

## 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  $0.05^\circ$  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  $2000 \text{ \AA}^3$ ; (c) no cell length is longer than  $20 \text{ \AA}$ .

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  $2\theta$  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 =  $1.0 \text{ \AA}$  (where RES is the data resolution), synchrotron data, indexed by *Fzon* (Visser, 1969).

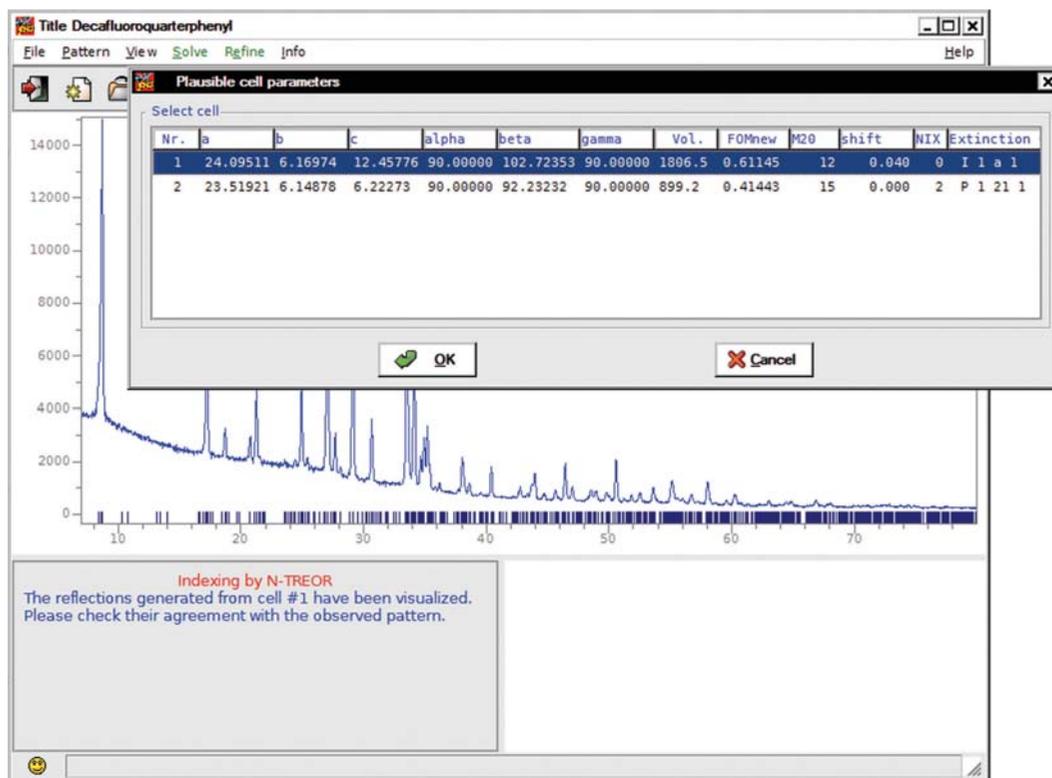


Figure 3.4.1

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

The  $2\theta$  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  $6\text{--}100^\circ$   $2\theta$ ,  $\lambda = 1.54056$  Å, RES = 1.22 Å, high-quality X-ray laboratory data, indexed by *DICVOL91*. The  $2\theta$  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  $7\text{--}80^\circ$   $2\theta$ ,  $\lambda = 1.79$  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 ( $2\theta$  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°  $2\theta$ , 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. & Nelmes, R. J. (1999). *A new approach for indexing powder*