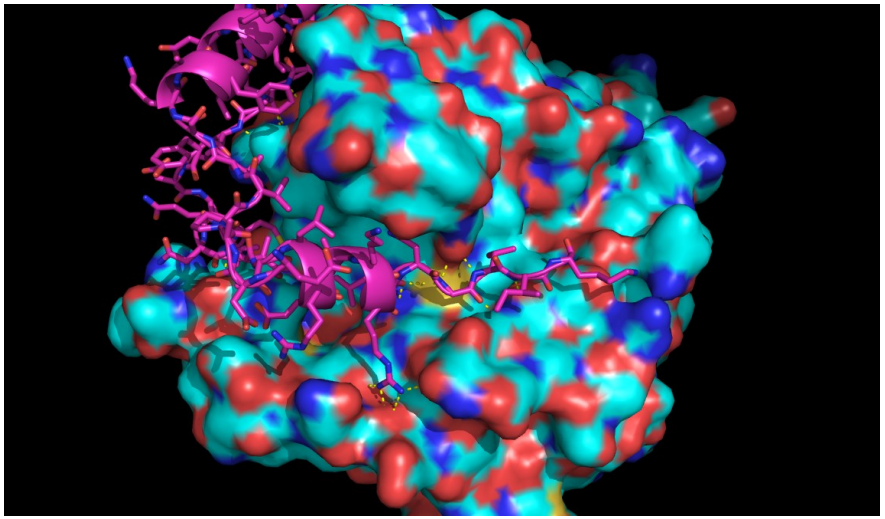
Introduction to CrystFEL

Data processing for serial crystallography



Thomas White

Serial Crystallography Data Analysis Workshop
ACA Annual Meeting
25 July 2015

HELMHOLTZ ASSOCIATION

CFEL SCIENCE

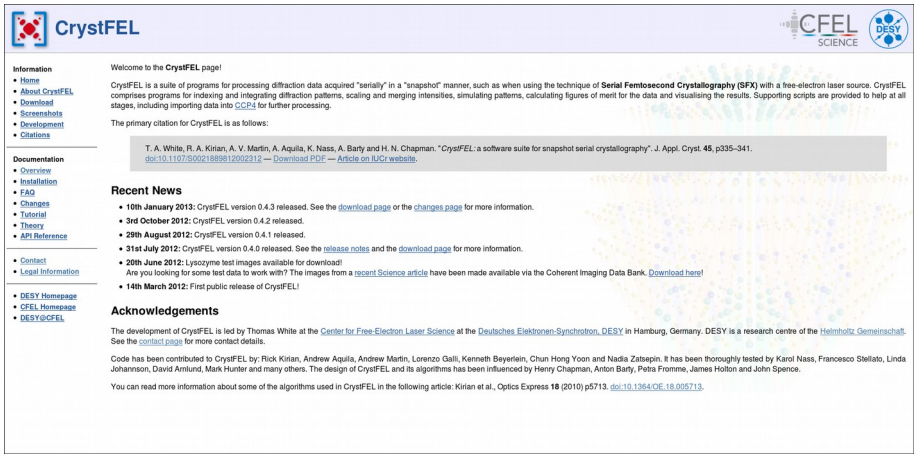
DESY

Contributors to CrystFEL

- Led by me, Thomas White (Center for Free-Electron Laser Science, DESY Hamburg).
- Code contributions from:
  - Valerio Mariani, Alexandra Tolstikova, Lorenzo Galli, Kenneth Beyerlein, Cornelius Gati, Anton Barty, Fedor Chervinskii, Oleksandr Yefanov (CFEL).
  - Takanori Nakane (MRC-LMB, now Tokyo University).
  - Keitaro Yamashita (SPring-8)
  - Wolfgang Brehm (U Konstanz)
  - Rick Kirian, Nadia Zatsepin (ASU).
  - Karol Nass, Andrew Aquila, Andrew Martin, Chun Hong Yoon (previously CFEL).
- Testing/feedback from the above people **and many more**.
- Words of wisdom from many people, in particular:
  - Henry Chapman (CFEL), Petra Fromme (CFEL), James Holton (ALS/UC Berkeley), John Spence (ASU/LBL), Kay Diederichs (U Konstanz).

What is CrystFEL?

- CrystFEL: data processing for “snapshot serial” crystallography (usually using an FEL, but not necessarily).
- Development since late 2009. Version 0.6.1 to be released very soon.
- Licence: GPLv3 (*free* and *open source* software)
- User-oriented development.
- **http://www.desy.de/~twhite/crystfel**



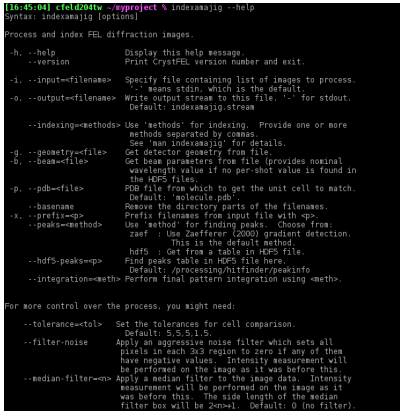
“Official” citation: White et al., J. Appl. Cryst. 45 (2012) p335

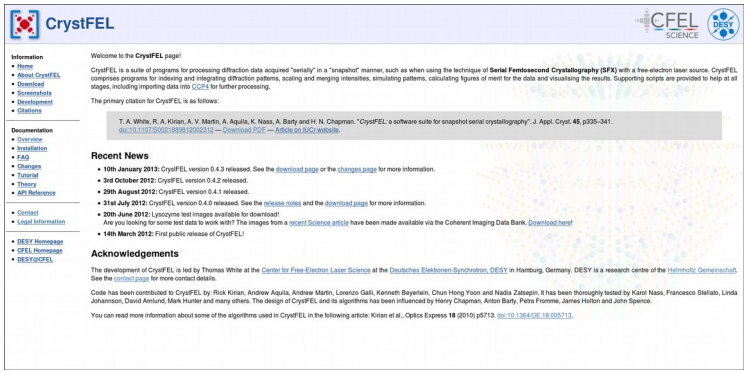
What do you get? Core programs

indexamajig	Bulk indexing and integration of patterns
ambigator	Resolve indexing ambiguities
process_hkl	Merge using “Monte Carlo” technique
hdfsee	View images in HDF5 format (with geometry)
cell_explorer	Examine unit cell parameter distributions
compare_hkl	Compare merged reflection data
check_hkl	Evaluate merged reflection data
pattern_sim	Simulate diffraction patterns
partial_sim	Simulate partial reflection intensities
get_hkl	Reflection data “Swiss army knife”
render_hkl	Render plane sections of reciprocal space
geoptimiser	Optimise multi-panel detector geometry
list_events	Generate event lists
whirligig	Detect rotation series
partialator	Merge using partialities and post-refinement

What do you get? Documentation

- **Manual pages:** `$ man indexamajig`
- **Help mesages:** `$ indexamajig --help`
- **Website, particularly: tutorial, best practice, FAQ**
- **Mailing list for announcements**
- **User support**





What do you get? Scripts

> Auxiliary scripts for plotting graphs, collating results etc.

> Difference from core programs: you're meant to copy them to your working directory and modify them

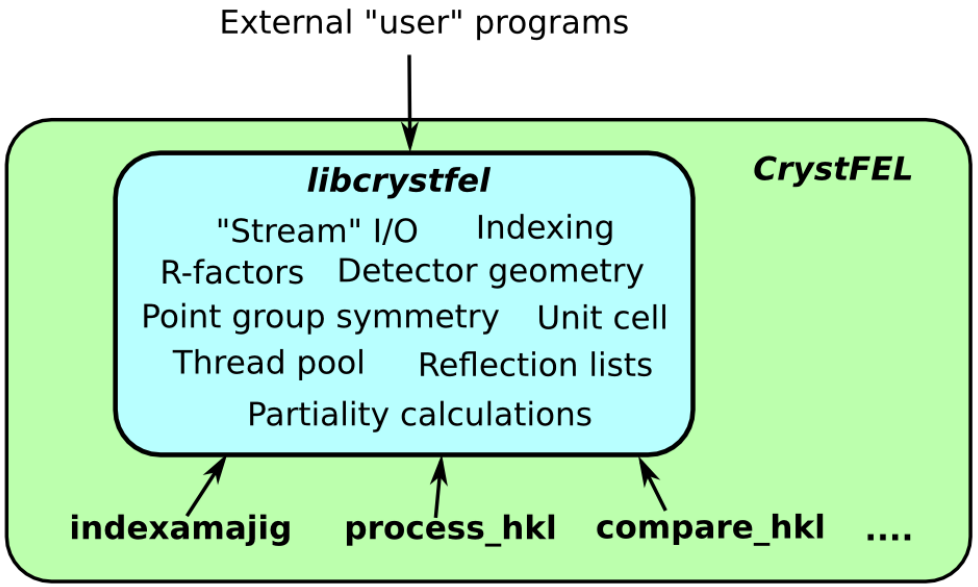
```
1 #!/usr/bin/perl -w
2
3 use strict;
4 use File::Basename;
5
6 my $args = join(" ", splice(@ARGV, 1, scalar(@ARGV)-1));
7 if ( !($args eq "") ) {
8         printf("Extra arguments for hdfsee: %s\n", $args);
9 } else {
10        # Default arguments - feel free to override!
11        $args = "--binning=2 --int-boost=10";
12        printf("Using default arguments for hdfsee: %s\n", $args);
13 }
14
15 open(FH, $ARGV[0]);
16 open(TMP, "> list.tmp");
17
18 my $in_image = 0;
```

What do you get? libcrystfel

> libcrystfel: shared library exposing high-level functions, e.g. “index this pattern” or “calculate these partialities”.

> Possible application: use CrystFEL for online data processing within a facility's monitoring software.

> API docs on website



Where can you get it?

- From the website: **http://www.desy.de/~twhite/crystfel**
  (or just type “crystfel” into your favourite search engine)

- “Release” versions: tested with a series of regression tests

- Git repository: very latest code, maybe broken, maybe awesome, slightly harder to install (need autotools installed).

- Packaged snapshots: like Git repository, but slightly easier to install.

- Ubuntu, Red Hat, Debian packages, MacPort (back soon).

Where can you get it?

- GNU General Public License (GPL):
  "*free*" as in "*freedom*" (also as in "*free beer*")

- Freedom to use the software *for any purpose* (even commercially).
- Freedom to study how the software works.
- Freedom to change the software, e.g. to fix bugs or add features.
- Freedom to share the software, even a modified version.

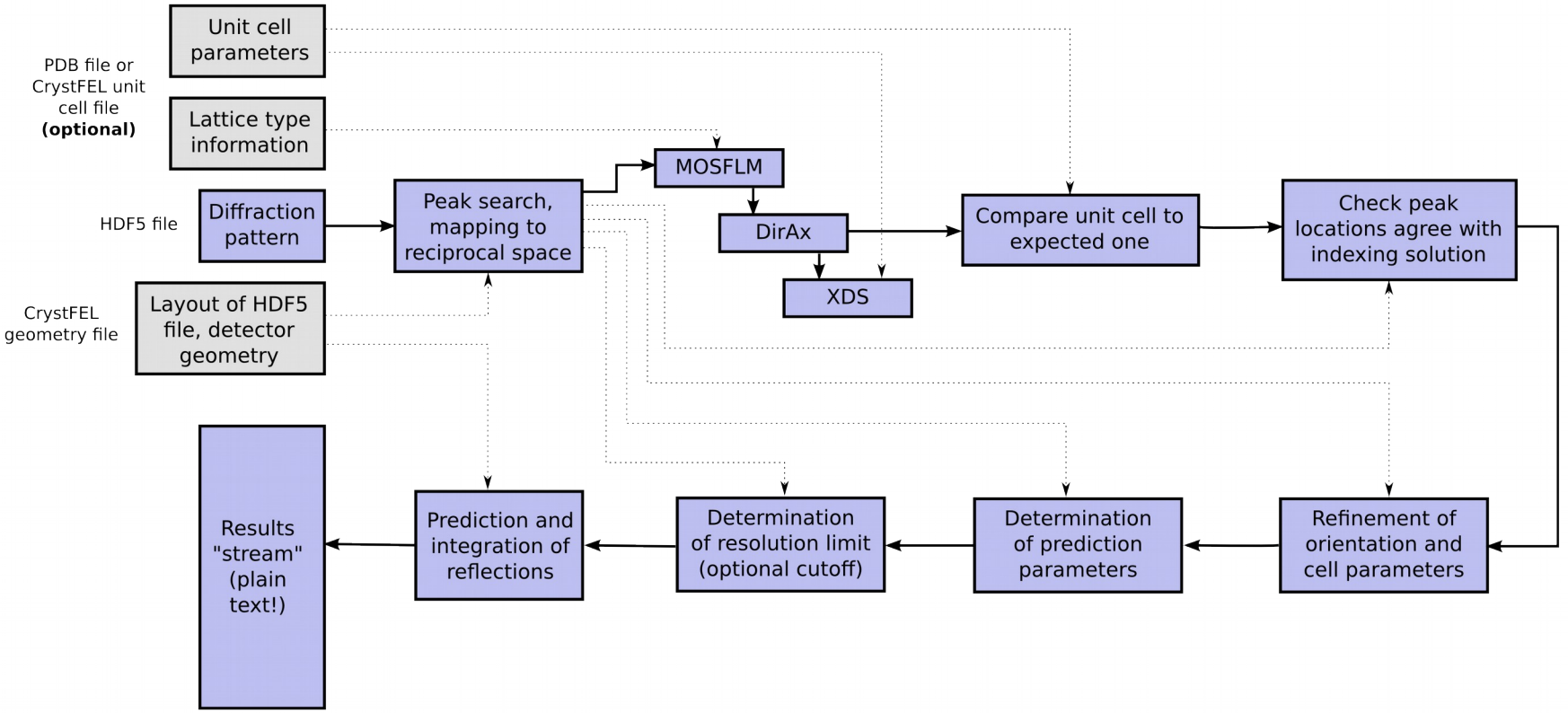
- No license forms to sign, send etc.

- The GPL ensures that these rights are never restricted.

Fundamental basis of CrystFEL

Every diffraction pattern is a new experiment and must be processed individually.

Pipeline in “indexamajig”



- Input: HDF5 (very common scientific format, many language bindings)
- Note: choice of methods for **peak search**, **indexing** and **integration**.
- Output: plain text “stream” - don't be afraid to look inside!
  (tip: use “less”, not “gedit” - file can be very large)

Peak search methods

> `indexamajig [....] --peaks=hdf5 [....]`
> **`....` means "get peak list from HDF5 file".**

> `indexamajig [....] --peaks=cxi [....]`
> **`....` means "get peak list from CXI file".**

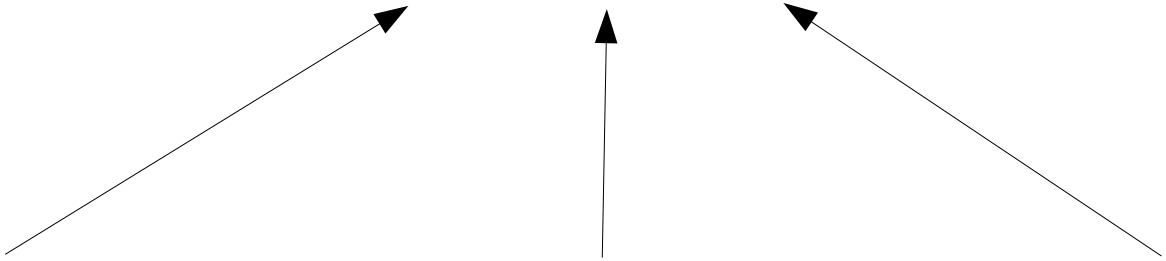
> `indexamajig [....] --peaks=zaef [....]`
> **`....` means "use internal peak search algorithm".**

> More peak search algorithms will probably be added in the future.

Indexing methods

> `indexamajig [...] --indexing=method1,method2,.. [....]`
> **List of indexing methods will be tried in order.**

> **Example method:** `mosflm-raw-latt`



Invoke mosflm to index the pattern using the peaks from the peak search

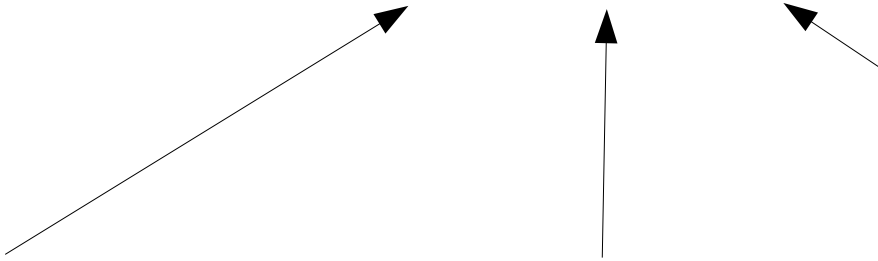
Skip the step where the unit cell parameters would be checked against the reference

Take the Bravais lattice information from the reference into account, e.g. “look for tetragonal P cells only”

Indexing methods

> `indexamajig [...] --indexing=method1,method2,.. [....]`
> **List of indexing methods will be tried in order.**

> **Example method:** `mosflm-axes-latt`



Invoke mosflm to index the pattern using the peaks from the peak search

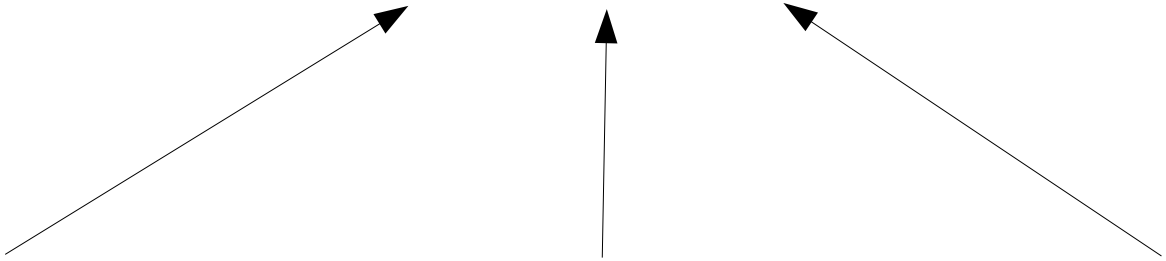
Check the unit cell parameters against the reference

Take the Bravais lattice information from the reference into account, e.g. “look for tetragonal P cells only”

Indexing methods

> `indexamajig [...] --indexing=method1,method2,.. [....]`
> **List of indexing methods will be tried in order.**

> **Example method:** `mosflm-raw-nolatt`



Invoke mosflm to index the pattern using the peaks from the peak search

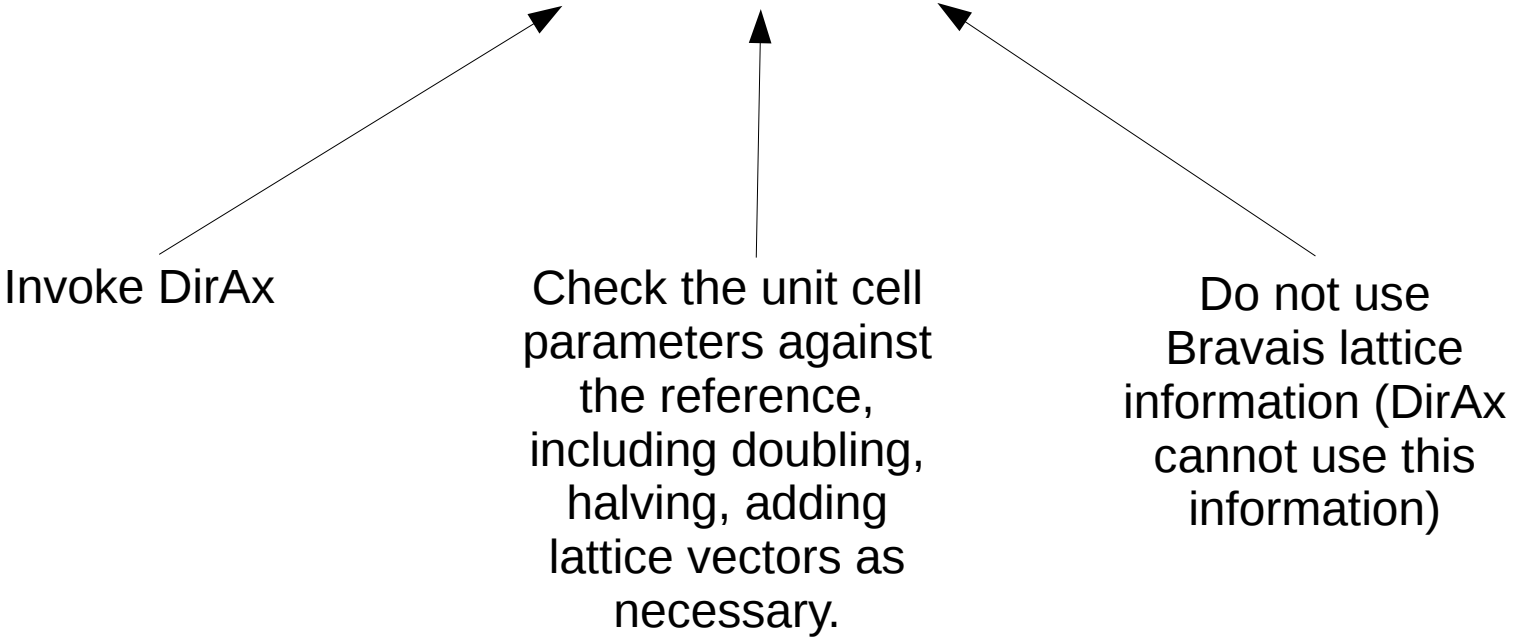
Do not check the unit cell parameters

Do not use Bravais lattice information

Indexing methods

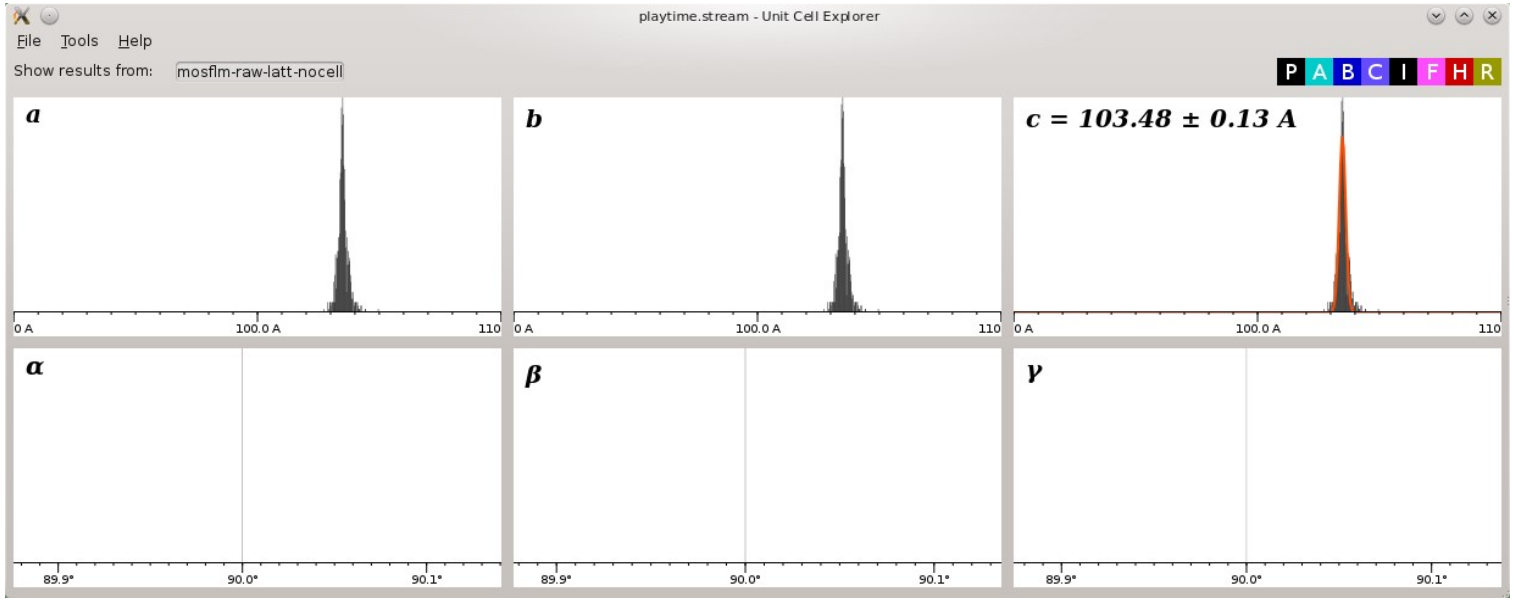
> `indexamajig [...] --indexing=method1,method2,.. [....]`
> **List of indexing methods will be tried in order.**

> **Example method:** `dirax-comb-nolatt`



The Unit Cell Explorer

```
> cell_explorer output.stream
```



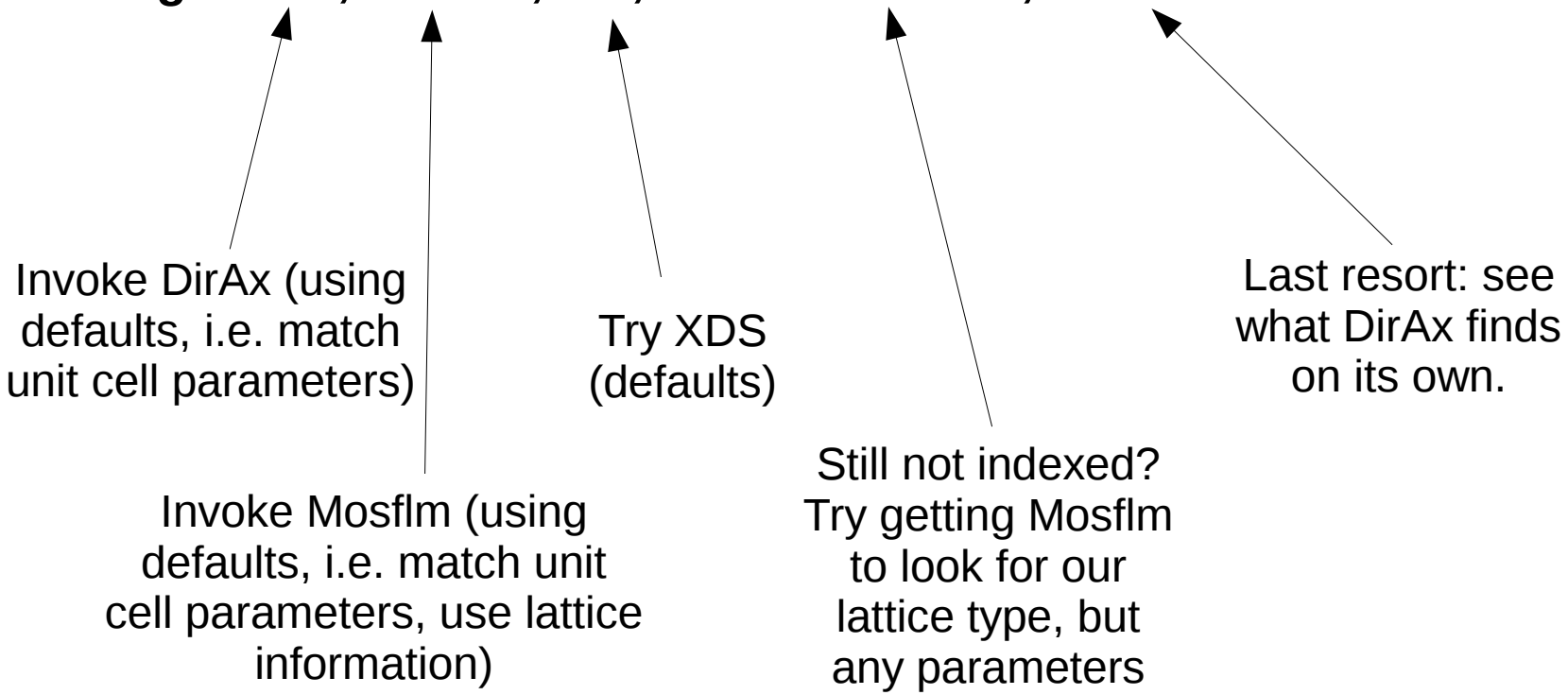
a = b = c = 103.5 Å
α = β = ɣ = 90.0°
Body centered (I).

Possible space groups: I23, $I2_13$, I432, $I4_132$, I422 $I4_122$, $I4_1$, I4, I222

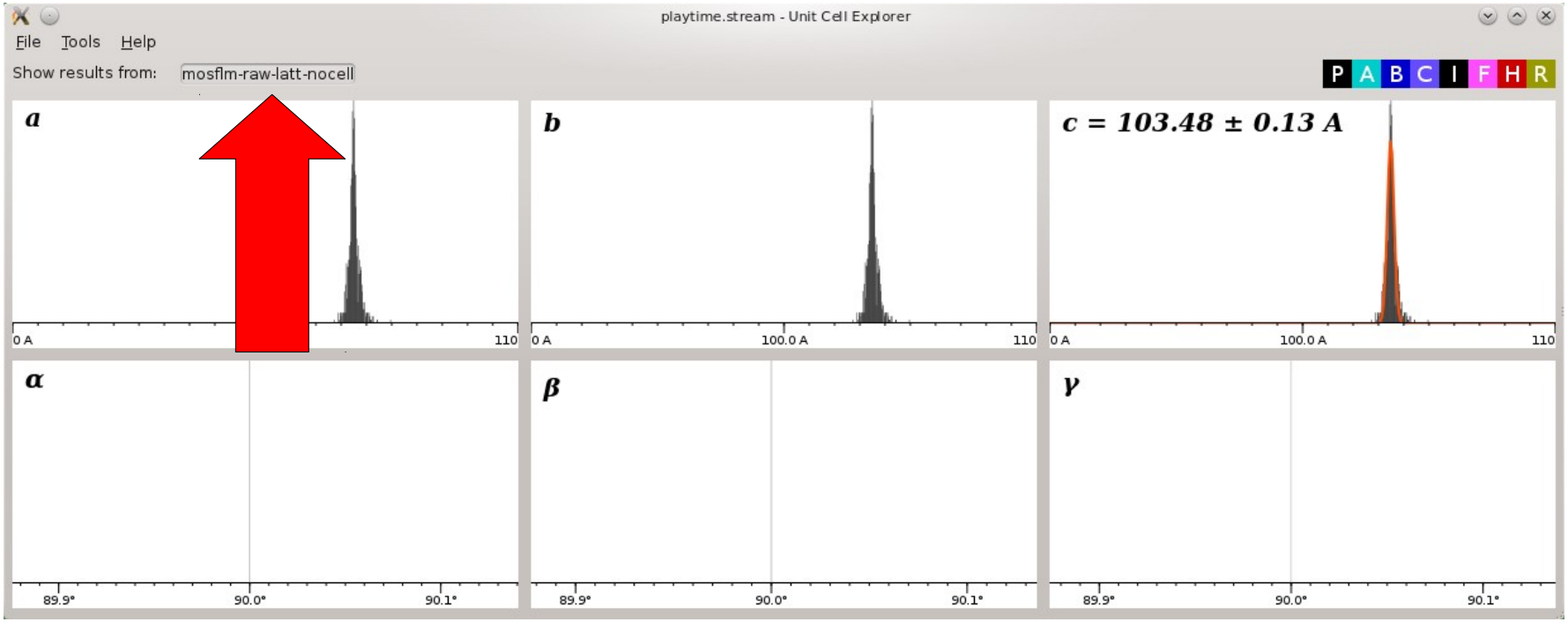
Indexing methods

> `indexamajig [...] --indexing=method1,method2,.. [....]`
> **List of indexing methods will be tried in order.**

> **--indexing=dirax,mosflm,xds,mosflm-raw-latt,dirax-raw**



The Unit Cell Explorer



Geometry files

```
> indexamajig [...] --geometry=cspad.geom [....]
```

Geometry file describes the layout of data in the file,
as well as the physical layout of the detector.

"man crystfel_geometry", templates distributed with CrystFEL, or ask a friendly PCS collaborator, beamline scientist or us.

Analysis in CrystFEL can be surprisingly robust to inaccurate geometry.

"Calibration mode" in hdfsee.

Geometry refinement tool (geoptimiser).

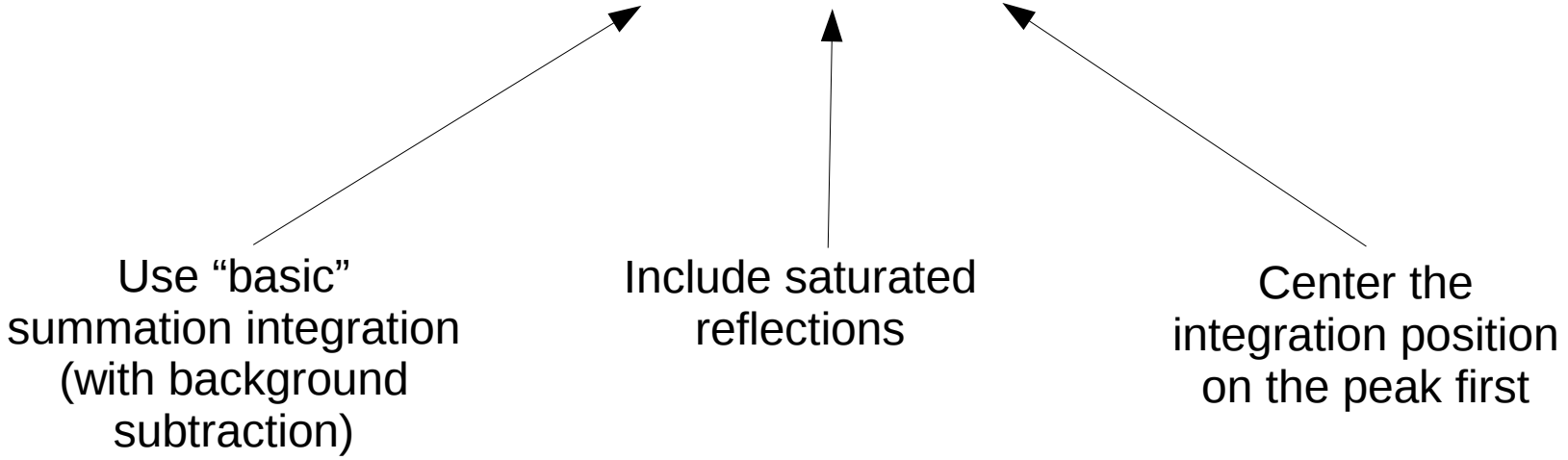
"Prediction refinement" stage updates central beam position.

("Beam files" are no longer needed as of CrystFEL 0.6.0)

Integration methods

> `indexamajig [...] --integration=method [....]`
(only one integration method allowed)

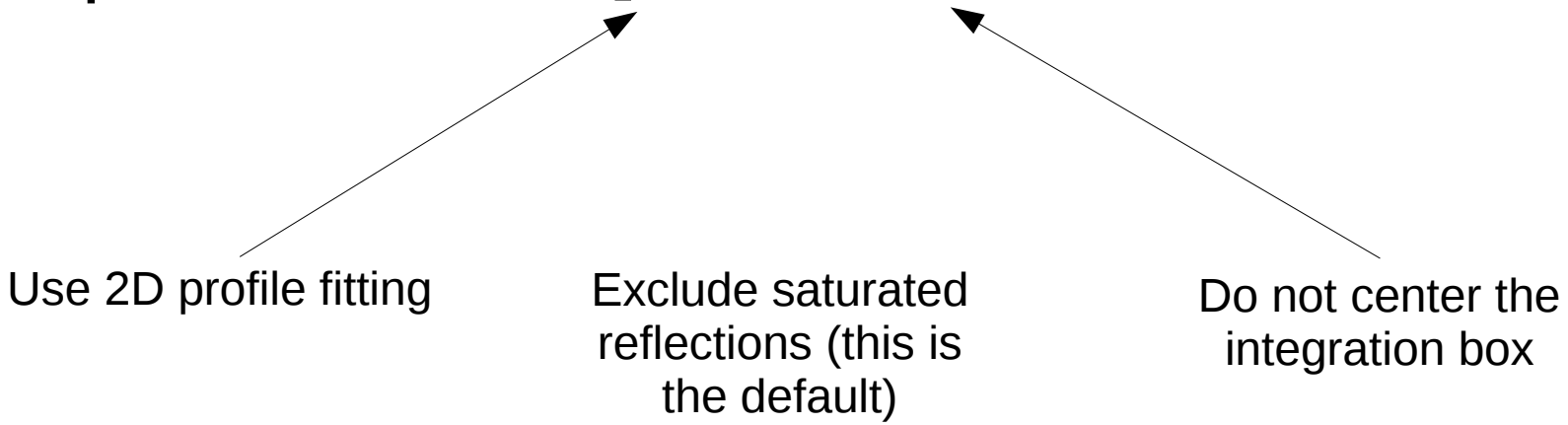
> **Example method:** `rings-sat-cen`



Integration methods

> `indexamajig [...] --integration=method [....]`
(only one integration method allowed)

> **Example method:** `prof2d-nocen`

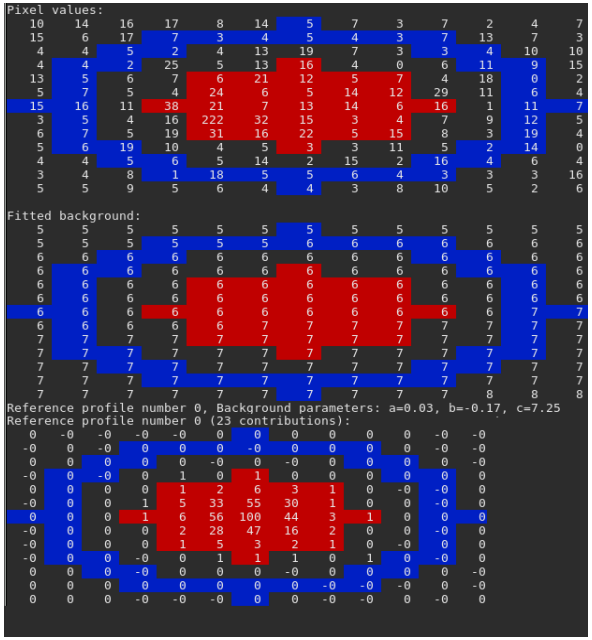


Integration diagnostics

> `indexamajig [...] --int-diag=<something>`

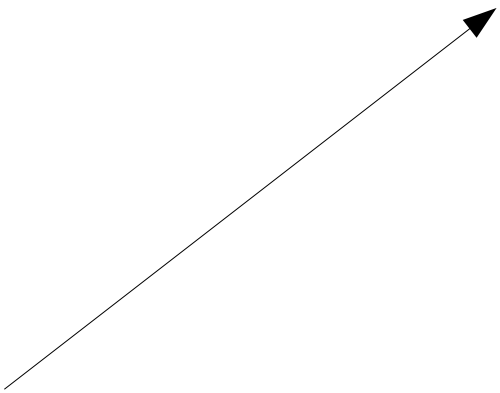
> <something> can be:

- `none` (this is the default)
- A set of indices, e.g. `4,3,5`
- `random` – show a random 1% of reflections
- `implausible` – show reflections with I < -5sigma
- `negative` – show reflections with I < -3sigma



Merging

```
> process_hkl -i input.stream -o output.hkl -y 6/m
```



Output: text file, can be imported into most structure solution programs. Alternatively: “create-mtz” and “create-xscale” scripts.

Actual symmetry (note: centre of symmetry added, i.e. merge Friedel pairs)

Confused about point group notation? “man crystfel” for a list of all possibilities.

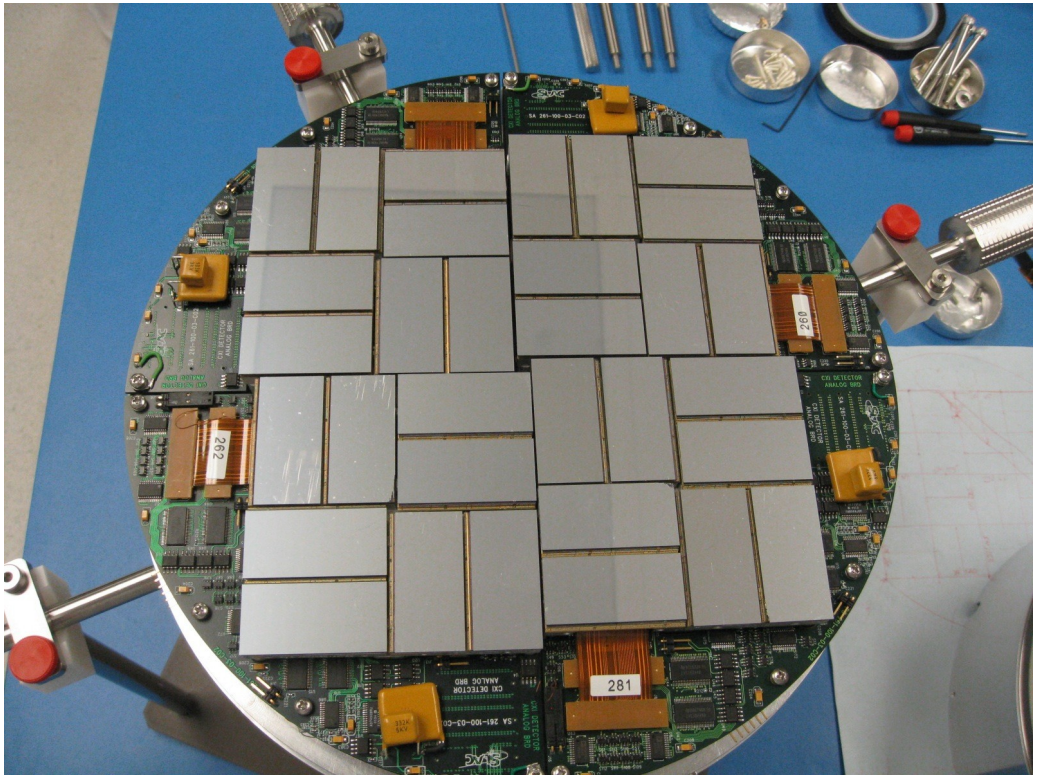
Merging (advanced/experimental/dangerous/awesome)

```
> partialator -i input.stream -o output.hkl -y 6/m
              --model=unity -iterations=1

> partialator -i input.stream -o output.hkl -y 6/m
              --model=unity -iterations=3

> partialator -i input.stream -o output.hkl -y 6/m
              --model=scsphere -iterations=1

> partialator -i input.stream -o output.hkl -y 6/m
              --model=unity -iterations=0
```

CrystFEL 0.6.0 - detector geometry refinement

```
> geoptimiser -i input.stream -g input.geom -o output.geom
```

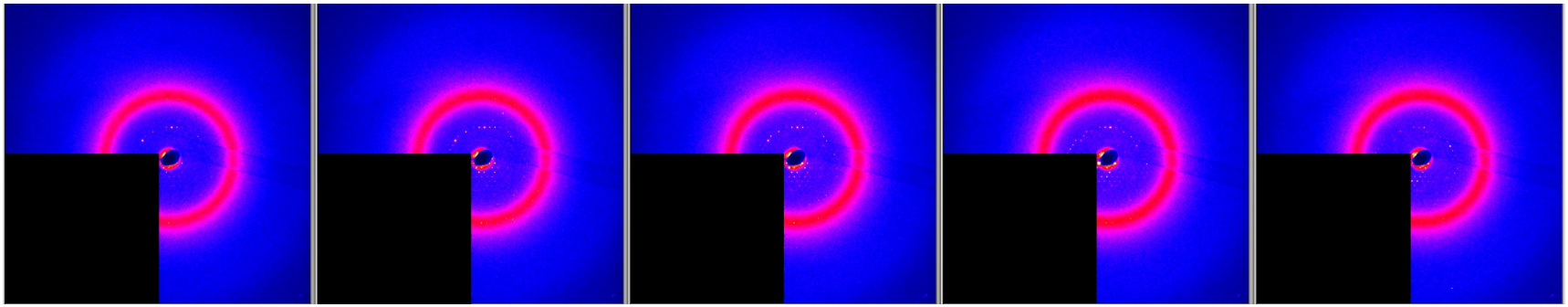
- Refines 3D position each part of the detector individually, or at a level of granularity you can specify (e.g. all tiles individually, quadrants as one, whole detector).



Finding “mini rotation series”

```
> whirligig input.stream
```

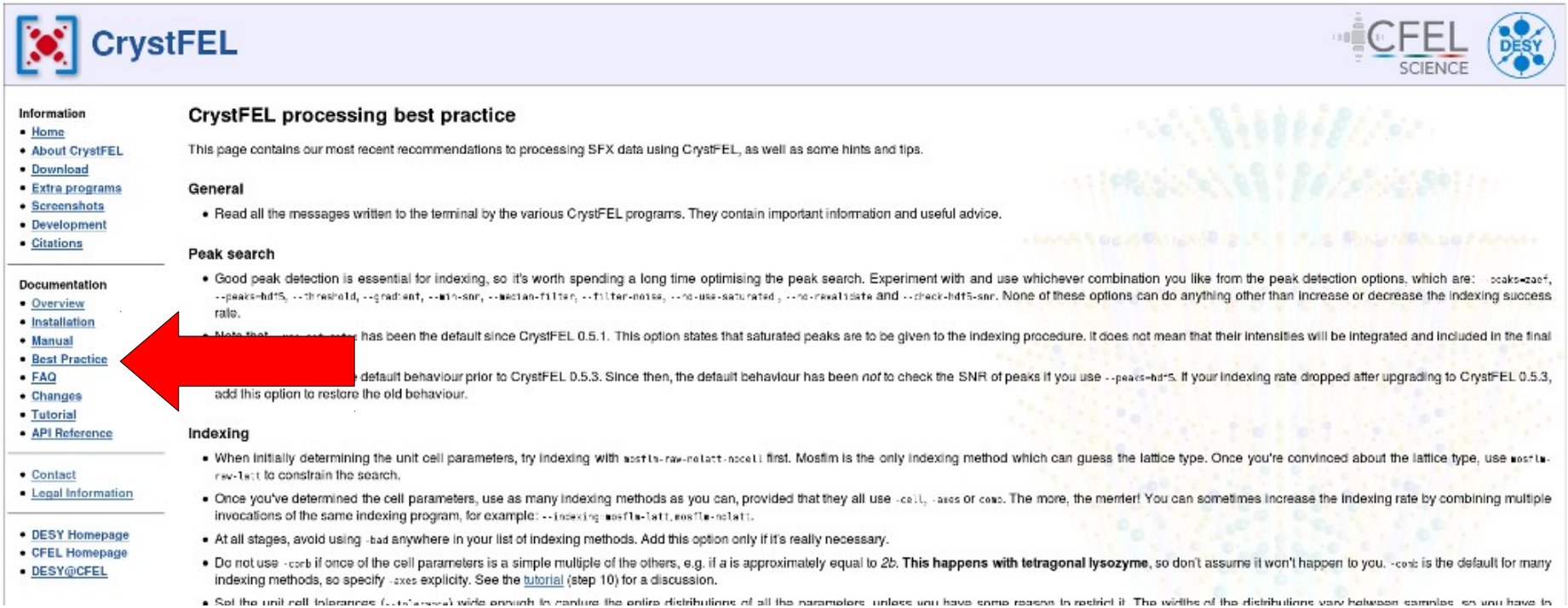
- Compares orientations of consecutive crystals, taking into account symmetry and indexing ambiguities. Outputs lists of events forming series.



P. Nogly, D. James, D. Wang, T. A. White et al., IUCrJ 2 (2015):

> Occasionally, two or three consecutive hits were recorded on some of the larger (40–50 µm) crystals. Bragg spots appeared and disappeared within this sequence of consecutive images, indicating that the rotational diffusion of crystals in the LCP within the 80 ms between two exposures is larger than their mosaic spread. To investigate this further, a computer program was written to compare the orientations of crystals in adjacent frames using the data stream output from *CrystFEL*. For the fraction of data acquired with a 25 ms exposure time (81% of the total frames), 1088 frames (26% of the successfully indexed patterns) were found to be part of a rotation series. The mean series length was 2.2 frames and the maximum series length was 4 frames.

“Best practice” guidelines



CrystFEL

CFEL SCIENCE DESY

Information
- Home
- About CrystFEL
- Download
- Extra programs
- Screenshots
- Development
- Citations

Documentation
- Overview
- Installation
- Manual
- Best Practice
- FAQ
- Changes
- Tutorial
- API Reference

- Contact
- Legal Information

- DESY Homepage
- CFEL Homepage
- DESY@CFEL

CrystFEL processing best practice

This page contains our most recent recommendations to processing SFX data using CrystFEL, as well as some hints and tips.

General

- Read all the messages written to the terminal by the various CrystFEL programs. They contain important information and useful advice.

Peak search

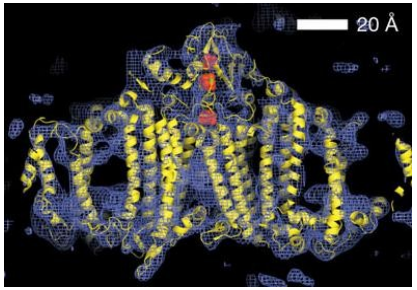
- Good peak detection is essential for indexing, so it's worth spending a long time optimising the peak search. Experiment with and use whichever combination you like from the peak detection options, which are: `--peaks=zaef`, `--peaks=hdf5`, `--threshold`, `--gradient`, `--min-snr`, `--median-filter`, `--filter-noise`, `--no-use-saturated`, `--no-revalidate` and `--check-hdf5-snr`. None of these options can do anything other than increase or decrease the indexing success rate.
- Note that `--no-sat-corr` has been the default since CrystFEL 0.5.1. This option states that saturated peaks are to be given to the indexing procedure. It does not mean that their intensities will be integrated and included in the final
- default behaviour prior to CrystFEL 0.5.3. Since then, the default behaviour has been *not* to check the SNR of peaks if you use `--peaks=hdf5`. If your indexing rate dropped after upgrading to CrystFEL 0.5.3, add this option to restore the old behaviour.

Indexing

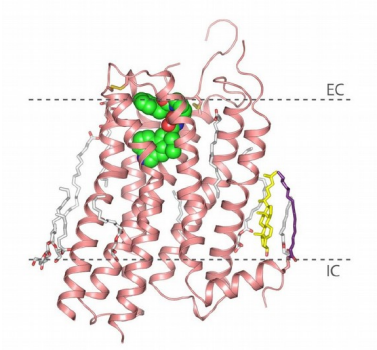
- When initially determining the unit cell parameters, try indexing with `mosflm-raw-nolatt-nocell` first. Mosflm is the only indexing method which can guess the lattice type. Once you're convinced about the lattice type, use `mosflm-raw-latt` to constrain the search.
- Once you've determined the cell parameters, use as many indexing methods as you can, provided that they all use `-cell`, `-axes` or `cons`. The more, the merrier! You can sometimes increase the indexing rate by combining multiple invocations of the same indexing program, for example: `--indexing mosflm-latt,mosflm-nolatt`.
- At all stages, avoid using `-bad` anywhere in your list of indexing methods. Add this option only if it's really necessary.
- Do not use `-corb` if once of the cell parameters is a simple multiple of the others, e.g. if *a* is approximately equal to *2b*. **This happens with tetragonal lysozyme**, so don't assume it won't happen to you. `-cons` is the default for many indexing methods, so specify `-axes` explicitly. See the tutorial (step 10) for a discussion.
- Set the unit cell tolerances (`--tolerance`) wide enough to capture the entire distributions of all the parameters, unless you have some reason to restrict it. The widths of the distributions vary between samples, so you have to

“Definitely use this option”
“Don't use this option except for testing”
“Try this option and see what happens, might be good or bad”
“Watch out for this 'gotcha'”
... etc ...

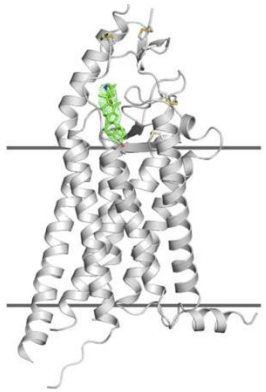
CrystFEL in the PDB (highlights)



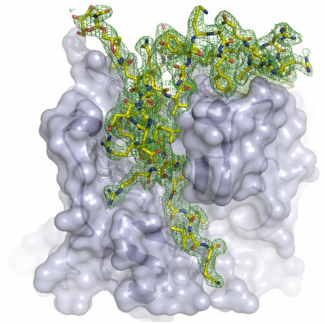
3PCQ
Photosystem I (the first SFX experiment)
Chapman et al., Nature 2011



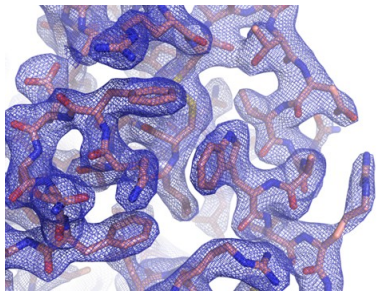
4NC3
Serotonin receptor bound to ergotamine
Lui et al., Science 2013



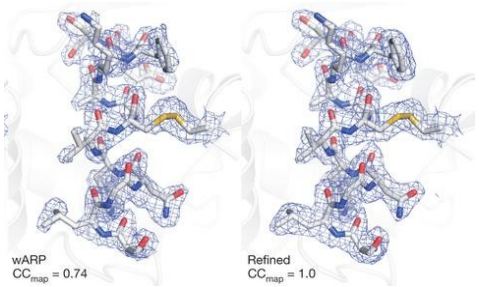
4O9R
Smoothened receptor using LCP injector
Weierstall et al., Nature Communications 2014



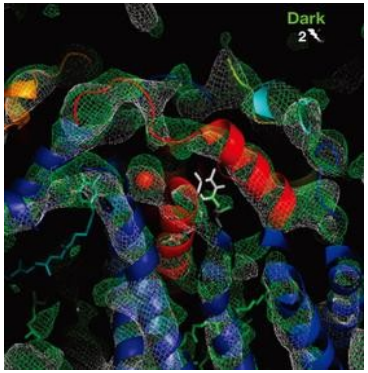
4HWY
Natively inhibited Cathepsin B
Redecke, Nass et al., Science 2013



4O34
Serial crystallography using a synchrotron beamline
Stellato et al., IUCrJ 2014



4N5R
Lysozyme (Gd derivative) ab initio phasing using SAD
Barends et al., Nature 2013



4Q54
Photosystem II in putative S_3 excited state
Kupitz, Basu et al., Nature 2014

CrystFEL in the PDB / selected citations

2015

Y. Kang, X. E. Zhou, X. Gao, Y. He et al. "Crystal structure of rhodopsin bound to arrestin by femtosecond X-ray laser". Nature (2015). **4ZWJ**.
C. E. Conrad, S. Basu, D. Janes, D. Wang et al. "A novel inert crystal delivery medium for serial femtosecond crystallography". IUCrJ 2 (2015) p421. **4Z8K**.
M. Bublitz, K. Nass, N. D. Drachmann, A. J. Markvardsen et al. "Structural studies of P-type ATPase-ligand complexes using an X-ray free-electron laser". IUCrJ 2 (2015) p409. **4XOU**.
N. Coquelle, A. S. Brewser, U. Kapp, A. Shilova et al. "Raster-scanning serial protein crystallography using micro- and nano-focused synchrotron beams". Acta Cryst. D71 (2015) p1184. **4WG1**, **4WG7**, **4WL6** and **4WL7**.
H. Zhang, H. Unal, C. Gati, G. W. Han et al. "Structure of the angiotensin receptor revealed by serial femtosecond crystallography". Cell (2015) p833. **4YAY**.
S. Boutet, L. Foucar, T. R. M. Barends, S. Botha et al. "Characterization and use of the spent beam for serial operation of LCLS". Journal of Synchrotron Radiation 22 (2015) p634. **4RW1** and **4RW2**.
K. Tono, E. Nango, M. Sugahara, C. Song et al. "DAPHNIS: application to serial protein crystallography using an X-ray free-electron laser". Journal of Synchrotron Radiation 22 (2015) p532. **3WUN**.
G. Fenalti, N. A. Zatsepin, C. Betti, P. Giguere et al. "Structural basis for bifunctional peptide recognition at human δ-opioid receptor". Nature Structural & Molecular Biology (2015) p265. **4RWD**.
P. Nogly, D. James, D. Wang, T. A. White et al. "Lipidic cubic phase serial millisecond crystallography using synchrotron radiation". IUCrJ 2 (2015). **4X31**..

2014

J. Tenboer, S. Basu, N. Zatsepin, K. Pande et al. "Time-resolved serial crystallography captures high-resolution intermediates of photoactive yellow protein". Science 346 (2014) p1242. **4WL9** and **4WLA**.
M. Sugahara, E. Mizohata, E. Nango, M. Suzuki et al. "Grease matrix as a versatile carrier of proteins for serial crystallography". Nature Methods (2014). **3WUL**, **3WXT**, **3WXU**, **3WUM**, **4W4Q**, **3WXS** and **3WXQ**.
M. R. Sawaya, D. Cascio, M. Gingery, J. Rodriguez et al. "Protein crystal structure obtained at 2.9 Å resolution from injecting bacterial cells into an X-ray free-electron laser beam". PNAS 111 (2014) p12769.
M. S. Hunter, B. Segelke, M. Messerschmidt, G. J. Williams et al. "Fixed-target protein serial microcrystallography with an x-ray free electron laser". Scientific Reports 4 (2014) 6026.
C. Kupitz, S. Basu, I. Grotjohann, R. Fromme et al. "Serial time-resolved crystallography of photosystem II using a femtosecond X-ray laser". Nature (2014). **4PBU** and **4Q54**.
M. Caffrey, D. Li, N. Howe and S. T. A. Shah. "'Hit and run' serial femtosecond crystallography of a membrane kinase in the lipid cubic phase". Phil Trans. Roy. Soc. B 369 (2014) 20130621.
O. Yefanov, C. Gati, G. Bourenkov, R. A. Kirian et al. "Mapping the continuous reciprocal space intensity distribution of X-ray serial crystallography". Phil. Trans. Roy. Soc. B 369 (2014) 20130333.
F. Stellato, D. Oberthür, M. Liang, R. Bean et al. "Room-temperature macromolecular serial crystallography using synchrotron radiation". IUCrJ 1 (2014). **4O34**.
K. Qu, L. Zhou and Y.-H. Dong. "An improved integration method in serial femtosecond crystallography". Acta Cryst. D70 (2014) p1202.
M. Frank, D. B. Carlson, M. S. Hunter, G. J. Williams et al. "Femtosecond X-ray diffraction from two-dimensional protein crystals". IUCrJ 1 (2014) p95.
C. Gati, G. Bourenkov, M. Klinge, D. Rehders et al. "Serial crystallography on in vivo grown microcrystals using synchrotron radiation". IUCrJ 1 (2014) p87. **4N4Z.**
U. Weierstall, D. James, C. Wang, T. A. White et al. "Lipidic cubic phase injector facilitates membrane protein serial femtosecond crystallography". Nature Communications 5:3309 (2014). **4O9R.**
C. Song, K. Tono, J. Park, T. Ebisu et al. "Multiple application X-ray imaging chamber for single-shot diffraction experiments with femtosecond X-ray laser pulses". J. Appl. Cryst. 47 (2014) p18.

2013

W. Liu, D. Wacker, C. Gati, G. W. Han et al. "Serial Femtosecond Crystallography of G Protein-Coupled Receptors". Science 342 (2013) p1522. **4NC3** and **CXIDB ID 21.**
L. C. Johansson, D. Arnlund, G. Katona, T. A. White et al. "Structure of a photosynthetic reaction centre determined by serial femtosecond crystallography". Nature Communications 4:2911 (2013). **4CAS.**
T. R. M. Barends, L. Foucar, S. Botha, R. B. Doak et al. "De novo protein crystal structure determination from X-ray free-electron laser data". Nature (2013). **4N5R** and **CXIDB ID 20.**
H. Dermirci,et al. "Serial femtosecond X-ray diffraction of 30S ribosomal subunit microcrystals in liquid suspension at ambient temperature using an X-ray free-electron laser". Acta Crystallographica F69 (2013) p1066.
T. R. M. Barends, L. Foucar, R. L. Shoeman, S. Bari et al. "Anomalous signal from S atoms in protein crystallographic data from an X-ray free-electron laser". Acta Crystallographica D69 (2013) p838.
L. Redecke, K. Nass et al. "Natively Inhibited Trypanosoma brucei Cathepsin B Structure Determined by Using an X-ray Laser". Science 339 (2013) p227. **4HWY.**

2012

S. Boutet, L. Lomb, G. J. Williams, T. R. M. Barends et al. "High-Resolution Protein Structure Determination by Serial Femtosecond Crystallography". Science 337 (2012) p362. **4ET8** and **4ET9**.
L. C. Johansson, D. Arnlund, T. A. White, G. Katona et al. "Lipidic phase membrane protein serial femtosecond crystallography". Nature Methods 9 (2012) p263. **4AC5**.

2011

H. N. Chapman, P. Fromme et al. "Femtosecond X-ray protein nanocrystallography". Nature 470 (2011) p73. **3PCQ.**

(full list on the website)