How to make structures better in the future and the use of published raw data in crystallographic software development

L.M.J. Kroon-Batenburg
Utrecht University
The Netherlands

DDDWG recommendations

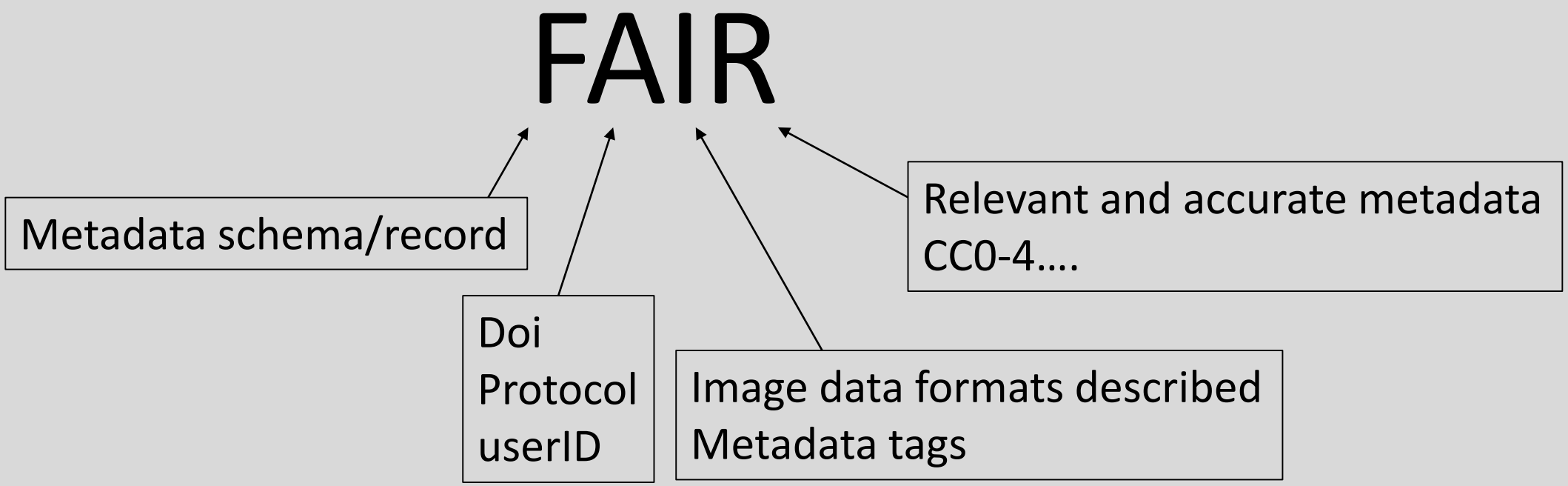
2011-2017

IUCr DDDWG Recommendations (top two)

•Authors should provide a **permanent and prominent link from their article to the raw data** sets which underpin their journal publication and associated database deposition of processed diffraction data (*e.g.* structure factor amplitudes and intensities) and coordinates, and which should obey the **'FAIR'** principles, that their raw diffraction data sets should be Findable, Accessible, Interoperable and Re-usable (https://www.force11.org/group/fairgroup/fairprinciples).

•A registered **Digital Object Identifier (doi)** should be the persistent identifier of choice (rather than a Uniform Resource Locator, url) as the most sustainable way to identify and locate a raw diffraction data set.

FAIR for raw data in crystallography



provides persistent identifiers (DOIs) for research data and other research outputs

DataCite: “x-ray diffraction” 47389 works

➔ **Raw images data, powder data, processed data or papers**

Raw data mostly:

- SBGrid
- IRRMC
- Zenodo
- CXI
- MX-RDR/RepOD

- Figshare
- Dryad
- Mendeley
- DataShare Edinburgh
- Universities of Manchester, Leeds, Bath, Aberdeen, Cambridge, St. Andrews, Strathclyde, Bristol, Cardiff, Utah
- Geological data

Raw data archive options for Chemical Crystallography

- Typical data set size (dependent on resolution and detector type): 100 MB – 4 GB

- Do (service) crystallographers have the infrastructure to archive **all** raw data?
  - University repository
  - Public repository
  - Neutron/Synchrotron facility
  - Hard drive/DVD/tape

- Funding agency adhere to Open Science policies:
  - all research data should become publicly available

Raw data archive options for Chemical Crystallography

- Typical data set size (dependent on resolution and detector type): 100 MB – 4 GB

- Do (service) crystallographers have the infrastructure to archive **all** raw data?
  - University repository
  - **Public repository -> Zenodo**
  - Neutron/Synchrotron facility
  - Hard drive/DVD/tape

- Funding agency adhere to Open Science policies:
  - all research data should become publicly available

Which raw data to archive?

- Data related to high-impact, complicated or remarkable structures with demanding structure solution and refinement
- Data with particular features, like (non-)merohedral twinning, multiple lattices, diffuse scattering, incommensurate modulation
- Data of from which the structure could not be solved

Where to archive?

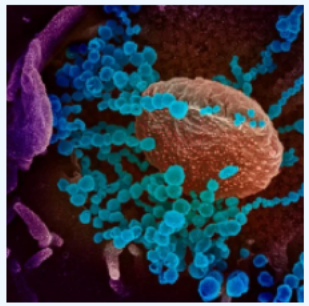
zenodo

Search

Upload Communities

Log in Sign up

Featured communities



Chicago COVID-19 Response

This repository community collects research outputs and i
efforts in Chicago. Users are encouraged to upload their re
discovery of information. Although Open Access articles ar

Curated by: saragon

- **Safe** — your research is stored safely for the future in CERN's Data Centre for as long as CERN exists.
- **Trusted** — built and operated by CERN and OpenAIRE to ensure that everyone can join in Open Science.
- **Citeable** — every upload is assigned a Digital Object Identifier (DOI), to make them citable and trackable.

Recent uploads

July 26, 2021 (v2) | Dataset | Open Access

View

Natural history specimens collected and/or identified and deposited.

Richard Pyle

Natural history specimen data collected and/or identified by Richard Pyle, https://orcid.org/0000-0003-0768-1286. Claims were made on Bionomia, https://bionomia.net using specimen data from the Global Biodiversity Information Facility, https://gbif.org.

Need help?

Contact us

Zenodo prioritizes all requested related to the COVID-19 outbreak.

We can help with:

- Uploading your research data, software, preprints, etc.

zenodo

Search

Upload

Communities

Log in

Sign up

Communities created and curated by Zenodo users

crystallography

Showing 0 to 10 out of 8309 communities.

Sort by

Coronavirus Disease Research Community - COVID-19

Featured

This community collects research outputs that may be relevant to the Coronavirus Disease (COVID-19) or the SARS-CoV-2. Scientists are encouraged to upload their outcome in this collection to facilitate sharing and discovery of information. Although Open Access articles and datasets are recommended, also closed and restricted access material are accepted. All types of research outputs can be included in this Community (Publication, Poster, Presentation, Dataset, Image, Video/Audio, Software, Lesson, Other).

Biodiversity Literature Repository

View

European Commission Funded Research (OpenAIRE)

View

Want your own community?

Sign Up

It's easy. Just click the button to get started.

- **Curate** — accept/reject what goes in your community collection.
- **Export** — your community collection is automatically exported via OAI-PMH
- **Upload** — get custom upload link to send to people

crystallography

Showing 0 to 5 out of 5 communities.

Sort by

Macromolecular Crystallography

View

Macromolecular crystallography community: people who use X-ray diffraction to gain insight into the structure and function of biological macromolecules such as proteins, DNA, carbohydrates & similar.

Curated by: graeme

Structural Genomics Consortium - Diamond Light Source I04-1 XChem Fragment screening by X-ray crystallography

View

Fragment-based screening is now well-established as a powerful approach to early drug, or 'lead', discovery. The principle is to identify weakly-binding compounds ('fragments') by screening a limited library of compounds, with resulting hits serving...

Curated by: SGC-DLS-XCHEM-I04-1-FS

Fragment based drug discovery by X-ray crystallography

View

Fragment based drug discovery (FBDD) has become an increasingly important tool for finding hit compounds for difficult targets. The technique utilises smaller than drug-like compounds to identify low potency, high quality leads. Libraries containing...

Chemical Crystallography

View

X-ray diffraction data for chemical crystallography, including but not limited to data from I19-1 and I19-2 at Diamond Light Source

Curated by: graeme

Chemical Crystallography

Recent uploads

Search Chemical Crystallography

December 11, 2020 (v1) | Dataset | Open Access

View

Diffraction data for CCDC 2031535–2031545 and 2031863–2031890. Claudetite I at various temperatures and pressures

Guńka, Piotr A.;

Diffraction data for claudetite I, monoclinic arsenic(III) oxide polymorph collected at room and low temperature (single-crystal diffraction data corresponding to CCDC 2031535-20311545) and at high pressure (powder diffraction data corresponding to CCDC 2031863–2031890).

Uploaded on December 14, 2020

Published in CrystEngComm, vol. 23, issue 3, pp. 638-644.

November 13, 2020 (v1) | Dataset | Open Access

View

Diffraction data for CCDC 1949980–1949982

Guńka, Piotr A.;

Diffraction data for: 10-hexyl-1-(10H-phenoxazin-10-yl)acridin-9(10H)-one 10-hexyl-2-(10H-phenoxazin-10-yl)acridin-9(10H)-one 10-hexyl-3-(10H-phenoxazin-10-yl)acridin-9(10H)-one

Uploaded on November 13, 2020

New upload

Chemical Crystallography

X-ray diffraction data for chemical crystallography, including but not limited to data from I19-1 and I19-2 at Diamond Light Source

Curated by:
graeme

Curation policy:
Not specified

Created:
April 14, 2016

Harvesting API:
OAI-PMH Interface

Want your upload to appear in this community?

- Click the button above to upload a record

November 13, 2020

Dataset Open Access

Diffraction data for CCDC 1515700

Guńka, Piotr A.

Project leader(s)

Kulszewicz-Bajer, Irena

Diffraction data for 8,16-dioctyl-8,16-dihydroacridino[2,1-a]acridine-5,13-dione

Polish National Science Centre, Grant No. 2015/17/B/ST5/00179

Files (3.6 GB)

Name	Size	
pg99_IKulszQA.7z	3.6 GB	Download

md5:cd88273f1871482ac1268306804c003e

Beta

Citations 0

Show only: Literature (0) Dataset (0) Software (0) Unknown (0)

Citations to this version

Search

No citations.

Publication date:
November 13, 2020

DOI:
DOI 10.5281/zenodo.4271549

Keyword(s):
X-ray diffraction data

Published in:
RSC Advances: 7 pp. 8627-8632 (14).

Related identifiers:
Supplement to
10.5517/ccdc.csd.cc1mw6k8 (Other)
10.1039/C6RA28567H (Journal article)

Communities:
Chemical Crystallography

CCDC FIZ Karlsruhe Leibniz Institute for Information Infrastructure

CSD Entry: JALNAO

Sign In

Simple Search | Structure Search | Unit Cell Search | Formula Search

Your query was: Identifier(s): 1515700 and the search returned 1 record.

Modify Search | New Search

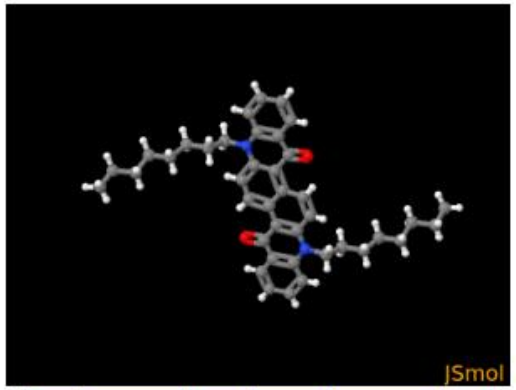
Results

	Database Identifier	Deposition Number
✔	JALNAO	1515700

Download

JALNAO : 8,16-dioctyl-8,16-dihydroacridino[2,1-a]acridine-5,13-dione
Space Group: P $2_1/c$ (14), **Cell:** *a* 12.6799(5)Å *b* 8.8317(3)Å *c* 15.5666(6)Å, *α* 90° *β* 109.439(4)° *γ* 90°

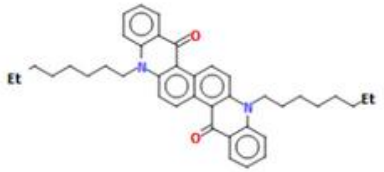
3D viewer



H | Disorder | Menu | Open

Style	Labels	Packing	Measure
Ball and Stick	No Labels	None	None

Chemical diagram



View group symbols key

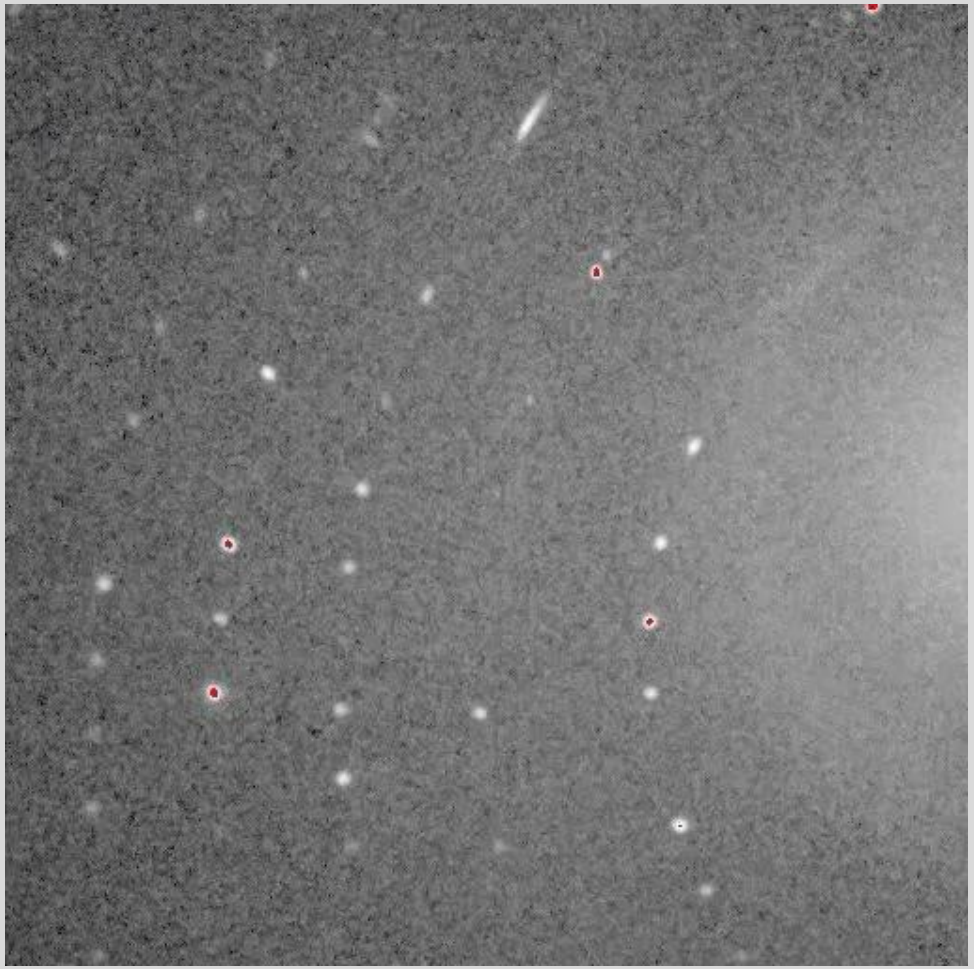
Additional details

Deposition Number	1515700
Data Citation	Piotr Kurzep, Łukasz Skórka, Malgorzata Zagorska, Piotr A. Guńka, Marzena Banasiewicz, Boleslaw Kozankiewicz, Irena Kulszewicz-Bajer CCDC 1515700: Experimental Crystal Structure Determination, 2017, DOI: 10.5517/ccdc.csd.cc1mw6k8
Deposited on	08/11/2016

Raw data DOI(s)

DOI	10.5281/zenodo.4271549

Carefully look at your data: Diaquobis(salicylato)copper(II)



- Reflections can be indexed in a orthorombic or monoclinic lattice
- Structures could be solved for h=even

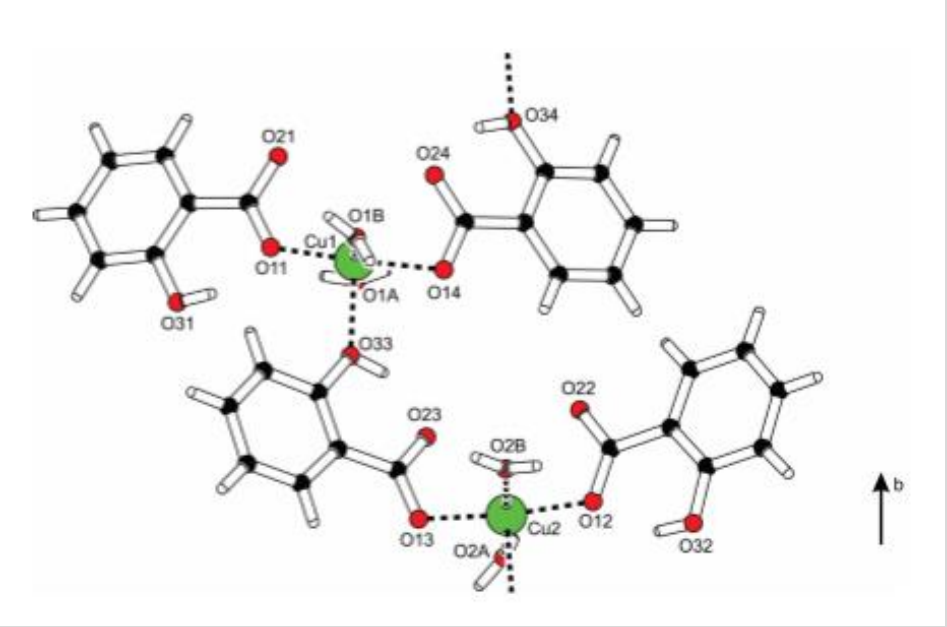
Not ignoring diffraction features

Table 3. Experimental details for the orthorhombic setting. Refined as an inversion twin. Contributions of the monoclinic structure removed from the structure factors

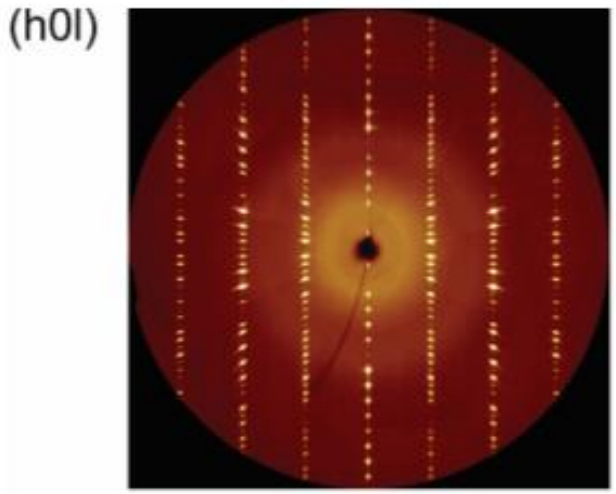
Molecular formula	$Cu(C_7H_5O_3)_2(H_2O)_2$
Formula weight	373.79
Crystal system	Orthorhombic
Space group	$Pca2_1$
a / Å	7.65385(11)
b / Å	11.7378(3)
c / Å	31.5707(6)

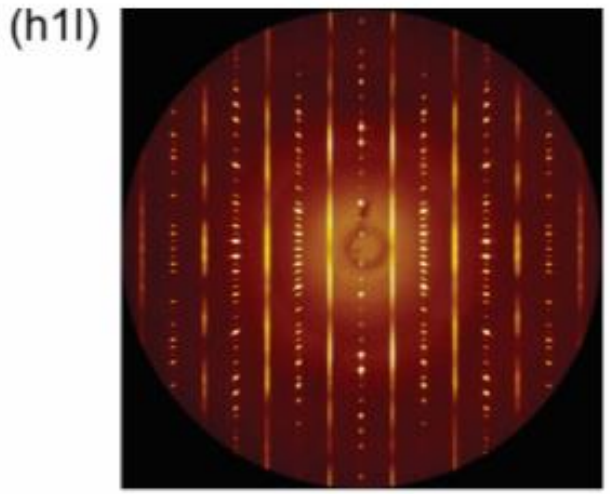
Table 4. Experimental details for the monoclinic setting. Refined as a rotation twin. Contributions of the orthorhombic structure removed from the structure factors.

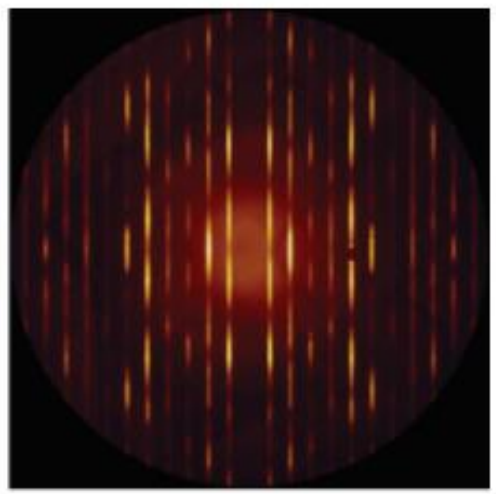
Molecular formula	$Cu(C_7H_5O_3)_2(H_2O)_2$
Formula weight	373.79
Crystal system	Monoclinic
Space group	$P2_1/a$
a / Å	7.6558(2)
b / Å	11.7408(3)
c / Å	31.8085(12)
β / °	96.931(2)



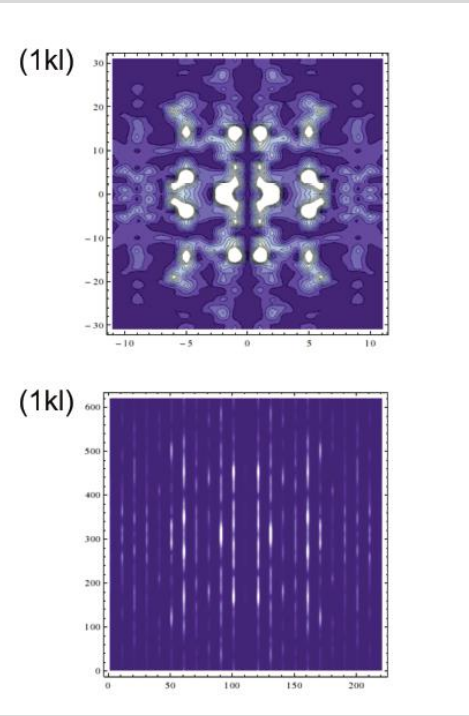
Reciprocal space





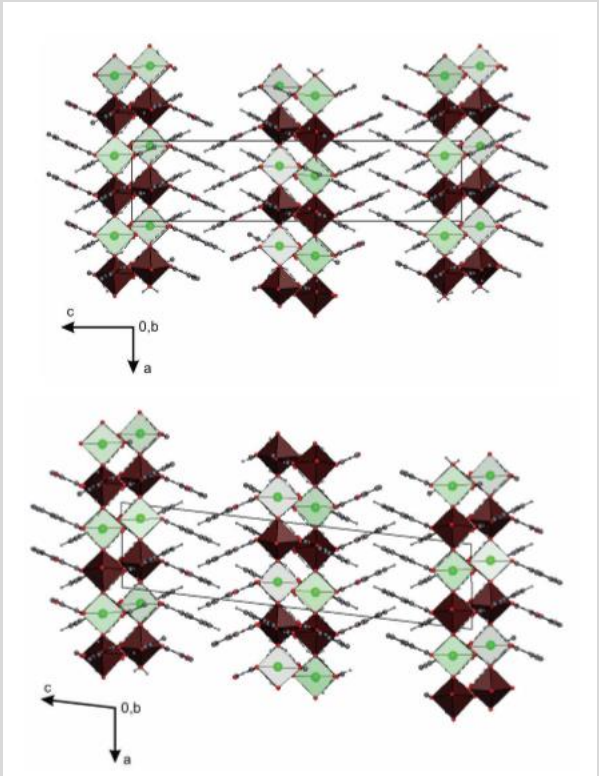


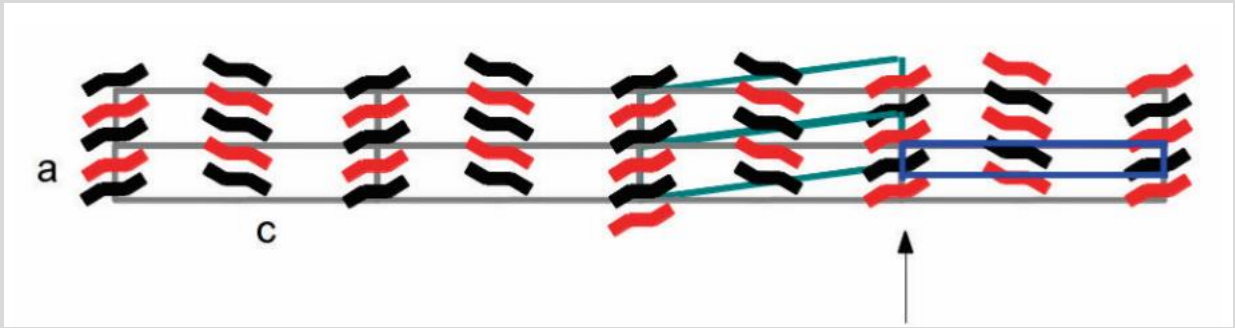
OD structure



ortho

mono





Random stacking faults

Croat. Chem. Acta 2018, *91*(2), 289–298

On the blue stacking lattice the structure is ordered

Structure and data validation

Raw data would complement the extensive archives of derived (coordinates) and processed (structure factors) data

- CheckCIF: syntax, consistency, completeness, validation
  Acta Cryst. C requires unmerged structure factors
- Macromolecular crystallography: PDB validation report wwPDB, Structure factors CIF and check with coordinates

- CheckCIF for raw data (CommDat/COMCIFS): *three levels*
  - ✓ Check for consistency, completeness of metadata
  - ✓ Reprocessing: sufficient metadata?
  - ✓ How much of the data is understood?
    indexing; lattice symmetry; pseudo merohedral twinning; incommensurate; data reduction; diffuse scattering

} Journal Editors/ Referees

March 27, 2019

Dataset | Open Access

Single-crystal X-ray diffractometry data for a sample of $NiCl_2$-dppe collected on beamline I19-2 at DIamond Light Source

Warren, Mark R; Allan, David R

Data curator(s)

Williams, Benjamin Heathcote

Single-crystal X-ray diffractometry data for a sample of [1,2-Bis(diphenylphosphino)ethane]dichloronickel(II) ($NiCl_2$-dppe, $[(C_6H_5)_2PCH_2CH_2P(C_6H_5)_2]NiCl_2$).

Data collected at Diamond Light Source I19-2 on 2015-05-18, publicly available for users to test data reduction routines. Data are known to produce good merging statistics and final refinements.

The sample was prepared as follows:
Nickel chloride (II) hexahydrate (1 g, 2 mmol) was heated under vacuum to produce anhydrous nickel chloride (II) with a visible colour change from green to yellow. The resulting solid was taken up in ethanol (5 ml) and added to 1,2-bis(dimethylphosphine)ethane (dppe) (0.837 g, 2 mmol) in ethanol (10 ml). The solution was refluxed for 3 hour after which the solvent was evaporated. The small red crystals were purified by recrystallisation in acetone (70% yield).

The sample was held at an approximate temperature of 150 K and the illuminating beam had a wavelength of 0.68890 Å (17.997 keV). The detector was held at 2θ = 25° throughout.

Inventory of data:
010_Ni_dppe_Cl_2_150K01 – 130° ω scan, 0.4° images, 0.4s per image, 325 images; κ = 45°, φ = 160°.
010_Ni_dppe_Cl_2_150K02 – 130° ω scan, 0.4° images, 0.4s per image, 325 images; κ = 45°, φ = 40°.
010_Ni_dppe_Cl_2_150K03 – 130° ω scan, 0.4° images, 0.4s per image, 325 images; κ = 45°, φ = -80°.
010_Ni_dppe_Cl_2_150K04 – 198° ω scan, 0.4° images, 0.4s per image, 495 images; κ = 0°, φ = -80°.

Files (315.2 MB)

Name	Size	
010_Ni_dppe_Cl_2_150K01.tar.xz md5:05339750aec15e311ef3af27ed901162	69.6 MB	Download
010_Ni_dppe_Cl_2_150K02.tar.xz md5:0c67738f6bdf79d66b6fff69c43a9ea5	69.7 MB	Download
010_Ni_dppe_Cl_2_150K03.tar.xz md5:c7425fe96d6bc852f588e0490900f9f9	69.7 MB	Download
010_Ni_dppe_Cl_2_150K04.tar.xz	106.2 MB	Download

Publication date:
March 27, 2019

DOI:
DOI 10.5281/zenodo.2611942

Keyword(s):
Diamond Light Source | I19-2 | NiCl2-dppe | data processing

Communities:
Chemical Crystallography

Versions

Version 2019-03-27 — Mar 27, 2019
10.5281/zenodo.2611942

Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.2611941. This DOI represents all versions, and will always resolve to the latest one. Read more.

Share

```
###CBF: VERSION 1.5, CBFlib v0.7.8 - PILATUS detectors

data_010_Ni_dppe_Cl_2_150K01_00001

_array_data.header_convention "PILATUS_1.2"
_array_data.header_contents
;
# Detector: PILATUS 300K, S/N 3-0104, Diamond
# 2015-05-18T21:41:55.427
# Pixel_size 172e-6 m x 172e-6 m
# Silicon sensor, thickness 0.000320 m
# Exposure_time 0.3970000 s
# Exposure_period 0.4000000 s
# Tau = 199.1e-09 s
# Count_cutoff 996974 counts
# Threshold_setting: 10000 eV
# Gain_setting: autog (vrf = 1.000)
# N_excluded_pixels = 31
# Excluded_pixels: badpixel_mask.tif
# Flat_field: FF_p3-0104_E20000_T10000_vrf_m0p100.tif
# Trim_file: p3-0104_E20000_T10000.bin
# Image_path: .../2015/cm12173-2/150518_bruker_test/010_Ni_dppe_Cl_2_150K/
# Wavelength 0.68890 A
# Energy_range (18, 18) eV
# Detector_distance 0.09529 m
# Detector_Voffset 0.00608 m
# Beam_xy (227.47, 310.68) pixels
# Flux 0.000000
# Filter_transmission 1.0000
# Start_angle -92.0000 deg.
# Angle_increment 0.4000 deg.
# Detector_2theta 25.0000 deg.
# Polarization 0.990
# Alpha 50.0000 deg.
# Kappa 45.0000 deg.
# Phi 160.0000 deg.
# Phi_increment 0.0000 deg.
# Omega -92.0000 deg.
# Omega_increment 0.4000 deg.
# Chi 45.0000 deg.
# Chi_increment 0.0000 deg.
# Oscillation_axis OMEGA
# N_oscillations 325
# CBF_template_file: /dls_sw/i19/software/gda/config/etc/i19.eh2.template.cbf
;

# This and all subsequent lines will appear in the header
# Template modified 08:20 27/JUN/2013
```

miniCBF

Full ImgCIF/CBF

```
loop_
_diffrn_scan_frame_axis.frame_id
_diffrn_scan_frame_axis.axis_id
_diffrn_scan_frame_axis.angle
_diffrn_scan_frame_axis.displacement
FRAME1 GONIOMETER_OMEGA -92.0000 0.0
FRAME1 GONIOMETER_KAPPA 45.0000 0.0
FRAME1 GONIOMETER_PHI 160.0000 0.0
FRAME1 DETECTOR_2THETA 25.0000 0.0
FRAME1 DETECTOR_Z 0.0 95.29
FRAME1 DETECTOR_Y 0.0 0.0
FRAME1 DETECTOR_X 0.0 0.0

loop_
_axis.id
_axis.type
_axis.equipment
_axis.depends_on
_axis.vector[1] _axis.vector[2] _axis.vector[3]
_axis.offset[1] _axis.offset[2] _axis.offset[3]
GONIOMETER_OMEGA rotation goniometer . 1 0 0 . . .
GONIOMETER_KAPPA rotation goniometer GONIOMETER_OMEGA 0.642788 -0.766044 0 . . .
GONIOMETER_PHI   rotation goniometer GONIOMETER_KAPPA 1 0 0 . . .
SOURCE           general source . 0 0 1 . . .
GRAVITY          general gravity . 0 -1 0 . . .
DETECTOR_2THETA  rotation    detector . 1 0 0 . . .
DETECTOR_Z       translation detector DETECTOR_2THETA 0 0 -1 0 0 0
DETECTOR_Y       translation detector DETECTOR_Z 0 -1 0 0 0 0
DETECTOR_X       translation detector DETECTOR_Y 1 0 0 0 0 0
ELEMENT_X        translation detector DETECTOR_X 1 0 0 -39.12 53.44 0
ELEMENT_Y        translation detector ELEMENT_X 0 -1 0 0 0 0

loop_
_array_structure_list.array_id
_array_structure_list.index
_array_structure_list.dimension
_array_structure_list.precedence
_array_structure_list.direction
_array_structure_list.axis_set_id
ARRAY1 1 487 1 increasing ELEMENT_X
ARRAY1 2 619 2 increasing ELEMENT_Y

loop_
_array_structure_list_axis.axis_set_id
_array_structure_list_axis.axis_id
_array_structure_list_axis.displacement
_array_structure_list_axis.displacement_increment
ELEMENT_X ELEMENT_X 0.0 0.1720
ELEMENT_Y ELEMENT_Y 0.0 0.1720

loop_
_array_element_size.array_id
```

How “Raw” can/should the data be?

March 15, 2019 (v1) | Dataset | Open Access

View

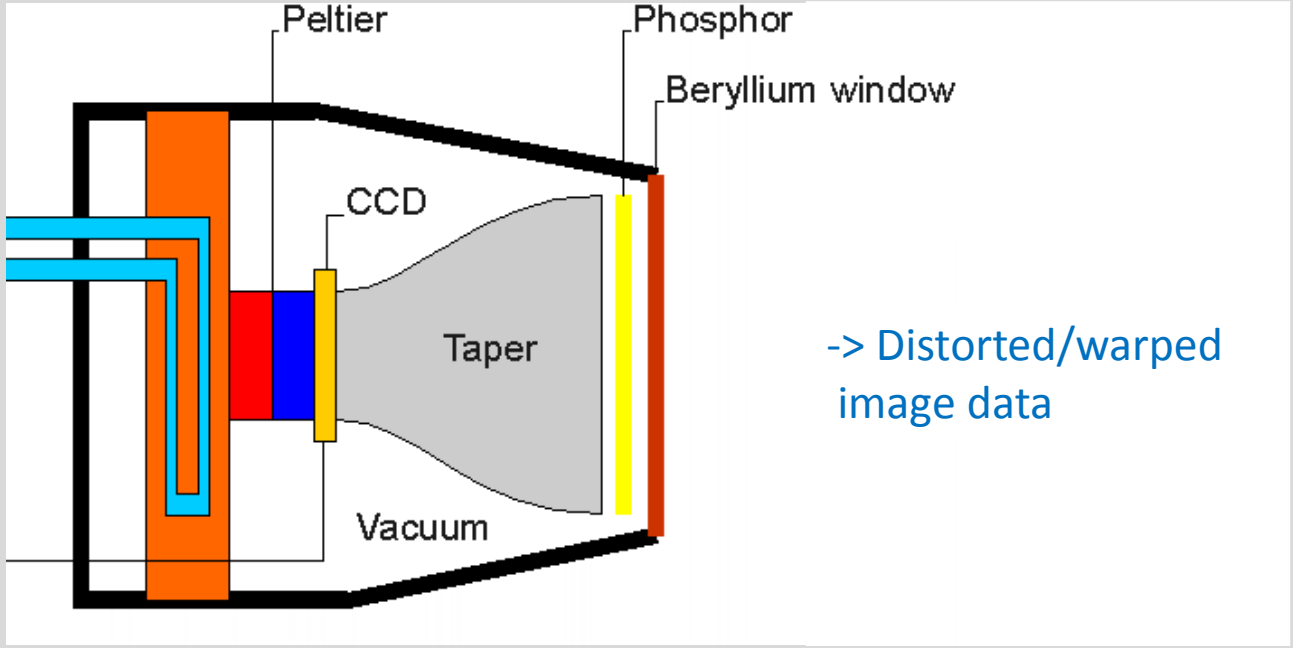
Single crystal diffraction raw data for Fampridine hydrochloride Phase 1

Coles, Simon; Montis, Riccardo; Horton, Peter; Hursthouse, Michael;

A set of diffraction images (in both Denzo .x and Bruker Nonius .kcd formats) collected on a Bruker Nonius KappaCCD diffractometer on an FR591 rotating anode generator equipped with confocal mirrors. The sample is Phase 1 of the Fampridine hydrochloride organic salt - a system that forms numerous c

Uploaded on March 21, 2019

BrukerNonius KappaCCD images
FR591 rotating anode



-> Distorted/warped image data

Data from in-house diffractometer with commercial software package

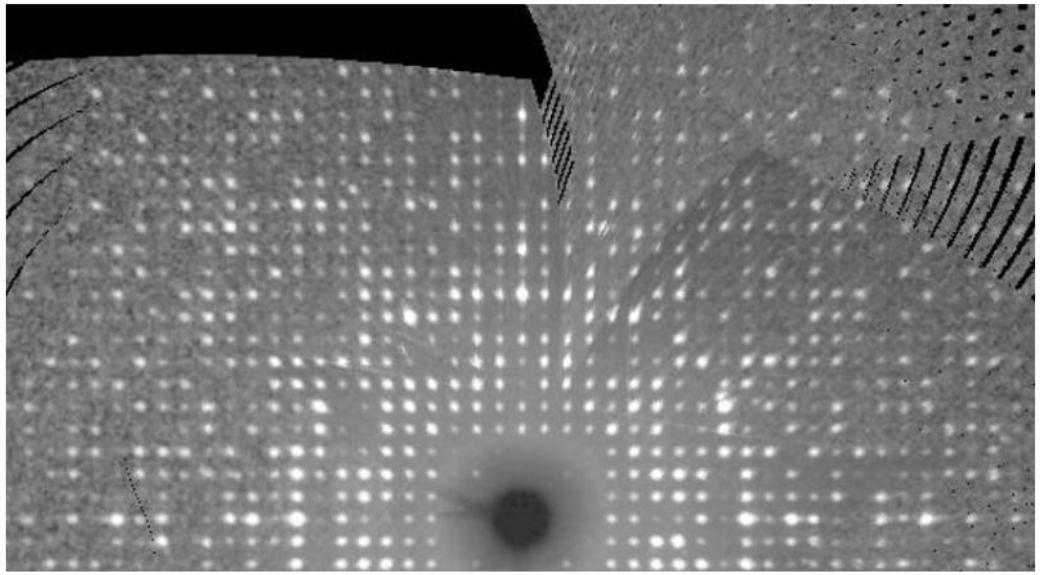
Single crystal diffraction raw data for Fampridine hydrochloride Phase 1

Coles, Simon; Montis, Riccardo; Horton, Peter; Hursthouse, Michael

A set of diffraction images (in both Denzo .x and Bruker Nonius .kcd formats) collected on a Bruker Nonius KappaCCD diffractometer on an FR591 rotating anode generator equipped with confocal mirrors.

The sample is Phase 1 of the Fampridine hydrochloride organic salt - a system that forms numerous complicated phases. The authors wish to make the raw data available so that those who wish to explore the modelling of these exceptionally complex systems may process the data themselves.

Preview



Files (384.3 MB)

Name

0kl.jpg

md5:afe35e52223ba22723fed25b18a1be19

0kl.syn

md5:c50261f7f4f2715170a01b4156108249

apex2-0625-00315-625.zip 1.4 MB

md5

Arc 282.8 M

md5

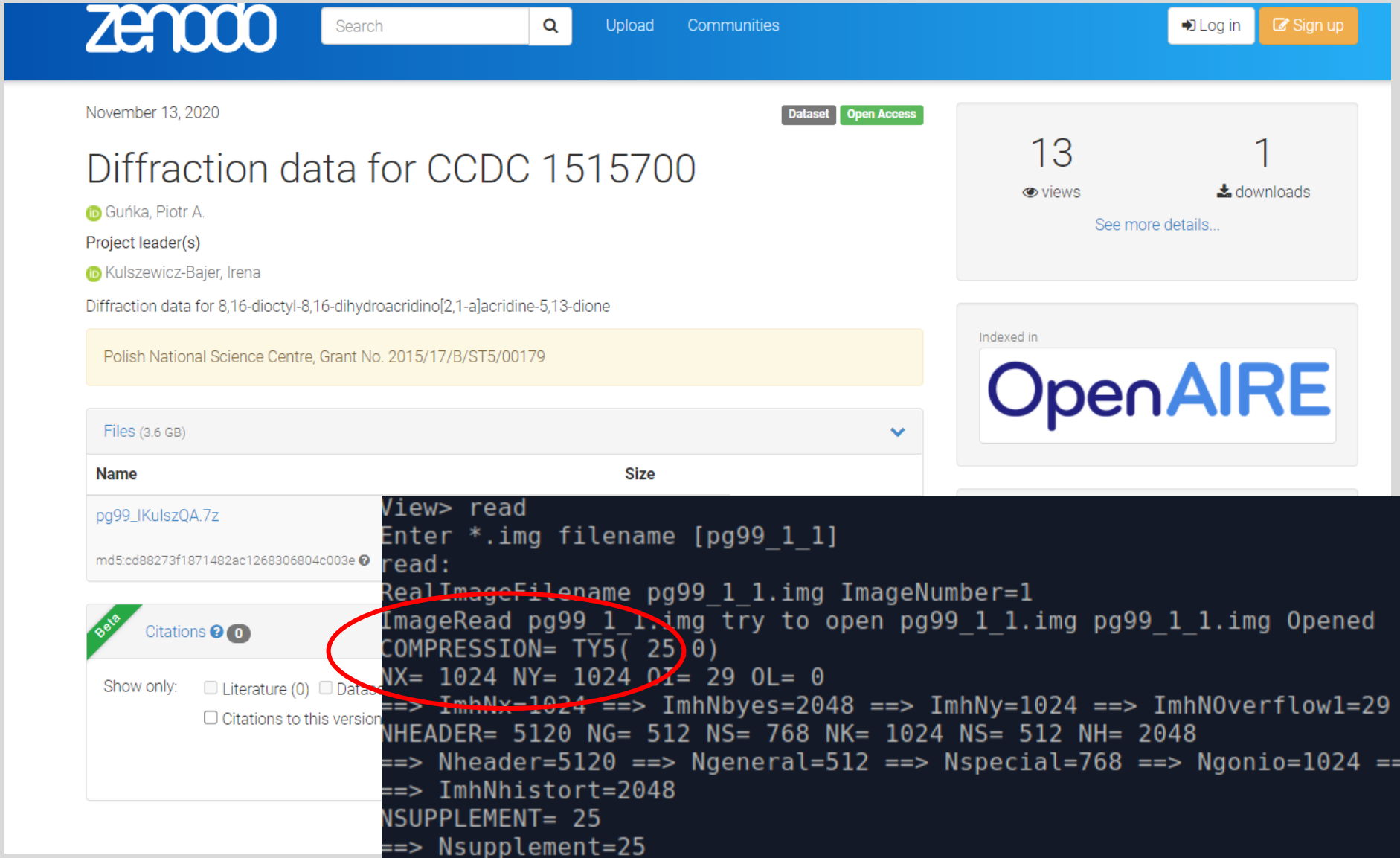
cell.drx 12.1 kB

md5:1bdc37dbbc960477f62239fb56c9dfd9

```
bate@phe:apex2-0625-00315-625> ll
total 3632
-rw-r--r-- 1 bate ks     237 Jul 27 15:18 MO-incidence.coefficient
-rw-r--r-- 1 bate ks   63898 Jul 27 15:18 badpixel.dark
-rw-r--r-- 1 bate ks    4223 Jul 27 15:18 badpixel.sdark
-rw-r--r-- 1 bate ks  724100 Jul 27 15:18 dark.kcd_frozen
-rw-r--r-- 1 bate ks     163 Jul 27 15:18 def.cal
-rw-r--r-- 1 bate ks     458 Jul 27 15:18 def.site
-rw-r--r-- 1 bate ks     287 Jul 27 15:18 detalign.vic
-rw-r--r-- 1 bate ks     372 Jul 27 15:18 distorpol.vic
-rw-r--r-- 1 bate ks  724100 Jul 27 15:18 responsMO.kcd
-rw-r--r-- 1 bate ks  724100 Jul 27 15:18 sdark.kcd_frozen
-rw-r--r-- 1 bate ks 1444100 Jul 27 15:18 sensi.kcd
-rw-r--r-- 1 bate ks      19 Jul 27 17:10 view.init
bate@phe:apex2-0625-00315-625>
```

```
DISTORDETID apex2-0625-00315-625
DISTORMAXMMX    30.72
DISTORMAXMMY    30.72
DISTORMAXPIXX    576.00
DISTORMAXPIXY    621.00
DISTORPOLORDER 1 1
DISTORMM2PX
 0.000000 -1.000000
 0.000000  0.000000
DISTORMM2PY
 0.000000  0.000000
 1.000000  0.000000
DISTORP2MMX
 0.000000  1.000000
 0.000000  0.000000
DISTORP2MMY
 0.000000  0.000000
-1.000000  0.000000
DISTORRMSERR 0.0800
```

Shouldn't the images be unwarped, corrected for sensitivity and dark current and dezingered to be re-usable?



EVAL software

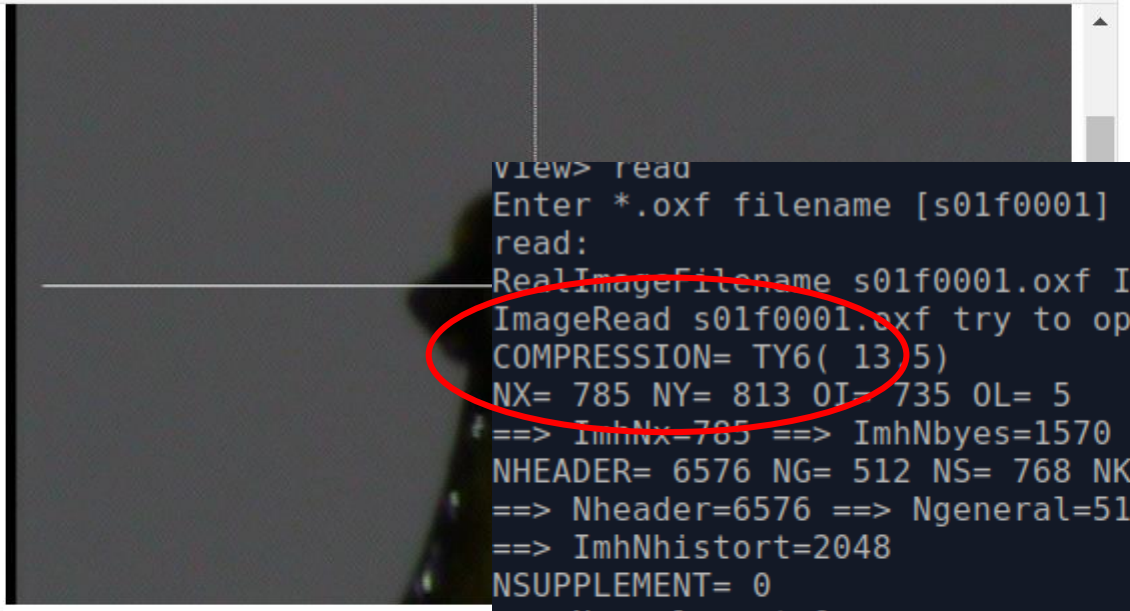
Single crystal diffraction raw data for Fampridine hydrochloride Phase 2

Coles, Simon; Montis, Riccardo; Horton, Peter; Hursthouse, Michael

A set of diffraction images in Rigaku Oxford Diffraction (img) format, collected on a Rigaku FRE+ diffractometer, equipped with HF Varimax confocal mirrors and an AFC12 goniometer and HG Saturn 724+ detector diffractometer.

The sample is Phase 2 of the Fampridine hydrochloride organic salt - a system that forms numerous complicated phases. The authors wish to make the raw data available so that those who wish to explore the modelling of these exceptionally complex systems may process the data themselves.

Preview



```
View> read
Enter *.oxf filename [s01f0001]
read:
RealImageFilename s01f0001.oxf ImageNumber=1
ImageRead s01f0001.oxf try to open s01f0001.oxf s01f0001.oxf Opened
COMPRESSION= TY6( 13 5)
NX= 785 NY= 813 OI= 735 OL= 5
==> ImhNx=785 ==> ImhNbyes=1570 ==> ImhNy=813 ==> ImhNOverflow1=735 ==> ImhNOverflow2=5
NHEADER= 6576 NG= 512 NS= 768 NK= 1024 NS= 512 NH= 2048
==> Nheader=6576 ==> Ngeneral=512 ==> Nspecial=768 ==> Ngonio=1024 ==> Nstatistics=512
==> ImhNhistort=2048
NSUPPLEMENT= 0
==> Nsupplement=0
TIME= Tue Nov 21 17:39:46 2017
==> ImhDatetime=Tue Nov 21 17:39:46 2017
General: ibx=1 iby=1 ix1=1 iy1=1 ix2=785 iy2=813 ikmtime=160 imode=0 iinttime=200 iyb=0
iyt=0 ixw=785 iyw=813 inx=785 iny=813 ixl=0 ixr=0 dwnofpoints=572413 inoofimages=1
==> ImhBinHor=1 ==> ImhBinVer=1 ==> ImhBinned=false ==> ImhNx=813 ==> ImhNy=785
==> ImhNBytes=638205
WARNING Main [read]: Nx=785 Ny=813 Nx*Ny=638205 differs from 572413
```

Images formats and compressions should be publically available

EVAL software

Re: [Fable-talk] [fabio] OXD images (with TY5 compression) (#19)

Re: [Fable-talk] [fabio] OXD images (with TY5 compression) (#19)

From: Jon Wright <wright@es...> - 2015-07-09 22:50:20

Vendors should make provisions for exporting raw data in a re-usable format

```
On 09/07/2015 21:14, Jerome Kieffer wrote:
> On Thu, 09 Jul 2015 05:57:12 -0700
> PhHarder <notifications@...> wrote:
>
>  > Hello,
>  > i also came across some OXD files with the TY5 compression.
>  > I think there is some additional information in the bitstream of the
> file, which produces artifacts in the resulting picture.
>  > If it helps i can provide some data and pictures.
>
> The big problem, to my understanding, is the fact that TY5 files we got
> up to now are raw (and wrapped) and uncompressed files have been
> unwrapped (i.e.
> corrected for geometric distortion).
> As we are searching bit-patterns to perform the reverse-engineering, if
> the two binary blobs are not exactly the same it is not worth spending
> time on it.
```

New section in IUCrData: Raw Data Letters

IUCrData, the peer-reviewed open-access data publication from the International Union of Crystallography (IUCr), is launching a new section for authors to describe their unprocessed or "raw" diffraction images. This is a collaborative innovation of IUCr Journals with the IUCr Committee on Data.

The new section will publish short descriptions of crystallographic raw data sets in the biological, chemical or materials science fields and provide a persistent link to the location of the raw data. This will allow researchers to attract attention to particular features of the data that could be of interest to methods and software developers or may be relevant to the structural interpretation.

The IUCr will adhere to the FAIR principles for which the correctness and completeness of the metadata are crucial, and these will be central to the reviewing process.

IUCrData

ISSN 2414-3146

Received 20 April 2021
Accepted 1 May 2021

Keywords: twinning; diffuse scattering; tetraspanin $CD9_{EC2}$.

Raw data reference: the doi will appear here
The raw data have been checked with checkCIF and an imgCIF file of metadata is available from iucrdata.iucr.org

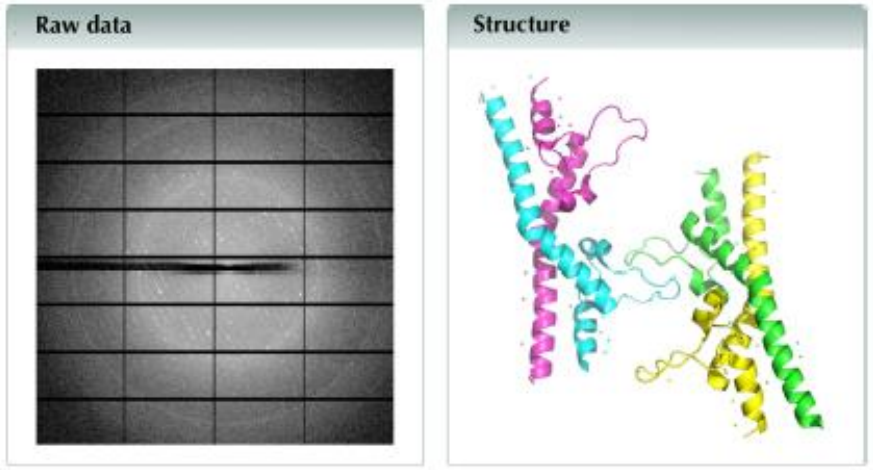
PDB reference: tetraspanin CD9, 6rlr

The extracellular domain of human tetraspanin CD9: twinning and diffuse scattering

Wout Oosterheert, Viviana Neviani, Martin Lutz, Piet Gros and Loes Kroon-Batenburg*

Department of Chemistry, Structural Biochemistry, Bijvoet Centre for Biomolecular Research, Faculty of Science, Utrecht University, Utrecht, The Netherlands. *Correspondence e-mail: l.m.j.kroon-batenburg@uu.nl

We describe remarkable features in the diffraction pattern produced by a crystal of tetraspanin $CD9_{EC2}$. $CD9_{EC2}$ crystallized in space group $P1$ and was twinned. Concurrent with the twinning, diffuse streaks were seen in the direction perpendicular to the twinning interface. We make preliminary conclusions on packing disorder and potential implications for the observed molecular structure. We envision that the raw diffraction images can be very useful for methods developers in trying to remove the diffuse scattering to extract accurate Bragg intensities or by using it to model the effect of packing disorder on the molecular structure.



The new section will be accepting submissions from the autumn and anyone wanting to know more should contact the IUCr Editorial Office (med@iucr.org).

Table 1
Experimental details.

The data for highest resolution shell are given in parentheses.

Data collection	
Beamline	DLS I-04
Detector type	
Data binary format and compression type	
Beam centre	
Wavelength (Å)	0.9795
Wavelength bandpass	
Orientation of goniometer and detector axes	
Rotation axis and increment per frame	
Pixel size	
Detectpor position	
Number of pixels	
Binning mode	
Resolution range (Å)	29.02–2.0 (2.07–2.0)
Crystal system, space group	Triclinic, *P*1
Cell dimensions *a*, *b*, *c* (Å)	39.986, 39.998, 63.643
Cell angles α, β, γ (°)	80.39, 76.29, 68.15
Total reflections	76175 (6985)
Unique reflections	22863 (2180)
Completeness	95.6 (93.0)
Multiplicity	0.08
$I/\sigma(I)$	4.8 (0.8)
R_{merge}	0.10 (1.15)
$R_{p.i.m.}$	0.07 (0.73)
$CC_{1/2}$	0.998 (0.322)

CheckCif for raw data

- Parser to extract meta data and write it in imgCIF
- Check validity of meta data

Conclusions

- It is strongly encouraged to archive raw data sets associated with publications
- It is possible to archive raw data sets for free: Zenodo
- The CX community has started publishing raw data
- The CX community should discuss what is meant with "Raw" data: unprocessed vs. re-usable