2 Structural Biology Research Center

2-1 Introduction

As reported in the previous year, the Structural Biology Research Group, Prof. Soichi Wakatsuki as a group leader, started in May 2000. The aims of the research group are user support in synchrotron radiation X-ray crystallography of macromolecules, highly advanced technical development and in house research of structural biology. The group has grown steadily during the last three years; the structural biology building extended from 429 m^2 to 643 m^2, and the members increased from four persons to about thirty (Fig. 1) including a professor, an associate professor and five research associates as a core staff. About the half of the members is engaged in beamline operation and development and the remaining half in biological research. Prior to the change of status of KEK from government institute to an agency in April, 2004, the Structural Biology Research Group has been promoted to the Structural Biology Research Center in May, 2003.

2-2 Protein 3000 Project, Individual Analyses Program

This is the second year of the 5-year national project. The Structural Biology Research Center serves as one of the 9 consortia of Protein 3000 project pursuing structural and functional analyses in the post-translational modification and transport. The consortium consists of seven universities and four research institutes (Table 1).

Cell signaling and intracellular trafficking are the means by which eukaryotic cells deliver cargo proteins to various organelles, cell membranes, and extracellular destinations (Fig. 2). Accurate distribution of these

Figuer 1
Photograph of the members of Structural Biology Research Center at the entrance of Structural Biology Building.

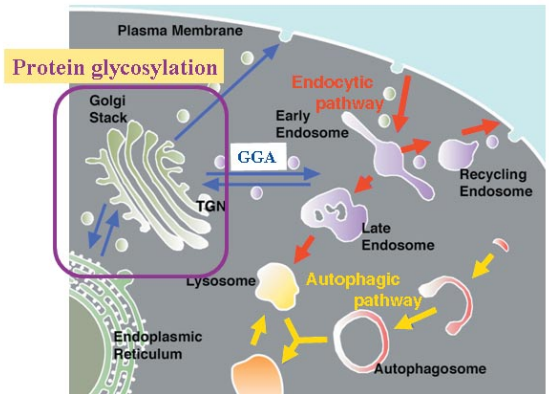

Figure 2
Protein glycosylation and transport in cells.

Table 1 Members of the "Posttranslational and Transport" network in Protein 3000 Project.

Functional Analyses	Intracellular trafficking	Akihiko Nakano (RIKEN, Univ. of Tokyo), Kazuhisa Nakayama (Kyoto Univ. Pharmaceutical), Hiroshi Ohno (Kanazawa Univ. Cancer Research Institute, RIKEN Laboratory of Epithelial Immunobiology), Hiroaki Kato (Kyoto Univ. Pharmaceutical), Masayuki Murata (University of Tokyo, Arts and Sciences), Soichi Wakatsuki (KEK-PF)
	Post-translational modification	Toshisuke Kawasaki (Kyoto Univ. Pharmaceutical), Naoyuki Taniguchi (Osaka Univ. Medicine), Yoshifumi Jigami (AIST), Koichi Kato (Nagoya City Univ. Pharmaceutical), Sumihiro Hase (Osaka Univ. Science), Soichi Wakatsuki (KEK-PF)
Structural Biology	X-ray crystallography	Takamasa Nonaka (Nagaoka Univ.of Technorogy), Nobutada Tanaka (Showa Univ. Pharmaceutical), Hiroaki Kato (Kyoto Univ. Pharmaceutical), Soichi Wakatsuki (KEK-PF)
	NMR, Bioinformatics	Koichi Kato (Nagoya City Univ. Pharmaceutical), Kei Yura (JAERI CCSE)

proteins is crucial for a range of cellular functions and activities. Mutations in genes encoding protein transport regulators underlie a number of genetic diseases. Accordingly, an increased understanding of the biological and biomedical function of transport proteins is indispensable for progress towards treating human diseases. Furthermore, more than half of human proteins undergo post-translational processing and modification such as glycosylation. Indeed, glycobiology has gained prominence as a post-genomic science recently for its role in modulating protein function and transducing cellular signals. Recent interest in proteomics, the study of structure-function relationships of proteins, further increases the relevance of structural exploration in glycobiology and protein transport. Post-translational modification and intracellular transport are closely related complex events. While these fields of study have historically remained unrelated, integration of the two research fields is a promising strategy for gaining a better understanding of the complexities of cellular events. Therefore, this project aims to coordinate researchers from both fields so that a concerted effort can be made towards the structure-function analyses of proteins involved in these processes.

To facilitate the current research program, the Structural Biology Research Center has taken the lead by coodiating the research network. As part of the project, we have sought to maximize the efficiency of large-scale expression, purification, and crystallization of proteins by trying to eliminate bottlenecks in each process. X-ray crystallographic studies using synchrotron X-ray radiation are conducted at KEK-PF, while NMR experiments are conducted at Nagoya City University. Each structural analysis project maintains close contacts with groups responsible for functional analyses. Our initial research plan was to accomplish structural and functional analyses of more than 70 proteins in 5 years. To date, we have completed structural analyses of 57 proteins and complexes and are currently working on about 100 additional targets, prompting us to increase the goal from 70 to 150 during the 5-year period. Furthermore, we filed seven domestic and two international patents, based on our R&D, crystallization, and structural and functional studies of the target proteins.

In addition, we have designed and constructed two new high-throughput beamlines (BL-5 and AR-NW12), and further improved the existing beamlines for protein crystallographic experiments at KEK-PF. To facilitate the Protein 3000 project, we have established an operation scheme to reserve about 30% of beam time of our beamlines for users of the eight university consortia of the project, for which we have developed a WEB-based beam time reservation system (Fig. 3).

As stated above, post-translational modification of proteins is closely associated with the intracellular transport systems, and new molecular interactions are being identified and characterized, thus providing new targets to our project almost weekly. A systematic approach combining structure determination and functional analysis

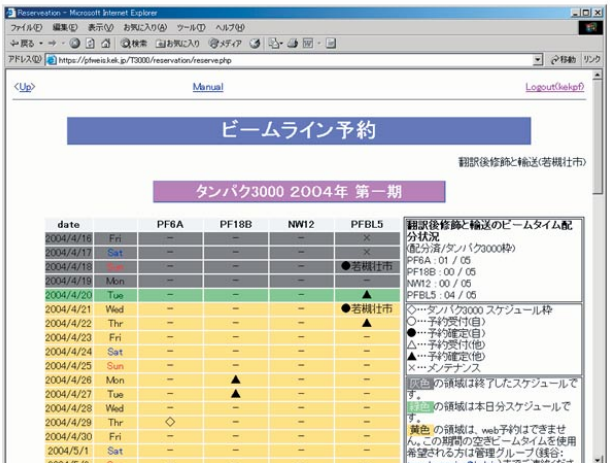

Figure 3
Protein 3000 WEB-based time reservation system.

of such proteins may lead to the development of a novel technology which allows for the synthesis of glycoproteins with a specific oligosaccharide structure. A logical extension of the current proteomics studies is the incorporation of the post-translational modification and transport machineries, and in particular, membrane complexes which are key players in membrane trafficking. Our future research plan, thus, includes structural analyses of these complexes. With this in mind, we will further develop our integrated and synergistic approaches within our consortia and establish new research collaborations with groups in other research networks.

2-3 Structural Biology Program

Structural/functional analyses of GGA adaptor protein for intracellular vesicle transport

The Golgi apparatus is the central delivery system in all eukaryotic cells, including those of human, where cargo proteins are transported via vesicles to lysosomes. This process involves binding of the proteins to a cargo receptor localized at the trans-Golgi network (TGN) membrane, sorting by the GGA adaptor protein, incorporation into the endocytic transport vesicle, and finally delivery to the endosome. The GGA proteins are a new class of monomeric adapter proteins which consist of three domains: VHS, GAT, and GAE (Fig. 4).

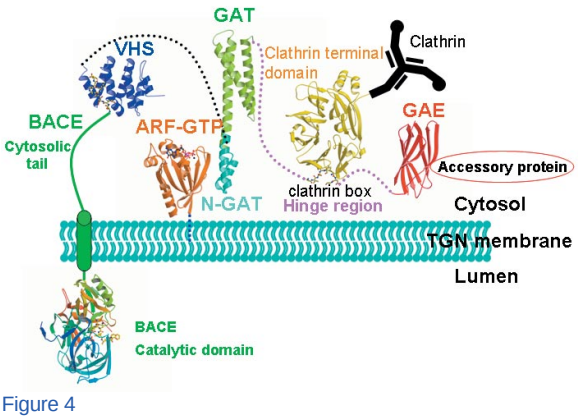

Figure 4
An interaction network between GGA and other proteins.

In year 2002, we reported the structures of the VHS domain, which binds cargo receptors [1]. In year 2003, we continued the research on the molecular recognition of the VHS domain and demonstrated that the VHS domain of GGA binds to β-secretase, which is involved in the synthesis of β amyloid, a molecule involved in the pathogenesis of Alzheimer's disease [2]. The complex structures of the VHS and the cytosolic signal peptide of β -secretase (unphosphorylated and phosphorylated) show the recognition mechanism by the VHS domain similar to the mannose 6-phosphate receptors. The resulting electrostatic landscapes of the complexes, however, are very different between the unphosphorylated and phosphorylated forms, which might serve as signals for different destinations.

The GAT-domain, the second domain following the N-terminal VHS domain, is responsible for docking of GGA onto the TGN membranes by interacting with ARF-family GTP-binding proteins. Determination of the structure of the GAT domain has become a focus of attention in structural biology worldwide, and four groups from the United States, England, and Japan reported its structure almost simultaneously. Our group, however, was the only one to have successfully elucidated not only the structure of the GAT domain alone but also the structure of the complex between the GAT domain and ARF, revealing the molecular mechanism underlying membrane recruitment of the GGA adaptor protein by ARF to the TGN membrane in vascular transport [3].

The hinge region between the GAT-domain and GAE-domain interacts with clathrin, a protein that coats transport vesicles. We previously solved the crystal structure of the γ-ear domain of the AP-1 complex, a homolog of the C-terminal GAE domain [4]. Subsequently, structural analysis of the complex between the C-terminal GAE domain and the accessory protein that interacts with the GAE domain and the γ-ear domain has been completed (paper in preparation), which sheds light on the additional regulation of the transport vesicles.

Structural/functional analyses of GlcAT-P, a gycotransferase which is required for HNK-1 carbohydrate biosynthesis

Carbohydrate molecules on cell surfaces modulate a variety of cellular functions, including cell-to-cell interactions. The HNK-1 carbohydrate epitope, which is found on many neural cell adhesion molecules, is thought to be required for development of brain and nerve systems. In mammalian cells, two glucuronyltransferases, GlcAT-P and GlcAT-S, which are associated with the biosynthesis of the HNK-1 carbohydrate epitope, have been found. Another glucuronyltransferase, GlcAT-I, whose acceptor substrate is proteoglycan, has also been characterized. These enzymes catalyze the transfer of glucuronic acid (GlcA) from a donor substrate, uridine diphosphoglucuronic acid (UDP-GlcA), to a reducing terminal residue of oligosaccharide chain in the presence of manganese.

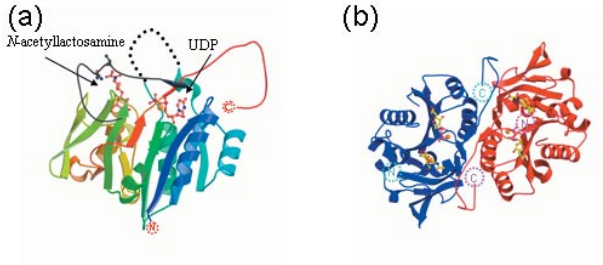

Figure 5
The structure of GlcAT-P in complex with Mn^{2+}, UDP and N-acetyllactosamine. (a) Monomer structure of the GlcAT-P quaternary complex. The C-terminal region of the neighbor molecule is shown in gray. (b) Dimer structure of the complex.

There are many differences in substrate specificity of acceptor sugars among them. This substrate specificity is thought to be important for cell differentiation and function.

To elucidate this, we solved the crystal structures of GlcAT-P with and without substrates [5]. The overall structure of GlcAT-P consists of two regions; the N-terminal UDP-GlcA binding region contains an α/β Rossmann-like fold and the C-terminal acceptor substrate binding region contains two β-sheets [Fig. 5(a)]. The C-terminal long loop, which extends to another molecule, is related by the non-crystallographic two-fold axis [Fig. 5(b)]. An UDP moiety of UDP-GlcA, the donor substrate, is recognized by the amino acid residues conserved among GlcAT-P, GlcAT-S and GlcAT-I not only on their primary sequences but also on their stereochemical dispositions. An acceptor sugar substrate analogue, N-acetyllactosamine (Galβ1-4GlcNAc), is recognized by conserved and nonconserved residues. The nonconserved amino acid residues in the C-terminal long loop of the neighboring molecule and stacking interaction between an aromatic residue and galactose ring account for the differences in the specificity of the substrate acceptor recognition. The details are described in elsewhere (see pp. 37).

2-4 Beamlines

A new beamline BL-5

In FY2003, construction and commissioning of a new high-throughput beamline, BL-5, was completed. We started the public use of BL-5 from spring 2004. BL-5 is designed for high-throughput protein crystallographic experiments at the PF 2.5 GeV-ring. The beamline is optimized for multiwavelength anomalous disperse (MAD) experiments for small size crystals and data collection from crystals with large cell dimensions (up to 1,000 Å). The details of BL-5 are described in elsewhere (see pp. 51).

Current status of the structural biology beamlines

We are now operating two bending magnet beamlines (BL-6A and BL-18B) and two insertion device beamlines (AR-NW12 and BL-5). The specifications of these

Table 2 Structural Biology Beamlines at Photon Factory.

	BL-6A	BL-18B	AR-NW12	BL-5
Starting year	1987	1994	2003	2004
Synchrotron ring	PF	PF	PF-AR	PF
X-ray source	BM	BM	Undulator	MPW
Wavelength range (Å)	0.9-1.3	0.5-2.0	0.7-1.7	0.7-1.9
Energy resolution (ΔE/E)	1×10^{-3}	2×10^{-4}	2×10^{-4}	2×10^{-4}
Photon flux at 1Å (photons/sec)	1.1×10^{10}	1.5×10^{9}	2×10^{11}	$> 10^{11}$
Slit size (mm)	0.1	0.2	0.2	0.2
Detector	ADSC Q4R	ADSC Q4R	ADSC Q210	ADSC Q315
Type	CCD	CCD	CCD	CCD
Active area (mm^2)	188 × 188	188 × 188	210 × 210	315 × 315
Pixel size (mm)	0.0816 × 0.0816	0.0816 × 0.0816	0.051 × 0.051	0.051 × 0.051
Pixel number	2304 × 2304	2304 × 2304	4096 × 4096	6144 × 6144
Frame data size (MB)	11	11	34	75
Rearout time (sec)	8	8	1	1
Typical exposure time (1.0° oscillation)	20 sec	60 sec	3 sec	5 sec
Typical data collection time (180 frames)	1.5 hr	3.5 hr	17 min	20 min
Buckup time for 1 data set (180 frames) (using the IEEE1394 interface)	3 min	3 min	6 min	6 min
Camera distance (mm)	60 ~ 400	50 ~ 500	60 ~ 1000	60 ~ 1000
Detector vertical offset	0 ~ 25°	0 ~ 45°	0 ~ 100 mm	0 ~ 180 mm
Software for image processing	HKL2000 DPS/mosflm	HKL2000 DPS/mosflm	HKL2000 DPS/mosflm	HKL2000 DPS/mosflm

four beamlines are summarized in Table 2.

AR-NW12 is designed for high-throughput diffraction measurements at the PF-AR. The beamline was constructed during FY2002, and has been made available for public users since May, 2003. In-vacuum tapered undulator is used as the source which provides high-flux X-ray beam optimized at around 12.7 keV using the 3rd harmonics. The mirror system focuses the X-ray beam at the sample position and the beam intensity is more than 10^{11} photons/sec at the sample position, with 200 x 200 microns slit and 0.5 horizontal mrad beam divergence. The beamline is equipped with a high-precision diffractometer and a high-speed CCD detector with 210 mm square active area, ADSC Q210. A combination of high-flux X-ray beam and high-speed, high-precision data collectioin system enable high-throughput structural biology experiments. It takes about one or two hours to complete a typical MAD experiment (three wavelengths, 180 frames, 1 degree oscillation and a few sec exposure). The system for the beamline control is common among the other structural biology beamlines (described bellow).

For the last few years, BL-6A has gone through a vigorous refurbishment program aimed at higher overall performance which includes a) installation of a CCD-based detector system, ADSC Q4R (FY1999), b) replacement of the optical bench for MAD experiments using the CCD system (FY2000), and c) replacement of the focusing mirror and its bending system which increased the beam intensity by about three times (FY2003). The common control system has been developed with the aim of fully automatic experiments (see bellow).

BL-18B was originally designed for both MAD experiments and the time-resolved Laue method using white X-ray experiments. At present, however, the beamline supports only monochromatic experiments and has been further developed to be optimized for MAD experiments. In year 2000, ADSC Q4R detector was installed for the purpose of automation and higher efficiency of data collection. In FY2003, we replaced the crystals for the monochromator from Si(111) to Ge(111), for low energy experiments, especially for the sulfur anomalous studies. After the replacement, the beam intensity at the wavelength longer than 1.12 Å is increased by 3 to 5 times. BL-18B is scheduled to be closed at the end of FY2004.

Construction of a new beamline BL-17

At the same time of the BL-18B closing, a new micro-focus MAD beamline, BL-17, using a mini-pole undulator as the insertion device, will be constructed. The beamline is now designed for the experiments with micro crystals and low energy experiments.

Development of the beamline control system

We have developed a common controlling system for the structural biology beamlines. The core of the system is named STARS, which was originally designed for the central control of the interlock system of the PF. We have developed new GUI software, as clients of STARS, for set up by beamline staff and for user experiments. The staff GUI is coded in Perl/Tk, thus transportable to almost all operating systems. This is a big advantage from the point of view of software development. The current version has three sets of GUIs for crystal alignment,

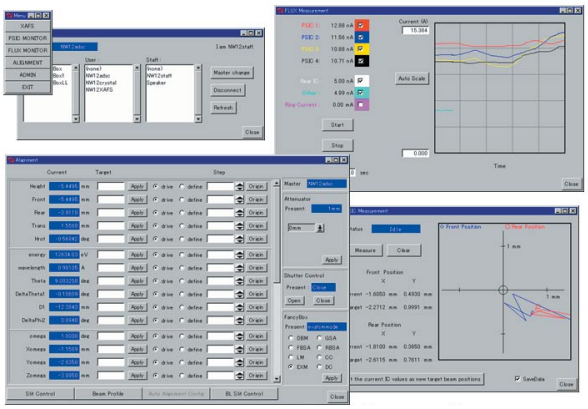

Figure 6
GUI for the beamline control.

XAFS measurement and diffraction experiments (Fig. 6). The systems are installed in BL-5, BL-6A and AR-NW12 where users can perform their experiments with the same operation. The integration of these GUIs for simpler and more convenient operation is being pursued.

As the next step, we have developed an integrated controlling system using a unified database for high throughput protein crystallography experiments. All information necessary to perform protein crystallography experiments (purification, crystallization, crystal harvesting, data collection, data processing) is stored in MySQL relational database. This control system communicates with experimental equipments via STARS server. In FY2004, this system will be tested and will be made available for user operation. In future, we will be able to use this system from outside of KEK via a high-speed network and the WEB server (see below)

.

Central server and high-speed network

A new central server system with high-speed network was installed (Fig. 7). It connects the PF structural biology beamlines and the experimental equipments in the associated laboratories via a Gbit network. The central server is a multi-CPU machine with redundant interfaces (Fig. 8). The server machine is an SGI Altix 3700, which has 16 Itanium2 processors. Each processor has a clock speed of 1.3 GHz, a 3-MB L3 cache and a peak performance of 5.2 Gflops. A total of 32-GB globally shared memory is available. A single SGI Linux operating system controls the machine. This presents an environment resembling a very large Linux workstation for users. An SGI TP9100 RAID storage system provides a 9-TB RAID5 hard disk, which is connected by two 2-Gbit fibre-channel storage area network to the server. This system is easily expandable; the architecture supports as many as 512 processors in the same system. We can increase the CPU, capacity of the RAID hard disks and the interface without any change in the operating system. Lots of numbers of large image data are sent with very high speed (Max 77 MB/sec) into the server from the beamlines. The very large bandwidth is indispensable for such mass transmission without any delay and trouble.

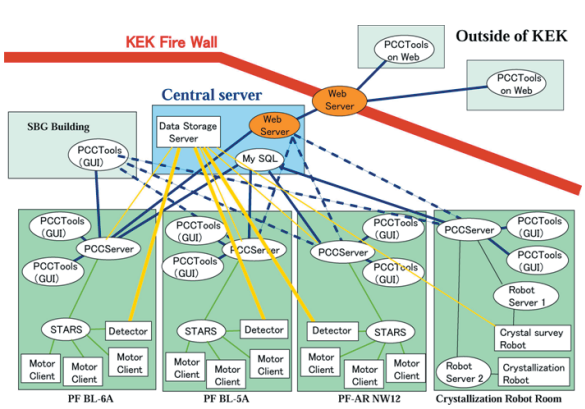

Figure 7
Schematic diagramm of the information network.

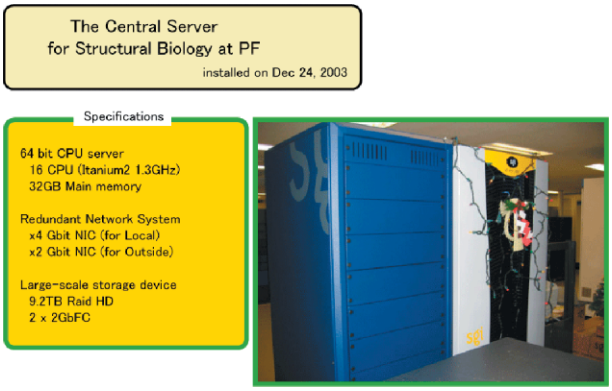

Figure 8
The exterior view of the central server.

2-5 Robotics for High-throughput and Automated Structural Analysis

As part of the high-throughput protein crystallographic analysis of a large number of proteins, we are developing a number of automated systems ranging from protein crystallization to crystal exchange in experimental hutches by using robotics technology.

High-throughput automated protein crystallization system

To circumvent one of the bottlenecks of protein crystallography, we developed a comprehensive screening system (Figs. 9, 10) for protein crystallization conditions that is unrivalled in speed compared with other approach-

Table 3 Specificications of the crystallization robot.

Method Sitting drop vapor diffusion Disposable tips Speed of making crystallization drops 36 sec / plate (96-well) 2,500 plates / day 240,000 drops / day Dispensing volume 0.5 μL + 0.5 μL Scale of one batch of crystallization 80 plates (96-well) 7,680 drops (8 protein samples) (960 crystallization conditions)	Observation speed (Phase 1) 80 sec / plate (96-well) (Phase 2) 5 sec / plate (96-well) Images 640 x 480 pixels (color) Three selectable magnifications Polarized optical system Image data server Storage capacity: 1 TB Linux based WEB server Incubators 1,100 plates / system 350 plates / incubator Temperature 16 – 25 (±0.5) ˚C

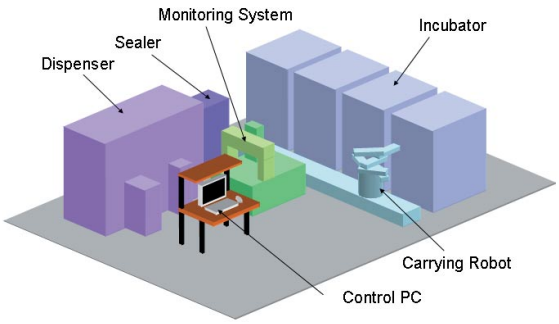

Figure 9
Schematic view of the protein crysallization system.

Figure 10
The exterior view of the protein crystallization system.

Figure 11
A crystal cassette developed by Stanford Synchrotron Radiation Laboratory.

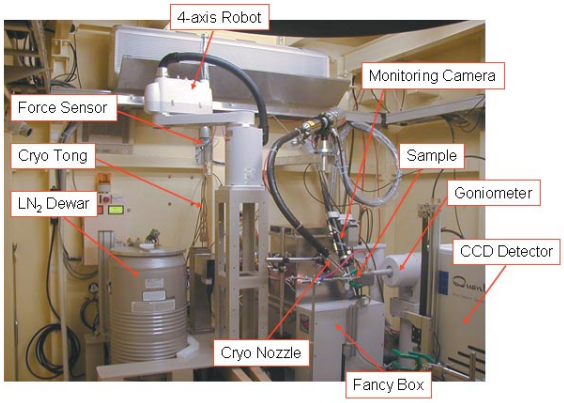

Figure 12
Crystal exchange robot installed on BL-5.

es (for specifications, see Table 3). The system can prepare over 200,000 crystallization conditions per day. Users can view images of the crystallization drops remotely using a WEB based viewing program and search quickly for appropriate crystallization conditions for a variety of proteins. In addition, to automate the process of harvesting protein crystals from the crystallization trays for X-ray diffraction experiments, we are currently working towards the development of a system for handling protein crystals, and have successfully built a remote control system that facilitates the capture of protein crystals by the use of loops and a micromanipulator.

Automated crystal sample exchanging system

The new high flux beamlines, BL-5 and AR-NW12, have significantly reduced exposure times of X-ray protein crystallographic experiments at the PF. Combined with fast readout CCD detectors, a high quality dataset can be collected within 10 to 30 minutes for 180 degree oscillations. This means each user group can collect 20 to 30 high quality X-ray diffraction datasets during 24 hours of beam time on these beamlines. To further improve the overall efficiency, it is necessary to reduce the time required for manual operations of the other steps as much as possible, notably crystal exchange in the experimental hutches. To this end, we have adopted an automated system developed by Stanford Synchrotron Radiation Laboratory for exchanging loops containing protein crystals from a liquid nitrogen Dewar, mounting it on the diffractometer, and centering the crystal. We selected this system based on its reliability with a 2 to 3 years of track record of user operation, large capacity (continuous operation of 286 crystals using three cassettes with 96 crystals each, Fig. 11), and compatibility with commercially available cryo-loops. In addition, we modified the Stanford system from single cryo-tong to double cryo-tong in order to reduce the number of trips between the Dewar and the diffractometer by a factor of 2. We have developed two exchange robots with double tongs and plan to install them on BL-5 (Fig. 12) and NW12 beamlines during the summer shutdown of 2004 for user operation in October. These robots will enhance the efficiency of the beamlines from 20 to 30 data sets a day to well over 100.

References

[1] T. Shiba, H. Takatsu, T. Nogi, N. Matsugaki, M. Kawasaki, N. Igarashi, M. Suzuki, R. Kato, T. Earnest, K. Nakayama and S. Wakatsuki, *Nature*, **415** (2002) 937.
[2] T. Shiba, S. Kametaka, M. Kawasaki, M. Shibata, S.Waguri, Y. Uchiama and S. Wakatsuki, *Traffic*, **5** (2004) 437.
[3] T. Shiba, M. Kawasaki, H. Takatsu, T. Nogi, N. Matsugaki, N. Igarashi, M. Suzuki, R. Kato, K. Nakayama and S. Wakatsuki, *Nature Struct. Biol.*, **10** (2003) 386.
[4] T. Nogi, Y.Shiiba, M. Kawasaki, T. Shiba, N. Matsugaki, N. Igarashi, M. Suzuki, R. Kato, H. Takatsu, K. Nakayama and S. Wakatsuki, *Nature Struct. Biol.*, **9** (2002) 527.
[5] S. Kakuda, T. Shiba, M. Ishiguro, H. Tagawa, S. Oka, Y. Kajiwara, T. Kawasaki, S. Wakatsuki, and R. Kato, *J. Biol. Chem.*, **279** (2004) 22693.