typically express various truncations of the N- and C-terminus and/or internal loop deletions. Precise truncations are defined with the results of bioinformatics analyses of target protein sequences and/or by experimental domain mapping via limited proteolysis combined with mass spectrometry.[24] For a typical target of unknown structure, a minimum of twenty to thirty constructs are prepared in multiple expression vectors (encoding N- and C-terminal hexahistidine tags and a removable N-terminal hexahistidine/Smt3 tag). Well-expressed, soluble versions of the target protein are purified in parallel and then tested for crystallizability, using a predetermined set of ~1,000 crystallization conditions at two temperatures (4° and 20°C). Rapid, fine sampling during the early stages of the process allows us to express the right truncated form(s) of a difficult target that are amenable to crystallization, thereby enabling structure determination.

Once an initial crystal structure is obtained, additional experiments are conducted to enable the target for fragment screening. This process encompasses the transition from small-scale crystal growth and data collection, required to determine a de novo crystal structure, to a robust large-scale process for x-ray screening. Typical requirements for crystallographic screening include the ability to routinely produce and soak crystals on a large scale (~300 diffraction quality crystals/screen) and to obtain reproducible diffraction data to better than 2.5Å resolution. In most cases, the crystal form used for de novo structure determination suffices for crystallographic screening of our core fragment library. In extreme cases, the process may require using information from the initial structure to engineer a new crystal form. Such protein reengineering may be necessary to improve crystal stability [particularly in the presence of dimethyl sulfoxide (DMSO)] and/or the packing of target molecules within the crystal lattice (to permit fragments to diffuse through solvent channels within the crystal and reach the enzyme active site).

After obtaining a suitable crystal form, the system is validated by soaking "control" compounds known to bind and/or inhibit the target of interest. In the absence of reference inhibitors, substrate analogs, cofactors, or other known ligands (i.e., ATP analogs and staurosporine for protein kinases) serve as controls. If the reference compound(s) is readily visible in difference electron density maps, the soaking system is considered validated. After validation of the crystal form, the ability to soak mixtures into the system is tested. In some cases, crystallization and/or soaking conditions must be further optimized to permit efficient soaking of mixtures.

Fragment x-ray screening

Once a target is enabled for crystallographic screening, crystals are prepared for data collection at our dedicated x-ray beamline at the Advanced Photon Source (APS; SGX-CAT). Duplicate crystals are soaked with mixtures of ten structurally dissimilar fragments (typically with each fragment present at ~5–10 mM, DMSO concentration ~5%), flash-frozen, and stored in liquid nitrogen. All experiments are tracked within the SGX LIMS system, which is accessible from SGX-CAT. Direct T3 line connectivity permits rapid data transfer between the two SGX facilities. Once the frozen crystals are transported to SGX-CAT by courier, pertinent sample information is accessed from the SGX LIMS system and the samples are manually loaded into data collection carousels. Multiple data collection carousels are then stored in liquid nitrogen and queued for automated data collection. When a carousel is ready for analysis, it is automatically transferred from the storage dewar to the crystal mounting robot. Figure 3.3 shows the SGX-CAT x-ray diffraction facility on the 31-ID beamline at the APS, which includes x-ray optical elements (for focusing and wavelength selection), beam carriage tubes, a crystal mounting robot, cryogenic nitrogen gas stream for crystal cooling, and a MarCCD detector. To facilitate unattended data collection, crystal centering software was developed by SGX in conjunction with Mar Research.

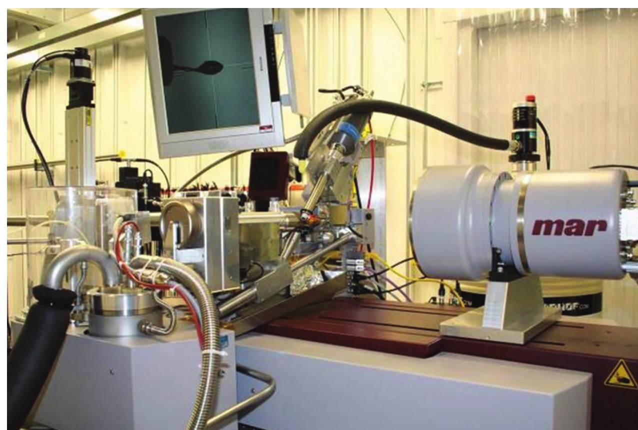

Figure 3.3. SGX-CAT beamline data collection apparatus at the Advanced Photon Source (APS). Shown are the x-ray beam carriage tubes, the cryogenic gaseous nitrogen stream, the sample stage, the Mar sample stage/automated sample changer, and the Mar CCD x-ray area detector.

Data collection/processing parameters are retrieved from the SGX LIMS system to control both the progress of the diffraction experiment and data processing in real time. Reduced diffraction data are automatically transferred back to SGX headquarters in San Diego via the T3 line and experimental parameters are captured by the SGX LIMS database. This system permits routine, unattended data collection from approximately fifty crystalline samples per day, enabling data acquisition for the entire SGX fragment library in about three days (recording diffraction data from the better of each pair of duplicate soaked crystals). Fragment screening results are analyzed automatically using a multi-CPU linux cluster located at SGX San Diego. Automated processing of diffraction data is performed using a system that combines proprietary SGX software and the CCP4[25] program package. For each screening attempt with