Ab initio structure determination via powder X-ray diffraction

DIGAMBER G POROB and T N GURU ROW*
Solid State and Structural Chemistry Unit, Indian Institute of Science,
Bangalore 560 012, India
e-mail: ssctng@sscu.iisc.ernet.in

Abstract. Structure determination by powder X-ray diffraction data has gone through a recent surge since it has become important to get to the structural information of materials which do not yield good quality single crystals. Although the method of structure completion when once the starting model is provided is facile through the Rietveld refinement technique, the structure solution *ab initio* is still not push-button technology. In this article a survey of the recent development in this area is provided with an illustration of the structure determination of a-$NaBi_3V_2O_{10}$.

Keywords. Structure determination; powder X-ray diffraction; pattern decomposition; structure refinements.

1. Introduction

Powder X-ray diffraction has routinely been used as a non-destructive fingerprinting technique in laboratory and industry for several decades. It has also been used in studies related to structural phase transitions at variable temperature and pressure conditions. Powder data is especially useful to deduce accurate cell parameters. Rietveld's refinement procedure[1,2] has revolutionized the application of powder X-ray diffraction by resulting in a large number of structures being refined in the last decade. If a suitable starting model is available, it has become routine to refine structures from a decent quality powder X-ray diffraction data using standard packages. Obviously, the generation of the starting model has taken priority over the last few years. With materials in the form of single crystals the generation of the starting model has become trivial for molecules with about a hundred atoms in the asymmetric unit. Use of direct methods and Patterson search methods basically solve the phase problem for small molecules, if single crystal data are available. However, in case of powder diffraction data on crystalline materials, the phase problem solution and the generation of the starting model becomes a challenge, even in case of structures with a few atoms. The fact that the three-dimensional reciprocal lattice information is condensed onto a one-dimensional intensity profile with respect to 2-theta in powder diffraction is the main concern. This leads to overlapping of the diffraction peaks, considerable (background that usually is not very accurately accounted for) and preferred orientation among the crystallites. Phase determination methods need accurately measured intensities from which structure factor modulus are evaluated for each clearly identified reflection. Several reviews and articles have appeared in recent years[3–6] in this area and the

---

*For correspondence

references are available on the personal web-site of Armel Le Bail[7] and a detailed tutorial for structure determination by powder diffraction (SDPD).

In this article, currently popular pattern decomposition methods will be discussed and, since successful application depends on the quality of the diffraction pattern, the approaches to obtain high-resolution data sets from powder X-ray diffraction techniques will be outlined. A step-by-step operational procedure to perform *ab initio* structure determination will be described with an example from our recent results[8]. Methods that utilize additional information about the structure combined with the Monte Carlo technique[9,10] or the genetic algorithm[11] are also now in use and will be described briefly.

2.   Data collection strategies

The sample preparation could turn out to be a severe rate-limiting step, especially if the grain size is variable. The compound to be investigated should be properly ground using a pestle and mortar or should be milled to uniform grain size. It is generally suggested that the sample be sieved through a fine grade mesh ($< 60 \mu$). The best approach is to fill the sample in a capillary for Debye–Scherrer mode of data acquisition, which reduces the preferred orientation errors.

Both conventional X-ray and Synchrotron X-rays can be employed. Synchrotron radiation is not necessary if the problem is very simple. For example, in the case of powder pattern for fullerene (space group *Im*3, 18 Å) there is absolutely no gain in using synchrotron data (10 times the best resolution data from conventional sources). After all, overlapping reflections continue to overlap. Also, if the sample is of average crystalline quality there is no gain in going for the expensive, time-involved synchrotron data. However, for complex systems the high-resolution data obtained with a synchrotron source would be definite advantage. At the synchrotron source in ESRF, the FWHM is as low as 0·008° in 2-theta. If the grain sizes of the sample are a few microns and the sample is free of structure imperfections this level of data quality is stupendous. It would result in about 77,500 data points in the range of 5 to 160° in 2-theta with the usual wavelength, and thus the structure of a small protein becomes a possibility.

It is recommended that on a typical laboratory machine, high quality powder diffraction data should be collected in order to obtain 0·02° accuracy in 2-theta. A typical good data set would be to collect from 3 to 100° in 2-theta in steps of 0·02° at 4–10 seconds per step. Use of Debye–Scherrer mode is preferred to reduce orientation effects and a PSD detector would be of immense use. Monochromatic $K\alpha_1$ radiation would be the ideal choice. It must be added that data collected by routine reflection mode geometry at its highest resolution on any conventional diffractometer using the step scan option is suitable, provided sufficient care is taken to apply appropriate corrections. It is also sufficient to collect data without $K\alpha_2$ stripping but utilising suitable software routes to handle the data.

3.   Indexing process

The problem of indexing and determination of the space group in powder diffraction analysis is also non-trivial. Several crystallographic packages are available to get the indexing done fairly automatically but certainly not without mistakes. The most used

programs are TREOR, ITO and DICVOL, which are present in the CCP14 [12–15] package. The merits and demerits of these packages and their success rates in indexing unknown samples are discussed in the tutorial package in the web-site mentioned above. It is generally possible to obtain an unique solution to the indexing of the peak and consequently the crystal system, space group and unit cell parameters, though this could end in a trap. *A priori* knowledge of similar systems, and information from the data bases help in getting over this step with reasonable success rates.

4. Pattern decomposition

The major step in the determination of the structure by *ab-initio* methods would be to resolve the powder pattern to recover individual integrated intensities from the diffraction pattern. This would ensure the assignment of a structure factor modulus for the application of direct methods or Patterson methods to these individual reflections. The success of pattern decomposition depends on two major factors, one the quality of the diffraction data and the other the efficacy of the mathematical approach on which it is based. There are two different philosophical approaches, one based on the whole pattern fitting with intrinsic cell constraints and the other with no restraints on the cell parameters.

   The basic idea here is to resolve the profile into individual reflections and consequently extract the integrated intensity information for each reflection from the peak profile analysis. This involves the generation of the profile in terms of analytical functions or convolutions of analytical functions such as Gaussian, Lorenzian, pseudo-Voight, and Pearson VII representations (table 1). An example of such pattern decomposition is illustrated in figure 1.

   The starting positions of the peaks are generated based on the refined cell parameters after the indexing is done. There are several approaches using this no constraint approach on cell parameters that are in practice, but the most widely used is that due to Rudolf and Clearfield [16] based on the MLE (maximum likelihood estimation) method.

Table 1. Profile functions.

Function	Name
$[C_0^{1/2}/H_k \boldsymbol{p}^{1/2}]\exp(-C_o(2\boldsymbol{q}_i - 2\boldsymbol{q}_k)^2/H_k^2)$	Gaussian (G)
$(C_1^{1/2}/\boldsymbol{p}H_k)1/[1 + C_1(2\boldsymbol{q}_i - 2\boldsymbol{q}_k)^2/H_k^2]^2$	Lorentzian (L)
$(2C_2^{1/2}/\boldsymbol{p}H_k)1/[1 + C_2(2\boldsymbol{q}_i - 2\boldsymbol{q}_k)^2/H_k^2]^2$	Mod 1 Lorentzian
$(C_3^{1/2}/2H_k)1/[1 + C_3(2\boldsymbol{q}_i - 2\boldsymbol{q}_k)^2/H_k^2]^{3/2}$	Mod 2 Lorentzian
$\boldsymbol{h}L + (1 - \boldsymbol{h})G$	pseudo-Voigt
The mixing parameter η can be refined as a linear function of $2\boldsymbol{q}$ wherein the refinable variables are NA and NB.	
$\boldsymbol{h} = NA + NB*2\boldsymbol{q}$	
$C_4/H_k[1 + 4*(2^{1/m} - 1)(2\boldsymbol{q}_i - 2\boldsymbol{q}_k)^2/H_k^2]^{-m}$	Pearson VII
m can be refined as a function of $2\boldsymbol{q}$	
$m = NA + (NB/2\boldsymbol{q}) + NC/(2\boldsymbol{q})^2$	
where NA, BA and NC are refinable parameters	

H_k is the full-width-at-half-maximum (FWHM) of the kth Bragg reflection; $C_0 = 4\ln2$; $C_1 = 4$; $C_2 = 4(2^{1/2}-1)$; $C_3 = 4(2^{2/3}-1)$; $C_4 = 2\sqrt{m}\ (2^{1/m}-1)^{1/2}/(m-0\cdot5)\boldsymbol{p}^{1/2}$
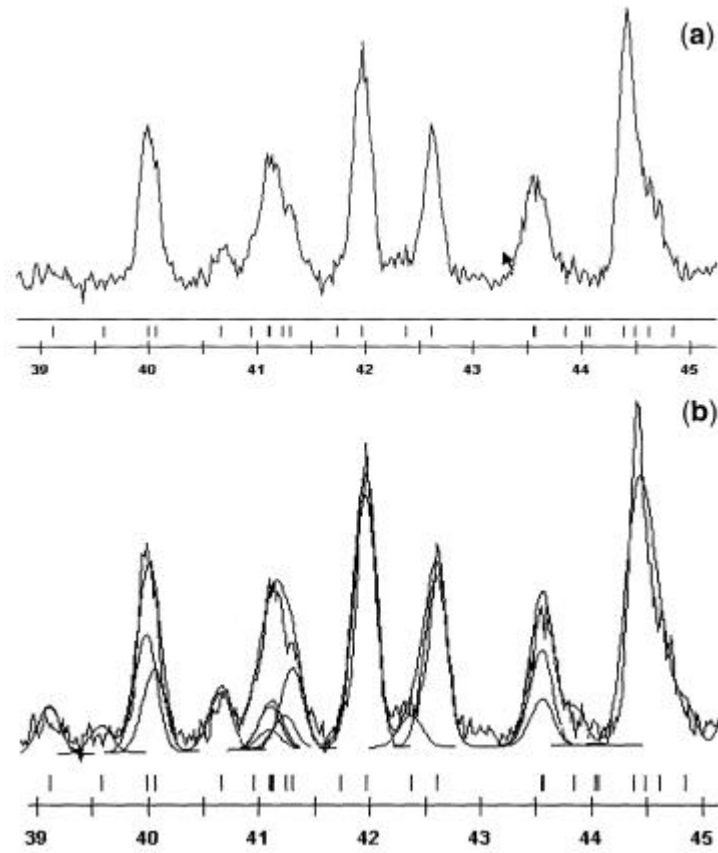
Figure 1. An illustrative example of pattern decomposition: (**a**) Pattern with the individual component reflections identified; (**b**) profiles of individual reflections compounding to the full pattern.

However, the two popular methods in use are based on the whole pattern decomposition with cell constraints and are due to Pawley and Le Bail[17–19]. The basic philosophy is to get the calculated profile, Y_{ci}, using the well-known Rietveld strategy,

$$y_{ci} = s \sum L_K |F_K|^2 f(2q - 2q_K) P_K A + y_{bi}.$$

The quantity minimized in the least-squares refinement is the residual, S_y:

$$S_y = \sum w_i (y_i - y_{ci})^2,$$

where $w_i = 1/y_i$, y_i = observed intensity at the ith step, y_{ci} = calculated intensity at the ith step, s is the scale factor, K represents the Miller indices, h, k, l for a Bragg reflection, L_k contains the Lorentz polarization and multiplicity factors, f is the reflection profile function, P_k is the preferred orientation function, A is an absorption

factor, F_k is the structure factor for the kth Bragg reflection, y_{bi} is the background intensity at the ith step.

In Pawley's approach, for each possible reflection an additional parameter gets added to the refinement strategy thereby resulting in n additional parameters for n reflections. Toraya[18] modified Pawley's approach to include changes in the profile function arising from specimen and instrumental causes. The advantage in this whole pattern fitting is that the profile shape is interpolated into regions of low resolution and of weak reflections such that all reflections are refined with equal weight irrespective of the intensities they carry. The reflections thus generated are now suitable for the application of routine phase determination processes as for example direct methods and Patterson search methods. Packages like ALLHKL (Pawley) and PROFIT (Toraya) are used to do the pattern decomposition and generate the structure factor moduli for input to packages like and GSAS[20] and SHELX[21].

In the Le Bail approach[19], the structure factors are extracted by iterating the Rietveld decomposition formula giving the structure factor moduli which are used to estimate R_B and R_F reliability indices (table 2). Here the only parameters refined are the cell and the profile parameters. At the starting stage, reflections as determined by the cell and the space group are given the same starting intensity. However, with every cycle of cell and profile parameter refinements, the intensities get re-determined with R_B and R_F as guidelines. The method is thus quite fast and efficient and hence is used more and more for the extraction of intensities from powder data. The packages EXPO[22,23], FULLPROF[24] provide the necessary software for this purpose and are often used for both organic and inorganic structures fairly routinely.

5. Structure determination

5.1 *Ab initio methods*

Application of single crystal methods for solving the phase problem either by direct methods or by the Patterson method needs to have all possible Bragg reflections with an atom to reflection ratio of one to ten in an ideal situation. Pattern decomposition never gives anything close to this ideal situation and the so-called overlap peaks remain uncertain. In fact, the closely overlapping peaks keep the same structure factor value till the end of the iterative scheme. In most of the cases, especially if there are unequal weight atoms in the structure, the heavier atoms can generally be located by the first

Table 2. Important numerical criteria of fit.

$R_F = \sum |(I_K(\text{obs}))^{1/2} - (I_K(\text{calc}))^{1/2}|/\sum (I_K(\text{obs}))^{1/2}$	R-structure factor
$R_B = \sum |(I_K(\text{obs}) - (I_K(\text{calc})|/\sum (I_K(\text{obs}))$	R-Bragg factor
$R_P = \sum |(y_i(\text{obs}) - (y_i(\text{calc})|/\sum (y_i(\text{obs}))$	R-pattern (profile)
$R_{wP} = \{\sum w_i|(y_i(\text{obs}) - (y_i(\text{calc})^2|/\sum w_i(y_i(\text{obs})^2\}^{1/2}$	R-weighted pattern (profile)
$R_E = [(N - P)/\sum w_i y_i(\text{obs})^2]^{1/2}$	R expected
where N and P are the number of profile points and refined parameters respectively	
Goodness-of-fit $S = (R_{wp}/R_E)^2$	Durbin–Watson statistics 'd'
$d = \sum ((\Delta_i/\boldsymbol{s}_i) - (\Delta_{i-1}/\boldsymbol{s}_{i-1}))^2 [\sum (\Delta_i/\boldsymbol{s}_i)^2]^{-1}$	
where $\Delta_i = y_i(\text{obs}) - y_i(\text{calc})$ and $\boldsymbol{s}_i$ is the standard deviation	

E-map from either the direct or the Patterson methods. It then becomes an iterative process of running difference Fourier calculations to locate the remaining atoms. When once a meaningful model emerges, application of the Rietveld refinement procedure completes the structure determination and refinement. Recent reviews[3–6] evaluate this methodology in great detail. It is generally observed that the success rate depends on how efficient the pattern decomposition is and how well the overlapping of peaks is handled. It also depends on the preferred orientation effects, the nature of the background and the quality of the data set.

The maximum entropy and likelihood methods adopt similar strategy to that in direct methods[25–27]. In this approach, groups of overlapping intensities (program MICE) are summed up to give combined group intensity. These are then put along with the non-overlapping intensities and normalized structure factors are obtained as input for the iterative application of entropy maximization.

5.2   *Other methods*

Considerable progress has been made in the generation of starting models from other experimental and theoretical sources. Besides, initial models for the molecular structure or fragment could be derived from methods other than crystallography as for example NMR, electron microscopy and other spectroscopic techniques. Additional knowledge combined with the Monte Carlo technique[9,10] can often help in getting off to a reasonable start for structure determination. Of particular use are methods developed based on the genetic algorithm[11]. If the geometry of a fragment of the molecule is known *a priori*, a method that employs translation and rotation of the known fragment in the asymmetric unit followed by bond-restraint refinements has been successfully applied to several organic structures.

The Monte Carlo method[4,9] differs from the traditional approaches as it operates in direct space rather than in reciprocal space as do most of the above mentioned methods. The structural models are postulated independently of the diffraction data and they are evaluated with respect to the matching between the calculated powder diffraction data and the experimentally observed data. Simulated annealing also uses the Monte Carlo algorithm to generate the structural models but the acceptance criteria differ for the starting model. Harris and Tremayne[4] have described the methodology and application of these procedures extensively in their review.

A method based on genetic algorithm has been developed by Shankland *et al*[11]. Genetic algorithms are a family of computational models inspired by evolution. These are essentially function optimisers. In powder diffraction, to start with, a random structural model is generated and is then provided reproduction opportunities such that the newly generated patterns which represent better solutions are given more chances to "reproduce" than poorer solutions. The "goodness" of the solution is typically defined with respect to the current proposed model. The philosophy is to fit the diffraction data generated from trial structures against the measured diffraction data and this approach has the ability to handle flexible molecules and multiple fragments. It is also highly efficient from the computational point of view and takes full advantage of the implicit parallelism of the genetic algorithm.

ESPOIR[7] is a program developed recently that essentially employs a reverse Monte Carlo and pseudo-simulated annealing code for *ab initio* crystal structure determination. This approach could be used for solving structures either from "scratch"

(from a completely random starting model) or by "molecule location" (analogous to the molecule replacement method). This is recommended as a last chance program to use after all the classical methods (specifically direct and Patterson methods) fail to produce the final solution.

6. Structure refinements

Any structure determination, either by single crystal methods or by powder methods, has to be done in two stages. The first stage is to obtain a starting model, which any of the methods described above provide. The second stage is the refinement stage, where the structure is refined to completion based on the full data set. In cases only a partial structure is obtained in the starting model, the rest of the positions of the missing atoms can usually be found by difference Fourier synthesis. The success here essentially depends on the percentage of the final structure represented in the starting model. In general, if 50% of the total electron density is located in the starting model correctly for an equal atom structure, the rest of the structural data can be generated by difference Fourier synthesis.

Rietveld refinement[1,2] would be sufficient in most cases to get to the best-fit situation. The calculated powder pattern is compared with the experimental pattern at each point on the diffractogram and selected parameters (both profile and structural) are refined by the least-squares technique to get the final structure. Several R-factors (table 2) are used to evaluate the correctness of the refinement process. The programs GSAS and FULLPROF are the most generally used packages for this purpose.

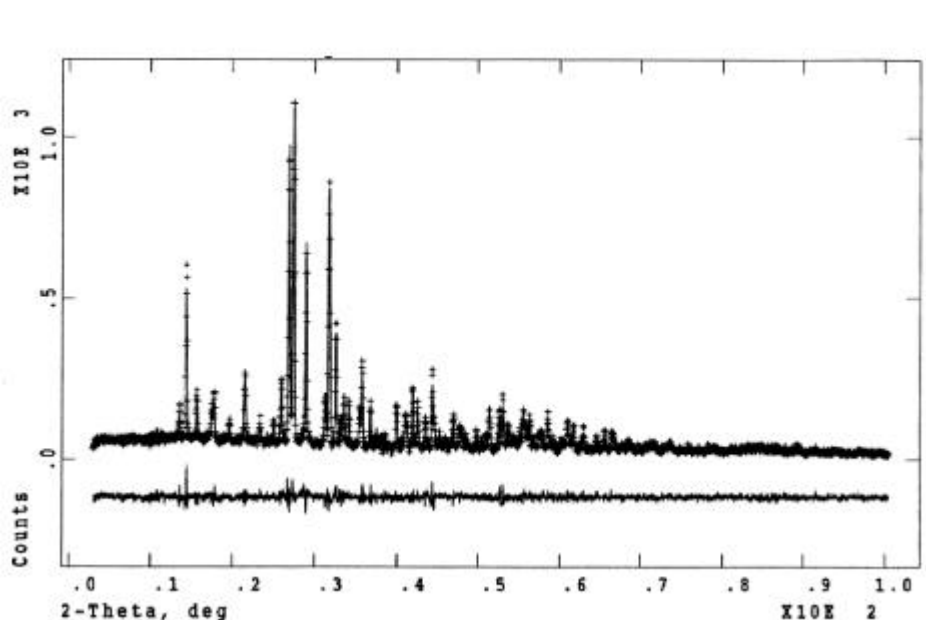## 7. Example: structure of $NaBi_3V_2O_{10}$

A compound $NaBi_3V_2O_{10}$ was recently isolated from the BINAVOX solid solution within $Na_2O–Bi_2O_3–V_2O_5$ ternary system and the structure was determined to be isostructural to $Pb_2Bi_2V_2O_{10}$[28,29]. We found that the simulated X-ray diffraction pattern of this structure does not match with the experimental one reported earlier. This led us to believe that the structure reported is of a different polymorph of $NaBi_3V_2O_{10}$. Synthetic protocol given by Sinclair *et al*[28,29] was thus used to synthesize $NaBi_3V_2O_{10}$ and the structure was determined *ab initio* using powder X-ray diffraction data. The structure established beyond doubt that a new polymorph of $NaBi_3V_2O_{10}$ (hereafter $\boldsymbol{a}$NBVO) is formed.

$NaBi_3V_2O_{10}$ was synthesized as per the reported procedure. Single crystals of this phase ($\boldsymbol{a}$NBVO) could not be grown because of the presence of a concomitant minor phase which forms at around 700°C and remains till the melting point (755°C). High resolution X-ray powder data were collected on a STOE/STADI-P X-ray powder diffractometer with germanium monochromated $CuK\boldsymbol{a}$ ($\boldsymbol{l}=1.54056\text{Å}$) radiation from a sealed tube X-ray generator (20 kV, 25 mA) in the transmission mode using a linear PSD ($2\boldsymbol{q}=3$ to $100.42°$ with a step size of $0.02°$ with 6 seconds/step exposure time) at room temperature. The sample was rotated during the data collection to minimize the preferred orientation effect, if any. The program ITO12 in CRYSFIRE package[30] was used to index the powder pattern, which gave a triclinic cell in agreement with the earlier report. The full pattern fitting and peak decomposition in the space group $P\bar{1}$ using the program EXTRA gave $R_p=11.55\%$, $R_{wp}=15.80\%$ for 498 independent reflections. SIRPOW 92 was used to locate the positional parameters of bismuth and

vanadium atoms, which were put in the starting model for preliminary Rietveld refinement using the GSAS program. In order to check the symmetry, the space group $P1$ was assigned and the above procedure was repeated. The heavy atoms located clearly indicated the presence of a centre of symmetry. The R_p and R_{wp} values were not significantly different in the two approaches. At this stage, the occupancy of Bi (1) atom refined to a lower value suggesting that Na atom must be located at the same site satisfying the overall stoichiometry. A subsequent difference Fourier map revealed the positions of the remaining oxygen atoms which were used for the final Rietveld refinement using GSAS.

The profile was fitted using the pseudo-Voigt function. The Chebyshev function consisting of 12 coefficients was used to define the background. Isotropic thermal parameters of all atoms were refined independently at the early stage of refinements. It was observed that the thermal parameter of all oxygens converged to approximately the same values, except that of O(3), which was higher than the rest. Hence the thermal parameter of O(3) was refined separately while that of the rest of the oxygens were constrained together and refined. Since the occupancy of O(3) showed a large deviation from unity it was also refined. The occupancies of Bi (1)/Na were constrained (to a total value of 1·0) while those of the other atoms were fixed.

The observed, difference and calculated patterns are shown in figure 2. The structure of $NaBi_3V_2O_{10}$ is built of $(Bi_2O_2)^{2+}$ chains extended along the *c*-axis with the vanadium tetrahedra acting as linkers between chains (figure 3). The structure depicts, for the first time, features which are not common to the Aurivillius family of compounds. The usual arrangement of BiO_4 units forming $(Bi_2O_2)^{2+}$ sheets in two dimensions in the Aurivillius family is restricted to a one-dimensional chain in $NaBi_3V_2O_{10}$. $(Bi_2O_2)^{2+}$ sheets in the doped Bi_2O_3–V_2O_5 system belonging to the Aurivillius family are known

Figure 2. Observed, calculated and difference X-ray diffraction patterns of $NaBi_3V_2O_{10}$.
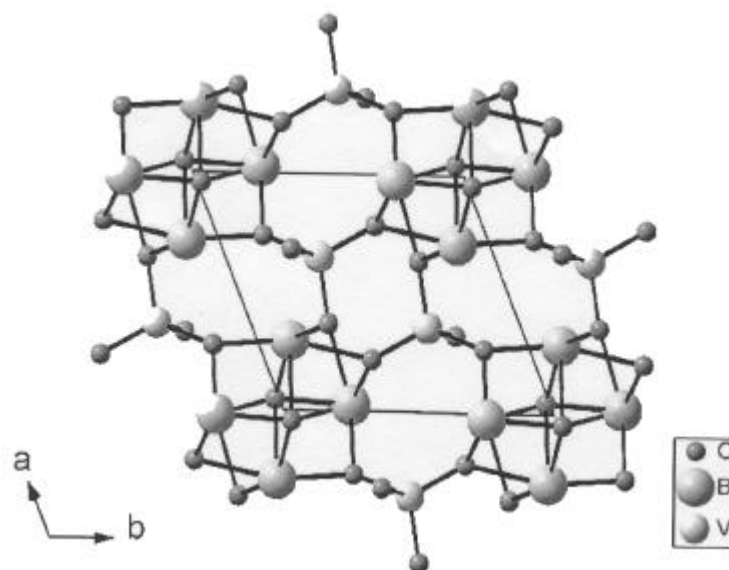
Figure 3.   Structure of $NaBi_3V_2O_{10}$ viewed down the *c* axis.

to produce an undesirable strong anisotropy in ionic conductivity. It may be speculated that the absence of these sheets in this compound might have significant impact on the conductivity behaviour.

8.   Conclusion

Determination of the crystal structure from powder X-ray diffraction data appears to hold promise for the future especially in materials science, and drugs and pharmaceuticals field where the growth of diffraction quality crystals becomes a serious rate limiting step. It may be mentioned that the methodologies described above are applicable even in case of samples having coexisting phases. These methods will be particularly useful in discerning structures which exhibit polymorphism. It is generally observed that the rate at which organic structure is determined via *ab-initio* methods is much lower than that of inorganic counterparts. The number of organic crystal structures deposited at the Cambridge Crystallographic database (CSD) per year exceed by a big margin the number of organic structures solved by powder diffraction. It may be either due to the fact that the growth of single crystals of organic compounds for X-ray diffraction is achieved with a higher rate of success or that inorganic chemists are more at home with powder diffraction.

References

1. Rietveld H M 1967 *Acta Crystallogr.* **22** 151
2. Rietveld H M 1969 *J. Appl. Crystallogr.* **2** 65
3. Giacovazzo C 1996 *Acta Crystallogr.* **A52** 331
4. Harris K D M and Tremayne M 1996 *Chem. Mater.* **8** 2554
5. Young R A 1995 *The Rietveld method* (IUCR, Oxford: University Press) p. 276
6. Poojary D M and Clearfield A 1997 *Acc. Chem. Res.* **30** 414

7. Le Bail A 2001 ALB Crystallography Home Page *http://www.cristal.org*
8. Porob D G and Row T N G 2000 *Chem. Mater.* **12** 3658
9. Tremayne M, Kariuki B M and Harris K D M 1996 *J. Appl. Crystallogr.* **29** 211
10. Andreev Yu G, Lightfoot P and Bruce P G 1997 *J. Appl. Crystallogr.* **30** 294
11. Shankland K, David W I F and Csoka T 1997 *Z. Kristallogr.* **212** 550
12. Werner P E 1964 *Z. Kristallogr.* **120** 375
13. Boultif A and Louer D 1991 *J. Appl. Crystallogr.* **24** 987
14. Visser J W 1969 *J. Appl. Crystallogr.* **2** 89
15. CCP14 Home page *http://www.ccp14.ac.uk*
16. Rudolf P R and Clarfield A 1989 *Inorg. Chem.* **28** 1706
17. Pawley G S 1981 *J. Appl. Crystallogr.* **14** 357
18. Toyara H 1986 *J. Appl. Crystallogr.* **19** 440
19. Le Bail A, Duroy H and Fourquet J L 1988 *Mater. Res. Bull.* **23** 447
20. Larson A C and von Dreele R B 1987 Los Alomos Laboratory Report, No. LA-UR-86-748
21. Sheldrick G K 1997 SHELX-97 Programs for crystal and structure solution and refinement, University of Göttingen, Germany
22. Altomare A, Burla M C, Cascarano G, Giacovazzo C, Guagliardi A, Moliterni A G G and Polidori G 1995 *J. Appl. Crystallogr.* **28** 842
23. Altomare A, Burla M C, Cascarano G, Giacovazzo C, Guagliardi A, Camalli M and Polidori G 1994 *J. Appl. Crystallogr.* **27** 435
24. Rodriguez-Carvajal J 1998 Program Fullprof (version 3.5), Laboratoire Leon Brillouin, France
25. David W I F 1987 *J. Appl. Crystallogr.* **20** 316
26. David W I F 1990 *Nature (London)* **346** 731
27. Gilmore C J, Bricogne G and Bannister C 1990 *Acta Crystallogr.* **A46** 297
28. Sinclair D C, Watson C J, Howie R A, Skakle J M S, Coats A M, Kirk C A, Lachowski E E and Marr J 1998 *J. Mater. Chem.* **8** 281
29. Sinclair D C, Marinou E and Skakle J M S 1999 *J. Mater. Chem.* **9** 2617
30. Shirley R 1999 *The CRYSFIRE system for automatic powder indexing: User's manual* (Lattice Press)