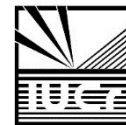


The vital role of Crystallographic Information Files in chemical and biological crystallography to underpin the databases' validation reports

Brian McMahon

International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, UK

Email: bm@iucr.org



The clue is in the name...

“

There is a traditional hierarchy of components of understanding: **data**, **information**, **knowledge**, **wisdom** (the DIKW model).

Crystallographic **information** is the component that bridges the gap between the raw experimental **data** and the global **knowledge** bases represented by the Protein Data Bank, Cambridge Structural Database, International Centre for Diffraction Data, Inorganic Crystal Structure Database, Crystallography Open Database etc. ”

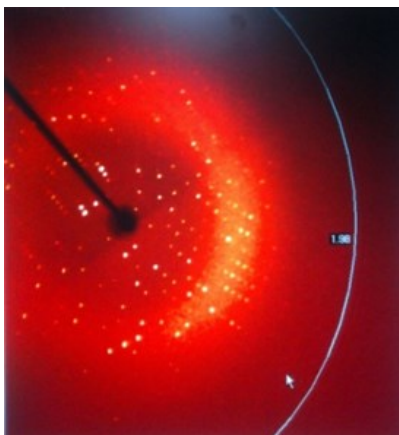
Summary: CIF and the *checkCIF* paradigm

- CIF for small molecules led to the *checkCIF* validation service
- *checkCIF* has a threefold approach:
 - Completeness
 - Correctness
 - Context
- CIF embraces all validation requirements (publication/database)
 - Handles metadata on same footing as data
 - Suitable for data query
 - Extensible

Types of crystallographic data described by CIF

'Raw' data

Numerical data collected directly
from an experimental apparatus



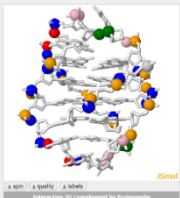
Annotation

Supporting information

Supplementary text, tables and figures. DOI: <https://doi.org/10.1101/52054746318100051/m5151supl.pdf>

Proteopedia molecular tour

DNA is a structurally plastic molecule, and its biological function is enabled by adaptation to its binding partners. We identify the DNA structural polytopisms by analyzing structures of about 1000 DNA-protein complexes. We developed programs that provide detail analysis and the protocol of separating distinct structural (conformational) classes called NIC (for Nucleotide Conformations) that describe the local DNA structure. To further facilitate understanding of the DNA structure, structurally similar NIC classes were grouped into 11 letters of a DNA "structural alphabet" called CANA (Conformational Alphabet of Nucleic Acids). Structural alphabet is a tool allowing to translate information about three-dimensional structure of a biopolymer into a linear sequence of letters of the alphabet. We applied the analysis of complex three dimensional object such as DNA analysis of a sequence of symbols.



Gallery of the CANA letters. Each letter is represented by 42 randomly selected structures of steps which represent one NRC class belonging to the letter in the golden set. A-Bia and mixed B-Bia conformers are drawn in pink and violet, B-like conformers in blue. The CANA letter AAA is represented by NRC 4403, R998 by R903, and R92 by R932.

'Processed' data

Reduced, calibrated, processed
numerical observations

```
h,k,l, Fo-squared, sigma(Fo-squared), sigma(Fo-squared) and status
#
data 6
_shelx_title ' O1SRC413 in F2(1)/n'
_shelx_refln_list_code 4
_shelx_F_calc_maximum 183.83
_exptl_crystal_F_000 1144.00
_reflns_d_resolution_high 0.7705

loop
_symmetry_equiv_pos_as_xyz
'x, y, z'
'-x+1/2, y+1/2, -z+1/2'
'-x, -y, -z'
'x-1/2, -y-1/2, z-1/2'

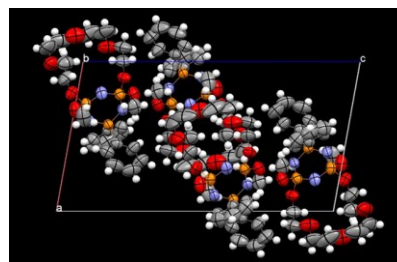
_cell_length_a 11.8293
_cell_length_b 10.3312
_cell_length_c 21.6318
_cell_angle_alpha 90.000
_cell_angle_beta 100.203
_cell_angle_gamma 90.000

_shelx_F_squared_multiplier 1.000

loop
_refln_index_h
_refln_index_k
_refln_index_l
_refln_F_squared_calc
_refln_F_squared_meas
_refln_F_squared_sigma
_refln_observed_status
2 0 0 772.37 856.47 28.20 0
4 0 0 1445.15 1446.80 39.55 0
6 0 0 1130.79 1097.08 30.62 0
8 0 0 1347.13 1490.27 55.41 0
10 0 0 3273.01 3545.44 154.91 0
12 0 0 48.20 40.50 4.56 0
14 0 0 79.87 63.02 7.91 0
2 1 0 2093.70 1975.83 47.36 0
3 1 0 33795.10 34884.29 1287.71 0
4 1 0 2298.16 2035.72 38.24 0
5 1 0 9.73 36.06 5.59 0
6 1 0 449.80 506.89 11.92 0
7 1 0 1.81 7.91 5.59 0
8 1 0 43.36 28.81 6.79 0
9 1 0 64.18 48.51 6.02 0
10 1 0 1412.22 1628.54 45.96 0
11 1 0 242.68 279.96 9.70 0
12 1 0 14.96 10.52 3.84 0
13 1 0 16.87 15.76 4.56 0
14 1 0 16.46 7.91 7.91 0
15 1 0 0.00 3.95 5.59 0
0 2 0 2443.71 2670.14 61.27 0
1 2 0 23397.80 23770.90 546.30 0
2 2 0 20572.37 19502.51 520.01 0
3 2 0 8854.88 8282.53 169.57 0
```

'Derived' data

Numerical description of the parameters of a calculated structure model



'Reference' data




 | [File](#) | [Edit](#) | [View](#) | [Tools](#) | [Help](#)







Bilbao Crystallographic Server - WYCKPOS - Wyckoff Positions

Help

Wyckoff positions of Group $P4/n\bar{c}$ (No. 126) [origin choice 2]

Multiplicity	Wyckoff letter	Site symmetry	Coordinates
16	k	1	(x,y,z) $(x+1/2,y+1/2,z)$ $(x,y+1/2,z)$ $(x+1/2,y,z)$ $(x,y,z+1/2)$ $(x+1/2,y,z+1/2)$ $(x,y+1/2,z+1/2)$ $(x+1/2,y+1/2,z+1/2)$
8	j	2	$(x,3/4,1/4)$ $(x+1/2,3/4,1/4)$ $(x,3/4,3/4)$ $(x+1/2,3/4,3/4)$
8	i	2	$(1/4,1/4,y)$ $(3/4,1/4,y)$ $(1/4,3/4,y)$ $(3/4,3/4,y)$
8	h	2	$(x,1/4,z)$ $(x+1/2,1/4,z)$ $(x,3/4,z)$ $(x+1/2,3/4,z)$
8	g	2	$(1/4,3/4,z)$ $(3/4,1/4,z)$ $(1/4,1/4,z+1/2)$ $(3/4,3/4,z+1/2)$
8	f	-1	$(0,0,z)$ $(1/2,1/2,z)$ $(0,0,z+1/2)$ $(1/2,1/2,z+1/2)$
4	e	4	$(1/4,1/4,1/4)$ $(1/4,1/4,3/4)$ $(3/4,1/4,1/4)$ $(3/4,1/4,3/4)$
4	d	-4	$(1/4,3/4,3/4)$ $(3/4,1/4,1/4)$ $(1/4,3/4,1/4)$ $(3/4,1/4,3/4)$
4	c	222	$(1/4,3/4,3/4)$ $(3/4,1/4,1/4)$ $(1/4,1/4,3/4)$ $(3/4,3/4,1/4)$
2	b	4	$(1/4,1/4,3/4)$ $(3/4,3/4,1/4)$
2	a	422	$(1/4,1/4,1/4)$ $(3/4,3/4,3/4)$

Wyckoff position and site symmetry group of a specific point

Specify the point by its relative coordinates (in fractions or decimals).
Variable parameters (x,y,z) are pre-assigned.

x =
y =
z =

If you want to see the Wyckoff position in other setting, click here

Bilbao Crystallographic Server
<http://www.crysl.bilbao.es>

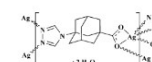
 For comments, please
administracion.bilbao@bilbao.es

'Interpretative' data

Variable parameters in the experimental set-up or numerical modelling and interpretation

Commentary

research communications



In present paper, we report the crystal structure of a new silver(I) coordination polymer $[\text{Ag}(\text{tr-ad-COO})_2]_n \cdot 2\text{H}_2\text{O}$ (**1**) based on the 1-(1,2,4-triazol-4-yl)-3-carboxyadamantane ($\text{C}_{14}\text{H}_9\text{N}_3\text{O}_2$; *tr-ad-COOH*) ligand.

2. Structural commentary

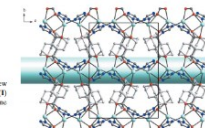


Figure 2
Projection on the *bc* plane showing the interconnection of sinusoidal A coordination chains by means of *n*-of-COO⁻ organic ligands into a three dimensional framework.

disordered over two adjacent sites (Fig. 1). The OH water molecule is situated on a crystallographic twofold axis passing through the O atom, while the OH water molecule is statistically disordered around the N atom. The hydrogen-bonding occupancy factor of 0.5. Thus, in the asymmetric unit, the total cationic center sums up to two water molecules. The 1,2,4-triazole functional group is coordinated by two Ag⁺ centers as a μ_2 -N,N' bridge and the carboxylate group connects two Ag⁺ centers in a chelating, bridging mode (μ_2 - η^2 - η^1), supporting the formation of sinusoidal chains with a periodicity of 13 Å.

In the case of compound 1b unusual situations with alternating of double triazoles and double carboxylate bridges are observed. The structure of compound 1bCO₃ is shown in a deprotonated form adopting a μ_2 - η^2 - η^1 coordination mode (Fig. 2) that yields a three-dimensional tetragonal pattern with open channels along the *c*-axis direction (Fig. 3).

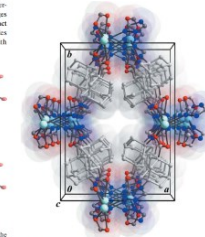


Figure 3

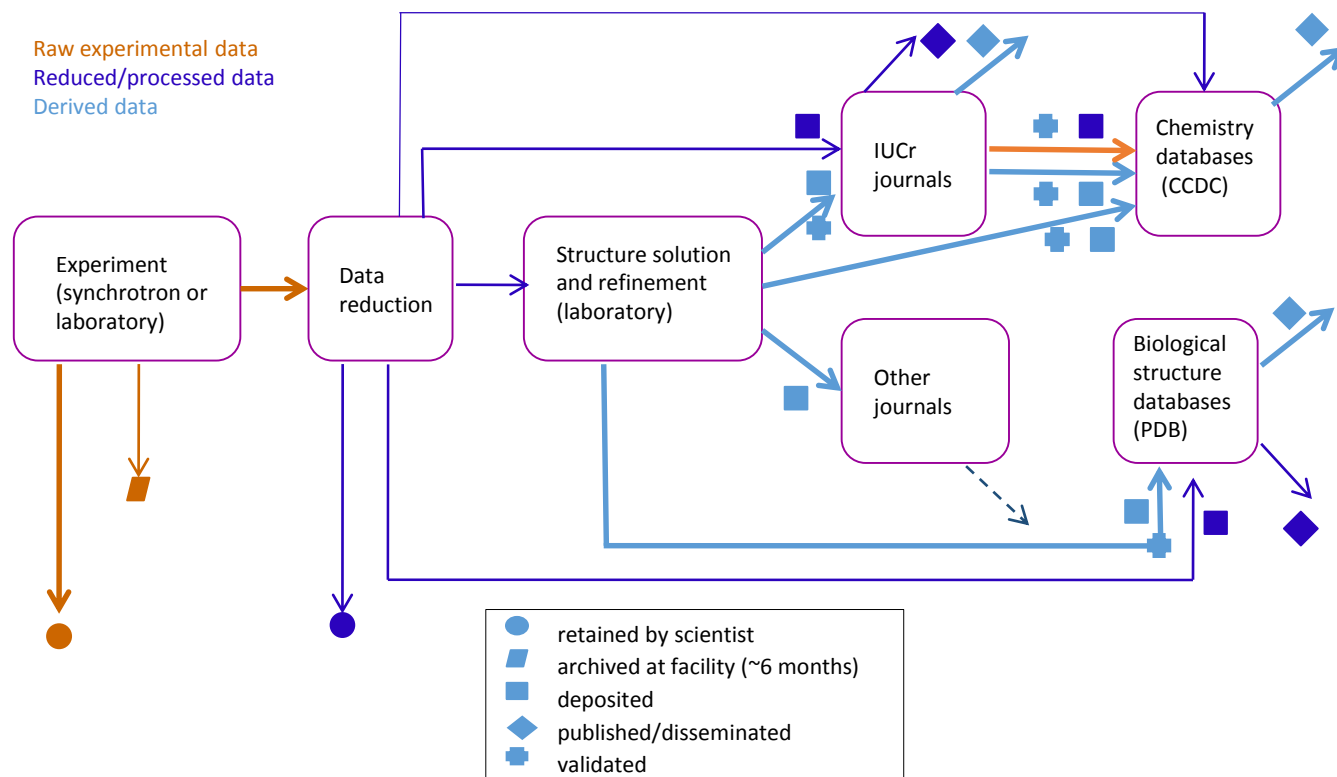
CIF dictionaries (COMCIFS)

- Crystallographic Core (coreCIF) – 1991
- Crystallographic Restraints – 2011
- Crystallographic Powder Diffraction (pdCIF) – 1997
- Modulated and Composite Structures (msCIF) – 2002
- Multipole Electron Density (rhoCIF) – 2003
- Crystallographic Twinning – 2014
- Magnetic Structures (magCIF) – 2016
- Lattice topology (topoCIF) – 2018
- Crystallographic Symmetry (symCIF) – 2001
- Diffraction Images (imgCIF) – 2000
- High pressure – under development
- Crystallographic Macromolecular Structure (mmCIF) – 1997

CIF dictionaries (wwPDB)

- Crystallographic Macromolecular Structure (mmCIF) – 1997
- PDB Exchange Dictionary (PDBx/mmCIF) – 1997 and ongoing
- Integrative/Hybrid (I/H) methods – 2017
- 3DEM Extension Dictionary – 2004
- NMRSTAR Dictionary – 2013
- Biological Small Angle Scattering– 1998
- Model Archive Extension Dictionary – 2018
- BIOSYNC Extension Dictionary – 2000
- NMR Exchange Format Dictionary – 2016

A coherent information flow



Early drivers for CIF

- Requirement to input raw data from many diffractometers
- Exchange data between software packages in the solution/refinement pipeline
- Provide a mechanism for electronic publication of articles
- Help referees to check the structural model
- Improve the accuracy of data capture by databases

Standardisation of raw data input

- Championed by Howard Flack for point detectors

J. Appl. Cryst. (1992). **25**, 455–459

DIFRAC, single-crystal diffractometer output-conversion software. By H. D. FLACK, *Laboratoire de Cristallographie, University of Geneva, 24 quai Ernest-Ansermet, CH-1211 Genève 4, Switzerland* and E. BLANC and D. SCHWARZENBACH, *Institut de Cristallographie, University of Lausanne, BSP Dorigny, CH-1015 Lausanne, Switzerland*

(Received 25 October 1991; accepted 9 January 1992)

Abstract

Software is described that converts single-crystal diffractometer output files as produced by manufacturers' software into a standardized instrument-independent form consisting of a clear, complete, well documented record of the sample and the diffraction measurements performed upon it. Information not already available in the manufacturers' diffractometer files is obtained in an interactive question-and-answer session. The software is written in a modular way. Available modules can deal with Enraf-Nonius CAD-4, Philips PW1100 and Siemens P_2 single-crystal diffractometers and produce output in CIF or SCFS format.

Introduction

Users of several models of four-circle single-crystal diffractometers may well have been struck by the diversity of form and content of the data files generated by diffractometer-manufacturers' software. The form of the file can create difficulties in its transfer to other computing equipment and further renders the corresponding computer programs specific to a certain type or types of diffractometer.

“The difficulties in the content of the files manifest themselves principally by **the paucity of the available information** necessitating additional input to the data-treatment software (e.g. type of radiation, wavelength of radiation, scan width, ...). The problem has become aggravated in recent times by the rapid development in electronic data exchange. ... The advent of machine-readable submission for publication and data-base and supplementary-material deposition further highlights **the problem of missing data.**”

Standardisation of raw data input

- Championed by Howard Flack for point detectors
- Relatively unenthusiastic uptake by vendors
- CIFs produced by commercial vendors often faulty
- imgCIF/CBF a second-wave attempt for diffraction images
- Perhaps slightly more effort from vendors
- 'mini-CBFs' showed significant diversity
- Reluctance to record essential experimental metadata

Data exchange between software packages

- STAR File modelled loosely on *Xtal* internal data structures
- Particular early adopters were collaborative packages (*NRCVAX*, *CCP4*) or general-purpose analysis programs (*PLATON*)
- Adoption by ‘market leader’ *SHELXL* crucial for success
- Reluctant uptake of mmCIF because of existing PDB format
- In general, has stood up well to the test of time
 - Generic presentations like XML not popular (but watch JSON!)
 - Limitations of PDB format eventually led to mmCIF as standard
 - Adoption of HDF5/NeXus in image capture driven by throughput needs
 - Relational nature of DDL dictionaries has become more important

Electronic publication

- Adoption by IUCr Journals (1991) very important
- Mandatory data format for *Acta C* by 1996
- *Notes for Authors* stipulate and validate mandatory metadata by listing required CIF data items
- Most journals reporting crystal structures now require CIFs as supporting information
- For online publication, availability of CIF makes the publication interactive

Technical peer review

- IUCr journals always required reviewers to check chemical structures
- CIF allowed ease of input to popular checking programs (*PARST*, *MISSYM*, *UNIMOL*, *CREDUC*)
- This led to semi-automated validation
 - IUCr checking reports
 - checkCIF dedicated tests
 - Consolidation with *PLATON*
- Elimination of 'Marshing'
- Subsequent ability to validate structure factors and re-do refinement

Improved quality of structures in databases

- Improved quality of small-unit-cell structures from IUCr journals
- Convergence of checking criteria with CCDC
- Adoption of *checkCIF* for CCDC direct depositions
- **wwPDB database structure isomorphic to mmCIF**
- PDB requirement for structure factors in CIF format
- Validation Reports developed alongside extension CIF dictionaries for novel techniques

mmCIF RDBMS

- CIF originally developed as tagged file format with one-dimensional loops
- York mmCIF workshop (1993) identified similarities with relational database structure
- DDL2 developed 1994/5 as a fully relational description of CIF data items
- The mmCIF/DDL2 schema is the basis for the relational database schema used by the wwPDB
- Latest iteration of DDLm retains this relational nature

mmCIF RDBMS

_computational
BIOLOGY

COMMENTARY

Overhauling the PDB

Amanda C Schierz, Larisa N Soldatova & Ross D King

The Brookhaven Protein Data Bank was once a pioneering database, but its organization of structural data is now outdated and in need of an upgrade.

Although structural biology was once a leader in both the development of standards for the preservation and sharing of scientific data and for database development, this lead has been lost. The main data standard used by PDB, mmCIF, does not meet state-of-the-art standards in biology for ontologies; this has serious repercussions for the analysis and sharing of data, and has led to problems in database design. The main database, the reengineered Brookhaven Protein Data Bank (PDB), although intended to be a relational database, does not conform to relational database design standards and principles; this diminishes the quality of information stored in PDB and has serious implications for data-storage capacity. If structural biology is to realize the full potential of its wealth of stored data and regain its lead in data standards and databases, then both mmCIF and PDB need to be extensively reengineered. In this article, we detail some of the more serious faults we have found in the mmCIF and PDB. We then sketch a way to

Since the early 1970s, structural biology has been at the forefront of the development of standards for the preservation and sharing of scientific data. The PDB was set up in 1971 (more than 10 years before the first sequence databases, such as EMBL-Bank and GenBank). Structural biology journals were among the first to require submission of data to international databases as a condition of publication, and the mmCIF dictionary³, published in 1997, was one of the first standards internationally agreed upon for reporting experimental results.

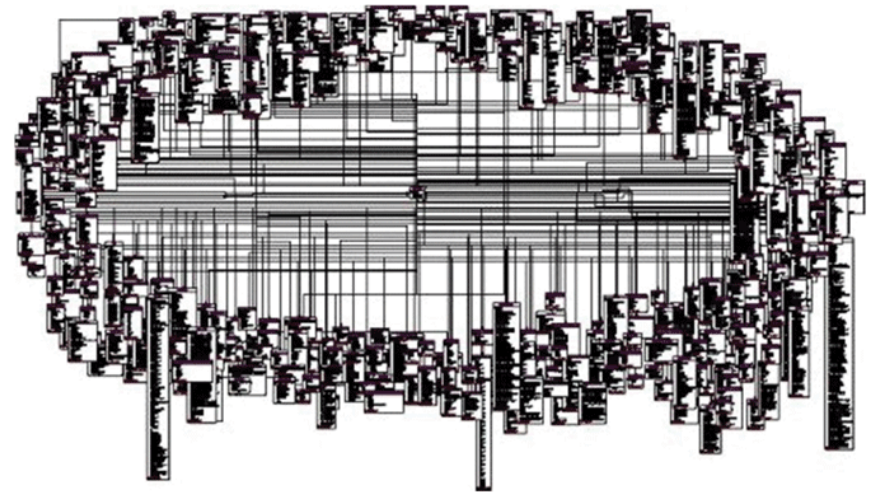
The PDB was recently reengineered to become a 'relational' database based on the mmCIF schema⁴. As wwPDB, it now includes the PDB at the Research Collaboratory for Structural Bioinformatics (RCSB, Brookhaven), the Macromolecular Structure Database (MSD) at the European Bioinformatics Institute (Hinxton Hall, UK), the Protein Data Bank (Osaka University, Japan) and more recently the Biological Magnetic Resonance Data Bank.

Our discussion here focuses mainly on the RCSB PDB, as we consider it to have the most problems.

Our analysis indicates that the reengineering process was not a success and has led to problems of data repetition, redundancy, inconsistency and integrity. These problems are important: they can result directly in incorrect answers to queries or more indirectly lead to erroneous results through the incomplete updating of the database due to the formation of uncontrolled redundancies. They also seriously inhibit the future possibility of designing intelligent analysis tools to analyze the data in PDB and aid the process of knowledge discovery.

Why mmCIF is not a true ontology

Several biological ontologies are now in use, including FMA (Foundational Model of Anatomy), which details the key concepts and relations in anatomy⁵, and FuGO (The



Global schema map of the entire PDB relational database; from Schierz, A. C., Soldatova, L. N. & King, R. D. (2007). *Nature Biotechnology*, **25**, 437-442

The Brookhaven Protein Data Bank was once a pioneering database, but its organization of structural data is now outdated and in need of an upgrade.

mmCIF RDBMS

- *Critique*: too many tables in the database empty or nearly empty
→ poorly defined schema
- *Reality*: each table represents a category of information needed to describe the structure completely
- *Problem*: depositors unwilling to provide that information (*i.e.* populate the tables)
- Most of the poorly populated tables capture experimental *metadata*

Metadata capture in macromolecular reports

← → ↻ ⓘ Not secure | journals.iucr.org/ff/services/structuralcommunications/

Acta Cryst F Acta Crystallographica Section F STRUCTURAL BIOLOGY COMMUNICATIONS

home archive editors for authors for readers submit subscribe open access

data for structural and crystallization communications

This page gives a list of recommended items for inclusion in structural and crystallization communications in *Acta Crystallographica Section F*.

Items that are given in a **magenta typeface are mandatory**.

The items will be published in tables that provide information on the sample and its treatment (including crystallization); data collection and structure refinement details.

An online tool is available for preparing these tables from an mmCIF.

Click here for more details on these individual items and information on supplying the recommended information in mmCIF format.

If your structure has been solved using NMR, click here for a separate set of recommendations.

1. Sample information

1.1. Macromolecule and source information

Structure name ⓘ
Component molecules
Additional molecular identifiers
Biological functional unit (BFU) or macromolecular assembly, numbers and types of chains
Mass of BFU (Da)

Macromolecule sequence and chemical configuration ⓘ
Sequence database reference code
Polymers (one-letter code sequence) or Polymer sequence as list of residues
Ligand, cofactor, ions, solvent
Mutations
Post-translational modifications
Formula weight of entity (Da)

Source organism
Scientific name
Strain
Details
Source gene
Scientific name
Strain
Details

1.2 Macromolecule production ⓘ

For each macromolecular entity
PCR protocol
Cloning protocol
Expression protocol
Purification protocol
Additional details

1.3. Crystallization ⓘ

Crystallization method
Temperature (K)
Additional details
Annarative

← → ↻ ⓘ Not secure | journals.iucr.org/ff/services/structuralcommunications/mmCIFreqditems.html

Acta Cryst F Acta Crystallographica Section F STRUCTURAL BIOLOGY COMMUNICATIONS

home archive editors for authors for readers submit subscribe open access

data for structural and crystallization communications

This page gives a list of recommended items for inclusion in structural and crystallization communications in *Acta Crystallographica Section F*.

The recommendations are tabulated below alongside the data names from the PDB mmCIF exchange dictionary available from the Protein Data Bank. These data items are provided in the mmCIF data sets created by the Protein Data Bank when a structure is deposited.

If you intend to submit to *Acta Crystallographica Section F*, you are recommended to use the *PDB_EXTRACT* utility available from the Protein Data Bank to extract as much as possible of this information from result and log files of most commonly-used macromolecular structure packages to create an initial mmCIF suitable for direct upload to the Protein Data Bank during the deposition process. A copy of the deposited mmCIF can then be uploaded to the journal to create standard tables for inclusion in your article.

The list below also includes examples to show how particular data will be organised in the mmCIF and how they will be arranged in the journal article.

Click here for a more compact summary of the recommendations.

- 1. Sample information
 - 1.1. Macromolecule and source information
 - 1.2. Macromolecule production
 - 1.3. Crystallization
 - 1.4. Crystal data
- 2. Data collection and structure solution statistics
 - 2.1. Data collection, refinement data set
 - 2.2. Phasing
 - 2.2.1. MAD/SAD data and structure solution statistics
 - 2.2.2. MIR/MIRAS/SIR/SIRAS data and structure solution statistics
 - 2.2.3. Molecular replacement data and structure solution statistics
- 3. Model generation and refinement
- 4. Model validation

1. Sample information

_symmetry.cell_setting
_symmetry.int_tables_number
_symmetry.space_group_name_H-M

Description	mmCIF items
-------------	-------------

1.1. Macromolecule and source information

Example 1: complex of *E. coli* glutamate decarboxylase α with glutarate
Example 2: a zinc-induced heterodimer of two isoforms of phospholipase A₂
Example 3: mutation
Example 4: mutation and modification

Structure name ⓘ	_struct.title
Component molecules	_entity.pdbx_description ENTITY_NAME_SYS ENTITY_NAME_COM _entity.pdbx_ec
Additional molecular identifiers	
Biological functional unit (BFU) or macromolecular assembly, numbers and types of chains	_struct_biol.details
Mass of BFU (Da)	_struct_biol.pdbx_formula_weight _struct_biol.pdbx_formula_weight_method
Macromolecule sequence and chemical configuration ⓘ	

Authoring tools to assist metadata capture (*publBio* publisher)

- For MX a particular shortcoming is details of sample preparation and crystallization
- *publBio* offers a helpful interface
- Still a drought of crystallization papers
- Perhaps a role for *IUCrData*?

publBio data collection - Mozilla Firefox

Precipitant solution

Volume Volume units pH

Components of the precipitant solution

Name	Concentration OR concentration range	in the precipitant solution
cadmium acetate		
polyethylene glycol 8000		
ammonium acetate		
ammonium bromide		
ammonium chloride		
Ammoniumchlorid		
ammonium citrate - ammonium hydroxide		
ammonium citrate - citric acid		
ammonium dihydrogen phosphate		
ammonium phosphate (monobasic)		
ammonium phosphate monobasic		
ammonium phosphate, monobasic		
ammonium fluoride		
ammonium formate		
ammonium iodide		
ammonium nitrate		
ammonium selenate		
ammonium sulfate		
Ammoniumsulfate		
ammonium sulphate		

Crystallization screens
(data kindly provided by Rigaku)

Vendor: Emerald BioSystems

Screen: PEG 8K Anomalous

Well number: 8

cadmium acetate (0.2 M)

polyethylene glycol 8000 (20 w/v)

Click any of the above components to add them to your table [the 'target' input box will be highlighted when you hover over the above items - click in any of the component fields (Name, Concentration...) to change the 'target']

The *checkCIF* paradigm

Three aspects to *checkCIF* validation:

1. Completeness (the 'metadata' problem)

The metadata problem

- *checkCIF* for small structures achieves this reasonably well
 - For IUCr journals: published requirements in Notes for Authors
 - For other journals: in principle, customised ‘request lists’; in practice, *de facto* acceptance of IUCr recommendations with *ad hoc* pick and choose
 - Possibility of better uptake through automated LIMS systems
- For protein structures
 - Reluctance in specifying mandatory data items
 - Inertia in design of standard experimental protocols in some facilities
 - Automated LIMS and other systems would help
- For raw data
 - CommDat sponsoring ‘*checkCIF* for raw data’ initiative

The *checkCIF* paradigm

Three aspects to *checkCIF* validation:

1. Completeness (the 'metadata' problem)
2. Internal self-consistency (relational methods)

Internal self-consistency

- IUCr *checkCIF* tests perform wide range of consistency checks
- Many of these also in *PLATON*
- Relational nature of DDL2 ensures category integrity within mmCIF
- DDLm offers a generic approach to specifying integrity relationships in dictionaries
- *dREL* a specific implementation of DDLm methods

The *checkCIF* paradigm

Three aspects to *checkCIF* validation:

1. Completeness (the 'metadata' problem)
2. Internal self-consistency (relational methods)
3. Comparison with related structures (the 'knowledge' problem)

The knowledge problem

- *PLATON* includes huge amount of crystallographic and chemical knowledge
- *Mogul* offers opportunity to compare new structures with existing curated ones in CSD
- PDB Validation Reports build on existing knowledge bases
- These tools are tuned to molecular geometry and other chemical properties
- Prospect of interrogating any ensemble of CIFs for any other defined properties through a non-specific API

Pain points for CIF in the 21st century

- Reluctance/difficulty in supplying experimental metadata
 - Reticence in revising core dictionaries
 - Lax approach to detailed standards in some software packages
 - Lack of resources in developing CIF2/DDIm applications
 - Lack of resources for validation software maintenance and development
-
- None are critical – many show some (but slow) progress
 - Important to keep momentum and train new generation

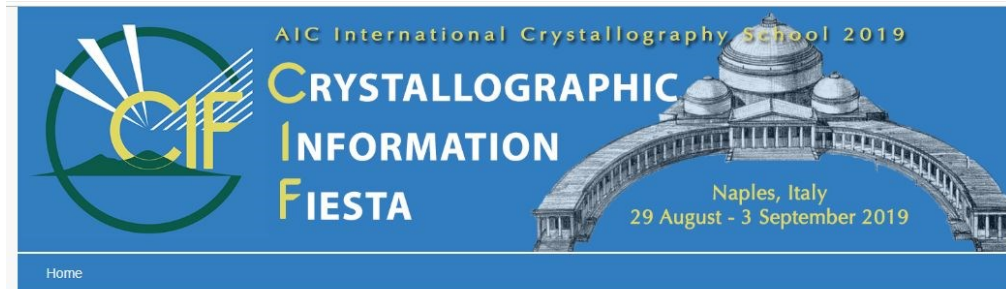
“

There is a traditional hierarchy of components of understanding: **data**, **information**, **knowledge**, **wisdom** (the DIKW model).

Crystallographic **information** is the component that bridges the gap between the raw experimental **data** and the global **knowledge** bases represented by the Protein Data Bank, Cambridge Structural Database, International Centre for Diffraction Data, Inorganic Crystal Structure Database, Crystallography Open Database etc.

This school will teach students to respect their raw data, extract the most reliable information they can, and disseminate that information in a complete and verifiable manner. In this way they will contribute to the sum total of scientific knowledge with rigour and integrity.

Crystallographic wisdom is outside the scope of this course. ”



AIC International Crystallography School 2019
CRYSTALLOGRAPHIC INFORMATION FIESTA
 Naples, Italy
 29 August - 3 September 2019

Home

- Aims
- Programme
- Timetable
- Lecturers
- Committees
- Bursaries
- Registration
- Venue
- AICS Policies

 Follow news and updates also on the AIC Facebook page!

For any queries please contact the Organizing Committee at aicschool@cristallografia.org

Organized by



Supporting Institutions



AIC International Crystallography School 2019

The AIC International Crystallography School 2019 offers an intensive course in Crystallographic Information, covering the extraction, dissemination and use of scientific knowledge from the structure determination experiment to database-driven discovery.



Download the AICS2019 First Circular
 Download the poster of AICS2019

There is a traditional hierarchy of components of understanding: data, information, knowledge, wisdom (the DIKW model). In a way, crystallographic information is the component that bridges the gap between the raw experimental data and the global knowledge bases represented by the Protein Data Bank, Cambridge Structural Database, International Centre for Diffraction Data, Inorganic Crystal Structure Database, Crystallography Open Database etc. It is typically thought of in terms of publications in the scientific literature, but in fact it is more than that. It includes the characterisation and understanding of the raw data collected in an experiment, an account of the methods applied to process and analyse those data to infer a structure, and a quantification of the extent to which features of the model should be trusted or treated with caution. In an era when crystal structure determination is considered routine and is often largely automatic, a critical appreciation of the quality of the results is increasingly important. **This school will teach students to respect their raw data, extract the most reliable information they can, and disseminate that information**

With the collaboration and support of



Sponsors



Crystallographic information in the FAIR era

