Biophysics and Physicobiology

Regular Article

A measure for the identification of preferred particle orientations in cryo-electron microscopy data: A simulation study

Ryota Kojima and Takashi Yoshidome

Department of Applied Physics, Graduate School of Engineering, Tohoku University, Sendai, Miyagi 980-8579, Japan

Cryo-electron microscopy (cryo-EM) is an important experimental technique for the structural analysis of biomolecules that are difficult or impossible to crystallize. The three-dimensional structure of a biomolecule can be reconstructed using two-dimensional electron-density maps, which are experimentally sampled via the electron beam irradiation of vitreous ice in which the target biomolecules are embedded. One assumption required for this reconstruction is that the orientation of the biomolecules in the vitreous ice is isotropic. However, this is not always the case and two-dimensional electron-density maps are often sampled using preferred biomolecular orientations, which can make reconstruction difficult or impossible. Compensation for under-represented views is computationally feasible for the reconstruction of three-dimensional electron density maps, but one must know whether or not there is any missing information in the sampled two-dimensional electron density maps. Thus, a measure to identify whether a cryo-EM data is obtained from the biomolecules adopting preferred orientations is required. In the present study, we propose a measure for which the geometry of manifold projected onto a low-dimensional space is used. To show the usefulness of the measure, we perform simulations for cryo-EM experiment of a protein. It is found that the geometry of manifold projected onto a two-dimensional space for a protein adopting a preferred biomolecular orientation is significantly different from that for a protein adopting a uniform orientation. This result suggests that the geometry of manifold projected onto a low-dimensional space can be used for the measure for the identification that the biomolecules adopt preferred orientations.

Key words: Cryo-electron microscopy, Manifold learning, Preferred particle orientation

Corresponding author: Takashi Yoshidome, Department of Applied Physics, Graduate School of Engineering, Tohoku University, 6-6-05, Aoba, Aramaki, Aoba-ku, Sendai, Miyagi 980-8579, Japan.
e-mail: t4yoshidome@camp.apph.tohoku.ac.jp

Introduction

Cryo-electron microscopy (cryo-EM) is a powerful experimental technique for the structural analysis of biomolecules [1–4]. In cryo-EM experiments, two-dimensional electron-density maps of a biomolecule are

---

◀ *Significance* ▶

Cryo-electron microscopy (cryo-EM) enables to reconstruct the three-dimensional electron density map of a biomolecule using two-dimensional electron-density maps, experimentally sampled via electron beam irradiation of vitreous ice in which the biomolecules were embedded. However, preferred biomolecular orientations in the vitreous ice often hamper the reconstruction. Thus, it is necessary to identify whether cryo-EM data is obtained from the biomolecules adopting a preferred orientation. Here we propose a measure for the identification based on the manifold learning. It is found that our measure can certainly identify that the biomolecules adopt preferred orientations, suggesting a usefulness of the measure.

sampled via the electron beam irradiation of vitreous ice containing several target biomolecular particles. A three-dimensional electron-density map of the biomolecule can be reconstructed from two-dimensional electron-density maps because each biomolecular particle has a different orientation relative to the electron beam. One difference between cryo-EM and X-ray crystallography is that in the former, the three-dimensional structures of biomolecules are reconstructed without crystallization. This facilitates the structural analysis of biomolecules that are difficult or impossible to crystallize.

One assumption required for the reconstruction is the isotropic orientation of biomolecules in vitreous ice. However, this is not always the case and two-dimensional electron-density maps are often sampled from a preferred biomolecular orientation [4]. The preferred orientation of a biomolecule is often manifested during the process of preparing vitreous ice. In this process, the solution in which the target biomolecules are solvated is injected into a hole in a cryo-EM grid. The biomolecules in the hole are typically absorbed by an air-water surface [4], causing them to adopt their preferred orientation. The preferred orientation of biomolecules results in resolution anisotropy of the three-dimensional electron density map [5] and hampers biomolecular reconstruction. Efforts to resolve this issue are ongoing [6–18]. One potential method is to tilt the vitreous ice [6–9]. For example, Tan *et al.* demonstrated that tilting vitreous ice improves directional isotropy in biomolecules exhibiting preferred orientations [9]. However, it is inevitable that the ice will become thicker in some areas after tilting [4,9]. Capturing the same field of view at multiple tilt angles has also been proposed [6–8], but this method suffers from the beam-induced movement of biomolecules [10], which limits the resolution of the three-dimensional electron density maps. Another method is to treat the cryo-EM grid to induce another orientation [11–16]. However, this method is generally impractical [9].

Compensation for under-represented views is computationally feasible for the reconstruction of three-dimensional electron density maps [17], but one must know whether or not there is any missing information in the sampled two-dimensional electron density maps [18]. Thus, a measure to identify whether a cryo-EM data is obtained from the biomolecules adopting preferred orientations is required. In the present study, we propose a measure for the identification. The manifold learning [19–22] provides the measure. Here, we will describe the concept of a manifold under the assumption of an isotropic orientation of biomolecules. Specifically, we consider a manifold in the following high-dimensional space [22]. A two-dimensional electron-density map is represented by a point in the high-dimensional space spanned by the set of $\rho(k, l)$, which represent the electron density at pixel (k, l) (Fig. 1(a)). Because the values of $\rho(k, l)$ of a map are completely determined by the orientations of a biomolecule (i.e., there is a one-to-one correspondence between the values of $\rho(k, l)$ of a map and the orientations of a biomolecule), all points will reside in a confined space, which is the *manifold*, within the high-dimensional space (Fig. 1(a)). The low-dimensional geometry of a manifold is reflected by the set of geodesic distances between the corresponding points [21]. For example, in the manifold shown in Figure 1(a), the geodesic distance $d_G(i, k)$ is less than $d_G(i, j)$ when the i-th map is more similar to the k-th map than it is to the j-th map. Similar maps indicate that they are obtained from the biomolecules with a similar orientation relative to the electron beam. Thus, all points on a manifold are arranged in accordance with the orientation of the corresponding biomolecule. Such characteristics cannot be obtained when using the Euclidean distance because the Euclidean distance $d_E(i, k)$ is larger than $d_E(i, j)$ as shown in Figure 1(a).

Because of the low-dimensional geometry of manifold, the manifold can be projected onto a low-dimensional space. Projection is performed using methods based on the geodesic distance [21]. After the projection, manifold can be visually analyzed in accordance with the distribution of the sites in the low-dimensional space (we refer to this type of analysis as *manifold learning*). In previous studies, we have performed simulations for a cryo-EM experiment of adenylate kinase (ADK) that adopts two states, open and closed states, respectively [21,22]. Through an analysis of the manifold projected onto a three-dimensional space, it was found that the points were arranged in accordance with the orientation with respect to the direction of the electron beam [21].

Regarding manifolds of two-dimensional electron density maps of biomolecules with preferred orientations in vitreous ices, the points for maps with failed sampling are missing in the high-dimensional space, but the positions of the points experimentally sampled should not be changed (Fig. 1(b)). However, because the geodesic distances between the maps change, the geometry of the manifold projected onto a low-dimensional space would change. Thus, the geometry of the manifold projected onto a low-dimensional space can be used for a measure whether the biomolecules in vitreous ices adopted preferred orientations in performing a cryo-EM experiment. To demonstrate this, in this study we performed simulations of cryo-EM experiments to investigate the geometry of projected manifolds for two-dimensional electron-density maps of a protein adopting preferred orientations. In our simulations, we first obtained two-dimensional electron-density maps of a protein adopting an isotropic orientation by rotating the protein. This was accomplished by randomly assigning polar coordinates θ and ϕ. Next, two-dimensional electron-density maps of a protein adopting a preferred orientation were obtained by limiting the range of θ or ϕ. To project the
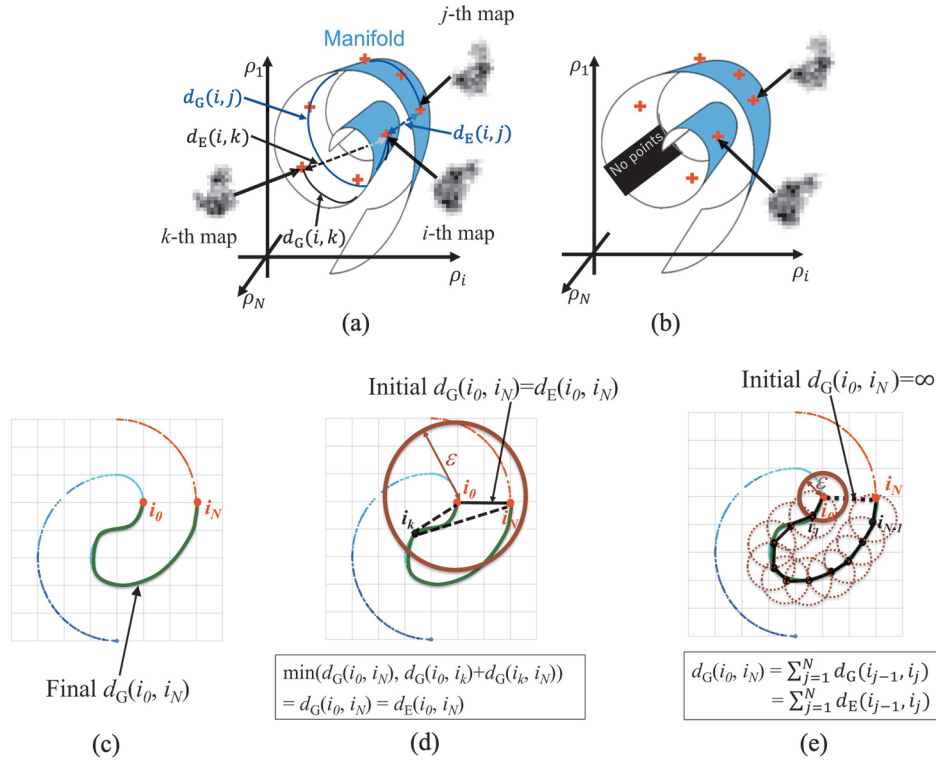
Figure 1 Schematic representations of the manifolds of two-dimensional electron-density maps of a protein adopting an isotropic (a) or preferred (b) orientation. Figures shown in Figures (c), (d) and (e) are schematic representation of a manifold on a two-dimensional space. It should be noted that this manifold is not the manifold projected onto a low-dimensional space. Suppose that the value of $d_G(i_0, i_N)$ between points i_0 and i_N is computed to project the manifold onto a one-dimensional space. Cases where ε is (d) larger than $d_E(i_0, i_N)$ and (e) smaller than $d_E(i_0, i_N)$ are illustrated. The range of $d_E(i, j) < \varepsilon$ is indicated by circles with solid and dotted lines. The black solid line in (d) and dotted line in (e) represent the Euclidean distances between points.

manifold, first the geodesic distances between the points were computed using the Isomap method [23]. Then, the projection onto a low-dimensional space was performed using the multidimensional scaling (MDS) [24] method with the geodesic distances computed with the Isomap method. We discussed how the geometry of the manifold projected onto a low-dimensional space changed by limiting the range of θ or ϕ.

Materials and Methods

Simulation of cryo-EM experiments

As a target biomolecule, we considered ADK in the open state. Its crystal structure model for ADK in an open state was taken from the Protein Data Bank [PDB ID: 4ake] [25]. We considered 2,000 conformations of the ADK protein that were sampled using molecular dynamics (MD) simulations in our previous study [21]. Although MD simulations were performed in an explicit solvent, only the protein conformations were used for subsequent calculations.

Two-dimensional electron-density maps were obtained using the following procedure [22]. First, each protein conformation was rotated around the center of the core domain by assigning random polar coordinate values (θ, ϕ) (Fig. 2(a)). No preferred orientation was assumed in this step. Next, two-dimensional electron-density maps projected along a fixed direction (z-axis in Fig. 2(a)) were calculated by counting the number of electrons in each pixel as follows. Suppose that the number of electrons for the i-th atom is $n_e(i)$, whose value is, for example, 1 for H, 8 for O, and so on. Also, the position of the i-th atom is denoted by $(x(i), y(i), z(i))$. The electron density at pixel (k, l) in the two-dimensional electron density map, denoted by $\rho(k, l)$ in Introduction section, is defined by

$$\rho(k, l) = \sum_{i=1}^{n_{Atoms}} n_e(i) F(i) \tag{1}$$

$$F(i) = \begin{cases} 1 & \left(x_P(k, l) - \dfrac{\delta}{2} \leq x(i) \leq x_P(k, l) + \dfrac{\delta}{2}, \right. \\ & \left. y_P(k, l) - \dfrac{\delta}{2} \leq y(i) \leq y_P(k, l) + \dfrac{\delta}{2}\right) \\ 0 & \text{(Otherwise)} \end{cases} \tag{2}$$
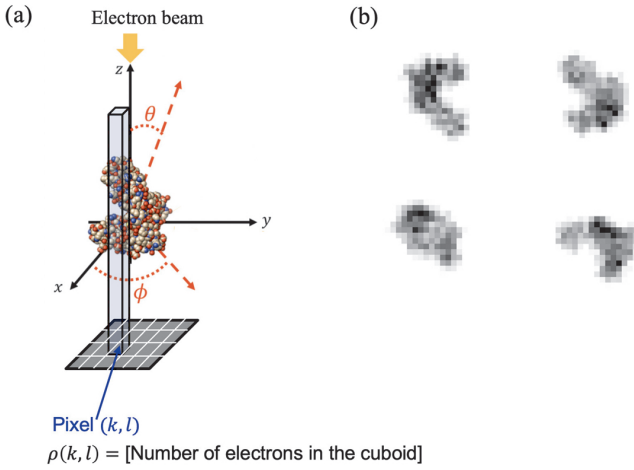
Figure 2 (a) Schematic representation of computing a two-dimensional electron-density map. (b) Some two-dimensional electron density maps obtained using the procedure described in the subsection "Simulation of cryo-EM experiment".

Here, the central position of the pixel (k, l) in the two-dimensional electron density map is denoted by $(x_P(k, l), y_P(k, l))$. Also, the pixel size is defined by $\delta \times \delta$. In the present study, δ was set to 4.0 Å and the number of pixels was 32×32. It is noted that no limitation was set to $z(i)$ in eq. (2), indicating that all electrons in the ADK within the pixel (k, l) were counted in eq. (2). Because the average of the root-mean squared displacement among the 2,000 conformations, D, defined by

$$D \equiv \frac{1}{2000 N_{C\alpha}} \sum_{i=1}^{2000} \sum_{j=1}^{N_{C\alpha}} \sqrt{|r(i, \ j) - r_{PDB}(j)|^2} , \quad (3)$$

where $N_{C\alpha}$ is the number of C_α atoms, and $r(i, j)$ and $r_{PDB}(j)$ are the position of the jth C_α atom of ith MD conformation, and that of the PDB structure, respectively, was less than 4 Å (1.53+0.28 Å (standard deviation)), the electron density maps would be regarded as those obtained from a fixed structure (The results described below would be obtained when using the PDB structure. A reason why the MD simulation data was used was that in the future study we extend the present study to the cases in which the protein fluctuation is taken into consideration by reducing the size of the pixel). Additionally, because water molecules were excluded, the maps contained no noise. We refer to the simulation using these 2,000 maps as "Simulation I" hereafter.

To investigate how the geometry of the manifold projected onto a low-dimensional space is changed by limiting the range of (θ, ϕ), Simulations II and III were performed (Table 1). In Simulation II, the range of θ was set to $0° \leq \theta \leq 90°$ and $90°+\Delta\theta \leq \theta \leq 180°$ ($\Delta\theta$=10°, 20°, or 30°). In Simulation III, the range of ϕ was set to $0° \leq \phi \leq 180°$ and $180°+\Delta\phi \leq \phi \leq 360°$ ($\Delta\phi$=10°, 20°, or 30°).

Isomap and MDS methods

Isomap calculates geodesic distance based on a set of Euclidean distances [23]. Let $\rho(i; k, l)$ denote the electron density at pixel (k, l) in the ith two-dimensional electron-density map. The Euclidean distance from point i to point j, denoted as $d_E(i, j)$, is calculated as follows:

$$d_E(i, j) = \sqrt{\sum_{k=1}^{32} \sum_{l=1}^{32} [\rho(i; k, l) - \rho(j; k + \Delta k_{Max}, l + \Delta l_{Max})]^2} . \quad (4)$$

Here, the values of Δk_{Max} and Δl_{Max} are those that yield the maximum correlation coefficient, which is defined as

$$C(i, j; \ \Delta k, \ \Delta l) = \frac{\sum_{k=1}^{32} \sum_{l=1}^{32} \rho(i; k, l)\rho(j; k + \Delta k, l + \Delta l)}{\sqrt{\left[\sum_{k=1}^{32} \sum_{l=1}^{32} \rho(i; k, l)^2\right]\left[\sum_{k=1}^{32} \sum_{l=1}^{32} \rho(j; k, l)^2\right]}} , \quad (5)$$

where both Δk and Δl vary between -3 and $+3$ [21]. To compute the geodesic distance from point i to point j, $d_G(i, j)$ is first initialized as $d_E(i, j)$ if $d_E(i, j)$ is less than ε. Otherwise, it is set to infinity. Then, all $d_G(i, j)$ values are updated as $\min(d_G(i, j), d_G(i, k)+d_G(k, j))$ for each value of k=1, 2, ..., N_{Maps}, where N_{Maps} denotes the number of two-

Table 1 Ranges of θ and ϕ, and the values of ε for Isomap, as well as the numbers of maps used for simulations

	Range of θ	Range of ϕ	ε	# of maps
Simulation I				
	$0° \leq \theta \leq 180°$	$0° \leq \phi \leq 360°$	447	2000
Simulation II				
(i)	$0° \leq \theta \leq 90°$ and $100° \leq \theta \leq 180°$ ($\Delta\theta$=10°)	$0° \leq \phi \leq 360°$	447	1876
(ii)	$0° \leq \theta \leq 90°$ and $110° \leq \theta \leq 180°$ ($\Delta\theta$=20°)	$0° \leq \phi \leq 360°$	470	1752
(iii)	$0° \leq \theta \leq 90°$ and $120° \leq \theta \leq 180°$ ($\Delta\theta$=30°)	$0° \leq \phi \leq 360°$	554	1645
Simulation III				
(i)	$0° \leq \theta \leq 180°$	$0° \leq \phi \leq 180°$ and $190° \leq \phi \leq 360°$ ($\Delta\phi$=10°)	520	1922
(ii)	$0° \leq \theta \leq 180°$	$0° \leq \phi \leq 180°$ and $200° \leq \phi \leq 360°$ ($\Delta\phi$=20°)	447	1870
(iii)	$0° \leq \theta \leq 180°$	$0° \leq \phi \leq 180°$ and $210° \leq \phi \leq 360°$ ($\Delta\phi$=30°)	447	1809

dimensional electron-density maps. The final $d_G(i, j)$ value represents the shortest path on the manifold.

ε is set to the smallest value such that the final $d_G(i, j)$ values for all pairs are finite (Table 1). This setting arises from the fact that the final geodesic distance value depends on the value of ε. Consider a manifold on a two-dimensional space (Fig. 1(c)), where the distance $d_G(i_0, i_N)$ between points i_0 and i_N is computed. If ε is set to a value greater than $d_E(i_0, i_N)$ (Fig. 1(d)), then the initial value of $d_G(i_0, i_N)$ is equal to $d_E(i_0, i_N)$. In this case, $\min(d_G(i_0, i_N), d_G(i_0, i_k)+d_G(i_k, i_N))$ is always $d_G(i_0, i_N)$ for any i_k (Fig. 1(d)) and is equal to $d_E(i_0, i_N)$. Therefore, the final $d_G(i_0, i_N)$ value is $d_E(i_0, i_N)$. However, this value is far from the correct $d_G(i_0, i_N)$ value presented in Figure 1(c). In contrast, when the initial ε value is set to be less than $d_E(i_0, i_N)$ (Fig. 1(e)), the initial $d_G(i_0, i_N)$ value is infinite. In this case, the final $d_G(i_0, i_N)$ value is the sum of the Euclidean distances between adjacent points [23] (Fig. 1(e)). Therefore, ε value should be set to a small value to avoid the situation presented in Figure 1(d).

By using the set of $d_G(i, j)$ values, the manifold was projected onto a low-dimensional space using the MDS method [24]. The first step in the MDS method is to determine the set of coordinate vectors denoted by $\{\vec{V}_i\}$ ($i=1, 2, ..., N_{\mathrm{Maps}}$) such that $d_G(i, j)$ is expressed as $d_G(i, j) = |\vec{V}_i - \vec{V}_j|$. To obtain $\{\vec{V}_i\}$, a matrix $\mathbf{B}$ defined by $\mathbf{VV}^t$ is considered. Here,

$$\mathbf{V} = \begin{pmatrix} \vec{V}_1^{\,t} \\ \vec{V}_2^{\,t} \\ \vdots \\ \vec{V}_{N_{\mathrm{Maps}}}^{\,t} \end{pmatrix}, \tag{6}$$

where $\vec{V}_i^{\,t}$ is the transposition of $\vec{V}_i$. The mth component of $\vec{V}_i$ is denoted as V_{im}. Assuming that $\sum_i V_{im}=0$ ($m=1, 2, ..., N_P$), where $N_P=32\times32$, all elements of $\mathbf{B}$ can be expressed using the set of $d_G(i, j)$ [23]. By using the diagonal matrix $\mathbf{\Lambda}$ of $\mathbf{B}$ and the matrix $\mathbf{X}$ formed by the eigenvectors of $\mathbf{B}$, $\mathbf{V}$ can be expressed as $\mathbf{X\Lambda}^{1/2}$. Using the components of $\mathbf{\Lambda}$ and $\mathbf{X}$, $d_G(i, j)$ can be expressed as

$$d_G(i, j) = \sqrt{\sum_{l=1}^{N_P} \lambda_l (X_{il} - X_{jl})^2}, \tag{7}$$

where the ith eigenvalue is denoted by λ_i and X_{il} is the lth component of $\vec{X}_i$. Here, the subscripts of λ_i and $\vec{X}_i$ are defined such that $\lambda_1 > \lambda_2 > ....$ If a large gap is identified between λ_k and λ_{i+1}, then $d_G(i, j)$ is approximated as follows:

$$d_G(i, j) \approx \sqrt{\sum_{l=1}^{k} \lambda_l (X_{il} - X_{jl})^2}. \tag{8}$$

In this case, the projection of the manifold onto a k-dimensional space can be performed using $\sqrt{\lambda_1}\vec{X}_1$, $\sqrt{\lambda_2}\vec{X}_2$, ..., and $\sqrt{\lambda_k}\vec{X}_k$ [23].

By using the maps of ADK adopting an isotropic orientation, we previously confirmed that the first three eigenvalues were sufficient for classification in terms of rotation angles (see Fig. 6 in Ref. 21). These results were confirmed for both the open and closed states of ADK [21]. However, in this study, the number of dimensions was derived from the eigenvalues because maps of ADK adopting a preferred orientation were considered, while a uniform orientation was assumed in the previous study [21].

Results

Simulation I

We first conducted Simulation I to analyze the manifold for the two-dimensional electron-density maps of ADK adopting an isotropic orientation. To project this manifold, first the $d_G(i, j)$ values between maps i and j were computed using Isomap method. Next, the λ_i values and the corresponding $\vec{X}_i$ for matrix $\mathbf{B}$ were obtained using the MDS method. The ε value was set to 447.

To determine the dimensions of the low-dimensional space onto which the manifold was projected, we first analyzed the eigenvalues of $\mathbf{B}$. As shown in Figure 3(a), there is a large gap between λ_2 and λ_3. A gap is also present between λ_3 and λ_4, although its size is smaller than that of the gap between λ_2 and λ_3. According to these results, the manifold should be projected onto the two-dimensional space spanned by $\sqrt{\lambda_1}\vec{X}_1$ and $\sqrt{\lambda_2}\vec{X}_2$ (hereafter referred to as the first and second Isomap coordinates, respectively). However, we will first discuss the manifold projected onto the three-dimensional space spanned by the first, second, and third Isomap coordinates (The third Isomap coordinate is defined by $\sqrt{\lambda_3}\vec{X}_3$), because this is necessary for the discussion of Simulation III.

The manifold projected onto a three-dimensional space is similar to a circular cylinder (Figs. 3(b) and 3(c)). To analyze the relationship between the positions of points and the projection directions, we classified points into six groups according to their corresponding θ and ϕ values. Figures 3(b) and 3(c) indicate that the positions of the points likely depend on θ and ϕ. Specifically, the third Isomap coordinate and arc are related to θ and ϕ, respectively. These results are consistent with those found in our previous study [21].

The manifold projected onto the two-dimensional space is similar to a circle (Figs. 3(d) and 3(e)). By classifying points into six groups according to the corresponding θ and ϕ values, it was determined that the radius and arc of the circle are related to θ and ϕ, respectively. It is noted that when the manifold shown in Figure 3(c) is projected onto
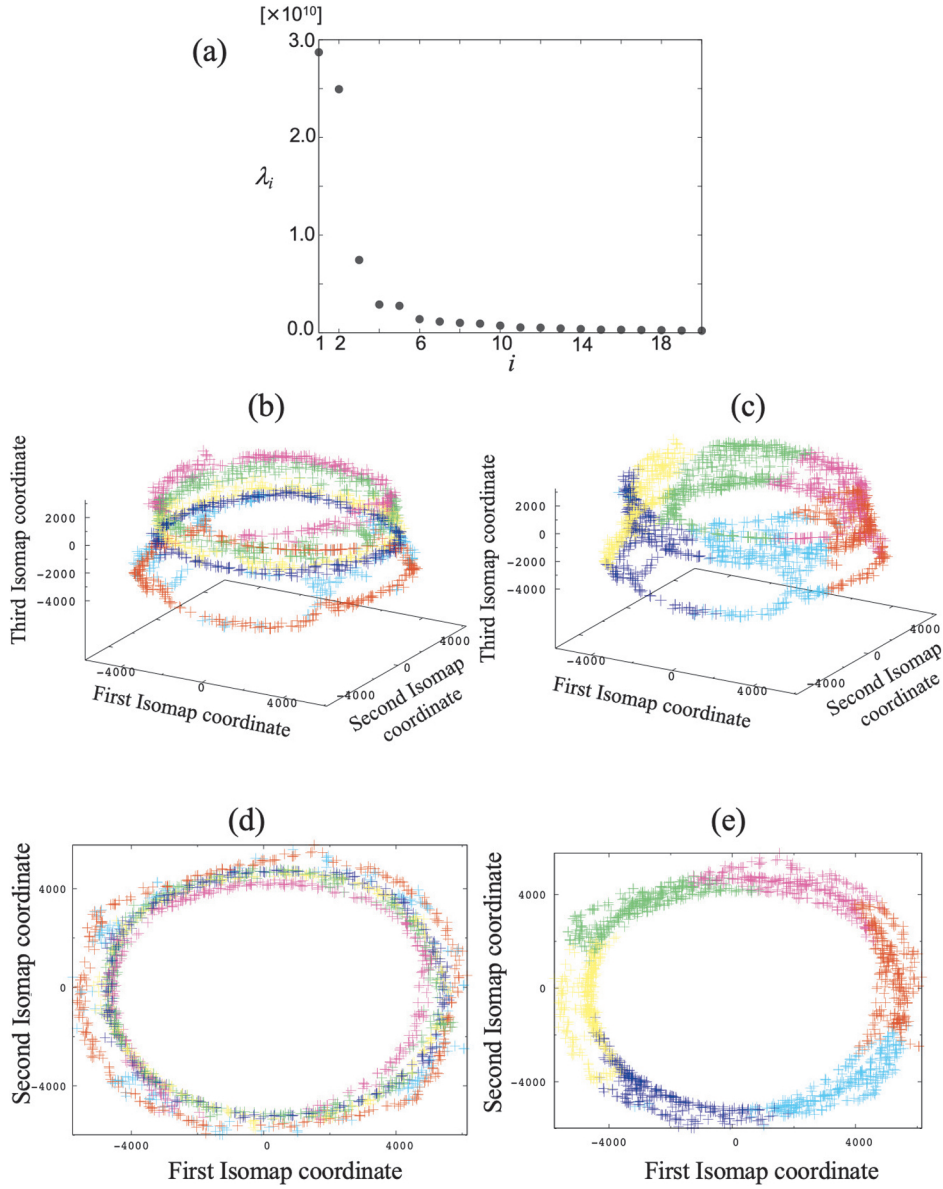
Figure 3 Computational results for Simulation I. (a) First 20 eigenvalues. (b) Manifold projected onto a three-dimensional space in which points are classified into six groups according to the value of θ. Their colors are defined as follows: red ($0°\leq\theta\leq30°$), cyan ($30°\leq\theta\leq60°$), blue ($60°\leq\theta\leq90°$), yellow ($90°\leq\theta\leq120°$), green ($120°\leq\theta\leq150°$), and pink ($150°\leq\theta\leq180°$). (c) Manifold projected onto a three-dimensional space in which points are classified into six groups according to the value of ϕ. Their colors are defined as follows: red ($0°\leq\phi\leq60°$), cyan ($60°\leq\phi\leq120°$), blue ($120°\leq\phi\leq180°$), yellow ($180°\leq\phi\leq240°$), green ($240°\leq\phi\leq300°$), and pink ($300°\leq\phi\leq360°$). (d) Manifold projected onto a two-dimensional space in which points are classified into six groups according to the value of θ. Their colors are the same as those in (b). (e) Manifold projected onto a two-dimensional space in which points are classified into six groups according to the value of ϕ. Their colors are the same as those in (c).

the plane at which the value on the third Isomap coordinate is zero, the manifold shown in Figure 3(e) is obtained. Thus, the radius of the circle in Figure 3(e) is related to the third isomap coordinate.

Simulation II

We performed the same procedure described in the previous subsection to project the manifolds for this simulation. The ε values for Isomap are summarized in Table 1. The manifolds were projected onto the two-dimensional space spanned by the first and second Isomap coordinates, although large gaps in eigenvalues were identified between λ_1 and λ_2 for both $\Delta\theta=20°$ and $30°$ (Figs. 4(a), 4(b), and 4(c)).

The geometry of the projected manifolds was found to be strongly dependent on the value of $\Delta\theta$ (Figs. 4(d), 4(e), and 4(f)). For $\Delta\theta=10°$ (Simulation II(i)), the projected manifold was essentially the same as the manifold from Simulation I.
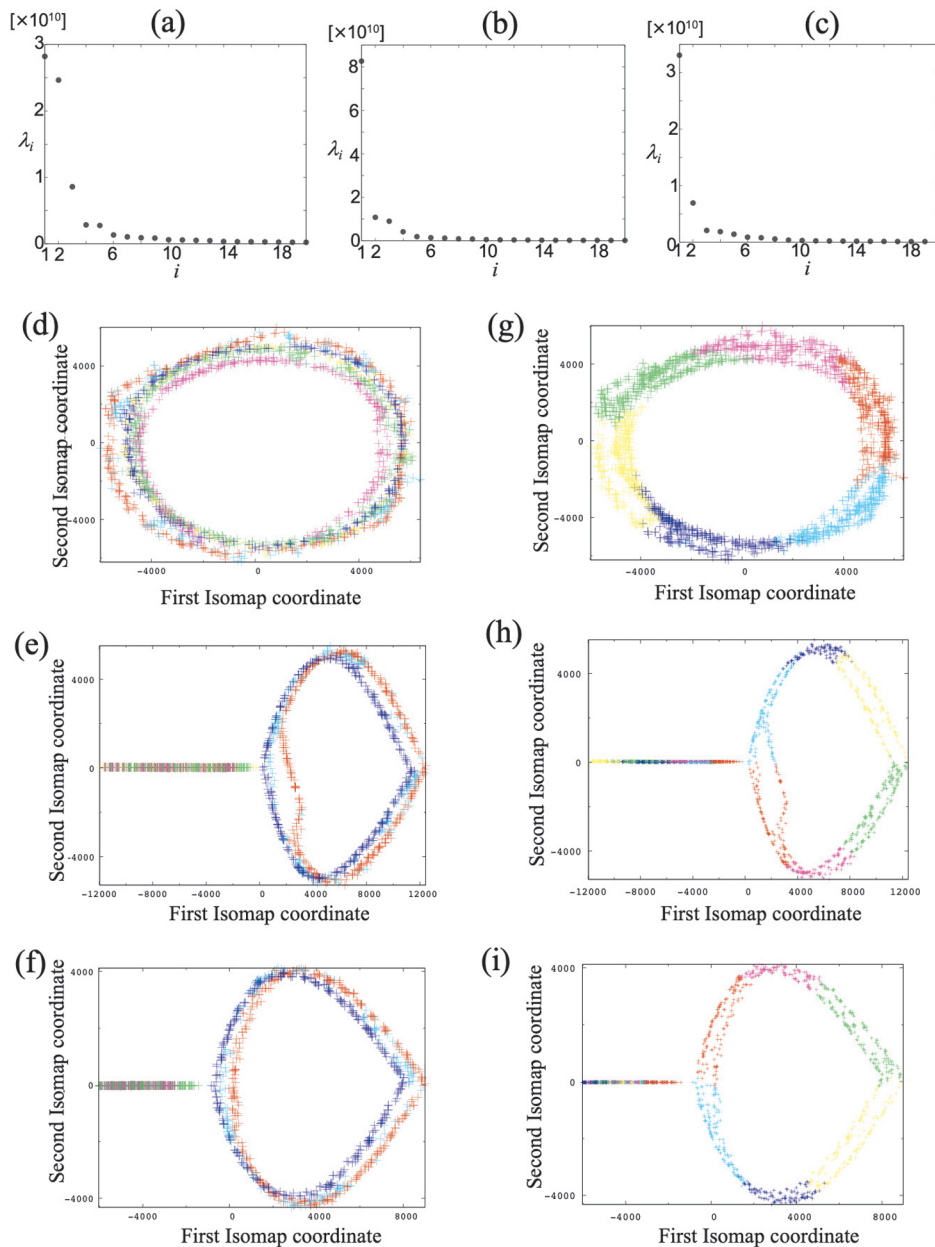
Figure 4   Computational results for Simulation II. First 20 eigenvalues for $\Delta\theta$ values of (a) 10°, (b) 20°, and (c) 30°. Manifolds projected onto a two-dimensional space are shown, where points are classified into six groups according to the value of θ for $\Delta\theta$ values of (d) 10°, (e) 20°, and (f) 30°, respectively. The colors are the same as those in Figure 3(b). In Figures (g), (h), and (i), the points are classified into six groups according to the value of ϕ for $\Delta\theta$ values of (g) 10°, (h) 20°, and (i) 30°, respectively. The colors are the same as those in Figure 3(c).

In contrast, the manifolds for $\Delta\theta$=20° and 30° (Simulation II(ii) and Simulation II(iii)) were separated into two. Therefore, the preferred orientation of the protein strongly affected the projected manifolds for $\Delta\theta$=20° and 30°. The geometry of manifold was highly sensitive to a slight limitation of the range of θ such as Simulation II(ii) ($\Delta\theta$=20°).

To visually discuss from the viewpoint of projection angle, the points were classifying into six groups according

to θ values (Figs. 4(d), 4(e), and 4(f)). For Simulation II(i), classification was successful (Fig. 4(d)). As for the other simulations, it was found that the points in the maps in the region from 0° to 90° and those in the remaining maps were separated (Figs. 4(e) and 4(f)). Failures in the sampling of electron-density maps for θ values between 90° and 90°+$\Delta\theta$ caused this separation. Because it was unclear from Figures 4(e) and 4(f) whether or not classification was successful for Simulations II(ii) and II(iii), we projected
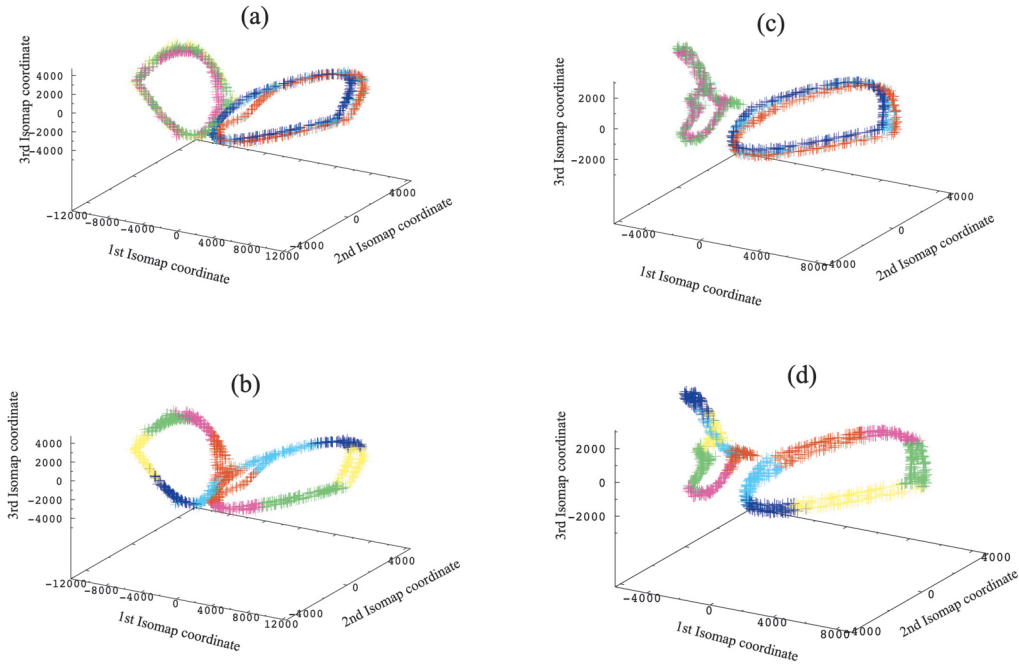
Figure 5   Manifolds projected onto a three-dimensional space for Simulation II(ii) [Figs. (a) and (b)] and for Simulation II(iii) [Figs. (c) and (d)]. In the figures points are classified into six groups according to the value of θ [(a) and (c)] and of ϕ [(b) and (d)]. The colors are the same as those in Figures 3(b) and 3(c).

these manifolds onto a three-dimensional space (Figs. 5(a) and 5(c)). While the points were roughly arranged in accordance with θ value, the points with $\theta \geq 90°$ was highly overlapped. As for the classifying into six groups according to ϕ values (Figs. 4(g), 4(h), 4(i), 5(b), and 5(d)), the points were arranged in accordance with ϕ value.

Simulation III

We performed the same procedure described in the previous subsections to project the manifolds for this simulation. The ε values for Isomap are summarized in Table 1. The manifolds were projected onto the two-dimensional space spanned by the first and second Isomap coordinates, although we identified a large gap in eigenvalues between λ_1 and λ_2 for both $\Delta\phi=20°$ and $30°$ (Figs. 6(a), 6(b), and 6(c)). We also projected the manifolds onto the three-dimensional space spanned by the first, second, and third Isomap coordinates in order to compare the manifold for Simulation II.

As shown in Figures 6(d), 6(e), and 6(f), the geometry of the manifolds projected onto the two-dimensional space strongly depends on the value of $\Delta\phi$. The manifold for $\Delta\phi=10°$ is essentially the same as that for Simulation I. In contrast, the manifolds for $\Delta\phi=20°$ and $30°$ exhibit completely different geometries compared to the manifold from Simulation I. The change in the geometry of manifold was also observed for Simulations III(ii) and III(iii) when the manifolds were projected onto the three-dimensional

space (Fig. 7): The manifolds are board-like shape spanned by the first and second Isomap coordinates. The geometry of manifold for Simulation III(i) was essentially the same as that for Simulation I. The geometry of manifold was also highly sensitive to a slight limitation of the range of ϕ such as Simulation III(ii) ($\Delta\phi=20°$).

To analyze how the points were arranged, the points on the manifolds were classified into six groups according to the values of ϕ. For Simulation III(i), the points were arranged in the same manner as that for Simulation I. On the other hand, for the other simulations, it was determined that the points were arranged in order of ϕ along the first Isomap coordinate (Figs. 6(e) and 6(f)). The same arrangement was also observed when the manifolds were projected onto the three-dimensional space (Figs. 7(a), 7(b), and 7(c)). Classification of the points into six groups according to the values of θ was also performed and the points were found to be arranged in accordance with θ (Figs. 6(g), 6(h), 6(i), 7(d), 7(e), and 7(f)). By comparing the results shown in Figures 6(e) and 6(f) to the results obtained from Simulation I (Fig. 3(c)), it was determined that the manifolds for $\Delta\phi=20°$ and $30°$ were obtained via the projection of the manifold from the three-dimensional space (Fig. 3(c)) onto the two-dimensional space spanned by the third Isomap coordinate and the circumference of the manifold.

Finally, the geometry of manifolds projected onto the three-dimensional space is compared between Simulations
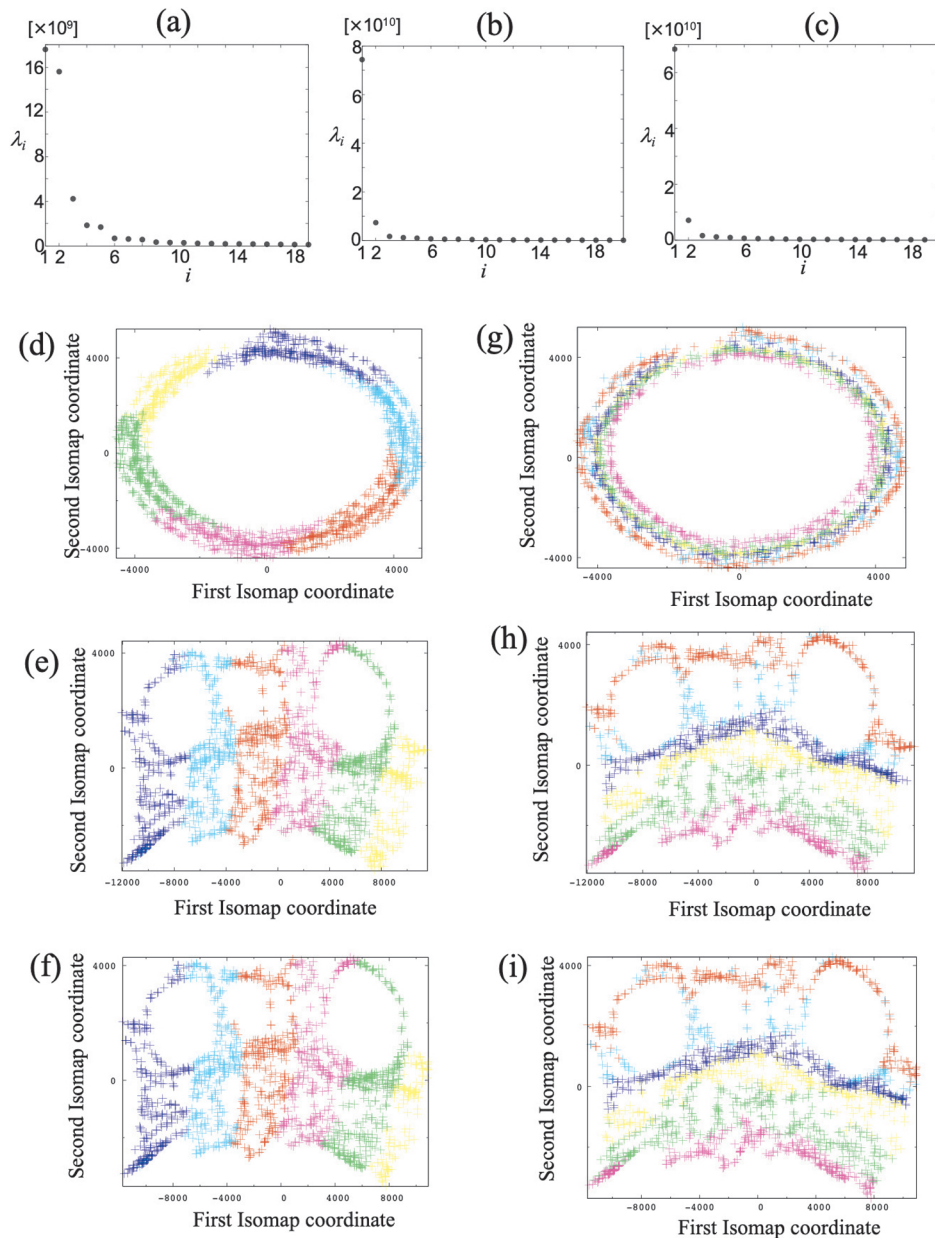
Figure 6   Computational results for Simulation III. First 20 eigenvalues for Δφ values of (a) 10°, (b) 20°, and (c) 30°. Manifolds projected onto a two-dimensional space are shown, where points are classified into six groups according to the value of φ for Δφ values of (d) 10°, (e) 20°, and (f) 30°, respectively. The colors are the same as those in Figure 3(b). In Figures (g), (h), and (i), the points are classified into six groups according to the value of θ for Δθ values of (g) 10°, (h) 20°, and (i) 30°, respectively. The colors are the same as those in Figure 3(c).

II and III (Figs. 5 and 7). When Δθ or Δφ is increased, the manifold was separated into two for Simulation II, while its shape was changed from circular cylinder to board for Simulation III. Such difference in the shape of manifolds can be explained as follows. In Simulations II(ii) and II(iii), an increase in the Δθ value leads to the separation of the range of θ into 0°≤θ≤90° and 90°+Δθ≤θ≤180°. This separation would bring the separation of the manifold for Simulations II(ii) and II(iii). For Simulations III(ii) and

III(iii), on the other hand, an increase in the Δφ value does *not* lead to the separation: For Δφ=30°, for example, the range of 0°≤φ≤180° and 210°≤φ≤360° is equal to −150°≤φ≤180°. Thus, the separation does not occur for Simulations III(ii) and III(iii).

Results for the closed state

We have also performed the same computations for the ADK in the closed state (See Supplementary Text S1
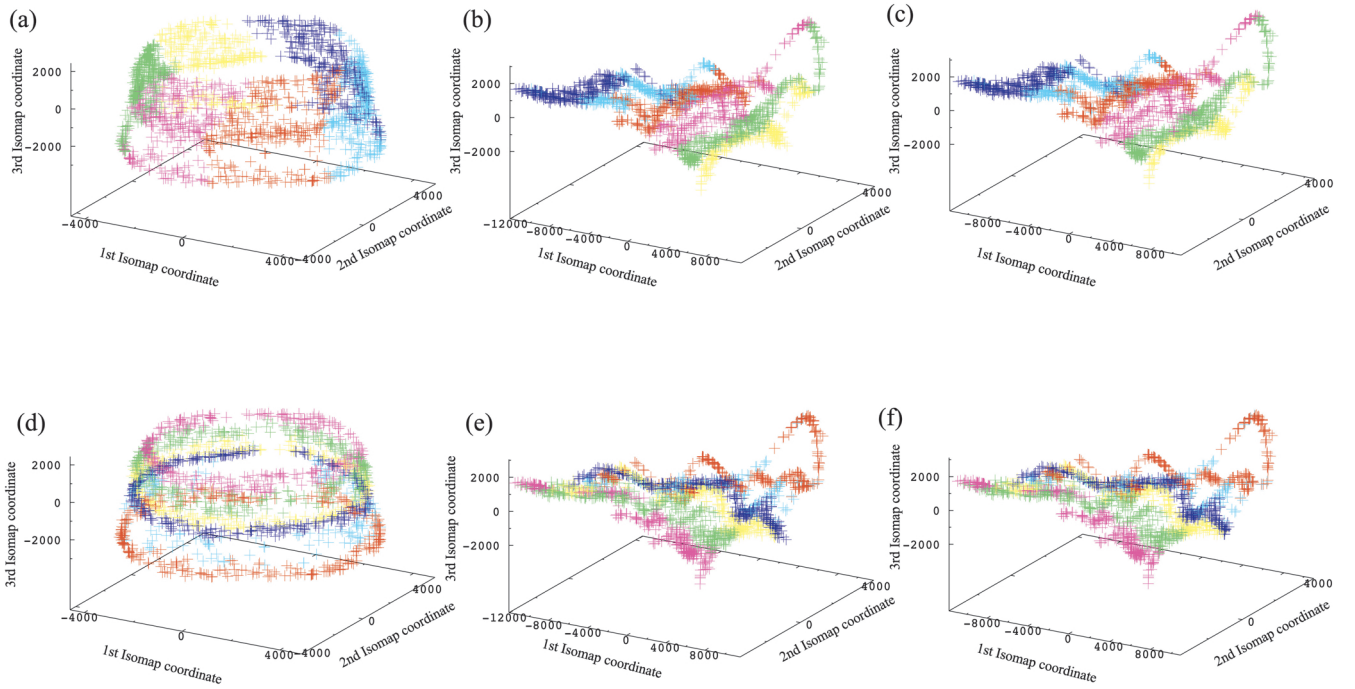
Figure 7   Manifolds projected onto a three-dimensional space for Simulation III(i) [Figs. (a) and (b)], for Simulation III(ii) [Figs. (c) and (d)], and for Simulation III(iii) [Figs. (e) and (f)]. In the figures points are classified into six groups according to the value of ϕ [(a), (c), and (e)] and of θ [(b), (d), and (f)]. The colors are the same as those in Figures 3(b) and 3(c).

and Supplementary Table S1). It was confirmed that essentially the same results as those for the open state were obtained (Supplementary Figs. S1-S5).

The results for the open and closed states leads to the following conclusions: (i) The geometry of manifold (circular cylinder) would be highly independent on protein conformation; and (ii) The change of the geometry of manifold from circular cylinder indicates that the cryo-EM data is obtained from the biomolecules adopting preferred orientations. Thus, the geometry of manifold can be a measure for the identification of preferred particle orientation in cryo-EM experimental data.

Discussion

We now discuss our method from the two points of view. One of the two is the scoring function with which the identification whether preferred orientations exist or not. The second point is a utility of the present method, which is discussed with maps of the biomolecule adopting realistic preferred orientations.

Scoring function for the identification whether preferred orientations exit or not

From the figures of the eigenvalues, we propose that λ_2/λ_1 can be used for the scoring function. When the two-dimensional electron density maps were obtained from the biomolecules adopting uniform orientation, λ_2/λ_1 is close to one (See Fig. 3(a)). On the other hand, λ_2/λ_1 for the two-dimensional electron density maps obtained from the biomolecules adopting a preferred orientation is far from one (For example, λ_2/λ_1=0.12 for Simulation II(ii) shown in Fig. 4(b)). Thus, using λ_2/λ_1 for a scoring function, we can identify whether preferred orientations exist or not.

A utility of the present method

To show a utility, we applied the present method to the two-dimensional electron density maps for the open conformation of ADK adopting a realistic preferred orientation. Because we could not find a realistic angular distribution of ADK, the distribution of the influenza hemagglutinin (HA) trimer shown in Figure 1(a) in Ref. 9 was adopted as a representative realistic angular distribution. Three high-population regions can be visually identified around polar angle (θ, ϕ) values of $(0°, 60°)$, $(90°, 0°)$, and $(90°, 120°)$.

To perform simulations for ADK adopting the angular distribution of the HA trimer, (θ, ϕ) values were first randomly generated according to a Gaussian distribution using the *np.random.normal* function implemented in the numpy library of the Python programming language. The central positions and variances of the Gaussian function are described in Table 2 and a scatter plot of the generated random numbers for (θ, ϕ) is presented in Figure 8(a).

Table 2 The parameters for generating the Gaussian random numbers and the number of images for the simulation in the subsection "A utility of the present method"

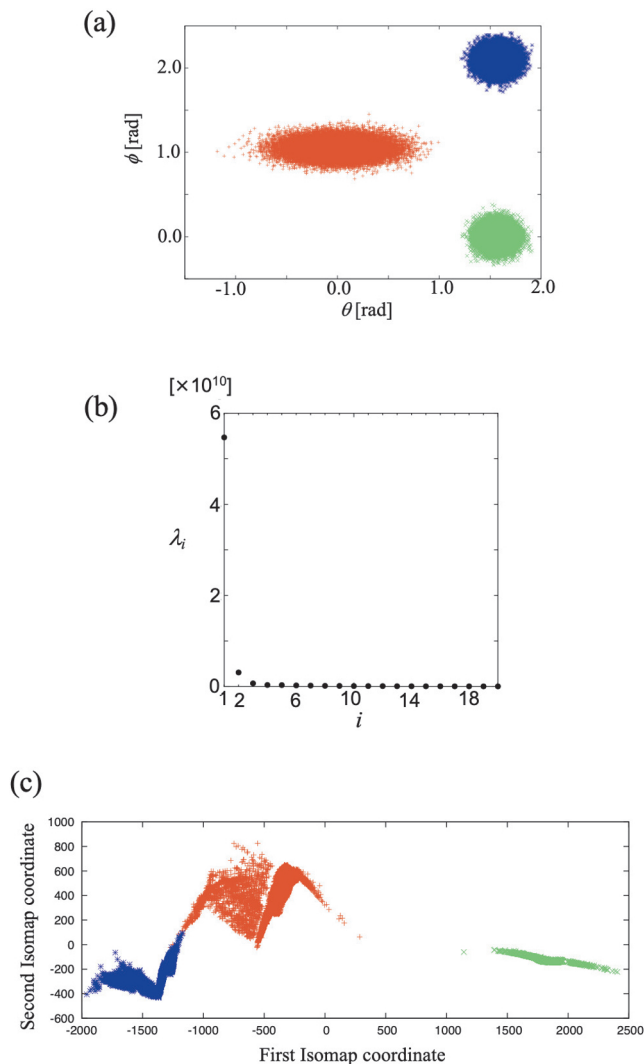| Region	Center position (θ, ϕ)	Variance (θ, ϕ)	# of maps
1	(0°, 60°)	(15°, 5°)	10,000
2	(90°, 0°)	(5°, 5°)	10,000
3	(90°, 120°)	(5°, 5°)	10,000

Figure 8 Computation results for the simulation of cryo-EM experiment for the ADK adopting a realistic preferred orientation: (a) Scatter plot of polar angles used for the simulation; (b) First twenty eigenvalues; and (c) projected manifolds. The colors in (a) and (c) represent the regions 1, 2, and 3, respectively in Table 2.

Next, 30,000 maps were computed using the random numbers as projection angles.

The manifold was projected using Isomap and the MDS method. An ε value of 845 was used for Isomap. As shown in Figure 8(b), there is a large gap between λ_1 and λ_2, suggesting the projection of the manifold onto the one-dimensional space spanned by the first Isomap coordinate. However, the manifold was projected onto the two-dimensional space spanned by the first and second Isomap coordinates in accordance with the simulations described above.

The manifold projected onto the two-dimensional space is presented in Figure 8(c). The points were classified according to their projection angles using the colors in Figure 8(a). It was found that the geometry of the projected manifold is different from that presented in Figure 3(d). In addition, λ_2/λ_1 was 0.05. These results indicate that our method successfully identified the present two-dimensional electron density maps that were obtained from the ADK adopting realistic preferred orientations.

Conclusion

In this study, we investigated the manifolds of two-dimensional electron-density maps of ADK adopting a preferred orientation. To this end, simulations of cryo-EM experiments were conducted to compute two-dimensional electron-density maps of ADK in the open and closed states. Next, Isomap and the MDS method were applied to the two-dimensional electron-density maps to project the manifold onto a low-dimensional space. We have found that the geometry of manifold was changed even when $\Delta\theta$ and $\Delta\phi$ are small such as Simulations II(ii) and III(ii). Thus, we have concluded that the geometry of manifold can be a measure for identification of preferred particle orientation in cryo-EM experimental data.

In our simulations, the rotation with respect to ϕ was performed after the rotation with respect to θ. We also performed simulations such that the rotation with respect to θ was performed after the rotation with respect to ϕ. Simulations were performed for Simulations I, II(iii), and III(iii). It was found that essentially the same results were obtained (Supplementary Fig. S6). Thus, the order of the rotation did not alter the present results.

Conflicts of Interest

The authors declare no conflicts of interest.

Author Contribution

R.K. and T.Y. performed the calculation. T.Y. wrote the paper.

References

[1] Bai, X. C., McMullan, G. & Scheres, S. H. How cryo-EM is revolutionizing structural biology. *Trends Biochem. Sci.* **40**, 49–57 (2015). DOI: 10.1016/j.tibs.2014.10.005

[2] Fernandez-Leiro, R. & Scheres, S. H. W. Unravelling biological macromolecules with cryo-electron microscopy. *Nature* **537**, 339–346 (2016). DOI: 10.1038/nature19948

[3] Cheng, Y. Single-particle cryo-EM—How did it get here and where will it go. *Science* **361**, 876–880 (2018). DOI: 10.1126/science.aat4346

[4] Lyumkis, D. Challenges and opportunities in cryo-EM single-particle analysis. *J. Biol. Chem.* **294**, 5181–5197 (2019). DOI: 10.1074/jbc.REV118.005602

[5] Noble, A. J., Dandey, V. P., Wei, H., Brasch, J., Chase, J., Acharya, P., *et al.* Routine single particle CryoEM sample and grid characterization by tomography. *Elife* **7**, e34257 (2018). DOI: 10.7554/eLife.34257

[6] Radermacher, M. Three-dimensional reconstruction of single particles from random and nonrandom tilt series. *J. Electron Microsc. Tech.* **9**, 359–394 (1988). DOI: 10.1002/jemt.1060090405

[7] Radermacher, M., Wagenknecht, T., Verschoor, A. & Frank, J. Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit of Escherichia coli. *J. Microsc.* **146**, 113–136 (1987). DOI: 10.1111/j.1365-2818.1987.tb01333.x

[8] Leschziner, A. E. & Nogales, E. The orthogonal tilt reconstruction method: an approach to generating single-class volumes with no missing cone for ab initio reconstruction of asymmetric particles. *J. Struct. Biol.* **153**, 284–299 (2006). DOI: 10.1016/j.jsb.2005.10.012

[9] Tan, Y. Z., Baldwin, P. R., Davis, J. H., Williamson, J. R., Potter, C. S., Carragher, B., *et al.* Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat. Methods* **14**, 793–796 (2017). DOI: 10.1038/nmeth.4347

[10] Henderson, R., Chen, S., Chen, J. Z., Grigorieff, N., Passmore, L. A., Ciccarelli, L., *et al.* Tilt-Pair Analysis of Images from a Range of Different Specimens in Single-Particle Electron Cryomicroscopy. *J. Mol. Biol.* **413**, 1028–1046 (2011). DOI: 10.1016/j.jmb.2011.09.008

[11] Frank, J., Penczek, P., Grassucci, R. & Srivastava, S. Three-dimensional reconstruction of the 70S Escherichia coli ribosome in ice: the distribution of ribosomal RNA. *J. Cell Biol.* **115**, 597–605 (1991). DOI: 10.1083/jcb.115.3.597

[12] Meyerson, J. R., Rao, P., Kumar, J., Chittori, S., Banerjee, S., Pierson, J., *et al.* Self-assembled monolayers improve protein distribution on holey carbon cryo-EM supports. *Sci. Rep.* **4**, 7084 (2014). DOI: 10.1038/srep07084

[13] Nguyen, T. H., Galej, W. P., Bai, X. C., Savva, C. G., Newman, A. J., Scheres, S. H. W., *et al.* The architecture of the spliceosomal U4/U6.U5 tri-snRNP. *Nature* **523**, 47–52 (2015). DOI: 10.1038/nature14548

[14] Liu, Y., Meng, X. & Liu, Z. Deformed grids for single-particle cryo-electron microscopy of specimens exhibiting a preferred orientation. *J. Struct. Biol.* **182**, 255–258 (2013). DOI: 10.1016/j.jsb.2013.03.005

[15] Chowdhury, S., Ketcham, S. A., Schroer, T. A. & Lander, G. C. Structural organization of the dynein-dynactin complex bound to microtubules. *Nat. Struct. Mol. Biol.* **22**, 345–347 (2015). DOI: 10.1038/nsmb.2996

[16] Lyumkis, D., Julien, J. P., de Val, N., Cupo, A., Potter, C. S., Klasse, P. J., *et al.* Cryo-EM structure of a fully glycosylated soluble cleaved HIV-1 envelope trimer. *Science* **342**, 1484–1490 (2013). DOI: 10.1126/science.1245627

[17] Penczek, P. A. Fundamentals of three-dimensional reconstruction from projections. *Methods Enzymol.* **482**, 1–33 (2010). DOI: 10.1016/S0076-6879(10)82001-4

[18] Urnavicius, L., Zhang, K., Diamant, A. G., Motz, C., Schlager, M. A., Yu, M., *et al.* The structure of the dynactin complex and its interaction with dynein. *Science* **347**, 1441–1446 (2015). DOI: 10.1126/science.aaa4080

[19] Schwander, P., Fung, R. & Ourmazd, A. Conformations of macromolecules and their complexes from heterogeneous datasets. *Philos. Trans. R. Soc. B* **369**, 20130567 (2014). DOI: 10.1098/rstb.2013.0567

[20] Dashti, A., Schwander, P., Langlois, R., Fung, R., Li, W., Hosseinizadeh, A., *et al.* Trajectories of the ribosome as a Brownian nanomachine. *Proc. Natl. Acad. Sci. USA* **111**, 17492–17497 (2014). DOI: 10.1073/pnas.1419276111

[21] Yoshidome, T., Oroguchi, T., Nakasako, M. & Ikeguchi, M. Classification of projection images of proteins with structural polymorphism by manifold: a simulation study for x-ray free-electron laser diffraction imaging. *Phys. Rev. E* **92**, 032710 (2015). DOI:10.1103/PhysRevE.92.032710

[22] Takano, N. & Yoshidome, T. Effects of a Low-Pass Filter on the Classification of Protein Electron-Density Maps in Terms of Structural Polymorphism using Manifold Learning: A Simulation Study for Cryo-Electron Microscopy Experiments. *J. Phys. Soc. Jpn.* **88**, 094801 (2019). DOI: 10.7566/JPSJ.88.094801

[23] Tenenbaum, J. B., de Silva, V. & Langford, J. C. A global geometric framework for nonlinear dimensionality reduction. *Science* **290**, 2319–2323 (2000). DOI: 10.1126/science.290.5500.2319

[24] Everitt, B. S. *An R and S-Plus® Companion to Multivariate Analysis* (Springer, New York, 2007).

[25] Müller, C. W., Schlauderer, G. J., Reinstein, J. & Schulz, G. E. Adenylate kinase motions during catalysis: an energetic counterweight balancing substrate binding. *Structure* **4**, 147–156 (1996). DOI: 10.1016/s0969-2126(96)00018-4

(Edited by Ryo Akiyama)