From raw SFX data to electron density using CASS, CrystFEL and PHENIX

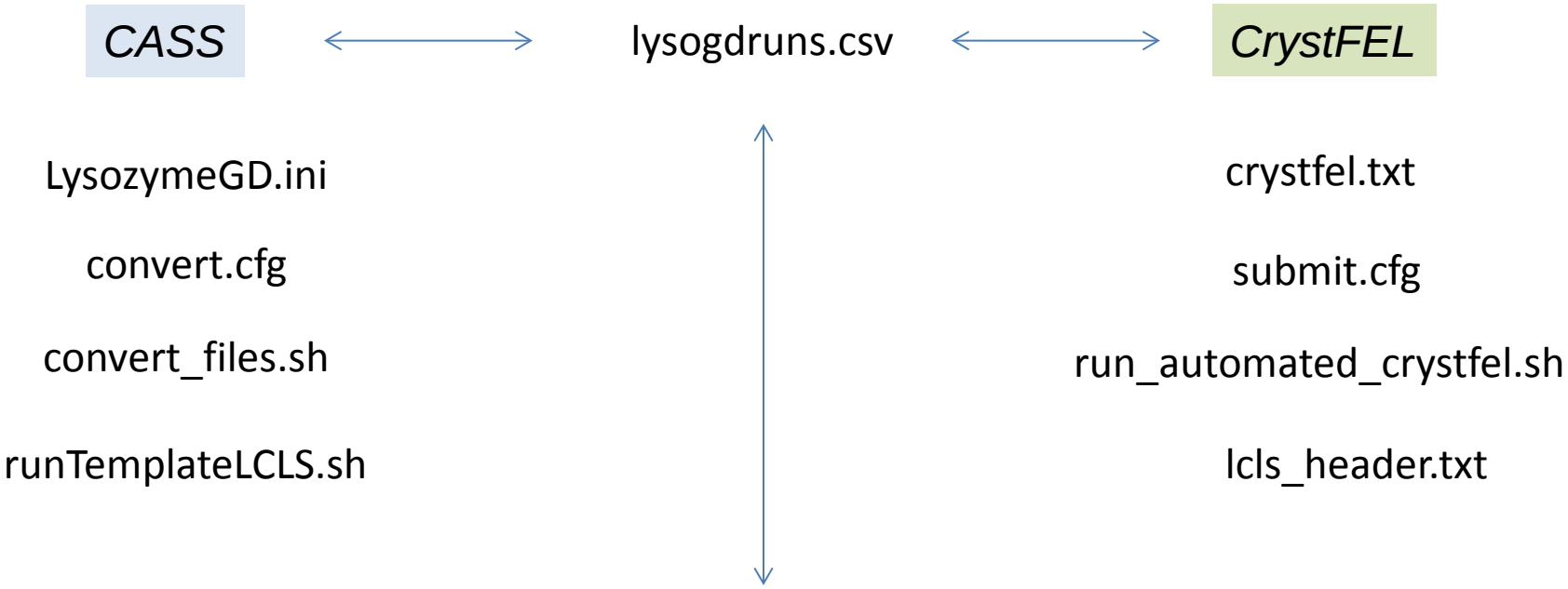
Karol Nass

MPI Medical Research Heidelberg

Objectives:

1. Convert raw data to HDF5 format using *CASS*
2. Use *CrystFEL* to index and merge
3. Solve the lysozyme-Gd structure using *PHENIX.autosol*

Scripts:


CASS
lysogdruns.csv
CrystFEL
LysozymeGD.ini
convert.cfg
convert_files.sh
runTemplateLCLS.sh
crystfel.txt
submit.cfg
run_automated_crystfel.sh
lcls_header.txt
001,dark
002,001,LysozymeGD.ini,,crystfel.txt
003,001,LysozymeGD.ini,,crystfel.txt
…
…

SFX Lysozyme-Gd data is avalible on-line at cxidb.org

CXIDB
Coherent X-ray Imaging Data Bank

Home | Mission | CXI File Format | Browse Data | Resources | Sponsors | Contact Us

CXIDB ID 22

Citation Details	
Title:	De novo protein crystal structure determination from X-ray free-electron laser data
Authors:	Thomas R. M. Barends et al.
Journal:	Nature
Year:	2013
DOI:	doi:10.1038/nature12773
Experimental Conditions	
Method:	Serial Femtosecond Crystallography
Sample:	Lysozyme Gadolinium
Wavelength:	1.46 Å
Lightsource:	LCLS
Beamline:	CXI
Deposition Summary	
Depositor:	Thomas R. M. Barends
Contact:	Thom...@mpimf-heidelberg.mpg.de
Deposition date:	2013-11-25
Last modified:	2013-11-25
Data Files	
Auxiliary Files	
Description:	Raw XTC Files
Alternative Formats	

Description

Serial femtosecond crystallography (SFX) data of microcrystals of a lysozyme gadolinium derivative. The data was used to demonstrate de-novo phasing by single anomalous dispersion.

Citing CXIDB

If you make use of CXIDB for your publication, please cite:

Maia, F. R. N. C. The Coherent X-ray Imaging Data Bank. *Nat. Methods* **9**, 854–855 (2012).

The continued growth of CXIDB depends on the support of institutions and research grants which ultimately use citations as a measure of quality.

Summary:
In this tutorial we will process SFX data from lysozyme Gadolinium derivative.
We will use CASS software suite on the complete experimental data set in XTC format
to convert it into many hdf5 files. Every hdf5 file will contain a diffraction
image with partial reflections and associated machine data eg. photon energy.

Such diffraction images can be used for analysis with CrystFEL, which will
index them, extract the partial intensities and merge them using Monte Carlo method.

Merged data set of lysozyme Gd structure factors can be used for succesisful ab initio
SAD phasing (Barends et al.).

Step by step tutorial:

1) Muchas gracias to Lutz Foucar for providing CASS and scripts that make the life easier.

2) Convert XTC files with Lysozyme-Gd diffraction data to single hdf5 files.
    a) use provided scripts, copy all of them to your 'scripts' directory,
    b) to start conversion edit files:
        .) 'convert.cfg' and provide the correct paths pointing to your directory,
        .) comment all lines in 'lysogdruns.csv' except 2-nd and 3-rd, we will start with
        converting only one run,
    c) type 'sh conver_files.sh lysogdruns.csv convert.cfg',
    d) watch the output while your jobs are beeing subbmitted to LCLS queue, monitor queue
        using 'bjobs -u "your user name",
    e) after few minutes the jobs should finish and lots of hdf5 files should appear in your
        'hdf5' directory.

3) Use 'run_automated_crystfel.sh' script to start CrystFEL indexing jobs.
    a) edit 'submit.cfg' file to enter paths that will point to your indexing directory,
    b) edit 'crystfel.txt' file to enter your favourite CrystFEL parameters,
    c) start 'run_automated_crystfel.sh lysogdruns.csv' script to submit indexing jobs,
    d) run 'merge-all.sh' script to merge indexing result from individual runs to one and
        calculate CrystFEL statistics.

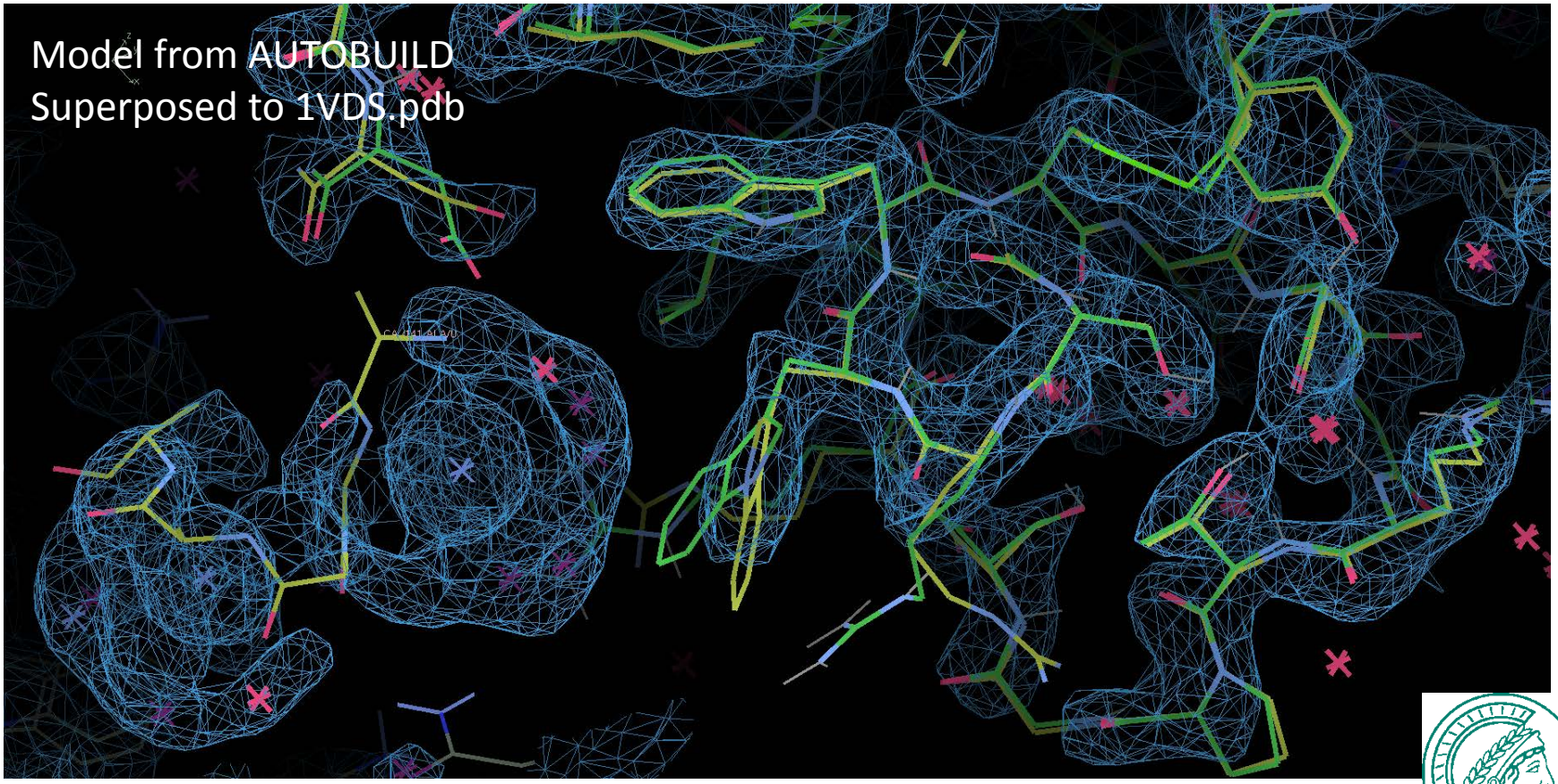
4) Run create-xscale-kn script to convert .hkl file to XDS format.
5) Use xdsconv to convert to mtz file
6) run it in Phenix.autosol

De novo phasing of Lysozyme-Gd structure in PHENIX

Wizard status

FINISHED

Top solution:	9	Sites:	4	Space group:	P 43 21 2	FOM:	0.600
BAYES-CC:	55.12	Residues:	130	Side-chains:	130	Chains: :	2
Model CC:	0.87	R-work:	0.2214	R-free:	0.2482		



```
ssh psana
cd /reg/d/psdm/cxi/cxi84914/scratch/
cd knass
mkdir CASS-tutorial
cd CASS-tutorial
cp /reg/d/psdm/cxi/cxi84914/scratch/knass/scripts/* .
vim lysogdruns.csv
vim convert.cfg
vim runTemplateLCLS.sh
vim convert.cfg
sh convert_files.sh lysogdruns.csv
cd cass/
ls
cd hdf5/
ls
cd run_002/
ls
cd slice_00/
ls
cd aa/
cd ../../../../
cd ../
ls
vim submit.cfg
vim submit.cfg
vim lcls_header.txt
vim lcls_header.txt
source /reg/g/cfel/crystfel/crystfel-dev/setup-sh
sh run_automated_crystfel.sh lysogdruns.csv
ll cass/
ll cass/indexing-ref/
ll cass/indexing-ref/run_002/
less cass/indexing-ref/run_002/run-002aa.stream
vim crystfel.txt
vim merge-all
./create-xscale-kn lyso-gd.hkl > lyso-gd.xscale
cp ../../../scripts/XDSCONV.INP .
vim XDSCONV.INP
~/XDS-INTEL64_Linux_x86_64/xdsconv
 f2mtz HKLOUT temp.mtz<F2MTZ.INP
 cad HKLIN1 temp.mtz HKLOUT output_file_name.mtz<<EOF
ls
mv output_file_name.mtz lyso-gd.mtz
```