



Direct space approach in action

Corrado Cuocci

Institute of Crystallography, Via G. Amendola 122/o, 70126 Bari, Italy.; corrado.cuocci@ic.cnr.it

Introduction

In the direct space (DS) methods the observed powder pattern is compared with the calculated pattern corresponding to a model of the crystal structure under studied. The model is described in terms of combination of scattering objects that are molecular fragments *e.g.*, isolated atoms, molecules, coordination polyhedra. For each molecular fragment, the position may be defined by the fractional coordinates (x, y, z) of the center-of-mass or a predefined pivot atom, the orientation may be defined by rotation angles (θ, φ, χ) around a set of orthogonal axes, and the intramolecular geometry may be specified by a set of n torsion angles $\{\tau_1, \tau_2, \dots, \tau_n\}$. These concepts may be extended to the case of two or more (identical or nonidentical) molecular fragments within the asymmetric unit (a.u.). Position, orientation and torsion angles are variable and they are optimized in such a way that the calculated pattern best fits the observed pattern. In general, bond lengths, bond angles and ring conformation are not considered as variables during the optimization procedure, but they should match as closely as possible that of the studied compound. The interested reader is referred to various reviews for details about this topic (Harris *et al.*, 2001; Florence *et al.*, 2005; Černý & Favre-Nicolin, 2007; Shankland *et al.*, 2010; Shankland *et al.*, 2013).

Building Molecules

One of the drawbacks of global optimization methods is to input an accurate 3D structural model of the a.u. Starting model for DS method can be prepared by the following two strategies, or by a combination of both approaches.

1) The user is strongly advised to check for similar molecules in databases of crystal structures, *e.g.*, Cambridge Structural Database (CSD) (organics & organometallics), Inorganic Crystal Structure Database (ICSD) (inorganics, elements, minerals & intermetallics), Crystallography Open Database (COD). If a new compound is being studied, one can frequently find significant molecular parts of the structure in existing crystal structures. Molecular fragment found in database can be modified and optimized by quantum-chemistry package to finally generate the desired structure. Alternatively, some open chemistry databases where to look up molecules are: NIST Chemistry WebBook (<http://webbook.nist.gov/chemistry/>), PubChem (<https://pubchem.ncbi.nlm.nih.gov/>), Drugbank (<http://www.drugbank.ca/>). Calculated 3D molecules in sdf, pdb, mol format can be found in these databases and imported. Computed Simplified Molecular-Input Line-Entry System (SMILES), when available, can be converted in 3D molecule by software.

2) Starting molecular model can be created by geometry optimization using quantum-chemistry package, *e.g.*, MOPAC (Stewart, 2016), Gaussian (Frisch *et al.*, 2009), Gamess (Schmidt *et al.*, 1993), NWChem (Valiev *et al.*, 2010), *etc.* Nowadays this is usually done by building the molecule with an interactive builder in a graphical user interface, then optimizing it with forcefield method by the click of mouse. The resulting structure is then subjected to an *ab initio*, semi-empirical, or density functional theory (DFT) calculations. Some free available software that can be used to sketch molecules, optimize the geometry by forcefield method and create input file for the quantum-chemistry calculations are:

Avogadro (Hanwell *et al.*, 2012), Gabedit (Allouche, 2011), ACD/ChemSketch (www.acdlabs.com), MarvinSketch (<https://www.chemaxon.com/products/marvin/marvinsketch/>).

If the input file does not contain information about the atomic isotropic thermal factors, default values are assigned: $B = 3.0 \text{ \AA}^2$ for non-H atoms and $B = 4 \cdot 6.0 \text{ \AA}^2$ for H atoms are typical values.

Organic materials

The description of crystal structure is straightforward in the case of molecular organic crystals: isolated molecules with known chemical connectivity are packed together by weak intermolecular forces. Low space group symmetry and accidental reflection overlap are typical aspects of these materials. Because the starting model should be a tridimensional representation of the a.u., it is important to know how many molecules are contained in the a.u. (Z'). In case of molecular compounds the a.u. usually contains a molecule or occasionally two or more molecules which differ from one another in orientation or conformation ($Z' > 1$). When a molecule has symmetry coincident with a crystallographic symmetry element, it may occupy a special position, and the a.u. will then be a half molecule or even some smaller fraction ($Z' < 1$). Given that H atoms contribute minimally to the scattering power, they could be ignored.

Inorganic materials

Inorganic materials often consist of connected polyhedra and the topology of this connectivity is not generally known a priori and thus building a model of inorganic crystal structure for global optimization is often less straightforward. Software provide tools to import regular polyhedra or you can find the type of polyhedra you need from literature or from database. The choice of the number of polyhedra to use must be made, taking into account how many atoms are expected per unit cell (volume of atom in the cell). These non-molecular crystals usually crystallize with higher symmetries, and atoms often occupy special positions. At the same time it is necessary to take into account the possibility that polyhedra share atoms at the corners. The use of dynamical occupancy correction (DOC) is able to automatically correct the occupancy of atoms if they are close to a special position or in case of overlap between atoms of the same atomic type. DOC could be very useful when the exact composition of the studied compound is *a priori* not exactly known. Theoretically this means that it is possible to add more atoms than initially deemed necessary, expecting DOC to artificially merge the excess atoms (Favre-Nicolin & Černý, 2002). It is also possible to solve structures putting into the cell independent atoms at random positions. It is advisable that in the optimization process the observation/parameter ratio ('observation' here means the number of reflections; 'parameter' corresponds to the number of variables which are the degrees of freedom) should be at least 8 or more depending on the quality of diffraction data.

Fundamentals of DS methods

DS techniques place a starting structure model in random locations in the asymmetric unit and look for the best set of structural parameters that optimize a cost function (CF). In EXPO2014 (Altomare *et al.*, 2013), two cost functions are available:

1. The default cost function is R_{wp} , which is the agreement factor usually used in the Rietveld refinement

$$R_{wp} = \sqrt{\frac{\sum_i w_i (y_{exp}(\theta_i) - y_{calc}(\theta_i))^2}{\sum_i w_i y_{exp}(\theta_i)^2}}$$

where $y_{exp}(\theta_i)$ and $y_{calc}(\theta_i)$ are the observed and calculated profile values at the θ value of the i -th experimental step, respectively, and $w_i = \frac{1}{y_{exp}(\theta_i)}$. If R_{wp} is used there is no need to extract the structure factor moduli (see the next point 2.) and a profile fitting procedure must be carried out. This task is automatically performed by the software before starting DS method.

2. R_B factor that compares the experimental integrated intensities I_h^{exp} with the intensities I_h^{calc} calculated by the model

$$R_B = \frac{\sum_h |I_h^{exp} - I_h^{calc}|}{\sum_h I_h^{exp}}$$

The preliminary extraction of experimental integrated intensities is automatically performed by using the Le Bail algorithm. The advantage of using R_B with respect to R_{wp} is that it offers a significant reduction of the computational time, but, if the overlap is severe, R_B can be unreliable because the I_h^{exp} values can be affected by large errors.

DS methods do not require the use of the entire pattern: usually the algorithm works well with data up to 2 Å resolution.

If the presence of preferred orientation is suspected, a March-Dollase correction (Dollase, 1986) can be included during SA.

A single global optimization run doesn't guarantee to locate the global minimum in a finite time. Consequently, multiple runs, from randomized starting points, should be performed to increase the chance to find the global minimum.

DS with low quality diffraction pattern

When the quantity of information available from a powder diffraction pattern is limited (due to, *e.g.*, severe peak overlap, broad peaks, preferred orientation, presence of weak scatters) and/or the number of degrees of freedom is large, it may be necessary to add extra chemical information to the optimization process in order to obtain the correct solution. In fact, in this situation, the correct structure may not correspond to the minimum of CF. The use of restraints on bond distances and angles or the application of bond valence restraints are approaches to increase the probability to obtain only chemically plausible models (Falcioni & Deem, 1999; Favre-Nicolin & Černý, 2002). In particular if the final global solution contains atoms colliding each other, anti-bumping restraints can be used to avoid this solution. They are relationships between atoms to prevent atomic group overlapping with other parts of structure. The restraint can be implemented as normal distance restraint, which is only applied if the interatomic distance becomes less than some threshold. The contribution of the anti-bumping restraints to the global cost function is measured by the expression

$$CF_{bump} = \sum_{ij}^n w_{ij} (d_{ij}^{min} - d_{ij}^{model})^{2k}$$

with the condition $d_{ij}^{min} < d_{ij}^{model}$, where d_{ij}^{min} and d_{ij}^{model} are the minimum ideal distances and the model distances, respectively, between pairs of atoms i and j , and summation is over n contacts, k is equal to 2 (Hendrickson & Konnert, 1980). d_{ij}^{min} depends on the atomic elements in contact and on the type of contact, in particular if

the two atoms are hydrogen-bonded atom pairs. d_{ij}^{min} in default are defined as the sum of the atomic radii of the two atoms multiplied by a scale factor ϵ (0.7 is the default value)

$$d_{ij}^{min} = \epsilon(R_i^{vdW} + R_j^{vdW})$$

and R_i^{vdW} and R_j^{vdW} are the van der Waals radii (Feng *et al.*, 2009). This minimum distance d_{ij}^{min} can also be input by the user for each pair of atom types. In principle, the weight ω_{ij} associated to restrains varies with the type of restrains, but in practice a uniform value is used (1.0 is the default value).

The application of restrains is relatively time-consuming because requires that all symmetry equivalent atoms are taken into account. Strictly speaking, it should be unnecessary and it should be use only when the quality of the diffraction pattern is insufficient to avoid that the final solution contains overlapping atoms.

Assessing the final solution

To decide if the best solution found corresponds to the global minimum you can examine the agreement between the calculated pattern and the observed data by using numerical indicators (*e.g.*, R_{wp}) or by a visual inspection of the final plot of the observed intensities, the calculated profile and the difference curve. The most important criteria for judging the quality of the optimization is the chemical sense of the final structural model: interatomic distances should be reasonable, large void spaces are unlikely, hydrogen bonds are expected in presence of nitrogen and/or oxygen atoms. Furthermore, the structure should be consistent with the results of other characterization techniques (*e.g.*, NMR, Raman, ultraviolet, electron microscopy, *etc.*) and with theoretical computations performed by using plane wave density functional theory with dispersion correction (DFT-D) (van de Streek *et al.*, 2010).

Common problems

The most frequent and general source of errors when DS methods fail to estimate the correct solution is in the definition of the starting model: bond lengths, bond angles, ring conformation, molecular stereochemistry may be affected by significant errors; not all scattering objects (*e.g.*, solvent molecules) are included in model, chemical formula is wrong. In this case the model should be improved by using information coming from structural databases, molecular building packages, theoretical computations. You can check the real chemical composition by combination of tools such as X-ray fluorescence spectroscopy (XRF), inductively coupled plasma mass spectrometry (ICP) and scanning electron microscopy/energy dispersive X-ray spectroscopy (SEM/EDS). You can verify that the bond lengths have reasonable values using a tables of standard bond lengths available in Volume C of International Tables for X-Ray Crystallography (Brown *et al.*, 2006).

The default conditions of the optimization algorithm, for complex structure, cannot be appropriate. You can increase the number of moves and/or runs.

Sometimes the initial assumption about thermal factors is invalid. Check thermal factors from similar structures or include them in the optimization process.

Experimental data may have systematic problem: preferred orientation, $K\alpha_2$ contributions, sample displacement, *etc.* Software usually have tools for the correction of these problems.

Space group and cell parameters could be not correct. Different potential space groups and/or unit cell should be tested in independent optimization procedures.

Allouche, A.R. (2011). *J. Comput. Chem.*, **32**, 174-182, <http://gabedit.sourceforge.net/>.

Altomare, A., Cuocci, C., Giacovazzo, C., Moliterni, A., Rizzi, R., Corriero, N. & Falcicchio, A. (2013). *J. Appl. Cryst.* **46**, 1231-1235.

Brown, P.J., Fox, A.G., Maslen, E.N., O'Keefe, M.A. & Willis, B.T.M. (2006). *International Tables for Crystallography*. Vol. C, ch. 9, pp. 774-811

Černý, R. & Favre-Nicolin, F. (2002). *Z. Kristallogr.* **222**, 105–113.

Dollase, W.A. (1986). *J. Appl. Cryst.* **19**, 267-272.

Falcioni, M., Deem, M.W. (1999). *J. Chem. Phys.* **110**, 1754-1766.

Favre-Nicolin, V. & Černý, R. (2002). *J. Appl. Cryst.* **35**, 734-743

Favre-Nicolin, V. & Černý, R. (2004). *Z. Kristallogr.* **219**, 847-856.

Feng, Z.J., Jia, R.R., Dong, C., Cao, S.X. & Zhang, J.C. (2010). *J. Appl. Cryst.* **43**, 179–180.

Florence, A. J., Shankland, N., Shankland, K., David, W. I. F., Pidcock, E., Xu, X., Johnston, A., Kennedy, A. R., Cox, P. J., Evans, J. S. O., Steele, G., Cosgrove, S. D. & Frampton, C. S. (2005). *J. Appl. Cryst.* **38**, 249–259.

Frisch M. J. *et al.* (2009). GAUSSIAN09, Revision A.02. Gaussian, Inc., Wallingford CT, USA.

Hanwell, M.D., Curtis, D.E., Lonie, D.C., Vandermeersch, T., Zurek, E., & Hutchison, G.R. (2012). *J. Cheminform.* **4**, 17, <http://avogadro.openmolecules.net>.

Harris, K.D.M., Tremayne, M. & Kariuki, B.M. (2001). *Angew. Chem. Int. Ed.* **40**, 1626–1651.

Hendrickson, W.A. & Konnert, J.H. (1980). *Computing in Crystallography*, edited by R. Diamond, S. Ramaseshan & K. Venkatesan, pp. 13.01-13.25. Bangalore: Indian Academy of Sciences and International Union of Crystallography.

Stewart, J.J.P., MOPAC2016, Stewart Computational Chemistry, Colorado Springs, CO, USA, (2016), <http://OpenMOPAC.net>.

Shankland, K., Markvardsen, A.J., Rowlatt, C., Shankland, N. & David, W.I.F. (2010). *J. Appl. Cryst.* **43**, 401–406.

Shankland, K., Spillman, M.J., Kabova, E.A., Edgeley, D.S. & Shankland, N. (2013). *Acta Cryst.* **C69**, 1251–1259

Schmidt M.W., Baldrige K.K., Boatz J.A., Elbert S.T., Gordon M.S., Jensen J.H., Koseki S., Matsunaga N., Nguyen K.A., Su S., Windus T.L., Dupuis M. & Montgomery J.A. (1993). *J. Comput. Chem.*, **14**, 1347-1363.

Streek, J. van de & A. Neumann, M.A. (2010) *Acta Cryst.* **B66**, 544–558.

Valiev M., Bylaska E.J., Govind N., Kowalski K., Straatsma T.P., van Dam H.J.J., Wang D., Nieplocha J., Apra E., Windus T.L. & de Jong W.A. (2010). *Comput. Phys. Commun.* **181**, 1477-1489.