

[e0001]

Unknowns in Chemical Company's Catalogs

Armel Le Bail

Laboratoire des Fluorures, Université du Maine, Avenue O. Messiaen, 72085 Le Mans cedex 9, France. Tel. +33 02 43 83 33 47, Fax +33 02 43 83 35 06, Email lebaill@univ-lemans.fr

Received: 28 May 1998 / Uploaded: 28 May 1998

Abstract: More explicit catalogs should be required from chemical companies. Chemists may have to search references in many databases before to decide to buy a product. For each solid state crystallized phase, the corresponding [CSD](#), [ICSD](#) or [PDB](#) entry number should be given in catalogs, in addition to the generally provided CAS number. This would be a kind of proof that the compound has been somewhere submitted to the highest characterization level: a crystal structure determination. This would save considerable time and improve confidence in the sample quality. Powder diffraction has a key role to play in this process, if ever realized.

Keywords: Structure Databases, Chemical Company Catalogs, Unknown Crystal Structure.

- **[Introduction](#)**
 - **[Discussion](#)**
 - [Unknowns](#)
 - [Crystallography-related databases without atomic coordinates](#)
 - [Structures determined from some chemical company's "unknowns"](#)
 - [Databases including atomic coordinates](#)
 - **[Conclusion](#)**
 - **[References](#)**
 - **[Note](#)**
-

Introduction

Chemical companies are essential to modern research in chemistry. Searchable structure and property databases, with Web access, or on compact disk (CD) support, are also absolutely essential for research efficiency. Indeed, most of chemical companies propose a searchable catalog on CD, if not online. On the other hand, a structure determination (generally from single crystal X-ray data) is well recognized as the highest characterization level attainable (though bad works are possible). The problem discussed here is that, in chemical catalogs, one would expect to find some references to the crystal structure, but usually there is none. The only link to chemical databases found in catalogs is the CAS number entry, which is not directly informative. The chemist wanting to know more about a particular compound before to buy it, has to search structural databases, by himself. As discussed below, chemists as well as companies may considerably benefit from more explicit chemical catalogs.

Discussion

Unknowns

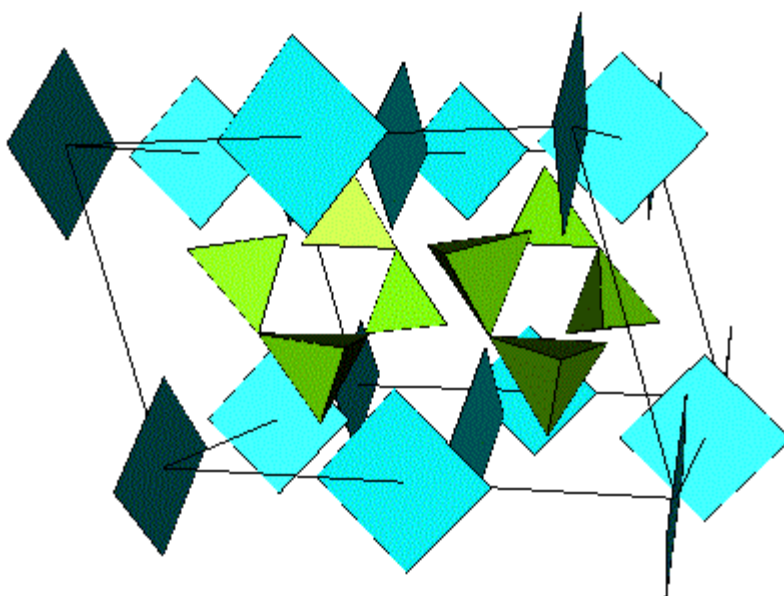
These comments originate from difficulties encountered during a search for unknown materials. The definition for "unknown" being here **"not included in the Cambridge Structural Database [1] (CSD), nor in the Inorganic Crystal Structure Database [2] (ICSD), nor in the Protein Data Bank [3] (PDB)"** meaning that the X-ray crystal structure has not been determined.

Most chemical company's catalogs do not indicate this essential information: was the crystal structure determined from X-ray (or neutron) single crystal (or powder) diffraction data? Therefore I asked some companies for their list of such "unknowns". I was searching for solid state stable compounds (preferably unavailable as crystals, only as fine powders) which were not included in the [CSD](#) or in the [ICSD](#). This question arose because a [Structure Determination by Powder Diffractometry Round Robin \(SDPDRR\)](#) was considered as timely. There was consequently a need for powder diffraction patterns of unknown compounds to be distributed to peoples wishing to compete (by different approaches in extracting structure factors from the powder pattern, solving by Patterson, direct methods, Monte Carlo...). Additionally, I suggested to these chemical companies that they might too be interested in the crystal structure determinations of their unknowns.

The standard reply was that **"Many of our solid state stable inorganic compounds are characterized by powder X-ray diffraction (we don't normally use X-ray diffraction for organic solids). We have some internal files, which we use as reference for these, to compare consistency from lot to lot. Also we use the JCPDS (Joint Committee of Powder Diffraction) as reference. Our systems are not set up to provide a list of products for which we do not have formal references, however, so we cannot provide this list for you."**

So that it seems that companies consider as sufficient a chemical analysis, or NMR results or that the sample X-ray powder pattern fits with one JCPDS-ICDD (International Center for Diffraction Data) card.

The problem is that many JCPDS cards even not give cell indexation (they also don't care with the information: is the crystal structure known?). Indeed, many "unknowns" which were not clearly identified as such in the ICDD [PDF-2](#) database (Powder Diffraction File), were further the subject of a structure determination from powder diffraction data. For instance, this is the case of [VO\(H₂PO₂\)₂•H₂O \[4\]](#), for which the unindexed 39-0057 JCPDS card has still not been actualized. Another variant is that an indexation is proposed in a JCPDS card, which has not been confirmed by a structure determination. This was the case of [\[Pd\(NH₃\)₄\]Cr₂O₇ \[5\]](#):



Two JCPDS cards were corresponding to this compound when the structure determination was decided: 40-1486 and 39-1422. These two cards were absolutely identical; both were proposed with a false triclinic cell. Since the publication of the structure, the 39-1422 card has been marked as "DELETED", remaining however in the package

with a "Quality: I" mark, meaning that it was indexed. Further adding of the true (monoclinic) cell parameters [5] rejuvenated the 40-1486 card. However, the old d(obs) (observed interplanar spacing) were kept and indexed according to the new cell! As a consequence, the corresponding FOM (Figure Of Merit) is of very low quality: $F_{30} = 8$ (0.028, 140). Really, this is amazing because the FOM given in the structure paper was six times better, $F_{20} = 51$ (0.0077, 48), corresponding to new d(obs) which were not inserted further in the modified card! The rules applied by ICDD for the publication of a JCPDS card remain a mystery for me. This case was discussed in the Rietveld Mailing List. The existence of double entries like the 39-1422 and 40-1486 cards, as they were before the palladium compound structure determination, indicates that new data are sometimes entered in [PDF-2](#) without even checking against the previous ones by a classical search-match procedure. There are a few other crystallography-related databases in which it is not clearly mentioned that the crystal structure was determined. Entries in these databases may include evaluated data on lattice parameters, crystal system, space group, Pearson's symbol, chemical name and formula, chemical class, density, references, and more, but no direct evidence that a structure determination has confirmed all these crystal data. The full list of these databases is given below.

Crystallography-related databases without atomic coordinates

[PDF-2](#) Powder Diffraction File Database [6]. The [PDF-2](#) Database commercialized by the ICDD is a collection of single-phase X-ray powder diffraction patterns in the form of tables of interplanar spacing and relative intensities and chemical name and formula as well as mineral name, if applicable. In addition, Miller indices, cell data and physical properties are listed, together with references for source information, where such data are available. As of Set 46, the total [PDF-2](#) database contains about 77,500 active patterns, the overwhelming majority of which represent unique phases.

[NIST](#) Crystal Data Identification File [7]. Produced in cooperation by ICDD with the National Institute of Standards and Technology ([NIST](#)), this compilation contains crystallographic and chemical data on more than 197,500 entries, representing approximately 60,000 unique phases. [NIST](#) Crystal Data covers the entire spectrum of well-characterized crystalline compounds including inorganic, organic, organo-metallic, metal, intermetallic, and mineral compounds.

[NIST/Sandia/ICDD](#) Electron Diffraction Database [8]. This database contains crystallographic and chemical information on over 817,200 crystalline materials, a large fraction of which is unique phases, for application to electron diffraction. Each entry, in addition to "R-spacing", contains space group data, unit cell data, chemical formula and name, literature references.

[CRYSDAT](#) [9] is the fusion of the above [NIST](#) Crystal Data and of CRYSMET [9] (database of intermetallic compounds). It contains about 250,000 entries for materials characterized by X-ray, neutron or electron diffraction whose unit cells and chemical compositions are known (organic, inorganic, mineral, biological, ionic, metallic, intermetallic, alloy, drug, antibiotic, pesticide).

Structures determined from some chemical company's "unknowns"

Examples of compounds of which the structures were solved from samples bought from chemical companies are not rare. The following structures were determined *ab initio* from powder diffraction data (examples could probably be found also, that were determined from a four-circle diffractometer study, using a sufficiently large single crystal). In some cases, the proposed formulae show a variable number of water molecules. It was the case of $Zr(OH)_2(NO_3)_2 \cdot 4.7H_2O$ [10] obtained from Aldrich as formulated $ZrO(NO_3)_2 \cdot xH_2O$, and of $NaAlO_2 \cdot 5/4H_2O$ [11] obtained from Merck with a $NaAlO_2 \cdot xH_2O$ formula. Incidentally, this means that sometimes you may buy rather undefined products. A list of more than 300 structures determined from powder diffraction data is available in the [SDPD-D](#) [12] (Structure Determination from Powder diffraction-Database). Among these, the following selected crystal structures were also determined from unknown samples as provided by chemical companies: 1-methylfluorene [13] from Aldrich, $KCaPO_4 \cdot H_2O$ [14] from Rhône Poulenc, toluene-*p*-sulfonhydrazide [15] from Aldrich, formylurea [16] from Lancaster, red fluorescein [17] from Aldrich, and probably many others. One compound was recrystallized from ethanol, to ensure that only a single powder phase was present: chlorothiazide [18], from Sigma. As a matter of fact,

one of the two samples selected for the [SDPD Round Robin](#) is tetracycline hydrochloride (from Aldrich), for which the [CSD](#) database mentions a structure determination but does not provide the atomic coordinates, because they were not listed in the reference publication...

It should now be clear that more should be asked from chemical companies. We, buyers, should require a clear mark in catalogs: "unknown (or known) crystal structure", for each crystalline product. This knowledge is not really difficult to obtain from the [CSD](#), [PDB](#) and [ICSD](#) databases. If the compound is an "unknown", then maybe the true composition is not as accurate as suggested by the vendor. Giving also the [CSD](#), [PDB](#) or [ICSD](#) entry number, in a way analogous to the CAS number, would be a plus. Nevertheless, for a chemical company, being sure that a compound corresponds really to some structure database entry could be a problem. For this purpose, either a calculated (from [CSD](#), [PDB](#) or [ICSD](#) atomic coordinates) or the tabulated JCPDS-ICDD powder diffraction pattern has to be checked against the sample experimental pattern. A good news is that the ICDD will soon add calculated patterns from [CSD](#) and [ICSD](#) data to its [PDF-2](#) database (proteins are too much complex, anyway). Many conflicts between calculated patterns and some obsolete JCPDS cards will be solved at this time, possibly. My feeling is that the different database owners communicate scarcely. None of the above databases had the (useful) appropriate link to those databases below, listing determined structures. The total of ~226000 determined structures suggests that some of the above databases might list a lot of unknowns (compare to the 817200 entries in the [NIST/Sandia/ICDD](#) electron diffraction database).

Databases including atomic coordinates

[CSD](#) [1] gathers now 175093 entries. Times are changing: the CAS registry number is no longer abstracted and all existing values have been deleted. The Inorganic Crystal Structure Database [2] presents more than 44000 compounds. The Protein Data Bank [3] is an archive of experimentally determined three-dimensional structures of biological macromolecules. Entries loaded on March 4, 1998: 7197 coordinate entries corresponding to 6655 proteins, 530 nucleic acids and 12 carbohydrates.

Conclusion

To be or not to be in crystal structure databases, this is the difference between well-characterized and hitherto unknown solid state compounds. Adding this information in chemical catalogs is timely. In case nothing is done before the next millenary, please contact me if you have identified some interesting "unknowns".

References

1. Cambridge Structural Database (CSD). <http://www.ccdc.cam.ac.uk/>
2. Protein Data Bank (PDB). <http://www.pdb.bnl.gov/>
3. Inorganic Crystal Structure Database (ICSD). <http://www.fiz-karlsruhe.de/home.html>
4. Le Bail, A., Marcos, M. D. and Amorós, P. *Ab initio* crystal structure determination of $\text{VO}(\text{H}_2\text{PO}_2)_2 \cdot \text{H}_2\text{O}$ from X-ray and neutron powder diffraction data. A monodimensional vanadium(IV) hypophosphite. *Inorg. Chem.* **1994**, 33, 2607-2613.
5. Lalignat, Y. and Le Bail, A. Structure of $[\text{Pd}(\text{NH}_3)_4]\text{Cr}_2\text{O}_7$. *Powder Diffraction* **1995**, 10, 159-164.
6. PDF-2 Powder Diffraction File Database. <http://www.icdd.com/>
7. NIST Crystal Data Identification File. <http://www.icdd.com/>
8. NIST/Sandia/ICDD Electron Diffraction Database. <http://www.icdd.com/>
9. CRYSDAT and CRYSTMET. <http://www.nrc.ca/programs/toth/>
10. Bénard, P., Louër, M. and Louër, D. Crystal structure determination of $\text{Zr}(\text{OH})_2(\text{NO}_3)_2 \cdot 4.7\text{H}_2\text{O}$ from X-ray powder diffraction data. *J. Solid State Chem.* **1991**, 94, 27-35.
11. Kaduk, J. A. and Pei, S. The crystal structure of hydrated sodium aluminate, $\text{NaAlO}_2 \cdot 5/4\text{H}_2\text{O}$, and its dehydration product. *J. Solid State Chem.* **1995**, 115, 126-139.
12. Structure Determination from Powder Diffraction-Database (SDPD-D). <http://fluo.univ-lemans.fr:8001/iniref.html>

13. Tremayne, M., Kariuki, B. M. and Harris, K. D. M. Solution of an organic crystal structure from X-ray powder diffraction data by a generalized rigid-body Monte Carlo method: crystal structure determination of 1-methylfluorene. *J. Mater. Chem.* **1996**, *6*, 1601-1604.
 14. Louër, M., Plévert, J. and Louër, D. Structure of $\text{KCaPO}_4 \cdot \text{H}_2\text{O}$ from X-ray powder diffraction data. *Acta Crystallogr.* **1988**, *B44*, 463-467.
 15. Tremayne, M., Lightfoot, P., Glidewell, C., Harris, K. D. M., Shankland, K., Gilmore, C.J., Bricogne, G. and Bruce, P.G. Application of the combined maximum entropy and likelihood method to the *ab initio* determination of an organic crystal structure from X-ray powder diffraction data. *J. Mater Chem.* **1992**, *2*, 1301-1302.
 16. Lightfoot, P., Tremayne, M., Harris K. D. M. and Bruce, P. G. Determination of a molecular crystal structure by X-ray powder diffraction on a conventional laboratory instrument, *J. Chem. Soc., Chem. Commun.* **1992**, 1012-1016.
 17. Tremayne, M., Kariuki, B. M. and Harris, K. D. M. Structure determination of a complex organic solid from X-ray powder diffraction data by a generalized Monte Carlo method: the crystal structure of red fluorescein. *Angew. Chem. Int. Ed. Engl.* **1997**, *36*, 770-772.
 18. Shankland, K., David, W. I. F and Sivia, D. S. Routine *ab initio* structure determination of chlorothiazide by X-ray powder diffraction using optimised data collection and analysis strategies. *J. Mater. Chem.* **1997**, *7*, 569-572.
-

Note : In order to see the [VO\(H₂PO₂\)₂•H₂O](#) and [\[Pd\(NH₃\)₄\]Cr₂O₇](#) crystal structures, you need a VRML viewer (preferably [Cosmo Player](#)).

Comments

During 1-30 September 1998, all comments on this poster should be sent by e-mail to ecsoc@listserv.arizona.edu with e0001 as the message subject of your e-mail. After the conference, please send all the comments and reprints requests to the author.
