International Tables for Crystallography (2018). Vol. H, Section 3.7.4, pp. 314–316.

searches can include the number of H atoms bonded to a particular atom, the charge, the number of bonded atoms and whether the atom is part of a ring. In addition, three-dimensional quantities can be defined, tabulated and analysed. These quantities can be analysed in *Mercury* (Macrae *et al.*, 2008) and *Mogul* (Bruno *et al.*, 2004). Such analyses are useful for defining geometrical restraints in a Rietveld refinement. A general practice is to use the mean and standard deviations directly output by *Mogul* for the restraints. It is important to understand the CSD conventions for defining bond types to obtain successful results.

In *ConQuest*, searches can be carried out on author and/or journal name, as well as the normal bibliographic characteristics. Compounds can be located by chemical and/or common names, but such searches should be complemented by chemical-connectivity searches. It is possible to limit the search universe by chemical class, including carbohydrates, nucleosides and nucleotides, amino acids, peptides and complexes, porphyrins, corrins and complexes, steroids, terpenes, alkaloids and organic polymers. Searches on elements and formulae are possible, as well as searches on space groups and crystal systems. Particularly useful in searching for structure analogues are reduced-cell searches. Queries on Z, Z' and density are useful in data mining. A wide variety of searches on experimental parameters are possible; there is an option to exclude powder structures. Searches on both pre-defined terms and general text searches are possible. Particularly convenient for users of the PDF-4/Organics database is the retrieval of individual refcodes; the refcode from the PDF entry can be input directly into *ConQuest*. Starting with the 2013 release of the PDF-4/Organics database this link is live; the display of a PDF entry will result in import of the coordinates from the CSD entry. Boolean operations can be used to combine search queries in many flexible ways.

In recent years, many (if not most) single-crystal structures have been determined at low temperatures, while most powder-diffraction measurements are made under ambient conditions. Thermal expansion (often anisotropic) can result in differences between the observed peak positions and those in a PDF-4/Organics entry calculated from a CSD entry. For successful phase identification, larger than default tolerances must often be used in the search/match process. Transparency effects in pure organic compounds can also lead to significant peak shifts to lower angles, as well as significant asymmetry, so wider search windows may be necessary for phase identification.

3.7.3.1. *Mercury*

The structure-visualization program *Mercury* (Macrae *et al.*, 2008) is available as a free version and as a version with the CSD which has additional capabilities that are useful for powder diffraction. *Mercury* reads and writes a variety of molecular and crystal structure file formats, but is most commonly used with CIFs. Structures can be edited, and among the edit options is Normalize Hydrogens. This is particularly useful to improve the approximate H-atom positions that are often used in the early and intermediate stages of a Rietveld refinement. It is always worth including the H atoms in the structure model (in at least approximate positions) because better residuals and improved molecular geometry are obtained.

The Display Symmetry Elements tool is particularly useful for teaching symmetry. The Display Voids tool is useful in validating structures after solution. For most materials (zeolites and metal–organic frameworks are notable exceptions) we do not expect empty spaces in the crystal structure, so the presence of voids suggests the presence of an incomplete structure model and/or errors.

Among the options in the Calculate menu is Powder Pattern. The calculation can be customized by the user to match the desired instrumental configuration. The calculated pattern can be saved in several formats for comparison in the user's instrument software. *Mercury* expects the displacement coefficients to be given as U values. CIFs can come from many sources and can use different conventions for the displacement coefficients or may be missing them entirely. Manual editing of the input CIF is often required, otherwise strange powder patterns can be calculated.

Among the CSD-Materials/Calculations options is BFDH morphology (Bravais, 1866; Friedel, 1907; Donnay & Harker, 1937). Although a simple calculation, it is often realistic enough to suggest the likelihood of profile anisotropy and preferred orientation, along with expected directions. The calculation can thus save guessing about preferred directions.

The Structure Overlay and Molecule Overlay options are very useful for comparing structures quantitatively. There is also an interface to the semi-empirical code *MOPAC*, which can also be a useful tool for assessing structural reasonableness. The H-Bonds and Short Contacts options are useful in completing a structure solved using powder data, as often the 'interesting' H-atom positions have to be deduced. There is a relatively new Solid Form menu, which contains several tools for analysing crystal structures.

3.7.4. Inorganic Crystal Structure Database (ICSD)

The Inorganic Crystal Structure Database (ICSD; https://icsd.fiz-karlsruhe.de; Bergerhoff & Brown, 1987; Belsky *et al.*, 2002; Hellenbrandt, 2004) strives to contain an exhaustive collection of inorganic crystal structures published since 1913, including their atomic coordinates. It is a joint project between FIZ Karlsruhe and NIST. The database is accessed through the online WebICSD or the locally-installed program *FINDIT*. Typical interatomic distances in inorganic compounds derived from the ICSD have been collected in Chapter 9.4 of *International Tables for Crystallography* Volume C (Bergerhoff & Brandenburg, 1999). Applications of the ICSD have been discussed by Kaduk (2002), Behrens & Luksch (2006) and Allmann & Hinek (2007).

The ICSD began as an inorganic crystal structure database of published structures with atomic coordinates. The scope was gradually extended to include intermetallic compounds. Since 2003, FIZ Karlsruhe has started to fill in the gaps, and the aim is for the ICSD to include all published intermetallic compounds. Originally the ICSD did not contain structures with C—H or C—C bonds. After 2003, this rule was modified so that new entries should not contain both C—H and C—C bonds; compounds containing tetramethylammmonium and oxalate ions are now included.

The ICSD contains fully determined structures with atomic coordinates. Coordinates of light atoms (such as H atoms) or extra-framework species (such as in zeolites) may be missing. Structures described as isotypic to other structures, but without determination of the atomic coordinates, are included using the coordinates from the corresponding structure-type prototype. Such entries get a special remark/comment: 'Cell and Type only determined by the author(s). Coordinates estimated by the editor in analogy to isotypic compounds.' Currently there are more than

26 000 entries with derived coordinates. At present, the ICSD contains more than 187 000 entries, including 2033 crystal structures of elements, 34 785 records for binary compounds, 68 730 records for ternary compounds and 68 083 records for quaternary and quintenary compounds. About 149 000 entries have been assigned a structure type; there are currently 9093 structure prototypes.

Most of the structures contained in the ICSD are from published journal articles, although private communications are also accepted. The entries are tested for formal errors, plausibility and logical consistency. The data are stored as published; the authors' settings of space groups are considered to be valuable information which should not be changed. Only some 'exotic' space groups are transformed. In addition, for each entry in the ICSD the structure is standardized using the program *STRUCTURE TIDY* by Gelato & Parthé (1987). The published cell, standardized cell and reduced cell are all searchable. Since 2003, FIZ Kalrsruhe has been assigning structure-type classifications (Allmann & Hinek, 2007). In the future, this feature will enable easier searches for compounds that are closely related in structure.

3.7.4.1. *General features of the ICSD*

The chemical name is given in English following IUPAC rules, with the oxidation state in roman numerals. The formula upon which the name is based is calculated from all atoms with defined coordinates. Phase (polymorph) designations are given after a hyphen. Mineral names and group names are given for all entries that correspond to minerals. Details of the origin are given after a hyphen. The formula is coded as a structural formula, which provides the opportunity to search for typical structure units (such as SiO_4). Such searches can be useful, but can easily miss structurally similar compounds, and should be used with caution.

The title of the publication is given in English, French or German. There can be several citations, but an author list is only given for the first reference. I have encountered truncated author lists. Authors' surnames can vary when the original publication uses a non-roman alphabet. In some cases, the first and last names of Chinese authors may be interchanged.

The Hermann–Mauguin space-group symbol is given according to the conventions of *International Tables for Crystallography* Volume A. If different origin choices are available, those space groups with the origin at a centre of symmetry (origin choice 2) are characterized by an additional '*z*', while an additional '*s*' is used for special origins (origin choice 1). Thus, the space group for magnetite may be reported as *Fd*-3*mz* or *Fd*-3*ms*, depending on which origin the authors used. Since all contemporary Rietveld programs use origin choice 2, care must be taken when importing coordinates.

Along with the fractional coordinates, atom identifiers are reported. These are principally running numbers and may differ from those reported by the authors. The oxidation state is given with a sign. When importing coordinates into a Rietveld program these oxidation states can influence which scattering factors are used, and so should be examined by the user. Both site multiplicities and Wyckoff positions are generated for all atoms.

The ICSD archives displacement coefficients (both isotropic and aniostropic) according to what the authors reported. Isotropic displacement coefficients can be given as either *B* or *U* values and anisotropic coefficients can be given as *β*, *B* or *U* values (or, in rare cases, using other conventions). Displacement coefficients imported into a Rietveld program should always be checked, as it is common for the program to interpret *B* as *U* and *vice versa*. Such wrong displacement coefficients can make Rietveld refinements hard to perform. There are a number of standard remarks and standard test codes; these text fields can be useful for limiting the universe of the search (such as for neutron-diffraction structures).

3.7.4.2. *Features particularly useful for powder crystallography*

A field which is particularly useful for identifying structural analogues is the ANX formula. This formula is generated according to the following rules:

(i) H^+ is not taken into account, even if coordinates are available.

(ii) The coordinates of all sites of all other atoms must be determined.

(iii) Different atom types on the same positions (for example, in solid solutions) are treated as a single atom type.

(iv) An exception: if cations and anions occupy the same site they will not be treated as one atom type.

(v) All sites occupied by the same atom type are combined, unless the oxidation state is different. Thus, $Fe^{2+}(Fe^{3+})_2O_4$ yields AB2X4, while $(Fe^{2.667+})_3O_4$ yields A3X4.

(vi) For each atom type, the multiplicities are multiplied by the site-occupancy factors and the products are added. The sums are rounded and divided by the greatest common divisor.

(vii) If the rounded sum equals zero, all sums are multiplied by a common factor so that the smallest sum equals unity, so no element will be omitted.

(viii) Cations are assigned the symbols A–M, neutral atoms are assigned N–R and anions are assigned X, Y, Z and S–W.

(ix) The symbols are sorted alphabetically and the characters are assigned according to ascending indices: AB2X4, not A2BX4.

(x) All ANX symbols with more than four cation symbols, three neutral atom symbols or three anion symbols are deleted.

The utility of these symbols is illustrated by the fact that the three garnets $Mg_3Al_2(SiO_4)_3$, $Ca_3(Al_{1.34}Fe_{0.66})Si_3O_{12}$ and $(Mg_{2.7}Fe_{0.3})(Al_{1.7}Cr_{0.3})Si_3O_{12}$ all yield ANX = A2B3C3X12.

Reduced-cell searches [see *International Tables for Crystallography* Volume A, Section 3.1.3 (de Wolff, 2016)] are particularly easy to carry out in the 'Cell' section of Advanced Searches. Once a unit cell has been determined by indexing the powder pattern, it is always worth carrying out a reduced-cell search to identify potential isostructural compounds using lattice-matching techniques. It is often wise to first carry out such a search using relatively narrow tolerances (say, 1% on the lattice parameters) and then carry out additional searches using larger tolerances. Systematic searches of the subcells and supercells of a given unit cell, as could be carried out using the *NBS*LATTICE* program with the NIST Crystal Data Identification File (Mighell & Himes, 1986; Mighell, 2003), are not yet implemented.

Under the 'Crystal Chemistry' section it is possible to search for crystal structures that contain bonds between particular atom types in a distance range. Such searches are particularly valuable in assessing the chemical reasonableness of crystal structures, such as the study by Sidey (2013) on the shortest $B^{III}{-}O$ bonds.

Because the ICDD Powder Diffraction File '01' entries contain the ICSD collection code in the comments, searching for the collection code of a hit in a search/match is particularly easy in the 'DB Information' section. In this way, the relevant ICSD entry can be located without any ambiguity and the best structure for the problem at hand can be used to start the Rietveld refinement.

3.7.5. Pearson's Crystal Data (PCD/LPF) (with Pierre Villars and Karen Cenzual)

3.7.5.1. *General information*

The Pearson's Crystal Data database (PCD; Villars & Cenzual, 2013) is an outgrowth of the (Linus) Pauling File (LPF; Villars *et al.*, 1998; http://www.paulingfile.com), which was designed to combine crystal structures, phase diagrams and physical properties under the same computer framework to form a tool useful for materials design. PCD is the result of a collaboration between Material Phases Data Systems (Vitznau, Switzerland) and ASM International (Materials Park, Ohio, USA). The retrieval software was developed by Crystal Impact (Bonn, Germany). As suggested by the name, Pearson's Crystal Data is a follow-up product to *Pearson's Handbook: Crystallographic Data for Intermetallic Phases* (Villars & Calvert, 1985, 1991; Villars, 1997). However, in contrast to the latter, it also covers oxides and halides, which represent about 80% of the compounds with more than four chemical elements.

The 2016/2017 release of Pearson's Crystal Data contains more than 288 000 data sets for more than 165 300 different chemical formulae, representing over 53 000 distinct chemical systems. To achieve this, the editors have processed over 93 500 original publications; recent literature is surveyed in a cover-to-cover approach, including about 250 journal titles. Over 153 000 database entries contain refined atom coordinates, as well as isotropic and/or anisotropic displacement parameters when published, whereas more than 72 000 data sets contain atom coordinates corresponding to the structure prototype assigned by the authors of the original publication or by the database editors. Approximately 15 000 data sets contain only crystallographic data such as the lattice parameters and possibly a space group.

When available in the original publications, each data set contains comprehensive information on the sample-preparation and experimental procedure, as well as on the stability of the phase with respect to temperature, pressure and composition. The presence of plots (cell parameters or diffraction patterns) in the original paper is indicated, and over 30 000 descriptions of the variation of the cell parameters as a function of temperature, pressure or composition are proposed. Roughly 18 300 experimental diffraction patterns are reported.

The Linus Pauling File was designed as a phase-oriented, fully relational database system. This required the creation of a 'distinct phases' table, with internal links between the three parts of the database. In practice, this means that the senior editors have evaluated the distinct phases existing in the system for every chemical system using all information available in the LPF. Each structure entry in Pearson's Crystal Data has been linked to such a distinct phase, which allows a rapid overview of a particular chemical system.

3.7.5.2. *Evaluation procedure*

Extensive efforts have been made to ensure the quality and reliability of the crystallographic data. Pearson's Crystal Data is checked for consistency by professional crystallographers, assisted by an original software package, *ESDD* (*Evaluation, Standardization and Derived Data*), containing more than 60 different modules (Cenzual *et al.*, 2000). The checking is carried out progressively, level by level. The following checks are made.

Individual database fields:

(i) order of magnitude of numerical values;
(ii) Hermann–Mauguin symbols, Pearson symbols;
(iii) consistency of journal CODEN, year, volume, first page, last page;
(iv) formatting of chemical formulae;
(v) neutrality of oxides and halides;
(vi) spelling.

Consistency within individual data sets:

(i) atom coordinates, Wyckoff letters, site multiplicities;
(ii) chemical elements in different database fields;
(iii) computed, published values (cell volume, density, absorption coefficient, *d*-spacings);
(iv) Pearson symbol, space group, cell parameters;
(v) Bravais lattice, Miller indices;
(vi) site symmetry, anisotropic displacement parameters.

Particular crystal-structure checks:

(i) interatomic distances, sum of atomic radii;
(ii) geometry of functional groups;
(iii) search for overlooked symmetry elements;
(iv) composition from refinement, chemical formula.

Consistency within the database:

(i) comparison of cell-parameter ratios for isotypic entries;
(ii) comparison of atom coordinates for isotypic entries with refined coordinates;
(iii) comparison of densities;
(iv) thorough search for duplicates, also considering translated references.

Wherever possible, misprints have been corrected based on arguments explained in remarks; as a result, more than 13 000 crystallographic data sets are accompanied by at least one erratum. In other cases remarks drawing the attention to discrepancies or unexpected features have been added.

The *ESDD* software package also produces derived data such as the Niggli reduced cell, equivalent isotropic displacement parameters, density and formula weight.

3.7.5.3. *Standardized crystallographic data*

The crystallographic data in Pearson's Crystal Data are presented as published, respecting the original site labels, but are also standardized following the method proposed by Parthé and Gelato (Parthé & Gelato, 1984, 1985; Parthé *et al.*, 1993). This second presentation of the same data has been further adjusted so that compounds crystallizing with the same prototype structure (isotypic compounds) can be easily compared. It is prepared in a three-step procedure as follows.

(i) The crystallographic data are checked for the presence of overlooked symmetry elements. Whenever it is possible to describe the structure in a higher-symmetry space group, or with a smaller unit cell, without any approximations, this is performed.
(ii) In the next step, the crystallographic data are standardized using the program *STRUCTURE TIDY* (Gelato & Parthé, 1987).
(iii) The resulting data are compared with the standardized data of the type-defining data set and, if relevant, adjusted using an *ESDD* module based on the program *COMPARE* (Berndt, 1994).

For data sets with no published coordinates, the cell parameters are standardized following the criteria defined for the unit-cell

references