

PDF-5+: a comprehensive Powder Diffraction File™ for materials characterization

Soorya N. Kabekkodu ^{a)}, Anja Dosen , and Thomas N. Blanton 

International Centre for Diffraction Data, Newtown Square, PA, USA

(Received 22 November 2023; accepted 19 March 2024)

For more than 80 years, the scientific community has extensively used International Centre for Diffraction Data's (ICDD®) Powder Diffraction File (PDF®) for material characterization, including powder X-ray diffraction analysis. Historically, PDF was made available for two major material types: one for inorganic analysis and the other for organic analysis. In the early years of the PDF, this two-material approach was implemented due to limited computer capabilities. With Release 2024, ICDD provides a comprehensive database consisting of the entire PDF in one database called PDF-5+, comprised of more than one million entries (1,061,898). The PDF-5+ with a relational database (RDB) construct houses extensive chemical, physical, bibliographic, and crystallographic data, including atomic coordinates and raw data, enabling qualitative and quantitative phase analysis. This wealth of information in one database is advantageous for phase identification, materials characterization, and several data mining applications in materials science. A database of this size needs rigorous data curation and structural and chemical classifications to optimize pattern search/match and characterization methods. Each entry in the PDF has an editorially assigned quality mark. An editorial comment will describe the reason if an entry does not meet the top-quality mark. The editorial processes of ICDD's quality management system are unique in that they are ISO 9001:2015 certified. Among several classifications implemented in PDF-5+, subfiles (such as Bioactive, Pharmaceuticals, Minerals, etc.) directly impact the search/match in minimizing false positives. Scientists with specific field expertise continuously review these subfiles to maintain their quality. This paper describes the features of PDF with an emphasis on the newly released PDF-5+.

© The Author(s), 2024. Published by Cambridge University Press on behalf of International Centre for Diffraction Data. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

[doi:10.1017/S0885715624000150]

Keywords: Powder X-ray diffraction, XRD database, quality mark, subfiles, PDF card

I. INTRODUCTION

Powder diffraction has become an indispensable tool in materials characterization owing to its wide variety of applications, from materials identification, quantification, and structure elucidation to texture analysis. Powder X-ray diffraction (PXRD) is one of the most widely used techniques in solid-state materials research, as the PXRD pattern is often considered a fingerprint of the material under study. PXRD is a vital characterization tool for phases where growing a single crystal is difficult or not attainable in practice or the phase of interest is an one member of a mixture of two or more polycrystalline materials.

The origin of the method for identifying the phase(s) present in a given sample using X-ray diffraction dates back to 1938 (Hanawalt et al., 1938). In this paper, using approximately a thousand PXRD patterns of crystalline phases, the authors presented a technique for identifying phases leading to the genesis of phase identification (Phase ID) using PXRD. The method, popularly known as *Hanawalt search*

involves developing an index using the 3 strongest lines (peaks) of a diffraction pattern and grouping them into 77 subdivisions (Hanawalt et al., 1938; Hanawalt, 1986). Strong PXRD lines of the unknown phases can be compared against the Hanawalt index to identify the phase. The importance of diffraction information was central to forming the Joint Committee for Chemical Analysis by Powder Diffraction Methods in 1941, co-sponsored by ASTM Committee E-4, the Crystallographic Society of America, and the British Institute of Physics (Faber and Fawcett, 2002). The first powder diffraction database collection was published in 1941 as Set 1 of the Powder Diffraction File (PDF) (Faber and Fawcett, 2002). By 1969, the ASTM Committee was constituted as a Pennsylvania nonprofit cooperation under the title of the Joint Committee on Powder Diffraction Standards (JCPDS). In 1978, the current name, International Centre for Diffraction Data (ICDD), was adopted. A summary of the formative years of the PDF was subsequently summarized by Messick, 2012. The PDF continues to be managed and maintained by the International Center for Diffraction Data (ICDD, www.icdd.com), a nonprofit scientific organization dedicated to collecting, editing, publishing, and distributing powder

^{a)} Author to whom correspondence should be addressed. Electronic mail: kabekkodu@icdd.com



diffraction data to identify materials. The membership of the ICDD consists of worldwide representation from academe, government, and industry. This paper is an update of a previously published paper (Gates-Rector and Blanton, 2019) on the PDF®. In this paper, we attempt to focus on the recent changes to the PDF database and relational database software and to minimize the duplication of features that were presented in the Gates-Rector and Blanton paper of 2019.

II. DATABASE DEVELOPMENT: PDF

In Set 1 of the PDF published in 1941, powder diffraction data were published on a 3 in × 5 in paper card (Figure 1(a)) listing interplanar spacings (d spacings) and relative intensities (I/I_0) characteristic of the compound (Jenkins et al., 1987). The original handwritten index book by Hanawalt et al. (1938) is shown in Figure 1(b). The first electronic version of PDF was introduced in the year 1967 in a flat file format enabling automated search/match (Johnson and Vand, 1967). This was the first electronic version of a crystallographic database. Over the years, ICDD has collaborated with various database organizations to expand the scope of materials coverage in the PDF to facilitate materials characterization using PXRD. A collaborative project with the National Bureau of Standards was started in 1953 and lasted more than 30 years (Wong-Ng et al., 2001). The growth of the PDF since its inception in 1941 to 2023 is shown in Figure 2. As seen in Figure 2, data spikes in the years 1998, 2002, and 2005 are due to the onset of collaborations with ICSD (Inorganic Crystal Structure Data, FIZ Karlsruhe; Zagorac et al., 2019), CSD (Cambridge Structural Data, Cambridge Crystallographic Data Centre; Groom et al., 2016; Bruno et al., 2017), and LPF (Linus Pauling File, Materials Phases Data System; Villars and Cenzual, 2018) respectively. ICDD has responded to the growing need for database requirements in materials characterization using powder X-ray diffraction by implementing a unique project targeting data acquisition, including raw data, and customized subfiles. The ICDD has a Grant-in-Aid program to acquire diffraction data on targeted materials and information about this grant is available on the ICDD website (<https://www.icdd.com/grant-in-aid/>).

The data management system for the PDF underwent a significant change in 2005 by adopting a relational database

management system (RDBMS). The PDF in relational database format (RDB) contains extensive chemical, physical, bibliographic, and crystallographic data, including atomic coordinates enabling characterization and computational analysis. The RDB construct provides pliable access to the database for data mining studies and enhances the PDF materials characterization capabilities (Faber and Fawcett, 2002; Kabekkodu et al., 2002). One of the profound benefits of RDB is the more straightforward implementation of FAIR (Findability, Accessibility, Interoperability, and Reusability) principles (Wilkinson et al., 2016) and better data interoperability. This type of construct is vital from an interoperability perspective, given the existence of multiple automated software routines (by various diffractometer manufacturers or software developers) used for phase identification and quantification.

Until the release of 2024, PDF database products were separated as PDF-4+ and PDF-4/Organics to focus on inorganic and organic phase identifications. With the growing number of crossover entries involving both inorganic and organic phases, the community using the PDF more often need both the inorganic and organic PDF content to successfully complete a characterization of the phases in the sample. To meet this growing challenge the new PDF-5+ combines the former inorganic and organic files into one single database containing 1 061 898 entries. There are many useful database filters (Figure 3, labels 1, 2, and 3) to optimize search mechanisms. In other words, if a user desires, it can mimic the function of PDF-4+ or PDF-4/Organics.

III. DATA CURATION

In phase identification or related characterizations, a user's powder pattern is compared against the database. A well-curated database is essential for such analysis to be efficient and reliable. In addition to diffraction patterns and crystallographic parameters, metadata curation plays a vital role in phase identification as one tries to associate all the peaks in a measured diffraction pattern with one or multiple entries in PDF-5+ database phase. Figure 4 depicts a general overview of the curated data in the PDF. Many of the metadata in Figure 4 are self-explanatory; the following section will briefly describe some key features in phase identification and data mining studies. All of these data fields can be searched, and multiple

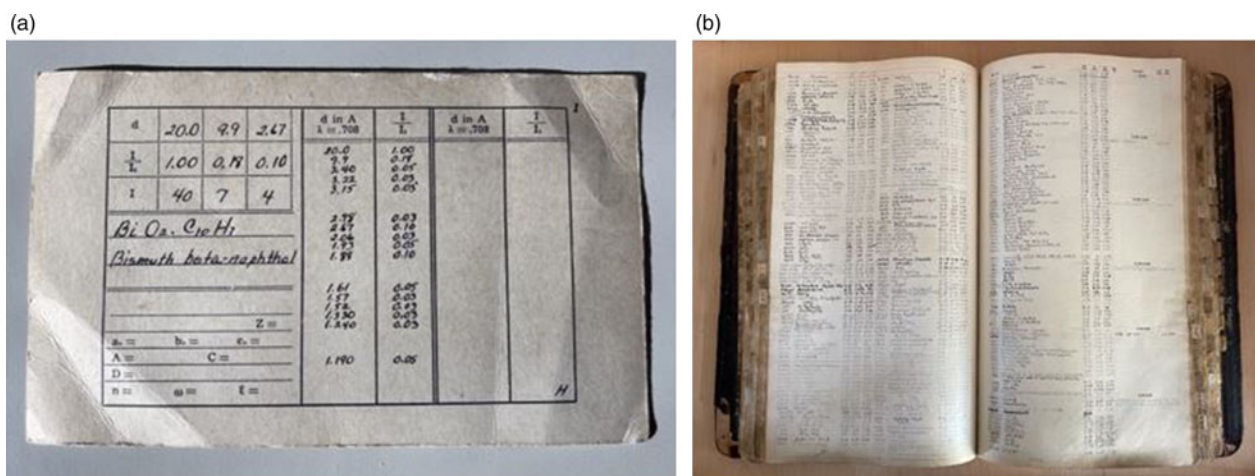


Figure 1. (a) Early PDF card and (b) original Hanawalt numerical index book (courtesy Dow chemical).

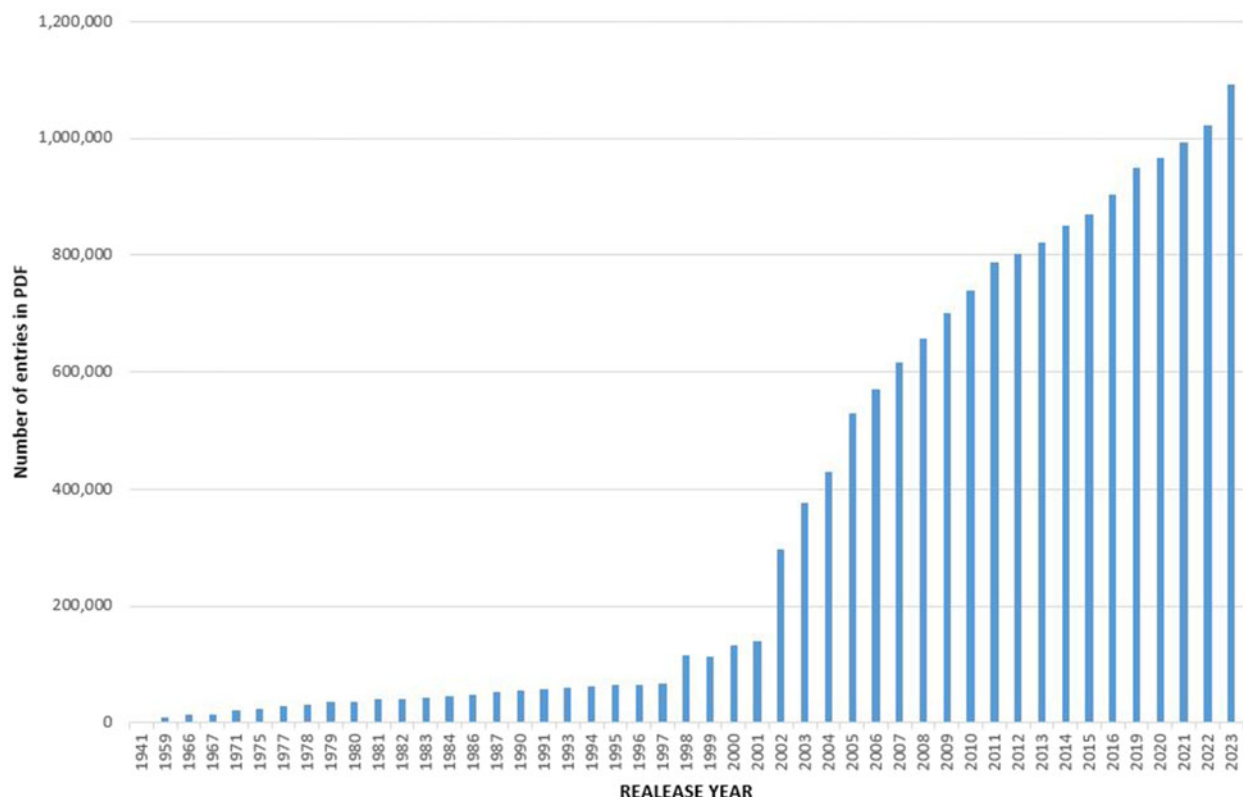


Figure 2. Growth of entries in the powder diffraction file (note: for a better clarity, only selected years during the period 1941–1993 are shown to illustrate the trend).

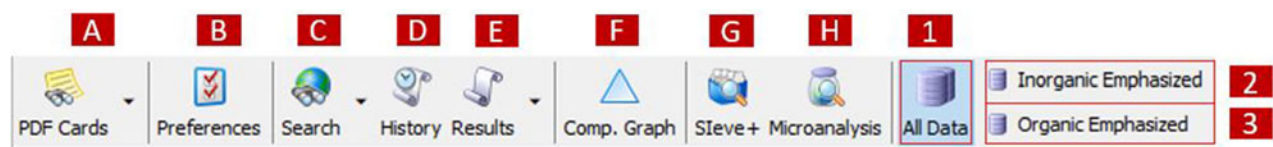


Figure 3. PDF-5+ tool bar with descriptions listed in Table I.

searches can be combined using Boolean operators to expedite search/match and data mining studies.

A. Quality mark and database status

The ICDD PDF is a curated database with each entry evaluated using a combination of computer and human editorial

review and presented using a quality mark (QM) nomenclature (Hubbard et al., 1980; Jenkins and Smith, 1987; Bruno et al., 2017; Gates-Rector and Blanton, 2019; Kaduk, 2019) that provides the user with a systematic process for understanding data entry quality and a consistent approach in assessing the goodness-of-match (GOM) for phase identification. All of the data in the PDF has gone through a multi-tier editorial

TABLE I. PDF-5+ tool bar descriptions.

| Label | Description |
|-------|--|
| 1 | All Data: Uses the entire database |
| 2 | Inorganic Emphasized: Mainly for inorganic phases with some common organics (DB code 02- and most of 05- are not included) |
| 3 | Organic Emphasized: Mainly Organic phases with some common inorganics (like excipients, pigments, etc.) |
| A | PDF Cards: Opens a PDF card (entry) by the given PDF number. Also has a feature to open PDF entries corresponding to the collaborating database (CSD, ICSD, NIST, LPF) identifier |
| B | Preferences: Display preferences such as desired fields and pattern simulation parameters can be customized. |
| C | Search: Brings up the main search window. The dropdown button can be used to select recent searches. |
| D | History: Contains previous search criteria. These searches can be combined using Boolean “and”, “or” operators |
| E | Results: Using the dropdown button, previous search results can be repopulated |
| F | Composition Graph: Any 2 or 3 elements combination can be graphed as a function of atomic% or weight%. For example, Li–Mn–O system |
| G | Sieve+: Launches the search/match program Sieve+ |
| H | Microanalysis: Elemental composition from a micro-XRF analysis, microprobe analysis, or other elemental analysis |

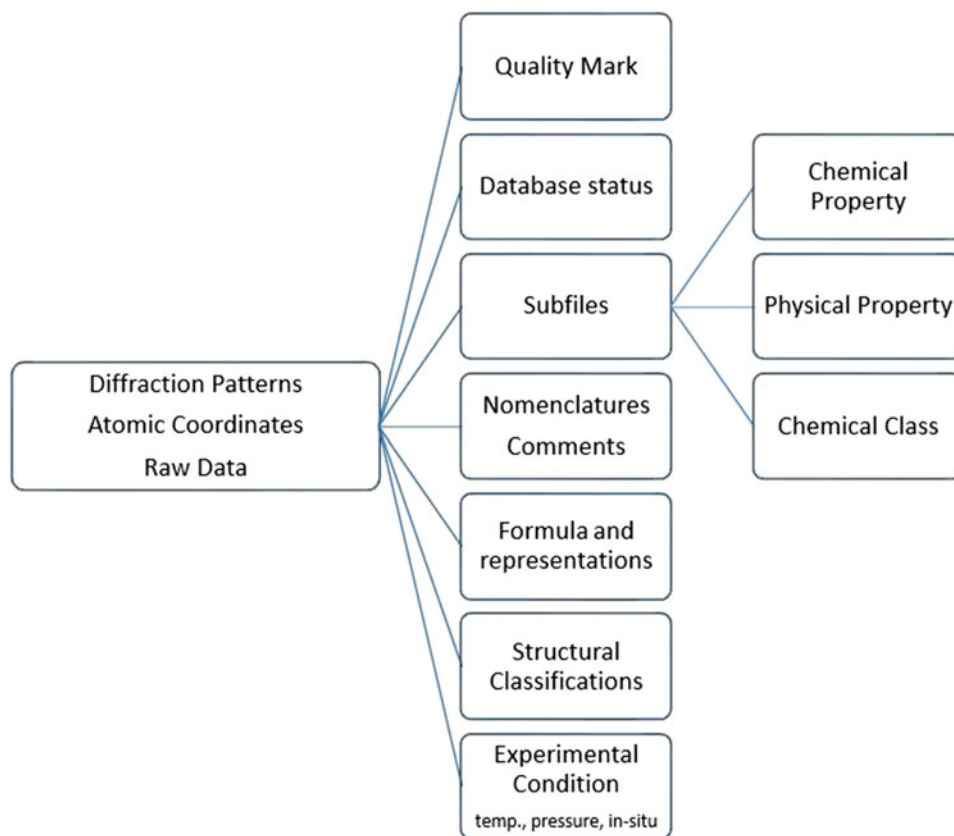


Figure 4. General overview of curated data in the powder diffraction file.



Figure 5. A schematic view of the quality mark (QM) assignment process. The top row corresponds to the breadth of checks while the second row of boxes provide an outline of the checks that are made. The assigned QM indicates high-quality (QM = * or G), medium-quality (QM = I, C, P, or M), low-quality (QM = B or O), or hypothetical (QM = H) patterns.

process. The ICDD quality management system is certified by the International Organization for Standardization (ISO 9001:2015, Certificate Number: 110409.01). Figure 5 briefly summarizes the QM evaluation criteria that are checked by computer routines and then reviewed by one of the PDF editors before the QM is assigned.

Every entry in the PDF has an editorially assigned QM. An editorial comment will describe the reason an entry does not meet the top-quality mark. The PDF has two types of PXRD patterns, experimental and calculated from published crystal structures. QM assignment considers them separately after going through more than 100 data validation checks as shown in Figure 5. It is important to note here that the QM assignment is strictly from the phase identification point of view. For example, PXRD calculated patterns using crystal structure data with missing atom(s), disorder, improper site

occupation factor, or questionable displacement parameters can have a significant impact on phase identification and thus would have a lower QM. Experimental PXRD patterns require a different type of treatment by considering phase purity, quality of the diffraction pattern, and indexing. QMs G and M are applicable only for poorly crystalline, nano material, and amorphous patterns. In this case, QM criteria emphasizes the chemical, data quality, and other characterizations (spectroscopy, pair distribution function, and microscopy). Table II describes the QMs used in the PDF.

It is common in chemical crystallographic databases to have multiple entries for a given phase, especially in the case of widely studied structures. The PDF uses the Database Status flag to categorize multiple entries as Primary, Alternate, or Deleted. Table III describes the meaning of these notations. The combination of quality

TABLE II. Major criteria for PDF quality mark assignment.

| Mark | $\Delta 2\theta$ | Crystallographic Information | Warnings | Other |
|-----------------|---|---|--|--|
| S or (*) | <i>Average</i> $\leq 0.03^\circ$ | <ul style="list-style-type: none"> ➤ High-quality diffractometer or Guinier data ➤ Known unit cell ➤ Complete indexing | ➤ None | $\Delta 2\theta$ specific to experimentally based patterns. For ICDD-calculated patterns, there should not be any crystallographic warnings. |
| G | <i>n/a</i> | <ul style="list-style-type: none"> ➤ Significant amorphous components present ➤ Good signal-to-noise ratio in digital diffraction pattern (PD3) provided ➤ Chemical analysis information provided that supports the specified composition of a material | <i>n/a</i> | (I.e. spectroscopy, pair distribution functions, commercial source, etc.). |
| I | <i>Average</i> $\leq 0.06^\circ$ <i>Absolute</i> $\leq 0.20^\circ$ for all reflection | <ul style="list-style-type: none"> ➤ Indexed pattern ➤ Known unit cell ➤ Reasonable range and uniform distribution in intensities | <ul style="list-style-type: none"> ➤ No serious systematic errors ➤ Maximum of two unindexed, space group extinct, or impurity reflections; none of these reflections are among the strongest eight lines ➤ For calculated patterns, only minor structural warnings are allowed | Completeness of the pattern is sensible. Reflections with <i>d</i> -value less than or equal to 2.000 Angstroms have at least three significant figures after the decimal point. |
| C | | <ul style="list-style-type: none"> ➤ Pattern calculated from single-crystal structural parameters for which the structural refinement R-factor was < 0.10. ➤ $F(\text{calc})$ data have been checked against the corresponding $F(\text{obs})$ <p>— OR —</p> <ul style="list-style-type: none"> ➤ Complete check of the bond distances and angles has been made along with density and formula checks ➤ Number of required significant digits is the same as for a “S” quality mark | ➤ If calculated pattern does not meet the “S” quality mark parameters, it is assigned QM = B | This QM refers to author’s published calculated patterns. Those with “C” should not be confused with ICDD-calculated patterns based on crystal structures from the collaborating databases. ICDD-calculated pattern will have a QM other than C/G/M. |
| M | | <ul style="list-style-type: none"> ➤ Amorphous (or poorly crystalline) as major component present ➤ Good signal-to-noise ratio in digital diffraction pattern (PD3) provided | | No chemical analysis data to support the materials composition provided (usually commercial samples). |
| B | | <ul style="list-style-type: none"> ➤ Unit cell maybe unknown ➤ If indexed, higher $\Delta 2\theta$ than QM “T” | ➤ For ICDD-calculated patterns, these have significant warnings on incomplete structure, symmetry warnings, nonpositive definite displacement parameters, etc. | Experimental pattern may have unindexed lines |
| P | | ➤ Applicable only for calculated patterns | ➤ For ICDD-calculated patterns based on structure prototype. These are the cases in which author or editor assigned the structure type | |
| O | | <ul style="list-style-type: none"> ➤ Poorly characterized material or the data are known (or suspected) to be of low precision. ➤ Number of unindexed, space group extinct, or impurity reflections ≥ 3 ➤ One of 3 strongest reflections is unindexed ➤ Poor chemical characterization ➤ No unit cell reported | ➤ Usually, the editor has inserted a comment to explain why the “O” was assigned. | A low-precision quality mark means that the diffraction data have been taken on questionable sample or one that contains significant impurities. |
| H | | ➤ Applicable only for calculated patterns | For ICDD-calculated patterns based on theoretical structures [e.g. Density Functional Theory (DFT) calculations] | |

TABLE III. Description of database status.

| Database status | Description |
|-----------------|--|
| Primary (P) | Editorially chosen as a representative entry for a given phase |
| Alternate (A) | Alternate choices for a given phase. This does not necessarily mean lower quality than primary |
| Deleted (D) | Usually duplicate entries, unresolved errors, replaced by entries with better data. |

marks with database status is extremely useful as a filter in searching the database, especially in cases where several entries exist for a given phase. For example, in release 2024 of the PDF, there are 77 quartz patterns, but if one uses the primary star quality pattern as a filter, the search result set will have one pattern.

B. Raw data archival

The interest in depositing to an archive the raw data associated with the reporting in a journal article has many positive aspects. First, is then data centers can retrieve the raw data and if desired archive it in a database. The IUCr has been promoting this initiative as has ICDD for many years. The review paper aptly titled Science in the Data (Helliwell et al., 2017 and references therein) elegantly encapsulates the need for raw data in crystallography. The challenges in archiving raw powder diffraction patterns are manifold due to phase impurities, data collection strategies, diffractometer geometry, sample preparation, systematic errors, and preferred orientation. In a traditional search/match algorithm, the user's d - I pairs are compared against

the database (Fawcett et al., 2017, 2020), which works well for crystalline phases due to their sharp (narrow) peaks. It is evident from Figure 6 that having raw data is essential to carry out phase identification in the case of poorly crystalline or amorphous patterns as they cannot be represented satisfactorily as a list of interplanar spacing (d values) and relative intensity (I values) due to broad peaks containing several d values. Matching the raw data using a similarity index (Hofmann and Kuleshova, 2005) is one of the best methods to perform search/match, particularly for cases where diffraction peaks are broad.

As of release 2024, more than 19,000 raw powder diffraction patterns (1D) in the PDF are available for users to analyze powder diffraction data. In an extended application, the crystal structure of Trandolapril was solved (Reid et al., 2016) using deposited raw data from the PDF. There is a growing number of examples where raw data is essential in characterizing pharmaceutical samples (Fawcett et al., 2019) and polymers (Gates et al., 2014).

C. Structural classifications

Structural classifications (Figure 7) play a seminal role in validating the quality of the published structure or powder diffraction patterns. It is an excellent tool for identifying similar phases that are otherwise not easily recognizable. It is also a good editorial tool to identify outliers in a supposedly related phase in a cluster. From the application point of view, structure prototypes are extremely useful in deriving a starting model for Rietveld refinement when the database is missing the exact chemistry [i.e. element(s)]. ICDD uses several structural classification symbols to meet the requirements of scientists from various fields who use X-ray diffraction for

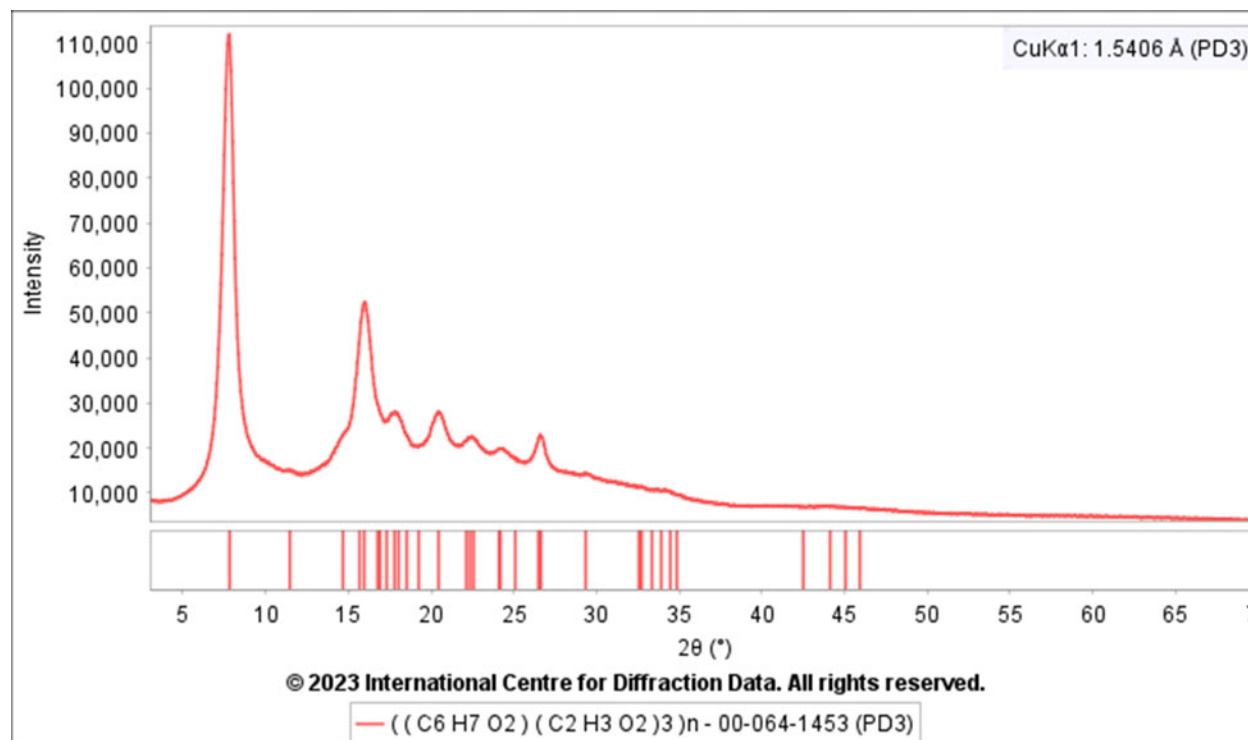


Figure 6. PXRD pattern of cellulose triacetate I with reported peak positions PDF entry 00-064-1453.

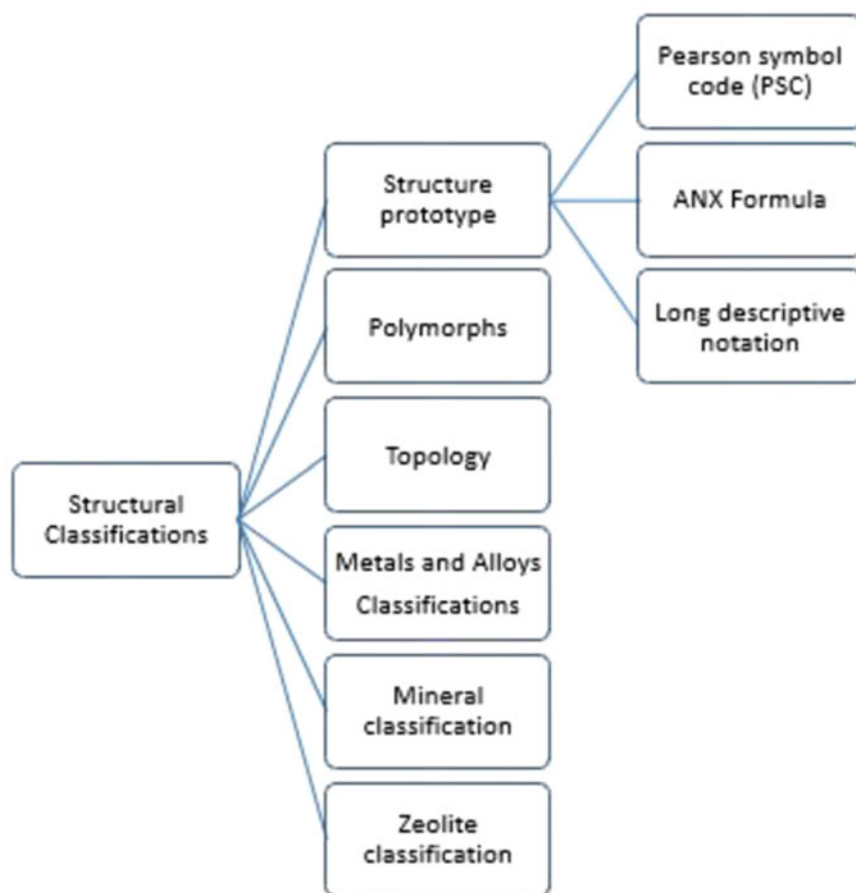


Figure 7. Structural classifications in the powder diffraction file.

materials characterization. Each of these has a specific usage in structural chemistry. The structural classifications currently used in the PDF are listed below.

- Traditional structure type notation is based on standardized unit cell, axial ratios, space group, Pearson symbol code (Pearson, 1967), and chemistry (Wallace et al., 1994). These descriptors are usually assigned manually, for example, the CaF_2 type structure. They are helpful in comparisons of phase-diagrams of systems that are expected to be similar (for example, Pb–Pu vs Pu–Sn) (Wallace et al., 1994), axial ratios and lattice parameters, which are extremely useful in controlling the quality of the database entries.
- ANX type is based on the type of ion and site occupancy. For example, CaTiO_3 is of ABX3 type and Fe_3O_4 is of AB2X4 type.
- Structure prototyping is based on the standardization method described by Gelato and Parthé, 1987, followed by a detailed crystallographic analysis of the atomic environment (Villars et al., 1998; Allmann and Hinek, 2007). For example, $\text{Cu}_3\text{As}, cI64, 220$ (structure type formula, Pearson symbol code, space group number). Standardization is crucial in maintaining the quality of the crystallographic data as some structures with the same space group and Pearson symbol code were erroneously considered as isotypic because their reported unit cell parameters are similar (Allmann and Hinek, 2007).
- Zeolite classification is based on framework type (Baerlocher et al., 2007). These are usually noted by a three-capital letter code describing a network of corner-sharing tetrahedrally coordinated framework atoms. The International Zeolite Association (IZA) publishes these codes and the codes are listed on their website (https://america.iza-structure.org/IZA-SC/ftc_table.php)
- Mineral classification is based on the crystal chemistry characteristics (Smith et al., 1997). The classification aims to exploit the crystal chemistry characteristics and similarity of the powder diffraction pattern among isostructural group members to enhance the quality of the mineral phases in the PDF.
- Topology: Unlike geometrical analysis, topological analysis provides a tool to use crystallographic data and explore the topological structure and the similarities in the crystal space (Blatov et al., 2014). From the database point of view, it is advantageous to look for isostructural or isorecticular series. Topological analyses can also be used to improve the quality of the database.

D. Subfiles

In phase identification, there are possibilities of having false positives due to similarity in diffraction patterns, often the case with isostructural entries. For example, with space group $Fm\bar{3}m$ and $a = 5.39(2) \text{ \AA}$, there are 577 entries in release 2024. Subfiles, a curated category based on chemistry, properties, or application, efficiently reduce the occurrence of such false positives. For example, a user interested in mineral analysis can limit their search/match to only minerals to eliminate unwanted result sets. To ensure the subfiles quality, field experts continuously review the subfiles. These experts are

usually members of a designated ICDD Subcommittee (<https://www.icdd.com/technical-subcommittees/>).

IV. THE PDF ENTRY

As shown in Figure 3, the exponential growth of PDF has long made it difficult to publish data in printed book format along with a digital version. The last printed book format was set 67 which was Release 2017. However, the historical nomenclature of cards still exists, and PDF entries are still referred to as PDF cards. These entries are identified by a numbering system *AA-BBB-XXXX* implemented in the PDF in 2003. The first two digits (*AA*) identify the data source: 00 = ICDD; 01 = ICSD (Inorganic Crystal Structure Database); 02 = CSD (Cambridge Structural Database); 03 = NIST (National Institute of Standards and Technology); 04 = LPF (Linus Pauling File); and 05 = ICDD (abstracted crystal data, including modulated structures). The remaining part, *BBB-XXXX*, corresponds to set and pattern numbers, respectively. The mineral Kaliophilite (Entry 00-071-0879) is used as an example in Figure 8 to illustrate a PDF entry.

Within a PDF entry, there are different types of data content and simulation features to explore, and these are

summarized in Figures 9 and 10. In addition to these features, PDF-5+ offers JADE toolkit extensions and these can be launched from a PDF entry as shown in Figure 11.

A. Bravais–Friedel–Donnay–Harker (BFDH)

This toolkit is for crystal morphology prediction. The BFDH model is based on the lattice geometry and symmetry and was proposed by Bravais, Friedel, Donnay, and Harker (Donnay and Harker, 1937). This model assumes that the energetically most stable and slowest growing faces are the ones with the highest density and largest spacing between adjacent layers and that the rate of growth of a particular surface is inversely proportional to the interplanar d spacing ($D \sim 1/d$). One can also choose methods for supersaturation/rapid growth with $D \sim \text{Exp}(-kd)$ (Dowty, 1976). Users can change the value of k if needed to obtain the growth rate curve (k is inversely proportional to the supersaturation). As an illustration, Figure 12 shows the correctly predicted (confirmed by the comment under “habit” on the PDF card) needle morphology using this BFDH module for a pharmaceutical PDF entry Eupatorin (PDF# 02-076-2266). The BFDH module can also be used to predict possible preferred orientations.

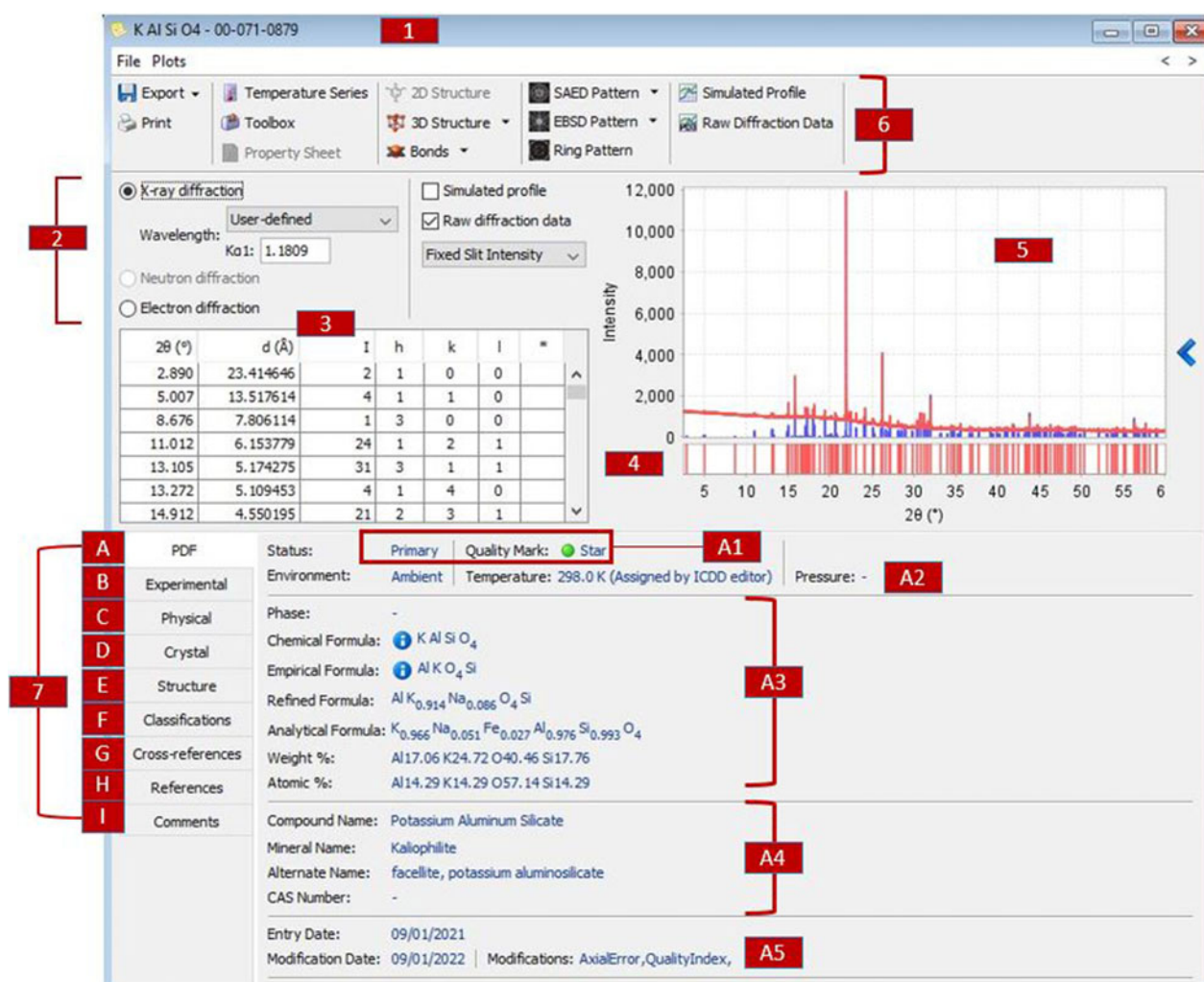


Figure 8. Screenshot of PDF card display. Example PDF# 00-071-0879: potassium aluminum silicate, KAlSiO₄ with designators in red corresponding to descriptions in Tables IV and V. Note that the wavelength on the card was customized to the raw data collection wavelength (synchrotron).

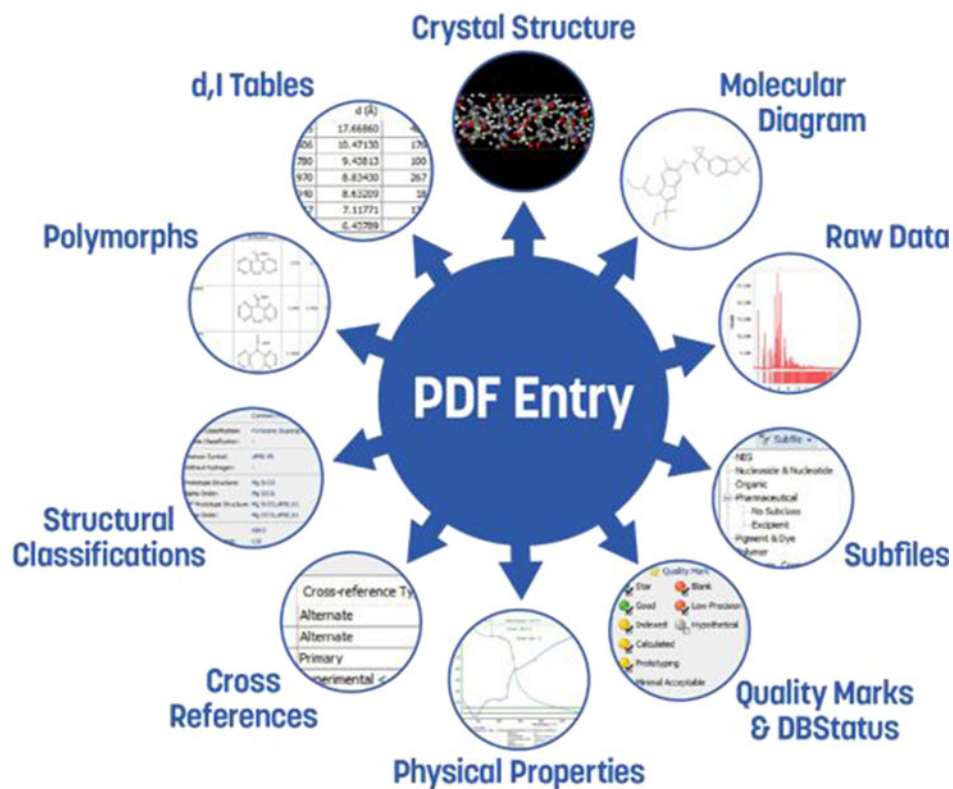


Figure 9. Overview of PDF entry data content.

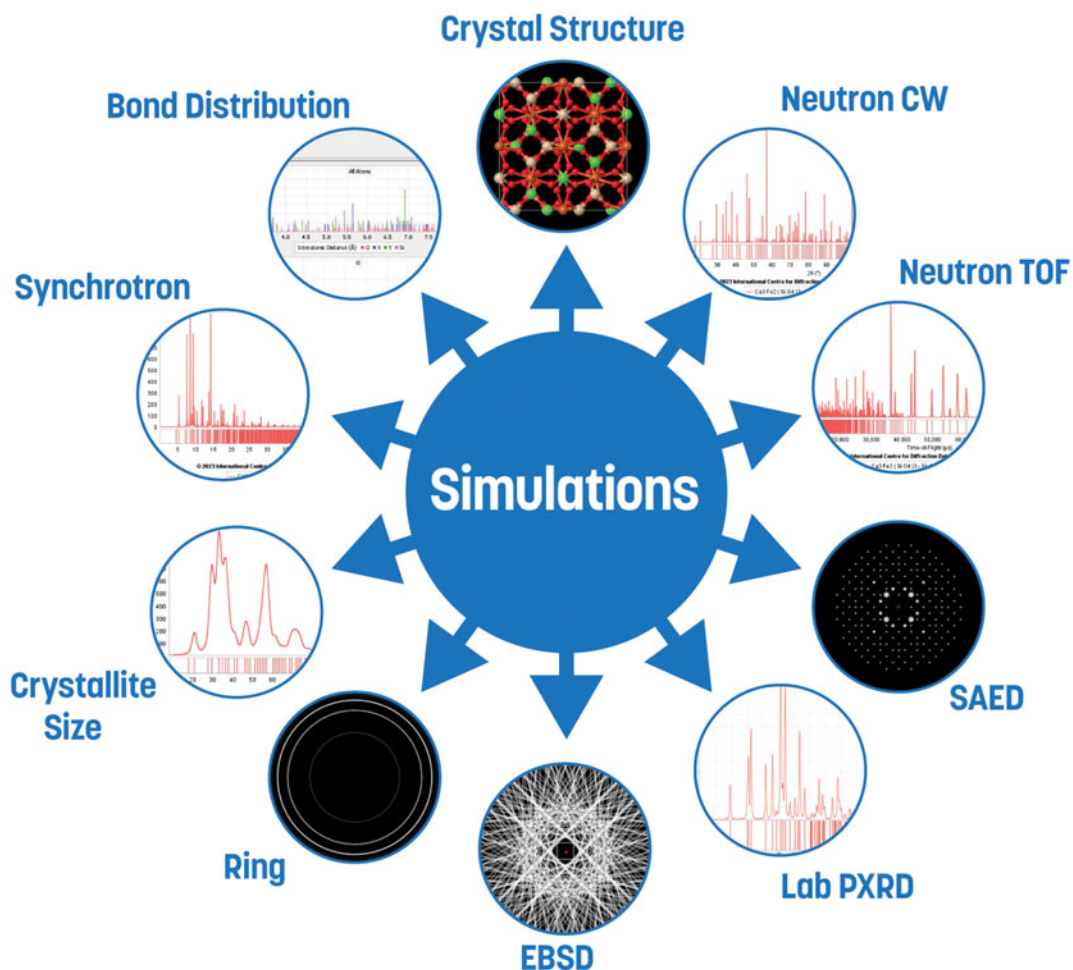


Figure 10. Simulation features on a PDF entry.

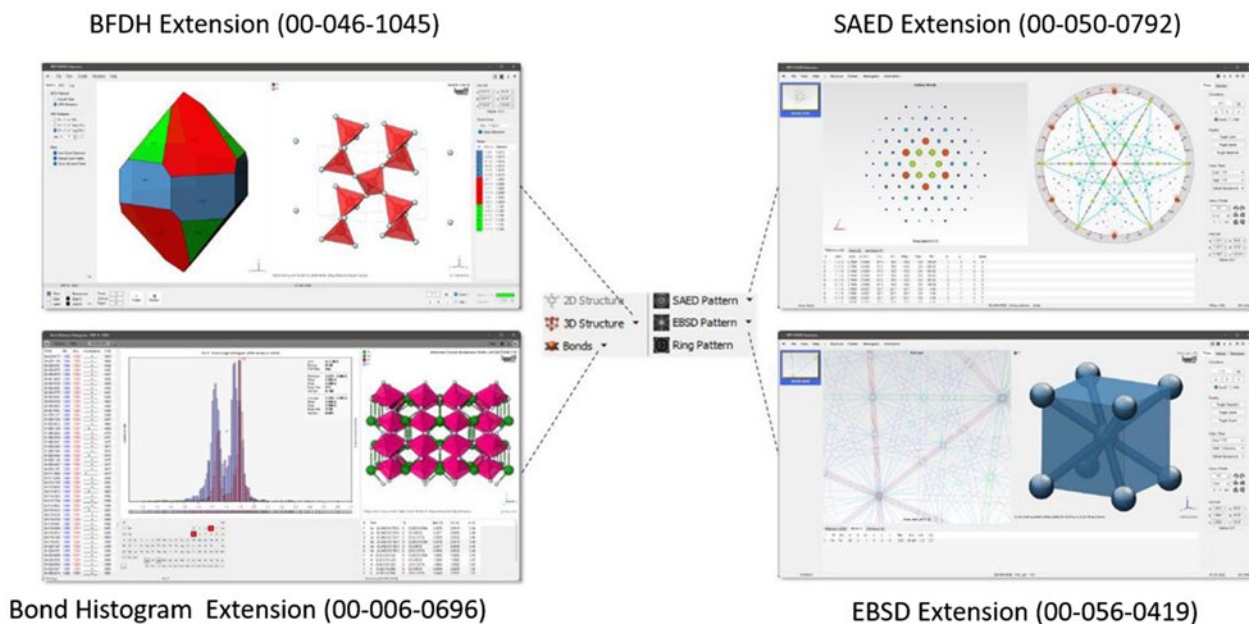


Figure 11. Launching JADE toolkit extension from a PDF entry using the dropdown arrow next to 3D structure, bonds, SAED pattern, and EBSD pattern icons. The PDF number of the entry used in each example is within parenthesis.

B. Selected area electron diffraction (SAED)

This module simulates SAED (Goehner and Michael, 1996) and offers a feature to import SAED images for comparison. The SAED extension feature offers kinematic simulations based on the crystal structure, SAED image import, processing, measurement tools, and grid indexing. This module offers powerful graphical features like synchronized structure viewing (Vector Graphics) and stereographic projections on a Wulff net. One of the strong features of this module is searching PDF-5+ database for 2D cells using an imported SAED image.

C. Electron back scattering diffraction

This simulation tool displays backscattered electron Kikuchi patterns (Goehner and Michael, 1996). The EBSD toolkit extension provides an interactive mode of displaying k-patterns (Kikuchi or Kossel) and their corresponding stereographic projections.

D. Bond histogram

The bond histogram module can be used to see the bond length distribution for a user-defined pair of atoms. Figure 13

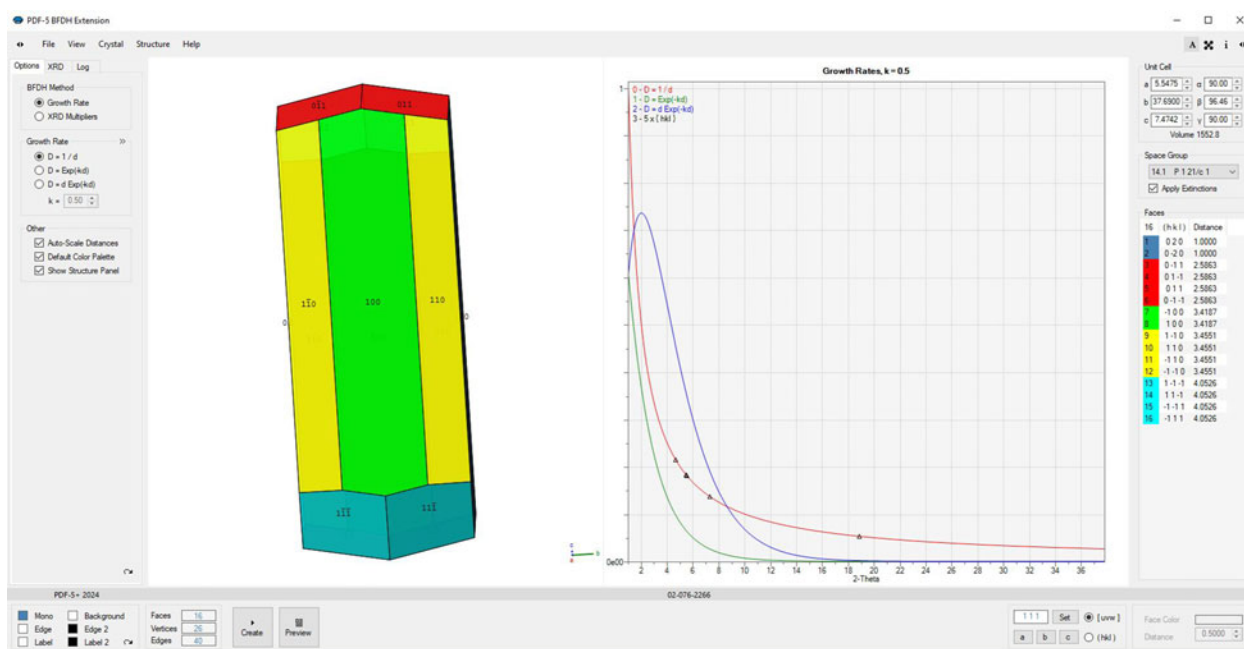


Figure 12. Crystal morphology prediction of Eupatorin (PDF# 02-076-2266) using the BFDH extension.

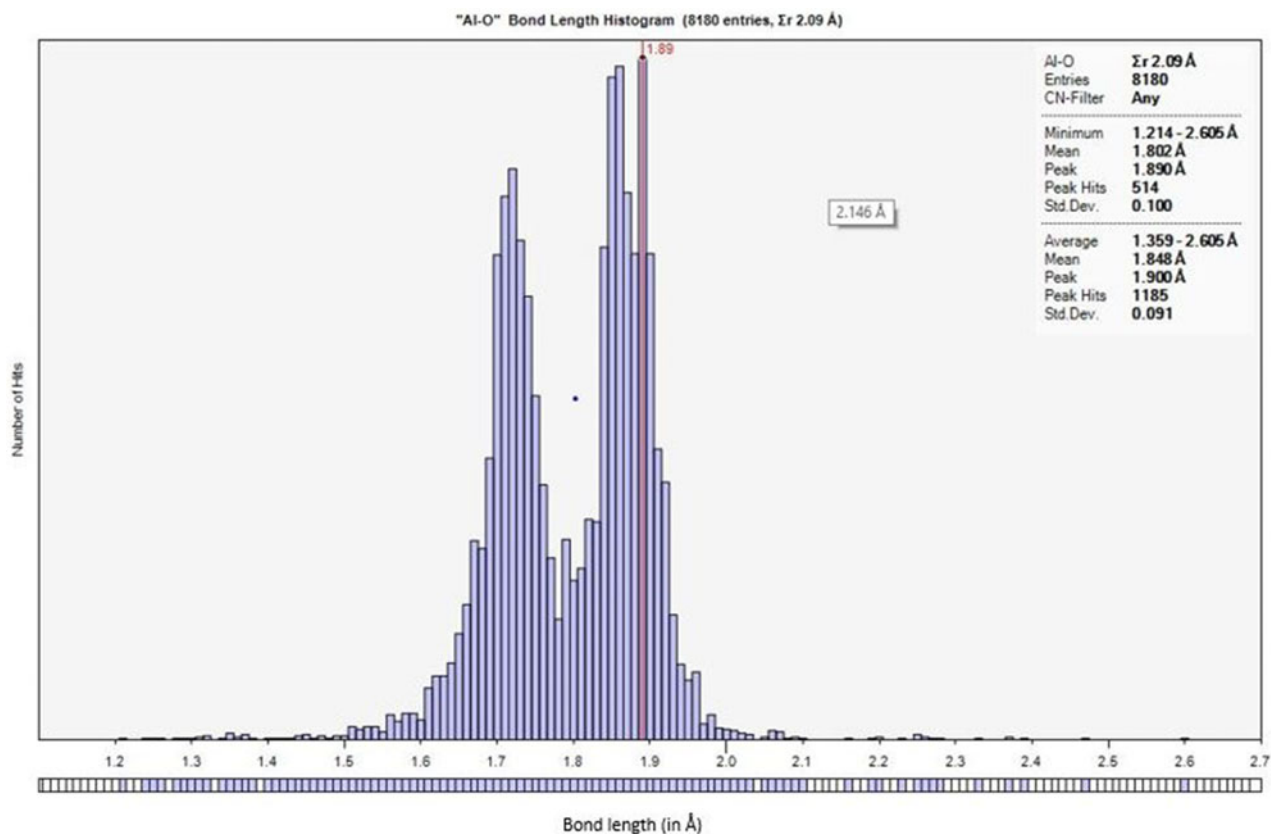


Figure 13. Al–O bond length histogram generated using crystal structures in PDF-5+ with Al–O bonds.



Figure 14. Example of phase identification using 2D diffraction data.

TABLE IV. Contents of a PDF entry.

| | | | |
|---|-------------------------------|---|---|
| 1 | PDFID | Entry number | |
| 2 | Radiation type | X-ray, electron, neutron; wavelength, and divergence of the beam, instrument apertures, etc. | |
| 3 | <i>d</i> -I list | Interplanar spacings (<i>d</i>) & intensities (I). <i>Miller indices are listed when available</i> | |
| 4 | Tick marks | Graphical representation of peak position | |
| 5 | Diffraction(s) | Experimental profile (raw diffraction data) or simulated profile diffractogram | |
| 6 | Function keys | Tools & simulations associated with the PDF entry. Gray icons indicate the tool or simulation is not available for this entry | |
| 7 | Supplemental Information tabs | Provides additional details pertaining to: | |
| | | (A) “ PDF ” – chemistry & general info | (F) “ Classification ” – subfile(s), structure type(s) |
| | | (B) “ Experimental ” – diffraction experiment | (G) “ Cross-references ” – correlated PDF entries |
| | | (C) “ Physical ” – unit cell data (<i>from author</i>) | (H) “ Reference ” – bibliographic references |
| | | (D) “ Crystal ” – ICDD-calculated unit cell data | (I) “ Comments ” – database comments |
| | | (E) “ Structure ” – atomic structure | |

TABLE V. Details for fields on the “PDF” supplemental information tab (Tab A in Figure 8).

| | | |
|----|------------------------|--|
| A1 | Editorial designations | Status – primary, alternate, or deleted Quality mark – see TABLE II: major criteria for PDF quality mark assignment |
| A2 | Sample Conditions | Specifies the environment, temperature, and pressure of data collection. |
| A3 | Chemistry | Indicates the polymorphic designation of a given phase, as well as the formula(e), weight percent, and atomic percent associated with entry; <i>as available</i> . |
| A4 | Compound Identifier(s) | Names and registry numbers used to identify the compound |
| A5 | History | Specifies pertaining to the entries initial publication, and if applicable, recent modifications |

shows the Al–O bond length distribution using 8180 PDF-5+ entries with Al–O bonds. It is clear from the histogram that the distribution is bimodal corresponding to tetrahedral and octahedral coordination. This module also has a feature to filter by coordination number. This information is useful in crystal structure validation as well as in setting up bond length restraints during Rietveld refinements.

The PDF-5+ database also comes with integrated data mining features and a search/match program (Sieve+) for phase identification. The recent development among many in Sieve+ is the capability of performing phase identification using 2D diffraction images/data (supports several formats). Sieve+ can process, integrate, and perform search/match seamlessly using 2D diffraction image/data as shown in Figure 14. One of the most used features of Sieve+ is quantitative phase analysis using Reference Intensity Ratio method (Hubbard and Snyder, 1988).

PDF-5+ has an extensive data mining capability facilitated with 137 display fields coupled with 83 searches. A very important distinction of PDF-5+ 2024 is that out of 1,061,800+ entries, 586 700+ entries have atomic coordinates to enable quantitative phase analysis using the Rietveld method. ICDD’s PDF goes through the continuous addition of significant data and classifications each year and it is important in any research or analytical laboratory to have the most up-to-date database.

V. AVAILABILITY

ICDD offers various database products at different levels to meet user’s requirements. Product summaries, licensing info, and operating specifications for each member of the PDF product line are available on the ICDD website (www.icdd.com).

Over the years, ICDD has been closely collaborating with licensed software developers, including equipment

manufacturers and independent developers, and works diligently to cultivate commercial data analysis programs that work impeccably with our databases.

VI. CONCLUSION

Well-curated databases play a pivotal role in any successful material characterization or any data-driven studies. The exponential growth and interest in data-driven research based on machine learning (ML) and artificial intelligence (AI) make it critical to have a database one would trust. ICDD, a nonprofit scientific organization founded by scientists has responded to the growing need for database requirements in materials characterization by providing a curated, quality database, the PDF. The advances in computer speed and digital storage media size and retrieval speed have enabled combining the PDF inorganic and organic subfiles into a comprehensive database product PDF-5+ designated Release 2024.

ACKNOWLEDGEMENTS

Editing, curating, and producing a quality database and software requires a significant team effort by the ICDD staff, ICDD members and the ICDD Board of Directors. We thank hundreds of researchers and scientists who have contributed to the Powder Diffraction File and ICDD organization over the past 80+ years.

REFERENCES

- Allmann, R., and R. Hinek. 2007. “The Introduction of Structure Types into the Inorganic Crystal Structure Database ICSD.” *Acta Crystallographica Section A Foundations of Crystallography* 63 (5): 412–17. doi:10.1107/S0108767307038081.

- Baerlocher, C. H., L. B. McCusker, D. Olson, & W. M. Meier. 2007. *Atlas of Zeolite Framework Types*. 6th rev. ed. Amsterdam Boston, Published on behalf of the Structure Commission of the International Zeolite Association by Elsevier.
- Blatov, V. A., A. P. Shevchenko, and D. M. Proserpio. 2014. "Applied Topological Analysis of Crystal Structures with the Program Package ToposPro." *Crystal Growth & Design* 14 (7): 3576–86. doi:10.1021/cg500498k.
- Bruno, I., S. Gražulis, J. R. Helliwell, S. N. Kabekkodu, B. McMahon, and J. Westbrook. 2017. "Crystallography and Databases." *Data Science Journal* 16: 38. doi:10.5334/dsj-2017-038.
- Donnay, J. D. H., and D. Harker. 1937. "A New Law of Crystal Morphology Extending the Law of Bravais" *American Mineralogist* 22: 446–67.
- Dowty, E. 1976. "Crystal Structure and Crystal Growth: I. The Influence of Internal Structure on Morphology." *American Mineralogist* 61: 448–59.
- Faber, J., and T. Fawcett. 2002. "The Powder Diffraction File: Present and Future." *Acta Crystallographica Section B Structural Science* 58 (3): 325–32. doi:10.1107/S0108768102003312.
- Fawcett, T. G., S. N. Kabekkodu, J. R. Blanton, and T. N. Blanton. 2017. "Chemical Analysis by Diffraction: The Powder Diffraction File™." *Powder Diffraction* 32 (2): 63–71. doi:10.1017/S0885715617000288.
- Fawcett, T. G., S. Gates-Rector, A. M. Gindhart, M. Rost, S. N. Kabekkodu, J. R. Blanton, and T. N. Blanton. 2019. "A Practical Guide to Pharmaceutical Analyses Using X-Ray Powder Diffraction." *Powder Diffraction* 34 (2): 164–83. doi:10.1017/S088571561900023X.
- Fawcett, T. G., S. Gates-Rector, A. M. Gindhart, M. Rost, S. N. Kabekkodu, J. R. Blanton, and T. N. Blanton. 2020. "Total Pattern Analyses for Non-Crystalline Materials." *Powder Diffraction* 35 (2): 82–8. doi:10.1017/S0885715620000263.
- Gates, S. D., T. N. Blanton, and T. G. Fawcett. 2014. "A New 'Chain' of Events: Polymers in the Powder Diffraction File™ (PDF ©)." *Powder Diffraction* 29 (2): 102–7. doi:10.1017/S088571561400013X.
- Gates-Rector, S., and T. Blanton. 2019. "The Powder Diffraction File: A Quality Materials Characterization Database." *Powder Diffraction* 34 (4): 352–60. doi:10.1017/S0885715619000812.
- Gelato, L. M., and E. Parthé. 1987. "STRUCTURE TIDY – A Computer Program to Standardize Crystal Structure Data." *Journal of Applied Crystallography* 20 (2): 139–43. doi:10.1107/S0021889887086965.
- Goehner, R. P., and J. R. Michael. 1996. "Phase Identification in a Scanning Electron Microscope Using Backscattered Electron Kikuchi Patterns." *Journal of Research of the National Institute of Standards and Technology* 101 (3): 301. doi:10.6028/jres.101.031.
- Groom, C. R., I. J. Bruno, M. P. Lightfoot, and S. C. Ward. 2016. "The Cambridge Structural Database." *Acta Crystallographica Section B Structural Science, Crystal Engineering and Materials* 72 (2): 171–79. doi:10.1107/S2052520616003954.
- Hanawalt, J. D. 1986. "Manual Search/Match Methods for Powder Diffraction in 1986." *Powder Diffraction* 1 (1): 7–13.
- Hanawalt, J. D., H. W. Rinn, and L. K. Frevel. 1938. "Chemical Analysis by X-ray Diffraction." *Industrial & Engineering Chemistry Analytical Edition* 10 (9): 457–512.
- Helliwell, J. R., B. McMahon, J. Mitchell Guss, and L. M. J. Kroon-Batenburg. 2017. "The Science Is in the Data." *IUCr J* 4 (6): 714–22. doi:10.1107/S2052252517013690.
- Hofmann, D. W. M., and L. Kuleshova. 2005. "New Similarity Index for Crystal Structure Determination from X-Ray Powder Diagrams." *Journal of Applied Crystallography* 38 (6): 861–66. doi:10.1107/S0021889805023484.
- Hubbard, C. R., and R. L. Snyder. 1988. "RIR – Measurement and Use in Quantitative XRD." *Powder Diffraction* 3 (2): 74–77. doi:10.1017/S0885715600013257.
- Hubbard, C. R., J. K. Stalick, and A. D. Mighell. 1980. "NBS* AIDS80: A Fortran Program to Evaluate Crystallographic Data." *Advances in X-Ray Analysis* 24: 99–109.
- Jenkins, R. & Smith, D. (1987). Powder Diffraction File in F. H. Allen, G. Bergerhoff, and R. Sievers (Eds.) *Crystallographic Databases* (pp. 159–74). Data Commission of the International Union of Crystallography.
- Jenkins, R., M. Holomany, and W. Wong-Ng. 1987. "On the Need for Users of the Powder Diffraction File to Update Regularly." *Powder Diffraction* 2 (2): 84–87. doi:10.1017/S0885715600012434.
- Johnson, G., and V. Vand. 1967. "A Computerized Powder Diffraction Identification System." *Industrial & Engineering Chemistry* 59 (8): 19–31.
- Kabekkodu, S. N., J. Faber, and T. Fawcett. 2002. "New Powder Diffraction File (PDF-4) in Relational Database Format: Advantages and Data-Mining Capabilities." *Acta Crystallographica Section B Structural Science* 58 (3): 333–37. doi:10.1107/S0108768102002458.
- Kaduk, J. A. (2019). Crystallographic Databases and Powder Diffraction in J. A. K. C. Gilmore and H. Schenk (Eds.) *International Tables for Crystallography Volume H: Powder Diffraction*. 1st ed. (Chap. 3.7, pp. 304–24). New York, Wiley.
- Messick, J. 2012. "The History of the International Centre for Diffraction Data." *Powder Diffraction* 27 (1): 36–44. doi:10.1017/S0885715612000085.
- Pearson, W. B. 1967. *Handbook of Lattice Spacings and Structures of Metals*. Oxford: Pergamon Press.
- Reid, J. W., J. A. Kaduk, and M. Vickers. 2016. "The Crystal Structure of Trandolapril, C₂₄H₃₄N₂O₅: An Example of the Utility of Raw Data Deposition in the Powder Diffraction File." *Powder Diffraction* 31 (3): 205–10. doi:10.1017/S0885715616000294.
- Smith, D. K., P. Bayliss, and A. C. Roberts. 1997. "Mineral Classification in the Mineral Powder Diffraction File." *Advances in X-Ray Analysis* 41: 606–13.
- Villars, P., & K. Cenzual. 2018. *Pearson's Crystal Data: Crystal Structure Database for Inorganic Compounds*. Materials Park, OH, USA: ASM International.
- Villars, P., N. Onodera, and S. Iwata. 1998. "The Linus Pauling File (LPF) and Its Application to Materials Design." *Journal of Alloys and Compounds* 279 (1): 1–7. doi:10.1016/S0925-8388(98)00605-7.
- Wallace, P. L., S. Weissmann, M. H. Mueller, L. D. Calveit, and R. Jenkins. 1994. "The New ICDD Metals and Alloys Indexes: Usefulness and Potentialities." *Powder Diffraction* 9 (4): 239–45. doi:10.1017/S0885715600018947.
- Wilkinson, M. D., M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, et al. 2016. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3 (1): 160018. doi:10.1038/sdata.2016.18.
- Wong-Ng, W., H. F. McMurdie, C. R. Hubbard, and A. D. Mighell. 2001. "JCPDS-ICDD Research Associateship (Cooperative Program with NBS/NIST)." *Journal of Research of the National Institute of Standards and Technology* 106 (6): 1013. doi:10.6028/jres.106.052.
- Zagorac, D., H. Müller, S. Ruehl, J. Zagorac, and S. Rehme. 2019. "Recent Developments in the Inorganic Crystal Structure Database: Theoretical Crystal Structure Data and Related Features." *Journal of Applied Crystallography* 52 (5): 918–25. doi:10.1107/S160057671900997X.