Lawrence Berkeley National Laboratory

Recent Work

Title
Web-Ice: Remote Access and Crystal Screening at the Berkeley Center

Permalink
https://escholarship.org/uc/item/0258z82f

Author
Gonzalez, Ana

Publication Date
2006-10-11

Web-Ice: Remote Access and Crystal Screening at the Berkeley Center

Synopsis: *Web-Ice*, a browser-based tool for characterizing macromolecular diffraction, is now available at beamlines 5.0.1, 5.0.2, and 5.0.3 of the Berkeley Center for Structural Biology (BCSB). Although originally prototyped at SSRL and specifically customized for the data-collection environment there, the application is organized so its main functions can be ported to other facilities with minor modifications. Many users are familiar with the beamlines at both SSRL and BCSB, and they will notice both similarities and differences in the two *Web-Ice* installations. The purpose of this primer is to give a general overview of *Web-Ice*, with particular emphasis on the BCSB implementation. Although we describe the software's current state as of October 2006, continuing improvements are expected in the coming months.

1. Program description

Web-Ice enables high-throughput crystal screening by automating the repetitive work of image analysis. It's role in data analysis is that of an application server, *i.e.* it permits the display of raw data and processing results, but delegates the actual calculations to underlying programs. *Web-Ice* employs *DISTL* to count and characterize candidate Bragg spots. The Bragg spot positions are used by *LABELIT* for autoindexing. Based on the probable Laue symmetry, *MOSFLM*, *BEST* and *RADDOSE* are scripted to determine proper φ-rotational ranges, detector distances, and exposure times needed to acquire a complete data set. Since the user interface is through a Web browser, it is possible to view information over any Internet connection. It is important to note that *Web-Ice's* role is completely passive, at least at BCSB. Data can be analyzed by remote access, but data acquisition is still controlled locally by the Beamline Operating System (*BOS*) at each endstation. A complete *Web-Ice* manual is posted on the SSRL Web site (see section 5).

2. The customized Web-Ice environment at BCSB

Covered beamlines. *Web-Ice* is installed at the three beamlines equipped with sample-handling robots: 5.0.1, 5.0.2, and 5.0.3.

Software location. Primary access to *Web-Ice* is through the following URLs:

Beamline	URL
5.0.1	http://bscb-ra1.als.lbl.gov/webice
5.0.2	http://bscb-ra2.als.lbl.gov/webice
5.0.3	http://bscb-ra3.als.lbl.gov/webice

These URLs are listed on the main BCSB Web page (bcsb.lbl.gov):

Beamline-specific data location. Beamline users typically write newly-acquired data to the following disk storage location:

> **/data2/dcsuser**

It is important to realize that this link points to different physical storage disks at each beamline. BCSB *Web-Ice* obeys these same disk-access rules: for example, to view data at Beamline 5.0.1, it is necessary to log in to the bcsb-ra1 server, and navigate to the /data2/dcsuser directory. It is generally not possible to cross-mount disks, *i.e.,* one cannot view Beamline 5.0.2 data from the bcsb-ra1 server. This is one reason why there are three separate BCSB *Web-Ice* servers[1]. SSRL *Web-Ice* is different because that facility has a single overarching disk storage namespace.

User accounts at BCSB. BCSB *Web-Ice* is linked to the same user-authentication database harnessed for general data processing at BCSB. Therefore there is no separate setup for *Web-Ice* accounts; you simply type in your normal unix username and password into the *Web-Ice* login prompt. This information is SSL-encrypted en route to the *Web-Ice* server, which then creates "child processes" on behalf of your own user account for data processing. If you keep your browser open, the session will time out after 20 hours of inactivity. *Web-Ice* data processing results are stored in your own home area in a special directory:

> **/home/useraccts/<username>/webice**

It is generally a good idea to avoid altering these directory contents (but see section 4 below for an important exception). User-specific preferences for your *Web-Ice* account are stored in this file:

> **/home/useraccts/<username>/.webice**

…these preferences can be edited from the "Preferences" tab within *Web-Ice*. Preferences can be reset to default values by deleting the ".webice" file and logging back in to *Web-Ice*. An important detail: although the /data2/dcsuser directory is specific for each Web server, the /home/useraccts mapping always refers to the same physical disk. Therefore, although the *Web-Ice* session will only be able to display the raw images acquired on one beamline, data processing results can be shown for projects conducted on all three beamlines. A final note: *Web-Ice* never alters the original data or writes any results to the /data2/dcsuser area.

Data organization paradigm. Differences in the way crystal samples are managed have caused BCSB and SSRL to develop different *Web-Ice* views of the crystal screening experiment. At SSRL, the high-throughput user arrives at the beamline with a 96-sample cassette, along with an Excel spreadsheet listing the cassette contents. The spreadsheet is uploaded into a *Sample Information List* (SIL) viewable in the Screening tab of SSRL *Web-Ice*. Rows in this table are labeled A1, A2,…,L8 to directly reflect positions in the cryocassette. (SSRL *Web-Ice* also supports the uni-puck format.) At BCSB there is a different sample-puck layout; also there is not a general facility to upload Excel spreadsheets. Therefore, BCSB *Web-Ice* takes a different

---

[1] The other reason pertains to load balancing.

approach to grouping samples together. BCSB users can take either of two paths to process a crystal screen:

1. *The Directory List.* In this paradigm, data are analyzed after the entire data acquisition process has been completed. In the Screening tab of BCSB *Web-Ice*, one clicks the [Directory List] button to access a directory browser:
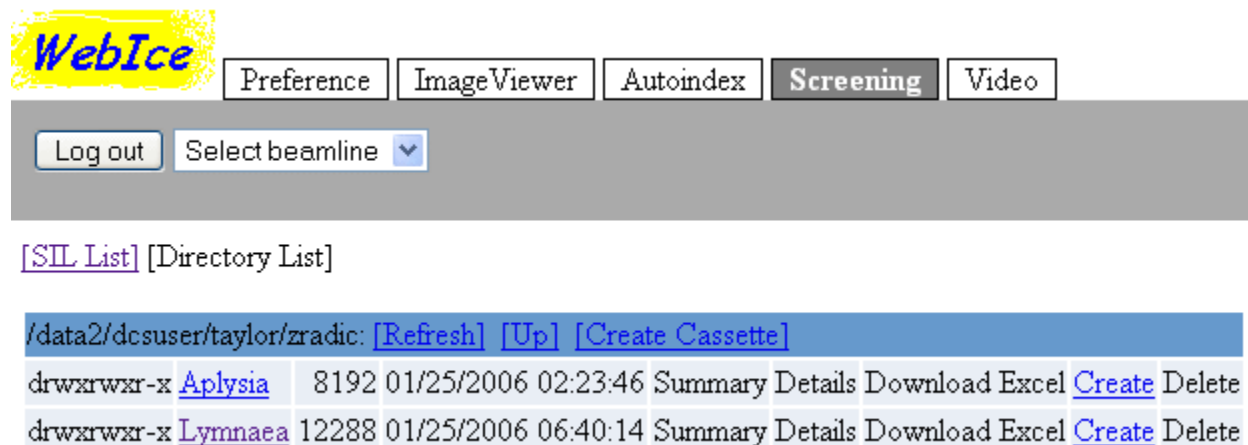
Clicking the "Create" button next to a directory name will create a new SIL and analyze the screening results. An important detail for the example shown:
- Clicking the "Create Cassette" button on the blue bar will create a SIL for the directory /data2/dcsuser/taylor/zradic.
- Clicking the "Create" button on the bottom line will create a SIL for the subdirectory /data2/dcsuser/taylor/zradic/Lymnaea.

SILs are created using the following procedure. A recursive file listing is produced for the chosen directory, as well as two levels of subdirectories. It is assumed that all data of interest will lie in this directory tree. A lexical analysis of filenames then determines which files appear to be grouped into datasets. For example, *lyso_1_001.img* and *lyso_1_002.img* are grouped together into one dataset, but *lyso_0_001.img* is classified as belonging to a separate dataset. Datasets are then examined to find the phi-rotational settings of each image. If possible, two images are chosen having goniometer settings as close as possible to 90° apart, which is the ideal input for autoindexing. Each pair of images is then entered into a new row of the SIL, and autoindexing is performed. The SIL rows in BCSB *Web-Ice* will simply be numbered 1,2,3,…; *i.e.,* not necessarily in order of puck location[2]. In the future BCSB will add features to optionally keep track of puck location, and to optionally add user comments to the table.

Once the new SIL is created, the user can return to it either by browsing again within the [Directory List], or by clicking on it directly in the [SIL List].

2. *Auto-triggered analysis.* In this paradigm, data are analyzed in real time during the acquisition process. Although the SIL is still viewed within BCSB *Web-Ice*, instructions for

---

[2] A1,A2,… numbering will be switched to 1,2,3,… numbering in Nov 2006.

creating the SIL are sent by *BOS*. This is analogous to the situation at SSRL, where the beamline graphical interface (*BLU-ICE*) sends signals to SSRL *Web-Ice* to trigger crystal analysis. At the time of this writing, details are still being worked out, and it is possible that the auto-triggered analysis paradigm will not be fully functional until Nov 2006. However, here is the vision: The *BOS* user will provide a username and password within the *BOS* graphical interface (this is a new feature at BCSB). A SIL name can be selected from a list of existing SILs, or a new SIL can be created. Then, all screening datasets subsequently acquired will be entered into the selected SIL in chronological order (until a new SIL is specified, or until the auto-trigger feature is turned off).

Strategy calculation. BCSB *Web-Ice* will be configured so that Strategy is calculated for all crystals in a SIL. The results can be viewed by clicking the View Strategy button in the Screening tab:

SSRL *Web-Ice* does not automatically trigger strategy calculation. In SSRL's view, the purpose of the screening experiment is to rapidly transfer many cyro-samples to the goniometer, collect test images, and then replace samples in to the storage Dewar. Once the best crystal is selected for further study and placed back on the the goniometer stage, the original orientation of the sample loop may have slipped significantly, so the phi rotation range calculated originally will have to be repeated anyway (new test images will need to be acquired to get the new orientation). BCSB takes a different viewpoint: in our experience, crystal slippage upon repeated transfer is minimal (~1°), so predicted phi ranges are not that bad. Furthermore, other values derived in the strategy calculation (optimal detector distance and exposure time) will not be affected by slippage. Finally, the ability to derive a suitable set of data collection parameters is in itself a criterion in the selection of the best crystal. So the user may want to see this analysis carried through, even if it has to be repeated prior to full data collection.

Having said this, the full details of Strategy calculation have not been implemented in BCSB *Web-Ice* at the time of this writing. *BEST* and *RADDOSE* have not yet been linked in with the server. Also, beamline-specific values (allowable ranges for detector distance, energy-dependent

beamline flux, detector size, etc.) still need to be programmed in.   Implementation will be complete by Nov 2006.

Image annotation. BCSB *Web-Ice* will be configured so that *DISTL* image characterization always runs in the background within the ImageViewer tab. In this example, *DISTL* marks Bragg spots yellow for display (spots having multiple maxima are coded white), and it also gives the table showing total spots, ice rings, and estimated resolution:

Here is what this site-wide default gains us: If the "Show spot overlay" box is checked, and the user clicks the "Next" button to advance to image #002, *DISTL* is automatically run, so that the new image can be properly marked up with Bragg spots. Therefore, the user can advance through the entire dataset without specifically asking for *DISTL* to run each time. This ability is determined by the "Automatically Analyze Image = true" setting in the Preference tab. When set to "false", the user needs to click on the "Analzye Image" button each time to run *DISTL*.

SSRL *Web-Ice* defaults to "Automatically Analyze Image = false" for two reasons. First, there is a noticeable lag time for running *DISTL*. Secondly, the markup information, written into the user's SSRL /data directory, currently takes up as much disk space as the original image.

In BCSB's view, these technical problems will eventually go away as the software improves. The software will be re-engineered to run in a shorter time and will cache results differently so as to require less disk space. The improvements can be expected sometime in 2007.

Features not present at BCSB. Since the *Web-Ice* package was developed at SSRL, certain features simply do not work in BCSB *Web-Ice*, even though there are corresponding buttons. These functions include the "Crystal Image" from the goniometer microscope, the hutch "Video" stream, the import of Excel spreadsheets into *Web-Ice*, and the automatic execution of data collection from within the Autoindex tab. There are no specific plans to implement these features at BCSB. There is currently an ability to *export* an Excel spreadsheet based on SIL data, but this feature will probably need improvement.

3. *A quick tour*

It is recommended that the user consult the *Web-Ice* home page (section 5) to see the full manual. But here are some not-to-be missed features:

Image Viewer. As shown in section 2 above, the image viewer can navigate through an entire dataset ("Prev" and "Next" buttons) with the same view settings applied to each image. The directory browser at the bottom can traverse the entire file system.

In the Screening tab, once a particular SIL is selected (with the "Summary" button shown in section 2) the "Cassette Details" button will again go to an image viewer. However, within the Screening tab, the bottom panel is a navigation bar to traverse all the datasets in the SIL. Therefore, diffraction from different crystals can be compared to help find the best samples.

Autoindex. This tab is a way to get some quick indexing results without going through the formality of a multi-crystal screen and a sample information list. Thus, once a single crystal is chosen for study (or if there is only one crystal) this is the entry point. A file browser is used to select two images for indexing; and then after a few more mouse clicks the autoindexing results are generated. In certain cases the Laue symmetry (or even the Bravais lattice) cannot be determined from the unit cell measurements alone. For such cases the autoindexing input permits you to choose a specific symmetry and unit cell:

…in this example it is useful to specify a monoclinic cell because the near-90° value of the β angle otherwise causes *LABELIT* to identify orthorhombic symmetry.

Data collection strategies are normally calculated for the triclinic setting (P1) and the highest-symmetry setting. If an intermediate-symmetry strategy is desired, wait until autoindexing is

finished, then go back to the "Setup" button and fill in the "Integrate other solutions" checkbox. In cases where the Laue setting is ambiguous (4/m vs. 4/mmm), both strategies are calculated, and there is a menu item to select which to view. The "Details" button provides a directory-view of the full processing results, as written by *DISTL*, *LABELIT*, and *MOSFLM*.   Of special value are the *LABELIT* output files "`integrationxx.csh`", which are command scripts to start an interactive session for *MOSFLM* integration. All of these files are in the user's unix area in the directory

        `/home/useraccts/<username>/webice/autoindex`

Screening. For the new user, it is important to realize that moving within the Screening tab is a matter of viewing information at three levels. This information hierarchy is not really shown well in the browser, but here are the essentials:

The top level is the [SIL List] | [Directory List] page, which shows all possible SILs ("cassettes" in SSRL terminology, or collections of screened crystals) or allows browsing the file system to find directory-based SILs.

The next level is the "Cassette Summary" view showing all crystals in a particular screen; with tabulated results from spot picking and autoindexing. Note: at this level, the "Analyze Crystal" button starts autoindexing all over again for the selected crystal; so don't push this button unless you mean it. Also, the "Refresh" button is *absolutely essential* here. Calculating the screening results takes a lot of time no matter what paradigm you are using (see section 2); so you will probably spend a lot of time at this level clicking on "Refresh".

The lowest level is the "Cassette Details" view, which is the special image viewer for individual SILs.

4.   *Frequently asked questions*

#1 Q: I've filled up the entire disk! What's going on?
   A: Yes, *Web-Ice* generates many data processing files in your own home directory, and disk space management is left to the user. There is a subtlety to this: Every Autoindexing result and Screening SIL is given a serial number or identifier. It is easy to find the corresponding subdirectories in your home area:

        `/home/useraccts/<username>/webice/autoindex/lyso1`
        `/home/useraccts/<username>/webice/screening/733`

The way it is now, autoindexing cases can be deleted within the *Web-Ice* browser, and the corresponding subdirectory will be deleted from your home area. However, deleting a SIL will only delete the SIL in *Web-Ice's* internal database, not in your home area. But to its credit, *Web-Ice* does give a screen message telling you what subdirectory you need to remove.

#2 Q: I went to the Image Viewer but couldn't navigate to see some images I stored in my home directory in /home/useraccts?

A: It looks like an automounting problem. Try doing something else first, like editing your Preferences or viewing some Screening results. Then your home directory will be mounted. We'll try to fix this soon.

#3 Q: My crystal is pseudo-cubic, with the true symmetry being rhombohedral. I want to use Web-Ice to calculate the phi-rotation strategy, so I tried entering the known Laue group and unit cell in to the Autoindex run setup. This failed. Autoindexing also failed when I gave just the Laue group. How can I get the right strategy?

A: In many cases entering the known symmetry and cell is a good way to handle pseudo-symmetry. However, this is a very special case. As shown by the *LABELIT* indexing log file, your crystal can be indexed in four possible rhombohedral settings, depending on which body diagonal of the primitive cell is taken as the threefold:

```
Solution Metric fit rmsd  crystal_system  unit_cell                                                 volume
:)  22    0.216 dg 0.120         cubic cP  112.6 112.6 112.6   90.0   90.0   90.0  1426035
:)  21    0.216 dg 0.118 rhombohedral hR   159.1 159.1 195.2   90.0   90.0  120.0  4277918
:)  20    0.216 dg 0.121 rhombohedral hR   159.1 159.1 195.1   90.0   90.0  120.0  4278512
:)  19    0.216 dg 0.123    tetragonal tP  112.6 112.6 112.5   90.0   90.0   90.0  1426059
:)  18    0.216 dg 0.120 orthorhombic oC   159.0 159.4 112.5   90.0   90.0   90.0  2852003
:)  17    0.216 dg 0.118   monoclinic mC   159.4 159.0 112.5   90.0   90.0   90.0  2851867
:)  16    0.201 dg 0.122    tetragonal tP  112.6 112.6 112.6   90.0   90.0   90.0  1426129
:)  15    0.201 dg 0.122    tetragonal tP  112.5 112.5 112.6   90.0   90.0   90.0  1426093
:)  14    0.201 dg 0.122 orthorhombic oP   112.5 112.6 112.6   90.0   90.0   90.0  1426093
:)  13    0.201 dg 0.121   monoclinic mP   112.5 112.5 112.6   90.0   90.2   90.0  1425961
:)  12    0.189 dg 0.115 rhombohedral hR   159.1 159.1 195.1   90.0   90.0  120.0  4278121
:)  11    0.189 dg 0.119 orthorhombic oC   159.0 159.4 112.6   90.0   90.0   90.0  2852171
:)  10    0.189 dg 0.118   monoclinic mC   159.4 159.0 112.6   90.0   90.0   90.0  2852168
:)   9    0.174 dg 0.106 orthorhombic oC   158.9 159.3 112.7   90.0   90.0   90.0  2851683
:)   8    0.174 dg 0.108   monoclinic mP   112.5 112.6 112.5   90.0   90.2   90.0  1425804
:)   7    0.171 dg 0.107   monoclinic mC   159.3 158.9 112.7   90.0   90.0   90.0  2851661
:)   6    0.172 dg 0.110   monoclinic mP   112.5 112.6 112.6   90.0   90.1   90.0  1426051
:)   5    0.089 dg 0.107 rhombohedral hR   159.4 159.4 194.6   90.0   90.0  120.0  4278868
:)   4    0.089 dg 0.104   monoclinic mC   159.0 159.5 112.5   90.0   90.2   90.0  2852939
:)   3    0.080 dg 0.100   monoclinic mC   159.1 159.3 112.5   90.0   90.2   90.0  2850854
:)   2    0.032 dg 0.102   monoclinic mC   158.9 159.4 112.7   90.0   90.2   90.0  2852415
:)   1    0.000 dg 0.093    triclinic aP  112.5 112.5 112.7   90.1   90.1   90.1  1425683
```

These four cases cannot be distinguished until the data are integrated and symmetry-related intensities are compared. Still, if radiation damage is a problem, there is a way to identify the threefold with minimal initial data. Follow these steps:

a. Collect 1°-oscillation images at 0°, 1°, 2°, 3°, and 90°.
b. Use *Web-Ice*'s Autoindex Tab to index on the 0° & 90° frames. Now navigate to the "Details" button to display the name of the Unix directory that contains the processing results.
c. Open a Unix command shell and change to that directory. Issue the command

`labelit.rsymop 1 4 #` this integrates the first 4° of data and helps identify the correct setting

Let us assume that in this case, solution #5 proves to be the correct rhombohedral setting.
d. Again in *Web-Ice*, return to the Autoindex results, and hit the "Setup" button. Request the additional integration of solution #5, and hit "Submit".

e. When this is complete, proceed to the "Strategy" button, select "solution 5" in the combo-box, and read off the optimal phi-rotation range from there.

5. *Resources*

More information about the ALS macromolecular beamlines is listed at the Berkeley Center home page. This site also gives the actual URLs for logging in to BCSB's *Web-Ice*.

The Web-Ice home page at SSRL gives a complete *Web-Ice* user manual, as well as the URL for logging in to an SSRL *Web-Ice* account.

Go to the LABELIT Web site (cci.lbl.gov/labelit) to learn more about the spotpicking toolbox (*DISTL*) and autoindexing software (*LABELIT*). The page provides electronic reprints of the original papers.

Contact information:

Nick Sauter, LBNL (510) 486-5713 nksauter@lbl.gov
Ana Gonzalez, SSRL (650) 926-8682 ana@slac.stanford.edu