Solution of the effects of twinning in femtosecond X-ray protein nanocrystallography *

ZHOU Liang(周亮) LIU Peng(刘鹏) DONG Yu-Hui(董宇辉)[1)]

Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, China

Abstract: With the development of the XFEL (X-ray free electron laser), high quality diffraction patterns from nanocrystals have been achieved. The nanocrystals with different sizes and random orientations are injected to the XFEL beams and the diffraction patterns can be obtained by the so-called "diffraction-and-destruction" mode. The recovery of orientations is one of the most critical steps in reconstructing the 3D structure of nanocrystals. There is already an approach to solve the orientation problem by using the automated indexing software in crystallography. However, this method cannot distinguish the twin orientations in the cases of the symmetries of Bravais lattices higher than the point groups. Here we propose a new method to solve this problem. The shape transforms of nanocrystals can be determined from all of the intensities around the diffraction spots, and then Fourier transformation of a single crystal cell is obtained. The actual orientations of the patterns can be solved by comparing the values of the Fourier transformations of the crystal cell on the intersections of all patterns. This so-called "multiple-common-line" method can distinguish the twin orientations in the XFEL diffraction patterns successfully.

Key words: coherent X-ray diffractive imaging, XFEL, oversampling technique, protein nanocrystallography, multiple-common-lines

PACS: 61.05.cp, 61.46.Hk, 87.59-e **DOI:** 10.1088/1674-1137/37/2/028101

1 Introduction

It is highly promising to determine the structure of nanocrystals using the coherent X-ray diffractive imaging (CXDI) technique with the development of the XFEL. The XFEL provides ultra-short and extremely intense coherent X-ray pulses with a peak brilliance 10–11 orders of magnitude higher than the synchrotron sources [1]. It has been proposed that the diffraction patterns may be obtained from the nanocrystals by using these XFEL pulses under an entirely new radiation damage mechanism [2, 3]. Recently, the first femtosecond X-ray protein nanocrystallography diffraction experiment has been carried out at the Linac Coherent Light Source. It has proved that high-quality nanocrystal diffraction data can be obtained by using the XFEL [4].

Due to the finite size of nanocrystals, there are many mid-Bragg peaks in the diffraction patterns of nanocrystals. The mid-Bragg peaks can solve the phase problem by using the oversampling technique [5, 6], which requires that the regular three-dimensional oversampled structure factors are obtained from tens of thousands of snapshot patterns. There is already a method to solve this problem by determining the mean shape transform [7], but this method requires the absolute orientations of each snapshot pattern. Presently, the orientations of diffraction patterns of nanocrystals are determined by the automated indexing software [4, 8]. The automated indexing algorithm only uses the positions of the Bragg peaks but not their intensities [9], if the symmetries of the Bragg lattices are higher than the point groups, such as the case of space group P63 used in the first femtosecond X-ray protein nanocrystallography diffraction experiments, the twin-related orientations cannot be distinguished [4, 10].

Here we describe a new approach to solve the problem. This method not only uses the positions of Bragg peaks but also uses the intensities of all Bragg peaks and mid-Bragg peaks. Our approach mainly consists of the following steps. The first one is to obtain a collection of possible orientations related by the symmetry of the lattice using the automated indexing software to each pattern; these orientations are classified as a set which consists of the actual orientation and twin-related orientation of the pattern. The second one is to calculate the shape transform (size) of nanocrystal and the Fourier

Received 19 March 2012, Revised 30 April 2012

∗ Supported by National Natural Science Foundation of China (10979005) and National Basic Research Program of China (2009CB918600)

1) E-mail: dongyh@ihep.ac.cn

transforms of the crystal cell (structure factor) in every orientation in the set. The third one is to determine the actual orientations of each pattern and remove the twinning ambiguity by comparing the values of the Fourier transformations of the crystal cell on the intersections of all patterns.

2 Determination of the shape transform and actual orientation of snapshot diffraction pattern

For a parallelepiped nanocrystal with random size and orientation, illuminated by a XFEL pulse with wavelength λ, its snapshot diffraction patterns can be expressed as:

$$I(n_1,n_2)=I(\boldsymbol{k})=J_0 r_e^2 |F(\boldsymbol{k})|^2 |G(\boldsymbol{k})|^2 \Omega_{\rm pix}(\boldsymbol{k}), \quad (1)$$

where $I(n_1,n_2)$ stands for the diffracted photon flux in detector pixel (n_1,n_2); $\boldsymbol{k}$ stands for the scattering vector corresponding to the detector pixel, mainly determined by the wavelength, pixel position, the sample-to-detector distance, and orientation of the nanocrystal; r_e is the classical electron radius; J_0 is the flux density of incoming X-ray pulse, and $\Omega_{\rm pix}(\boldsymbol{k})$ is the solid angle of the detector pixel; $F(\boldsymbol{k})$ is the structure factor; and $G(\boldsymbol{k})$ is the shape transform of nanocrystal.

According to Eq. (1), we can find that the diffracted photon flux is mainly determined by the structure factor $F(\boldsymbol{k})$ and the shape transform $G(\boldsymbol{k})$. The square of the modulus of $F(\boldsymbol{k})$ changes gently between the neighboring Bragg peaks, but the square of modulus of $G(\boldsymbol{k})$ changes rapidly between the neighboring Bragg peaks, as shown in Fig. 1. Therefore, we can conclude that the diffraction intensity difference between the neighboring pixels in the snapshot diffraction pattern is mainly caused by the shape transform, the number of unit cells $nx\times ny\times nz$ can be determined by calculating the average difference between each neighboring pixel.

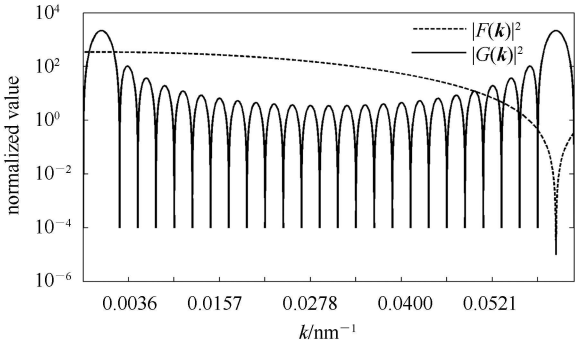

Fig. 1. The square of the modulus of $F(\boldsymbol{k})$ and $G(\boldsymbol{k})$ plotted against the scattering vector $\boldsymbol{k}$ between the neighboring Bragg peaks, where the squares of modulus of $F(\boldsymbol{k})$ and $G(\boldsymbol{k})$ are normalized, respectively, the Miller indices of neighboring Bragg peaks are (000) and (001).

To the $n^{\rm th}$ pattern, it is reasonable to assume that we have obtained a collection of orientations related by the symmetry of the lattice using the automated indexing software. These orientations are classified as set n. Then the square of modulus of shape transforms $G(\boldsymbol{k'})$ of the $n^{\rm th}$ pattern in pixel (n_1,n_2) corresponding to the orientation (α, β, γ), which is one of the orientations in the set n, can be expressed as:

$$G(n_1,n_2) = |G(\boldsymbol{k'})|^2 = \frac{\sin^2 0.5\times nx\times \boldsymbol{k'}\cdot\boldsymbol{a}}{\sin^2 0.5\times \boldsymbol{k'}\cdot\boldsymbol{a}} \times \frac{\sin^2 0.5\times ny\times \boldsymbol{k'}\cdot\boldsymbol{b}}{\sin^2 0.5\times \boldsymbol{k'}\cdot\boldsymbol{b}} \frac{\sin^2 0.5\times nz\times \boldsymbol{k'}\cdot\boldsymbol{c}}{\sin^2 0.5\times \boldsymbol{k'}\cdot\boldsymbol{c}}, \quad (2)$$

where (nx, ny, nz) is the number of unit cells, $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$ are the real-space unit cell vectors, $\boldsymbol{k'}$ stands for the scattering vector in the detector pixel corresponding to the orientation (α, β, γ). The average difference diff between the diffracted photon flux and the square of modulus of shape transforms can be expressed as:

$$\text{diff}=\frac{\sum\limits_{n_1,n_2}\left(\left(\frac{I(n_1+1,n_2)}{I(n_1,n_2)}-\frac{G(n_1+1,n_2)}{G(n_1,n_2)}\right)+\sum\limits_{a=-1}^{1}\left(\frac{I(n_1+a,n_2+1)}{I(n_1,n_2)}-\frac{G(n_1+a,n_2+1)}{G(n_1,n_2)}\right)\right)}{\sum\limits_{n_1,n_2}\left(1+\sum\limits_{a=-1}^{1}1\right)}. \quad (3)$$

An exhaustive search of (nx, ny, nz) is performed for all nanocrystal sizes. Theoretically, the average difference diff should be minimal in the actual nanocrystal size and the actual orientation. Therefore, if there is a global minimum of diff in one nanocrystal size (nx, ny, nz), this size will be considered as the correct size corresponding to the orientation (α, β, γ), then the shape transform in this orientation is obtained. Due to the errors in the orientations determined by the automated indexing software [10], only the low resolution data in the snapshot diffraction patterns are used to determine the shape transforms. The reason is that the $|G(\boldsymbol{k})|^2$ is sensitive to the orientation errors, especially at high resolution. In addition, the diffraction intensities are integrated by binning pixels; the binning can reduce the effects of saturated pixels and the noises. However, here we must note that the binning pixels can't change the number of mid-Bragg peaks; in this paper binning 11×11

pixels into 1 pixel. To further reduce the error, only the pixels whose photon counts are higher than 10 are used to calculate the average difference.

After obtaining the shape transforms in every orientation in the set n, the corresponding structure factor $|F(\boldsymbol{k})|$ in these orientations can be easily obtained according to Eq. (1). For convenience, we select the 1st pattern as the reference pattern, $|F(\alpha_1,\beta_1,\gamma_1,1)|$ is used to represent the reference structure factor, $(\alpha_1,\beta_1,\gamma_1)$ is the actual orientation of the 1st pattern; $|F(\alpha,\beta,\gamma,n)|$ is used to represent the structure factor of $n^{\rm th}$ pattern in the orientation (α, β, γ), where (α, β, γ) represents one of the possible orientations in the set n. Both the $|F(\alpha_1,\beta_1,\gamma_1,1)|$ and $|F(\alpha,\beta,\gamma,n)|$ are interpolated onto regular girds in reciprocal space using the nearest neighbor interpolation. There will be an intersection line between the $|F(\alpha_1,\beta_1,\gamma_1,1)|$ and $|F(\alpha,\beta,\gamma,n)|$, theoretically the value of $|F(\alpha_1,\beta_1,\gamma_1,1)|$ and $|F(\alpha,\beta,\gamma,n)|$ on the intersection line will be equal if the actual orientation of $n^{\rm th}$ patter is (α, β, γ), since the structure factors belong to the same crystal cell even calculating from nanocrystals with different sizes.

First we use the intersection between two patterns (common line), where the structure factors of the crystal cell should be equal, to determine the actual orientations of these snapshot diffraction patterns [11, 12]. The actual orientation of the $n^{\rm th}$ pattern can be solved by scanning the orientation (α, β, γ) in the set n and calculating the average difference among regular grids near the intersection line between $|F(\alpha_1,\beta_1,\gamma_1,1)|$ and $|F(\alpha,\beta,\gamma,n)|$. Rfactor($\alpha$, β, γ, n) is used to represent the average difference, it can be expressed as:

$$\begin{aligned}&\text{Rfactor}(\alpha,\beta,\gamma,n)\\ &=\frac{\sum\limits_{hkl}\dfrac{||F_{hkl}(\alpha,\beta,\gamma,n)|-|F_{hkl}(\alpha_1,\beta_1,\gamma_1,1)||}{|F_{hkl}(\alpha_1,\beta_1,\gamma_1,1)|}}{\sum\limits_{hkl}1},\end{aligned}\quad(4)$$

where the integer set hkl is the indices of regular girds near the intersection line. If there is a minimum of Rfactor(α, β, γ, n) in one of the orientations in the set n, then this orientation is considered as the determined actual orientation of the $n^{\rm th}$ pattern. The errors of determining orientation by a single common line is big, since the number of regular grids near the common line which are used to calculate Rfactor(α, β, γ, n) is small.

Therefore we apply the so-called "multiple-common-lines" method to re-determine the actual orientation of the snapshot diffraction pattern more accurately. The concept is that two arbitrary patterns in reciprocal space intersect on a line, not only the original reference pattern (the 1st pattern) and the $n^{\rm th}$ pattern. Thus not only the original reference pattern but also all the other patterns can be used to determine the orientation of $n^{\rm th}$ pattern. The orientations of each pattern can be re-determined using the following steps.

(1) The $n^{\rm th}$ pattern whose orientation is to be re-determined is interpolated onto regular grids in reciprocal space in orientation (α, β, γ), which is one of the orientations in the set n, $|F_{hkl}(\alpha,\beta,\gamma,n)|$ is used to represent the 3D data set.

(2) The new reference patterns, which consist of all patterns except the $n^{\rm th}$ pattern, are also interpolated onto the regular grids in reciprocal space. Another 3D data set $\langle|F_{hkl}|\rangle$ can be assembled from these patterns. The initial orientations of the new reference patterns are given by the single common-line method.

(3) Scanning the orientation (α, β, γ) in the set n and calculating the average difference among regular grids near the intersection lines between $|F_{hkl}(\alpha,\beta,\gamma,n)|$ and $\langle|F_{hkl}|\rangle$, multi-Rfactor (α, β, γ, n) is used to represent the average difference, it can be expressed as:

$$\begin{aligned}&\text{multi-Rfactor}(\alpha,\beta,\gamma,n)\\ &=\frac{\sum\limits_{hkl}\dfrac{||F_{hkl}(\alpha,\beta,\gamma,n)|-\langle|F_{hkl}|\rangle|}{\langle|F_{hkl}|\rangle}}{\sum\limits_{hkl}1}.\end{aligned}\quad(5)$$

If there is a minimum of multi-Rfactor (α, β, γ, n) in one of the orientations in the set n, then this orientation is considered as the re-determined actual orientation of the $n^{\rm th}$ pattern.

(4) Repeat Step 1–3 for all the patterns except the 1st pattern.

(5) Repeat Step 1–4 until the orientations of all patterns reach convergence. In this paper after 3 cycles the orientations of each pattern no longer change.

The multiple-common-lines method can significantly reduce the azimuth error relative to single common-line method. The main reason is that if there are enough patterns, the whole diffraction information of the pattern will be used to determine its own orientation.

3 Simulations

The simulated patterns can be obtained according to Eq. (1). The parameters are basically similar to the first femtosecond X-ray protein nanocrystallography diffraction experiment carried out at the LCLS [4]. We download the membrane protein (1JB0) whose space group is $P6_3$ and cell constants of a=b=28.1 nm, c=16.5nm from the Protein Data Bank, the orientations and the number of unit cells (nx, ny, nz) are randomly generated with a uniform distribution. The length of nanocrystal ranges from 300 nm to 500 nm. Then we simulate an XFEL pulse with λ=6.9 Å and 1×10^{12} photons per

pulse focused to 3 μm. The detector contains 1024×1024 75 μm pixels, for a low resolution pattern the sample-to-detector distance is 564 mm, for a high resolution pattern the sample-to-detector distance is 68 mm. We simulate a series of high resolution patterns and low resolution patterns containing Poisson noise with these parameters respectively and randomly choose 40 patterns.

Assuming we have obtained a collection of possible orientations related by the symmetry of the lattice using the automated indexing software with an RMS error of 0.06 degrees to each pattern [8]. For the $P6_3$ space group, only one twin possibility arises, an ambiguity existing in that each pattern could correspond to one of two possible orientations. The two possible orientations are related by the rotation of 180 degrees about the $h=k$ axis. This operation brings the lattice into coincidence with itself, but not the associated structure factors. It converts index *hkl* into *kh-l* [8]. Then we employ our method to solve the problem. First, we determine the shape transforms of each low resolution patterns in the two possible orientations. The result is shown in Fig. 2, Fig. 3 and Table 1. As shown in Fig. 2, there is indeed a minimum of diff in the size (11, 12, 19), which is actual nanocrystals size, in the actual orientation. As shown in Fig. 3, there is a minimum of diff in another size (12, 11, 19) in the twin-related orientation, this is caused by the

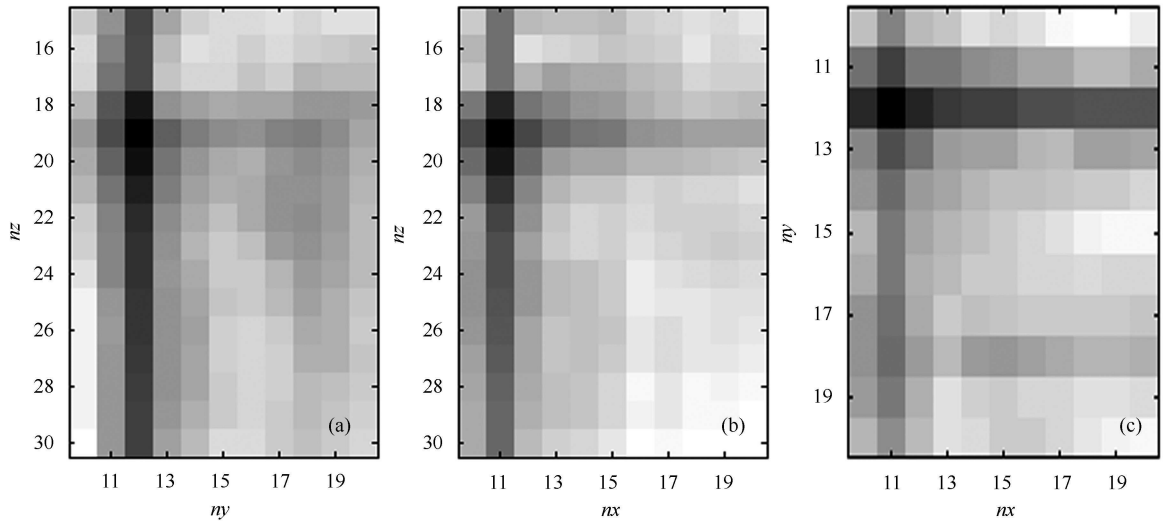

Fig. 2. The diff for pattern 5 in the actual orientation, whose actual nanocrystal size is (11, 12, 19). (a) The diff along the *ny*-*nz* plane, where nx=11. (b) The diff along the *nx*-*nz* plane, where ny=12. (c) The diff along the *nx*-*ny* plane, where nz=19.

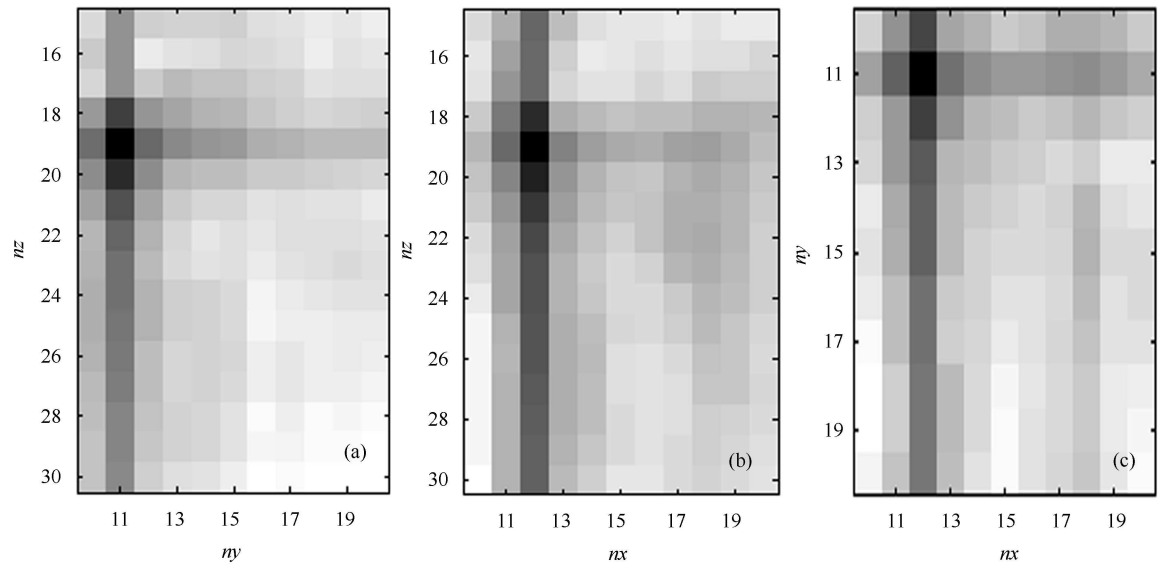

Fig. 3. The diff for pattern 5 in the twin-related orientation. (a) The diff along the *ny*-*nz* plane, where nx=12. (b) The diff along the *nx*-*nz* plane, where ny=11. (c) The diff along the *nx*-*ny* plane, where nz=19.

Table 1. The result for determining the shape transforms in the two possible orientations.

pattern	actual size	recovered in actual orientation	recovered in twin-related orientation
pattern 1	(11,11,30)	(11,11,30)	(11,11,30)
pattern 2	(14,14,22)	(14,14,22)	(14,14,22)
pattern 3	(17,15,29)	(17,15,29)	(15,17,29)
pattern 4	(14,15,29)	(14,15,29)	(15,14,29)
pattern 5	(11,12,19)	(11,12,19)	(12,11,19)
pattern 6	(18,15,27)	(18,15,27)	(15,18,27)
pattern 7	(15,13,29)	(15,13,29)	(13,15,29)
pattern 8	(13,12,21)	(13,12,21)	(12,13,21)
pattern 9	(14,17,18)	(14,17,18)	(17,14,18)
pattern 10	(16,16,28)	(16,16,28)	(16,16,28)
pattern 11	(18,11,24)	(18,11,24)	(11,18,24)
pattern 12	(14,16,24)	(14,16,24)	(16,14,24)
pattern 13	(18,12,21)	(18,12,21)	(12,18,21)
pattern 14	(13,11,22)	(13,11,22)	(11,13,22)
pattern 15	(18,16,27)	(18,16,27)	(16,18,27)
pattern 16	(18,15,19)	(18,15,19)	(15,18,19)
pattern 17	(15,14,25)	(15,14,25)	(14,15,25)
pattern 18	(16,11,20)	(16,11,20)	(11,16,20)
pattern 19	(17,11,19)	(17,11,19)	(11,17,19)
pattern 20	(18,13,29)	(18,13,29)	(13,18,29)
pattern 21	(11,11,23)	(11,11,23)	(11,11,23)
pattern 22	(15,12,24)	(15,12,24)	(12,15,24)
pattern 23	(12,18,24)	(12,18,24)	(18,12,24)
pattern 24	(18,12,28)	(18,12,28)	(12,18,28)
pattern 25	(16,15,21)	(16,15,21)	(15,16,21)
pattern 26	(17,18,29)	(17,18,29)	(18,17,29)
pattern 27	(11,11,26)	(11,11,26)	(11,11,26)
pattern 28	(15,16,23)	(15,16,23)	(16,15,23)
pattern 29	(14,12,22)	(14,12,22)	(12,14,22)
pattern 30	(13,12,25)	(13,12,25)	(12,13,25)
pattern 31	(16,17,20)	(16,17,20)	(17,16,20)
pattern 32	(15,13,24)	(15,13,24)	(13,15,24)
pattern 33	(18,13,24)	(17,13,24)	(13,17,24)
pattern 34	(11,18,29)	(11,18,27)	(18,11,27)
pattern 35	(17,11,27)	(17,11,27)	(11,17,27)
pattern 36	(16,18,25)	(16,18,25)	(18,16,25)
pattern 37	(13,16,22)	(13,16,22)	(16,13,22)
pattern 38	(17,13,26)	(17,13,26)	(13,17,26)
pattern 39	(14,17,21)	(14,17,21)	(17,14,21)
pattern 40	(17,15,19)	(17,15,19)	(15,17,19)

rotation of 180 degrees about the $h=k$ axis. As shown in Table 1, the actual shape transforms of most snapshot diffraction patterns can be determined in the actual orientations, except the diffraction pattern 33 and 34, but this error is not large.

After getting the shape transforms in the two possible orientations, we then employ the so-called "multiple-common-lines" method to determine the actual orientation of each pattern. The result for determining the actual orientation is shown in Fig. 4 and Fig. 5. As shown in Fig. 4, the actual orientations of most patterns can be determined by a single common line. As shown in Fig. 5, the actual orientations of all patterns can be determined using multiple-common-lines methods.

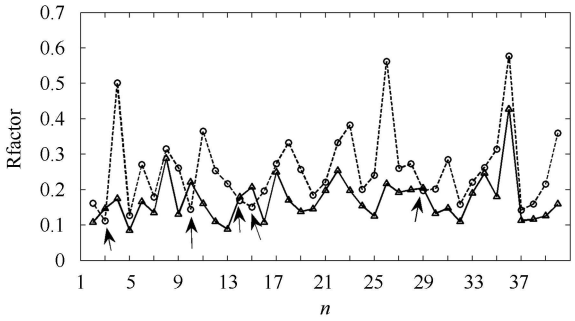

Fig. 4. The $F(\alpha_1,\beta_1,\gamma_1,1)$ is chosen as reference structure factor, the solid line represents Rfactor($\alpha_{\rm actual}$,$\beta_{\rm actual}$,$\gamma_{\rm actual}$,n) in actual orientations of each pattern, and the dotted line represents Rfactor($\alpha_{\rm twin}$,$\beta_{\rm twin}$,$\gamma_{\rm twin}$,n) in twin-related orientations of each pattern, where (α_1,β_1,γ_1) represents the actual orientation of the 1st pattern. Except the n=3, 10, 14, 15, 29, all Rfactor($\alpha_{\rm actual}$,$\beta_{\rm actual}$,$\gamma_{\rm actual}$,n) are less than Rfactor($\alpha_{\rm twin}$,$\beta_{\rm twin}$,$\gamma_{\rm twin}$,n), which means that the actual orientations of most patterns have been determined.

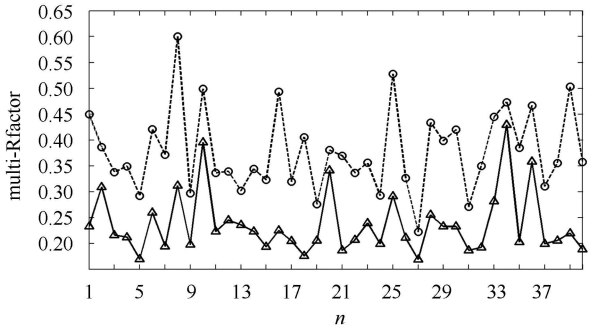

Fig. 5. The solid line represents multi-Rfactor ($\alpha_{\rm actual}$,$\beta_{\rm actual}$,$\gamma_{\rm actual}$,n), and the dotted line represents multi-Rfactor ($\alpha_{\rm twin}$,$\beta_{\rm twin}$,$\gamma_{\rm twin}$,n). All multi-Rfactor ($\alpha_{\rm actual}$,$\beta_{\rm actual}$,$\gamma_{\rm actual}$,n) are less than multi-Rfactor ($\alpha_{\rm twin}$,-$\beta_{\rm twin}$,$\gamma_{\rm twin}$,n), which means that the actual orientations of all patterns have been determined.

4 Conclusions

In this paper, we neglect beam divergence and spectral width, because, for small nanocrystals and beam divergence, the crystal-size effects dominate the diffracted intensity rather than beam divergence and the spectral width [10]. In addition, we assume that the nanocrystals are perfect crystals. In summary, we propose a new method which may solve the shape transforms and actual orientations of snapshot diffraction patterns taken from nanocrystals varying in size and orientation. After that the regular three-dimensional structure factor which satisfies oversampling can be easily obtained, then the phasing problem can be solved using iterative phasing techniques and the 3D structure of the nanocrystal can be reconstructed.

References

1 Hajdu J. Current Opinion Structural Biology, 2000, **10**: 569–573
2 Neutze R, Wouts R, Hajdu J et al. Nature, 2000, **406**: 752–757
3 Caleman C, Huldt G, Ortiz C et al. ACS Nano, 2011, **5**(1): 139–146
4 Chapman H N, Hajdu J, Spence J C H et al. Nature, 2011, **470**: 73–77
5 Fienup J R. Appl. Opt., 1982, **21**(15): 2758–2769
6 MIAO J, Ishikawa T, Anderson E H et al. Phys. Rev. B, 2003, **67**(17): 174104
7 Spence J C H, Kirian R A, Chapman H N et al. Opt. Express, 2011, **19**: 2866–2873
8 Kirian R A, Weierstall U, Spence J C H et al. Opt. Express, 2010, **18**: 5713–5723
9 Leslie A G W. Acta Cryst. D, 2006, **62**: 48–57
10 Kirian R A, White T A, Spence J C H et al. Acta Cryst. A, 2011, **67**: 131–140
11 Huldt G, Szoke A, Hajdu J. Journal of Structural Biology, 2003, **144**: 219–227
12 Shneerson V L, Ourmazd A, Saldin D K. Acta Cryst. A, 2008, **64**: 303–315