00 | 1.00 | 360 | 1.65 | 0.046 | 1.243 | 0.078 | 0.096 | 0.033 | | 0.495 | 886.290 | | | | 2018-05-03 17:01 GMT+8 |
| pus9_14 | P 21 21 2 | 121.72 | 213.81 | 33.83 | 90.00 | 90.00 | 90.00 | 1.00 | 360 | 1.84 | 0.047 | 2.129 | 0.093 | -0.039 | 0.049 | | 0.506 | 993.683 | | | | 2018-05-03 16:54 GMT+8 |
| pus9_13 | P 21 21 2 | 122.04 | 213.78 | 33.97 | 90.00 | 90.00 | 90.00 | 1.00 | 360 | 1.78 | 0.053 | 1.416 | 0.093 | -0.063 | -0.051 | | 0.505 | 647.934 | | | | 2018-05-03 16:47 GMT+8 |
| pus9_12 | P 21 21 2 | 122.01 | 214.23 | 33.96 | 90.00 | 90.00 | 90.00 | 1.00 | 360 | 1.75 | 0.071 | 1.617 | 0.102 | 0.642 | 0.362 | | 0.501 | 661.245 | | | | 2018-05-03 16:40 GMT+8 |
| pus12_5 | P 21 21 21 | 63.82 | 85.61 | 110.73 | 90.00 | 90.00 | 90.00 | 1.00 | 360 | 1.94 | 0.062 | 1.543 | 0.158 | 0.073 | 0.077 | | 0.486 | 347.179 | | | | 2018-05-03 16:33 GMT+8 |
| pus12_3 | P 21 21 21 | 64.18 | 85.58 | 110.75 | 90.00 | 90.00 | 90.00 | 1.00 | 360 | 1.55 | 0.039 | 1.524 | 0.106 | 0.083 | 0.001 | | 0.485 | 823.993 | | | | 2018-05-03 16:34 GMT+8 |
| pus12_16 | P 2 2 21 | 85.45 | 108.22 | 64.33 | 90.00 | 90.00 | 90.00 | 1.00 | 360 | 1.74 | 0.149 | 0.966 | 0.394 | -0.008 | 0.615 | | 0.481 | 244.002 | | | | 2018-05-03 16:17 GMT+8 |
| pus12_15 | P 21 21 2 | 85.64 | 108.74 | 64.27 | 90.00 | 90.00 | 90.00 | 1.00 | 360 | 1.64 | 0.124 | 1.617 | 0.262 | -0.037 | 0.062 | | 0.476 | 314.364 | | | | 2018-05-03 16:10 GMT+8 |
| pus12_14 | P 21 21 2 | 106.36 | 64.42 | 85.27 | 90.00 | 90.00 | 90.00 | 1.00 | 360 | 1.57 | 0.083 | 1.333 | 0.156 | 0.552 | 0.112 | | 0.435 | 575.410 | | | | 2018-05-03 16:03 GMT+8 |
| pus12_13 | P 21 21 21 | 64.11 | 85.65 | 110.86 | 90.00 | 90.00 | 90.00 | 1.00 | 360 | 1.65 | 0.055 | 1.616 | 0.125 | 0.296 | -0.051 | | 0.486 | 595.125 | | | | 2018-05-03 15:57 GMT+8 |
| pus12_12 | P 2 2 2 | 84.28 | 83.73 | 108.43 | 90.00 | 90.00 | 90.00 | 1.00 | 360 | 1.90 | 0.102 | 1.271 | 0.24 | 0.039 | -0.124 | | 0.498 | 229.342 | | | | 2018-05-03 15:50 GMT+8 |
| pus12_11 | P 2 2 21 | 85.29 | 110.02 | 64.21 | 90.00 | 90.00 | 90.00 | 1.00 | 360 | 1.51 | 0.144 | 1.696 | 0.297 | -0.081 | -0.071 | | 0.497 | 446.642 | | | | 2018-05-03 15:43 GMT+8 |
| pus12_10 | P 2 2 21 | 110.40 | 63.90 | 85.68 | 90.00 | 90.00 | 90.00 | 1.00 | 360 | 1.87 | 0.096 | 1.524 | 0.208 | 0.032 | -0.014 | | 0.503 | 284.857 | | | | 2018-05-03 15:35 GMT+8 |
| pus12_9 | P 21 21 21 | 64.46 | 85.44 | 111.05 | 90.00 | 90.00 | 90.00 | 1.00 | 360 | 1.47 | 0.043 | 1.381 | 0.103 | 0.029 | -0.024 | | 0.501 | 1,008.016 | | | | 2018-05-03 15:28 GMT+8 |
| pus12_8 | P 21 21 2 | 85.67 | 111.05 | 64.05 | 90.00 | 90.00 | 90.00 | 1.00 | 360 | 1.38 | 0.042 | 0.962 | 0.078 | -0.17 | -0.053 | | 0.485 | 1,326.530 | | | | 2018-05-03 15:21 GMT+8 |
| pus12_6 | P 21 21 21 | 64.50 | 85.91 | 111.46 | 90.00 | 90.00 | 90.00 | 1.00 | 360 | 1.47 | 0.048 | 1.292 | 0.104 | -0.091 | 0.019 | | 0.493 | 1,027.059 | | | | 2018-05-03 15:11 GMT+8 |
| pus13_2 | P 21 21 2 | 86.38 | 111.12 | 64.46 | 90.00 | 90.00 | 90.00 | 1.00 | 360 | 2.00 | 0.066 | 1.569 | 0.177 | 0.268 | 0.046 | | 0.475 | 270.848 | | | | 2018-05-03 14:49 GMT+8 |
| pus13_1 | P 2 2 2 | 64.02 | 85.88 | 111.40 | 90.00 | 90.00 | 90.00 | 1.00 | 360 | 1.96 | 0.076 | 1.608 | 0.273 | -0.333 | 0.033 | | 0.486 | 237.608 | | | | 2018-05-03 14:39 GMT+8 |
| pus13_16 | P 21 21 2 | 247.21 | 80.27 | 127.06 | 90.00 | 90.00 | 90.00 | 1.00 | 360 | 5.99 | 0.065 | 1.804 | 0.11 | 0.549 | 0.3 | | 0.509 | 14.309 | | | | 2018-05-03 14:25 GMT+8 |
| pus13_13_sht2 | | | | | | | | 1.00 | 360 | 7.34 | | | | | | | | | | | | 2018-05-03 14:02 GMT+8 |
| pus13_13_sht1 | | | | | | | | 0.10 | 3600 | 7.40 | | | | | | | | | | | | 2018-05-03 13:53 GMT+8 |
| pus13_13 | P 2 2 2 | 80.15 | 127.14 | 247.26 | 90.00 | 90.00 | 90.00 | 0.20 | 1800 | 6.20 | 0.097 | 1.792 | 0.21 | -0.372 | -0.156 | | 0.462 | 3.941 | | | | 2018-05-03 13:45 GMT+8 |
| pus13_9 | P 2 2 21 | 127.58 | 245.88 | 80.37 | 90.00 | 90.00 | 90.00 | 0.50 | 720 | 6.39 | 0.086 | 1.297 | 0.183 | -0.308 | 0.013 | | 0.507 | 9.490 | | | | 2018-05-03 13:19 GMT+8 |
| pus9_11 | P 21 21 2 | 121.60 | 213.77 | 33.88 | 90.00 | 90.00 | 90.00 | 0.30 | 720 | 1.76 | 0.03 | 1.782 | 0.063 | -0.098 | -0.033 | | 0.503 | 805.414 | | | | 2018-05-03 13:16 GMT+8 |

First Previous 1 2 Next Last

Figure S1. A summary of datasets of a visit. R_{meas} , AnormCC and Anorm Limit Res come from Aimless; Lstat is from CTRUNCATE; scores were calculated using the number of unique reflections, the completeness, the number of molecules per asymmetric unit, I/σ and the volume of unit cell. The scoring function is identical with that adopted in XChemExplorer; the built residues, built ratio and NCS come from porpoise_sad.

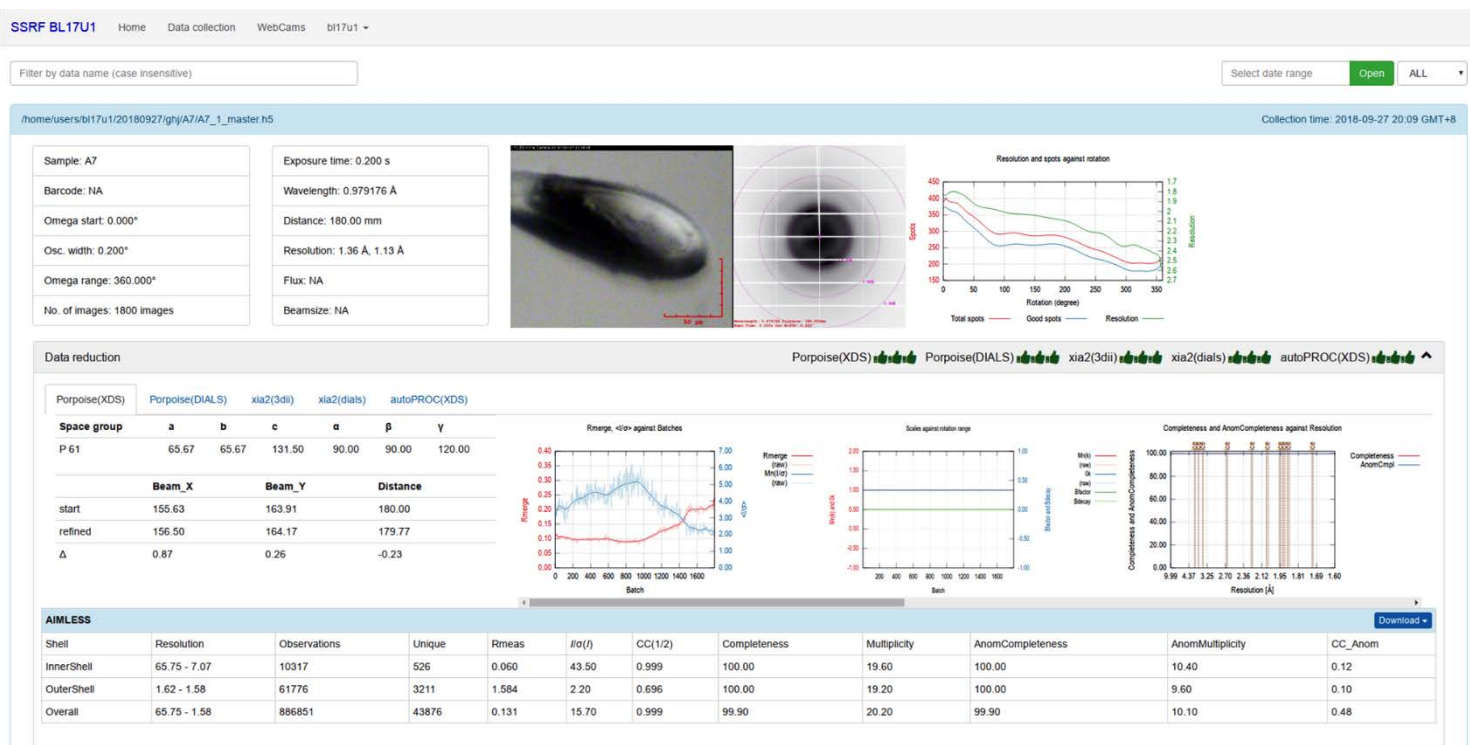


Figure S2. The detailed overview of the dataset, including the sample name, wavelength, distance, oscillation width, exposure time, etc. The snapshot, diffraction image and DISTL results show in the right side. Results from data reduction and SAD pipeline can be inspected through dropdown menu. SealWeb applies the thumb icons to represent the quality of data. The more thumbs-up means the better quality of the data (Table S3 of supplementary materials).

1. A test example of Porpoise from data reduction to mainchain building

Aquarium runs on a small HPC with Linux system. There are 8 computing nodes, each with two Intel Xeon 2640-V4 CPUs and 64G memory. The storage contains 27 TB flash memory and 48 TB hard disk drive.

A selenium derivatised crystal is used as test example for Porpoise. The dataset contains 1200 frames of raw images. The space group is $P4_122$; the unit cell constants are $a = 42.67 \text{ \AA}$, $b = 42.67 \text{ \AA}$, $c = 136.02 \text{ \AA}$. There are 123 residues in an asymmetric unit, 6 of which are methionine. The data collection, data reduction and SAD phasing taken about 390 s, 93.281 s and 104.644 s, respectively. Finally, five Se atoms were found, and 98 of 123 residues were built.

Table S1. The data processing time from data collection to main-chain building










| | |
|--|--|
| Data collection | 1200 frames, omega range 360° and 0.3 second/frame. 360 s for exposure, and extra about 30 s waiting for data transport. |
| Initialization (porpoise_dr) | 0.164 s |
| Spots finding (all frames) and indexing | 29.177 s |
| The first-pass integration as $P1$ (the first 30° of raw images) | 9.815 s |
| Laue group determination, the second-pass integration (all frames), merging and XSCALE | 44.614 s |
| Systematic absences determination | 5.555 s |
| Reindex and Scaling | 2.592 s |
| Convert intensities to structure factors | 1.362 s |
| Initialization (porpoise_sad) | 0.156 s |
| Checking twinning and anomalous signals | 0.785 s |
| Finding anomalous scatters | 44.613 s |
| Chirality determination | 1.939 s |
| Phasing | 2.572 s |
| Density modification | 37.665 s |
| NCS finding | 3.426 s |
| Model building | 13.488s |
| Total | 587.925 s (390s for data collection, 93.281 s for data reduction and 104.644 s for SAD phasing and model building) |

Table S2. Computing resources, run time and results of porpoise_sad

| Protein and dataset information | | | |
|--|-----------------------------------|---|--|
| The number of residues in the ASU | | 123 | |
| The number of methionine in the ASU | | 6 | |
| Solvent content | | 47% | |
| High-resolution limit | | 2.64 Å | |
| The number of unique reflections | | 4155 | |
| High preference cluster | | | |
| Nodes | | 8 nodes | |
| Echo node | | 2 × Intel Xeon 2640-V4 CPUs (20 physics cores) 64 G memory | |
| SAD Phasing and model building | | | |
| Programs | Computing resource | Run time | Results |
| SHELXC | Run concurrently on a single node | 4.150 s | - |
| SHELXD | Run concurrently on eight nodes | 41.404 s | CFOM = 92.52 CCall = 56.19 CCweak = 36.33 10 Se atoms found. |
| SHELXE | Run concurrently on a single node | 1.939 s | Contrast: Original hand is 0.412. Inverted hand is 0.367. Pseudo-free CC: Original hand is 53.33 %. Inverted hand is 51.93 %. |
| Phaser | Run on a single node | 2.572 s | 5 Se atoms after refinement |
| PARROT and SOLOMON | Run concurrently on eight nodes | 37.665 s | PARROT: solvent fraction 60% SOLOMON: solvent fraction 50% |
| PARROT (NCS) | Run concurrently on a single node | 3.426 s | No NCS found |
| BUCCANEER | Run concurrently on a single node | 13.488 s | 98 residues built |

2. The icon meaning of data process results in SealWeb

Table S3. The icon meaning of data process results in SealWeb

| | |
|---|---|
|  | Data reduction: R_{meas} of inner resolution shell $\leq 7\%$, and high-resolution limit ≤ 2.0 Å. SAD phasing: built residues/estimated residues $\geq 85\%$. |
|  | Data reduction: R_{meas} of inner resolution shell $\leq 10\%$, high-resolution limit ≤ 3.0 Å, but excluding above. SAD phasing: built residues/estimated residues $\geq 70\%$, but excluding above. |
|  | Data reduction: R_{meas} of inner resolution shell $\leq 15\%$, high-resolution limit ≤ 3.5 Å, but excluding above. SAD phasing: built residues/estimated residues $\geq 60\%$, but excluding above. |
|  | All other successful jobs |
|  | Job finished, but there was error happened when parsing the log files. |
|  | Job failed |
|  | Job is in the queue and waiting for executing |
|  | Job is executing |
|  | Unknown |