



Structure model optimization

Angela Altomare

Institute of Crystallography-CNR, Bari, Italy
angela.altomare@ic.cnr.it

Introduction

The outcome of the *ab-initio* crystal structure solution process by Direct Methods depends on not only the quality of the extracted integrated intensities but also the experimental resolution [1]. In fact, the limited experimental resolution distorts the Fourier synthesis even at atomic resolution: when the resolution is larger than 1 Å, the electron density map is an imperfect representation of the true density. This is particularly true in case of organic compound for which, because of the low atomic scattering factors of the light atoms constituting the molecule, the experimental diffraction intensity at high diffraction angle value is very weak and unusable. In case of powder diffraction data, the resolution bias errors combine with uncertainty on moduli and/or phases of structure factors and often provide an approximate electron density map to which a structure model far from the correct one corresponds: some peaks are missed, false positions occur, some other positions are misplaced from the true ones. Such a quality of structure model requires an optimization step preliminary to the final refinement by the Rietveld method. In fact, the low convergent rate of the Rietveld method does not guarantee the attainment of the correct solution if the starting model to be refined is very rough. Therefore, the *ab-initio* crystal structure solution process by powder diffraction data usually includes procedures aiming at improving the model: recovering the missed atoms and locating atom peaks in better positions.

Several methods have been developed for improving a structure model. They have been introduced in the software package EXPO [2], the program devoted to execute all the steps of the powder solution pathway. Herein, we quote the following procedures [3]:

- a. Weighted Least-Squares Fourier Recycling method (WLSQ-FR) [4];
- b. Resolution Bias Modification algorithm (RBM) [5-7];
- c. COVMAP approach [8];
- d. Random-model-based method (RAMM) [9].

Weighted Least-Squares Fourier recycling (WLSQ-FR)

The Weighted Least-Squares Fourier recycling procedure consists of cyclic combination of two steps:

1. least-squares calculation aiming at improving the atom positions;
2. computation of $2|F_h|_o - |F_h|_c$ Fourier map for recovering the missing atoms, where $|F_h|_o$ is the observed structure factor modulus and $|F_h|_c$ is that calculated from the current model.

Least-squares use $|F_h|_o$ values as observations. The structure factor amplitudes are modulated by a suitable weighting scheme, which is able to take into account the errors on the integrated intensity estimates of the overlapping reflections and to drive the refinement towards effective convergence. The weighting factor depends on two terms: the first factor hinders that least-squares are dominated by the largest intensities while the second factor takes into account the major difficulty to carefully model peak overlapping and background at high diffraction angle regions of the experimental pattern.

The weighted least-squares refine the atomic fractional coordinates and the isotropic thermal factors.

During the WLSQ-FR procedure, to increase the possibility of converging towards the correct model, the observed intensities corresponding to clusters of overlapped reflections are partitioned proportionally to the calculated intensity of each reflection in the cluster. The effect of the partition strongly depends on the quality of the model to be improved. The partitioned intensities, which should be closer to the correct ones, are then used in the Fourier map calculation.

The WLSQ-FR method is effective in particular for the optimization of models corresponding to inorganic compounds and is able to discard false atomic positions and move misplaced positions towards the true ones.

Resolution Bias Correction Algorithm (RBM)

It is well known that the limited experimental data resolution (RES, the smallest interplanar distance value) modifies the Fourier synthesis, introducing severe series truncation errors. In general, the worse the experimental resolution the less accurate the Fourier map: the map is negative in some regions, peaks are broadened and surrounded by positive and negative ripples of decreasing amplitude, peaks are misplaced from their correct positions and have distorted intensities. In case of powder data, the diffraction atomic resolution is seldom attained because peak overlapping limits the efficiency of the decomposition process. In addition, for organic compounds, due to their rapid scattering factor decay, RES is usually far away the atomic one ($RES > 1 \text{ \AA}$). Three different kinds of RBM strategies, working in direct and reciprocal space, have been developed to complete and improve the structure models, in particular of organic compounds. Based on Fourier map calculation, they are able to correct the electron density map by typical resolution bias errors. The RBM procedure working in direct space has been chosen as default strategy of model optimization in the EXPO program. It is below briefly described:

Five cyclic blocks are carried out; in each block the following steps are executed:

1. the $2|F_o| - |F_c|$ Fourier map is computed ($|F_o|$ is the observed and $|F_c|$ the structure factor calculated by the current model, respectively);
2. one atom is positioned at the origin of the unit cell and the corresponding structure factors are calculated and used as coefficients for a Fourier map calculation. The electron density map so obtained is suitably normalized and then subtracted from each peak in the $2|F_o| - |F_c|$ Fourier map obtained at point 1. This step aims at removing the resolution bias errors;
3. each peak in the map is well described by the best fitting Gaussian function. The map so modified is more reliable: peaks are moved on more correct positions and intensities are better resolved;
4. a peak search in the corrected map is executed for providing an improved structure model, whose fractional coordinates and isotropic thermal factors are further optimized by least-squares cycles.

COVMAP approach

Inspired from the theory, which calculates the covariance and the correlation between two points of the electron density map, the COVMAP approach recovers the missing atoms by modifying the electron density map according to some basic crystal chemistry rules. It exploits three kinds of information:

- The chemical interpretation of the peaks of the model, which must be optimized.

- Some basic crystal chemistry rules, essentially the bond distances expected for pairs of atoms.

The electron density map corresponding to the model is considered. If a peak is present at \mathbf{r} , and no other model peak is at bond distance from \mathbf{r} , very likely the model is inadequate in the region surrounding \mathbf{r} . Indeed, if the peak in \mathbf{r} corresponds to a correct atomic position, it is very unlikely that the atom is isolated (atoms are usually connected in molecules). COVMAP, tentatively, assumes that the peak at \mathbf{r} is correct and establishes a non-vanishing covariance between the electron density at \mathbf{r} and the electron density of the points contained in the coordination sphere. The electron density in this sphere is modified to generate additional peaks.

- The efficiency of the crystallographic residual $R_F = \frac{\sum_h \|F_{oh} - |F_{ch}|\|}{\sum_h |F_{oh}|}$, where $|F_{oh}|$ is the observed structure factor modulus extracted from the experimental profile and $|F_{ch}|$ is obtained from the current model. R_F is expected to be smaller for the model including the additional atomic positions suggested by COVMAP than for the original model.

The method has been properly combined with WLSQ-FR and RBM for strengthening its capacity to improve the structure model. In this way, the correct solution is obtained even when the starting model is very approximate.

Random-model-based method (RAMM)

A surprising result was obtained from the combined procedure COVMAP-WLSQ-FR-RBM. It is able to improve a rough structure model and, sometimes, to reach the complete and correct solution even starting from a structure fragment which is far from the true one. This result laid the foundation of the RAMM solution method.

The RAMM strategy skips the phasing step by Direct Methods and substitutes the Direct Methods model by a starting fully random model driven towards the correct solution by the COVMAP-WLSQ-FR-RBM approach. The process of random generation is cyclically carried out and the procedure stops when the correct solution is identified. RAMM is able, on average, to automatically recognize the correct solution and to stop the cycling procedure before the maximum number of cycles is performed. RAMM has been proposed as a valid alternative to the standard Direct Methods-based approach in the EXPO program.

Conclusion

The process of model optimization is usually a necessary step when the *ab-initio* structure solution is carried out. The reason is that, for powder diffraction data, the model obtained after the classical phasing process by Direct Methods is rarely completed and well located. The computer program EXPO, devoted to the *ab-initio* solution of microcrystalline compounds, starts from the minimal information of the experimental pattern and the chemical formula. It is able to carry out automatically all the steps of the process, from the indexation to the structure model refinement. In particular, it automatically applies specific strategies for optimizing the structure model, depending on the kind of compound under study: the WLSQ-FR procedure runs in case of inorganic structure and the RBM approach for organic and metal-organic structure. Moreover, EXPO makes available the

additional COVMAP-based and RAMM methods, which can be applied when the good outcome is not attained in a default way.

References

- [1] Altomare, A., Cuocci, C., Giacovazzo, C., Kamel, G.S., Moliterni, A. & Rizzi, R. (2008): Minimally resolution biased electron-density maps. *Acta Cryst. A* **64**: 326–336.
- [2] Altomare, A., Cuocci, C., Giacovazzo, C., Moliterni, A., Rizzi, R., Corriero, N. & Falcicchio, A. (2013): EXPO2013: a kit of tools for phasing crystal structures from powder data. *J. Appl. Cryst.* **46**:1231-1235.
- [3] Altomare A., Cuocci C., Moliterni A. & Rizzi R. (2013): Single crystal and powder XRD techniques: An overview *In* Dibenedetto, A. & Aresta, M. (2013) (eds.): Inorganic Micro- and Nanomaterials: Synthesis and Characterization, Germany: De Gruyter, chapter 4, 57-91.
- [4] Altomare A., Cuocci C., Giacovazzo C., Moliterni A.G.G. & Rizzi R. (2006): Powder diffraction: the new automatic least-squares Fourier recycling procedure in EXPO2005. *J. Appl. Cryst.* **39**: 558–562.
- [5] Altomare, A., Cuocci, C., Giacovazzo, C., Moliterni, A. & Rizzi, R. (2008): Correcting resolution bias in electron density maps of organic molecules derived by direct methods from powder data. *J. Appl. Cryst.* **41**: 592–599.
- [6] Altomare, A., Cuocci, C., Giacovazzo, C., Moliterni, A. & Rizzi, R. (2009): Correcting electron-density resolution bias in reciprocal space. *Acta Cryst. A* **65**: 183–189.
- [7] Altomare, A., Cuocci, C., Giacovazzo, C., Moliterni, A. & Rizzi, R. (2010): The dual-space resolution bias correction algorithm: applications to powder data. *J. Appl. Cryst.* **43**: 798–804.
- [8] Altomare, A., Cuocci, C., Giacovazzo, C., Moliterni, A. & Rizzi, R. (2012): COVMAP: a new algorithm for structure model optimization in the EXPO package. *J. Appl. Cryst.* **45**: 789-797.
- [9] Altomare, A., Cuocci, Moliterni, A. & Rizzi, R. (2013): RAMM: a random-model-based method for solving *ab initio* crystal structure using EXPO package. *J. Appl. Cryst.* **46**: 476-482.