RESEARCH
@INDIANA

INDIANA UNIVERSITY · PURDUE UNIVERSITY · UNIVERSITY OF NOTRE DAME

The Xport Collaboratory for X-ray Crystallography

RESEARCH @INDIANA
INDIANA UNIVERSITY

Determining the structure of large biological molecules by X-ray crystallography is made possible in the US through several synchrotron X-ray sources, notably the Advanced Light Source (ALS) at the Lawrence Berkeley National Laboratory and the Advanced Photon Source (APS) at the Argonne National Laboratory.

A critical problem is access to and management of beam time at these national high-brilliance X-ray sources. *Xport*, a DoE NGI funded project at Indiana University, Argonne National Laboratory, and the Lawrence Berkeley National Laboratory targets revolutionary improvements in telepresence for major scientific instrumentation systems to improve throughput and user access.

Our goal is to exploit a combination of advanced networking, middleware services, and remote instrumentation technologies to achieve interactive "better-than-being-there" capabilities for remote experiment planning, instrument operation, data acquisition, reduction and analysis. These capabilities are being deployed and evaluated at several X-ray crystallography facilities including the ALS, APS and the Indiana University Molecular Structure Center (IUMSC).
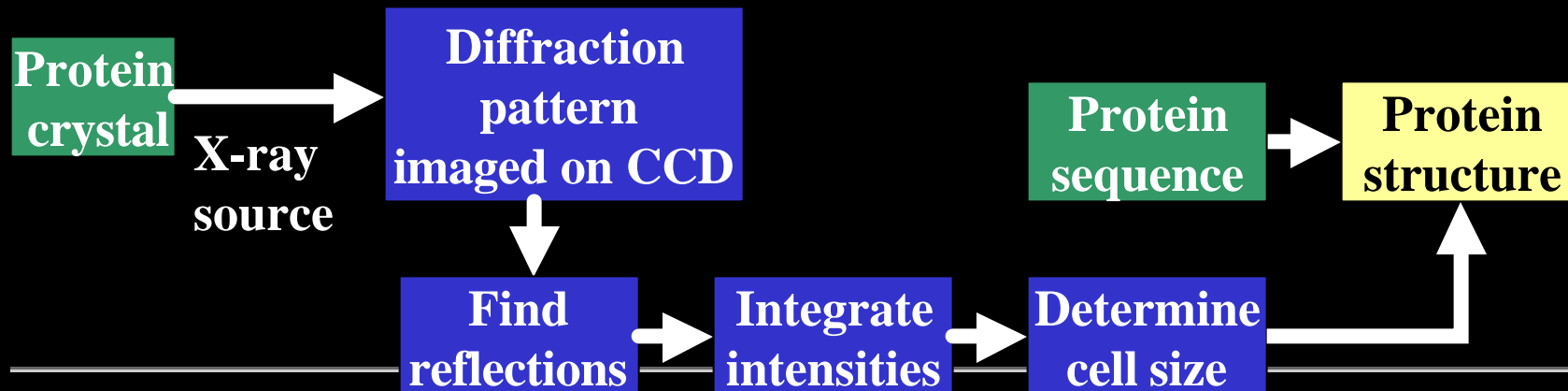
A primary aim of the Xport project is to develop a set of components for high brilliance X-ray crystallography that will simplify the design and execution of macromolecular crystal structure determinations.

The Xport Collaboratory for X-ray Crystallography
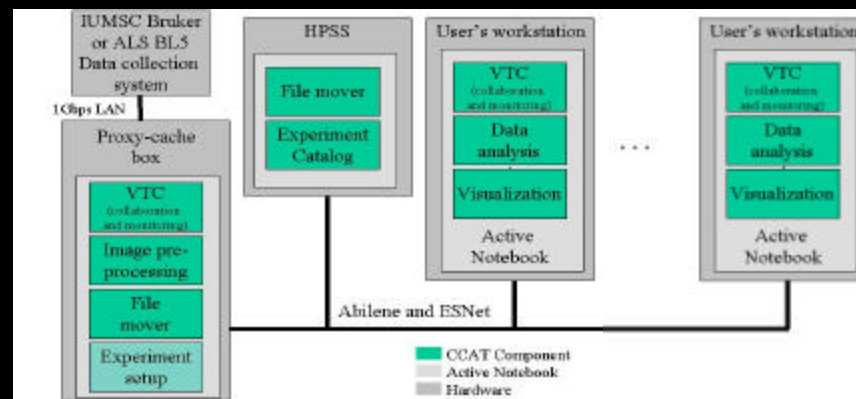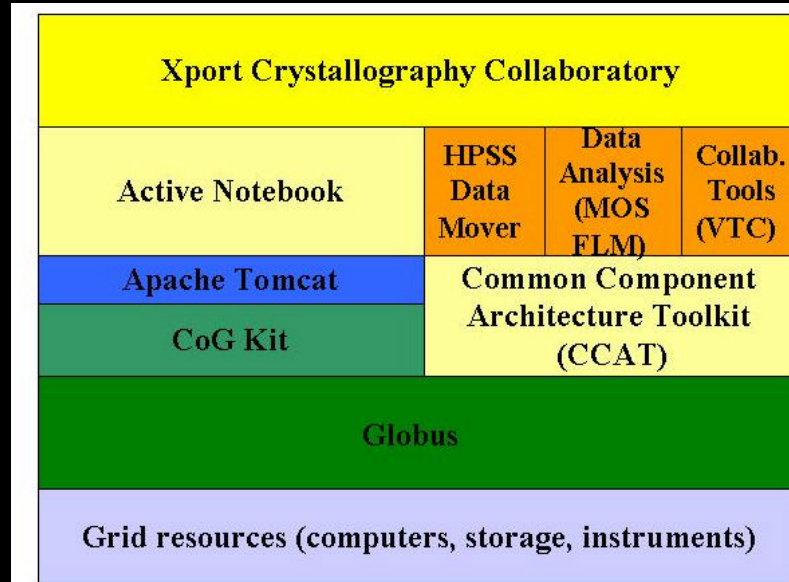
Macromolecular Crystallography

High-intensity synchrotron X-ray sources provide unparalleled spatial resolution for macromolecular crystallography, and have become the standard for large molecule structure determination. In a typical crystallographic experiment a protein crystal at low temperature is placed in a high intensity X-ray beam. The regular arrangement of the atoms in the crystal causes the X-ray beam to be diffracted and as the crystal is rotated in the beam a number of these "reflections" (diffraction spots) can be imaged on a CCD array. The position of the reflections in space as determined by the absolute position of the crystal can be used to determine the unit cell size of the crystal and position of heavy atoms in the cell. This information is used in turn along with known information about the sequence of amino acids in the protein to determine the three-dimensional structure of the protein. High resolution imaging (large CCD images) is needed to provide spatial resolution for macromolecular crystallography. The large data sets from each experiment, often hundreds of images in the 5 to 20MB range, provides challenges for data storage and management as well as distributed processing.

Protein crystal → X-ray source → Diffraction pattern imaged on CCD → Find reflections → Integrate intensities → Determine cell size → Protein structure ← Protein sequence

The Xport Collaboratory for X-ray Crystallography

RESEARCH @INDIANA
INDIANA UNIVERSITY

Xport Architecture

Xport is built on current and emerging middleware standards for grid computing. The user interface is the CCAT Active Notebook. Several plug-in CCAT components provide data management, data reduction and collaboration tools. These components in turn use the grid computing services provided by Globus and the CoG Toolkit for coordinating compute, storage and network services when users access the beamline. The diagram to the right shows the layout of CCAT components on the grid during a typical collaborative session. Users can add their own components to the notebook, and a generic component is used to wrap existing codes. The current implementation uses the crystallography code MOSFLM for data analysis and visualization. Another important aspect is the integration of mass storage into the data acquisition phase. This provides both permanent archiving at the time of the run and universal access by collaborators during and after the experiment.

The Xport Collaboratory for X-ray Crystallography

Testbed and data transport requirements

The APS, the Molecular Structure Center at Indiana University, and the ALS were selected as the initial testbed sites. Data transport requirements for current generation CCD detectors (9-32MB/image every 2 sec.) were estimated at 128Mbps sustained for ~12 h. per experiment. In addition to CCD image transport we expect each experiment to require two or more channels of streaming video and audio at 384Kbps each by which researchers at remote sites can monitor the progress of the experiment and consult with local expertise at the beamline.

Network Quality of Service

A key element of the architecture for this project is the use of a "network storage" paradigm to eliminate file system dependencies in data reduction and visualization component codes. To prevent local cache overflow and stalling of the experiment some level of QoS is needed in the link between instrument and network storage. At present we are characterizing the GARA bandwidth reservation API from Argonne using the Emerge testbed. QoS requirements between functional blocks is indicated in the diagram above. Paths that need some level of QoS (bandwidth guarantees primarily) are indicated in green. Components for data transfer to an HPSS server and for video are being modified to support bandwidth reservation.

For additional information: http://www.cs.indiana.edu/ngi

mcmullen@indiana.edu, bramley@cs.indiana.edu, huffman@indiana.edu