

MANCHESTER
1824

The University of Manchester

IUCr Diffraction Data Deposition Working Group



*Access to raw diffraction data:
A revolution in the making*

John R Helliwell

School of Chemistry, University of
Manchester, M13 9PL, UK

Polish Crystallographic conference

john.helliwell@manchester.ac.uk

Overview

Towards a future structural science based on raw data archiving: the IUCr's
Diffraction Data Deposition Working Group (DDDWG)

- *IUCr Commissions* are actively working on
"defining their commission's metadata for raw diffraction data"
namely:-
 - Commission on EXAFS;*
 - Commission on Small angle scattering;*
 - Commission on High pressure;*
 - Commission on Biological Macromolecules.*
- *The ICDD has been active on the harnessing of raw powder diffraction data sets* for some time and reported to us at ECM30 in Rovinj that they now have incorporated 10,000 raw powder diffraction data sets into their powder diffraction file. *The Commission on Powder Diffraction* is planning further work on neutron powder diffraction raw data and will liaise with the Commission on Neutron Scattering as appropriate.
- The *Commission on Structural Chemistry* had enthusiastic participants in Madrid, Bergen and Rovinj DDDWG events.

2014

international year of
crystallography



United Nations
Educational, Scientific and
Cultural Organization



International
Union of
Crystallography

Partners for the International Year of Crystallography 2014

*Crystallography
in
modern life*

*2014 is proclaimed International
Year of Crystallography*

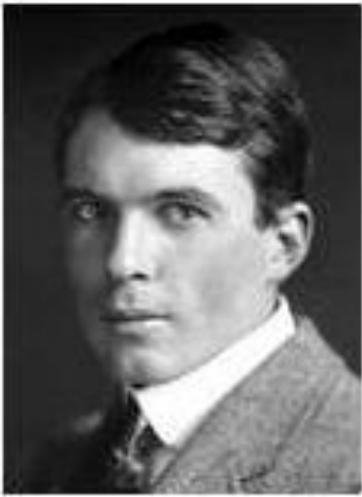


United Nations

*“Crystallography has an important place
as we work for inclusive sustainable
development – policies that are good for
people and the planet”*

*Ban Ki-Moon, UN Secretary-General • IYCr 2014
Opening Ceremony*

***Some history; let's go back to the first
crystal structures.....***



William Lawrence Bragg

Raw data 1913

Bragg, W. L. (1913). The Structure of Some Crystals as Indicated by their Diffraction of X-rays. Proc. R. Soc. London Ser. A, 89, 248-277; contained a lot of raw data (Laue patterns)

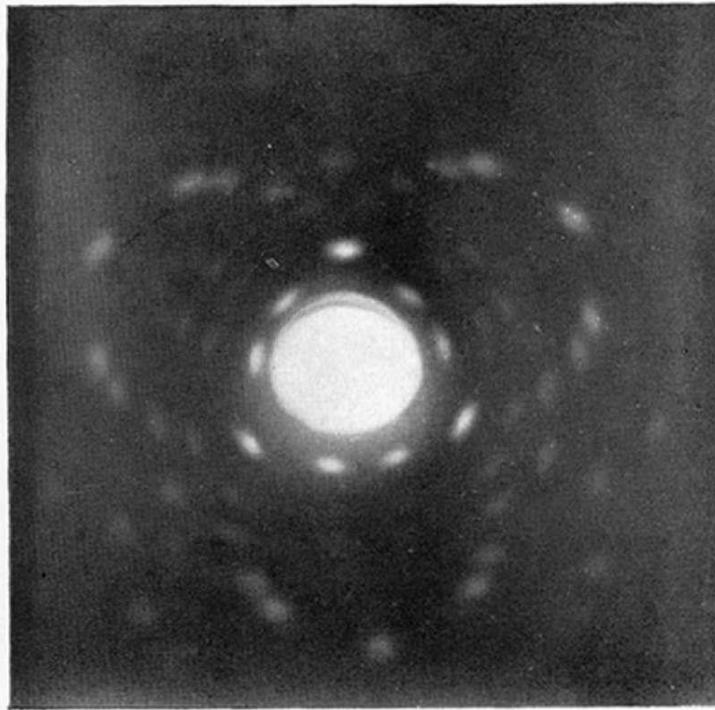


FIG. 11.—Fluorspar.

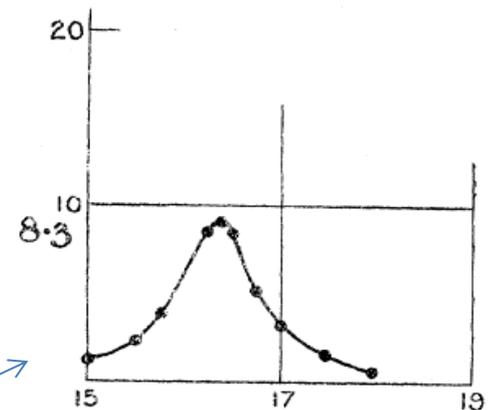
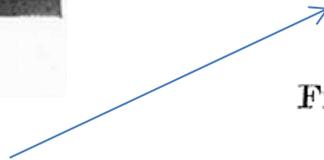


FIG. 1.—Regular reflection from cleavage face of rock-salt, glancing angle 8.3° .



W. H. Bragg and W. L. Bragg The Reflection of X-rays by Crystals Proc. R. Soc. Lond. A 1913 88, 428-438 using the WHB X-ray spectrometer



Sir William Henry Bragg

Processed diffraction data 1913

now assume a face centred lattice

		Calc ^d	Actual value found
$d_{100} = \frac{a \cdot 951}{2} = 3.04 \times 10^{-8}$	$\sin \theta_{100} = \frac{\lambda}{2d} = .0947$	$\theta = 5.43$	5.35
$d_{1\bar{1}0} = \frac{a \cdot 779}{2} = 2.48 \times 10^{-8}$	$\sin \theta_{1\bar{1}0} = .1161$	$\theta = 6.67$	6.6
$d_{110} = \frac{a \cdot 601}{2} = 1.917 \times 10^{-8}$	$\sin \theta_{110} = .1502$	$\theta = 8.63$	8.70
$d_{111} = \frac{a \cdot 437}{2} = 1.79 \times 10^{-8}$	$\sin \theta_{111} = .2064$	$\theta = 11.92$	5.71
$d_{2\bar{1}\bar{1}} = \frac{a \cdot 437}{2} = 1.43 \times 10^{-8}$	$\sin \theta_{2\bar{1}\bar{1}} = .2064$	$\theta = 11.92$	11.65

Plane.	Calculated Spacing.	Calculated Angle.	Observed Angle.
100	$d_{(100)} = 3.04.$	$5^{\circ} 26'$	$5^{\circ} 21'$
$1\bar{1}0$	$d_{(1\bar{1}0)} = 2.48.$	$6^{\circ} 40'$	$6^{\circ} 36'$
110	$d_{(110)} = 1.92.$	$8^{\circ} 38'$	$8^{\circ} 42'$
111	$d_{(111)} = 2.79.$	$5^{\circ} 55'$	$5^{\circ} 46'$
$2\bar{1}\bar{1}$	$d_{(2\bar{1}\bar{1})} = 1.43.$	$11^{\circ} 35'$	$11^{\circ} 39'$

The upper figure is from the Braggs' notebook, showing observed angular locations of diffracted beams from different crystal planes, and their calculated values. The lower figure is a similar published tabulation from the 1915 book X-rays and Crystal Structure.



The Nobel Prize in Chemistry 1964
Dorothy Crowfoot Hodgkin

The Nobel Prize in Chemistry 1964



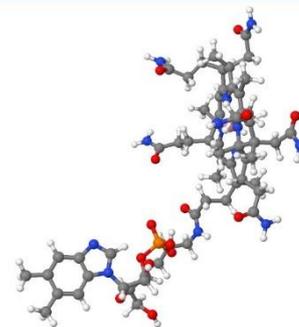
***Dorothy Hodgkin (1910-1994):
still a major inspiration today***

Dorothy Crowfoot
Hodgkin

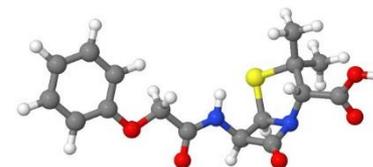
Prize share: 1/1

The Nobel Prize in Chemistry 1964 was awarded to Dorothy Crowfoot Hodgkin *"for her determinations by X-ray techniques of the structures of important biochemical substances"*.

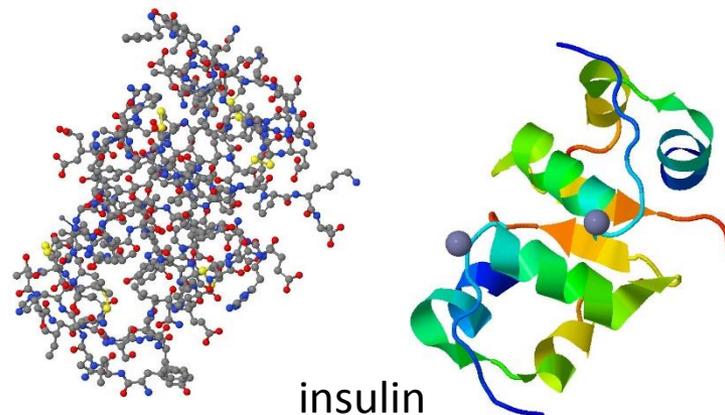
Photos: Copyright © The Nobel Foundation



vitamin B12 (cobalamin)

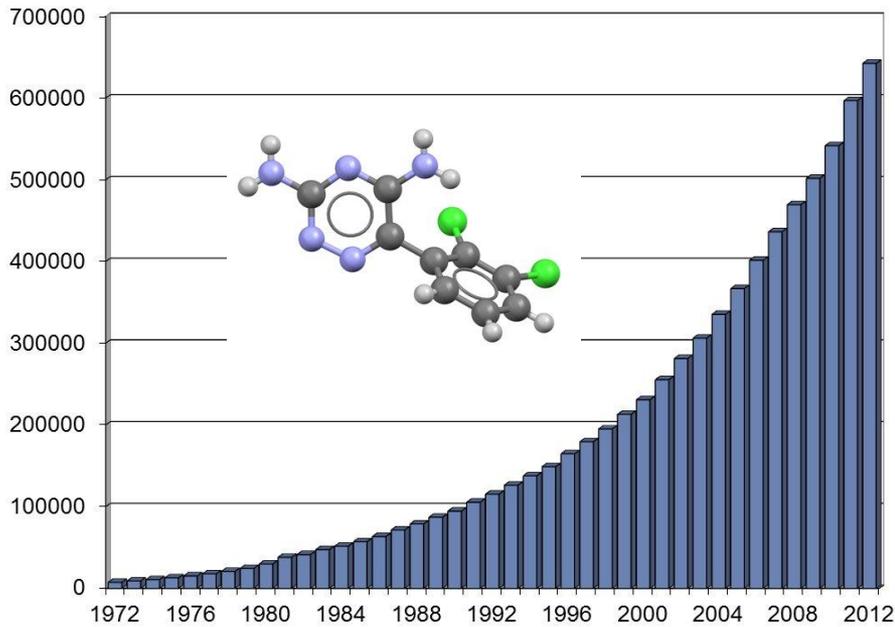


penicillin

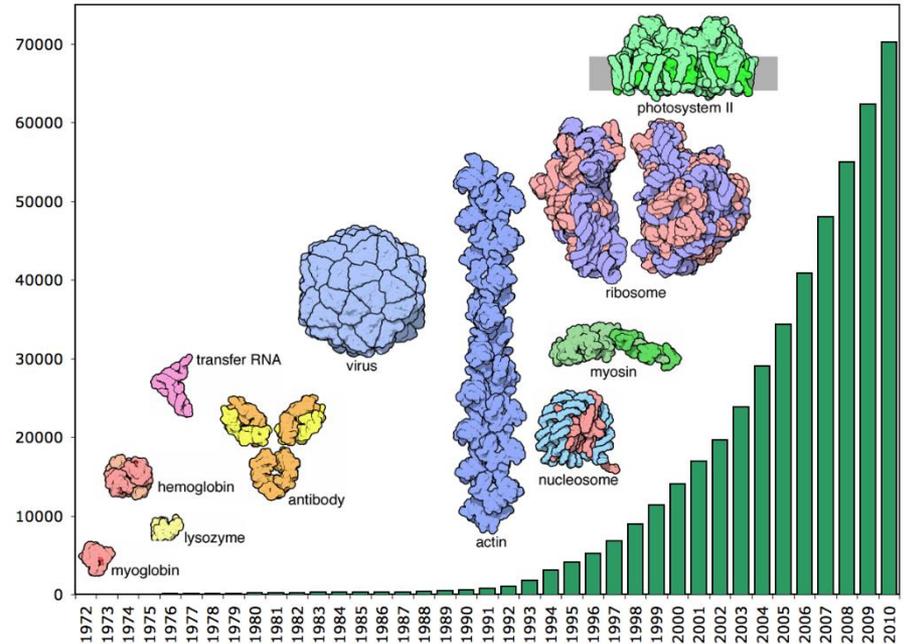


insulin

Structures large and small

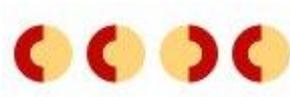


Cambridge Structural Database

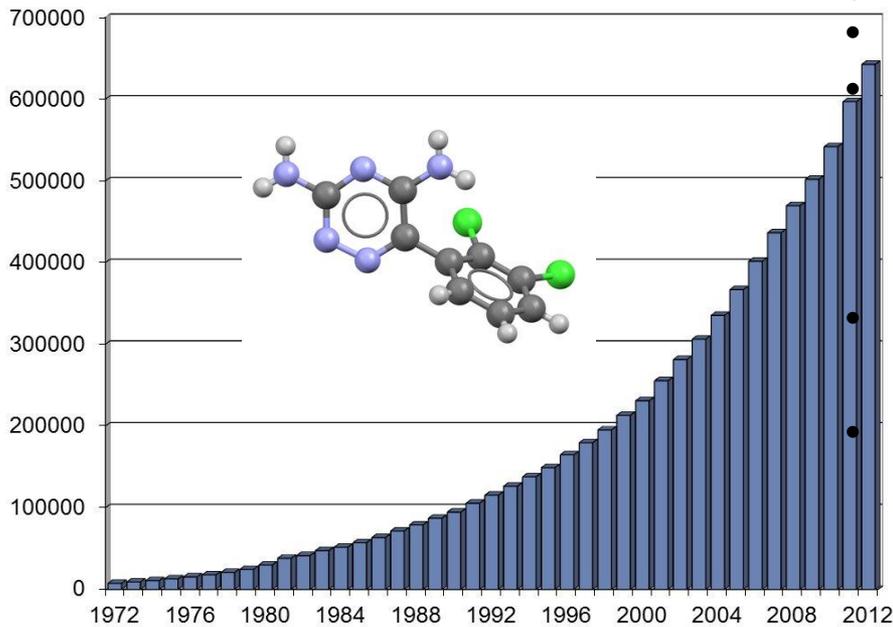


Protein Data Bank

Interactions between large and small

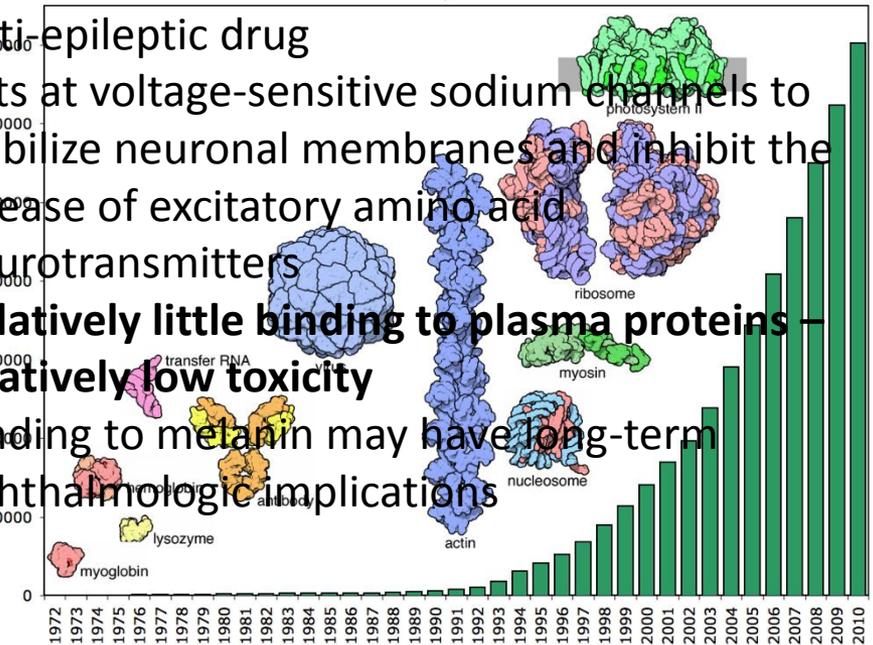


Lamotrigine
(500,000th structure in CCDC)



Cambridge Structural Database

- Anti-epileptic drug
- Acts at voltage-sensitive sodium channels to stabilize neuronal membranes and inhibit the release of excitatory amino acid neurotransmitters
- **Relatively little binding to plasma proteins – relatively low toxicity**
- Binding to melanin may have long-term ophthalmologic implications



Protein Data Bank

What has this all to do with “data”?

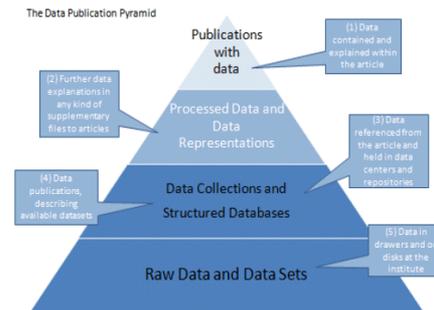
- *Structural models, stored in curated databases, have immense value for comparative studies, new compound discovery etc.*
- *Experimental data, stored as reduced and processed data sets, are invaluable for validating models and re-refinement of structures*
- *Raw experimental data has the potential for unleashing new methods and new science*

” Ideally, the full scientific record should provide access to the raw data.....the IUCr is beginning to consider longer-term approaches to archiving the raw data”

P. Strickland, B. McMahon and J. R. Helliwell, Learned Publishing 21 (2008) 63.

Raw diffraction images offer the opportunity of

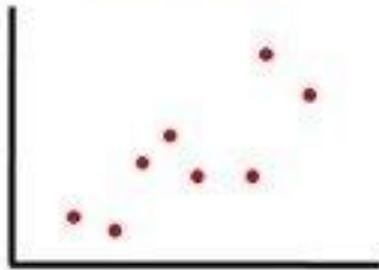
- *analysing data at higher resolution than used in the original work*
 - *serving as benchmarks in developing improved methods of analysis*
 - *checking the interpretation of the symmetries of the crystals*
 - *analysing diffraction from multiple lattices present in the crystals*
 - *analysing the diffuse scattering that reflects correlated motions or disorder of atoms in the crystals*
- **Nb the *philosophical view* of the importance of access to raw diffraction data; namely analysis through one's own eyes not the lens of someone else**



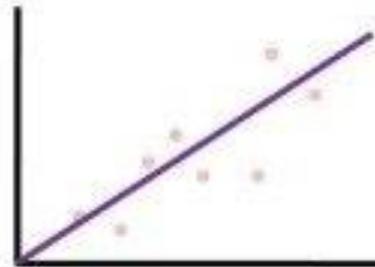
Data seen through the lens of others!

PhD tips # 22

Actual data



How you saw it:
Perfectly linear, as expected



How your supervisor saw it:
Perfectly exponential, amazing new stuff



How the referee saw it:
Pure noise

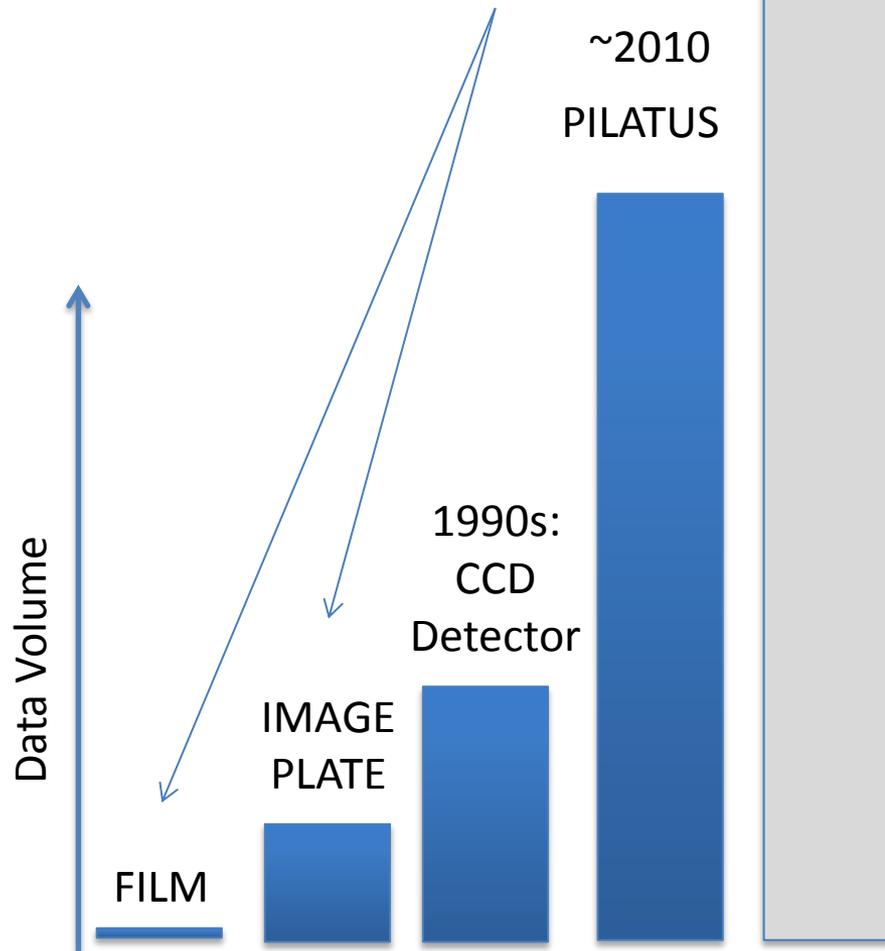


Modern data flows in biological crystallography are extreme>>>>

Would keeping all these raw data be “worth the pain”?

*A group of 4 articles in Acta Cryst D October 2014 Biological Crystallography led by **Tom Terwilliger** explained why keeping raw data is a natural next step for crystallography*

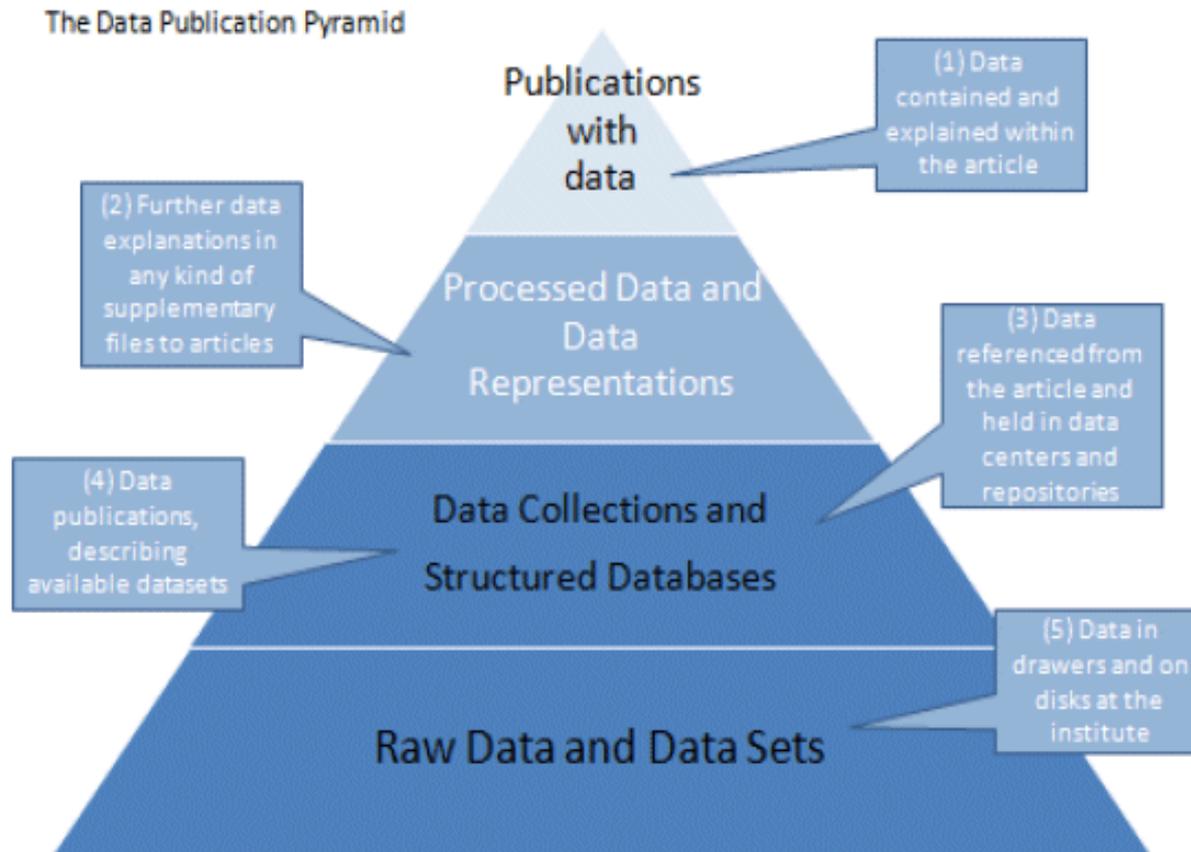
*Biological crystallography Labs at SR Sources;
data flows have
increased dramatically since the 1980s*



The University of Manchester
Manchester Institute of Biotechnology

The data publication pyramid

- the publishers' view

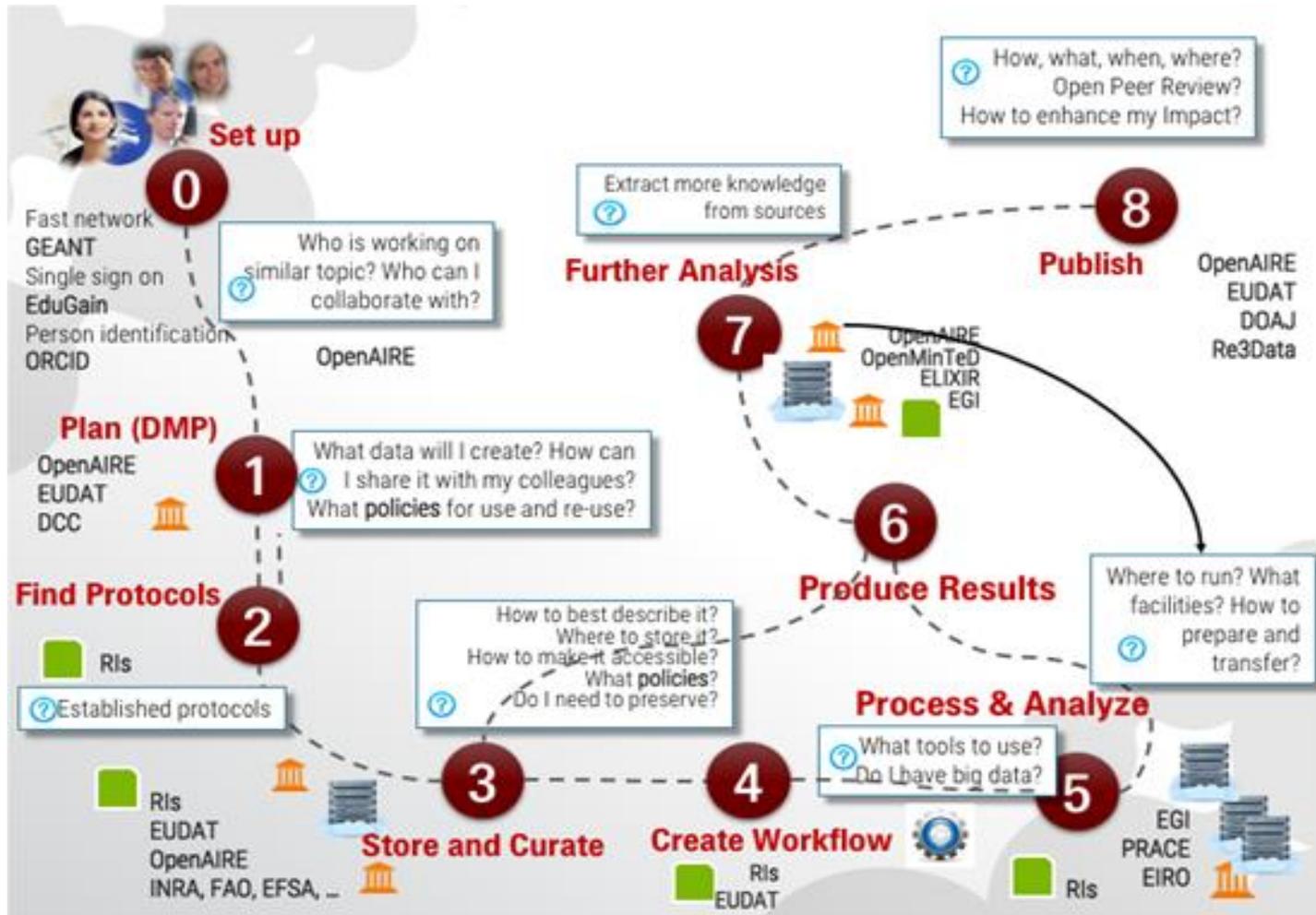


Reilly, S., Schallier, W., Schrimpf, S., Smit, E. & Wilkinson, W. (2011). Report on Integration of Data and Publications.

Available from <http://www.stm-assoc.org/integration-of-data-and-publications/>

Context

A call for legal interoperability of data and for (a more) open science



Concept for a European Open Science Cloud for Research

Context

‘FAIR Data’ (Findable, Accessible, Interoperable and Reusable)

www.nature.com/scientificdata

SCIENTIFIC DATA

OPEN

SUBJECT CATEGORIES

» Research data

» Publication

characteristics

Comment: The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson *et al.*[#]

Box 2 | The FAIR Guiding Principles

To be Findable:

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
 - A1.1 the protocol is open, free, and universally implementable
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available

To be Interoperable:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta)data

To be Reusable:

- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
 - R1.1. (meta)data are released with a clear and accessible data usage license
 - R1.2. (meta)data are associated with detailed provenance
 - R1.3. (meta)data meet domain-relevant community standards

Examples of crystallography in the life sciences and urgent societal problems demand an open science approach



The screenshot shows a web browser window displaying the IUCr website. The address bar shows the URL www.iucr.org/news/research-news/protein-secrets-of-ebola-virus. The page features the IUCr logo and navigation menus. The main content area is titled "Protein secrets of Ebola virus" and includes a text paragraph, a photograph of medical staff in protective suits, and a 3D ribbon diagram of a protein structure with labels for various domains and helices.

Protein secrets of Ebola virus

The current Ebola virus outbreak in West Africa, which has claimed more than 2000 lives, has highlighted the need for a deeper understanding of the molecular biology of the virus that could be critical in the development of vaccines or antiviral drugs to treat or prevent Ebola hemorrhagic fever. Now, a team at the University of Virginia (UVA), USA – under the leadership of Dr Dan Engel, a virologist, and Dr Zygmunt Derewenda, a structural biologist – has obtained the crystal structure of a key protein involved in Ebola virus replication, the C-terminal domain of the *Cryst.* D70, 2420-2429;

Cryst. D70, 2420-2429;

The International Union of Crystallography is a non-profit scientific union serving the world-wide interests of crystallographers and other scientists employing crystallographic methods.

home | advanced search | contact us | site index | journals
© International Union of Crystallography

Based on IUCr website 11 September 2014

Benefits of retaining derived data

- *Scientific record*
- *Database-driven discovery*
- *Protein-ligand interactions*
- *New pathways to synthesis, manufacturing, energetics...*
- *Identification/indexing (e.g. forensic science)*



www.iucr.org/resources/data

International Union of
CRYSTALLOGRAPHY

IUCR Journals | International Tables | World Directory

lucr journals books news education people resources lycr2014

world directory | other directories | data | cif | lists | blogs | forums | commissions | nexus | symmetry font

Home > resources > data

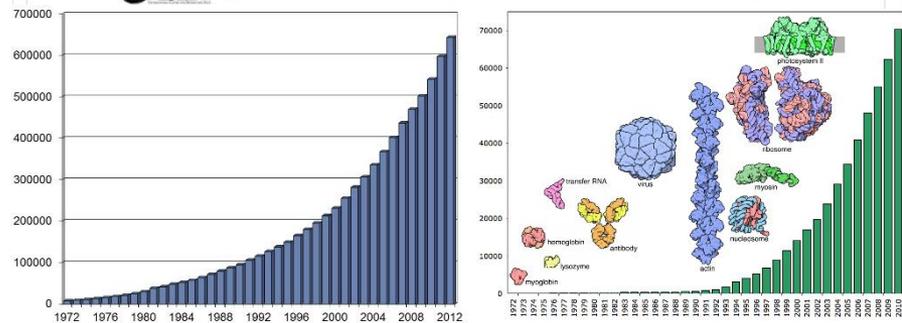
Data activities in crystallography

Databases

Primary crystallographic databases

These are the major public databases of crystal structure and related data. They are generally maintained by large organisations and are valuable resources for the benefit of science as a whole.

-  BCS: Bilbao Crystallographic Server of crystallographic symmetry information
-  BMCD: Biological Macromolecule Crystallization Database
-  CRYSTMET: Metal and intermetallic structures
-  CSD: Cambridge Structural Database of organic and metal-organic structures
-  ICSD: Inorganic Structural Database
ICSD: Web interface to Inorganic Structural Database: ICSD-for-Web
-  NDB: Nucleic Acid Database
-  The Pauling File
-  PDB: Protein Data Bank
-  ICDD: PDF: Powder Diffraction File of the International Centre for Diffraction Data



Benefits of retaining processed data

- *Structure validation*
- *Re-refinement*
- *Systematic bias, methods development*
- *Guard against structures associated with incorrect data sets*
- *Help guard against 'bad apples' in the databases (Minor et al. 2016)*



Article 2

Continuous improvement of macromolecular crystal structures

Thomas C. Terwilliger

Summary

Accurate crystal structures of macromolecules are of high importance in biological and biomedical fields. Models of crystal structures in the Protein Data Bank (PDB) are in general of very high quality, but methods for modeling protein structures and for determination of structures are still improving. We suggest that it is both desirable and feasible to carry out small and large-scale efforts to continuously further improve the models deposited in the PDB. Small-scale efforts could focus on optimizing structures that are of interest to specific investigators. Large-scale efforts could focus on systematic optimization of all structures in the PDB, on redetermination of groups of related structures, or on redetermination of groups of structures focusing on specific questions. All the resulting structures could be made generally available, with various views of the structures available depending on the types of questions that users are interested in answering.

1. Introduction

1.1 Crystal structures of macromolecules

The three-dimensional structures of biological macromolecules such as proteins, DNA and RNA are of high importance in many areas of biology and biotechnology. Structures of proteins and of complexes between proteins, between proteins and small molecules, and between proteins and nucleic acids are all crucial for understanding how these molecules function to catalyze chemical reactions and to control metabolism, growth and development. Structures of proteins bound to candidate drug molecules are highly useful in the development of new pharmaceuticals. Structures of natural and engineered proteins are crucial for rational engineering of these molecules to give them new functions or altered properties.



IUCr Diffraction Data Deposition Working Group (DDDWG) report to the IUCr General Assembly Montreal

John R Helliwell on behalf of the
DDDWG



Members of the DDDWG 2011 to 2017

- *John R Helliwell and Brian McMahon (UK),
Chair and Co-Chair;*
- *Steve Androulakis (Australia)*
- *Sol Gruner (USA)/Dolothia Szebenyi (USA)*
- *Loes Kroon-Batenburg (Netherlands)*
- *Tom Terwilliger (USA)*
- *John Westbrook (USA)*
- *Heinz-Josef Weyer (Switzerland) †*

Recommendations from the DDDWG for the upcoming Triennium

- IUCr Commissions to define their metadata;
- *J. Appl. Cryst.* to introduce a 'Difficult Raw Data' Section (Loes Kroon-Batenburg);
- A centralised crystallographic repository of raw data set metadata should be scoped, inc a search interface, leading to a pilot service;
- With a viable pilot metadata registry **authors should** provide a permanent and prominent link from an article to their raw data sets underpinning a journal publication.

Issues for the IUCr

- The IUCr's science involves 'Big data' up towards the level of the data-deluge of the Square Kilometre Array radio telescope; we may have to consider subsets of data retention or limited time periods for retention;
- Rights of access to publicly funded, but unpublished, crystallographic research data after *e.g.* 3 to 5 years.

Charge to the ECM29 Rovinj IUCr Workshop Participants

- *Define your Metadata or at the least Define the Challenges you face*
- *We have provided a **template form** for you to supply information about metadata for your specific research field / IUCr Commission*
- *We consider the challenge of aligning scientific metadata with generic standards like the **'Dublin Core of Metadata descriptors'***
- *We have to understand each other's fields; this will assist the philosophical challenge of **'seeing a data set through a new-user's eyes'***

Complete video record of Rovinj Workshop: <http://tinyurl.com/diffraction-metadata>

Session I: Introduction

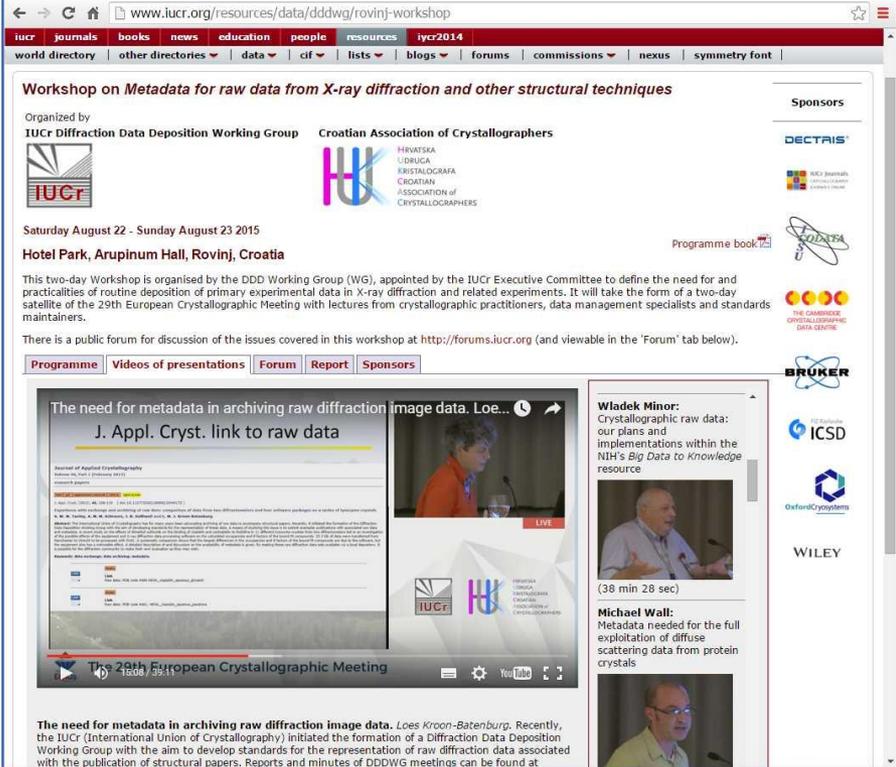
Session II: Diffraction images - what can we get out?

Session III: Metadata for diffraction images and other experimental methods

Session IV: Data in the Wider World - From Laboratory to Database

Session V: What new metadata items are needed?

Session VI: Metadata schemas



The screenshot shows a web browser window with the URL www.iucr.org/resources/data/dddwg/rovinj-workshop. The page title is "Workshop on Metadata for raw data from X-ray diffraction and other structural techniques". It is organized by the IUCr Diffraction Data Deposition Working Group and the Croatian Association of Crystallographers. The workshop took place on Saturday August 22 - Sunday August 23, 2015, at Hotel Park, Arupinum Hall, Rovinj, Croatia. The page includes a navigation menu with options like "Programme", "Videos of presentations", "Forum", "Report", and "Sponsors". A video player is embedded, showing a presentation titled "The need for metadata in archiving raw diffraction image data. Loe... J. Appl. Cryst. link to raw data". The video player shows a speaker at a podium. To the right of the video player, there is a list of sponsors including DECTRIS, IUCr journals, HRVATSKA UDRUGA KRIŠTALOGRAFA, CROATIAN ASSOCIATION OF CRYSTALLOGRAPHERS, BRUKER, ICSD, Oxford Cryosystems, and WILEY. Below the video player, there is a text block that reads: "The need for metadata in archiving raw diffraction image data. Loes Kroon-Batenburg. Recently, the IUCr (International Union of Crystallography) initiated the formation of a Diffraction Data Deposition Working Group with the aim to develop standards for the representation of raw diffraction data associated with the publication of structural papers. Reports and minutes of DDDWG meetings can be found at".

Significant pioneering developments

- Australian synchrotron MX raw data archive
https://store.synchrotron.org.au/public_data/
- The USA NIH funded various structural genomics projects with raw data archives
- The Institut Laue Langevin and ISIS are exemplar at preserving all data and with dois

eg



The screenshot displays the ISIS Data Journal interface. At the top, there are logos for the Science & Technology Facilities Council and the ISIS Data Journal, which is described as 'The archive for ISIS research data'. Below the header, a navigation bar contains 'ISIS' and 'ISIS Data' links. The main content area features the dataset ID 'RB920486' and the following details:

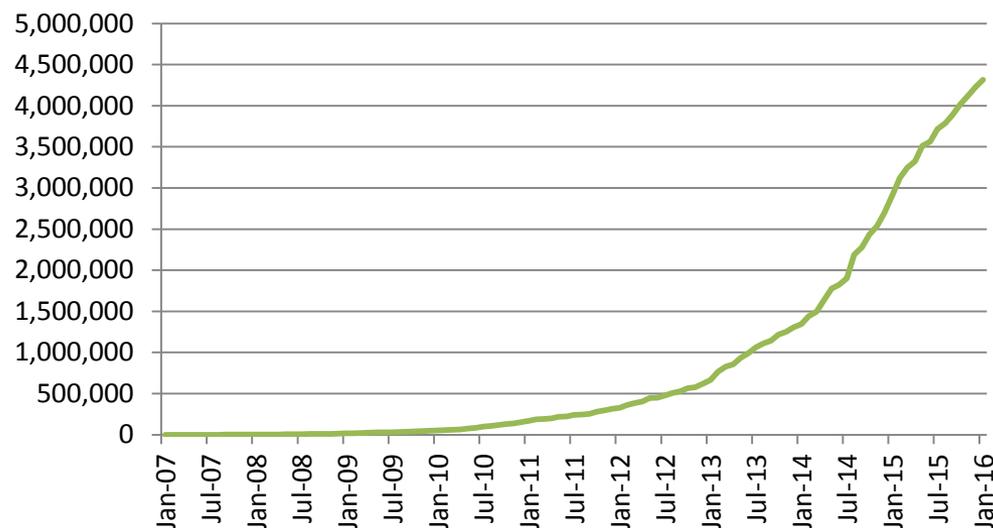
- Investigation title:** Electric field effect on the interfacial uncompensated spins in the Co/BiFeO₃/STO exchange bias system.
- Release date:** Fri Jul 26 09:06:29 BST 2013
- Creator:** Dr Nina-Juliane Steinke
- DOI:** 10.5286/ISIS.E.24079627
- Date of Experiment:** Fri Jul 23 08:52:43 BST 2010
- Publisher:** STFC ISIS Facility
- Data format:** RAW/Nexus

A note indicates: 'Select the data format above to find out more about it.' Below this, a 'Data Citation' section provides the recommended format for citing the dataset in a research publication: [author], [date], [title], [publisher], [doi].

On the right side of the page, there is a 'DOWNLOAD' button with a download icon and the text 'download the dataset'. Below the button is a small photograph of a laboratory setting with the caption 'Data collected CRISP line at the ISIS'.

A pioneer SR facility retaining all data since start up

Total Data On Diamond Archive (Gb)



diamond.ac.uk/Users/UserGuide/Data-User-Guide/Accessing-Data/Data-Policy.html#Owns

diamond About Us For Users Industry Public Science Beamlines Careers

For Users

Users User Guide Applying for Beamtime Diamond Users Committee FAQ New User Administration

Users / User Guide / Your Data: User Guide / Accessing Your Data / Experimental Data Management Policy

In This Section Experimental Data Management Policy

Users conducting Peer Reviewed Research will own the Experimental Data that they produce.

“...Following the initial 30 day storage period, Diamond will create a single archive copy of the Experimental Data on tape.”

“...Users of Diamond Facilities are responsible for meeting any third-party data management obligations that may be applicable.”

Recent developments

- University data repositories with doi registrations.
- EC's Zenodo science data archive (free of charge!)
- The University of Virginia BD2K for MX, led by Wladek Minor <http://www.proteindiffraction.org/> (USA)
- The PDB now requests the information on raw data and metadata for raw data during a deposition ie their dois
- ESRF Data Archive (“every raw data set measured with a registered doi”)
- IUCrData (initially, *derived* data sets)
- The Structural Biology Data Grid has been launched (Nature Comms Meyer et al 2016)

Search resources

Academic support

Our services

The John Rylands Library

Special Collections

About Us

My account

▲ The University of Manchester Library

▲ Search resources

Manchester eScholar

▶ Search

▶ Browse by

▶ Help

▶ About

▶ Contact us

FEEDBACK AND ENQUIRIES



4XAN_Carboplatin_NaBr_Diamond_I04

Tanley, Simon

[Experimental data]. 2014. The University of Manchester.

Access to files

[carboplatin_NaBr_diamond_1_1-180.ZIP](#) (x-zip-compressed)

[carboplatin_NaBr_diamond_1_181-360.ZIP](#) (x-zip-compressed)

Abstract

Carboplatin is a second-generation platinum anticancer agent used for the treatment of a variety of cancers. Previous X-ray crystallographic studies of carboplatin binding to histidine (in hen egg-white lysozyme; HEWL) showed the partial conversion of carboplatin to cisplatin owing to the high NaCl concentration used in the crystallization conditions. HEWL co-crystallizations with carboplatin in NaBr conditions have now been carried out to confirm whether carboplatin converts to the bromine form and whether this takes place in a similar way to the partial conversion of carboplatin to cisplatin observed previously in NaCl conditions. Here, it is reported that a partial chemical transformation takes place but to a transplatin form. Thus, to attempt to resolve purely carboplatin binding at histidine, this study utilized co-crystallization of HEWL with carboplatin without NaCl to eliminate the partial chemical conversion of carboplatin. Tetragonal HEWL crystals co-crystallized with carboplatin were successfully obtained in four different conditions, each at a different pH value. The structural results obtained show carboplatin bound to either one or both of the N atoms of His15 of HEWL, and this particular variation was dependent on the concentration of anions in the crystallization mixture and the elapsed time, as well as the pH used. The structural details of the bound carboplatin molecule also differed between them. Overall, the most detailed crystal structure showed the majority of the carboplatin atoms bound to the platinum centre; however, the four-carbon

Related resources

Full-text held externally

DOI: 10.15127/1.266906

<http://scripts.iucr.org/cgi-bin/paper?S2053230X16000777>

University researcher(s)

Academic department(s)

[Faculty of Life Sciences' website](#)

[Search resources](#)

[Academic support](#)

[Our services](#)

[The John Rylands Library](#)

[Special Collections](#)

[About Us](#)

[My account](#)

▲ [The University of Manchester Library](#)

▲ [Search resources](#)

Manchester eScholar

▶ [Search](#)

▶ [Browse by](#)

▶ [Help](#)

▶ [About](#)

▶ [Contact us](#)

FEEDBACK AND
ENQUIRIES



HEWL_cisplatin_5percentDMSO_RT: 4g4a

Tanley, Simon

[Experimental data] version online. 2012. The University of Manchester.

Access to files

[HEWL_cisplatin_5percentDMSO_RT_01_0001to01_0347.zip](#) (zip)

[HEWL_cisplatin_5percentDMSO_RT_02_0001to02_0200.zip](#) (zip)

[HEWL_cisplatin_5percentDMSO_RT_03_0001to03_0235.zip](#) (zip)

[HEWL_cisplatin_5percentDMSO_RT_04_0001to04_0303.zip](#) (zip)

[HEWL_cisplatin_5percentDMSO_RT_05_0001to05_0314.zip](#) (zip)

[HEWL_cisplatin_5percentDMSO_RT_06_0001to06_0348.zip](#) (zip)

[HEWL_cisplatin_5percentDMSO_RT_07_0001to07_0200.zip](#) (zip)

[HEWL_cisplatin_5percentDMSO_RT_07_0201to07_0377.zip](#) (zip)

[HEWL_cisplatin_5percentDMSO_RT_08_0001to08_0300.zip](#) (zip)

[HEWL_cisplatin_5percentDMSO_RT_08_0301to08_0584.zip](#) (zip)

[HEWL_cisplatin_5percentDMSO_RT_09_0001to09_0310.zip](#) (zip)

[HEWL_cisplatin_5percentDMSO_RT_09_0311to09_0658.zip](#) (zip)

Abstract

Abstract: The anticancer complexes cisplatin and carboplatin are known to bind to both the N and the N atoms of His15 of hen egg-white lysozyme (HEWL) in the presence of dimethyl sulfoxide (DMSO). However, neither binds in aqueous media after 4 d of crystallization and crystal growth, suggesting that DMSO facilitates cisplatin/carboplatin binding to the N atoms of His15 by an unknown mechanism. Crystals of HEWL cocrystallized with cisplatin in both aqueous and DMSO media, of HEWL cocrystallized with carboplatin in DMSO medium and of HEWL cocrystallized with cisplatin and N-acetylglucosamine (NAG) in DMSO medium were stored for between seven and 15 months. X-ray diffraction studies of these crystals were carried out on a Bruker APEX II home-source diffractometer at room temperature. Room-temperature X-ray diffraction data

Related resources

Full-text held externally

DOI: [10.15127/1.215887](#)

DOI: [doi:10.1107/S1744309112042005](#)

DOI: [doi:10.1107/S1744309112042005](#)

University researcher(s)

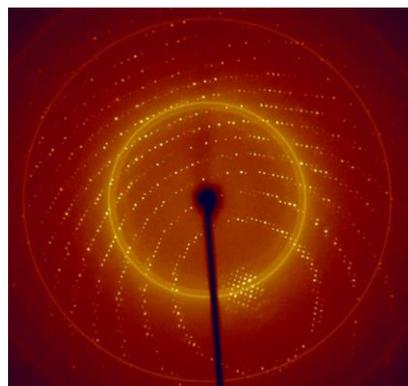
Academic department(s)

Faculty of Engineering and Physical Sciences' website

Faculty of Life Sciences' website

School of Chemistry 's website

Each entry also includes a variety of metadata



*The Interdisciplinary Centre for Mathematical and Computational Modelling
(ICM) University of Warsaw*

- In 2015 initiated archiving of raw diffraction images with assigned DOI numbers in their RepOD (<https://repod.pon.edu.pl/>) open science repository;*
- At present there are just a few deposits, but macromolecular crystallographers in Poland are encouraged to archive their raw data there.*



Mariusz Jaskolski personal communication

wwPDB Deposition & Annotation System

Identifying Primary Data



wwPDB Deposition: D_8000200025 -- Requested ID: PDB

FAQ

Tutorial

Welcome to the Worldwide Protein Data Bank

Navigation

- ✓ Instructions
- ✓ Communication
- ✓ Re-upload files
- ✓ Upload summary
- Admin
 - ✓ Contact information
 - ✓ Grant information
 - ✓ Release status
 - ✓ Entry title & author
 - ✓ Citation information
- Macromolecules
 - ✓ 1) Interleukin-1 beta
- Data collection
 - ✓ Crystal Information
 - ✓ Collection Source
 - ✓ Software Used
 - ✓ Collection Statistics
- Refinement
 - ✓ Refinement
 - ✓ Ligands
 - ✓ Biological assembly
 - Validation reports
 - Summary & conditions

Log out

Database ID:

Database name:

Details:



Related external experimental data sets

DOI for the related experimental data set:

DOI for additional metadata describing the related data set:

The type of experimental data:



DOI for data set

DOI for descriptive metadata

Continue to next section

Big Research needs Big Machines leads to Big Data



World leading

*ESRF premiere
X-ray source*

*Institut Laue Langevin
Neutron source*

The European Synchrotron Radiation Facility (ESRF) in Grenoble, in which the UK has a 14% share, and the Institut Laue Langevin nuclear reactor is to the right (UK share 25-33%)

ESRF DATA POLICY SINCE LAST PROPOSAL ROUND

- **ESRF Council officially adopted a Data Policy (1/12/2015)**
- **ESRF is custodian of data and metadata**
- **ESRF to collect high quality metadata to facilitate reuse of data**
- **ESRF will keep raw (or reduced) data for 10 years + metadata for ever**

- **Data will be registered in a data catalogue (icat) + published with a Digital Object Identifier (DOI)**
- **Principal investigators have exclusive access to data during the embargo period (3 years but can be extended)**
- **Data will be made public after the embargo period under CC-BY**
- **Data Policy will be implemented on all beamlines by 2020**

<http://www.esrf.eu/home/UsersAndScience/UserGuide/esrf-data-policy-implementation.html>

Editor, referee and reader/user

- In the early 1990s *Acta Cryst. C* pioneered refereeing of articles as well as their structure factors and coordinates; any one or all of these could be revised!
- Databases can then harvest the fruits of that peer review thoroughness;
- In 2002 JRH as IUCr EinC proposed at IUCr Geneva that the *Acta Cryst. C* method should be harnessed for *Acta Cryst. D* submissions (sadly, was rejected);
- Today I still commend the *Acta Cryst. C* method to the benefit of journals, databases, authors and readers;
- ***Now we also have the technology and organisation to have access to the raw diffraction data as well.***

IUCr Journals

Provides file upload options that include:- the .doc word file, the PDB validation reports, the PDB files and e.g. mtz processed diffraction data files

The article's words → ~~xxxxxx~~ Word file

PDB validation report → PDB validation report* ⓘ

PDB file → mmcif or PDB file ⓘ

Diffraction data file → other supporting information ⓘ

Upload your source files for this submission ⓘ

1. Select the role of the file, its number and part number as appropriate
2. Locate the file on your local system using the 'Browse' button
3. Click on the 'Upload file' button

The article text can also obviously provide the dois for the raw diffraction images if the authors wish to

IUCrData

- *Data publishing platform launched 2016*
- *Addresses immediate need for concise reporting of peer-reviewed structural model data sets*
- *Intends to develop new synergies between journal publications and data sets*

The screenshot shows the IUCrData website interface. At the top, there is a navigation bar with links for home, archive, editors, for authors, for readers, submit, and open access. The main content area is divided into several sections:

- Issue contents:** Volume 1 | Part 3
- March 2016 issue:** Early view articles. A large image of a crystal structure is displayed.
- metal-organic compounds:** A list of articles with their titles and authors. The first article is titled "[μ -N,N,N',N'-Tetrakis(pyridin-2-ylmethyl)butane-1,4-diamine]bis-[(dimethanol- κ O)(perchlorato- κ O)copper(II)] bis(perchlorate)" by X.-H. Zhu, P. Li, X.-W. Chen, W.-S. Ke, F. Chen and H.-X. Zhang. The article includes a small image of the crystal structure, the CCDC reference (1439047), and a "Read article" button.
- publCIF:** A section for the publCIF software, described as "free software to edit and preview a CIF for publication".
- IUCr outreach activities:** A section listing various outreach activities, including "Crystallography in Africa initiative", "Worldwide crystal-growing competition", and "IUCr sponsorship schemes".
- Information on IUCr Journals:** A section providing information on how to find more information on any of the journals by clicking on a journal icon.
- Meetings:** A section listing upcoming meetings, such as the "National Workshop on Theory and".

In conclusion

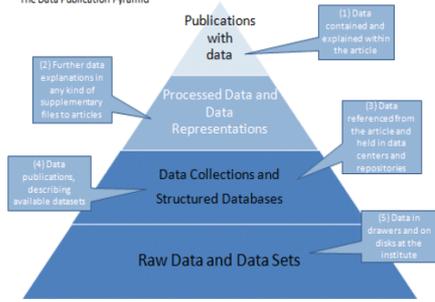
- New modus operandi for *published funded research*; access to all data and should be open access.
- A *limited time for funded researchers to analyse their data and publish*; typically 3 years. Then the raw data are put on open access.
- Policy makers are now discussing new ways to '*speed up science and discovery for tax payers to reap quicker benefits*'; *Open Science*. New rules of conduct for funded research would be essential!

- Next actions of the IUCr DDDWG; articles are in preparation from ECM29 Rovinj Workshop; with IUCr COMCIFS a checkcif for raw data; planned workshop on *metadata for raw data at* ACA New Orleans; sessions at IUCr Hyderabad.....

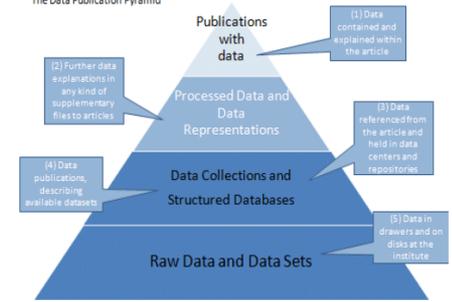
- Current actions of JRH as a researcher; inviting other *platinis with proteins* and nucleic acids researchers to contribute to the *fully open access raw data, SFs, coordinates and publications*.....

***Join in with the raw diffraction data
revolution***

The Data Publication Pyramid

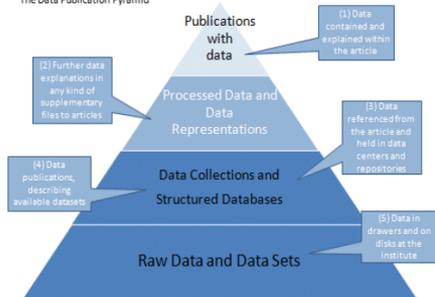


The Data Publication Pyramid



Thank you

The Data Publication Pyramid



The Data Publication Pyramid

