

Precise improvement of ISAF reconstruction algorithm based on the computational radius of density function

Gongming Wang^{a,*} and Fan Xu^b

^a *Department of Information Art and Design, Academy of Art and Design, Tsinghua University, Haidian District, 100084, Beijing, China*

^b *Institute of Computing Technology, Chinese Academy of Science, Haidian District, 100190, Beijing, China*

Abstract. The output of icosahedral symmetry-adapted functions (ISAF) reconstruction algorithm is a virus capsid, where the density function of every outside point is theoretically zero. However, original ISAF algorithm takes reconstructed map as a solid sphere and density function of every point in the whole sphere is calculated. The density functions outside virus capsid may not be always zero, which may sometimes produce noise. In this study, Fourier shell correlation (FSC) is used for proving that density function calculation of original ISAF algorithm can reduce reconstruction precision. Subsequently, the range of virus capsid is determined with radial density function curve and the original ISAF algorithm is improved by calculating density function inside virus capsid. The experimental results on Cryo-electron microscopy (Cryo-EM) data of cytoplasmic polyhedrosis virus (CPV) indicate that the improved ISAF algorithm can generate more details of virus capsid and achieve a higher resolution.

Keywords: Single-particle reconstruction, icosahedral symmetry-adapted functions, virus capsid, density function, Fourier shell correlation

1. Introduction

During the second part of 1960s, DeRosier and Klug [1] reported the principle of reconstruction of molecule from a set of electron microscopy (EM) images. After many years of research, this method has been expanded into three major techniques: electron crystallography [2], electron tomography [3] and single particle technology [4], out of which, single-particle technology has more number of advantages. It can capture the instantaneous images of biological molecule at different states and is used for studying both symmetrical/asymmetrical molecules. Most importantly, there is no limitation of molecule size nor requirement for any crystallization [5].

*Corresponding author: Gongming Wang, Department of Information Art and Design, Academy of Art and Design, Tsinghua University, No 1 Tsinghua Park, Haidian District, Beijing, ZIP 100084, China. Tel.: +86 15001012036; Fax.: +86 10 62798933; E-mail.: gongmingwang@126.com.

Generally speaking, the icosahedral virus is the main area of research for single-particle technology. There are three types of single-particle technologies: direct Fourier inversion algorithm [6], Fourier-Bessel algorithm [1], and icosahedral symmetry-adapted functions (ISAF) algorithm [7-10]. For ISAF algorithm, its basis function has icosahedral symmetry and the calculation region is a sphere that is close to the shape of icosahedral virus. Thus, ISAF algorithm can easily get the structure near-atomic resolution compared to Fourier-Bessel synthesis method and direct Fourier inversion method. Recently, several near-atomic resolution structures of icosahedral viruses are reconstructed with ISAF method, such as 3.9Å cytoplasmic polyhedrosis virus (CPV) [11], 3.6Å bovine papillomavirus (BPV) [12] and 3.6Å human adenovirus type 5 (Ad5) [13].

The original ISAF algorithm takes molecule as a solid sphere and calculates density function of every point. However, reconstruction results only a virus capsid and density function of every point outside virus capsid is theoretically zero. Thus, the calculation result may not be always zero, but sometimes may produce noise. In this study, it is observed that density function calculation of original ISAF algorithm can reduce reconstruction precision. To solve this issue, the range of virus capsid (inner and outer radiuses) must be determined with radial density function curve at first. Subsequently, the density function outside virus capsid should be assigned to zero.

The above mentioned technique of calculating density function is applied to original ISAF algorithm, which is based on the computational radius of density function and called as improved ISAF algorithm. To validate this algorithm, it is compared with original ISAF algorithm by using Cryo-electron microscopy (Cryo-EM) data of CPV. The effective resolution of original ISAF algorithm is 8.6Å (0.5 Fourier shell correlation (FSC) criterion [14]) and 6.8Å (0.143 FSC criterion). However, the one of improved ISAF algorithm is 7.9Å (0.5 FSC criterion) and 6.3Å (0.143 FSC criterion). Further, protrusion, wrinkle and symmetrical details of virus capsid which was not clearly seen by original ISAF algorithm is clearly visible by the modified algorithm.

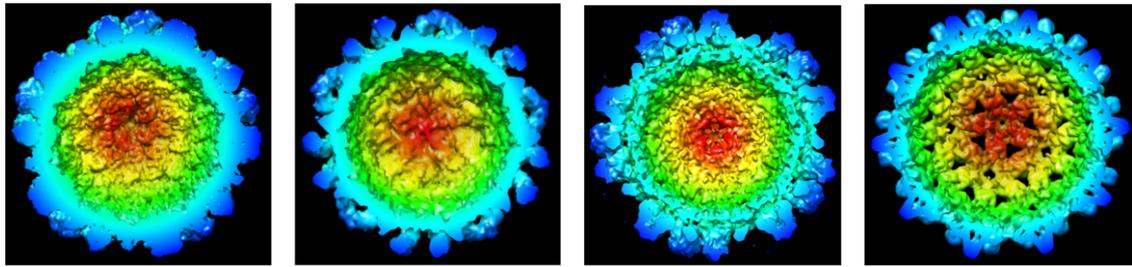
2. ISAF algorithm based on the computational radius of density function

ISAF algorithm [7-10] can achieve the near-atomic resolution structures of icosahedral viruses, and therefore, further study of this algorithm shall enhance the result.

2.1. Description of ISAF algorithm

ISAF algorithm is composed of four stages [15] i.e. Mapping molecular images, Constructing and solving fitting equation set, Calculating density function, and Storing reconstructed map into Medical Research Council (MRC) file. In the third stage, the density function $\rho(x, y, z)$ of mesh point inside asymmetric unit [15] is calculated by Eq. (1) and then mapped onto other 59 symmetrical areas.

$$\rho(r, \theta, \varphi) = \int F(R, \Theta, \Phi) \exp(-2\pi i \vec{R} \cdot \vec{r}) R^2 \sin\Theta dR d\Theta d\Phi = 4\pi \left[\sum_{l=0 \bmod 2}^{L_M} (-i)^l \sum_{u=1}^{u(l)} \left(\int_0^\infty f_{l,u}(R) j_l(2\pi R r) R^2 dR \right) S_{l,u}(\theta, \varphi) + \sum_{l=1 \bmod 2}^{L_M} (-i)^l \sum_{u=1}^{u(l)} \left(\int_0^\infty f_{l,u}(R) j_l(2\pi R r) R^2 dR \right) S_{l,u}(\theta, \varphi) \right] \quad (1)$$



(a) 2ZAH

(b) 3E5S

(c) RHDV

(d) 1QGT

Fig. 1. Four typical molecular structures.

2.2. Precise reduction of original ISAF algorithm for calculating density function

For original ISAF algorithm, the density function $\rho(r, \theta, \varphi)$ of every point in the entire sphere must be calculated. However, the reconstruction result is only a virus capsid as shown in Figure 1. Thus, calculating $\rho(r, \theta, \varphi)$ inside virus capsid is enough and can improve the resolution of reconstructed map. This statement can be proved as follows:

Proposition: S denotes size of molecular image, $r_M = \frac{1}{2}S$. The inner and outer radiuses of virus capsid are r_S and r_E . It is clear that $0 \leq r_S \leq r_E \leq r_M$. FSC_{Old} is FSC curve [14] of reconstructed map by calculating $\rho(r, \theta, \varphi)$, $0 \leq r \leq r_M$. And FSC_{New} is the one by calculating $\rho(r, \theta, \varphi)$, $r_S \leq r \leq r_E$. R_M is the maximum Fourier radius. Thus, for any $R \in [0, R_M]$, $FSC(R)_{New} \geq FSC(R)_{Old}$ is held.

Proof: The structural factor of point $\vec{R} = (R, \Theta, \Phi)$ is Fourier transform of density function $\rho(\vec{r}) = \rho(r, \theta, \varphi)$, as shown by:

$$F(\vec{R}) = F(R, \Theta, \Phi) = \int \rho(\vec{r}) \exp(2\pi i \vec{R} \cdot \vec{r}) dv = \int_0^{2\pi} d\varphi \int_0^\pi \sin\theta d\theta \int_0^{r_p} \rho(r, \theta, \varphi) \exp(2\pi i \vec{R} \cdot \vec{r}) r^2 dr \quad (2)$$

where $r \in [0, r_p]$. Let $r_p = r_M$, the Eq. (2) can be written as accumulation:

$$F(\vec{R})_O = \sum_{\varphi=0}^{2\pi} \sum_{\theta=0}^{\pi} \sum_{r=0}^{r_M} \rho(r, \theta, \varphi) \exp(2\pi i \vec{R} \cdot \vec{r}) r^2 \sin\theta dr d\theta d\varphi \quad (3)$$

Let $I(\vec{R}, r_1, r_2) = \sum_{\varphi=0}^{2\pi} \sum_{\theta=0}^{\pi} \sum_{r=r_1}^{r_2} \rho(r, \theta, \varphi) \exp(2\pi i \vec{R} \cdot \vec{r}) r^2 \sin\theta \Delta r \Delta\theta \Delta\varphi$, and Eq. (3) is expressed as:

$$F(\vec{R})_O = I(\vec{R}, 0, r_S - 1) + I(\vec{R}, r_S, r_M) + I(\vec{R}, r_E + 1, r_M) \quad (4)$$

The $\rho(r, \theta, \varphi)$, $r \notin [r_S, r_E]$ may not be zero in real $F(\vec{R})_O$, which produces noise. Thus, noise component $N(\vec{R})$ denotes $I(\vec{R}, 0, r_S - 1) + I(\vec{R}, r_E + 1, r_M)$, and signal component $F(\vec{R})_N$ denotes $I(\vec{R}, r_S, r_M)$. So, Eq. (4) is written for short:

$$F(\vec{R})_O = F(\vec{R})_N + N(\vec{R}) \quad (5)$$

It is clear that $F(\vec{R})_N$ and $N(\vec{R})$ are signal and noise component of $F(\vec{R})_O$.

In the opinion of FSC curve [14], the reciprocal space is divided into a series of shells. Initially, the full data set is split into two half data sets to produce two independent density maps. Secondly, the correlations of them at different shells are computed and finally, the resolution is determined. With this method, $FSC(R)_{Old}$ is given by:

$$FSC(R)_{Old} = \sum_{\vec{R}_i \in R} F_1(\vec{R}_i)_O \cdot F_2^*(\vec{R}_i)_O \bigg/ \sqrt{\sum_{\vec{R}_i \in R} |F_1(\vec{R}_i)_O|^2 \cdot \sum_{\vec{R}_i \in R} |F_2(\vec{R}_i)_O|^2} \quad (6)$$

where $F_1(\vec{R}_i)_O$ and $F_2(\vec{R}_i)_O$ are structure factors, and "*" denotes complex conjugation.

The numerator in Eq. (6) is as follows:

$$\begin{aligned} \sum_{\vec{R}_i \in R} F_1(\vec{R}_i)_O \cdot F_2^*(\vec{R}_i)_O &= \sum_{\vec{R}_i \in R} (F_1(\vec{R}_i)_N + N_1(\vec{R}_i)) \cdot (F_2^*(\vec{R}_i)_N + N_2^*(\vec{R}_i)) = \sum_{\vec{R}_i \in R} F_1(\vec{R}_i)_N \cdot F_2^*(\vec{R}_i)_N \\ &+ \sum_{\vec{R}_i \in R} F_1(\vec{R}_i)_N \cdot N_2^*(\vec{R}_i) + \sum_{\vec{R}_i \in R} F_2^*(\vec{R}_i)_N \cdot N_1(\vec{R}_i) + \sum_{\vec{R}_i \in R} N_1(\vec{R}_i) \cdot N_2^*(\vec{R}_i) \end{aligned} \quad (7)$$

With [16], the sums of cross terms between signal and noise are small compared to the one of other terms and can be ignored, i.e. $\sum_{\vec{R}_i \in R} F_1(\vec{R}_i)_N \cdot N_2^*(\vec{R}_i) \approx 0$, $\sum_{\vec{R}_i \in R} F_2^*(\vec{R}_i)_N \cdot N_1(\vec{R}_i) \approx 0$

The numerator in Eq. (6) can be approximately expressed as:

$$\sum_{\vec{R}_i \in R} F_1(\vec{R}_i)_O \cdot F_2^*(\vec{R}_i)_O = \sum_{\vec{R}_i \in R} F_1(\vec{R}_i)_N \cdot F_2^*(\vec{R}_i)_N + \sum_{\vec{R}_i \in R} N_1(\vec{R}_i) \cdot N_2^*(\vec{R}_i) \quad (8)$$

With the same approximation in [16], the denominator in Eq. (6) are as follows:

$$\sum_{\vec{R}_i \in R} |F_1(\vec{R}_i)_O|^2 = \sum_{\vec{R}_i \in R} F_1(\vec{R}_i)_O \cdot F_1^*(\vec{R}_i)_O = \sum_{\vec{R}_i \in R} |F_1(\vec{R}_i)_N|^2 + \sum_{\vec{R}_i \in R} |N_1(\vec{R}_i)|^2 \quad (9)$$

$$\sum_{\vec{R}_i \in R} |F_2(\vec{R}_i)_O|^2 = \sum_{\vec{R}_i \in R} |F_2(\vec{R}_i)_N|^2 + \sum_{\vec{R}_i \in R} |N_2(\vec{R}_i)|^2 \quad (10)$$

Thus, $FSC(R)_{Old}$ is shown as follows:

$$FSC(R)_{Old} = \frac{\sum_{\vec{R}_i \in R} F_1(\vec{R}_i)_N \cdot F_2^*(\vec{R}_i)_N + \sum_{\vec{R}_i \in R} N_1(\vec{R}_i) \cdot N_2^*(\vec{R}_i)}{\sqrt{\sum_{\vec{R}_i \in R} |F_1(\vec{R}_i)_N|^2 + \sum_{\vec{R}_i \in R} |N_1(\vec{R}_i)|^2} \cdot \sqrt{\sum_{\vec{R}_i \in R} |F_2(\vec{R}_i)_N|^2 + \sum_{\vec{R}_i \in R} |N_2(\vec{R}_i)|^2}} \quad (11)$$

With [14], $FSC(R)_{New}$ is expressed as:

$$FSC(R)_{New} = \sum_{\vec{R}_i \in R} F_1(\vec{R}_i)_N \cdot F_2^*(\vec{R}_i)_N \bigg/ \sqrt{\sum_{\vec{R}_i \in R} |F_1(\vec{R}_i)_N|^2 \cdot \sum_{\vec{R}_i \in R} |F_2(\vec{R}_i)_N|^2} \quad (12)$$

So:

$$\frac{FSC(R)_{Old}}{FSC(R)_{New}} = \frac{\left(\sum_{\vec{R}_i \in R} F_1(\vec{R}_i)_N \cdot F_2^*(\vec{R}_i)_N + \sum_{\vec{R}_i \in R} N_1(\vec{R}_i) \cdot N_2^*(\vec{R}_i) \right) \cdot \sqrt{\sum_{\vec{R}_i \in R} |F_1(\vec{R}_i)_N|^2 \cdot |F_2(\vec{R}_i)_N|^2}}{\sqrt{\sum_{\vec{R}_i \in R} |F_1(\vec{R}_i)_N|^2 + \sum_{\vec{R}_i \in R} |N_1(\vec{R}_i)|^2} \cdot \sqrt{\sum_{\vec{R}_i \in R} |F_2(\vec{R}_i)_N|^2 + \sum_{\vec{R}_i \in R} |N_2(\vec{R}_i)|^2}} \cdot \sum_{\vec{R}_i \in R} F_1(\vec{R}_i)_N \cdot F_2^*(\vec{R}_i)_N \quad (13)$$

From [17], it can be seen that:

$$\sum_{\vec{R}_i \in R} F_1(\vec{R}_i)_N \cdot F_2^*(\vec{R}_i)_N \approx \sqrt{N} \cdot \overline{F_1}(\vec{R}_i)_N \overline{F_2^*}(\vec{R}_i)_N \quad \sum_{\vec{R}_i \in R} N_1(\vec{R}_i) \cdot N_2^*(\vec{R}_i) \approx \sqrt{N} \cdot \overline{N_1}(\vec{R}_i)_N \overline{N_2^*}(\vec{R}_i)_N \quad (14)$$

where N is the number of sampling points in shell whose radius is R , $\overline{F_1}(\vec{R}_i)_N = \sqrt{\frac{1}{N} \sum_{\vec{R}_i \in R} |F_1(\vec{R}_i)_N|^2}$, $\overline{F_2^*}(\vec{R}_i)_N = \sqrt{\frac{1}{N} \sum_{\vec{R}_i \in R} |F_2^*(\vec{R}_i)_N|^2}$, $\overline{N_1}(\vec{R}_i)_N = \sqrt{\frac{1}{N} \sum_{\vec{R}_i \in R} |N_1(\vec{R}_i)_N|^2}$, and $\overline{N_2^*}(\vec{R}_i)_N = \sqrt{\frac{1}{N} \sum_{\vec{R}_i \in R} |N_2^*(\vec{R}_i)_N|^2}$.

Taking above equations into Eq. (13), we can get:

$$\frac{FSC(R)_{Old}}{FSC(R)_{New}} = \frac{\sqrt{\sum_{\vec{R}_i \in R} |F_1(\vec{R}_i)_N|^2} \cdot \sqrt{\sum_{\vec{R}_i \in R} |F_2(\vec{R}_i)_N|^2} + \sqrt{\sum_{\vec{R}_i \in R} |N_1(\vec{R}_i)|^2} \cdot \sqrt{\sum_{\vec{R}_i \in R} |N_2(\vec{R}_i)|^2}}{\sqrt{\sum_{\vec{R}_i \in R} |F_1(\vec{R}_i)_N|^2 + \sum_{\vec{R}_i \in R} |N_1(\vec{R}_i)|^2} \cdot \sqrt{\sum_{\vec{R}_i \in R} |F_2(\vec{R}_i)_N|^2 + \sum_{\vec{R}_i \in R} |N_2(\vec{R}_i)|^2}} \quad (15)$$

The numerator and denominator are divided by $\sqrt{\sum_{\vec{R}_i \in R} |N_1(\vec{R}_i)|^2} \cdot \sqrt{\sum_{\vec{R}_i \in R} |N_2(\vec{R}_i)|^2}$, i.e.

$$\begin{aligned} \frac{FSC(R)_{Old}}{FSC(R)_{New}} &= \frac{\sqrt{\sum_{\vec{R}_i \in R} |F_1(\vec{R}_i)_N|^2} / \sqrt{\sum_{\vec{R}_i \in R} |N_1(\vec{R}_i)|^2} \cdot \sqrt{\sum_{\vec{R}_i \in R} |F_2(\vec{R}_i)_N|^2} / \sqrt{\sum_{\vec{R}_i \in R} |N_2(\vec{R}_i)|^2} + 1}{\sqrt{\sum_{\vec{R}_i \in R} |F_1(\vec{R}_i)_N|^2} / \sqrt{\sum_{\vec{R}_i \in R} |N_1(\vec{R}_i)|^2} + 1 \cdot \sqrt{\sum_{\vec{R}_i \in R} |F_2(\vec{R}_i)_N|^2} / \sqrt{\sum_{\vec{R}_i \in R} |N_2(\vec{R}_i)|^2} + 1} \quad (16) \\ &= \frac{\sqrt{SNR_1(R)} \cdot \sqrt{SNR_2(R)} + 1}{\sqrt{SNR_1(R)} + 1 \cdot \sqrt{SNR_2(R)} + 1} \end{aligned}$$

where $SNR_1(R)$ and $SNR_2(R)$ are signal-to-noise ratios at the Fourier radius of R . Generally speaking, $SNR_1(R) + SNR_2(R) \geq 2\sqrt{SNR_1(R) \cdot SNR_2(R)}$ is held. Adding $1 + SNR_1(R) \cdot SNR_2(R)$ on each side, we can get:

$$SNR_1(R) + SNR_2(R) + 1 + SNR_1(R) \cdot SNR_2(R) \geq 2\sqrt{SNR_1(R) \cdot SNR_2(R)} + 1 + SNR_1(R) \cdot SNR_2(R) \tag{17}$$

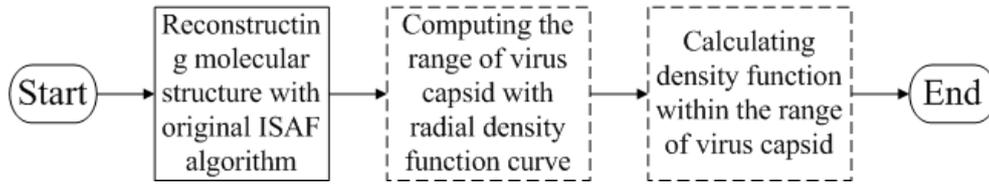


Fig. 2. The flow chart of improved ISAF algorithm.

That is $[1 + SNR_1(R)] \cdot [1 + SNR_2(R)] \geq [\sqrt{SNR_1(R)} \cdot \sqrt{SNR_2(R)} + 1]^2$

The square-root of above inequality is $\sqrt{SNR_1(R) + 1} \cdot \sqrt{SNR_2(R) + 1} \geq \sqrt{SNR_1(R)} \cdot \sqrt{SNR_2(R)} + 1$

Taking above inequality into Eq. (16), we can get: $\frac{FSC(R)_{Old}}{FSC(R)_{New}} = \frac{\sqrt{SNR_1(R)} \cdot \sqrt{SNR_2(R)} + 1}{\sqrt{SNR_1(R) + 1} \cdot \sqrt{SNR_2(R) + 1}} \leq 1$

Therefore, original ISAF algorithm may produce noise and reduce reconstruction precision. But the loss of accuracy can be regained by calculating density function of point inside the virus capsid. Therefore, the range of virus capsid must be estimated in advance.

2.3. Measurement of inner radius and outer radius of virus capsid

The radial density function curve is suitable for calculating inner radius r_S and outer radius r_E of virus capsid. Generally speaking, density function outside virus capsid is noise and its average approaches zero. However, the inside part of virus capsid is composed of signal and noise and its average is far from zero. Thus, the mean density functions of inside and outside virus capsid are different. The range of virus capsid can be determined as follows.

(1) Two arrays $DenVal[]$ and $NoVal[]$ are constructed. $DenVal[r]$ and $NoVal[r]$ are mean density function and the number of sampling points at shell "r". The initial values are both zero.

(2) The molecular structure is reconstructed by calculating $\rho(r, \theta, \varphi)$, $0 \leq r \leq r_M$.

(3) $\rho(r, \theta, \varphi)$ is loaded and carried out with the following steps.

a. The distance between (x, y, z) and origin $(0, 0, 0)$ is $d = \sqrt{x^2 + y^2 + z^2}$. If $d \geq r_M$ is held, $\rho(x', y', z')$ of next mesh point (x', y', z') is loaded and go to label "a".

b. The integral part is $Id = [d]$ and fractional part is $Fd = d - Id$.

c. The density information of point (x, y, z) is distributed to $DenVal[]$ and $NoVal[]$.

$$DenVal[Id] = DenVal[Id] + (1 - Fd)\rho(x, y, z), NoVal[Id] = NoVal[Id] + 1 - Fd,$$

$$DenVal[Id + 1] = DenVal[Id + 1] + Fd\rho(x, y, z), NoVal[Id + 1] = NoVal[Id + 1] + Fd.$$

d. $\rho(x', y', z')$ of next mesh point (x', y', z') is loaded and go to label "a".

(4) The mean density functions in different shells are $DenVal[r] = DenVal[r]/NoVal[r]$, $0 \leq r \leq r_M$.

(5) The radial density function curve " $DenVal[r] - r$ " is drawn. The inner radius r_S and outer radius r_E can be determined by analyzing this curve.

2.4. ISAF algorithm based on the computational radius of density function

If the range of virus capsid is calculated and then applied to original ISAF algorithm, the reconstruction precision is improved. For brevity in expression, this algorithm is called improved ISAF algorithm in the following paper. Its flowchart is shown in Figure 2. In this figure, the boxes with dashed line are additional operations compared with the original ISAF algorithm.

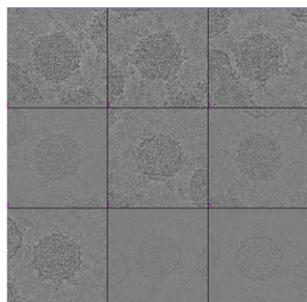


Fig. 3. Cryo-EM images of CPV.

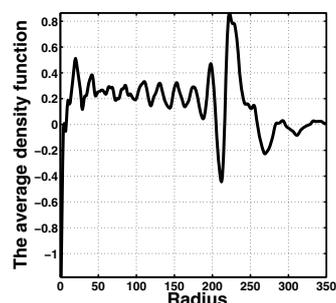
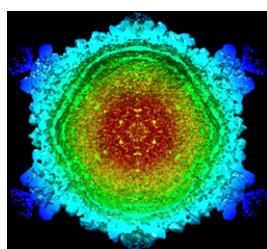
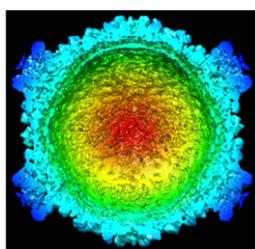


Fig. 4. The radial density function curve.

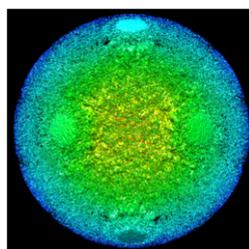


(a) Original one

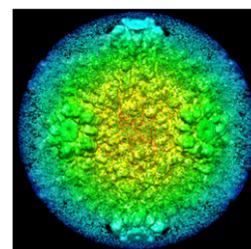


(b) Improved one

Fig. 5. The internal structures of two methods.



(a) Original one



(b) Improved one

Fig. 6. The external structures of two methods.

3. Results and discussion

CPV belongs to genus cypovirus and has icosahedral symmetrical property. The structure of CPV at resolution of 3.9\AA was reconstructed and the full atomic model of CPV capsid has been built by Lingpeng Cheng, etc [9]. We are provided 2000 images of this data set honourably. Figure 3 shows 9 images at different defocuses.

The density map was reconstructed with original ISAF algorithm at target resolution of 6.09\AA . And its radial density function curve was shown in Figure 4. This curve is unstable before 25 and become stable from 25 to 183.

Original ISAF algorithm and improved ISAF algorithm are both tested. The internal and external structures are shown in Figures 5 and 6. Inside the molecule, original ISAF algorithm has produced lots of noise and wrong structures. Thus, the protrusion and wrinkle look blurred. In addition, the one from original ISAF algorithm has lots of noise outside virus capsid. The five-fold axis could not be seen clearly. However, the one from improved ISAF algorithm has avoided above weakness, protrusion, wrinkle and symmetry can be seen clearly.

It is reported that the sharp mask of reconstructed map from original ISAF algorithm may produce an artificially improvement. But the improvement of improved ISAF algorithm is not due to this effect. To validate this point and test accuracy, the FSC curves [14] at target resolution of 6.09\AA are shown in Figure 7. On the one hand, with 0.5 FSC criterion, the effective resolution of original ISAF algorithm and improved ISAF algorithm are 8.6\AA and 7.9\AA . On the other hand, with 0.143 FSC criterion, they are 6.8\AA and 6.3\AA . In addition, the FSC curve of improved ISAF algorithm is always above the one of original ISAF algorithm, and is different from the one with sharp mask.

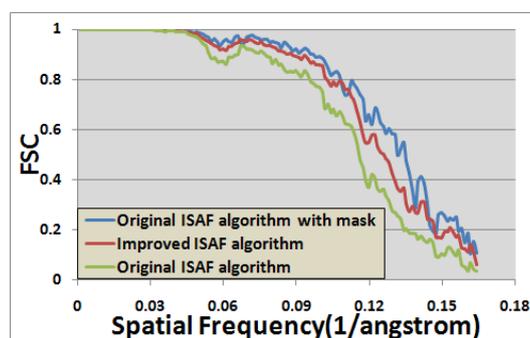


Fig. 7. The FSC curves of reconstructed maps at the target resolution of 6.09Å.

4. Conclusion

For ISAF reconstruction algorithm, the calculation precision of density function affects resolution. In fact, the reconstructed map is virus capsid and density function of point outside virus capsid is zero. However, original ISAF algorithm takes reconstructed map as a solid sphere and calculates density function of all points. The density functions outside virus capsid may not be all zero, which could produce noise. And the resolution is reduced. To solve this problem, the range of virus capsid must be determined and the density functions of points within this range only are calculated.

In this study, the range of virus capsid is determined with the radial density function curve. The improved ISAF algorithm only calculates density function within this range and can improve reconstruction precision.

The Cryo-EM data of CPV was used for validation. The results show that noise was reduced sharply and details was observed clearly. In addition, the FSC curve of improved ISAF algorithm was better than the one of original ISAF algorithm.

The improved ISAF algorithm can be extended to other single-particle reconstruction algorithms by changing method of calculating density function. Of course, calculating density function is only one aspect that affects reconstruction precision. Except for this element, several factors can affect resolution, such as initial template selection [6], CTF correction [18], structure factor sampling [19], Fourier radius initialization, sampling step etc. Thus, further work will study other aspects to improve the reconstruction precision.

Acknowledgement

The authors are grateful to Zhongjun Hu, Kai Zhang, and others for solving relative problems in structural biology, and also thankful to Lingpeng Cheng for providing the CPV data. Furthermore, this work was supported by National Basic Research Program of China (Grant No. 2012CB725300), National Natural Science Foundation of China (Grant No. 61373072) and China Postdoctoral Science Foundation (Grant No. 2013M530035).

References

- [1] R.A. Crowther, D.J. DeRosier and A. Klug, The reconstruction of a three-dimensional structure from projections and its application to electron microscopy, *Proceedings of the Royal Society of London, Series A, Mathematical, Physical and Engineering Sciences* **317** (1970), 319–340.

- [2] D.J. DeRosier and A. Klug, Reconstruction of three-dimensional structures from electron micrographs, *Nature* **217** (1968), 130–134.
- [3] J.G. Colsher, Iterative three-dimensional image reconstruction from tomographic projections, *Computer Graphics and Image Processing* **6** (1977), 513–537.
- [4] J. Frank, W. Goldfarb, D. Eisenberg and T.S. Baker, Reconstruction of glutamine synthetase using computer averaging, *Ultramicroscopy* **3** (1978), 283–290.
- [5] W. Chiu, Electron microscopy of frozen, hydrated biologicals specimens, *Annual Review of Biophysics and Biophysical Chemistry* **15** (1986), 237–257.
- [6] Z.H. Zhou, Towards atomic resolution structural determination by single particle cryo-electron microscopy, *Current Opinion in Structural Biology* **18** (2008), 218–228.
- [7] J. Navaza, On the three-dimensional reconstruction of icosahedral particles, *Journal of Structure Biology* **144** (2003), 13–23.
- [8] H.R. Liu, L.P. Cheng, S.J. Zeng, C.Y. Cai, Z.H. Zhou and Q.B. Yang, Symmetry-adapted spherical harmonics method for high-resolution 3D single-particle reconstructions, *Journal of Structural Biology* **161** (2008), 64–73.
- [9] L.F. Estrozi and J. Navaza, Ab initio high-resolution single-particle 3D reconstructions: the symmetry adapted functions way, *Journal of Structure Biology* **172** (2010), 253–260.
- [10] S. Trapani and J. Navaza, Fourier-space TEM reconstructions with symmetry adapted functions for all rotational point groups, *Journal of Structure Biology* **182** (2013), 87–92.
- [11] L.P. Cheng, J.C. Sun, K. Zhang, Z.J. Mou, X.X. Huang, G. Ji, F. Sun, J.Q. Zhang and P. Zhu, Atomic model of a cy-povirus built from Cryo-EM structure provides insight into the mechanism of mRNA capping, *Proceedings of the National Academy of Sciences of the United States of America* **108** (2011), 1373–1378.
- [12] M. Wolf, R.L. Garcea, N. Grigorieff and S.C. Harrison, Subunit interactions in bovine papillomavirus, *Proceedings of the National Academy of Sciences of the United States of America* **107** (2010), 6298–6303.
- [13] H.R. Liu, L. Jin, S.B. S Koh, I. Atanasov, S. Schein, L. Wu and Z.H. Zhou, Atomic structure of human adenovirus by Cryo-EM reveals interactions among protein networks, *Science* **329** (2010), 1038–1043.
- [14] G. Harauz and M.V. Heel, Exact filters for general geometry three dimensional reconstruction, *Optik* **73** (1986), 146–156.
- [15] G.M. Wang, F. Zhang, Q. Chu, L.Y. Fan, F. Sun and Z.Y. Liu, A fast calculation strategy of density function in ISAF reconstruction algorithm, *SCIENCE CHINA, Series F, Information Sciences* **56** (2013), 1–12.
- [16] D. Sousa and N. Grigorieff, Ab initio resolution measurement for single particle structures, *Journal of Structural Biology* **157** (2007), 201–210.
- [17] M.V. Heel and M. Schatz, Fourier shell correlation threshold criteria, *Journal of Structural Biology* **151** (2005), 250–262.
- [18] H.P. Erickson and A. Klug, The fourier transform of an electron micrograph: effects of defocussing and aberrations, and implications for the use of underfocus contrast enhancement, *Berichte der Bunsengesellschaft für physikalische Chemie* **74** (1970), 1129–1137.
- [19] S. Lanzavecchia and P.L. Bellon, A moving window shannon reconstruction algorithm for image interpolation, *Journal of Visual Communication and Image Representation* **5** (1994), 255–264.