Indexing and space group determination

Anna Moliterni

Institute of Crystallography, Bari, Italy

Structure solution by powder diffraction data is a multi-step process which needs the identification of the unit cell parameters and the determination of the space group; the correct definition of cell and/or space group is a mandatory request for the success of the structure solution process. An outline of the main methods aimed at finding unit cell parameters and space group is thereinafter provided, with particular attention to the approaches implemented in the *EXPO2014* program (Altomare *et al*., 2013). Applications of *EXPO2014* to a practical case is briefly described. For more details on theoretical methods and algorithms developed for indexing and space group determination see, *e.g*., Altomare *et al*. (2008).

Indexation

The primary step of the solution process by powder diffraction data is the indexation, whose main goal is to recover the reciprocal unit cell parameters (a^*, b^*, c^*, α^*, β^*, γ^*), and, from them, the direct cell parameters (a, b, c, α, β, γ), by assigning the appropriate triple of Miller indices (hkl or $\mathbf{h}$) to each observed interplanar spacing ($d_{\mathbf{h}}$). Indeed, the indexation step enables to reconstruct the three-dimensional elementary cell by the information on the $d_{\mathbf{h}}$ values extracted from the experimental one-dimensional pattern. In case of single crystal the data are three-dimensional and the diffraction effects are usually well separated, consequently, the determination of the unit cell parameters is often a trivial task. The main limitation of powder data, with respect to single crystal data, is due to the collapse of the three-dimensional reciprocal space into the one-dimensional powder diffraction profile. The experimental information is the interplanar spacing $d_{\mathbf{h}}$, related to the diffraction angle $\theta_{\mathbf{h}}$ by the Bragg law:

$$d_{\mathbf{h}}=\lambda/2\ \sin\theta_{\mathbf{h}}, \qquad (1)$$

where λ is the measurement wavelength.

Consequently, in spite of the advances in theoretical and experimental methods and computer programs, the indexing process by powder diffraction data can be still nowadays a challenge. In order to derive the cell parameters from the $d_{\mathbf{h}}$ values let us assume (see, *e.g*., Altomare *et al*., 2008): $Q(hkl)=\frac{10^4}{d_{\mathbf{h}}^2}$. (2)

The expression of Q(*hkl*) in terms of reciprocal cell parameters and (*hkl*) indices changes with the crystal system (see Table 1).

Table 1. The expression of Q(*hkl*) *vs* the crystal system.

Crystal System	**Q(*hkl*)**
Cubic	$(h^2+k^2+l^2)\ A_{11}$
Tetragonal	$(h^2+k^2)\ A_{11}+l^2\ A_{33}$
Hexagonal	$(h^2+hk+k^2)\ A_{11}+l^2\ A_{33}$
Orthorhombic	$h^2\ A_{11}+k^2\ A_{22}+l^2\ A_{33}$
Monoclinic	$h^2\ A_{11}+k^2\ A_{22}+l^2\ A_{33}+hl\ A_{13}$
Triclinic	$h^2\ A_{11}+k^2\ A_{22}+l^2\ A_{33}+hk\ A_{12}+hl\ A_{13}+kl\ A_{23}$

Let us consider the more general case (the triclinic system):

$Q(hkl)=h^2 A_{11} +k^2 A_{22} +l^2 A_{33} +hk A_{12} +hl A_{13} +kl A_{23}$, (3)

where

$A_{11}=10^4a^{*2}$, $A_{22}=10^4b^{*2}$, $A_{33}=10^4c^{*2}$, $A_{12}=2\cdot10^4a^*b^*\cos\gamma^*$, (4)

$A_{13}=2\cdot10^4a^*c^*\cos\beta^*$, $A_{23}=2\cdot10^4b^*c^*\cos\alpha^*$.

Eq. (3) is the basic equation of the indexing process. Indexation is based on the determination of the A_{ij} values, whose number decreases increasing the symmetry of the crystal system, and the Miller indices (*hkl*) triple, in such a way that Eq. (3) is satisfied for each Q*(hkl)* value, within a given tolerance. This last is adopted to take into account the unavoidable errors on experimental peaks positions, due to multiple and often concomitant causes (*i.e.*, peak overlap, and/or zero-point error, and/or bad crystallinity of the sample, and/or poor peak resolution, and/or low peak/background ratio,..). The solution of (3) is not unique, at the end of the indexing process more than one plausible cell is usually available. In order to assess the physical reliability of the solutions a figure of merit (FOM) is associated to each of them. The most commonly used FOM is M_{20} (de Wolff, 1968) that uses the twenty lowest 2θ angle peaks (this number is sufficient for carrying out a feasible indexing process and avoids to involve high angle peak positions, usually more affected by errors). M_{20} increases with i) the agreement between calculated and observed peak positions and ii) the decreasing of the candidate cell volume. As a basic rule, if the number of unindexed peaks, among the first twenty ones, is not larger than 2 and $M_{20} > 10$ the indexing process is physically reliable (de Wolff, 1968; Werner, 2002). Alternative and/or more recent FOMs have been proposed (*e.g.*, Smith & Snyder, 1979; Bergmann, 2007; Le Bail, 2008; Altomare *et al.*, 2009).

A careful check of the largest FOM cells, in terms of compatibility with the experimental data (*i.e.*, agreement between the reflection $2\theta_{\mathbf{h}}$ values, calculated using the candidate cell, and the experimental peak positions), is strongly suggested. Among the sources of indexing failure: the inaccuracy of the experimental peak positions and/or the presence of impurity peaks in the experimental pattern, *e.g.*, due to the presence of more than one chemical phase in the compound under study.

Traditional indexing methods are based on the Eq. (3) and are applied by the most widely used indexing computer programs, among them: a) DICVOL (Boultif & Louër, 1991) and its updated version DICVOL06 (Boultif & Louër, 2004); b) TREOR (Werner *et al.*, 1985) and its evolution *N-TREOR09* (Altomare *et al.*, 2009); ITO (Visser, 1969).

Recently, approaches alternative to the traditional methods and based on global optimization methods (*e.g.*, Monte Carlo, Genetic Algorithm) have been proposed, searching in direct space for the unit cell parameters (a, b, c, α, β, γ) providing the best agreement between the experimental and the calculated patterns in terms of a suitable figure of merit (*e.g*, the agreement factor R_p), corresponding to the global minimum of the hypersurface R_p [a, b, c, α, β, γ] (Altomare et *al.*, 2008). For an overview of the indexing computing programs, based on different strategies, see Cranswick (2008).

Let us consider one of the computer programs adopting the traditional indexing method, *N-TREOR09*. It is the default indexing program of *EXPO2014*. After that a peak-search procedure has located the experimental peaks, *N-TREOR09* searches for the correct cell by 1) selecting a subset of low resolution observed Q(*hkl*) values (*i.e.*, the basis lines) among the set of the experimental $d_{\mathbf{h}}$. The number of basis lines depends on the investigated

crystal system symmetry; 2) assigning trial Miller indices to the basis lines; 3) solving Eq. (3).

Let us consider the results of the indexation by *EXPO2014* in case of the 2,6-diamino-5-hydroxy-3-nitro-4*H*-pyrazolo[1,5-*a*]-pyrimidin-7-one monohydrate compound (Chernyshev *et al*., 1999), with code name AND2 (published cell parameters: a=17.576 Å, b=10.900 Å, c=4.6738 Å, α=90.00°, β=92.87°, γ=90.00°, space group: $P2_1/n$).

56 lines of the diffraction pattern (synchrotron data, 2*theta* range: 3°- 65.5°), with intensity values greater than a default threshold, were automatically selected by *EXPO2014* for indexing [the first 25 low-angle peaks were actively used for solving Eq. (3)]. Three cells (all of them monoclinic and similar) were provided and ranked according to a figure of merit that takes into account not only the M_{20} value but also the agreement between the calculated (obtained by assuming the space group with the largest Laue symmetry compatible with the candidate cell and no extinction conditions) and the observed powder pattern, the degree of overlap in the powder pattern, the symmetry and the number of unindexed lines. For them, the first one in the list (cell parameters: a=17.5849 Å, b=10.9026 Å, c=4.6760 Å, α=90.00°, β=92.82°, γ=90.00°, the most similar to the published ones), has no unindexed lines and is the most probable candidate, having a reliable value of M_{20} (M_{20} = 53), much greater than those ones corresponding to the second and the third cells, for which M_{20} is equal to 13 and 10, respectively, and, in addition, one observed line is unindexed.

Space Group Determination

At the end of the indexation process, once the crystal system and the unit-cell have been identified, the next step of the *ab initio* structure solution process is the determination of the space group. The statistical analysis of the diffraction intensities provides information on the systematically absent reflections (Altomare *et al*., 2008), that correspond to null integrated intensities, and, combined with the information about the Laue group, enables to determine the extinction symbol (ES). The list of the extinction symbols varies with the crystal system (Altomare *et al*., 2004): 14 extinction groups for the monoclinic, 111 for the orthorhombic, 31 for the tetragonal, 12 for the trigonal-hexagonal and 18 for the cubic system. The determination of the ES represents a critical point in case of powder data because it is based on the integrated intensities of each individual reflection, affected by unavoidable errors. The overlap problem as well as the background can hinder the correct identification of the systematically absent reflections, in particular in case of reflections lying at the large 2θ angle value regions of the pattern, characterized by weak diffraction signal. Probabilistic approaches have been developed (Markvardsen *et al*., 2001; Altomare *et al*., 2004, 2005, 2007) able to provide a probability value for each extinction symbol compatible with the crystal system determined by the indexation step.

Not always the ES defines unambiguously the space group; it can occur that more than one space group corresponds to the same ES, as, for example, in the monoclinic crystal system, in case of the '*P* 1_1' extinction symbol to which three space groups ($P2$, Pm, $P2/m$) correspond. To overcome the space group ambiguity it is possible: a) to carry out the structure solution process for each possible space group and finally choose the correct solution; b) to select the space group which best agrees with the molecular properties of the material under investigation.

The method proposed by Altomare *et al*. (2004, 2005, 2007) has been introduced in *EXPO2014*. As an example, let us consider the main steps carried out by *EXPO2014* in order to determine the space group in case of AND2. First of all, a full pattern decomposition

step is performed in the largest Laue symmetry compatible with the crystal system and with no extinction conditions (*i.e.*, in case of AND2, in the space group $P2/m$) in order to extract the integrated intensities. Then, a statistical analysis of the integrated intensities is carried out. For each of the 14 extinctions symbols compatible with the monoclinic crystal system a probability value is calculated as the product of the probabilities of each symmetry element concerning the group. The low accuracy of the diffraction integrated intensities estimates can invalidate the statistical analysis, consequently, not always the ES corresponding to the largest probability is the correct one. A careful check of the space group determination results is recommended; useful graphical tools have been introduced in *EXPO2014* to help the user to easily recognize wrong extinction symbol(s) and identify the correct one.

In Fig. 1, in case of AND2, the list of extinction groups graphically ranked by *EXPO2014* according to the calculated probability value (the ideal probability is 1) is given. The choice of 'P 1 2_1/n 1' (corresponding to the largest value of the estimated probability) as the correct ES is supported by the careful inspection of the experimental pattern. The solution process can be carried out by assuming $P2_1/n$ as space group of the structure under investigation.

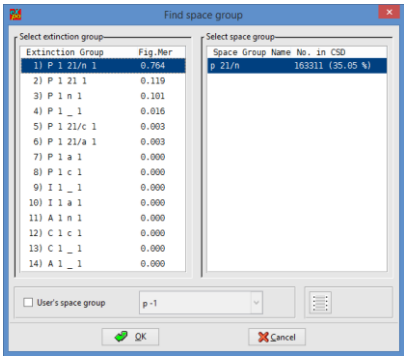

Figure 1. AND2 structure. The 14 extinction groups ranked by *EXPO2014* according to the calculated probability value.

References

Altomare, A, Caliandro, R, Camalli, M., Cuocci, C., da Silva, I., Giacovazzo, C., Moliterni, A.G.G. & Spagna, R. (2004). *J. Appl. Cryst*. **37**, 957–966.

Altomare, A, Camalli, M, Cuocci, C., da Silva, I., Giacovazzo, C., Moliterni, A.G.G. & Rizzi, R. (2005). *J. Appl. Cryst*. **38**, 760–767.

Altomare, A., Camalli, M., Cuocci, C., Giacovazzo, C., Moliterni, A.G.G. & Rizzi, R. (2007). *J. Appl. Cryst*. **40**, 743–748.

Altomare, A., Campi, G., Cuocci, C., Eriksson, L., Giacovazzo, C., Moliterni, A., Rizzi, R. & Werner, P.-E. (2009). *J. Appl. Cryst*. **42**, 768–775.

Altomare, A., Cuocci, C., Giacovazzo, C., Moliterni, A., Rizzi, R., Corriero, N. & Falcicchio, A. (2013). *J. Appl. Cryst.* **46**, 1231-1235.

Altomare, A., Giacovazzo, C. & Moliterni, A. (2008). *In* Dinnebier, R.E. & Billinge, S.J.L. (eds.): *Powder Diffraction Theory and Practice*. Cambridge: RSCPublishing, chapter 7, 206-226.

Bergmann J. (2007). *Z. Kristallogr*. *Suppl*. **26**, 197–202.
Boultif, A. & Louër, D. (1991). *J. Appl. Cryst*. **24**, 987-993.
Boultif, A. & Louër, D. (2004). *J. Appl. Cryst*. **37**, 724-731.
Chernyshev, V. V., Fitch, A. N., Sonneveld, E. J., Kurbakov, A. I. , Makarov, V. A. & Tafeenko, V. A. (1999). *Acta Cryst*. B**55**, 544-562.
Cranswick, L.M.D. (2008). *In* Dinnebier, R.E. & Billinge, S.J.L. (eds.). *Powder Diffraction Theory and Practice*. Cambridge: RSCPublishing, 494-570.
Le Bail A. (2008). *In* Clearfield, A., Reibenspies, J.H. & Bhuvanesh, N. (eds.). *Principles and Applications of Powder Diffraction*. Oxford, England, Wiley-Blackwell, 261–309.
Markvardsen, A.J., David, W.I.F., Johnson, J.C. & Shankland, K. (2001). *Acta Cryst*. A57, 47–54.
Smith G.S. & Snyder R.L.J. (1979). *J. Appl. Cryst*. **12**, 60–65.
Visser, J. W. (1969). *J. Appl. Cryst*. **2**, 89-95.
Werner, P.-E., Eriksson L. & Westdahl, M. (1985). *J. Appl. Cryst*. **18**, 367-370.
Werner, P.-E. (2002). *In* David, W.I.F., Shankland, K., McCusker, L.B. and Baerlocher, Ch., eds. *Structure Determination from Powder Diffraction Data.* Oxford University Press, Oxford, 233-251.
Wolff, P.M. de (1968). *J. Appl. Cryst*. **1**, 108-113.