(*a*) If the default indexing process fails, the unit-cell search is automatically repeated by changing some default choices, *e.g.*, increasing the tolerance value on the observed *d* values. If still no solution is obtained, the maximum (*hkl*) Miller indices assigned to the orthorhombic or monoclinic base lines are increased and the tolerance limits of the default values are halved in order to avoid the generation of wrong large unit cells.

(*b*) At the end of the first run, whatever the obtained results, a possible 2θ zero-position shift is taken into account: the indexing process starts again by applying positive and negative 2θ zero-position shifts to the original peak search result.

(*c*) An exhaustive triclinic search is performed. The dominant-zone tests that are usually carried out for the monoclinic system have been extended to include the triclinic case.

(*d*) A new figure of merit, WRIP20, more powerful than the classical M_{20}, is used. It is calculated when more than one possible cell is found and takes into account the M_{20} value, the full experimental pattern, the degree of reflection overlap, the systematically absent reflections and the number of unindexed lines (see Section 3.4.2.1).

This program is also able to index powder patterns from small proteins: see Example 4 in Section 3.4.4.6.2.

3.4.4.2.2. *DICVOL06 (Louër & Boultif, 2006, 2007) and DICVOL14 (Louër & Boultif, 2014)*

The most recent of a series of versions, *DICVOL14* is the successor of *DICVOL04* (Boultif & Louër, 2004) and *DICVOL06*. *DICVOL06* includes *DICVOL04* with its optimized search procedure and an extended search in shells of volumes. *DICVOL04* represented an improvement of *DICVOL91*. Among the features of *DICVOL06* are:

(*a*) A tolerance for unindexed lines that can result from the presence of unwanted additional phases or inaccurately measured peaks. The program can tolerate a user-defined number of unindexed lines. Care must be taken when using this option to avoid the possibility of generating erroneous cells. It is worth noting that the inclusion of the possibility of at least one unindexed peak has markedly increased the success rate of *DICVOL06*.

(*b*) A correction of the zero-point error in the measured data. *Via* an *a priori* zero-origin evaluation, two different approaches can be followed: (i) if there is a non-negligible zero shift (*i.e.*, ~0.1°), the reflection-pair method is adopted (Dong *et al.*, 1999); (ii) if the shift is small (<0.03°), a refinement of the experimental data zero point together with the cell parameters is carried out as soon as a solution is found. In the monoclinic and triclinic systems, a reduced-cell analysis is performed to choose among equivalent solutions.

(*c*) When a solution is found in a 400 Å^3 shell of volume, the exhaustive search is extended to the whole domain.

No formal limits on the number of input Bragg peaks have been established but, for reliable indexing, it is recommended that 20 or more peaks (in the low-2θ region) are used.

Compared to *DICVOL04*/*DICVOL06*, *DICVOL14* includes: an optimization of filters in the final stages of the convergence of the successive dichotomy process; an optimization and extension of scanning limits for the triclinic case; a new approach for zero-point offset evaluation; a detailed review of the input data from the resulting unit cells; and cell centring tests. *DICVOL14* has been improved particularly for triclinic cases, which are generally the most difficult to solve with the dichotomy algorithm.

3.4.4.3. *Non-traditional indexing programs*

The indexing programs described above are based on using, for a limited number of lines, the measured positions of peak maxima as directly obtained from the experimental powder diffraction pattern. *Conograph* (Oishi-Tomiyasu, 2014*b*), which has been more recently proposed, also belongs to that group of programs. A brief description of *Conograph* follows

3.4.4.3.1. *Conograph: indexing via the topographs method*

Conograph is based on the topographs method, and its main functions are the determination of the primitive unit cell and lattice symmetry, and refinement of lattice parameters. Among the main features we note:

(1) A new Bravais-lattice determination algorithm (Oishi-Tomiyasu, 2012), which has been proved to be stable with respect to peak-position errors under very general conditions. The algorithm applies the Minkowski reduction to primitive cells and the Delaunay reduction (Delaunay, 1933) to face-centred, body-centred, rhombohedral and base-centred cells in such a way that the computational efficiency of the process is better than the Andrews & Bernstein (1988) method.

(2) The two figures of merit $M_n^{\rm Rev}$ and $M_n^{\rm Sym}$ proposed by Oishi-Tomiyasu (2013) are used for for selecting the true unit cells. They are also used to estimate the zero-point shift.

(3) The use of many observed peaks in the default setting, which aims to make *Conograph* robust against dominant zones and missing or false peaks (Oishi-Tomiyasu, 2014*b*).

(4) The method for exhaustively searching unit cells that involve geometrical ambiguity (Oishi-Tomiyasu, 2014*a*, 2016). The geometrical ambiguities that are detected also include lattices with very similar calculated lines, because of the error tolerance in the *d* spacings.

Programs that use only the measured positions of peak maxima are particularly vulnerable to experimental errors in the measured peak positions and to the presence of impurity peaks. For these reasons, at the end of the 1990s new indexing strategies were developed that do not require the peak locations in the experimental pattern. These approaches are completely different from the methods described above because they use the whole diffraction profile. They try to explore the parameter space (direct space) exhaustively by applying different optimization techniques in order to find the cells in best agreement with the experimental powder diffraction pattern. Some of the most widely used indexing programs in direct space are described here.

3.4.4.3.2. *GAIN: indexing via a genetic-algorithm search method*

The use of genetic algorithms (GAs) for indexing powder diffraction data by exploiting the diffraction geometry (as in the traditional indexing methods) was firstly proposed by Tam & Compton (1995) and Paszkowicz (1996). Subsequently, Kariuki *et al.* (1999) applied GA techniques by using whole profile fitting with the aim of exploring the parameter space $\{a, b, c, \alpha, \beta, \gamma\}$ and finding the global minimum of the *R*-factor $\{a, b, c, \alpha, \beta, \gamma\}$ hypersurface, yielding the parameter set able to generate the best agreement between the observed and calculated powder diffraction patterns.

This new strategy has been implemented in the program *GAIN* (Harris *et al.*, 2000), whose main features are:

(1) Starting from a population of N_p sets of lattice parameters and using the evolutionary operations of mating, mutation and natural selection, the population is allowed to evolve

through several generations, with the aim of generating sets of possible trial cell parameters.
(2) The search procedure, using a GA, is performed in restricted, sensible cell-volume ranges consistent with the knowledge of the system under study.
(3) For each set of trial parameters a calculated powder diffraction pattern is constructed. The peak positions and parameters describing the shape and width of each peak are used in the Le Bail profile-fitting procedure (Chapter 3.5).
(4) The pattern is split into different regions (defined by the user), and the weighted profile R factor is calculated for each region; all the values are summed to obtain the overall $R'_{\rm wp}$:

$$R'_{\rm wp} = \sum_{\rm regions} \left[\frac{\sum_i w_i (y_i - y_{ci})^2}{\sum_i w_i y_i^2} \right]^{1/2},$$

where the summation is over the regions, i runs over the experimental points belonging to each region and y_i and y_{ci} are the observed and calculated profile at the ith experimental step, respectively. *Via* the $R'_{\rm wp}$ formula the residual for each region is scaled according to the total intensity in the region, so a region with only low-intensity peaks can make an important contribution to $R'_{\rm wp}$.

This approach is robust at handling the problems that may affect the experimental powder pattern: peak overlap, (*hkl*)-dependent effects and zero-point errors. It is time consuming (particularly in the case of low symmetry) but not very sensitive to the presence of minority impurity phases.

3.4.4.3.3. *McMaille: indexing via a Monte Carlo search method*

The information in the whole powder diffraction profile is exploited by the program *McMaille* (Le Bail, 2004), which is based on the random generation of cell parameters and uses the Monte Carlo optimization technique. Once the trial cell parameters have been generated and the Miller indices and the peak positions have been calculated, the quality of the cell is assessed by using, as figure of merit, the conventional Rietveld profile reliability factor R_p (Young, 1993) or McM_{20} (see Section 3.4.2.1). The program uses some tricks that can increase the success of the Monte Carlo algorithm:
(1) Only the trial cells corresponding to a value of R_p that is smaller than a user-defined value (∼50%) are retained for successive refinement.
(2) If all the observed peaks, except for a user-defined number of tolerated impurity peaks, are 'explained' whatever the R_p value, the cell is retained for successive examination.
(3) If either of the conditions (1) or (2) is fulfilled, the cell parameters are randomly changed in 200 to 5000 attempts (for cubic to triclinic cases, respectively) in which small random parameter variations *via* the Monte Carlo algorithm are carried out. The new parameters are preserved if an improvement of R_p is verified in 85% of the attempts.

This procedure is not sensitive to impurity lines, provided that the sum of their intensities is less than 10–15% of the total intensity. A zero-point error up to 0.05° is tolerated. To reduce the long computing time required to successfully complete the procedure, a significant increase in speed has been obtained by using idealized profiles generated by applying simplified line profiles to extracted line positions. A parallelized version of *McMaille* has also been developed. The indexing problem can usually be solved in few minutes if: (*a*) no triclinic symmetry is handled (because this requires more computing time); (*b*) the cell volume is less than 2000 Å^3; (*c*) no cell length is longer than 20 Å.

3.4.4.4. *Crysfire: a suite of indexing programs*

The *Crysfire* suite (Shirley, 2002) is a multi-program indexing facility. It can perform a self-calibration, which is aimed at detecting and correcting 2θ zero errors, and is able to strip out weak lines. Its single unified user interface and data-file format make a wide set of indexing packages accessible with minimal effort (especially to non-specialists). *Crysfire* provides a list of the possible cells suggested by each indexing program, suitably ranked. The *Crysfire* 2003 suite supports a total of 11 programs (Bergmann *et al.*, 2004), among which are *ITO*, *TREOR90*, *DICVOL91* and *McMaille*. The possibility of using different indexing programs, working in parameter space or index space and adopting different indexing approaches increases the probability of finding the correct cell.

3.4.4.5. *Two commercial programs*

3.4.4.5.1. *SVD-Index*

This commercial indexing program (Coelho, 2003*a*), which uses the Monte Carlo method, is part of the *TOPAS* (Coelho, 2003*b*) suite from Bruker AXS. The reciprocal-cell parameters in equation (3.4.2) are found by using, in an iterative way, the singular value decomposition (SVD) approach (Nash, 1990) to solve linear equations relating (*hkl*) values to d spacings. The method is particularly useful in cases for which there are more equations than variables. All the observed lines in the powder pattern are involved in the indexing procedure. It is claimed that the program is relatively insensitive to impurity peaks and missing high d spacings; it performs well on data with large diffractometer zero errors.

More recently, two indexing methods have been introduced in *TOPAS*: LSI (least-squares iteration), an iterative least-squares process which operates on the d-spacing values extracted from reasonable-quality powder diffraction data, and LP-Search (lattice parameter search), a Monte Carlo based whole-powder-pattern decomposition approach independent of the knowledge of the d-spacings (Coelho & Kern, 2005).

3.4.4.5.2. *X-CELL*

This commercial program is part of the *Materials Studio* suite from Accelrys (Neumann, 2003). To perform an exhaustive search, like *DICVOL*, the program uses the successive-dichotomy approach. Its principal features are:
(1) the user can define how many impurity lines can be tolerated;
(2) a search for the zero-point shift of the diffraction pattern; and
(3) systematic absences are taken into account.

The program is described as 'virtually exhaustive'; it is expected to work well when faced with missing lines, impurities and errors.

3.4.4.6. *Examples of applications of indexing programs*

3.4.4.6.1. *Indexing using DICVOL06*

The program *DICVOL06*, as implemented in the *WinPLOTR*/*FULLPROF* suite (Roisnel & Rodríquez-Carvajal, 2001) and recently introduced into *EXPO*, was applied to two experimental diffraction patterns.

Example 1

Norbornene (Brunelli *et al.*, 2001). Published information: C_7H_{10}, monoclinic, a = 7.6063 (9), b = 8.6220 (1), c = 8.749 (1) Å, β = 97.24 (1)°, $P2_1/c$, experimental range 5–60° 2θ, λ = 0.85041 Å, RES = 1.0 Å (where RES is the data resolution), synchrotron data, indexed by *Fzon* (Visser, 1969).