Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 12 August 2008
Accepted 12 September 2008

Deconvolution of the interatomic vector set using a convolution table

Jianglin Feng

Department of Molecular Physiology and Biological Physics, University of Virginia, Charlottesville, VA 22908, USA. Correspondence e-mail: jf4x@virginia.edu

The deconvolution of the interatomic vector set (the ideal Patterson function) with the superposition technique is not complete because of the vector overlaps: multiple images and false peaks usually exist in the superposition map. Here, a new method for the deconvolution of the interatomic vector set is presented. This method involves constructing a table termed the 'convolution table' from vectors in a superposition map and then sorting the table so that vectors belonging to different images are separated, and thus the overlaps are naturally solved. This method does not use the symmetry information.

1. Introduction

The Patterson function $P(\mathbf{r})$ is the convolution of the electron density function $\rho(\mathbf{r})$ with its inverse $\rho(-\mathbf{r})$. Assuming that $\rho(\mathbf{r})$ is a set of N atomic vectors, then $P(\mathbf{r})$ is a set of N^2 interatomic vectors, or N shifted and superposed copies (called images) of $\rho(\mathbf{r})$. Patterson approaches are devised to deconvolute $P(\mathbf{r})$ to obtain a single image. Superposition (Wrinch, 1939; Buerger, 1959) is a basic technique for Patterson interpretation. However, a single image can rarely be obtained by a superposition because of the overlapping problems: the superposition map usually contains multiple images and many false peaks. The problem of multiple images has been extensively analyzed (Simonov, 1965; Germain & Woolfson, 1966; Kuzmin *et al.*, 1973). Overlapping problems are usually more serious for higher-symmetry crystals and symmetry analysis of interatomic vectors plays an important role in most Patterson analyses (Harker, 1936; Mighell & Jacobson, 1963; Borisov, 1965; Simpson *et al.*, 1965; Ilyukhin *et al.*, 1972; Lenstra & Schoone, 1973; Tollin, 1975; Luger & Fuchs, 1986; Richardson & Jacobson, 1987; Sheldrick, 1991; Pavelčík, 1994; Burla *et al.*, 2006). Here we present a general method to solve the overlapping problems without using the symmetry information.

2. The convolution-table method

If we use $\{\}$ to denote a set, then $\rho(\mathbf{r}) = \{\mathbf{r}_i\}$ and $P(\mathbf{r}) = \{\mathbf{r}_{ij}\}$ ($i, j = 1, 2, \ldots, N$), where $\mathbf{r}_i$ represents the ith atomic vector and $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$ is an interatomic vector. As a set of images, $P(\mathbf{r}) = \{I_1, I_2, \ldots, I_N\}$ with image $I_i = \{\mathbf{r}_{i1}, \mathbf{r}_{i2}, \ldots, \mathbf{r}_{iN}\}$, or $P(\mathbf{r}) = \{I_{-1}, I_{-2}, \ldots, I_{-N}\}$, with $I_{-i} = \{\mathbf{r}_{1i}, \mathbf{r}_{2i}, \ldots, \mathbf{r}_{Ni}\}$. All these images are connected at the origin. The superposition map $S(\mathbf{r})$ can be defined as the intersection of vector set $P(\mathbf{r})$ and its shift $P(\mathbf{r} + \mathbf{u})$,

$$S(\mathbf{r})|\mathbf{u} = P(\mathbf{r}) \cap P(\mathbf{r} + \mathbf{u}) = \{\mathbf{r}_{ij}\} \cap \{\mathbf{r}_{ij} + \mathbf{u}\}, \quad (1)$$

where the shift vector $\mathbf{u}$ is an interatomic vector in $P(\mathbf{r})$, and the intensity of the peak in the superposition map takes the minimum of the two superposed peaks. Similarly, a superposition with multiple shifts is defined as the intersection of $P(\mathbf{r})$ and its multiple shifts $P(\mathbf{r} + \mathbf{u}_1)$, $P(\mathbf{r} + \mathbf{u}_2)$ *etc.* Without overlaps among normally unrelated vectors, a superposition map contains two images, *e.g.* $S(\mathbf{r})|\mathbf{r}_{23} = \{I_2, I_{-3}\}$. Overlaps may be classified into two types: (1) two or more vectors in $P(\mathbf{r})$ coincide, for example, $\mathbf{r}_{23} = \mathbf{r}_{41}$, and (2) a vector coincides with the sum of two other vectors, for example, $\mathbf{r}_{56} = \mathbf{r}_{71} + \mathbf{r}_{23}$. The first type of overlap results in extra images if they are used as the shift vector (Buerger, 1959; Kuzmin *et al.*, 1973; Pavelčík, 1994). If $\mathbf{u} = \mathbf{r}_{23} = \mathbf{r}_{41}$, then a map of four images is obtained: $S(\mathbf{r})|\mathbf{r}_{23} = \{I_2, I_{-3}, I_4, I_{-1}\}$. The second type of overlap results in extra peaks (false peaks) if one of the involved vectors is used as the shift vector. For example, if $\mathbf{u} = \mathbf{r}_{56}$ and $\mathbf{r}_{56} = \mathbf{r}_{71} + \mathbf{r}_{23}$, which implies that $\mathbf{r}_{17} + \mathbf{r}_{56} = \mathbf{r}_{23}$, $\mathbf{r}_{32} + \mathbf{r}_{56} = \mathbf{r}_{71}$, $\mathbf{r}_{37} + \mathbf{r}_{56} = \mathbf{r}_{21}$ and $\mathbf{r}_{12} + \mathbf{r}_{56} = \mathbf{r}_{73}$, then four extra peaks $\{\mathbf{r}_{23}, \mathbf{r}_{71}, \mathbf{r}_{21}, \mathbf{r}_{73}\}$ appear because they belong to both $P(\mathbf{r} + \mathbf{r}_{56})$ and $P(\mathbf{r})$ [equation (1)], and $S(\mathbf{r})|\mathbf{r}_{56} = \{I_5, I_{-6}, \mathbf{r}_{71}, \mathbf{r}_{23}, \mathbf{r}_{73}, \mathbf{r}_{21}\}$. If $\mathbf{r}_{23} = \mathbf{r}_{41}$, $\mathbf{r}_{56} = \mathbf{r}_{71} + \mathbf{r}_{23}$ and $\mathbf{u} = \mathbf{r}_{23}$ then both types of overlaps are involved, and the superposition map contains four images and four false peaks: $S(\mathbf{r})|\mathbf{r}_{23} = \{I_2, I_{-3}, I_4, I_{-1}, \mathbf{r}_{56}, \mathbf{r}_{57}, \mathbf{r}_{16}, \mathbf{r}_{17}\}$.

The interatomic vector set $P(\mathbf{r})$ can be completely deconvoluted by the following procedure:

(*a*) Obtain a superposition map $S(\mathbf{r})|\mathbf{u}$, which contains one or several complete images and some false peaks.

(*b*) Use the vectors (peaks) in $S(\mathbf{r})|\mathbf{u} = \{\mathbf{v}_k\}$ ($k = 1, 2, \ldots, m$) to construct an $m \times m$ convolution table: if $\mathbf{v}_{pq} = \mathbf{v}_p - \mathbf{v}_q$ ($p, q = 1, 2, \ldots, m$) is an interatomic vector, *i.e.* a peak of $P(\mathbf{r})$, then the corresponding table element X_{pq} takes the value one, otherwise, zero. Each row represents a consecutive superposition map with two shifts (first $\mathbf{u}$ and then $-\mathbf{v}_k$ for the kth row). The number of ones in the row is the number of peaks in the superposition. The diagonal vectors are origin vectors, and hence the items X_{kk} are all ones. Furthermore, rows corresponding to $\mathbf{v}_k = \mathbf{0}$ and $\mathbf{v}_k = \mathbf{u}$ are all ones.

(*c*) Sort the convolution table. If vectors $\mathbf{v}_{k1}, \mathbf{v}_{k2}, \ldots, \mathbf{v}_{kN}$ construct an image I_k, then the vectors generated by their convolution construct the entire vector set $P(\mathbf{r})$, and in the convolution table they form an $N \times N$ all-one square domain if they are put next to each other. By definition, this is the largest all-one square domain, and the vectors of the domain are, therefore, from a single image. The following is a simple algorithm to sort the convolution table to decompose the superposition map $S(\mathbf{r})|\mathbf{u}$. We define (1) the *product* of row p and row q as a new row (*product row*), whose element is $X_{pk}X_{qk}$ ($k = 1, 2, \ldots, m$), and (2) the *rank* of a row as the number of ones in the row. Row $\mathbf{v}_k = \mathbf{0}$ and row $\mathbf{v}_k = \mathbf{u}$ are common to all images so we do not consider them. Start with row $k1$, find among the rest a row that has the highest-ranked product with row $k1$, and record the product row.

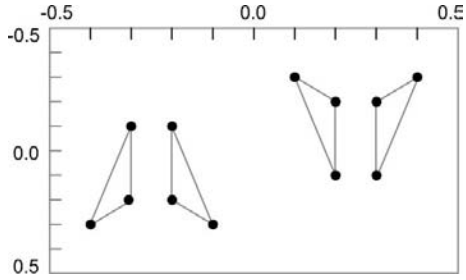

Figure 1
The unit cell of a two-dimensional structure with symmetry *P2mg* (from Simpson *et al.*, 1965).

Then find the third row that has the highest-ranked product with the product row, and update the product row. Repeat this till the rank of the product row is equal to the number of rows that the product row is generated from. This final product row represents an all-one square domain in the convolution table, and it is largely determined by the starting row $k1$. By starting from row $k2$, $k3, \ldots, km$, we can find all representative product rows in the same way. Among all these product rows, those (usually several) having the highest rank determine the largest all-one domains and each of them represents a single image. The number of different highest-ranked product rows is the number of images contained in superposition map $S(\mathbf{r})|\mathbf{u}$. The convolution table can now be reordered so that vectors belonging to different images are grouped together, and multiple images and false peaks can be easily recognized.

3. An example

Fig. 1 shows the unit cell of a two-dimensional structure with 12 equal atoms and symmetry *P2mg*, which was used by Simpson *et al.* (1965) for illustrating the symmetry-minimum-function method. Fig. 2(*a*) is the Patterson function generated by Fourier transforms. It has 52 non-origin peaks: four are single (smallest or weakest peaks) and all others are two-to-fourfold overlapped. The circled peak (a twofold overlap) in Fig. 2(*a*) is used as a shift vector to obtain the superposition map (Fig. 2*b*). There are 26 peaks in Fig. 2(*b*) and they are labeled as **a**, **b**, . . . , **z** in decreasing order of intensity. Fig. 2(*c*) shows the convolution table constructed from all 26 vectors, and Fig. 2(*d*) is the sorted convolution table. The top-left all-one domain has 12 vectors {**a**, **u**, **j**, **d**, **i**, **g**, **y**, **b**, **n**, **q**, **k**, **m**}, which form exactly a single

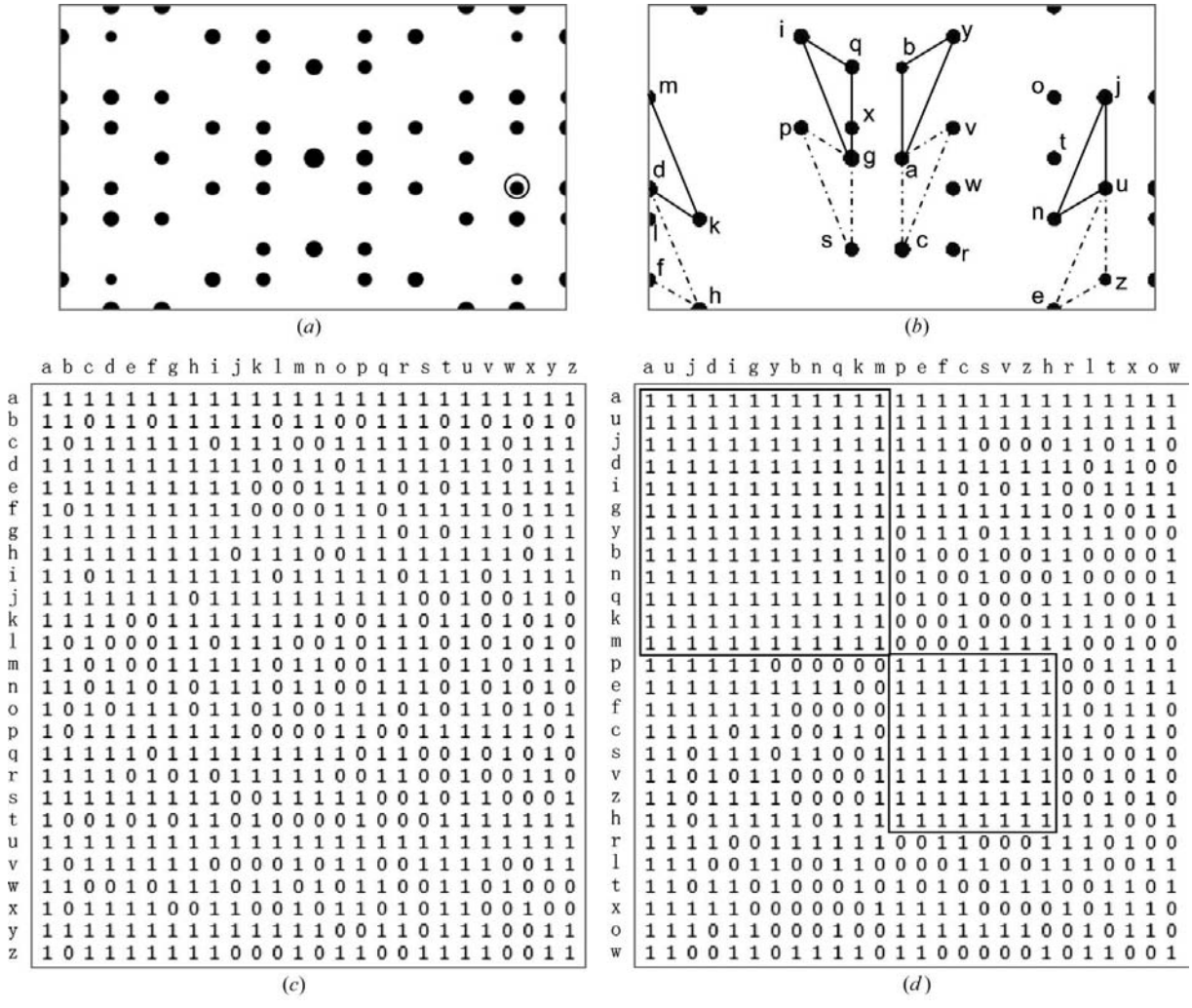

Figure 2
(*a*) The Patterson function of the example in Fig. 1. (*b*) The superposition map with the shift vector circled in (*a*). (*c*) The unsorted and (*d*) the sorted convolution table. The 12 vectors in the top-left box of (*d*) form a single image [shown by solid lines in (*b*)]; vectors **a**, **u**, **d**, **g** and the eight vectors in the second box form another single image [dotted lines in (*b*)]. There are six false peaks: **r**, **l**, **t**, **x**, **o** and **w**.

image and are shown by solid lines in Fig. 2(*b*). By examining the sorted table, one can easily find that vectors **a**, **u**, **d** and **g** in the top-left boxed domain and the eight vectors in the second boxed domain construct another 12-atom single image {**a**, **u**, **d**, **g**, **p**, **e**, **f**, **c**, **s**, **v**, **z**, **h**}, which is shown by dotted lines in Fig. 2(*b*). These two images share a substructure {**a**, **u**, **d**, **g**}. Peaks **r**, **l**, **t**, **x**, **o** and **w** are false peaks. The Patterson map is thus completely deconvoluted without using the symmetry information. Two-dimensional model structures of different symmetries and with equal or non-equal atoms have been successfully analyzed with this method.

The convolution-table method may be used to analyze small structures or large structures containing a small number of heavy atoms. Systematic applications of this method to real three-dimensional crystals are in progress.

References

Borisov, S. V. (1965). *Sov. Phys. Crystallogr.* **9**, 515.
Buerger, M. J. (1959). *Vector Space.* New York: John Wiley.
Burla, M. C., Caliandro, R., Carrozzini, B., Cascarano, G. L., De Caro, L., Giacovazzo, C., Polidori, G. & Siliqi, D. (2006). *J. Appl. Cryst.* **39**, 527–535.
Germain, G. & Woolfson, M. M. (1966). *Acta Cryst.* **21**, 845–848.
Harker, D. (1936). *J. Chem. Phys.* **4**, 381.
Kuzmin, E. A., Drozdov, Iu. N., Iliukhin, V. V. & Belov, N. V. (1973). *Dokl. Akad. Nauk SSSR*, **209**, 344–347.
Lenstra, A. T. H. & Schoone, J. C. (1973). *Acta Cryst.* A**29**, 419–423.
Ilyukhin, V. V., Borisov, S. V., Chernov, A. N. & Belov, N. V. (1972). *Sov. Phys. Crystallogr.* **17**, 227.
Luger, P. & Fuchs, J. (1986). *Acta Cryst.* A**42**, 380–386.
Mighell, A. D. & Jacobson, R. A. (1963). *Acta Cryst.* **16**, 443–445.
Pavelčík, F. (1994). *Acta Cryst.* A**50**, 467–474.
Richardson, J. W. & Jacobson, R. A. (1987). *Patterson and Pattersons*, edited by J. P. Glusker, B. K. Patterson & M. Rossi, pp. 310–317. Oxford University Press.
Sheldrick, G. M. (1991). *Crystallographic Computing 5. From Chemistry to Biology*, edited by D. Moras, A. D. Podjarny & J. C. Thierry, pp. 145–157. Oxford University Press.
Simonov, V. I. (1965). *Sov. Phys. Crystallogr.* **10**, 116–120.
Simpson, P. G., Dobrott, R. D. & Lipscomb, W. N. (1965). *Acta Cryst.* **18**, 169–179.
Tollin, P. (1975). International Summer School on Crystallographic Computing, Prague, Czech Republic.
Wrinch, D. M. (1939). *Philos. Mag.* **27**, 490–507.