

XIA2 – a brief user guide

*Graeme Winter, STFC Daresbury Laboratory Warrington WA4 4AD, United Kingdom
August 2007*

Introduction

xia2 is an expert system for reducing diffraction data from macromolecular crystals, which makes use of existing data reduction software, including Mosflm (Leslie, 1992), Labelit (Sauter et Al, 2004), Pointless (Evans, 2006), CCP4 (Bailey, 1994) and XDS (Kabsch, 1993). There are two main programs of interest –

- xia2setup
- xia2

xia2setup is used to create the configuration file – which can of course be composed by hand, while xia2 performs the actual data reduction. In this article I will describe each of these programs, what they do and how they are used.

xia2setup

This program was written in response to feedback that writing the input file for xia2 was too time consuming. The setup program will read all of the headers of the images using diffdump – a program written using the Diffraction Image library (Remacle and Winter, 2006). This may take a moment or two for substantial data sets, so if you know what you are doing it may be simpler to just write the input file yourself. Come on – you know you want to!

The input file looks a little like this:

```
BEGIN PROJECT DEMO
BEGIN CRYSTAL INSULIN

BEGIN HA_INFO
ATOM S
NUMBER_PER_MONOMER 6
END HA_INFO

BEGIN WAVELENGTH SAD
WAVELENGTH 1.488
END WAVELENGTH SAD

BEGIN SWEEP SAD
WAVELENGTH SAD
DIRECTORY c:\data\insulin\images
IMAGE insulin_1_001.mar2300
END SWEEP SWEEP1

END CRYSTAL INSULIN
END PROJECT DEMO
```

This starts by establishing the project and crystal information – these will be used for harvesting and will end up in the resulting MTZ files. For each crystal we have some information about the wavelengths of data which were collected (the names of which will end up as WAVELENGTH_IDs in the harvest files and DATASETS in the MTZ files) and the sweeps which were measured for these wavelengths. If more than one sweep is assigned to a wavelength the reflections will be merged to form a single dataset – this is appropriate for low and high resolution passes, for instance. This example is for sulphur SAD data measured from cubic insulin on SRS station 7.2. This input file can be created with the command:

```
xia2setup -project DEMO -crystal INSULIN -atom s c:\data\insulin\images
```

which will write the results to the screen – you may want to pipe this to a file called e.g. demonstration.xinfo. If you have MAD data and you have left a scan file in this directory xia2setup will run Chooch and have a guess at the wavelength names, e.g. LREM, INFL, PEAK, HREM. If you do not specify a heavy atom and there is only one wavelength of data then xia2setup will guess that this is native data. If you have Labelit installed, this will be run to refine the beam centre prior to processing as well, so that you can ensure that the results look correct.

The way I would recommend using this is to run xia2setup to generate an input file, then open this file in your favourite editor to check that all of the names etc. are sensible. Note well – xia2setup will assume that all images in a directory come from the same crystal so if this is not the case you will need to edit the .xinfo file appropriately!

Now that we have the input file written we can move on to actually running the program.

xia2

xia2 is run very simply – you just need to run

```
xia2 -xinfo demonstration.xinfo
```

There are, however, a couple of useful command line options.

By default xia2 will run with the “2d” pipeline – that is, using Mosflm to perform the integration which uses two-dimensional profile fitting. This can be made explicit by adding “-2d” to the command line. Alternatively, if you wish to use XDS in the place of Mosflm use “-3d” on the command line. This tends to do a better job for highly mosaic data and lower resolution data (e.g. less than 2.5Å resolution.)

The full breakdown of useful options is:

<code>[-quick]</code>	hurry – don’t make too much effort to be thorough
<code>[-migrate_data]</code>	move the data to e.g. /tmp on the local machine
<code>[-2d]</code> or <code>[-3d]</code>	select the integration program to use
<code>-xinfo foo.xinfo</code>	specify the input file

In addition, if you are using XDS you can specify a number of processors your machine has (xia2 will then use xds_par) – this is done with "`-parallel N`" where N is e.g. 2, 4. I do not recommend using `-quick` with `-3d`.

`-migrate_data` is useful if you are accessing the images over a relatively slow NFS connection – this will move them to a local disk which should result in a net improvement in processing time. Also useful if you are having problems with auto-mounted disks.

`-quick` cuts out some of the refinement of the resolution parameters but is useful if you want to get a data set while the sample is still available to confirm the pointgroup, quality of diffraction etc.

What does it do?

In a nutshell it will process all of your data for you, merging multiple sweeps within wavelengths together and scaling together data from multiple wavelengths e.g. for MAD experiments. The result is a reflection file suitable for immediate use in your favourite phasing pipeline, e.g. Mr BUMP, Happy, Phenix, Crank...

Along the way xia2 will figure out the correct pointgroup (with a *lot* of help from pointless) and have a good guess at the spacegroup. If it looks like the spacegroup is e.g. P 2 21 21 it will also reindex the data to the standard setting – P 21 21 2 in this case.

If you do not specify `-quick` – and you therefore allow xia2 to "do it's stuff" – each sweep will be integrated once to get an idea of where the observations are reasonable to, then again with this resolution limit to ensure that all of the profiles used are ok. Once all sweeps are integrated a preliminary scaling takes place, which may involve a number of "data shuffling" steps where the correct pointgroup and setting is assigned to each set of integrated intensities. From this a more reasonable resolution limit is determined – currently where I/σ is about 2, which is then fed back to the integration if necessary to reprocess to a lower resolution. If necessary, the integration is repeated and then the full scaling takes place, where the scaling parameters are adjusted to optimise the error parameters and so on.

Once the integration and scaling steps are finished, the data are converted to F's using Truncate, merged together with CAD, have a single unit cell applied, twinning tests and so on are performed and a FreeR column is added. This is very helpful for MAD datasets as it can be quite time consuming and fiddly to do by hand.

If the results of the pointgroup analysis at the beginning of scaling indicate that the lattice used for integration was wrong, this will be eliminated from the set of possible indexing solutions and the processing repeated from the beginning. This is useful for cases where the lattice symmetry appears higher than the real crystal symmetry. Alternatively, for cases where you have e.g. a monoclinic cell with beta nearly but not exactly 90 degrees, the indexing solution may have already been eliminated based on the results of postrefinement.

Although it may take only a few minutes for a small data set (say 90 degree 1.8Å native set) it has been known to take several hours for huge sets, for example with 2000 frames...

Output

While xia2 is working it will provide a running commentary of what is going on to the screen. If you specify “-debug” on the command line you will see an awful lot more of this! I don’t recommend this unless you are reporting a bug ;o)

The single most important part of the output is the list of citations you should include in your paper – xia2 is using lots of programs written by people who work really hard on them, and they should be acknowledged. This looks something like:

```
XIA2 used... ccp4 distl labelit mosflm pointless scala
Here are the appropriate citations (BIBTeX in xia-citations.bib.)
Bailey, S. (1994) Acta Crystallogr. D 50, 760--763
Evans, P.R. (1997) Proceedings of CCP4 Study Weekend
Evans, Philip (2006) Acta Crystallographica Section D 62, 72--82
Leslie, AGW (1992) Joint CCP4 and ESFEACMB Newsletter on Protein
Crystallography 26
Leslie, Andrew G. W. (2006) Acta Crystallographica Section D 62, 48--57
Sauter, Nicholas K. and Grosse-Kunstleve, Ralf W. and Adams, Paul D. (2004)
Journal of
Applied Crystallography 37, 399--409
Zhang, Z. and Sauter, N.K. and van den Bedem, H. and Snell, G. and Deacon, A.M.
(2006)
J. Appl. Cryst 39, 112--119
```

For those who use LaTeX, these citations are also provided in BibTeX format in xia2-citations.bib. In addition to this xia2 includes a summary of the diffraction information from each wavelength – your standard “Table 1” stuff which Scala produces at the end of the log file:

For DEMO/INSULIN/SAD			
High resolution limit	1.78	5.64	1.78
Low resolution limit	24.67	24.67	1.88
Completeness	97.8	98.9	84.9
Multiplicity	20.5	19.9	15.8
I/sigma	48.0	80.4	15.6
Rmerge	0.048	0.027	0.182
Rmeas(I)	0.052	0.032	0.195
Rmeas(I+/-)	0.051	0.028	0.194
Rpim(I)	0.011	0.007	0.047
Rpim(I+/-)	0.015	0.008	0.065
Wilson B factor	19.49		
Partial bias	-0.008	-0.008	-0.007
Anomalous completeness	97.3	81.9	
Anomalous multiplicity	10.6	8.2	
Anomalous correlation	0.369	0.037	
Anomalous slope	1.473		
Total observations	153956	5223	14854
Total unique	7514	263	940

There is also a very concise summary of the integration for each run:

Integration status per image (60/record):

```
oooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooo
oooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooo
oooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooo
"o"=> ok           "%" => iffy rmsd "!" => bad rmsd
"O"=> overloaded   "#" => many bad  "." => blank
```

This is nice for checking that the integration has gone well – for good data all you should see is "o" with the odd "%". If you have lots of "!" then there is probably something wrong.

For each step of processing the "final" log file is recorded in the LogFiles directory. The final reflection files can be found in the DataFiles directory, and all of the harvesting information is placed in the harvest directory.

Availability

xia2 is available from the ccp4 server at <http://www.ccp4.ac.uk/xia> - and is free to download and use, though you are responsible for supplying and correctly licensing CCP4 and optionally XDS and Labelit. To make installation more straightforward I include an "extras" package which includes the correct version of pointless and ipmosflm binaries – this is almost mandatory to install, unless you are in the habit of installing things from the CCP4 prerelease pages. These are distributed separately as xia2 is licensed as BSD, while the extras are licensed as per CCP4 license. New versions of xia2 are announced on the xia2bb – if you are interested in keeping up-to-date I would recommend you subscribe. The mailing list is also the appropriate forum for general xia2 chit-chat, although there is very little traffic.

Platforms

The following platforms are supported:

- 32 and 64 bit x86 linux
- Windows 2000/XP
- OS X PPC and Intel

All share the same packages – the setup scripts deal with the details.

Installation

Installation is relatively straightforward – unpack the tarballs and set XIA2_HOME correctly in the setup files. That's all! You will need to add

BASH	. XIA2_HOME/setup.sh
(T)CSH	source XIA2_HOME/setup.csh
Windows	call XIA2_HOME/setup.bat

to your environment, but this is usually pretty easy. For Windows Francois Remacle has made a binary installer, which is probably the way to go ;o)

Acknowledgements and References

Development of xia2 is supported by e-HTPX, BioXHit and CCP4, and would not have happened without input from Harry Powell, Andrew Leslie, Nick Sauter, Phil Evans, Eleanor Dodson, Wolfgang Kabsch and many patient users. Also the following publications were critical in the development:

- Bailey, S. (1994) Acta Crystallogr. D 50, 760--763
- Evans, P.R. (1997) Proceedings of CCP4 Study Weekend
- Evans, Philip (2006) Acta Crystallographica Section D 62, 72--82
- Kabsch, W. (1988) Journal of Applied Crystallography 21, 67--72
- Kabsch, W. (1988) Journal of Applied Crystallography 21, 916--924
- Kabsch, W. (1993) Journal of Applied Crystallography 26, 795--800
- Leslie, AGW (1992) Joint CCP4 and ESF/EACMB Newsletter on Protein Crystallography 26
- Leslie, Andrew G. W. (2006) Acta Crystallographica Section D 62, 48--57
- Sauter, Nicholas K. and Grosse-Kunstleve, Ralf W. and Adams, Paul D. (2004) Journal of Applied Crystallography 37, 399--409
- Zhang, Z. and Sauter, N.K. and van den Bedem, H. and Snell, G. and Deacon, A.M. (2006) J. Appl. Cryst 39, 112--119