

## TABLE OF CONTENTS

<b>PREFACE</b>	<b>4</b>	
<b>HKL-2000 OVERVIEW</b>	<b>5</b>	
<b>THE HKL-2000 TABS</b>	<b>6</b>	
The Project Tab		6
The Data Tab		7
The Auto Tab		8
The Summary Tab		9
The Index Tab		10
The Integrate Tab		12
The Scale Tab		13
The Merge Tab		14
The Macros Tab		15
<b>CONVENTIONS USED IN THIS MANUAL</b>	<b>16</b>	
<b>INSTALLING HKL-2000</b>	<b>17</b>	
<b>STARTING HKL-2000</b>	<b>18</b>	
Site Definitions		19
The Site Included option		20
Creating a new site		20
<b>DATA, RESULTS, AND DATASETS</b>	<b>25</b>	
The Directory Tree Window		25
Select the raw data directory		25
Choose an output file directory		26
Selecting datasets		27
Removing sets		30
Integrate, Scale, and Collect checkboxes		31
Splitting sets		31
Checking and Editing Sets		32
Start working before all frames are collected		35
<b>CREATE, SAVE, AND LOAD PROJECTS</b>	<b>36</b>	
Create and edit projects and crystals		36
Saving and Loading Projects		40
<b>DISPLAY AND PEAK SEARCH THE DATA</b>	<b>42</b>	

## TABLE OF CONTENTS

Automatic Peak Search	45
Manual Peak Searches	46
Secondary Lattices	48
Adding and Removing Peaks	49
Peak search on different frames.	51
<b>AUTO-INDEXING</b>	<b>52</b>
The Pending Sets window	52
Initial Parameters for Indexing	53
Peaks used for autoindexing	54
<b>LATTICE PARAMETER REFINEMENT</b>	<b>56</b>
When refinement goes wrong	57
Initial 3D window setting	57
Choosing the correct lattice	58
Primitive triclinic crystals	60
Positional and Partiality $\chi^2$ values	60
Starting refinement over	64
Resolution Limits, Fitting Radius and Spot Size	65
The Profile Fitting Radius	65
Setting the Spot Size	67
What is the Elongation Limit?	69
Setting the integration resolution limit	69
The Crystal Rotation X, Crystal Rotation Y, and Crystal Rotation Z parameters	71
The Reference Zone	73
The 3D Window and Mosaicity	74
Masking the beamstop and other shadows	75
<b>STRATEGY AND SIMULATION</b>	<b>78</b>
Strategy	78
The I/Sigma and Crystal Dose buttons	81
Simulation	82
Combining simulation runs to generate a cohesive strategy	84
<b>INTEGRATING ALL FRAMES</b>	<b>85</b>
The Integrate button is red. Is this a problem?	85
Integration	85
The Crystal Slippage and Special integration buttons.	85
Integration Information Plots	86
What is an acceptable range for $\chi^2$ ?	87
Drastically larger $\chi^2$ values	87
The Mosaicity histogram	88
How long should this take? How do I know if the program is stuck?	89
Waiting for image frames.	89
Updating the site configuration after processing data	90

## TABLE OF CONTENTS

How do I stop and start over?	90
<b>SCALING</b>	<b>91</b>
Scaling Options	91
Global Refinement Options	96
The three <i>Scalepack</i> output Files	97
<b>OUTPUT STATISTICS AND GRAPHS</b>	<b>98</b>
Working with anomalous data	102
The Scaling Controls Box	103
Space Group Identification	111
Scaling * . x files without reintegrating diffraction images	114
Statistics and <i>Scalepack</i> .	115
The Resolution of Diffraction Data.	118
<b>THE PROJECT TAB AFTER SCALING</b>	<b>120</b>
<b>AUTOMATIC DATA PROCESSING</b>	<b>122</b>
<b>MERGING MULTIPLE SCALED FILES</b>	<b>125</b>
<b>USING MACROS</b>	<b>127</b>
<b>APPENDIX – HKL-2000 INSTALLATION</b>	<b>ERROR! BOOKMARK NOT</b>
DEFINED.	
Obtaining HKL-2000 installation files	129
License key	129
Using the HKL Installer	131
Installing HKL-2000 manually	134
Starting HKL-2000	135
<b>APPENDIX - TROUBLE SHOOTING</b>	<b>137</b>
When refinement goes wrong	137
<b>APPENDIX – INDEXING AND INTEGRATION KEYWORDS</b>	<b>145</b>
<b>APPENDIX - SCALING KEYWORDS</b>	<b>173</b>
<b>TABLE OF FIGURES</b>	<b>197</b>

# Preface

The *HKL-2000* suite is a package of programs intended for fast indexing, data reduction, and scaling of X-ray diffraction data collected from a single crystal.

*HKL-2000* contains extended versions of the programs included in the HKL package: **Denzo**, **XdisplayF**, and **Scalepack**, as well as several additional 'utility' programs. The strategy and simulation features can help plan the experiment in real-time. The experimenter can adjust the data collection strategy to achieve the desired data completeness. A proper data collection strategy can help avoid later problems in data reduction and structure solution. There is also mosaicity refinement during processing, variable spot size, easy change of the space group, 3-D processing, direct beam position search, report generation, and other tools.

The graphical command center coordinates all the steps of an experiment and enables performing an interactive experiment with on-line data analysis, where results are automatically updated when new data are collected.

The graphical user interface (GUI) greatly simplifies interaction with the underlying suite of programs and improves the convenience of working with various data sets. The interface was designed to be intuitive and allows the user to easily adjust a variety of parameters. It provides a guided path through all steps from data collection to scaling.

This manual aims to describe how to use the GUI to efficiently and accurately process your data. The original manuals for the underlying suite can be found on the HKL Research website. The figures in this manual are for version 717 of the *HKL-2000* program. Since the program continues to undergo constant revision and development, the version of the *HKL-2000* package you are using may not be the same one described here.

To find more about the features of and algorithms used in *Denzo* and *Scalepack*, please read the article: Z. Otwinowski and W. Minor "**Denzo & Scalepack**" *International Tables for Crystallography Volume F: Macromolecular Crystallography*, p. 226-235, 2001, M. G. Rossmann, Kluwer Academic Publishers (Dordrecht/Boston/London).

Question about the program described here should be directed to Zbyszek Otwinowski or Wladek Minor at the addresses below, with CC: to HKL Research:

HKL WWW homepage:	<a href="http://hkl-xray.com">http://hkl-xray.com</a>
HKL support:	<a href="mailto:support@hkl-xray.com">support@hkl-xray.com</a>
HKL Research e-mail	<a href="mailto:hkl@hkl-xray.com">hkl@hkl-xray.com</a>
Zbyszek Otwinowski:	<a href="mailto:zbyszek@work.swmed.edu">zbyszek@work.swmed.edu</a>
Wladek Minor:	<a href="mailto:wladek@iwonka.med.virginia.edu">wladek@iwonka.med.virginia.edu</a>

To reference the *HKL-2000* package for publication, please use the following citation:  
Z. Otwinowski and W. Minor, "**Processing of X-ray Diffraction Data Collected in Oscillation Mode**," *Methods in Enzymology*, Volume **276**: Macromolecular Crystallography, part A, p.307-326, 1997, C. W. Carter, Jr. & R. M. Sweet, Eds., Academic Press (New York).

**Denzo®**, **HKL®**, and **HKL-2000®** are registered trademarks of HKL Research, Inc.



# HKL-2000 Overview

There is a logical path from raw X-ray diffraction images to a crystal structure, and HKL aims to guide you through these steps. The default settings have been selected to work well for most cases, and the manual has been designed to emphasize the most critical parameters that you should pay attention to. A Quick Start document is available on the HKL Research web site that can guide you through data processing with example data.

---

## *The general workflow of HKL*

- *Select the detector used to collect the data.*
  - *Select the data and results directories.*
  - *Display and Peak Search the data to find reflections.*
  - *Auto-index the data in all crystal classes and select the best guess.*
  - *Index the data in the selected crystal class (assign Miller Indices).*
  - *Integrate the data. This indexes each image, refines relevant parameters, and combines partial reflections for each set of consecutive images. Integration generates \*.x files, which contain Miller Indices, intensities, and errors for every recorded reflection.*
  - *Scale the data. In simplistic terms, this ensures that the data is merged together properly with respect to image to image variations, and that symmetrically equivalent reflections are merged properly.*
  - *Analyze the data. HKL presents many informative graphs that let you instantly know the quality of your data.*
-

# The HKL-2000 Tabs

The next few pages show the layout of the different tabs in HKL-2000, briefly describe the tasks performed on each tab, and provide information about where in the manual to look for more detailed information.

## The Project Tab

File Options Site Configuration Crystal Information Report About Help

Project Data Auto Summary Index Strategy Integrate Scale Merge Macros

**Project**

Name: APC5771  
 Crystal: crystal1  
 Experimenter: dcoop  
 Date: Jun 26, 2018

Load Save

Edit Project Edit Crystal

New Project New Crystal

Evaluate Model

Principal Component: Protein

**Protein Data**

Name:  
 Number of Residues: 94  
 Organism:  
 Description:  
 Molecular Weight: 10715.48  
 NCBI accession code:  
 Swiss-Prot accession code:  
 No. of Homologues:  
 Last check in PDB:

GHMPAYVFSKESFLKFLEGHLEDDVWWWSSDVTFCKKLSSES  
 MVGEKEYCFAEFAPADIFDAEDEIDEMMKYAIIVFEKEKLSE  
 AGRNAIR

Phasing Method: SAD/MAD  
 Source of Anomalous Signal: Se 4

Ala (A): 8	Gly (G): 4	Met (M): 4	Ser (S): 7
Cys (C): 2	His (H): 2	Asn (N): 1	Thr (T): 1
Asp (D): 9	Ile (I): 4	Pro (P): 2	Val (V): 10
Glu (E): 13	Lys (K): 8	Gln (Q): 0	Trp (W): 0
Phe (F): 9	Leu (L): 5	Arg (R): 2	Tyr (Y): 3

Anomalous Wizard

**Scaled Data**

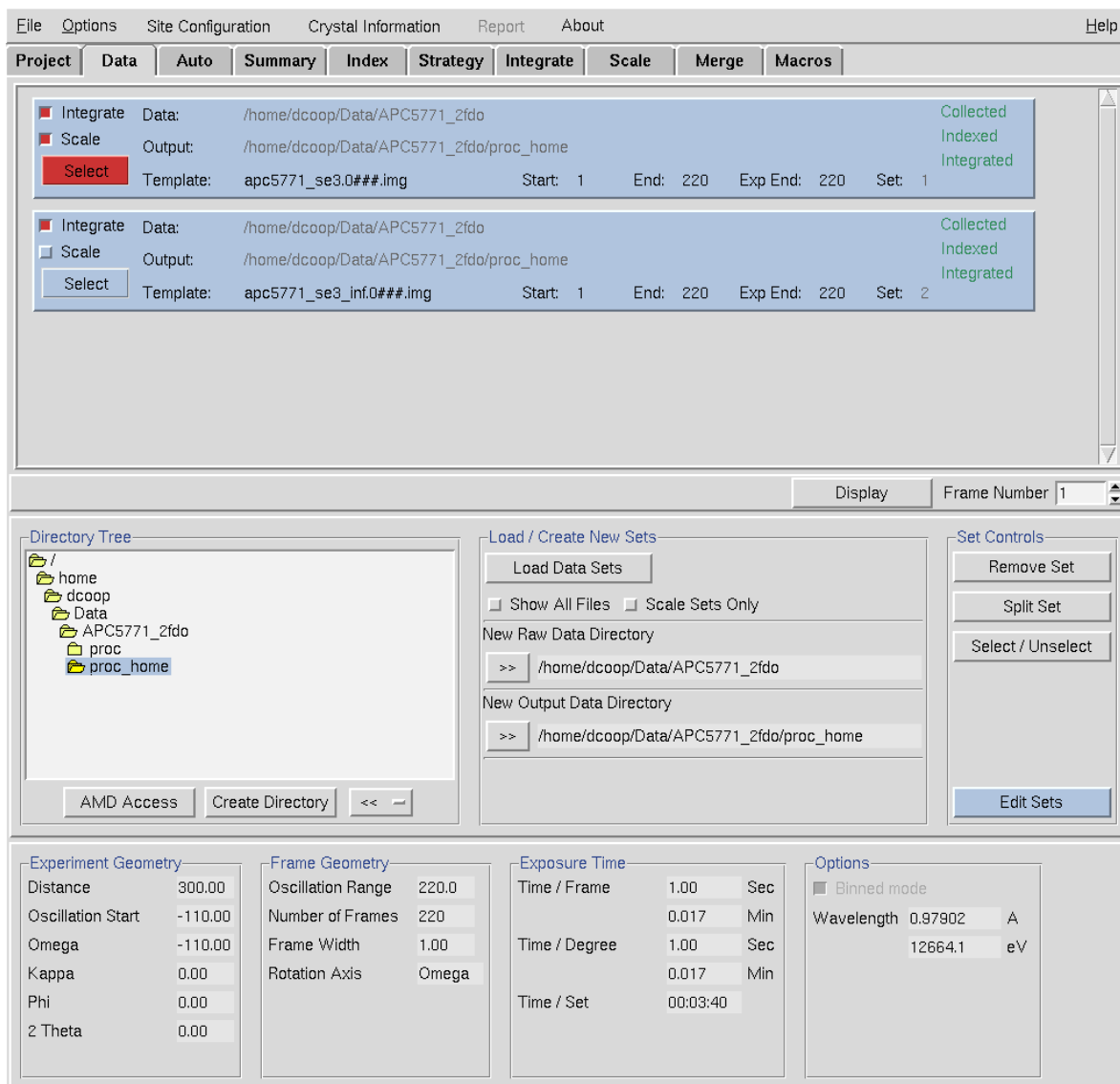
Structure	File Name:	Cell:	Space Group:	Wavelength:	Resolution:	Mode:	Order:
<input checked="" type="checkbox"/> Structure	/home/dcoop/Data/APC5771_2fdo/aproc_1/output.sca	45.459 48.856 86.337 90.000 90.000 90.000	P212121	0.97902		Native	1

Select

Add File  
 Remove File  
 Statistics

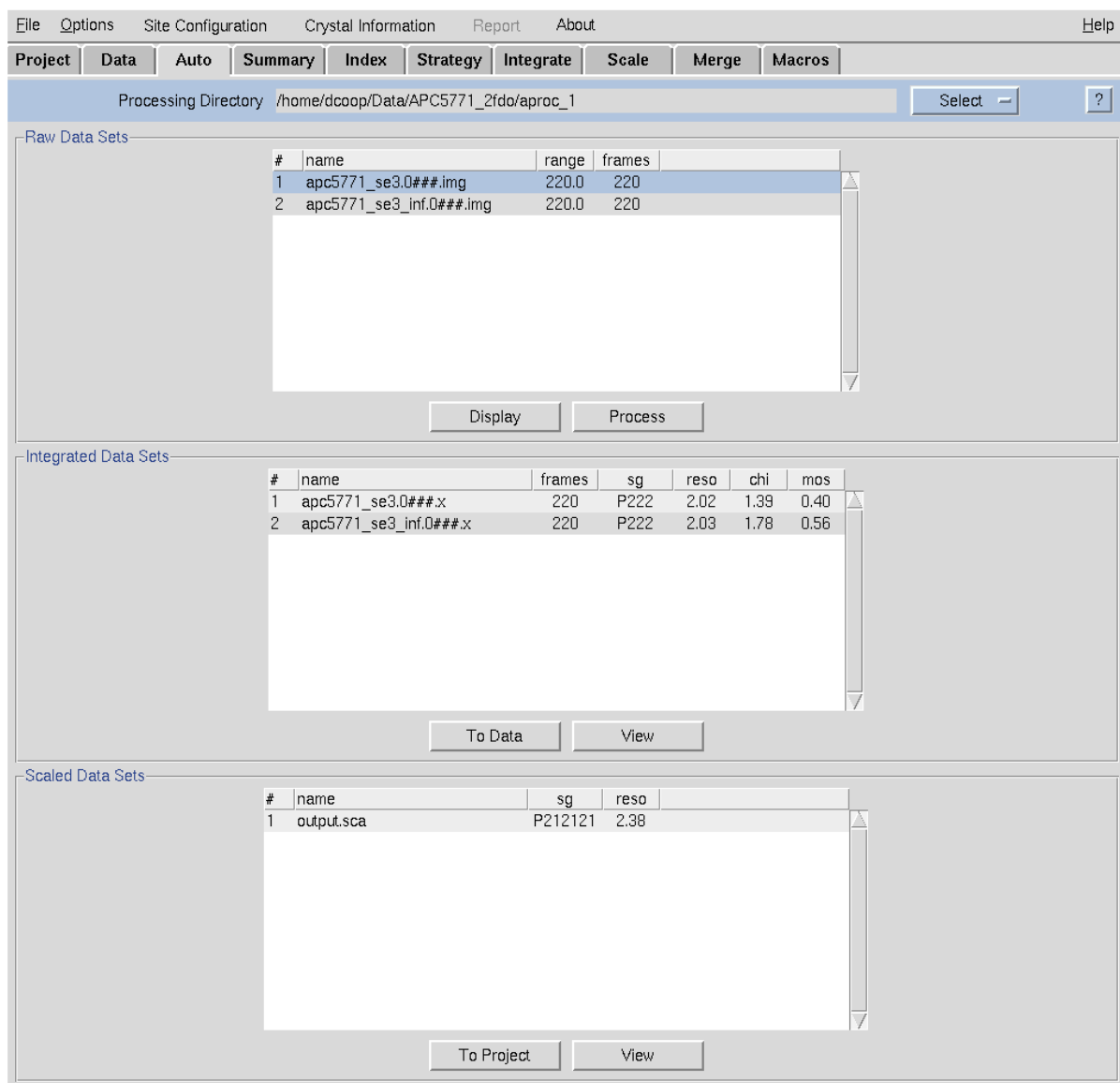
**Figure 1.** The *Project* tab contains information about the macromolecule. Using this tab, you can load or save session information, including which data sets to work with. Some of the features in the Project tab are more useful in database linked versions of HKL-2000 and in all versions of HKL-3000, which requires a database module license. See **Create, Save, and Load Projects** on page 36 and **The Project Tab After Scaling** on page 120.

## The Data Tab



**Figure 2.** The **Data** tab displays information about input and output files, displays information about individual datasets, and has control buttons to work with the datasets. The top of this tab shows the datasets that you are working with and what you intend to do with them. The middle of the page allows you to load and edit the sets you want to use, and the lower part of the display gives information about the currently selected set. See **Data**, **Results**, and **Datasets** starting on page 25.

## The Auto Tab



**Figure 3.** The *Auto tab* provides you with one-click processing and scaling your data. See **Automatic Data Processing** starting on page 122.

## The Summary Tab

Set	Raw Data Template	Start	End	Osc Start	Osc Width	Osc Range	Phi	Distance	Res Min	Res Max	3D Window	Profile Fitting Radius	Wavelength
	Default Parameters			0.00			0.00		50.00	edge	3	31.0	
1	apc5771_se3.0###.img	1	220	-110.00	1.00	220.0	0.00	300.00	50.00	edge	3	31.0	0.97902
2	apc5771_se3_inf.0###.img	1	220	-110.00	1.00	220.0	0.00	300.00	50.00	edge	3	31.0	0.97919

**View Controls**

<input checked="" type="checkbox"/> Set	<input checked="" type="checkbox"/> End	<input type="checkbox"/> Time/ Frame	<input type="checkbox"/> 2 Theta	<input type="checkbox"/> Mosaicity
<input checked="" type="checkbox"/> Raw Data Template	<input checked="" type="checkbox"/> Osc Start	<input type="checkbox"/> Time/ Degree	<input checked="" type="checkbox"/> Distance	<input checked="" type="checkbox"/> 3D Window
<input type="checkbox"/> Output Data Template	<input checked="" type="checkbox"/> Osc Width	<input type="checkbox"/> Kappa	<input checked="" type="checkbox"/> Res Min	<input checked="" type="checkbox"/> Profile Fitting Radius
<input checked="" type="checkbox"/> Start	<input checked="" type="checkbox"/> Osc Range	<input checked="" type="checkbox"/> Phi	<input checked="" type="checkbox"/> Res Max	<input checked="" type="checkbox"/> Wavelength
<input type="checkbox"/> Spot Size M.axis	<input type="checkbox"/> Spot Size m.axis	<input type="checkbox"/> Box Size X	<input type="checkbox"/> Box Size Y	<input type="checkbox"/> Spot Angle
<input type="checkbox"/> Energy	<input type="checkbox"/> Elongation Limit			

**Figure 4.** The *Summary tab* provides detailed information about your datasets and refinement parameters. Parameters that vary from set to set are highlighted in cyan. The selector buttons can be used to change which information is displayed in the table. In this case, these two sets have been collected using the same strategy, but at different wavelengths. See **Checking and Editing Sets** on page 32.

## The Index Tab

File Options Site Configuration Crystal Information Report About Help

Project Data Auto Summary Index Strategy Integrate Scale Merge Macros

Pending Sets

1. apc5771\_se3.0###.img from 1 to 220  
2. apc5771\_se3\_inf.0###.img from 1 to 220

☒ Keep Current Values For All Sets

Resolution

Edge Half Corner

Min 50.00 Max edge

☐ Limit Resolution in Pre-refinement

40.0 20.0 8.0 4.0 2.0 Draw Circles

Refinement Options

Crystal	Detector	Crossfire
<input checked="" type="checkbox"/> Rot X	<input checked="" type="checkbox"/> Rot X	<input checked="" type="checkbox"/> X
<input checked="" type="checkbox"/> Rot Y	<input checked="" type="checkbox"/> Rot Y	<input checked="" type="checkbox"/> Y
<input checked="" type="checkbox"/> Rot Z	<input type="checkbox"/> Rot Z	<input checked="" type="checkbox"/> XY

Other parameters

<input checked="" type="checkbox"/> X Beam	<input checked="" type="checkbox"/> Y Beam	Fit All
<input checked="" type="checkbox"/> Cell	<input checked="" type="checkbox"/> Distance	Fix All
<input checked="" type="checkbox"/> Mosaicity		Fit Basic

Integration Box

Profile Fitting Radius 31.0

Box Size 36

Spot Size 0.30

Elongation ☐ Limit 0.7

More Options

Refinement Information

Space Group: P222

Resolution: 50.00 - 2.01

Positional:	861	X- $\chi^2$ :	1.61	Y- $\chi^2$ :	0.97	(0)
Partiality:	898	$\chi^2$ :	1.40			(0)

X Beam: 156.184 Y Beam: 160.877

a:	45.38	b:	48.82	c:	86.16
$\alpha$ :	90.00	$\beta$ :	90.00	$\gamma$ :	90.00

Crystal Rotation X: -121.977 0.000 0.015

Crystal Rotation Y: -170.983 0.000 0.003

Crystal Rotation Z: 4.347 0.000 0.004

Detector Rotation X: -0.003 0.000 0.022

Detector Rotation Y: 0.092 0.000 0.018

Detector Rotation Z:

Crossfire X: 0.051 0.000 0.012

Crossfire Y: -0.044 0.000 0.013

Crossfire XY: 0.023 0.000 0.017

Distance: 303.110 0.000 0.086

Mosaicity: 0.481 0.000 0.009

Ready

Controls

Index Peaks Limit 300 Sigma Cutoff Index 5.0 Refinement 5.0

3D Window 3

Peak Search on Set 1 Frame 1

Display Change Display to Frame 1

Index Frame 1

Refine for 5 Cycles on Frame 1

Bravais Lattice Check Mosaicity

Abort Refinement Reference Zone Crystal Alignment

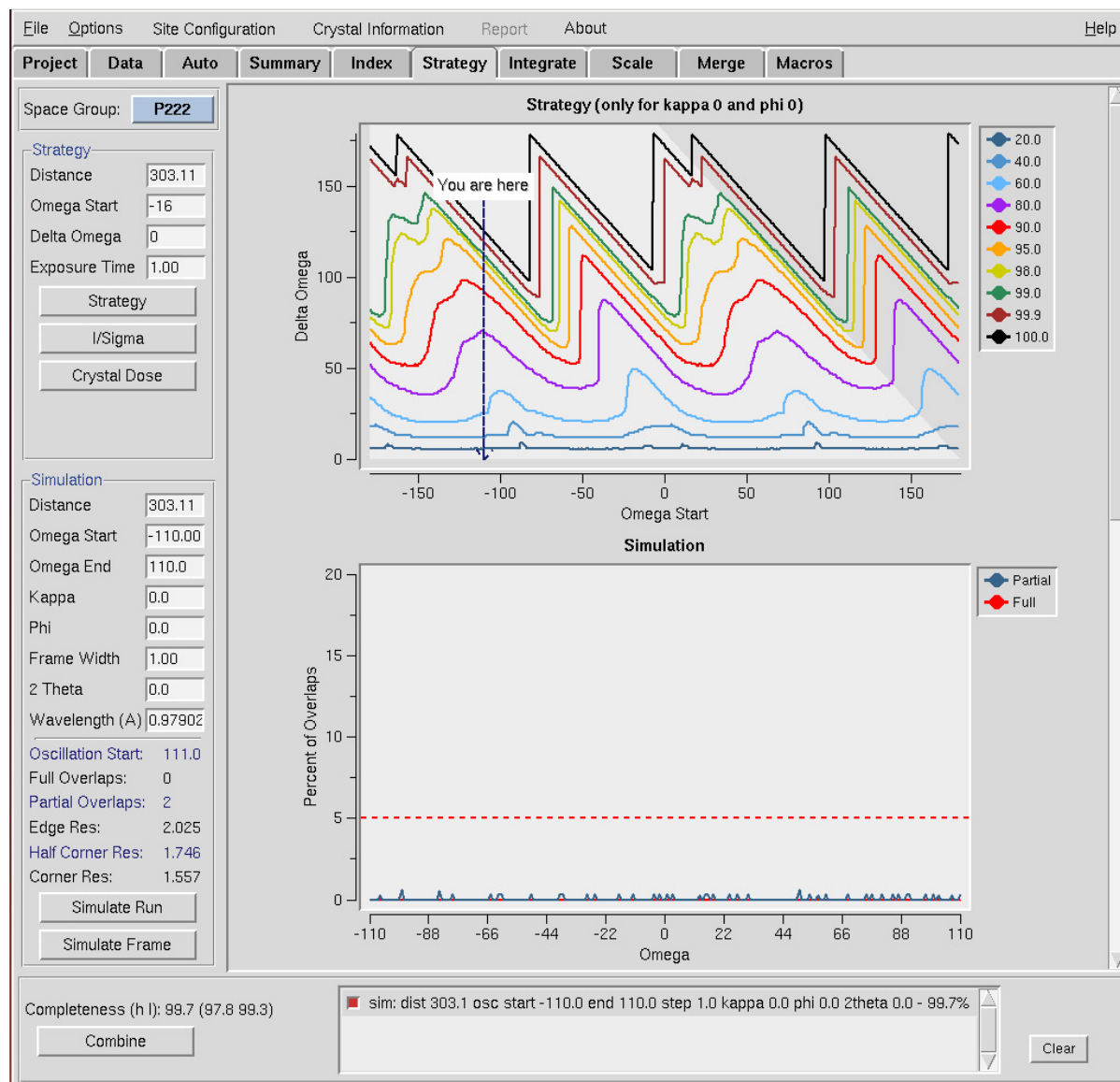
Integrate Integration Setup Cell Divider

Set Beam Position Find Beam Position Set Blind Region

Reject Criteria Update Site Check in PDB

**Figure 5.** The *Index tab* allows you to index your data and refine various parameters. This tab also includes buttons to mask the beamstop and specify the beam center. In this figure, the refinement has been completed, and this data is ready to be integrated. The left side of the display contains control buttons and parameters. The top right shows indexing statistics and crystal characteristics. The bottom right contains action control buttons and a few more parameters. See **Auto-Indexing** starting on page 52.

## The Strategy Tab



**Figure 6.** The Strategy tab can help you design your data collection experiment. An optimal initial orientation will give you a complete dataset before collecting redundant data, thereby minimizing complications due to crystal decay. Simulating a run will let you know if you should expect overlapping reflection, which can severely degrade the quality of your data. See **Strategy and Simulation** on page 78.

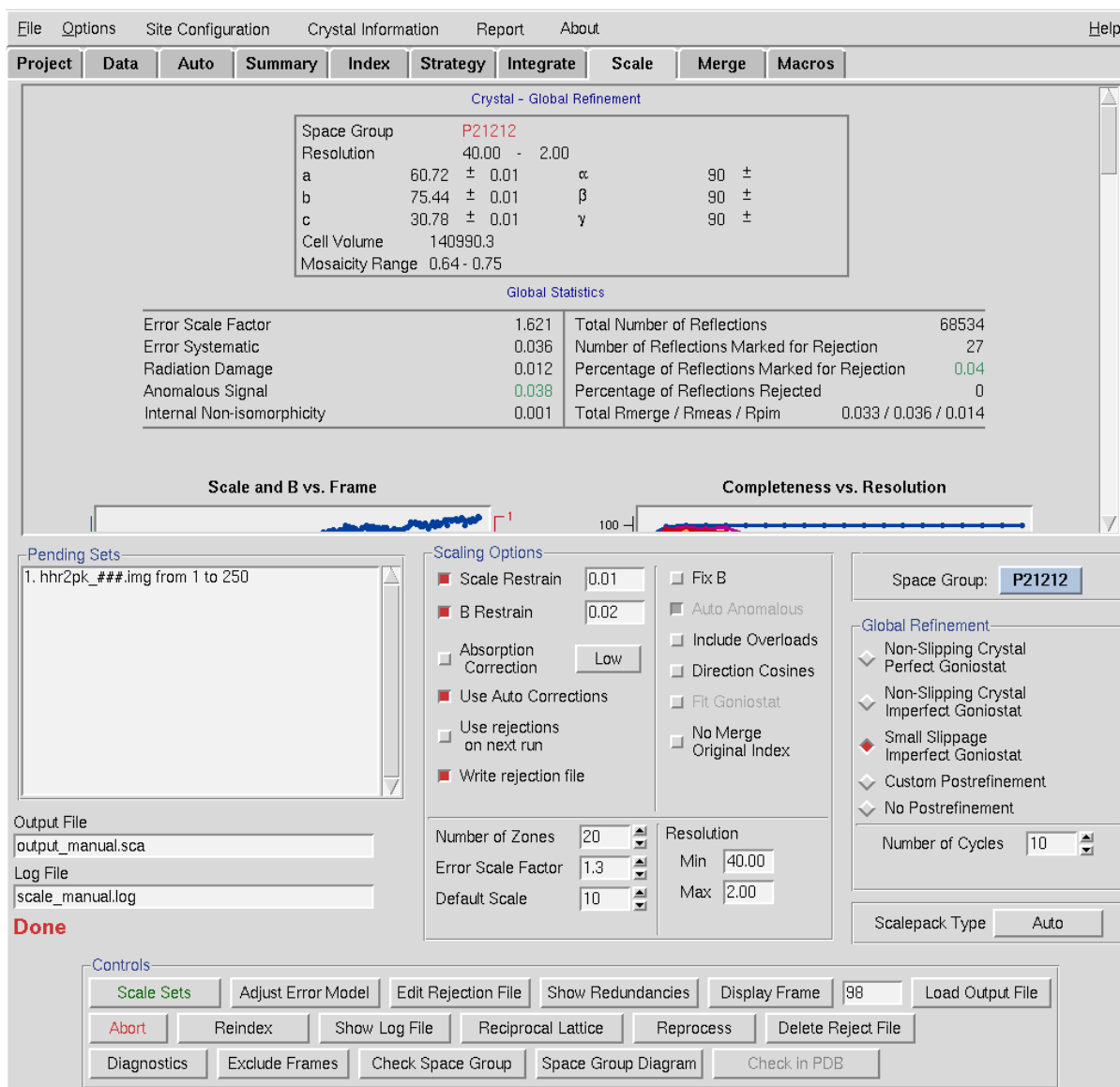
## The Integrate Tab



**Figure 7.** The *Integrate* tab updates while processing the individual frames. Refinement information and mosaicity distribution for each frame are shown along with a graph of how refinement parameters change over the course of the dataset. See **Integrating** on page 85.

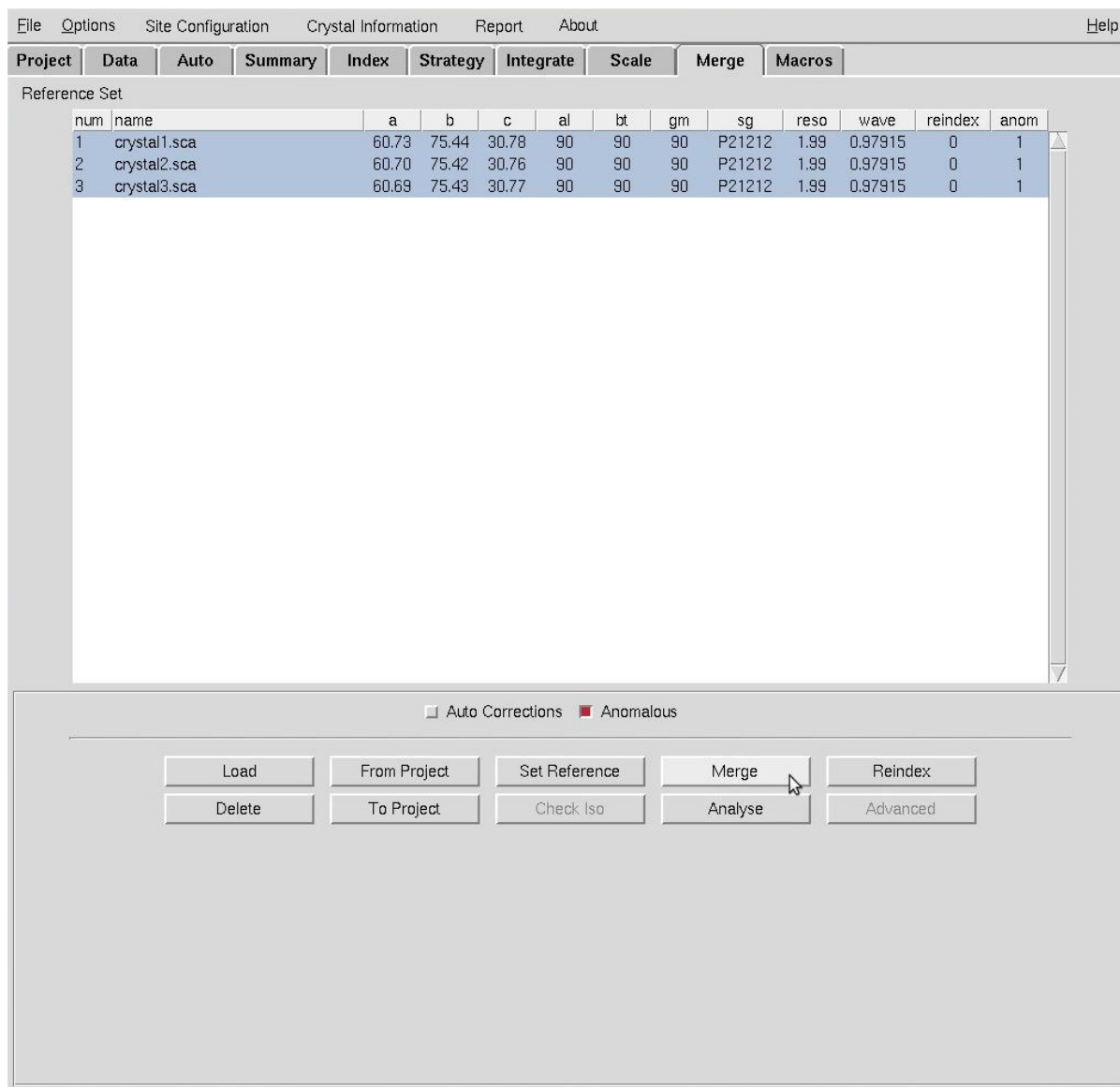


## The Scale Tab



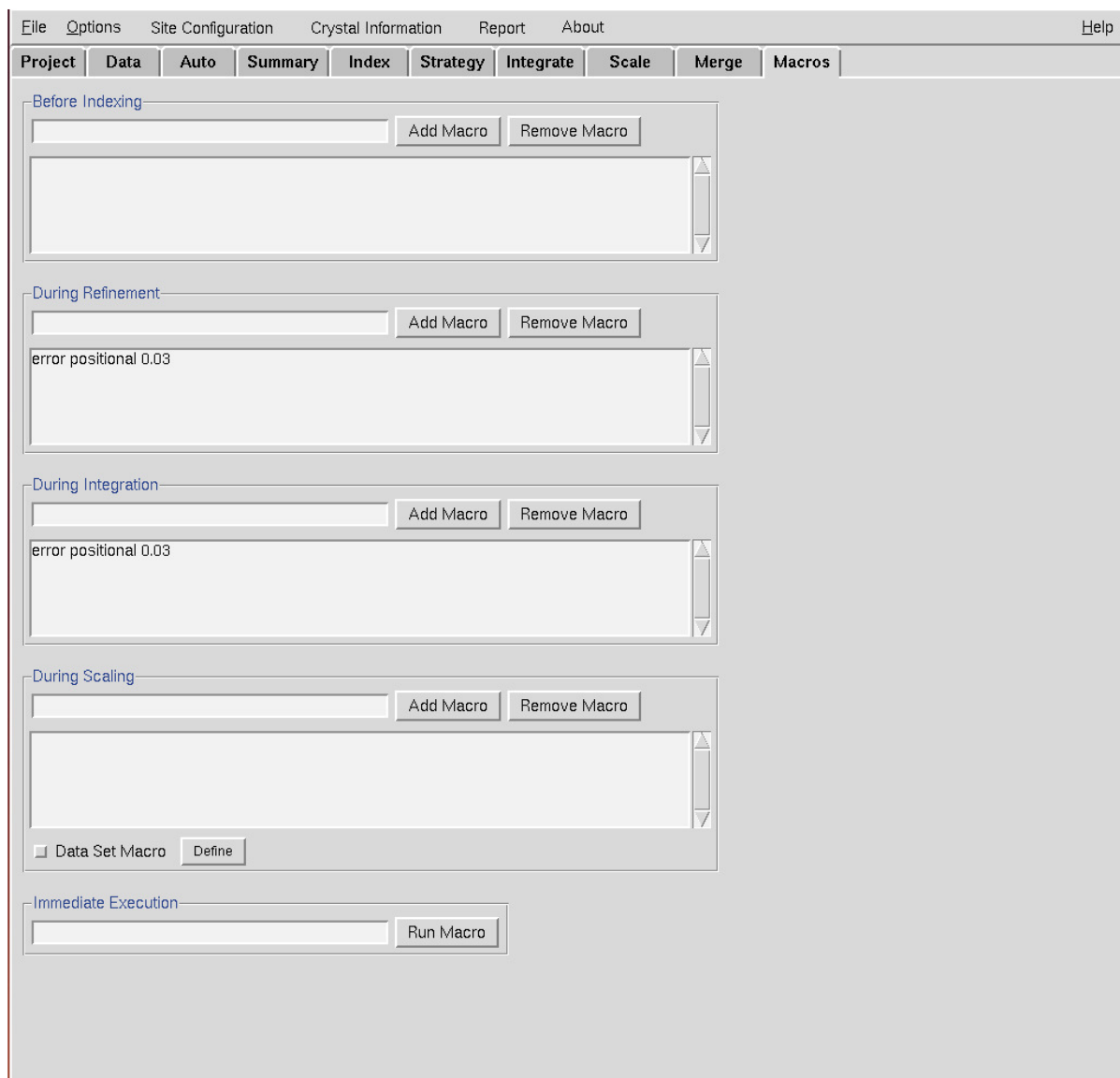
**Figure 8.** The *Scale tab* is where you combine all the information from the individual frames into one Scalepack file (called output\_manual.sca in this case), which contains a unique list of all reflections. Symmetry related reflections have been merged. The top panel is a scrollable selection of graphs that provide information about the data quality. The bottom pane lets you control various parameters or let you inspect results that are not initially displayed. See **Scaling** on page 91.

## The Merge Tab



**Figure 9.** The *Merge tab* can be used to scale and merge multiple individually scaled data sets. This allows you to merge data from different crystals or data that could not otherwise be scaled together. See **Merging Multiple Scaled Files** on page 125.

## The Macros Tab



**Figure 10.** The *Macros tab* allows you to give specific commands to HKL during various parts of data processing. The “error positional” macro is one of the more commonly used macros. It can significantly lower your positional  $\chi^2$  by allowing more “wobble room” for your reflections. This is often a detector dependent necessity. See **Using Macros** on page 127.

# *Conventions Used in This Manual*

The most critical steps and information that should not be overlooked are bold, blue, and underlined.

**HKL-2000** BUTTONS ARE IN A BLUE, SMALL CAPS FONT.

File names and paths are in a fixed-width font similar to a Linux console.

***Tabs*** and ***Options*** are bold and italic.

---

## **Deeper Information**

Some explanations and details are contained in text boxes like this one. Read the heading to decide if you need to read the information in the box.

## *Installing HKL-2000*

Installation of HKL-2000 uses an installer program that guides you through the installation process. The HKL installer and executables of all programs can be obtained from the HKL Research website (<https://hkl-xray.com>). Details of the installation process are contained in **Error! Reference source not found.** beginning on page **Error! Bookmark not defined.**

To install and run HKL-2000, you will need to obtain a license key from HKL Research. Details of the licenses are available on the website. Academic users who collect data at a synchrotron beamline where HKL-2000 is installed may obtain a 30-day license free of charge to reprocess their data. The license can be released for one computer up to 3 times per year.

To obtain the license key, you need to run the program (`access_prod`) program on each machine that you plan to install the *HKL-2000* programs. Details of this are also in the **Error! Reference source not found.**

# Starting HKL-2000

The installer places an HKL-2000 icon (Figure 11) on the desktop that will launch the program. The icon will set important environmental variables and launch a terminal window that will display HKL-2000's runtime information.



*Figure 11. The HKL-2000 Start Icon*

You may start HKL-2000 from a terminal if you wish. If the *HKL-2000* suite has been installed in the default location (`/usr/local`) you can typically just type `HKL2000`. The *HKL-2000* suite must be configured for your system, and a license key must be installed before it can be run. Installation and configuration of the suite are described in the Appendix. You must have permission to write files in the directory you start HKL-2000 because the program writes several files to the working directory.

## The Installation Directory

HKL-2000 can be installed in any location, but we strongly recommend the directory `c:\HKL` for Windows installations. The HKL Installer will ask where you would like HKL installed and will create an icon that will launch HKL-2000.

If you are working with an older version of the suite, you can use a non-standard directory if you set the environmental variable `$HKLPATH`. For instance, if all programs are installed in the directory `/home/janedoe/hkl2000` the following commands can simplify launching HKL. You may want to put these in your login scripts.

For Bourne-type shells (sh, bash, dash, etc.):

```
$ export HKLPATH=/home/janedoe/hkl2000
$ alias HKL2000="$HKLPATH/HKL2000"
```

For C-type shells (csh or tcsh):

```
$ setenv HKLPATH /home/user/prog2000
$ alias HKL2000 $HKLPATH/HKL2000
```

Then for both

```
$ HKL2000
```

## Site Definitions

To work with diffraction data, HKL needs to know details about the experimental setup. This information is stored as a data collection “site.” Therefore, the first dialog to appear asks you to choose a site definition (detector and other site information) from the menu (Figure 12). You must select a site definition (or use the SITE INCLUDED option) before continuing to the program.

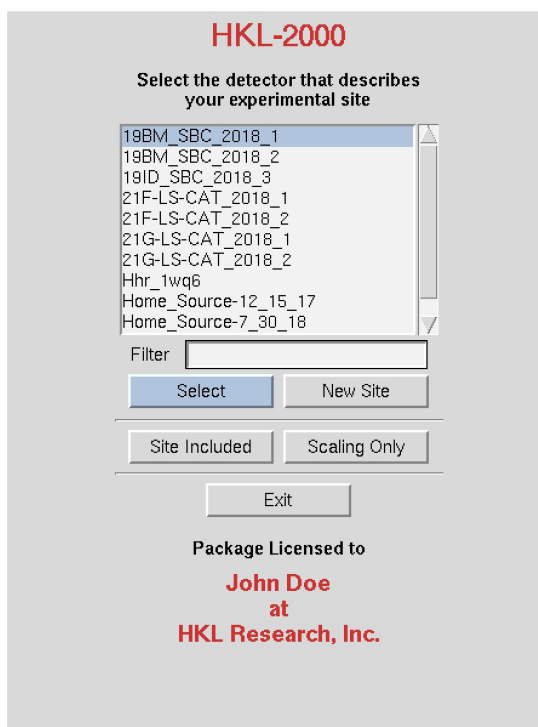
A site definition contains information about the experimental setup, such as the X-ray detector, goniostat, and X-ray source. The definition includes the detector name and type, the axes of rotation and misalignment parameters of the goniostat, the X-ray beam properties, and properties of the diffraction image files collected, especially the expected suffix for these files (e.g., \*.img, \*.osc, etc.). It contains the X and Y coordinates of the direct beam position on the detector, which are crucial pieces of information. The site files are traditionally stored in /usr/local/hklint but can be placed in any directory if the environmental variable \$HKLDIR is set. If your installation of HKL-2000 was created using the installation program, then the icon it creates will automatically set this variable. This directory will contain one subdirectory for each site. Each site definition directory contains a “site file” called def.site where most of the settings are stored. These subdirectories may also contain other files such as detector calibration data (in the form of files called def.cal and

---

### Files created by HKL

- **Filename.sca** contains the scaled and reduced data. This is the end product of data reduction and the input for structure solution programs. It is important for structure deposition!
- **scale.log** is important for structure deposition! It contains detailed information about data scaling. The most important tables are at the bottom of the file.
- **peaks.file** is the file that contains peaks selected by the Peak Search routine
- **reject** is the list of rejected reflections created in scaling
- **hklpredictions** is a binary file that contains a listing of predicted reflection positions
- **den.in** is a listing of the commands used to run the indexing
- **den.out** is a modified log file of the indexing
- **scl.in** is the *Scalepack* input file
- **xdisp.in** and **xdisp.out** are input and output files for the display program.
- **filename\_###.x** (dot X files) contain tabulated data describing reflections for each individual frame. Some experimental details are found at the bottom of these files (i.e. x-beam, y-beam, cell parameters, refinement parameters for this frame.)

denzo\_nonunf) which are necessary for processing uncalibrated data. For synchrotron data processing, these files should be obtained from the synchrotron beamline.



*Figure 12. The HKL-2000 launch and site selection screen.*

## The Site Included option

The **SITE INCLUDED** option tells HKL-2000 to use a **def.site** that is located in the same directory as the data. This option can be convenient when sharing data or downloading data from ProteinDiffraction.org.

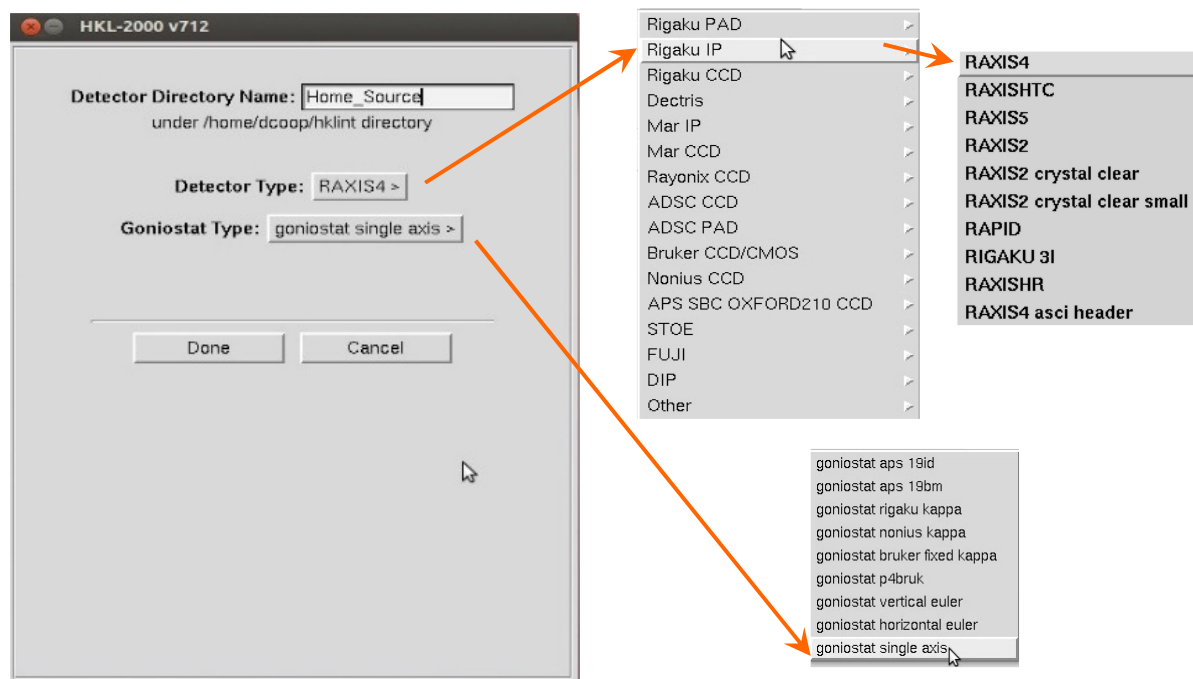
## Creating a new site

To make a new site, click on the new site button. A new window with a field to give the new site name will open. This will create a new subdirectory in /usr/local/hkclint or \$HKCLINT with the name you give (note that the user running the program will have to have permission to write to this directory). It's best to use a descriptive name that does not contain spaces. It can be helpful if the name indicates a time period, like home\_Raxis\_fall\_2018 or 19-ID\_2018\_run3 (Figure 13). Now click on the **DETECTOR TYPE** button, and select your detector. Some detector manufacturers have different models, and these are listed under each general detector heading. If your detector is not listed



## STARTING HKL-2000

at all (unlikely, but theoretically possible), you will have to contact HKL Research (or your beamline coordinator) for the appropriate site file.



**Figure 13.** The "New Site" selection panes

The most important values in the site file are the X-beam and Y-beam positions and goniostat type. Be sure to record them every time you collect data because they can fluctuate. The default values for newly created `def.site` files will almost always have incorrect values! You can change these by clicking on **SITE CONFIGURATION** (see below) and make changes using the dialog. While not recommended, it is also possible to edit the appropriate line in the `def.site` file in the newly created subdirectory and then restart the program. Collecting and processing data with a well-characterized crystal (such as lysozyme) can ensure you are using a correct site definition.

Choose the appropriate goniostat type from the predefined list. Presently the following types of goniostats are available:

- single-axis – most common
- vertical euler – goniostat used with Rigaku Rapid, Saturn and Raxis HTC detector
- horizontal euler -
- nonius kappa – goniostat used by Nonius KappaCCD and Kappa2000 systems

## STARTING HKL-2000

- rigaku kappa -
- bruker fixed kappa – used with Smart 1000 and Smart 6000 Bruker systems
- p4bruk – P4 goniostat used with Smart 6000

Once you've selected the correct detector and goniostat, [click OK](#). The new site definition directory will be created with the appropriate files, including a `def.site` file. You will now get the main **HKL-2000** window.

**Make sure the site file is correct!** Click on **SITE CONFIGURATION** on the top menu bar, and this will bring up a window with the site parameters (Figure 14 and Figure 15 ). You can verify that the detector is the right one, and you can **specify important parameters like the X Beam and Y Beam Positions** and the Wavelength and so on.

### Features of the site creation dialog:

1. You can load in data from an existing site file by clicking on the **LOAD SITE INFO** button. This brings up a file navigation window where you can choose a site file (`def.site`). This is very useful for just changing the beam position.
2. The anode type (and thus the wavelength) for rotating anode X-ray sources is specified in this window, while the wavelength for synchrotron X-rays is specified by data set in the main **HKL-2000** window (and is hopefully, but not guaranteed, to be correctly specified in the image file header – check to make sure!).
3. Clicking on **SAVE SITE INFO** means that you are permanently editing or updating the file called `def.site` (in `/usr/local/hklint/sitewhatever`). For this you will need a password (which is provided to the “*Contact Person*” your group has specified with HKL Research). This is not a typical user option. Also, the user running the program must have permission to write to the directory that contains the site file (usually `/usr/local/hklint/sitewhatever`).
4. You can use your updated site parameters without saving them to the `def.site` file by clicking the **CLOSE** button. The altered site parameters will remain in effect until you quit the program, after which time they will revert to the original values.

**Site Configuration**

**General** | **Advanced** | **Description**

Detector: Q4/Q4R transformed  
 Detector Directory: HHR

Rotation X: 0.013  
 Rotation Y: 0.041  
 Rotation Z: 0.00

☐ Use Calibration File  
☐ Use Sensitivity File

Beam Position X: 96.503  
 Beam Position Y: 93.645  
 Crossfire X: -0.018  
 Crossfire Y: 0.031  
 Crossfire XY: -0.066

☒ Synchrotron ☐ Neutron Beam  
☐ Cu ☐ Mo ☐ Cr ☐ Ag  
 Polarization: 0.99  
☐ Graphite ☐ Mirrors

**Goniostat Single Axis**  
 Alignment: 0 0 0  
 2Theta Direction: ☒ + ☐ -

Predefined File Template: \*.img

☒ Blind ☐ Shadow Region

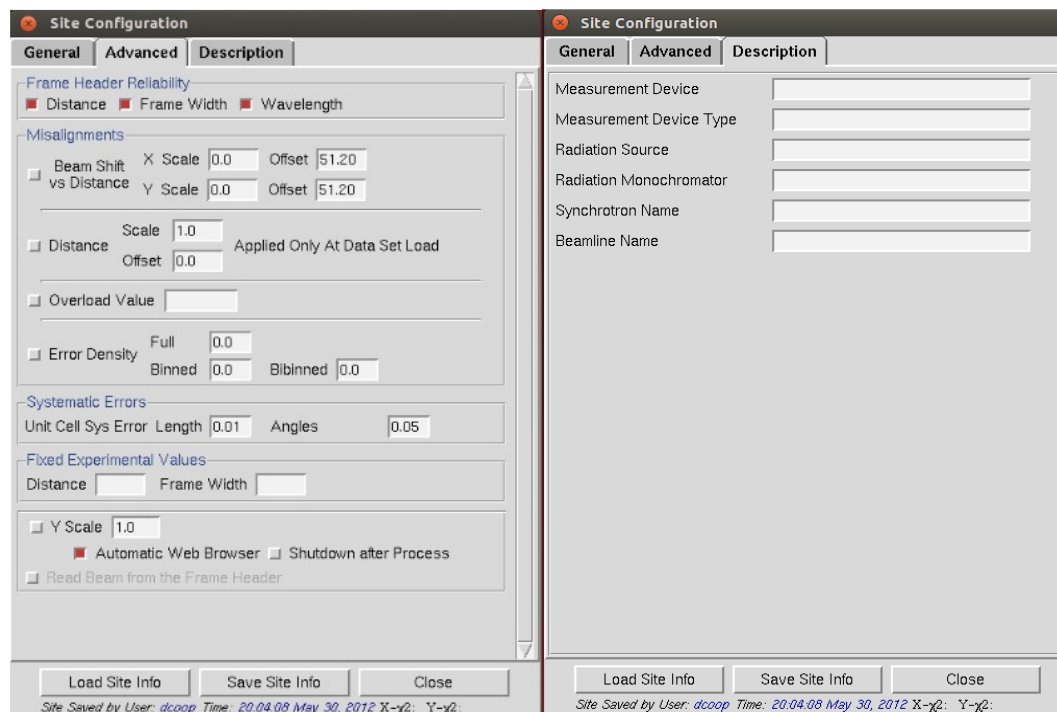
**first area:**  
☒ 92.8 0.3 92.6 95.3  
☐ 101.4 95.8 104.8 0.5  
☒ 97.0 97.0 8.1

**second area:**  
☒  
☐  
☒

Load Site Info | Save Site Info | Close

Site Saved by User: dcoop Time: 20:04:08 May 30, 2012 X- $\chi^2$ : Y- $\chi^2$ :

*Figure 14. The main dialog for site creation.*



**Figure 15.** The Advanced and Description tabs of the site configuration window.

### What if I choose the wrong site because I wasn't sure?

If you select the wrong site file, the display most likely will be distorted, which is what happens if you have selected the wrong detector type. If you select a site that specifies the same detector type as your images, the display will look normal, but the indexing and integration will fail if other parameters (i.e. the beam position) are incorrect. If you are lucky and the detector is correct and the goniostat description is compatible, you may be able to process your data.

### What is the meaning of the goniostat alignment windows?

The windows describe the six mis-setting angles of a particular goniostat. The choice of the single axis goniostat allows only the first two mis-setting angles to be entered:

- 1) Angle around the beam
- 2) Angle describing the non-perpendicularity of the angle between beam and omega/spindle axis

The other goniostat types require all six mis-setting angles to be defined. The four remaining angles are:

- 3) Omega zero
- 4) Deviation between nominal and actual kappa (chi) – omega angle
- 5) Kappa zero
- 6) Deviation between nominal and actual kappa (chi) – phi angle

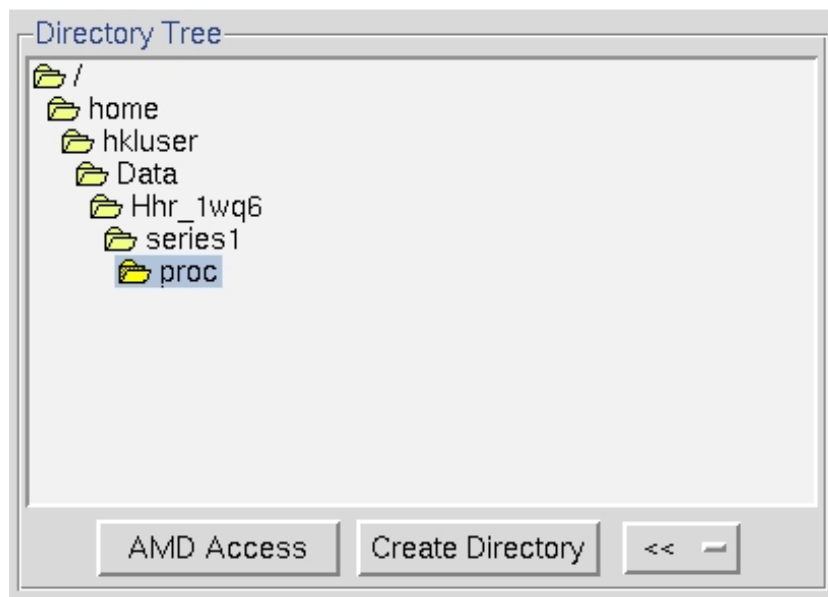
A change of the spindle axis direction requires a corresponding change of the first mis-setting angle by 180 degrees. This is equivalent to the Denzo command **film rotation 180**.

# Data, Results, and Datasets

Select your data files using the Directory Tree and the New Raw Data Directory field.

## The Directory Tree Window

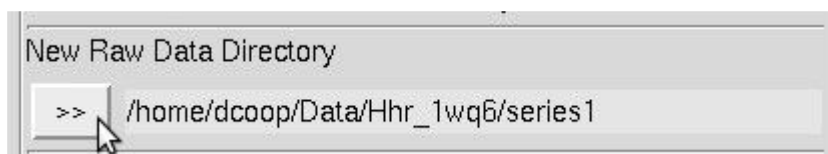
Double clicking on a file folder displays the subdirectory contents of that directory. It will not display the files in that directory, so don't be puzzled when you can't see your data files. To go up one or more directories, double click on a parent directory. To make a new subdirectory, single click on the folder under which you want to create a new directory, click on **CREATE DIRECTORY** and a dialog box will open with the path displayed. Append your subdirectory name onto the end of this path and click **CREATE**. If your data is in a directory configured to auto mount but is not currently mounted, click **AMD ACCESS** (auto mount directory), type the name of the directory and click **ACCESS** to mount (Figure 16).



*Figure 16. The directory tree*

## Select the raw data directory

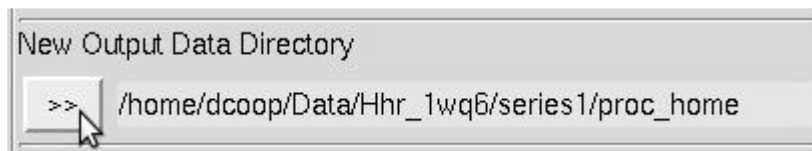
Single click the subdirectory that contains your data (which will highlight it), then click the “»” button to the right under New Raw Data Directory. The full path will be written in the field next to “»” button (Figure 17).



*Figure 17. The Raw Data directory*

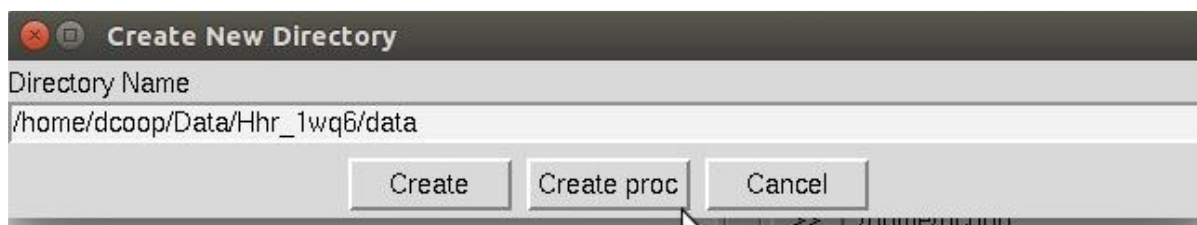
## Choose an output file directory

Select a directory for your output files using the Directory Tree and NEW OUTPUT DATA DIRECTORY (Figure 18). The suggested convention is to have a subdirectory for your output (historically called proc) in the directory that holds your frames. The nice thing about having proc subdirectories is that if you reprocess the data multiple times (say, in different space groups) you can have a different proc directory for each attempt (e.g., proc\_home, proc\_p222, proc\_p21, proc\_p6, etc.) and not worry about overwriting your work or having to make new file name patterns.



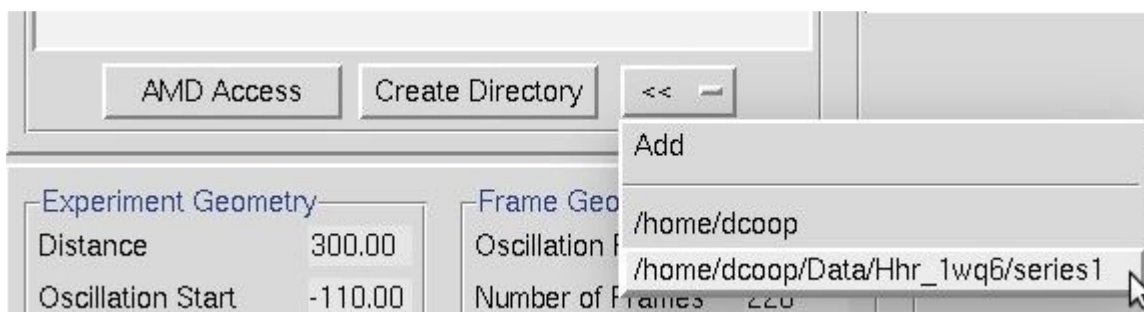
*Figure 18. The Output Data directory*

To make a new subdirectory, single click on the folder under which you want to create a new directory, click on **CREATE DIRECTORY** and a dialog box will open with the path displayed (Figure 19). Append your subdirectory name onto the end of this path and click **CREATE**. This will create the directory and select it in the directory tree. A shortcut button **CREATE PROC** will create a directory named “proc” (the traditionally named processing directory) in the selected directory.



*Figure 19. The Create New Directory dialog*

The double arrow button (<<) can be used to create a shortcut to a directory to facilitate quickly selecting a directory. This may be useful if you have to select multiple data sets that are in various places. To add a shortcut, click the **ADD** button. To switch to one of the shortcuts, select the shortcut from the list.



**Figure 20.** The << button can make directory shortcuts.

### Do I need a "proc" directory or can I put the files anywhere?

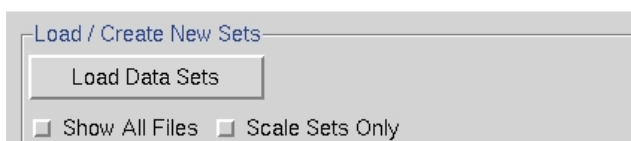
As long as you have permission to write to a directory, HKL can write your output there. However, **it is a good practice to have the diffraction images separate from the output files**. Some labs keep their diffraction images on a networked drive, yet save the processed files on individual workstations.

### How much disk space am I going to need for this?

Each indexed raw data file yields a new file, called an \*.x file, with the extracted reflection data from the frame in it. The size of the \*.x files is therefore linearly proportional to the number of reflections on the image. As you might imagine, the \*.x files for a lysozyme data set would be quite a bit smaller than those for a ribosomal subunit data set if both crystals diffract well. Generally, the \*.x file size depends on the unit cell dimensions and resolution. You always should have at least a few gigabytes of free disk space.

## Selecting datasets

*Select the frames you wish to work with using the load data sets button (Figure 21) and clicking on one or more sets of frames.*

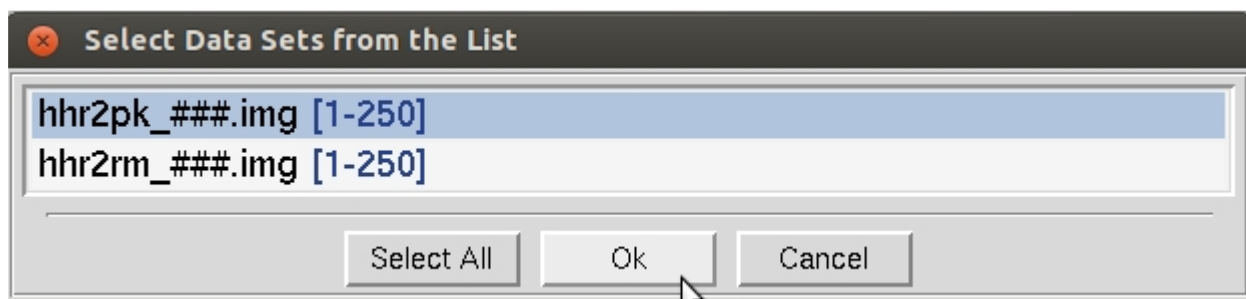


**Figure 21.** The Load/Create New Sets window

**Avoid the naming pitfall**

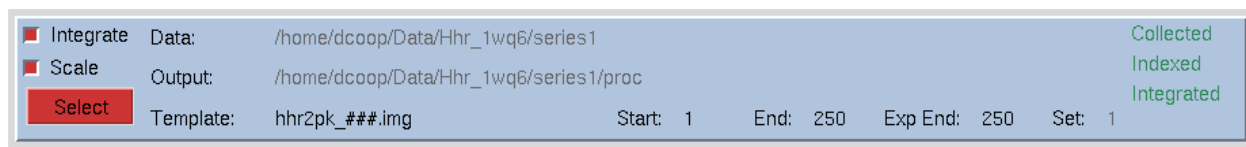
When collecting diffraction data, avoid creating frames with names ending in numbers, because the set parser cannot distinguish a contiguous data set among the frames. For example, if the root of your data file names is nc401, and then you collected frames 001 through 140, you might expect files called nc401001.img through nc401140.img. Unfortunately, it is very difficult for a computer to parse this type of file name and distinguish between: 1-140, 1001-1140, or 401,001-401,140 frames. You may keep the root nc401, but **make sure to have non-numeric characters before the sequential part of the file name** (for example: nc401\_0001.img or nc401\_set 2\_0001.img )

If you have not clicked on one of the checkboxes below the **LOAD DATA SETS** button, you should see what the program considers to be the sets of data frames in the **New Raw Data Directory**, with # signs standing for digits of the frame numbers (Figure 22). **Single click on the set of frames you want to process** - it will turn blue - then hit **OK**. The “Select All” button has been added for convenience.



*Figure 22. The Load Data Sets window*

When the set has been added, a blue box will appear in the main window summarizing the properties of the set of frames you have added (Figure 23). Once you have loaded the different sets of data, the **SELECT** button at the bottom left of the set box allows you to see more details about the set or manipulate the set.



*Figure 23. The Single Set window*

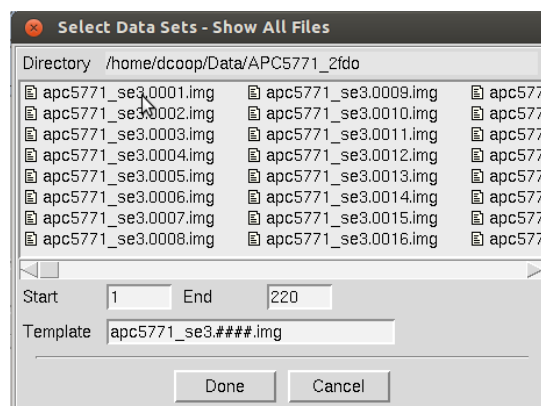


### What if I don't see any frames listed?

This means that either you have no data frames in the **New Raw Data Directory**, or else they are not named and numbered in such a way that the program can identify them as data frames. For example, your files may be of the form \*.mccd, but the **Site Configuration** may specify data files end with the suffix \*.img. This is controlled by the “**File Template.**” You can edit the **File Template**, which describes the naming convention for data files, in the **Site Configuration** dialog. Alternatively, you can select files using the **show all sets** dialog.

### What does Show All Sets do?

This option will open a file browser in the **New Raw Data Directory** (see [Figure 21](#)). If no sets appear in the default **Load Data Sets**, but you see your files when you select **show all sets**, then the **File Template** is probably wrong. You can use this option to temporarily override the “**File Template**” set in the **Site Configuration**. When you click on a file, HKL-2000 will deduce a **File Template** and determine the range of files that exist. You may edit this range to only work with a portion of the files.



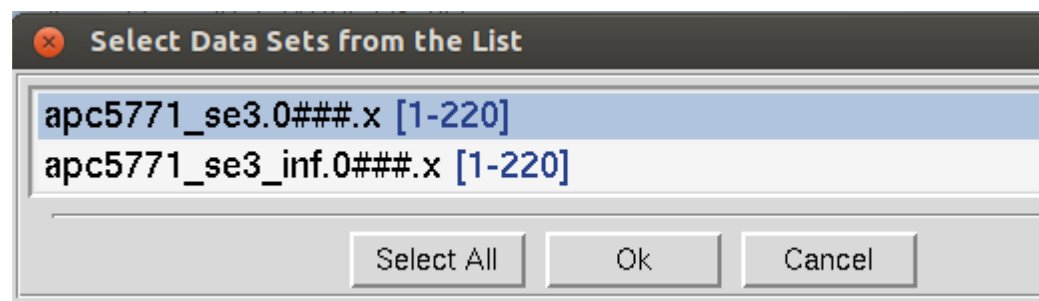
*Figure 24. The Show All Files dialog can be used for file selection.*

### What if the File Template isn't right?

If you've already collected data with ambiguous file names, you don't need to manually change all the file names! Consider the files ranging from nc401001.img to nc401140.img described above. Load in the data set as described above, even if the parser does not extract the range of frame numbers properly. Go to the *Edit Set(s)* dialog, change the name in **File Template** from nc#####.img to nc401###.img. The program will parse the range of frames and a dialog box will ask for approval. After selecting **ok**, the proper values of **First Frame Number** and **Number of Frames** should appear and **File Template** will be automatically adjusted according to the new scheme.

### What is the Scale Sets Only option?

This will display only the \*.x files that are in the **New Raw Data** or **New Output Data** directory paths (Figure 25). You might want to do this to scale data already indexed, instead of reindexing and reintegrating the raw frames. You may also want to do this to rescale the data in a different space group within the same crystal class.



*Figure 25. Files listed after activating the Scale Sets Only button*

## Removing sets

If you would like to remove a set, click the **SELECT** button of the set you want to remove. Then click **REMOVE SET** in the *Set Controls* panel on the right side of the main window. The program will ask if you really want to do this, and if so, the set will be removed.

## Integrate, Scale, and Collect checkboxes

The checkboxes on the left of the data set window tell the program what you want to do with the data in this set. Most installations of HKL-2000 will have checkboxes to index and integrate the frames (**INTEGRATE**) and scale the data (**SCALE**). Some installations will also have a checkbox to collect the data (**COLLECT**). The default is for all the available options to be selected when the set panel is displayed. Control of data collection using *HKL-2000* is currently only supported at the Structural Biology Center (SBC-CAT) beamlines at APS and for a few selected detectors. **If your configuration of HKL-2000 does not support direct data collection, the collect checkbox may be missing.** If you want to scale frames that have already been indexed and integrated, just unselect the **INTEGRATE** checkbox, but leave the **SCALE** checkbox selected, and the program will know that you are passing that data set directly to *Scalepack*. If you are working with data from more than one wavelength, you can select **INTEGRATE** for all the sets, but you have to scale the sets independently. Integrating all the sets at once will ensure that they have the same crystal orientation parameters.

### Working with multiple data sets

You can load multiple sets of data within the **SELECT** data window. Clicking once selects, clicking again deselects. You do not have to hold down any additional keys to select more than one – just click on each set you want.

Selecting multiple sets is extremely useful when you have collected multiple data sets from a single crystal and want to index and process them all together. Examples of this include:

1. multi-wavelength experiments,
2. inverse-beam experiments,
3. data sets which differ by the  $2\theta$  offset of the detector, or
4. low-pass and high-pass data sets that differ in the crystal-to-detector distance.

In all cases, if the goniostat is aligned (and this alignment is described in the site configuration [def.site file]), the crystal has not been moved relative to the goniostat, and the header information of the image files is reasonably accurate (so that detector distance and position information is present), then you should be able to process all of the frames together from a single auto-indexing. This will ensure that the indexing from all sets is consistent.

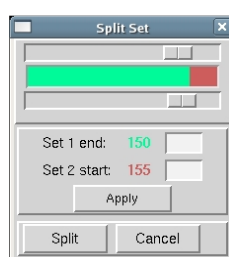
## Splitting sets

Notice that under the panel called **Set Controls**, there is an option called **SPLIT SET**. This splits your data set into two or more groups, which can then be integrated consecutively, independently, or selectively. One reason to do this is if you had a bad or blank frame

## DATA, RESULTS, AND DATASETS

somewhere in the series of frames. Additionally, when screening crystals, you may have two orthogonal images which results in non-consecutive images with consecutive file numbers. In general, sets are split whenever there is an interruption in a contiguous series of data collection. To split a set, select the desired set and then click on the **SPLIT SET** button in the **Set Controls** panel. A new box will appear that has two sliders (Figure 26). Move the top slider to define the end of the first set, and then move the bottom slider to define the start of the second set. If you want to skip one or more frames, just make sure that they are in the “gap” between the two sets. When done, click on the **SPLIT** button, and a new set will appear (Figure 27). Make sure that the **Distance** and **Oscillation Start** values are correct for each set (they should be read from the header of the data files, but you never know).

In principle, you can split a data set as many times as you need to; however, keep in mind that 3D integration cannot work across the boundaries between split sets.



**Figure 26.** *The Split Set window*

<input type="checkbox"/> Collect	Detector:	RAXIS4	Collected
<input checked="" type="checkbox"/> Integrate	Data:	/home/anna/Data/RaxisIVpp	Indexed
<input checked="" type="checkbox"/> Scale	Output:	/home/anna/Data/RaxisIVpp/proc1	Integrated
<input checked="" type="checkbox"/> Select	Template:	Lysozyme0###.osc	
	Start:	1	End: 150
	Exp.End:	150	Set: 1

<input type="checkbox"/> Collect	Detector:	RAXIS4	Collected
<input checked="" type="checkbox"/> Integrate	Data:	/home/anna/Data/RaxisIVpp	Indexed
<input checked="" type="checkbox"/> Scale	Output:	/home/anna/Data/RaxisIVpp/proc1	Integrated
<input type="checkbox"/> Select	Template:	Lysozyme0###.osc	
	Start:	155	End: 180
	Exp.End:	180	Set: 2

**Figure 27.** *Data set divided into two sets removing four images in the middle.*

## Checking and Editing Sets

When you select each data set, the program will display the set’s parameters at the bottom of the window **under Experiment Geometry, Frame Geometry, Exposure Time and Options** information panels. Make sure the parameters for each set are correct. These values are taken from the image’s header, but the header is not always correct. Changes can be made in the **Summary** tab or the **Edit Sets** dialog, both described below. There are three ways you can do this. The quickest method to see the parameters is to look at the experimental details at the bottom of the Data tab (Figure 28). This information will correspond to the “selected” set. The information in this frame is not editable.

## DATA, RESULTS, AND DATASETS

Experiment Geometry		Frame Geometry		Exposure Time		Options	
Distance	127.38	Oscillation Range	90.0	Time / Frame	181.0 Sec	Wavelength	0.71074 A
Oscillation Start	0.00	Number of Frames	90		3.017 Min		17444.4 eV
Omega	0.00	Frame Width	1.00	Time / Degree	181.0 Sec	<input type="checkbox"/> Narrow Frame	
Chi	20.00	Rotation Axis	Omega		3.017 Min		
Phi	0.00			Time / Set	04:31:30		
2 Theta	0.00						

**Figure 28.** The experimental details panel

The second way that you can view set parameters is to switch to the **Summary** tab (Figure 29). The information in this tab is editable. The **View Controls** panel is beneath the sets you have and allows you to select the information you want to be displayed in the panel at the top.

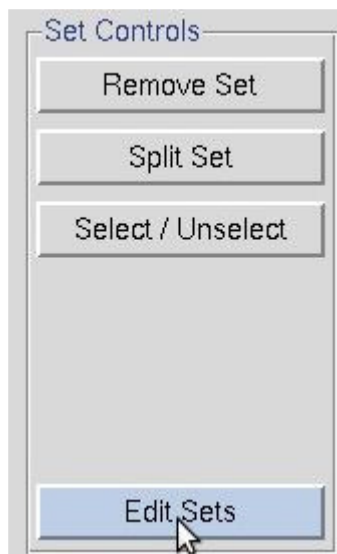
File Options Site Configuration Crystal Information Report About Help													
Project Data Auto Summary Index Strategy Integrate Scale Merge Macros													
Set	Raw Data Template	Start	End	Osc Start	Osc Width	Osc Range	Phi	Distance	Res Min	Res Max	3D Window	Profile Fitting Radius	Wavelength
	Default Parameters			0.00			0.00		50.00	edge	3	30.0	
1	apc5771_se3.0###.img	1	220	-110.00	1.00	220.0	0.00	300.00	50.00	edge	3	30.0	0.97902
2	apc5771_se3_inf.0###.img	1	220	-110.00	1.00	220.0	0.00	300.00	50.00	edge	3	30.0	0.97919

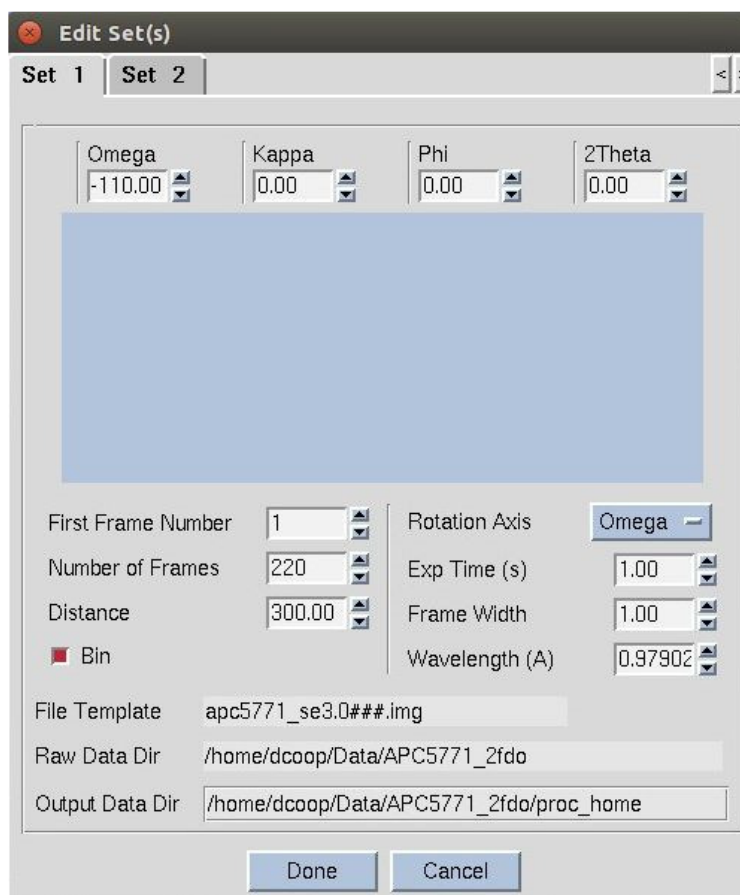
View Controls									
<input checked="" type="checkbox"/> Set	<input checked="" type="checkbox"/> End	<input type="checkbox"/> Time/ Frame	<input type="checkbox"/> 2 Theta	<input type="checkbox"/> Mosaicity					
<input checked="" type="checkbox"/> Raw Data Template	<input checked="" type="checkbox"/> Osc Start	<input type="checkbox"/> Time/ Degree	<input checked="" type="checkbox"/> Distance	<input checked="" type="checkbox"/> 3D Window					
<input type="checkbox"/> Output Data Template	<input checked="" type="checkbox"/> Osc Width	<input type="checkbox"/> Kappa	<input checked="" type="checkbox"/> Res Min	<input checked="" type="checkbox"/> Profile Fitting Radius					
<input checked="" type="checkbox"/> Start	<input checked="" type="checkbox"/> Osc Range	<input checked="" type="checkbox"/> Phi	<input checked="" type="checkbox"/> Res Max	<input checked="" type="checkbox"/> Wavelength					
<input type="checkbox"/> Spot Size M.axis	<input type="checkbox"/> Spot Size m.axis	<input type="checkbox"/> Box Size X	<input type="checkbox"/> Box Size Y	<input type="checkbox"/> Spot Angle					
<input type="checkbox"/> Energy	<input type="checkbox"/> Elongation Limit								

**Figure 29.** The Summary tab showing two datasets

The third way in see the experimental details of your sets is the **Edit Sets dialog in the Data tab**. The **EDIT SETS** button (Figure 30) will bring up a dialog with a different tab for each set of data (Figure 31). If you edit information about the sets, click **DONE** to save the information. Use the **CANCEL** button if you have made changes but do not want the information saved.



*Figure 30. The Edit Sets button is in the Set Controls box*



*Figure 31. The Edit Sets dialog*

## Start working before all frames are collected

You don't have to wait until you have all the frames to get started processing your data, which can be useful if you are processing data as it is being collected or transferred to your computer. In the *Edit Set(s)* window, enter the correct **Number of Frames** you are planning to process in this batch. You will notice that as you enter this number, the **Number of Frames** value will increase in the *Frame Geometry* window, as well as the **Exp. End** in **SELECT** panel at the top of the page. Notice as well that the **End** value in these two fields does not change -- this is the number of frames that the program knows is on hand at the moment. For example, suppose you are collecting 100 frames of data, and you have 20 collected so far. You would enter 100 in the **Number of Frames** box before you started the integration of the 20 frames collected so far. You could then start the integration job, and the first 20 frames will be integrated. When the last frame on hand is processed, the program will wait -- giving the message "Waiting for Frames" -- and when the new frames are written to disk, the program will continue to process them. This is especially helpful at places like synchrotrons, where the frame integration rate is a substantial fraction of the data collection rate, and you want to maximize your efficiency.

Another way you can use the **Number of Frames** value is to stop the integration before all the frames are processed. This is helpful if you have a lot of frames that have lots of spots and so take a long time to integrate. Perhaps you just want to process a wedge of this data and get some idea of its quality. For example, say you have 300 frames but just want to look at the first 30. Enter 30 in the **Number of Frames** box and start processing. The first 30 frames will be integrated, and then the program will stop. This is easier than splitting the data set and more civilized and nicer than hitting the **ABORT** button.

# Create, Save, and Load Projects

## Create and edit projects and crystals

HKL-2000 provides a convenient way to keep track of the proteins, data sets, and results of your projects. Although some of the features in the Project tab are more useful in database linked versions of HKL-2000 and in all versions of HKL-3000, which requires a database module license and/or an HKL-3000 license, the basic functionality of the Project tab can make working with your data easier.

When you first open the Project tab, most of the tab will be empty. The Project box will contain the default project name (project 1), the default crystal name (crystal 1), your user name, and the date (Figure 32). To create a new project, click either the **NEW PROJECT** or the **EDIT PROJECT** button. This will allow you to fill in information about the name of the project, a description of the project, what type of molecule you are working with (protein, small molecule, or virus), the proteins sequence, and the source of any anomalous signal (if present). If you are using a data collection capable version of HKL-2000, your collected images will be stored in a directory named for this project.

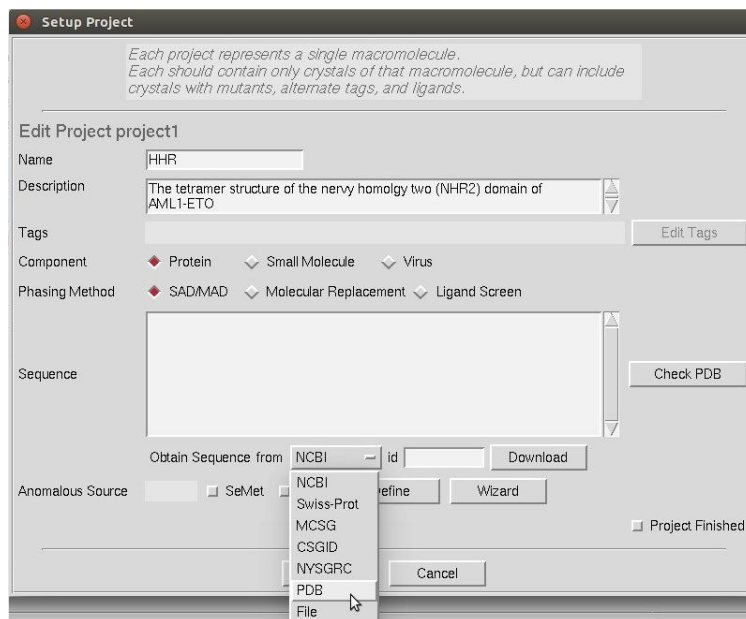
Project	Data	Auto	Summary
Project			
Name	project1		
Crystal	crystal1		
Experimenter	dcoop		
Date	Aug 27, 2018		
<div>Load Save</div> <div>Edit Project Edit Crystal</div> <div>New Project New Crystal</div> <div>Evaluate Model</div>			

*Figure 32. Starting a project*

The protein's sequence can be imported from several databases. Select the database using the drop-down menu labeled *Obtain Sequence from* and type the id associated with that database in the *id* field (Figure 33). Click **DOWNLOAD** to fetch the sequence.



## CREATE, SAVE, AND LOAD PROJECTS



**Figure 33.** Obtain a sequence from a database.

### Pasting in HKL-2000

If you need to paste information into HKL-2000, you may find that Ctrl-v doesn't work depending on your operating system. If you run into this, you will probably be able to paste using Shift-Insert key combination.

The Project window also has to be used to specify the source of an expected anomalous signal. There is an Anomalous Wizard button in both the **Edit/Create Project** dialog and the main **Project** tab. If you select SeMet or Sulfur, the number of those residues present in the sequence will be filled in. The Define button will display a periodic table that will let you select any element, but then you should enter the number of expected 'heavy' atoms. The Wizard button will display a periodic table with the atomic absorption edges of elements that are accessible at synchrotrons (Figure 34). Selecting an element will display its X-ray anomalous scattering plot (Figure 35). Move the mouse in the window to display the anomalous scattering coefficients ( $f'$  and  $f''$ ) for the energy (and wavelength) at the position of the crosshair (Figure 35). This can help you determine the wavelength to collect your data or the expected anomalous signal for the data you have already collected.

## CREATE, SAVE, AND LOAD PROJECTS

X-ray Absorption Edges v122

H																	He
Li	Be											B	C	N	O	F	Ne
Na	Mg											Al	Si	P	S	Cl	Ar
K	Ca 3.0704	Sc	Ti	V 2.2687	Cr 2.0701	Mn 1.8961	Fe 1.7433	Co 1.6083	Ni 1.4879	Cu 1.3808	Zn 1.2837	Ga 1.1959	Ge	As 1.0448	Se 0.9795	Br 0.9202	Kr 0.8655
Rb 0.8157	Sr	Y	Zr	Nb	Mo 0.6199	Tc	Ru	Rh	Pd	Ag	Cd 3.0857	In	Sn	Sb 2.6389 2.8304 3.0004	Te 2.5102 2.6883 2.8559	I 2.3898 2.5553 2.7207	Xe 2.2738 2.4293 2.5926
Cs 2.1697 2.3134 2.4738	Ba 2.0703 2.2047 2.3630	La 1.9786 2.1048 2.2614	Hf 1.1001 1.1545 1.2968	Ta 1.0614 1.1134 1.2548	W 1.0247 1.0740 1.2147	Re	Os 0.9561 1.0011 1.1405	Ir 0.9240 0.9668 1.1055	Pt 0.8933 0.9341 1.0722	Au 0.8638 0.9028 1.0402	Hg 0.8355 0.8726 1.0093	Tl 0.8079 0.8436 0.9795	Pb 0.7817 0.8157 0.9511	Bi	Po	At	Rn 0.8481 2.7663 2.9811
Fr	Ra	Ac	Rf	Ha													
		Ce 1.8932 2.0114 2.1663	Pr 1.8140 1.9251 2.0788	Nd 1.7399 1.8446 1.9972	Pm 1.6692 1.7680 1.9195	Sm 1.6025 1.6957 1.8460	Eu 1.5398 1.6277 1.7771	Gd 1.4803 1.5634 1.7118	Tb 1.4238 1.5025 1.6500	Dy 1.3706 1.4449 1.5916	Ho 1.3198 1.3903 1.5362	Er 1.2715 1.3383 1.4834	Tm 1.2257 1.2892 1.4337	Yb 1.1823 1.2426 1.3863	Lu 1.1406 1.1981 1.3412		
		Th 2.3925 2.5667 3.0643	Pa	U 2.2348 2.3925 2.8811	Np	Pu	Am	Cm	Bk	Cf	Es	Fm	Md	No	Lr		

Close

Figure 34. The anomalous element wizard.

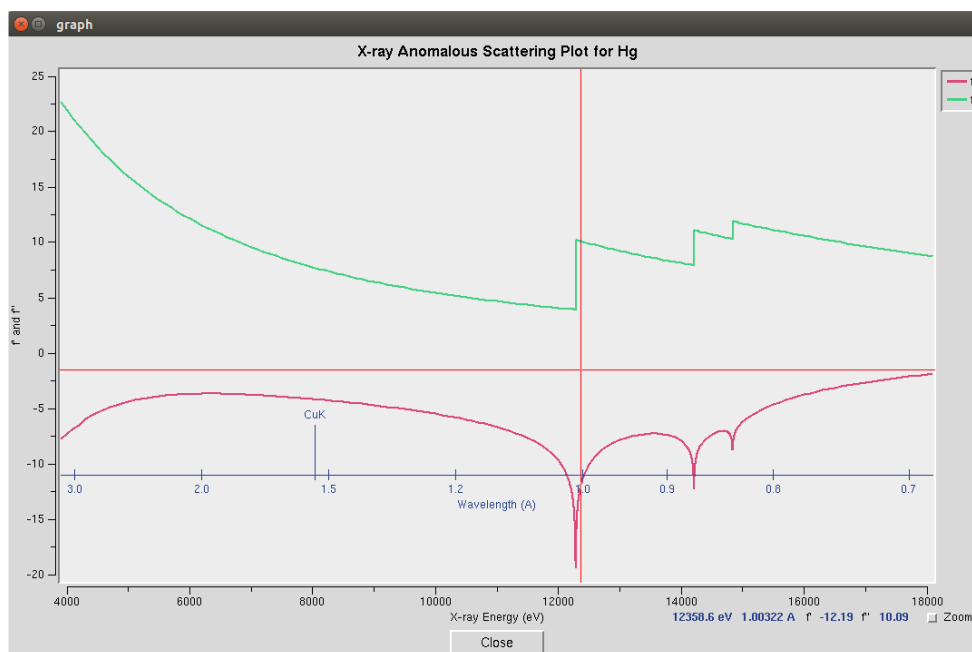
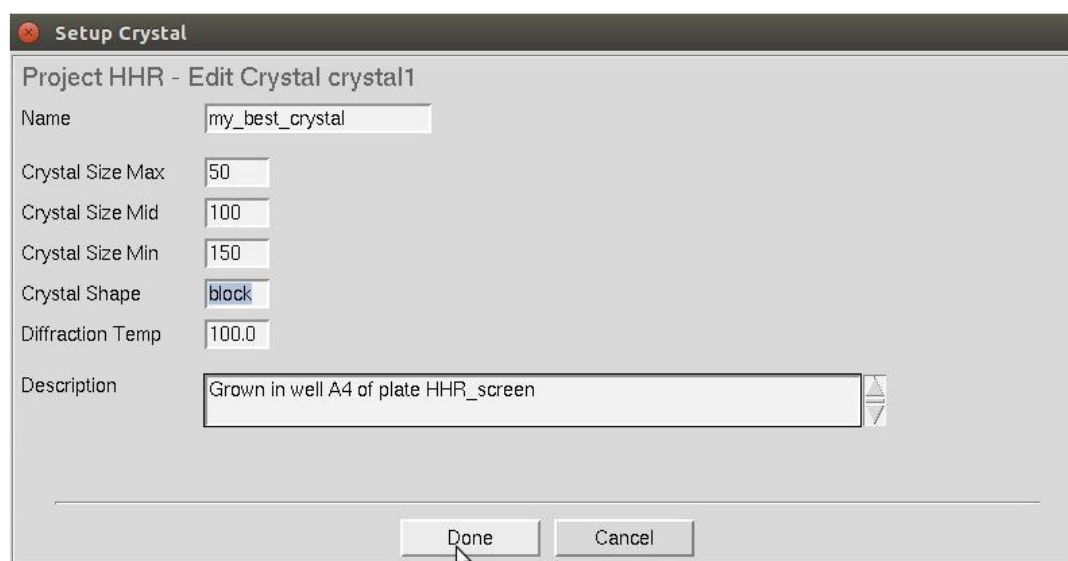


Figure 35. The anomalous scattering plot for mercury

## CREATE, SAVE, AND LOAD PROJECTS

When you have finished entering the project information, click **DONE**. If you entered this dialog using the create project button, you will be taken directly to the **Edit Crystal** dialog (Figure 36). If you entered this dialog using the edit project button, you will return to the Project tab, which will now display the information about your project (Figure 37).

The **edit crystal** dialog allows you to enter information about the crystal, including the crystal name, size, and description (Figure 36). If you are using a data collection capable version of HKL-2000, your collected images will be named based on the crystal name.



**Figure 36.** *The Create/Edit crystal dialog*

When you have finished editing the project, you will return to the Project tab. The sequence will be displayed, and you will see a tabulation of the number of each residue in your sequence. In database connected versions of HKL-3000, information from the database will appear in the Protein Data box.

## CREATE, SAVE, AND LOAD PROJECTS

The screenshot shows the 'Project' tab in the HKL-2000 software. The interface is divided into several sections:

- Project Information:**
  - Name: HHR
  - Crystal: my\_best\_crystal
  - Experimenter: dcoop
  - Date: Aug 25, 2018
  - Buttons: Load, Save, Edit Project, Edit Crystal, New Project, New Crystal, Evaluate Model
- Protein Data:**
  - Principal Component: Protein
  - Name:
  - Number of Residues: 72
  - Organism:
  - Description:
  - Molecular Weight: 8950.37
  - NCBI accession code:
  - Swiss-Prot accession code:
  - No. of Homologues:
  - Last check in PDB:
- Amino Acid Counts:**

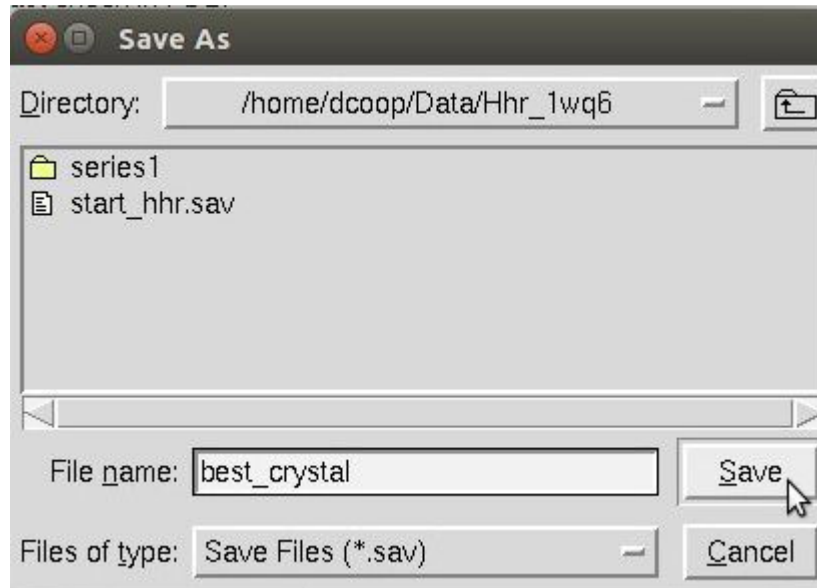
Ala (A): 5	Gly (G): 0	Met (M): 4	Ser (S): 2
Cys (C): 2	His (H): 3	Asn (N): 2	Thr (T): 4
Asp (D): 7	Ile (I): 3	Pro (P): 0	Val (V): 2
Glu (E): 10	Lys (K): 3	Gln (Q): 2	Trp (W): 3
Phe (F): 0	Leu (L): 8	Arg (R): 10	Tyr (Y): 2
- Phasing Method:** SAD/MAD
- Source of Anomalous Signal:** Se 4
- Anomalous Wizard:** Button
- Scaled Data:**
  - Buttons: Add File, Remove File, Statistics

**Figure 37.** The Project tab after creating a project.

## Saving and Loading Projects

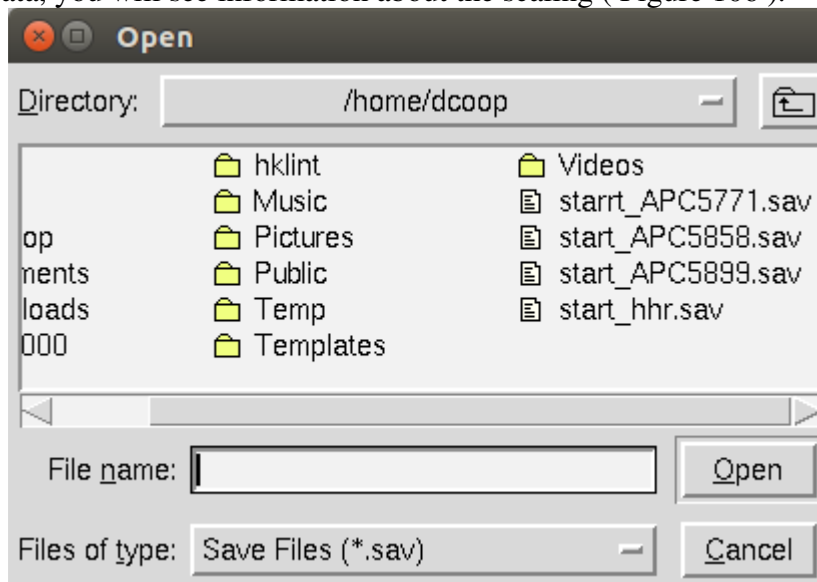
The primary advantage of creating projects in HKL-2000 is that you can save your project, which can simplify the process of re-starting or continuing where you left off. To save your project, select the **SAVE** button in the **Project** box. You will be presented with a dialog to specify where to save your project (Figure 38). The file name will automatically be given a “.sav” extension.

## CREATE, SAVE, AND LOAD PROJECTS



**Figure 38. The Save Project dialog**

Loading a project is also straight forward. Use the **LOAD BUTTON** to display the file selection dialog (Figure 39). After loading a project, not only will it load the information about the project, but will also populate the data tab with the data sets you have. If you have already scaled some data, you will see information about the scaling ( Figure 106 ).



**Figure 39. The Load Project dialog**

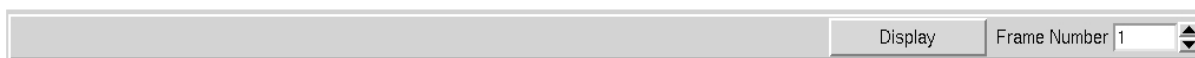
## Display and Peak Search the Data

The HKL suite now includes two XDisp programs. The default program is the newer version, XDispQt. You can select the older version, XDispX, using the XDisp selector located in the Options menu.



**Figure 40. The XDisp selector**

There are three ways to display your images. The most convenient method is the display button (Figure 41) in the Data tab. There is also a display button in the Index tab. Both of the display buttons have an adjacent frame selector box next that allows you to select a particular frame to view. The default frame is the first frame of the set.



**Figure 41. The Display image button in the data tab**

The layout of XDispQt opens with the image displayed in the center window (see Figure 42. The XDispQt display). You can zoom and scroll in this window to see more details of the image. Action buttons and display altering buttons are located on the left. A small view of the image, information about the cursor position, and an intensity histogram are located on the right. The XDispQt window can also be resized using standard window manipulation methods. Many modern Linux systems allow you to “Alt-middle-click and drag” to resize a window (some systems use “Ctrl-middle-click”). For this, your mouse just needs to be inside the window somewhere near a corner, which is usually much easier than either clicking and dragging the exact edge or right-clicking in the window's top panel and selecting “resize.”

## DISPLAY AND PEAK SEARCH THE DATA

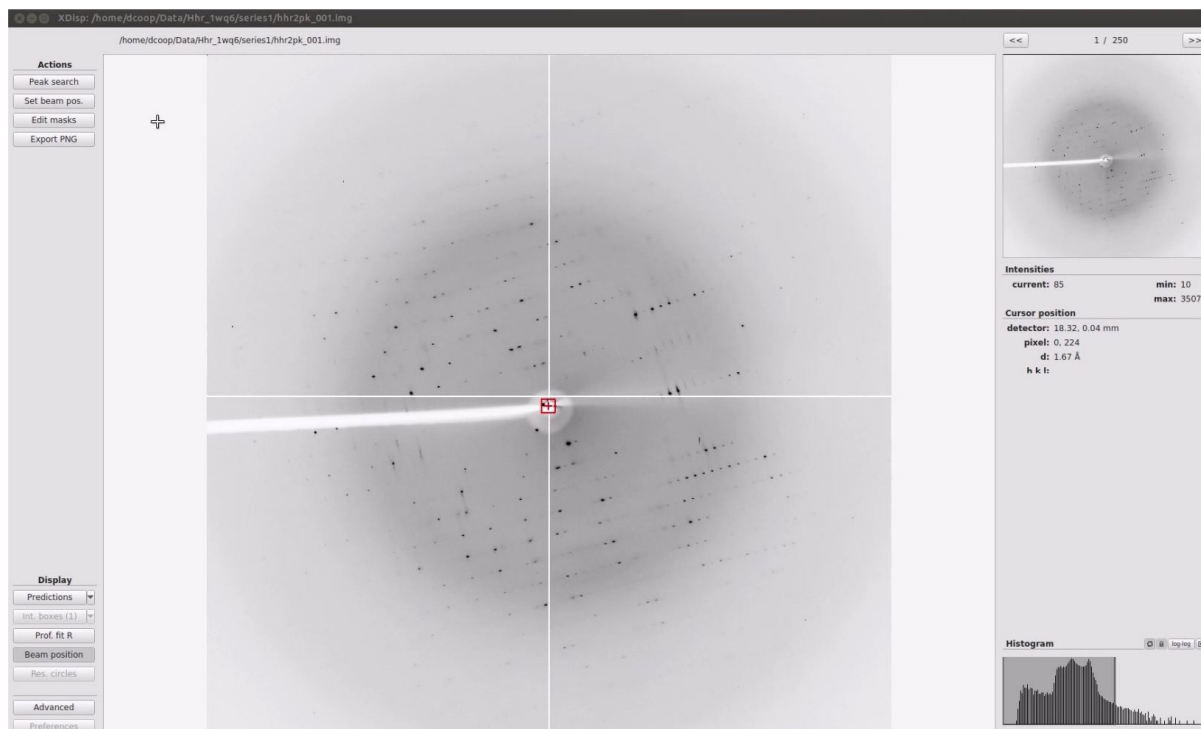


Figure 42. The XDispQt display

To view different frames, use the frame up “>>” and frame down “<<”.

Zooming and selecting areas to view are controlled by the mouse. Zoom in using the mouse wheel. Once zoomed in, you can click-and-drag within the main display to move around. You may also click-and-drag the black frame representing the current view in the whole-image display on the top right. Zoom in far enough and the numeric intensity value for each pixel will be displayed. Click and drag with the right mouse button to show a cross-hair and intensity distribution for the vertical and horizontal lines displayed on their respective axis.

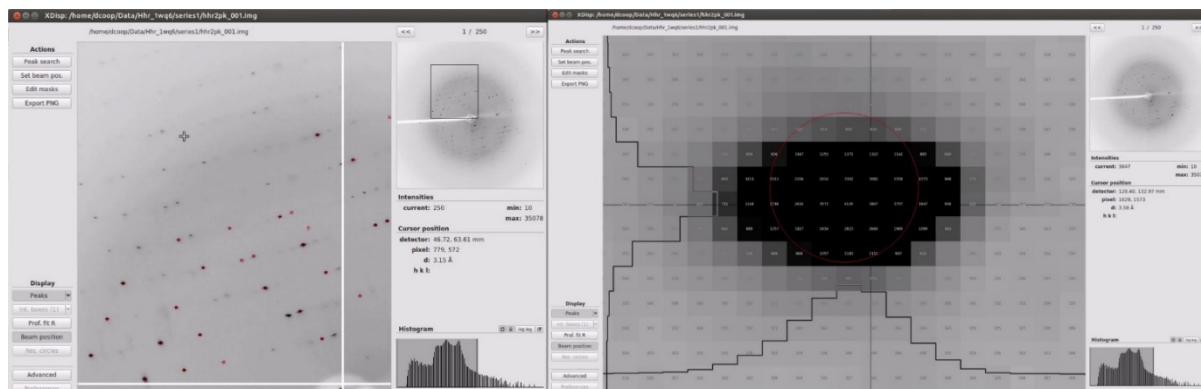


Figure 43. Two zoom levels are shown

## DISPLAY AND PEAK SEARCH THE DATA

The frame up and frame down arrows are on the top right of XDispQt (Figure 44). Change the intensity maximum and minimum using the *intensity histogram* on the bottom right (Figure 45). Click and drag the shaded box to change the range of intensities or drag an edge when it turns yellow to change just the maximum or minimum intensity. If your screen update is too slow, turn off the *automatic refresh* button (the cycle button next to the word “Histogram”).

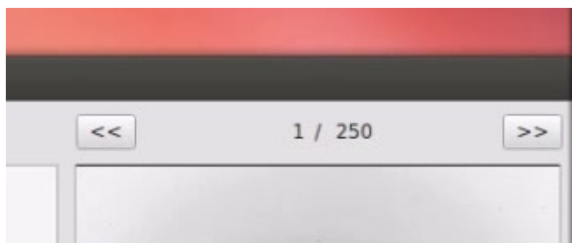


Figure 44. Frame up and down buttons

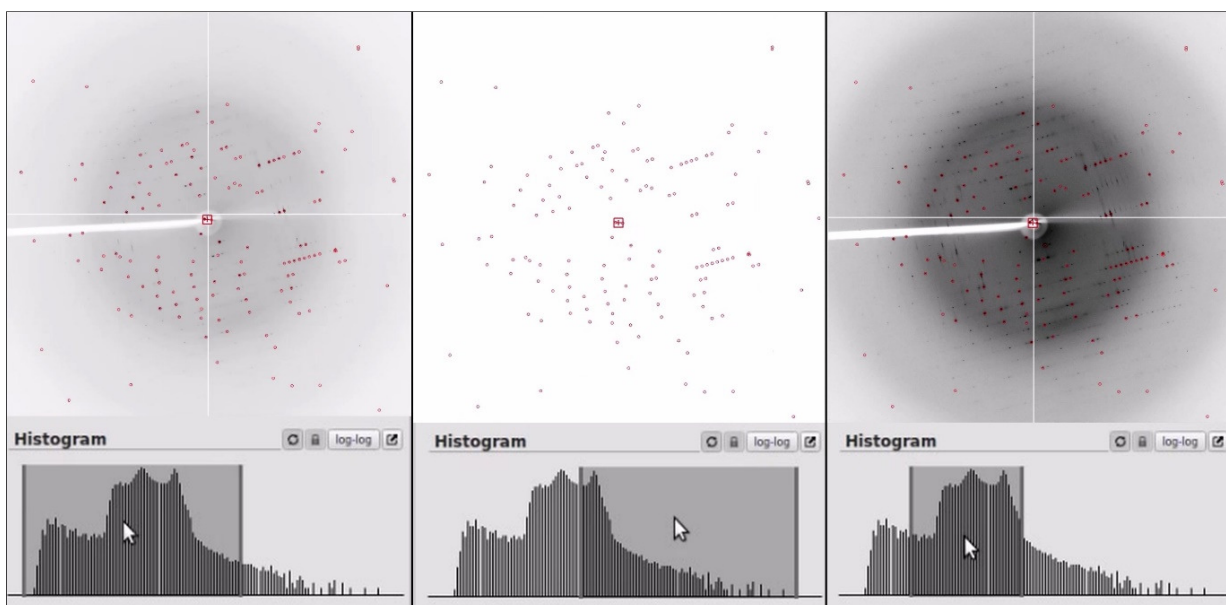


Figure 45. The Histogram controls the display contrast.



### Working with frames in XDispX

To see subsequent images within this dataset, middle-click the frame button. By default, the **FRAME** button will increase the frame number, as noted by the up arrow. Clicking the **FRAME** button with the left and right mouse buttons will change this to a down arrow and back to an up arrow, respectively. The middle mouse button will increase or decrease the frame number based on the direction of the arrow. The **FRAME** button will only increment one frame at a time. If you want to quickly go to a frame that is far away from the current one, enter the frame number in the box by **Frame Number** in the **Main** page (Figure 46).



Figure 46. The Frame button from the Display window

### XDispX - Adjusting the display contrast

The displayed image can be made darker or lighter by clicking on the **BRIGHT** or **DIM** buttons in the **Image Display** window. Small increments in either are made by clicking with the left mouse button. Larger jumps are made by clicking with the middle or right mouse buttons. To change the displayed dynamic range, use the **FLOOR UP** or **FLOOR DOWN** buttons. **FLOOR UP** narrows the dynamic range, while **FLOOR DOWN** increases it. If you have fairly weak diffraction you can enhance the contrast (i.e. make it look “better” than it actually is) by alternately clicking the **BRIGHT** and **FLOOR UP** buttons (Figure 47).

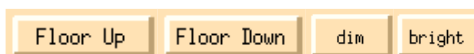


Figure 47. The buttons for adjusting the image

## Automatic Peak Search

The default peak selection parameters of HKL are very often sufficient for indexing diffraction images, so if your image looks reasonable, you should give automatic peak search a try. To use this feature, switch to the **Index** tab, and click **PEAK SEARCH**. If you haven’t displayed the image yet, this feature will open the display for you and peaks will automatically be selected. The selected peaks will be indicated by red circles. If at least 100 peaks are selected using the automatic peak search, you can skip to the indexing step.

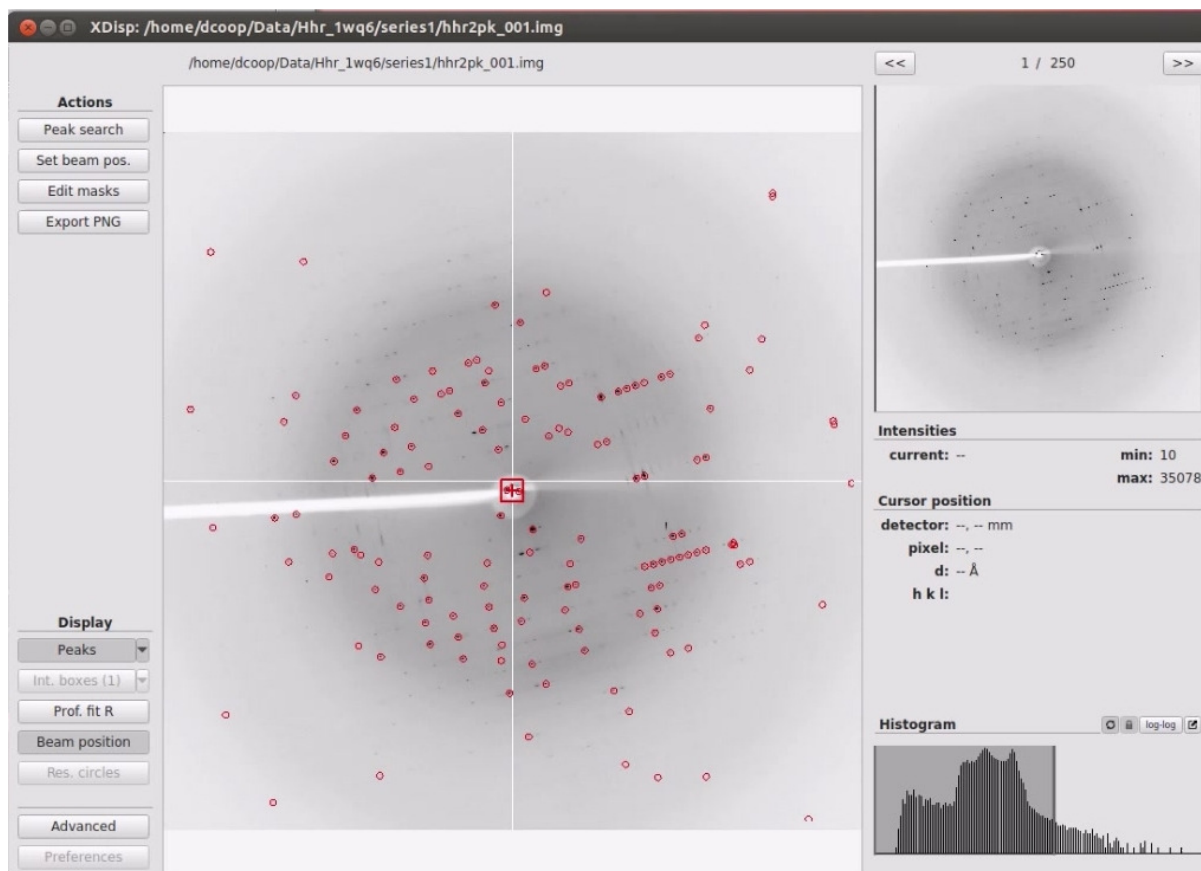
## Manual Peak Searches

In some cases (weak data, noisy data, multiple crystals, etc.), the automatic peak search may not work as well as a manual search. In these cases, display an image and click **PEAK SEARCH**. Peaks will be selected based on their intensity, and a dialog box will appear on the right side of the image display. The goal of the peak search is to pick peaks for auto-indexing, and around 100 peaks should be sufficient if they clearly belong to the major lattice (i.e., not from a second crystal or ice rings). Sometimes the number of peaks will be low. In most cases, you can increase the number of peaks using the **MORE PEAKS** button. If you click **MORE PEAKS** too many times (especially with weak data), noise peaks will be selected, and you should decrease the number of peaks using **FEWER PEAKS**. Sometimes the peaks will not be selected if they are too large (or small) for the default peak search size. In these cases, the **BIGGER PEAKS** or **SMALLER PEAKS** (**SIZE PEAK UP** or **SIZE PEAK DOWN** in XDispX) buttons will allow more peaks to be selected.

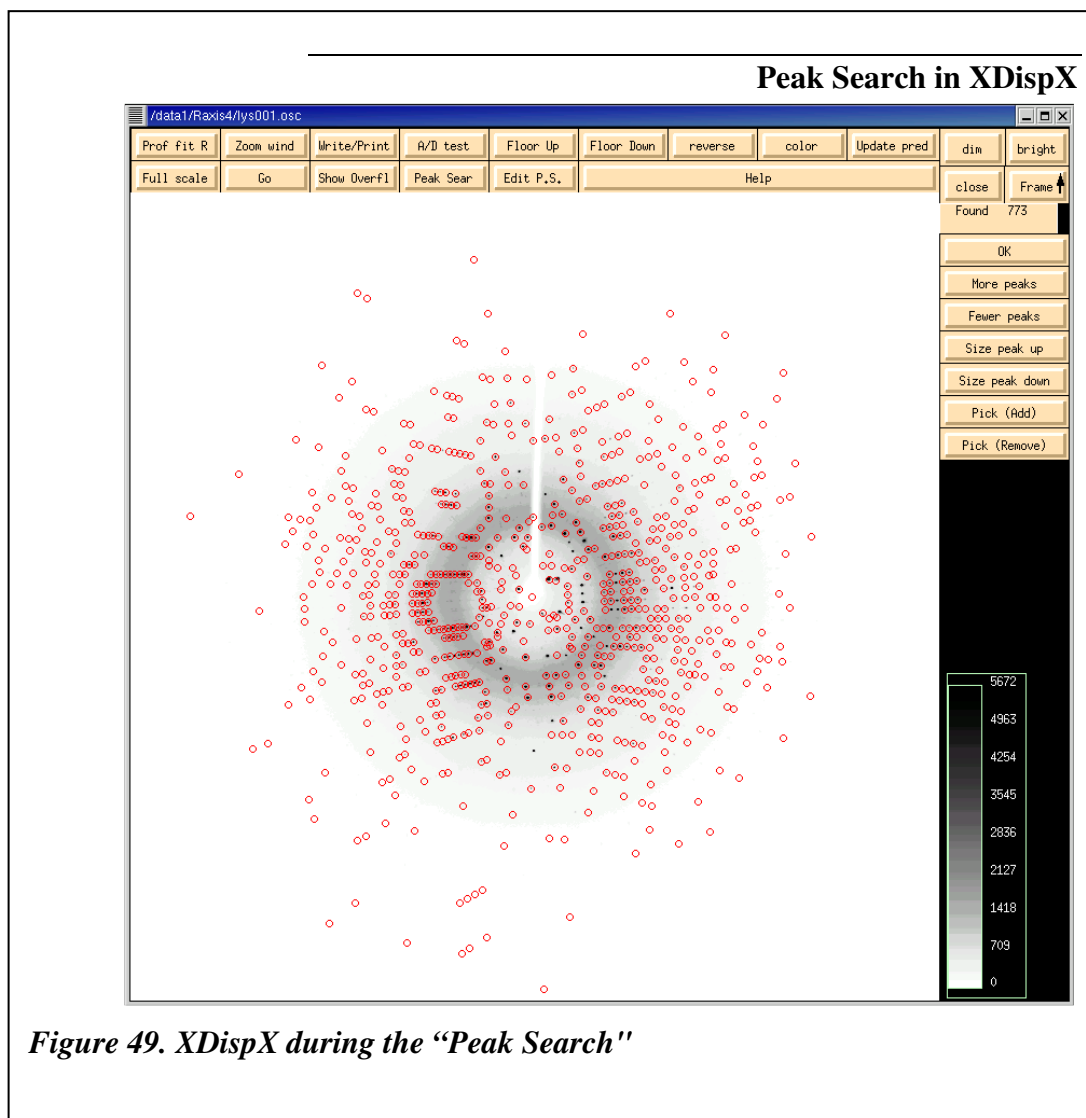
Another way of adding more peaks to the peak list is to include peaks from multiple frames. To do this, middle-click on the “>>” button (**FRAME UP** in XDisp QT) button to display the next frame, and the peak search will be repeated on the next frame using the same criteria (e.g., **MORE PEAKS**, **FEWER PEAKS**, etc.). These peaks will be added to the peak search from the first frame. This is a way to index poorly diffracting crystals and may help with certain crystal orientations.

When you are satisfied with the peak search results, click **FINISH** on the bottom left (**OK** in XDispQt) (Figure 48).

## DISPLAY AND PEAK SEARCH THE DATA



*Figure 48. After peak search*



## Secondary Lattices

*Denzo*, the HKL suite program that indexes the data, assesses the quality of each peak before using it in auto-indexing, so most spurious peaks, such as those coming from ice rings, can be ignored by the user. However, if you have a second distinct lattice contaminating the diffraction pattern, you will have to make some effort to assure that the selected peaks come mainly from the primary lattice. The first way is to see if *Denzo* can automatically distinguish between the primary lattice and any secondary lattices. If the secondary lattice is weak or at a significantly different orientation from the primary lattice, this is not unreasonable. Alternatively, you can try indexing on a different frame, where the secondary lattice may not

## DISPLAY AND PEAK SEARCH THE DATA

be so pronounced. However, if this does not work, you can assist the program by editing the peak search list in the *Image Display* window, as described below.

### Ice Rings

Ice rings usually appear as complete rings at the resolution listed below. However, if individual crystals of frost have formed on the outside of the crystal and surrounding vitrified cryo-solution, you may get individual reflections that result from these ice crystals. These reflections will complicate the indexing process more than complete rings will. This type of frost can often be rinsed off of the crystal by drizzling a stream of liquid nitrogen on the crystal or by gently wiping it off with a paper wick soaked in liquid nitrogen. Frost on the outside of the crystal can occur during shipping if lid of the shipping dewar has a gap between the neck of the dewar and the dewar's cap. Securing the cap well can help reduce the chance of this type of frost.

**Ice Ring Resolutions:** Regardless of whether ice rings occur from an insufficient concentration of cryo-protectant or frost that forms on the outside of a properly "frozen" crystal, the diffraction resulting from the ice will appear at the following resolutions:

3.90 Å  
3.67 Å  
3.44 Å  
2.67 Å  
2.25 Å  
2.07 Å  
1.92 Å

You can display circles corresponding to these resolutions using the Resolution Circles option shown in Figure 53. Since ice rings are centered at the direct beam position, they can be used to verify the beam center. Set the resolution circles to ice ring distances to see they match up.

If your ice rings are picked up by automatic peak picking, you may want to set the Max Resolution to 4 during the indexing step. This resolution is usually sufficient to index the data, but will avoid ice ring complications. A convenient ***Limit Resolution in Pre-refinement*** option in the Resolution control window can do this for you, yet let you process your data to a better resolution.

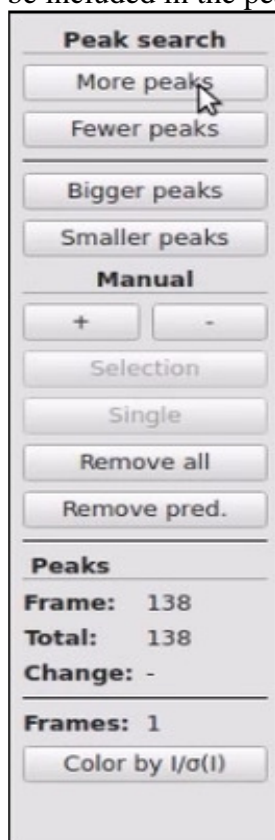
## Adding and Removing Peaks

The presence of multiple crystals, cracked crystals, bent crystals, and certain types of twinned data can give rise to multiple diffraction patterns. These problems may only exist in sections of your data, or it may be persistent throughout the data. If it is only part of the data, see the *Peak Search on Different Frames* section. If the peak search process identifies reflections from more than one crystal, you may find it difficult or impossible to index and process the data. If you clearly have one diffraction pattern that diffracts stronger than the other, you may isolate the primary lattice by adjusting the sigma level of the reflections used

## DISPLAY AND PEAK SEARCH THE DATA

in indexing. If this doesn't work, HKL provides you with the ability to change the results of a peak search manually. Add peaks that seem to be part of the primary lattice and remove peaks that don't. (If the crystal survives, it may be interesting to go back and use manual peak picking to characterize and possibly collect the "sister crystal.")

To perform a manual peak search, use the peak search button in the top left. The buttons **MORE PEAKS**, **FEWER PEAKS**, **BIGGER PEAKS**, and **SMALLER PEAKS** work as described above, and have a pop-up description if you hover over the buttons. Use the frame up button to increase the number of frames that will be included in the peak search.



**Figure 50. Peak Search options.**

If you want to change the peaks that are picked manually, use the **+** and **-** buttons to add or remove peaks. Once the **+** button has been selected, click on reflections to add them to the peak list. With the **-** button, you can either remove reflections one at a time (Single will be selected as default) or use the **SELECTION** button to define a rectangle of reflections to be removed from the peak list. To use this feature, click selection then and click twice (not click-and-drag) to define a region you wish to omit.

You also have the options of removing all the peaks so you can start over or coloring the peaks by their **I/σ** levels. **RESET**, **CANCEL**, and **FINISH** buttons are located at the bottom left. When performing a manual peak search, it is important to click the **FINISH** button to save the peaks to a file so they can be used for indexing.

**Manually picking peaks in XDispX**

Peaks can be added or removed with **PICK (ADD)** or **PICK (REMOVE)**. **PICK (ADD)** can be very helpful if you have weak data and it is obvious that peaks from the main lattice were not selected, and **PICK (REMOVE)** is helpful if many peaks from a secondary lattice were selected. These buttons will change the cursor to a cross missing its center. In **PICK (ADD)** mode, clicking the mouse will add a green circle at the crosshair. You can add more than one peak at a time, but you need to click on **PICK (ADD)** or **PICK (REMOVE)** to complete the selection process. In **PICK (REMOVE)** mode, peaks will be marked with a green 'x'. As with **PICK (ADD)**, your peak selection will remain active until you click **PICK (ADD)** or **PICK (REMOVE)**. The **OK** button will also accept your additions or deletions, but it will also end the peak search process

**Peak search on different frames.**

Sometimes it is advantageous to index on a frame other than the first frame. This can be helpful if a crystal axis is closely aligned with the X-ray beam. Choosing a frame with easily recognizable diffraction loons may help index your data. There are two ways to index your data on a different frame. The first way is to change the frame specified next to the **PEAK SEARCH** button in the **Index** tab. After setting the frame number, use the peak search button to conduct an automatic peak search. Changing the peak search frame selection will automatically change the frame used for indexing.

The second method of peak searching on a different frame is to divide the data into multiple sets using the **SPLIT SET** button, as described above. This method will allow you to perform a manual peak search if you would like. Select the set that starts with the desired frame and conduct a peak search. If you do split the sets, be sure to select the proper set in the **Pending Sets** box of the **Index** tab before indexing, or you will get an error message that says, "Error – Peaks none."

## Auto-Indexing

Before you can index diffraction data properly, you first have to characterize the underlying symmetry of the crystal. HKL will attempt to index the data within each Bravais Lattice type. For historical reasons, the process indexing the data in all crystal classes is known as auto-indexing. To auto-index your data, select the Index tab of the main HKL-2000 window. If you have more than one set of data, they will appear in the *Pending Sets* window. Select the one you used for indexing. You may also want to change the **Resolution Limits** (the default is to the edge; changing it later is also OK), then index the data by clicking the **INDEX** button. The program will try to fit the diffraction pattern in each of the 14 Bravais lattices and generate an index describing the fit of each lattice to the observed peaks. The Bravais Lattice dialog will automatically open to allow you to select a different lattice. If you are working with unfamiliar crystals, it is a good idea to refine some parameters in the primitive triclinic lattice (space group P1) and revisit this dialog later. Close the *Bravais Lattice* window using the **APPLY AND CLOSE** button. Check to make sure the green circles identified by indexing line up with real spots by examining the *Image Display* window (click the **DISPLAY** button if it's not already open). You can toggle the predicted spots on and off using the **UPDATE PRED** button.

## The Pending Sets window

All of the data sets that you are integrating will appear in the *Pending Sets* box in the upper left corner of the **Index** page. Which sets are present is controlled by selecting the **Integrate** checkbox for each data set on the **Data** tab. You can choose a particular set to index and/or refine by selecting it, which will highlight it with a blue background (Figure 51). Once you have selected a set, you can change the parameters of the integration for that set, while leaving the parameters for the other set(s) unchanged. For example, set 1 may be set for processing to 3 Å resolution, while you may want to process set 2 to 1.5 Å. Similarly, perhaps the **Spot Size** is 0.3 for the first set, but 0.25 for the second set. Perhaps the **3D Window** is five frames for the first set but only 3 for the second. The *Pending Sets* box allows you to toggle between the pending data sets in preparation for integrating them in the same run. You can also accomplish the same thing by going to the **Summary** tab page and editing the parameters for each set (Figure 52). It is worth noting that if your datasets are from the same crystal, with approximately the same orientation, you can probably integrate all of the data at the same time. You only need to index one of the pending sets before integrating them all. It is wise to keep an eye on reflection markings during integrations. A sudden increase of rejected and/or unpredicted reflections after the transition to another set, then you may have to index, refine, and integrate the sets individually.



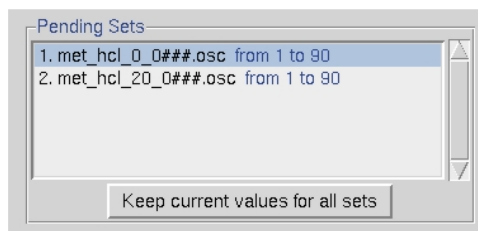


Figure 51. The Pending Sets box

Set	Raw Data Template	Start	End	Osc Start	Osc Width	Osc Range	Phi	Distance	Res Min	Res Max	3D Window	Profile Fitting Radius	Wavelength
	Default parameters			0.00	1.00		0.00	100.00	50.00	edge	1	30.0	
1	lys###.osc	1	20	198.00	1.00	20.0	198.00	75.14	50.00	edge	1	30.0	1.5418
2	lys###.osc	25	60	222.0	1.00	36.0	198.00	75.14	50.00	edge	1	30.0	1.5418

Figure 52. The summary information about processed sets

## Initial Parameters for Indexing

Most of the default setting are sufficient for indexing, and there are only a few that you may want to consider before indexing your data. If your data looks good close to the edge of the plate, then leave the default settings as they are (Figure 53). However, there are certain situations where it may be appropriate to change the **Resolution**. For example, if your data does not extend to the edge of the detector, or if your data has a combination of weak reflections and ice rings you may want to limit (by raising) the resolution. For example, if you have an ice ring, try setting the **Max** value to 4 Å. If you have a secondary lattice which is more pronounced at higher resolutions, try indexing at low resolution. For your convenience, a **Limit Resolution in Pre-refinement** option will limit the resolution to 4 Å for the initial pass at indexing before extending the resolution to the resolution limit you have set. This can help with ice rings or weak resolution data. Do not truncate your resolution pre-maturely, however. HKL can detect reflections that your eyes may not. Additionally, you can truncate your data at the scaling step.

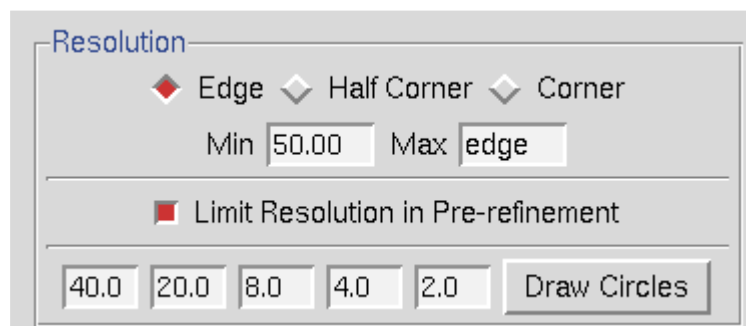


Figure 53. The Resolution limits box

Other parameters you may want to change before indexing are at the top of the **Controls** box (Figure 54). If you have a well-diffracting crystal with a large unit cell, you may

## AUTO-INDEXING

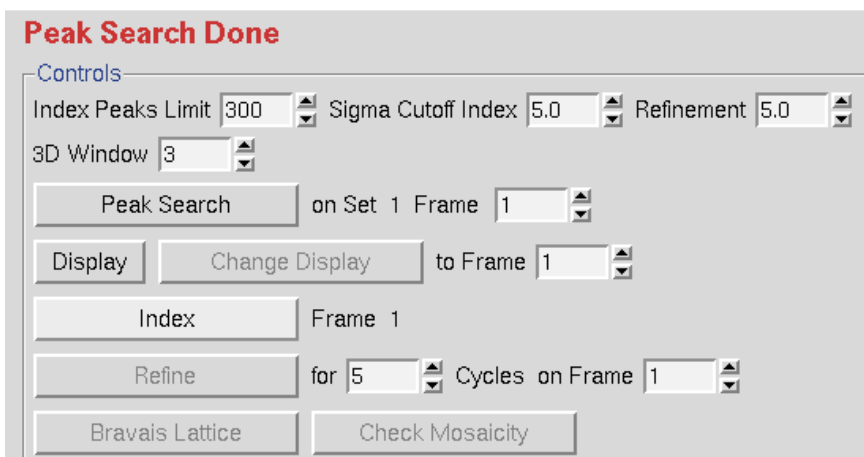
want to increase the number of reflections that are used for indexing. The *Sigma Cutoff Index* limits the signal to noise ratio that is used for indexing. If a secondary lattice is present, you may want to increase this. On the other hand, if your data is very weak, you may need to decrease the limit.

The **INDEX** button starts the auto-indexing process and will attempt to index the data in every crystal system.

---

*The **INDEX** button starts the auto-indexing process and will attempt to index the data in every crystal system.*

---



**Figure 54.** Other pre-indexing options.

## Peaks used for autoindexing

The green circles that appear during indexing denote the subset of the peak search peaks that were used in the auto-indexing. The criteria for usability include falling within the resolution limits specified, high enough signal to noise, sufficiently low background (which helps to eliminate spots in ice rings and overlaps), and identification of a lune. If there aren't any green circles after auto-indexing, one or more of the following problems may be the cause:

1. The **Peak Search** was done on a different frame from that being used to auto-index. In this case, you should have already gotten the error message "Error – Peaks none." The remedy is to make sure these are the same. The rightmost column of the file `peaks.file`, generated during peak searching, tells which frame the peak search peaks were derived from.

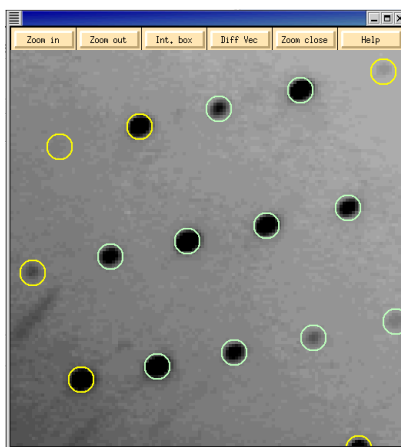
## AUTO-INDEXING

2. None of the spots are strong enough to qualify for use in the auto-indexing routine. The remedy is to take a longer exposure, pick more peaks on the frame or on several frames, or lower the **Refinement Sigma Cutoff**, which is normally set at five  $\sigma$ . Lowering the sigma cutoff is a dangerous option, however, and is not recommended because weak reflections are not high-quality reflections, and therefore may be noise.

*Denzo* assesses the quality of the background of each spot against criteria for slope and uniformity, and if a certain fraction of the pixels in the background fails the test, then the whole spot is rejected. This is controlled by the keyword **reject fraction**, which can be entered in the **Macros Tab**. This default value for this rejection is 0.75, i.e., the 75% of the background pixels must be acceptable. If you have an ice ring, you can increase **reject fraction** until most of the spots in the ice ring are no longer picked up. Useful increments are 1-2% at a time. Do this while you are looking at the display, just as you would do for spot shape and mosaicity.

3. No lunes could be detected. This can result from either too many or too few spots. At certain crystal orientations, especially with high mosaicity crystals, the entire image can be filled with reflections, which can make it impossible to discern where one lune ends and the next starts. In these cases, it may be necessary to index the data on a different frame where the lunes can be more easily identified. On the other hand, if there weren't enough spots to describe a lune (sparse lattices, low mosaicity, and low resolution), the remedy is to pick more peaks or include more frames in the peak search. Sometimes manually selecting reflections that are not included in the automatic peak search process may be necessary.
4. The spot size is too small. If you have large reflections, the reflections may extend beyond the boundary defined for the reflection into the background for the reflection, causing the reflection to be rejected. The Integration Box dialog in the bottom left can be used to ensure that the reflection boundaries are appropriate.

[See the Trouble Shooting appendix has an in-depth look at the causes of indexing failures.](#)



*Figure 55. The Zoom window with image after auto-indexing*

# Lattice Parameter Refinement

Click the refine button to refine the lattice parameters. The initial parameters are those of **FIT BASIC**, and these are appropriate for auto-indexing and the first refinement step. When the Refine button is clicked, the parameters selected in the Refinement Options box (Figure 56) will be refined the number of times selected in the selector box next to the **REFINE** button. The overall measurements of how well the selected crystal class fits the data are the *Positional and Partiality Chi-Square* values (Figure 57). These are also color-coded from dark green to light green to red from good to bad. The other refinement parameters that are refined are presented with their current value, how much they have changed during the refinement cycle, and the error associated with the value (from left to right). Click Refine several times, and the values should become stable.

The 'Refinement Options' dialog box contains the following sections:

- Crystal:** Rot X, Rot Y, Rot Z (all checked with red squares).
- Detector:** Rot X, Rot Y, Rot Z (all unchecked).
- Crossfire:** X, Y, XY (all unchecked).
- Other parameters:**
  - X Beam (checked with red square), Y Beam (checked with red square)
  - Yscale (unchecked), Skew (unchecked)
  - Cell (checked with red square), Distance (unchecked)
  - Mosaicity (unchecked)
- Buttons:** Fit All (green), Fix All (red), Fit Basic (grey).

Figure 56. The Refinement Options box with the Fit Basic parameters selected

Refinement Information					
Space Group: P1					
Resolution: 50.00 - 2.01					
Positional:	1031	X- $\chi^2$ :	1.35	Y- $\chi^2$ :	1.13 (0)
Partiality:	662	$\chi^2$ :	2.05		(0)
		X Beam:	96.498	Y Beam:	93.663
a:	31.12	b:	62.01	c:	76.31
$\alpha$ :	89.66	$\beta$ :	89.90	$\gamma$ :	89.79
Crystal Rotation X:			-9.152	-0.001	0.003
Crystal Rotation Y:			-163.924	0.000	0.003
Crystal Rotation Z:			-14.690	0.000	0.008

Figure 57. Refinement statistics generated using Fit Basic parameters

At this point, the predicted reflections displayed on the image are color-coded. Reflections that are recorded completely on this image will be green, whereas reflections that are split between more than one image will be yellow. These “partial” reflections will be combined during the scaling process. After a few rounds of refinement using the default

## LATTICE PARAMETER REFINEMENT

parameters, check the image to make sure the predicted reflections sit on real spots. Make sure the direct beam position is correct and looks reasonable.

Click the **FIT ALL** button and **REFINE** again. Click the **BRAVAIS LATTICE** button, select the appropriate lattice for your crystal, and then click the **APPLY & CLOSE** bar. HKL uses eigenvalue filtering to ensure that any parameters that are too highly correlated will be restrained by the program. The days of selective manual fitting and fixing are over. The program almost always does a better job.

## When refinement goes wrong

Occasionally the auto-indexing routine does not give clear-cut results. Symptoms range from utter failure to difficulty scaling to something that is just odd. Please refer to the Trouble Shooting appendix for detailed exploration of common causes of indexing problems.

The Trouble Shooting section contains information about :

- When default settings aren't good enough
- Macros that can save the day
- Most common sources of problems
- Finding the beam position

## Initial 3D window setting

A good rule of thumb is to set the number of oscillation degrees covered by the **3D Window** (after the resolution limits and lattice have been determined) to be at least twice the refined mosaicity. For example, if the mosaicity is about  $1.0^\circ$ , and the frame width is  $0.5^\circ$ , a value of 5 is a good choice. The default value of 5 is reasonable for macromolecular data, with a frame width of  $1.0^\circ$ . For small-molecule structure determinations, where the reciprocal lattice is sparsely populated, a **3D Window** of 3-5 frames is a good choice (Figure 58).

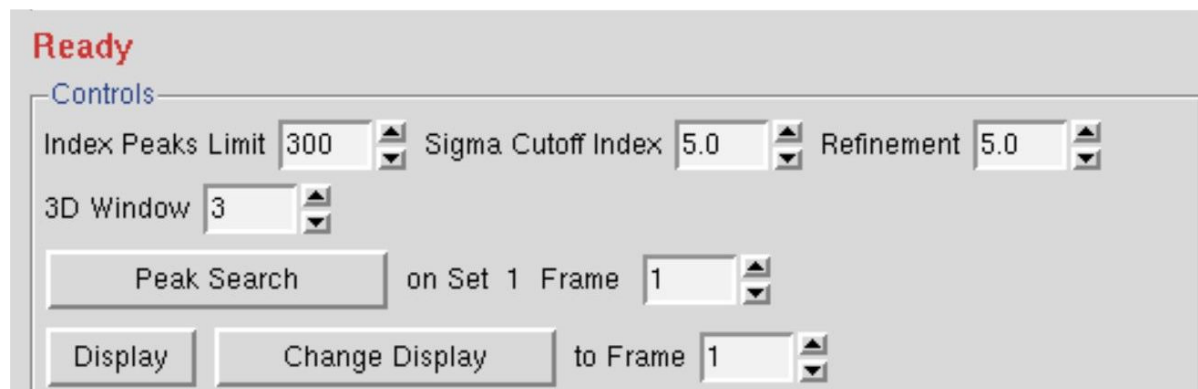


Figure 58. The 3D window in Controls box

## Choosing the correct lattice

The lattice is best chosen after the first frames of the **3D Window** have been refined in space group *P1*, but before the whole set of frames is integrated. Since primitive triclinic is the lowest possible symmetry for your crystal, if you can't succeed in refining in *P1*, choosing a higher symmetry lattice with less refinable parameters will be even worse.

Clicking on the **REFINE** button starts the refinement of the auto-index lattice. After auto-indexing, a small set of crystal parameters will be selected for refinement. Refine the indexing a few times using **FIT BASIC**, then select the **FIT ALL** regime – you do not want to be fixing refinement parameters. During refinement, you should notice the number of spots used in the refinement go way up, and the  $\chi^2$  errors in X and Y positions approach 1.0. Before proceeding further, go back to your diffraction **Image Display** window and examine the positions of the yellow or green circles with respect to the diffraction spots. If indexing and refinement worked, the predicted reflections (circles) should still match up with the diffraction pattern. If the predictions do not line up with the diffraction spots, check out the Trouble Shooting Appendix on page 137. Also consider the possibility that you are looking at reflections from more than one crystal. HKL-2000 can only predict reflections from one lattice.

Assuming that the predicted reflections still match up with the diffraction pattern and the  $\chi^2$  values are not red colored (i.e., above 5), now is the time to select the proper Bravais lattice. Click on the **BRAVAIS LATTICE** button and examine the table ([Error! Reference source not found.](#)). The Bravais Lattice table shows two unit cell values for each lattice type (or three for primitive rhombohedral lattice). The top cell is a primitive cell that is close to one suitable for that lattice, and the lower cell shows the changes that need to be made to the primitive unit cell parameters (a, b, c,  $\alpha$ ,  $\beta$  and  $\gamma$ ) in order to force that unit cell into each. For example, tetragonal unit cells will have all angles set to 90°. A color-coded distortion index is shown to the left of the unit cell parameter and reflects how distorted the resulting geometry is with respect to the primitive triclinic geometry. The lower the distortion index, the more likely the given lattice is correct; by definition, the distortion index for the primitive triclinic lattice is 0%. If you know (really know, not “kind of” know) the space group ahead of time, you can select the correct lattice immediately. If not, select the lattice that has the highest symmetry with a low distortion index. Once you have selected the lattice, by clicking on the little radio-button by the one you chose, click the **APPLY** button to input the new lattice. Refine the selected cell using the **FIT ALL** button. If the **Positional and Partiality Chi-Square** values do not dramatically increase, then you can assume that this is the right lattice.

## LATTICE PARAMETER REFINEMENT

Bravais Lattice Table

Autoindexing performed for unit cell between 1.6 to 43 Angstroms

◇ primitive cubic	44.26%	24.26	5.29	7.04	89.90	90.02	90.01
		12.19	12.19	12.19	90.00	90.00	90.00
◇ I centred cubic	62.60%	25.26	8.80	24.83	97.33	20.24	102.86
		19.63	19.63	19.63	90.00	90.00	90.00
◇ F centred cubic	61.33%	8.80	49.31	8.81	100.27	106.17	87.16
		22.30	22.30	22.30	90.00	90.00	90.00
◇ primitive rhombohedral	15.95%	25.26	24.83	24.26	12.30	16.18	20.20
		24.78	24.78	24.78	16.23	16.23	16.23
		7.92	7.92	73.31	90.00	90.00	120.00
◇ primitive hexagonal	15.29%	7.04	5.29	24.26	90.01	89.98	90.10
		6.16	6.16	24.26	90.00	90.00	120.00
◇ primitive tetragonal	6.78%	7.04	5.29	24.26	89.99	89.98	89.90
		6.16	6.16	24.26	90.00	90.00	90.00
◇ I centred tetragonal	8.00%	7.04	5.29	49.31	96.16	98.20	89.90
		6.16	6.16	49.31	90.00	90.00	90.00
◇ primitive orthorhombic	0.04%	5.29	7.04	24.26	89.98	90.01	90.10
		5.29	7.04	24.26	90.00	90.00	90.00
◇ C centred orthorhombic	2.57%	5.29	48.80	7.04	89.99	89.90	96.21
		5.29	48.80	7.04	90.00	90.00	90.00
◇ I centred orthorhombic	4.27%	5.29	7.04	49.31	98.20	83.84	90.10
		5.29	7.04	49.31	90.00	90.00	90.00
◇ F centred orthorhombic	8.06%	8.80	8.81	49.31	92.83	100.27	73.83
		8.80	8.81	49.31	90.00	90.00	90.00
◇ primitive monoclinic	0.01%	5.29	24.26	7.04	90.02	90.10	89.99
		5.29	24.26	7.04	90.00	90.10	90.00
◇ C centred monoclinic	2.56%	5.29	48.80	7.04	90.01	90.10	96.21
		5.29	48.80	7.04	90.00	90.10	90.00
◇ primitive triclinic	0.00%	5.29	7.04	24.26	89.98	89.99	89.90

If you would like to change the crystal lattice: select desired bravais lattice, press Apply button and close window, otherwise just close window.

Apply Apply & Close

Figure 59. The Bravais Lattice table

If a lower symmetry lattice has a lower distortion index (e.g., primitive monoclinic at 0.13%) than a higher symmetry lattice (e.g., primitive tetragonal at 0.57%) then sometimes this may be the better choice, even if the higher symmetry lattice (e.g., tetragonal) is colored green. Why is this? Thanks to the use of the **3D Window** in the refinement process, the indexing now uses many more reflections and encompasses a thicker slice of diffraction space than the previous incarnation of auto-indexing and refinement in *Denzo*. As a result, the distortion indices are much more accurate and may reflect real deviations from crystallographic symmetry, rather than just an incomplete refinement of crystal parameters.

If the highest symmetry lattice with a low distortion index does not refine well, a systematic way to choose the correct lattice is to march your way up the list of candidate lattices (candidate lattices being the green ones) and look for a significant change in the distortion index. For example, let's say you think you have a tetragonal crystal, and the candidate lattices are primitive triclinic, C-centered monoclinic, primitive monoclinic, C-centered orthorhombic, primitive orthorhombic, and primitive tetragonal. After primitive triclinic, you would select primitive monoclinic, refine, and examine the **Bravais Lattice** table again. The distortion index for monoclinic (and all lattices of lower symmetry – take note!) should be zero. Are the distortion indices for the higher symmetry lattices still close? If there is a significant jump (say

> 0.2%) to go to a higher symmetry lattice, then perhaps those lattices are not correct, and you should stick with the lower symmetry lattice.

Keep in mind that a crystal misindexed to a lower symmetry lattice will be characterized by a high degree of “non-crystallographic” symmetry and can reliably be solved and refined as long as complete data was collected, whereas a crystal misindexed to a higher symmetry lattice will cause you no end of grief as you try to force idealized symmetry onto a system that lacks it. That’s why it is important to accurately determine the lattice of your crystal early on in a data collection session, so that you will not be surprised and shocked when your best so-called “tetragonal” crystal turns out to be monoclinic, and you’ve only collected half a data set, and the crystal is lying in a puddle on the floor of the synchrotron hutch. Considering how fast modern synchrotrons are, you should always collect more than 180 degrees of any uncharacterized crystal, even if the indications from the first few frames make you think you have a higher symmetry crystal. However, if you suspect your symmetry is something higher than monoclinic, it is worth using the strategy prediction functionality of HKL-2000 (see Strategy and Simulation starting on page 78).

### Primitive triclinic crystals

If no crystal class besides *P1* is colored green, you may indeed have a triclinic lattice. For these crystals, you will have to collect 360° (go “around the world”) to get a complete anomalous data set. Alternatively, your data may be misindexed (see the discussion about misindexing above and the Troubleshooting Appendix). If you are not able to see a particular pathology that would contribute to misindexing (e.g., incorrect direct beam position or incorrectly specified goniostat or detector parameters), then the best thing to do is to continue in *P1* and integrate the frames. If the frames process well in *P1* and the **Bravais Lattice** table continues to show high distortion indices for lattices other than primitive triclinic, then that is your lattice. If the frames do not process well in *P1*, then there is indeed something wrong with your indexing or with your data. In other words, if *P1* indexes and integrates well, the lattice could be triclinic. If *P1* does not index and integrate well, then the data or their processing is bad, and you cannot rule out other lattices yet. When you scale the data, you can check the space group. If you have chosen the wrong lattice, you will need to return to this step and re-index the data.

### Positional and Partiality $\chi^2$ values

The  $\chi^2$  values represent the average ratio, squared, of the error in the fitting divided by the expected error. In other words, it represents how close your observed errors are to those you predicted you would observe.  $\chi^2$  analysis is the foundation of Bayesian statistics. The numbers in parentheses are the number of overlaps for each class of reflections.



Refinement Information					
Space Group: P4					
Resolution: 50.00 - 1.46					
Positional:	1098	X- $\chi^2$ :	0.78	Y- $\chi^2$ :	0.71 (0)
Partiality:	1270	$\chi^2$ :	1.10		(84)
X Beam: 150.055 Y Beam: 148.844					
a:	79.06	b:	79.06	c:	38.06
$\alpha$ :	90.00	$\beta$ :	90.00	$\gamma$ :	90.00
Crystal Rotation X:			140.382	0.011	0.004
Crystal Rotation Y:			1.722	-0.002	0.003
Crystal Rotation Z:			2.408	0.002	0.008
Detector Rotation X:			0.010	-0.008	0.008
Detector Rotation Y:			0.013	-0.003	0.009
Detector Rotation Z:					
Crossfire X:			0.223	0.022	0.034
Crossfire Y:			0.228	-0.022	0.033
Crossfire XY:			-0.009	-0.001	0.049
Distance:			75.138	0.026	0.016
Mosaicity:					
Y Scale:			0.99898	0.00026	0.00024
Skew:			0.00033	-0.00003	0.00017

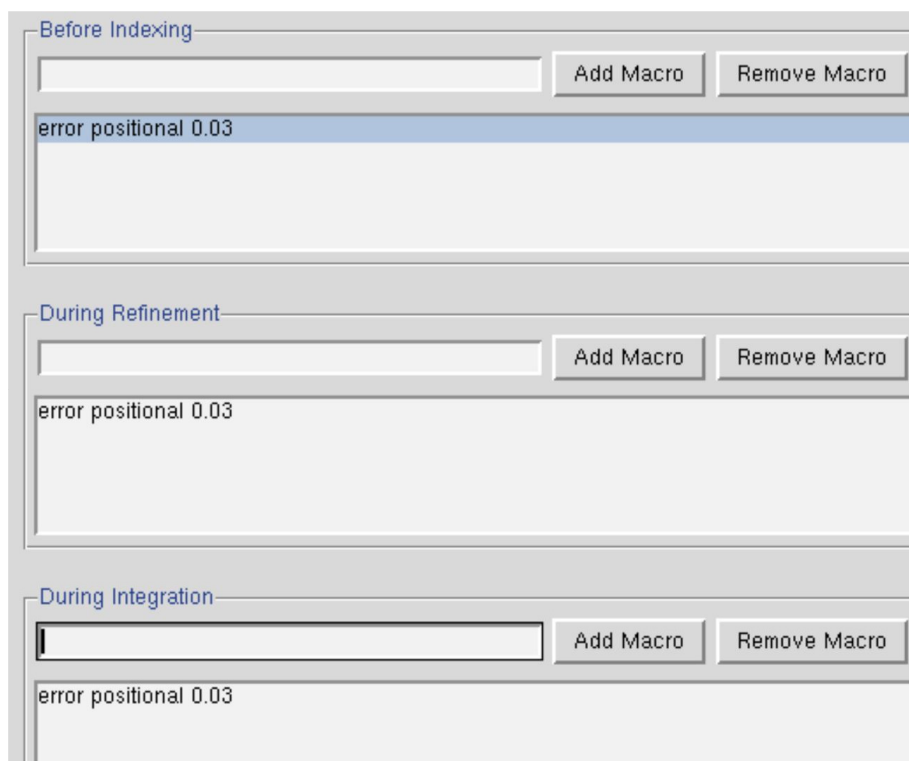
Figure 60. The refinement information using Fit All

A good refinement will have  $\chi^2$  values near 1.0. In the early stages of refinement, these values can be high, but towards the end they should be below 2 or so. All other things being equal,  $\chi^2$  values for x and y positions are sensitive to the **Spot Size** - a larger **Spot Size** will result in larger  $\chi^2$  values. Choosing a very large spot allows you considerable latitude when choosing the center of the reflection since small changes in position won't appear to have much effect on the integrated spot, and thus you will not have "pinned down" the center of the spot very well. A very small spot size, on the other hand, will be quite sensitive to the center of the reflection, since small changes here would have you integrating the sides of the reflection and not the center, and so you will detect that one place is good to be and another is worse.

The magnitude of the  $\chi^2$  values is not a critical test for the success of *Denzo* since they represent only the comparison of the spatial differences between the observed and predicted reflections to an error model that is typically quite strict in the prediction of small errors. Assuming default values of the error estimates,  $\chi^2$  values of 2 or even 3 are acceptable, because the position of the predicted reflection, and hence the intensity, is still very accurate. Thus, the  $\chi^2$  values may not reflect errors in the integration of the reflections.

Note that you can "improve" your  $\chi^2$  values (get them closer to 1) by changing the error model. The most likely candidate for change is the value for "error positional." This value represents the error, in mm, in the positional measurement of the spot center. The default values of **Error Positional** are set in the program for "typical" crystal for a particular detector. This can be changed by going to the **Macros** tab and entering the keyword "**ERROR POSITIONAL** [value]" under the fields *Before Indexing*, *During Refinement*, and *During Integration* where [value] is replaced by your new error, e.g., 0.03 or 0.04 (Figure 61). Some detectors (and therefore some beamlines) will almost always benefit from using this macro.

## LATTICE PARAMETER REFINEMENT



**Figure 61. The Macros dialog**

You should be cautious about increasing the *error positional* value, especially for a well-tested and characterized detector, because the high  $\chi^2$  values may reflect underlying problems with your crystal, rather than an actual error with the detector. For example, your crystal may be slightly cracked, with incomplete spot splitting, and the spot shape will be enlarged and distorted, reflecting the contribution from the multiple lattices. This will lead to ambiguities in determining the centers of the diffraction spots, which will be reflected in larger  $\chi^2$  values.

It is a good idea to keep your error model roughly the same from crystal to crystal so that you can use the  $\chi^2$ 's to pick up any other variations in your data collection. For example, if you are getting  $\chi^2$ 's in the range of 2 to 5 consistently and your crystal is good, you probably have problems with detector calibration. Note that most people think that their crystals, like their children, are outstanding. At least for crystals, this is not always the case.

If the  $\chi^2$  values are very high, (> 10 or so for default values of the error parameters) something is seriously wrong with the indexing, refinement, or the detector. This will be apparent from visual inspection of the displayed image in any case.

## **LATTICE PARAMETER REFINEMENT**

Since the goal of *Denzo* is to produce a list of *hkl*'s and unscaled intensities, it does not matter that the errors in the positions differ a bit from the standard expected error. On the other hand,  $\chi^2$  values substantially different from 1 in scaling should be investigated, because this directly affects a critical result of the scaling procedure, namely the  $\sigma$  value assigned to each scaled intensity.

---

### **Refinement Options**

#### **FIT BASIC** parameters

- X beam and Y beam position of the direct beam at the detector (in mm)
- Crystal unit cell length (in Angstroms) and angles
- Crystal orientation rotations (in degrees)

#### **FIT ALL** parameters

- All the **FIT BASIC** parameters
- Detector rotation X, Y, Z
- Crossfire X, Y, XY
- Detector to crystal distance (in mm)
- Mosaicity (when **3D Window** is > 1)
- Y scale
- Skew

### Refinement parameters are often correlated

Someone told me to gradually increase the resolution and alternate cycles of fitting a few parameters and fixing things. I'm patient, and I want to be careful. **Why shouldn't I do this?**

This is the wrong way to go about indexing and integrating your images. Back in the old days, people would get an initial index at low resolution, fixing all parameters except for, say, the crystal rotations. Then they would fix the rotations and refine the distance. Then they would fix the distance and refine the unit cell parameters. Then maybe they would increase the resolution a bit and repeat the process. The problem with this approach is that single parameter refinement may not allow correlated parameters to be sufficiently refined. In other words, you could get trapped into overfitting one variable and underfitting another, resulting in both variables being incorrectly determined. So unless you know with near-absolute certainty that a variable is correct, cycles of fixing and fitting is the wrong way to go.

People used to take these kinds of approaches because refinements were unstable and simultaneous fitting of multiple correlated variables led to instability in the refinement. High correlation makes it possible for the errors in one parameter to compensate partially for the errors in other parameters. If the compensation is 100%, the parameter would be undefined, but the error compensation by other parameters would make the predicted pattern correct.

*HKL-2000* overcomes this problem in two main ways. First, eigenvalue filtering (described in *Numerical Recipes* as Singular Value Decomposition) is employed to remove the most correlated components from the refinement and make it numerically stable. Eigenvalue filtering works reliably when the starting parameters are close to the correct values but it may fail to correct large errors in the input parameters. Therefore, once the whole data set is integrated, the global refinement (sometimes called postrefinement) can refine crystal parameters more precisely. The unit cell used in further calculations should come from the global refinement (in *Scalepack*) and not from the *Denzo* refinement.

Second, refinements are generally carried out at the maximum resolution of the diffraction data, and, with the 3D Window, over multiple successive frames. More data allows more parameters to be fit simultaneously, and increasing the dimensionality (by adding a third dimension) further constrains variables that might be poorly fit in only 2 dimensions (like the distance, or the third unit cell dimension). These advances are made possible by improvements in the program as well as improvements in computer speed, data storage, and memory.

## Starting refinement over

To start over completely, you can go to the **File** menu and quit the program. However, if you just want to try auto-indexing again, you can click on the **ABORT REFINEMENT** button, go

## LATTICE PARAMETER REFINEMENT

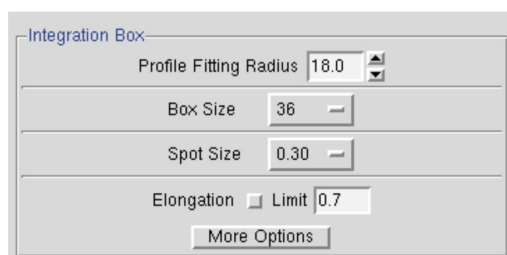
back to the image and re-do the peak search, and proceed as before. To re-set the space group without aborting the refinement, you can re-open the **Bravais Lattice** table and choose another lattice, like primitive triclinic. Click **APPLY & CLOSE**.

## Resolution Limits, Fitting Radius, and Spot Size

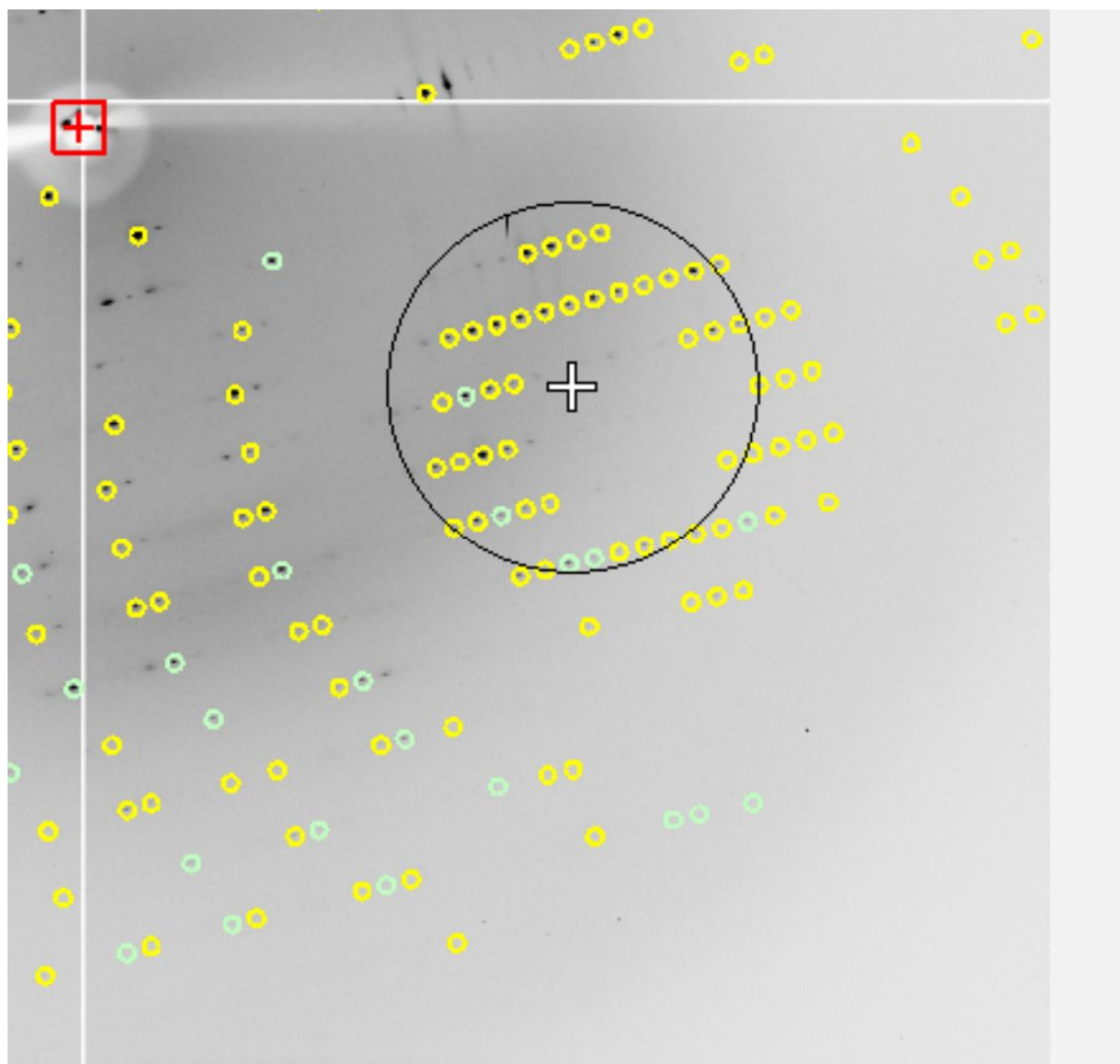
While referring to the indexed image, carefully set the correct **Resolution** limits, **Profile Fitting Radius**, and **Spot Size**. Click the **REFINE** button to update the display after changing any of these parameters. If the Crystal Rotation Y value displayed in the Refinement Information panel is colored red (i.e., for  $450 < \text{Crystal Rotation Y} < 1350$ ) click on the reference zone button in the Controls panel and choose a new zone with a better value for rot<sub>y</sub>.

## The Profile Fitting Radius

The **Profile Fitting Radius** is the radius of the area around each particular spot, in mm, containing neighboring spots, used to calculate the average spot profile (Figure 63). The spot in question is fitted to the average profile of all the spots within the specified radius. Generally, the radius is set so those spots on roughly 3-5% of the area of the detector are included in the averaging. You can display the profile fitting circle using the **PROF FIT R** button



**Figure 62** The Profile Fitting Radius is set in the Integration Box



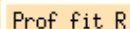
**Figure 63.** *The profile fitting radius can be displayed and moved around.*

The particular size of the area chosen depends on how much the profiles vary across the film and the density of the spots on the film. For example, if the spot profiles vary a lot across the film, then you would choose a smaller radius. If you have a small lattice and the spots are widely separated, you might want to choose a larger radius.

The calculation of the average profile is a time-consuming task, and it is proportional to the number of spots in the profile fitting radius circle, but too small a radius will not capture enough spots and lead to a noisy average profile. Too large a radius will average out significant profile variations.

### Profile Fitting Radius in XDispX

To see the **Profile Fitting Radius** in the *Image Display* window, click the **PROF FIT R** button (Figure 64). A thin circle corresponding to the profile fitting radius will appear and will move around with the cursor. The pred display changes now too. Instead of seeing all of the predicted reflections, you see only those preds that correspond to the reflections that are stronger than the refinement sigma cutoff (usually set at  $5\sigma$ ). Only this subset is shown because these are the only reflections used in the calculation of the average spot profile within the radius. (Note: to return to the normal mode, click **PROF FIT R** button again).



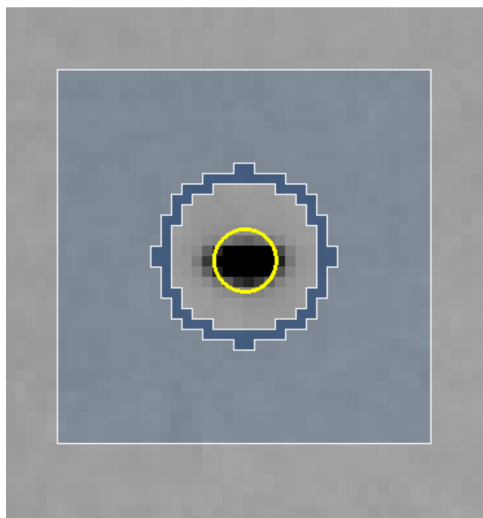
**Figure 64. Profile fitting radius button**

A good rule of thumb is that the Profile Fitting Radius should enclose about 10 to 50 spots whose intensities are above the weak level. Useful increments of the profile fitting radius are 2.5% of the detector size (e.g., 5 mm for a 200 mm wide IP).

The profile fitting radius can be set in the **Index** window under the *Integration Box* panel in the lower-left quadrant of the window (Figure 63). The value is given in millimeters, and you can adjust it iteratively, hitting the **REFINE** button in between your guesses. In general, it is preferable to enlarge the radius to make sure that there are no orphan reflections (reflections that, when placed in the center of the circle, have no other non-weak reflections within the circle). However, keep in mind that if you are using a **3D Window** greater than 1, that the **Profile Fitting Radius** is actually the “Profile Fitting Sausage,” since it encompasses the stack of frames in the **3D Window**, and more spots may be used than first seems from examining a single frame.

## Setting the Spot Size

The final parameter to set is the **Spot Size**. This is generally a function of the size of your crystal and the quality of the X-ray beam (collimation, crossfire, parallelism, etc.), but it must be set for proper integration. Each reflection can be thought of as a buffered region (usually a circle) within a box. As seen in Figure 65, each reflection can be divided into three different regions. Here, the light blue region is the background region, and the measured reflection is in the unshaded region in the center. The dark blue region is a guard region, which is not used to calculate either the reflection or the background. The yellow circle is the predicted position of the reflection. Its size is not used in calculations.



**Figure 65. A Single Reflection**

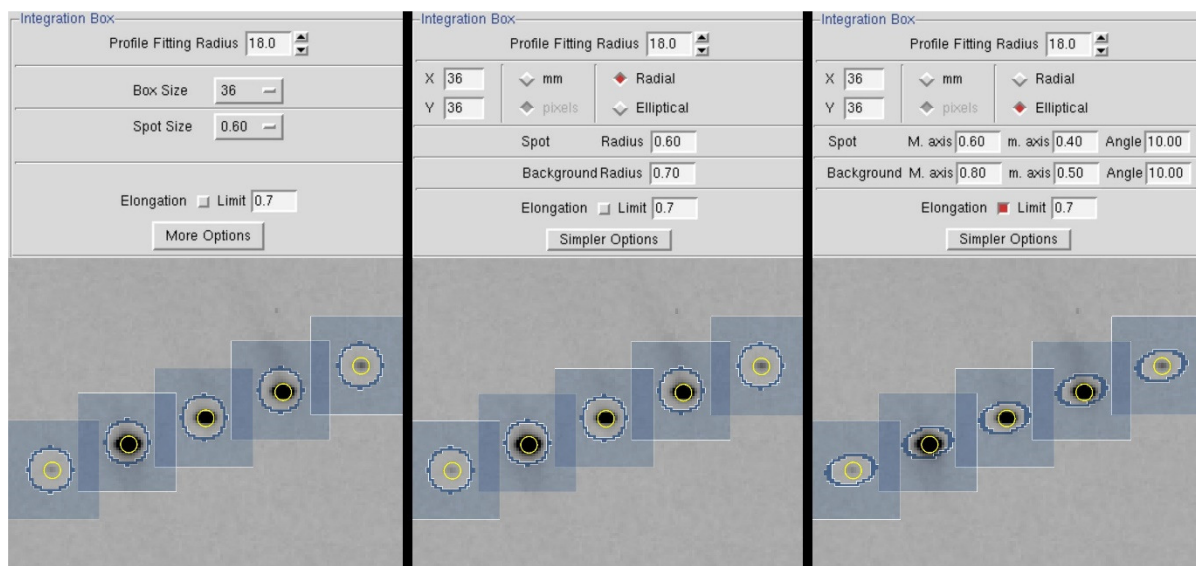
If you choose too small a **Spot Size**, significant data is lost because the reflections will be rejected because of the reflections extending into the background region. Choosing too large a **Spot Size** runs the risk of 1) rejecting reflections due to overlaps, 2) including portions of the neighboring reflection in the measured intensity, and 3) including too much of the background in your intensity measurement.

Setting the **Spot Size** is somewhat of a tradeoff, since not all the spots on an image are of the same size, and one spot size and shape must suffice for the entire set of images (with the exception of correcting for the elongation caused by the intersection of reflections at high angle to the detector face, which is corrected by the parameter **Elongation Limit**). However, it is unlikely that your data will suffer too much because profile fitting limits the damage caused by including the background in the spot. As a general rule of thumb, it is best to try to adjust the spot shape to fit some of the medium to strong spots on the image.

The spot can be either circular or elliptical. Clicking on the **MORE OPTIONS** button gives you more control of the spot and background limits. You also have the option of defining elliptical spots if your spot shapes are sufficiently elongated (Figure 66).

After making the changes to your spot shape, or turning on the **Elongation Limit** option, you can verify that the spot shape has changed by examining the image and the predicted spots in the **Zoom** window of the **Image Display** window (you may have to click **UPDATE PRED** to show the predictions).





**Figure 66.** The Integration Box and Zoom windows are shown with simple options (left), radial options (center), and elliptical options (right).

## What is the Elongation Limit?

Note that *Denzo* now has an option to take into account the radial elongation of the spots that arise from the intersection of the diffracted X-rays at a high angle with the detector. This is activated by checking the **Elongation Limit** button in the [Integration Box](#) panel. The default value for **Elongation Limit** (expressed as the fraction of the box size the spot is allowed to elongate) is 0.7. The range is from 0 to 1.

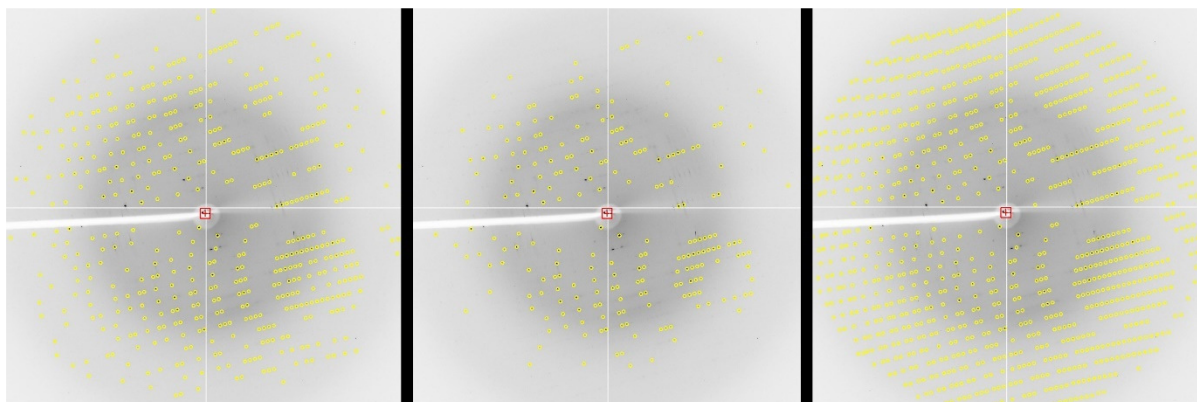
## Setting the integration resolution limit

Examine the indexed diffraction pattern carefully. The cursor can be used to determine the resolution of a particular spot. In general, you want to be on the generous side (say 0.2 Å) with setting the resolution limits, since they can always be lowered in **Scaling** if you were too optimistic. Conversely, if you are too stingy with the resolution limits in *Denzo*, you will have to go back and re-integrate all your frames again since *Scalepack* cannot increase the resolution limits beyond what you have set in *Denzo*. You may have to darken the image using the histogram slider to see the high-resolution reflections clearly. Using the **Zoom** function also helps. Once you have determined the proper resolution limits, go back to the **Index** window and input the new resolution limits into the proper box. Then hit the [REFINE](#) button again, and the new predicted reflections will appear in the diffraction pattern window. You may have to do this a couple of times until you get it right. Look in all quadrants of the diffraction image, (and if possible, multiple frames) as crystals often diffract anisotropically.

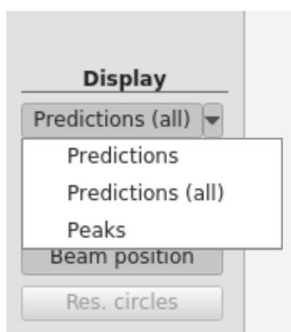
Another feature that you can use to estimate the resolution of your data is the predicted reflection circles. By default, reflections that are above the sigma cutoff for indexing will be displayed. If you set this limit to 2 sigma, you can get a good feel for the resolution you can get by checking the resolution of the reflections closest to the edge (or corner). You can also compare the predicted reflections above the sigma cutoff with all predictions to get a feel for

## LATTICE PARAMETER REFINEMENT

the fraction of reflections above the cutoff limit. To toggle which reflections are shown, use the selector found next to the prediction toggle button. As mentioned above, it is good practice to process the diffraction images to a resolution that is slightly better than you think your ultimate resolution.



**Figure 67.** *Predicted reflections at 2 sigma, 20 sigma, and all reflections.*



**Figure 68** *The Prediction selector*

One more thing you can do to get a feel for the resolution of your diffraction is to display the resolution circles. Found in the resolution dialog box, you can set the resolution of five circles to display (Figure 69). These can be toggled on and off using the **RES CIRCLES** button.

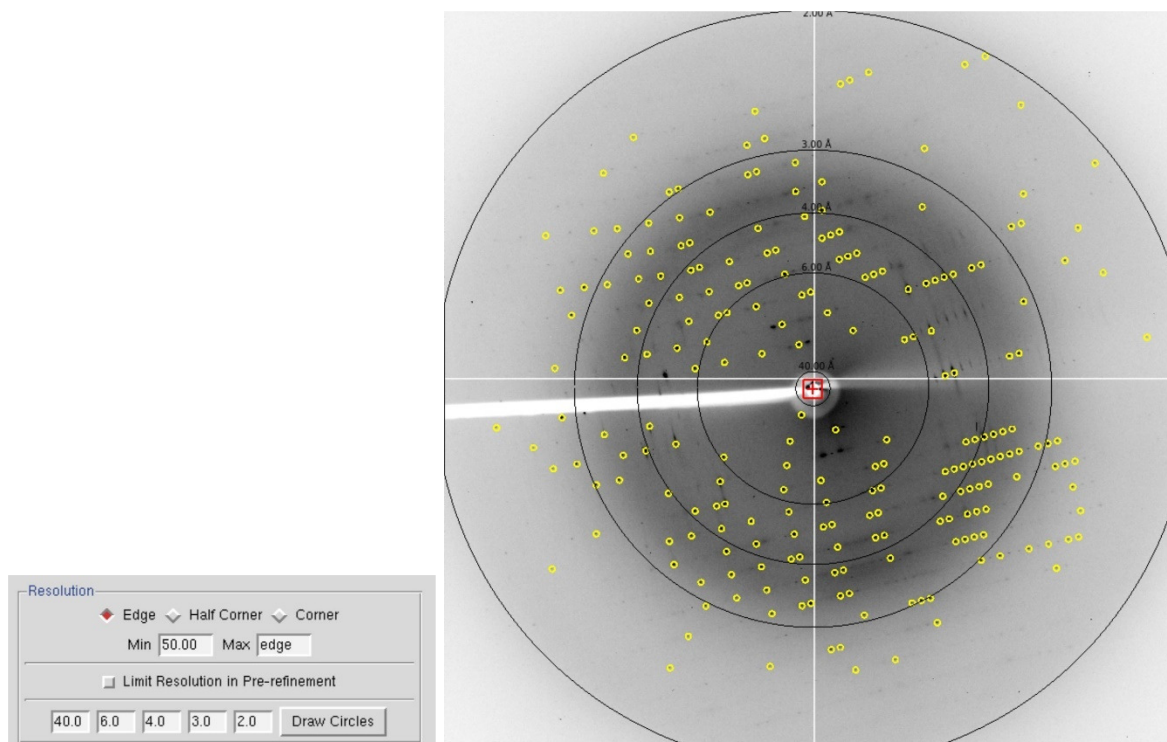


Figure 69. Resolution circles dialog and the displayed circles.

## The Crystal Rotation X, Crystal Rotation Y, and Crystal Rotation Z parameters

The crystal *Crystal Rotation X*, *Crystal Rotation Y*, and *Crystal Rotation Z* angles define the orientation of the crystal lattice in space for the calculation of the predicted reflections. It is inconvenient and non-intuitive to work with images in the data conventions because we are used to thinking about diffraction geometry in terms of the X-ray beam, the spindle axis, and the laboratory frame of reference (horizontal, vertical). There are two possible conventions within this "intuitive" set:

*Gravity/beam system (used in XDisp, but not used in Denzo)*

In this system, the X-ray beam is defined as being perpendicular to the direction of the force vector due to gravity ("down"). The beam is the  $z$  axis, gravity is parallel to the  $y$  axis, and the  $x$  axis is perpendicular to both of these. In terms of crystal orientation rotations

- *Crystal Rotation Z* would denote rotations of the crystal around the **beam** axis,
- *Crystal Rotation Y* would denote rotations of the crystal about the **vertical** axis
- *Crystal Rotation X* would denote rotations of the crystal about the **horizontal** axis.

This convention has the advantage that it does not change with different camera spindle geometry. In addition, it corresponds directly to the common orthogonal detector directions, in that the detector directions  $x$  and  $y$  on the film or IP are perpendicular to the X-ray beam. A disadvantage of this system is that the  $x$  axis may or may not correspond to the spindle axis,

## LATTICE PARAMETER REFINEMENT

depending upon the  $\chi$  (chi) and  $\Omega$  (omega) setting angles of the 3-axis goniostat. For the purposes of aligning your crystal, this makes life more complicated, as we are used to thinking about aligning crystals by moving the arcs on a goniometer head. For  $\chi$  equal 90 or 270 degrees and  $\Omega$  equal to zero. However, the spindle axis is parallel to the  $x$  axis, and thus **Crystal Rotation X** corresponds to rotations about the spindle axis.

Although this system will not (formally) work in outer space (where the gravity vector is essentially zero), nor in the event that the X-ray beam is vertical, the fact remains that (so far) all macromolecular crystallography is done with horizontal X-ray beams on the surface of good old planet Earth, and thus it is a viable option. While *Denzo* does not use this convention, *XDisp* does.

### *Spindle/beam system (used in Denzo)*

The spindle/beam system is the *Denzo* crystal and cassette orientation convention. In this system,

- the  $z$  axis is again parallel to the **beam**,
- the  $x$  axis is parallel to the **spindle** axis and
- the  $y$  axis, also termed the "**vertical**" axis (for want of a better word), is perpendicular to the spindle and beam axes.

In terms of crystal orientation rotations:

- **Crystal Rotation Z** would again denote rotations of the crystal around the **beam** axis,
- **Crystal Rotation X** would denote rotations of the crystal about the **spindle** axis and
- **Crystal Rotation Y** would denote rotations about the axis perpendicular to the beam and the spindle.

The advantage of this system is that it is very intuitive to the crystallographer. In addition, with Eulerian  $\chi$  values of 0, 90, 180 or 270 degrees the spindle and vertical axes are parallel to the  $x$  or  $y$  axes of the detector. The main disadvantage of this system is that it is dependent on knowing the geometry of the camera.

While the **crystal** and **cassette** orientations follow the spindle/beam convention for *Denzo*, the beam, boxprintout, spot, margin, film width and length in the *Denzo* log file follow the data convention.

The **X-beam** and **Y-beam** values are the distance from the edge of the detector data collection area to the beam spot, in mm. These are perhaps the most important parameters.

**Crossfire** is a measure of the X-ray beam divergence and focusing as it leaves the collimator and illuminates the crystal. Crossfire, being a symmetric tensor, has  $x$ ,  $y$ , and  $xy$  components. It affects the prediction of partial reflections and their positions, not their angular width. It is expressed as angular divergence of the beam. The default value is zero crossfire, i.e., a perfectly parallel beam or a beam focused on the detector.

## The Reference Zone

The **Reference Zone** shows the current zone and other zones which are crystallographically equivalent to the current zone. Keep in mind that upon indexing the crystal for the first time (which is just another way of stating that the crystal orientation parameters are being determined) *Denzo* makes an arbitrary choice of reference zone (Figure 70). Being able to choose an equivalent zone is useful in a number of ways.

1. If the crystal *Crystal Rotation Y* happens to fall between 45 and 135 degrees, the refinement of the crystal orientation is more difficult for the underlying algorithm (although it will still work). Choosing an equivalent zone from the list is an easy way to overcome this. [Change the reference zone if the Crystal Rotation Y parameter is colored red.](#)
2. If two or more data sets are collected from the same crystal, without removing the crystal or changing its orientation relative to the goniostat, and the data sets are indexed independently of one another, then it is possible that *Denzo* will arbitrarily assign a different reference zone to the two data sets. Unfortunately, this will result in a different original index for reflections which otherwise should be equivalent, and so merging reflections between data sets will not work. However, by choosing the same reference zone for both data sets, scaling and merging reflections can be done correctly and accurately.





















Reference Zone			
 h0l	 hk0	 l0h	 kh0
 h-hl	 0kl	 0lk	 0k-l
 hhl	 0-k	 k-l	 k-h
Equivalent Orientations			
	X	Y	Z
Current:	-9.205	-163.786	-14.628
	170.795	-16.214	-14.628
	-9.205	-163.786	165.372
	-9.205	16.214	14.628
	170.795	163.786	-165.372
	-9.205	16.214	-165.372
	170.795	163.786	14.628
	-9.205	-163.786	-14.628
	170.795	-16.214	165.372
Close			

Figure 70. The Reference Zone window

## The 3D Window and Mosaicity

The **3D Window** specifies the number of consecutive frames from which the reflections will be integrated simultaneously. Set the 3D Window so that the oscillation range encompassed by this sector of reciprocal space is at least twice as large as the refined mosaicity. Typically for CCD data, five frames are used, but if small oscillation angles are used, then this value will need to be increased. One principal advantage of the 3D window is that binning frames together can include more reflections in the refinement, thereby increasing the accuracy of indexing on frames that have very few reflections (e.g., small molecule data sets, low mosaicity / small unit cell macromolecular data sets, low-resolution macromolecular data sets). The disadvantage of grouping frames together is that local variations in frames are averaged out (e.g., "bad" frames, frames where the crystal slipped, etc.)

### The 3D Window and Image Plate Detectors

Unlike CCD detectors, which are solid state, image plate detectors move the IP relative to the scanning head in the course of data acquisition, data read-out, and plate erasure. While these devices are very well made, there can be a small non-reproducibility in their operation that results in slight variations in the actual position of the direct beam position and the distance. Until it has been established that this variation is small relative to the pixel size, the 3D Window should be kept at 1 to avoid inaccurate mapping of the pixels of sequential frames onto each other. That's why the default value for the 3D Window is 1 for these sites.

The limitation for 3D Window usage is primarily computer memory. All the frames of the 3D Window must be held in memory simultaneously; however, most modern computers have sufficient memory to handle reasonable values of the 3D Window parameter.

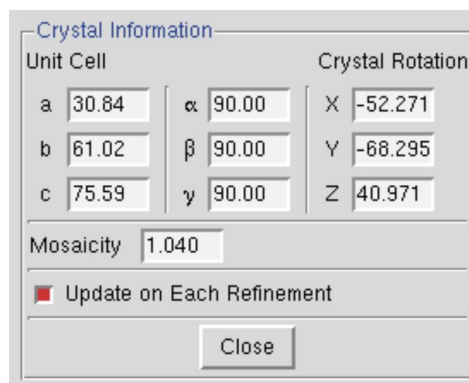
For small molecule data collection, the reciprocal lattice is very sparsely populated and it is reasonable to set the 3D Window to 30 or more frames.

**Mosaicity** is defined in *Denzo* as the rocking angle, in degrees, in both the vertical and the horizontal directions, which would generate all the spots seen on a still diffraction photograph. It not only includes contributions due to X-ray bandwidth, beam crossfire, etc. but is also a characteristic of the crystal itself. Mosaicity is refinable when integrating your frames if the **3D Window** is set to be at least twice as large as the refined mosaicity. For example, if your crystal has a mosaicity of about  $1^\circ$ , then a **3D Window** encompassing  $2^\circ$  or more of data should allow for proper refinement.



## LATTICE PARAMETER REFINEMENT

The best way to set the mosaicity is to use the 3D Window and let Denzo refine it in the course of integrating the reflections. If you are unable to use a **3D Window** greater than 1, and the mosaicity is greater than half your oscillation range (which is usually the case), then you will have to set the mosaicity manually. You do this by going to the **Crystal Information** button and entering the **Mosaicity** in the field provided. If you have to do this, remember to uncheck *Mosaicity* in the *Refinement Options*.



Unit Cell			Crystal Rotation		
a	30.84	$\alpha$	90.00	X	-52.271
b	61.02	$\beta$	90.00	Y	-68.295
c	75.59	$\gamma$	90.00	Z	40.971

Mosaicity: 1.040

☒ Update on Each Refinement

Close

*Figure 71. The Crystal Information window*

A correctly set mosaicity will not always result in predictions covering all of the reflections on an image. In particular, very strong low order reflections may require unreasonably large values of the mosaicity for coverage. These reflections may not represent the intrinsic mosaic spread of the crystal, but rather may result from particularly intense diffraction from that lattice point. A mosaicity that is set too high is usually better than one that is set too low, but setting the mosaicity too high could result in an unnecessarily high number of overlaps.

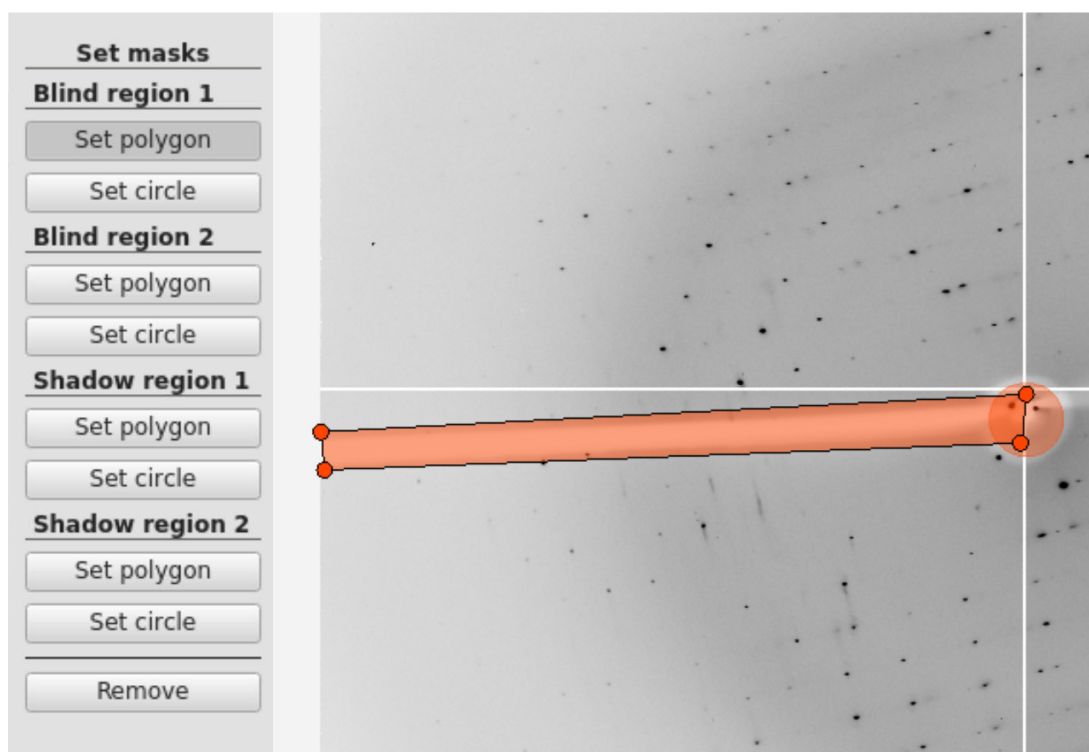
## Masking the beamstop and other shadows

It is important to define the area of the beamstop shadow. If the beamstop shadow region is not defined explicitly, *Denzo* may fail under certain circumstances of high background to reject potential reflections, which are obscured by the beamstop. These few inaccurately measured “reflections” will severely degrade the quality of your anomalous signal. XDispQt allows you to set two distinct blind regions and two shadow regions. Shadow regions are regions where diffraction is not completely blocked, but the background and overall intensity is lower than the rest of the image. This can result from X-rays being blocked by the goniometer. The background will be treated differently in shadow regions.

To open the Masks dialog, click the **EDIT MASKS** button on the left of the display. The dialog (Figure 72) allows you to set to polygon and circular region for each blind region or shadow region. Polygons will be shown with red circles at the vertices, that can be grabbed and moved with the mouse (or stylus). Circular masks have one red circle that is used to set the radius of the mask. Both types of masks can be moved with a click-and-drag motion of the mouse. You can select with which mask to work with by either clicking on the corresponding button in the mask dialog or by clicking within the mask you want to work with. A mask can

## LATTICE PARAMETER REFINEMENT

be removed by selecting it, then clicking on **REMOVE** in the dialog. If you have more than one mask or shadow region, the polygon and circular region will be color coordinated. When you are finished working with the masks, click **FINISH** on the bottom left to remove the mask dialog and return to the normal viewing mode. The mask regions will not be displayed in the normal viewing mode, but the information about the masks will update the site configuration dialog. The masks you set will be used during your current HKL session. If you want to save this information to the site definition, click **SAVE SITE INFO** in the site configuration dialog.

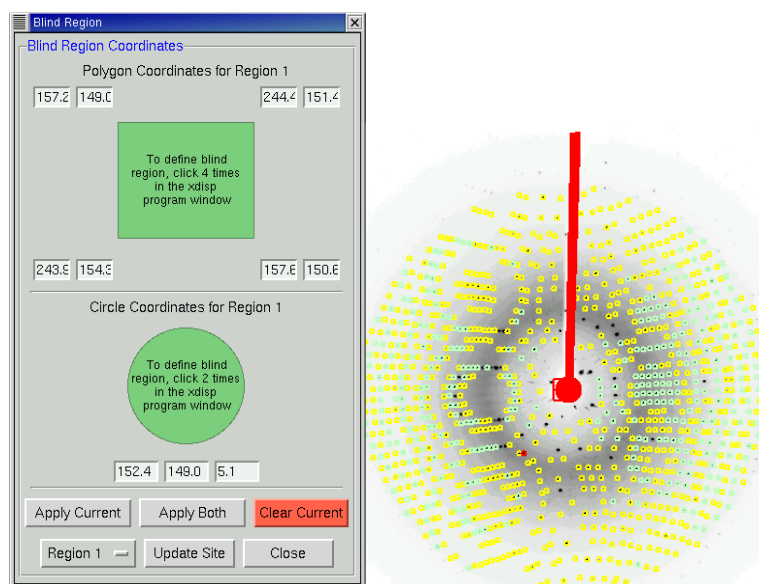


*Figure 72. The Mask dialog (left) and setting the polygon of Blind region 1 (right)*



### Setting the Blind Region in XDispX

To define the beamstop shadow, keep the *Image Display* window open and return to the **Index** page. Click on **Set Blind Region** and then click the top large green box in the new window that appears. Now go to the *Image Display* window and click with the left mouse button four times to define the four corners of the rectangle that corresponds to the beamstop arm. After the fourth click, the blocked out region will appear red. If you don't like the results you can repeat the process, and after selecting your four new vertices the original rectangle will be replaced by your new choice. To define the circular area of the beamstop shadow, click on the big green circle, then move to your *Image Display* window and click once on the center of the circular area. After this you can move the mouse to enlarge or reduce the blind region. You don't have to hold down a mouse button to do this. When you are satisfied, click a second time with the left mouse button and a filled red circle will appear. Going back to the **Blind Region** dialog box, click **apply both**. If you don't like your selection, click **clear current**. If you want to begin integrating, click **apply current**. If you want to store these blind regions permanently, click **update site** and later this information will be forwarded to your **Site Configuration**, so you will be able to use that for all subsequent processing. Please note that this may or may not be appropriate given the geometry of your experimental setup—for example, the size of the beamstop shadow may change as the detector distance changes. Click **close** when you are done.



**Figure 73.** The *Blind Region* definition window and resulting blind region

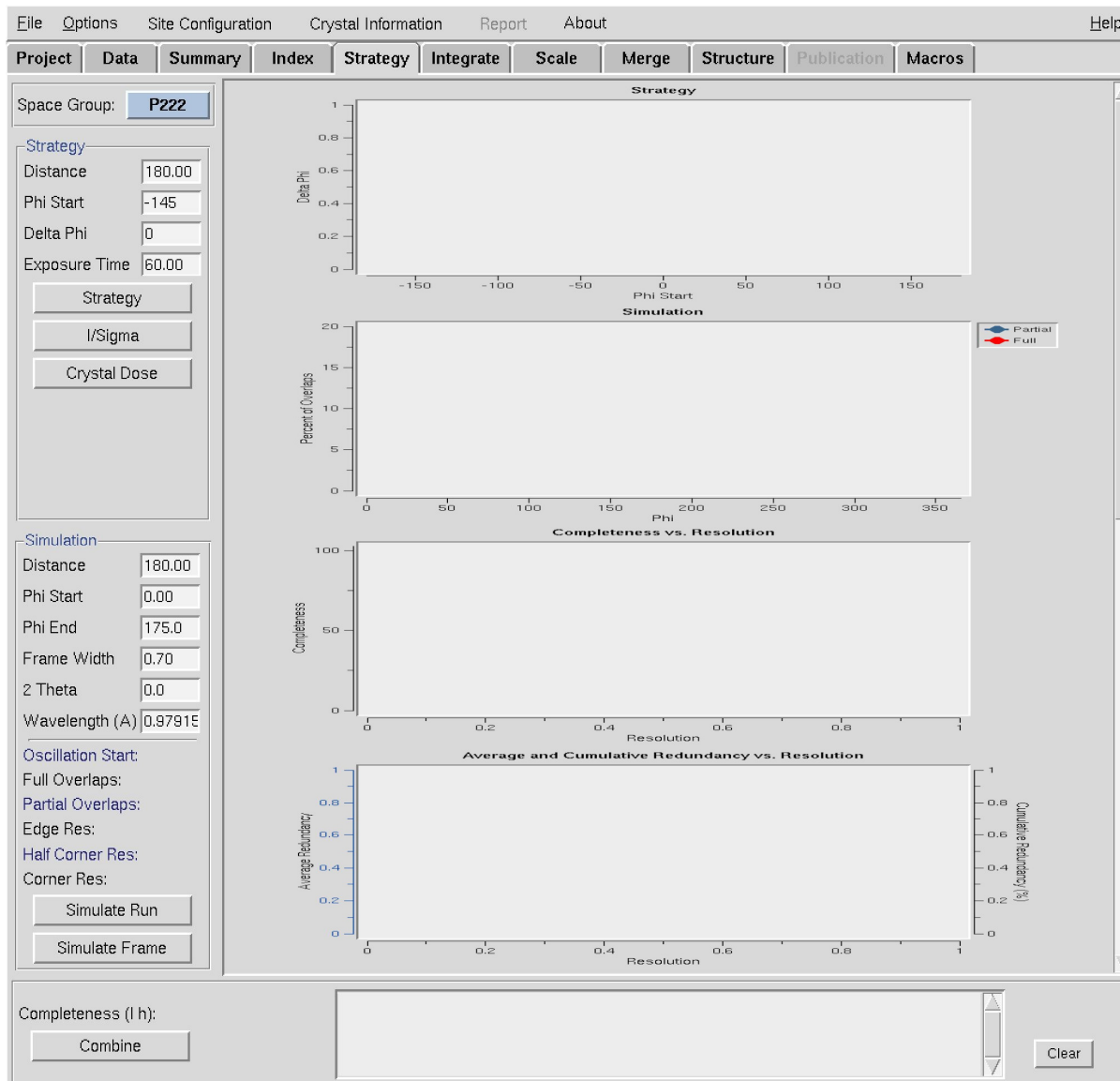
# Strategy and Simulation

## Strategy

Data collection is not an automatic process. In a perfect world, crystals would be perfect and have negligible mosaicity, crystal lifetime would be unlimited, crystal exposures would be short, data collection time would be unlimited, and disk storage capacity would be infinite. In this case, one would just choose a tiny oscillation range and collect 360 degrees of data without worrying about the consequences. However, as we all know, crystals are imperfect and decay fast, data collection time is not unlimited, and storage space is not infinite. Thus, we have to make strategic decisions about how we will go about collecting a diffraction data set to avoid having an incomplete data set or one that is ruined by overlapped spots. This involves making decisions about the oscillation range to collect, the width of each frame, and the exposure time per frame. **The strategy and simulation utilities in HKL-2000 can help you determine an optimal data collection strategy.**

In order to calculate a strategy, you first need to collect and index a few frames. **The strategy module will work with just one frame but will work better if a few frames are available** due to the increased accuracy of the mosaicity calculation. Index a frame or set of frames as usual, making sure that the correct Bravais lattice is selected, the resolution is properly chosen, and refine with the **FIT ALL** regime (including mosaicity with the use of **3D window**). Do not proceed to the integration step.

Once the first frames are indexed, click on the **Strategy** tab. This tab is divided into three regions. Three control boxes are on the left: space group, strategy, and simulation. On the right, there are four initially empty chart areas. The top two are the **Strategy** and **Simulation** charts. A scroll bar on the right is necessary to see the remaining charts, **Completeness vs. Resolution**, and **Average and Cumulative Redundancy vs. Resolution**. The control box at the bottom of the screen is used to combine the results of different simulations (Figure 74).



**Figure 74.** The blank Strategy tab (chart areas compressed to show them all).

By default, the space group of the Bravais lattice will be selected in the Space Group control window, and the parameters of the strategy window will correspond to the values of the frames that have been indexed.

Start the strategy prediction with the **STRATEGY** button. In a few seconds you will see a series of colored lines in the Strategy plot, which shows the number of degrees of data you would have to collect to achieve specific levels of completeness by charting **Delta Omega** vs. **Omega Start**—or depending on the goniostat, the rotation angle (Kappa, Chi, Phi) used for the oscillation (Figure 75). The different lines indicate various levels of completeness, with a color key indicated on the right. At a given **Omega** (Kappa, Chi, Phi) **Start**, the value of **Delta Omega** for each line is the number of degrees of Omega you have to collect starting

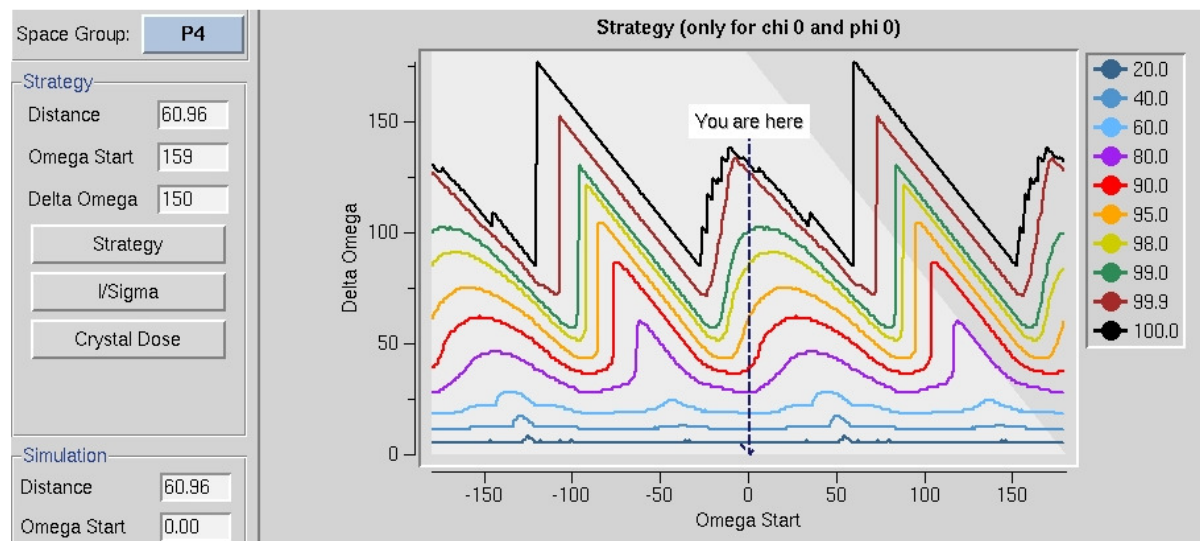
## STRATEGY AND SIMULATION

at **Omega Start** to get the percent completeness represented by the line. In other words, each line corresponds to the percentage of a complete data set that can be collected for a chosen starting oscillation angle (**Omega Start**) and the total number of degrees collected (**Delta Omega**).

Thus, considering the example in Figure 75, if only a 20% complete (dark blue line) data set is desired (although why you'd want that is somewhat mystifying), only a  $6^\circ$  wide oscillation needs to be collected, regardless of starting angle. On the other hand, most people want 95+% complete data sets, so the lines toward the top of the chart are more relevant. The triangular gray region present in Figure 75 marks the region that is not available for data collection. Usually, these regions are inaccessible due to limits set by beamline staff or by hardware limitations. Sometimes at home sources, it is possible to rotate into these regions, but doing so could cause damage to the detector and should be avoided.

Without clicking the mouse, move it around in the window, and a crosshair will appear. Let say you want to collect a 100% complete data set. Line the crosshairs up on the line corresponding to 100% complete. As you move it around, notice that the **Omega Start** and **Delta Omega** values in the control box are updated. To use and collect a complete data set with the smallest possible oscillation range, choose a point on the 100% line at an Omega Start value where the Delta Omega is at a minimum.

Note that for most high symmetry space groups the plots will resemble a sawtooth with a rapidly rising leading (left-hand side) edge. This jagged shape is a result of the symmetry. As you rotate a crystal in a beam, it is possible to collect some data, then collect some symmetry-related data before collecting more unique data. It is better to err slightly on the side of caution and select an **Omega Start** value a few degrees less than the point with the minimum **Delta Omega**, to avoid the high side of the sawtooth. If you do not have a high symmetry space group, the plot will look more like a horizontal line or a gentle wave. In those cases, there is not much strategy to ponder, since the lattice does not allow for much economy in data collection. Just be sure to select a strategy that does not venture into the inaccessible region.

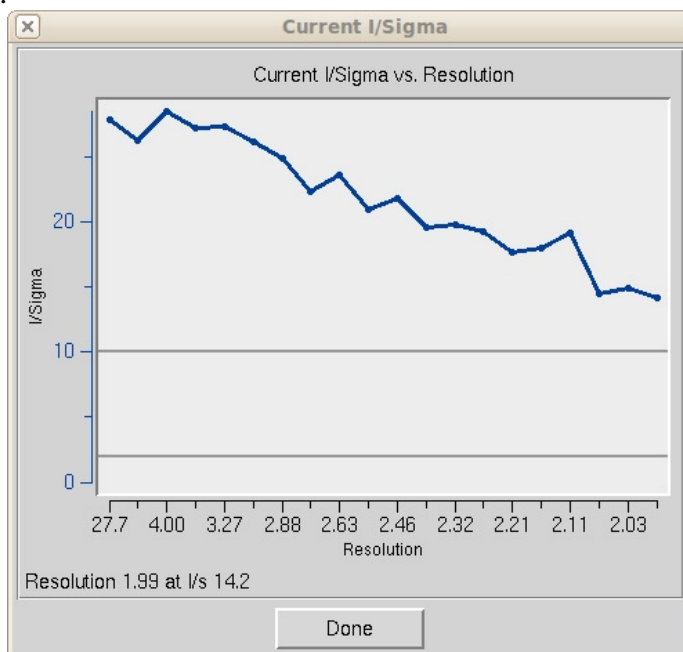


**Figure 75. The Delta Omega vs. Omega Start plot**

Once you have settled on a **Delta Omega** and **Omega Start** point on the plot, click the mouse and the values will be entered into the appropriate boxes in the **Simulation** control panel.

## The I/Sigma and Crystal Dose buttons

You need to collect at least 3 frames to use the **I/SIGMA** button, and the program will estimate (after a quick integration and scaling)  $I/\sigma$  up to the resolution value at the edge of the detector for the specified detector distance. A new window **Current I/Sigma** opens with an **I/Sigma vs. Resolution** plot (Figure 76). Two gray lines are drawn on the plot, at  $I/\sigma = 10$  and  $I/\sigma = 2$  respectively.



**Figure 76. The I/Sigma vs. Resolution plot**

The **CRYSTAL DOSE** button is to be used for X-ray sources with very high flux (typically synchrotron beamlines). It opens a new window **Estimated Crystal Dose** where you insert a correct **Beam Flux** value. The program runs a simulation estimating how radiation damage will cause the  $I/\sigma$  values to decrease at the end of data collection and shows the result as on the  **$I/\sigma$  vs. resolution for Current and Last Frame** plot (Figure 77). Note that this analysis requires the sequence of the protein in the crystal, which can be added by clicking the **EDIT PROJECT** button on the **Project** tab.

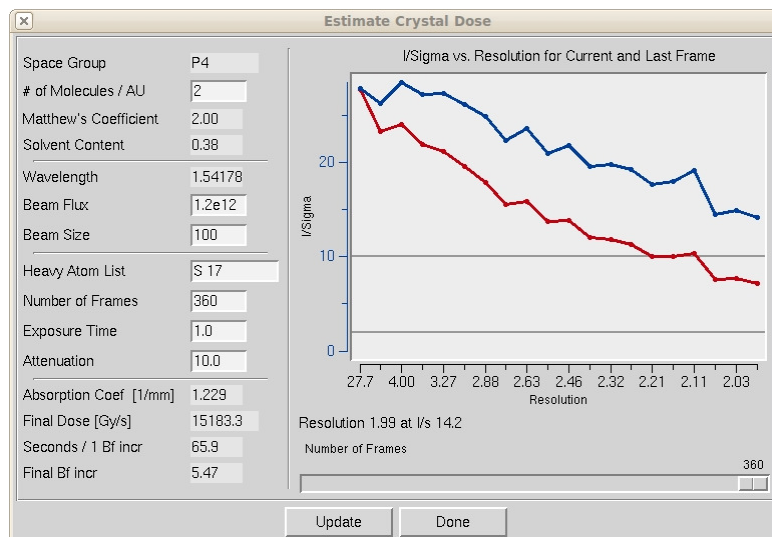
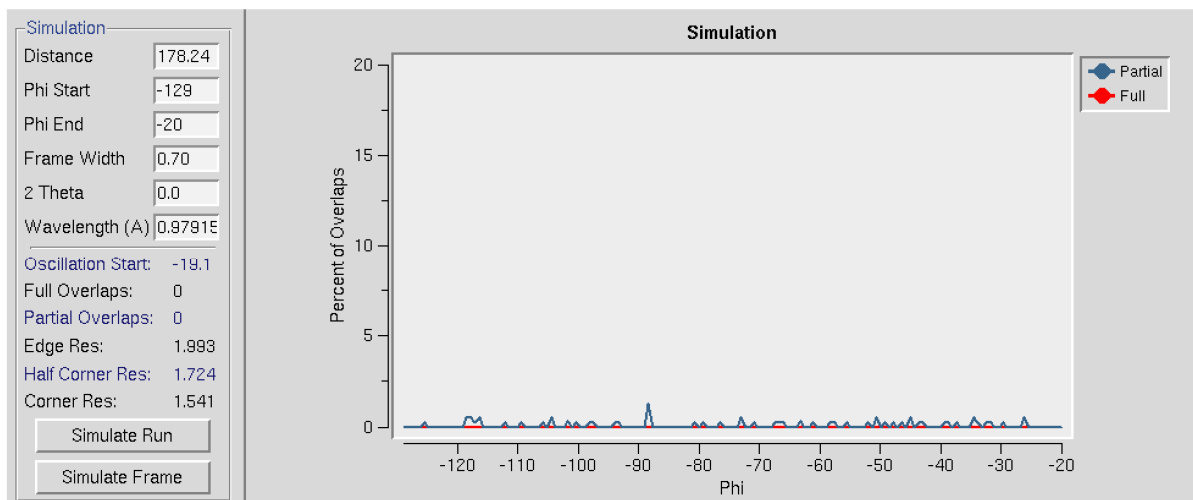


Figure 77. The Estimated Crystal Dose window

## Simulation

The simulation utility predicts the number of overlaps for a data collection strategy as a function of frame width, distance, and oscillation range. This is particularly important if your crystal has high mosaicity or long unit cell axes. In those cases, the spots from some frames may have adequate separation, but then subsequent frames may be rendered useless by overlapping spots as (for example) a long axis that is perpendicular to the spindle axis rotates into a position parallel to the beam. The simulation utility allows you to simulate an entire data collection run, to view the potential overlaps, and then to adjust your data collection parameters before the crystal is exposed and time and valuable crystals are wasted. It is really quite useful.

Once the **Omega Start** and **Omega End** are input, click on the **SIMULATE RUN** button. A mouse click on the Strategy plot will change these values to those that correspond to the cursor position in the plot. You will see two things. First, a plot will be drawn with red and blue lines (which may overlap), corresponding to the percentage of fully and partially recorded reflections that are predicted to be overlapped for each frame you plan on collecting (Figure 78). Remember, if a reflection is overlapped on another, it will not be useful for your data set and be excluded during integration and scaling. Second, on your **Image Display** image window, a new series of predicted reflections will be displayed, corresponding to each frame in your simulated data collection run. Obviously, since you haven't yet collected the data, the underlying diffraction image will not change or correspond to the predicted reflections, but that is OK (this is just a simulation). This will show you visually how your diffraction image will index over the data set collection.



**Figure 78. The Simulation window with almost no overlaps**

The thing to pay attention to is the percentage of overlaps. Typically, you'd like to lose no more than 5% of the reflections to overlap. This is where the simulation is useful. In the control box, you can replace the original frame width with a new value. Got a lot of overlaps? [Decreasing the frame width can reduce problems with overlaps](#), so decrease this value and simulate the run again by clicking on the **SIMULATE RUN** button. Examine the results; the overlaps should decrease. If you only have a few overlaps and want to increase the percentage of fully-recorded reflections, increase the frame width, and simulate the run again. Typically at some point, you will find a reasonable frame width ( $0.2$ - $1.0^\circ$  for macromolecules and  $1.0$ - $5.0^\circ$  for small molecules) that yields an acceptably low number of overlapped reflections. This is not guaranteed, of course, since large unit cell axes and high mosaicity can conspire to ruin a run despite your best intentions.

[Another way to decrease the number of overlaps is to increase the crystal-to-detector distance](#). As the detector is moved back, the spot separation should increase, and the percentage of overlaps will decrease. Naturally, this will limit how high the resolution of the data that can be collected, but it is worth trading off higher resolution data to ensure you can determine the structure. High-resolution data that can't be processed is useless. Perhaps you are one of the unlucky people that have the triple whammy of high resolution, large unit cell, and high mosaicity. In that case, you will be perpetually frustrated, but at least you will be able to develop the most rational data collection plan before exposing your crystals to an excessive amount of radiation. At some point, an acceptable compromise should be reached. You can also consider collecting multiple oscillations of data with varied data collection parameters and then scaling the sets together.

[Once you have settled on an Omega Start, Delta Omega, Distance, and Frame Width, you can input these values into the control software for the data collection run and begin your data collection](#). Typically, the whole strategy and simulation exercise should take only a couple of minutes or less. It's actually easier to do it than to read about it!

If you are working with a precious crystal, it may be a good idea to collect a few frames using the strategy you have devised and check the strategy again. If anything is wrong, you

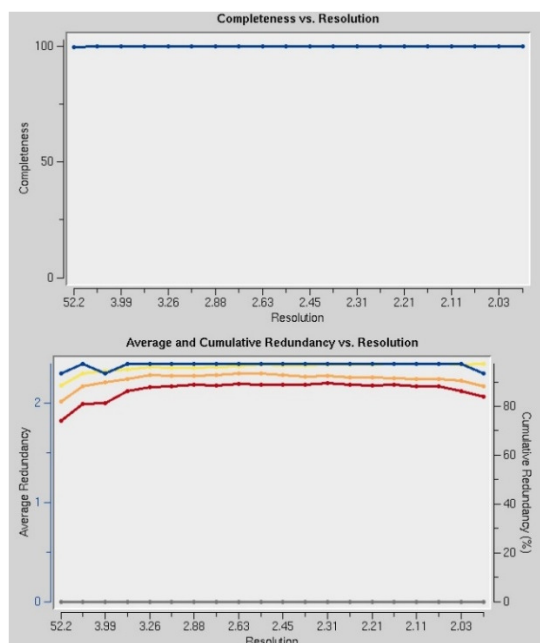


## STRATEGY AND SIMULATION

can stop the data collection before you've irradiated the crystal too much. Errors or inaccuracies in the strategy or simulation, when they do occur, usually arise from an inaccurate estimation of the mosaicity or the unit cell axes, especially when these are derived from just a few frames.

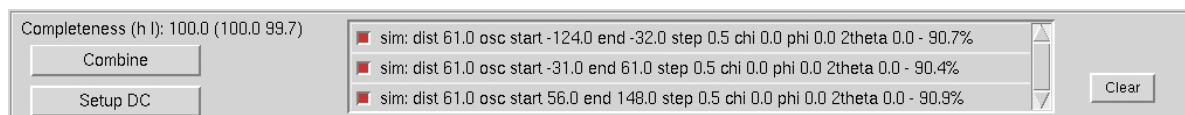
### Combining simulation runs to generate a cohesive strategy

The lower two plots, **Completeness vs. Resolution**, and **Average and Cumulative Redundancy vs. Resolution** (Figure 79), initially show statistics for a single simulation.



**Figure 79. Completeness and Redundancy vs. Resolution plots**

Sometimes due to hardware limitations, it is not possible to collect a complete data set while using a simple strategy, and you will have to combine several “wedges” of data to get a complete dataset. A short description of each simulation run will appear in the panel next to the **COMBINE** button (Figure 80). The completeness in low (l) and high (h) resolution shells is shown above the **COMBINE** button. You may select any combination of simulated runs by checking or unchecking the appropriate boxes next to each simulation run (each run will be selected by default). By clicking the **COMBINE** button, you will obtain statistics and plots for a combination of the selected runs. You may remove all simulated runs from the last panel by clicking the **CLEAR** button. If your version of *HKL-2000* is configured with the data collection module, the **SETUP DC** button exports a chosen strategy to the **Collect** tab of the main *HKL-2000* window.



**Figure 80. The Combine panel**



# Integrating All Frames

## The Integrate button is red. Is this a problem?

The lettering in the button should be colored green. If it is orange, that means that either the  $\chi^2$  values are suspiciously high and should be checked, or else you have failed to select a lattice other than primitive triclinic. If the lettering is red, this warns you that the  $\chi^2$  values are unacceptably high and should be investigated (Figure 81). You can continue, but there is probably a problem you will have to address it eventually.



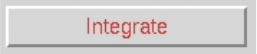
	Positional: 1634 X- $\chi^2$ : 0.98 Y- $\chi^2$ : 0.77 (0) Partiality: 1420 $\chi^2$ : 1.10 (2)
	Positional: 754 X- $\chi^2$ : 1.53 Y- $\chi^2$ : 2.47 (0) Partiality: 754 $\chi^2$ : 0.66 (0)
	Positional: 268 X- $\chi^2$ : 13.71 Y- $\chi^2$ : 4.95 (0) Partiality: 825 $\chi^2$ : 7.35 (16)

Figure 81. The Integrate buttons colored green, orange, red

## Integration

After indexing your data, you need to process all of the images to get a complete dataset, a process known as integration. There are two ways to start integration: clicking on the **INTEGRATE** button in the *Index* tab or the **INTEGRATE** button in the *Integration* tab. Monitor the progress of the integration by examining the  $\chi^2$ , **Cell Constants**, **Crystal Rotations**, **Mosaicity**, and **Distance vs. Frame** plots, as well as the agreement between the predicted reflections and the spots on each image. Ideally, all of these plots will yield horizontal lines, indicating little or no change of the refinement parameters for different 3D Windows in the data set.

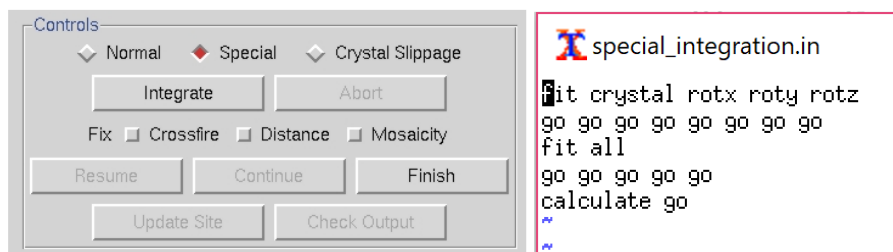
## The Crystal Slippage and Special integration buttons.

The **Special** radio-button opens up the command file used in the integration of the diffraction images. If you want to customize the integration settings passed to *Denzo* (for example, putting in many more “go” statements or some other modification to the standard procedure), this is the place to make those changes ([Figure 82](#)). You can also make these changes under the **Macros** tab in the *HKL-2000* interface.

The **Crystal Slippage** button is a shorthand way of performing the integration in two resolution stages. The first stage is done at low resolution (4 Å) and then is expanded to the final resolution. Normally, integration is done to the highest resolution immediately. However, if the crystal is slipping in position during data collection, the crystal orientation parameters

## INTEGRATING ALL FRAMES

may not be consistent enough from frame to frame to process frames automatically. Using only the low order reflections at first allows a larger radius of convergence, which helps account for slipping crystals. Although crystal slippage is traditionally associated with capillary mounted crystals, it can occur even with cryo-mounted crystals. Sometimes liquid nitrogen can cause the glue or epoxy that keeps the pin in the base to come loose. Additionally, long, thin cryo loops have been known to bend from the pressure of the cryo-stream.



**Figure 82.** The special integration option and the default integration commands

## Integration Information Plots

After starting integration, the display will change and present two graphs at the bottom of the tab. On the left is a mosaicity histogram refinement for the frame that is being refined; on the right is a graphical plot of the integration progress. The graph on the right, contained within the **Integration Information** plots, has radio button style selectors at the bottom that change the display between five different graphs  $\chi^2$ , **Unit Cell Constants**, **Crystal Orientation**, **Distance**, and **Mosaicity**, each shown as a graph of the parameter vs. frame number. One of the most useful plots (diagnostically speaking) is  $\chi^2$  vs. **Frame** (Figure 83). Ideally, both lines (the positional  $\chi^2$  for X and Y) will be in the range of 1 to 2 over the set of processed images. All of the plots should be examined, with particular attention paid to detecting bad frames, which will show up as strong outliers in the **Integration Information** plots. All of the plots should be relatively flat as a function of frame number. You'll also notice that the points for each plot appear in bunches that correspond to the number of frames in the 3D window. If some frames in the  $\chi^2$  vs. **Frame** plot have unreasonably high  $\chi^2$  values, it makes sense to exclude those bad frames from scaling. Methods to do this are described below.

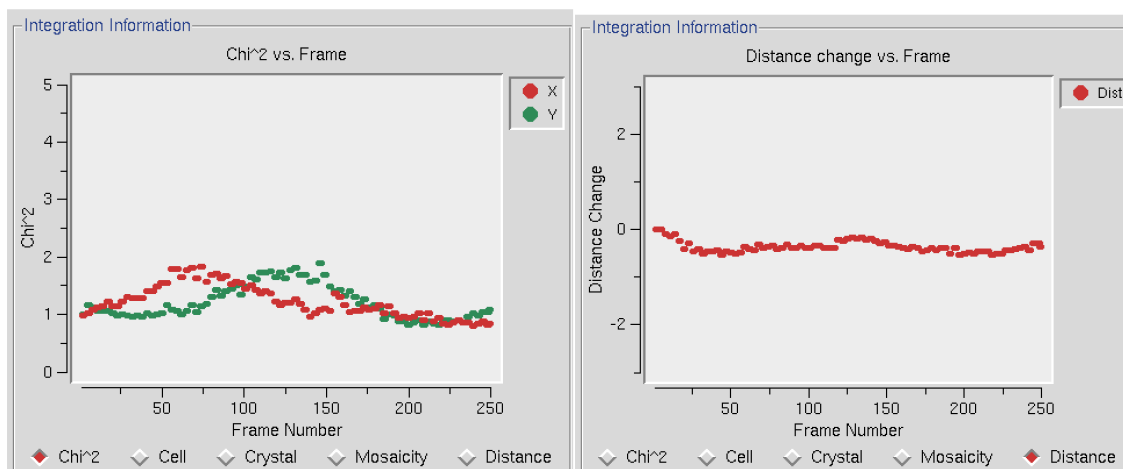


Figure 83. The diagnostic plots:  $\chi^2$  vs. Frame and Distance Change vs. Frame

## What is an acceptable range for $\chi^2$ ?

Ideally, the  $\chi^2$  x and y values should be around 1 (i.e., colored green in the Refinement Information panel; see Figure 60). Sometimes they rise up into the 3 to 4 range. As the integration proceeds, the Image Display window will continually update. If the predicted reflections continue to superimpose on the observed reflections as the integration proceeds, then oscillating  $\chi^2$  values are no immediate cause for alarm. However, if the mosaicity values start refining to high values (greater than  $2^\circ$  for frozen crystals), or the crystal orientation, cell, or distance parameters are changing dramatically, you should consider re-indexing or choosing another lattice. In some particularly problematic integrations, especially when the calculated unit cell parameters are changing substantially, you may try fixing the mosaicity by clicking the **Fix Mosaicity** button on the **Controls** panel (Figure 82). Fixing the mosaicity is particularly relevant for cases where there are several lattices in the beam at the same time. In the best-case scenario, only one of the lattices would dominate the diffraction pattern and be indexed, but sometimes as the crystal rotates other lattices become more prominent and the program may start to confuse them. By restricting the mosaicity, you may be able to effectively limit the number of nearby spots that can be included in the integration.

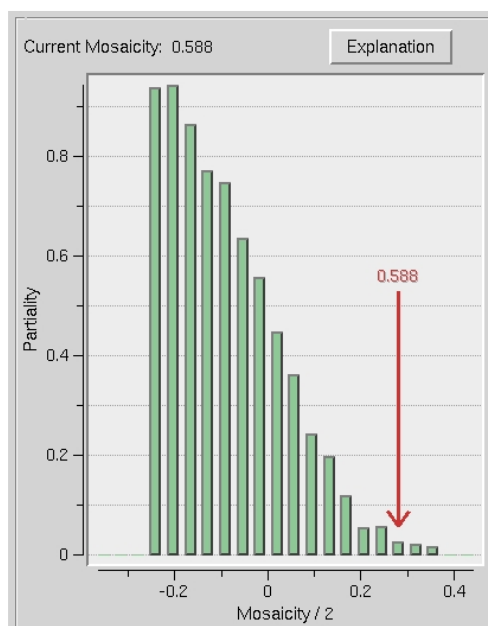
## Drastically larger $\chi^2$ values

Ideally, the  $\chi^2$  values will be close to 1, but values up to 3 and even 4 are sometimes normal and may not indicate a serious problem. However,  $\chi^2$  values that are off the plot that is displayed during integration (i.e.,  $>5$ ) probably mean that something is wrong. The most likely culprit for otherwise “normal” data sets (i.e., where the diffraction pattern on the poorly indexed frames still looks normal, and there weren’t any shutter or goniostat problems), is misindexing due to an incorrect **X Beam** or **Y Beam** value. Experimenting with different beam position values may solve the problem. This is relatively easy to do using the **SET BEAM POSITION** button found in the **Index** panel (described in the Troubleshooting Appendix). Crystal idiosyncrasies (such as slight splitting, which may not be apparent from an inspection

of the diffraction images) or shutter, spindle or goniostat problems can also lead to high  $\chi^2$  values. Problems with the experimental setup will often manifest themselves suddenly, and a sharp change in the  $\chi^2$  values will pinpoint the frames that have been corrupted. It is always a good idea to verify the experimental setup by collecting a data set from a known, high-quality crystal in order to benchmark your setup.

## The Mosaicity histogram

The mosaicity histogram reflects the reflections of the frame or of the frames of the **3D Window**. The reflections are sorted into 20 zones, which range from plus or minus half of the input (or refined) mosaicity (Figure 84). Each zone represents the shortest angular distance of the center of the reflection from the surface of the Ewald sphere at the end of the oscillation range and is formally defined as the average observed partiality of the reflections in each zone. In other words, if, at the end of the oscillation range a particular reflection was  $-0.3^\circ$  away from (i.e., past) the surface of the Ewald sphere, then it would contribute to the “bar” in the  $-0.3^\circ$  zone.



**Figure 84. The mosaicity histogram**

Only reflections that are single partials are included in the analysis. The thickness of the bars is proportional to the number of single partial reflections being integrated in the 3D Window. More reflections will lead to thicker bars. A reflection in negative zones means that the center of the reflection has already passed through the Bragg condition. In all cases, no matter what mosaicity you have input to *Denzo*, the histogram should pass through 50% in the zero zone. If the mosaicity you guessed was too small, then the histogram will not descend to zero at  $+\text{mos}/2$ . If the mosaicity you guessed was too large, then you will see a histogram which looks like a step function, falling to zero just past the zero zone.

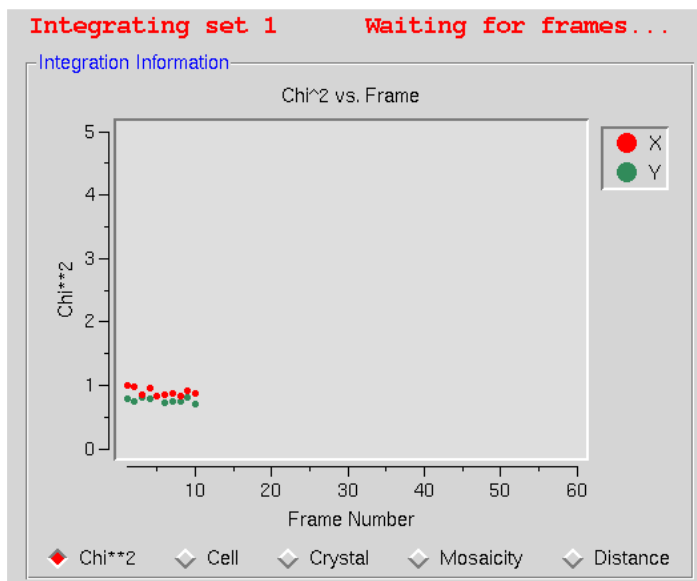
If the mosaicity was chosen correctly or slightly overestimated (which is the preferred side to err on) the histogram will descend smoothly to zero at  $+\text{mos}/2$ . Frequently this is not the case and the histogram tails off due to the effects of diffuse scattering. If there are only a few reflections, the histogram may be quite choppy. The number of reflections can be increased by increasing the size of the **3D Window**. If the histogram does not resemble any of the shapes mentioned here, it may be indicative of motor, spindle, or shutter problems.

### How long should this take? How do I know if the program is stuck?

When the integration is complete, the display will show the message “Integration Complete.” The speed of the integration depends on both hardware and data-specific factors. Among the hardware ones are CPU speed, amount of physical memory, and the speed at which the computer can access the raw data from disk (local disks are faster than remote ones). Data-specific factors include the number of observed and predicted reflections on each image, the size of each image file (larger frames take longer to read from disk), and the number of frames in the **3D Window** (integrating 5 frames at a time will take about 5 times as long as integrating a single frame). For most modern computers and typical macromolecular data sets, this process usually takes 1-3 minutes at most. In my experience, if several minutes go by with no additions to the plots, something is probably amiss. The most likely culprit is slowness in reading the image files from disk, particularly if your frames are on a remote file server.

### Waiting for image frames.

A very nice feature of *HKL-2000* is the program’s ability to wait for image frames during processing. This is indicated by the message “Waiting for Frames,” which appears above the **Integration Information** window during integration ([Figure 85](#)). The program knows to wait for frames based on the value you entered in the **Number of Frames** entry of the **Edit Set(s)** dialog on the **Data** panel. For example, let’s say you are collecting 60 images for a data set, but only 10 have been collected so far, and the others are on the way (e.g., are still being collected, read from the detector, being copied from a file server or CD/DVD, coming over FTP, etc.). When you first set up the data files, only 10 frames will be seen by the program. However, if you enter 60 for **Number of Frames**, the program will know that eventually 60 frames will constitute the complete set of diffraction images. You can start the indexing and integration, and these will proceed through the first 10 (or however many you have) frames currently on the disk. When those have been integrated, the message “Waiting for Frames” will appear. At this point, you don’t have to do anything. As more frames appear, they will be automatically read and integrated. This allows you to monitor the progress of your data collection as it occurs. It also allows you to use your data collection time (e.g., at the synchrotron) more efficiently, by working on another task while you wait for integration to be completed. You can even proceed to scale the files that have been processed and monitor data collection as it happens, giving you the option of cutting data collection short if your crystal dies in the beam.



*Figure 85. The program is waiting for new frames while processing the data*

## Updating the site configuration after processing data

Ordinarily, you would not update the site unless:

- 1) you have the password to do so,
- 2) the refined values of the site parameters (**X and Y Beam Positions, Crossfire, Y Scale**) are substantially different from the original site values, and
- 3) the values you've refined are reliable.

Site parameters determined from weak, poorly diffracting crystals may not be much of an improvement (if at all!) from the default values. Generally, you want to use a superior-diffracting crystal to determine updated site parameter values.

## How do I stop and start over?

To stop the integration hit **ABORT**, and you will get a message box with the question: "Which crystal orientation and cell parameters would you like to keep?" You can choose between three options: "original values" (values as read from the header of data files), "indexed values" (values after the last auto-indexing round), and "current values." Now you can reindex if you'd like.

## Scaling

The *Scale* tab provides the interface to the *Scalepack* program. This window is divided into three sections. The top of the window is reserved for the results and analysis of the scaling process and will be blank until after the first round of scaling. The middle section contains a control panel with options that can be changed before or (and after) scaling the initial round of scaling. The bottom of the display contains the **SCALE SETS** button and a set of controls that only make sense after at least one round of scaling (they are actually not available at first). It is important to know that scaling is an iterative process. To get high-quality data, you will have to scale the data multiple times, adjusting some parameters based on the results of the last round of scaling.

Under most circumstances, the default options will be sufficient to start the scaling process, although you should make sure the Anomalous option is checked if you are conducting a SAD or MAD experiment. It will be set by default if you have selected SAD/MAD as the Phasing Method for the project, but will not be selected if you have jumped into data processing without defining the project. The default **Global Refinement** choice (Small Slippage and Imperfect Goniostat) should be suitable for most cases, but you do have the option of selecting a different option (see the Post Refinement box below for more details).

## Scaling Options

These options affect how the scaling of the integrated diffraction intensities is performed. The default options are: **Scale Restrain** 0.01, **B restrain** 0.1, **Write Rejection File**, **Ignore Overloads**, **Number of Zones** 10, **Error Scale Factor** 1.3, **Default Scale** 10, and **Resolution Limits** as determined from the \*.x file headers (Figure 86). What follows is a discussion of each of the options listed in this panel.

## SCALING AND SCALING OPTIONS

The screenshot shows a software window titled "Scaling Options". It contains the following controls:

- ☒ Scale Restrain: 0.01
- ☒ B Restrain: 0.02
- ☐ Absorption Correction: Low
- ☐ Use Auto Corrections
- ☐ Use rejections on next run
- ☒ Write rejection file
- ☐ Fix B
- ☒ Anomalous
- ☐ Scale Anomalous
- ☐ Include Overloads
- ☐ Direction Cosines
- ☐ Fit Goniostat
- ☐ No Merge Original Index
- Number of Zones: 20 (with up/down arrows)
- Error Scale Factor: 1.3 (with up/down arrows)
- Default Scale: 10 (with up/down arrows)
- Resolution:
  - Min: [Yellow box]
  - Max: [Yellow box]

**Figure 86. The Scaling Options panel**

**Scale restrain** can be used to restrain scale factor differences from consecutive films or batches. The value entered in the box represents the amount you will allow the scale factors to differ from consecutive frames or batches. It adds a factor of  $(\text{scale1} - \text{scale2})^2 / (\text{scale restrain})^2$  to the target function minimized in scaling. This only applies to batches between which you add partials. The value should roughly represent the expected relative change in scale factors between adjacent frames.

For very thin frames, or frames where most of the reflections are partially recorded (yellow, as opposed to green, spots on the image during integration, see [Figure 48](#)), or in sectors whose width is less than the mosaicity, or from data which consists of only a few frames, **SCALE RESTRAIN** is almost obligatory. In these cases, there may not be enough intersections between frames to get accurate scale and B factors. Indeed, what you may see is both scale and B factors ranging all over the place. If things get really bad, the program may crash due to floating-point arithmetic exceptions when taking the exponent of unreasonable B factors. This does not mean that the data is unusable; it simply means that the scale and B factors must be restrained. Note that the restraints only apply to the frames over which you are adding up partially recorded reflections.

It is not correct to try to “bin” the individual frames into larger batches to try to overcome the problem of few intersections between frames. This is because you then lose the ability to **ADD PARTIALS** between the new “bins.” You can, however, overcome the problem in another way, by adding the macro **number of iterations 0** to the **During Scaling** box on the **Macros** panel. In this case, the data will not be scaled but simply merged. This has its drawbacks.

**B Restrain** can be used to restrain B factor differences from consecutive frames or batches. The value which follows the flag represents the amount in  $\text{\AA}^2$  you will allow the B factors to differ from consecutive frames or batches.



## SCALING AND SCALING OPTIONS

**Absorption Correction** applies additional parameters that correct for variations in the degree of X-ray absorbance by the crystal in different orientations. For example, more X-rays are absorbed when the longer dimensions of the crystal are parallel to the incident beam. In essence, this procedure adjusts for the fact that most crystals are not perfect spheres! The absorption correction algorithm parameterizes this effect by an expansion of real spherical harmonics terms (the **Low** or **High** option controls how many terms are included in the correction). This procedure, along with several other scaling corrections applied by *Scalepack*, is described in more detail in Otwinowski et al. Acta Cryst A59: 228-234 (2003).

In the majority of cases, there is likely to be the improvement of the quality of the scaling results (and no harm!) in including at least **Low** absorption corrections. The only serious drawback is in the speed of the scaling process, as *Scalepack* requires additional processing time to calculate the spherical harmonic expansions.

**Use Auto Corrections** can help determine difficult structures, but the resulting reduced data should not be used for refinement or deposition. This option automatically adjusts the error model, which can be especially useful with weak anomalous signals or anisotropic diffraction. This option reliably models correlated errors and is generally better than manually adjusting the '**Error Scale Factor**' and '**ADJUST ERROR MODEL**' options. It is a good practice to add "auto" to the results filename to remind yourself not to refine against this data.

**Use Rejections On Next Run** means that the reflections that were written to the file `reject` are not used in subsequent scaling runs. Unless this is explicitly checked, reflections will only be flagged for rejection, not actually rejected. Generally, one uses this on the second and all subsequent runs of scaling. It does not hurt to have this option selected if there is no `reject` file (i.e., your first run or after deleting the file).

**Write Rejection File** tells the program to create a list of reflections, stored in the file called `reject` that meet the criteria for rejection. They are applied, i.e., rejected, when the **USE REJECTIONS ON NEXT RUN** button is checked. If you make changes to the error model or make a change the space group that would merge different reflections, you will need to remove the reject file using the **DELETE REJECT FILE** in the control box at the bottom of the page.

**Fix B** tells the program not to fit B factors at all. Usually, it is combined with the input of the B factors you want to apply but do not wish to refine anymore, or it is used for frozen crystals where you do not expect significant decay. This is in contrast to the default procedure, where the B factors are fit only after the convergence of the scaling. In the default procedure, if scaling does not converge in 20 (default) cycles of refinement, B factors will not be fitted.

**Anomalous.** This flag keeps Bijovets (I+ and I-) separate in the output file. If the **Anomalous** flag is on, anomalous pairs are considered equivalent when calculating scale and B factors and when computing statistics but are merged separately and output as I+ and I- for

## SCALING AND SCALING OPTIONS

each reflection. See the next section for a discussion of how to treat data that contains an anomalous signal.

**Scale Anomalous.** This is the flag for keeping Bijvoets (I+ and I-) separate both in scaling and in the output file. If the **Scale Anomalous** flag is on, anomalous pairs are considered non-equivalent when calculating scale and B factors and when computing statistics, are merged separately, and output as I+ and I- for each reflection.

This is a dangerous option because scaling may be unstable due to the reduced number of intersections between images. The danger is much greater in low symmetry space groups. **Scale Anomalous** will always reduce  $R_{\text{merge}}$ , even in the absence of an anomalous signal, because of the reduced redundancy. However,  $\chi^2$  values will not be affected in the absence of an anomalous signal.

**Include Overloads.** “Overloads” are reflections that contain pixels that exceed the maximum measurable intensity and will have an intensity is set at the maximum intensity possible for the detector. Since the intensity of these pixels cannot be determined accurately, the default behavior is to ignore reflections that contain overloads. In some cases, you may want to retain the reflections that if only a few pixels exceed the maximum. If this option is selected, fitted profiles with missing pixels (typically due to overload) will be included in the scaling.

**Direction Cosines** produce information that can be read by an outside absorption correction program, such as *Shelx*. Generally, this option is not needed as *HKL-2000* is capable of applying absorption corrections intrinsically (see above).

**Fit Goniostat** gives the goniostat misalignment angles. It should only be used by beamline staff or the maintainers of X-ray diffractometers.

**No Merge Original Index** will disable the default behavior of merging reflections; thus, reflections with the same unique hkl will not be combined. Furthermore, the output will also contain the original (not unique) hkl for each reflection. This option can be useful if you would like to examine the indexed, but unmerged data, in an external program. This may be useful when determining the space group is difficult. The output will consist of the original hkl, unique hkl, batch number, a flag (0 = centric, 1 = I+, 2 = I-), another flag (0 = hkl reflecting above the spindle, 1 = hkl reflecting below the spindle), the asymmetric unit of the reflection, I (scaled, Lorentz and Polarization corrected), and the  $\sigma$  of I.

**Number of Zones.** This sets the number of resolution shells the data is divided into for the basis of calculating the statistics summarized on the plots on the **Scale** panel. This input is required and will be equivalent to the number of error zones in the dialog launched by the **ADJUST ERROR MODEL** button. If you use the **Macros** panel to set the error model manually using the **ESTIMATED ERROR** keyword, you must specify an equivalent number of shells. The default value is 20, but for higher resolution data, a larger value allows to you detect small crystals of ice or salt not always easily visible on the diffraction image.

**Are fewer zones better?**

No. Combining those resolution shells where Rmerge is rapidly changing into a single zone yields misleading statistics. In this case, the shell will be dominated by the strong data at the low resolution end of the zone and give the impression that the high resolution limit of the zone has better statistics than it really does. For example, if you combined all your data into a single zone, the Rmerge in the final shell would be pretty good (i.e. equal to the overall Rmerge), when in fact it was substantially worse. It is more sensible to divide your zones into equal volumes and have enough of them so that you can accurately monitor the wholly expected decrease in the quality of the data as resolution increases.

**Error Scale Factor.** This is a single multiplicative factor which is applied to the input  $\sigma_I$ . This should be adjusted so the normal  $\chi^2$  (goodness of fit) value that is printed in the final table of the output comes close to 1. By default, the input errors are used (**Error Scale Factor** = 1.3). Reasonable values of the error scale factor are between 1 and 2. If you have to resort to higher values (i.e., >2) to get your overall  $\chi^2$  values to 1, then you should be suspicious of some problem with your crystal, indexing, space group/lattice assignment, or detector or goniostat. That doesn't mean that your data is useless, but it does mean that there are problems with it.

Increasing the Error Scale Factor can also decrease the number of reflections that will be rejected. Ideally, you would not reject any reflections; however, in practice even excellent data sets have a few outliers that need to be rejected. Generally, if fewer than 0.5% of the reflections are rejected, this is considered normal. Rejections significantly greater than 1% usually indicates some problem with the indexing, integration, space group assignment, or the data itself (e.g., spindle or shutter problems).

**Default Scale.** This is the overall scale factor used in the absence of an initial scale factor. This is useful if the data are too intense, which is sometimes the case with small molecules. It will reduce the output intensities by the factor entered.

**Resolution.** This is the maximum and minimum resolution (or **d-spacing**) of the reflections used in this run. The defaults are the full resolution range found in the input integration data. Note that if you integrated your raw data correctly, the maximum resolution should slightly exceed the resolution at which your signal drops below  $2\sigma$ . Therefore, in subsequent rounds of scaling it is usually important to enter a new value for the maximum resolution over which scaling is performed. Here's an example: You collected data and integrated it over the resolution range of 50 to 2.2 Å. The first round of scaling was performed over this resolution range, and you noted that the average  $I/\sigma$  dropped below 2 at 2.3 Å resolution. Therefore, you should do all subsequent scaling with a minimum resolution of 50 Å and a maximum resolution of 2.3 Å.

Don't change the resolution until after the first round of scaling with a properly adjusted error scale factor and rejected outliers. It is only at this point that you can determine where the signal drops below  $2\sigma$ . This is your high-resolution limit. Once this has been determined, you should delete the reject file using the **DELETE REJECT FILE** button in the Controls panel at the bottom. Then start scaling over again, this time with the proper resolution limits.

## Global Refinement Options

The values of the unit cell parameters obtained from a single image may be quite imprecise due to the correlations between crystal and detector parameters. This lack of precision is of little significance during integration as long as the predicted positions are on target. However, at the end of the data reduction process, it is important to use all the available information to get precise unit cell values. This is done in a procedure referred to as **Global Refinement** or **Postrefinement**. The implementation of this method in *Scalepack* allows for separate refinement of the orientation of each image, but with the same unit cell value for the whole data set. In each batch of data (a batch is typically one image) different unit cell parameters may be poorly determined. However, in a typical data set, there are enough orientations to determine precisely all unit cell lengths and angles.

**Global Refinement** is also more precise than the processing of a single image in the determination of crystal mosaicity and the orientation of each image. This is particularly important when deciding which reflections are fully recorded and which are partially recorded, which is why you should routinely do the **Global Refinement**.

Within the global refinement, the following parameters can be refined: the unit cell lengths and angles, the crystal orientation parameters (**Crystal Rotation X**, **Crystal Rotation Y**, and **Crystal Rotation Z**), and the mosaicity. These can either be refined over the "Crystal" (generally this means over the whole data set) or over the "Batch" (generally this means each image). You can see these listed if you click on the 4<sup>th</sup> choice under the **Global Refinement** menu: **Custom Postrefinement**.

The defaults are set so that the unit cell parameters and the mosaicity are refined to be the same over the entire data set. This makes sense for most data collections these days when the crystal is frozen and the entire data set is collected from a single crystal. However, if something happened to the crystal in the middle of the data collection (e.g., a temperature change that caused the unit cell constants to change, or the crystal began to decay severely), then one would want to refine these for each batch rather than over the entire data set.

Similarly, the defaults for the crystal orientation parameters, in this case, are refined for each batch, or frame, rather than over the entire data set. This allows you to spot any problems with the goniostat. If you do refine this for each batch and it doesn't change much, then on subsequent rounds of refinement you can refine the orientation once for the entire data set (crystal).

In general, it is best to assume that the crystal is slipping and that the goniostat is imperfect (**Small Slippage Imperfect Goniostat**) when choosing **Global Refinement** options, and only restrict the refinement with the other options if the data warrants it.

## SCALING AND SCALING OPTIONS



**Figure 87.** The global refinement options

**Number of cycles.** The default for the **Number of Global Refinement Cycles** is 10. You can change this by changing the number in the box under **Global Refinement** (Figure 87).

**Scalepack Type.** For very large or small data sets, you may need to specifically set which version of Scalepack should be used. For example, the virus version of Scalepack has memory allocations that are compatible with images that have a very high number of reflections, as would be expected for viruses of other crystals that have extremely large unit cells. In most cases, the default “Auto” option will be sufficient.

## The three *Scalepack* output Files

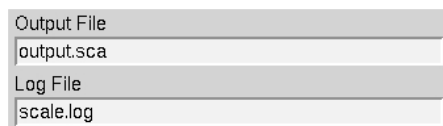
The three files and their default names are `output.sca`, `scale.out`, and `scale.log` (Figure 88).

`output.sca` is the file that contains your scaled reflections. The header lists the unit cell and space group, and the five columns correspond to  $h$ ,  $k$  and  $l$ ,  $I$ , and  $\sigma(I)$ . You may rename this file as you wish, although the convention is to retain the `*.sca` suffix to show that Scalepack created it. Many other programs that read *Scalepack* files expect this suffix. If you are trying to choose between space groups, it is wise to include the space group in file names.

`scale.log` is a log of the scaling operations. It is often quite long once the list of rejected reflections is included. You can view this file within HKL-2000 using the **SHOW LOGFILE** button.

`scale.out` (not visible in an **Output File** window) is a file that contains the data for the plots seen in the user interface. As this file is only used internally by *HKL-2000*, you shouldn't rename it.

The files are all written to the output data directory you specified in the **Main** page. However, if you fill in the path explicitly in the box provided, you can write the files anywhere you'd like.



**Figure 88.** The *Scalepack* output files

Use the **SCALE SETS** button in the Scaling Controls Box to scale the data. Expect to click this button numerous times, since scaling is an iterative process.

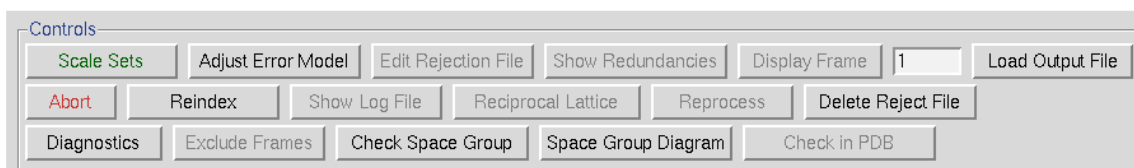


Figure 89. The Scaling Control Box will have some options disabled before scaling

## Output Statistics and Graphs

After each round of scaling, several statistics and graphs will be displayed in the top half of the window. This section contains 1) Postrefinement statistics, 2) overall statistics, and 3) ten diagnostic graphs.

The top of the screen will have global refinement statics, including the results from post-refinement, including the unit cell parameters and information about the number of reflections (Figure 90). If you have selected the **Use Auto Corrections** option, more statistics will be available on the left side of the display (Figure 91). If you have selected the **No Postrefinement** option, only information about the reflections will be available (Figure 92).

Crystal - Global Refinement					
Space Group	P21212				
Resolution	50.00 - 1.99				
a	30.78	± 0.01	α	90	±
b	60.72	± 0.01	β	90	±
c	75.44	± 0.01	γ	90	±
Cell Volume	140992.9				
Mosaicity Range	0.64 - 0.74				
Global Statistics					
Error Scale Factor				Total Number of Reflections	68920
Error Systematic				Number of Reflections Marked for Rejection	65
Radiation Damage				Percentage of Reflections Marked for Rejection	0.09
Anomalous Signal				Percentage of Reflections Rejected	0
Internal Non-isomorphism				Total Rmerge / Rmeas / Rpim	0.034 / 0.040 / 0.021

**Figure 90. Global Refinement with post-refinement information**

## SCALING OUTPUT STATISTICS AND GRAPHS

Crystal - Global Refinement					
Space Group		P21212			
Resolution		50.00 - 1.99			
a	30.78	± 0.01	α	90	±
b	60.72	± 0.01	β	90	±
c	75.44	± 0.01	γ	90	±
Cell Volume		140989.5			
Mosaicity Range		0.64 - 0.75			

Global Statistics			
Error Scale Factor	1.619	Total Number of Reflections	68932
Error Systematic	0.034	Number of Reflections Marked for Rejection	33
Radiation Damage	0.016	Percentage of Reflections Marked for Rejection	0.05
Anomalous Signal	0.038	Percentage of Reflections Rejected	0
Internal Non-isomorphism	0.005	Total Rmerge / Rmeas / Rpin	0.033 / 0.036 / 0.014

**Figure 91. Global Refinement with post-refinement information and auto-corrections**

Crystal - Global Refinement					
NO POSTREFINEMENT					
Global Statistics					
Error Scale Factor	Total Number of Reflections	68848			
Error Systematic	Number of Reflections Marked for Rejection	1250			
Radiation Damage	Percentage of Reflections Marked for Rejection	1.82			
Anomalous Signal	Percentage of Reflections Rejected	0.00			
Internal Non-isomorphism	Total Rmerge / Rmeas / Rrim	0.053 / 0.057 / 0.022			

**Figure 92. Global statistics with no post-refinement**

### Rejected vs. Marked for Rejection

The table that appears at the top of the page after each round of scaling has two categories: **Reflections Marked for Rejection**, and **Reflections Rejected**. The **Reflections Marked for Rejection** are the percentage of all reflections that have been written to the `reject` file. The **Reflections Rejected** represent the percentage of reflections, which have actually been omitted from the data in the latest round of scaling. Rejection of reflections is an iterative process that should converge after only a few rounds of rejection and refinement. For example, on the first round of scaling, 100 reflections may be flagged for rejection and written to the `reject` file. If **Use Rejections on Next Run** is checked, then the reflections in the `reject` file will be removed from the set of reflection data on the next round of scaling. Having done this, other reflections may now be flagged as outliers, and written to the end of the `reject` file. Perhaps there are 20 more this time. Repeating this procedure should converge to a point when no more reflections are being rejected (and thus the two percentages are identical).

Ten diagnostic graphs will appear below global refinement statistics. A scroll bar on the right allows you to view graphs that do not originally fit in the window. These graphs are:

- Scale and B vs. Frame
- Completeness vs. Resolution
- $\chi^2$  and Rmerge vs. Frame
- $\chi^2$  and Rmerge vs. Resolution
- $I/\sigma$  and  $CC_{1/2}$  vs. Resolution
- $I$  vs. Resolution
- $\chi^2$  and Rmerge vs. Average  $I$
- Average and Cumulative Redundancy vs. Resolution
- Mosaicity vs. Frame
- Low-Resolution Completeness vs. Resolution

Most of the graphs have an **EXPLANATION** button that describes the graph. For example, the explanation for the Completeness vs. Resolution describes what the different colored lines mean (different sigma levels). Graphs that are initially displayed as logarithmic graphs also have a **linear scale** button that will change the scale of the Y-axis. Note that some plots have two vertical columns; in most cases, the data displayed are color-coded to match its axis. For example, blue data are plotted with respect to the blue axis and red data to the red axis. Some plots also have a button to toggle between linear and log representations.

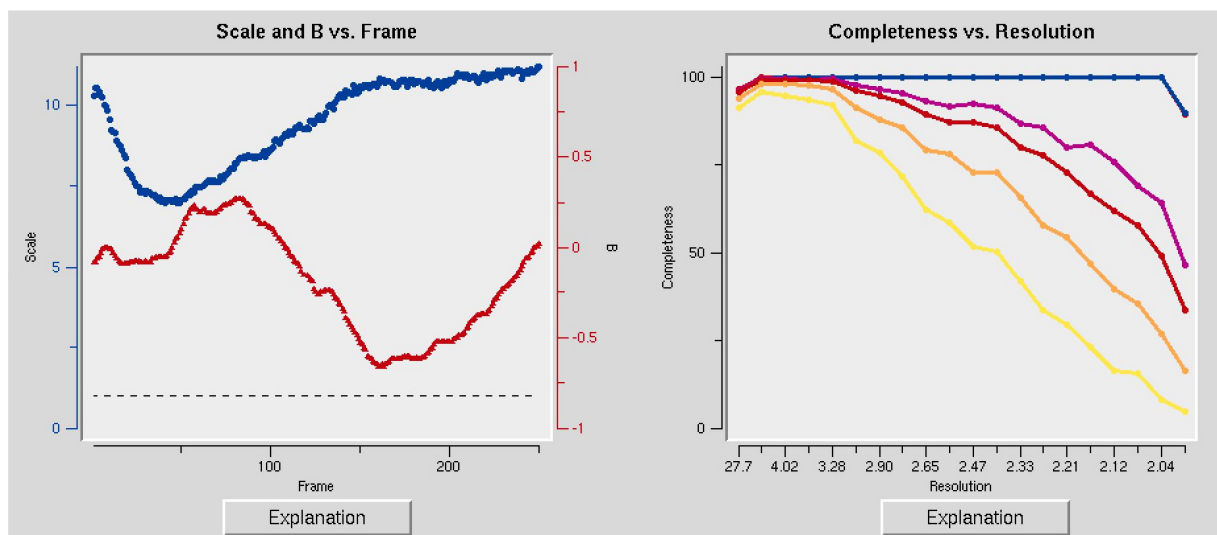
When the initial round of scaling is finished, examine the output plots and adjust the **Error Scale Factor** to bring the overall  $\chi^2$  down to around 1 (the overall  $\chi^2$  is also explicitly specified at the end of the log file). After making changes to the **Error Scale Factor**, you must delete the `reject` file by clicking on the **DELETE REJECT FILE** button in the lower **Controls**



## SCALING OUTPUT STATISTICS AND GRAPHS

panel, and then click **SCALE SETS** again. Repeat these steps until the overall  $\chi^2$  is acceptable then proceed to subsequent rounds of scaling by activating **USE REJECTIONS ON NEXT RUN** and scale again. Repeat the cycle of scaling and rejecting until the number of new rejections has declined to just a few. At this point, the scaling is completed.

The first two charts can help you determine if your data collection strategy was appropriate and if the overall data collection went well. The **Scale and B vs. Frame** (Figure 93, right) to see if there are significant variations in intensity and B-factor of frames across the data set, which could indicate the crystal rotating out of the beam. Variations can also result from the crystal morphology. For example, plates will have more intense reflections when the crystal's edge is along the beam than when the face of the plate is perpendicular to the beam. A dramatic rise of the B-factor (to around 10) that does not decrease is usually an indication of radiation damage. The **Completeness vs. Resolution** plot (Figure 93, left) describes the completeness of the data for different resolution shells. The blue line represents all reflections, the purple reflections with  $I/\sigma$  greater than 3, the red reflections with  $I/\sigma$  greater than 5, the orange reflections with  $I/\sigma$  greater than 10, and the yellow reflections with  $I/\sigma$  greater than 20. This chart can be very informative for deciding if you have collected sufficient data to solve your structure. It is not necessary to achieve 100% completeness, but datasets with low completeness can severely hinder or thwart the process. Low-resolution reflections are particularly important for molecular replacement solutions.



**Figure 93. The Scale and B vs. Frame and the Completeness vs. Resolution and  $I/\sigma$  and  $CC_{1/2}$  vs. Resolution graphs**

The  **$I/\sigma$  and  $CC_{1/2}$  vs. Resolution** shows the ratio of the intensity to the error of the intensity ( $I/\sigma$  – blue line) and the  $CC_{1/2}$ , the correlation coefficient between intensity estimates of half data sets (upper red line), as a function of resolution (Figure 94). The traditional approach of selecting the maximum resolution was to set it to the resolution where your data falls  $2\sigma$ , but recent analysis suggests that you can obtain useful information as long

## SCALING OUTPUT STATISTICS AND GRAPHS

as the  $CC_{1/2}$  is above 0.2. This graph will also include an  $I/\sigma$  plot for anomalous data if this data is present.

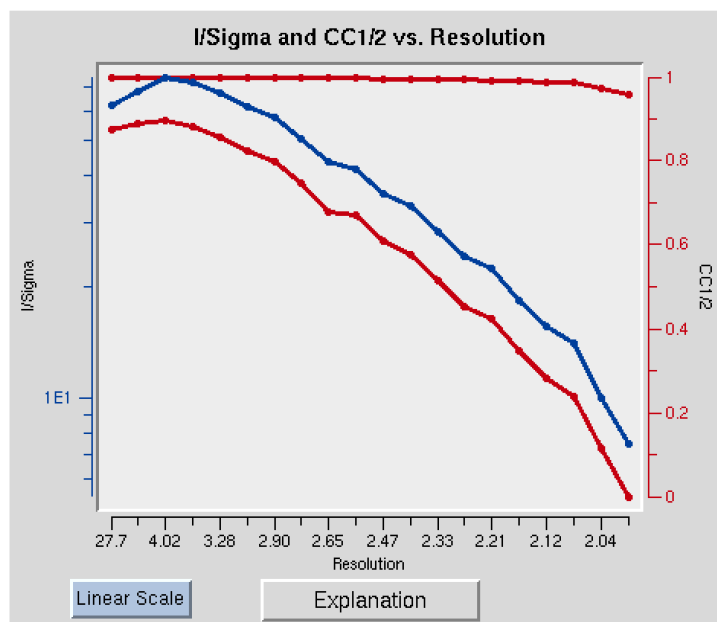


Figure 94.  $I/\sigma$  and  $CC_{1/2}$  vs. Resolution

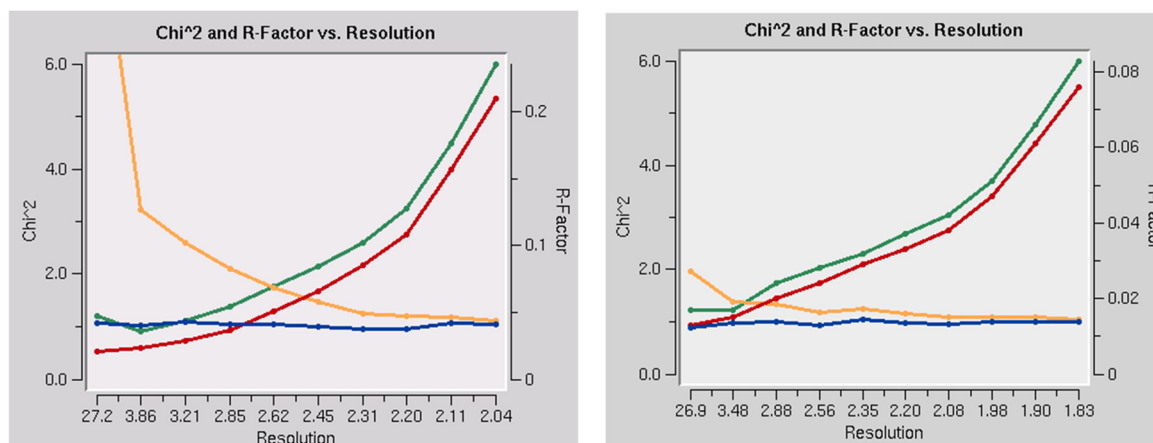
## Working with anomalous data

Detecting an anomalous signal is very easy. In the first step, the data are scaled and postrefined normally with the **ANOMALOUS** option selected (*not* the **SCALE ANOMALOUS** option!) This tells the program to output a \*.sca file where the I+ and I- reflections are separated. In this case, *Scalepack* will treat the I+ data and the I- data as two separate measurements within a data set and the statistics that result from merging the two will reflect the differences between the I+ reflections and the I- reflections. Obviously, for a centric reflection, there is no I-, so the merging statistics will only reflect the non-centric reflections.

The presence of an anomalous signal is detected by examining the  $\chi^2$  values on the **Chi^2 and R-Factor vs. Resolution** plot ([Figure 95](#)). The two  $\chi^2$  curves show the  $\chi^2$  values for each resolution shell, calculated for merged and unmerged Friedel pairs, in orange and blue respectively. Assuming that the errors scale factors were reasonable, and there is no useful anomalous signal in your data, both curves showing the  $\chi^2$  resolution dependence should be flat and average around 1 when scaling with either merged or unmerged Friedel pairs. On the other hand, if  $\chi^2 > 1$  and you see a clear resolution dependence of the  $\chi^2$  for scaling with merged Friedel pairs (i.e., the  $\chi^2$  increases as the resolution decreases), there is a strong indication of the presence of an anomalous signal. The resolution dependence allows you to determine where to cut off your resolution to calculate an anomalous difference Patterson map with an optimal signal to noise. Please note, however, that this analysis assumes that the error model is reasonable and gives you a  $\chi^2$  close to 1 when the anomalous signal is not present. You can

## SCALING OUTPUT STATISTICS AND GRAPHS

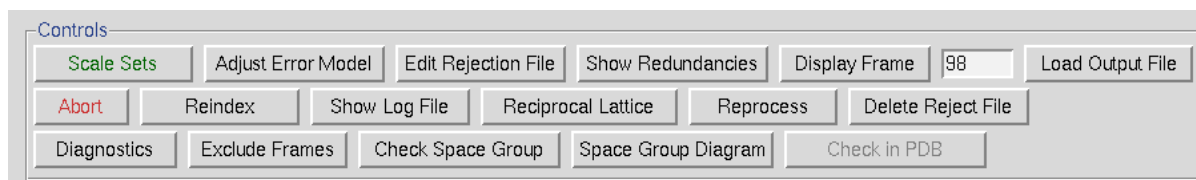
only use **SCALE ANOMALOUS** when you have enough redundancy to treat the F+ and F- completely independently



**Figure 95.** Anomalous signal detection for a strong (left) and moderate (right) anomalous signal.

## The Scaling Controls Box

The control box in the scaling panel (Figure 96) contains the **SCALE SETS** button, and several other options that can only be performed after at least one round of scaling has been performed.



**Figure 96.** The Scaling Control panel

**SCALE SETS** is the button that starts the scaling process using the parameters selected in the *Scaling Options* and *Global Refinement* boxes.

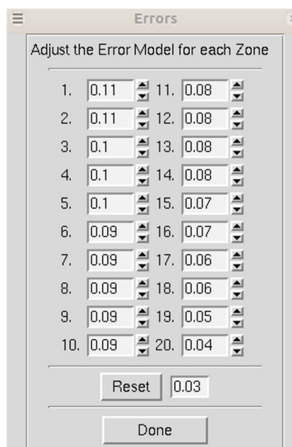
**ADJUST ERROR MODEL** allow you to set the different error levels for each resolution shell, giving you more control over the error model than the overall effect of changing the *Error Scale Factor*.

## THE SCALING CONTROLS BOX

The error model is the estimate of the systematic error for each of the resolution shells. There will be exactly the same number of error estimates here as there are **NUMBER OF ZONES**.

The error estimates do not all have to be the same. The estimated error applies to the data which are read **after** this keyword, so you can apply different error scale factors to subsequent batches by repeating this input with different values. This is an important point if you enter data from a previous *Scalepack* output that does not need its  $\sigma$  to be increased.

The error estimates should be approximately equal to the R-factors for the resolution shells **where statistical errors are small**, namely the lower resolution shells where the data is strong. This is a crude estimate of the systematic error (to be multiplied by  $I$ ) and is usually invariant with resolution. The default is 0.03 (i.e. 3%) for all zones. Examine the difference between the total error and the statistical error in the final table of statistics (this can be viewed by clicking the **SHOW LOGFILE** button and scrolling to the bottom). The difference between these numbers tells you what contribution statistical error makes to the total error ( $\sigma$ ). If the difference is small, then reasonable changes in the estimated error values following the will not help your  $\chi^2$  much. This is because the estimated errors represent your guess and/or knowledge of the contribution of the systematic error to the total error, and a small difference indicates that systematic error is not contributing much. If the difference between total and statistical error is significant, and the  $\chi$ 's are far from 1, then consider adjusting the estimated error values in the affected resolution shells.

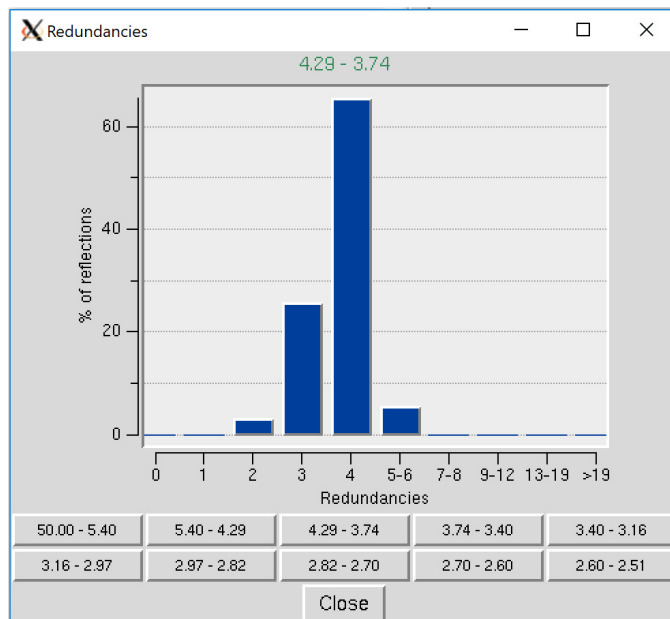


**Figure 97. Adjusting the error model**

**EDIT REJECTION FILE** will open the rejects file in an external editor, allowing you to remove particular reflections from the list of rejected reflections. This is rarely necessary and not advised.

**SHOW REDUNDANCIES** will display a window that presents a histogram showing the percent of reflections with a specific number of redundancies. A redundancy value is the number of times a unique reflection has been measured after symmetry-related reflections have been merged. Buttons for different resolution ranges are below the histogram, allowing you to view your redundancies in different bins of data (Figure 98).

## THE SCALING CONTROLS BOX



**Figure 98. Redundancies Window**

The **DISPLAY FRAME** button works in conjunction with the number input box next to it. It will open the frame specified in the XDisp program.

The **LOAD OUTPUT FILE** will open a file selection dialog and allow you to see the output from previous scaling rounds.

The **ABORT** button will stop the current round of scaling. This may come in handy if you realize you have set a parameter wrong.

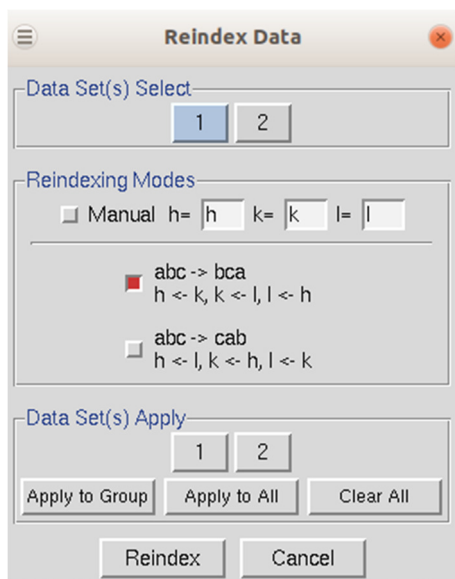
**REINDEX** allows you to change the reference frame for the Miller indices, effectively changing the orientation of the lattice. For example, if the data has been indexed such that a screw symmetry is on the wrong axis, you can apply a permutation to get the symmetry element on the correct axis. The **Reindexing dialog** lets you transpose one unit cell axis to another axis by reassigning the Miller indices. For example, if you have indexed your data in the primitive orthorhombic crystal class, the analysis of systematic absences may indicate that you have one screw axis. By convention, this screw axis should be on the c axis. The original indexing might have this screw axis along a different axis; therefore, you will need to reindex the data to make it conform to space group conditions.

Reindexing may also be necessary when comparing two or more data sets that were collected and processed independently. *Denzo*, when confronted with a choice of more than one possible indexing convention, makes a random choice. This is not a problem unless it makes a different choice for a second data set because the two will not be compatible for scaling together without first reindexing. One cannot distinguish non-equivalent alternatives without scaling the data, which is why this is not done in *Denzo* (i.e., at the indexing or integration step). You can tell if you need to reindex one of two data set if the  $\chi^2$  values upon

## THE SCALING CONTROLS BOX

merging the two are very high (e.g., 50). This makes sense when you consider that scaling two or more data sets together involves comparing reflections with the same *hkl* or index.

To reindex the data, use the **REINDEX** button. You will be presented with different reindexing options and a manual assignment option (Figure 99). If you are working with more than one data set, you can select which datasets will be reindexed, or you can reindex all of the data sets. For example, if you have a screw axis on the b & c axes, but they should be on the a & b axes, you should select the option “abc → bca: h←k, k ←l, l ←h”. After choosing reindexing options, click **REINDEX**. This example is consistent with the POINTLESS results in Figure 104.



**Figure 99.** The Reindexing dialog

**No reindexing, no new auto-indexing, and nothing except changing the sign of y scale in Denzo can change the sign of the anomalous signal.**

There are two ways to manually check whether or not you have systematic absences. The first is to look at the *Intensities of systematic absences table* at the bottom of the log file, which can be shown using the **SHOW LOG FILE** button. If you have scaled your data in a space group that is expected to have systematic absences, then the reflections that should be missing will be listed at the bottom of the log file. The table lists these reflections’ intensity, sigma, and I/sigma. The other way to evaluate the systematic absences is to view particular slices of reciprocal space using the **RECIPROCAL LATTICE** button.

**SHOW LOG FILE** will open the log file that resulted from the last round of scaling. Much of the data in this file is displayed in the graphs as described below. Most of the important information is in the tables at the end of the log. The *Scalepack* log file can be very long since it contains the list of all \*.x files read in and a record of every cycle of scaling. The log file is divided into several sections. In order these are:

1. the list of \*.x files read in, and the list of reflections from each of these files that will be rejected;

## THE SCALING CONTROLS BOX

2. the output file name;
3. the total number of raw reflections used before combining partial reflections;
4. the initial scale and B factors for each frame, goniostat parameters, and the space group;
5. the initial scaling procedure;
6. analysis and results tables;
7. systematic absences (if expected)
8. automatic corrections statistics (if used)

The log file should be examined after each iteration. In particular, the errors,  $\chi^2$  and both  $R$  factors should be checked

**RECIPROCAL LATTICE** will open a display one of three reciprocal lattices. There are two good reasons you may want to view the reciprocal lattice: 1) If you would like to verify the space group, especially the presence of a screw axis or centering and 2) if you want to visualize the completeness of the data. Clicking the on the **RECIPROCAL LATTICE** button will reveal the choice of three different slices of the reciprocal lattice to view:  $h,0,l$ ;  $h,k,0$ ;  $0,k,l$ . Looking at these slices can make it apparent if you have systematic absences or, conversely, if you are missing the data necessary to determine if you have systematic absences. The  $h,0,l$  slice of the  $P2_12_12_1$  data in Figure 100 clearly shows a screw axis along  $h$  and  $l$  because the odd reflections are systematically absent.

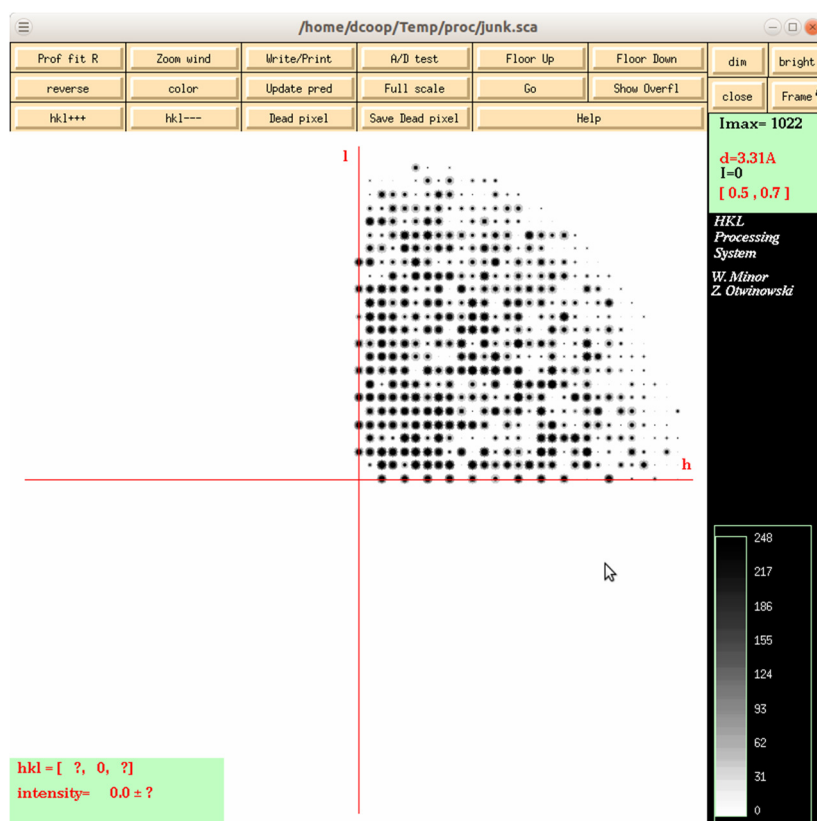
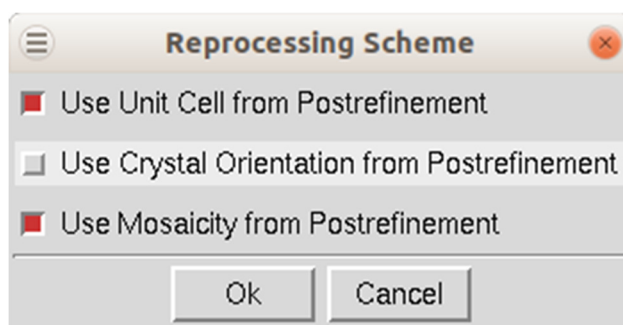


Figure 100. The reciprocal lattice viewer



**REPROCESS** lets you take a step back and reindex all of the images using values calculated during postrefinement. You can use 1) the unit cell from prostrefinement, 2) the crystal orientation from postrefinement, and/or 3) the mosaicity from postrefinement (Figure 101). This dialog box allows you to select more than one of these parameters. This could be beneficial if there were large changes in these values as the data was initially processed. This option will overwrite your original \*.x files.



**Figure 101. Reprocessing options**

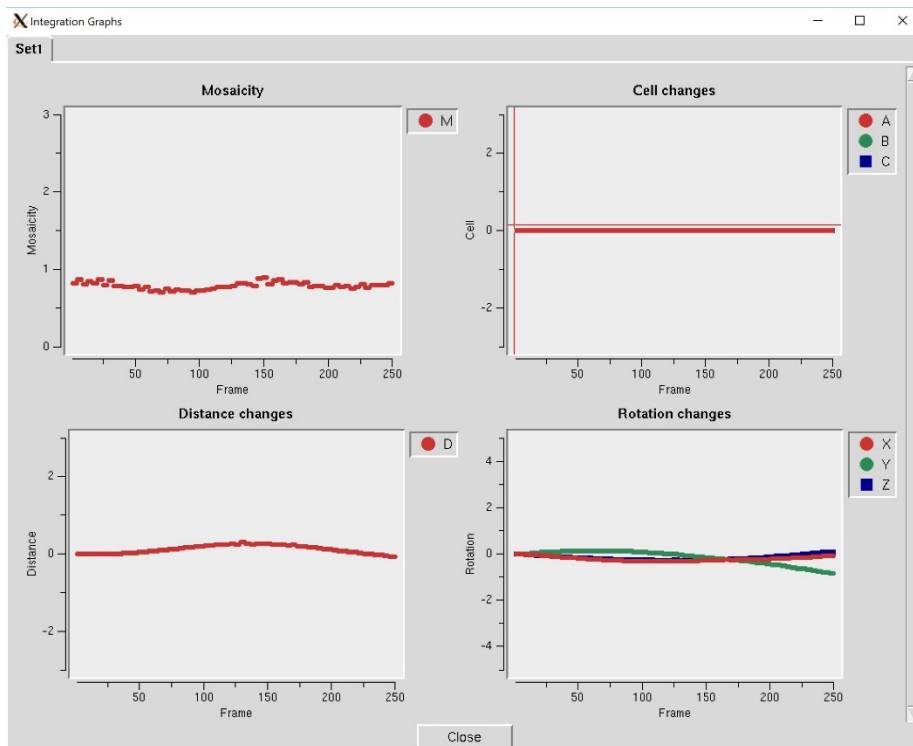
**DELETE REJECT FILE** removes the file that contains the reflections that are excluded from scaling if you have the Use *rejections on next run* option selected. It is important to delete the reject file if you change any of the options that affect the error model.

**DIAGNOSTICS** will present you with a selector panel that will allow you to run three types of diagnostics: 1) *Check Integration*, 2) *Check Spindle Movement*, and *Check Shutter Timing* 3). The *Check Integration* option will display four of the graphs generated while the data was being integrated (Figure 102). If you have more than one data set, each set will have a tab. The resulting graphs should be relatively flat. If they are not, you may benefit from the **REPROCESS** option described above.

The *Check Spindle Movement* calculates the Rmerge of your data using a macro that corrects for uneven rotations of the spindle axis. This is not a common problem but does happen. If no problems are found, you should see the message “No problems with uneven spindle movement were detected.” Otherwise, you will get the message, “The decrease in Rmerge indicates uneven spindle movement.” You should report this type of problem to the site administrator where you collected your data.

The *Check Shutter Timing* calculates the Rmerge of your data using a macro that corrects for uneven shutter timings. This is not a common problem but does happen. If no problems are detected, you should see the message “No problems with synchronization between shutter and spindle axis rotation were found,” otherwise you will get the message “The decrease in Rmerge indicates poor synchronization between the shutter and the spindle axis.” You should report this type of problem to the site administrator where you collected your data.



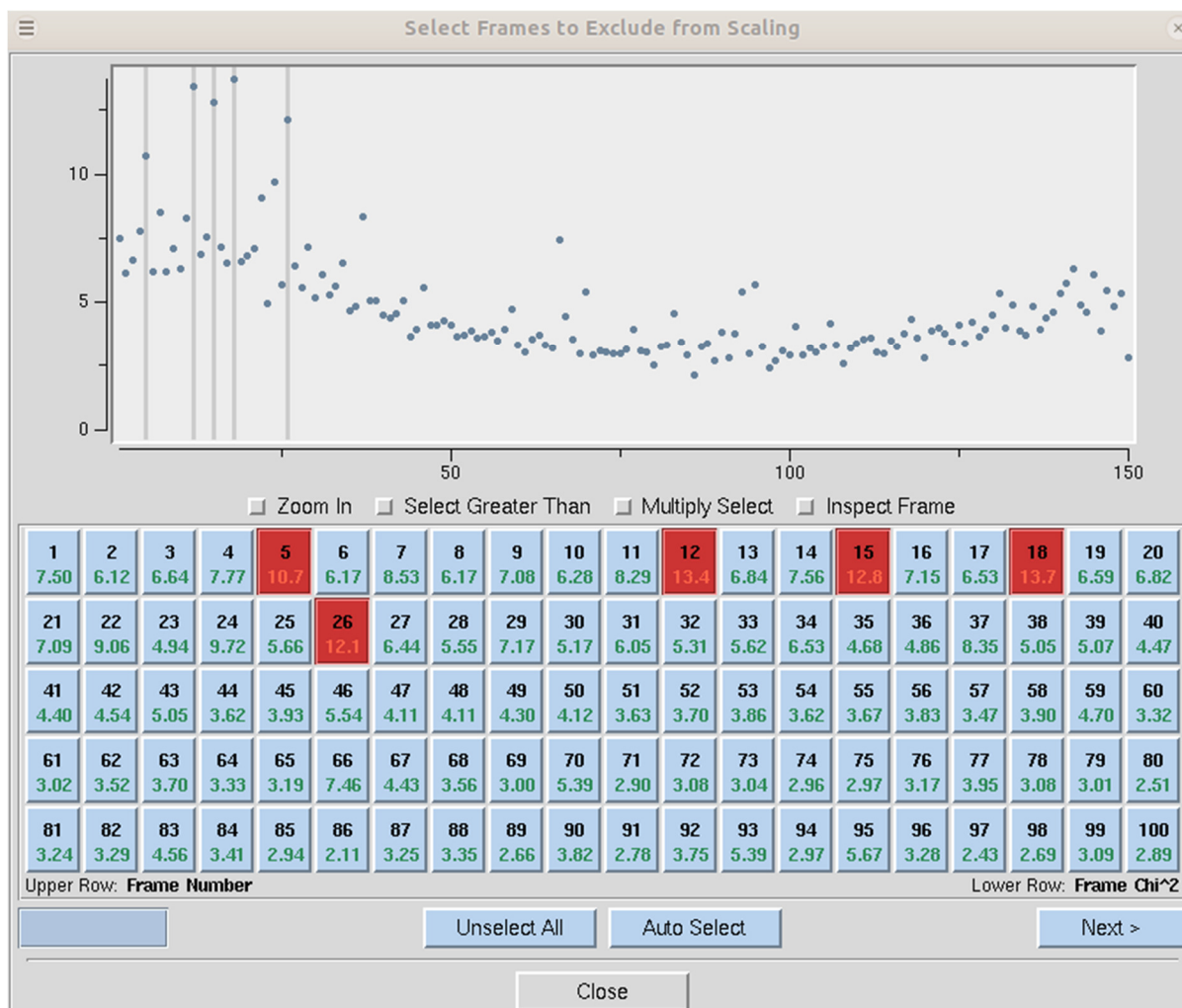


**Figure 102. The Integration Graphs from the Inspect Integration option**

**EXCLUDE FRAMES** provides you with the flexibility to omit particular images from the scaling process. After integration, some frames may appear to have strong outliers in the integration information plots, like unacceptably high  $\chi^2$  values – i.e., far above the mean value. You can exclude those frames from scaling by clicking on the **EXCLUDE FRAMES** button. A table with the list of frames will appear (Figure 103). Select a single frame by clicking on its box or on a point on the plot. Select a range of frames by checking **Multiple Select** and clicking twice, once on the first frame and once on the last. You may have to move to a different page with the **PREVIOUS** or **NEXT** buttons to see all frames. You may exclude frames above a certain limit by selecting the **Select Greater Than** option. This will turn on crosshairs controlled by the mouse. When you click, all frames with a  $\chi^2$  higher than the horizontal crosshair. Frames marked in red will be excluded when you close the dialog and rescale. There is also an Auto Select button that will automatically select images with a  $\chi^2$  higher than 10.

If you would like to see an individual frame, select the **Inspect Object** button, and then click on an image tile.

## THE SCALING CONTROLS BOX



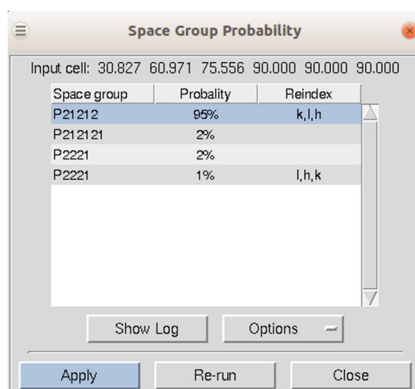
**Figure 103.** Frames selected to be excluded from scaling are marked red

**CHECK SPACE GROUP** and **SPACE GROUP DIAGRAM** are described in detail in the next section.

**CHECK IN PDB** will check the PDB to see if any deposited structures have a similar unit cell. It is unlikely, although not impossible, that two different proteins will crystallize with the same space group. This option was added to help alert you to the possibility that you have crystallized one of the proteins that have been known to co-purify with recombinant proteins, such as maltose-binding protein or trypsin.

## Space Group Identification

If your installation of HKL has been linked to CCP4, then the first approach to space group determination is to use the **CHECK SPACE GROUP** button. This button runs the CCP4 program POINTLESS and will present the results in an easy to read table. To use this button, first scale the data in the primary space group for the crystal lattice (the default space group selected after integration). The dialog will not only provide the probability for each possible space group but will also indicate if re-indexing the data is necessary to ensure compatibility with space group conventions (Figure 104). The logfile from POINTLESS can be viewed. If you select the **APPLY** button, the space group you select will be applied and put in the Space Group selection box. This will not, however, reindex your data, even if reindexing is suggested. To do this, see the section on reindexing box on page 106.



Space group	Probability	Reindex
P21212	95%	k,l,h
P212121	2%	
P2221	2%	
P2221	1%	l,h,k

**Figure 104.** The Space Group Probability results from POINTLESS.

If your installation is not linked to CCP4, or if you would like to verify the suggested space group, *Scalepack* can be used to determine the space group of your crystal. What follows is a description of how you would continue from the lattice type given by *Denzo* to determine your space group. This whole analysis only applies to enantiomorphic compounds such as proteins. It does not apply to small molecules, necessarily, which may crystallize in centrosymmetric space groups. If you expect a centrosymmetric space group, you should use any space group which is a subgroup of the Laue class to which your crystal belongs. You also need enough data for this analysis to work so that you can see systematic absences.

To determine your space group using just HKL, follow these steps:

1. Determine the lattice type in *Denzo*.
2. Scale by the primary space group in *Scalepack*. The primary space groups are the first space groups in each Bravais lattice type in the table that follows this discussion. In the absence of lattice pseudosymmetries (e.g., monoclinic with  $\beta \approx 90^\circ$ ) the primary space group will not incorrectly relate symmetry related reflections. Note the  $\chi^2$  statistics. Now try a higher symmetry space group (next down the list) and repeat the scaling,

## SPACE GROUP IDENTIFICATION

keeping everything else the same. If the  $\chi^2$  is about the same, then you know that this is OK, and you can continue. If the  $\chi^2$  are much worse, then you know that this is the wrong space group, and the previous choice was your space group. The exception is primitive hexagonal, where you should try  $P6_1$  after if  $P3_121$  and  $P3_112$  fail.

3. Examine the bottom of the log file (the **SHOW LOGFILE** button) or the simulated reciprocal lattice picture (the **RECIPROCAL LATTICE** button) for systematic absences. If this was the correct space group, all of the systematic absence reflections should be absent (or with very small values). Compare this list with the listing of reflection conditions by each of the candidate space groups. The set of absences seen in your data that corresponds to the absences characteristic of the listed space groups identifies your space group or pair of space groups. Note that you cannot do any better than this (i.e., get the handedness of screw axes) without phase information.
4. If it turns out that your space group is orthorhombic and contains one or two screw axes, you may need to reindex to align the screw axes with the standard definition. If you have one screw axis, your space group is  $P222_1$ , with the screw axis along *c*. If you have two screw axes, then your space group is  $P2_12_12$ , with the screw axes along *a* and *b*. If the *Denzo* indexing is not the same as these, then you should reindex using the **REINDEX** button.
5. So far, this is the way to index according to the conventions of the International Tables. If you prefer to use a private convention, you may have to work out your own transformations. One such transformation has been provided in the case of space groups  $P2$  and  $P2_1$ .

To manually select a space group, use the drop-down box. A selection list will appear with the 14 Bravais lattices. When you click on a lattice type, it will expand to show the available space groups within that lattice type. Click on a space group to select it.

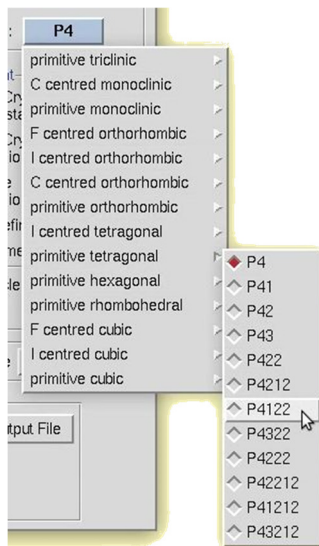


Figure 105. Manual space group selection

# SPACE GROUP IDENTIFICATION

Bravais Lattice	Primary assigned Space Groups	Candidates	Reflection Conditions along screw axes
Primitive Cubic	$P2_13$	195 $P23$	
		198 $P2_13$	$(2n,0,0)$
	$P4_132$	207 $P432$	
		208 $P4_232$	$(2n,0,0)$
		212 $P4_332$	$(4n,0,0)^*$
		213 $P4_132$	$(4n,0,0)^*$
I Centered Cubic	$I213$	197 $I23$	*
		199 $I213$	*
	$I4132$	211 $I432$	
		214 $I4132$	$(4n,0,0)$
F Centered Cubic	$F23$	196 $F23$	
	$F4_132$	209 $F432$	
		210 $F4_132$	$(2n,0,0)$
Primitive Rhombohedral	$R3$	146 $R3$	
	$R32$	155 $R32$	
Primitive Hexagonal	$P3_1$	143 $P3$	
		144 $P3_1$	$(0,0,3n)^*$
		145 $P3_2$	$(0,0,3n)^*$
	$P3_112$	149 $P312$	
		151 $P3_112$	$(0,0,3n)^*$
		153 $P3_212$	$(0,0,3n)^*$
	$P3_121$	150 $P321$	
		152 $P3_121$	$(0,0,3n)^*$
		154 $P3_221$	$(0,0,3n)^*$
	$P6_1$	168 $P6$	
		169 $P6_1$	$(0,0,6n)^*$
		170 $P6_5$	$(0,0,6n)^*$
		171 $P6_2$	$(0,0,3n)^{**}$
		172 $P6_4$	$(0,0,3n)^{**}$
		173 $P6_3$	$(0,0,2n)$
	$P6_122$	177 $P622$	
		178 $P6_122$	$(0,0,6n)^*$
		179 $P6_522$	$(0,0,6n)^*$
		180 $P6_222$	$(0,0,3n)^{**}$
		181 $P6_422$	$(0,0,3n)^{**}$
		182 $P6_322$	$(0,0,2n)$
Primitive Tetragonal	$P4_1$	75 $P4$	
		76 $P4_1$	$(0,0,4n)^*$
		77 $P4_2$	$(0,0,2n)$

<b>SPACE GROUP IDENTIFICATION</b>			
	$P4_12_12$	78	$P4_3$ (0,0,4n)*
		89	$P422$
		90	$P4212$ (0,2n,0)
		91	$P4_122$ (0,0,4n)*
		95	$P4_322$ (0,0,4n)*
		93	$P4_222$ (0,0,2n)
		94	$P4_22_12$ (0,0,2n),(0,2n,0)
		92	$P4_12_12$ (0,0,4n),(0,2n,0)**
I Centered Tetragonal	$I4_1$	96	$P4_32_12$ (0,0,4n),(0,2n,0)**
		79	$I4$
		80	$I4_1$ (0,0,4n)
Primitive Orthorhombic	$P2_12_12_1$	97	$I422$
		98	$I4_122$ (0,0,4n)
		16	$P222$
		17	$P222_1$ (0,0,2n)
C Centered Orthorhombic	$C222_1$	18	$P2_12_12$ (2n,0,0),(0,2n,0)
		19	$P2_12_12_1$ (2n,0,0),(0,2n,0), (0,0,2n)
		20	$C222_1$ (0,0,2n)
I Centered Orthorhombic	$I2_12_12_1$	21	$C222$
		23	$I222$ *
F Centered Orthorhombic	$F222$	24	$I2_12_12_1$ *
Primitive Monoclinic	$P2_1$	22	$F222$
		3	$P2$
C Centered Monoclinic	$C2$	4	$P2_1$ (0,2n,0)
		5	$C2$
Primitive Triclinic	$P1$	1	$P1$

Note that for the pairs of similar candidate space groups followed by the \* (or \*\*) symbol, scaling and merging of diffraction intensities cannot resolve which member of the possible pair of space groups to which your crystal form belongs. This is inherent to crystallography.

## Scaling \*.x files without reintegrating diffraction images

If you have already integrated your data, you may want to rescale the data without starting from scratch, either during the same session or later on. If you have just finished processing and scaling the data, changing the output file name and changing the parameters is all that is necessary to re-scale the data in a different space group. If you are doing this to try to determine the space group, it's a good idea to put the space group in the output file (the .sca file) and the log file.

If you want to scale \*.x files that you have previously processed in a new session of HKL, select **SCALING ONLY** when the initial site selection dialog opens, then set the **Output Data Directory** in the **Data** panel to be the directory where your \*.x files are located using the directory tree. It is not necessary to set the **Raw Data Directory**. Next, click the **Scale Sets Only** button followed by **LOAD DATA SETS**. You will see a list of \*.x files in the dialog box.

## SPACE GROUP IDENTIFICATION

Select the set you want to scale, followed by **OK**. If you want to scale additional sets of \*.x files, repeat this process and add them to the list. The individual sets do not have to be in the same directory. Each set will have the file locations associated with it. Make sure the **SCALE** button is selected for each set you want to scale together. For example, if you want to scale two sets of \*.x files together (say, from two different crystals, from high and low resolution passes of data collection, or from a native and a derivative, etc.) make sure that **SCALE** is clicked for each one. If you don't want them to be scaled together, then make sure that **SCALE** is not clicked. You don't have to worry about the *Image Display* or the *Experiment Geometry* since this was done when you first generated the \*.x files.

## Statistics and Scalepack.

The quality of X-ray data is initially assessed by statistics. In **small-molecule** crystallography, there is almost always a large excess of strong data, so this allows the crystallographer to discard a substantial amount of suspect data and still accurately determine a structure. Compared to small molecules, however, **protein crystals** diffract poorly. Moreover, important phase information comes from weak differences, and we must be sure these differences do not arise from the noise caused by the limitations of the X-ray exposure and detection apparatus. As a result, we cannot simply throw away or statistically downweight marginal data without first making a sophisticated judgment about which data are good and which are bad.

To accurately describe the structure of a protein molecule, we often need higher resolution data than the crystal provides. That is life. One of the main judgments the crystallographer makes in assessing the quality of his data is thus the **resolution** to which his crystal diffracts. In making this judgment, we wish to use the statistical criteria which are most discriminatory and which are the least subjective. In practice, there are three ways of assessing the high-resolution limit of diffraction. The first is **the ratio of the intensity to the error of the intensity**, i.e.,  $I/\sigma$ . The second way, which is traditional, but inferior, is the agreement between symmetry-related reflections, i.e.,  $R_{\text{merge}}$ . The most recently adopted convention of selecting a data cutoff is to use data that has a  $CC_{1/2} > 0.2$ .

From a statistical point of view,  $I/\sigma$  is a *superior criterion*, for two reasons. First, it defines a resolution "limit" since by definition  $I/\sigma$  is the *signal to noise* of your measurements. In contrast, the  $R_{\text{merge}}$  is an unweighted statistic that does not take into account the measurement error.

Second, the  $\sigma$  assigned to each intensity derives its validity from the  $\chi^2$ 's, which represent the *weighted* ratio of the difference between the observed  $I$  and the average value  $\langle I \rangle$  squared, divided by the square of the error model and multiplied by a factor correcting for the correlation between  $I$  and  $\langle I \rangle$ . Since it depends on an explicit declaration of the expected error in the measurement, the user of the program is part of the Bayesian reasoning process behind the error estimation.

The essence of Bayesian reasoning in *Scalepack* is that you bring  $\chi^2$  (or technically speaking, the goodness-of-fit, which is related to the total  $\chi^2$  by a constant) close to 1.0 by manipulating the parameters of the error model.  $R_{\text{merge}}$ , on the other hand, is an *unweighted* statistic which is independent of the error model. It is sensitive to both intentional and unintentional manipulation of the data used to calculate it, and may not correlate with the quality of the data. An example of this is seen when collecting more and more data from the same crystal. As the redundancy goes up, the final averaged data quality definitely improves, yet the  $R_{\text{merge}}$  also goes up. As a result,  $R_{\text{merge}}$  is only really useful when comparing data which has been accumulated and treated the same. This will be discussed again later.

In short, ***I/σ*** is the preferred way of assessing the quality of diffraction data because it derives its validity from the  $\chi^2$  (likelihood) analysis. Unless all of the explicit and implicit assumptions which have been made in the calculation of an  $R_{\text{merge}}$  are known, this criterion is less meaningful. This is particularly true when searching for a “number,” which can be used by others to critically evaluate your work.

There are two modes of analysis of data using  $\chi^2$ s. The first mode keeps the  $\chi^2$  (or more precisely, the goodness-of-fit) constant and compares the error models. This means that you are adjusting your estimates of the errors associated with the measurement until the deviations within observations agree with the expectation based on the error model.

The second mode keeps the error model constant and compares  $\chi^2$ s. This mode is computationally much faster and is used in refinement procedures. Of the two modes, the first is more informative, because it forces you to consider changes in the error model. Which mode you use generally depends on what you are comparing. When assessing the general quality of your detector, the first mode is used. When comparing a *derivative* to a *native*, the second mode is used due to an incomplete error model which does not take into account important factors like *non-isomorphism*. Thus, the  $\chi^2$  of scaling between a native and a derivative provides a measure of non-isomorphism, assuming the detector error is accurately modeled for both samples.

$R_{\text{merge}}$  was historically used as an estimate of the non-isomorphism of data collected on film using several different crystals, and for this purpose, it still has some validity because we do not account for non-isomorphism in our error model. It is not so important now that complete X-ray data sets are collected from single, frozen crystals.

One of the drawbacks of using  $R_{\text{merge}}$  as a measure of the quality of a data set is that it *can be intentionally and unintentionally manipulated*. Unintentional factors which can artificially lower  $R_{\text{merge}}$  generally have the effect of reducing the redundancy of the data or eliminating weaker observations. In crystallography, the greater the redundancy of the data is, the worse is the  $R_{\text{merge}}$ , because of the correlation between  $I$  and  $\langle I \rangle$  which reduces the  $R_{\text{merge}}$ . The greater the redundancy, the lower is the correlation. For two measurements with the same  $\sigma$ , the correlation is 50%, so  $R_{\text{merge}}$  is underestimated by  $\sqrt{2}$  compared to the case of no correlation. Known unintentional factors which lower  $R_{\text{merge}}$  include the following:

1. Data collected so that lower resolution shells, where the data is strong, have a higher redundancy than the higher resolution shells, where the data is generally weaker. This can be accomplished by collecting data on detectors where  $2\theta \neq 0$ ,



or including data from the corners of rectangular or square image plates. There is nothing wrong with using this data; it will just artificially lower the  $R_{\text{merge}}$ .

2. Inclusion of single measurements in the calculation of  $R_{\text{merge}}$  in one widely used program, which is why a table using this erroneous calculation used to be presented in the *Scalepack* output. Although the bug in the program was unintentional, it nonetheless reduced the  $R_{\text{merge}}$ , and this may have accounted for the parameter's longevity. A second, more subtle bug that reduced  $R_{\text{merge}}$  prompted the introduction of the keyword **BACKGROUND CORRECTION**. Fortunately, both bugs have now been fixed, but the point is that errors of this type can persist.
3. The omission of negative or weak reflections from the calculation of  $R_{\text{merge}}$ . This is often undocumented behavior of crystallographic data scaling/merging software. Examples include:
  - a) elimination from the  $R_{\text{merge}}$  calculation the reflections that have negative intensities.
  - b) conversion of  $I < 0$  to  $I = 0$  before the calculation of  $R_{\text{merge}}$  and inclusion of this reflection in the data set (the statistics of such a type are included in the *Scalepack* output for reasons of comparison. This is the first  $R_{\text{merge}}$  table in the log file, not the final one).
  - c) omitting reflections with  $\langle I \rangle < 0$  from the calculation of  $R_{\text{merge}}$  but the inclusion of these reflections in the output data set. Default  $\sigma$  cutoffs set to unreasonable values, like 1. This is, in fact, the default of the software used commonly to process image plate data.
4. Use of unreasonable default or recommended rejection criteria in other programs. These eliminate individual  $I$ 's which should contribute to  $R_{\text{merge}}$  and yet are still statistically sound measurements.
5. Use of the eigenvalue filter to determine the overall B factor of a data set collected on a non-frozen, decaying crystal. In this case, the eigenvalue filter will calculate an overall B factor, which is appropriate for the middle of the data set, yet apply this to all data. As a result, the high-resolution data will be down-weighted compared to data processed with the first, least decayed frame as the reference. The high-resolution data is generally weaker than the low-resolution data, and as a result, is more likely to result in higher  $R_{\text{merge}}$ . By down-weighting the high-resolution data, the  $R_{\text{merge}}$  is artificially lowered. Any program which does not allow the option of setting the reference frame will have this problem. Of course, there is no problem with non-decaying crystals.

There are also intentional ways of lowering your  $R_{\text{merge}}$ . Like those ways listed above, they generally result from the statistically invalid elimination of weak reflections, reduction of the redundancy of the data, or de-emphasis of weak data. The difference between these methods and those listed above is that they are generally under the control of the user.

1. Use of an unreasonable **SIGMA CUTOFF** (e.g.  $\geq 0$ ). The rejection of weak data will always improve  $R_{\text{merge}}$ . There is a further discussion of **SIGMA CUTOFF** in the *Scalepack Keywords Descriptions* section.
2. Use of a resolution limit cutoff. Again, the omission of weak data will improve  $R_{\text{merge}}$ . A reasonable resolution cutoff is the zone where  $I/\sigma < 2$ .
3. Combining into a single zone for the purposes of calculations those resolution shells where  $R_{\text{merge}}$  is rapidly changing. In this case, the shell will be dominated by the strong data at the low-resolution end of the zone and give the impression that the high-resolution limit of the zone has better statistics than it does. For example, if you combined all your data into a single zone, the  $R_{\text{merge}}$  in the *final* shell would be pretty good ( $R_{\text{merge}}$  overall), when in fact it was substantially worse. As described above, it is more sensible to divide your zones into at least 10 (or preferably more) equal volumes.
4. Omitting partially recorded reflections. This has the effect of a) reducing the redundancy, and b) eliminating poorer reflections. Partially recorded reflections will always have a higher  $\sigma$  associated with them because they have a higher total background, due to the inclusion of background from more than one frame in the reflection.
5. Scaling I+ and I- reflections separately in the absence of a legitimate anomalous signal (**SCALE ANOMALOUS**). This has the effect of reducing the redundancy.
6. Ignoring overloaded reflections using the **IGNORE OVERLOADS** in *Scalepack*. The intensity of overloaded or saturated, reflections cannot be directly measured because some of the pixels are saturated. Profile fitting only measures these reflections indirectly, by fitting a calculated profile to the spot using the information contained in the *wings* or *tail* of the spot. Ignoring the inaccuracies inherent in this method by ignoring overloads may have a dramatic effect on  $R_{\text{merge}}$ . Note that in case of molecular replacement the option **INCLUDE OVERLOADS** should be used.

**IGNORE OVERLOADS** is often a useful tool. For example, when calculating anomalous differences, you do not want to use overloaded reflections because you are looking for very, very small differences and want to use only the most accurate data. Another time you might ignore overloads is when you collect multipass data. In this case, a crystal is exposed twice, once for a short time, the other for a longer time. The longer exposure is to sufficiently darken the high resolution reflections but will result in saturated low-resolution reflections. Since the low-resolution reflections can be obtained from the short exposures, the overloaded ones can be ignored in the long exposures.

## The Resolution of Diffraction Data.

A statistically sensible cutoff for the resolution of a diffraction data set is that shell where the average  $I/\sigma$  is 2 after correctly integrating and scaling the data. The resolution of the data is a distinct criterion from its completeness. The best data will be 100% complete. Completeness may suffer due to anisotropic diffraction, overlapped reflections, or geometrical

constraints (data from the corners of the detector was used, beam stop or cooling nozzle shadows, etc.). In order to properly estimate the resolution of the data, you have to take into account the  $I/\sigma$ ,  $R_{sym}$  and the completeness of the data.

.

## The Project Tab After Scaling

After scaling your data, the Project Tab will contain information about your scaled sets (Figure 106). A new set will appear every time you rename the output file. This can be useful if you are scaling the data in different space groups or want to check scaling results using different parameters.

If you have a different results file from a previous session (that was not saved to your project), you can load it here using the **ADD FILE** button. Conversely, if you do not want to keep a file in your set history, you can use the **REMOVE FILE** button. For example, if you have scaled the data in two data sets, you may want to remove the file corresponding to the incorrect space group.

**HKL-2000 v717** Package Licensed to John Doe at HKL Research, Inc.

File Options Site Configuration Crystal Information Report About Help

**Project** Data Auto Summary Index Strategy Integrate Scale Merge Macros

**Project**

Name: APC5771  
 Crystal: crystal1  
 Experimenter: dcoop  
 Date: Aug 29, 2018

Load Save

Edit Project Edit Crystal

New Project New Crystal

Evaluate Model

Principal Component: Protein

**Protein Data**

Name:  
 Number of Residues: 94  
 Organism:  
 Description:  
 Molecular Weight: 10715.48  
 NCBI accession code:  
 Swiss-Prot accession code:  
 No. of Homologues:  
 Last check in PDB:

GHMPAYVFSKESFLKFLEGHLEDDVWWWSSDVTDFCKKLSES  
 MVGEKEYCFAEFAPADIFDADEDEIDEMIMKYAMFVEKEKLSE  
 AGRNAIR

Phasing Method: SAD/MAD  
 Source of Anomalous Signal: Se 4

Ala (A): 8	Gly (G): 4	Met (M): 4	Ser (S): 7
Cys (C): 2	His (H): 2	Asn (N): 1	Thr (T): 1
Asp (D): 9	Ile (I): 4	Pro (P): 2	Val (V): 10
Glu (E): 13	Lys (K): 8	Gln (Q): 0	Trp (W): 0
Phe (F): 9	Leu (L): 5	Arg (R): 2	Tyr (Y): 3

Anomalous Wizard

**Scaled Data**

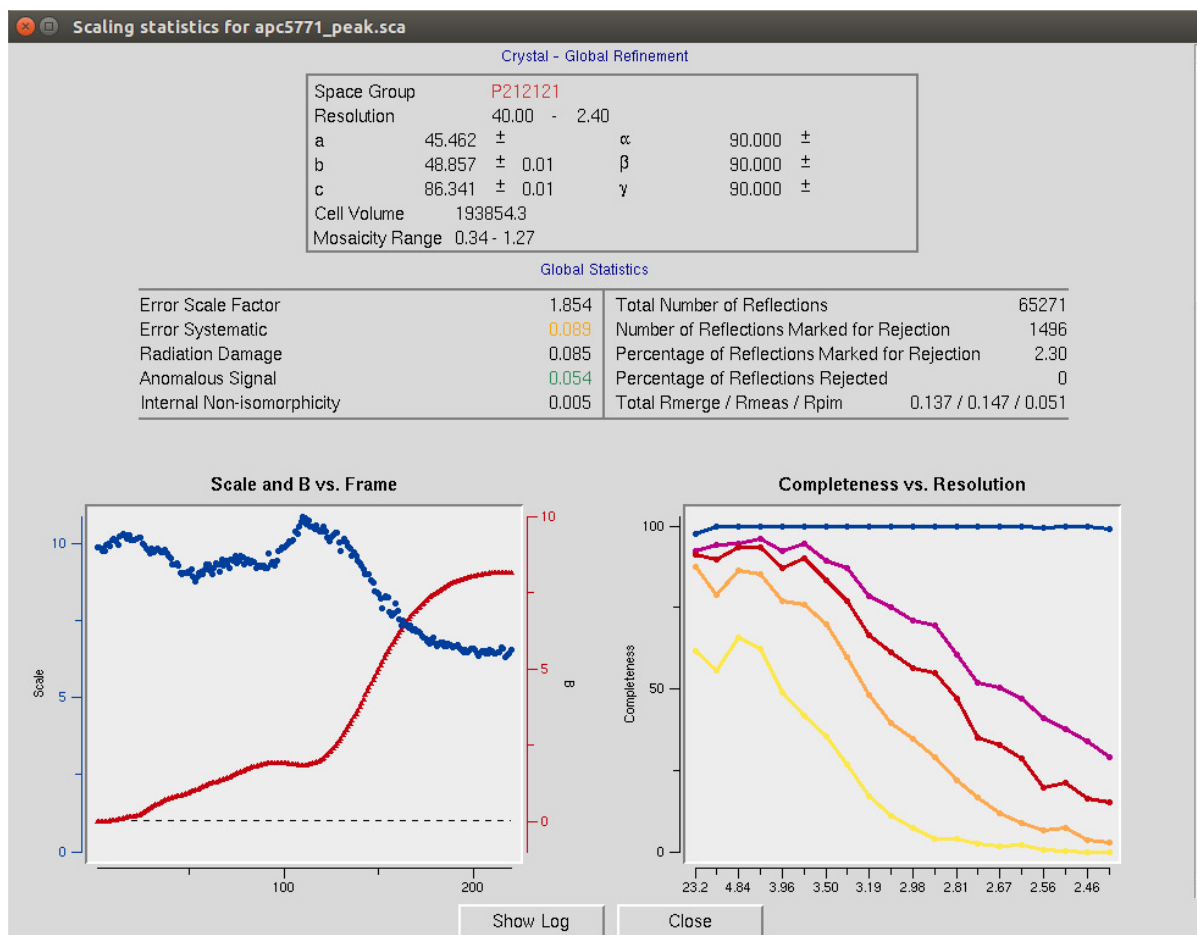
Structure	File Name	Cell	Space Group	Wavelength	Resolution	Mode	Order
Select	/home/dcoop/Data/APC5771_2fdo/proc_home/apc5771_peak.sca	45.462 48.857 86.340 90.000 90.000 90.000	P212121	0.97902	2.40	Native	1
Select	/home/dcoop/Data/APC5771_2fdo/proc_home/apc5771_inf.sca	45.722 48.889 86.724 90.000 90.000 90.000	P212121	0.97919	2.40	Native	2

Add File  
 Remove File  
 Statistics

Figure 106. The Project Tab after scaling two data sets

## THE PROJECT TAB AFTER SCALING

You can use the **STATISTIC** button to display the results associated with the selected file. This display also includes a **SHOW LOG** button that will display the log file associated with the scaling.



**Figure 107.** Use the Statistics button to see scaling results

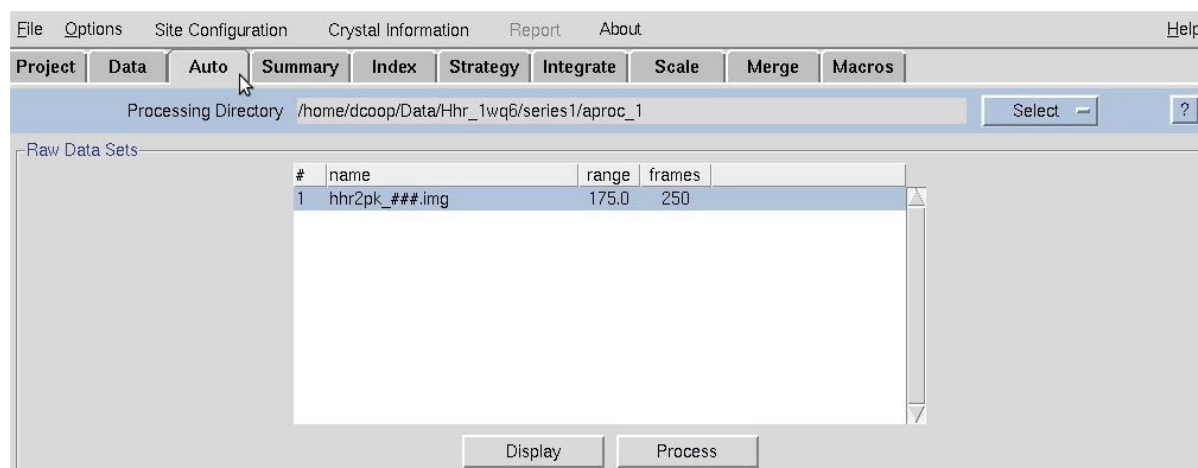
# Automatic Data Processing

## HKL-2000 now has the ability to integrate and scale many data sets automatically.

Although this process is not always successful, it has a high rate of success. This process has been tested using over 6000 datasets from the Integrated Resource for Reproducibility in Macromolecular Crystallography (<http://proteindiffraction.org>). Data indexing and integration were successful for 83% of the data. The correct space group was identified for 76% of the datasets.

Automatic processing uses the CCP4 program POINTLESS to identify the space group and therefore requires that CCP4 is installed and HKL is aware of the location of the CCP4 programs. CCP4 requires its own license, and the installation of CCP4 is not covered in this manual. For more information about CCP4, please see the CCP4 website (<http://www.ccp4.ac.uk>).

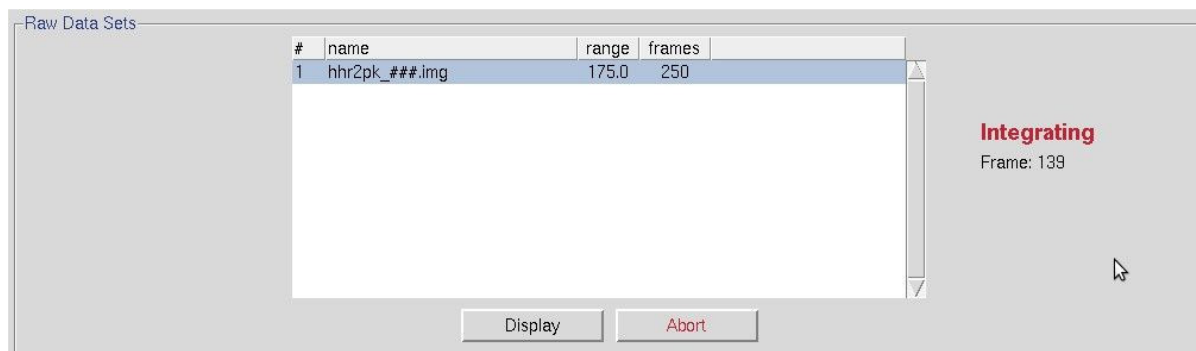
To begin the automatic processing, select the Auto tab after loading the datasets using one of the methods outlined above, using either the **LOAD DATA SETS** button on the **Data** tab or load a project that already has data sets you would like to process. The **Auto** tab will open with the set of images in the **Raw Data Sets** box (Figure 108). Use the **DISPLAY** button see the first raw diffraction image. You do not need to peek search the data. To start the automatic processing procedure, click the **PROCESS** button.



**Figure 108. The Raw Data Sets window**

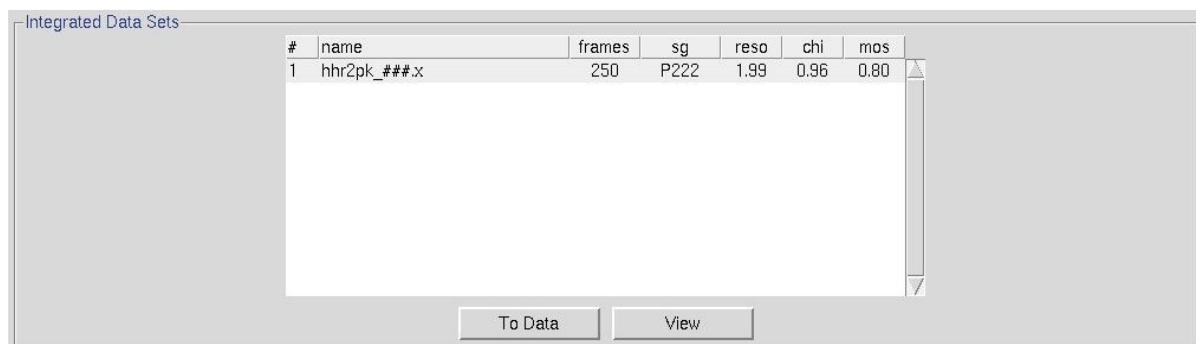
HKL-2000 will process every file in the data set. During the first set, the word “Integrating” will flash, and the frame number being processed will update (Figure 109). This does not update for every frame, but rather is used as an indication that the process is ongoing. If the process seems to freeze, you can use the **ABORT** button to stop the process.

## AUTOMATIC DATA PROCESSING

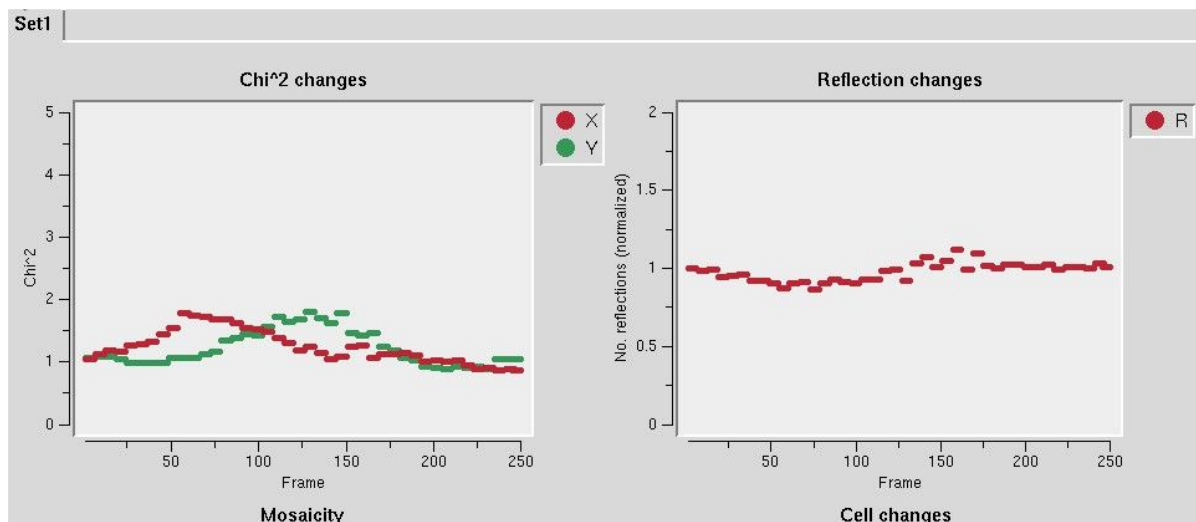


**Figure 109.** During automatic data processing.

Once all of the individual diffraction images have been processed, each set will be listed with some basic information about the integration process (Figure 110). The **TO DATA** button can be used to create a set within the Data tab. This can be useful if you would like to manually scale the data after the process has finished. The view button will show you six diagnostic graphs about the integration process, including graphs of the frame number versus  $\chi^2$ , Unit Cell Changes, Normalized Number of Reflections, Changes, Distance Changes, Crystal Orientation Changes, and Mosaicity (Figure 111). These graphs should generally be flat, but some variability is OK. If you have more than one data set, the resulting graphs for each data set will be in a different tab.



**Figure 110.** Automatic integration results



**Figure 111. Automatic integration diagnostic graphs**

After the integration process has finished, the reflections from all the frames need to be scaled together. During this process, the flashing word “Integrating” will change to “Scaling”. Once scaling is complete, the name of the output file will appear, along with the crystals space group and the resolution chosen for scaling (Figure 112. Automatic scaling results). You can transfer this output file to the Project tab using the **TO PROJECT** button (see The Project Tab After Scaling on page 120). The **VIEW** button can be used to display the typical output from scaling (see Output Statistics and Graphs on page 98).

Scaled Data Sets

#	name	sg	reso
1	output.sca	P21212	1.99

To Project View

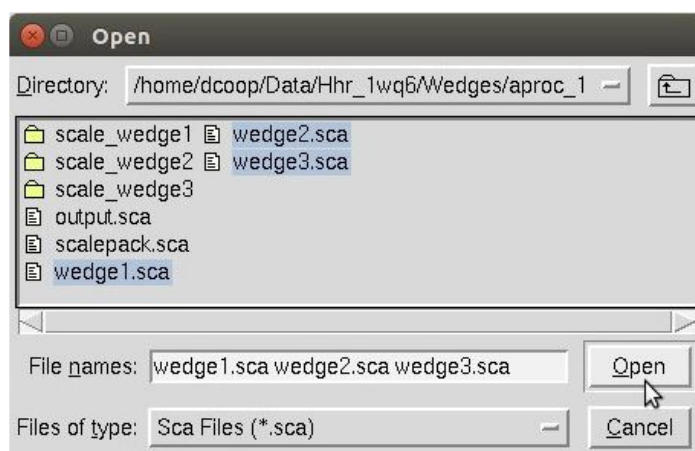
**Figure 112. Automatic scaling results**



## Merging Multiple Scaled Files

You may come across situations that necessitate combining data sets that have been scaled separately. This most obvious application of this is when you are combining data from multiple crystals to obtain a complete data set, but there are other situations that may require this. Perhaps you have collected some data from a crystal, returned it to a storage dewar, and subsequently realized you had to collect more data on it. On many beamlines, the crystal mounting robot will often return the crystal to an orientation that is close enough that you can process and scale the data all at once, but scaling the different parts of the data separately will be more appropriate in some situations.

The **Merge** tab provides a straightforward interface to combine various scaled files. To start the process, you need to load the individual scaled files into the interface. There are two methods to do this. First, you may use the **LOAD** button to bring up a file selection dialog. If the files are in the same directory, you may select multiple files at once (Figure 113). In most cases, the files will be in separate directories and will have to be loaded one at a time. The other method of loading the files into the merge dialog is to copy them from the Project tab, using the **FROM PROJECT** button. Each scaled file in the will be copied to the list of scaled files. If you have some files that were not already stored in your project, select them and click **To PROJECT**, and those files will be added to your project.



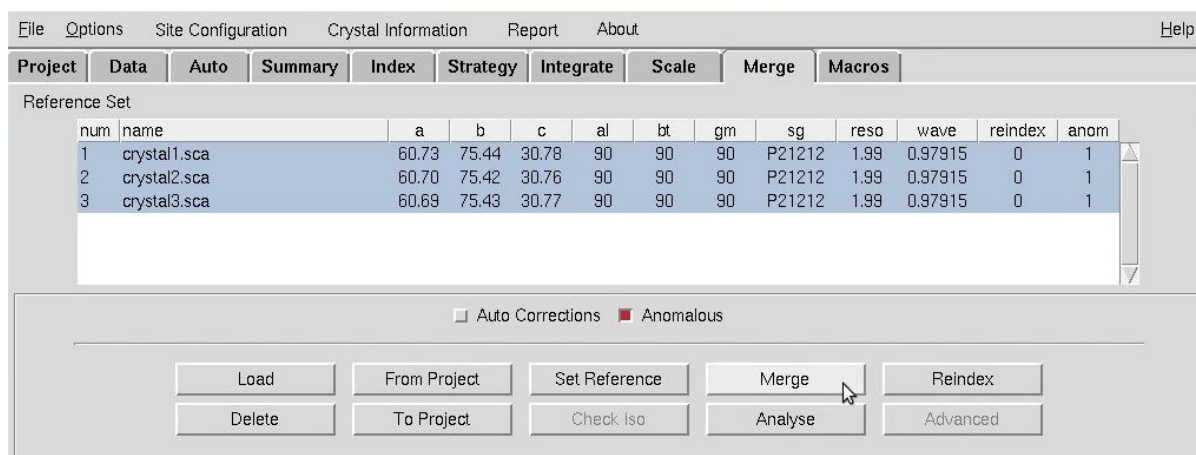
**Figure 113.** *Selecting several files to merge.*

To merge files together, you must first select at least two files you want to combine. Select a file by clicking within its row, and it will turn blue (Figure 114). The **ANALYSE** button can be used to look at the scaling results from a single set that is selected. By default, HKL will choose the unit cell from the first (highest on the list) file that is selected. If you would like to use the unit cell from another file, select that file first and click the **SET REFERENCE** button.

If any of the files you have loaded into the screen have anomalous data, by default the **Anomalous** option will be selected, preserving any anomalous data you have. You may also

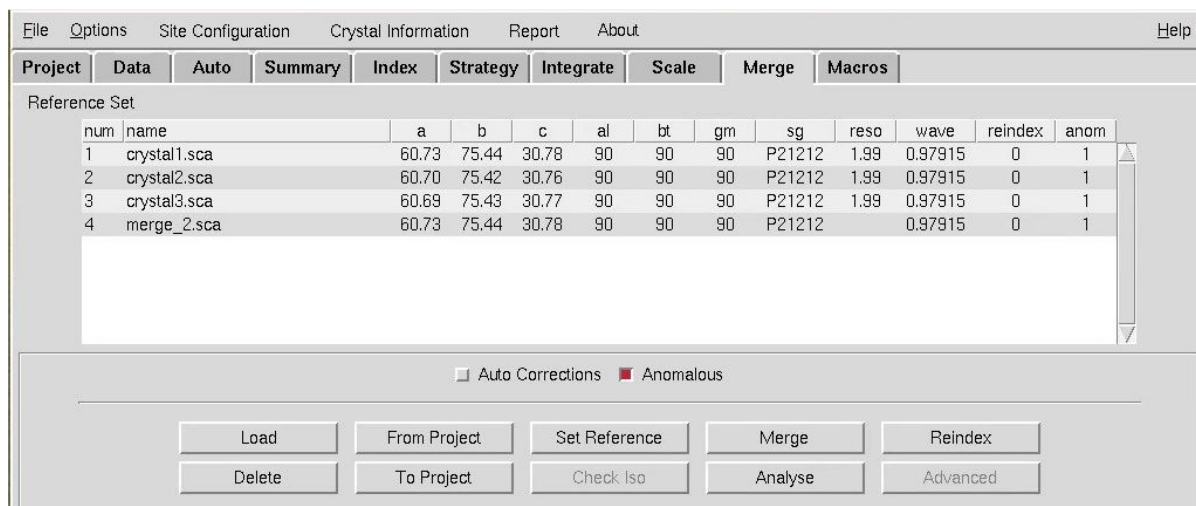
## MERGING MULTIPLE SCALED FILES

choose the Auto Corrections option, which will decide for you if anomalous data is present. To merge the files, click the **MERGE** button.



**Figure 114.** Files selected to merge.

You will immediately see the scaling results open in a new window, and the resulting file name will now appear in the sets window. You will notice that the graphs that plot values against the frame number will only have as many “frames” as the number of files you were merging. The other graphs will show the cumulative results. If the scaling results are not good, you may need to reindex one or more of the data sets. The **REINDEX** button opens the reindex dialog and allows you to reindex the sets independently.



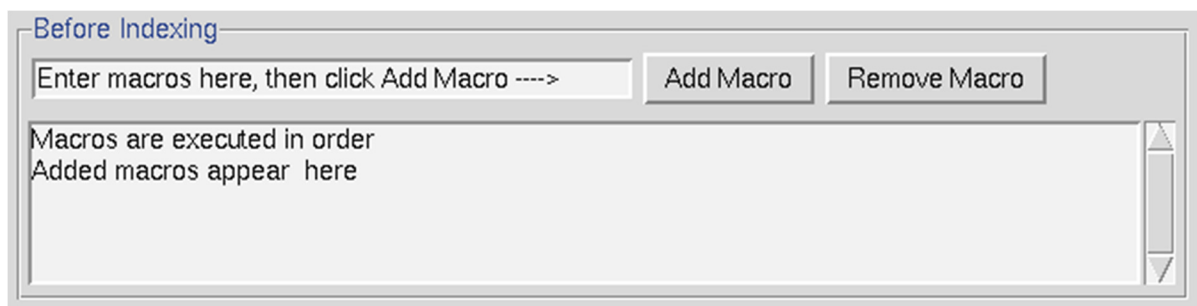
**Figure 115.** The merged file will appear in the list.

## Using Macros

In HKL-2000, macros provide you with a mechanism of interacting with your data in a way that is sometimes not possible using the interface alone. Before HKL-2000, the HKL package consisted of individual programs (i.e., Denzo and Scalepack) that were run using command scripts that were composed of keywords specific to those programs. These individual commands are still at the heart of HKL-2000, but the interface removes the burden of remembering and ordering the keywords. These keywords can still be used to interact with your data via the **Macros** tab (Figure 116). Keywords can have modifiers, arguments (usually values), or both. A macro is a combination of a keyword and necessary modifiers and/or argument.

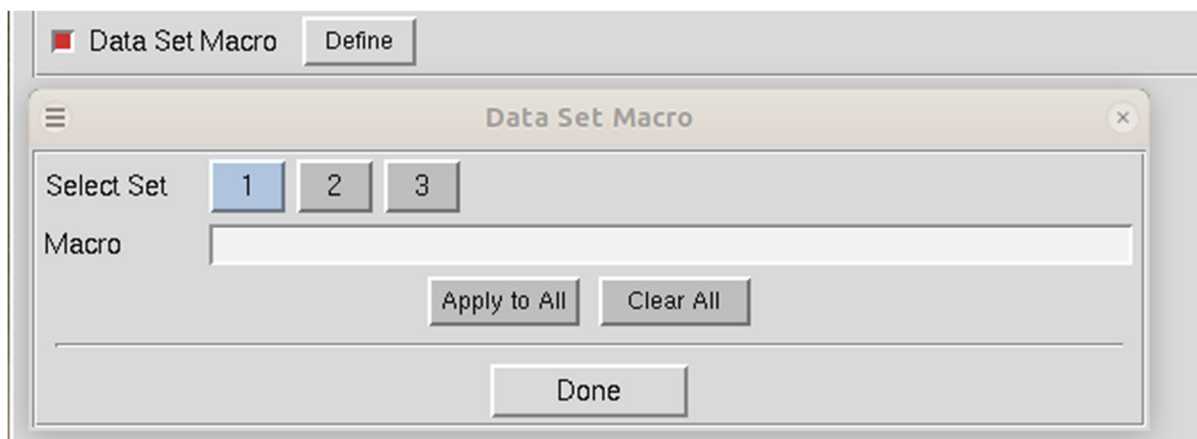
The **Macros** tab is divided into five sections: **Before Indexing**, **During Refinement**, **During Integration**, **During Scaling**, and **Immediate Execution**. Each section has a text box at the top where the macros can be entered. After you have typed a complete macro (keyword/modifiers/values) click **ADD MACRO**, and the macro will appear in the macro list for that section. The name of the section indicates when the macros in that section are executed. Macros will be executed in the order they appear in the list.

You can select a macro by clicking on it, which will give it a blue background. This has the same effect as if you had selected the text and issued a “copy” command and allows you to paste it into the text box of a macro section using the middle mouse button. For example, many macros that are useful **During Refinement** are also helpful **During Integration**. If you need to remove a macro, select it and click the **REMOVE MACRO** button.



**Figure 116. The Macros Tab**

The **During Scaling** section has an additional option. If you select “Data Set Macro” and click **DEFINE**, you will a new dialog that allows you to specify to which data sets the macro applies (Figure 117). Choose one or more data sets by clicking on the numbered buttons. The macro will apply to the selected (blue) sets. Click **DONE** when you have entered all the macros you want.



**Figure 117. Data Set specific macros**

The bottom section of the *Macros* tab, *Immediate Execution*, only has one button, **RUN MACRO**, which does what it says.

The list of keywords that can control HKL-2000 are listed in *Appendix – Indexing and Integration Keywords* and *Appendix - Scaling Keywords* starting on page 145. Many of these keywords are more conveniently issued using the HKL-2000 interface but are presented for completeness.

## Appendix – HKL-2000 Installation

---

*The following section on installation is included for completeness, but users are strongly encouraged to get up to date information about installation from the HKL Research, Inc website - <https://hkl-xray.com/>. The website also includes information about installing HKL-2000 on Windows systems.*

---

### Obtaining HKL-2000 installation files

An HKL installer and executables of all programs can be obtained from <https://hkl2000.hkl-xray.com>. To download the programs, use the login id: hkl2000ftp. The password can be obtained through the automatic password system (see the web page) or by contacting HKL Research, Inc. via email.

Several versions of HKL-2000 are available. The latest version, which is described in this manual, is version 717. If the program has been updated since the initial version release, then the version may have a version extension, such as v717.1. The directory corresponding to this version has versions for several operating systems, as well as copies of the HKL Installer for each operating system. Each file will be compressed using standard compression used for that operating system. For example, if you are using Linux, use the command

```
> tar zxf HKL2000_v717.1_installer-Linux-x86_64.tar.gz
```

This will create the directory named after the downloaded file. The files contained within this directory will depend on the file downloaded. For example, the HKL installer file will contain a file `access_prod`, the `HKL_installer`, a `README.txt` file, and a directory of `setup_files`. The compressed file containing the Linux executables will contain a `README.txt` file, setup files for Bash or CShell (`hkl_setup.sh` and `hkl_setup.csh`, respectively), and `bin`, `data`, and `lib` directories.

### License key

In order to run HKL-2000 or use the HKL Installer, you will need to obtain a license from HKL Research Inc. The license key is contained in a file `cr_info`. The traditional location of this file is in the `/usr/local/lib` directory. The HKL Installer will look for the file in this location but allows you to specify the location. To obtain the license key, you

## APPENDIX – INSTALLATION

need to run the `access_prod` program on each machine that you plan to use for running the *HKL-2000* programs. The `access_prod` program is main directory if you downloaded the installation file and in the `bin` directory if you downloaded the executables. Executing the `access_prod` results in the following text-based dialog (example answers to the prompts are shown in red):

- enter the name of your institution: **State University**
- enter the name of PI: **John Doe**
- enter the name of person responsible for HKL maintenance: **Jane Doe**
- enter names of all frame formats (detectors) you want to use: **mar raxis4**
- enter the number of detectors in your lab which are currently supported by HKL package: **2**
- enter the name of company from which you received your license: **HKL Research, Inc.**
- enter HKL version number: **717**
- if this is temporary license, enter the number of days you request, for permanent license enter 'noexpir': **noexpir**
- do you want to send automatically e-mail to [hkl@hkl-xray.com](mailto:hkl@hkl-xray.com) in order to activate your license (y) ? **n**

If the answer is y (yes), the `access_prod` program may automatically send an e-mail with a request for an access file: `cr_info`. It does not always work, though (depending upon how the email system is configured on your computer. If the answer is n (no) the program will produce the following info file. You may save the info file and afterwards "manually" send it to [hkl@hkl-xray.com](mailto:hkl@hkl-xray.com).

Please:

- ✓ Do not edit or alter the **info** file - it is processed automatically after verification.
- ✓ Do not send an **info** file as an attachment.
- ✓ Note that the **access\_prod** program does not work properly if the setup of the computer operating system is not consistent.

The example of an info file created by the `access_prod` program is given below (comments in italic):

HOST-NAME: p4.med.state.edu	(extracted by the program - <b>do not edit!</b> )
	NOTE: the field is limited in length to 30 characters
HOST-ID: -2035379618	(extracted by the program - <b>do not edit!</b> )
HOSTNAME: state.edu	(extracted by the program - <b>do not edit!</b> )
HW_PROV: sgi	(extracted by the program - <b>do not edit!</b> )
CPU-SERIAL: Linux, eth1, 94DE...	(extracted by the program - <b>do not edit!</b> )
INSTITUTION: State University	Licensee's Institution name
PI: John Doe	Principal Investigator or Group Leader
CONTACT: Jane Doe	Contact Person

## APPENDIX – INSTALLATION

DETECTORS: mar raxis4  
NO-OF-DET: 2  
PROVIDER: HKL Research, Inc  
VERSION: 717  
EXPIR: noexpir

List detectors you expect to use, like Eiger or Rigaku  
Number of Detectors at Home Laboratory  
License Provider: Rigaku/MSD, HKL Research, etc.  
Requested Version of the Program

### Using the HKL Installer

The default location of the HKL-2000 installation is in `/usr/local/HKL/` in a directory named for the version number. This directory almost always requires elevated privileges, so you will have to start the program with sudo privileges (`sudo ./HKL_installer`). The installer will first verify that your license key is valid. The traditional location of the license key file (`cr_info`) is in `/usr/local/lib`. The text area at the bottom will provide details about your license (Figure 118).

The buttons at the bottom of the installer let you select which HKL version you would like to install. Select HKL-2000 to install the program described in this manual. Click “Continue” to see version specific license. You must accept the license to proceed (Figure 119).

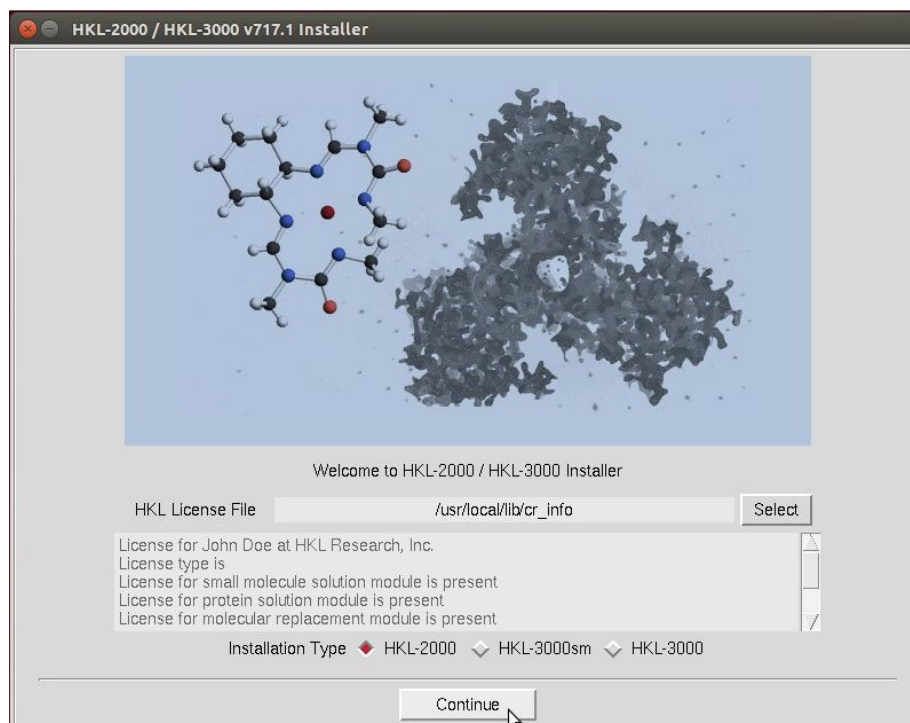
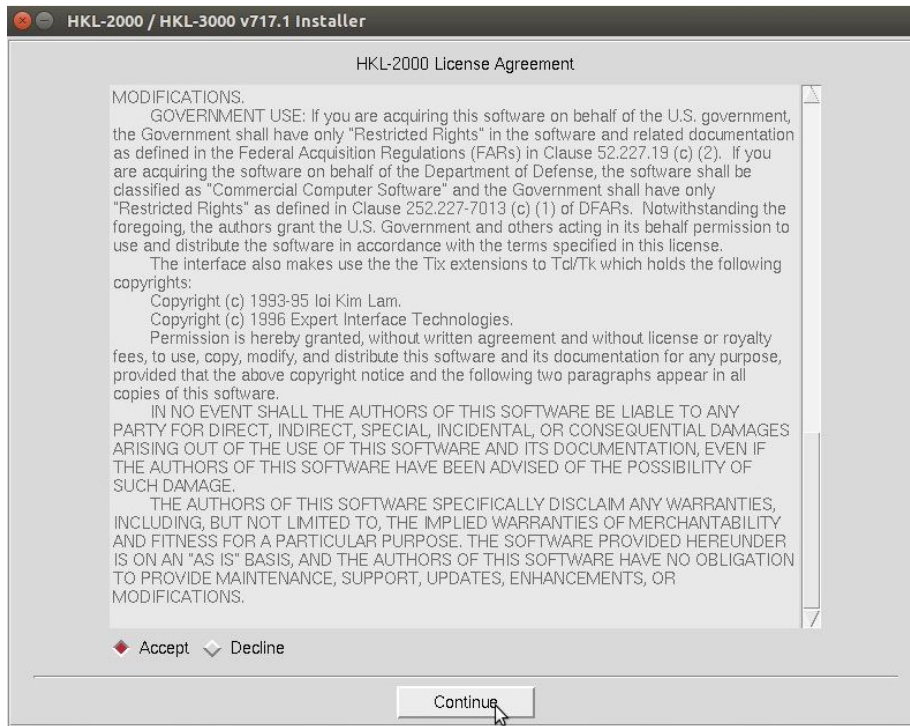


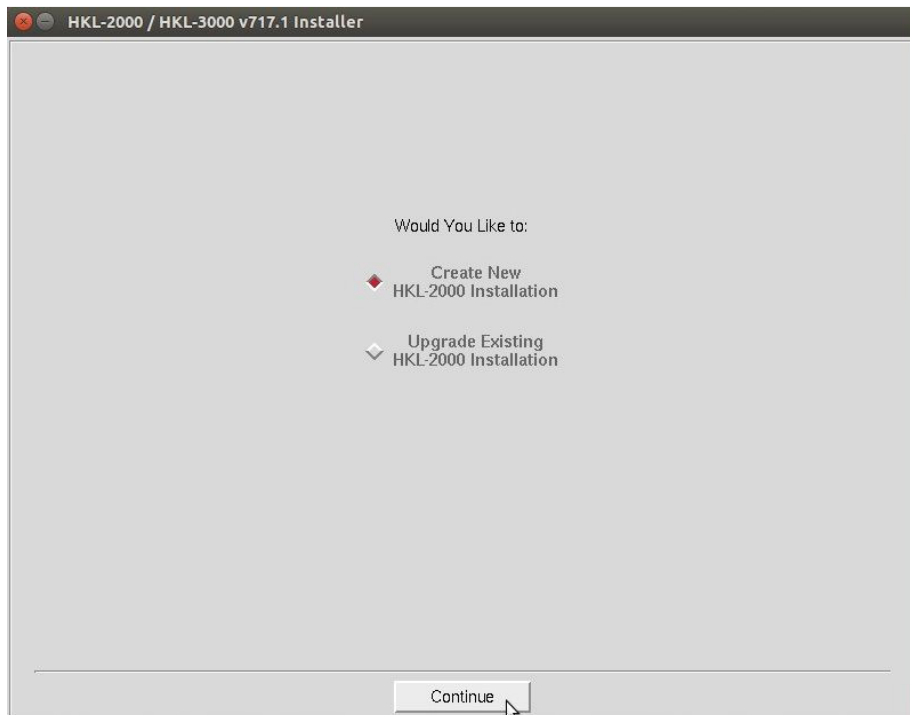
Figure 118. Select to install HKL-2000





**Figure 119. Read and Accept the HKL-2000 License**

The following screen will let you select if you want to create a new HKL installation or upgrade an existing installation (Figure 120).



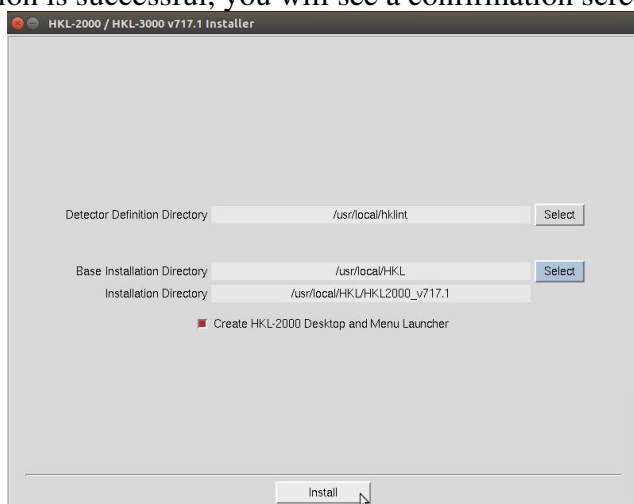


**Figure 120. Create a new installation or upgrade an existing one**

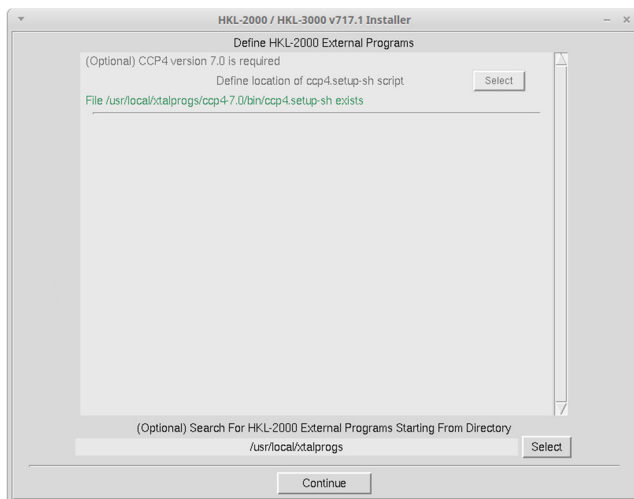
The next dialog allows you to specify the location of the program and the detector definition directory (Figure 121). The default locations for these are `/usr/local/HKL` and `/usr/local/hklint`, respectively. You may change these if you want to use another directory or do not have write permission for these directories. The default folder for HKL will include the version number. This page will also ask if you want to create a desktop icon to launch HKL-2000. Once you have things set up how you would like, click **INSTALL** to finish the setup. The installer will then check if you have permission to write to these directories and install all the necessary files.

HKL-2000 can optionally use some CCP4 programs, and the HKL Installer will let you select the location of CCP4 (Figure 122).

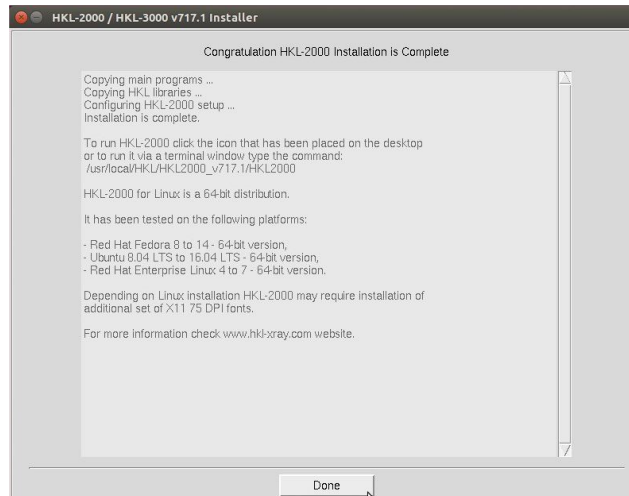
If the installation is successful, you will see a confirmation screen (Figure 123).



**Figure 121. Select the installation directory**



**Figure 122. External program selection**



**Figure 123. Finish the installation**

### **Installing HKL-2000 in Windows**

HKL-2000 can be installed in Windows 10 (build 1703 or greater). This requires the activation of Windows Subsystem for Linux (WSL), which allows you to run a complete Bash shell in a sandbox alongside your Windows operating system, and installation of an X-windows implementation. Because this requires changes to your operating system, the procedure is beyond the scope of this manual. However, the instructions can be found on the HKL Research, Inc. website ( <https://hkl-xray.com/windows-installation> ). Once WSL and an X-windows client has been setup, installation of HKL-2000 essentially follows the instructions contained in this manual.

## **Installing HKL-2000 manually**

If you would like to install HKL-2000 using the traditional method used before the HKL Installer was available, download and uncompress the executable directory in a suitable location. Typically, the executables are placed in the directory `/usr/local/HKL`, which requires system administrator privileges. You can also place executables in an arbitrary directory, for instance, `/home/joe/programs/hkl`. In that case, you have to set the environmental variable `$HKLPATH` to that directory. The exact command will depend on the shell you are using to run programs (this can be determined by typing `echo $SHELL`). For C-type shells (`csh`, `tcsh`, etc.) and the example above, the command is:

```
setenv HKLPATH /home/joe/programs/hkl
```

For Bourne-type shells (e.g., `sh`, `bash`, `dash`, etc.), the command is:

```
export HKLPATH=/home/joe/programs/hkl
```

HKL-2000 also requires a directory `hklint` for site files. Its default location is:

```
/usr/local/hklint
```

Typically, accessing this location requires administrator privileges, but it can be changed to an arbitrary location, by setting the environmental variable `$HKLDIR`. For example:

```
setenv HKLDIR /home/joe/programs/hkl
```

(for C-type shells, and `export HKLDIR=<path>` for Bourne-type shells as above).

The `anom.dat` file should be placed in this directory together with subdirectories containing the site files. The site file (`def.site`) of each detector (home or synchrotron) should be stored in its own directory. Each such directory should be appropriately named to identify the specific facility and the detector easily.

Example:

```
/usr/local/hklint/Raxis4  
/usr/local/hklint/Spring8-BL23-Quantum4  
/usr/local/hklint/BNLX25-Quantum4
```

You can check how the environmental variables were set by typing the command:

```
env | grep HKL
```

To keep the values of the HKL environmental variables for subsequent logins, you should modify the configuration files for the shell you are using (`~/.tcshrc`, `~/.cshrc`, `~/.bashrc`, etc.).

## Starting HKL-2000

After you have installed *HKL-2000* and placed the license file `cr_info` either in the `/usr/local/lib` directory or the directory you specified during installation, there are two ways to start the program. If you have used the HKL Installer, the program can be started from a terminal window by typing:

```
HKL2000
```

## **APPENDIX – INSTALLATION**

If you receive an error message "no display name and no \$DISPLAY environment variable", it means that the X windows system is not working or not configured properly. On Mac OS X, you need to start the X11 application (from the Finder's **Go** menu, choose **Applications**, then choose **Utilities** and select **X11**). You should see the **X11** icon in the dock, double click on it to open an **xterm** window and start *HKL-2000* from this window.

If you receive an error message "Command not found", it means that HKLPATH is not in the executable command search path. Modify the environmental variable PATH or type a full path e.g. /usr/local/bin/HKL2000.

### **What files make up the *HKL-2000* suite, and what do they do?**

HKL2000 – The main program with the graphical user interface  
access\_prod – The program that prepares file necessary for license key generation.  
It has to be executed on each computer that will run the *HKL-2000* suite.  
anomalous – Anomalous signal analysis  
cdirt – Directory manipulation tools  
denzo\_3d – Auto-indexing, refinement, and integration; 3-D processing  
hklfilename – Filename manipulation tools  
hklsysinfo – Program that gathers information about computer resources  
readheader – File header reader  
scalepack – Scaling data comprising up to 2000 frames and 2 million reflections  
scalepack16m – Scaling data comprising up to 2000 frames and 16 million reflections  
scalepack8m – Scaling data comprising up to 2000 frames and 8 million reflections  
scalepacklvirus – Scaling data comprising up to 3000 frames and 50 million reflections  
scalepackmanyframes – Scaling data comprising up to 4000 frames and 2 million reflections  
scalepackmf6500 – Scaling data comprising up to 6500 frames and 3 million reflections  
scalepackribo – Scaling data comprising up to 2000 frames and 32 million reflections  
scalepackvirus – Scaling data comprising up to 2000 frames and 32 million reflections  
strat – Strategy and simulation  
suser – Internal communication program  
xdark – Dark current calculation  
xdisp – Image display and measurement

# Appendix - Trouble Shooting

## When refinement goes wrong

Occasionally the auto-indexing routine does not give clear-cut results. There are four classes of data-reduction problems:

### Sometimes it's your crystal

Below are things that can cause problems processing and scaling your data. However, there is only so much that can be done to compensate for problems with crystal. Disorder, split crystals,  $\text{CuK}\alpha_1/\text{CuK}\alpha_2$  splitting, multiple crystals, twinned crystals, etc can all make it difficult to get good data. Bad crystals are still going to give bad data.

1. Utter failure
2. Misprediction
3. Kind-of-failed
4. The predicted positions agree perfectly with the peaks, but the data scale poorly

**1. Utter failure** means that there is no reasonable predicted lattice, the crystal unit cell parameters are unreasonable, and the predicted diffraction pattern does not correspond to the observed image. In most cases, failure can be ascribed to problems in three areas:

- A. experimental conditions,
- B. auto-indexing parameters, and/or
- C. the peak search (rarely).

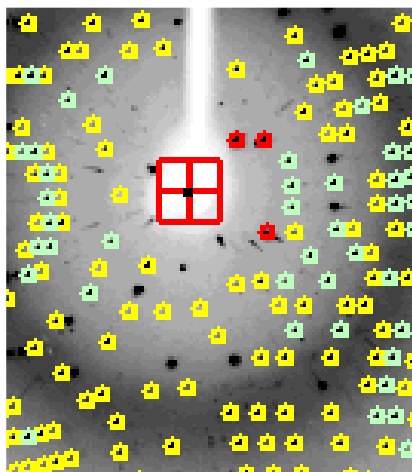
A. **Experimental conditions that can cause indexing failure:** Problems with experimental site parameters account for the vast majority of the problems with auto-indexing single-crystal data. If any of the following parameters is widely off, your auto-indexing will also be off (listed in order of importance):

- i. The **X BEAM** and **Y BEAM** positions. The error of the direct beam coordinates should be less than half of the distance between the closest diffraction spots; therefore, small unit cells are more forgiving about beam position errors.

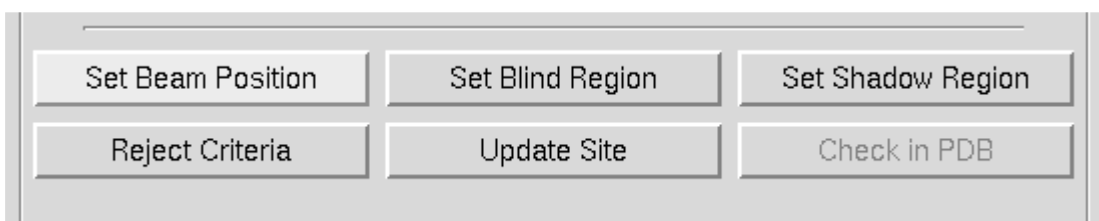
There are several methods of setting the **X BEAM** and **Y BEAM** positions. The simplest is to measure the coordinates directly from the *Image Display*. The position of the cursor is displayed in the little green box on the right. The best case for this is if the beam stop permits a tiny amount of the direct beam to pass through and make a spot on each frame (Figure 124), but this method can also be used if you have a combination of a small beam stop and a relatively small unit cell. After determining

## APPENDIX – TROUBLE SHOOTING

the position of the direct beam with the cursor, it can either be entered using the **Site Configuration** dialog (Figure 14) or using the window launched by the **SET BEAM POSITION** button (Figure 125).

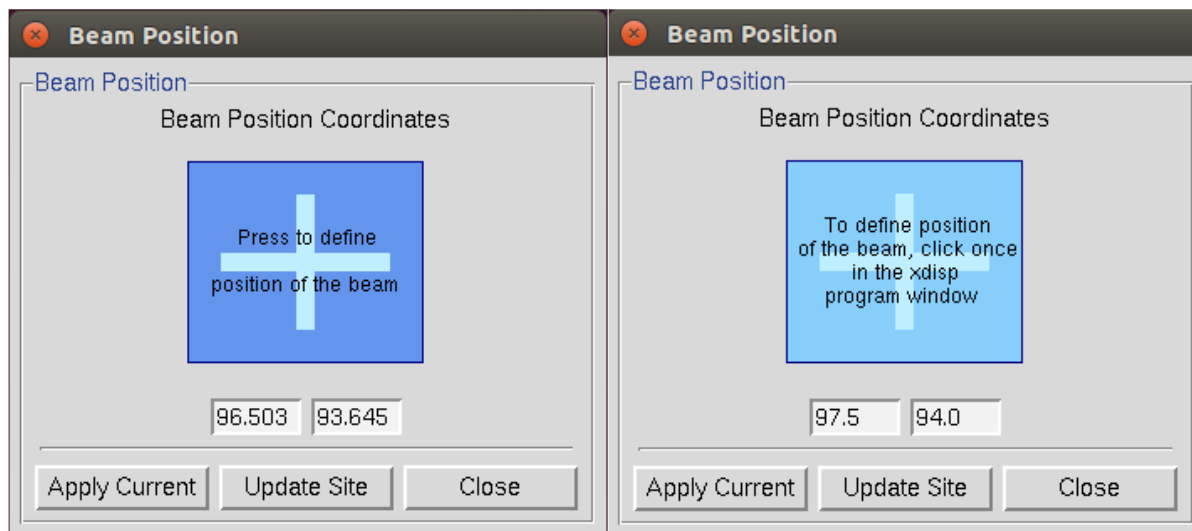


*Figure 124. The beam stop thinned in the center shows the direct beam position.*



*Figure 125. The **SET BEAM POSITION** button opens the beam position dialog.*

The set beam position dialog (Figure 126) provides an interactive method of setting the direct beam position. Clicking on the big blue square will activate the cursor such that the coordinates of wherever you click in the main **Image Display** window will be entered into the fields for the new **X** and **Y Beam Positions**. At this time, only the main **Image Display** window is active for this operation, not the **Zoom** window. These new positions will take effect after you click on the button that says **APPLY CURRENT**. Clicking here does two things. First, it enters new **Beam Position** coordinates into the refinement, and second, it gives a **GO** command to *Denzo*, which starts the refinement (equivalent to clicking the **REFINE** button). This allows you to see the effect of the new **Beam Position** on the predicted reflections right away. (You may have to click **ABORT REFINEMENT** and try to **INDEX** again.) You can also enter these new values into the **Site Configuration** file by clicking the **UPDATE SITE** button. However, just as with the Site Configuration dialog, you can only update the site if you have permission to write to the site definition file. Even if you do not update the site file, the new beam coordinates will be in use until you close the program. If you think that the beam position has changed, it may be prudent to create a new site with the updated beam position.



*Figure 126. The Beam Position Dialog in the normal and activated state.*

- ii. Specification of the **detector format** (believe it, it's happened).
- iii. The crystal-to-detector **DISTANCE**. The distance should be known accurately (within about 1%) and can be measured with a tape measure as a last resort.
- iv. The **Detector Rotation X** and **Detector Rotation Y** values. Misset **Detector Rotation X** and **Detector Rotation Y** values may be due to  $2\theta$  angles other than zero.
- v. **Oscillation range**. Very rarely auto-indexing will fail because too large or too small an oscillation range was collected. In the case of **large** oscillations, the lunes may severely overlap, and the auto-indexing program will have a problem with the non-uniqueness of the mapping from the detector to reciprocal space. In the case of **small** oscillations or stills collected on non-mosaic crystals with a very parallel beam, too few reflections may be present to describe a lune. As a rule of thumb, the following types of oscillation ranges should be used:
  - **viruses:**  $0.25^\circ$  oscillations,
  - **proteins:**  $1.0^\circ$  oscillations,
  - **small molecules:**  $2.5^\circ$  -  $5.0^\circ$  oscillations

There is a fairly wide margin on either side, by at least a factor of 2. The main determinant of the oscillation range is whether or not the diffraction spots are well resolved.

## APPENDIX – TROUBLE SHOOTING

Knowing the detector parameters only makes sense. If you want to deduce some unknowns (crystal lattice, cell and orientation parameters), you need some ‘knowns’ too.

- B. Auto-indexing parameter causes of indexing failure:** Before auto-indexing *Denzo* edits the peak list produced by the Peak Search algorithm of the *XdisplayF* program. Problems with the auto-indexing parameters account for a few of the remaining problems with auto-indexing. These may have to be adjusted in the case of non-single crystals, very large or small unit cells, lots of noise peaks, or ice rings. These parameters (with the Denzo keywords in parentheses) are the longest vector (**LONGEST VECTOR**), resolution limits (**RESOLUTION LIMITS**), refinement sigma cutoff (**WEAK LEVEL**), integration box size (**BOX**) and integration box spot size (**SPOT**). All of these parameters except for the longest vector can be changed through the *HKL-2000* window (the longest vector may be changed via a macro on the *Macros* tab).
- i. The **LONGEST VECTOR** may need to be adjusted if the peaks are sparse due to a small unit cell or low mosaicity. In this case, it should be set to equal the longest vector of the reduced primitive real space unit cell. If you did not understand that last sentence, simply use the longest unit cell axis (which is the same in primitive space groups, anyway). Better to err on the side of larger rather than smaller values.
  - ii. **Resolution limits** can be adjusted, if you have:
    - An **ice ring**, which interferes with the indexing. You can set the resolution limits to, say, 100 - 4.0 Å to eliminate these reflections.
    - A **small satellite crystal**, which contributes some reflections at low resolution but not at high resolution. You can set the resolution limits to, say, 8 - 2 Å, thereby excluding the strongest contributors from the contaminating lattice.
    - A **cracked crystal**, where the two lattices are the same space group only 2-3 degrees apart. You can often adjust the resolution limits, as in the case of an ice ring, so that you index at low resolution, where the two lattices are approximately identical and then use manual refinement to zero in on both lattices. The idea here is to index each lattice separately and to combine the data later in *Scalepack*.
    - A **twinned crystal**, where the two lattices have a non-random orientation relative to one another. The program may fail spectacularly by finding a “**superlattice**,” which indexes both lattices simultaneously but cannot be used to integrate the data because the superlattice does not usually fit both lattices exactly. Then it's time to call an expert or grow new crystals.
  - iii. **Resolution sigma cutoff.** If you have a number of spurious peaks the image distributed randomly over the image, you can adjust the value of the resolution sigma cutoff. This will affect the number of peaks accepted by the auto-indexing routine. By increasing the resolution sigma cutoff, you can exclude weak, spurious peaks,



## APPENDIX – TROUBLE SHOOTING

then run auto-indexing, and examine the superimposition of the auto-indexing predictions (green circles) on the image. If most of them cover real reflections and not noise, then you are succeeding.

- iv. **Box** and **spot** sizes may have to be adjusted if too many spots are being removed by the overlap or background rejection criteria. Auto-indexing does not use spots which *Denzo* determines are overlapping.

C. **Peak search causes of indexing failure.** Finally, you can try changing the number of peaks you select in the peak picking routine. However, since *Denzo* edits the list of peaks anyway, this is unlikely to yield much improvement if the suggestions above do not work. In that case, *do you have a better crystal?*

**2. Misprediction** results in the location of reflection masks not corresponding to the positions of the Bragg peaks. It is visibly obvious and is disastrous. It may be due to:

- Mis-indexing,
- Forcing the wrong space group symmetry, or
- Serious detector misconfiguration or malfunction.

**Mis-indexing causes of mispredictions.** Auto-indexing is based on the assumption that the diffraction spots are correctly mapped from detector coordinates to diffraction (reciprocal) space. The origin of the diffraction space is defined by the position of the direct beam on the detector. A substantial error in the beam position can shift the indexing of the diffraction pattern by an integer vector. Such misindexing can be totally self-consistent until the stage when symmetry-related reflections are compared. For any assumed (starting) value of the beam position, the origin of the diffraction space during indexing will be shifted to the nearest grid point of the best primitive lattice. An initial error in the direct beam position by 0.48 times the distance between reflections will lead to correct indexing, while an error of 0.52 times the same distance will cause a misindexing of the diffraction pattern by one index. **Misindexing by one is never corrected by the subsequent refinement** of the crystal and detector parameters. Misindexing often produces a poor agreement between the predicted and the observed positions of the reflections, but for some crystal orientations, the agreement between the predicted and the observed positions can be equally good for both correctly indexed and misindexed cases. This property of the diffraction geometry creates a potential trap for the unwary crystallographer.

Sometimes the **crystal's asymmetric unit** may have molecules related by an approximate translation by a fraction (typically one half) of a unit cell edge or diagonal. The resulting diffraction pattern will have odd index reflections much weaker than even index reflections. Auto-indexing may find one of the two possible solutions, the choice depending on whether odd reflections are weak enough to be assumed (within experimental error) systematically absent or not. This depends on the fraction of odd reflections in the peak search. If there are only a few odd reflections, then most of the peak search result may be explained with a smaller real-space unit cell. To prevent auto-indexing from finding such a smaller cell, you should enhance the fraction of odd

## APPENDIX – TROUBLE SHOOTING

reflections used in auto-indexing by changing the peak-search criteria or by using only high-resolution reflections in auto-indexing. If you still cannot index odd reflections, then you should consider ignoring them altogether. In such a case, you can solve the structure in a smaller unit cell, and the resulting structural error will not be very significant if odd reflections are much weaker than even ones.

- 3. Kind-of-failed** means that a lattice other than primitive triclinic was predicted, but it had a relatively high distortion index, say 1-1.5% (the lattice distortion index is described in more detail under “

## APPENDIX – TROUBLE SHOOTING

Choosing the correct lattice” on page 58). This may not be a “failure” but simply an indication that further refinement of parameters is needed, because the initial detector parameters in the Site Configuration were very far off. You should just run a few cycles of refinement in *P1* and change **Bravais Lattice**. Unfortunately, misindexing may give similar results. A quick check of the superposition of the predicted spot positions on the reflections should tell you if the refinement is likely to succeed. Note that refinement may converge to low  $\chi^2$  values even with misindexed crystals, but eventually the problem will surface and you will have to index correctly. This is now less of a problem with the **3D Window** option because the requirement to integrate with consistent values over a range of frames is much more stringent a test. You are much less likely to refine misindexed data to low  $\chi^2$  values using a multiple-frame **3D Window**.

Incorrect parameters in the Site Configuration can cause predictions to not line up with the reflections. For example, the spindle rotation direction, or even orientation, may be specified incorrectly – or the Frame Width. An easy way to test for this is to reduce the size of the 3D Window from 5 frames (the default for CCD detectors) to 1 frame, hit REFIN, and note the  $\chi^2$  values again. If they improve (move closer to 1.0 and change to green in color) this tells you that there is a problem with specifying the relationship between successive oscillation frames.

Two parameters affect the relationship between successive frames:

- a. The most likely candidate is an incorrect Frame Width (perhaps the header was written incorrectly). This can be changed in the Frame Geometry panel of the main page.
- b. A second candidate, especially with data collected from customized goniostats, is that the spindle rotation direction is incorrectly specified. To change the goniostat misalignment parameters, go to the Site Configuration box and change the sign of the motor axis value. For example, if the axis is specified as 0 1 0, you could try changing it to 0 –1 0. Changing the order of the numbers, e.g., to 1 0 0 means that you are specifying a different rotation axis orientation, a severe error

- 4. Predicted positions agree perfectly with the peaks** and no detector or diffraction artifacts are visible, **but data scale poorly** and produce many outliers. This problem may be either a simple mistake in data processing (such as using the wrong file format or wrong goniostat definitions) or a non-uniform exposure during crystal oscillation.

The non-uniform exposure may be caused by spindle motor backlash, shutter malfunction (opening too early or too late), ionization chamber electronics failure (if used), decay or variation of the X-ray beam intensity (if an ionization chamber is not used), variable speed of the spindle motor, etc. Non-uniform exposure is best diagnosed by exclusion of other problems that may affect data quality. Graphical feedback provides confidence that the problem cannot be at the indexing/integration stage.

Large variations in absorption of X-rays by the crystal will make data scale poorly and will produce a visible variation of the background; however, it will not affect positional

## APPENDIX – TROUBLE SHOOTING

agreement. The variation in the absorption can be avoided easily by properly mounting the crystal. The correction for absorption is a whole field in itself.

Errors in detector orientation will produce a distorted mapping from the detector to reciprocal space. A wrong crystal-to-detector distance will increase (or decrease) the apparent reciprocal space separation between Bragg peaks. This error will not be uniform in three directions; in the first approximation, along the beam direction the error will be proportional to the square root of the error in the distance, and in the other two directions the error will have linear dependence. In most cases, a significantly wrong crystal-to-detector distance (say an error of 10%) will not make the auto-indexing step fail immediately; however, the calculated unit cell will be quite wrong. The length of the unit cell along the beam direction will be 5% shorter, in the perpendicular directions 10% shorter. If the crystal is diagonally oriented (no principal axis along the beam direction) then the apparent angles between axes will violate the lattice symmetry. Incorrectly defined angles between the detector and the beam will result in incorrect angles between crystal axes. Therefore, the interpretation of the lattice symmetry is dependent, to some extent, on how precisely the detector position parameters are known *a priori*.

## Appendix – Indexing and Integration Keywords

This is a listing of all of the Denzo keywords and their modifiers, if any:

AIR ABSORPTION LENGTH	FORMAT	PRINT (continued)
ANGULAR OFFSET	<i>dectector specific info</i>	<i>solver; no solver</i>
BACKGROUND	GO	<i>statistics; no statistics</i>
<i>radius; elliptical</i>	IBOX	PROFILE FITTING RADIUS
BAD VALUE	IGNORE	QUIT
BOX	IMAGING PLATE	RADIAL OFFSET
<i>print</i>	INPUT / @	RASTER
CALCULATE	INTENSITY AVERAGING	RAW DATA FILE
CASSETTE	WINDOW	RECORD LENGTH
<i>flat; radius;</i>	LAMBDA	REFINE PARTIALITY
<i>rotx, roty, rotz</i>	LATTICE	REFINE NO PARTIALITY
CELL	LEFT MARGIN	REJECT
COUPLING	LIST	<i>fraction</i>
CROSSFIRE	LONGEST VECTOR	<i>cutoff</i>
<i>x; y; xy</i>	MONOCHROMATOR	<i>slope</i>
CRYSTAL	<i>filter</i>	<i>systematic</i>
<i>rotx, roty, rotz</i>	<i>graphite</i>	RESOLUTION LIMITS
DISTANCE	MOSAICITY	SECTOR
EIGENVALUE FILTER	NO BEAM	SKEW
END FIT	NUMBER OF RECORDS	SPACE GROUP
END OF JOB	OMIT	SPINDLE AXIS
END OF PACK	<i>radius</i>	SPOT
ERROR	<i>elliptical</i>	<i>radius</i>
<i>density</i>	ORIENTATION AXIS 1	<i>elliptical</i>
<i>overload</i>	ORIENTATION AXIS 2	START REFINEMENT
<i>partiality</i>	OSCILLATION	STOP
<i>positional</i>	<i>start</i>	TITLE
<i>systematic</i>	<i>end</i>	TOP MARGIN
<i>y/x factor</i>	<i>range</i>	UNIT CELL
FIDUCIALS	<i>step</i>	USE PARTIALS POSITION
FILM	OVERLAP	USE BEAM
<i>absorption</i>	<i>spot</i>	USE FIDUCIALS
<i>length</i>	<i>guard</i>	USE NO FIDUCIALS
<i>output file</i>	<i>none</i>	USE PARTIALS POSITION
<i>rotation</i>	OVERLOAD VALUE	USE PARTIALS NO
<i>width</i>	PLATE ROTATION	POSITION
FIT/FIX	PEAK SEARCH FILE	VERTICAL AXIS
<i>all; a*; b*; c*;</i>	peaks.file	WAVELENGTH
<i>alpha*; beta*; gamma*;</i>	POLARIZATION	WEAK LEVEL
<i>crystal rotx roty rotz;</i>	PRINT	WRITE PREDICTIONS
<i>cassette rotx roty rotz;</i>	<i>chi square; no chi square</i>	WRITE NO PREDICTIONS
<i>distance; x beam; y beam;</i>	<i>fitted</i>	X BEAM
<i>y scale; skew ;</i>	<i>hkl</i>	Y BEAM
<i>film rotation; radius</i>	<i>memory; no memory</i>	Y SCALE
<i>crossfire x y xy</i>	<i>profiles; no profiles</i>	YORK OUTPUT FILE
	<i>shifts; no shifts</i>	

## AIR ABSORPTION LENGTH

Length of air or helium, in mm, over which the scattered X-rays are attenuated by a factor of  $e$  (2.718)

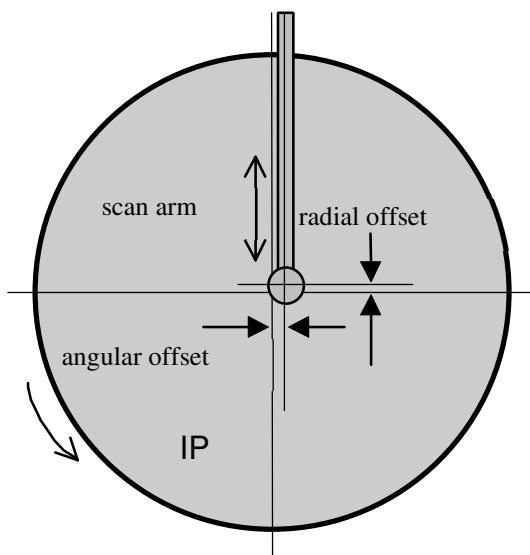
<i>format</i>	AIR ABSORPTION LENGTH <i>value_</i> (mm)	
<i>default</i>	AIR ABSORPTION LENGTH 860	
<i>example</i>	AIR ABSORPTION LENGTH 200	[for 2.5 Å X-rays]

## ANGULAR OFFSET

This is one part of the alignment correction needed for the spiral MacScience and MAR scanners. It is the distance, in mm, that the reading head is offset in the non-radial scanning direction from the center of the imaging plate. See also RADIAL OFFSET.

<i>format</i>	ANGULAR OFFSET <i>value_</i> (mm)	
<i>default</i>	ANGULAR OFFSET 0.0	
<i>example</i>	ANGULAR OFFSET 1.2	

## BACKGROUND



**Figure 127. Parameters of a reflection**

Defines the shape and size of the background area where the background measurement takes place. Usually defined to be outside of the spot area. However, when no background is specified, the default area is the area just outside of the spot border.

## APPENDIX – INDEXING AND INTEGRATION KEYWORDS

<i>modifiers</i>	<i>RADIUS</i>	Defines the background to be circular with a radius value, in mm. The center of the background area and of the box coincide.
	<i>ELLIPTICAL</i>	Defines the background to be ellipsoidal with characteristic major and minor semi-axis lengths (in mm), along with the clockwise angle between a horizontal line and the major semiaxis. Refer to the diagram under the keyword SPOT.
<i>formats</i>	BACKGROUND radius	<i>value_(mm)</i>
	BACKGROUND elliptical	<i>value-major(mm) value-minor_(mm) value-angle_(degrees)</i>
<i>default</i>		a ring just outside of the spot border
<i>examples</i>	BACKGROUND radius	0.45
	BACKGROUND elliptical	0.80 0.70 0.0

---

### BAD VALUE

The value in the image file which does not represent a measurement. Typically 0. When set to 40000 it disables 0. Used only on DIP series scanners to process small molecule data.

### BOX

Defines the size of the box where both the spot and the background are measured, given as two lengths, in mm. If the BOX size is larger than the average separation between spots (i.e., the boxes overlap), then the program will be slowed down by the spot overlap calculation. The current limit on the box size is 90 pixels in each direction, for a total of 8100 pixels.

The box size can alternatively be specified by the IBOX, where the size is given in pixels instead of mm.

---

<i>modifier</i>	<i>PRINT</i>	tells the program to print the spot definition at the beginning of every refinement cycle (i.e. for every GO statement)
<i>formats</i>	BOX	<i>value1_(mm) [slow scan direction] value2_(mm) [fast scan direction]</i>
	BOX print	<i>value1 value2</i>
<i>default</i>		none
<i>example</i>	BOX	2.4 2.4

---

## APPENDIX – INDEXING AND INTEGRATION KEYWORDS

### CALCULATE

This command tells the program that it is the last cycle, when the spot intensities are calculated and written to the film output file. It is followed by the command GO. After the cycle both the RAW DATA FILE and the FILM OUTPUT FILE are closed. Signals the SECTOR argument to be incremented by 1.

### CASSETTE

Defines the orientation and the shape of the detector, historically, a film cassette.

modifier	<i>ROTZ, ROTY, ROTX</i>	the three angular deviations from zero
	<i>RADIUS</i>	radius of the film in a curved cassette
		The radius of curvature does not have to be equal to the film-to-crystal distance and may be different than the radius of the cassette due to the thickness of the film or imaging plate.
	<i>FLAT</i>	turns off radius
formats	CASSETTE rotx value1 roty value2 rotz value3_(degrees)	
	CASSETTE radius value1_(mm)	
default	all angles = 0.0	
example	CASSETTE rotx 2.0 roty -1.0	

You rarely have to enter these values since the program can refine them once the crystal parameters are established. However, some computer time can be saved if previously determined values are used. Usually, they change significantly only after the camera is realigned. There is no limit on values.

The cassette rotations are the primary means by which a non-zero  $2\theta$  angle is specified (along with DISTANCE, X BEAM, and Y BEAM). For example, a  $2\theta$  setting of, say, +20 degrees (R-Axis convention), would be described by CASSETTE rotx -20.0 (note change in sign). The DISTANCE will have to be changed as well since it is the distance from the crystal to the point where it intersects the IP surface. This can be calculated by the formula: Distance (*Denzo*) = distance (R-Axis) /  $\cos(2\theta)$ . X BEAM and Y BEAM are the position of the direct beam on the IP during exposure, not at  $2\theta = 0$ . For large values of  $2\theta$  the Y BEAM may be off the detector. The correct value may be negative or more than 200 in such cases.

CRYSTAL rotz and CASSETTE rotz are strongly correlated and **cannot be refined simultaneously**.

### CELL

Alias for the keyword UNIT CELL



## APPENDIX – INDEXING AND INTEGRATION KEYWORDS

### COUPLING

This value reflects how fast the detector moves during exposure for Weissenberg images. Given as mm per degree. This is the **inverse** of the Photon Factory parameter degree/mm, and is only useful for Weissenberg images.

---

<i>format</i>	COUPLING <i>value_(mm/degree)</i>
<i>default</i>	0.0
<i>example</i>	COUPLING 0.057

---

### CROSSFIRE

Crossfire is a measure of the X-ray beam divergence and focusing as it leaves the collimator and illuminates the crystal. Crossfire, being a symmetric tensor, has x, y, and xy components. It affects the prediction of partial reflections and their positions, not their angular width. It is expressed as angular divergence of the beam. The default value is zero crossfire, i.e., a perfectly parallel beam or a beam focused on the detector.

---

<i>modifiers</i>	<i>X</i>	angular spread of the beam in the x-direction
	<i>Y</i>	angular spread of the beam in the y-direction
	<i>XY</i>	correlated xy component of the beam spread. This tends to be zero within error.
<i>format</i>	CROSSFIRE x <i>value1_(arbitrary_units)</i>	
	CROSSFIRE y <i>value2_(arbitrary_units)</i>	
	CROSSFIRE xy <i>value3_(arbitrary_units)</i>	
<i>defaults</i>	0 0 0	
<i>example</i>	CROSSFIRE x 0.023 <i>value taken from previous refinement</i>	
	CROSSFIRE y 0.345	
	CROSSFIRE xy 0.0004	
	CROSSFIRE x 0.023 y 0.345 xy 0.0004	

---

### CRYSTAL

Defines the crystal orientation for the calculation of predicted spots.

---

<i>modifiers</i>	<i>ROTZ, ROTY, ROTX</i> the three angular deviations from the reference orientation specified by SPINDLE AXIS and VERTICAL AXIS. See the diagram and further description in Denzo orientation conventions.
<i>format</i>	CRYSTAL rotz <i>value1</i> roty <i>value2</i> rotx <i>value3_(degrees)</i>
<i>defaults</i>	0 0 0
<i>example</i>	CRYSTAL rotx 2.34 roty -4.566 rotz 34.560

---

***DISTANCE***

The distance from the crystal to the beam spot on the film, in mm, along the X-ray beam path, regardless of the  $2\Theta$  angle.

---

<i>format</i>	DISTANCE <i>value_(mm)</i>
<i>default</i>	DISTANCE 75.0
<i>example</i>	DISTANCE 155.453

---

***EIGENVALUE FILTER (USE WITH GREAT CAUTION)***

When attempting to simultaneously refine two or more highly correlated parameters with *Denzo*, an eigenvalue filter is triggered, and an error message appears in the log file stating that the parameters being fit were too highly correlated to be fit simultaneously. The eigenvalue filter value provides a way of defeating this check, by lowering the value. Unfortunately, just because you change the value doesn't mean that the correlated parameters are going to become uncorrelated, so changing the value is unlikely to help you and may lead to some horrible result.

---

<i>format</i>	EIGENVALUE FILTER value
<i>default</i>	0.001
<i>example</i>	EIGENVALUE FILTER 0.0005

---

***END OF JOB***

Signals the end of the *Denzo* job. Placed at the end of the command file or used to exit from *Denzo* during manual fitting and indexing.

***ERROR***

The keyword ERROR and its modifiers define the model for the error in the measurement of the spot positions. **The default values should be adequate in most cases**, and you should not worry too much about them. The error model does, however, directly affect the  $\chi^2$  values printed out at the end of each cycle of refinement. Since this is usually the only check you can reasonably make as to the success of your data reduction in *Denzo*, it is wise to know something about how these are calculated. The names of the modifiers contributing to the error definition should not necessarily be taken literally since they are shorthand notation for complicated effects on the accuracy of the measurement. The model has been developed and black boxed (i.e., given the Z.O. Good Housekeeping seal of approval) to give a realistic estimate of the weight each spot should contribute to the refinement of the crystal and camera coordinates. This model gives  $\chi^2$  values close to one

## APPENDIX – INDEXING AND INTEGRATION KEYWORDS

for good (non-mosaic) crystals. For mosaic crystals either error positional needs to be increased or higher values than 1 of  $\chi^2$  should be expected.

<i>modifiers</i>	<i>POSITIONAL</i> The <b>minimum</b> error, in mm, in the positional measurement. It is usually a fraction of the pixel size, typically one half or one-quarter of a pixel. Its squared value is added to the other components of the error estimate. Enlarging this value will lower your $\chi^2$ values, but will not necessarily improve your data.	
	<i>DENSITY</i> Models the error measurement of the optical density of the film. The error equals the square root of the corrected reading times the density value. This is used only to estimate the error in spot positions and in the definition of strong reflections. The error density is affected by the voltage on the photomultiplier. The higher the voltage, the higher the error density, with a very steep dependence.	
	<i>OVERLOAD</i> Refers to the increase in the modeled positional error due to the existence of pixels with intensity overflow. For every overflowed pixel the modeled error squared is increased by the overload value squared.	
	<i>Y/X FACTOR</i> This reflects the increase in the positional error of the spot due to its being close to the oscillation axis. The error estimate is by $1 +$ the ratio of y/x factor to the absolute value of the tangent of the angle between the oscillation axis and the line connecting the spot with the beam spot. This prevents the spot distortion that occurs close to the oscillation axis from affecting the refinement very much.	
	<i>PARTIALITY</i> the expected error of the partiality determination The value for the error is expressed as a decimal, e.g., $\pm 10\%$ would be written 0.1.	
	<i>SYSTEMATIC</i> the weight given the positional refinement compared to the partiality refinement when refining crystal orientation parameters. The higher the value, the more weight is given to the positional refinement. Typical values range from 1 to 20. A value of 5 seems to work almost all the time.	
<i>format</i>	ERROR modifier1 value modifier2 value ...	
<i>defaults</i>	overload	0.005
	density	detector and wavelength dependent
	positional	0.030
	y/x factor	0.1
	partiality	0.10
	systematic	5.0

## APPENDIX – INDEXING AND INTEGRATION KEYWORDS

*example*      ERROR systematic 5.00 partiality 0.100 positional 0.050

Avoid modifying the error parameters other than positional and density. Moreover, error density should be constant for a detector (site) and changed only with caution.

### FIDUCIALS

Defines the position of the fiducial marks on the film relative to the beam spot, before FILM rotation and Y SCALE are applied. Given as a series of x,y coordinates, in mm. These days this is only used with Photon Factory Weissenberg images.

<i>format</i>	FIDUCIALS    xvalue_(mm)    yvalue_(mm)    xvalue_(mm) yvalue_(mm) ...
<i>default</i>	none
<i>example</i>	FIDUCIALS -90.0 -90.0 90.0 -90.0

### FILM

This keyword tells the program about the characteristics of the data collection medium. Identical to the keyword IMAGING PLATE.

<i>modifiers</i>	<b>ABSORPTION</b> Total absorption of the film. Default is 100 for image plates
	<b>LENGTH</b> Distance from the top of the detector to the last useful pixel in the <b>fast</b> direction. Depending on the detector, on the computer display it could be the apparent width of the detector.
	<b>WIDTH</b> Width of the detector in the <b>slow</b> direction. Depending on the detector, on the computer display it could be the apparent height of the detector.
	<b>ROTATION</b> Describes the orientation of the <b>spindle</b> relative to the detector. The value of the FILM rotation is a function of which file storage convention (see figure 5 in <i>Detector Conventions analysis</i> ) is being used. For the manual scanner, adding 180° to film rotation can compensate for putting plate upside-down into the scanner. Small changes (refined by <i>Denzo</i> ) to film rotation can compensate for small rotation slop that is inherent when manually handling plates.
	<b>OUTPUT FILE</b> This gives the name of the processed output file
<i>formats</i>	FILM modifier1 <i>value</i> ...

## APPENDIX – INDEXING AND INTEGRATION KEYWORDS

	output file format is a string of characters
<i>defaults</i>	absorption 100 length <i>detector dependent</i> rotation <i>detector dependent</i> output file (none)
<i>examples</i>	FILM rotation 270.0 FILM absorption 2

---

### ***FIT, FIX, END FIT, END FIX***

These command words tell the program to FIT (refine) or FIX (not refine) a parameter specified by a *Denzo* keyword. The FIT command is followed by the command GO, which calls one cycle of refinement. One GO command calls one cycle of refinement, 5 GO commands call for 5 cycles of refinement, etc. If a keyword parameter is fixed, it will retain its current value through the refinement. **The program assumes that all of the keyword parameters are fixed at the beginning of every new frame.**

The effect of these commands is cumulative and carries over through subsequent cycles of refinement. For example, if the command FIT CRYSTAL rotx roty rotz GO is given, then on every subsequent refinement cycle (called by the GO statement), the CRYSTAL rotx roty and rotz parameters will be refined. Typically a refinement starts by fitting just a few parameters, like CRYSTAL rotx roty rotz, allowing the refinement to proceed towards convergence for a few cycles, and then fitting more parameters a few at a time. The keyworded parameters that can be FIT are given below.

All of the keyworded parameters can be refined simultaneously, so care should be taken not to fit parameters which are too strongly correlated at low resolution. For example, the crystal to film distance and the unit cell lengths are highly correlated, so one should be allowed to refine before refining the other (usually the unit cell lengths are refined first).

The END FIT and END FIX command words mark the end of a list of fitted parameters. This is especially useful for inputting more parameters after you have called a FIT command.

Consider the following case: you want to refine the crystal rotation values, so you type FIT CRYSTAL rotx roty rotz. Now let's say that you also want to enter a new value for the distance before giving a GO command. You might think to try FIT CRYSTAL rotx roty rotz DISTANCE 200 GO, but this won't work! The program will not recognize the 200, and it will give you an error message, plus it will try to refine the distance value as well. The solution is to use the END FIT command, as follows: FIT CRYSTAL rotx roty rotz END FIT DISTANCE 200 GO. The same goes for the END FIX command.

The option FIT cell, when used in conjunction with the SPACE GROUP or LATTICE command, tells the program to fit the unit cell lengths and angles. The program is smart enough to know which lengths and angles are to be fit, and which are to be fixed. For

## APPENDIX – INDEXING AND INTEGRATION KEYWORDS

example, it knows enough to keep all angles at 90 degrees in an orthorhombic space group. Likewise, it also knows enough to keep  $a = b$  in tetragonal space groups, etc. The LATTICE or SPACE GROUP must be declared before FIT CELL can be used.

In practice, the most often option used is FIT all. This will fit all parameters that can be sensibly fit given the circumstances (space group, detector). Fitting all parameters is now stable even at low resolution.

---

	FIT- and FIX- able keywords
	all, a*, b*, c*, alpha*, beta*, gamma*, cell, CRYSTAL rotx roty rotz, CASSETTE rotx roty rotz, radius, DISTANCE, X BEAM, Y BEAM, Y SCALE, SKEW, FILM rotation, CROSSFIRE x, y, xy
<i>format</i>	FIT KEYWORD modifier
<i>default</i>	all keyword parameters fixed (i.e., type GO and nothing will be refined)
<i>examples</i>	FIT CRYSTAL rotx roty rotz FIT X BEAM Y BEAM Y SCALE SKEW DISTANCE GO GO GO

---

**Do not run alternative cycles of fit and fix commands!** This is the old way and does not take advantage of the fact that the *Denzo* refinement is very stable and can refine many parameters simultaneously. Simultaneous refinement is the best way to avoid getting into a false minimum.

### FORMAT

A critical keyword. The modifier following this keyword tells the program about a series of defaults to using for the detector parameters. This keyword can thus substitute for many of the parameters called by, for example, the FILM (or IMAGING PLATE) keyword.

A detailed table of all the keywords specified by the format keyword can be found in the *Defaults Values* and *Detector Specific Information* appendixes.

---

<i>modifiers</i>	Detector Specific Information
<i>format</i>	FORMAT modifier
<i>default</i>	this is a required input!
<i>example</i>	FORMAT dip2000

---

### GO

This command tells the program to execute one iteration cycle of refinement. After the cycle the fitted parameters will have new values. You can specify more than one GO in

## APPENDIX – INDEXING AND INTEGRATION KEYWORDS

one line. This will tell *Denzo* to print results of the fit only for the last cycle from the series of GOs. See also FIT/FIX.

### **IBOX (REDUNDANT)**

IBOX is exactly the same as the keyword BOX, except that instead of giving the dimensions of the box in mm, the dimensions of the box are given in pixels. The I in IBOX stands for *integer*. The maximum dimensions of the box are 90 pixels in each direction.

---

<i>format</i>	IBOX <i>integer integer_(pixels)</i>
<i>default</i>	none
<i>example</i>	IBOX 25 25

---

### **IGNORE**

Tells the program to ignore a specific region of the X-ray image. For example, this is useful for excluding the shadow due to the beam stop. Don't try to use this for masking out ice rings. That is accomplished using the REJECT fraction keyword. It is useful to think of IGNORE as describing a masked-out area of the film. The mask is provided by a series of geometrical descriptions, followed by their coordinates. For example, the beam stop shadow could be masked out by a combination of a quadrilateral and a circle (i.e., a lollipop shape).

---

<i>modifiers</i>	<i>QUADRILATERAL</i> A quadrilateral as defined by four <i>x,y</i> coordinates, in mm, picked in a <b>counterclockwise</b> (if Y SCALE > 0) or <b>clockwise</b> (if Y SCALE < 0) direction. The best way to do this is to use the cursor to pick the points off of the display window. The <i>x, y</i> coordinates in mm are listed in the bottom row of the green box of the display.
	<i>CIRCLE</i>
	<i>TRIANGLE</i>
<i>format</i>	IGNORE
<i>default</i>	nothing ignored      defined by the <i>x,y</i> coordinates of the center, and the radius of the circle, all in mm. defined by three <i>x,y</i> coordinates. Use the display program to pick these.
	quadrilateral <i>x1 y1 x2 y2 x3 y3 x4 y4</i> (all values in mm)
	circle <i>x5_(mm) y5_(mm) r_(mm)</i>
	triangle <i>x6 y6 x7 y7 x8 y8</i> (all values in mm)
<i>example</i>	IGNORE quadrilateral 50 0 52 0 52 100 50 100

---

## INPUT / @

This keyword is the same as the @ keyword. Signals to the program to change reading of the control data to an auxiliary file. At the beginning of the program, the auxiliary file is closed, so the keyword INPUT has to be followed by the filename of the auxiliary file. The auxiliary file is read until a semicolon is reached (or an end of file mark is detected). The control is then returned to the original input stream. In the case of reaching a semicolon, the auxiliary file remains open, and the next time the keyword INPUT appears it must not be followed by a file name or else the program will become confused since it still wants to read from the open auxiliary file. On the other hand, if control is returned due to an end of file mark, then the next time input is encountered, it must have an auxiliary file name next to it. If you give it the same auxiliary file name as before, then it will begin reading the auxiliary file from the beginning. Nesting of auxiliary files is not allowed. Auxiliary files are useful to keep the data common to many films in a separate file.

---

<i>format</i>	INPUT <i>filename</i>
<i>default</i>	none
<i>examples</i>	INPUT subbatch1.dat @ refine

---

## INTENSITY AVERAGING WINDOW

Sets the number of reflections in a resolution shell used to define that resolution shell for the purpose of refining partiality. Default is 100.

## LAMBDA (REDUNDANT)

X-ray wavelength. Same as keyword WAVELENGTH.

---

<i>format</i>	wavelength in Ångstroms	
<i>default</i>	1.5418	[Cu Kα]
<i>example</i>	lambda 1.5418	

---

## LATTICE (REDUNDANT)

Lattice type. Note that this command is superseded by the SPACE GROUP command for most routine use.

---

<i>format</i>	single letter or single letter with name. Can also specify lattice type. Possible choices are: P, C, A, B, F, I and R
<i>default</i>	P triclinic

---



*example*      LATTICE P monoclinic

---

## LEFT MARGIN

The left margin, in mm from the left edge, of useful data. It is defined in the raw data convention, not in the display program convention.

---

<i>format</i>	LEFT MARGIN <i>value_(mm)</i>
<i>default</i>	0.0
<i>example</i>	LEFT MARGIN 2.0

---

## LIST

This command is normally used in the manual mode. It tells the program to provide a list of the current parameters. This is convenient if you forget where you were in the refinement, or you want to copy down the latest parameters. To use, just type LIST.

## LONGEST VECTOR

In the autoindexing subroutine of *Denzo*, LONGEST VECTOR denotes the longest real-space vector, in Å, the program expects to find in the reduced, primitive, unit cell. If the autoindexing fails, and you are confident of your X BEAM, Y BEAM, and DISTANCE values, then this is the next parameter to try changing. Typically you would either input a value close to or somewhat larger (say  $\approx 50\%$ ) than what you really suspect is the longest vector in the unit cell.

---

<i>format</i>	LONGEST VECTOR <i>value_(<math>\approx</math>)</i>
<i>default</i>	LONGEST VECTOR $\approx \lambda \cdot \text{distance} / 2 \cdot \text{spot radius}$
<i>example</i>	LONGEST VECTOR 300

---

## MONOCHROMATOR

Defines the polarization of the X-ray beam. Same as the keyword POLARIZATION. This correction is not a big deal except for synchrotron radiation, where it is a minor deal unless you are collecting high angle data. The *Denzo* definition of polarization is as follows:

$$\frac{(I_s - I_p)}{(I_p + I_s)}$$

## APPENDIX – INDEXING AND INTEGRATION KEYWORDS

where  $I_s$  is the intensity of the electric field squared of the X-ray beam along the spindle axis, and  $I_p$  is the intensity perpendicular to the spindle axis. Values range from +1.0 to -1.0.

Generally speaking, this is not a number you can measure very easily and is often calculated theoretically. It is really the responsibility of beamline operators to know this value for their beamlines. Note that the sign of the correction depends on whether the spindle is horizontal or vertical. As a rule of thumb, if the polarization correction is  $> 0$ , then the intensity of the X-ray scattering is stronger perpendicular to the axis of the spindle than it is to the axis parallel to the spindle. You can check the background of your image to confirm this. Typical values range from  $\pm 0.11$  for graphite to  $+ 0.9$  for synchrotron beams (the spindle axis is typically oriented at synchrotrons to make the polarization positive). For a **laboratory source**, use modifiers: graphite (if graphite monochromator is installed) or filter (if X-ray beam is filtered or mirrors are installed).

<i>modifiers</i>	FILTER	No polarization, value = 0
	GRAPHITE	Reflecting plane perpendicular to the plane formed by the spindle and the beam, CuK $\alpha$ and MoK $\alpha$ radiation. Value of $\pm 0.11$ for CuK $\alpha$ or $\pm 0.03$ for MoK $\alpha$ .
<i>format</i>	MONOCHROMATOR modifier	
<i>default</i>	0.0	
<i>example</i>	MONOCHROMATOR graphite POLARIZATION -0.11	

The old keyworded values for the synchrotrons are probably wrong now. Always ask the synchrotron staff what the correct number is. A good guess for a bending magnet source is polarization 0.9. For a wiggler source, it can vary a lot from station to station, in the range of 0.0 to 0.9.

### MOSAICITY

Defined in *Denzo* as the rocking angle, in degrees, in both the vertical and the horizontal directions which would generate all the spots seen on a still diffraction photograph. It includes contributions due to X-ray bandwidth, beam crossfire, etc.

<i>format</i>	MOSAICITY <i>value_(degrees)</i>
<i>default</i>	this is a required input!
<i>example</i>	MOSAICITY 0.5

## NO BEAM

Tells the program not to use the beam spot in the refinement. If the option USE BEAM has been invoked earlier in the refinement, it should be revoked with the NO BEAM keyword in the later stages.

---

<i>format</i>	NO BEAM
<i>default</i>	this is the default option. see USE BEAM

---

## NUMBER OF RECORDS

Specifies the size of the data file. Equal to the number of vertical (fast scanning direction) lines. The default value is now 'hard-wired' into the FORMAT modifier. In other words, once you specify the detector, the NUMBER OF RECORDS value is input from a look-up table.

---

<i>format</i>	NUMBER OF RECORDS <i>integer</i>
<i>default</i>	comes along with whatever FORMAT is specified

---

## OMIT

Defines the area **outside** a specified circle or ellipse, but **within** the box, to be omitted from any use. By default, there is no omitted area.

---

<i>modifiers</i>	<i>RADIUS</i> or <i>RAD</i>	Defines the omit boundary to be circular with a radius value, in mm. The center of the omit boundary and that of the box coincide
	<i>ELLIPTICAL</i> or <i>ELL</i>	Defines the omit boundary to be ellipsoidal with characteristic major and minor semi-axis lengths (in mm), along with the clockwise angle between a horizontal line and the major semi-axis. Refer to the diagram under the keyword SPOT.
<i>formats</i>	OMIT radius value	
	OMIT elliptical valuemajor valuemminor valueangle	
<i>default</i>	no area omitted	
<i>examples</i>	OMIT rad 0.35	
	OMIT cell 0.70 0.60 0.0	

---

## APPENDIX – INDEXING AND INTEGRATION KEYWORDS

### OSCILLATION

This keyword tells the program that the oscillation method is being used.

<i>modifiers</i>	<i>START</i>	Starting oscillation angle		
	<i>END</i>	Ending oscillation angle		
	<i>STEP</i>	Angle between 1st oscillation start and 2nd oscillation start		
	<i>RANGE</i>	The actual oscillation angle		
<i>formats</i>	OSCILLATION	modifier1	<i>value1_(degrees)</i>	modifier2
		<i>value2_(degrees)</i>		
<i>default</i>	end = start + range    step = range			
<i>examples</i>	OSCILLATION range 1.0 start 30.0			
	OSCILLATION start 30.0 end 31.0			
	OSCILLATION range 1.0 step 0.8 start 30.0			

### OVERLAP

Overlap defines which pixels of any spot are checked for overlap with the pixels of a neighboring reflection box. The choices are the spot pixels, the guard pixels (found between the spot circle and the background circle) or none. For example, if you specify OVERLAP spot and if the spot pixels overlap with the spot pixels of another reflection, then the two reflections are flagged as overlaps.

---

<i>modifiers</i>	<i>SPOT</i>
	<i>GUARD</i>
	<i>NONE</i>
<i>format</i>	OVERLAP modifier
<i>default</i>	None
<i>example</i>	<b>OVERLAP</b> spot

---

### OVERLOAD VALUE

The lowest optical density reading considered unreliable due to saturation of the film or image plate. The default value corresponds to the full dynamic range being useful, with only the highest value being considered overflow. A value that is higher than the maximum possible in the data tells the program to ignore the intensity overflow. The range of the values themselves is determined by the analog to digital converter used to write the data. For 8-bit ADC's (found on film scanners) the highest value is  $2^8-1=255$ . For the 16-bit ADC's found on most on-line detectors, the range is from 1 to  $2^{16}-1=65535$ . The particular OVERLOAD VALUE that is considered unreliable has been determined for several detectors and is input as the default when you specify the detector

## APPENDIX – INDEXING AND INTEGRATION KEYWORDS

using the FORMAT keyword. *Denzo* applies the overload table in Mar scanners so that the overload value should represent the highest reliable entry in the Mar overload table.

---

<i>format</i>	OVERLOAD VALUE integer (units are function of detector type)
<i>default</i>	separate values for different IP formats, contained in FORMAT defaults
<i>example</i>	OVERLOAD VALUE 255 (film)

---

- One-of-a-kind detectors, as well as old Mar scanners, may have the wrong default value.
- The overload value is a data representation value before applying the data format dependent correction for strong intensities.
- Default values assume the full dynamic range of the instruments (Mac Science, Mar, R-Axis).

### PEAK SEARCH FILE PEAKS.FILE

Tells the program to perform **autoindexing**.

### PLATE ROTATION (REDUNDANT)

This is exactly the same as the keyword FILM rotation.

### POLARIZATION (REDUNDANT)

Same as MONOCHROMATOR.

### PRINT / END PRINT

The PRINT command tells the program to print the output specified by subsequent keywords to the log file. It does not affect the results of the refinement. Default values of the print sub-options can be of several types. Some options are printed by default at every cycle, some are printed on the last cycle of every film of the refinement, and some are never used unless specified. See specifically the various print descriptions below.

---

<i>modifiers</i>	<i>CHI SQUARE</i> print the $\chi^2$ statistics, including the number of spots fitted, the $\chi^2$ of x and y (a measure of how well the spots were fitted based on what sort of error you expected - see also ERROR), and the predicted decrease in the $\chi^2$ 's after the current cycle of refinement. For those of you who just want numbers, $\chi^2$ 's <b>should be close to 1.0</b> (between 0.5 and 2.0), and the predicted decrease
------------------	---

---

## APPENDIX – INDEXING AND INTEGRATION KEYWORDS

should be near zero, showing that the refinement has converged.

**FITTED** Print to the log file the observed and fitted profiles of those spots having the specified *hkl* value. No density readings in the background will be printed. By default, no fitted spots printed.

**HKL** Print to the log file the density readings and other statistics of those spots having the specified *hkl* value. By default, no *hkls* printed.

**MEMORY** Print statistics of the usage of variables.

**PROFILES** The profiles modifier tells the program to print the average spot profile in a specified number of regions to the log file.

It is best to refrain from printing profiles until the very last cycle of refinement, or else your log file will become truly gargantuan. You can also specify PRINT no profiles at the beginning of your refinement, just to make sure it is **off** until you want to turn it **on**.

**SHIFTS** PRINT shifts tells the program to print the value, the change from the previous value, and the estimated error of all the keyworded parameters fitted on a current cycle of refinement. PRINT no shifts turns this off.

**SOLVER** This little-used command prints out correlation matrices calculated in the refinement.

**STATISTICS** Print the *hkl* statistics, including the number of whole and partially recorded reflections matched with preds.

**ZONES** Print out a list of equivalent zones, given as alternative values for CRYSTAL rotx roty and rotz.

One advantage to invoking this option is to use rotx, roty, and rotz values that are more convenient to remember or type. For example, instead of using crystal rot values of, say, -213.6, 137.4, -13.27, you could use the equivalent values of, say, (I'm just making these up) 18.2, -15.4, 63.9. If you wanted to orient your crystal for a particular data collection scheme, these alternative zones could make life easier. PRINT zones lists all of these equivalent values.

*format:*

PRINT chi square / no chi square

PRINT fitted *integer integer integer*

PRINT *hkl integer integer integer*

PRINT memory / no memory

PRINT profiles *integer integer* (the first integer specifies the number of equal-sized horizontal regions to divide the frame into,

## APPENDIX – INDEXING AND INTEGRATION KEYWORDS

the second specifies the number of equal sized vertical regions. The total number of regions is the product of the first and second integers). The more profiles you specify, the larger the log file will be

PRINT no profiles

PRINT shifts / no shifts

PRINT solver / no solver

PRINT statistics / no statistics

*defaults:* chi squares printed after every cycle of refinement

no fitted spots printed

no hkl's printed

print memory statistics on the last cycle

no profiles printed except on the last cycle, when a 3 x 3 grid is printed all shifts printed no solver

statistics turned on on the last cycle only

*examples* PRINT fitted 12 8 0

PRINT profiles 3 3 (tells the program to divide the frame into 9 equal regions and print out 9 average spot profiles).

PRINT hkl 12 8 0

**You can easily generate gigabytes of log files if you use some of the print options, so be careful!**

### PROFILE FITTING RADIUS

The radius of the area around each particular spot, in mm, containing neighboring spots, used to calculate the average spot profile. The spot in question is fitted to the average profile of all the spots within the specified radius. Generally, **the radius is set so that spots on roughly 35% of the area of the detector are included in the averaging.** Only on the last cycle of the refinement is the profile fitting applied, via the keyword CALCULATE.

The particular size of the area chosen depends on how much the profiles vary across the film and the density of the spots on the film. For example, if the spot profiles vary a lot across the film, then you would choose a smaller radius. If you have a small lattice and the spots are widely separated, you might want to choose a larger radius.

The calculation of the average profile is a time-consuming task, proportional to the number of spots in the profile fitting radius circle, but too small a radius will not capture enough spots and lead to a noisy average profile. Too large a radius will average out significant profile variations.

To see the PROFILE FITTING RADIUS in *ImageWindow* of *XdisplayF*, click **PROF\_FIT\_R**. A thin white circle corresponding to the profile fitting radius will appear and will move around with the cursor; only those spots whose intensity is above the weak

## APPENDIX – INDEXING AND INTEGRATION KEYWORDS

level will have predictions displayed in this mode and only these reflections are used in the calculation of the average spot profile (to return to the normal mode, click **PROF\_FIT\_R** again).

A good rule of thumb is that the PROFILE FITTING RADIUS should enclose about 10 to 50 spots *whose intensities are above the weak level*. Useful increments of the profile fitting radius are 2.5% of the detector size (e.g., 5 mm for a 200 mm wide IP).

---

<i>format</i>	PROFILE FITTING RADIUS value_(mm)
<i>default</i>	20.0
<i>example</i>	PROFILE FITTING RADIUS 30.0

---

### QUIT

Exit from program *Denzo*. Same as STOP or a semicolon (;).

### RADIAL OFFSET

This is one part of the alignment corrections needed for the spiral MacScience and MAR scanners. It is the radial departure, in mm, of the scanning heads from perfect concentricity with the imaging plate. A perfectly aligned scanner would have a RADIAL OFFSET of 0.0. The radial offset is perpendicular to the angular offset. A diagram of these offsets is shown under the ANGULAR OFFSET keyword.

---

<i>format</i>	RADIAL OFFSET value_(mm)
<i>default</i>	0.0
<i>example</i>	RADIAL OFFSET -0.43

---

### RASTER

Pixel size in the horizontal (slow scanning) dimension

---

<i>format</i>	real number in mm
<i>default</i>	defined by the values in the FORMAT statement
<i>example</i>	RASTER 0.1

---

### RAW DATA FILE

This defines the detector data file name template. If the file name contains a run of # characters, then they will be substituted with the numerical value of the SECTOR argument. This allows for easy processing of a large series of sequential images.



## APPENDIX – INDEXING AND INTEGRATION KEYWORDS

---

<i>format</i>	RAW DATA FILE <i>file_name</i>
<i>default</i>	none
<i>examples</i>	RAW DATA FILE /vms/data/dir/xtal###.osc

---

You must not repeat this command before the next GO statement.

### RECORD LENGTH

Defines the number of pixels in the data record. It is equal to the number of horizontal (slow scanning direction) pixels.

---

<i>format</i>	RECORD LENGTH <i>value_(pixels)</i>
<i>default</i>	FILM length / (RASTER / Y SCALE) (this is defined by the FORMAT statement)
<i>example</i>	RECORD LENGTH 2000

---

### REFINE PARTIALITY / REFINE NO PARTIALITY (DO NOT USE!)

Tells the program to determine the fraction partiality of the reflections on the image. It is critical that this be on during the refinement of the crystal rotation parameters since CRYSTAL rotx and roty cannot be determined without knowing the partiality of the reflections on the image.

---

<i>format</i>	REFINE PARTIALITY
<i>defaults</i>	REFINE PARTIALITY

---

**Do not use REFINE NO PARTIALITY !**

### REJECT

Describes a condition whereby a reflection is rejected by *Denzo*. This is especially useful for eliminating spurious measurements due to ice rings.

---

<i>modifiers</i>	<b>FRACTION</b>	The fraction of <b>acceptable</b> pixels in the background, which are needed for the reflection to be accepted. Default = 0.75 (i.e. 75%). <b>Very sensitive to small changes</b> , so increment by 1-2% if you are messing with this. To get rid of spots in an ice ring, increase this value by 1-2% until the preds in the ice ring are colored red. After all, the defining feature of an ice ring is a horrible background. The acceptability of a background pixel is determined by the next three modifier words.
------------------	-----------------	--

---

## APPENDIX – INDEXING AND INTEGRATION KEYWORDS

**CUTOFF** The  $\sigma$  **deviation** from the best fit to the background plane at which a pixel has a 50% chance of being incorrect and thus rejected. Default is  $3\sigma$ .

The background of a spot can be described by a plane and should be uniform and flat. Sometimes it is not; for example, it may have a slope to it, which can be calculated. If a pixel is still out of the best-fit background plane by (default)  $3\sigma$ , then it is flagged as a reject. It is best not to mess with this unless you really know what you are doing.

**SLOPE** The slope of the background plane above which the entire background and spot are considered unacceptable and rejected. Default is 50 units. Flat = 0, vertical = infinite. It is sometimes useful to increase this to 100 for very low background images.

**SYSTEMATIC** This is a way of changing the background rejection criteria when you have a major problem with diffuse scattering. Default is 0.01. The range is from 0 to 1, with 0 meaning rely totally on the cutoff criteria, and 1 meaning don't reject anything. As you can see from the default, you normally use the  $3\sigma$  cutoff criterion.

<i>defaults</i>	REJECT fraction 0.75
	REJECT cutoff 3.0
	REJECT slope 50
	REJECT systematic 0.01
<i>example</i>	REJECT fraction 0.80

### RESOLUTION LIMITS / RES

Defines the resolution range used in the program, in Ångströms.

<i>format</i>	RESOLUTION LIMITS <i>valuelower valueupper_</i> (Å)
<i>default</i>	none
<i>example</i>	RESOLUTION LIMITS 20.0 2.5

### SECTOR

This is the **counter and range delimiter** for a series of frames when processing in batch mode. The number in the sector argument is substituted for the ### (or #, or ##, ...) variable in the RAW DATA FILE and FILM output file commands. Make sure there are enough # marks to accommodate all of the digits in the sector argument, e.g., if you have sector values in the 100's then there must be at least three # marks in the file names specified in RAW DATA FILE and FILM output file. The sector value can start at any

## APPENDIX – INDEXING AND INTEGRATION KEYWORDS

non-negative integer, and is incremented by 1 every time an END OF PACK statement is reached. The first number in the sector argument is the first frame, and the last number is the last frame. See also

### START REFINEMENT.

---

<i>format</i>	SECTOR integer to integer
<i>default</i>	none
<i>example</i>	SECTOR 1 to 50

---

### SKEW

SKEW refers to the **non-orthogonality** of the vertical and horizontal scanning directions, measured in radians. For a perfectly orthogonal system SKEW is defined to be zero. Generally, SKEW values will be very close to this number. It can be fitted in the later rounds of parameter refinement.

---

<i>format</i>	SKEW value_(dimensionless)
<i>default</i>	0.0000
<i>example</i>	SKEW 0.00169 (value taken from the earlier refinement)

---

Spiral scanners and CCD detectors should use the default value of zero, which should not be refined.

### SPACE GROUP

The space group symbol or number as given in the International Tables. Note that for rhombohedral space groups, autoindexing only works for the primitive choice, i.e., H3 or H32, not R3 or R32. The SPACE GROUP must be declared before the FIT cell command can work.

*Denzo* does not actually care about the SPACE GROUP per se. It indexes films by Bravais lattice and does not take into account systematic absences arising from two-folds, screws, etc. These are accounted for in Scalepack when the data are reduced. So for the purposes of *Denzo*, the resulting .x file will be the same whether you declare the space group to be P222 or P212121, or, P3 and P6122.

---

<i>format</i>	SPACE GROUP modifier
<i>default</i>	P1
<i>example</i>	SPACE GROUP P212121

---

### SPINDLE AXIS / ORIENTATION AXIS 1

The reciprocal lattice vector parallel to the spindle axis. Same as keyword ORIENTATION AXIS 1

## APPENDIX – INDEXING AND INTEGRATION KEYWORDS

---

<i>format</i>	3 integers representing the reciprocal vector
<i>default</i>	0 0 1 (c* axis)
<i>example</i>	SPINDLE AXIS 1 0 0 (a* axis) SPINDLE AXIS 1 0 1 (a non-major axis)

---

### ***SPOT***

Defines the shape and area of the region in the box where the spot is to be found.

---

<i>format</i>	SPOT radius value_(mm) SPOT elliptical value major value minor_(mm) value angle_(degrees)
<i>defaults</i>	none
<i>examples</i>	SPOT radius 0.35
<i>modifiers</i>	radius Defines the spot to be circular with a radius value, in mm. The center of the spot area and of the box coincide

---

### ***START REFINEMENT***

START REFINEMENT denotes the start of the series of commands given to index the diffraction image. The end of the START REFINEMENT loop is denoted by the CALCULATE GO statement. The START REFINEMENT statement must come after the SECTOR *x* to *y* statement, and the loop is run for every sector value.

### ***STOP (REDUNDANT)***

This command stops the *Denzo* job. Same as QUIT or a semicolon (;).

### ***TITLE***

Data set title

---

<i>format:</i>	string. If it contains any spaces, commas, equal signs, or brackets ( <i>DENZO</i> recognized separators), it must be enclosed in single quotes.
<i>default:</i>	none
<i>example:</i>	TITLE 'my crystal'

---

### ***TOP MARGIN***

The top margin, in mm from the top edge, of useful data, i.e., the beginning of every record. Do not set the top margin to be larger than the film length. Note that commands

## APPENDIX – INDEXING AND INTEGRATION KEYWORDS

like top margin are working in the data coordinate system. It is very important to check the results of these commands in the display window.

---

<i>format</i>	TOP MARGIN value_(mm) default 0.0
<i>example</i>	TOP MARGIN 2.0

---

### UNIT CELL

Unit cell parameters in either real or reciprocal space.

---

<i>format</i>	UNIT CELL a* b* c* $\alpha^*$ $\beta^*$ $\gamma^*$ UNIT CELL a b c $\alpha$ $\beta$ $\gamma$ . Lengths in Ångstroms or reciprocal Å, angles in degrees.
<i>default</i>	none
<i>example</i>	UNIT CELL 120.92 120.92 131.65 90 90 90

---

If the space group has been declared, the program is smart enough to know which cell lengths should be equivalent, and which angles are variable, and it will hold those lengths equal or fixed throughout the refinement.

### USE BEAM

Tells the program to use the beam spot in the refinement. This keyword has the opposite effect of the keyword NO BEAM. This is useful in the initial steps of the refinement, when not too many spots are matched by preds.

---

<i>modifier</i>	only Tells the program to use only the beam spot in the refinement. Has the effect of turning off the RESOLUTION LIMITS keyword.
<i>format</i>	USE BEAM USE BEAM only
<i>default</i>	This is not the default option. See NO BEAM.

---

### USE FIDUCIALS / USE NO FIDUCIALS

Tells the program to use the **fiducials** in the refinement. Turned off with USE NO FIDUCIALS. Fiducial positions are entered with the keyword FIDUCIALS.

<i>modifier</i>	only Tells the program to use only the fiducial positions in the refinement.
<i>format</i>	USE FIDUCIALS USE FIDUCIALS only

## APPENDIX – INDEXING AND INTEGRATION KEYWORDS

*default* not the default. Default is USE NO FIDUCIALS.

---

### USE PARTIALS POSITION / USE PARTIALS NO POSITION

Normally, data reduction programs use only fully recorded reflections in the refinement of the crystal and detector positional parameters. The keyword USE PARTIALS POSITION tells the program to use a reflection that is predicted to be only partially recorded in the positional refinement. A spot is predicted to be partially recorded based on the values entered for the oscillation range, as well as the crystal mosaicity and beam crossfire. This is a very useful feature of *Denzo* because it allows one to refine the orientation of the crystal even if none of the reflections are fully recorded. This occurs if one collects either very thin oscillation sectors (in an attempt to minimize background), has a crystal of relatively high mosaicity, or some combination of both. USE PARTIALS NO POSITION is mostly of historical interest.

---

<i>default</i>	USE PARTIALS POSITION
<i>examples</i>	USE PARTIALS POSITION
	USE PARTIALS NO POSITION

---

### VERTICAL AXIS / ORIENTATION AXIS 2

The reciprocal lattice vector perpendicular to both the spindle axis and the X-ray beam. Equivalent to ORIENTATION AXIS 2.

---

<i>format</i>	3 integers representing the reciprocal vector
<i>default</i>	1 0 0 (a* axis)
<i>example</i>	VERTICAL AXIS 0 1 0 (b* axis)

---

### WAVELENGTH

X-ray wavelength. Same as LAMBDA.

---

<i>format</i>	WAVELENGTH value_(Ångstroms)
<i>default</i>	1.5418 (CuK $\alpha$ ) <i>example</i> WAVELENGTH 1.5418

---

### WEAK LEVEL

This is the intensity rejection level, expressed as an I/ $\sigma$  ratio. Only strong spots with an integrated intensity (i.e., I) above the WEAK LEVEL times the intensity error measurement (i.e.,  $\sigma$ ) are used in the positional refinement and as reference profiles in profile fitting.

## APPENDIX – INDEXING AND INTEGRATION KEYWORDS

The intensities of all reflections determined to be *weak* are still measured and included in the .x file. **They're just not used in the profile fitting.** You can use **PROG FIT R** in the *XdisplayF* window to see which reflections are above the weak level.

---

<i>format</i>	WEAK LEVEL value_(dimensionless)
<i>default</i>	5.0
<i>example</i>	WEAK LEVEL 5.00

---

### WRITE PREDICTIONS / WRITE NO PREDICTIONS

This tells *Denzo* to write a file called hklpredictions. This file contains the positions of the predicted reflections and is read by *XdisplayF* to update the predictions in the display. Once the flag is turned on, the file is updated every time a new cycle of refinement is executed. This way you can follow the course of the refinement in the display program. Turned off with WRITE NO PREDICTIONS.

---

<i>format</i>	WRITE PREDICTIONS
<i>default</i>	WRITE NO PREDICTIONS
<i>example</i>	WRITE PREDICTIONS

---

### X BEAM

Defines the distance from the edge of the data to the beam spot, in mm. You can use coordinates of the cursor in the green box of the *XdisplayF* window to find the beam position.

---

<i>format</i>	X BEAM lue_(mm) default FILM width/2
<i>example</i>	X BEAM 98.961

---

### Y BEAM

Defines the distance from the edge of the data to the beam spot, in mm. Again, use coordinates of the cursor in the green box of the *XdisplayF* window to find the beam position.

---

<i>format</i>	Y BEAM value_(mm)
<i>default</i>	FILM length/2
<i>example</i>	Y BEAM 100.146

---

**Y SCALE**

Y SCALE in *Denzo* serves two purposes. First, it is a dimensionless correction factor for anisotropy in the pixel dimensions. It is applied to the fast scanning, or *y*, dimension of the pixel. If the pixel is larger in *y* than in *x*, then Y SCALE is smaller than 1.

It is possible to create a data file which has the inverted direction of scanning along one of the axis - similarly as in the case of film scanned from the backside. *Denzo* handles this aberration by changing the sign of the Y SCALE keyword to negative. If this is not done, then the assignment of the Friedel mates (Bijvoet differences) will be inverted, and thus the hand of the structure will be similarly inverted - see *Determining the values of FILM ROTATION and Y SCALE* section in '*Denzo Orientation Conventions*' chapter. The sign of Y SCALE is a detector property, while the absolute value of Y SCALE seems stable for all detectors. Once the correct sign has been determined, the calibration of Y SCALE is done by refining with a high symmetry crystal (trigonal, tetragonal, hexagonal, or cubic).

For CCD and spiral scanners, the value of Y SCALE should be exactly  $\pm 1$  and not refined!

---

<i>format</i>	Y SCALE value_(dimensionless)
<i>default</i>	detector dependent, defined by the FORMAT statement
<i>example</i>	Y SCALE 1.03152

---

**YORK OUTPUT FILE**

Substitutes for the OUTPUT FILE command, and tells the program to write the '.x' file in the **Eleanor Dodson** York file format. Note that Scalepack reads the *Denzo* YORK OUTPUT FILE as format denzo\_york1.



# Appendix - Scaling Keywords

This is a listing of all of the Denzo keywords and their modifiers, if any:

ADD	NO ANOMALOUS
ADD PARTIALS	NO MERGE
<i>all</i>	<i>include partials</i>
ANOMALOUS	<i>include no partials</i>
B RESTRAIN	<i>original index</i>
BACKGROUND CORRECTION	NO PROFILE TEST
DEFAULT B FACTOR	NUMBER OF ITERATIONS
DEFAULT SCALE	NUMBER OF ZONES
DO NOT REJECT OUTLIERS	ORIENTATION AXIS 1
ERROR SCALE FACTOR	ORIENTATION AXIS 2
ESTIMATED ERROR	ORIGINAL WINDOW
EXTEND PARTIALS / DO NOT EXTEND	OUTPUT ANOMALOUS
PARTIALS	POLARIZATION
FILE	POSTREFINE
FIT	PRINT
<i>crystal</i>	total chi2
<i>batch</i>	single chi2
<i>film</i>	correlation / no correlation
FIT B	solution / non solution
FIX B	shifts / no shifts
FIXED WINDOW	profile test
FORMAT	PROFILES FITTED
<i>denzo_ip</i>	PROFILES SUMMED
<i>denzo_york1</i>	RECSQ
<i>scalepack</i>	REFERENCE BATCH
<i>rigaku raxis</i>	REFERENCE BATCHES, FILM, FILMS
<i>xds profit.hkl</i>	REJECT HKL
<i>madnes procor intout</i>	REJECT OUTLIERS
<i>madnes procor ascii</i>	REJECTION PROBABILITY
<i>xengen urf</i>	RESOLUTION
<i>archive</i>	RESOLUTION STEP
FRAME WIDTH	SCALE ANOMALOUS
HKL MATRIX	SCALE RESTRAIN
HKL SCALE	SECTOR
HKL SHIFT	SECTOR WIDTH
IGNORE OVERLOADS	SIGMA CUTOFF
INCLUDE OVERLOADS	SPACE GROUP
INITIAL B FACTOR	SPINDLE AXIS
INITIAL SCALE	UNIT CELL
INPUT / @	VERTICAL AXIS
MERGE	WRITE BADDIES
MOSAICITY	WRITE REJECTION FILE

**ADD**

Increments batch numbers by a constant to every batch from this point on until another add command is read. Useful to make unique batch numbers from two or more files which have the same batch numbers inside. For example, **Denzo\_york1** format embeds the batch number in the .x file.

---

<i>format</i>	ADD value
<i>default</i>	nothing added
<i>example</i>	ADD 1000     This will add 1000 to each batch number.

---

**ADD PARTIALS**

Tells the program to add partially recorded reflections among consecutive batches, even if the batches do not have consecutive numbering. **Essentially obligatory.**

---

modifier	all     partials are added over all consecutive batches
<i>format</i>	ADD PARTIALS start1 to end1 start2 to end2 ... etc
default	do not ADD PARTIALS
example	ADD PARTIALS 0 to 49 51 to 99 100 to 149

---

**Be sure that the ranges of numbers do not overlap.**

**ANOMALOUS**

Flag for keeping Bijovets (I+ and I-) separate in the output file. If the ANOMALOUS flag is on, anomalous pairs are considered equivalent when calculating scale and B factors and when computing statistics, but are merged separately and output as I+ and I- for each reflection.

---

<i>format</i>	ANOMALOUS
<i>default</i>	not turned on (I+ and I- are combined)
<i>example</i>	ANOMALOUS

---

**B RESTRAIN**

Can be used to restrain B factor differences from consecutive films or batches. The value which follows the flag represents the amount in Å<sup>2</sup> you will allow the B factors to differ from consecutive frames or batches. See also the keyword SCALE RESTRAIN.

---

<i>format</i>	B RESTRAIN <i>value_</i> (Å <sup>2</sup> ) <i>default</i> not turned on
<i>example</i>	B RESTRAIN 0.5

---

**BACKGROUND CORRECTION**

The `BACKGROUND CORRECTION` command specifies the amount per frame to increase background. Corrects for errors in SDMS (Hamlin) data integration in the software distributed in the late 1980s and early 1990s. The value given after the keyword increases the background by #counts/frame. Valid **for archive file format only**.

---

<i>format</i>	BACKGROUND CORRECTION value_(#counts/frame)
<i>default</i>	not turned on
<i>example</i>	BACKGROUND CORRECTION 15

---

**DEFAULT B FACTOR**

Overall B used only in the absence of an `INITIAL B FACTOR`. You can apply a higher value to lower your  $R_{\text{merge}}$ . Does not affect the quality of the data.

---

<i>format</i>	DEFAULT B FACTOR value
<i>default</i>	0
<i>example</i>	DEFAULT B FACTOR 5

---

**DEFAULT SCALE**

Overall scale factor used in the absence of an initial scale factor. This is useful if the data are too strong, which is sometimes the case with small molecules. It will reduce the output intensities by the factor entered.

---

<i>format</i>	DEFAULT SCALE value
<i>default</i>	1
<i>example</i>	DEFAULT SCALE 10

---

Useful to reduce the overall scale of the data set. If the numbers in the output file are too large, `DEFAULT SCALE 10` will reduce them 10-fold.

**DO NOT REJECT OUTLIERS**

Turns off the reject outliers flag.

---

<i>format</i>	DO NOT REJECT OUTLIERS
<i>default</i>	DO NOT REJECT OUTLIERS
	outliers not rejected automatically, but see <code>REJECT OUTLIERS</code> for more discussion.

---

**ERROR SCALE FACTOR**

This is a single multiplicative factor which is applied to the input  $\sigma_i$ . This should be adjusted, so the normal  $\chi^2$  (goodness of fit) value that is printed in the final table of the output comes close to 1. By default, the input errors are used (ERROR SCALE FACTOR = 1). It applies to the data which are read after this keyword, so you can apply a different error scale factor to subsequent batches by repeating this input with different values.

---

<i>format</i>	ERROR SCALE FACTOR value
<i>default</i>	ERROR SCALE FACTOR 1
<i>example</i>	ERROR SCALE FACTOR 1.3 good starting value for format denzo_ip

---

**ESTIMATED ERROR**

The estimate of the systematic error for each of the resolution shells. **There must be exactly the same number of error estimates here as there are** NUMBER OF ZONES. So if you have 10 zones, you need 10 numbers following the keyword estimated error - one for each zone.

The error estimates do not all have to be the same. The estimated error applies to the data which are read **after** this keyword, so you can apply a different error scale factor to subsequent batches by repeating this input with different values. This is an important point if you enter data from a previous Scalepack output that does not need its  $\sigma$  to be increased.

The error estimates should be approximately equal to the R-factor in the table at the end of the output for resolution shells **where statistical errors are small**, namely the earlier resolution shells where the data is strong. This is a crude estimate of the systematic error, to be multiplied by I, and is usually invariant with resolution. Default = 0.06 (i.e. 6%) for all zones.

---

<i>format</i>	ESTIMATED ERROR value1 value2 value3 value4 ...etc
<i>default</i>	each value set to 0.06
<i>example</i>	ESTIMATED ERROR 0.02 0.03 0.03 0.03 0.04 0.04 (6 zones)

---

**EXTEND PARTIALS / DO NOT EXTEND PARTIALS**

Some partially recorded reflections may be predicted by *Denzo* or *Scalepack* to start or end their Bragg condition in between consecutive frames due to small variations in crystal orientation from frame to frame. For these reflections only, there are two choices of defining where the reflection started (or ended): including the extra frame, or not. EXTEND PARTIALS tells

Scalepack to include this extra frame. It only affects a very small fraction of the reflections. Opposite of DO NOT EXTEND PARTIALS.

---

<i>format</i>	EXTEND PARTIALS
<i>default</i>	this is the default

---

## FILE

This specifies the files read by *Scalepack*. The input has two components. The first is a number. The second is a file name, which usually contains wildcard characters (###) that are incremented by the `SECTOR` command. The number which follows `FILE` specifies the **starting batch number**. A batch, previously called a **film**, can be as small as a single .x file (or the equivalent). It can be a group of .x files, even an entire data set. The most frequent content of a batch is a single .x file.

This conversion of files into batches is particularly useful if you want to scale more than one data set together. For example, let's say you want to scale 10 oscillation frames (numbered 1 through 10) from first data set with 37 oscillation frames (numbered 1 through 37) from another one. The `FILE` statement will take each of the individual .x files and assign them a batch number. Thus, you would enter something like this: `FILE 1 'setone###.x'` and `FILE 101 'settwo###.x'` (see Scenario 5d). Thus, batch numbers 1 - 10 will correspond to files `setone001.x`, `setone002.x`, etc. (assuming you used the `SECTOR 1 to 10` command above `FILE` so that the wildcards would be substituted with numbers). The batch numbers 101 - 137 will correspond to files `settwo001.x`, `settwo002.x`, etc. the ### is replaced by the sector argument, not by the batch number.

---

<i>format</i>	FILE value 'filename' (see also note below)
<i>defaults</i>	None
<i>example</i>	FILE 101 '/frames/scale/lysoz###.x'

---

The `FILE` must come after `FORMAT` because the syntax depends on which input format is being read. The `FILE` must not be followed by a number in the case of the archive, `denzo_york`, and `denzo_york1` formats, where the number after the word `FILE` is not given because the batch numbers are already stored in the file. If you want to change the batch numbers in these file formats, see the `ADD` command described above.

## FIT

Tells the program what parameters to fit in **post-refinement** and specifies the group of files over which the fitting is to be performed. Postrefinement fit can be applied to an

entire set of batches (one batch being the entire set of frames, for example) using the modifier `crystal`, or to each individual input file using the modifier `film` or `batch`.

---

<i>modifiers</i>	<i>CRYSTAL</i>	specifies that the fit operation is over the entire set of frames specified by a range and restrains the fit parameter to have exactly the same value over this range.
	<i>BATCH</i>	specifies that the fit operation is performed on each member of the set of frames specified by a range.
<i>parameters</i>	<i>FILM</i>	alias for batch
	<i>a*, b*, c*</i>	unit cell lengths. Values returned are real space
	<i>alpha*</i>	unit cell angles. Values returned are real space.
	<i>beta*</i>	
	<i>gamma*</i>	
	<i>rotx</i>	crystal orientation parameters deduced in Denzo
	<i>roty</i>	
	<i>rotz</i>	
	<i>wavelength</i>	
	<i>mosaicity</i>	incident wavelength..
<i>format</i>	<i>fit modifier parameter1 filmnumber to filmnumber</i>	mosaicity as defined in Denzo, in degrees
	<i>the default is that nothing is fit unless specified</i>	
<i>default</i>	<i>FIT crystal roty 1 to 137</i>	
	<i>FIT crystal mosaicity 1 to 5 7 to 10 102 to 104 to 137</i>	
	<i>FIT batch rotx 1 to 137</i>	

---

Most mistakes in Scalepack can be attributed to errors in FIT commands because the program is very sensitive to mistakes in the batch numbers. If you input non-existent batch numbers or define overlapping ranges (e.g., 1 to 10 5 to 20), the program is likely to fail in a strange way. If you specify a range of numbers, the program will only use the batch numbers that exist within the range. For example, if your batch numbers go from 1 to 40 and 70 to 90, you can get away with saying, say, FIT batch (parameter) 1 to 90, which is the same as FIT batch (parameter) 1 to 40 70 to 90. For FIT crystal, these two inputs are not equivalent. In the case of FIT crystal (parameter) 1 to 90, one value will be fit for all batches. In the case of FIT crystal (parameter) 1 to 40 70 to 90, two values will be fit, one for each range. Note that different parameters may be fit over different ranges and either over batch or crystal. You can also mix batch and crystal for the same parameters.

Do not fit unit cell parameters restrained by space group symmetry. For example, if you have space group P3, you must not fit *b\**.

Do not FIT `batch rotx` because this parameter is very poorly defined by the intensities of observed partial reflections. This is another very common mistake.

Unless you know what you are doing, do not FIT `crystal rotx roty rotz`, because if the spindle is even slightly misaligned, the assumption that there is only one crystal

orientation parameter for a large sweep of data will force incorrect restraints on the crystal orientation refinement.

About fitting rotations: Changes in rotations, like crystal *rotx*, *roty*, and *rotz*, are expressed as small rotations, call them  $\Delta$ , about the laboratory frame of reference. These  $\Delta$ :  $\Delta_x$ ,  $\Delta_y$ , and  $\Delta_z$ , are used because to a first approximation, they commute with one another (commute means that the order in which they are applied is irrelevant). This, in turn, is because these  $\Delta$  are small, typically less than one degree. The crystal or cassette rotations,  $R_x$ ,  $R_y$ , and  $R_z$ , on the other hand, do not commute with one another because their values tend to be large (much greater than one degree). So when you ask *Scalepack* postrefinement or *Denzo* to fit these rotations, what is actually happening is that the  $\Delta$  are being refined. After each refinement cycle, the  $\Delta$  are converted into changes in  $R_x$ ,  $R_y$ , and  $R_z$  by a (complicated) algebraic relation. Those of you with sharp eyes will have noticed in *Denzo* that the shifts reported by the program when fitting the crystal rotations do not correspond to the changes in the *rot* values. This is because the shifts reported are the  $\Delta$ , not the changes in *rotx*, *roty*, and *rotz*. The other reason for fitting  $\Delta$  rotations the way they are defined in *Denzo* is to make them have a more intuitive correlation with the other parameters. Otherwise, changes in crystal *rotz* would not correlate with cassette *rotz*. This is of more importance in *Scalepack*, where only *rotx* and *roty*, and not *rotz*, are refined.

About fitting unit cell parameters: In both *Denzo* and *Scalepack*, the unit cell is fitted in reciprocal space, not real space. This means that for a non-orthogonal space group, refining the value of  $a^*$  may end up changing the values of  $b$  and  $c$ , even though  $b^*$  and  $c^*$  remain the same. The same is true for the angles: fitting  $\alpha^*$  may end up changing  $\beta$  and  $\gamma$ , even though  $\beta^*$  and  $\gamma^*$  remain the same. So what you may notice sometime if you are not careful is that when you ask the program to fit crystal  $a^*$ , but not  $b^*$ ,  $c^*$ , and the angles, then  $a$  will not be a constant.  $a^*$  will be constant, but when converted back to  $a$  when the other unit cell parameters have not been changes,  $a$  will not. The moral: when fitting unit cells, fit all the relevant parameters.

## ***FIT B***

This flag tells the program to refine B factors of every batch from the very first cycles of refinement. This is in contrast to the default procedure, where the B factors are fit only after the convergence of the scaling. In the default procedure, if scaling does not converge in 20 (default) cycles of refinement, B factors will be not be fitted. The `FIT B` command can override this. Not to be confused with the postrefine `FIT B*` command described above. **You cannot postrefine B (temperature) factors.**

---

<i>format</i>	FIT B
<i>default</i>	this flag is turned off
<i>example</i>	FIT B

**FIX B**

This flag tells the program not to fit B factors at all. Usually, it is combined with the input of the B factors you want to apply but do not wish to refine anymore, or it is used for frozen crystals where you do not expect significant decay.

---

<i>format</i>	FIX B
<i>default</i>	turned off; B factors are fit after the scale factor refinement converges.
<i>example</i>	FIX B

---

**FIXED WINDOW**

For **Hamlin** archive files a fixed window of 3, 5, 7, or 9 frames or the original Hamlin determined window of frames may be used for summing a reflection. **Valid for archive file format only.**

---

<i>format</i>	FIXED WINDOW value
<i>default</i>	uses the Hamlin definition of the window
<i>example</i>	FIXED WINDOW 7

---

**FORMAT**

This keyword specifies the format of the input *hkl* and intensity data. Input data can come from any of the nine types of files. This program requires this keyword to properly read the files.

---

<i>modifiers</i>	<i>DENZO_IP</i> from frames processed with Denzo
	<i>DENZO_YORKI</i> output created with Denzo option york
	<i>SCALEPACK</i> from Scalepack output file
	<i>RIGAKU RAXIS</i> binary R-Axis software output
	<i>XDS PROFIT.HKL</i> binary output from XDS output file
	<i>MADNES PROCOR INTOUT</i> binary madness output
	<i>MADNES PROCOR ASCII</i> ascii madness output
	<i>XENGEN URF</i> binary output from xengen. Also have to supply the program with info about frame width, in degrees, using frame width keyword.

---



<i>format</i>	<i>ARCHIVE</i> binary output from Hamlin software
<i>default</i>	FORMAT modifier
<i>example</i>	denzo_ip FORMAT denzo_ip * if you use one of the binary formats that the data must be scaled on the same type of computer that created the binary files due to incompatibilities in number representations between computers.

**FRAME WIDTH**

**Only for URF files** created by *Xengen*. Other formats do not need this specification because *Scalepack* can read this information off the file header. The oscillation range for each frame.

<i>format</i>	FRAME WIDTH value
<i>default</i>	there is no default, this input is required for this format
<i>example</i>	FRAME WIDTH 0.2

**HKL MATRIX**

Matrix for re-indexing. This matrix is applied to the hkl's as they are read in (applies to data read in after this command):

$$h' = (1)*h + (2)*k + (3)*l$$

$$k' = (4)*h + (5)*k + (6)*l$$

$$l' = (7)*h + (8)*k + (9)*l$$

Input is in the order (1....9). HKL MATRIX works with postrefinement. The program will not accept a matrix which has a negative determinant

<i>format</i>	HKL MATRIX	value11 value12 value13 value21 value22 value23 value31 value32 value33
<i>default</i>	the default is the unit matrix	
<i>example</i>	HKL MATRIX 0 1 0 1 0 0 0 0 -1 transforms hkl into k h -l, has the effect of switching a and b. -1 is to keep the determinant positive.	

**HKL SCALE**

Divides each *h*, *k*, and *l* by the input values. Useful for reducing the unit cell volume, particularly after *hkl* matrix transformation. Rarely needed. Example, index data

originally in C222 and want your data in, say, P3, you would apply `HKL SCALE` of 2 2 1, and `HKL MATRIX` of 1 1 0 1 -1 0 0 0 -1 so that the new indices would have values of  $h + k/2$ ,  $h - k/2$ , and  $l$ .

---

<i>format</i>	HKL SCALE value1 value2 value3
<i>default</i>	1 1 1
<i>example</i>	HKL SCALE 2 2 1

---

### HKL SHIFT

Used to make a quick test of **misindexing**. Adds the specified integer vector to each original  $h$ ,  $k$ , and  $l$ . If you are successful in using this, congratulations, you now have to reprocess your data in *Denzo* with the correct indexing.

---

<i>format</i>	HKL SHIFT integer integer integer
<i>default</i>	0 0 0
<i>example</i>	HKL SHIFT 0 1 0      we'll assume you weren't too far off

---

### IGNORE OVERLOADS

Opposite of `INCLUDE OVERLOADS`. This is **not the default**. Useful if you collect data at low and high exposures and is useful to ignore the saturated reflections at the high exposures. This macro applies to data read in after this command.

---

<i>format</i>	IGNORE OVERLOADS
<i>default</i>	This is NOT the default → See Include Overloads

---

### INCLUDE OVERLOADS

`INCLUDE OVERLOADS` is a flag for whether fitted profiles with some pixels missing (typically due to overload) should be included in the scaling. Affects only Denzo image plate output files (formats `denzo_ip`, `denzo_york1`). Note that for summed profiles this does not apply because only profile fitting can estimate the value of the overloaded pixels.

---

<i>format</i>	INCLUDE OVERLOADS
<i>default</i>	This is the default

---

**INITIAL B FACTOR**

Table of starting B factors, one per batch, beginning on the record after the title. After running the program once, the output table of B factors can be cut out and pasted in here, for example, when you have set the FIX B flag and are not refining B values anymore.

---

<i>format</i>	INITIAL B FACTOR batch no. b value batch no. b value ...
<i>default</i>	value of the default B factor, which is defaults to 0
<i>example</i>	INITIAL B FACTOR 1 0.0 2 0.1 3 0.1 4 0.1 5 -0.2 ...etc

---

**INITIAL SCALE**

Table of starting scale factors, one per batch, beginning on the record after the title. After one run of the program, the output table of scale factors can be cut out and pasted in here. This table is not required. If it is not included in the control input, then the DEFAULT SCALE is used (this is 1 unless otherwise specified).

If the initial scale factor is set to zero, that frame is ignored in the scaling and refinement.

---

<i>format</i>	batch no. scale value batch no. scale value ...
<i>default</i>	value of the default scale factor, which is defaulted to 1
<i>example</i>	INITIAL SCALE FACTOR 1 1.0 2 0.9 3 0.0 4 0.85 5 1.1 ...etc.

---

**INPUT / @**

Redirection. Tells the program to read the file which follows the keyword. Same as in *Denzo*.

---

<i>format</i>	@filename
<i>default</i>	no redirection
<i>example</i>	@reject

---

**MERGE**

Flag that tells the program to merge (combine, average) reflections with the same unique index. This is the *Scalepack* default.

---

<i>format</i>	MERGE
<i>default</i>	This is the default and need not be specified

---

**MOSAICITY**

Allows you to input the value of the mosaicity of the data set. This is normally read from the header of the .x file.

---

<i>format</i>	MOSAICITY value
<i>default</i>	read from the header of the .x file
<i>example</i>	MOSAICITY 0.5

---

**NO ANOMALOUS**

Opposite of the keyword ANOMALOUS. Causes I+ and I- to be merged. This is the default.

---

<i>format</i>	NO ANOMALOUS
<i>default</i>	This is the default

---

**NO MERGE**

Flag for the output of **unmerged** (reflections with the same unique *hkl* are not combined) data. Opposite of MERGE. This is a very handy for specialized work. This flag has subsidiary modifiers `include partials`, `original index` or as **default** `include no partials`

---

<i>modifiers</i>	<b>INCLUDE PARTIALS</b>	all observations, both fully and partially recorded, are included in the output. The output will consist of the unique <i>hkl</i> , batch number, asymmetric unit, <i>I</i> , $\sigma$ , and fractionality of the reflection. There is no information about I+ and I-, although it may be possible to get this in subsequent versions of the program.
	<b>INCLUDE NO PARTIALS</b>	only fully recorded reflections and those fully recorded reflections created by the summation of partials are included in the output. Partial reflections which cannot be summed to a fully recorded reflection are lost. The output will consist of the unique <i>hkl</i> , batch number, asymmetric unit, <i>I</i> , and $\sigma$ . This is the default for NO MERGE. There is no information about I+ and I-.
	<b>ORIGINAL INDEX</b>	the output will also contain the original (not unique) <i>hkl</i> for each reflection. This is designed for MAD/local scaling work. The original index modifier only works with the default INCLUDE NO PARTIALS. The output will consist of the original <i>hkl</i> , unique <i>hkl</i> , batch number, a flag (0 = centric, 1 = I+, 2 = I-), another flag (0 = <i>hkl</i> reflecting above the spindle, 1 = <i>hkl</i> reflecting below the

spindle), the asymmetric unit of the reflection, I (scaled, Lorentz and Polarization corrected), and the  $\sigma$  of I. The format is (6i4, i6, 2i2, i3, 2f8.1).

<i>format</i>	NO MERGE modifier
<i>example</i>	NO MERGE original index

---

This command is also useful if you just want to combine all of the information contained in multiple .x files into a single file. Simply read in all of the .x files, don't do any scale or B factor refinement, and output NO MERGE include partials.

### NO PROFILE TEST

This is a flag which operates only on **ARCHIVE** format files. Tells *Scalepack* that reflections with weird profiles should not be rejected. Opposite of **PROFILE TEST**.

<i>format</i>	NO PROFILE TEST
<i>default</i>	This is not the default

---

### NUMBER OF ITERATIONS

The number of **cycles for refinement** of scale and B-factors. The default value is 20 cycles. If it is set to 0, the program computes statistics, merges reflections with the same unique *hkl*, and writes the output file based on the initial scale and B factors. Normally you do not have to specify this unless you want to avoid scaling.

<i>format</i>	NUMBER OF ITERATIONS integer
<i>default</i>	20 iterations
<i>example</i>	NUMBER OF ITERATIONS 0

---

### NUMBER OF ZONES

The number of **resolution shells** the data is divided into for the basis of calculating statistics. This input is required and must match the number of zones specified under the **ESTIMATED ERROR** keyword. Handy tip: it's nice to set up the **NUMBER OF ZONES** to equal the number of zones used by *X-plor* for the output of refinement statistics by shell, for when you get around to publishing your data and refinement statistics together in a paper.

<i>format</i>	NUMBER OF ZONES integer
<i>default</i>	This input is required
<i>example</i>	NUMBER OF ZONES 10

---

**ORIENTATION AXIS 1**

Three integer vector which describes the orientation of the vertical axis of the crystal. Equivalent to *Denzo* and *Scalepack* keywords VERTICAL AXIS. This information is usually not inputted by the user, but subsequent versions will read it from the header of the .x files. This command does not affect scaling or postrefinement. However, it does affect the values of the misorientation angles reported in the *Scalepack* log file.

---

<i>format</i>	ORIENTATION AXIS 1 integer integer integer
<i>default</i>	1 0 0
<i>example</i>	ORIENTATION AXIS 1 0 1 0

---

**ORIENTATION AXIS 2**

Same as ORIENTATION AXIS 1 except describes the spindle axis of the crystal. Equivalent to *Denzo* and *Scalepack* keywords SPINDLE AXIS. Otherwise the same as above.

---

<i>format</i>	ORIENTATION AXIS 2 integer integer integer
<i>default</i>	0 0 1
<i>example</i>	ORIENTATION AXIS 2 1 0 0

---

**ORIGINAL WINDOW**

Refers to the window size of ARCHIVE format files. Opposite of FIXED WINDOW.

**OUTPUT ANOMALOUS**

Alias for the keyword ANOMALOUS.

**OUTPUT FILE**

Name of file for output of scaled measurements. A new file will be created if none exists, but a pre-existing file will be overwritten. Maximum of 80 characters allowed in the file name. Depending on whether the ANOMALOUS flag is set, there are one or two sets of I and  $\sigma(I)$  per reflection. The output is h, k, l, I,  $\sigma(I)$  (NO ANOMALOUS flag) or h, k, l, I+,  $\sigma(I+)$ , I-,  $\sigma(I-)$  (ANOMALOUS flag) in format (3I4, 4F8.1). If the NO MERGE flag has been set, unmerged data are output as h, k, l, asymmetric unit #, I,  $\sigma(I)$  in format (4I4, 2F8.1) and the ANOMALOUS flag has no effect. For NO MERGE original index the format is: original hkl, unique hkl, batch (film) number, a flag (0 = centric, 1 = I+, 2 = I-), another flag (0 = hkl reflecting above the spindle, 1 = hkl reflecting below the spindle),

the asymmetric unit of the reflection, I (scaled, Lorentz and Polarization corrected), and the  $\sigma(I)$ . The format is (6i4, i6, 2i2, i3, 2f8.1).

---

<i>format</i>	OUTPUT FILE name use single quotes for safety
<i>default</i>	you must specify the output file name
<i>example</i>	OUTPUT FILE '/people/myname/LYSOZ.sca'

---

## **POLARIZATION**

This command allows you to correct a mistaken polarization/monochromator value entered when you ran *Denzo*. It saves you from the chore of reprocessing all of your original images once you have learned of a mistake. This is mainly for synchrotron users who collect very high-resolution data.

---

<i>format</i>	POLARIZATION Denzo value corrected value
<i>default</i>	The polarization is read off the header of the .x file
<i>example</i>	POLARIZATION Denzo 0.0 corrected 0.9

---

## **POSTREFINE**

**The number of cycles of postrefinement** to be done. Works in conjunction with the `FIT` commands described above. If you set the number of postrefinement cycles to zero, then postrefinement will be skipped even if you have `FIT` commands in your control file.

---

<i>format</i>	POSTREFINE integer
<i>default</i>	Postrefinement is not done unless you specify it
<i>example</i>	POSTREFINE 10

---

## **PRINT**

This flag tells the program **to print the results** of the specified calculation **after every refinement cycle**. Useful in the case where one scales non-isomorphous data and wants to make the outlier list short by using the statement `PRINT total chi2 200 PRINT single chi2 100`.

---

<i>modifiers</i>	<i>TOTAL CHI2</i>	cutoff for printing poor reflections. Total $\chi^2$ for all measurements of a reflection. Default = 52
	<i>SINGLE CHI2</i>	cutoff for printing poor reflections. $\chi^2$ for any single measurement of a reflection. Default = 3.52.
	<i>CORRELATION / NO CORRELATION</i>	prints correlation matrix. Default: no correlation

---

	<i>SOLUTION / NON SOLUTION</i>	prints new scale and B factors.
	Default: no solution	
	<i>SHIFTS / NO SHIFTS</i>	prints changes to scale and B factors.
<i>format</i>	PRINT modifier	
<i>defaults</i>	PRINT total chi2	
	PRINT single chi2	
	PRINT no correlation	
	PRINT no solution	
	PRINT no shifts	

---

New scale and B factors will always be printed if the number of fitted parameters exceeds 300.

### ***PROFILE TEST***

PROFILE TEST is a flag that can be turned on and off as desired. This is the default. Bad profiles can be used to reject a DCREDUCE summed hkl. Valid for ARCHIVE file format only. Opposite of NO PROFILE TEST.

### ***PROFILES FITTED***

This flag tells Scalepack how the profiles were treated by the indexing program (e.g., Denzo). The two choices for profiles are fitted and summed. Valid for denzo\_ip, Denzo\_york1, and xengen urf formats. Most of the time, people fit profiles, so this is the default and need not be specified in the command file.

---

<i>format</i>	PROFILES FITTED
<i>default</i>	This is the default

---

### ***PROFILES SUMMED***

The opposite of PROFILES FITTED. This is not the default and must be specified if true.

### ***RECSQ***

**Metric tensor** description of the reciprocal space unit cell. Why would you ever use this when you can specify the real space constants?

---

<i>format</i>	RECSQ value1 value2 .... value6
<i>default</i>	same as UNIT CELL
<i>example</i>	RECSQ 0.0001 0 0 0.0004 0 0.0009 [to describe unit cell 100 50 33 90 90 90.]

---



**REFERENCE BATCH**

Specifies which batch or film or set of batches or films will be the reference for the scaling and B refinement. The scale and B factor for these are not refined. More than one film or batch may be used as the reference. This is important only for crystals which decay during data collection. If the crystal is frozen and does not decay, then the default may be used, which is to let the eigenvalue filter define the overall scale and B factor. With a large number of batches, reliance on the eigenvalue filter is a little bit dangerous, so you should consider using a reference batch number in those cases.

This keyword is entirely equivalent to the keywords REFERENCE BATCHES, REFERENCE FILM, AND REFERENCE FILMS. The others exist because some people have a grammatical hangup about using the singular to describe more than one object.

---

<i>format</i>	REFERENCE BATCH integer integer integer ... etc.
<i>default</i>	no reference batch. Eigenvalue filter defines the overall scale and B
<i>example</i>	REFERENCE BATCH 1 3 4 5

---

**REFERENCE BATCHES, FILM, FILMS**

Same as reference batch

**REJECT HKL**

This flag tells the program to reject the list of individual h, k, l's which follow. This is useful for iterative rejection cycles since the file containing the rejected reflections can be reread. Each record after the title contains one reflection with the variables h, k, l (original index) and film number in free format. This information is most easily edited from the printed output of a previous run, can contain the whole line. One can read it from a reject file with command @reject.

---

<i>format</i>	REJECT HKL
<i>default</i>	This is not the default

---

**REJECT OUTLIERS**

Automatic rejection of outliers. Not yet reliably implemented. This is supposed to replace multiple rounds of running Scalepack. The idea is that when this keyword is set, Scalepack runs once, makes a reject.dat file, then runs again, reads the reject.dat file and scales the data based on the reduced set of observations.

**REJECTION PROBABILITY**

Applies Bayesian rejection of outliers. Rejected observations are written to the reject.dat file (see WRITE REJECTION FILE). The whole hkl (all original hkl's with the same unique hkl) with at least one observation with a probability of being an outlier greater than or equal to the value specified after the WRITE REJECTION FILE keyword (default 0.9) are written to the log file. This is an estimation on your part of how frequently you expect any observation to be an outlier. In principle, the rejection probability should be about equal to the number of outliers divided by the number of observations. A good value to use is 1/10,000, i.e. 0.0001, for normal good data. If you have a non-random signal in your background (e.g., satellite crystals, malfunctioning detector, ice rings), then you will probably have to increase the rejection probability. If you do not want to generate a reject list in the log file, then omit the keyword. A comparison of R-linear and R-squared is often helpful for deciding whether to increase the rejection probability. See the discussion following Scenario 1 for more on this.

The rejection algorithm used in Scalepack is the most sensible and statistically sound outlier rejection algorithm. Unlike some other programs, the Scalepack outlier rejection is based on comparing differences to  $\sigma$ . So it will typically reject, say 4 or 5  $\sigma$  outliers. Some reflections with a large discrepancy from the average value of I may simply represent a lack of adequate statistics in measurement and not a "mistake" (non-random error) in measurement. If a reflection has a large discrepancy from the average and a large  $\sigma$ , its contribution to the average will be very small anyway.

Here's an example of how the outlier rejection algorithm works:

Input data (after scaling)		
Obs #	I	$\sigma$
1	10.0	0.1
2	10.1	0.1
3	9.9	0.1
4	11.0	0.1
5	20.0	10.0

Although you may think that {5} is the outlier, in fact the consistent set of observations will be {1, 2, 3, 5}, and {4} will be the outlier. Why is this? A consistent average will be 10.0 (or more exactly 10.000003), because observation 5 has a 1/1000000 lower statistical weight than the other observations, due to its large  $\sigma$ . Observation 4, on the other hand, which was measured more accurately, will be 8.7  $\sigma$  deviations from the average of the remaining observations because the expected error of the difference  $4 - \langle\{1,2,3,5\}\rangle$  is 0.115, and  $8.7 = (11.0 - 10.000003) / 0.115$ . So the observation with the largest deviation from the average is not necessarily a statistical outlier.

$R_{\text{-merge}}$  (after outlier rejection) will be very bad, 25%, even if the average is measured extremely well, with a sigma of 0.057%! This has to do with  $R_{\text{-merge}}$  being an unweighted statistic, and  $I/\sigma$  being a weighted statistic. If some other program rejected observation 5 (for no good reason) and observation 4 (for good reason),  $R_{\text{-merge}}$  would be 0.067%. Practice shows that most programs would reject observation 5 because many users want a good  $R_{\text{-merge}}$ . And sometimes these programs would not even flag the true outlier - observation 4!

---

<i>format</i>	REJECTION PROBABILITY value
<i>default</i>	0.0001
<i>example</i>	REJECTION PROBABILITY 0.0001

---

## RESOLUTION

Minimum **d-spacing** for this run. Default is the maximum resolution found in the input data. One can supply two numbers in any order, and they will be minimum and maximum d-spacings.

---

<i>format</i>	RESOLUTION value [value]
<i>default</i>	highest resolution detected in .x file
<i>examples</i>	RESOLUTION 10 2.2 RESOLUTION 2.2

---

## RESOLUTION STEP

Changes the number of reflections per shell for the purposes of printing out statistics. Formally, it represents the exponent of the zone volume calculation. Normally, this is 3, because the volume of a sphere goes as radius cubed, so all the statistics shells will have the same number of reflections. Changing the resolution step could be useful to prepare a table of statistics to compare with other programs which may print out the statistics differently.

---

<i>format</i>	RESOLUTION STEP value
<i>default</i>	RESOLUTION STEP 3
<i>example</i>	RESOLUTION STEP 2.5

---

## SCALE ANOMALOUS

This is the flag for keeping Bijvoets (I+ and I-) separate both in scaling and in the output file. If the SCALE ANOMALOUS flag is on, anomalous pairs are considered non-

equivalent when calculating scale and B factors and when computing statistics, and are merged separately and output as I+ and I- for each reflection.

This is a dangerous option because scaling may be unstable due to the reduced number of intersections between images. The danger is much larger in low symmetry space groups. SCALE ANOMALOUS will always reduce R-merge, even in the absence of an anomalous signal, because of the reduced redundancy. However,  $\chi^2$ 's will not be affected in the absence of an anomalous signal.

---

<i>format</i>	SCALE ANOMALOUS
<i>default</i>	This is not the default. Use with extreme caution.

---

### SCALE RESTRAIN

Can be used to restrain scale factor differences from consecutive films or batches. The value which follows the flag represents the amount you will allow the scale factors to differ from consecutive films or batches. It adds a factor of  $(\text{scale1} - \text{scale2})^2 / (\text{scale restrain})^2$  to the target function minimized in scaling. This only applies to batches between which you add partials. For very thin frames, this is almost obligatory. The value should roughly represent the expected relative change in scale factors between adjacent frames.

---

<i>format</i>	SCALE RESTRAIN value
<i>default</i>	not turned on
<i>example</i>	SCALE RESTRAIN 0.01 (expect 1% change between adjacent frames)

---

### SECTOR

Substitutes for the ### wildcard to specify a group of files to be read. See FILE keyword.

---

<i>format</i>	SECTOR integer to integer
<i>default</i>	no default
<i>example</i>	SECTOR 1 to 40

---

### SECTOR WIDTH

Sector width can be specified in degrees. A pseudofilm is a sector width's worth of data from one detector. Valid for area detector data only, default value 5 degrees.

---

<i>format</i>	SECTOR WIDTH value
<i>default</i>	5 degrees
<i>example</i>	SECTOR WIDTH 3.0

---

**SIGMA CUTOFF**

Cutoff for rejecting measurements on input. Default = -3.0. Be careful if you increase this!

What is the rationale for using  $\sigma$  cutoff -3.0 in Scalepack? Wouldn't you want to reject all negative intensities? Why shouldn't you use a  $\sigma$  cutoff 1.0 or zero? The answer is as follows: The best estimate of  $I$  may be negative, due to background subtraction and background fluctuation. Negative measurements typically represent random fluctuations in the detector's response to an X-ray signal. If a measurement is highly negative ( $\leq -3\sigma$ ) than it may be more likely the result of a mistake, rather than just random fluctuation. If one eliminates negative fluctuations, but not the positive ones before averaging, the result will be highly biased. In Scalepack, SIGMA CUTOFF is applied before averaging. If one rejects all negative intensities before averaging a number of things would happen:

1. The averaged intensity would always be positive;
2. For totally random data with redundancy 8, in a shell where there was no signal, there would be on average 4 positive measurements, with average intensity one  $\sigma$ . This is because the negative measurements had been thrown out. So the average of the four remaining measurements would be about 2  $\sigma$ ! This would look like a resolution shell with a meaningful signal!
3.  $R_{\text{-merge}}$  would always be less than the  $R_{\text{-merge}}$  with negative measurements included
4. A SIGMA CUTOFF of 1 would improve  $R_{\text{-merge}}$  even more, by excluding even more valid measurements!

Why should this worry you? Exclusion of valid measurements will deteriorate the final data set. One may notice an inverse relationship between  $R_{\text{-merge}}$  and data quality as a function of SIGMA CUTOFF So much for using  $R_{\text{-merge}}$  as any criterion of success.

Even the best (averaged) estimate of intensity may be negative. How to use negative  $I$  estimates in subsequent phasing and refinement steps? The author of Scalepack suggests the following:

1. You should never convert  $I$  into  $F$
2. You should square  $F_{\text{calc}}$  and compare it to  $I$ . Most, but not all of the crystallography programs do not do this. That is life. In the absence of the proper treatment, one can make approximations. One of them is provided by French and also by French and Wilson and implemented in the CCP4 program TRUNCATE. A very simplified and somewhat imprecise implementation of TRUNCATE is this: if  $I > \sigma(I)$ ,  $F = \sqrt{I}$  if  $I < \sigma(I)$ ,  $F = \sqrt{\sigma(I)}$

---

<i>format</i>	SIGMA CUTOFF value
<i>default</i>	-3
<i>example</i>	SIGMA CUTOFF -2.5

**SPACE GROUP**

Space group symbol from the list below. This input is required! The space group may be entered as a name (e.g., P212121) or as a number (e.g., 19, for the same space group). Most of the numbers correspond to those of the International Tables. The numbers above 230 are nonstandard definitions of space groups.

1	P1	89	P422	154	P3221	207	P432
3	P2	90	P4212	155	R32	208	P4232
4	P21	91	P4122	168	P6	209	F432
5	C2	92	P41212	169	P61	210	F4132
16	P222	93	P4222	170	P65	211	I432
17	P2221	94	P42212	171	P62	212	P4332
18	P21212	95	P4322	172	P64	213	P4132
19	P212121	96	P43212	173	P63	214	I4132
20	C2221	97	I422	177	P622	303	P2C
21	C222	98	I4122	178	P6122	305	B2
22	F222	143	P3	179	P6522	318	P21221
23	I222	144	P31	180	P6222	401	C1
24	I212121	145	P32	181	P6422	403	P21C
75	P4	146	R3	182	P6322	446	H3
76	P41	149	P312	195	P23	455	H32
77	P42	150	P321	196	F23	501	I1
78	P43	151	P3112	197	I23	503	I2
79	I4	152	P3121	198	P213	505	C21
80	I41	153	P3212	199	I213		

Notes to particular space groups:

146 R3 R3 in hexagonal setting

446 H3 R3 in primitive setting

155 R32 R32 in hexagonal setting

455 H32 R32 in primitive setting

401 C1 Non-standard, but useful to make angles close to 90.

501 I1 Non-standard, but useful to make angles close to 90.

303 P2C P2, C axis unique

403	P21C	P21, C axis unique
305	B2	like C2, B face centered, c axis unique
503	I2	Non-standard, but useful to make beta angle close to 90.

**SPINDLE AXIS**

Alias for ORIENTATION AXIS 2.

**UNIT CELL**

Real cell specified as a, b, c, alpha, beta, gamma. **UNIT CELL** is included in the header of *Denzo\_IP*, *Denzo\_york1*, *ARCHIVE*, and *Scalepack* files. Otherwise, it must be included.

---

<i>format</i>	UNIT CELL value value value angle angle angle
<i>default</i>	the value from the first header encountered
<i>example</i>	UNIT CELL 50 62 100.3 90 90 90

---

The postrefined value of the unit cell constants is not used, nor is it output in the *Scalepack* output file. You must get this information from the log file if you are interested in it.

**VERTICAL AXIS**

Same as ORIENTATION AXIS 1

**WRITE BADDIES**

Writes \*.xrej files so that reflections from the reject file may be displayed by XdisplayF. Not fully implemented. Will terminate the program and prevent scaling and postrefinement.

Note that the command @reject must precede the input of the .x files.

**WRITE REJECTION FILE**

Creates a file with *hkl*'s to be rejected. UNIX file name: *reject*, VMS file name:

REJECT.DAT. Reject file is created if it does not exist. If the file exists, it is overwritten.

You can specify the threshold for what you consider to be a **rejectable probability**. The default is 0.9, which is fairly safe, but you may want to decrease this to, say, 0.5 on later rounds of rejection.

---

<i>format</i>	WRITE REJECTION FILE value
<i>default</i>	Does not write the rejection file, but if it does, the default value is 0.9
<i>example</i>	WRITE REJECTION FILE 0.5

---



# ***Table of Figures***

Figure 1. The Project tab.....	6
Figure 2. The Data tab .....	7
Figure 3. The Auto tab.....	8
Figure 4. The Summary tab .....	9
Figure 5. The Index tab.....	10
Figure 6. The Strategy tab.....	11
Figure 7. The Integrate tab.....	12
Figure 8. The Scale tab .....	13
Figure 9. The Merge tab.....	14
Figure 10. The Macros tab.....	15
Figure 11. The HKL-2000 Start Icon.....	18
Figure 12. The HKL-2000 launch and site selection screen. ....	20
Figure 13. The "New Site" selection panes.....	21
Figure 14. The main dialog for site creation.....	23
Figure 15. The Advanced and Description tabs of the site configuration window.....	24
Figure 16. The directory tree .....	25
Figure 17. The Raw Data directory.....	26
Figure 18. The Output Data directory.....	26
Figure 19. The Create New Directory dialog.....	26
Figure 20. The << button can make directory shortcuts.....	27
Figure 21. The Load/Create New Sets window .....	27
Figure 22. The Load Data Sets window.....	28
Figure 23. The Single Set window.....	28
Figure 24. The Show All Files dialog can be used for file selection. ....	29
Figure 25. Files listed after activating the Scale Sets Only button .....	30
Figure 26. The Split Set window .....	32
Figure 27. Data set divided into two sets removing four images in the middle. ....	32
Figure 28. The experimental details panel.....	33
Figure 29. The Summary tab showing two datasets .....	33
Figure 30. The Edit Sets button is in the Set Controls box.....	34
Figure 31. The Edit Sets dialog.....	34
Figure 32. Starting a project.....	36
Figure 33. Obtain a sequence from a database. ....	37
Figure 34. The anomalous element wizard. ....	38
Figure 35. The anomalous scattering plot for mercury.....	38
Figure 36. The Create/Edit crystal dialog .....	39
Figure 37. The Project tab after creating a project.....	40
Figure 38. The Save Project dialog.....	41
Figure 39. The Load Project dialog .....	41
Figure 40. The XDisp selector .....	42
Figure 41. The Display image button in the data tab.....	42
Figure 42. The XDispQt display .....	43

## TABLE OF FIGURES

Figure 43. Two zoom levels are shown .....	43
Figure 44. Frame up and down buttons .....	44
Figure 45. The Histogram controls the display contrast .....	44
Figure 46. The Frame button from the Display window .....	45
Figure 47. The buttons for adjusting the image .....	45
Figure 48. After peak search .....	47
Figure 49. XDispX during the "Peak Search" .....	48
Figure 50. Peak Search options.....	50
Figure 51. The Pending Sets box .....	53
Figure 52. The summary information about processed sets .....	53
Figure 53. The Resolution limits box .....	53
Figure 54. Other pre-indexing options.....	54
Figure 55. The Zoom window with image after auto-indexing .....	55
Figure 56. The Refinement Options box with the Fit Basic parameters selected.....	56
Figure 57. Refinement statistics generated using Fit Basic parameters.....	56
Figure 58. The 3D window in Controls box .....	57
Figure 59. The Bravais Lattice table.....	<b>Error! Bookmark not defined.</b>
Figure 60. The refinement information using Fit All .....	61
Figure 61. The Macros dialog .....	62
Figure 62 The Profile Fitting Radius is set in the Integration Box .....	65
Figure 63. The profile fitting radius can be displayed and moved around. ....	66
Figure 64. Profile fitting radius button .....	67
Figure 65. A Single Reflection .....	68
Figure 66. The Integration Box and Zoom windows are shown with simple options (left), radial options (center), and elliptical options (right).....	69
Figure 67. Predicted reflections at 2 sigma, 20 sigma, and all reflections. ....	70
Figure 68 The Prediction selector .....	70
Figure 69. Resolution circles dialog and the displayed circles. ....	71
Figure 70. The Reference Zone window.....	73
Figure 71. The Crystal Information window .....	75
Figure 72. The Mask dialog (left) and setting the polygon of Blind region 1 (right).....	76
Figure 73. The Blind Region definition window and resulting blind region.....	77
Figure 74. The blank Strategy tab (chart areas compressed to show them all). ....	79
Figure 75. The Delta Omega vs. Omega Start plot.....	81
Figure 76. The I/Sigma vs. Resolution plot .....	81
Figure 77. The Estimated Crystal Dose window .....	82
Figure 78. The Simulation window with almost no overlaps .....	83
Figure 79. Completeness and Redundancy vs. Resolution plots .....	84
Figure 80. The Combine panel.....	84
Figure 81. The Integrate buttons colored green, orange, red .....	85
Figure 82. The special integration option and the default integration commands.....	86
Figure 83. The diagnostic plots: $\chi^2$ vs. Frame and Distance Change vs. Frame .....	87
Figure 84. The mosaicity histogram .....	88
Figure 85. The program is waiting for new frames during processing the data .....	90
Figure 86. The Scaling Options panel.....	92
Figure 87. The global refinement options.....	97

## TABLE OF FIGURES

Figure 88. The Scalepack output files.....	97
<i>Figure 89. The Scaling Control Box will have some options disabled before scaling</i> .....	98
Figure 90. Global Refinement with post-refinement information .....	98
Figure 91. Global Refinement with post-refinement information and auto-corrections .....	99
Figure 92. Global statistics with no post-refinement .....	99
Figure 93. The Scale and B vs. Frame and the Completeness vs Resolution and I/Sigma and CC <sub>1/2</sub> vs. Resolution graphs .....	101
Figure 94. I/Sigma and CC <sub>1/2</sub> vs. Resolution .....	102
Figure 95. Anomalous signal detection for a strong (left) and moderate (right) anomalous signal. ....	103
Figure 96. The Scaling Control panel .....	103
Figure 97. Adjusting the error model.....	104
Figure 98. Redundancies Window .....	105
Figure 99. The Reindexing dialog .....	106
Figure 100. The reciprocal lattice viewer .....	107
Figure 101. Reprocessing options.....	108
Figure 102. The Integration Graphs from the Inspect Integration option.....	109
Figure 103. Frames selected to be excluded from scaling are marked red .....	110
Figure 104. The Space Group Probability results from POINTLESS.....	111
Figure 105. Manual space group selection .....	112
Figure 106. The Project Tab after scaling two data sets .....	120
Figure 107. Use the Statistics button to see scaling results .....	121
Figure 108. The Raw Data Sets window .....	122
Figure 109. During automatic data processing. ....	123
Figure 110. Automatic integration results .....	123
Figure 111. Automatic integration diagnostic graphs.....	124
Figure 112. Automatic scaling results .....	124
Figure 113. Selecting several files to merge.....	125
Figure 114. Files selected to merge. ....	126
Figure 115. The merged file will appear in the list.....	126
Figure 116. The Macros Tab.....	127
Figure 117. Data Set specific macros .....	128
Figure 118. Select to install HKL-2000.....	131
Figure 119. Read and Accept the HKL-2000 License.....	132
Figure 120. Create a new installation or upgrade an existing one .....	133
Figure 121. Select the installation directory .....	133
Figure 122. External program selection.....	133
Figure 123. Finish the installation .....	134
Figure 124. The beam stop thinned in the center shows the direct beam position. ....	138
Figure 125. The <b>SET BEAM POSITION</b> button opens the beam position dialog.....	138
Figure 126. The Beam Position Dialog in the normal and activated state.....	139
Figure 127. Parameters of a reflection.....	146