Powder *Shake-and-Bake* Method

H. Xu[*], Ch. M. Weeks, R. H. Blessing

Hauptman-Woodward Medical Research Institute and Department of Structural Biology,
State University of New York at Buffalo, 700 Ellicott Street, Buffalo, NY 14203, USA
[*] xu@hwi.buffalo.edu

Keywords: powder diffraction, direct methods, *Shake-and-Bake*, structure determination

Abstract. The dual-space *Shake-and-Bake* procedure, one of the most successful direct methods for phasing single-crystal diffraction data, has been adapted and modified for structure determination from microcrystalline powder samples. The new method, termed *Powder Shake-and-Bake* and implemented in the computer program *PowSnB*, incorporates *extracted* and *overlapped* reflections in the *Shake-and-Bake* procedure. In each *SnB* cycle, it performs (i) an overlapped-reflection partition (*via* partial structural information from the previous cycle), (ii) a reciprocal-space phase refinement (*via* the reduction of the values of a statistical minimal function), and (iii) a real-space density modification (*via* peak picking). The successful *PowSnB* applications with experimental powder diffraction data are described.

Introduction

The collapse of the three dimensions of reciprocal space to the single dimension of a powder diffraction pattern, with the resultant Bragg-peak overlap becoming particularly severe at shorter *d*-spacings, results in a very substantial loss of information. When powder patterns exhibit clear discrimination between Bragg peaks and background as well as clear resolution of Bragg peaks from one another, individual structure-factor modulus can be reliably obtained by integrated intensity extraction methods such as the Le Bail method [1] or the Pawley method [2]. The extracted reflections are treated as *non-overlapped* or `well-determined' reflections. When powder diffraction patterns do not exhibit clear peak-background discrimination and peak-peak resolution, individual intensities obtained from intensity extraction methods must be considered unreliable, and are treated as *overlapped* reflections. The overlap problem highlights most clearly the loss of information in the powder diffraction measurement and limits the applicability of powder diffraction as a routine tool for determining and refining crystal structures from microcrystalline powders.

Adaptations of traditional direct methods, specifically tailored to deduce the phase-angle information from the non-overlapped reflection intensities, have been particularly encouraging in recent years. The best known direct-methods package is *EXPO2004* [3], which has evolved continually and now incorporates indexing, space-group determination, structure solution and refinement capabilities [4]. Recently, the charge flipping method [5] has been adapted to powder diffraction data with some very promising results [6]. Despite these important advances, the overlap problem is not, in general, solved. As a consequence, *ab initio* direct-methods applications to powder diffraction data have limited success in terms of structure size and complexity.

Direct methods work best when large numbers of well-determined reflection intensities have been measured to atomic resolution (better than 1.2Å). However, the collapse of the three dimensions of reciprocal space to the single dimension of a powder diffraction results in a very substantial loss of information. For moderately sized crystal structures, even with the highest-resolution powder diffractometers, it is almost impossible to obtain individual integrated intensities at atomic resolution. In this paper, we propose novel constrained global optimization models to tackle simultaneously the overlap problem and the crystallographic phase-angle problem as well as to develop efficient computational algorithms to find the constrained global minimum of the objective function for both modulus extraction and phase-angle estimation.

Powder *Shake-and-Bake* method

The *Shake-and-Bake* algorithm [7], specifically designed for finding the constrained global minimum from single-crystal diffraction data, has been adapted and modified for structure determination from powder diffraction data. The new algorithm, termed Powder *Shake-and-Bake* and shown by the flow diagram in figure 1, combines overlapped-reflection partitioning, minimal-function phase refinement and real-space filtering. It is an iterative process that is repeated until a solution is achieved or a designated number of trial structures have been processed. For present purposes, 'solution' does not necessarily imply a complete structure but rather a set of atomic positions that can be readily refined and extended by the standard Rietveld [8] refinement method. With reference to figure 1, the detailed *Shake-and-Bake* algorithm (Steps 1-3, 5-7) can be found in many publications (*e.g.*, [7]), and only the major differences between *Shake-and-Bake* and *Powder Shake-and-Bake* (steps 1 & 4) are described next.

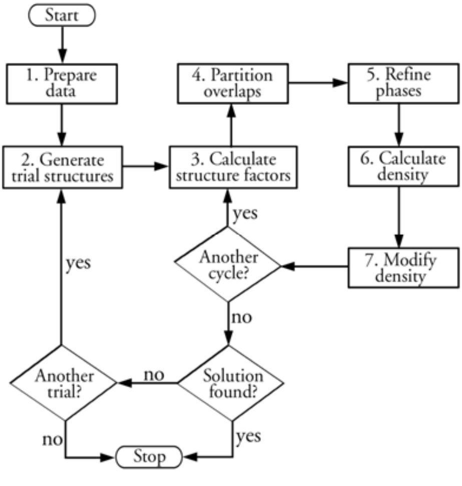

Figure 1. Powder Shake-and-Bake algorithm.

Data preparation. Reflections are divided into a non-overlapped set G_0 and overlapped sets G_j ($j = 1, 2, \ldots, M$) containing L_j reflections with overlap constraints

$$\left|F_{G_j}\right|^2 = \sum_{l=1}^{L_j} m_l \left|F_{\mathbf{H}_l}\right|^2 = \sum_k \left[obs(k) - back(k)\right], \tag{1}$$

where m_l and $\left|F_{\mathbf{H}_l}\right|$ are the l^{th} reflection multiplicity and structure-factor modulus, respectively, and $obs(k) - back(k)$ is the background-subtracted peak contribution at the k^{th} point in the diffraction pattern. For reflections $\mathbf{H} \in G_0$, the structure-factor moduli $|F_\mathbf{H}|$ are obtained directly from the extracted integrated intensities. For reflections $\mathbf{H} \in G_j$, the initial values of the structure-factor moduli are obtained from the overlap-constraint equation (1) with equipartitioning, *i.e.*, $|F_\mathbf{H}| = \langle |F_\mathbf{H}| \rangle_{\mathbf{H} \in G_j}$. Normalized structure-factor moduli ($|E|$'s) are generated from the structure-factor moduli ($|F|$'s) by standard scaling methods such as a Wilson plot. Reflections are then sorted in descending order of $|E|$ values, and a predetermined number of the top reflections is selected to generate a predetermined number of structure invariants $\Phi_{\mathbf{HK}} = \varphi_\mathbf{H} + \varphi_\mathbf{K} + \varphi_{-\mathbf{H}-\mathbf{K}}$ having the largest values of the modulus factors $A_{\mathbf{HK}} = 2N^{-1/2} |E_\mathbf{H} E_\mathbf{K} E_{\mathbf{H}+\mathbf{K}}|$.

Overlapped-reflection partition. Let the sum

$$\sum_{l=1}^{L_j} m_l |E_{\mathbf{H}_l}|^2 = |E_{G_j}|^2 \tag{2}$$

be the j^{th} overlap constraint, and the quantities $|E_{\mathbf{H}_l}^{cal}|$ $(1 \le l \le L_j)$ be the individual normalized structure-factor modulus calculated using the trial structure from the previous cycle. Then a new partition for the overlapped reflections is given by

$$|E_{\mathbf{H}_l}| = \left(\sum_{l=1}^{L_j} m_l |E_{\mathbf{H}_l}^{cal}|^2 \right)^{-1/2} |E_{\mathbf{H}_l}^{cal}| |E_{G_j}|. \tag{3}$$

The newly partitioned reflections are used to generate a new trial structure (*via* steps 5-7). This iterative re-partitioning of the overlapped reflections is the distinctive feature of the *Powder Shake-and-Bake* algorithm.

Applications

The traditional and powder *Shake-and-Bake* procedures have been tested successfully using experimentally measured atomic-resolution powder diffraction data. Here, we report solution efficacy measured by the success rate (percentage of trial structures that converge to solutions) as well as solution quality measured by (i) the mean phase error (MPE) or average absolute value of the deviations of the phase angles from their values calculated using the final refined coordinates and thermal parameters; (ii) the atom-match rate (AMR) or percentage of atoms whose coordinates are within a 0.5Å radius of the final refined coordinates; and (iii) the average inter-atomic distance (DIST) in pairs of matched atoms. Unless clearly stated otherwise, the results reported here are based on 1000 trial structures and produced by computer software *SnB* [9] and *PowSnB* (traditional and powder *Shake-and-Bake* procedures, respectively) implementing constrained minimization of the statistical minimal function [10, 11].

Overlap-constraint Protocol P(M). In *PowSnB* experiment, the extracted intensities are divided into non-overlapped group ($d = \sin\theta/\lambda > d_0$, where d_0 is a resolution parameter) and overlapped group ($d_{max} < d < d_0$, where d_{max} is the highest data resolution). The extracted intensities in the overlapped group are sorted in decreasing order of the |E| values and grouped into subgroups containing *M* reflections. For each subgroup, the average value of the normalized structure factors, ⟨|E|⟩, is calculated and assigned to each reflection in the subgroup. The new set of structure factors, which consists of non-overlapped and overlapped reflections having largest initial |E| values, is subsequently input to *PowSnB* to test the validity, accuracy and robustness of the *Powder Shake-and-Bake* procedure.

Cimetidine application

The crystal structure of the drug molecule cimetidine [12] ($C_{10}H_{16}N_6S$) is monoclinic, has space group $P2_1/n$, and Z=4. The high-resolution, synchrotron X-ray powder pattern was collected on station 8.3 at Daresbury Laboratory with a wavelength of $\lambda = 1.4599$Å over a 2θ range of 6 to 105°. The integrated intensities were extracted from the indexed profile to a resolution of $d_{max} = 1.13$Å using a modified form of the Pawley pattern decomposition program [2] implemented in the computer software *GSAS* [13]. The test results of *SnB*, as well as the results from *PowSnB* with overlap-constraint protocol P(4) and a series of resolution parametr $d_0 = 1.2, 1.3, \ldots, 3.0$Å, are reported in table 1. These results demonstrate the following:

- *SnB* effectively yields high-quality solutions as indicated by 17.0% success rates, low mean phase errors, high average percentage of matched atoms, and small average distances between matched atoms.
- *PowSnB* yields an average of 26.2% success rate using overlap-constraint protocol P(4) for a series of resolution parameter $d_0 = 1.2, 1.3, \ldots, 3.0$Å, which ranges from 16.1% at $d_0 = 1.3$Å to 42.0% at $d_0 = 2.3$Å. The quality of solutions is similar to that from *SnB*.
- The minimal function histogram is no longer a bimodal distribution for *PowSnB* as commonly observed in single-crystal applications [9]. This is shown in table 1 by the overlapping figure-of-merit (FOM) ranges for solutions and non-solutions. The unimodal histogram is the result of unavoidable errors in the structure-factor moduli that become more severe as the resolution increases.

Table 1. Comparison of SnB and PowSnB runs for cimetidine powder diffraction data. † *Percentage of trial structures having mean phase error (MPE) < 25°.* § Overlap-constraint protocol P(4) and *resolution parameter $d_0 = 2.3$Å. * Values of the statistical minimal function.*

Phasing Method	Success Rate†	Figure-of-merit Ranges*		Solution Averages		
		soln	non-solns	⟨MPE⟩	⟨AMR⟩	⟨DIST⟩
SnB	17.0%	0.124-0.141	0.161-0.355	4.2°	82%	0.18Å
PowSnB§	42.0%	0.063-0.089	0.072-0.273	9.7°	81%	0.20Å

For unknown structures, in the absence of bimodal distributions to distinguish solutions from non-solutions, we propose an atom-match-based FOM to identify Pow*SnB* structure solutions:

- The distances between pairs of atoms from two *PowSnB* trial structures can be calculated after taking the possibility of different origins and enantiomorphs into account [14]. Two atoms are considered matched if their distance is less than 0.5Å.
- The percentage of matched atoms (> 70% for solutions and < 50% for non-solutions) is used as a figure of merit to identify all solution structures from non-solution structures.

Solutions for cimetidine identified by the atom-match-based FOM were perfectly correlated with solutions identified by low mean phase errors (< 25°). Unless clearly stated otherwise, the atom-match-based FOM was used to identify potential solutions discussed later in this paper.

Fluorescein diacetate **application**

Fluorescein diacetate [15] ($C_{24}H_{16}O_7$) crystallizes in the triclinic space group P-1 with two molecules in the unit cell. The high-resolution, synchrotron X-ray powder pattern was measured at the Swiss-Norwegian beam line (BM1) at the European Synchrotron Radiation Facility (ESRF) in Grenoble with a wavelength $\lambda = 0.6006$Å over the 2θ range 2-50°. A total of 2905 individual values of $|F|^2$ were extracted from the powder data to $d_{max} = 0.88$Å using the Le Bail pattern decomposition method [1] implemented in the computer software *EXTRA* [16]. The results of *SnB* applications, reported in the first block of table 2, demonstrate that:

- *SnB* yielded high quality solutions with fragments of the molecule comprising 79-87% of the atoms.
- The structure quality yielded at 1.0Å is better than that at 0.88Å as indicated by the higher atom-match rate and the shorter distances between matched atoms. This indicates large errors in the extracted intensities from the high-angle part of the powder diffraction pattern.
- *SnB* failed to yield any structure solution at a resolution of 1.2Å.

Table 2. Comparative results for the crystal structure of fluorescein diacetate.

Using Overlap	Overlap Protocol	Data Resolution	Success Rate	Solution Averages ⟨MPE⟩	⟨AMR⟩	⟨DIST⟩
No (*SnB*)	N/A	0.88Å	2.6%	34.2°	79%	0.21Å
		1.0	2.4	24.5	87	0.18
		1.2	0.0	N/A	N/A	N/A
Yes (*PowSnB*)	P(4)	1.2→1.0Å	3.1	22.7	88	0.18
		1.3→1.0Å	3.2	21.6	89	0.17
		1.4→1.0Å	3.3	23.6	87	0.17

The results of *PowSnB* applications, reported in the second block of table 2, demonstrate that:

- *PowSnB* is a very efficient program for producing high-quality structure solutions from powder diffraction data, as indicated by low mean phase errors (<25°), high atom-match rates (≥87%), and short matching distances (≤0.18Å).
- Iterative re-partitioning of overlapped reflections is the key to successful *PowSnB* applications to powder diffraction data. This is the case when individual reflection intensities can be extracted at lower-than-atomic resolution, provided that signal-to-noise in the powder pattern is sufficient to allow overlap constraints to be reliably obtained up to atomic resolution.
- The change of resolution parameter $d_0 = 1.2$, 1.3 and 1.4Å has little impact on success rates or the quality of *PowSnB* solutions as evidenced by only small changes in all figure-of-merit values.

Concluding remarks

Powder Shake-and-Bake is an *ab initio* direct-phasing method for structure determination from powder diffraction data. It is aimed to improve structure solutions from subatomic resolution data or atomic resolution data with large errors in extracted intensities from powder profile. *PowSnB*, the computer program implementing the *Powder Shake-and-Bake* algorithm, is capable of providing high-quality structure models for further structure refinement.

References

1. Le Bail, A., Duroy, H. & Fourquet, J.L., 1988, *Mat. Res. Bull.,* **23**, 447.
2. Pawley, G.S., 1981, *J. Appl. Cryst*., **14**, 357.
3. Altomare, A., Caliandro, R., Camalli, M., Cuocci, C., Giacovazzo, C., Moliterni, A.G.G. & Rizzi, R., 2004, *J. Appl. Cryst*., **37**, 1025.
4. Altomare, A., Cuocci, C., Giacovazzo, C., Moliterni, A.G.G. & Rizzi, R., 2006, *J. Appl. Cryst*., **39**, 145.
5. Oszlányi, G. & Sütö, A., 2008, *Acta Cryst*., **A64**, 123.
6. Baerlocher, C., Gramm, F., Massuger, L., McCusker, L.B., He, Z.B., Hovmoller, S. & Zou, X.D., 2007, *Science*, **315**, 1113.
7. Weeks, C.M., DeTitta, G.T., Hauptman, H.A., Thuman, P. & Miller, R., 1994, *Acta Cryst*., **A50**, 210.
8. Rietveld, H.M., 1969, *J. Appl. Cryst.*, **2**, 65.
9. Miller, R., Gallo, S.M., Khalak, H.G. & Weeks, C.M., 1994, *J. Appl. Cryst*., **27**, 613.
10. Xu, H. & Hauptman, H.A., 2004, *Acta Cryst*., **A60**, 153.
11. Xu, H., Weeks, C.M. & Hauptman, H.A., 2005, *Acta Cryst*., **D61**, 976.
12. Cernik, R.J., Cheetham, A.K., Prout, C.K., Watkin, D.J., Wilkinson, A.P. & Willis, B.T.M., 1991, *J. Appl. Cryst*., **24**, 222.
13. Larson, A.C. & Von Dreele, R.B., 2004, *Technical Report LAUR 86-748*, Los Alamos National Laboratory, Los Alamos, NM.
14. Smith, G.D., 2002, *J. Appl. Cryst*., **35**, 368.
15. Knudsen, K.D., Pattison, P., Fitch, A.N. & Cernik, R.J., 1998, *Angew. Chem. Int. Ed*., **37**, 2340.
16. Altomare, A., Burla, M.C., Cascarano, G., Giacovazzo, C., Guagliardi, A., Moliterni, A.G.G. & Polidori, G., 1995, *J. Appl. Cryst*., **28**, 842.

Acknowledgements. The authors would like to thank Robert Von Dreele (ANL) and Kenneth Knudsen (Institute for Energy Technology, Norway) for the testing data. This research was financially supported by NIH grant GM-72023.