

3.4. INDEXING

defines a domain D and, by taking into account the current limits for the parameters A , B , C and β , a calculated pattern is generated, not in terms of discrete $Q(hkl)$ values but of allowed intervals $[Q_-(hkl), Q_+(hkl)]$. D is retained only if the observed Q_i values belong to the range $[Q_-(hkl) - \Delta Q_i, Q_+(hkl) + \Delta Q_i]$, where ΔQ_i is the absolute error of the observed lines (*i.e.*, impurity lines are not tolerated). If D has been accepted, it is divided into 2^4 subdomains by halving the original intervals $[A_-, A_+]$, $[B_-, B_+]$, $[C_-, C_+]$ and $[\beta_-, \beta_+]$ and new limits $[Q_-(hkl), Q_+(hkl)]$ are calculated; if a possible solution is found, the dichotomy method is applied iteratively. In case of triclinic symmetry the expression for $Q(hkl)$ in terms of direct cell parameters is too complicated to be treated *via* the successive-dichotomy method; therefore the basic indexing equation (3.4.2) is used. In this case, the $[Q_-(hkl), Q_+(hkl)]$ intervals are set in reciprocal space according to the A_{ij} parameters of (3.4.2). To reduce computing time the following restrictions are put on the (hkl) Miller indices associated with the observed lines: (1) maximum h , k , l values equal to 2 in case of the first five lines; (2) $h + k + l < 3$ for the first two lines.

The outcome of the successive-dichotomy method is not strongly influenced by the presence of a dominant zone. New approaches have been devoted to overcome the limitations of the method with a strict dependence on data accuracy and on impurities (Boultif & Louër, 2004; Louër & Boultif, 2006, 2007), see Section 3.4.4.2).

3.4.3.2. Non-traditional indexing methods

New indexing procedures that provide alternatives to the traditional approaches outlined in Section 3.4.3.1 have recently been proposed.

3.4.3.2.1. The topographs method

This method (Oishi *et al.*, 2009) is based on the Ito equation (de Wolff, 1957):

$$Q(\mathbf{h}_1 + \mathbf{h}_2) + Q(\mathbf{h}_1 - \mathbf{h}_2) = 2[Q(\mathbf{h}_1) + Q(\mathbf{h}_2)], \quad (3.4.9)$$

where $Q(\mathbf{h})$ is the length of the reciprocal vector \mathbf{r}_{hkl}^* corresponding to the Miller index vector $\mathbf{h} = (hkl)$. It uses Conway's topograph (Conway & Fung, 1997), a connected tree obtained by associating a graph to each equation of type (3.4.9) and consisting of infinite directed edges. According to Ito's method, if quadrupoles (Q_1, Q_2, Q_3, Q_4) detected among the observed Q_i values satisfy the condition $2(Q_1 + Q_2) = Q_3 + Q_4$, two Miller-index vectors \mathbf{h}_1 and \mathbf{h}_2 are expected to exist such that $Q_1 = Q(\mathbf{h}_1)$, $Q_2 = Q(\mathbf{h}_2)$, $Q_3 = Q(\mathbf{h}_1 - \mathbf{h}_2)$ and $Q_4 = Q(\mathbf{h}_1 + \mathbf{h}_2)$. If an additional value Q_5 satisfying the condition $2(Q_1 + Q_4) = Q_2 + Q_5$ is found, the graph of the quadrupole (Q_1, Q_2, Q_3, Q_4) grows *via* the addition of the Q_5 contribution; this procedure is iterated. If topographs share a Q value that corresponds to the same reciprocal-lattice vector, then a three-dimensional lattice is derived containing the two-dimensional lattices associated with the original topographs. Three-dimensional lattices are also obtained by combining topographs. The probability that topographs correspond to the correct cell increases with the number of edges of the graph structure. The method is claimed by the authors to be insensitive to the presence of impurity peaks.

3.4.3.2.2. Global-optimization methods

Global-optimization methods, widely adopted for solving crystal structures from powder data, have also been successfully

applied to indexing. Among them, we provide brief descriptions of genetic algorithms, and Monte Carlo and grid-search methods.

3.4.3.2.2.1. Genetic-algorithm search method

The use of genetic algorithms (GAs) for solving the indexing problem was proposed by Tam & Compton (1995) and Paszkowicz (1996). Since then, Kariuki and co-workers (Kariuki *et al.*, 1999) have combined GAs with a whole-profile-fitting procedure for indexing powder diffraction patterns. This approach exploits the information of the full powder diffraction pattern. It is inspired by the Darwinian evolutionary principle based on mating, mutation and natural selection of the member of a population that survives and evolves to improve future generations. The initial population consists of a set of trial cell parameters, chosen randomly within a given volume range; a full pattern-decomposition process is performed using the Le Bail algorithm (Chapter 3.5) and the agreement between the calculated and observed profiles is derived and used for assessing the goodness of an individual member (*i.e.*, a set of unit-cell parameters). The most plausible cell is therefore found by exploring a six-dimensional hypersurface $R'_{wp}(a, b, c, \alpha, \beta, \gamma)$ and searching for the global minimum of R'_{wp} (see Section 3.4.4.3.2). In contrast to the main traditional methods, whose outcomes depend on the reliability of a set of peak positions, this procedure has the advantage of being insensitive to the presence of small impurity peaks that have a negligible influence on the agreement factor between the experimental and calculated profiles: the global minimum of R'_{wp} is reached if the majority phase is correctly indexed. The main disadvantage of the method is the computing time required, in particular in the case of low symmetry.

3.4.3.2.2.2. Monte Carlo search method

The Monte Carlo approach has also been applied to indexing powder diffraction patterns (Le Bail, 2004; Bergmann *et al.*, 2004; Le Bail, 2008). It exploits all the information contained in the full pattern, randomly generates and selects trial cell parameters, and calculates peak positions to which it assigns the corresponding Miller indices. An idealized powder pattern consisting of peak positions d and extracted intensities I is considered to test the trial cell. The cell reliability is assessed by suitable figures of merit (*e.g.* R_p and McM_{20} , see Section 3.4.2.1). The main drawback of this approach is the significant computing time required, in particular for triclinic systems.

3.4.3.2.2.3. Grid-search method

This performs an iterated 'step-and-repeat search' in the parameter space. It has the advantage of being flexible, exhaustive and not particularly sensitive to impurities or errors, and the disadvantage of being slow (Shirley, 2003).

3.4.4. Software packages for indexing and examples of their use

The different strategies and methods described in Section 3.4.3 have been implemented in a variety of automatic indexing programs (Bergman *et al.*, 2004). Almost all use one of the two different approaches working in parameter space (*i.e.*, unit-cell parameters) or index space (*i.e.*, reflection indices). Only the *EFLECH/INDEX* program (Bergman, 2007), applying the scan/covariance strategy, works in both spaces: in parameter space from cubic down to monoclinic, switching to index space for triclinic. The different indexing methods are classified according to Shirley (2003) in Table 3.4.3. Alternative classifications can be

3. METHODOLOGY

Table 3.4.3

Classification of indexing methods

Method (strategy/search)	Space	Exhaustive
Zone indexing	Parameter	No
SIW heuristic	Parameter	Semi
Scan/covariance	Parameter (for cubic to monoclinic); index (for triclinic)	To monoclinic
Index heuristics	Index	Semi
Index permutation	Index	Yes
Successive dichotomy	Parameter	Yes
Grid search	Parameter	Yes
Genetic algorithms	Parameter	No
Monte Carlo	Parameter	No

made by considering whether a program works in direct or reciprocal space, or uses Bragg diffraction line positions or the whole experimental diffraction profile. Using the whole experimental diffraction profile requires a lot of computing time, but has become possible as a consequence of recent increases in the speed of computers. In the following, descriptions of the principal (default) steps of the most widely used indexing programs are given. Several non-default options are available for each program. The chances of success of the indexing step increase if more than one program is used.

3.4.4.1. Traditional indexing programs

3.4.4.1.1. ITO (Visser, 1969)

This program is based on the zone-indexing strategy and uses the Runge–Ito–de Wolff–Visser method of decomposition of the reciprocal space into zones, as described in Section 3.4.3.1.1.

The following steps are executed by the program:

- (1) The potential zones are found always using the 20 lowest Bragg angle peaks. The program does not work with fewer peaks.
- (2) All possible combinations of the six best zones (including the combination of each zone with itself) are found by searching for trial zones that share a row of common points. For every pair of such zones, the angle between them is found, thus giving a trial reciprocal lattice.
- (3) The reduction of the resulting unit cells is carried out using the Delaunay–Ito method (Pecharsky & Zavalij, 2009).
- (4) The program tries to index the first 20 lines and repeats this check after least-squares refinement of the unit-cell parameters.
- (5) The figures of merit are calculated to assess the quality of each trial unit cell and the four best lattices are provided.

ITO is very efficient at indexing patterns with low symmetry and is only weakly sensitive to impurity peaks, if they occur at high angles. The most frequent causes of failure are inaccuracy or incompleteness of the input data.

3.4.4.1.2. TREOR90 (Werner et al., 1985)

Classified by Shirley (1980, 2003) as semi-exhaustive, *TREOR90* is based on the index-heuristics strategy (see Section 3.4.3.1.3) and uses a trial-and-error approach. It performs the following steps:

- (1) Some basis lines are selected among the experimental d values, generally from the low- 2θ region of the powder diffraction pattern. Five sets of basis lines are generally sufficient for orthorhombic tests, whereas more than seven

sets may be necessary for the monoclinic system. At least 20–25 experimental d values are potentially required.

- (2) The trial unit cells in the index space are searched by varying the Miller indices that are tentatively assigned to the basis lines.
- (3) The analysis starts with cubic symmetry and, in a stepwise manner, tests for lower-symmetry crystal systems are performed. In the case of monoclinic symmetry, a special short-axis test is carried out.
- (4) The solution of the linear system in equation (3.4.8) gives the possible cell parameters. Different combinations of the basis lines are tested.
- (5) Each possible solution is checked by using the full list of experimental lines.
- (6) The quality of the trial cell parameters is mainly assessed by using the M_{20} figure of merit [see equation (3.4.4)]. An effective rule for identifying a reliable solution is $M_{20} > 10$ and no more than one unindexed line.

The success of the program is related to the use of some suitable standard sets of parameter values (maximum unit-cell volume, maximum cell axis, tolerance of values *etc.*) arising from the accumulated experience of the authors; they can be easily changed by the user *via* suitable keywords in the input file.

3.4.4.1.3. DICVOL91 (Boultif & Louër, 1991)

This program works in direct space (down to the monoclinic system) by using the successive-dichotomy search method (see Section 3.4.3.1.5), which was introduced for the automatic indexing of powder diffraction patterns by Louër & Louër (1972). *DICVOL91* has been defined as exhaustive by Shirley (1980, 2003). Its main steps are:

- (1) The unit-cell volume is partitioned by moving from high to low symmetry. Shells of 400 \AA^3 of volume are scanned for all the symmetry systems except for triclinic; for the triclinic system the shells are based on the definition of volume proposed by Smith (1977) (see Section 3.4.3.1.5).
- (2) A search using the successive-dichotomy method, based on suitable intervals (see Section 3.4.3.1.5), is carried out. In the case of the triclinic system the general expression for $Q(hkl)$ as function of the direct cell parameters is too complicated, and the reciprocal-space parameters in equation (3.4.2) are used for setting the intervals.
- (3) The derived cell parameters are refined using the least-squares method.
- (4) The quality of each trial unit cell is evaluated by using the M_N and F_N figures of merit (see Section 3.4.2.1).

The program is fast at performing exhaustive searches in parameter space (except for the triclinic case); on the other hand, its efficiency is strongly related to the quality of the data and to the presence of impurities (in fact, impurities are not permitted).

3.4.4.2. Evolved indexing programs

3.4.4.2.1. N-TREOR09 (Altomare et al., 2009)

Implemented in the *EXPO* program (Altomare et al., 2013) to perform the powder pattern indexing step, *N-TREOR09* is an update of *N-TREOR* (Altomare et al., 2000), which in turn is an evolution of *TREOR90*, and preserves the main strategies with some changes introduced to make the program more exhaustive and powerful. In particular:

3.4. INDEXING

- (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., $\sim 0.1^\circ$), the reflection-pair method is adopted (Dong *et al.*, 1999); (ii) if the shift is small ($< 0.03^\circ$), 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 \AA^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, 2014b), 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^{Rev} and M_n^{Sym} proposed by Oishi-Tomiyasu (2013) are used 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, 2014b).
- (4) The method for exhaustively searching unit cells that involve geometrical ambiguity (Oishi-Tomiyasu, 2014a, 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

3. METHODOLOGY

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'_{wp} :

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

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 i th experimental step, respectively. Via the R'_{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'_{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 ($\sim 50\%$) 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 \AA^3 ; (c) no cell length is longer than 20 \AA .

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, 2003a), which uses the Monte Carlo method, is part of the *TOPAS* (Coelho, 2003b) 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 & Rodríguez-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, $a = 7.6063(9)$, $b = 8.6220(1)$, $c = 8.749(1) \text{ \AA}$, $\beta = 97.24(1)^\circ$, $P2_1/c$, experimental range $5\text{--}60^\circ 2\theta$, $\lambda = 0.85041 \text{ \AA}$, RES = 1.0 \AA (where RES is the data resolution), synchrotron data, indexed by *Fzon* (Visser, 1969).

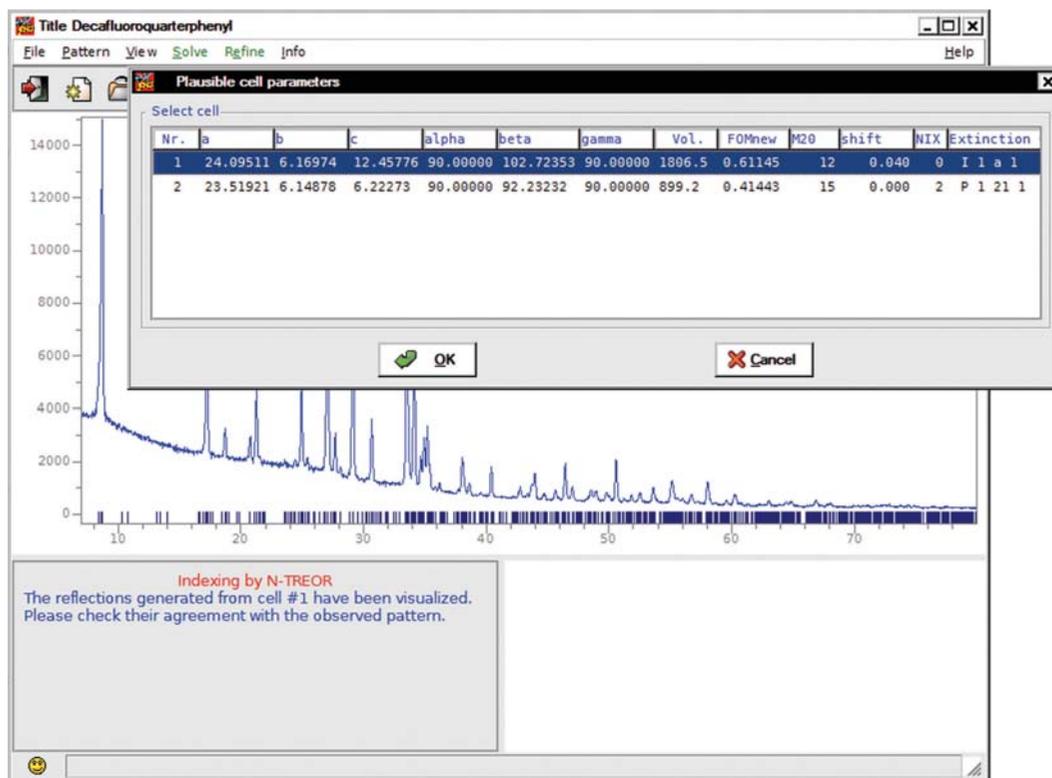


Figure 3.4.1

The list of possible cells for the decafluoroquarterphenyl structure automatically found using *N-TREOR09*.

The 2θ values of the first 25 peaks, in the range $5\text{--}25^\circ$, were determined by *WinPLOTR* and supplied to *DICVOL06*. The first 20 peaks were used for searching for the solution. No plausible cell was found when assuming that no impurity was present and exploring all the systems (from cubic to triclinic). *DICVOL06* was also unsuccessful when the non-default strategies of extended search and data correction for zero-point error were considered (by setting some flags to 1 in the input file). If it was supposed that two impurity lines might be present among the peaks (by setting the flag corresponding to the maximum number of accepted impurity/spurious lines to 2), *DICVOL06* was able to find the following monoclinic cell: $a = 8.7480$ (36), $b = 8.6313$ (32), $c = 7.6077$ (26) Å, $\beta = 97.201$ (33)°, with two unindexed lines, $M_{18} = 41.5$, $F_{18} = 125(0.0041, 35)$. The refinement of the cell by considering all the 25 lines gave $a = 7.6087$ (26), $b = 8.6295$ (30), $c = 8.7459$ (34) Å, $\beta = 97.201$ (34)°, which is very similar to the published one; 23 indexed lines, $M_{20} = 30.1$, $F_{20} = 102.6(0.0048, 41)$. The presence of the two impurity lines has been ascribed by the authors to a small amount of hexagonal plastic phase.

Example 2

Cu(II)–Schiff base complex (Banerjee *et al.*, 2002). Published information: $\text{Cu}(\text{C}_{15}\text{H}_{12}\text{NO}_2)_2$, triclinic, $a = 11.928$ (4), $b = 12.210$ (5), $c = 9.330$ (5) Å, $\alpha = 102.54$ (4), $\beta = 111.16$ (5), $\gamma = 86.16$ (4)°, $P\bar{1}$, experimental range $6\text{--}100^\circ$ 2θ , $\lambda = 1.54056$ Å, RES = 1.22 Å, high-quality X-ray laboratory data, indexed by *DICVOL91*. The 2θ values of the first 30 peaks, in the range $6\text{--}25^\circ$, were determined by *WinPLOTR* and supplied to *DICVOL06*. The first 20 peaks were used for searching for the solution. If it was assumed that no impurity was present, no plausible cell was found down to the monoclinic system. When the triclinic system was explored, *DICVOL06* suggested only one plausible solution: $a = 12.2157$ (73), $b = 12.2031$ (77), $c = 9.3071$ (41) Å, $\alpha = 65.798$ (46), $\beta = 102.572$ (59), $\gamma =$

95.711 (61)°, with no unindexed lines, $M_{20} = 27.0$, $F_{20} = 77.0(0.010, 26)$. The refinement of the cell considering all the 30 lines gave $a = 12.2125$ (65), $b = 12.1989$ (61), $c = 9.3016$ (32) Å, $\alpha = 65.826$ (33), $\beta = 102.569$ (40), $\gamma = 97.755$ (44)°, no unindexed lines, $M_{20} = 27.9$, $F_{20} = 72.8(0.0106, 26)$. For this, the corresponding conventional cell is $a = 11.93313$ (61), $b = 12.2125$ (65), $c = 9.3016$ (32) Å, $\alpha = 102.569$ (40), $\beta = 111.152$ (33), $\gamma = 86.151$ (44)°, similar to the published one.

3.4.4.6.2. Indexing using *N-TREOR09*

Two examples of powder diffraction pattern indexing by using *N-TREOR09*, as implemented in the *EXPO* program, will be described. To activate the procedure some specific instructions must be given to *EXPO* via the input file or the graphical interface. As a first step, the peak-search procedure is automatically performed on the experimental powder pattern and the list of corresponding d values are supplied to *N-TREOR09*. During the indexing process a correction for zero-point error is automatically carried out (positive and negative shifts are taken into account). Both the examples below were successfully indexed by a default run of *EXPO*.

Example 3

Decafluoroquarterphenyl (Smrčok *et al.*, 2001). Published information: $\text{C}_{24}\text{H}_8\text{F}_{10}$, monoclinic, $a = 24.0519$ (9), $b = 6.1529$ (3), $c = 12.4207$ (5) Å, $\beta = 102.755$ (2)°, $I2/a$, experimental range $7\text{--}80^\circ$ 2θ , $\lambda = 1.79$ Å, RES = 1.39 Å, medium-quality X-ray laboratory data. The first 43 peaks (in the range $7\text{--}67^\circ$) with intensities greater than a default threshold were selected (an intensity-based criterion is automatically adopted). The first 25 lines were used to find a possible cell that was then refined by considering all the 43 peaks. At the end of the automatic indexing procedure, *N-TREOR09* suggested two possible cells ranked according to WRIP20 [equation (3.4.5)], as shown in Fig. 3.4.1 (WRIP20 is denoted as

3. METHODOLOGY

FOMnew in *N-TREOR09*). The first one in the list is the correct cell. It is worth mentioning that the classical M_{20} figure of merit was not able to pick up the solution. The best cell parameters, found according to FOMnew, were $a = 24.0951$ (50), $b = 6.1697$ (21), $c = 12.4578$ (37) Å, $\beta = 102.724$ (18)°, similar to those reported in the literature, with FOMnew = 0.61, $M_{20} = 12$; all the lines in the pattern were indexed. The program provided the solution thanks to its automatic check for a zero-point correction (2θ zero shift = 0.04°) and was able to correctly identify the extinction group (I_{a-}). For the second suggested cell (the wrong solution) FOMnew = 0.41, $M_{20} = 15$, and two lines were unindexed.

Example 4

Hexagonal turkey egg-white lysozyme (Margiolaki *et al.*, 2005). Published information: hexagonal, $a = 71.0862$ (3), $c = 85.0276$ (5) Å, $P6_122$, experimental range 0.4–12° 2θ , RES = 3.35 Å, synchrotron data. The first 94 peaks (in the range 0.4–6°, $\lambda = 0.700667$ Å) with intensities greater than a default threshold were selected. An intensity-based criterion was automatically adopted. The first 25 lines were used to find possible cells that were then refined by considering all 94 peaks. Five possible unit cells were automatically suggested by the program in the following systems: hexagonal (1), orthorhombic (1) and monoclinic (3). The highest value for WRIP20 was 0.99, and was for the correct hexagonal cell parameters: $a = 71.0922$ (4), $c = 85.0269$ (7) Å, which are similar to those reported in the literature; all the 94 selected lines in the pattern were indexed. For this cell, the program detected a geometrical ambiguity (see Section 3.4.2.2) between hexagonal and orthorhombic lattices and automatically selected the higher-symmetry one.

3.4.5. Conclusion

Indexing a powder diffraction pattern is sometimes described as a ‘gateway technology’, because the determination of the cell parameters is so fundamental: if no cell has been identified the execution of the subsequent steps of the structure solution process is impossible, and if a wrong cell has been used the correct solution is unreachable. Therefore extremely close attention must be paid to the indexing step of the process. From the early 1970s, the increasing interest in powder pattern indexing and the progress seen, in terms of both methods and algorithms, have strongly contributed to opening the door to modern applications of powder diffraction techniques. The availability of a quite large number of software packages, based on different indexing strategies, enables the scientist interested in solving crystal structures to switch from one program to another when the first fails, so increasing the possibility of success. In some cases indexing is still a challenging process. Good-quality data are necessary and careful inspection of each indexing step, in particular in the selection of the experimental peak positions to be used, is advisable.

References

Altomare, A., Caliandro, R., Camalli, M., Cuocci, C., da Silva, I., Giovacazzo, C., Moliterni, A. G. G. & Spagna, R. (2004). *Space-group determination from powder diffraction data: a probabilistic approach*. *J. Appl. Cryst.* **37**, 957–966.

Altomare, A., Camalli, M., Cuocci, C., da Silva, I., Giovacazzo, C., Moliterni, A. G. G. & Rizzi, R. (2005). *Space group determination: improvements in EXPO2004*. *J. Appl. Cryst.* **38**, 760–767.

Altomare, A., Campi, G., Cuocci, C., Eriksson, L., Giovacazzo, C., Moliterni, A., Rizzi, R. & Werner, P.-E. (2009). *Advances in powder diffraction pattern indexing: N-TREOR09*. *J. Appl. Cryst.* **42**, 768–775.

Altomare, A., Cascarano, G., Giovacazzo, C., Guagliardi, A., Moliterni, A. G. G., Burla, M. C. & Polidori, G. (1995). *On the number of statistically independent observations in a powder diffraction pattern*. *J. Appl. Cryst.* **28**, 738–744.

Altomare, A., Cuocci, C., Giovacazzo, C., Moliterni, A., Rizzi, R., Corriero, N. & Falcicchio, A. (2013). *EXPO2013: a kit of tools for phasing crystal structures from powder data*. *J. Appl. Cryst.* **46**, 1231–1235.

Altomare, A., Giovacazzo, C., Guagliardi, A., Moliterni, A. G. G., Rizzi, R. & Werner, P.-E. (2000). *New techniques for indexing: N-TREOR in EXPO*. *J. Appl. Cryst.* **33**, 1180–1186.

Altomare, A., Giovacazzo, C. & Moliterni, A. (2008). *Indexing and space group determination*. In *Powder Diffraction Theory and Practice*, edited by R. E. Dinnebier & S. J. L. Billinge, pp. 206–226. Cambridge: RSC Publishing.

Andrews, L. C. & Bernstein, H. J. (1988). *Lattices and reduced cells as points in 6-space and selection of Bravais lattice type by projections*. *Acta Cryst.* **A44**, 1009–1018.

Banerjee, S., Mukherjee, A., Neumann, M. A. & Louër, D. (2002). *Ab-initio structure determination of a Cu(II)-Schiff base complex from X-ray powder diffraction data*. *Acta Cryst.* **A58**, c264.

Bergmann, J. (2007). *EFLECH/INDEX – another try of whole pattern indexing*. *Z. Kristallogr. Suppl.* **26**, 197–202.

Bergmann, J., Le Bail, A., Shirley, R. & Zlokazov, V. (2004). *Renewed interest in powder diffraction data indexing*. *Z. Kristallogr.* **219**, 783–790.

Boultif, A. & Louër, D. (1991). *Indexing of powder diffraction patterns for low-symmetry lattices by the successive dichotomy method*. *J. Appl. Cryst.* **24**, 987–993.

Boultif, A. & Louër, D. (2004). *Powder pattern indexing with the dichotomy method*. *J. Appl. Cryst.* **37**, 724–731.

Brunelli, M., Fitch, A. N., Jouanneaux, A. & Mora, A. J. (2001). *Crystal and molecular structures of norbornene*. *Z. Kristallogr.* **216**, 51–55.

Buerger, M. J. (1957). *Reduced cells*. *Z. Kristallogr.* **109**, 42–60.

Buerger, M. J. (1960). *Note on reduced cells*. *Z. Kristallogr.* **113**, 52–56.

Coelho, A. A. (2003a). *Indexing of powder diffraction patterns by iterative use of singular value decomposition*. *J. Appl. Cryst.* **36**, 86–95.

Coelho, A. A. (2003b). *TOPAS. Version 3.1 User's Manual*. Bruker AXS GmbH, Karlsruhe, Germany.

Coelho, A. A. & Kern, A. (2005). *Discussion of the indexing algorithms within TOPAS*. *IUCr Commission on Powder Diffraction Newsletter*, **32**, 43–45.

Conway, J. H. & Fung, F. Y. C. (1997). *The Sensual (Quadratic) Form*. Washington, DC: The Mathematical Association of America.

Delaunay, B. (1933). *Neue Darstellung der geometrischen Kristallographie*. *Z. Kristallogr.* **84**, 109–149.

Dong, C., Wu, F. & Chen, H. (1999). *Correction of zero shift in powder diffraction patterns using the reflection-pair method*. *J. Appl. Cryst.* **32**, 850–853.

Giovacazzo, C. (2011). *Crystallographic computing*. In *Fundamentals of Crystallography*, 3rd ed., edited by C. Giovacazzo, pp. 66–156. Oxford: IUCr/Oxford University Press.

Harris, K. D. M., Johnston, R. L., Chao, M. H., Kariuki, B. M., Tedesco, E. & Turner, G. W. (2000). *Genetic algorithm for indexing powder diffraction data*. University of Birmingham, UK.

Ishida, T. & Watanabe, Y. (1967). *Probability computer method of determining the lattice parameters from powder diffraction data*. *J. Phys. Soc. Jpn*, **23**, 556–565.

Ishida, T. & Watanabe, Y. (1971). *Analysis of powder diffraction patterns of monoclinic and triclinic crystals*. *J. Appl. Cryst.* **4**, 311–316.

Ito, T. (1949). *A general powder X-ray photography*. *Nature*, **164**, 755–756.

Ito, T. (1950). *X-ray Studies on Polymorphism*. Tokyo: Maruzen Company.

Karen, V. L. & Mighell, A. D. (1991). *Converse-transformation analysis*. *J. Appl. Cryst.* **24**, 1076–1078.

Kariuki, B. M., Belmonte, S. A., McMahon, M. I., Johnston, R. L., Harris, K. D. M. & Nemes, R. J. (1999). *A new approach for indexing powder*