

## 3. METHODOLOGY

possible and the desired level of structural studies (*e.g.* complete) may be specified.

Many more searches can be carried out in the complete Search Dialog. Particularly useful are searches on atomic environments and interatomic distances. The atomic environment is defined by the coordination number, the geometry of the coordination polyhedron and the identities of the central and peripheral atoms. Searches on the number of different atomic environment types in the same structure can also be carried out. Specifying a pair of elements makes it possible to select a range of interatomic distances to be included in the search. These histograms are also useful in assessing the reasonableness of a particular distance in a Rietveld refinement.

## 3.7.5.8. Particular software features

All searches use the 'Perpetual Restraining' feature, which updates the selection set in real time as a new query is introduced, so that the progress of the search scheme can easily be monitored. The complete Search Dialog offers a large variety of features that make the retrieval and presentation of information extremely flexible.

The Chemical System Matrix View makes it easy to locate phases in binary, ternary, quaternary and pseudo-quaternary systems. The Phases List View collects a selection set into its 'distinct' phases. From the individual database entry it is particularly easy to find all database entries with the same prototype structure and plot the unit-cell volume as a function of selected atomic radii. The standard display of an entry includes a short summary about the phase, structural, bibliographic, experimental and editorial data, as well as a structure drawing, a powder pattern and a table of interatomic distances.

The software for producing structure drawings offers the visualization of atomic environments (coordination polyhedra), the statistics of interatomic distances and the calculation of selected distances and angles. Four different models are available (ball and stick, wires, sticks and space-filling), with on-the-fly rotation controlled by the mouse. The nearest-neighbour histogram of a selected atom is compared with a statistical plot containing all distances in the database involving the same chemical elements, and the atomic environments can be instantaneously modified by clicking on the nearest-neighbour histogram.

Powder-diffraction patterns can be computed for any user-defined wavelength and the visualization includes a tool for zoom-in/out tracking. Patterns based on published lists of interplanar spacings can also be visualized. It is further possible to export database entries as CIF files, tables (*e.g.* powder-diffraction pattern, distances and angles) or graphics (*e.g.* structure drawings; BMP, GIF, JPG, PNG, TIFF or Diamond documents), and individually tailored dossiers can be designed and printed.

## 3.7.6. Metals data file (CRYSTMET)

CRYSTMET (White *et al.*, 2002) began as a database of critically evaluated crystallographic data for metals, including alloys, intermetallics and minerals, and has grown to include inorganic compounds in general. It was started in 1960 by Cromer and Larson at Los Alamos National Laboratory, and its development was continued by the National Research Council of Canada. In 1996, the production and dissemination was transferred to Toth Information Systems.

CRYSTMET contains chemical, crystallographic and bibliographic data, together with comments regarding experimental details for each study. Using these data, a number of associated data files are generated, with the major one being a file of calculated powder patterns. Entry into CRYSTMET is *via* a number of search screens, including chemistry, bibliographic information, unit cell and reduced cell, powder patterns (using the positions of the strongest peaks as input), formula, structure type, Pearson symbol and space group. The results of queries reside in sets, which can be further manipulated using logical operations.

The results are displayed as a series of screens, which include crystallographic data, distances and angles, and the powder pattern. There is some ability to customize the calculation of the powder pattern of an entry; the calculation is performed for Debye–Scherrer geometry. Included on the Results tabs is a direct interface to the *MISSYM* program (Le Page, 1987, 1988), which searches the reported structure for additional symmetry elements. This is a very useful tool for detecting missed symmetry.

## 3.7.7. Protein Data Bank (PDB)

The Protein Data Bank is described in Chapter 24.1 of *International Tables for Crystallography* Volume F (Berman *et al.*, 2011). Current information is available on the web at <https://www.wwpdb.org/>.

## 3.7.7.1. Powder diffraction by proteins

Although powder-diffraction techniques had been applied to proteins as long ago as 1936 (Wyckoff & Corey, 1936; Corey & Wyckoff, 1936), and proof-of-principle experiments had been carried out (Rotella *et al.*, 1998, 2000), real progress in protein powder crystallography began with the work of Von Dreele (Von Dreele, 1998, 1999, 2003; Von Dreele *et al.*, 2000).

Progress in powder crystallography on macromolecules has been reviewed by Margiolaki & Wright (2008) and is also discussed in Chapter 7.1 of this volume. Notable studies include the characterization of the binding of *N*-acetylglucosamine oligosaccharides to hen egg-white lysozyme (Von Dreele, 2007*a*) and determination of the second SH3 domain of ponsin (Margiolaki *et al.*, 2007).

As with all powder diffraction, peak overlap ultimately limits the information available. Multi-pattern strategies to overcome the overlap problem have been investigated by Von Dreele (2007*b*). Multiple-pattern resonant-diffraction experiments have enabled study of the binding of  $\text{PtBr}_6^{2-}$  ions to lysozyme (Helliwell *et al.*, 2010). A bootstrap approach has been used to determine the structure of bacteriorhodopsin to 7 Å resolution (Dilanian *et al.*, 2011). Parametric resonant-scattering experiments have been used to determine the secondary structures of lysozyme derivatives (Basso *et al.*, 2010). Powder-diffraction experiments have also been used to gain insight into the general features of a nonstructural protein 3 (nsp3) macro domain (Papageorgiou *et al.*, 2010).

The structure of a five-residue peptide has been determined *ab initio* using laboratory powder data (Fujii *et al.*, 2011). We can expect further useful results at this interface between small-molecule and protein powder crystallography.

As is typical in other areas of science, powder diffraction has proven to be useful in more practical features of protein processing. It has been used to identify insulin (Norrman *et al.*, 2006) and GB1 (Frericks Schmidt *et al.*, 2007) polymorphs and