

## 3. METHODOLOGY

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 tetramethylammonium 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 isotopic 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 isotopic compounds.' Currently there are more than