CrossMark

TECHNICAL ARTICLE

Evaluating the electron density model by applying an imaginary modification

Hui Li,[1,a)] Meng He,[2,3,a)] and Ze Zhang[4]
[1]*Beijing University of Technology, Beijing 100124, China*
[2]*CAS Key Laboratory of Nanosystem and Hierarchical Fabrication, CAS Center for Excellence in Nanoscience, National Center for Nanoscience and Technology, Beijing 100190, China*
[3]*School of Physical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China*
[4]*Zhejiang University, Hangzhou 310014, China*

(Received 7 June 2017; accepted 18 October 2017)

A function has been proposed to evaluate the electron density model constructed by inverse Fourier transform using the observed structure amplitudes and trial phase set. The strategy of this function is applying an imaginary electron density modification to the model, and then measuring how well the calculated structure amplitudes of the modified model matches the expected structure amplitudes for the modified correct model. Since the correct model is not available in advance, a method has been developed to estimate the structure amplitudes of the modified correct model. With the estimated structure amplitudes of the modified correct model, the evaluation function can be calculated approximately. Limited tests on simulated diffraction data indicate that this evaluation function may be valid at the data resolution better than 2.5 Å. © *2017 International Centre for Diffraction Data.*
[doi:10.1017/S0885715617001154]

Key words: crystallographic phase problem, figure of merit, electron density modification

I. INTRODUCTION

In X-ray crystallography, the structure of a crystal is actually the electron density distribution in the crystal. The electron density distribution of a crystal is the inverse Fourier transform of its structure factors. Unfortunately, only structure amplitudes can be derived from the experimentally recorded diffraction intensity data; phases of the structure factors usually cannot be measured directly. Various techniques have been developed to derive the phases and then build the electron density map, such as the direct methods (Karle and Hauptman, 1950; Woolfson, 1987; Woolfson and Fan, 1995), which have been dominating the small-molecule X-ray crystallography. In a previous paper (Li *et al.*, 2015), we have proposed a new strategy to derive the phases for observed structure amplitudes: the correctly retrieved electron density map is identified by an evaluation function (or a figure of merit) from a tremendous number of electron density models, which are constructed by assigning all possible phase sets (within certain accuracy) to the observed structure amplitudes. Two evaluation functions were proposed for this purpose. Recently, Burla *et al.* (2017) developed a multipurpose figure of merit, MPF, which can also be used in recognizing the quality of a current electron density model. Here, we put forward another evaluation function (or figure of merit) for identifying the correct electron density map. Instead of examining the electron density model itself, this evaluation function assesses the agreement between the calculated and expected structure amplitudes of the modified electron density model.

Indeed, iterative electron density modification methods for structure determination, such as the charge flipping methods (Oszlányi and Sütő, 2004, 2008), have been developed a long time ago, and are gaining their popularity. In the process of structure determination using the iterative electron density modification method, the initial electron density model is usually obtained by the inverse Fourier synthesis using observed structure amplitudes and random phases. The obtained electron density map is modified in a particular way in each following iterative cycle, and the correct electron density map may be obtained finally. Oszlányi and Sütő have proposed an R factor to monitor the convergence of charge flipping iterations (Oszlányi and Sütő, 2004). This R factor measures how well the calculated structure amplitudes of the modified current electron density model agree with the observed ones. Inspired by the R factor proposed by Oszlányi and Sütő, we conceived an evaluation function, which hopefully can be used to identify the correct electron density model. The details of this evaluation function will be presented in the following sections.

II. DEFINITION AND CALCULATION

A. Notation

EDM: electron density modification.
FT: Fourier transform.
IFT: inverse Fourier transform.
ρ_{cryst}: the electron density distribution in a real crystal.
ρ_o: the image of ρ_{cryst} at a certain resolution, which is calculated by IFT using the observed structure amplitudes and correct phases.
ρ: the current electron density model which is calculated by IFT using the observed structure amplitudes and the current trial phase set.
ρ_o*: the resultant electron density map after ρ_o has been modified in a particular way.

a)Author to whom correspondence should be addressed. Electronic mail: huilicn@yahoo.com, mhe@nanoctr.cn

https://doi.org/10.1017/S0885715617001154 Published online by Cambridge University Press

ρ^*: the resultant electron density map after the current electron density model ρ has been modified in a particular way.
$|F_{obs,h}|$: the observed structure amplitudes.
$F_{cal,h}$: the calculated structure factors of the current electron density model, ρ.
$F^*_{cal,h}$: the calculated structure factors of ρ^*.
$F^*_{o,cal,h}$: the calculated structure factors of ρ_o^*.
ϕ: the current trial phase set for $|F_{obs,h}|$.
f_i: atomic scattering factor of the ith atom in the unit cell.
h: Miller indices.

B. Definition of the evaluation function

Traditionally, the so-called crystallographic residual is generally used to evaluate how well the structure model agrees with the observed structure amplitudes. The crystallographic residual is defined as

$$\mathrm{RES} = \frac{\sum_h ||F_{obs,h}| - |F_{cal,h}||}{\sum_h |F_{obs,h}|} \quad (1)$$

RES is usually used in cases where an atomic structural model is available, and then $|F_{cal,h}|$ can be calculated using the atomic scattering factors and atomic coordinates. Generally, RES cannot be used to evaluate the quality of an electron density model which is built up by IFT using a trial phase set ϕ and the observed structure amplitudes because $|F_{cal,h}| = |F_{obs,h}|$ always holds regardless of whatever the phase set is assigned to the observed structure amplitudes.

An R factor slightly different from RES is proposed by Oszlányi and Sütő, and defined as

$$R = \frac{\sum_h ||F_{obs,h}| - |F^*_{cal,h}||}{\sum_h |F_{obs,h}|} \quad (2)$$

$|F^*_{cal,h}|$ is the structure amplitude of the modified current electron density model, therefore is always different from $|F_{obs,h}|$ even if the model is correct. When the high-resolution diffraction data are available and the trial phase set ϕ is correct (within certain accuracy), an EDM implemented in a particular way (e.g. electron density flipping in low-density regions as implemented in the charge flipping algorithm) results in very little difference between ρ and ρ^*, therefore a small R factor is an indicator of the correctness of the electron density model. Nevertheless, when the resolution of the diffraction data is poor, electron density ripples because the truncation effect of Fourier synthesis will lead to considerable electron density out of the nuclear sites even the electron density model is correctly constructed using the correct phase set. Then an EDM, which usually take effect in low-density regions, will lead to significant difference between ρ and ρ^*. We expect that the R factor will no longer be capable of indicating the correct model in such cases. This conjecture was confirmed by examining the evolution of R factor in successive EDM iterative cycles using simulated diffraction data at various resolutions. As shown in Figure 1, with the deterioration of the data resolution, R values for the correct electron density models increase, and so do the finally resultant R factors of iterative EDM cycles. In the case where the data resolution is 2.5 Å, the R value for the correct electron density model is significantly higher than the finally resultant one obtained from the iterative cycles. Clearly, in such a case, R factor defined in Eq. (2) is no longer capable of indicating the correct electron density model. We also note that R factor works well when the data of high resolution are available, and the correct model can be retrieved from the diffraction data by using the iterative EDM method. This implies that it might be unnecessary to find a better figure of merit for iterative EDM algorithms to monitor the evolution of electron density model in the successive iterative cycles, but we do need a new figure of merit different from R factor for identifying the correct electron density model when the data resolution is poor.

We propose R_{Tian} factor which is defined as

$$R_{Tian} = \frac{\sum_h ||F^*_{o,cal,h}| - |F^*_{cal,h}||}{\sum_h |F^*_{o,cal,h}|} \quad (3)$$

where $F^*_{cal,h}$ and $F^*_{o,cal,h}$ represent the calculated structure factors obtained from the modified current and correct electron density models, respectively, to which the same EDM has been applied. When the trial phase set is correct within a certain accuracy, R_{Tian} will approach to zero. This holds true even for low-resolution data. The difficulty lies in that $|F^*_{o,cal,h}|$ cannot be obtained before the correct electron density model is available. If a reasonable approximation of $|F^*_{o,cal,h}|$ can be established using the available data, then R_{Tian} can be used as a figure of merit to identify the correct electron density model.

C. The estimation of R_{Tian} factor

We suppose that the lattice parameters, the constituent atoms and their numbers in the unit cell are known together with the observed structure amplitudes before the structure determination. The number of atoms in the unit cell of the investigated crystal is assumed to be N, and the atomic scattering factor of the ith atom is supposed to be f_i.

For a hypothetical structure which has the same unit cell as that of the investigated crystal and only one atom with an atomic scattering factor of f_i at the origin, its structure factor can be readily calculated using the following equation

$$F_{i,h} = \sum f_i \times \exp(-2\pi i \vec{h} \cdot \vec{r}) = f_i. \quad (4)$$

The electron density map of the hypothetical structure ρ_i can be calculated by a straightforward IFT, namely,

$$\rho_i = \mathrm{IFT}(F_{i,h}) = \mathrm{IFT}\ (f_i). \quad (5)$$

The Fourier synthesis in Eq. (5) is terminated at a resolution identical to that of the observed data of the investigated crystal. We suppose ρ_i changes into ρ_i^* after applying an imaginary EDM, then the structure factor of ρ_i^*, which is denoted as $F^*_{i,h}$, can be obtained by Fourier transform. We suppose the modification applied to ρ_i is V_i, then we have

$$\rho_i^* = \rho_i \times V_i \quad (6)$$

and

$$F^*_{i,h} = \mathrm{FT}(\rho_i^*) = \mathrm{FT}\left(\rho_i \times V_i\right) = \mathrm{FT}(\rho_i) \otimes \mathrm{FT}(V_i) = f_i \otimes v_i \quad (7)$$

https://doi.org/10.1017/S0885715617001154 Published online by Cambridge University Press

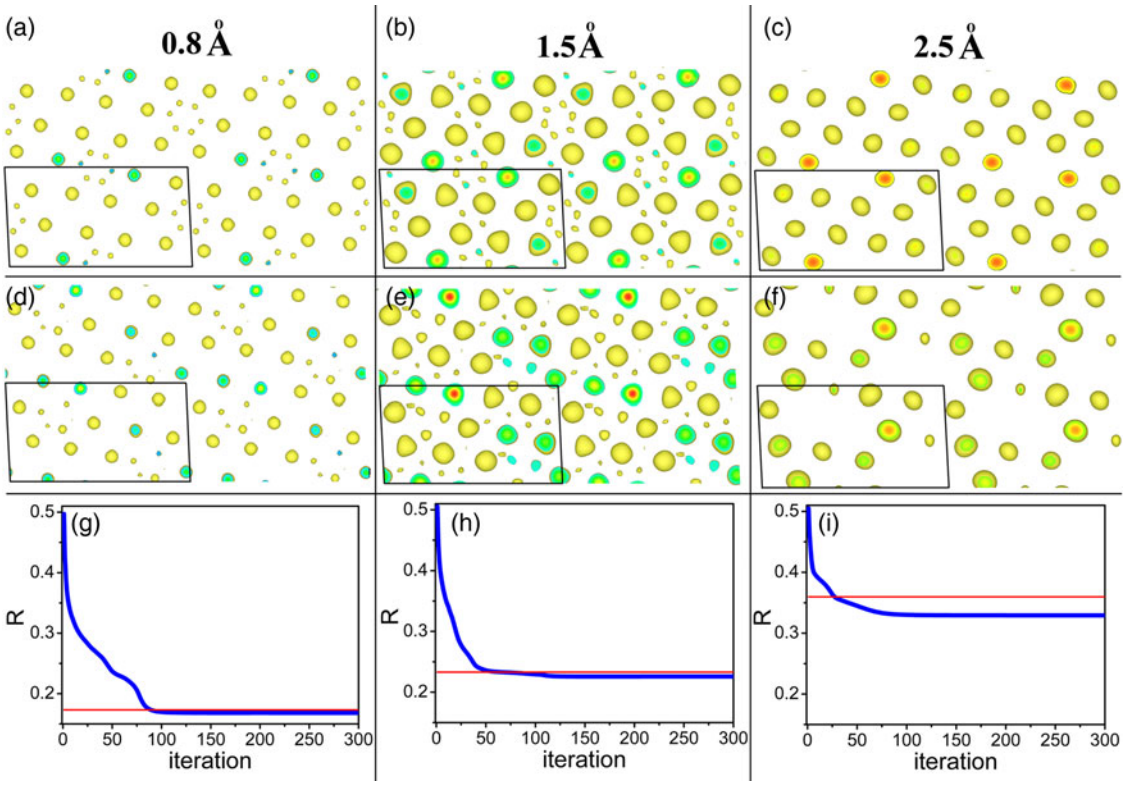


Figure 1. (color online) Electron density maps of C_6Br_6 at the resolutions of 0.8 (a), 1.5 (b), and 2.5 Å (c) constructed by Fourier synthesis using the structure amplitudes and correct phases. Shown in (d–f) are the resultant electron density models at the resolutions of 0.8, 1.5, and 2.5 Å, respectively, which are obtained from successive iterative EDM cycles. The structure amplitudes and correct phases are calculated from the structure data of C_6Br_6 reported by Wu *et al.* (2004). All electron density maps in (a–f) are projections of C_6Br_6 along [010] direction. The unit cell is represented by a parallelogram in each panel. Note that electron density models shown in (d–f) have origins different from that of the original structural model. Only electron density above a certain positive threshold is illustrated for clarity. (g–i) The evolution of R factor with the iterative EDM cycles. In each iterative cycle, non-negative electron densities are retained and the negative ones are reset to zero. The red horizontal line indicates the R value corresponding to the correct structure model. Panels (g–i) correspond to the iterative cycles using the simulated diffraction data of 0.8, 1.5, and 2.5 Å, respectively. Crystal data of C_6Br_6: $P2_1/n$, $a = 8.381$, $b = 4.0192$, $c = 15.3939$ Å, $\beta = 92.674°$, $Z = 2$.

where $\otimes$ represents convolution, and v_i represents FT(V_i). In analogy to ρ_i, ρ_i^* may be considered as the electron density map built up using a pseudo-atom at the origin with an atomic scattering factor $f_i^* = f_i \otimes v_i$, which is hereinafter referred to as pseudo-atomic scattering factor.

The imaginary EDM applied to ρ_i is assumed to be

$$V_i = \begin{cases} 1, & \rho_i \geq \delta_i \\ 0, & \rho_i < \delta_i \end{cases} \tag{8}$$

where δ_i is a real number close to 0. When δ_i is appropriately selected, ρ_i^* remains the same as ρ_i only in a small region around the center of the atom and is set to zero in the rest regions. An example is given in Figure 2 to illustrate how ρ_i^* and f_i^* vary with δ_i. It can be clearly seen from Figure 2 that positive ρ_i^* is limited to the region very proximate to the atomic position even when δ_i is quite a small positive value. In addition, it is found that f_i^* varies significantly with δ_i, especially in the low $\sin\theta/\lambda$ range. When δ_i is set to an appropriate value, f_i^* coincides with the original atomic scattering factor very well in the low $\sin\theta/\lambda$ range, and only in the high $\sin\theta/\lambda$ range significant difference between f_i^* and f_i can be observed. Since f_i value at $\theta = 0$ represents the number of electrons of the ith atom, f_i^* coinciding with f_i in the low $\sin\theta/\lambda$ range implies that the pseudo-atom has the same number of electrons as that of the ith atom at this δ_i value.

The "observed" electron density distribution of a crystal, ρ_o, may be considered as the superposition of electron density of each constituent atom in the unit cell, which can be represented by

$$\rho_o = \sum_i^N \rho_i(r_i) \tag{9}$$

where $\rho_i(r_i)$ is ρ_i translated from the origin to the atomic position r_i.

We assume

$$\rho_o^* = \rho_o \times \sum_i^N V_i(r_i) = \sum_i^N \rho_i(r_i) \times \sum_j^N V_j(r_j) \tag{10}$$

where $V_i(r_i)$ is V_i translated from the origin to the atomic position r_i.

It is achievable that the region with $V_j(r_j) = 1$ of each atom does not overlap with that of adjacent atoms by selecting appropriate δ_i. In such cases,

$$\rho_i(r_i) \times V_j(r_j) = 0$$

for $i \neq j$

https://doi.org/10.1017/S0885715617001154 Published online by Cambridge University Press

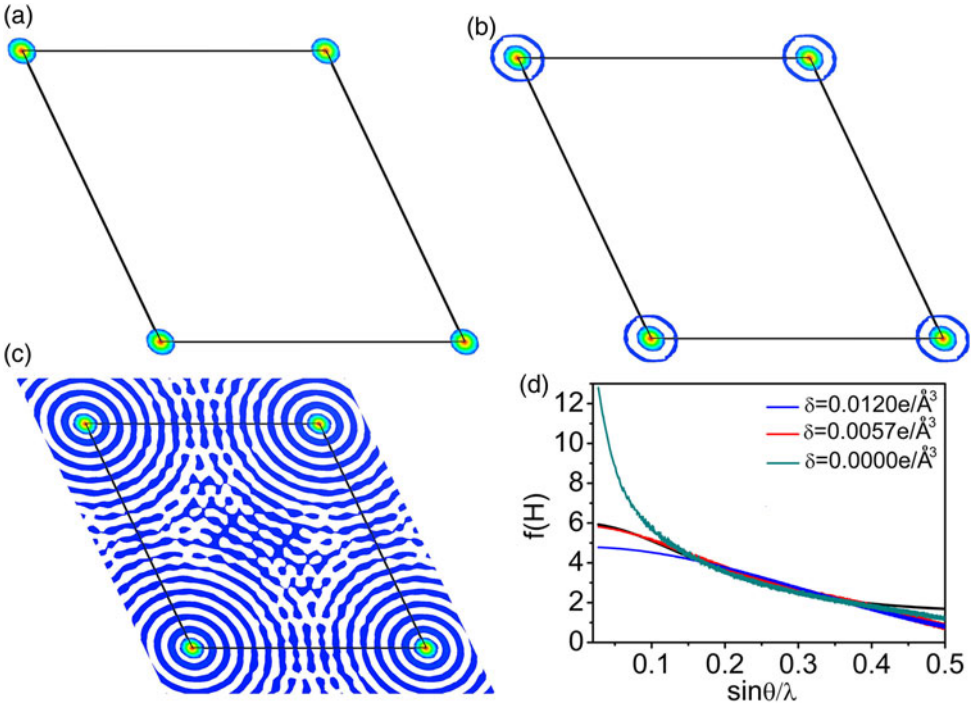


Figure 2. (color online) The dependence of ρ_i^* and f_i^* on δ_i. (a–c) The modified electron density distributions on the (001) plane (at $z = 0$) of a hypothetical structure with only one carbon atom at origin. The colored regions have electron densities greater than δ_i. The original electron density map of the hypothetical structure is constructed at the resolution of 1 Å by using Eq. (5), and then modified according to Eqs (6) and (8). Maps shown in (a–c) correspond to the resultant electron density distributions modified with $\delta = 0.012$, 0.0057, and 0.0 e/Å^3, respectively. The hypothetical structure is derived from a known organic compound $C_{252}H_{326}O_{19}$ reported by Czugler *et al.* (2003). (d) f_i^* corresponding to $\delta = 0.0$, 0.0057, and 0.012 e/Å^3 and the atomic scattering factor of carbon (shown in black).

Then we obtain

$$\rho_o^* = \sum_i^N \rho_i(r_i) \times V_i(r_i) = \sum_i^N \rho_i^*(r_i) \tag{11}$$

and

$$F_{o,cal,h}^* = \mathrm{FT}(\rho_o^*) = \sum_i^N f_i^* \exp(-2\pi i\vec{h} \cdot \vec{r_i}) \tag{12}$$

The experimental observed diffraction intensity of a crystal is given by

$$\begin{aligned} |F_{obs,h}|^2 &= K\sum_i^N\sum_j^N f_i f_j \exp\left[2\pi i\vec{h}\cdot\left(\vec{r_i}-\vec{r_j}\right)\right] \\ &= K\{\sum_i^N f_i^2 + \sum_i^N \sum_{j(i\neq j)}^N f_i f_j \exp\left[2\pi i\vec{h}\cdot\left(\vec{r_i}-\vec{r_j}\right)\right]\} \\ &= K\sum_i^N f_i^2 \left\{1+\frac{\sum_i^N\sum_{j(i\neq j)}^N f_i f_j \exp\left[2\pi i\vec{h}\cdot\left(\vec{r_i}-\vec{r_j}\right)\right]}{\sum_i^N f_i^2}\right\} \end{aligned} \tag{13}$$

where K is a scale factor independent of θ. According to Eq. (13), the intensity of the imaginary crystal with the electron density distribution ρ_o^* should be

$$\begin{aligned} |F_{o,cal,h}^*|^2 &= K\sum_i^N f_i^{*2}\left\{1+\frac{\sum_i^N\sum_{j(i\neq j)}^N f_i^* f_j^* \exp\left[2\pi i\vec{h}\cdot\left(\vec{r_i}-\vec{r_j}\right)\right]}{\sum_i^N f_i^{*2}}\right\} \\ &= K\sum_i^N f_i^{*2}\left\{1+\frac{\sum_i^N\sum_{j(i\neq j)}^N k_i k_j f_i f_j \exp\left[2\pi i\vec{h}\cdot\left(\vec{r_i}-\vec{r_j}\right)\right]}{\sum_i^N k_i^2 f_i^2}\right\} \end{aligned} \tag{14}$$

where $k_i = f_i^*/f_i$ and $k_j = f_j^*/f_j$. If $k_i = k_j$, then we obtain

$$|F_{o,cal,h}^*|^2 = \frac{\sum_i^N f_i^{*2}}{\sum_i^N f_i^2}|F_{obs,h}|^2 \tag{15}$$

As discussed above, f_i^* can coincide with f_i in the low θ range when δ_i is appropriately set. This implies that in such cases $k_i = k_j = 1$ and Eq. (15) is valid. At high θ angle, $k_i = k_j$ seems to be a prerequisite hard to meet. Nevertheless, we note that Eq. (15) agrees well with results derived from Wilson distribution (Wilson, 1942), which is described by

$$<|F_{obs,h}|^2> = K'\sum_i^N f_i^2 \tag{16}$$

where $<|F_{obs,h}|^2>$ is the averaged $|F_{obs,h}|^2$ over a given range of $\sin^2\theta$, and K' is a scale factor independent of θ. Eq. (16) is

https://doi.org/10.1017/S0885715617001154 Published online by Cambridge University Press

valid when $\lambda/\sin\theta$ is small compared with the inter-atomic distances. Similarly, the following equation

$$<\left|F^*_{\mathrm{o,cal,h}}\right|^2> = K' \sum_{\mathrm{i}}^{N} f_{\mathrm{i}}^{*2} \tag{17}$$

is valid when $\lambda/\sin\theta$ is small. Combining Eqs (16) and (17), we obtain

$$<\left|F^*_{\mathrm{o,cal,h}}\right|^2> = \frac{\sum_{\mathrm{i}}^{N} f_{\mathrm{i}}^{*2}}{\sum_{\mathrm{i}}^{N} f_{\mathrm{i}}^{2}} <\left|F_{\mathrm{obs,h}}\right|^2> \tag{18}$$

Although the validity of Eq. (18) does not mean that Eq. (15) is valid, it implies that estimating $\left|F^*_{\mathrm{o,cal,h}}\right|^2$ using Eq. (15) is statistically reasonable.

Combining Eqs (3) and (15), we obtain

$$\begin{aligned} R_{\mathrm{Tian}} &= \frac{\sum_{\mathrm{h}} \left||F^*_{\mathrm{o,cal,h}}| - |F^*_{\mathrm{cal,h}}|\right|}{\sum_{\mathrm{h}} |F^*_{\mathrm{o,cal,h}}|} \\ &= \frac{\sum_{\mathrm{h}} \left|\sqrt{\frac{\sum_{\mathrm{i}}^{N} f_{\mathrm{i}}^{*2}}{\sum_{i}^{N} f_{i}^{2}} \left|F_{\mathrm{obs,h}}\right|^2} - |F^*_{\mathrm{cal,h}}|\right|}{\sum_{\mathrm{h}} \sqrt{\frac{\sum_{\mathrm{i}}^{N} f_{\mathrm{i}}^{*2}}{\sum_{\mathrm{i}}^{N} f_{\mathrm{i}}^{2}} \left|F_{\mathrm{obs,h}}\right|^2}} \end{aligned} \tag{19}$$

Note that EDM applied to both the correct and the current electron density models should be the same in Eq. (3). In Eq. (19), we actually have assumed that the multiplier applied to modify the correct electron density model is $\sum_{\mathrm{i}}^{N} V_{\mathrm{i}}(r_{\mathrm{i}})$ by taking $\sqrt{\left(\sum_{\mathrm{i}}^{N} f_{\mathrm{i}}^{*2} / \sum_{\mathrm{i}}^{N} f_{\mathrm{i}}^{2}\right)\left|F_{\mathrm{obs,h}}\right|^2}$ as the approximation of $|F^*_{\mathrm{o,cal,h}}|$. Therefore, the current electron density model should be modified by multiplying it with $\sum_{\mathrm{i}}^{N} V_{\mathrm{i}}(r_{\mathrm{i}})$. Unfortunately, $\sum_{\mathrm{i}}^{N} V_{\mathrm{i}}(r_{\mathrm{i}})$ contains the atomic position r_{i} and cannot be determined prior to the structure determination. An alternative way to modify the current electron density model is multiplying it with

$$W = \begin{cases} 1, & \rho \geq \delta_{\mathrm{cryst}} \\ 0, & \rho < \delta_{\mathrm{cryst}} \end{cases}$$

where δ_{cryst} is a real number. Clearly, $|F^*_{\mathrm{cal,h}}|$ varies with both the trial phase set and δ_{cryst}. When the modified current electron density map $\rho^* = \rho \mathrm{W}$ is close to the modified correct electron density map $\rho^*_{\mathrm{o}} = \sum_{\mathrm{i}}^{N} \rho_{\mathrm{i}}(r_{\mathrm{i}}) \times V_{\mathrm{i}}(r_{\mathrm{i}})$, R_{Tian} calculated with Eq. (19) reaches the minimum, ideally, zero.

III. DISCUSSION

A. The validity of the method to estimate $\left|F^*_{\mathrm{o,cal,h}}\right|$

The validity of the method to estimate $\left|F^*_{\mathrm{o,cal,h}}\right|$ using Eq. (15) is tested on simulated diffraction data, which are obtained from the known structure of C_6Br_6 (Wu *et al.*, 2004). Electron density maps with the resolution of 1.0, 1.5, and 2.5 Å are calculated by IFT. At each data resolution, δ_{C} and δ_{Br} are appropriately set to satisfy the condition that f_{i}^* coincides with f_{i} approximately in the low $\sin\theta/\lambda$ region. Then the theoretical values of $\left|F^*_{\mathrm{o,cal,h}}\right|$ are calculated using Eq. (14) and shown in Figure 3 as red crosses. The estimated values of $\left|F^*_{\mathrm{o,cal,h}}\right|$ are calculated using Eq. (15) and presented in Figure 3 as

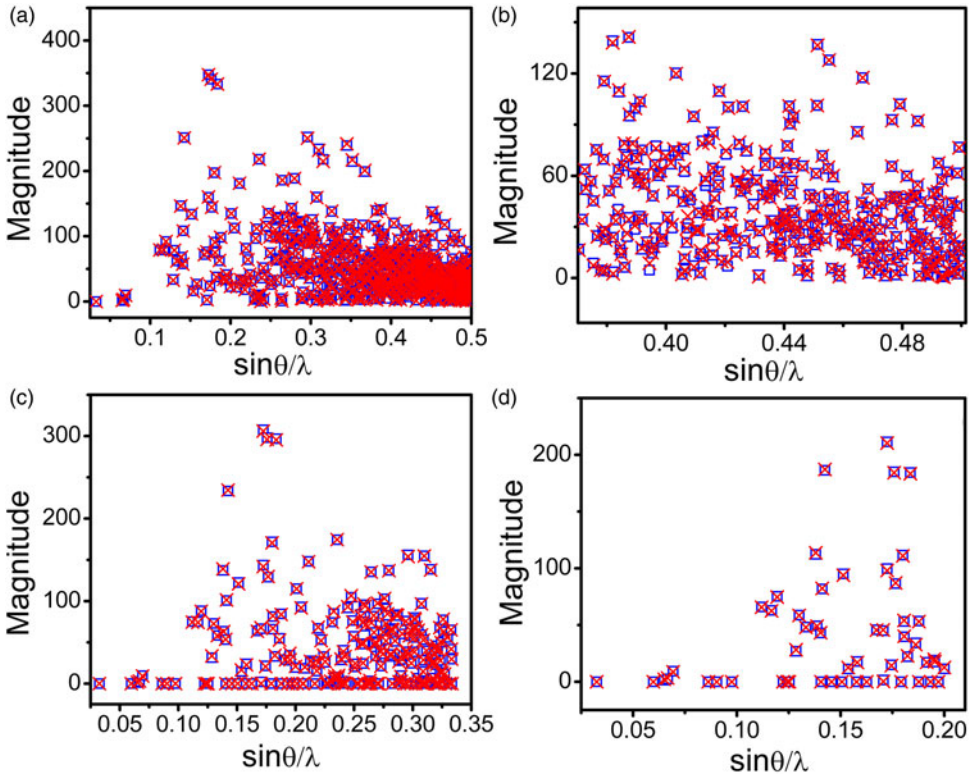


Figure 3. (color online) $\left|F^*_{\mathrm{o,cal,h}}\right|$ calculated using Eq. (14) (red cross) and (15) (blue square) of C_6Br_6. (a) Data resolution: 1.0 Å, $\delta_{\mathrm{C}} = 0.0018$ e/Å^3, $\delta_{\mathrm{Br}} = 0.125$ e/Å^3; (b) is an enlarged view of the bottom-right region of (a). (c) Data resolution: 1.5 Å, $\delta_{\mathrm{C}} = 0.003$ e/Å^3, $\delta_{\mathrm{Br}} = 0.07$ e/Å^3; (d) Data resolution: 2.5 Å, $\delta_{\mathrm{C}} = 0.0024$ e/Å^3, $\delta_{\mathrm{Br}} = 0.021$ e/Å^3.

https://doi.org/10.1017/S0885715617001154 Published online by Cambridge University Press

blue squares. As shown in Figure 3, at each data resolution, the estimated $\left|F^{*}_{o,cal,h}\right|$ values agree well with the theoretical ones. To derive Eq. (15) from Eq. (14), we have assumed that atomic scattering factors of different atoms vary by the same proportion as a result of EDM. It seems to be a condition hard to be satisfied. Nevertheless, tests on simulated diffraction data of C_6Br_6 indicated that Eq. (15) is valid in estimating $\left|F^{*}_{o,cal,h}\right|$ even there are atoms with significantly different scattering power, such as C and Br, in the unit cell. On the other hand, we also would like to explicitly state that Eq. (15) can only be applied to estimate the structure amplitudes of an electron density map modified in a specific way, which has been described in details in Section II C.

B. The validity of R_{Tian} in identifying the correct phase set

The factor R_{Tian} is aimed to identifying the correct phase set from a great number of possible phase sets assigned to the observed structure amplitudes. The capability of R_{Tian} in evaluating the trial phase sets is tested on simulated diffraction data of known structure $C_{252}H_{326}O_{19}$ (Czugler *et al.*, 2003). Two data sets at the resolution of 1.0 and 2.5 Å, respectively, are generated. Four hundred iterative EDM cycles are performed to retrieve the correct electron density model from the simulated data set at the resolution of 1.0 Å. The initial electron density models are constructed by FT using the "observed" structure amplitudes and random phases. In each following iterative cycle, electron densities greater than δ_{cryst} are retained, and the others are set to zero. The variation of R and R_{Tian} factors with the iterative EDM cycles is shown in Figure 4(a). Both R and R_{Tian} drop drastically at the early stage and then keep stable in the following iterative cycles. The convergent R and R_{Tian} values coincide with each other, and both are much higher than the expected values for the correct model. It is noteworthy that the expected value of R_{Tian} is lower than that of R factor. In this case, the correct model cannot be retrieved by iterative EDM cycles, as indicated by the high convergent R and R_{Tian} values. Then a partial structure including 10 correctly located carbon atoms is build up, and a phase set is derived from this partial structure. This phase set is combined with the "observed" structure amplitudes to generate the initial electron density model. Starting with this initial electron density model, iterative EDM cycles converge to the correct electron density map. The evolution of R and

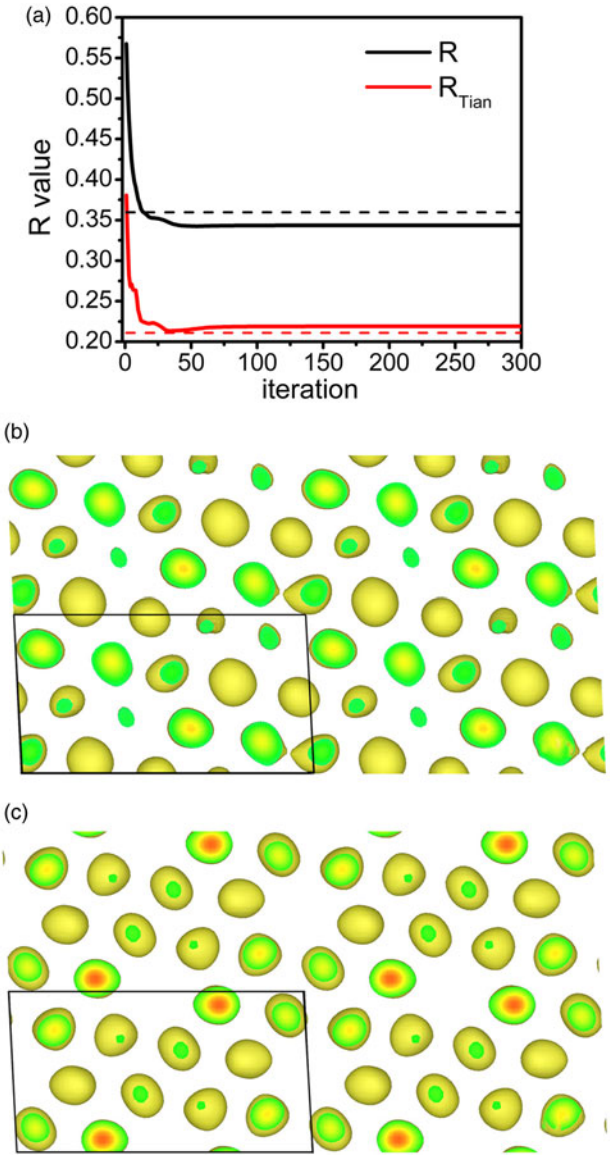


Figure 5. (color online) (a) The variation of R and R_{Tian} factor with the iterative EDM cycles at the resolution of 2.5 Å. δ_{cryst} is set to zero in the iterative EDM cycles. Black and red dotted lines indicate the R and R_{Tian} values for the correct electron density map, respectively. (b) The electron density model corresponding to the minimum value of R_{Tian}. (c) The correct electron density map of C_6Br_6 at the resolution of 2.5 Å. Both (b) and (c) are projections along [010] direction. Only electron densities above 2.5 e/Å^3 are illustrated for clarity.

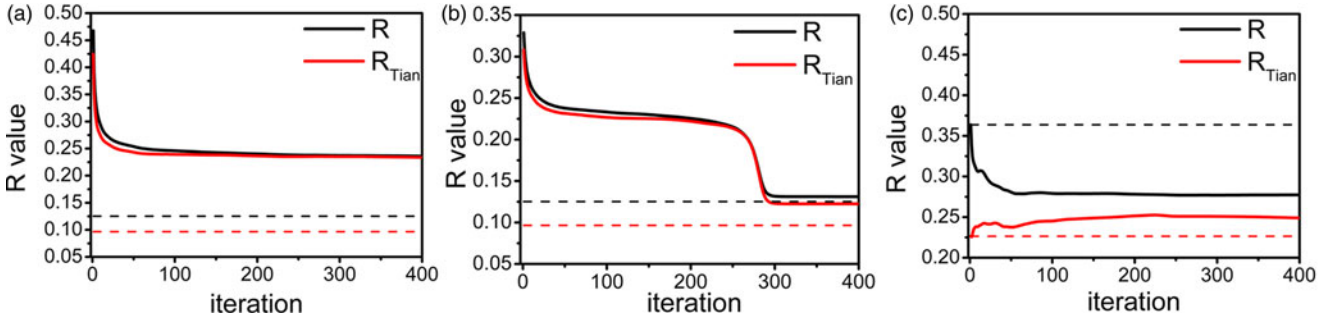


Figure 4. (color online) The variation of R and R_{Tian} factors with the iterative EDM cycles at the resolution of 1.0 (a, b) and 2.5 Å (c). The initial electron density model in (a) is generated using random phases while in (b) it is generated using phase set derived from the partial structure. In (c) the correct electron density map of $C_{252}H_{326}O_{19}$ is used as the initial model. In each panel, the target values of R and R_{Tian} corresponding to the correct electron density maps are shown as black and red horizontal dotted lines, respectively. δ_{cryst} is set to zero in all iterative EDM cycles.

https://doi.org/10.1017/S0885715617001154 Published online by Cambridge University Press

R_{Tian} factors in this attempt is shown in Figure 4(b). In this case, the convergent R value approaches to the expected value for the correct model, while the convergent R_{Tian} value is still significantly higher than the expected value for the correct model. The discrepancy between the convergent and the expected value of R_{Tian} is because of the difference between the convergent and the correct electron density models. This indicates that R_{Tian} is more sensitive than R to the variation of electron density models. At the resolution of 2.5 Å, we use the correct phase set to establish the initial electron density model for the iterative EDM cycles. The evolution of R and R_{Tian} in this case is shown in Figure 4(c). The R factor decreases considerably with the iterative cycles, while R_{Tian} increases significantly in the same process. This indicates clearly that R factor is no longer valid at this resolution, but R_{Tian} is still valid for identifying the correct phase set. In addition, we note that R_{Tian} is lower than R for the correct model at both data resolution, and the latter increases faster than the former when the data resolution deteriorates.

Tests are also performed on the simulated data of C_6Br_6 at the resolution of 2.5 Å. The variation of R_{Tian} and R with the iterative EDM cycles is shown in Figure 5(a). The initial electron density model was generated by IFT using the "observed" structure amplitudes and the random phases. As discussed in Section II B, the R factor is no longer valid in identifying the correct electron density model in this case. It is interesting to note that R_{Tian} reaches a minimum during the iterative process, which is close to the expected value for the correct electron density model. The electron density model corresponding to the minimum value of R_{Tian} is presented in Figure 5(b). Comparison between this model and the correct electron density map [shown in Figure 5(c)] reveals that it is a close resemblance to the correct map. This example further confirms that R_{Tian} is still valid in identifying the correct electron density model at a resolution of 2.5 Å. Moreover, R_{Tian} seems to be more sensitive than R factor because it varies more significantly than R, especially when an approximately correct model has been obtained.

C. Application of R_{Tian} in structure determination

In Section III B, we have checked the capability of R_{Tian} in identifying correct phase set by investigating its variation with the iterative EDM cycles. This does not mean that the R_{Tian} factor is designed for iterative EDM algorithms. Usually, iterative EDM algorithms work only in cases where the high-resolution diffraction data are available. In such cases, R_{Tian} is only slightly different from R factor numerically, and R factor can be more readily calculated than R_{Tian}. R_{Tian} factor is independent of iterative EDM algorithms. For an arbitrary trial phase set assigned to the observed structure amplitudes, R_{Tian} can be estimated approximately. The correct phase set is supposed to have the lowest R_{Tian} value. Tests on simulated diffraction data indicate that R_{Tian} is valid even at the resolution of 2.5 Å. Clearly, since the number of all possible phase sets is tremendous, it is not realistic to estimate R_{Tian} values for all possible phase sets and then pick out the phase set with the lowest R_{Tian} value. Potentially, R_{Tian} may be applied as a goal function of global optimization algorithms for retrieving the electron density map from data set of poor resolution.

Acknowledgements

This work was financially supported by the National Natural Science Foundation of China (Grant Nos. 11574060 and 11474281). HL and MH are very grateful to Guang-Lian Tian for her consistent support and encouragement.

Burla, M. C., Carrozzini, B., Cascarano, G. L., Giacovazzo, C. and Polidori, G. (**2017**). "MPF, a multipurpose figure of merit for phasing procedures," Acta Crystallogr. A **73**, 69–76.

Czugler, M., Weber, E., Parkanyi, L., Korkas, P. P. and Bombicz, P. (**2003**). "Supermolecular [6]chochin and 'Big mac' made from chiral piedfort assemblies," Chem. A Eur. J. **9**, 3741–3747.

Karle, J. and Hauptman, H. A. (**1950**). "The phases and magnitudes of the structure factors," Acta Crystallogr. A **3**, 181–187.

Li, H., He, M. and Zhang, Z. (**2015**). "Image definition evaluation functions for X-ray crystallography: a new perspective on the phase problem," Acta Crystallogr. A **71**, 526–533.

Oszlányi, G. and Sütő, A. (**2004**). "*Ab initio* structure solution by charge flipping," Acta Crystallogr. A **60**, 134–141.

Oszlányi, G. and Sütő, A. (**2008**). "The charge flipping algorithm," Acta Crystallogr. A **64**, 123–134.

Wilson, A. J. C. (**1942**). "Determination of absolute from relative X-Ray intensity data," Nature **150**, 151–152.

Woolfson, M. M. (**1987**). "Direct methods – from birth to maturity," Acta Crystallogr. A **43**, 593–612.

Woolfson, M. M. and Fan, H. F. (**1995**). *Physical and non-Physical Methods of Solving Crystal Structure* (Cambridge University Press, Cambridge).

Wu, J. S., Spence, J. C. H., O'Keeffe, M. and Groy, T. L. (**2004**). "Application of a modified Oszlányi and Sütő *ab initio* charge-flipping algorithm to experimental data," Acta Crystallogr. A **60**, 326–330.

https://doi.org/10.1017/S0885715617001154 Published online by Cambridge University Press