Ab initio crystal structure analysis based on powder diffraction data using PDXL

Akito Sasaki*, Akihiro Himeda*, Hisashi Konaka* and Norihiro Muroyama*

1. Introduction

Physical and chemical properties of a crystalline solid depend strongly on the molecular arrangement, that is, on both the crystal structure and the composition of the molecule comprising the solid. In order to understand the mechanisms and developing properties of a crystalline solid, it is essential to know the crystal structure. Typically, crystal structure analysis has been performed using hundreds or thousands of X-ray intensity data collected from a single crystal. The data is collected with a four-circle diffractometer or a diffractometer equipped with an image plate or other 2D detector. Just 10 years ago, single crystals several hundred microns in diameter were needed. Recent improvements in X-ray sources and detectors enable the collection of intensity data which can be used in the analysis of crystal structures from crystal specimens measuring only a few microns across.

There are many substances which cannot be grown to a single crystal of quality and size sufficient for single crystal diffraction measurements. Inorganic compounds have often had their crystal structures analyzed using the Rietveld method[1] devised in 1969. Fundamentally, the Rietveld method is used to refine crystal structure parameters such as lattice constants, atomic coordinates, occupancies, temperature factors, etc. based on powder diffraction data. There are many groups of inorganic compounds which have almost identical composition and crystal structure. In these cases, the crystal structure can usually be solved using the Rietveld method using the structure parameters of an analogous compound as the initial structural model.

On the contrary, since organic compounds are formed molecular crystals, their crystal structures are affected by even very small changes in composition. This makes it exceptionally difficult to perform *ab initio* crystal structure analysis of organic compounds using only the Rietveld method.

Then, how can we determine the *ab initio* crystal structure of organic compounds from powder diffraction data? As shown in Fig. 1, there are several steps in the analysis procedure. While intensity data of independent diffraction spots are collected from single crystal diffraction measurements, with powder diffraction measurements, the diffraction spots, 3-dimensionally arranged in the reciprocal space, are compressed to 1-dimensional diffraction patterns. As a result, the number of intensity readings from independent diffraction spots decreases. For this reason, crystal structure analysis based on powder diffraction data collected with the methods used for single crystal diffraction data is often unsuccessful.

This paper describes the analytical process used for the *ab initio* crystal structure determination of, in particular, organic compounds based on powder diffraction data. It also introduces *PDXL*[2], the integrated X-ray powder diffraction software package in which all these features are implemented.

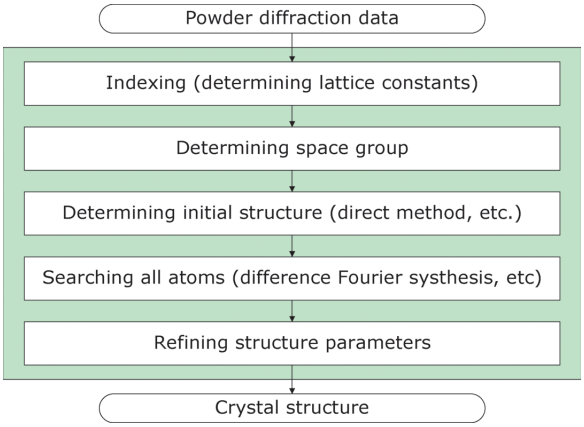


Fig. 1. *Ab initio* crystal structure analysis workflow.

2. Individual steps

2.1. Indexing

"Indexing" refers to the process of assigning Miller indices to each diffraction peak as well as the determination of lattice constants. ITO, DICVOL, TREOR are very popular indexing programs, and recent versions of each of these programs are integrated into the PDXL package (ITO13[3], DICVOL06[4], and N-TREOR[5]). Each of these programs searches for the unit cell based on about 10 to 30 peak positions picked up from the low angles of the powder diffraction data. As described above, the 3-dimensional diffraction spots are compressed into a 1-dimensional diffraction pattern, which makes it very hard to obtain a correct and unique set of lattice constants. Regardless of which program is used, several candidates for lattice constants will be listed along with the degree of reliability for each. From among the candidates, you have to select a set of lattice

* Application & Software Development Department, Rigaku Corporation.

constants which can generate all the peak positions in higher angles. Needless to say, the more accurately the peak positions used for indexing, the higher the probability correct lattice constants can be found. On this basis it is strongly recommended that powder diffraction data are collected using optics which provide the highest-possible resolution. It is also recommended to use a very high flux X-ray source to detect all low-angle peaks even if the peak intensities are low. If indexing fails and correct lattice constants cannot be found, the correct crystal structure cannot be obtained. Scrupulous attention should be given to the correct determination of lattice constants.

2.2. Space group determination

Space group is determined based on the systematic absences obtained from Miller indices vs. diffraction intensity data. Since only Miller indices are assigned to each peak at the indexing step, the intensity of overlapping diffraction peaks need to be decomposed and distributed to each diffraction line. PDXL performs intensity decomposition using Pawley's method[6]. After decomposition, which distributes diffraction intensity to each diffraction line, PDXL determines space groups. Like indexing, space groups typically cannot be determined uniquely, that is, several space group candidates are listed with the degree of reliability for each. PDXL expresses the reliability of a space group candidate in terms of N and R. N is the number of diffraction lines calculated from the space group and the R value is the degree of coincidence between the calculated N and observed diffraction lines in measurement data. The smaller the N and R values are, the higher the probability that the space group is correct. Figure 2 shows the space group candidates calculated from the powder diffraction data of Al_2O_3. A candidate whose N and R values are both small should be selected from among these as described above. In some cases, taking the small N candidate rather than the small R candidate will result in a higher probability of obtaining the correct space group. In the case of Fig. 2, the correct space group is R-3c belonging to R**c with R=0% and N=53. In this way, the user selects one space group thought to be correct from among candidates and goes on to the next step.

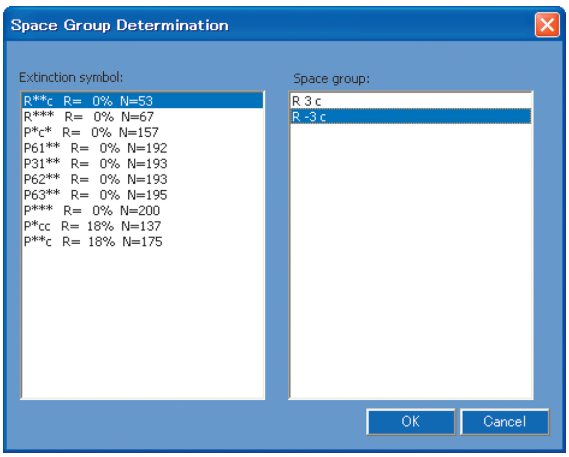


Fig. 2. Space group candidates.

2.3. Initial structure determination

Information on diffraction spots is compressed in powder diffraction data as described above, therefore, some innovation has been introduced to the initial structure determination process. Three methods implemented in PDXL for determining initial structures are introduced below. Since either method requires as precise diffraction intensity for each set of Miller indices as possible, pattern decomposition has to be performed after the space group determination and before the initial structure determination.

2.3.1. Direct method

In the direct method, phases are predicted based on the tangent formula[7] and the initial structure is determined using sets of Miller indices vs. diffraction intensity data. Although the data obtained from powder diffraction is limited, there are some cases in which the initial structure can successfully be determined by the direct method. PDXL allows the user to make use of the direct method with EXPO2009[8] developed by Giacovazzo et al. part of the group that developed SIR[9], the direct method program for single crystal structure analysis. Their direct method programs have become a tried and trusted part of the analysis process for many users. Many sets of Miller indices vs. diffraction intensity data are necessary to achieve the initial structure determination using the direct method. Therefore, powder diffraction data at high angles and with as high intensity as possible are required.

2.3.2. Charge flipping

Electron density distribution is calculated by the Fourier synthesis of Miller indices vs. diffraction intensity data. Since electron density peaks in an electron density distribution map indicate atomic positions, an initial structure can be determined from the electron density distribution map. However, the Fourier synthesis cannot be performed without phases because information on phases is missing in diffraction data.

Recently, Oszlányi et al. have developed a new method called "charge flipping[10] (CF)". In this method, (1) Diffraction lines are assigned randomly determined phases. (2) Fourier synthesis is performed on measurement data to obtain an electron density distribution map. (3) The signs of charges below a certain threshold are flipped. (4) A Fourier transform is performed on the modified charge density map to obtain new sets of intensity data. (5) Phases of measured intensity data are updated based on the phases of the calculated intensity data. (6) Steps (2) through (5) are repeated. Advantages of the CF method include the relatively short amount of time needed to calculate and improve the electron density distribution map and the ability to start without any knowledge on the structure at all (Fig. 3).

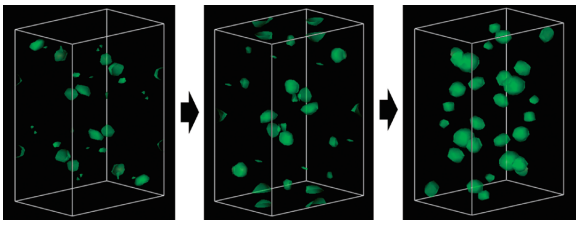


Fig. 3. Transition diagram of electron density map by CF.

2.3.3. Direct space method

When molecular crystals of organic compounds are the main targets of structure analysis, there is a method in which crystal structure can be obtained by constructing a structural model and searching for the position and orientation of the molecular model in the unit cell. This is called the "direct space method[11]", or "real space method". In the case of organic molecules, a molecular orbital calculation program such as MOPAC[12] can calculate the stable conformation of a molecule. The obtained molecular structural model is set as an initial state for the direct space method, and the position and possible orientations of the molecule within the unit cell are searched for. At the same time, rotation (torsion) angles around free rotation bonds are optimized to determine the molecular conformation while the other rigid parts of the molecule are fixed. The method is called "global optimization method". When the initial structural model is somewhat different from the real structure, the global optimization methods can give the real solution (almost global minimum), whereas the Rietveld method may give a false solution (local minimum). While the genetic algorithm[11] and simulated annealing[13] are popular as a global optimization methods, PDXL adapts the parallel tempering algorithm[14].

The direct space method is effective when the structural formula of a molecule is given and the molecular conformation is known to some extent. In the case of organic compounds, the structural formula can be given by spectroscopic methods such as NMR, IR, etc. The theoretically-stable conformation of a molecule can be calculated using molecular orbital or dynamics calculation. Therefore, the direct space method is very powerful for determining the initial crystal structure of organic compounds.

2.4. Structure refinement

After determining the initial structure, the structure is refined using the Rietveld method. In the *ab initio* crystal structure analysis, while lattice constants and the preferred orientation function are normally fixed, structure parameters such as atomic coordinates, temperature factors, etc. are generally refined. In the case of organic compounds, the molecular structure may break apart, which means that, if atomic coordinates are refined individually, bond lengths and angles will deviate from their theoretical values. To avoid this, PDXL has a feature to express atomic coordinates in Z-matrix so that crystal structure can be refined in a manner such that the structure is retained to some degree. Since atomic coordinates expressed in Z-matrix are constrained by bond lengths and angles, the structure of an organic molecule comprised of light atoms can be refined without disassembling. In addition, using PDXL the refinement process can be followed by the crystal structure transitions as well as the numerical values, that is, the degree of coincidence between measured and calculated data.

2.5. Searching missing atoms

In both powder and single crystal structure analysis, it is often the case that some atoms will be missing from the initial structure determination. In these cases, the remaining atoms may be found using an electron density distribution analysis method such as MEM[15],[16], difference Fourier synthesis etc. The electron density distribution map will include the atoms whose positions have already been determined. If there are electron density peaks in the distribution map which have not been assigned atoms, it can be determined that the peaks are the atoms which were not found at the initial structure determination step. Using difference Fourier synthesis, which performs a Fourier synthesis on the measured data from which the structure factors based on the known atoms have been subtracted, an electron density distribution map can be created that shows only the unassigned areas. The peak positions in the resulting map show the residual atom positions. Whenever either method is applied, it is strongly recommended that pattern decomposition be performed again before electron density distribution analysis.

It is often the case that very light atoms like hydrogen cannot be found. In the case of organic compounds, the positions of hydrogen atoms, in particular, will be determined through calculation. When all the atoms have been found, including hydrogen, the crystal structure will be refined using the Rietveld method before it is considered complete.

2.6. Validation of the crystal structure

The obtained crystal structure has to been evaluated to determine whether it is appropriate as an actual representation of the crystal structure. There are several indicative points of validation, as follows:

(1) *R*-factor is small, *S* nearly equals one.

Needless to say, these are very important. But please note that these only give an indication that the assumed model (crystal structure) explains the measured data. In general, use $R_{wp}<10\%$ and $S<2$ as measuring sticks for successful analysis.

(2) Obtained molecular structures are reasonable.

Confirm the obtained molecular structures are reasonable based on the bond lengths and angles.

(3) Density calculated from the crystal structure equals observed density.

This is extremely important. If the calculated density differs from the observed, it is a strong indication that

elements or the number of atoms in the unit cell is wrong.

(4) Interatomic distances are not too short.

Confirm that any non-bonded interatomic distance is not much shorter than the sum of Van der Waals radius of each atom.

(5) There is no big void.

Draw the crystal structure as a space-filling model, in which atom sizes are expressed by the Van der Waals radius, and confirm there are not any big voids between molecules. If there are, confirm whether it is possible that another molecule such as a solvent molecule may exist in the void using a thermal analysis method.

(6) Intermolecular forces are appropriate.

If in the comprising molecules there exist any groups like the O–H, N–H, C=O groups which will form hydrogen bonds, confirm that intermolecular hydrogen bonds are formed properly, also taking angles and distances into consideration. And if there exist any other functional groups with large polarization, confirm that the molecules are arranged such that the dipole moment is cancelled.

PDXL includes various drawing tools to create graphical representations of obtained molecular structures (the position and orientation of the molecule(s) in the asymmetric unit of unit cell). You can select the display style of a molecule, display molecular packing, expand/collapse the molecular/crystal structure, display interatomic distances, bond angles, torsion angles, change the perspective, and so forth. Making use of these functionalities, you can confirm the above points one by one, then you can easily validate whether the obtained crystal structure is correct. As a result of validation, if there are some uncertainties remaining, trace back through the structure analysis steps. The analysis can be re-performed from the indexing step if necessary.

3. Structure analysis examples

Figure 4 summarizes the *ab initio* structure analysis flow based on powder diffraction data using PDXL.

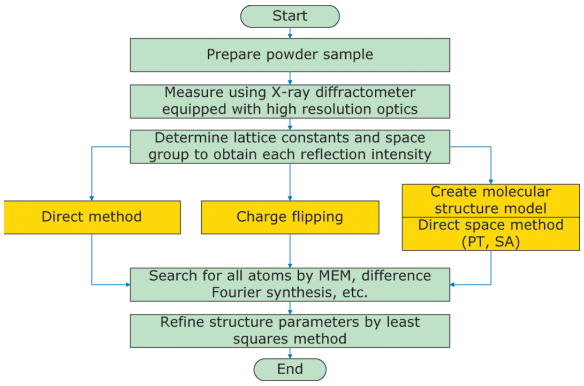


Fig. 4. *Ab initio* powder crystal structure analysis flow.

This chapter describes two analysis examples: one using the direct method, the other the direct space method.

3.1. Cimetidine

Cimetidine is well known as a medicine for the treatment of gastric ulcers. A cimetidine molecule is comprised of 17 non-hydrogen atoms, and indicative of very high crystallinity. Here, *ab initio* structure analysis was attempted using the direct method with EXPO2009. To make the direct method successful, pattern decomposition needs to be performed with high accuracy. Owing to the 2-bounce Ge(220) crystals in the incident optics and CALSA[17] in the receiving optics, it was possible to collect high-intensity and high-resolution data (Fig. 5), which was then used for the structure analysis.

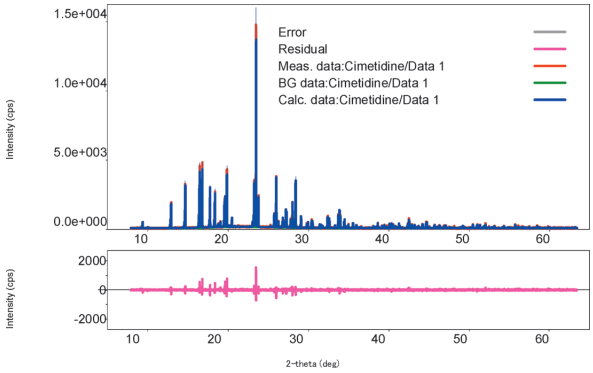
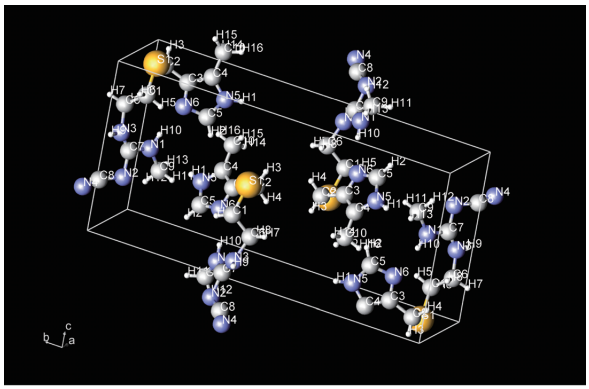


Fig. 5. Powder diffraction pattern of cimetidine.



Space group	$P2_1/c$	
Lattice constants	a=6.82828(7) Å	α=90°
	b=18.82828(19) Å	β=106.4437(4)°
	c=10.39992(11) Å	γ=90°
	R=12.60%、S=1.7645	

Fig. 6. Crystal structure of cimetidine.

3.2. Nicotinamide

Nicotinamide is also known as vitamin B3. The molecular structure is shown in Fig. 7. The molecular structure optimized by MOPAC was used as the initial structural model for the direct space method. The initial crystal structure was obtained giving freedom of rotation to the carbon-carbon bond between the benzene ring and the amide group. In the crystalline environment, four types of structure are possible, depending on the orientations of the amide group and the pyridine ring.

The structure shown in Fig. 7(b) was determined by the direct space method of PDXL. However, based on the schematic presentation of molecular packing, for the structures shown in Fig. 7(b), (c), and (d) it was determined that there existed steric hindrance between the hydrogen atoms and/or there were not any intermolecular hydrogen bonds supposed to exist. As a result, we concluded that the most probable structure was (a) (Fig. 8).

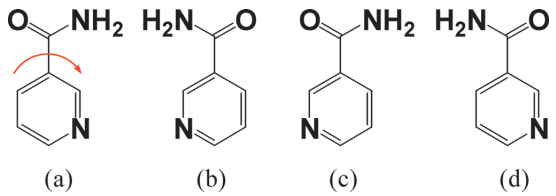
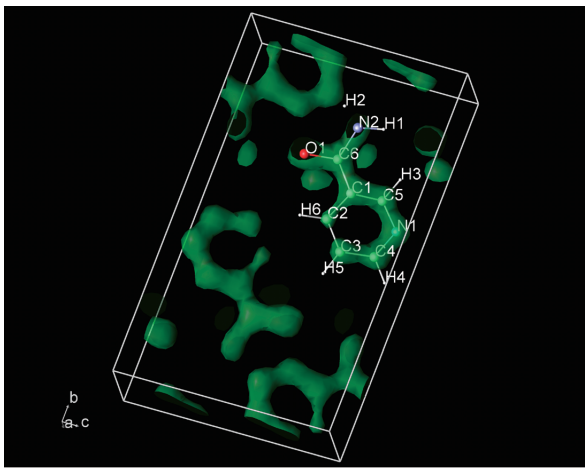


Fig. 7. Possible molecular structure of nicotinamide.



Space group	$P2_1/c$	
Lattice constants	a=3.9817(6) Å	α=90°
	b=15.649(2) Å	β=99.011(2)°
	c=9.4400(10) Å	γ=90°
	R=9.05%、S=1.9563	

Fig. 8. Crystal structure of nicotinamide.

4. Summary

Recent years have seen a rapid increase in the number of people who utilize *ab initio* crystal structure analysis based on powder diffraction data. However, unlike for single crystal structure analysis, there is no standard procedure for powder structure analysis. The user has to perform the analysis through a process of trial and error. Therefore, an easy-to-use tool for the powder structure analysis has been eagerly anticipated. PDXL provides for this need by offering multiple algorithms for indexing and the initial structure determination, simplifying the normally complex *ab initio* crystal structure analysis steps. Preparing multiple analysis paths, as described above, is strongly associated with a high probability of the success in crystal structure analysis.

We hope that the application of this structure analysis process will lead to the crystal structures of various powder materials being solved, resulting in significant advances in structural and properties sciences.

References

(1) H. M. Rietveld: *J. Appl. Cryst.*, **2** (1969), 65–71.
(2) *The Rigaku Journal,* **26** (2010), No. 1, 23–27.
(3) J. W. Visser: *J. Appl. Cryst.*, **2** (1969), 89–95.
(4) A. Boultif and D. Louer: *J. Appl. Cryst.*, **37** (2004), 724–731.
(5) A. Altomare, C. Giacovazzo, A. Guagliardi, A. G. G. Moliterni, R. Rizzi and P.-E. Werner: *J. Appl. Cryst.*, **33** (2000), 1180–1186.
(6) G. S. Pawley: *J. Appl. Cryst.*, **14** (1981), 357–361.
(7) J. Karle and I. Karle: *Acta Cryst.,* **21** (1966), 849–859.
(8) A. Altomare, M. Camalli, C. Cuocci, C. Giacovazzo, A. Moliterni and R. Rizzi: *J. Appl. Cryst.*, **42** (2009), 1197–1202.
(9) M. C. Burla, M. Camalli, G. Cascarano, C. Giacovazzo, G. Polidori, R. Spagna and D. Viterbo: *J. Appl. Cryst.*, **22** (1989), 389–393.
(10) G. Oszlányi and A. Sütő: *Acta Cryst.,* **A60** (2004), 34–141.
(11) K. D. M. Harris, R. L. Johnston and B. M. Kariuki: *Acta. Cryst.,* **A54** (1998), 632–645.
(12) http://software.fujitsu.com/jp/scigress/mocompact/
(13) Y. G. Andreev, G. S. MacGlashan and P. G. Bruce: *Phys. Rev.*, **B55** (1997), 12011–12017.
(14) V. Favre-Nicolin and R. Černý: *J. Appl. Cryst.*, **35** (2002), 734–743.
(15) D. M. Collins: *Nature*, **298** (1982), 49–51.
(16) M. Sakata, R. Mori, S. Kumazawa, M.Takata and H. Toraya: *J. Appl. Cryst.*, **23** (1980), 526–534.
(17) H. Toraya: *J. Appl. Cryst.*, **42** (2009), 485–489.