



Phasing *via* Direct Methods

Rosanna Rizzi

Institute of Crystallography, Bari, Italy.
rosanna.rizzi@ic.cnr.it

Introduction

One of the most challenging target of the X-ray powder diffraction technique (XRPD) is the structure solution. In the last 25 years, interest for powder solution is remarkably increased and the number of crystal structures solved by XRPD is surprisingly growing. The reasons of this progress can be found in the evolution of diffractometric instrumentation and in the development of new and efficient computational, methodological and theoretical approaches for data analysis.

Structure solution by powders is carried out by a complex pathway, which requires first the determination of cell parameters and space group; after that the structure model can be accomplished by using different approaches: 1) the reciprocal space methods [1-3]; 2) the direct space methods (see [4] for an overview); 3) the hybrid methods [5-7].

In this chapter, the solution process of polycrystalline compounds, *via* Direct Methods, will be described.

Direct Methods solution

A diffraction experiment provides information on the diffraction intensities and, therefore, on structure factors moduli only ($I_h \propto |F_h|^2$). In order to determine the electron density $\rho(\mathbf{r})$ of a crystal, it is necessary to recover the lost information on reflection phases, *i.e.*, to solve the so called *phase problem*, a well-known problem in Crystallography. Direct Methods (DM) greatly contributed to the solution of the phase problem. Their role is explained by the name: to derive the phase of reflections *directly* from the observed integrated intensities, through mathematical relationships.

DM are probabilistic methods based on three main hypotheses:

- the positivity of $\rho(\mathbf{r})$ (*i.e.*, $\rho(\mathbf{r})$ is everywhere positive);
- the atomicity of $\rho(\mathbf{r})$ (*i.e.*, the electron density is concentrated around the nuclei);
- the uniform and independent distribution of atoms in the unit cell.

Under the above conditions, the DM procedure performs, in sequence, the following fundamental steps:

1) Normalization.

• The observed diffraction intensities provided by X-ray diffraction experiments, are on relative scale *i.e.*, $I_h = K |F_h^0|^2 \exp(-2Bs^2)$, where K is the scale factor, $|F_h^0|$ is the observed structure amplitude on absolute scale for atoms at rest, B is the average isotropic thermal factor (assumed equal for all the atoms) and $s = \sin \theta / \lambda$.

Wilson overcame this dependence by proposing a method, the Wilson plot method [8], able to provide an estimate of the scale factor k and of the overall isotropic thermal factor B by the following equation:

$$\ln \left(\frac{\langle |F_h|^2 \rangle_s}{\sum_s^0} \right) = \ln K - 2Bs^2 \quad (1)$$

The plot of the left side of (1) vs. S^2 represents a straight line whose intercept with the vertical axis gives $\ln K$ and its slope the value of $-2B$.

• The previous DM hypothesis c) enables to derive the expected intensity as follows:

$$\langle |F_h|^2 \rangle = \sum_{j=1}^N f_j^2 \quad (2)$$

where f_j^0 and $f_j = f_j^0 \exp(-B_j \frac{\sin^2 \theta}{\lambda^2})$ are the scattering factors of the j -th atom at rest and thermally moved, respectively.

Due to f_j , $\langle |F_h|^2 \rangle$ depends both on the thermal motion and on data resolution ($\sin \theta / \lambda$). This means that reflections measured at different θ values cannot be compared directly. In order to overcome this drawback, normalized structure factors $|E_h|$ can be calculated according to:

$$|E_h| = \left[|F_h|^2 / k \cdot \exp(-2B \sin^2 \theta / \lambda^2) \sum_{j=1}^N f_j^{0^2} \right]^{1/2} \quad (3)$$

On the contrary of $|F_h|$, $|E_h|$ is independent on $\sin \theta / \lambda$ (it can be demonstrated that $\langle |E_h|^2 \rangle = 1$ for all values of θ) and corresponds to an ideal structure with point shaped atom not thermally moved. The largest $|E_h|$ reflections (the strong reflections) play a relevant role in the phasing process, as described below.

2) Estimate of the structure invariants.

Before passing to the step of phasing, it is important to consider that the structure factor referred to a new origin \mathbf{x}_0 is given by $F'_h = F_h \exp(-2\pi i \mathbf{h} \cdot \mathbf{x}_0)$, that is, the phase ϕ_h changes according to $\phi'_h = \phi_h - 2\pi \mathbf{h} \cdot \mathbf{x}_0$ (is not a structure invariant), while the modulus remains unchanged ($|F'_h| = |F_h|$) (i.e., it is structure invariant). The consequence is that from the moduli we can obtain information only on combinations of phases which are structure invariants: i.e., products of n structure factors which remain invariant whatever the origin. They may be written as:

$$F_{h_1} F_{h_2} \dots F_{h_n} = |F_{h_1} F_{h_2} \dots F_{h_n}| \exp[i(\phi_{h_1} + \phi_{h_2} + \dots + \phi_{h_n})] \quad (4)$$

with vectorial sum of Miller indices equal to zero ($\mathbf{h}_1 + \mathbf{h}_2 + \dots + \mathbf{h}_n = \mathbf{0}$).

The most important structure invariant is the triplet invariant $F_h F_k F_{-h-k}$ and, in terms of normalized structure factors, $E_h E_k E_{-h-k}$, that has a crucial role in the phasing process. Infact, after the normalization the reflections are sorted according to their $|E|$ values and only those larger than a threshold ($|E| > 1$, they are called N_{LARGE}) are involved in the generation of triplet relationships and actively used in the phasing process.

3) Phasing by tangent formula.

Generally, each \mathbf{h} reflection takes part in more than one triplet relationship. Random phases are initially assigned to the N_{LARGE} reflections; after that, the tangent formula proposed by Hauptman and Karle [9], is applied to refine and extend the N_{LARGE} phases. Because the tangent procedure is not always able of making converge the random phases to their correct values, it is cyclically repeated, starting from different initial random value, by following a multi-solution approach [10].

4) Electron density map calculation.

At the end of the phasing process more sets of phases are available whose reliability is assessed by a suitable figure of merit (the CFOM: a special properly weighted combination of figures of merit). The set corresponding to the largest CFOM is automatically selected for calculating a Fourier map (the $|E|$ -map), given by:

$$\rho(\mathbf{r}) \cong \frac{1}{V} \sum_{\mathbf{h}} |E_{\mathbf{h}}| \exp(i\phi_{\mathbf{h}}) \exp(-2\pi i \mathbf{h} \cdot \mathbf{r}). \quad (5)$$

The symbol \cong in (5) has been introduced in order to emphasize that, in real cases, the summation is not extended to an infinite number of reflections but to the limited experimentally available one.

In case of powder data, the structure model determined by the $|E|$ -map is often far from being a chemically reasonable structure model, corresponding only to a rough approximation of the final structure: some atoms are imperfectly located, others are in wrong positions and/or some atoms are missing. This is essentially due to the unavoidable *powder problems*: peak overlapping, difficulty in correct estimation of the background, presence of preferred orientation effects, limited experimental resolution. As a consequence, the experimental diffraction moduli are only approximately estimated (in particular for overlapping reflections) and, consequently, the structure factor phases may be not correctly determined. Since DM are very sensitive to the reliability of the observed moduli, some problems may occur:

- 1) the efficiency of DM is reduced so depressing the full phasing process. In other words, the probabilistic evaluation of the phases may be not fully reliable and/or the identification of the best phasing set, by means of the largest CFOM figure of merit, may be not successful;
- 2) the optimization of the structural model, before running the final Rietveld refinement, owing to the modest convergence rate of this technique, is mandatory.

Recently, several methods have been developed for overcoming the limits above mentioned. They have been promptly transformed in computing algorithms and implemented in software packages increasing the probability of success of the solution process.

As an example, in the EXPO [11] software, a package devoted to the structure solution from powder diffraction data, procedures have been introduced to:

- 1) complete and optimize the structure model (for a general overview see [12-16]);
- 2) explore all the sets of phases generated by DM. The application of this approach is advisable when CFOM is not significant: *f.e.*, when it is equal to its maximum value (say 1) for all the stored phasing sets, or when the largest CFOM values are nearly equivalent. In these cases, the exploration of all the 20 stored trials can be performed by EXPO through the non-default *ALLTRIALS* procedure. It automatically carries out, for each of the stored trials, the structure refinement process. The quality of each final structure model is assessed by calculating the reliability parameter:

$$R_F = \frac{\sum ||F|_{oss} - K|F|_{calc}|}{\sum |F|_{oss}}$$

where $|F|_{oss}$ and $|F|_{calc}$ are the observed and the calculated structure factor moduli, respectively and K a scale factor. The trials are ranked according to increasing values of R_F , and the model corresponding to the smallest R_F value is automatically selected and graphically shown.

Conclusions

The solution process is quite trivial for single crystal, much more complex for microcrystalline and nanocrystalline powder. The great progress, reached in the last twenty five years, in terms of theories, computing strategies, experimental devices and, particularly, software packages has strongly improved the solution process by powder, so reducing the gap with respect to the single crystal case.

References

- [1] Giacovazzo, C. (1998). IUCr/Oxford University Press, Oxford.
- [2] Rius, J. & Frontera, C. (2007). *J. Appl. Cryst.* **40**, 1035-1038.
- [3] Gilmore, C., Dong, W. & Bricogne, G. (1999). *Acta Cryst.* **A55**, 70-83.
- [4] David, W.I.F., Shankland, K., McCusker, L.B., Maerlocher, C. (2002). IUCr/Oxford University Press, Oxford. Pp. 252-285.
- [5] Baerlocher, C., McCusker L. & Palatinus, L. (2007). *Z. Kristallogr.* **222**, 47-53.
- [6] Altomare, A., Caliandro, R., Giacovazzo, C., Moliterni, A.G.G., Rizzi, R. (2003). *J. Appl. Cryst.* **36**, 230-238.
- [7] Altomare, A., Caliandro, R., Cuocci, C., Giacovazzo, C., Moliterni, A.G., Rizzi, R., Platteau, C. (2008). *J. Appl. Cryst.* **41**, 56-61.
- [8] Wilson A.J.C. (1942). London, *Nature*, **150**, 151-152.
- [9] Karle, J., Hauptman, H.A. (1956). *Acta Cryst.* **9**, 635-651.
- [10] Chernishev, V.V. (2000). *Acta Cryst.* **A56**, s132.
- [11] Altomare, A., Cuocci, C., Giacovazzo, C., Moliterni, A., Rizzi, R., Corriero, N., Falcichio, A. (2013). *J. Appl. Cryst.*, **46**, 1231-1235.
- [12] Altomare, A., Cuocci, C., Giacovazzo, C., Moliterni, A.G.G., Rizzi, R. (2006). *J. Appl. Cryst.* **39**, 558-562.
- [13] Altomare, A., Camalli, M., Cuocci, C., Giacovazzo, C., Moliterni, A.G.G., Rizzi, R. (2007). *J. Appl. Cryst.* **40**, 344-348.
- [14] Altomare, A., Cuocci, C., Giacovazzo, C., Kamel, G.S., Moliterni, A., Rizzi, R. (2008). *Acta Cryst.* **A64**, 326-336.
- [15] Altomare, A., Cuocci, C., Giacovazzo, C., Moliterni, A., Rizzi, R. (2008). *J. Appl. Cryst.* **41**, 592-599.
- [16] Altomare, A., Cuocci, C., Giacovazzo, C., Moliterni, A., Rizzi, R. (2009). *Acta Cryst.* **A65**, 183-189.
- [17] Altomare, A., Cuocci, C., Giacovazzo, C., Moliterni, A., Rizzi, R. (2010). *J. Appl. Cryst.* **43**, 798-804.