Faster, Better, Cheaper: MacCHESS and High-throughput Crystallography

Marian Szebenyi

Macromolecular Crystallography Division of CHESS, Cornell University

Determination of a macromolecular crystal structure is now, in most cases, routine. Investigators are turning to studies involving determination of a series of structures, such as a single protein with different ligands bound, or homologous proteins from different species. Solution of each structure in the series, although routine, requires time and money. Reduction in these, while maintaining high data quality, increases scientific productivity, and is needed to make some large projects feasible. MacCHESS is engaged in a number of activities to increase crystallographic throughput and decrease users' expenses at CHESS, while maintaining or improving the quality of data collected.

Let's look at the steps involved in solving a series of structures:

1) Produce crystals of the various materials to be studied.
2) Obtain time at a synchrotron source, send crystals and workers to the facility.
3) Set up to do the experiments.
4) Mount a crystal on the camera.
5) Center it.
6) Collect a few images and evaluate crystal quality. If bad, go back to step 4. If acceptable, go on.
7) Collect data set(s) from the crystal.
8) Process the raw images, evaluate quality of the resulting data set.
9) Solve the structure, evaluate quality of the solution.
10) Repeat steps 4-9 for each crystal.
11) Archive data and results.

Steps 8 and 9 need not be completed before proceeding to the next crystal, but it is advisable to do at least part of step 8 while at the beamline, as processing may reveal problems that require recollecting some data.

Step 1 is the responsibility of the user, but developments at MacCHESS have helped, and will continue to help, with all the other steps.

Steps 2-3: Arrival and Set up
The majority of requests for routine data collection time now utilize the Express Mode Proposal system, which allows much more rapid scheduling of beamtime than the CHESS Standard Proposal. Increased use of Web-based user training has allowed users to become familiar with CHESS, and to complete much of the required safety training, before arriving in Ithaca. This reduces the time needed for user check-in and training, and hence the lag time between arrival at CHESS and the start of experiments. "FedEx" mail-in data collection, introduced in the spring of 2002, carries reduction of setup time to an extreme; in this mode, users send only their crystals to CHESS. Data are collected by experienced staff, who don't need to travel to Ithaca and are already trained. Moreover, shipping costs for crystals are generally less than food, lodging, and travel expenses for personnel, so FedEx saves users money as well as time. The expertise of MacCHESS staff, and remote consultation with users when necessary, insures the highest possible quality of data collected using this mode. Users and staff have been well pleased with the operation of FedEx mode over the past several months, and its use is expected to increase in the future.

Step 4: Crystal Mounting
Various crystal mounting techniques have been investigated by MacCHESS personnel, and tools and advice for oil drop mounting, use of extended arc goniometers, and so forth, are available to users. Further optimization of the crystal mounting step will require automation, and a robotic crystal mounting system, similar to that at Lawrence Berkeley Laboratory (reported in a News article in Science, Vol. 292, pp. 187-188 (2001)), is currently under construction; it is scheduled for completion and testing in 2003. This system will allow multiple crystals to be evaluated, and data collected from them, without the need for users to run in and out of the hutch whenever a new crystal is needed.

Step 5: Crystal Centering
In the past, crystal centering has been done manually by the users; CHESS has merely provided the necessary tools: a telescopic video camera focused on the crystal, a flexible light to illuminate it, crosshairs on the video image showing the beam position, a set of goniometer wrenches, and occasional advice (the cover image shows most of these hardware parts in action at the F1 station). Incremental improvements have been made to the crystal-viewing optics, as necessitated by the use of ever-smaller crystals. For the future, automated

crystal centering will be needed to go with automated crystal mounting. Even now, a more automatic method of centering a crystal would speed up the process, reduce the time spent opening and closing hutch doors, and lead to more precise placement of the crystal in the beam. In theory, centering a crystal requires only acquisition of three images at well-separated values of the spindle angle, determination of the position of the crystal in each image, and translation of the crystal to place it on the center of rotation and in the beam. The system must be calibrated in order to map position in the image to position in real space; knowledge of the beam position is necessary to determine crystal placement along the spindle axis but not perpendicular to it. In practice, automatic determination of the crystal's position in an image is often difficult, and we have opted for a semi-automated system in which the user chooses the spindle angles at which to view the crystal and informs the system of the crystal's location in each image.

A prototype crystal centering system was designed and assembled in the summer of 2001. In this system, the user was required to provide instructions as to the crystal's motion, as well as its position, so that the accuracy of crystal centering was about the same as for fully manual operation. Even so, there were advantages in terms of reduced opportunity for a dropped wrench or a bumped beam stop, and fewer hutch openings. The system is shown schematically in Figure 1. A standard goniometer head, bearing a crystal, was mounted on a special rotation stage equipped with two motorized slides for translation perpendicular to the rotation axis. The signal from the standard crystal-viewing video camera was fed into an Axis 2400 video server, which made the video stream available over the Internet. A Java application running on a workstation presented the view of the crystal to the user, along with control buttons for rotating the spindle and translating the crystal. Button presses were sent back to the Axis 2400, which passed commands along to a Compumotor controller for actual motion of the motors. The prototype demonstrated feasibility of the semi-automated system, but deficiencies in operation of the translation stage, and in communication between components, indicated a need for further development. This is now underway; a better translation stage has been obtained, and improved Java code is being written to enhance reliability. In addition, use of a high-resolution digital video camera will improve the quality of the images and reduce the overhead involved in digitizing the video signal.

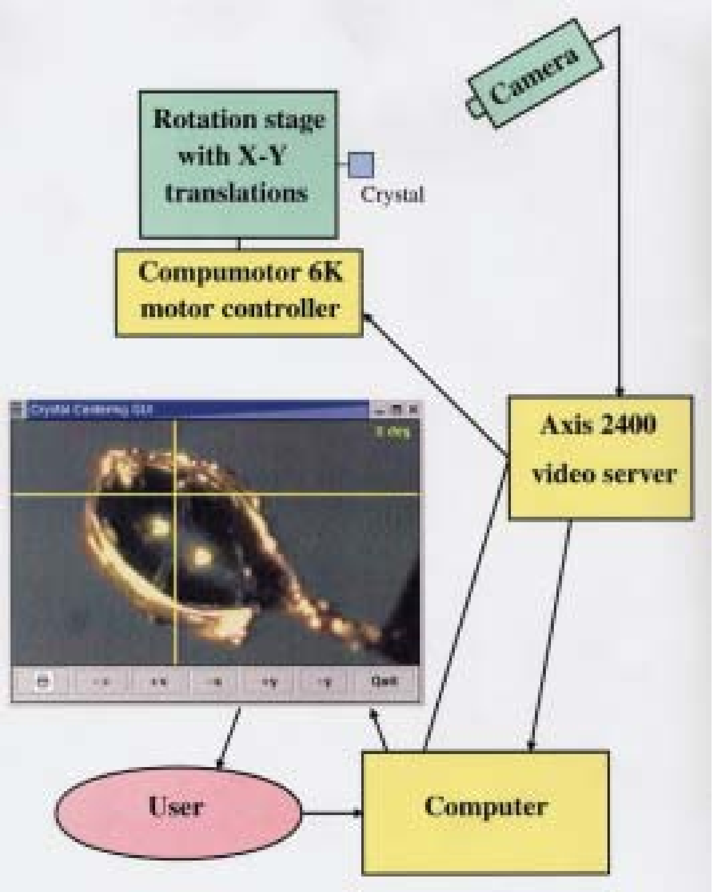


Fig. 1 Schematic of the prototype crystal-centering system developed at MacCHESS. The video signal from the crystal-viewing camera is fed to an Axis 2400 device, which includes a free-standing Web server to export the video stream to the Internet. A workstation runs a Java program to display the image, in real time, in a window equipped with control buttons for rotating and translating the crystal. Commands entered using the buttons are sent back to the Axis 2400, which transmits them over a serial line to a Compumotor controller, which moves the motors as requested. The system was developed by Chris Heaton, Joe Oosterhout, and Richard Gillilan (who provided the image displayed in the GUI).

Steps 6-7: Data Collection

An increase in the speed of taking diffraction images has been obtained by installing CCD detectors (ADSC Quantum-210's) with very short readout times. The large area of these detectors also means that it is seldom necessary to offset a CCD; this eliminates the need for taking extra frames to complete a dataset from an offset detector. A dual-CCD device allows even very high-resolution data to be collected in a single run. This device is shown in Figure 2. It incorporates two Quantum-4 detectors in a unique custom-built housing which allows them to be operated either as a single large flat detector or in a "V" configuration. The latter arrangement reduces the radial elongation of spots at high resolution, due to the oblique incidence of x-rays, thus reducing spot overlap and improving spot profiles. When the large area is not needed, the device is placed so that the bottom detector is centered on the beam, and this detector is used alone. When a larger area is needed, a simple

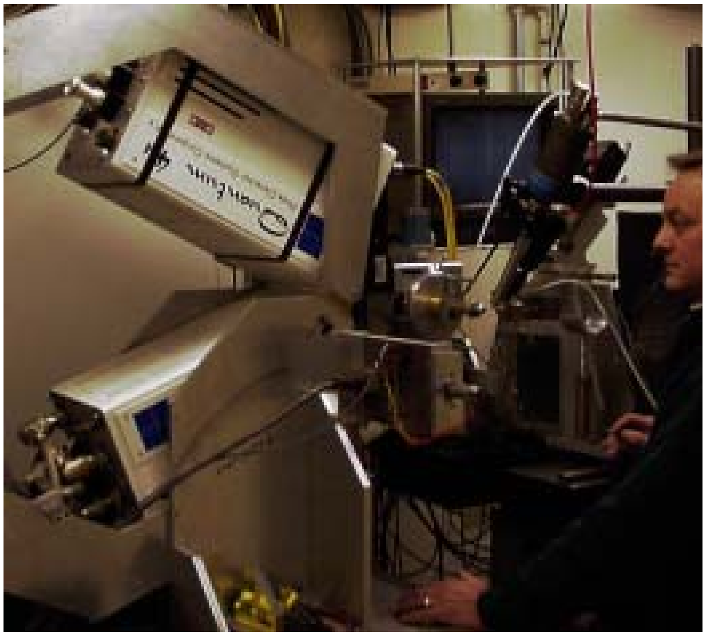


Fig. 2 The dual Quantum-4 detector system installed at F1 station by Bill Miller (pictured) and other MacCHESS staff members. The two detector assembly, designed by Mike Cook, is capable of independent rotation of each detector by up to 30 degrees; the total surface area of the device is 380 x 190 mm.

vertical translation places one detector below the beam and one above. The **ADX** data collection software allows use of either detector alone, or both together, by setting an environment variable. Figure 3 shows a pair of images taken in a single exposure, using the dual-CCD system. If low-resolution data falling in the gap between detectors are required, a second data set must be acquired using a single centered detector, but this set will typically require shorter exposure times. Hence a complete data set can be produced faster with the dual system than with a single offset detector.

In addition to the new detectors, other CHESS/MacCHESS features make data collection convenient and reliable. With the current x-ray optics and feedback system the beam position is quite stable, and when it does shift a single button click realigns the oscillation camera with the new position. A Compumotor controller is used to operate the shutter and spindle motor for oscillations; this dedicated device is very reliable and (in conjunction with a DC spindle motor) supports a very wide range of oscillation speeds. An automated liquid nitrogen filling system for the crystal cooling apparatus eliminates the time previously needed to manually fill Dewars. Upgrades to computers and networks, and the addition of massive amounts of RAID storage, have made it easier to deal with the many gigabytes of data produced. Improvements to the **ADX** data collection software include better reliability, easier restarts of interrupted data collection, and increased flexibility. An enlarged MacCHESS staff has reduced the time needed to solve problems arising during data collection. Finally, improvements to the user environment, such as reduced noise, more desk space, and better chairs, have made users more comfortable and (we hope) more productive.

Step 8: Processing of Raw Images

Rapid processing of the initial data obtained from a crystal is important, in order to evaluate crystal quality and avoid collection of useless data sets. When exposures are long or beam time is short, immediate processing also allows determination of the best strategy for data collection, taking into account the crystal symmetry (which is not always known initially for the particular crystal being studied), mosaicity,

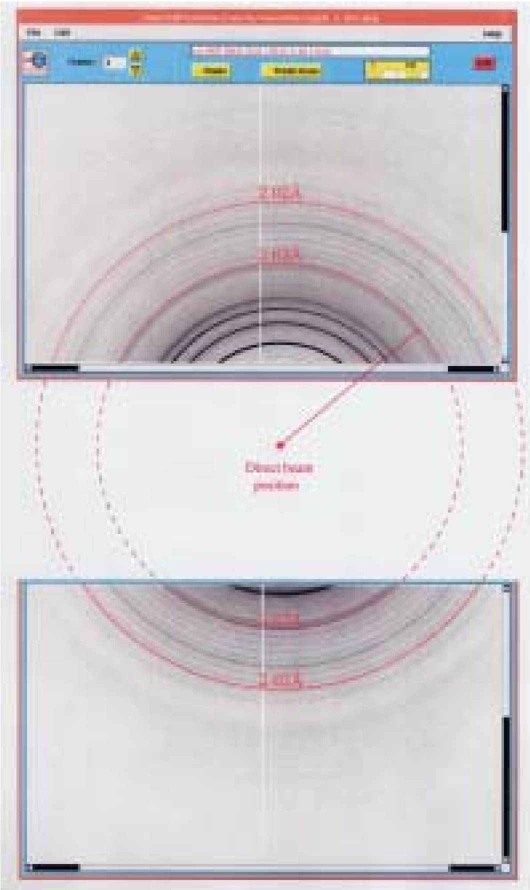


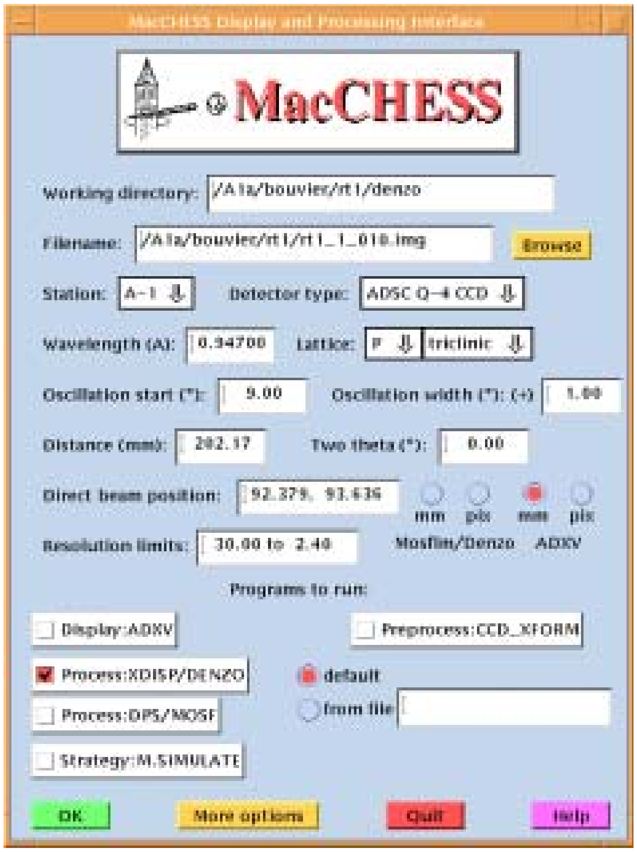
Fig. 3 A pair of images from a sample of silver behenate powder, taken using the dual-CCD detector with the two CCD's configured to form a large flat surface. Only the lower resolution portions of the images are shown (note scroll bars). These images were used to determine the direct beam position: the red circles generated about the estimated beam position were interactively adjusted to match observed diffraction rings; the center of the circles then defines the beam position. In addition, the calculated resolution of strong rings, in comparison with the known diffraction peaks from silver behenate, provides a check on the sample-to-detector distance. The direct beam position obtained by this method was used to successfully process a series of image pairs taken on a lysozyme crystal. When the detectors are tilted, diffraction rings become ellipses rather than circles, but the same basic method may be used to determine the beam center; the **dps_display** program supports the display of such ellipses.

and orientation. To make data processing fast and easy, it is important to (1) automatically provide as much information as possible to the software, while allowing for user verification of this information, and (2) make available software that the users are accustomed to work with.

The **mccview** interface, shown in Figure 4, fills this need. Selection of the proper CHESS station sets the correct detector, direction of spindle rotation, and default wavelength. Once the user has selected one of the files in the current data set (which may be a single snapshot or a complete series) using the *Browse* button, clearing a field (distance, beam position, etc.) and typing a C/R fills in the value from the image file header. Other basic parameters, such as beam polarization, are not visible on the interface but are set to appropriate values for CHESS. The data processing software to use may be set to either **denzo/xdisp** (the HKL package) or **DPS/Mosflm**; most users are familiar with one or the other of these. Recently, the **XDS** data processing program has been installed at CHESS as a third option, with a **WebXDS** graphical interface designed by Art Weaver; an **XDS** selection may easily be added to **mccview** in the future. Not yet available, but under development, is an automatic data processing option, which will be activated from the data collection program, to perform preliminary data reduction using only the information available from **ADX** and the images themselves. This software is intended to keep the user informed about the status of the current data set (completeness, redundancy, R_{sym}, etc.) and warn of processing difficulties.

Step 9: Structure Solution

To be useful, a data set must have both good internal quality (low R_{sym}, high completeness and redundancy, sufficient resolution; evaluated by the initial processing discussed in the preceding section) and lead to an interpretable molecular structure. With the current short data collection times and availability of excellent software, it has become feasible to evaluate a data set in terms of its usefulness for structure solution, as well as its internal consistency, at CHESS. For this purpose, standard crystallographic software (including the **CCP4** suite, **CNS**, **Shake-n-Bake**, **SOLVE** and **RESOLVE**, **O**, and others) is available on fast Alpha and Athlon computers. Some programs in the **CCP4** suite, including **SAPI**, **OASIS**, and **FSEARCH**, were written by Quan Hao, the Assistant Director of MacCHESS; support for these routines is particularly extensive.



Working directory: /A1a/bouvier/rt1/denzo
Filename: /A1a/bouvier/rt1/rt1_1_010.img
Station: A-1
Detector type: ADSC Q-4 CCD
Wavelength (A): 0.94700
Lattice: P triclinic
Oscillation start (°): 9.00
Oscillation width (°): (+) 1.00
Distance (mm): 202.17
Two theta (°): 0.00
Direct beam position: 92.379, 93.636 ○ mm ○ pix (Mosflm/Denzo) ● mm ○ pix (ADXV)
Resolution limits: 30.00 to 2.40

☐ Display:ADXV
☐ Preprocess:CCD_XFORM
☑ Process:XDISP/DENZO
● default
○ from file
☐ Process:DPS/MOSF
☐ Strategy:M.SIMULATE

Fig. 4 The **mccview** graphical interface, ready to start processing. As soon as the user clicks the OK button, **denzo** and **xdisp** will be started up using the information visible in the interface plus appropriate defaults for CHESS A-1 data. To index the image, it is then only necessary to do a peak search in the **xdisp** window and type 'go' in the **denzo** window.

Steps 10-11: Cycling through Crystals, Archiving Data and Results

All the improvements noted above have contributed to increasing throughput at CHESS and allowing the collection and processing of multiple data sets during a user visit. Options for making final disposition of all that information have also expanded; a user now has a choice of four types of tape backup, transfer to a FireWire or other type of hard disk temporarily attached to a MacCHESS computer (several types of computers are available for this purpose), transfer to a laptop computer temporarily plugged in to the CHESS internal network, or transfer to a remote computer over a 100 MB Ethernet connection. Moreover, the installation of several terabytes of RAID storage means that data can be maintained locally for months, in case of difficulties in transfer or archiving.

In summary, CHESS and MacCHESS now provide good facilities for high throughput crystallography. Future developments in hardware and software, and the eventual conversion of CHESS to a dedicated machine, will only improve this capability.