IUCrData - update on data publication and practices at the IUCr

Gillian Holmes
IUCr
5 Abbey Square, Chester CH1 2HU
checkin@iucr.org
‘it is essential that the methods by which the results have been gained, and the data on which they are founded, should be fully published so that they may be subjected to the expert criticism necessary to assess their reliability.’

Editorial preface, *Acta Cryst.* (1948). 1, 1
History of data publishing at the IUCr

• **1983** Launch of *Acta Crystallographica Section C* (incorporating *Crystal Structure Communications* published by University of Parma since 1972)

• **1987** Vote at Perth Congress to establish Working Party for Crystallographic Information

• **1991** CIF: facilitated machine submission and automatic article generation

• **1996** Mandatory CIF submission

• **1997** CIF-access papers introduced in *Acta C*

• **1999** *Crystallography Journals Online*

• **2000** *Acta C* included in *Crystallography Journals Online*, electronic papers replace CIF-access

5-Phenylidibenzophosphate, Polymorph 2
P. R. Meehan, E. C. Alyea and G. Ferguson

Abstract
Polymorph 1 of the title compound, C_20H_{13}P, has been described previously (Alyea, Ferguson & Gallagher, 1992) and crystallizes in space group P2_1_2_1_2_1 with two independent molecules in the asymmetric unit. The present polymorph 2 was obtained serendipitously, it crystallizes in space group P2_1_2_1_2_1 with one molecule in the asymmetric unit. The pendant phenyl ring of polymorph 2 adopts a slightly different orientation than that reported for the two independent molecules in polymorph 1.

Comment
X-ray analysis of crystals from what proved to be a wrongly labelled sample bottle has shown that they were a second polymorph of 5-phenylidibenzophosphate (I), whose structure we have reported previously (Alyea, Ferguson & Gallagher, 1992) in space group P2_1_2_1_2_1 with two independent molecules in the asymmetric unit. The polymorph reported here crystallizes in the chiral space group P2_1_2_1_2_1 with one molecule in the asymmetric unit.

A view of the molecule is in Fig. 1. Molecular dimensions are very similar to those reported for polymorph 1, with P–C in the range 1.808 (3) to 1.828 (6) Å (1.808 (9) to 1.846 (10) Å in polymorph 1). The orientation of the phenyl ring relative to the benzophosphate ring is defined by e.g. the torsion angle C11–C1–C21–C22 68.3 (3)°. In polymorph 1, the corresponding angles for the two independent molecules are 32.0 (5) and 34.1 (6)° respectively. Examination of the structure with PLATON (Spek, 1990) showed that there were no solvent accessible voids in the crystal lattice.

Experimental
The compound was prepared as described previously (Alyea, Ferguson & Gallagher, 1992) and recrystallized from dichloromethane.

Refinement
Molecules of (I) are achiral but crystallize in the orthorhombic system in the chiral space group P2_1_2_1_2_1. For the data collection, reflections in the range 0 to 44, 0 to 44 and 0 to 44 were measured to ensure a full Friedel set. During data
History of data publishing at the IUCr

• **2001** *Acta Crystallographica Section E: Structure Reports Online* launched
• **2007** Transition to open access, new shorter format introduced
• **2010** Editorial published reporting systematic scientific fraud in *Acta E*
• **2012** *Acta E* delisted from Science Citation Index
• **2014** Relaunch of *Acta E* with new paper formats, longer *Research Communications* and short *Data Reports*, and new subtitle *Crystallographic Communications*
• **2016** *IUCrData* launched with aim of ‘providing short descriptions of crystallographic datasets and datasets from related scientific disciplines’, the first phase providing a home for the short crystal structure reports previously published in *Acta E*
FAIR data

IUCr policy was always to allow free access to data

- Coordinates, anisotropic thermal parameters, structure factors
- In print days, knowledge of the existence of these depended on subscribing to the journal
- Since approximately 1991, these were exposed via online tables of contents (gopher predating invention of Web!)
- Since 1999 (Crystallography Journals Online) fully accessible including historical content
- Structure factors mandatory as electronic files since 1997
- Recommendation for links to primary data following DDDWG activities (2017)
IUCrData objectives

• To provide a new source of revenue
• To develop data publishing at the IUCr
• To get as much data into the public domain as possible
• To create a new home for Acta E and other IUCr datasets
• To try to ensure the quality of data in the public domain
• To promote data and journal articles already published by the IUCr
• To investigate ideas for deposition of novel types of data and large data sets
• To provide another mechanism to access the data held by the IUCr
Features of a data article in IUCrData

- Title
- Authors
- Abstract
- Keywords
- Peer-reviewed short text
- Figure(s) including an ellipsoid plot
- Chemical scheme
- Tabular information
- Acknowledgements
- References

- Supporting information, e.g. text files, videos etc.
- PDF version
- Interactive HTML version with integrated supporting information
- CIF or other data set (e.g. mmCIF, crystallization data etc.)
- Structure factors
- checkCIF report
- 3D visualisation tool
- Search
- Deposition with relevant databases
- Open access
Crystal structure of a salt with a cobalt(II) complex anion and a protonated sugar cation: (K(H₂O)₆)[(M(nitro)salt)glycine]·6H₂O.
A number of services are available to authors of data articles in IUCrData. The aim is to provide a publication process that is as smooth and rapid as possible.

These pages give instructions on how to prepare your data article for submission and tools to help you do this, details of what to do about supporting information, and utilities to track the status of your data article once it has been accepted and is in production.

PREPARE
- Notes for authors
- CIF information
- Author guide
- Standard software references
- Advantages of upgrading from SHELX-97 to SHELX-2014
- Prepare an enhanced figure (small molecule)
- Submit your article

SUBMIT
- Submission instructions
- Submit your data article
- Peer-review process

FOLLOW
- Status of your data article
- Download your proofs
- Download your electronic report
- Order reprints of your data article
- Publish your data article

POLICIES
- Article correction and retraction policy
- Ethics
- Author rights
- Copyright policy
- Open access

The title compound, [Co(2CN3)(C5H5H2N2O)2], crystallizes with two half molecules in the asymmetric unit, which are completed by crystallographic...
ICrData

home archive editors for authors for readers submit

| N2-C14 | 1.467Å |

Structure description

N-Methyl-N-propargyltryptamine (MNPT) is a structural analog of N,\textsubscript{\textalpha}-dimeethyltryptamine (DMT), which is a well-known \textit{psychedelic} molecule found in a variety of naturally occurring organisms, including plants, animals, and fungi, including mushrooms. In humans, DMT is the only known endogenous mammalian \textit{N,\textalpha}-dimeethyltryptamine (\textit{N,\textalpha}-DMT) naturally occurring \textit{tryptamines} (e.g., \textit{psilocybin}, \textit{\textalpha}-\textit{psilocin}, \textit{\textalpha}-\textit{psilocybin}, \textit{\textalpha}-\textit{psilocybin}) and their synthetic \textit{derivatives} (e.g., \textit{psilocin}, \textit{MNPT}) have garnered considerable attention of late due to new evidence demonstrating their efficacy in treating mood disorders (e.g., anxiety and depression) and post traumatic stress disorders (PTSDs) (Aubert et al., 2016\textsuperscript{1}, Cameron et al., 2019\textsuperscript{1}).

Polysaccharides isolated from the so-called \textit{magic} mushroom, is perhaps the best known \textit{product} of the \textit{psilocybin} \textit{2} against polianic (methyl, 2016\textsuperscript{1}).

Recent studies indicate that \textit{psilocin} and its precursors (\textit{psilocycin} and \textit{3,4,5-trimethoxyphenyl-2H-pyrimidine}) could provide effective treatment for mood disorders...
Development of data publishing

- Launch of *IUCrData*: a peer-reviewed service providing a home for the very short articles previously published in *Acta E*
- Enhancement of search and visualisation features across structural data sets housed on the IUCr servers
- Extend data management services to novel types of data
- Host or provide discovery mechanisms for experimental raw data sets (primarily diffraction images)
Crystallization and X-ray analysis of D-threonine aldolase from Chlamydomonas reinhardtii

Yuuki Hirasato, Masaru Gotou, Masami Takahashi, Mirona Tanigawa and Katashi Nishimura

*Department of Materials and Applied Chemistry, College of Science and Technology, Nihon University, 1-8-15, Funabashi, Chiba 274-8501, Japan
**Department of Biomolecular Science, Faculty of Science, Tokyo University, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. (Correspondence e-mail: hayasato@chem.nthu.edu.tw)

D-threonine aldolase from the green alga Chlamydomonas reinhardtii (CDTA) catalyzes the interconversion of several β-hydroxy-α-amino acids (e.g., D-threonine) and glycine plus the corresponding aldehydes. Recombinant CDTA was overproduced in Escherichia coli and purified to homogeneity; it was subsequently crystallized using the hanging-drop vapour-diffusion method at 25 °C. Data were collected and processed at 1.85 Å resolution. Analysis of the diffraction pattern showed that the crystal belonged to space group P21, with unit-cell parameters a = 64.79 Å, b = 73.92 Å, c = 80.44 Å, α = 77.87°, β = 89.34°, γ = 71.93°. The asymmetric unit contained four molecules of CDTA. The Matthews coefficient was calculated to be 2.32 Å³ Da⁻¹ and the solvent content was 41.9%.

Crystal

| Protein | Size (kDa) | Molar extinction coefficient (ε, M⁻¹ cm⁻¹) | pI | Solvent pH | Percentage identity (%) | Contact between residues (Å) | Secondary structure (%)
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CDTA</td>
<td>30.5</td>
<td>650</td>
<td>5.5</td>
<td>7.0</td>
<td>64.79</td>
<td>73.92</td>
<td>80.44</td>
</tr>
</tbody>
</table>

Sequence

1. Introduction

Threonine contains two chiral centres and exists as four stereoisomers (L- and D- isomers). The interconversion of several β-hydroxy-α-amino acids (e.g., threonine) and glycine plus the corresponding aldehydes (e.g., acetaldehyde) is catalyzed by threonine aldolases (lim, 2004). In general, threonine aldolases are classified into L- and D-type enzymes according to their stereocatalysticity at the α-carbon. Depending on their stereocatalysticity at the β-carbon of threonine, (α-type threonine aldolases can be further classified into three types (Dücker et al., 2008).
Ambiguity of DOIs

• DOI = digital object identifier
• Resolution services (e.g. CrossRef) allow long-term reference to a digital object
• Does not guarantee that the **content** is persistent
• IUCr journals assign DOI to each supporting data set (CIFs, s.f.s etc.)
• These are characterised as ‘part of’ the parent publication (*i.e.* not ‘first class’ primary research outputs)
Bi(2-phenyl-4-methylthiobis(carbamato)-N,S)(pyridine-2-yl)cadmium(II)

M. A. Elhan, M. Mahbar and F. R. T. Thakrar

The title compound, \(\text{[Bi(II)(CH}_2\text{CN}(\text{pyridine-2-yl})\text{]}_2\text{Cd}^{2+}\), features a five-coordinate Cd\(^{2+}\) atom, being coordinated by two nearly symmetrally disposed thiacarbamato ligands and a pyridine \(\text{N}\) atom. The resulting four donor set defines a distorted coordination geometry bending toward square pyramidal in all packing, and is isomorphous to \(\text{Cu}^{2+}\), whereas the coordination is as a result of methyl-C-H...\(\pi\) (phenyl) interactions.

Keywords: crystal structure; cadmium; thiacarbamato.

Supporting information

Crystallographic Information File (CIF) https://doi.org/10.1107/S2414314615024293
\(\text{/sj4002sup1.cif}\)
Contains datablocks 1, global

Structure factor file (CIF format) https://doi.org/10.1107/S2414314615024293
\(\text{/sj4002sup2.hkl}\)
Contains datablock 1

CCDC reference: 851623

* checkCIF report

3D view

* Full crystallographic data
Where does the DOI point?

• Following a CrossRef link associated with an IUCr data set DOI downloads the data file immediately
• Same is true of PDB files

BUT

• https://doi.org/10.1107/S2414314619007880/tk4058sup1.cif downloads a CIF directly from IUCr journals
• http://dx.doi.org/10.2210/pdb5ZZ0/pdb downloads a compressed PDB file directly from PDB
• https://dx.doi.org/10.5517/ccdc.csd.cc20vdhs takes the user to the WebCSD landing page
IUCr journals:
https://doi.org/10.1107/52414314619007880/kek4058sup1.cif

Protein Data Bank:
http://dx.doi.org/10.2210/pdb5ZZ0/pdb

Cambridge Structural Database:
https://dx.doi.org/10.5517/ccdc.csd.cc20vdhs
Towards an API

• Regardless of the decision on the canonical presentation of a DOI, it would be useful to develop an application programming interface:
  • Keyed on DOI
  • Load data in multiple (arbitrary?) formats
    • CIF1
    • CIF2
    • CIF-JSON
    • PDB and alternative macromolecular formats
  • Retrieve data sets matching specific criteria
Pilot project

• IUCr Chester will work on a prototype data harvesting API
• This will inform the direction of development effort in related fields
• Any suggestions for useful features of such an API welcomed
Summary

• The IUCr has always recognised the importance of data
• The journals require deposition of supporting data for published crystal structures - these data sets are thoroughly assessed as part of the publication process.
• IUCrData currently publishes short structure reports and is looking to expand into new areas
• The IUCr is participating in a pilot project on a data harvesting API
• We wish to work with the community and welcome suggestions for future developments