



# *The Future of Structural Science Session at BCA 2016 Nottingham, UK*

*Access to raw diffraction data; current practice in  
article linking to raw diffraction data*

John R Helliwell<sup>1,\*</sup> and Brian McMahon<sup>2</sup>  
School of Chemistry, University of Manchester, M13 9PL, UK  
IUCr 5 Abbey Square, Chester CH1 2HU, UK

[john.helliwell@manchester.ac.uk](mailto:john.helliwell@manchester.ac.uk)



*The arrow of time...the future*

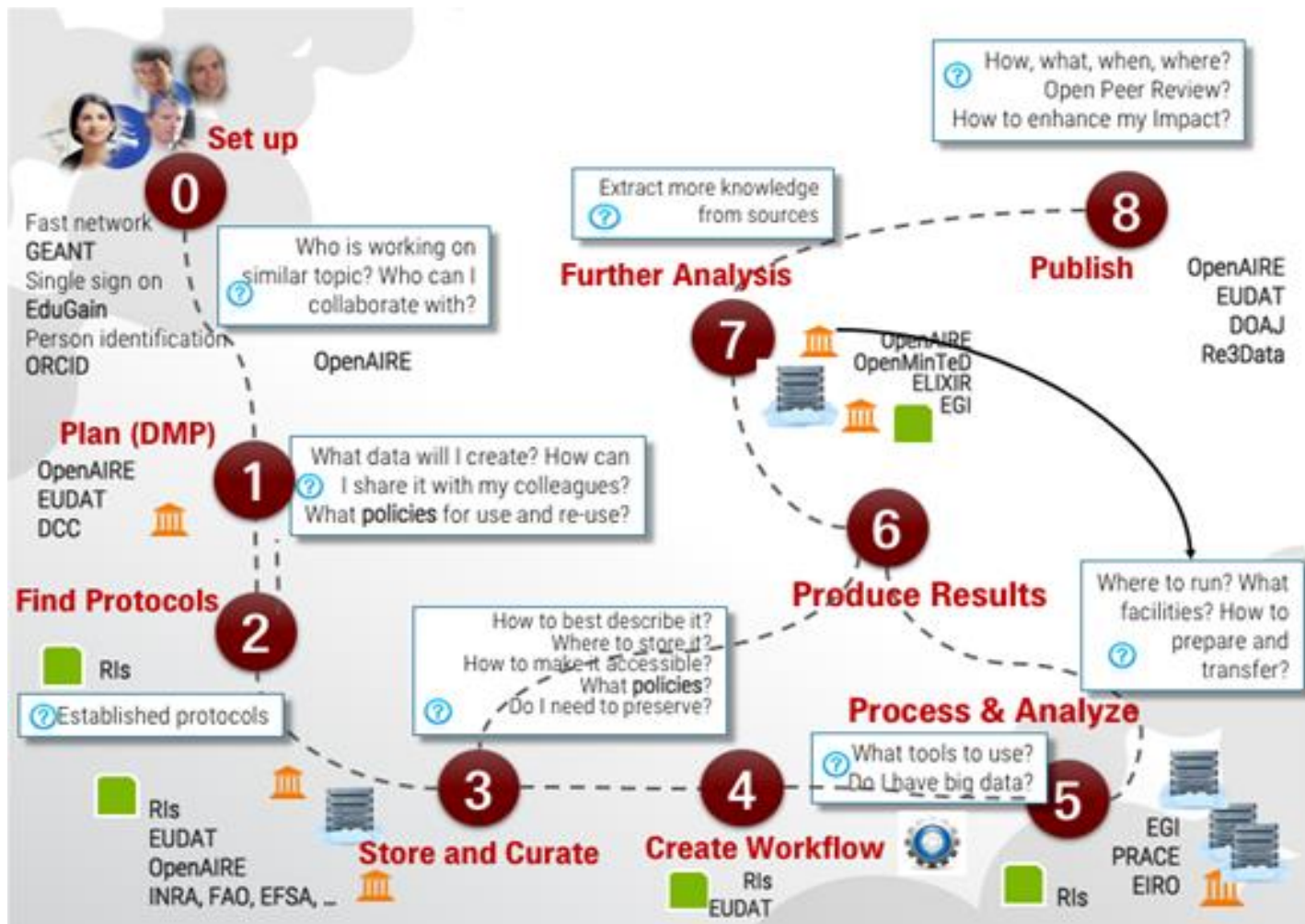
# Overview

## Towards a future structural science based on raw data archiving

- *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.

# Context

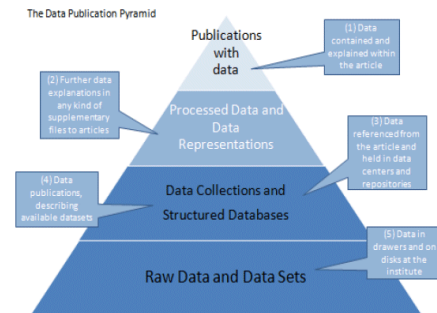
A call for legal interoperability of data and for (a more) open science



*Concept for a European Open Science Cloud for Research*

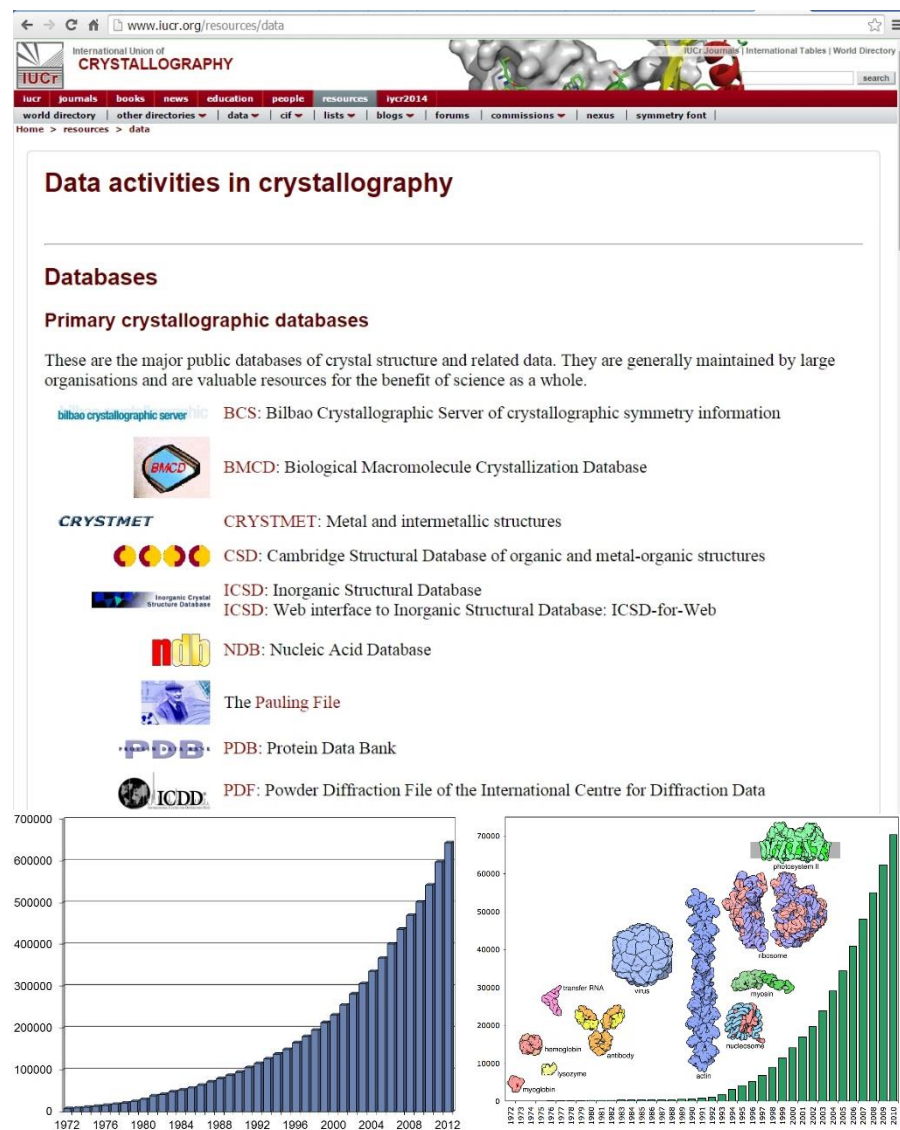
# *Raw diffraction images offer the opportunity of*

- 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*



# 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)*





# 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)*

The screenshot shows the Acta Crystallographica Section B website. The header includes the journal title, ISSN (2052-5206), and navigation links. The main content area displays the article title "Continuous improvement of macromolecular crystal structures" by Thomas C. Terwilliger. The article is categorized as "Article 2". The summary text discusses the importance of accurate crystal structures in biological and biomedical fields, mentioning the Protein Data Bank (PDB) and the need for continuous improvement. The article is divided into sections, with "1. Introduction" and "1.1 Crystal structures of macromolecules" visible.

Acta Crystallographica Section B  
STRUCTURAL SCIENCE, CRYSTAL ENGINEERING AND MATERIALS

home archive editors for authors for readers submit subscribe open access

research papers

Volume 58 | Part 1 | February 2002 | Pages 62-77  
doi:10.1107/S0108768101017128

Some 60 new space-group corrections

ICSEI Insights

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



# 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.

# 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 the IUCr website for the Rovinj Workshop. The page is titled "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 is scheduled for Saturday August 22 - Sunday August 23 2015 at Hotel Park, Arupinum Hall, Rovinj, Croatia. The page includes a list of sponsors such as Dectris, IUCr Journals, and the European Crystallographic Meeting. A video player is embedded on the page, showing a presentation titled "The need for metadata in archiving raw diffraction image data. J. Appl. Cryst. link to raw data". The video player shows a speaker, J. Appl. Cryst. link to raw data, and a video player interface with a progress bar and controls. The video is titled "The 29th European Crystallographic Meeting".

www.iucr.org/resources/data/dddwg/rovinj-workshop

**Workshop on Metadata for raw data from X-ray diffraction and other structural techniques**

Organized by  
IUCr Diffraction Data Deposition Working Group  
Croatian Association of Crystallographers

**Saturday August 22 - Sunday August 23 2015**  
**Hotel Park, Arupinum Hall, Rovinj, Croatia**

This two-day Workshop is organised by the DDD Working Group (WG), appointed by the IUCr Executive Committee to define the need for and practicalities of routine deposition of primary experimental data in X-ray diffraction and related experiments. It will take the form of a two-day satellite of the 29th European Crystallographic Meeting with lectures from crystallographic practitioners, data management specialists and standards maintainers.

There is a public forum for discussion of the issues covered in this workshop at <http://forums.iucr.org> (and viewable in the 'Forum' tab below).

**Programme** | **Videos of presentations** | **Forum** | **Report** | **Sponsors**

**The need for metadata in archiving raw diffraction image data. J. Appl. Cryst. link to raw data**

**Wladek Minor:**  
Crystallographic raw data: our plans and implementations within the NIH's Big Data to Knowledge resource

**Michael Wall:**  
Metadata needed for the full exploitation of diffuse scattering data from protein crystals

**The 29th European Crystallographic Meeting**

**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/](https://store.synchrotron.org.au/public_data/)
- The USA NIH funded various structural genomics projects with raw data archives
- ISIS is an exemplar at preserving all data and with dois eg

 Science & Technology  
Facilities Council

ISIS Data Journal  
The archive for ISIS research data

ISIS	ISIS Data
------	-----------

RB920486

**Investigation title:** Electric field effect on the interfacial uncompensated spins in the Co/BiFeO<sub>3</sub>/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  
Select the data format above to find out more about it.

**Data Citation**  
The recommended format for citing this dataset in a research publication is as:  
[author], [date], [title], [publisher], [doi]

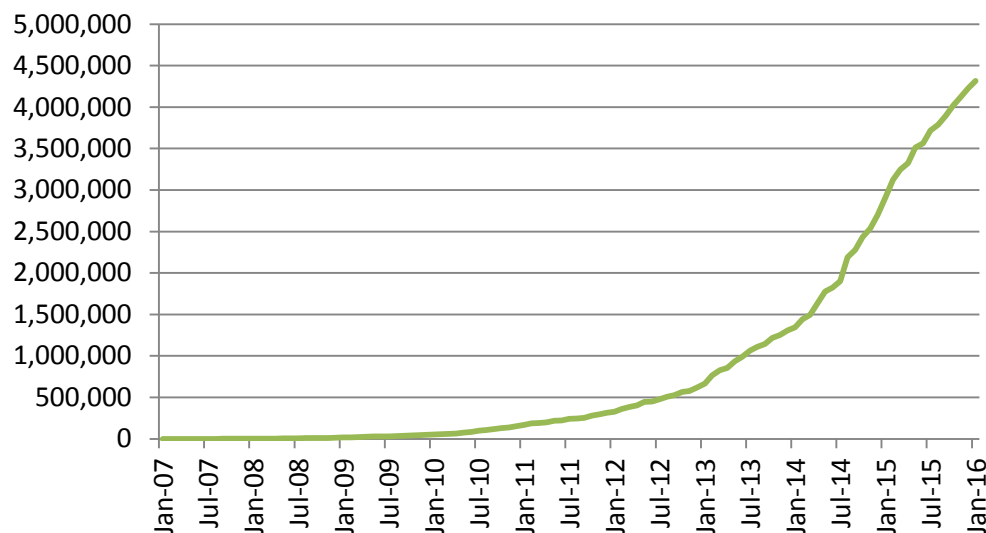
 DOWNLOAD  
download  
the dataset




Data collected  
CRISP ins  
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



## 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](#)

[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

# 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:10.000/100/da.dat

DOI for additional metadata describing the related data set:



doi:10.000/100/md.cif

The type of experimental data:



diffraction image data



DOI for data set

DOI for descriptive metadata

Continue to next section

# 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.*



# IUCrData

The screenshot displays the IUCrData website interface. At the top, the URL is [iucrdata.iucr.org/x/issues/2016/03/00/](http://iucrdata.iucr.org/x/issues/2016/03/00/). The page features a navigation bar with links: home, archive, editors, for authors, for readers, submit, and open access. The main content area is titled "issue contents Volume 1 | Part 3" and shows the "March 2016 issue" with ISSN: 2414-3146. A featured article is highlighted with a molecular structure image: "[ $\mu$ -N,N,N',N'-Tetrakis(pyridin-2-ylmethyl)butane-1,4-diamine]bis-[(dimethanol- $\kappa$ O)(perchlorato- $\kappa$ 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 is available for open access. The right sidebar contains sections for "publCIF" (free software to edit and preview a CIF for publication), "IUCr outreach activities" (besides publishing top-quality research, the IUCr is committed to a series of outreach activities), "IUCr Legacy", "Crystallography in Africa initiative", "Worldwide crystal-growing competition", "IUCr sponsorship schemes", "Information on IUCr Journals" (to find out more information on any of our journals, click on a journal icon and follow the links for authors), and "Meetings" (National Workshop on Theory and

- 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

# 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 *platins with proteins* and nucleic acids researchers to contribute to the *fully open access raw data, SFs, coordinates and publications*.....

# 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) †*

*Join in with the raw diffraction data revolution*



*Thankyou*