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

A Satellite Workshop to the 29th European Crystallographic Meeting

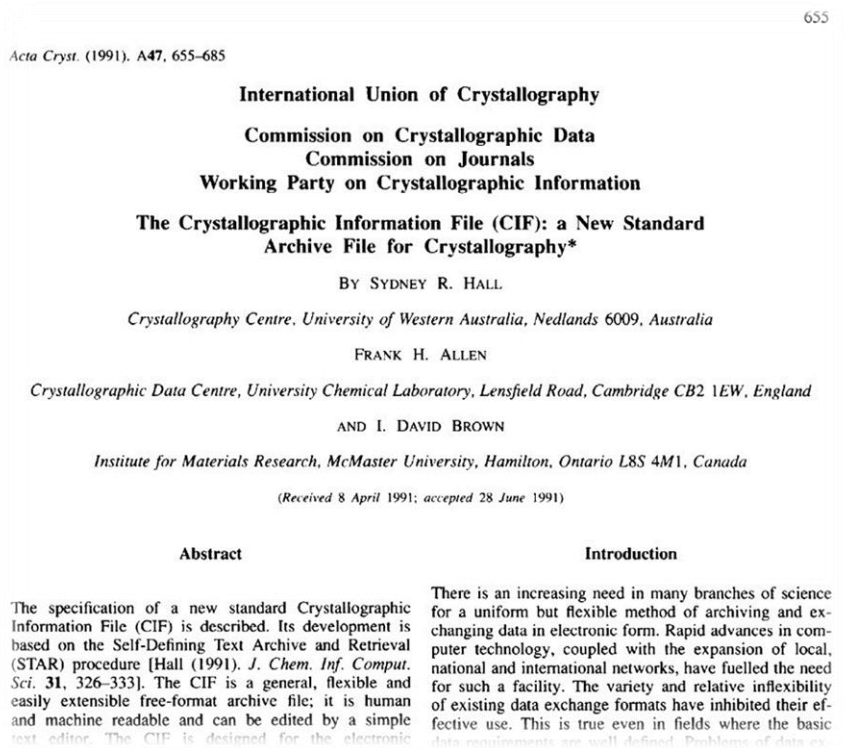
The Crystallographic Information Framework as a metadata library

Brian McMahon

International Union of Crystallography
5 Abbey Square
Chester CH1 2HU
UK
bm@iucr.org

CIF as a file format

1991 – Introduction of Crystallographic Information File format



655

Acta Cryst. (1991). A**47**, 655–685

International Union of Crystallography

Commission on Crystallographic Data
Commission on Journals
Working Party on Crystallographic Information

The Crystallographic Information File (CIF): a New Standard Archive File for Crystallography*

BY SYDNEY R. HALL

Crystallography Centre, University of Western Australia, Nedlands 6009, Australia

FRANK H. ALLEN

Crystallographic Data Centre, University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, England

AND I. DAVID BROWN

Institute for Materials Research, McMaster University, Hamilton, Ontario L8S 4M1, Canada

(*Received* 8 *April* 1991; *accepted* 28 *June* 1991)

Abstract

The specification of a new standard Crystallographic Information File (CIF) is described. Its development is based on the Self-Defining Text Archive and Retrieval (STAR) procedure [Hall (1991). *J. Chem. Inf. Comput. Sci.* **31**, 326–333]. The CIF is a general, flexible and easily extensible free-format archive file; it is human and machine readable and can be edited by a simple

Introduction

There is an increasing need in many branches of science for a uniform but flexible method of archiving and exchanging data in electronic form. Rapid advances in computer technology, coupled with the expansion of local, national and international networks, have fuelled the need for such a facility. The variety and relative inflexibility of existing data exchange formats have inhibited their effective use. This is true even in fields where the basic

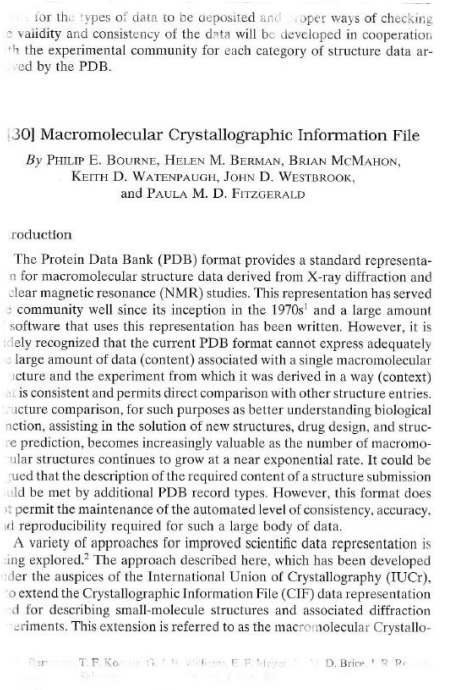
Hall, S. R., Allen, F. H. & Brown, I. D. (1991). The Crystallographic Information File (CIF): a New Standard Archive File for Crystallography. *Acta Cryst.* A**47**, 655-685

CIF as a file format

```
data_I

_chemical_name_systematic   'Biphenyl-2,4,4',6-tetracarboxylic acid monohydrate'
_chemical_formula_moiety          'C16 H10 O8, H2 O'
_chemical_formula_sum             'C16 H12 O9'
_chemical_formula_weight           348.26
_symmetry_cell_setting             monoclinic
_symmetry_space_group_name_H-M     'P 21/c'
_symmetry_space_group_name_hall    '-p 2ybc'
loop_
    _symmetry_equiv_pos_as_xyz
    'x, y, z'     '-x, y+1/2, -z+1/2'     '-x, -y, -z'     'x, -y-1/2, z-1/2'
_cell_length_a         5.638(4)
_cell_length_b        16.160(11)
_cell_length_c        16.798(12)
_cell_angle_alpha     90.00
_cell_angle_beta      92.524(12)
_cell_angle_gamma     90.00
_cell_volume        1528.9(19)
```

CIF as a file format

1991 – Introduction of Crystallographic Information File format
1996 – Introduction of mmCIF



for the types of data to be deposited and proper ways of checking
validity and consistency of the data will be developed in cooperation
the experimental community for each category of structure data ar-
ved by the PDB.

[30] Macromolecular Crystallographic Information File

By PHILIP E. BOURNE, HELEN M. BERMAN, BRIAN MCMAHON, KEITH D. WATENPAUGH, JOHN D. WESTBROOK, and PAULA M. D. FITZGERALD

roduction

The Protein Data Bank (PDB) format provides a standard representa-
n for macromolecular structure data derived from X-ray diffraction and
clear magnetic resonance (NMR) studies. This representation has served
community well since its inception in the 1970s[1] and a large amount
software that uses this representation has been written. However, it is
idely recognized that the current PDB format cannot express adequately
large amount of data (content) associated with a single macromolecular
ucture and the experiment from which it was derived in a way (context)
is consistent and permits direct comparison with other structure entries.
ucture comparison, for such purposes as better understanding biological
nction, assisting in the solution of new structures, drug design, and struc-
re prediction, becomes increasingly valuable as the number of macromo-
ular structures continues to grow at a near exponential rate. It could be
ued that the description of the required content of a structure submission
uld be met by additional PDB record types. However, this format does
t permit the maintenance of the automated level of consistency, accuracy,
d reproducibility required for such a large body of data.

A variety of approaches for improved scientific data representation is
ing explored.[2] The approach described here, which has been developed
der the auspices of the International Union of Crystallography (IUCr),
o extend the Crystallographic Information File (CIF) data representation
d for describing small-molecule structures and associated diffraction
eriments. This extension is referred to as the macromolecular Crystallo-

Bourne, P. E., Berman, H. M., McMahon, B., Watenpaugh, K. D., Westbrook, J. & Fitzgerald, P. M. D. (1997). The Macromolecular Crystallographic Information File (mmCIF). *Methods Enzymol.* **277,** 571-590

CIF as a file format

1991 – Introduction of Crystallographic Information File format

1996 – Introduction of mmCIF

2000 – Introduction of imgCIF/CBF

2.3. Specification of the Crystallographic Binary File (CBF/imgCIF)

BY H. J. BERNSTEIN AND A. P. HAMMERSLEY

2.3.1. Introduction

The Crystallographic Binary File (CBF) format is a complementary format to the Crystallographic Information File (CIF) (Hall *et al.*, 1991) supporting efficient storage of large quantities of experimental data in a self-describing binary format. The image-supporting Crystallographic Information File (imgCIF) is an extension to CIF to assist in ASCII debugging and archiving of CBF files and to allow for convenient and standardized inclusion of images, such as maps, diagrams and molecular drawings, into CIFs for publication. The binary CBF format is useful for handling large images within laboratories and for interchange among collaborating groups. For smaller blocks of binary data, either format should be suitable. The ASCII imgCIF format is appropriate for interchange of smaller images and for long-term archiving.

CBF is designed to support efficient storage of raw experimental data (images) from area detectors with no loss of information, unlike some existing formats intended for this purpose. The format enables very efficient reading and writing of raw data, and encourages economical use of disk space. It may be coded easily and is portable across platforms. It is also flexible and extensible so that new data structures can be added without affecting the present definitions.

These goals are achieved by a simple file format, combining a CIF-like file header with compressed binary information. The file header consists of ASCII text giving information about the binary data as CIF tag–value pairs and tables. Each binary image is presented as a text-field value, either as raw octets of binary data in a CBF data set, or as an ASCII-based encoding of the same binary information in a true ASCII imgCIF data set. The ASCII-based encoded format uses e-mail MIME (Multipurpose Internet Mail Extensions) conventions to encode the binary data (Freed & Borenstein, 1996*a*,*b*,*c*; Freed *et al.*, 1996; Moore, 1996). The present version of the format tries to deal only with simple Cartesian data. These are essentially the 'raw' diffraction data that typically are stored in commercial formats or individual formats internal to particular institutes. Other forms of binary image data could be accommodated. It is hoped that CBF will replace individual laboratory or institute formats for 'home-built' detector systems, will be used as an inter-program data-exchange format, and will be offered as an output choice by commercial detector manufacturers specializing in X-ray and other detector systems. In this chapter we discuss the basic framework within which binary data and images are stored. The categories and data items that are used to describe beam and equipment axes, rastering methodologies, and image compression techniques are described in Chapter 3.7. The CBF/imgCIF dictionary is given in Chapter 4.6. An application programming interface (API) for the manipulation of image data is described in Chapter 5.6.

Affiliations: HERBERT J. BERNSTEIN, Department of Mathematics and Computer Science, Kramer Science Center, Dowling College, Idle Hour Blvd, Oakdale, NY 11769, USA; ANDREW P. HAMMERSLEY, ESRF/EMBL Grenoble, 6 rue Jules Horowitz, France.

2.3.2. CBF and imgCIF

CBF and imgCIF are two aspects of the same format. Since CIFs are pure ASCII text files, it was necessary to define a separate binary format to allow the combination of pseudo-ASCII sections and binary data sections. In the binary-file CBF format, the ASCII sections conform closely to the CIF standard but must use operating-system-independent 'line separators'. In order to facilitate interchange of files, an API that writes CBF files should use `\x0d\x0a` (carriage return, line feed) for the line separator. Use of this line separator allows the ASCII sections to be viewed with standard system utilities (*e.g.* 'more', 'pg') on a very wide range of operating systems (*e.g.* Unix, MacOS and Windows). However, an API that reads CBF format must accept any of the following three alternative line terminators as the end of an ASCII line: `\x0d`, `\x0a` or `\x0d\x0a`. As for all CIF data sets, an imgCIF file conforms to the normal text file-writing conventions of the system on which it is written. imgCIF is also one of the two names of the CIF dictionary (see Chapter 4.6) that contains the terms specific to describing image data in both CBF and imgCIF data sets. Thus a CBF or imgCIF data set uses data names from the CBF/imgCIF dictionary and other CIF dictionaries.

The general structure of a CBF or imgCIF data set is shown in Example 2.3.2.1. After a special comment to identify the file type (a so-called 'magic number') and any other initial comments, the data set begins with a '`data_biochname`' which gives the name of the data block. Tags and values that describe the image data and how they were collected come next. For efficiency in processing, it is recommended that all the descriptive tags come before the actual image data. This recommendation is a *requirement* for the binary CBF format. It is *optional* for the ASCII imgCIF format. The image data are given as the value of the tag '`_array_data.data`'. The image data are given in a text field, using MIME conventions to describe the encoding.

2.3.2.1. A simple example

Before describing the format in full, we start by showing a simple but important and complete use of the format: that of storing a single detector image in a file together with a small amount of auxiliary information. This is intended to be a useful example that can be understood without reference to the full definitions. It also serves as an introduction or overview of the format definition. This example uses CIF DDL2-based dictionary items (see Chapter 2.6).

Example 2.3.2.2 relates to an image of 768×512 pixels stored as 16-bit unsigned integers, in little-endian byte order (this is the native byte ordering on a PC). The pixel sizes are 100.5×99.5 μm.

The example will be presented and discussed in three sections. The circled numerals (*e.g.* ①) are included to allow us to comment on portions of the example. They are not part of the CBF/imgCIF format.

The line marked by ①, starting with a hash character (#), is a CIF and CBF comment line. As a first line, the pattern of three hashes followed by 'CBF' helps to identify the data set as a CBF. It is a so-called 'magic number'. The text '`###CBF: VERSION` must be present as the very first line of every CBF file. Following

Bernstein, H. J. & Hammersley, A. P. (2006). Specification of the Crystallographic Binary File (CBF/imgCIF). In *International Tables For Crystallography Vol. G: Definition and exchange of crystallographic data*. Dordrecht: Springer

CIF as a file format

1991 – Introduction of Crystallographic Information File format
1996 – Introduction of mmCIF
2000 – Introduction of imgCIF/CBF
2016 – Introduction of CIF2.0

Specification of the Crystallographic Information File (CIF) format, version 2.0

Herbert J. Bernstein,[a] John C. Bollinger,[b]* I. David Brown,[c] Saulius Gražulis,[d] James R. Hester,[e] Brian McMahon,[f] Nick Spadaccini,[g] and Simon P. Westrip[h]

[a]Rochester Institute of Technology, 85 Lomb Memorial Drive, Rochester, NY 14623, USA; [b]St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, TN 38105, USA; [c]BIMR, McMaster University, 1280 Main Street West, Hamilton, Ontario, Canada L8S 4M1; [d]Vilnius University Institute of Biotechnology, Graičiuno 8, LT-02241 Vilnius, Lithuania; [e]Australian Nuclear Science and Technology Organisation, New Illawarra Road, Lucas Heights, NSW 2234, Australia; [f]International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England; [g]The University of Western Australia, Crawley 6009, Australia; [h]The Walled Garden, Horton Green, Cheshire SY14 7EY, England.

Correspondence e-mail: John.Bollinger@StJude.org

Synopsis - Version 2.0 of the CIF format is described, and a formal specification is provided.

Abstract - Version 2.0 of the CIF format incorporates novel features implemented in STAR 2.0. Among these are an expanded character repertoire, new and more flexible forms for quoted data values, and new compound data types. The CIF 2.0 format is compared with both CIF 1.1 and STAR 2.0, and a formal syntax specification is provided.

Keywords: CIF; CIF 2.0

1. Introduction

The Crystallographic Information File (CIF, Hall *et al.*, 1991, Hall *et al.*, 2005) is a well-established format for data exchange and archiving in crystallography. Since its debut, CIF and CIF applications have come to support an extensive, ontology-based, global framework for crystallographic data exchange and processing, sometimes called the Crystallographic Information Framework (also CIF, Hall & McMahon, 2005).

Although CIF version 1.1 (Hall *et al.*, 2005) and its parent format, STAR 1.0 (Hall, 1991; Hall & Spadaccini, 1994), have broad expressive power, their designs incorporate limitations that were common at the time of their introduction. Those restrict the characters and therefore languages that can be readily represented, and they make presentation of vectors,

Bernstein, H. J., Bollinger, J. C., Brown, I. D., Gražulis, S., Hester, J. R., McMahon, B., Spadaccini, N. & Westrip, S. P. (2015). Specification of the Crystallographic Information File (CIF) format, version 2.0. *J. Appl. Cryst.* Submitted

The irrelevance of format

```
    'C' 'C' 0.0033 0.0016
                        'International Tables Vol C Tables 4.2.6.8 and 6.1.1.4'
    'H' 'H' 0.0000 0.0000
                        'International Tables Vol C Tables 4.2.6.8 and 6.1.1.4'
    'N' 'N' 0.0061 0.0033
                        'International Tables Vol C Tables 4.2.6.8 and 6.1.1.4'
    'O' 'O' 0.0106 0.0060
                        'International Tables Vol C Tables 4.2.6.8 and 6.1.1.4'
    'S' 'S' 0.1246 0.1234
                        'International Tables Vol C Tables 4.2.6.8 and 6.1.1.4'
    'Br' 'Br' -0.2901 2.4595
                        'International Tables Vol C Tables 4.2.6.8 and 6.1.1.4'
_computing_data_collection      '<i>CrystalClear</i> (Rigaku, 2008)'
_computing_cell_refinement      '<i>CrystalClear</i> (Rigaku, 2008)'
_computing_data_reduction      '<i>CrystalClear</i> (Rigaku, 2008)'
_computing_structure_solution      '<i>SIR97</i> (Altomare  <i>et al.</i>, 1999)'
_computing_structure_refinement
        '<i>SHELXL97</i> (Sheldrick, 2008) within <i>WinGX</i> (Farrugia, 2012)'
_computing_molecular_graphics
 '<i>ORTEP-3 for Windows</i>  (Farrugia, 2012) and <i>POV-RAY</i> (Cason, 2004)'
_computing_publication_material
          '<i>SHELXL97</i> (Sheldrick, 2008) and <i>publCIF</i> (Westrip, 2010)'
loop_
    _atom_site_type_symbol
    _atom_site_label
    _atom_site_fract_x
    _atom_site_fract_y
    _atom_site_fract_z
    _atom_site_U_iso_or_equiv
    _atom_site_adp_type
    _atom_site_calc_flag
    _atom_site_refinement_flags
    _atom_site_occupancy
    _atom_site_symmetry_multiplicity
    _atom_site_disorder_assembly
    _atom_site_disorder_group
    C C1 0.4149(12) 0.8677(5) 0.3436(3) 0.0254(13) Uani d . 1 1 . .
    H H1 0.4019 0.9100 0.2958 0.030 Uiso calc R 1 1 . .
    C C2 0.5672(11) 0.7710(5) 0.3635(3) 0.0223(12) Uani d . 1 1 . .
    C C3 0.7454(12) 0.6007(5) 0.3458(3) 0.0244(13) Uani d . 1 1 . .
    H H3A 0.8736 0.5675 0.3141 0.029 Uiso calc R 1 1 . .
    H H3B 0.5491 0.5663 0.3489 0.029 Uiso calc R 1 1 . .
    C C4 0.8723(12) 0.5774(5) 0.4210(3) 0.0238(12) Uani d . 1 1 . .
    H H4A 0.8982 0.4962 0.4395 0.029 Uiso calc R 1 1 . .
    H H4B 1.0666 0.6133 0.4177 0.029 Uiso calc R 1 1 . .
    C C5 0.5503(10) 0.7210(5) 0.4403(3) 0.0188(11) Uani d . 1 1 . .
```

Mitchell, L. A. & Holliday, B. J. (2014). 6-Bromo-*N*-(6-bromopyridin-2-yl)-*N*-[4-(2,3-dihydrothieno[3,4-*b*][1,4]dioxin-5-yl)phenyl]pyridin-2-amine. *Acta Cryst.* E**70**, o797.

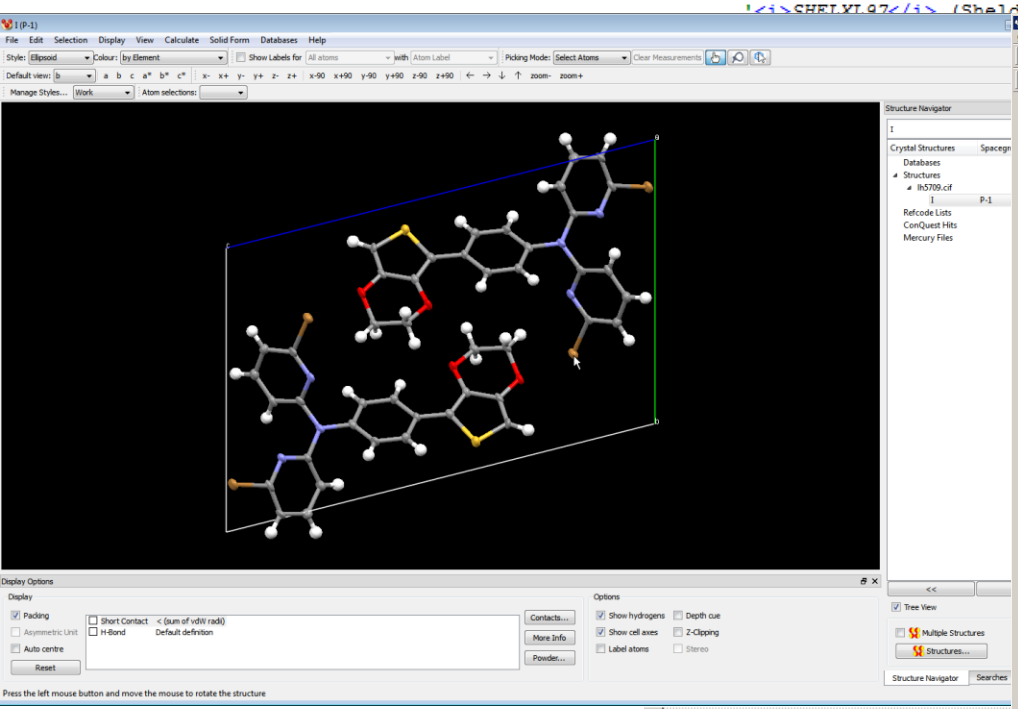
Rendered using *Mercury* [Macrae, C. F. *et al.* (2009). *J. Appl. Cryst.* **41**, 466-470]

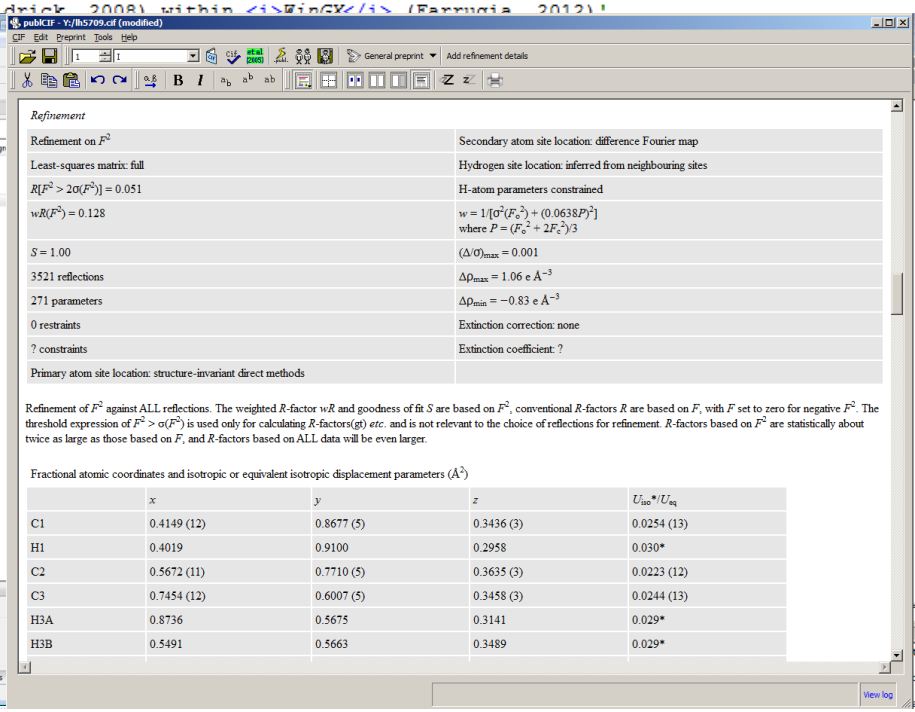
Rendered using *publcif* [Westrip, S. P. (2010). *J. Appl. Cryst.* **43**, 920-935]

The irrelevance of format

```
'C'  'C'   0.0033  0.0016
                    'International Tables Vol C Tables 4.2.6.8 and 6.1.1.4'
'H'  'H'   0.0000  0.0000
                    'International Tables Vol C Tables 4.2.6.8 and 6.1.1.4'
'N'  'N'   0.0061  0.0033
                    'International Tables Vol C Tables 4.2.6.8 and 6.1.1.4'
'O'  'O'   0.0106  0.0060
                    'International Tables Vol C Tables 4.2.6.8 and 6.1.1.4'
'S'  'S'   0.1246  0.1234
                    'International Tables Vol C Tables 4.2.6.8 and 6.1.1.4'
'Br' 'Br' -0.2901  2.4595
                    'International Tables Vol C Tables 4.2.6.8 and 6.1.1.4'
_computing_data_collection      '<i>CrystalClear</i> (Rigaku, 2008)'
_computing_cell_refinement      '<i>CrystalClear</i> (Rigaku, 2008)'
_computing_data_reduction      '<i>CrystalClear</i> (Rigaku, 2008)'
_computing_structure_solution      '<i>SIR97</i> (Altomare  <i>et al.</i>, 1999)'
_computing_structure_refinement
        '<i>SHELXL97</i> (Sheldrick, 2008) within <i>WinGX</i> (Farrugia, 2012)'
```

Mitchell, L. A. & Holliday, B. J. (2014). 6-Bromo-*N*-(6-bromopyridin-2-yl)-*N*-[4-(2,3-dihydro-thieno[3,4-*b*][1,4]dioxin-5-yl)phenyl]pyridin-2-amine. *Acta Cryst.* E**70**, o797.





Refinement

Refinement on F^2	Secondary atom site location: difference Fourier map
Least-squares matrix: full	Hydrogen site location: inferred from neighbouring sites
$R[F^2 > 2\sigma(F^2)] = 0.051$	H-atom parameters constrained
$wR(F^2) = 0.128$	$w = 1/[\sigma^2(F_o^2) + (0.0638P)^2]$ where $P = (F_o^2 + 2F_c^2)/3$
$S = 1.00$	$(\Delta/\sigma)_{max} = 0.001$
3521 reflections	$\Delta\rho_{max} = 1.06$ e Å^{-3}
271 parameters	$\Delta\rho_{min} = -0.83$ e Å^{-3}
0 restraints	Extinction correction: none
? constraints	Extinction coefficient: ?
Primary atom site location: structure-invariant direct methods	

Refinement of F^2 against ALL reflections. The weighted *R*-factor *wR* and goodness of fit *S* are based on F^2, conventional *R*-factors *R* are based on *F*, with *F* set to zero for negative F^2. The threshold expression of $F^2 > \sigma(F^2)$ is used only for calculating *R*-factors(gt) *etc.* and is not relevant to the choice of reflections for refinement. *R*-factors based on F^2 are statistically about twice as large as those based on *F*, and *R*-factors based on ALL data will be even larger.

Fractional atomic coordinates and isotropic or equivalent isotropic displacement parameters (Å^2)

	x	*y*	*z*	U_{iso}*/U_{eq}
C1	0.4149 (12)	0.8677 (5)	0.3436 (3)	0.0254 (13)
H1	0.4019	0.9100	0.2958	0.030*
C2	0.5672 (11)	0.7710 (5)	0.3635 (3)	0.0223 (12)
C3	0.7454 (12)	0.6007 (5)	0.3458 (3)	0.0244 (13)
H3A	0.8736	0.5675	0.3141	0.029*
H3B	0.5491	0.5663	0.3489	0.029*

Rendered using *Mercury* [Macrae, C. F. *et al.* (2009). *J. Appl. Cryst.* **41**, 466-470]

Rendered using *publcif* [Westrip, S. P. (2010). *J. Appl. Cryst.* **43**, 920-935]

The irrelevance of format

Gao, Y. R., Feng, N., Chen, T., Li, D. F. & Bi, L. J. (2015). Structure of the MarR family protein Rv0880 from *Mycobacterium tuberculosis*. *Acta Cryst.* F**71**, 741-745.

PDB structure 4YIF

```
HEADER    DNA BINDING PROTEIN                     02-MAR-15   4YIF
TITLE     CRYSTAL STRUCTURE OF RV0880
COMPND    MOL_ID: 1;
COMPND   2 MOLECULE: UNCHARACTERIZED HTH-TYPE TRANSCRIPTIONAL REGULATOR RV0880;
COMPND   3 CHAIN: A, B, C, D, E, F;
COMPND   4 FRAGMENT: UNP RESIDUES 2-143;
COMPND   5 ENGINEERED: YES
SOURCE    MOL_ID: 1;
SOURCE   2 ORGANISM_SCIENTIFIC: MYCOBACTERIUM TUBERCULOSIS;
SOURCE   3 ORGANISM_TAXID: 83332;
SOURCE   4 STRAIN: ATCC 25618 / H37RV;
SOURCE   5 GENE: RV0880, MTCY31.08;
SOURCE   6 EXPRESSION_SYSTEM: ESCHERICHIA COLI BL21(DE3);
SOURCE   7 EXPRESSION_SYSTEM_TAXID: 469008;
SOURCE   8 EXPRESSION_SYSTEM_STRAIN: BL21(DE3);
SOURCE   9 EXPRESSION_SYSTEM_VECTOR_TYPE: PLASMID;
SOURCE  10 EXPRESSION_SYSTEM_PLASMID: PET-28A
KEYWDS    MARR FAMILY, DNA BINDING PROTEIN, REPRESSOR, MYCOBACTERIUM
KEYWDS   2 TUBERCULOSIS
EXPDTA    X-RAY DIFFRACTION
AUTHOR    Y.R.GAO,N.FENG,D.F.LI,L.J.BI
REVDAT   1   10-JUN-15 4YIF    0
JRNL        AUTH    Y.R.GAO,N.FENG,T.CHEN,D.F.LI,L.J.BI
JRNL        TITL    STRUCTURE OF THE MARR FAMILY PROTEIN RV0880 FROM
JRNL        TITL 2 MYCOBACTERIUM TUBERCULOSIS
JRNL        REF     ACTA CRYSTALLOGR.,SECT.F      V.  71   741 2015
JRNL        REFN                   ESSN 2053-230X
JRNL        DOI     10.1107/S2053230X15007281
REMARK   2
REMARK   2 RESOLUTION.    2.00 ANGSTROMS.
REMARK   3
REMARK   3 REFINEMENT.
REMARK   3   PROGRAM     : PHENIX (PHENIX.REFINE: 1.8.4_1496)
REMARK   3   AUTHORS     : PAUL ADAMS,PAVEL AFONINE,VINCENT CHEN,IAN
REMARK   3                : DAVIS,KRESHNA GOPAL,RALF GROSSE-KUNSTLEVE,
REMARK   3                : LI-WEI HUNG,ROBERT IMMORMINO,TOM IOERGER,
REMARK   3                : AIRLIE MCCOY,ERIK MCKEE,NIGEL MORIARTY,
REMARK   3                : REETAL PAI,RANDY READ,JANE RICHARDSON,
REMARK   3                : DAVID RICHARDSON,TOD ROMO,JIM SACCHETTINI,
REMARK   3                : NICHOLAS SAUTER,JACOB SMITH,LAURENT
REMARK   3                : STORONI,TOM TERWILLIGER,PETER ZWART
REMARK   3
REMARK   3    REFINEMENT TARGET : ML
REMARK   3
REMARK   3  DATA USED IN REFINEMENT.
REMARK   3   RESOLUTION RANGE HIGH (ANGSTROMS) : 2.00
REMARK   3   RESOLUTION RANGE LOW  (ANGSTROMS) : 37.11
REMARK   3   MIN(FOBS/SIGMA_FOBS)              : 1.960
REMARK   3   COMPLETENESS FOR RANGE        (%) : 94.0
REMARK   3   NUMBER OF REFLECTIONS             : 54971
```

```
data_4YIF
#
_entry.id   4YIF
#
_audit_conform.dict_name       mmcif_pdbx.dic
_audit_conform.dict_version    4.054
_audit_conform.dict_location   http://mmcif.pdb.org/dictionaries/ascii/mmcif_pdbx.dic
#
loop_
_database_2.database_id
_database_2.database_code
PDB   4YIF
WWPDB D_1000207521
#
_database_PDB_rev.num            1
_database_PDB_rev.date           2015-06-10
_database_PDB_rev.date_original  2015-03-02
_database_PDB_rev.status         ?
_database_PDB_rev.replaces       4YIF
_database_PDB_rev.mod_type       0
#
_pdbx_database_status.status_code                     REL
_pdbx_database_status.status_code_sf                  REL
_pdbx_database_status.status_code_mr                  ?
_pdbx_database_status.entry_id                        4YIF
_pdbx_database_status.date_begin_release_preparation  .
_pdbx_database_status.SG_entry                        N
_pdbx_database_status.deposit_site                    RCSB
_pdbx_database_status.process_site                    RCSB
_pdbx_database_status.methods_development_category    ?
_pdbx_database_status.pdb_format_compatible           Y
#
loop_
_audit_author.address
_audit_author.name
_audit_author.pdbx_ordinal
? 'Gao, Y.R.' 1
? 'Feng, N.'  2
? 'Li, D.F.'  3
? 'Bi, L.J.'  4
#
_citation.abstract                  ?
_citation.abstract_id_CAS           ?
_citation.book_id_ISBN              ?
_citation.book_publisher            ?
_citation.book_publisher_city       ?
_citation.book_title                ?
_citation.coordinate_linkage        ?
_citation.country                   US
_citation.database_id_Medline       ?
_citation.details                   ?
_citation.id                        primary
_citation.journal_abbrev            'Acta Crystallogr.,Sect.F'
```

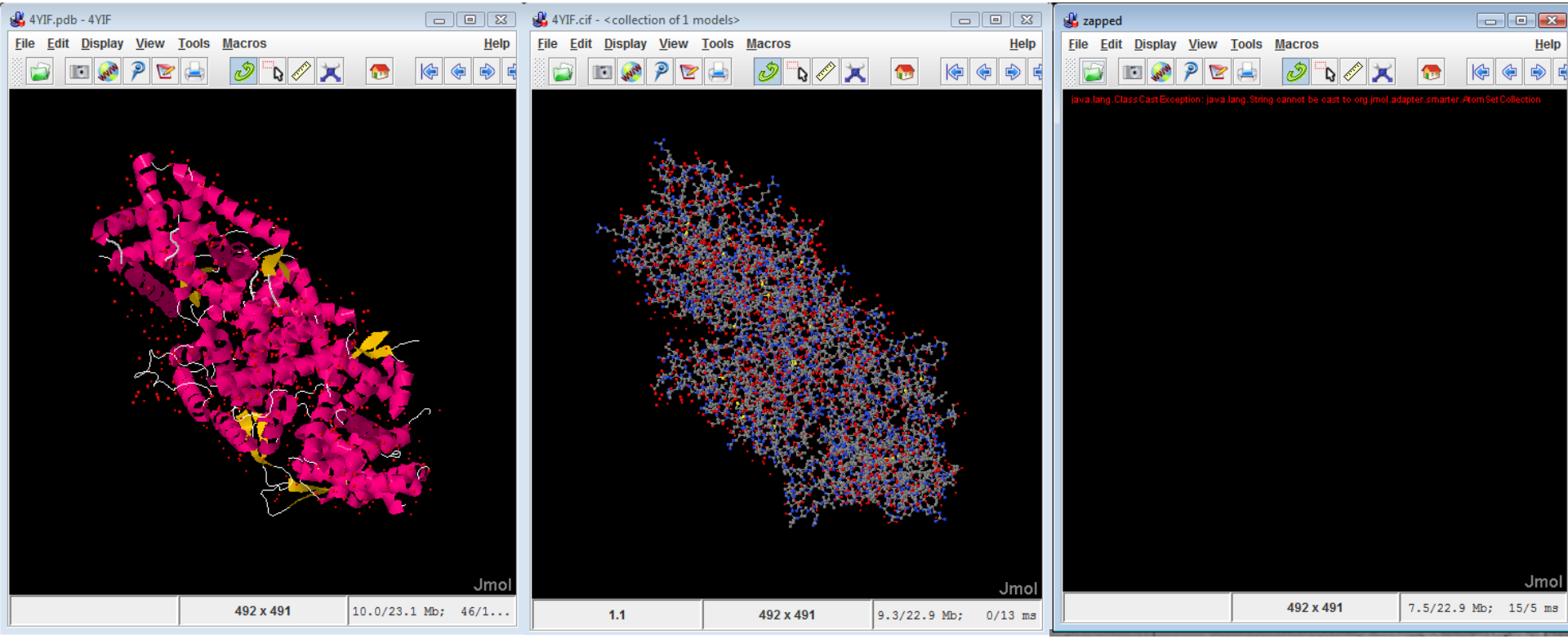
```
<?xml version="1.0" encoding="UTF-8" ?>

<PDBx:datablock datablockName="4YIF"
   xmlns:PDBx="http://pdbml.pdb.org/schema/pdbx-v42.xsd"
   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
   xsi:schemaLocation="http://pdbml.pdb.org/schema/pdbx-v42.xsd pdbx-

   <PDBx:atom_siteCategory>
      <PDBx:atom_site id="1">
         <PDBx:B_iso_or_equiv>61.04</PDBx:B_iso_or_equiv>
         <PDBx:Cartn_x>30.198</PDBx:Cartn_x>
         <PDBx:Cartn_y>-17.284</PDBx:Cartn_y>
         <PDBx:Cartn_z>47.960</PDBx:Cartn_z>
         <PDBx:auth_asym_id>A</PDBx:auth_asym_id>
         <PDBx:auth_atom_id>N</PDBx:auth_atom_id>
         <PDBx:auth_comp_id>SER</PDBx:auth_comp_id>
         <PDBx:auth_seq_id>4</PDBx:auth_seq_id>
         <PDBx:group_PDB>ATOM</PDBx:group_PDB>
         <PDBx:label_alt_id xsi:nil="true" />
         <PDBx:label_asym_id>A</PDBx:label_asym_id>
         <PDBx:label_atom_id>N</PDBx:label_atom_id>
         <PDBx:label_comp_id>SER</PDBx:label_comp_id>
         <PDBx:label_entity_id>1</PDBx:label_entity_id>
         <PDBx:label_seq_id>25</PDBx:label_seq_id>
         <PDBx:occupancy>1.00</PDBx:occupancy>
         <PDBx:pdbx_PDB_model_num>1</PDBx:pdbx_PDB_model_num>
         <PDBx:type_symbol>N</PDBx:type_symbol>
      </PDBx:atom_site>
      <PDBx:atom_site id="2">
         <PDBx:B_iso_or_equiv>60.75</PDBx:B_iso_or_equiv>
         <PDBx:Cartn_x>30.886</PDBx:Cartn_x>
         <PDBx:Cartn_y>-17.907</PDBx:Cartn_y>
         <PDBx:Cartn_z>49.084</PDBx:Cartn_z>
         <PDBx:auth_asym_id>A</PDBx:auth_asym_id>
         <PDBx:auth_atom_id>CA</PDBx:auth_atom_id>
         <PDBx:auth_comp_id>SER</PDBx:auth_comp_id>
         <PDBx:auth_seq_id>4</PDBx:auth_seq_id>
         <PDBx:group_PDB>ATOM</PDBx:group_PDB>
         <PDBx:label_alt_id xsi:nil="true" />
         <PDBx:label_asym_id>A</PDBx:label_asym_id>
         <PDBx:label_atom_id>CA</PDBx:label_atom_id>
         <PDBx:label_comp_id>SER</PDBx:label_comp_id>
         <PDBx:label_entity_id>1</PDBx:label_entity_id>
         <PDBx:label_seq_id>25</PDBx:label_seq_id>
         <PDBx:occupancy>1.00</PDBx:occupancy>
         <PDBx:pdbx_PDB_model_num>1</PDBx:pdbx_PDB_model_num>
         <PDBx:type_symbol>C</PDBx:type_symbol>
      </PDBx:atom_site>
      <PDBx:atom_site id="3">
         <PDBx:B_iso_or_equiv>57.54</PDBx:B_iso_or_equiv>
         <PDBx:Cartn_x>29.948</PDBx:Cartn_x>
         <PDBx:Cartn_y>-18.061</PDBx:Cartn_y>
```

Rendered using *Jmol* [Hanson, R. M. (2010). *J. Appl. Cryst.* **43**, 1250-1260]

The irrelevance of format

Gao, Y. R., Feng, N., Chen, T., Li, D. F. & Bi, L. J. (2015). Structure of the MarR family protein Rv0880 from *Mycobacterium tuberculosis*. *Acta Cryst.* F**71**, 741-745.

PDB structure 4YIF



Rendered using *Jmol* [Hanson, R. M. (2010). *J. Appl. Cryst.* **43**, 1250-1260]

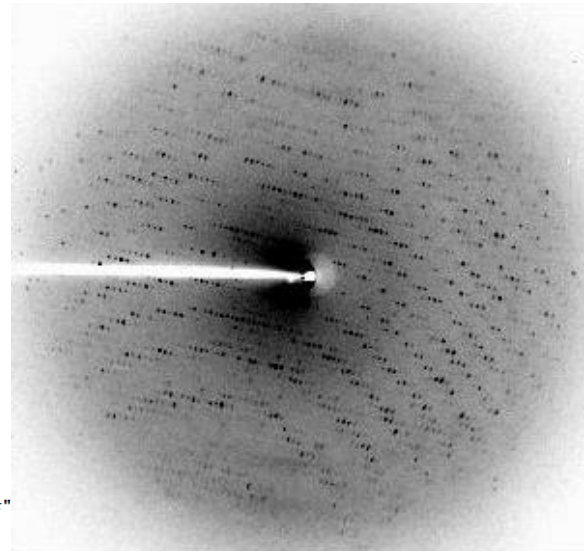
The irrelevance of format

Example MAR image file example.mar2300

```
loop_
_array_element_size.array_id
_array_element_size.index
_array_element_size.size
 image_1 1 150.0e-6
 image_1 2 150.0e-6

loop_
_array_intensities.array_id
_array_intensities.binary_id
_array_intensities.linearity
_array_intensities.gain
_array_intensities.overload
_array_intensities.undefined
 image_1 1 linear 1.55 240000 0

loop_
_array_data.array_id
_array_data.binary_id
_array_data.data
 image_1 1
;
--CIF-BINARY-FORMAT-SECTION--
Content-Type: application/octet-stream;
     conversions="x-CBF_PACKED"
Content-Transfer-Encoding: BASE64
X-Binary-Size: 3745758
X-Binary-ID: 1
X-Binary-Element-Type: "signed 32-bit integer"
Content-MD5: 1zsJjWPfol2GYl2V+QSXrw==

ELhQAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAADHcRzHcRxGQQwCZsGuAKUFAIhS93U8
/91rMvpiEXw1pwoceMIBYHj78x7u9nszkeh7qm3XK6jk/Aa4x3Ecx3Ecx3Ecx3EcBzEEgApW
/y8xGar1BaqZXkcCow74Aw77fp8W5Sf2vP6O6A/SD8ZnixLf4/WMOzCgEAhqVnnv3wsk8oO9
EFa5G/3Gfq94GwLjHNE+gd8ndgf1foI+GN2LQIAneVRf9rXyCkwIyc/y/ILuHsdxHMdxHMdx
HMdxGAURAGraeHhJh6ccYbjVMKH5Xxjq3wx6VNQ4sPPCAxNNMrb/BEsADjdtYzBLtS+VJMtJ
+YsPmIJ/MPhvQe2Guq6xK7XFt74PXl+Oqivhz8A3zpxUjbvWDz/r85c8f6aYofNL3WwXB8vq
nSlYClkRjj4jdn8zxNvzx04wyrbs8xAQLzw6DIojGAAAHMdxHMdxHMdxGASgvsmn+2/6
gNiVYecjJWyI5LsifRdreREU9oLAyenEjqvQirPfh3ELks1TMc0BxyPcePTr/VtTStFvQJ0H
hz4x5hp1/Qfmm0McDl/JB368u7B41P/4wZ2/eCTaY1i8f/fbJQHBfxEXOM4XR1vDrAAY5Uq+
Qx3JIXwCV2a7fkqIoD7Kq0WtqHI7hNzPCkznmOL4KUD1gyI+m9hAHkcx3Ecx3Ecx3Ecx1EQ
QwCsWQXyHOcl8/yb/RYQnSDmJMAXV4nN6UA87TdTZwHiEkhvk3aL3q/meB+FBvhUAMDzm58Q
2plbK6W5Uh90sCq2J5C/Ia9Fs49rgKYRMH7A2aADfYpsNtqSHhfejjx9ecQ5G6I3v86uF7j9
HIHijajmFloLvfxL8gR8giAicwchtmFCX3v4+L098njzBXrBO3DqZsAOhvzGgFnnNxsmHHFK
iigogWysj14gxMC8az6O4ziO4ziO4ygE1UtBvF9JOiZH8AHDLreQKwGAnE6gi7jGVhEp
XpW/SMx9UBxRDQ+gNoEfbr2v0PER4IMj+8KIW8OmSpKRj4qEd5lcGPMTEbzObtAnDzzH5YYN
E5sgT466JIi4h3EEYOpG6pXxn/JmGMT+0h7fr4Pg+Mq8kOlyvN4TQXCjYK2XeCRQ0bWC6Fp
ZQhwgzc4PrLtwp2I40yQfED67a8LBn//hESglgxL/76HYCULQaLv5wLfZS+5cPI7SvCAGrU8
lnYA9T2O4ziO4ziO4ziO4yDSt/Fg5HT5k/UhUdlEP8r5YOHPSzjxwHcVCUcdfg+AHtxD3rJ+
XyBsjPiybY3vjj1o7o9kBwnIQF9f2xkGv6fuZgAEB/kjwR/09XfDQNBh+Z/mBN7Nj+wx+2em
QU+Q3n33/35A8Ok2PFi+oHwfg/dCohbF1/9UnS2olB3xRkJQhzcaAe4OAr3Nm9miFUYXbHx8
SOCU715fmEROiwPAN4O2oX/pg8ngIVWpeLwfDDH2f/7g2rvYkUP3l5JApPqeR0MIbdB30vMt
4XNi2H9v4j0q7E1AX4/jOI7jOI7jOI7jMApiEIBTR5Ew3HdEgJ4uyQkFR1yPz1sIAfpgdK5H
BGu5JLMXgBN1CfnvOOADDXJD8m38K+WFZnwnFCyPOP7e+KsREDnEvtYzFnUM53rnyBdg/8fP
9w8K9oT88QhJqne5u3/w6xUIAGJe0ff3XhAm7j7kHXnH+aOeQQ8nl8kD3ymhmHFeuDgaH9np
ZxaEZJed4G5MyUZ4mKB+fqrtoacNJFWOfrl7P1oplEJ+YmVeZ09mOiAuyBUY/tStYZaMf+QZ
RaD7f8kdoJ95Q0ekM6Crwbh7/UeU3V1B2iPQG+oh8o/jOI7jOI7jOI7jMIpBULUg/Os5YRc3
```

imgCIF format

```
loop_
_array_element_size.array_id
_array_element_size.index
_array_element_size.size
 image_1 1 150.0e-6
 image_1 2 150.0e-6

loop_
_array_intensities.array_id
_array_intensities.binary_id
_array_intensities.linearity
_array_intensities.gain
_array_intensities.overload
_array_intensities.undefined
 image_1 1 linear 1.55 240000 0

loop_
_array_data.array_id
_array_data.binary_id
_array_data.data
 image_1 1
;
--CIF-BINARY-FORMAT-SECTION--
Content-Type: application/octet-stream;
     conversions="x-CBF_PACKED"
Content-Transfer-Encoding: BINARY
X-Binary-Size: 3745758
X-Binary-ID: 1
X-Binary-Element-Type: "signed 32-bit integer"
Content-MD5: 1zsJjWPfol2GYl2V+QSXrw==
```



CBF format

CIF as a data model

Format innovation	Data description evolution
1991: CIF	Initial decoupling of syntax/semantics
1995:	DDL1: machine-readable semantics
1996: mmCIF	Relational data model
2000: imgCIF	(Handling of binary data)
2013:	DDLm: dynamic (methods) data model
2016: CIF2.0	(Extended character set, data types)

CIF as a data model

- **DDL** (dictionary definition language) is the mechanism in the CIF world for describing relationships between defined data objects
- **DDL** (data description language) is the mechanism in the relational database world for characterising relationships between items in the database
- The two perform very similar functions

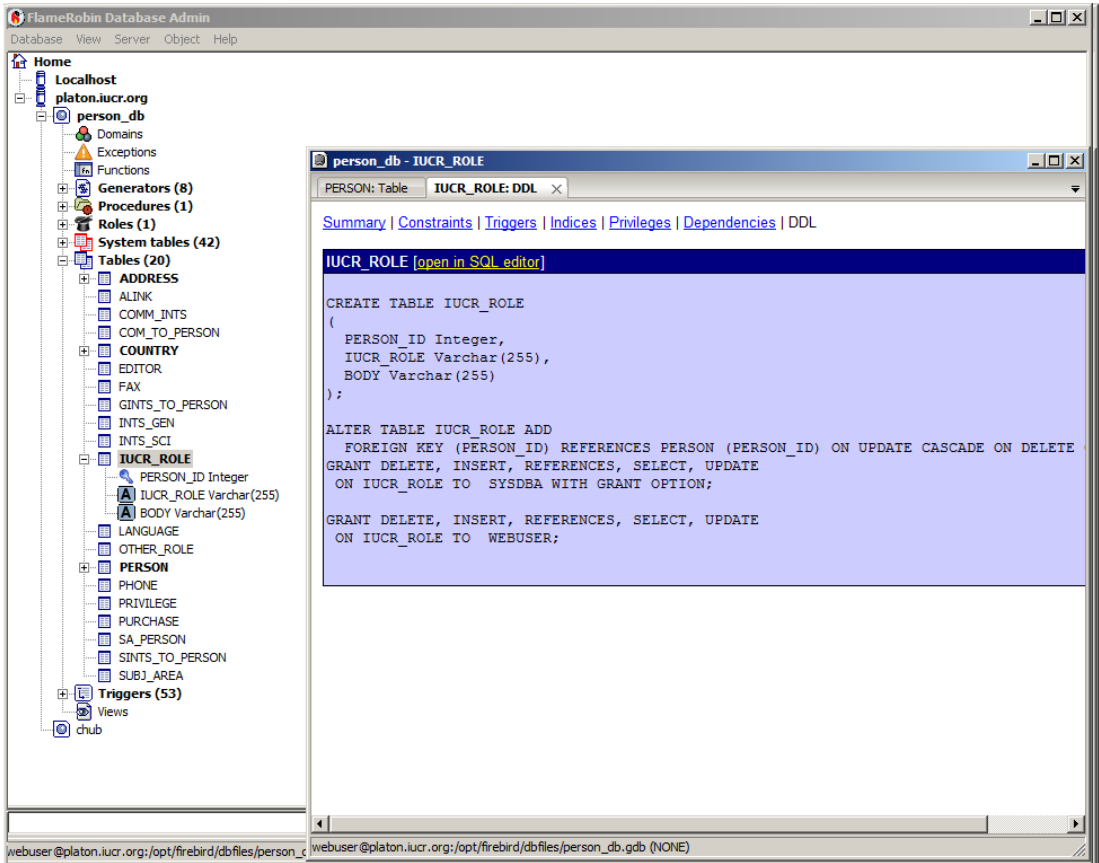
CIF as a relational data model

World Database of Crystallographers (WDC)

WDC initially implemented as a STAR database

Updated as an InterBase RDBMS

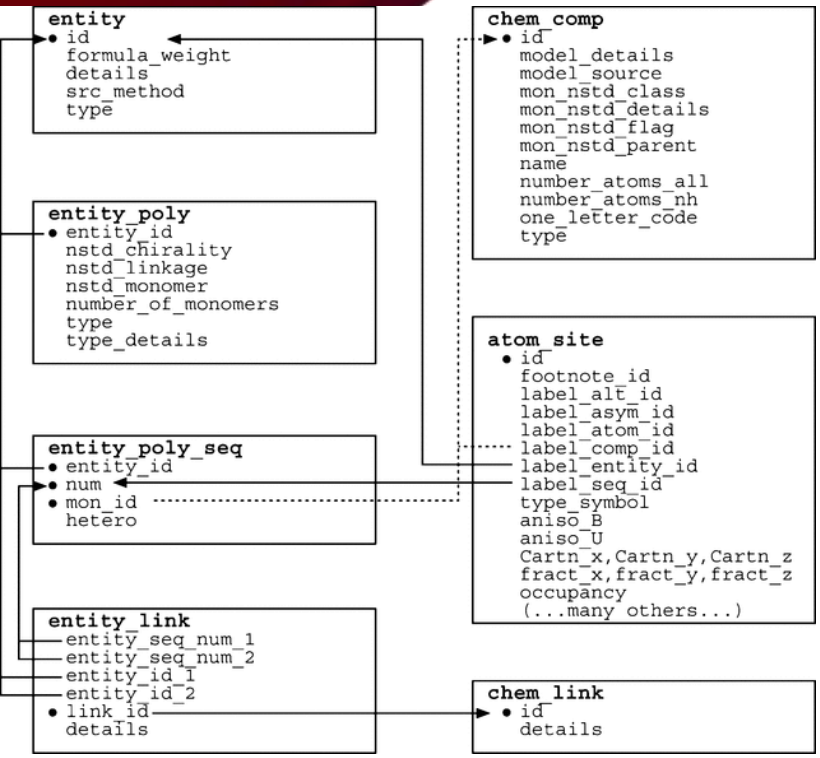
```
####################
## WDC_IUCR_ROLE ##
####################

save_WDC_IUCR_ROLE
    _category.description
;             The WDC_IUCR_ROLE records the holders of offices or named
              positions within the International Union of Crystallography,
              its Committees and Commissions.
;
    _category.id                   wdc_iucr_role
    _category.mandatory_code       no
    _category_key.name            '_wdc_iucr_role.id'
     loop_
    _category_group.id             'wdc_group'
                                   'person_group'
     loop_
    _category_examples.detail
    _category_examples.case
# - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
;
    Example 1 - based on IUCr staff entries for the World Database 11th edition
;
;
     loop_
    _wdc_iucr_role.person_id
    _wdc_iucr_role.position
    _wdc_iucr_role.body

    002456   President   .
    006282   Member   'Promotions Committee'
    006189   'Coordinating Secretary'
                      'Committee for the Maintenance of the CIF Standard'
    006189   Member   'Electronic Publishing Committee'
    006189   Editor
    ;    International Tables for Crystallography Vol. G: Crystallographic
         Information
    ;
;
# - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
     save_

save__wdc_iucr_role.body
    _item_description.description
;             The name of an IUCr Committee, Commission or other body.
;
    _item.name                    '_wdc_iucr_role.body'
    _item.category_id               wdc_iucr_role
    _item.mandatory_code            no
    _item_type.code                 char
     loop_
    _item_enumeration.value         'Executive Committee'
                                    'Finance Committee'
                                    'Sub-committee on the Union Calendar'
;                                    Sub-committee on Electronic Publishing,
                                     Dissemination and Storage of Information
;
```



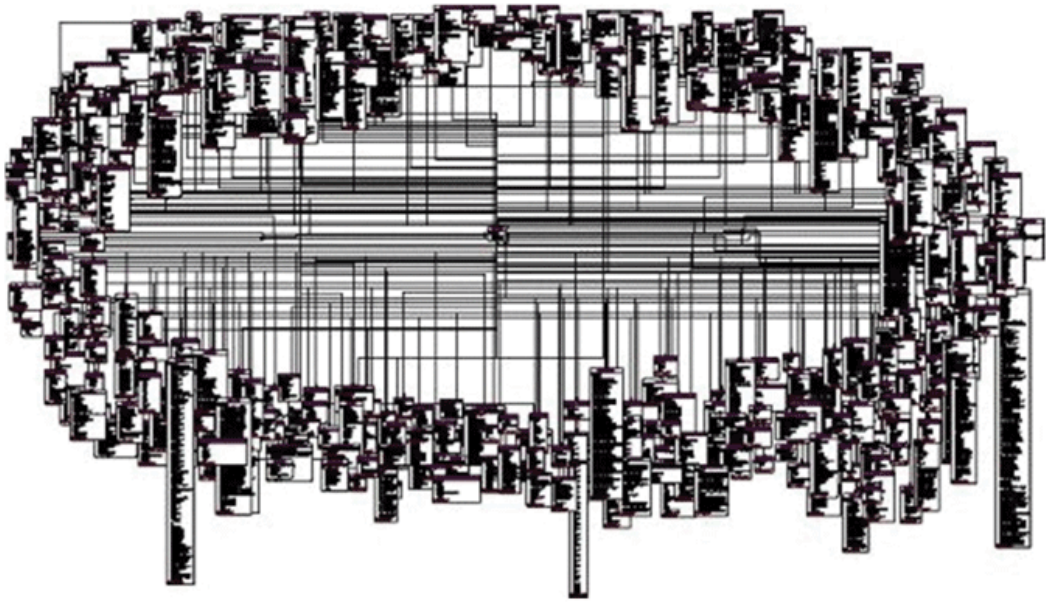
CIF as a relational data model

Protein Data Bank

PDB schema based on the mmCIF dictionary



The family of categories (in the mmCIF dictionary) used to describe polymer chemical entities; *from* Fitzgerald, P. M. D. et al. (2006). Classification and use of macromolecular data. In *International Tables For Crystallography Vol. G: Definition and exchange of crystallographic data*. Dordrecht: Springer

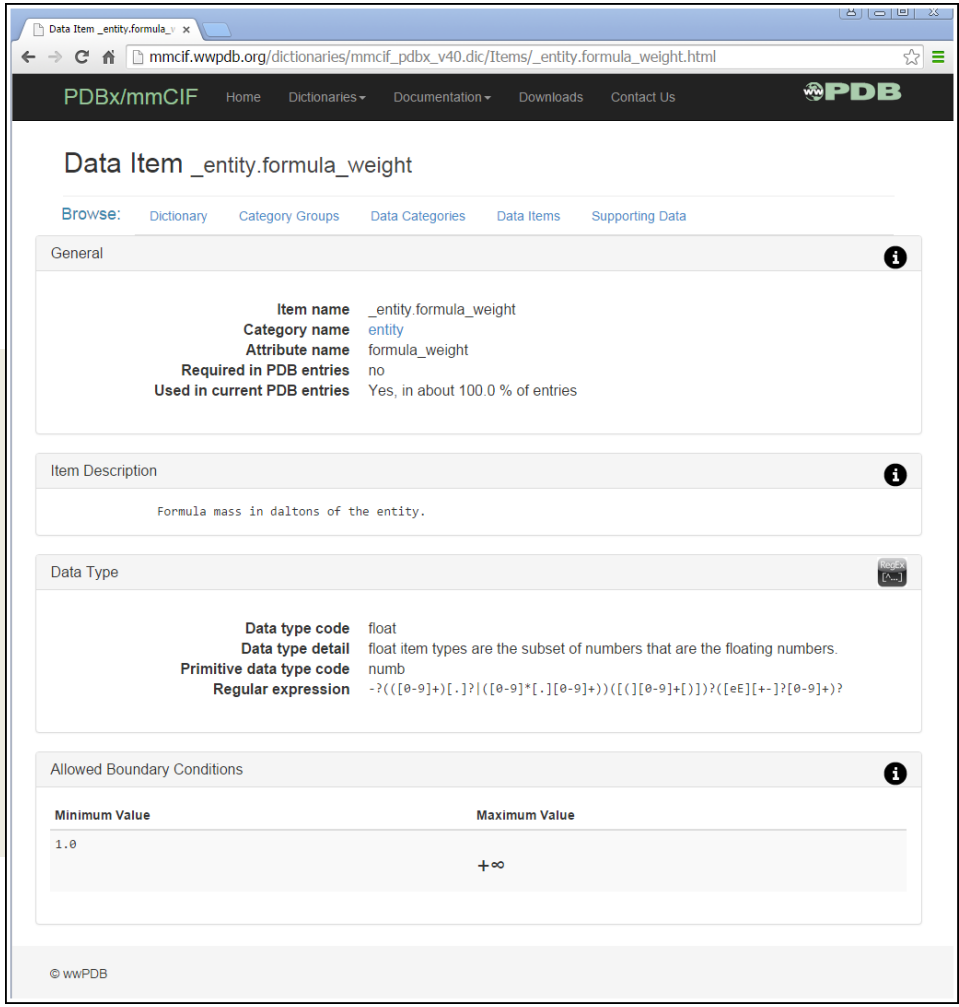


Global schema map of the entire PDB relational database; *from* Schierz, A. C., Soldatova, L. N. & King, R. D. (2007). *Nature Biotechnology*, **25**, 437-442

Structure of a CIF dictionary definition

```
save__entity.formula_weight
    _item_description.description
;               Formula mass in daltons of the entity.
;
    _item.name                    '_entity.formula_weight'
    _item.category_id               entity
    _item.mandatory_code            no
    loop_
    _item_range.maximum
    _item_range.minimum              .    1.0
                                    1.0   1.0
    _item_type.code                 float
     save_
```

As it appears in the mmCIF dictionary (easy for computers to read)



As it appears on the PDB website (easier for people to read?)

Structure of a CIF dictionary definition

```
data_cell_volume
    _name                       '_cell_volume'
    _category                     cell
    _type                         numb
    _type_conditions              esd
    _enumeration_range            0.0:
    _units                        A^3^
    _units_detail               'cubic angstroms'
    _definition
;               Cell volume V in angstroms cubed.

                V = a b c [1 - cos^2^(alpha) - cos^2^(beta) - cos^2^(gamma)
                              + 2 cos(alpha) cos(beta) cos(gamma) ] ^1/2^

                a     = _cell_length_a
                b     = _cell_length_b
                c     = _cell_length_c
                alpha = _cell_angle_alpha
                beta  = _cell_angle_beta
                gamma = _cell_angle_gamma
;
```

A more complex example. The relationships with other data items are described – but not in a form that a computer program can do anything with.

Structure of a CIF dictionary definition

```
save_cell.volume
    _definition.id              '_cell.volume'
     loop_
    _alias.definition_id        '_cell_volume'
    _definition.update           2013-03-07
    _description.text
;
     Volume of the crystal unit cell.
;
    _name.category_id            cell
    _name.object_id              volume
    _type.purpose                Measurand
    _type.source                 Derived
    _type.container              Single
    _type.contents               Real
    _enumeration.range           0.0:
    _units.code                  angstrom_cubed
     loop_
    _method.purpose
    _method.expression
     Evaluation
;
      With v  as  cell_vector

      _cell.volume =  v.a * ( v.b ^ v.c )
;
     save_
```


- Spadaccini, N. & Hall, S. R. (2012). Extensions to the STAR File Syntax. *J. Chem. Inf. Model.* **52**, 1901-1906.
- Spadaccini, N. & Hall, S. R. (2012). DDLm: A New Dictionary Definition Language. *J. Chem. Inf. Model*. **52**, 1907-1916.
- Spadaccini, N., Castleden, I. R., Du Boulay, D. & Hall, S. R. (2012). dREL: A Relational Expression Language for Dictionary Methods. *J. Chem. Inf. Model.* **52**, 1917-1925.

But in the new DDLm, **methods** are introduced that can be interpreted and executed by computer programs: allows validation of data *or* retrieval of absent data provided more primitive data values are present.

Structure of a CIF dictionary definition

```
save_refln.F_complex
                                    . . .

     loop_
    _method.purpose
    _method.expression
     Evaluation
;
      With r as refln            # reflection packet in scope
     fc  =  Complex (0., 0.)
     h   =  r.hkl
     Loop a as atom_site {       # Summation over atom sites
          x  = a.fract_xyz
          f  = (r.form_factor_table[a.type_symbol]*
             a.symmetry_multiplicity*a.occupancy)
          B  = a.matrix_beta
          Loop s as symmetry_equiv { # Summation over symmetry
               fc += f * Exp(-h*s.R*B*s.R*h) *
                ExpImag(TwoPi*(h*(s.R*x+s.T)))
          }
     }
     refln.F_complex = fc        # evaluate defined item
;
     save_
```

Another, more complicated example: structure factor $F(h) = \sum_{j}^{asymmetric} f_j(s) \sum_{k}^{symmetry} e^{-B_j(s_k)} e^{-2\pi i h \bullet R_k r_j}$

Existing CIF dictionaries

'Official': managed by COMCIFS

Dictionary name	DDL	Purpose
cif_core.dic	1.4.1	crystallographic core
cif_core_restraints.dic	1.4.1	restraints or constraints in least-squares refinement of crystal structures
cif_pd.dic	1.4.1	powder diffraction
cif_ms.dic	1.4.1	incommensurately modulated crystal structures
cif_rho.dic	1.4.1	results of electron density studies
cif_twinning.dic	1.4.1	crystallographic twinning
cif_img.dic	2.1.3	diffraction images
cif_sym.dic	2.1.3	crystallographic symmetry
meta-dictionaries		
ddl_core.dic	1.4.1	dictionary definition language
ddl_core_2.1.6.dic	2.1.6	relational dictionary definition language
cifdic.register	1.4	register of dictionaries distributed by IUCr

Existing CIF dictionaries

‘Official’: managed by wwPDB

Dictionary name	DDL	Purpose
PDBx/mmCIF	2.1.15	PDB Exchange Dictionary supporting data files in the PDB archive
NMR-STAR.dic	2.1.*x*	NMR structures in the BioMagResBank archive
mmcif_nef.dic	2.1.*x*	NMR exchange format
mmcif_sas.dic	2.1.*x*	small-angle scattering
mmcif_em.dic	2.1.*x*	3D electron microscopy data
mmcif_img.dic	2.1.*x*	PDB maintained version of diffraction image data description
mmcif_sym.dic	2.1.*x*	PDB maintained version of crystallographic symmetry
mmcif_biosync.dic	2.1.*x*	features of synchrotron facilities and beamlines
mmcif_mdb.dic	2.1.*x*	Homology models and homology modelling methodologies
meta-dictionaries		
mmcif_ddl.dic	2.1.15	relational dictionary definition language

Existing CIF dictionaries

‘Private’: distributed by IUCr

Dictionary name	DDL	Purpose
cif_iucr.dic	1.4.1	private data items used by the IUCr in journal publishing
cif_ccdc.dic	1.4.1	private data items used by the Cambridge Crystallographic Data Centre
meta-dictionaries		
cif_compat.dic	1.4.1	legacy CIF Dictionary of deprecated terms

Putative CIF dictionaries

Reserved namespaces registered with IUCr

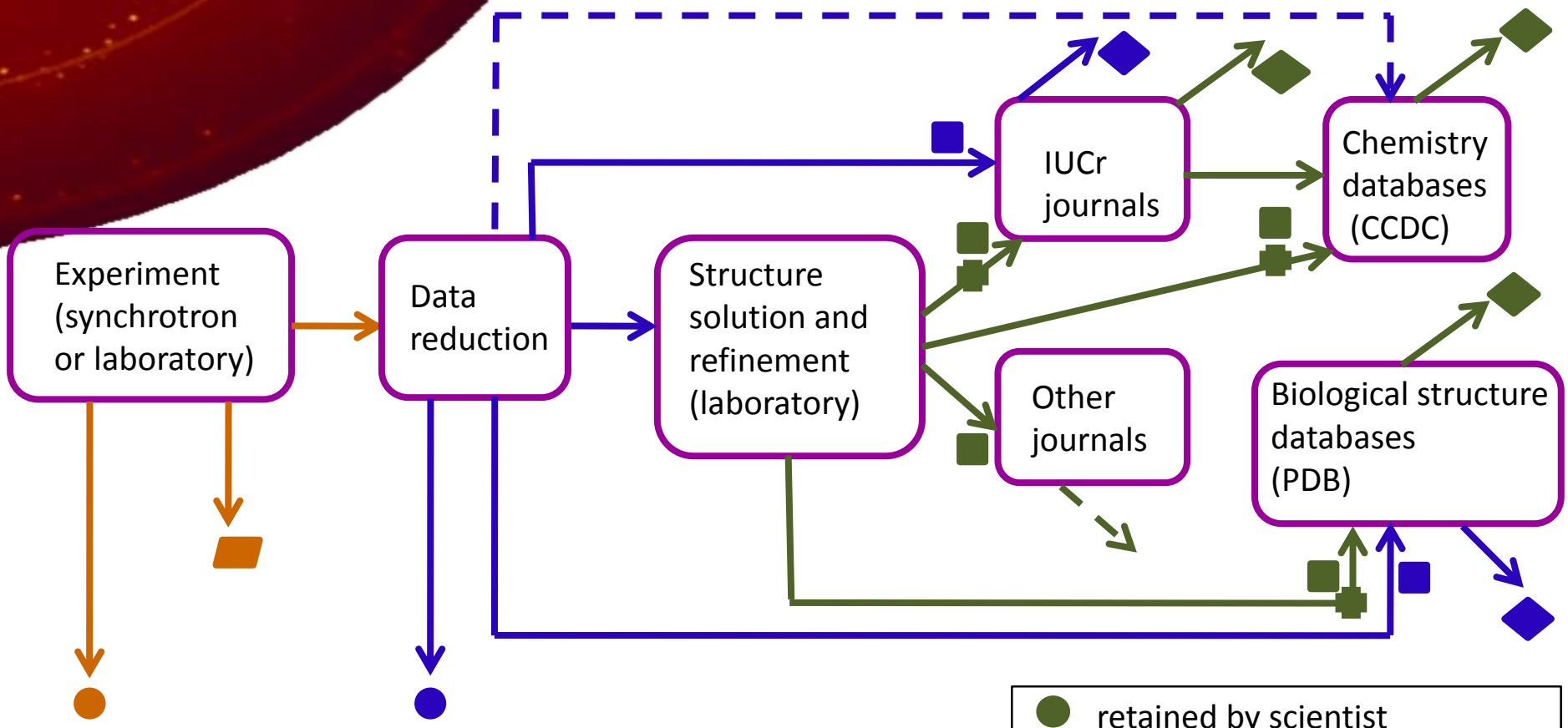
Prefix	Purpose/owner	Prefix	Purpose/owner	Prefix	Purpose/owner
B+S, BplusS, pdb2cif	H. J. Bernstein	crystmol	CrystMol software	msd	EBI Molecular Structure Database Group
CCP4	CCP4 software suite	csd	Cambridge Structural Database	ndb	Nucleic Acids Database
H5, NX	support of HDF5 and NeXus integration	dft	density functional theory calculations	nottingham	University of Nottingham
NIEHS	Natl Inst. of Environmental Health Sciences	ebi	EBI macromolecular harvest deposition file	oxford	U. Oxford *CRYSTALS* software
SSAD	Sulfur SAD Database	edchem	Edinburgh University Chemistry	parvati	PARVATI validation server
acihd	ACI Heidelberg	gsas	GSAS powder refinement system	pdb, pdbx	Protein Data Bank
amcsd	*American MIneralogist* Crystal Structure Database	gsk	GlaxoSmithKline	phenix	*PHENIX* software suite
anbf	Australian National Beamline Facility	iims	Integration of 3D Electron Microscopy with X-ray and NMR methods	publcif	*publCIF* editor software
asd	Active Site Database	itqb	Inst. Tecnologia Quimica e Biologica da Univer. Nova de Lisboa, Oeiras, Portugal	rayonix	Rayonix (Mar USA) instruments
bruker	Bruker AXS	iucr	IUCr journal use	rcsb	Research Collaboratory for Structural Bioinformatics
ccdc	Cambridge Crystallographic Data Centre	mdb	database of model structures of biological molecules	shelx	*SHELXL* solution and refinement programs
cgraph	Oxford Cryosystems *Crystallographica* package	montpellier	University of Montpellier	tcod	Theoretical Crystallography Open Database
cod	Crystallography Open Database	mpod, prop	Material Properties Open Database	vrf	checkCIF validation reply form entries
wdc	entries in *World Directory of Crystallographers*	xtal	*Xtal* program system		

Potential CIF dictionaries

Proposed, under development or abandoned

Dictionary name	Purpose
magCIF	magnetic structures; under development by Commission on Magnetic Structures
Xasformat, xasCIF	efforts to define a data standard for X-ray absorption spectroscopy; input from Commission on XAFS
MPOD	Materials Properties Open Database ontology
Lattice topology	*e.g.* following descriptive model of *TOPOS* software

Data flow in crystallography



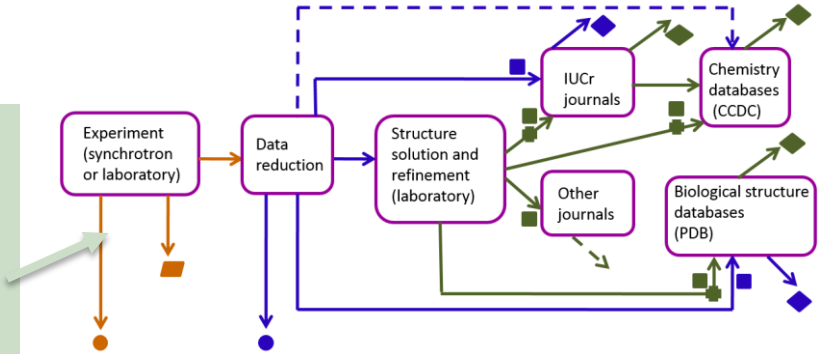
Raw experimental data (e.g. diffraction images)
Reduced/processed data (e.g. structure factors)
Derived data (e.g. coordinates, a.d.p.s)

- retained by scientist
- archived at facility (~6 months)
- deposited
- published/disseminated
- validated

Data flow beyond crystallography

Scientific resource

- Technique
- Probe (X-ray, neutron, electron)
- Equipment
- Beamline
- Funding
- Time
- Scale



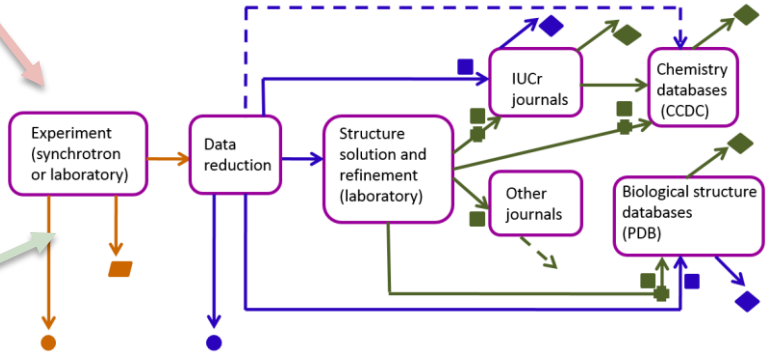
Data flow beyond crystallography

Scientific context for the experiment

- Theory
- Invention
- Development
- Synthesis
- Preparation

Scientific resource

- Technique
- Probe (X-ray, neutron, electron)
- Equipment
- Beamline
- Funding
- Time
- Scale



Data flow beyond crystallography

Motivation
- Curiosity
- Mission
- Exploitation
- Serendipity
- Literature
- Patents

Management
- Funder
- Research Framework
- Principal Investigator
- Accountability

IT infrastructure
- Word size, Byte order
- Encoding
- Lossiness
- Volume
- Software dependencies
- Bit-rot

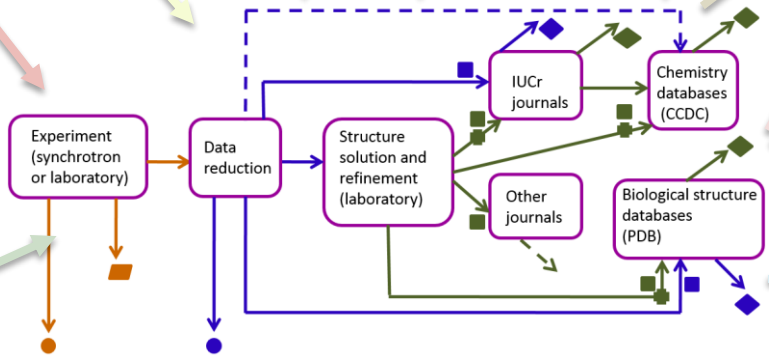
Knowledge
- Semantic Web
- Literature
- Text mining
- Theory generation

Scientific context for the experiment
- Theory
- Invention
- Development
- Synthesis
- Preparation

Reliability
- Equipment maintenance
- Calibration
- Regression testing
- Statistical analysis

Archiving
- Repositories
- Digital Libraries
- Curated databases

Credit
- Citation
- Reputation
- Recognition, Prizes



Scientific resource
- Technique
- Probe (X-ray, neutron, electron)
- Equipment
- Beamline
- Funding
- Time
- Scale

Trust
- Verification
- Validation
- Re-use
- Authorisation

Data flow beyond crystallography

IT infrastructure
- Word size, Byte order
- Encoding
- Lossiness
- Volume
- Software dependencies
- Bit-rot

Management
- Funder
- Research Framework
- Principal Investigator
- Accountability

Knowledge
- Semantic Web
- Literature
- Text mining
- Theory generation

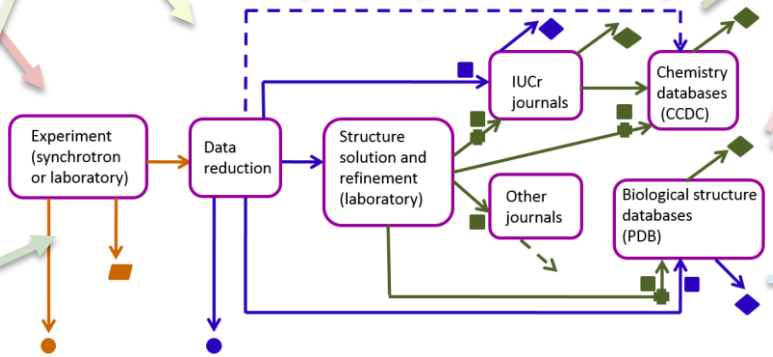
Motivation
- Curiosity
- Mission
- Exploitation
- Serendipity
- Literature
- Patents

Reliability
- Equipment maintenance
- Calibration
- Regression testing
- Statistical analysis

Archiving
- Repositories
- Digital Libraries
- Curated databases

Scientific context for the experiment
- Theory
- Invention
- Development
- Synthesis
- Preparation

Credit
- Citation
- Reputation
- Recognition, Prizes



Scientific resource
- Technique
- Probe (X-ray, neutron, electron)
- Equipment
- Beamline
- Funding
- Time
- Scale

Trust
- Verification
- Validation
- Re-use
- Authorisation

Acknowledgements

The activities described in the talks in today's Symposium owe much to many collaborators over the years:

Alan Mighell, Alex Renshaw, Alexei Vagin, Allen Larson, Alun Ashton, Andre Authier, Andy Hammersley, Andy Howard, Arie Van Der Lee, Ashley Buckle, Ben Watts, Bill Clegg, Bob Hanson, Bob Sweet, Brian Matthews, Brian Toby, Charlie Bugg, Chris Nielsen, Colin Groom, Curt Haltiwanger, Dale Tronrud, Dave Duchamp, Dave Stampf, David Brown, David Watkin, David Watson, Doug Du Boulay, Doug Greer, Eldon Ulrich, Eleanor Dodson, Enrique Abola, Eric Gabe, Erica Yang, Ethan Merritt, Frances Bernstein, Frank Allen, George Ferguson, George Sheldrick, Gerard Bricogne, Gerard Kleywegt, Gotzon Madariaga, Greg Shields, Gunter Bergerhoff, Helen Berman, Herbert Bernstein, Howard Einspahr, Howard Flack, I. David Brown, Ian Bruno, James Hester, Jan Zelinka, Jean Richelle, John Huffman, Jim Kaduk, Joe Krahn, Joel Sussman, John Bollinger, John Helliwell, John Westbrook, Keith Watenpaugh, Kim Henrick, Lachlan Cranswick, Liz Lyon, Liz Potterton, Lynn Ten Eyck, Manfred Weiss, Mario Nardelli, Mark Koennecke, Martyn Winn, Matt Towler, Michael Scharf, Mike Dacombe, Mike Hoyland, Mike Hursthouse, Mois Aroyo, Nick Day, Nick England, Nick Spadaccini, Owen Johnson, Paul Edgington, Paul Mallinson, Paula Fitzgerald, Peter Grey, Peter Keller, Peter Murray-Rust, Peter Strickland, Phil Bourne, Phil Coppens, Ralf Grosse-Kunstleve, Richard Ball, Robert Downs, Sameer Velankar, Sandy Blake, Saulius Grazulis, Shoshana Wodak, Sidney Abrahams, Simon Coles, Simon Hodson, Simon Parsons, Simon Westrip, Sine Larsen, Steve Androulakis, Steve Bryant, Syd Hall, Ted Maslen, Tom Koetzle, Tom Terwilliger, Ton Spek, Tony Linden, Vicky Karen, Vivian Stojanoff, Weider Chang, Wolfgang Bluhm, Yvon Le Page

… and many more besides

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

A Satellite Workshop to the 29th European Crystallographic Meeting

Sponsors

DECTRIS®

IUCr Journals
CRYSTALLOGRAPHY JOURNALS ONLINE

THE CAMBRIDGE CRYSTALLOGRAPHIC DATA CENTRE

ICODATA ISU

FIZ Karlsruhe
ICSD

WILEY

OxfordCryosystems

BRUKER