

Soma's Data Processing Notes

Notes on Data Processing-1

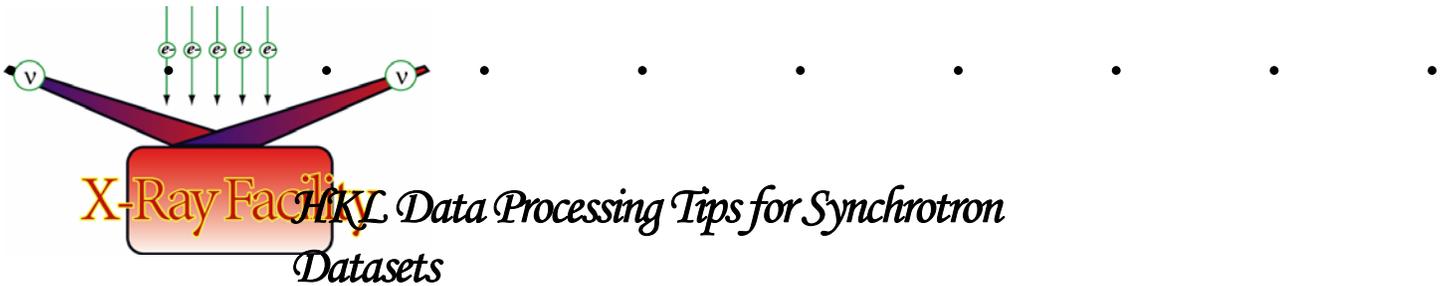


TABLE OF CONTENTS

Introduction	1
License	1
HKL.1.97.9	2
Synchrotron data	2
<i>Wavelength</i>	2
<i>Keyword 'step'</i>	2
<i>'Fixing' the distance</i>	3
<i>Keyword 'y scale'</i>	1
<i>Keyword 'film rotation'</i>	2
Conclusions	3

© 2000-2004 Thayumanasamy Somasundaram
414 Kasha Laboratory
Institute of Molecular Biophysics, Florida State University,
Tallahassee, FL 32306-4380
E-mail: soma@sb.fsu.edu • URL: <http://www.sb.fsu.edu/~soma>
Phone: 850.644.6448 (Office) • 850 645 1333
Fax: 850.644.7244
August 10, 2004

Data Processing Tips-1

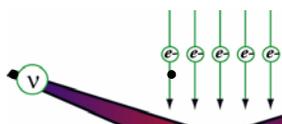
HKL Suite Data Processing Tips for Synchrotron Datasets

Version: Aug.10, 2004; Original: May 10, 2003

Introduction

[HKL Research Inc](#) has recently released a new version of HKL Suite (Denzo, XdisplayF, and Scalepack). This version is called HKL 1.97.9 and has several modifications including, slightly modified Graphical User Interface (GUI), ability to read and process several new detector formats including Quantum 4 CCD, and handle large number of reflections for scaling. The new executables have new names, especially those ones used for scaling.

Processing of synchrotron data necessitates the ability to read charge-coupled device (CCD) detectors. Most of the synchrotron sites have [Area Detector Systems Corporation](#)'s (ADSC) Quantum Q4 detector. Some synchrotrons have [Mar USA](#)'s MarCCD165 or 225. Our own CCD detector is a MarCCD165. In order to read these formats, specific keywords have to be used in Denzo and XdisplayF ([see below](#))



X-Ray Facility

X-Ray Facility has acquired one new license for version 1.97.9 (currently XRF has one license for version 1.97.2, one for 1.96.9). Each version can be used by ten (10) different people or in different computers simultaneously. This means that synchrotron data can be processed in HP Alphas and Linux operating systems.

License

List of computers with licenses: HKL executable will not run on a computer that does not have a license to run the program since licenses are tied to specific hardware found on the computer. The following table shows the list of computers that have licenses to run various versions of HKL suite. Please let [Soma](#) know if there is an error in this table.

No	Computer	HKL 1.97.9	1.97.2	HKL.1.96.9
1	Arg.sb.fsu.edu	Yes	Yes	Yes
2	Lys.sb.fsu.edu	Yes	Yes	Yes
3	Leu.sb.fsu.edu	Yes	Yes	Yes
4	Tyr.sb.fsu.edu	Yes	Yes	Yes
5	Raccoon.sb.fsu.edu	Yes	Yes	Yes
6	Neptune.sb.fsu.edu	Yes	Yes	Yes
7	Tampa.sb.fsu.edu	Yes	Yes	Yes

8	Orlando.sb.fsu.edu	Yes	Yes	No
9	Mozart.sb.fsu.edu	Yes	Yes	No
10	Dallas.sb.fsu.edu	Yes	Yes	No
11	Miami.sb.fsu.edu	Yes	Yes	No

Table 1. List of computers with HKL licenses. **Orange:** Alphas; **Green:** Linux machines; **Yellow:** X-Ray PI Linux machines

All executables are located in one central directory (see below) and you are NOT required to copy any of the executables to your home directory to run these programs. Instead, create an alias that points to one of the executables.

```
Alpha version: /tyr/e/users/soma/HKL.1.97.9/
Alpha version: /tyr/e/users/soma/HKL.1.96.9/
Linux version: neptune:/usr/local/xray/HKL.1.97.9/
Linux version: neptune:/usr/local/xray/HKL.1.96.6/
Linux version: raccoon:/usr/local/xray/HKL.1.97.9/
Linux version: raccoon:/usr/local/xray/HKL.1.96.6/
```

HKL.1.97.9

Version 1.97.9 executables have slightly different names. The new names are given below along with their capabilities:

```
xdisp
denzo
scalepack
```

```
scalepack8m
scalepack16m
scalepackmanyframes
```

Scalepack8m can handle 8×10^6 observations, 60,000 hkl pairs, 2000 frames and 50,000 rejections. For scalepack16m, the corresponding numbers are: 16×10^6 , 60,000 pairs, 2000 frames and 100,000 rejections. For scalepackmanyframes the relevant numbers are: 2×10^6 , 100,000 pairs, 4000 frames and 50,000 rejections.

Version 1.6.0 and upward require only one display program called 'xdisp'. Xdisp executable along with a modifier can read various formats. Given below are the various modifiers needed to read home and synchrotron data. Add the relevant modifier to 'xdisp' to read a specific format, e.g., 'xdisp raxis myxtal001.osc' for R-axis 105 μm data collected at home.

```
xdisp ↵ raxis ↵ myxtal001.osc | Regular r-axis data
```

```
xdisp ↵ raxis ↵ 210 ↵ myxtal001.osc | Small r-axis format
```

```
xdisp ↵ raxis2n ↵ myxtal001.osc | New r-axis data (note 2n)
```

```
xdisp ↵ raxis2n ↵ 210 ↵ myxtal001.osc | New small r-axis format (note 2n)
```

```
xdisp ↵ ccd ↵ unsupported-m165 ↵ xtal01.001 | MarCCD165 format
```

```
xdisp ↵ ccd ↵ adsc ↵ unsupported-q4 ↵ xtal01.001 | Quantum 4 format
```

↵ : indicates a required empty space.

The same modifiers are required while processing the data using Denzo with the keyword 'format'. In the following section, some tips for processing data sets collected specifically at synchrotron are discussed.

Synchrotron data

As mentioned in the Introduction, synchrotron data sets are usually collected in Quantum 4 CCD or MariCCD225 formats which are different from Image Plate detector format. Since only Denzo and XdisplayF deal with the different formats, most attention should be given to `auto.dat`, `site.dat`, and `expr.dat` (or the equivalent of these files). There are several important differences in the data set collected at synchrotrons. The first one is the wavelength of x-rays. This value can be anywhere between 0.9 to 1.5Å depending upon the experimental station. The second major difference is that the phi axis or spindle axis is usually horizontal (in R-Axis it is vertical and in Mar it is horizontal). The third difference is rotation of spindle can be clockwise or anti-clockwise. The fourth difference is 'Y Scale' which is the ratio of pixel size on fast reading direction (say y) to slow reading direction. The next difference is the film rotation.

Wavelength

Home sources (rotating anodes) are limited to a few distinct wavelengths depending upon the target used (copper, molybdenum, gold, or iron). In macromolecular crystallographic facilities, it is usually a copper target and therefore the wavelength λ is 1.541Å.

However, in synchrotron it is possible to get wavelengths in the range from 0.9 to 1.5Å and the user has to modify the Denzo keyword 'wavelength' appropriately. Given below is an illustration where a data set collected at 0.948Å is processed with wrong wavelength of 1.541Å. For comparison, data processed with the correct wavelength of 0.948Å is also shown. Note that in the wrong processing the predictions (overlaps) are much closer compared to the correct processing. The log file will also show that crystal cell dimensions are about 1.6 times ($1.541/0.948=1.62$) larger for the wrong one compared to the correct one.

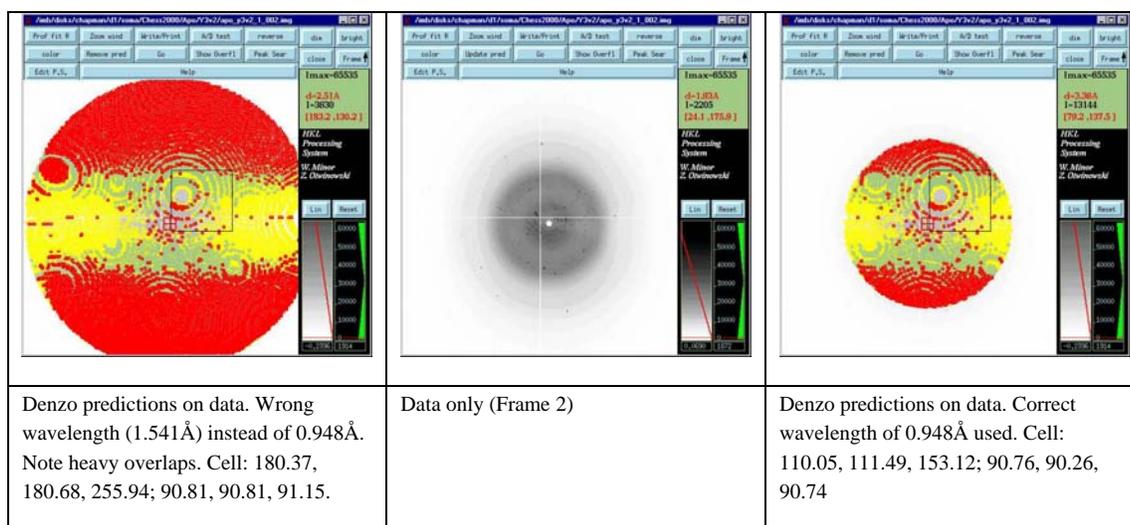


Figure 1. Incorrect & correct wavelength usage during synchrotron data processing

Keyword 'step'

Synchrotron phi axis rotation can be different from that we usually see at home. One such example is at [F1 Station](#) at [CHESS](#) where they use anti-clockwise rotation to achieve the movement of phi axis. To distinguish this we have to add a special Denzo keyword 'step' as

```
oscillation start 0.0 step -1.0 range 1.0
```

Note the special keyword `step= -1.00` specifying different rotation convention. If you don't use this keyword, your first image will be indexed nicely, since you are not rotating the crystal very much. However, subsequent images will be terribly off due to the rotation of the data opposite to that of predictions.

Given below are example images of this problem. In the first set, there are three images (the very first frame of the data set). Of this set the middle, one is the real data, left side is predictions on top of data with the wrong convention and right side is predictions on top of data with the correct convention. Of course, you do not see much of a problem. However, the second set of three images shows the same series but for the fifth frame in the data set. That means the crystal has rotated at least several degrees (here five degrees in the opposite direction). Here one can clearly see the mismatch between the data and predictions.

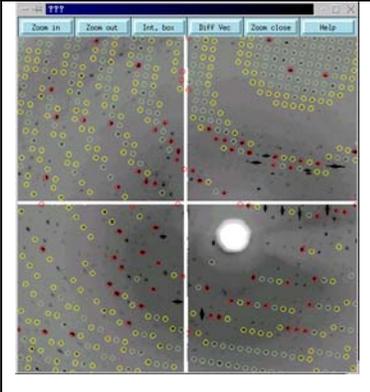
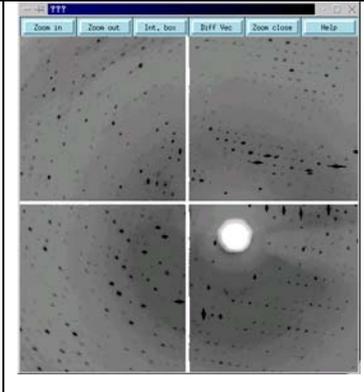
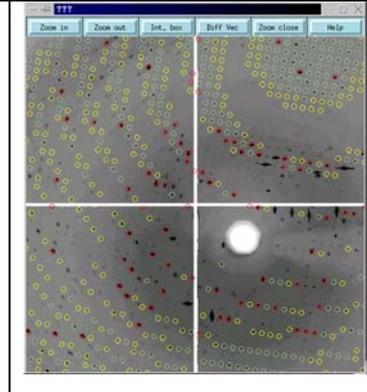
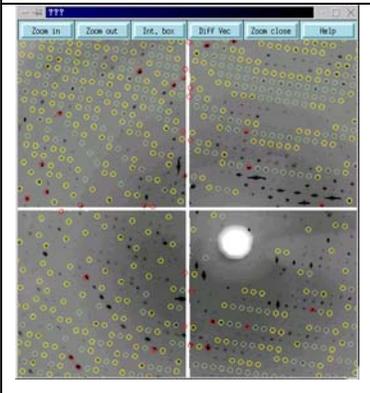
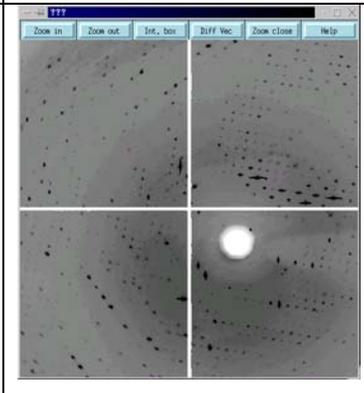
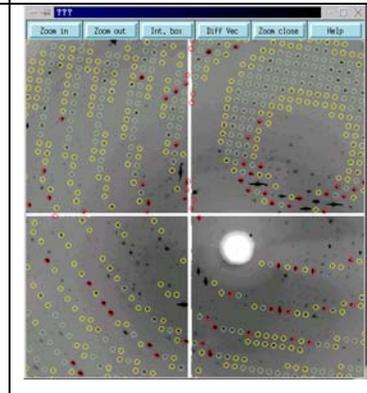
		
Denzo predictions on data (Frame 1). Looks okay!	Data only (Frame 1)	Denzo predictions on data (Frame 1). Looks okay!
		
Denzo predictions on data (Frame 5). Wrong convention. Predictions are off.	Data only (Frame 5)	Denzo predictions on data (Frame 5). Correct convention. Reasonably good predictions.

Figure 2. Incorrect and correct predictions while using the keyword 'step'

'Fixing' the distance

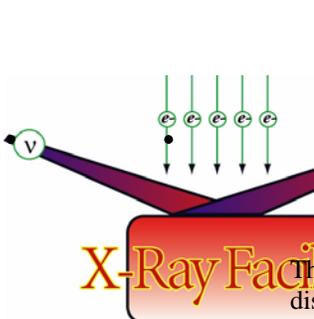
For some data sets, it may become necessary to fix the distance (we have determined the distance, wavelength very accurately and we know the correct cell dimensions) while refining other parameters such as crystal `rotx`, `roty`, `rotz`, and `mosaicity`. If you use the regular `refine.dat` file, it is impossible to keep the distance fixed even if you use key words like:

```
Fit rotx roty rotz
Go go go
Distance 125.0          (you trying to keep it constant)
Go go go
```

However, the program will not recognize the fact that you are trying to keep the distance constant. Instead, it will refine the distance. This is shown in the log file excerpts below:

```
Auto index unit cell 111.05, 111.49, 153.12, 90.76, 90.2, 90.74
Real space unit cell, a 123.18 b 122.66 c 188.87
Real space unit cell, a 124.50 b 123.80 c 193.04
Detector to crystal distance 125.00 / Your input value
Distance      138.950 shift 13.950 error 0.531
Detector to crystal distance 138.95 / But program has refined to a different value
Distance      140.351 shift 1.401 error 0.476
```

HKL authors recommend use of the following compact `refine.dat` for these special cases:



```
Start refinement
Resolution limits 50.0 2.50
Fit all fix distance go (Keep distance constant)
Calculate go
End of pack
Go go go
```

This will produce the expected result of reasonable variation in cell dimension and fixed distance value as shown from the excerpts of log file:

```
Auto index unit cell 111.0, 111.49, 153.12, 90.76, 90.26, 90.74
Real space unit cell, a 110.74 b 111.14 c 153.74
Real space unit cell, a 111.17 b 111.11 c 154.54
Detector to crystal distance 125.00
Detector to crystal distance 125.00
Detector to crystal distance 125.00
```

Keyword 'y scale'

As mentioned in the beginning, keyword 'y scale' is used to correct the anisotropy in the pixel dimensions. Y scale for r-axis has a dimensionless value of 1.03448 (105 μm /101.5 μm). However, most CCD have a value of exactly one (actually it is -1 and is not refined). Now let us see what happens if this value is accidentally entered as 1.54 (home wavelength value).

You can immediately notice the center beam has moved way off to the right in the incorrect usage. You will not notice the difference in the log file. The log file will continue to show the 'correct x beam and y beam' but if you tried to auto index, it will fail.

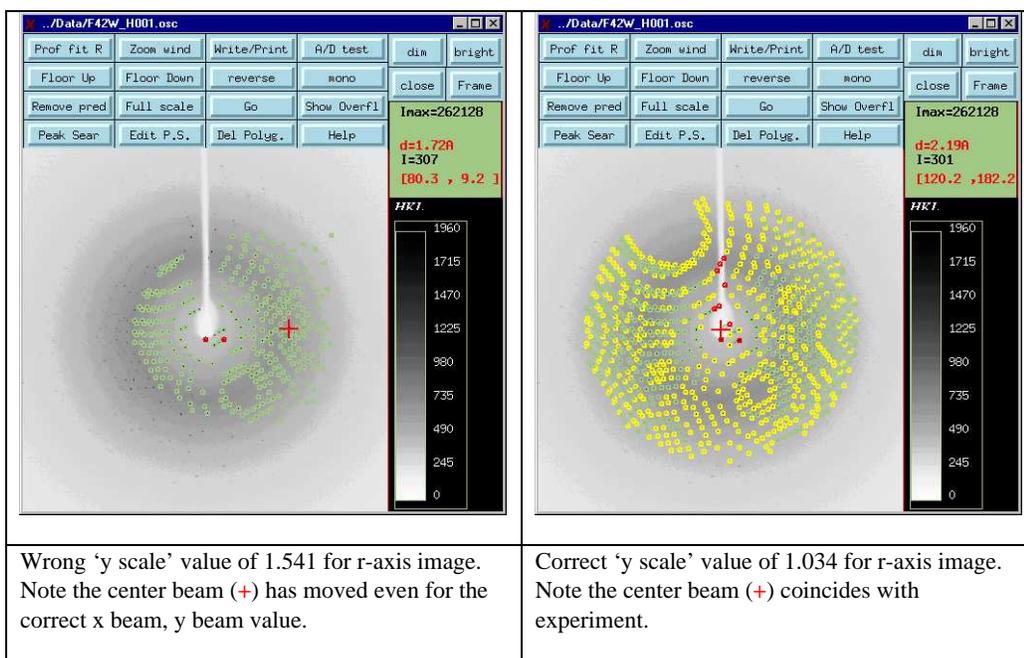


Figure 3. Incorrect and correct value for 'y scale' for r-axis image

Keyword 'film rotation'

This keyword describes the orientation of the spindle relative to the detector. Spindle orientation is usually fixed for a particular home detector and therefore it is tied to the keyword format. However, it may not be so for synchrotron detectors. So, pay attention to the correct rotation. If you are in doubt, consult the HKL manual.

Given below are example images of this problem and it may be immediately obvious that there is a problem. In this set, there are three images. In this set, the middle image is the real data, left side is predictions on top of data with the wrong convention (film rotation of 180°, usual for home source r-axis) and right side is predictions on top of data with the correct convention (film rotation 90° for ADSC CCD Q4).

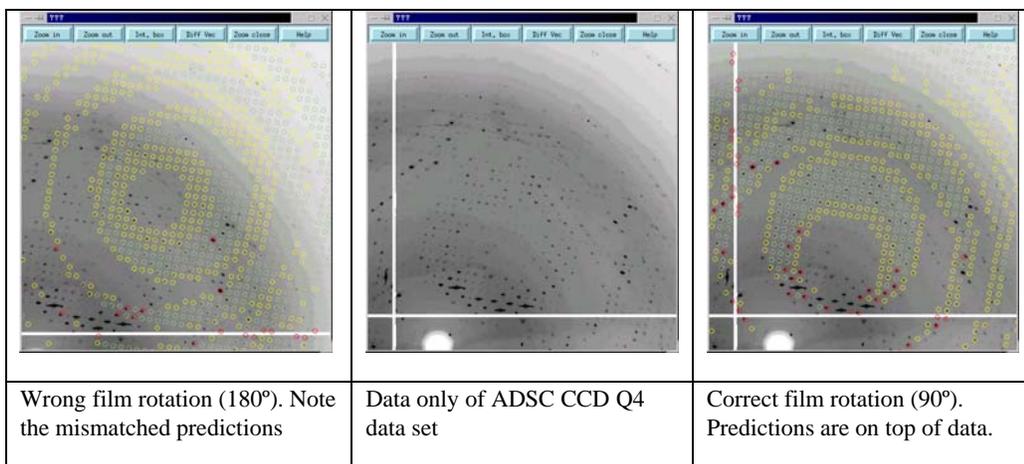


Figure 4. Incorrect and correct film rotation convention for ADSC CCD Q4 data

Check the log file to see what film rotation convention is being 'assumed' by Denzo, since it is possible to start the auto indexing **without** specifying the film rotation. In this case, Denzo will 'assume' it based on the format keyword.

Conclusions

For processing synchrotron data therefore, one has to pay attention to make sure that the correct wavelength (check before leaving the synchrotron), correct Y scale, correct oscillation step, and correct film rotation values are entered. Another point is that several display programs associated with different formats also specify the 'x beam' and 'y beam' differently. So, double-check the correct values. For example, x and y values in R-axis format is exactly opposite of the Denzo convention (R-Axis x >> Denzo y; R-Axis y >> Denzo x).

More tips will be provided soon. Some of the materials used in this write-up can be found in Linux PC (flame.sb.fsu.edu) under the following directory:

rtgpc3:/imb/d1/users/soma/X-RayTutorial

Thanks to Mohammad for letting me play with his data.

List of figures included in this article and their locations:

Top level directory: E:\My Documents\Images

Figure 1. apo_fr2_pw.jpg, apo_fr2_x.jpg, apo_fr2_pw1.jpg

Figure 2. apo-fr1-px.png, apo-fr1-x.png, apo-fr1-x1.png; apo-fr5-px.png, apo-fr5.x.png, apo-fr5-px1.png

Figure 3. y_scale_incorrect.jpg, y_scale_correct.jpg

Figure 4. apo_fr5-pr.png, apo_fr5-r.png, apo_fr5-pr1.png