automateD collectioN of datA (DNA) Software Package

A quick tutorial for version 1.1
23.05.2007

Evangelia Kapetaniou[1], Kristian Koski[2], Kyriakos Petratos[1], Rikkert Wierenga[2]

[1] IMBB, FORTH, Heraklion, Crete, EU-BIOXHIT-TID centre, Greece
[2] Biocenter Oulu and Department of Biochemistry, University of Oulu, EU-BIOXHIT-TID centre, Finland

DNA developers

Karen Ackroyd (a)
Alun Ashton (b)
Gleb Bourenkov (c)
Sandor Brockhauser (d)
Steve Kinder (a)
Pierre Legrand (e)
Karl Levik (b)
Romeu Pieritz (f)
Harry Powell (g)
Darren Spruce (f)
Olof Svensson (f)
Graeme Winter (a)

DNA Exec. Committee:

Gérard Bricogne (h)
Andrew Leslie (g)
Sean McSweeney (f)
Colin Nave (a)
Alexander Popov (c)
Raimond Ravelli (d)
Andrew Thompson (e)

(a) CLRC Daresbury, UK
(b) Diamond light source, UK
(c) EMBL Hamburg, Germany
(d) EMBL Grenoble, France
(e) Synchrotron Soleil, France
(f) ESRF, France
(g) MRC LMB, Cambridge, UK
(h) Global Phasing, Cambridge, UK

This manual was written by a collaboration of the EU-BIOXHIT-TID centres in Oulu and Crete and the program developers of DNA at the ESRF.

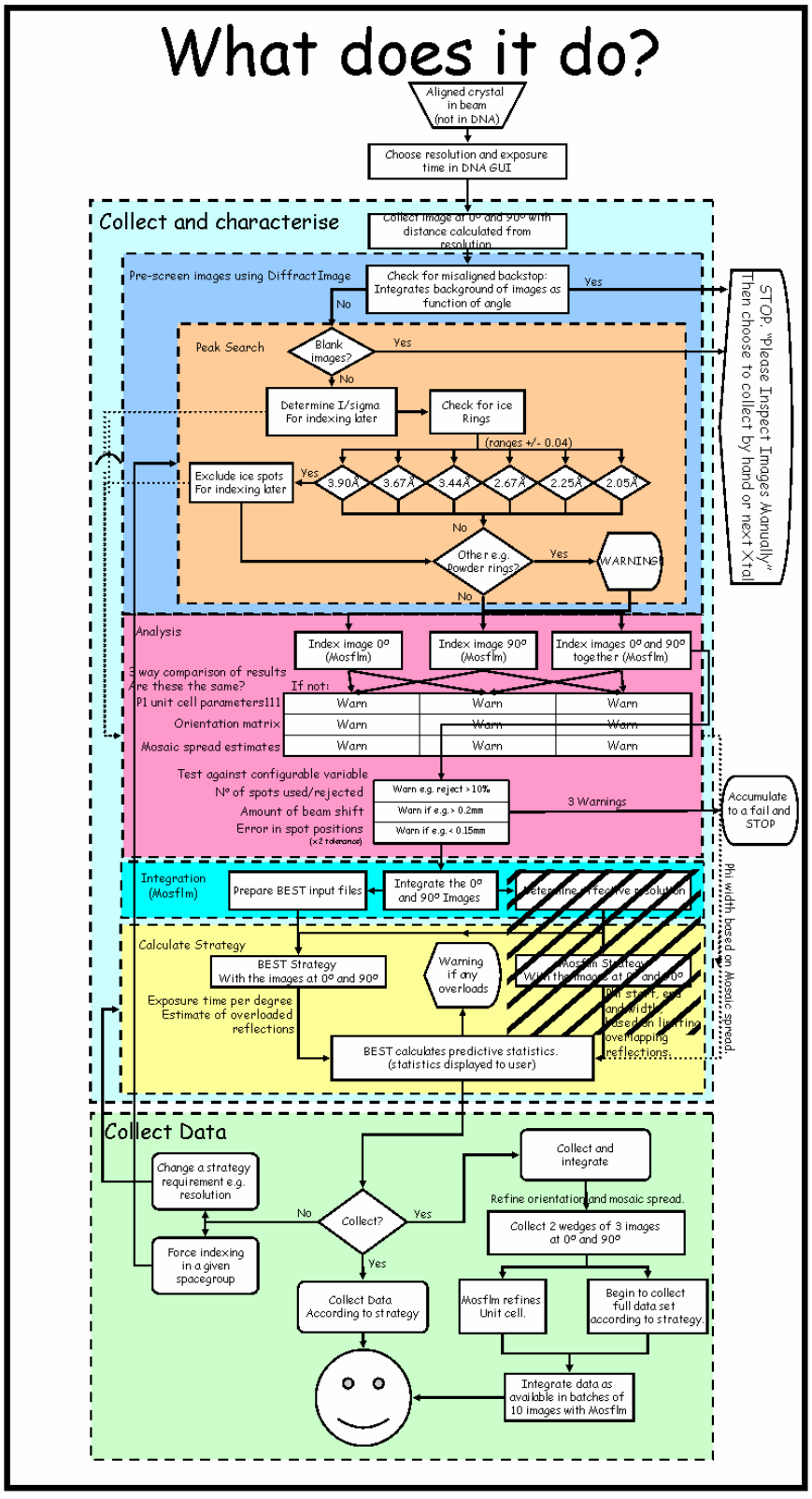

Figure 1. The DNA flow diagram (http://www.dna.ac.uk/).
In current version (1.1) only BEST strategy is used.

Short Overview

The DNA software package (Leslie *at al*, 2002) has been developed for the control of the automatic collection and the processing of x-ray diffraction data from crystals of macromolecules. It consists of several independent software modules that communicate with each other (Fig. 1 and 2). Input and output data are passed to data base (for example ISPyB - Information System for Protein CrystallographY Beamlines) and the Graphical User Interface (GUI), from which experiments are launched and monitored. The remainder of data handling and processing is carried out by underlying layers of software:

- MOSFLM (Leslie, 2005): data processing program.
- BEST (Bourenkov and Popov, 2006): data collection strategy calculations.
- SCALA (Evans, 2006): data reduction.
- POINTLESS (Evans, 2006): determination of point group/space group.
- ISPyB (https://wwws.esrf.fr/ispyb/ispyb/security/logon.do).

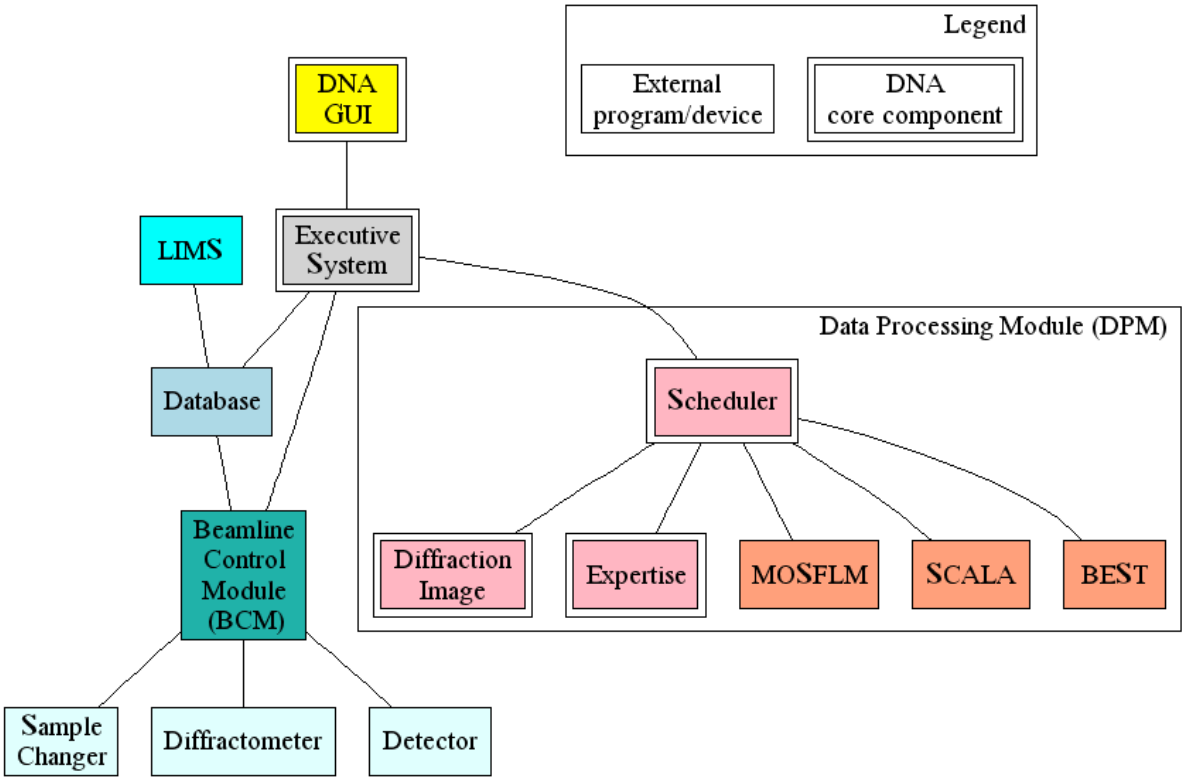

Figure 2: The DNA system contains a number of modules, where each module is specialised in a particular task (http://www.dna.ac.uk/).

DNA GUI

1. Getting started

The DNA GUI (Fig. 3) consists of three separate windows:

- Top Main window exhibits the main options needed for the high-throughput synchrotron data collection experiment: *Sample Screening*, *Sample Ranking*, *Collect Reference Images*, *Auto Index, Strategy* and *Results*.
- Middle Main window states the status of the experiment. *Help*, *Submit feedback* and *Abort* options are available.
- Lowest Main window shows on the fly the log files from DNA (*Executive Output*) or from MOSFLM (*MOSFLM Output*).

When a new experiment is started at ESRF MX beamlines, the DNA interactive window with the experimenter's MX code is on display (Fig. 3).

- click on *OK*.

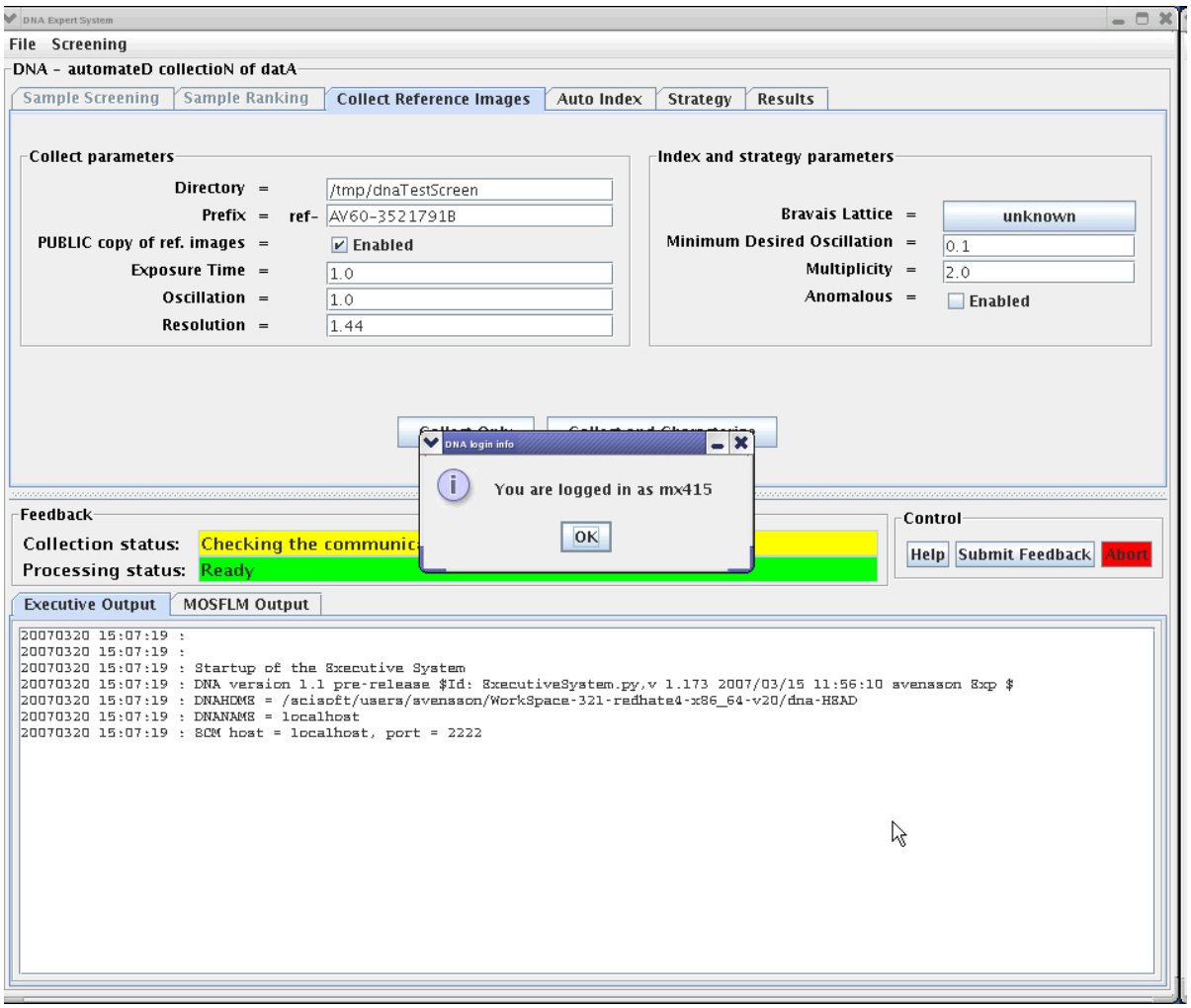

Figure 3. Starting window.

In the following, a number of scenarios, corresponding to the typical applications of DNA will be presented:

List of Scenarios:

- Screening crystals (2) and ranking them (2.1)
  - User has multiple crystals of the same protein in the sample changer.
- Characterize a single crystal (2.2)
  - User just wants to collect a dataset from a crystal already mounted on the beamline.
- Strategy calculation (3).
- Data collection and integration (4).
- Quick scaling (5).

2. Screening Crystals

Once crystals have been transferred into the baskets and placed in the sample changer and the correct shipment has been selected for experiment in the beamline LIMS (e.g. ISPyB) interface and activated, the next step is to automatically characterize and rank them for data collection. Within the DNA system sample screening procedure consists of the collection of two reference diffraction images, collected 90° apart for each sample. At ESRF beamlines, the sample changer can take up to 50 crystals (in five baskets), which all can be selected for one screening experiment (*Select All Samples* button).

- Select *Screening* → *Screening/Ranking* from top menu and press *Sample screening* option. A Screening window will appear (Fig. 4).

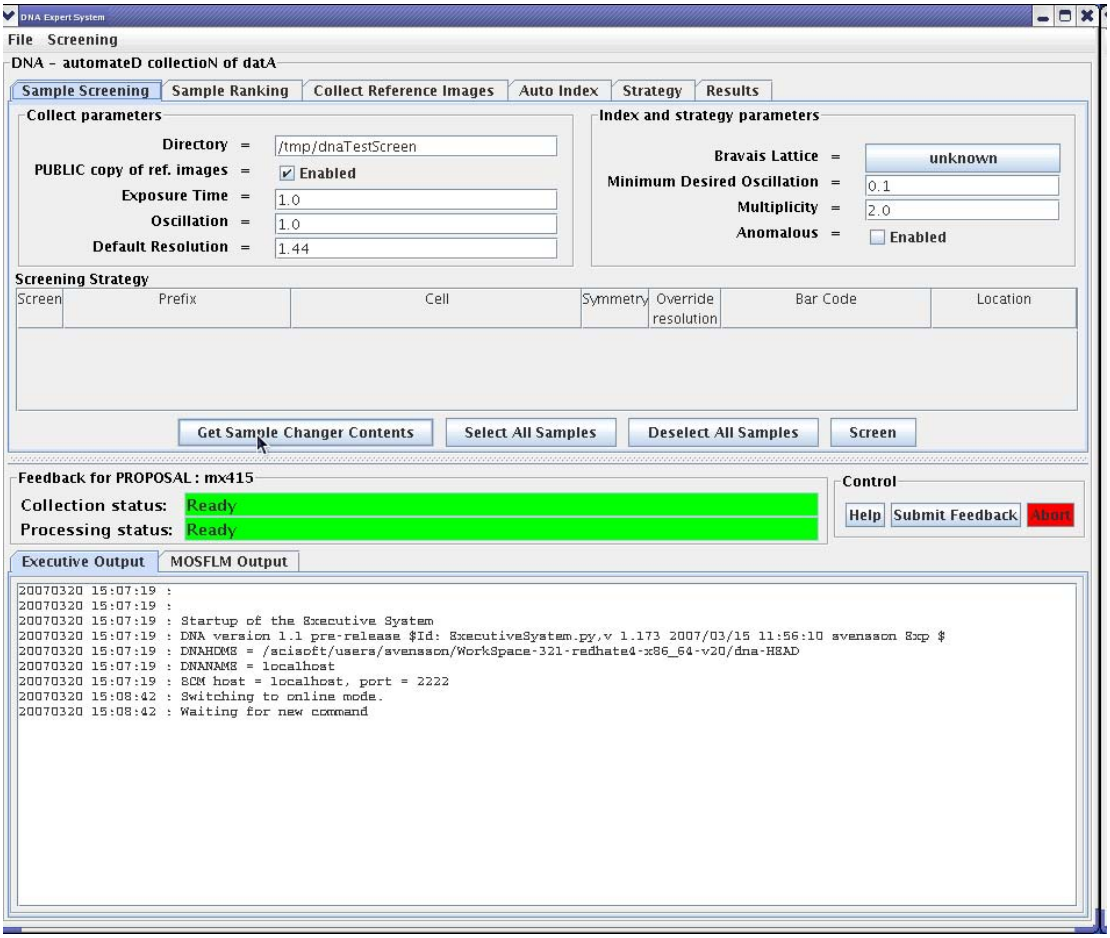

Figure 4: Screening window.

— Fill up mandatory information: directory, exposure time, oscillation and resolution. The resolution for screening the crystals can be provided by the diffraction plan (2.8 Å for 2nd and 3rd crystal in Figure 5). If no resolution is available in the diffraction plan, DNA either uses the default resolution (2 Å in Figure 5) or the user can enter an override resolution for each crystal individually.
— Select Bravais lattice, if known, and click on *Anomalous Enabled*, if you perform SAD/MAD experiments.
— Click on the *Get Sample Changer Contents* option (Screening window, middle). Sample changer starts to read the bar codes from the top of each sample holder and compares them with the corresponding information in the data base. Crystal data will appear in the Screening window (Fig. 5).
— Select the samples for screening (press *Select All Samples* option, Screening window middle, if you wish to screen all the samples).
— Click on *Screen* option (Screening window, middle).
— The program will prompt you for rank file input (Fig. 5). If similar crystals have not been screened previously click on *Create a New Rank Project File*. The program will then automatically produce a .pdrp file containing all the information about the screened crystals. If the user wants to compare screening results to previous screening, a *Load Existing Rank Project File* option should be pressed and then the proper rank file (.pdrp) ashould be selected.

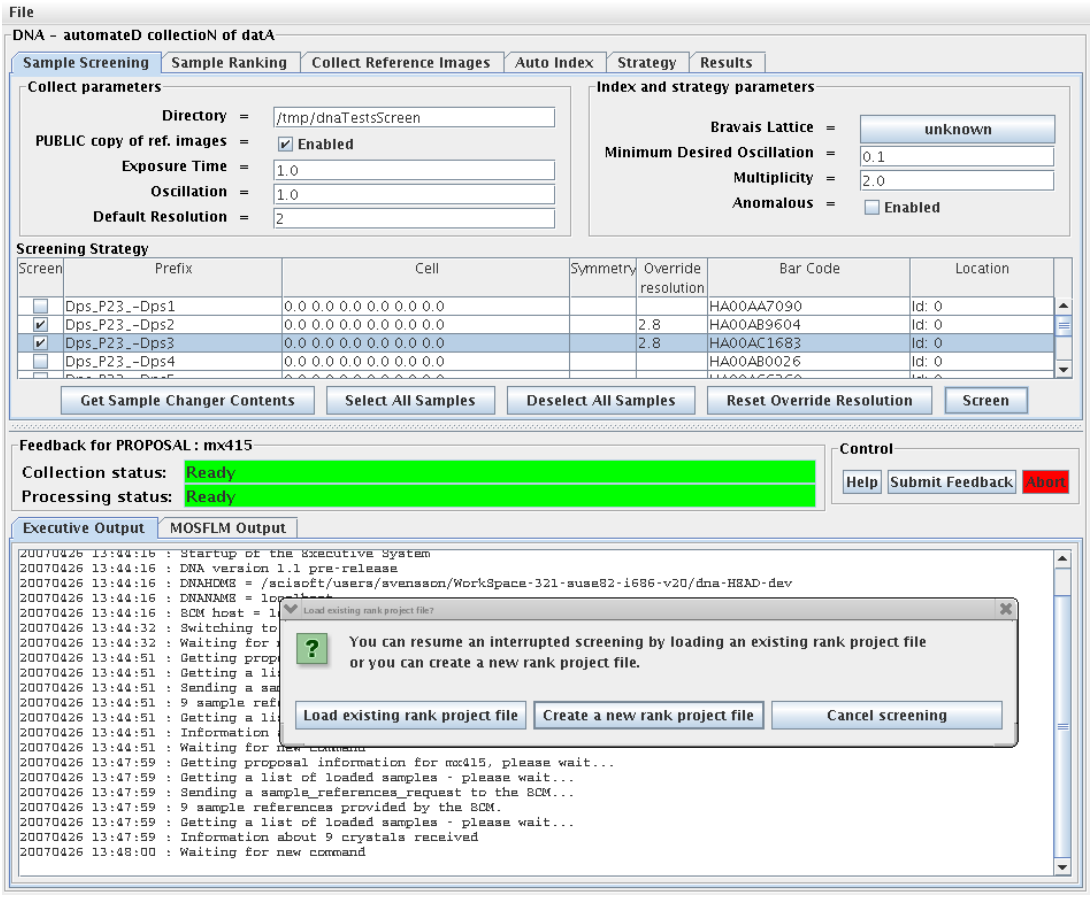

Figure 5. Screening window with rank project question.

2.1. Crystal ranking

In principle, the information gathered by DNA from the two reference diffraction images is sufficient to allow the automatic ranking of samples. Criteria to be used in a system of crystal ranking are currently under discussion and include: total exposure time, ranking resolution, crystal mosaic spread, number of images and spot deviation. New ranking schemes are easy to implement thanks to the flexible ranking module and will vary according to the experiment goals and user personal style.

- Click on the *Sample Ranking* menu option (starting window, top left; Fig. 2). A Sample Ranking window will appear (Fig. 6) with space group and unit cell parameters information for each crystal.
- The crystals are ranked as a default by the total exposure time needed for collecting a full data set as calculated by BEST. The user can select an alternative ranking criterion from the Sample Ranking window, middle and click on *Redo Rank* (Fig. 7).

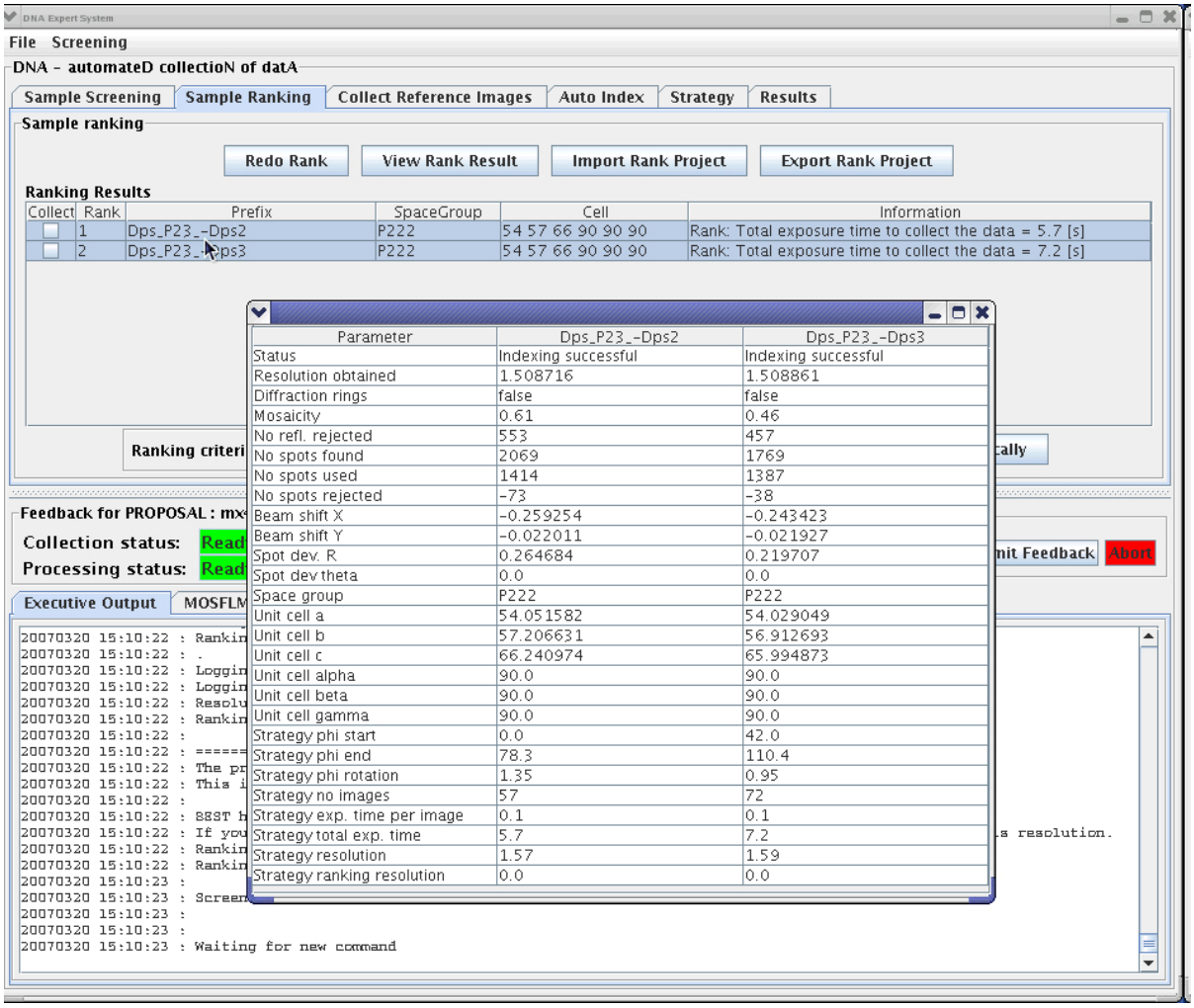

Figure 6. Crystal ranking window.

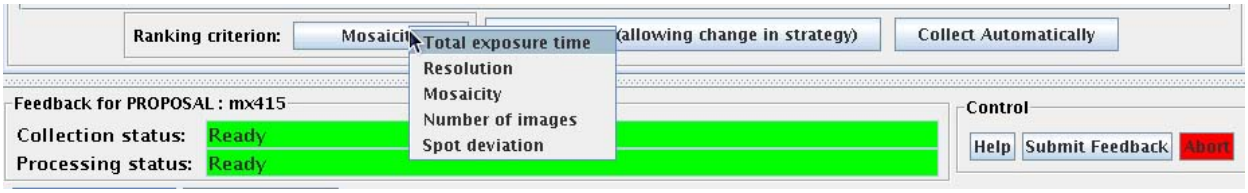

Figure 7. Redo ranking.

- To display the table containing the more detailed statistics for all screened crystal (as shown in Fig. 6), select the crystals with the left mouse button and then click the right mouse button to see the table.
- Click on *View Rank Result* and observe the diffraction images (Fig. 8), with or without predictions.

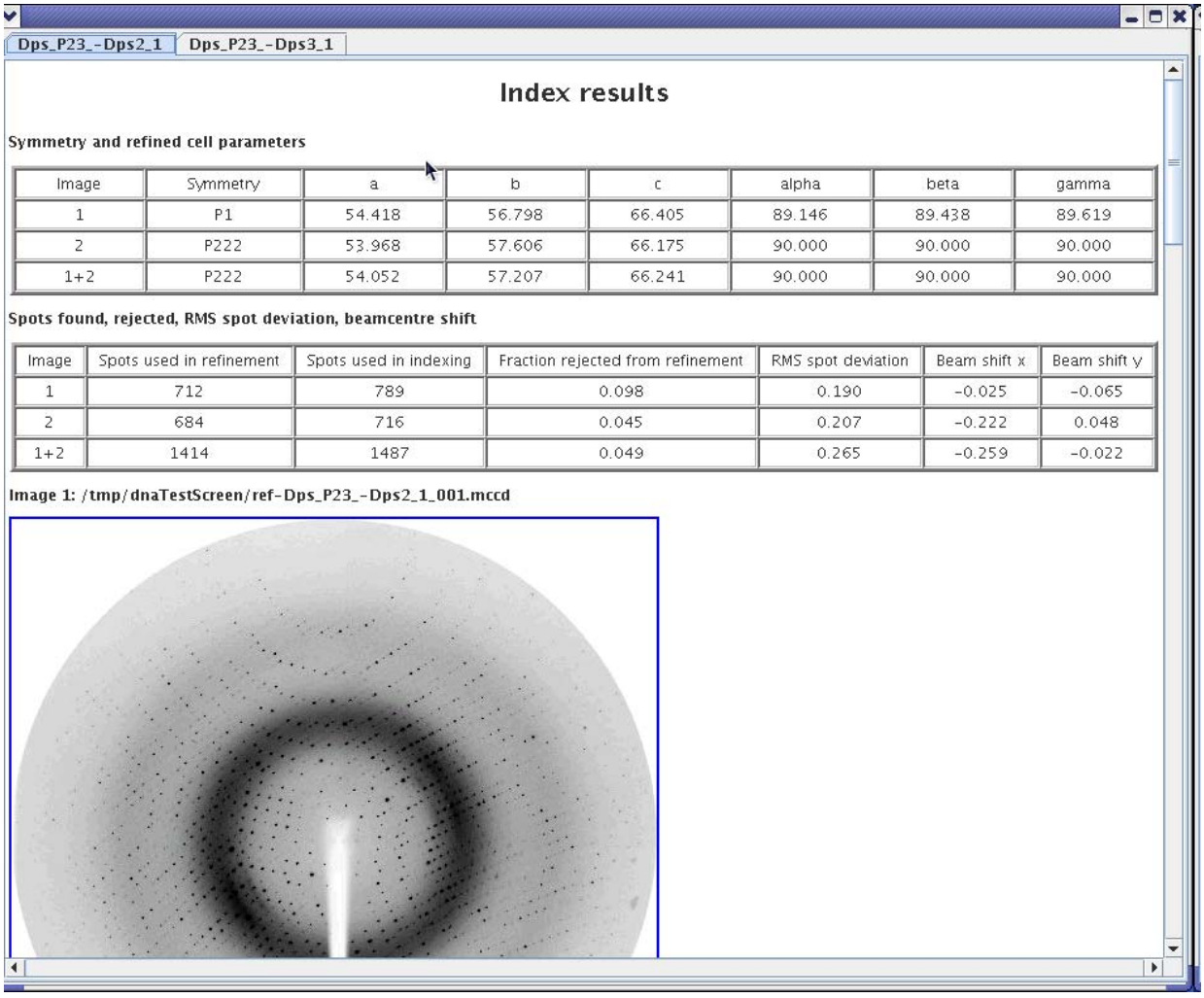

Figure 8. Index results.

- After selecting the best crystal press *Collect data (allowing change in strategy)*. A Strategy window will appear showing the data collection strategy proposed by BEST. See sections 3 and 4 (below) for details on how to proceed to data collection (and processing).
- *Collect Automatically* button allows the user to collect the data automatically from all selected crystals using the strategy calculated by BEST (However, it is recommended to

use the option *Collect data (allowing change in strategy)*, because strategy calculated by BEST is not always the most convenient one (may have very small oscillation angle, etc.).

2.2. Characterization of a single crystal

If the user wants to collect a data set from a single crystal which is already mounted on the beam:

- Click on the *Collect Reference Images* menu option.
- Fill up mandatory information: directory, exposure time, oscillation, expected resolution.
- Select Bravais lattice, if known, and click on *Anomalous Enabled*, if you perform SAD/MAD experiments.
- Press *Collect and Characterize.*

Two reference images are collected and processed. If both have been indexed in the same space group and with comparable unit cell parameters, the *Auto index* window will appear (Fig. 7). From Auto Indexing window, DNA will continue automatically to *Strategy* window, which is described in next chapter.

The *Collect Only* button is only collecting two reference images without starting the data processing. This option can be used to test the data collecting hardware or just to check the diffraction images.

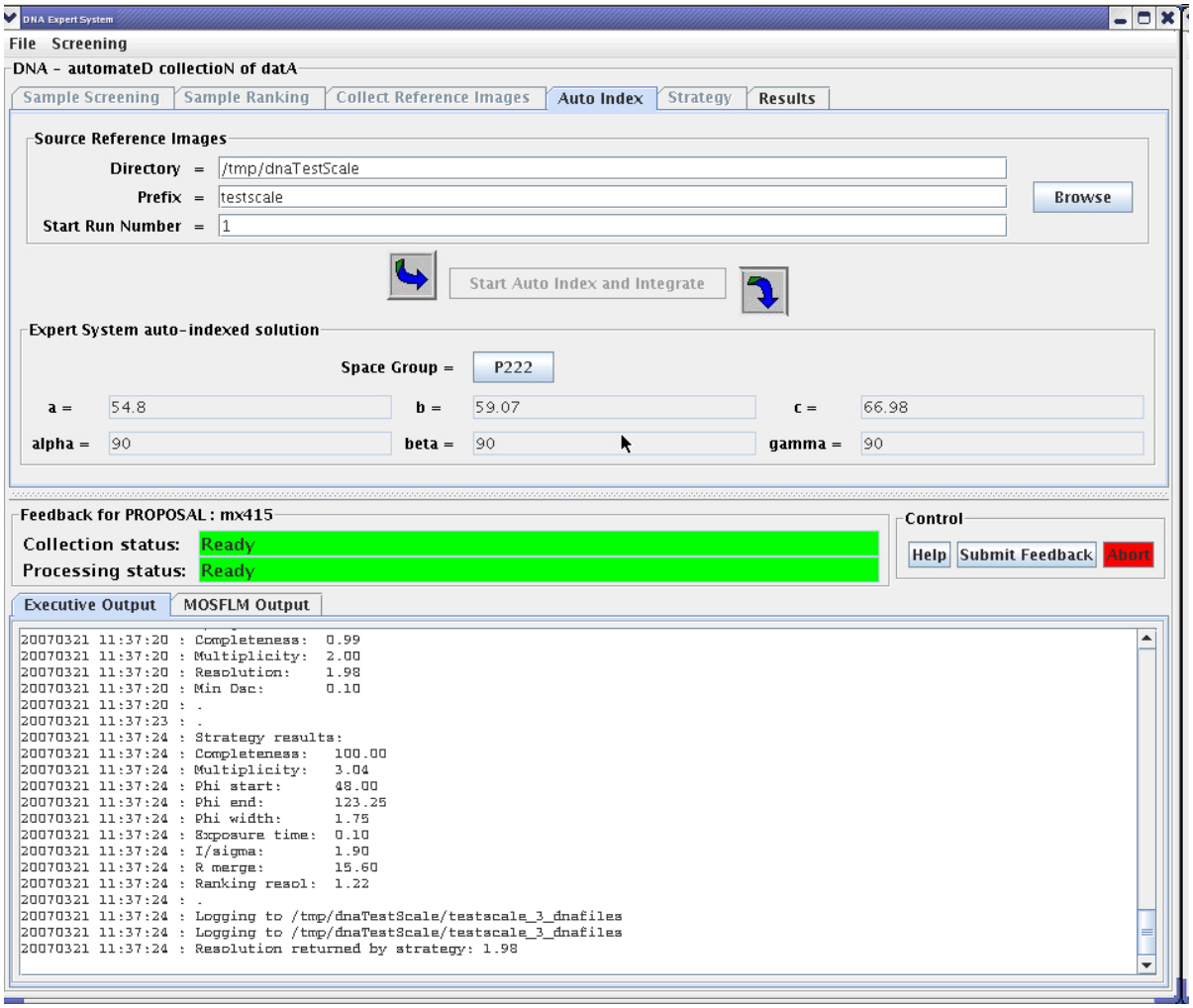

Figure 9. Auto index window.

3. Calculating strategy

After auto indexing the best data collection strategy is calculated (by BEST) and the Strategy window will appear automatically. The reference strategy is seen on the left of the Strategy window (Fig. 10.). The actual data collection strategy on the right side of the window can be freely edited by the user by simply clicking the parameters that user wants to change.

However, the user has to keep in mind that changing the parameter may cause changes in the optimal data collection strategy. So, if the data collection parameters have been changed the user may have to recalculate the strategy by pressing the *Calculate Strategy* button in the middle left. A new strategy will be then calculated.

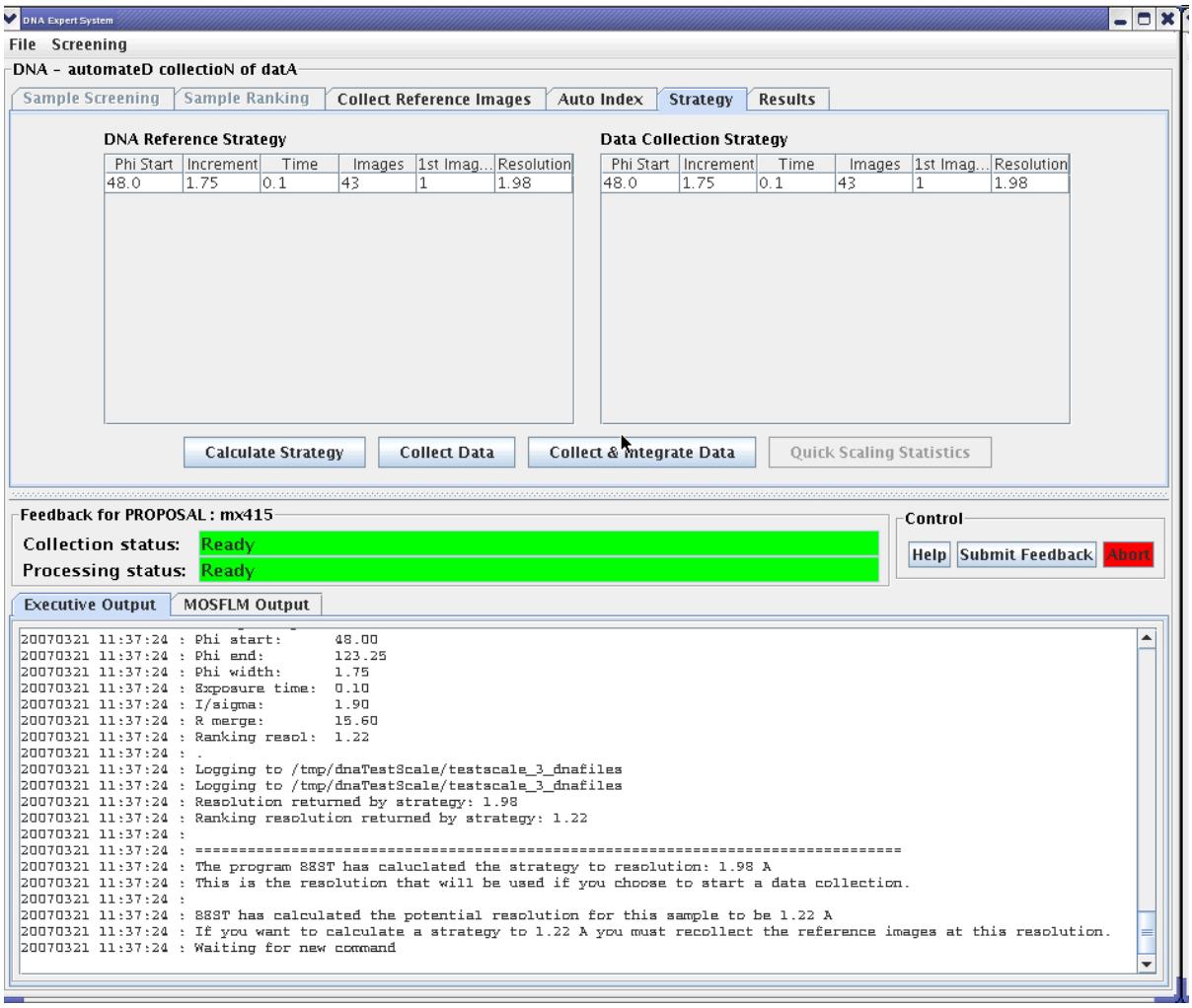

Figure 10. Strategy window.

4. Data Collection and Integration

After the Strategy Calculation there are two options:

— Click on *Collect Data* (Strategy window, middle; Fig. 10).

Data collection will be carried out using the given strategy.

— Click on *Collect and Integrate Data* (Strategy window, middle; Fig. 10).

Data collection will be carried out using the given strategy and the data set is also integrated by MOSFLM, while it is collected.

DNA will first collect three images starting from at a phi value 90° away from the starting phi for data collection. These 3 images, together with the first 3 images of the data collection, are used for unit cell refinement (requirement of MOSFLM).

During integration, DNA will use POINTLESS to determine the point group of the crystal (Fig. 11).

```
20070321 12:42:02 : .                                          .
20070321 12:42:02 :       IMPORTANT INFORMATION!!
20070321 12:42:02 : .                                          .
20070321 12:42:02 :  With confidence 1.000000
20070321 12:42:02 :  Pointless thinks the spacegroup should be P 2 2 2
20070321 12:42:02 : .                                          .
20070321 12:42:02 : Though it may have screw axes...
20070321 12:42:02 : .                                          .
20070321 12:42:03 : ------------------------------------------------------
```

Figure 11. Pointless results from *Executive output window* (Lowest window of the DNA GUI).

5. Quick Scaling

This option is only possible if the *Collect and Integrate* method is used during data collection. The quick scaling is meant ONLY for rapid data analysis. The data is scaled only quickly to save more time for actual data collection (during scaling the data collection is stopped).

— Click on *Quick Scaling Statistics* (Strategy window, middle; Fig. 10).

The following statistics are shown after a quick scaling run: resolution, R merge, I/sigma, completeness, multiplicity, number of reflections, unique reflections. The values in the highest resolution shells are shown in brackets (Fig. 12).

```
20070321 12:42:25 : .                                                .
20070321 12:42:25 :          IMPORTANT INFORMATION!!
20070321 12:42:25 : .                                                .
20070321 12:42:25 : Resolution    41.96 -  1.98 ( 2.09 -  1.98)
20070321 12:42:25 : R merge         0.040     (0.065)
20070321 12:42:25 : I/sig           12.4      (  9.9)
20070321 12:42:25 : Completeness    27.53      ( 29.80)
20070321 12:42:25 : Multiplicity    1.31       ( 1.24)
20070321 12:42:25 : Reflections     5491         (816)
20070321 12:42:25 : Unique          4197         (656)
20070321 12:42:25 : .                                                .
20070321 12:42:25 : .                                                .
```

Figure 12. Quick scaling results from *Executive output window* (Lowest window of the DNA GUI).

6. Results

— Click on the *Results* menu option (Starting window, top right).

The following information is displayed in the Results window

— Indexing results with symmetry information and unit cell parameters (Fig. 13).
— Diffraction images of the two reference images with and without predictions (Figs. 8 and 14).
— MOSFLM log file.
— Integration results shown as graphics.
— Strategy results.

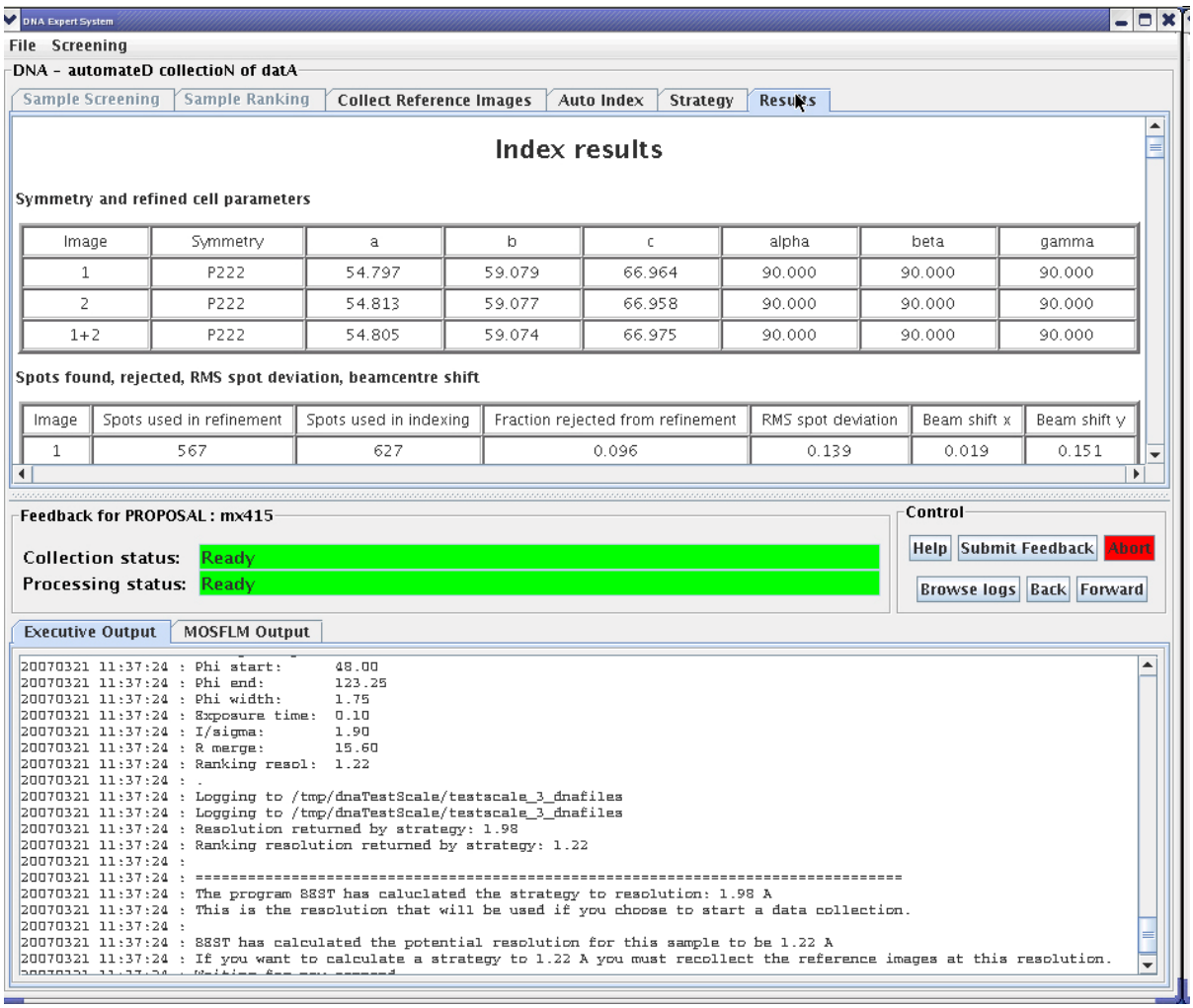

Figure 13.

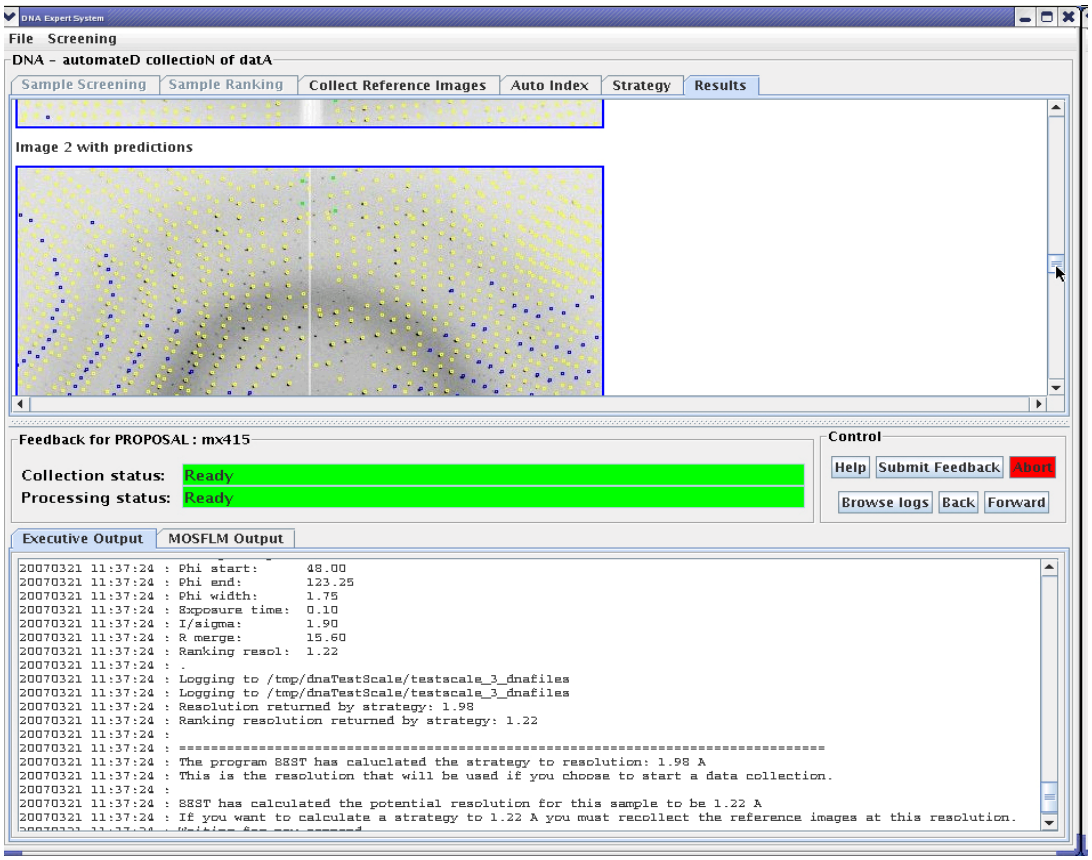

Figure 14.

All the results files are also written to the working directory. Also input files for integrating the images using XDS and MOSFLM are generated by DNA.

The indexing results and the diffraction patterns of the reference images (with and without predictions) are also written to the database.

7. References

- Evans, P. (2006) Scaling and assessment of data quality, *Acta Cryst*. **D62**:72-82.
- Leslie, A. (2002) Automation of the collection and processing of X-ray diffraction data – a generic approach, *Acta Cryst*. **D58**:1924-1928.
- Leslie, A. (2005) The integration of macromolecular diffraction data, *Acta Cryst*. **D62**:48-57.
- Bourenkov, G. and Popov, A. (2006) A quantitative approach to data-collection strategies, *Acta Cryst*. **D62**:58-64.