1. Introduction

Two approaches can be followed for solving \textit{ab-initio} crystal structures by powder diffraction data: the \textit{traditional approach} and the \textit{direct space approach}. Both of them require the knowledge of the following minimal information: a) experimental diffraction profile; b) cell parameters; c) space group; d) unit cell content. In addition, the direct space methods (simulated annealing, genetic algorithm, Monte Carlo techniques) need the knowledge of the structure molecular geometry also. When this information is not available, the traditional method is the obligatory choice.

The main steps of the traditional approach consist of: 1) indexing the powder pattern; 2) determining the space group; 3) solving the structure by direct methods, Patterson methods or Maximum Entropy; 4) refining the structure by the Rietveld method.

About point 3), direct methods are based on statistical and probabilistic calculations. They use the experimental $|F_h|$ structure factor modulus corresponding to each $h$ reflection (see Ref. [2] for direct methods theory) and aim at solving the phase problem. The inverse Fourier transform of the $F_h$ structure factors provides the electron density map whose maxima correspond to the atomic positions. Direct methods are successfully applied to single crystal data. In case of powder solution, the extraction of the integrated intensities $I_h$ ($I_h \propto |F_h|^2$) for each reflection from the experimental pattern is preliminary to the application of direct methods. It requires the decomposition of the experimental profile into the single peaks and the area under each peak gives the wanted $I_h$ value. Unfortunately, some problems occur and make the extraction procedure very critical for the success of the structure solution step. The main problems regard:

a) the peak overlap. It depends on the experimental resolution and on the structure complexity. It increases at large $2\theta$ angular values of the observed pattern;

b) the background. Its estimate is not trivial. It adds to the peak overlap effect so that, especially at large $2\theta$
values, it is difficult to estimate the noise contribution correctly;
c) the preferred orientation. The crystallities are not always randomly oriented. This behaviour modifies the ratios of the experimental intensities.

For the above mentioned problems the $|F_h|$ estimate process reveals itself as a crucial point in the powder ab-initio solution: the more reliable the extracted integrated intensities values, the larger the success probability of solving the structure.

2. The Integrated Intensity Extraction Process

Two methods are widely used for extracting the integrated intensities from the powder pattern: the Pawley method [3] and the Le Bail method [4].

The Pawley method is based on a non linear least squares procedure. The integrated intensities are refinable variables in addition to the profile parameters. Because of the peak overlap, the least squares are often unstable and they provide negative integrated intensity values which must be discarded. For this reason the method needs positivity constraints [5], [6].

The Le Bail method is an iterative decomposition algorithm following the Rietveld formula [7]. The integrated intensity value is calculated according to:

$$I_h = \sum_i \left( y_{\text{obs}}(i) - y_b(i) \right) y_{\text{calc}}(i, h) / \sum_k y_{\text{calc}}(i, k).$$

where the summation is over the peak range, $y_{\text{obs}}(i)$ is the experimental count in the $2\theta$ angular value, $y_b(i)$ is the background contribution, $y_{\text{calc}}(i, h)$ is the calculated count in $2\theta$, due to the $h$ reflection contribution. The Le Bail method starts with arbitrary but fixed integrated intensity values and the formula is cyclically applied. It is rapidly convergent; it provides positive values if the background is properly estimated, but it tends to equiportion the intensity of a group of reflections strongly overlapping.

3. The Direct Methods Efficiency With Powder Data

In order to assess which of the two above mentioned methods is more suitable to be combined with direct methods, it proves useful to take into account the reliability parameter $R_F$ about the extracted amplitudes:

$$R_F = \frac{\sum |F_{\text{true}}| - |F_{\text{extracted}}|}{\sum |F_{\text{true}}|}$$

where the summation is over the number of reflections, $|F_{\text{true}}|$ is the structure factor modulus calculated by using the published atomic parameters. The reliability parameter $R_p$ about the profile:

$$R_p = \frac{\sum |y_{\text{obs}}(i) - y_{\text{calc}}(i)|}{\sum |y_{\text{obs}}(i)|}$$

is considered also. The summation is extended to the number of profile counts. $y_{\text{obs}}(i)$ and $y_{\text{calc}}(i)$ are the observed and the calculated counts, respectively.

In Table 1, crystal chemical information are given for some test structures (the code name, the space group, the unit cell content, the $2\theta$ experimental range and the number of reflections in the range). They cover a quite large variety of cases.

In Table 2, for each of some test structures, the $R_F$ and the $R_p$ values are shown. They are calculated by using the integrated intensities extracted by the EXTRA program [8] and the ALLHKL program [3], respectively. EXTRA is a Le Bail based package, ALLHKL uses the Pawley method. In Table 2 the $R_p$ values are small but the $R_F$ values are large (0.4 is the average value). This means that: a) low $R_p$ value is necessary and not sufficient condition for a reliable extraction; b) the integrated intensity accuracy is very low and this is the reason for which the powder ab-initio solution is not straightforward. Moreover, the $R_F$ values by EXTRA are always smaller than the $R_F$ values by ALLHKL so that we could conclude that the Le Bail method should be preferred but that behaviour may depend on the equipartition tendency of the Le Bail approach. However we proved that the statistical efficiency of direct methods improved by using the Le Bail extracted intensities [9].

4. The Le Bail Method Advantage

The Le Bail method offers a great advantage: it is very sensitive to the starting point. This aspect is shown in Table 3 where, for some test structures, the $R_F$ values are shown. Protocol 1 corresponds to the traditional Le Bail and Pawley extraction cases respectively, while Protocol 2 corresponds to the case when the true inte-
grated intensities are used as starting values in the Le Bail and Pawley methods, respectively. The values in Table 3 suggest that the integrated intensity estimate is not sensitive to the starting point if the Pawley method is adopted. On the contrary, it is not so for the Le Bail approach. This means that if the starting integrated intensities are less arbitrary and closer to the true ones the amplitude estimate is improved. On the other hand, the values of Protocol 2 represent the maximal accuracy level we can reach. The great advantage preserved by the Le Bail method can be exploited. From this last consideration the EXPO program was developed [1].

5. The EXPO Program

EXPO is the integration of EXTRA and SIR-POW[10] programs. This last is devoted to the structure solution by direct methods. EXPO needs the minimal information about the experimental powder pattern, the cell parameters, the space group and the unit cell content (Fig. 1 shows an example of the minimal EXPO input). Its main steps are:

1) Extraction of the integrated intensities (EXTRACTION routine);
2) Normalization of the extracted intensities (NORMALIZATION routine);
3) Calculation of the structure invariant relationships (triplets and quartets) (INVARIANT routine);

where \( \langle |E_h|^2 \rangle = 1 \)

The normalization rule restrains that:

The statistical analysis of the normalized structure factors can reveal the presence of pseudo-translational symmetry and/or preferred orientation.

Table 1. Code and crystal chemical information for the test structures. X: home diffractometer data; N: neutron data; S: synchrotron data

| Code      | Space Group | Unit cell content  | 2\(\theta\) Range | Nr. reflection |
|-----------|-------------|--------------------|--------------------|---------------|
| AGPZ (X)  | \(P\)bc\(a\) | \(\text{Ag}_{8}\text{N}_{16}\text{C}_{24}\text{H}_{24}\) | \(5.0\text{–}80.0\) | 258           |
| AND1 (S)  | \(P\) \(2\_n\) | \(\text{Ca}_{2}\text{Na}_{2}\text{O}_{4}\text{H}_{14}\) | \(4.0\text{–}50.0\) | 896           |
| BACO (N)  | \(C\)2\(m\) | \(\text{Ba}_{6}\text{C}_{12}\text{O}_{28}\) | \(20.7\text{–}150.0\) | 272           |
| BAMO (X)  | \(P\) \(2\_1\) | \(\text{Ba}_{2}\text{Mo}_{2}\text{O}_{24}\) | \(10.0\text{–}119.0\) | 1220          |
| BENZ (N)  | \(P\)2\(\_1\)/\(a\) | \(\text{Co}_{2}\text{F}_{12}\) | \(5.0\text{–}100.0\) | 716           |
| CF3BR (N) | \(P\)2\(\_1\)/\(a\) | \(\text{C}_{2}\text{B}_{12}\text{F}_{12}\) | \(6.0\text{–}150.0\) | 375           |
| CFCL (N)  | \(F\)\(dd\)2 | \(\text{C}_{2}\text{F}_{16}\text{Cl}_{16}\) | \(5.0\text{–}150.0\) | 203           |
| CFI (N)   | \(C\)mc\(a\) | \(\text{C}_{6}\text{F}_{12}\) | \(10.0\text{–}150.0\) | 428           |
| CIME (S)  | \(P\)2\(\_1\)/\(n\) | \(\text{S}_{2}\text{C}_{6}\text{N}_{2}\text{H}_{6}\) | \(8.01\text{–}84.99\) | 924           |
| CROX (X)  | \(P\) 1 | \(\text{Cr}_{2}\text{O}_{21}\) | \(6.0\text{–}80.0\) | 657           |
| CUPZ (X)  | \(P\)bc\(a\) | \(\text{Cu}_{6}\text{N}_{16}\text{C}_{24}\text{H}_{24}\) | \(5.0\text{–}80.0\) | 243           |
| DADA (X)  | \(P\) \(2\_1\)2\(\_1\)\(2\_1\) | \(\text{Ti}_{4}\text{K}_{12}\text{Si}_{12}\text{O}_{40}\) | \(10.0\text{–}95.0\) | 518           |
| EMT (S)   | \(P\)6\(\_2\)m\(\_m\)\(c\) | \(\text{Si}_{12}\text{Al}_{12}\text{Na}_{28}\text{O}_{294}\) | \(4.5\text{–}63.0\) | 670           |
| GAPO (S)  | \(P\) \(b\)\(c\)\(a\) | \(\text{Ga}_{12}\text{P}_{12}\text{O}_{32}\text{F}_{3}\text{C}_{5}\) | \(7.0\text{–}63.69\) | 1235          |
| LAMO (X)  | \(P\)2\(\_1\)/\(a\) | \(\text{La}_{2}\text{Mo}_{2}\text{O}_{12}\) | \(11.0\text{–}69.0\) | 271           |
| LASI (N)  | \(P\)2\(\_1\)/\(c\) | \(\text{La}_{8}\text{Si}_{10}\text{O}_{28}\) | \(10.0\text{–}115.72\) | 253           |
| LEV (S)   | \(R\)3\(m\) | \(\text{Si}_{12}\text{O}_{18}\text{C}_{8}\text{N}_{8}\) | \(8.0\text{–}85.6\) | 323           |
| MCM (S)   | \(P\) \(6\_3m\)\(m\)\(m\) | \(\text{Si}_{12}\text{N}_{9}\text{O}_{12}\) | \(2.2\text{–}50.0\) | 480           |
| MES (X)   | \(P\)2\(\_1\)/\(c\) | \(\text{Si}_{2}\text{N}_{2}\text{O}_{2}\text{H}_{2}\) | \(5.0\text{–}88.0\) | 719           |
| METYL (S) | \(I\)2\(2\_2\)2\(m\) | \(\text{Na}_{16}\text{C}_{7}\text{H}_{6}\) | \(5.2\text{–}70.0\) | 318           |
| NBPO (S)  | \(C\)2\(c\) | \(\text{Na}_{2}\text{O}_{16}\text{P}_{28}\) | \(3.0\text{–}60.0\) | 1201          |
| NIZR (S)  | \(P\)2\(\_1\)/\(n\) | \(\text{Ni}_{2}\text{Zr}_{12}\text{O}_{16}\) | \(8.0\text{–}52.0\) | 627           |
| PBS (S)   | \(P\)bc\(a\) | \(\text{Pb}_{5}\text{Si}_{2}\text{O}_{21}\) | \(7.5\text{–}79.8\) | 477           |
| SAPO (S)  | \(P\)\(n\)\(m\)\(m\) | \(\text{Si}_{12}\text{O}_{4}\text{Na}_{4}\text{C}_{48}\) | \(5.0\text{–}79.98\) | 716           |
| SBPO (S)  | \(P\)2\(\_2\)/\(n\) | \(\text{Si}_{8}\text{P}_{14}\text{O}_{48}\) | \(6.0\text{–}100.0\) | 1071          |
| GTG (S)   | \(I\)4\(\_1\)\(m\)\(d\) | \(\text{Si}_{4}\text{O}_{12}\text{C}_{104}\) | \(8.5\text{–}92.96\) | 451           |
| UTM1 (S)  | \(C\)2\(m\) | \(\text{Si}_{12}\text{O}_{38}\) | \(2.5\text{–}49.97\) | 1133          |
| VFI (S)   | \(P\) \(6\_3\) | \(\text{Al}_{12}\text{P}_{12}\text{O}_{114}\) | \(5.0\text{–}90.0\) | 787           |
| VNI (S)   | \(P\) \(4\_2\)\(1\_2\) | \(\text{Rb}_{12}\text{K}_{12}\text{Si}_{12}\text{Zn}_{12}\text{O}_{288}\) | \(5.0\text{–}60.0\) | 1345          |
| YONO (S)  | \(P\) \(2\_1\) | \(\text{Y}_{2}\text{O}_{2}\text{Zn}_{2}\text{H}_{2}\) | \(7.0\text{–}80.0\) | 680           |
| YURI (X)  | \(P\)2\(\_1\)/\(c\) | \(\text{Na}_{2}\text{S}_{2}\text{O}_{16}\text{C}_{12}\text{F}_{2}\) | \(8.0\text{–}63.96\) | 243           |
4) Phasing the reflections (PHASE routine);
Random phases are given to few pivotal reflections and the phase information is expanded to the large \( |E| \) value reflections by using the most reliable triplets. The phasing trial corresponding to the largest combined figure of merit (CFOM) is selected.

5) Calculation of the Fourier map (FOURIER routine).
The selected phases are used for calculating the Fourier map whose maxima are searched and chemically interpreted. The map is optimized by combining successive structure factor calculations with preliminary least squares cycles.

Therefore, EXPO is a program able to reach the structure solution starting from minimal experimental information. Thanks to the Le Bail tendency to be very sensitive to the starting point, EXPO is more than the trivial combination of the two programs. It is able to exploit information becoming available during the structure solution process itself in the extraction routine to improve the structure factor modulus estimate.

6. The Use of Prior Information

The following types of information provided by the solution process can be used as prior information for improving the extraction of the integrated intensities:

a) Pseudo-translational symmetry information [11].

When a structure is affected by pseudo translational symmetry, a percentage of its electron density repeats itself after an \( u \) vector shift. This means that

\[
\rho_p(r) = \rho(r + u)
\]

where \( \rho_p \) is a \( p \) percentage of the electron density and \( u \) is the pseudo-symmetry vector. In EXPO, the statistical \( |E| \) value analysis is able to reveal the presence of pseudo-symmetry, to recognize the percent-

| Code  | EXTRA (Le Bail based) | ALLHKL (Pawley based) |
|-------|-----------------------|-----------------------|
| AGPZ  | 0.53 0.60 0.15       | 0.53 0.60 0.15       |
| BACO  | 0.35 0.39 0.05       | 0.35 0.39 0.05       |
| BENZ  | 0.43 0.81 0.22       | 0.43 0.81 0.22       |
| CF3BR | 0.33 0.41 0.12       | 0.33 0.41 0.12       |
| CFCL  | 0.22 0.34 0.06       | 0.22 0.34 0.06       |
| CFI   | 0.51 0.78 0.05       | 0.51 0.78 0.05       |
| CROX  | 0.39 0.60 0.13       | 0.39 0.60 0.13       |
| CUPZ  | 0.50 0.86 0.11       | 0.50 0.86 0.11       |
| LAMO  | 0.39 0.46 0.23       | 0.39 0.46 0.23       |
| LASI  | 0.40 0.47 0.11       | 0.40 0.47 0.11       |
| LEV   | 0.60 0.77 0.06       | 0.60 0.77 0.06       |
| MES   | 0.50 0.77 0.06       | 0.50 0.77 0.06       |
| METYL | 0.33 0.53 0.10       | 0.33 0.53 0.10       |
| NBPO  | 0.43 0.52 0.18       | 0.43 0.52 0.18       |
| NIZR  | 0.41 0.61 0.18       | 0.41 0.61 0.18       |
| PBS   | 0.43 0.48 0.09       | 0.43 0.48 0.09       |
| SAPO  | 0.47 0.87 0.12       | 0.47 0.87 0.12       |
| SBPO  | 0.51 0.82 0.10       | 0.51 0.82 0.10       |
| SULPH | 0.35 0.47 0.08       | 0.35 0.47 0.08       |
| YONO  | 0.33 0.44 0.10       | 0.33 0.44 0.10       |

| Code  | EXTRA Protocol 1 | ALLHKL Protocol 1 | EXTRA Protocol 2 | ALLHKL Protocol 2 |
|-------|------------------|------------------|------------------|------------------|
| AGPZ  | 0.53 0.60 0.15   | 0.62 0.61       | 0.26 0.23       | 0.53 0.60 0.15   |
| BACO  | 0.34 0.39 0.05   | 0.34 0.39       | 0.21 0.23       | 0.34 0.39 0.05   |
| BENZ  | 0.43 0.81 0.22   | 0.52 0.80       | 0.22 0.23       | 0.43 0.81 0.22   |
| CF3BR | 0.33 0.41 0.12   | 0.59 0.46       | 0.15 0.23       | 0.33 0.41 0.12   |
| CFCL  | 0.22 0.34 0.27   | 0.12 0.27       | 0.12 0.27       | 0.22 0.34 0.27   |
| CFI   | 0.50 0.74 0.84   | 0.32 0.46       | 0.15 0.23       | 0.50 0.74 0.84   |
| CROX  | 0.39 0.55 0.60   | 0.16 0.23       | 0.16 0.23       | 0.39 0.55 0.60   |
| CUPZ  | 0.50 0.86 0.74   | 0.24 0.45       | 0.24 0.45       | 0.50 0.86 0.74   |
| LAMO  | 0.39 0.39 0.45   | 0.22 0.45       | 0.22 0.45       | 0.39 0.39 0.45   |
| LASI  | 0.39 0.45 0.47   | 0.16 0.23       | 0.16 0.23       | 0.39 0.45 0.47   |
| LEV   | 0.60 0.77 0.06   | 0.23 0.45       | 0.23 0.45       | 0.60 0.77 0.06   |
| MES   | 0.49 0.73 0.77   | 0.26 0.50       | 0.26 0.50       | 0.49 0.73 0.77   |
| METYL | 0.32 0.37 0.54   | 0.25 0.54       | 0.25 0.54       | 0.32 0.37 0.54   |
| NBPO  | 0.44 0.52 1.11   | 0.12 0.23       | 0.12 0.23       | 0.44 0.52 1.11   |
| NIZR  | 0.43 0.61 0.55   | 0.24 0.45       | 0.24 0.45       | 0.43 0.61 0.55   |
| PBS   | 0.43 0.48 0.43   | 0.26 0.48       | 0.26 0.48       | 0.43 0.48 0.43   |
| SAPO  | 0.47 0.87 0.86   | 0.19 0.23       | 0.19 0.23       | 0.47 0.87 0.86   |
| SBPO  | 0.51 0.82 0.79   | 0.15 0.23       | 0.15 0.23       | 0.51 0.82 0.79   |
| SULPH | 0.34 0.44 0.56   | 0.20 0.44       | 0.20 0.44       | 0.34 0.44 0.56   |
| YONO  | 0.31 0.32 0.42   | 0.18 0.23       | 0.18 0.23       | 0.31 0.32 0.42   |
Concerning the use of prior information are shown. The results become more reliable. In Table 4, the results extraction is more efficient and the structure solution the starting point closer to the true one. In this way, the considering different kinds of prior information to make sensitive to the starting point can be exploited by considering:

(a) The located fragment information [14].
(b) Probabilistic estimate information [12].
(c) The Patterson information [13].
(d) The located fragment information [14].

If a traditional Le Bail extraction EXPO run is able to locate a fragment in correct way the structure factor moduli can be exploited as starting point in a new Le Bail extraction. The LEs statistical estimates as starting values.

Therefore, the use of prior information can help when the obtained traditional Le Bail extraction solution is not reliable. The following suggestions can be taken in consideration for optimise its use and for avoiding the bad combined use of prior information because of their correlation: 1) if pseudo-symmetry is revealed, and especially when the detected percentage is large, it is convenient to use it; 2) if no pseudo-symmetry effect is detected, but the structure contains heavy atoms, the use of Patterson information can improve the results.

| Code | \(R_D\) | \(R_{PSEUD}\) | \(R_{PATT}\) | \(R_{PROB}\) | \(R_{TRUE}\) |
|------|--------|---------|--------|--------|--------|
| AGPZ | 52.28  | 33.51   | 38.93  | 47.05  | 24.18  |
| BACO | 31.32  | 28.09   | 28.22  | 16.94  |
| BENZ | 41.42  | 35.83   | 36.81  | 21.28  |
| CF3BR| 29.89  | 27.55   | 27.43  | 10.95  |
| CFCL | 21.14  | 15.64   | 19.29  | 9.99   |
| CFI  | 49.29  | 45.04   | 46.23  | 30.25  |
| CDOX | 36.65  | 33.11   | 31.06  | 15.71  |
| CUPZ | 47.07  | 34.79   | 34.49  | 21.90  |
| LAMO | 35.59  | 34.99   | 35.11  | 25.93  |
| LASI | 37.81  | 37.07   | 35.68  | 12.99  |
| LEV  | 58.60  | 51.70   | 55.70  | 22.38  |
| MES  | 46.39  | 44.08   | 42.09  | 25.65  |
| METYL| 28.98  | 27.13   | 27.31  | 22.74  |
| NBO  | 38.95  | 32.73   | 24.36  | 29.85  | 8.27   |
| NIZR | 42.17  | 41.65   | 37.07  | 36.06  | 21.25  |
| PBS  | 40.82  | 38.10   | 38.32  | 26.50  |
| SAPO | 45.33  | 41.09   | 41.77  | 17.37  |
| SBPO | 48.56  | 30.85   | 28.88  | 31.69  | 13.17  |
| SULPH| 32.53  | 27.42   | 30.12  | 19.31  |
| YONO | 31.95  | 27.42   | 25.54  | 16.75  |

Table 4. \(R_F\) reliability parameters (×100). \(R_D\) is in the prior pseudo-symmetry information case; \(R_{PSEUD}\) is in the prior pseudo-symmetry information case; \(R_{PATT}\) is in the prior Patterson information case; \(R_{PROB}\) is in the prior probabilistic estimate information case, \(R_{TRUE}\) when the true structure factor moduli are used.
3) the probabilistic estimate information can be used in all the cases;
4) if a fragment is located it can be exploited.

The structure solutions supplied by EXPO are shown in Table 6, where for each test structure we have: the maximum \((\sin \theta / \lambda)^2\) value, the number of reflections, the number of independent observations (see [15] for details), the number of atoms to find in the asymmetric unit and the number of atoms found by EXPO (in a traditional Le Bail extraction or intensity-recycled run). Most of the structures are completely solved. This doesn’t occur when the data quality is poor (small \((\sin \theta / \lambda)^2\) value and/or a large overlapping degree).

### 7. The Random Approach

When no prior information is available, or when it is poor, a recently developed procedure can be attempted [16]. It is based on a random approach and it works so that, for each cluster of overlapping reflections, some random partitions of the cluster overall intensity are considered. The partition corresponding to the best fit (the lowest \(R_F\) value) in the cluster local range is selected as the most reliable one and it provides the integrated intensity values to use as starting ones in the Le Bail formula. The random procedure is applied before each Le Bail cycle. The merit of the new approach is to break the Le Bail tendency to equipartition the intensity of a group of overlapping reflections. Its aim is to modify the equipartitioned intensities: a necessary goal, if the modified intensities correspond to the pivotal reflections in the phasing process. The results of the random procedure are shown in Table 7, where the phase error in the traditional Le Bail extraction case

| Code      | Selected fragment (%) | \(R_D\) | \(R_{FRAG}\) |
|-----------|-----------------------|--------|-------------|
| AGPZ      | 1 Ag (97.2 %)         | 51.12  | 29.47       |
| BAMO      | 2 Ba (37.8 %)         | 42.35  | 38.06       |
| CUPZ      | 1 Cu (92.3 %)         | 46.98  | 26.53       |
| DADA      | 1 Ti 2 K (54.4 %)     | 33.65  | 31.92       |
| LAMO      | 1 La 2 Mo (55.8 %)    | 35.14  | 32.75       |
| LASI      | 2 La (34.1 %)         | 37.68  | 33.55       |
| NBPO      | 3 Nb (82.7 %)         | 40.19  | 23.99       |
| NIZR      | 2 Zr (68.6 %)         | 41.85  | 36.09       |
| SBPO      | 2 Sb (97.9 %)         | 49.52  | 23.25       |
| YONO      | 4 Y (95.8 %)          | 31.93  | 20.15       |

### Table 6.
For each test structure: the maximum \((\sin \theta / \lambda)^2\), the number of reflections, the number of independent observations, the number of atoms to locate (NATS1) and the number of atoms correctly located by EXPO in the asymmetric unit (NATS2) are given

| Code    | \((\sin \theta / \lambda)^2\) | \(M\) | \(M_{ind}\) | NATS1 | NATS2 |
|---------|-------------------------------|------|------------|-------|-------|
| AGPZ    | 0.17                          | 258  | 72         | 6     | 4     |
| BACO    | 0.26                          | 272  | 127        | 6     | completed |
| BAMO    | 0.32                          | 1220 | 396        | 28    | 27    |
| BENZ    | 0.30                          | 716  | 258        | 9     | completed |
| CF3BR   | 0.25                          | 375  | 141        | 3     | completed |
| CFCL    | 0.37                          | 203  | 106        | 3     | completed |
| CFI     | 0.37                          | 429  | 149        | 3     | completed |
| CINE    | 0.19                          | 924  | 484        | 17    | completed |
| CROX    | 0.21                          | 657  | 202        | 15    | completed |
| CUPZ    | 0.17                          | 243  | 72         | 6     | 5     |
| DADA    | 0.23                          | 518  | 197        | 16    | completed |
| LAMO    | 0.13                          | 271  | 126        | 14    | 12    |
| LASI    | 0.13                          | 253  | 105        | 11    | 8     |
| LEV     | 0.19                          | 323  | 103        | 17    | 8     |
| MES     | 0.20                          | 719  | 229        | 13    | 11    |
| METYL   | 0.27                          | 318  | 169        | 5     | completed |
| NBPO    | 0.25                          | 1201 | 481        | 22    | completed |
| NIZR    | 0.18                          | 628  | 239        | 18    | 11    |
| PBS     | 0.27                          | 477  | 179        | 6     | 5     |
| SAPO    | 0.17                          | 717  | 183        | 21    | 9     |
| SBPO    | 0.28                          | 1071 | 337        | 17    | 13    |
| SULPH   | 0.26                          | 220  | 93         | 3     | completed |
| YONO    | 0.27                          | 680  | 203        | 18    | completed |

### Table 7.
For some test structures the phase error (the difference between the direct methods phases and the true phases) corresponding to the EXPO traditional Le Bail extraction run (ERR1) and to the use of the random procedure (ERR2) are given

| Code    | ERR1 (°) | ERR2 (°) |
|---------|----------|----------|
| AND1    | 29.87    | 20.68    |
| DADA    | 48.33    | 23.32    |
| GAPO    | 37.18    | 29.90    |
| LEV     | 73.55    | 32.53    |
| UTM1    | 66.80    | 23.02    |
| YURI    | 28.24    | 22.50    |

(Err1) and in the random case (ERR2) are given for some test structures. The values corresponding to ERR2 are always much better than ERR1. This means that the power of the new procedure, to modify a small number of reflections that are very important in the phasing process, remarkably improves the phasing process, even though, on average, no more accurate structure factor moduli estimates are obtained.
8. The POLPO Procedure

The solution provided by direct methods is frequently incomplete. In particular, this happens in the case of heavy atom structure when the heavy atoms are easily located, but the light atoms are hardly recognised. The traditional approach for completing a partial solution consists of combining Fourier map calculations with Rietveld refinement. The trend is not trivial, not automatic, not fast. The new POLPO procedure [17] has been introduced in EXPO for completing the structure when the structure cations are located. The procedure uses the polyhedral information and it is based on the Monte Carlo technique. The starting point is the cation positions supplied by direct methods. The user gives the polyhedral information by using directives about the polyhedron type, the corresponding cation label, the expected polyhedral average distance, the distance tolerance and the angle tolerance. The procedure automatically calculates the cation connectivity [17]. Several configurations obeying the requested polyhedral and connectivity rules are built. The geometrical construction takes into account the tolerance about the distances and angles. Some configurations are rejected because they are chemically inconsistent. Among the remaining possible configurations, the model corresponding to the best fit between the observed and the calculated profile (the lowest $R_p$ value) is selected. Table 8 shows the POLPO results: the number of feasible obtained solutions, the lowest $R_p$ value corresponding to the chosen model, the number of anions located in the asymmetric unit (in parentheses the true number), the average distances between the POLPO positions and the true ones and the CPU time are given. It can be seen that all the structures are completed in few time. The discrepancy with regard to the number of anions depends on the imperfectly located positions of the starting cations and on the fact that the construction by POLPO is carried out in a geometrically perfect way. The POLPO procedure is currently being enhanced with the aim of completing a structure when only some cations are positioned.

9. Conclusions

Thank to its graphical interface, EXPO is a very user-friendly program. It is able to give different opportunities for overcoming the difficulties in solving ab initio crystal structures by powder diffraction data. The next version of EXPO will include N-TREOR [18], a modified and updated version of the program for indexing TREOR90 [19], the POLPO procedure and new strategies for optimising the Fourier map.

10. References

[1] A. Altomare, M. C. Burla, B. Carrozzini, G. Cascarano, C. Giacovazzo, A. Gagliardi, A. G. G. Moliterni, G. Polidori, and R. Rizzi, J. Appl. Cryst. 32, 339-340 (1999).
[2] C. Giacovazzo, H. L. Monaco, D. Viterbo, F. Scordari, G. Gilli, G. Zanotti, and M. Catti, Fundamentals of Crystallography, Oxford University Press (1992).
[3] G. S. Pawley, J. Appl. Cryst. 14, 357-361 (1981).
[4] A. Le Bail, H. Duroy, and J. L. Fourquet, Math. Res. Bull. 23, 447-452 (1988).
[5] I. Jansen, R. Peschar, and H. Schenk, J. Appl. Cryst. 25, 231-236 (1992).
[6] D. S. Sivia and W. I. F. David, Acta Cryst. A50, 703-714 (1994).
[7] H. M. Rietveld, J. Appl. Cryst. 2, 65-71 (1969).
[8] A. Altomare, M. C. Burla, G. Cascarano, C. Giacovazzo, A. Guagliardi, A. G. G. Moliterni, and G. Polidori, J. Appl. Cryst. 28, 842-846 (1995).
[9] A. Altomare, B. Carrozzini, C. Giacovazzo, A. Gagliardi, A. G. G. Moliterni, and R. Rizzi, J. Appl. Cryst. 29, 667-673 (1996).
[10] A. Altomare, G. Cascarano, C. Giacovazzo, A. Guagliardi, M. C. Burla, G. Polidori, and M. Camalli, J. Appl. Cryst. 27, 435-436 (1994).
[11] A. Altomare, J. Foadi, C. Giacovazzo, A. Guagliardi, and A. G. G. Moliterni, J. Appl. Cryst. 29, 674-681 (1996).
[12] B. Carrozzini, C. Giacovazzo, A. Guagliardi, R. Rizzi, M. C. Burla, and G. Polidori, J. Appl. Cryst. 30, 92-97 (1997).
[13] A. Altomare, J. Foadi, C. Giacovazzo, A. G. G. Moliterni, M. C. Burla, and G. Polidori, J. Appl. Cryst. 31, 74-77 (1998).

Table 8. For each test structure the number of feasible solutions (NFS), the best $R_p$ value, the number of located anions (NA) respect to the true number (in parentheses), the average distance from the published atomic positions $\langle d \rangle$ and the CPU time are given.

| Code | NFS | $R_p$ | NA | $\langle d \rangle$ (Å) | CPU time (s) |
|-----|-----|-------|-----|-------------------|--------------|
| CROX | 5   | 0.22  | 11 (11) | 0.24            | 3            |
| EMT  | 2   | 0.12  | 12 (12) | 0.23            | 103          |
| MCM  | 3   | 0.23  | 15 (13) | 0.23            | 257          |
| NIZR | 2   | 0.22  | 13 (12) | 0.23            | 34           |
| SAPO | 6   | 0.29  | 10 (10) | 0.27            | 38           |
| SGT  | 6   | 0.49  | 11 (7)  | 0.45            | 98           |
| UTMI | 9   | 0.23  | 15 (13) | 0.34            | 94           |
| VFI  | 8   | 0.25  | 14 (14) | 0.37            | 81           |
| VNI  | 4   | 0.19  | 35 (30) | 0.24            | 181          |

$^a$ Compaq\textsuperscript{1} personal Workstation 500au.

\textsuperscript{1} Certain commercial equipment, instruments, or materials are identified in this paper to foster understanding. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.
[14] A. Altomare, C. Giacovazzo, A. Guagliardi, A. G. G. Moliterni, and R. Rizzi, J. Appl. Cryst. 32, 963-967 (1999).
[15] A. Altomare, G. Cascarano, C. Giacovazzo, A. Gagliardi, and A. G. G. Moliterni, J. Appl. Cryst. 28, 738-744 (1995).
[16] A. Altomare, C. Giacovazzo, A. G. G. Moliterni, and R. Rizzi, J. Appl. Cryst. 34, 704-709 (2001).
[17] A. Altomare, C. Giacovazzo, A. Guagliardi, A. G. G Moliterni, and R. Rizzi, J. Appl. Cryst. 33, 1305-1310 (2000).
[18] A. Altomare, C. Giacovazzo, A. Guagliardi, A. G. G Moliterni, R. Rizzi, and P.-E. Werner, J. Appl. Cryst. 33, 1180-1186 (2000).
[19] P.-E. Werner, L. Eriksson, and M. Westdahl, J. Appl. Cryst. 18, 367-370 (1985).

About the authors: Angela Altomare, Anna Grazia, Giuseppina Moliterni, and Rosanna Rizzi are physicists, and Carmelo Giacovazzo is a Full Professor of Mineralogy at the National Council of Research, Bari, Italy.