

# Mail-in data collection at SPring-8 protein crystallography beamlines

Nobuo Okazaki,<sup>a\*</sup> Kazuya Hasegawa,<sup>a</sup> Go Ueno,<sup>b</sup> Hironori Murakami,<sup>b</sup>  
Takashi Kumasaka<sup>a</sup> and Masaki Yamamoto<sup>b,a\*</sup>

<sup>a</sup>Japan Synchrotron Radiation Research Institute/SPring-8, 1-1-1 Kouto, Sayo-cho, Sayo-gun, Hyogo 679-5198, Japan, and <sup>b</sup>RIKEN SPring-8 Center, 1-1-1 Kouto, Sayo-cho, Sayo-gun, Hyogo 679-5148, Japan.  
E-mail: okazaki@spring8.or.jp, yamamoto@postman.riken.go.jp

A mail-in data collection system makes it possible for beamline users to collect diffraction data without visiting a synchrotron facility. In the mail-in data collection system at SPring-8, users pack crystals into sample trays and send the trays to SPring-8 *via* a courier service as the first step. Next, the user specifies measurement conditions and checks the diffraction images *via* the Internet. The user can also collect diffraction data using an automated sample changer robot and beamline control software. For distant users there is a newly developed data management system, D-Cha. D-Cha provides a graphical user interface that enables the user to specify the experimental conditions for samples and to check and download the diffraction images using a web browser. This system is now in routine operation and is contributing to high-throughput beamline operation.

**Keywords:** mail-in data collection; high-throughput data collection; beamline automation; web application; database system.

## 1. Introduction

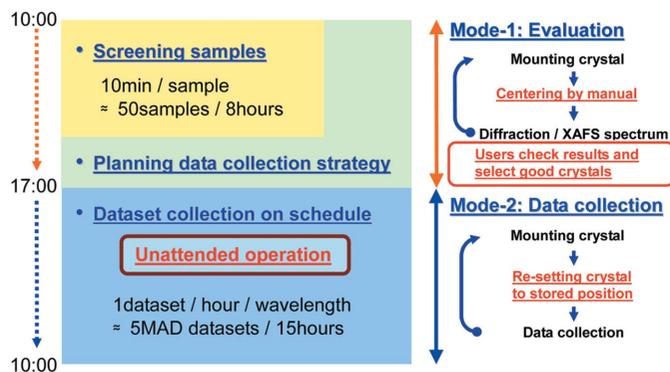
At SPring-8, the RIKEN structural genomics beamlines I and II (BL26B1 and BL26B2) have been constructed to contribute to structural genomics research. To achieve high-throughput protein crystallography, we have developed two special components: an automated sample changer robot, SPACE (SPring-8 precise automatic cryo-sample exchanger) (Ueno, Hirose *et al.*, 2004), which can change up to 100 crystals in a diffractometer, and the beamline control software BSS (beamline scheduling software) (Ueno *et al.*, 2005), which can perform successive data collection by controlling beamline devices, including SPACE, and managing the data collection schedule. The combination of SPACE and BSS enables unmanned overnight data collection and allows the beamline to operate with high efficiency.

To better accommodate the new system to the needs of distant users and a laboratory information management system (LIMS) for beamline operation and experiments, we have developed a new operation system that enables users to use the beamlines from their own laboratories. In general, the methods of conducting experiments using the synchrotron facilities from remote locations fall into two categories: (i) mail-in data collection and (ii) remote-controlled data collection. The first method is used by the Swiss Light Source at the Paul Scherrer Institut (SLS/PSI) and the Advanced Photon Source (Advanced Photon Source, 2006). These users send their samples to a beamline, and datasets collected by beamline staff are returned to the users. The users, by the remote-controlled method, can actually control their experiments from remote locations *via* a network program after sending their samples to a beamline. For example, at the Stanford Synchrotron Radiation Laboratory remote data can be collected by using the beamline control application *Blu-Ice* and the

*Distributed Control System (DCS)* (McPhillips *et al.*, 2002). At the European Synchrotron Radiation Facility, this kind of service is available *via* their remote access control system (European Synchrotron Radiation Facility, 2007). We have developed a mail-in data collection system which is a combination of the mail-in and remote access methods. In our system, users can decide their own measurement conditions and benefit from the beamline operator's assistance. Here we discuss in detail the mail-in data collection at SPring-8 using the web database application D-Cha (database for crystallography with home-lab arrangement), which enables remote operation and mail-in data collection.

## 2. Automated operation

The beamline automation system at BL26B2 (Ueno *et al.*, 2006) has been in operation since 2003. There are two modes of operation: mode-1 is the evaluation phase and mode-2 is the data collection phase (Fig. 1). In mode-1, the user or beamline operator interactively centers the crystal on the X-ray beam path and a few diffraction images and optionally the XAFS spectrum of the crystal are automatically measured by BSS. The centering position of each crystal is recorded in addition to the diffraction images and XAFS spectrum. Before conducting mode-2, the user selects crystals from which to collect datasets, based on the mode-1 evaluation, and enters the measurement conditions into BSS. In mode-2, each crystal is automatically restored to the centering position recorded in mode-1, and measurement is performed automatically and continuously. These operations are made possible by accurate sample mounting with SPACE. Thus, mode-1 requires the user to inspect the crystals and diffraction images, whereas mode-2 is an unmanned operation and



**Figure 1**  
Schematic diagram of the two modes of operation. In a typical procedure, mode-1 is conducted in the day time and mode-2 is operated during the night.

the user simply waits for completion of data collection. Therefore, mode-1 and mode-2 are usually conducted during the day and night, respectively.

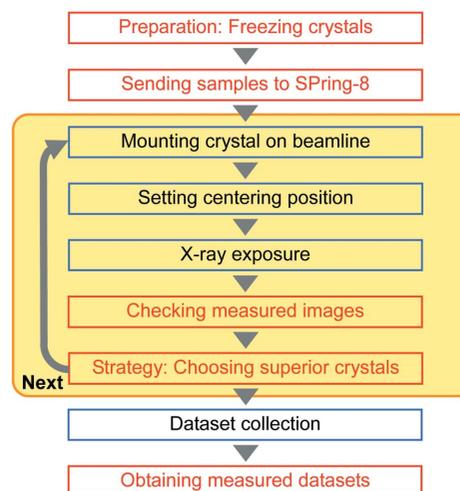
### 3. Mail-in data collection

#### 3.1. Aims

Synchrotron facilities are usually at distant locations from the user's laboratory, which necessitates considerable expenditure of time and money in traveling and carrying samples. Remote users would therefore benefit from being able to conduct their experiments without visiting the synchrotron facility. Although an outsourcing service is a possible solution, another possibility is remote beamline operation by users. The aim of our mail-in system is to provide such a service with flexible use of the beamline for distant users. Such a system would both improve beamline usability for researchers and increase the efficiency of data collection. Our mail-in data collection system was designed so that (i) the user need not visit SPring-8; (ii) the user can request particular measurement conditions; (iii) the user can check their measurement results and have the collected data returned to them with storage media or *via* the Internet; and (iv) the system performs as a LIMS. To establish this system an additional component is required: a remote-access user interface to enable the user to record sample information, edit the measurement conditions, browse and acquire measurement data.

#### 3.2. Mail-in data collection cycle and database

A typical experimental protocol is shown in Fig. 2. At the first stage of mail-in data collection the remote users store samples in trays designed for SPACE. This operation is performed using an offline type SPACE, which is dedicated to packing the sample. The offline SPACE stores the samples to the sample tray. Another way of sample preparation is to use a compact toolkit designed for mounting crystals by hand (Fig. 3). After the trays are prepared, the user sends the samples to SPring-8 and enters the measurement schedule from their laboratory using D-Cha. The beamline operators load the sample trays sent by the users into SPACE and perform the centering of each crystal. The centering position is stored in the sample information database and can be retrieved by SPACE and BSS. In evaluation mode (mode-1), 10 min are required for each sample on average, therefore the evaluation of 50 samples can be completed in about 8 h of beamline operators' working time. At this time, the user can browse the measurement results (diffraction images and XAFS spectra) at their remote laboratory using the D-Cha web interface.



**Figure 2**  
Typical scheme of our mail-in data collection. Red and blue boxes show operations performed by users and beamline operators, respectively. Procedures in the yellow box are repeated for each sample.

When evaluation mode is completed, the data collection mode (mode-2) is conducted automatically overnight without an operator. The merit of this scheme is that human manipulation and inspection are completed during the daytime.

### 4. D-Cha: a web database application

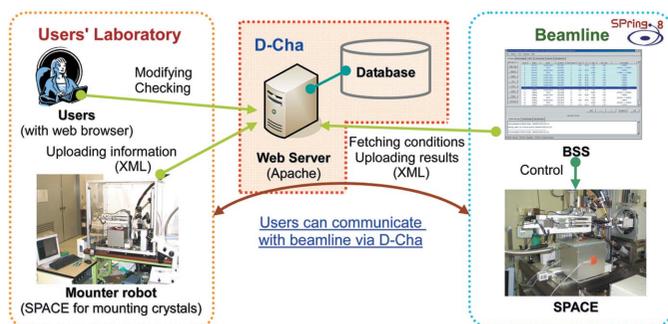
#### 4.1. Overview

D-Cha is a web application consisting of a relational database and a web server. The user can use D-Cha *via* the Internet. Fig. 4 shows the relationships between the user, the beamline and D-Cha. At the user's laboratory the user can access D-Cha with a web browser. They can enter and modify measurement conditions, and check the results at their remote laboratory. Offline SPACE is also used to upload information on the samples to D-Cha using the Hypertext Transfer Protocol (HTTP). BSS fetches measurement conditions from D-Cha and uploads experimental results. BSS and offline SPACE commu-

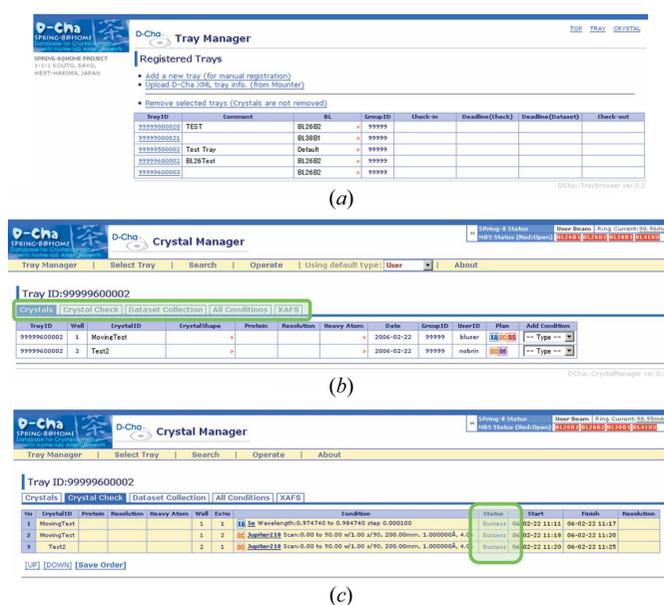


**Figure 3**  
Overview of the compact toolkit designed for mounting crystals by hand.

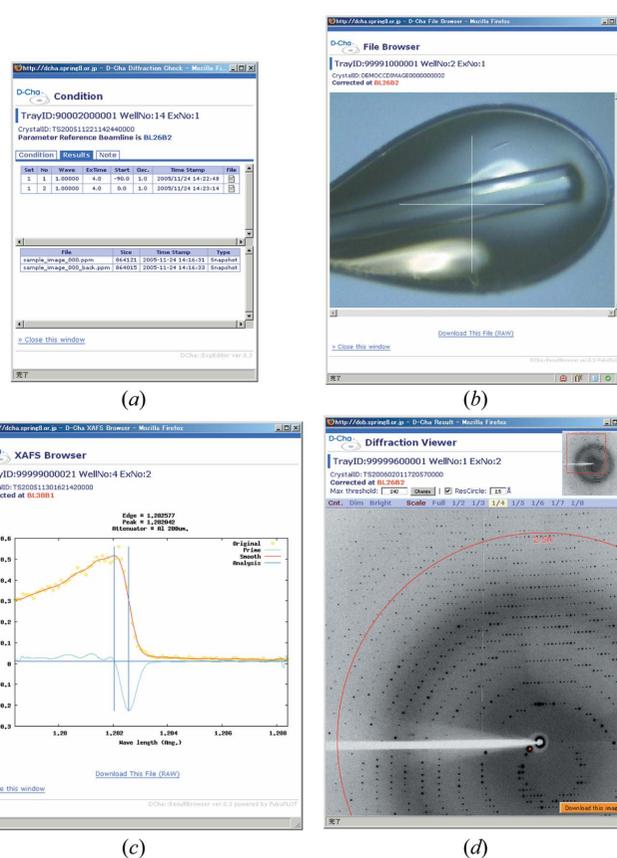
# diffraction structural biology



**Figure 4**  
Schematic diagram of relationships among D-Cha, the user and the beamline. Light green lines represent connections over HTTP or secure HTTP (HTTPS).



**Figure 5**  
GUI for registering crystals and measurements. (a) Tray Manager: registering SPACE trays. (b) Crystal Manager-1: registering and browsing crystals. The box highlighted in green is a tab for switching between crystals and measurements. (c) Crystal Manager-2: entering and browsing measurement conditions. The box highlighted in green shows measurement status. Users can check the progress of measurements. (d) Condition dialog window.



**Figure 6**  
Dialog windows for browsing measurement results. (a) Condition dialog with result file listing. The user can open the result browser by clicking the filename. (b) Crystal snapshot view. (c) XAFS spectrum browser. (d) Diffraction image viewer.

icate with D-Cha in the extensible markup language (XML) format. BSS controls SPACE and other beamline components, such as the detector, according to the recorded schedule. In this way the user and the beamline can communicate using D-Cha. Thus, the user can conduct experiments with the synchrotron facility at their laboratory via D-Cha.

## 4.2. Usage and graphical user interface of D-Cha

The user first accesses D-Cha with a web-based graphical user interface (GUI) and logs in with a registered account name and password. In the next step the user enters the sample information and experimental conditions using the GUI of the web browser. Two types of sample registration scheme are available on D-Cha. One uses the Tray Manager and Crystal Manager windows (Figs. 5a and 5b); alternatively, sample information can be entered via the offline SPACE control GUI while the samples are being mounted on the trays. When the samples are all registered, the user enters the measurement schedule for each crystal using the Crystal Manager window (Fig. 5c). Experiments are classified into the following four types: (i) diffraction checking, (ii) XAFS, (iii) single dataset collection and (iv) multiwavelength anomalous dispersion (MAD) dataset collection. The first two measurement modes are used in the evaluation phase (mode-1) and the last two modes are in the data collection phase (mode-2). Details of measurement conditions are entered via the Condition dialog window (Fig. 5d). The appearance of the window is similar to those of BSS, so that BSS users can easily operate this GUI interface.

**Table 1**

Summary of mail-in data collection with distant users at BL26B2 in 2007A operation period.

Operating days	26	
Evaluated crystals	652	Average 25 crystals/daytime
Single datasets	188	
MAD experiments	17 (48 datasets)	
Total datasets	236	Average 9 datasets/night

The user can browse the measurement results on D-Cha. The Condition dialog displays a list of the result files (Fig. 6a). The user can browse the photographs, diffraction images and XAFS spectra of each crystal (Figs. 6b–6d) and download their raw data through these dialog windows.

#### 4.3. Data management

D-Cha was designed to manage massive amounts of data, including sample information, measurement conditions and the results. Trays, crystals and measurement conditions are identified by a unique tray ID, crystal ID and experiment number for each crystal ID, respectively. All information is accessible only by its owner and illegal access is prohibited to ensure privacy for each user.

D-Cha provides a simple experimental database that can store any information that the user has independently defined for the crystals and measurements (for example, sample name, cryo condition, heavy atoms and comments). The user can use these data freely and relate records in D-Cha to their own database and to the LIMS.

#### 5. Operation result

At BL26B2, mail-in data collection runs in a routine manner. This system has been used from distant facilities like the Genomic Science Center (RIKEN Yokohama), the Synchrotron Radiation System Biology Research Group (RIKEN SPring-8 Center) and the Advanced Protein Crystallography Research Group (RIKEN SPring-8 Center). A summary of mail-in data collection with distant users for the first half of the year 2007 (2007A period) is shown in Table 1.

Trials of mail-in data collection were conducted at the Structural Biology Beamline III for the public (BL38B1). Eleven academic users have conducted their measurements with the mail-in data collection system.

#### 6. Platform, language and availability

D-Cha was developed as a web application suitable for Mozilla Firefox (Mozilla Foundation, <http://www.mozilla.org/>). For ease of maintenance, the application is written in the Perl scripting language (<http://www.perl.org/>). It runs with mod\_perl on the Apache Web Server (Apache Software Foundation, <http://www.apache.org/>). mod\_perl is the Perl interface for Apache, used widely in web application development to improve the performance of applications. D-Cha consists of several modules: CGI::Application (a web application framework), DBIx::Class (an object-relational mapper), and Template Toolkit for HTML processing. These modules are obtainable from the Comprehensive Perl Archive Network (CPAN, <http://www.cpan.org/>). D-Cha uses a relational database, PostgreSQL (<http://www.postgresql.org/>).

D-Cha is currently customized to work with SPACE and BSS at SPring-8 and is used at the Pharmaceutical Industry Beamline (BL32B2), BL26B1, BL26B2 and BL38B1.

We would like to thank all D-Cha users for very useful suggestions, which led to the system improvements.

#### References

- Advanced Photon Source (2006). *APS User News-at-a-Glance*, Issue 40, [http://www.aps.anl.gov/Users/Communications/User\\_News/2006/user\\_news\\_40.html#01](http://www.aps.anl.gov/Users/Communications/User_News/2006/user_news_40.html#01).
- European Synchrotron Radiation Facility (2007). Remote Access, [http://www.esrf.eu/UsersAndScience/Experiments/MX/How\\_to\\_use\\_our\\_beamlines/remote-access](http://www.esrf.eu/UsersAndScience/Experiments/MX/How_to_use_our_beamlines/remote-access).
- McPhillips, T. M., McPhillips, S. E., Chiu, H.-J., Cohen, A. E., Deacon, A. M., Ellis, P. J., Garman, E., Gonzalez, A., Sauter, N. K., Phizackerley, R. P., Soltis, S. M. & Kuhn, P. (2002). *J. Synchrotron Rad.* **9**, 401–406.
- Ueno, G., Hirose, R., Ida, K., Kumasaka, T. & Yamamoto, M. (2004). *J. Appl. Cryst.* **37**, 867–873.
- Ueno, G., Kanda, H., Hirose, R., Ida, K., Kumasaka, T. & Yamamoto, M. (2006). *J. Struct. Funct. Genom.* **7**, 15–22.
- Ueno, G., Kanda, H., Kumasaka, T. & Yamamoto, M. (2005). *J. Synchrotron Rad.* **12**, 380–384.