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

HRVATSKA
UDRUGA
KRISTALOGRAFA
CROATIAN
ASSOCIATION of
CRYSTALLOGRAPHERS

Sponsors

DECTRIS®

IUCr Journals
CRYSTALLOGRAPHY
JOURNALS ONLINE

THE CAMBRIDGE
CRYSTALLOGRAPHIC
DATA CENTRE

ICSU CODATA

BRUKER

FIZ Karlsruhe
ICSD

OxfordCryosystems

WILEY

Introduction, Welcome and Update on behalf of the IUCr Diffraction Data Deposition Working Group (DDDWG)

from

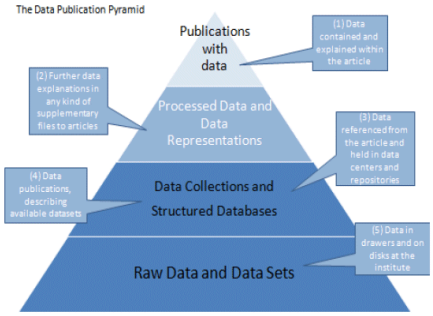
John R Helliwell & Brian McMahon

A Hearty Big Thankyou to Our Sponsors
&
To the Croatian Association of Crystallographers
&
All Our Speakers from Around the World!

Contents

- *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*
- *Brief recap of 2011 to 2014; our DDDWG Report to the IUCr Montreal 2014 General Assembly*
- *The challenge of the sheer volume of our raw diffraction data*
- *The challenge of achieving a new-user depth to raw diffraction data metadata description i.e. can a new-user successfully understand and process any raw data set he/she wishes to analyse?*

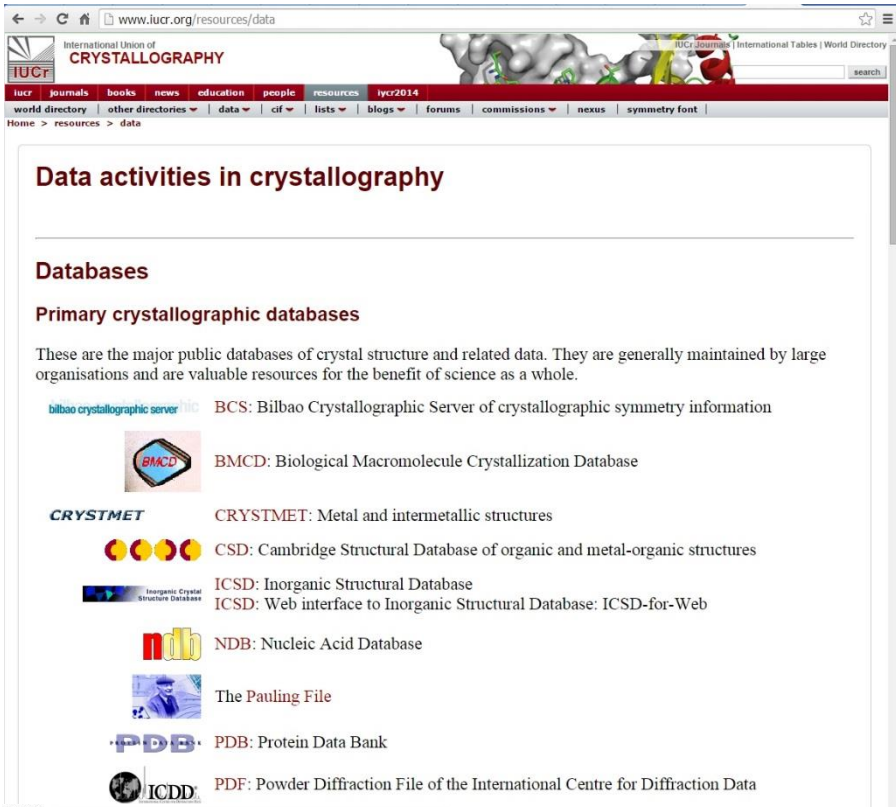
Raw diffraction images offer the opportunity of

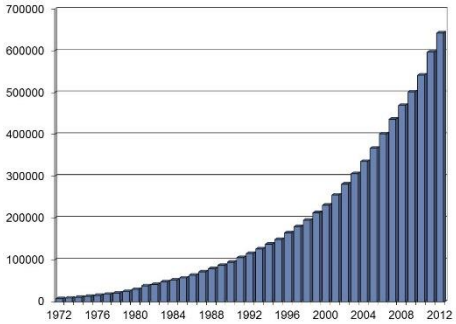


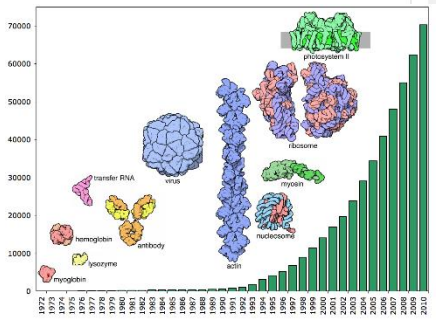
- *analysing data at higher resolution than used in the original work*
- *serving as benchmarks in developing improved methods of analysis*
- *checking the interpretation of the symmetries of the crystals*
- *analysing diffraction from multiple lattices present in the crystals*
- *analysing the diffuse scattering that reflects correlated motions or disorder of atoms in the crystals*

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*

journals.iucr.org/b/issues/2002/01/00/bm0047/index.html

Structural Science
Crystal Engineering
and Materials

contents of issue
search
subscribe
help

navigate
pdf
supplementary files
next in issue

research papers

SHARE highlight terms

Structural Science

Volume 58
Part 1
Pages 62-77
February 2002

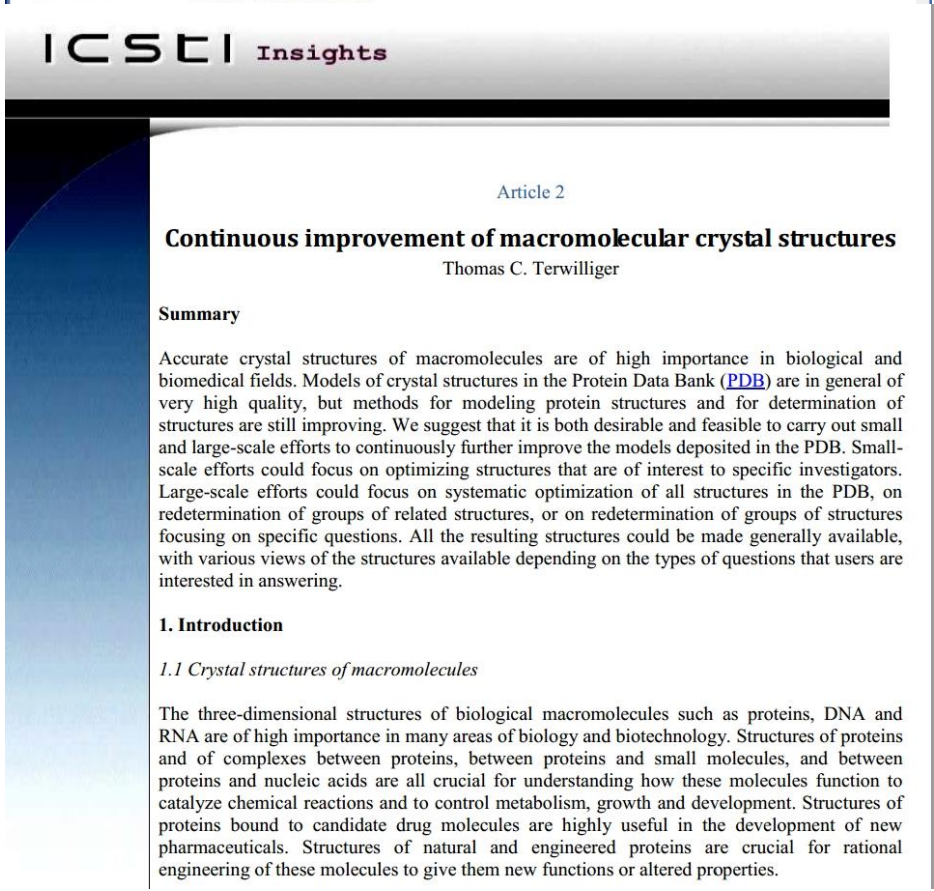
Received 2 June 2001
Accepted 12 October 2001

Some 60 new space-group corrections

Richard E. Marsh,[a] Moshe Kapon,[b] Shengzhi Hu[c] and Frank H. Herbstein[b]*

[a]The Beckman Institute, California Institute of Technology, Pasadena, CA 91125, USA, [b]Department of Chemistry, Technion-Israel Institute of Technology, Haifa 32000, Israel, and [c]Department of Chemistry, Xiamen University, Xiamen, Peoples' Republic of China
Correspondence e-mail: chr03fk@tx.technion.ac.il

CrossMark



ICSTI 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.

CONGRÈS ET ASSEMBLÉE GÉNÉRALE
DE L'UNION INTERNATIONALE DE CRISTALLOGRAPHIE

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

John R Helliwell on behalf of the DDDWG

CONGRESS AND GENERAL ASSEMBLY
OF THE INTERNATIONAL UNION OF CRYSTALLOGRAPHY

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, including 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.

Members of the DDDWG 2011 to 2014

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

Members of the DDDWG 2014 to 2017

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

The Cloud to help solve the raw data storage challenge?

nature International weekly journal of science

NATURE | EDITORIAL

Cloud cover

Opposition to storing vast scientific data sets on cloud-computing platforms is weakening.

08 July 2015

Charge to the 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 core challenge of 'seeing a data set through a new-user's eyes'***

Our Workshop Sessions

- *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*

Let’s go to it!