



An Auto-Encoder to Reconstruct Structure with Cryo-EM Images via Theoretically Guaranteed Isometric Latent Space, and Its Application for Automatically Computing the Conformational Pathway

Kimihiro Yamazaki¹, Yuichiro Wada^{1,2}, Atsushi Tokuhisa³, Mutsuyo Wada¹, Takashi Katoh¹, Yuhei Umeda¹, Yasushi Okuno^{3,4}, and Akira Nakagawa¹(✉)

¹ Fujitsu Ltd., Kanagawa, Japan
anaka@fujitsu.com

² RIKEN Center for Advanced Intelligence Project, Tokyo, Japan

³ RIKEN Center for Computational Science, Hyogo, Japan

⁴ Graduate School of Medicine, Kyoto University, Kyoto, Japan

Abstract. Structural analysis by cryo-electron microscopy (Cryo-EM) has become well-established in the field of structural biology. Recently, cutting-edge methods have been proposed for the purpose of reconstructing either a small set of structures or a conformational pathway (continuous structural change), where a 3D density map represents the structure. However, we usually perform heavy manual labor to define the plausible pathway related to biological significance. In this study, for automatizing such manual labor, we propose a deep Auto-Encoder (AE) with a trainable prior. The AE is trained using only a set of single particle Cryo-EM images. The trained AE reconstructs the corresponding structures for the latent variables of the Cryo-EM images. The latent distribution can not only be theoretically proportional to a distribution of the structure but also consistent with the trained prior. Taking advantage of this property, we can automatically compute the pathway by only accessing the latent space as follows: i) generating a ridgeline on the latent distribution and ii) defining the conformational pathway as a sequence of the reconstructed structures along the ridgeline using the trained decoder. In our numerical experiments, we evaluate the computed pathways by comparing them with existing ones that were manually determined by other researchers, and confirm that they are sufficiently consistent.

Keywords: Cryo-EM · Conformational Pathway · Deep Learning · 3D Reconstruction · Structural Biology

K. Yamazaki, Y. Wada, A. Tokuhisa and M. Wada—Represents equal contribution

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43907-0_38.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14220, pp. 394–404, 2023.
https://doi.org/10.1007/978-3-031-43907-0_38

1 Introduction

Over the past few years, Cryo Electron Microscopy (Cryo-EM) has made remarkable progress in biomolecule structure analysis, becoming a major structural analysis technique along with X-ray crystallography. Based on a brief history of the early period of the technique, researchers had proposed methods that reconstruct the static structures of protein molecules from a set of single particle Cryo-EM images, e.g., EMAN2 [24], RELION [20, 21, 31], and cryoSPARC [17, 18]. Methods that reconstruct nonstatic structures have also been proposed, e.g., cryoDRGN [28–30], e2gmm [2], 3DVA [16], and 3DFlex [15]. In these methods, a 3D density map expresses the structure and helps to find biological significance [25, 27].

Recently, taking advantage of method capturing the nonstatic structures, researchers have tried to define a plausible conformational pathway (continuous change with 3D density map) from only single particle Cryo-EM images [10, 26] to efficiently determine the biological significance (e.g., new drug). The pathway is usually defined via the latent space of a trained deep Auto-Encoder (AE) by the Cryo-EM images. For example, in [10], the authors first collect the latent variables, which are the outputs of the trained encoder in cryoDRGN [28], as inputs of the Cryo-EM images. Then, the variables are mapped into a 2D space by UMAP [13]. Thereafter, a sequence of the 2D points is manually constructed via qualitatively evaluating the corresponding structures to those points. The structure is reconstructed by the trained decoder of cryoDRGN. Lastly, the conformational pathway is defined as a sequence of the reconstructed structures along the sequence of the 2D points. Although the protocol of [10] is state-of-the-art, it can be tedious as it involves the manual construction with qualitative evaluation. The following are two of the most important causes: i) for a trained AE of cryoDRGN, there are no theoretical insights into the relation between the latent distribution and the distribution of the structure and ii) the latent distribution of the trained AE is not consistent with the fixed prior, i.e., standard normal Gaussian.

In this study, we propose a deep AE with a trainable prior that is expressed by a Gaussian Mixture Model (GMM). We name the AE *cryoTWIN*¹. CryoTWIN is trained by single particle Cryo-EM images with the estimated pose orientations under an objective inspired by RaDOGAGA [8]. The trained AE reconstructs the nonstatic structures for the latent variables of the Cryo-EM images. A property of the trained AE is that its latent space is theoretically isometric with a space of the structure, where the latent distribution can be proportional to the distribution of the structure. Additionally, the trained prior can fit the latent distribution. This useful property helps us compute a ridgeline on the trained GMM as a sequence of the latent variables and automatically define the conformational pathway by a sequence of the reconstructed structure along the ridgeline.

Our main contributions are as follows: i) we propose cryoTWIN: a deep AE model with the beneficial property for computing the conformational pathway and ii) in our numerical experiments, we confirm that the pathway computed using cryoTWIN is sufficiently consistent with an existing one, that was manually determined by researchers.

¹ The name “TWIN” comes from the isometricity.

In Sect. 2, we describe representatives of methods introduced at the beginning of this section in detail. Furthermore, we explain the differences of cryoTWIN compared with RaDOGAGA and cryoDRGN, as our method is partly inspired by them. In Sect. 3, we give a detailed account of cryoTWIN with theoretical guarantees. Additionally, we introduce an algorithm to compute the conformational pathway. In Sect. 4, we present the results obtained in our numerical experiments using a ribosome dataset. Finally, in Sect. 5, we conclude this study and discuss our future work.

2 Related Work

2.1 Existing Reconstruction Methods

The representative method from the static structure category is cryoSPARC [17, 18]. Given a set of single particle Cryo-EM images and the number of structures, N , this method reconstructs N structures using an efficient stochastic gradient descent technique. Cryo-SPARC has a resolution similar to that of RELION but a smaller computational complexity [20, 21, 31]. The representative method from the nonstatic structure category is cryoDRGN [28–30], whose statistical model is based on spatial-VAE [1] (a variant of VAE [9]). The training objective is to maximize the variational lower-bound. After the training, cryoDRGN can reconstruct continuous structures using the continuous latent variables and the trained decoder.

2.2 RaDOGAGA Revisit

RaDOGAGA [8] is an AE whose latent space is isometric [5] to the input space, inspired by the rate-distortion theory [3]. Let f_θ , g_ϕ , and P_ψ denote an encoder, a decoder, and a trainable prior distribution of latent variables, respectively, where θ , ϕ , and ψ are a set of trainable parameters. Then, the training objective is $\arg \min_{\theta, \phi, \psi} \frac{1}{n} \sum_{i=1}^n \{ \|\mathbf{x}_i - \tilde{\mathbf{x}}_{z_i}\|_2^2 - \beta \log Q_{z_i} \}$, where $\|\mathbf{x}_i - \tilde{\mathbf{x}}_{z_i}\|_2$ (resp. $-\log Q_{z_i}$) corresponds to the distortion (resp. rate). Here, \mathbf{x}_i , ($i = 1, \dots, n$) is the i -th original data. Additionally, $\tilde{\mathbf{x}}_{z_i} = g_\phi(\mathbf{z}_i + \epsilon)$, $\mathbf{z}_i = f_\theta(\mathbf{x}_i)$, and each element of ϵ is uniformly sampled from $[-T/2, T/2]$. Moreover, $\beta > 0$ is a hyper-parameter, and Q_{z_i} is given by

$$Q_{z_i} = \int_{-\infty}^{\infty} U(\mathbf{z} - \mathbf{z}_i) P_\psi(\mathbf{z}) d\mathbf{z}, \quad (1)$$

where $U(\mathbf{z})$ is a rectangular window function: $U(\mathbf{z}) = 1$ if $-T/2 \leq z_j \leq T/2$ holds for all j (z_j is the j -th element of \mathbf{z}), and $U(\mathbf{z}) = 0$ otherwise. Let δ_1 and δ_2 be infinitesimal vectors with arbitrary directions. Then, the optimally trained AE, whose decoder is g_{ϕ^*} , has the following isometric property at all \mathbf{z} for any δ_1 and δ_2 :

$$\underbrace{\langle (\mathbf{z} + \delta_1) - \mathbf{z}, (\mathbf{z} + \delta_2) - \mathbf{z} \rangle}_{\text{abbreviated as } d\mathbf{z}_1 \cdot d\mathbf{z}_2} \propto \underbrace{\langle \hat{\mathbf{x}}_{\mathbf{z}+\delta_1} - \hat{\mathbf{x}}_{\mathbf{z}}, \hat{\mathbf{x}}_{\mathbf{z}+\delta_2} - \hat{\mathbf{x}}_{\mathbf{z}} \rangle}_{\text{abbreviated as } d\hat{\mathbf{x}}_1 \cdot d\hat{\mathbf{x}}_2}, \quad (2)$$

where $\hat{\mathbf{x}}_{\mathbf{z}} = g_{\phi^*}(\mathbf{z})$. Because of Eq. (2), $P(\mathbf{z}) \propto P(\hat{\mathbf{x}}_{\mathbf{z}})$ holds, where $P(\mathbf{z})$ (resp. $P(\hat{\mathbf{x}}_{\mathbf{z}})$) is the latent distribution (resp. the probability density function of $\hat{\mathbf{x}}_{\mathbf{z}}$). The following equation is derived by applying $\delta_1 = \mathbf{z}' - \mathbf{z}$ and $\delta_2 = \delta_1$ to Eq. (2):

$$\|\mathbf{z} - \mathbf{z}'\|_2 \approx 0 \Rightarrow \|\mathbf{z} - \mathbf{z}'\|_2 \propto \|\hat{\mathbf{x}}_{\mathbf{z}} - \hat{\mathbf{x}}_{\mathbf{z}'}\|_2.$$

2.3 Differences Between cryoTWIN and Existing Methods

To understand the difference between cryoTWIN and RaDOGAGA, let V (resp. I) denote a 3D density map (resp. the Cryo-EM image). For cryoTWIN, only I is required to make the latent space isometric to a space of V , whereas V (usually inaccessible) is required to make RaDOGAGA have the same isometricity. Differences between cryoTWIN and cryoDRGN are i) the latent distribution of cryoTWIN is theoretically proportional to a distribution of V , whereas cryoDRGN does not hold such property and ii) cryoTWIN can fit the prior to the latent distribution, whereas cryoDRGN can not.

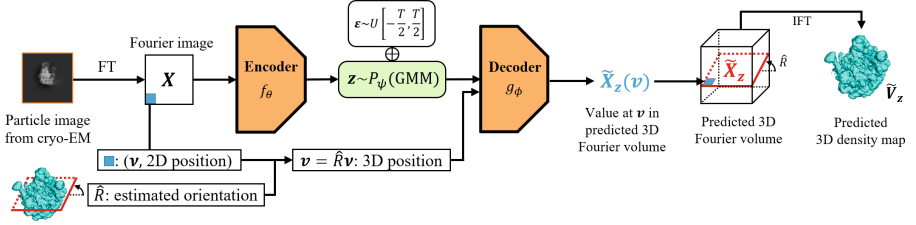


Fig. 1. Diagram of cryoTWIN during training.

3 Proposed Method

We first describe our deep AE, cryoTWIN, with the theoretical guarantees before we explain how to compute the conformational pathway using cryoTWIN. Given a set of single particle Cryo-EM images $\mathcal{I} = \{I_i\}_{i=1}^n$, the AE is trained via two steps: the first step is preprocessing and the second one is to train the AE under a rate distortion theory-based objective, such as RaDOGAGA (see Sect. 2.2). The diagram of cryoTWIN during the training is shown in Fig. 1. In the preprocessing, we estimate the orientation of the Cryo-EM image I_i in the 3D density map, as our training requires the estimated orientation $\hat{R}_i \in \mathbb{R}^{3 \times 3}$. Such \hat{R}_i can be obtained by setting one as the number of models in RELION or cryoSPARC. For the prediction after training, given the latent variable z_i of the image I_i , the trained decoder predicts the corresponding 3D density map to I_i . Further details for the diagram, the training objective, and the prediction after training are presented in Sect. 3.1. The theoretical guarantees are presented in Sect. 3.2. In Sect. 3.3, we describe an algorithm to compute the conformational pathway.

3.1 cryoTWIN

As shown in Fig. 1, a deep AE of cryoTWIN consists of P_ψ , f_θ , and g_ϕ , which denote a trainable GMM (prior), an encoder, and the decoder, respectively. Symbols ψ , θ , and ϕ represent a set of trainable parameters. The parameter ψ is expressed as $\psi := \{(\pi_c, \mu_c, \Sigma_c)\}_{c=1}^C$, where μ_c , Σ_c respectively, indicate the mean and variance of the c -th Gaussian distribution and $\pi_c \in [0, 1]$, $\sum_{c=1}^C \pi_c = 1$ is a weight for the c -th distribution. Additionally, C denotes the number of components.

Diagram During Training (Fig. 1): Let \mathcal{F} be Fourier Transform (FT), and let $X_i = \mathcal{F}I_i \in \mathbb{R}^{D \times D}$ denote the Fourier image of the Cryo-EM image $I_i \in \mathcal{I}$. Given X_i , the encoder f_θ firstly returns $(\mathbf{z}_i, \mathbf{w}_i) = f_\theta(X_i)$, $\mathbf{z}_i \in \mathbb{R}^d$, $\mathbf{w}_i \in \mathbb{R}^C$, where \mathbf{z}_i is the latent variable of X_i . The parameter \mathbf{w}_i is used to enforce the latent empirical distribution to follow P_ψ . Secondly, using \mathbf{z}_i , \hat{R}_i , and a 2D position $\boldsymbol{\nu} = (s, t, 0)^\top$ in X_i , the decoder g_ϕ returns $\tilde{X}_{\mathbf{z}_i}(\mathbf{v}) = g_\phi(\mathbf{z}_i + \boldsymbol{\epsilon}, \mathbf{v})$, $\mathbf{v} = \hat{R}_i \boldsymbol{\nu}$, where $\tilde{X}_{\mathbf{z}_i}(\mathbf{v})$ means the predicted value at the 3D position \mathbf{v} in a 3D Fourier volume and $\boldsymbol{\epsilon} \in \mathbb{R}^d$ is a random noise vector, each of whose elements is uniformly sampled from $[-T/2, T/2]$. By computing $\tilde{X}_{\mathbf{z}_i}(\mathbf{v})$, $\mathbf{v} = \hat{R}_i \boldsymbol{\nu}$ for all 2D positions $\boldsymbol{\nu}$ (i.e., for all pairs (s, t)), the predicted Fourier image $\tilde{X}_{\mathbf{z}_i}$ is defined via the set $\{\tilde{X}_{\mathbf{z}_i}(\mathbf{v})\}_{s,t}$.

Training Objective: We introduce the ℓ -th ($\ell \geq 1$) training with a mini-batch $\tilde{\mathcal{B}} = \{X_i\}_i$, $|\tilde{\mathcal{B}}| = m$, where $X_i = \mathcal{F}I_i$, $I_i \in \mathcal{B} \subset \mathcal{I}$. Firstly, partially inspired by [32], we update ψ of the GMM using $\{(\mathbf{z}_i, \mathbf{w}_i)\}_i$, where $(\mathbf{z}_i, \mathbf{w}_i) = f_\theta(X_i)$, $X_i \in \tilde{\mathcal{B}}$. Secondly, we update θ and ϕ based on an objective inspired by rate-distortion theory, i.e., minimization of i) the distortion (reconstruction error), and ii) the rate. The first minimization problem is defined by $\min_{\theta, \phi} \frac{1}{m} \sum_{i=1}^m \sum_{s,t} (X_i(\mathbf{v}) - \tilde{X}_{\mathbf{z}_i}(\mathbf{v}))^2 \sqrt{s^2 + t^2}$. Let us simplify the problem by $\min_{\theta, \phi} \frac{1}{m} \sum_{i=1}^m \|W \odot (X_i - \tilde{X}_{\mathbf{z}_i})\|_2^2$, where \odot is the Hadamard product, and the (s, t) -th element of $W \in \mathbb{R}^{D \times D}$ is $\sqrt[4]{s^2 + t^2}$. The second minimization problem is defined by $\min_\theta -\frac{1}{m} \sum_{i=1}^m \log Q_{\mathbf{z}_i}$, where $Q_{\mathbf{z}_i}$ is given by Eq. (2). Thus, we solve the combined minimization problem of Eq. (3):

$$\arg \min_{\theta, \phi} \frac{1}{m} \sum_{i=1}^m \left[\underbrace{\|W \odot (X_i - \tilde{X}_{\mathbf{z}_i})\|_2^2}_{\text{distortion (reconstruction error)}} + \underbrace{\beta (-\log Q_{\mathbf{z}_i})}_{\text{rate}} \right], \quad (3)$$

where $\beta > 0$ is a hyper-parameter whose appropriate value depends on \mathcal{I} . For the solver, we use the RAdam optimizer [11].

Prediction After Training: Let ψ^* , θ^* , and ϕ^* be the trained parameters in the deep AE, where $\psi^* := \{(\pi_c^*, \boldsymbol{\mu}_c^*, \Sigma_c^*)_{c=1}^C\}$. Let \mathcal{F}^{-1} denote the Inverse FT (IFT). Given a latent variable \mathbf{z} , the trained decoder g_{ϕ^*} predicts the corresponding 3D density map similar to cryoDRGN [28] by the following procedure. Firstly, $g_{\phi^*}(\mathbf{z}, \mathbf{v})$ is computed for all possible 3D positions $\mathbf{v} \in \mathbb{R}^3$. Secondly, using $\{g_{\phi^*}(\mathbf{z}, \mathbf{v})\}_{\mathbf{v}}$, the predicted 3D Fourier volume is defined before applying \mathcal{F}^{-1} to the Fourier volume. We define the predicted 3D density map by $\hat{V}_{\mathbf{z}}$. Note that the reconstruction error used in Eq. (3) is not a naive one, since we theoretically need to guarantee the isometric property between the latent space and a space of $\hat{V}_{\mathbf{z}}$ in line with Theorem 1.

3.2 Analysis of CryoTWIN

Theorem 1. *If we assume i) $n \gg 1$ for $\mathcal{I} = \{I_i\}_{i=1}^n$ (a set of Cryo-EM images), then Eq. (3) is approximately equivalent to $\arg \min_{\theta, \phi} \frac{1}{m} \sum_{i=1}^m \{\|V_{\mathbf{z}_i} - \hat{V}_{\mathbf{z}_i}\|_2^2 - \beta' \log Q_{\mathbf{z}_i}\}$, where $V_{\mathbf{z}}$ denotes the true 3D density map with a latent variable \mathbf{z} , and $\beta' > 0$ is a hyper-parameter. The symbol $\hat{V}_{\mathbf{z}_i}$ means the predicted 3D density map during the training and it is defined by $\{g_\phi(\mathbf{z}_i + \boldsymbol{\epsilon}, \mathbf{v})\}_{\mathbf{v}}$ and \mathcal{F}^{-1} . Therefore, from Sect. 2.2, the*

Algorithm 1: Pseudocode for computing conformational pathway

Input: Two means of the GMM P_{ψ^*} : μ_i^* as the start and μ_j^* as the end point,
Hyper-parameters: $\omega > 1$ (Larger ω returns less continuous pathway), $K, M \in \mathbb{N}$

Output: A sequence of reconstructed structures from $\hat{V}_{\mu_i^*}$ to $\hat{V}_{\mu_j^*}$

- 1 Set μ_i^* and μ_j^* as to $\mathbf{z}^{i \rightarrow j}(0)$ and $\mathbf{z}^{i \rightarrow j}(K)$, respectively.
- 2 **for** $k = 1, \dots, K - 1$ **do**
- 3 Fix α_i and α_j to $\cos(\frac{\pi k}{2K})$ and $(1 - \cos(\frac{\pi k}{2K}))^\omega$ respectively. Let $\alpha_{-i,j}$ denote $(C - 2)$ -dimensional vector made by removing the i -th and j -th element from the C -dimensional vector $(\alpha_1, \dots, \alpha_C)^\top$. Then, generate M samples for $\alpha_{-i,j}$ under the constraint $\sum_{c \neq i,j} \alpha_c = 1 - \alpha_i - \alpha_j$ and $\alpha_c \geq 0$. Let $\alpha_{-i,j}^{(m)}$ ($m = 1, \dots, M$) denote the m -th sample. For all m , define $\mathbf{z}^{(m)}$ using $\alpha_i, \alpha_j, \alpha_{-i,j}^{(m)}$ by

$$\mathbf{z}^{(m)} = \left(\sum_{c=1}^C \alpha_c \Sigma_c^{*-1} \right)^{-1} \left(\sum_{c=1}^C \alpha_c \Sigma_c^{*-1} \mu_c^* \right).$$
 Then,

$$\mathbf{z}^{i \rightarrow j}(k) = \underset{\mathbf{z}^{(m)}}{\operatorname{argmin}} \underbrace{\frac{P_{\psi^*}(\mathbf{z}^{i \rightarrow j}(0))}{P_{\psi^*}(\mathbf{z}^{(m)})} \|\mathbf{z}^{(m)} - \mathbf{z}^{i \rightarrow j}(k-1)\|_2}_{\text{Cost function for } \mathbf{z}^{(m)} (m=1, \dots, M) \text{ in the latent space}}.$$
- 4 Using $\{\mathbf{z}^{i \rightarrow j}(k)\}_{k=0}^K$ and the trained decoder g_{ϕ^*} , compute the following sequence:

$$\hat{V}_{\mu_i^*} (= \hat{V}_{\mathbf{z}^{i \rightarrow j}(0)}) \rightarrow \hat{V}_{\mathbf{z}^{i \rightarrow j}(1)} \rightarrow \dots \rightarrow \hat{V}_{\mathbf{z}^{i \rightarrow j}(K)} \rightarrow \dots \rightarrow \hat{V}_{\mu_j^*} (= \hat{V}_{\mathbf{z}^{i \rightarrow j}(K)}).$$

latent space of the trained cryoTWIN can be isometric with a space for the predicted 3D density map, i.e., $\forall(\mathbf{z}, \mathbf{z}'); \|\mathbf{z} - \mathbf{z}'\|_2 \approx 0 \Rightarrow \|\mathbf{z} - \mathbf{z}'\|_2 \propto \|\hat{V}_{\mathbf{z}} - \hat{V}_{\mathbf{z}'}\|_2$ & $\forall \mathbf{z}; P(\mathbf{z}) \propto P(\hat{V}_{\mathbf{z}})$ hold, where $P(\mathbf{z})$ (resp. $P(\hat{V}_{\mathbf{z}})$) is the latent distribution (resp. the probability density function of $\hat{V}_{\mathbf{z}}$). Additionally, after the training, if we assume ii) $\forall \mathbf{z}; P_{\psi^*}(\mathbf{z}) \approx \hat{P}(\mathbf{z})$ holds, and iii) $\forall \mathbf{z}; \hat{V}_{\mathbf{z}} \approx V_{\mathbf{z}}$ holds, where $\hat{P}(\mathbf{z})$ is an empirical latent distribution, then, as $\hat{P}(\mathbf{z}) \approx P(\mathbf{z})$ holds due to $n \gg 1$, the following equation approximately holds:

$$\forall(\mathbf{z}, \mathbf{z}'); \|\mathbf{z} - \mathbf{z}'\|_2 \approx 0 \Rightarrow \|\mathbf{z} - \mathbf{z}'\|_2 \propto \|V_{\mathbf{z}} - V_{\mathbf{z}'}\|_2 \text{ \& } \forall \mathbf{z}; P_{\psi^*}(\mathbf{z}) \propto P(V_{\mathbf{z}}). \quad (4)$$

A brief derivation of the theorem is given in Appendix A. The assumption ii) of the theorem is realizable if we set a large integer as the number of components C in the GMM P_{ψ} . This is empirically confirmed in Sect. 4. For assumption iii), as Eq. (3) implies the minimization of $\frac{1}{m} \sum_{i=1}^m \|V_{\mathbf{z}_i} - \tilde{V}_{\mathbf{z}_i}\|_2^2$ if $n \gg 1$, Eq. (3) tends to lead the AE to have the property of $\forall \mathbf{z}; \hat{V}_{\mathbf{z}} \approx V_{\mathbf{z}}$. Thus, if we collect a sufficient amount of the Cryo-EM images, Eq. (4) is realizable for cryoTWIN with large C .

3.3 Computation for Conformational Pathway

We compute the plausible conformational pathway that is defined as a sequence of 3D density maps. If we assume that an AE of cryoTWIN with large C is trained by the Cryo-EM images $\{I_i\}_{i=1}^n$, $n \gg 1$, then because of Eq. (4), the pathway can be defined by the following two steps: i) generating a ridgeline from μ_i^* to μ_j^* on the GMM P_{ψ^*} as a sequence of the latent variables (see the second and third lines in Algorithm 1) and ii) reconstructing the corresponding 3D density maps along the ridgeline using the trained decoder g_{ϕ^*} (see the fourth line in Algorithm 1).

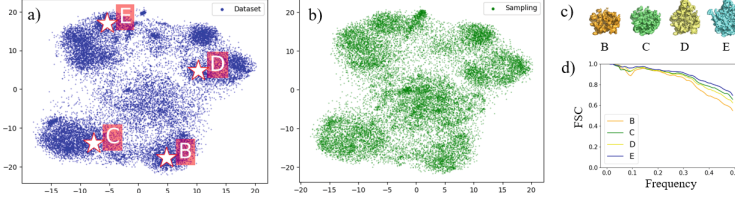


Fig. 2. a) The latent variables (blue) from the trained encoder of cryoTWIN using the ribosomal dataset are visualized by t-SNE. The white star indicates one of means with the trained GMM P_{ψ^*} , and the “B” to “E” indicate the known structural label. We annotate the labels to the corresponding structures (as shown in c)) to the stars via visual examination. b) The sampled points (green) from P_{ψ^*} of cryoTWIN as indicated by t-SNE. d) The similarity score between the two reconstructed structures by cryoTWIN and cryoDRGN with the same label as measured by Fourier Shell Correlation (FSC).

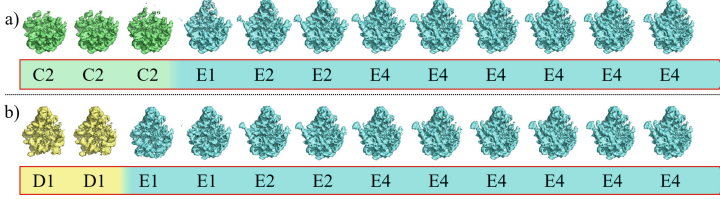


Fig. 3. Two conformational pathways via Algorithm 1 using the ribosomal dataset are shown. In a), the pathway starts from C2 and ends at E4; C2 and E4 indicate the known structural label, and annotated to the structure via visual examination.

A brief explanation of the algorithm is as follows for a sample $z^{(m)}$ of the third line, the sample always ranges a set of candidate points achieving a local maximum with P_{ψ^*} , according to the results of [19]. Additionally, the cost function used to define $z^{i \rightarrow j}(k)$ is inspired by the objective of constructing Max-Flux path [7]. Moreover, from Eq. (4), the cost is proportional to $\frac{P(V_{z^{i \rightarrow j}(0)})}{P(V_{z^{(m)}})} \|V_{z^{(m)}} - V_{z^{i \rightarrow j}(k-1)}\|_2$. Thus, an output of Algorithm 1 can be interpreted as a sequence of 3D density maps, which is generated directly in the space of the 3D density map under Max-Flux objective. We usually cannot access the space of the 3D density map.

4 Numerical Experiment

Using common single particle Cryo-EM images of ribosomes², we conduct two experiments: Expt1 and Expt2. The motivation of Expt1 (resp. Expt2) is to examine whether our theoretical claims work for the ribosomal dataset (resp. whether our computed pathways using the ribosomal dataset are consistent with the existing manually constructed pathways [4]). For both experiments, we employ cryoDRGN [28] as our baseline method. Let $\{I_i\}_{i=1}^n$ denote a set of the Cryo-EM images, where $n = 131899$,

² <https://www.ebi.ac.uk/empair/EMPIAR-10076/>.

and the image size is 128×128 . A set of the corresponding estimated orientations $\{\hat{R}_i\}_{i=1}^n$ is available here³, which are the same ones used in the cryoDRGN. Throughout both experiments, we set $(100, 3, 1/8) = (C, T, \beta)$ with cryoTWIN of Sect. 3.1. The dimension of the latent space is fixed to eight. For Algorithm 1, we set $(11, 2, 1000) = (K, \omega, M)$. Our computational environment is four NVIDIA V100 GPU accelerators with two Intel Xeon Gold 6148 processors.

Expt1: Setting): Firstly, we evaluate the isometricity of cryoTWIN by the correlation coefficient between $d\mathbf{z}_1 \cdot d\mathbf{z}_2$ and $d\hat{V}_1 \cdot d\hat{V}_2$ as described by [8] (see Eq. (2)). Secondly, we evaluate the gap between $P_{\psi^*}(\mathbf{z})$ of cryoTWIN and $\hat{P}(\mathbf{z})$ using both Kullback-Leibler (KL) divergence based on KLIEP [23] and t-SNE visualization [12]. **Results):** Firstly, cryoTWIN achieves a correlation coefficient of 0.77, showing closeness to isometricity (by contrast, cryoDRGN has a correlation coefficient of only 0.52); see details in Fig 4 of Appendix B. Secondly, the KL divergence of cryoTWIN (resp. cryoDRGN) is only 0.25 (resp. 5.7), indicating that $P_{\psi^*}(\mathbf{z}) \approx \hat{P}(\mathbf{z})$ holds (resp. does not hold). Additionally, as shown in Fig. 2 a) and b), the visualized distributions with $P_{\psi^*}(\mathbf{z})$ and $\hat{P}(\mathbf{z})$ of cryoTWIN seem similar, whereas those of cryoDRGN are different; see Fig. 5 of Appendix B. Notably, the large KL divergence with cryoDRGN implies that the conformational pathways based on the fixed prior of cryoDRGN are implausible.

Expt2: Setting): Firstly, as the preliminary experiment of Expt2, using structural labels of [4], we examine how diversified the reconstructed structures of cryoTWIN are by comparing⁴ them with those of cryoDRGN. We annotate the labels to reconstructed structures of cryoTWIN via visual examination and then quantitatively evaluate the accuracies of the annotated labels via Fourier Shell Correlation (FSC) between the two structures reconstructed by cryoTWIN and cryoDRGN with the same label. The annotation is performed by our experts using PyMOL [22]. Additionally, if the FSC is close to one in high and low frequency areas, the two structures are similar enough; see details in the third footnote of [29]. Secondly, we visually examine whether some of the conformational pathways computed by Algorithm 1 can be sufficiently consistent with the sub-paths in the four manually constructed pathways presented in [4]. For the computation, we prepare several pairs (μ_i^*, μ_j^*) satisfying i) both π_i^* and π_j^* are large, and ii) $\hat{V}_{\mu_i^*}$ and $\hat{V}_{\mu_j^*}$ have different labels. The motivation to focus on the sub-paths comes from the following preliminary observations: Given “B” and “E” in Fig. 2 a) as start and end points, Algorithm 1 estimated only one of the four (B-C2-E1-E2-E4-E5 of Fig. 7 in [4]) as the most probable ridgeline. Thus, to reproduce the four pathways, we needed to aggregate sub-paths generated by the algorithm, whose start and end points were significant. **Results):** Firstly, by observing both i) the labeled structures shown in Fig. 2 c) and Fig. 3 (and Fig. 6 of Appendix B), and ii) the labeled structures presented in the original study of cryoDRGN [29] (and Fig. 5 c) of Appendix B), we confirm that cryoTWIN is as diversified as cryoDRGN in terms of the reconstructed structure. Additionally, considering the large FSC value in Fig. 2 d), the accuracies of the annotated labels are high. Secondly, considering the four pathways in [4], we confirm that

³ https://github.com/zhongge/cryodrgn_empiar/tree/main/empiar10076.

⁴ CryoSPARC was used for manually constructing the pathways in [4], but we here employ cryoDRGN, which has the higher structural diversity.

the computed two pathways in Fig. 3 are consistent with the sub-paths in the main four pathways. Note that all the four pathways were successfully reproduced by the aforementioned aggregation in our preliminary experiments.

Time and Memory Complexities: The training time of cryoTWIN (resp. cryoDRGN) is eleven hours (resp. four hours), whereas their memory complexities are comparable. After training cryoTWIN, the running time to compute the pathways including the evaluation time is around one hour, which is much shorter than the running time of the state-of-the-art protocol of a few days [10].

5 Conclusion and Future Work

We propose cryoTWIN for computing plausible pathways from Cryo-EM images, and the efficiency is demonstrated in our numerical experiments. For further research, it would be better to estimate the orientation of the image simultaneously in the training of cryoTWIN, since the preliminary estimation gives an bias to the predicted structure, as explained in [28]. Additionally, it is interesting to combine cryoTWIN and molecular dynamics simulators such as flexible fitting [14], as cryoTWIN provides envelopes for molecular structures in various intermediate states. The combined method could be a powerful tool for more practical applications, such as drug discovery.

Acknowledgements. This work was supported by the FOCUS Establishing Supercomputing Center of Excellence project subject6; MEXT as Simulation- and AI-driven next-generation medicine and drug discovery based on “Fugaku” (JPMXP1020230120). This work used the supercomputer Fugaku provided by RIKEN R-CCS through the HPCI System Research Projects (IDs: hp220078, hp230102, hp230216, ra000018); the supercomputer system at Hokkaido University through the HPCI System Research Projects (IDs: hp220078, hp230102); ABCI provided by AIST was also used.

References

1. Bepler, T., Zhong, E., Kelley, K., Brignole, E., Berger, B.: Explicitly disentangling image content from translation and rotation with spatial-VAE. In: *Advances in Neural Information Processing Systems*, vol. 32. Curran Associates, Inc. (2019)
2. Chen, M., Ludtke, S.J.: Deep learning-based mixed-dimensional gaussian mixture model for characterizing variability in cryo-EM. *Nat. Methods* **18**(8), 930–936 (2021)
3. Cover, T.M.: *Elements of information theory*. John Wiley & Sons (1999)
4. Davis, J.H., Tan, Y.Z., Carragher, B., Potter, C.S., Lyumkis, D., Williamson, J.R.: Modular assembly of the bacterial large ribosomal subunit. *Cell* **167**(6), 1610–1622 (2016)
5. Han, Q.: Isometric embedding of Riemannian manifolds in Euclidean spaces. *American Mathematical Society* (2006)
6. Hsieh, J.: *Computed tomography: principles, design, artifacts, and recent advances*. SPIE Press (2003)
7. Huo, S., Straub, J.E.: The MaxFlux algorithm for calculating variationally optimized reaction paths for conformational transitions in many body systems at finite temperature. *J. Chem. Phys.* **107**(13), 5000–5006 (1997)

8. Kato, K., Zhou, J., Sasaki, T., Nakagawa, A.: Rate-distortion optimization guided auto-encoder for isometric embedding in euclidean latent space. In: Proceedings of the 37th International Conference on Machine Learning, ICML 2020, vol. 119, pp. 5166–5176 (2020)
9. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. In: International Conference on Learning Representations, ICLR 2014 (2014)
10. Kinman, L., Powell, B., Zhong, E., Berger, B., Davis, J.: Uncovering structural ensembles from single-particle cryo-em data using cryodrgn. *Nat. Protoc.* **18**(2), 319–339 (2023)
11. Liu, L., et al.: On the variance of the adaptive learning rate and beyond. In: International Conference on Learning Representations, ICLR 2020 (2020)
12. van der Maaten, L., Hinton, G.: Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008)
13. McInnes, L., Healy, J., Saul, N., Großberger, L.: UMAP: uniform manifold approximation and projection. *J. Open Source Softw.* **3**(29), 861 (2018)
14. Orzechowski, M., Tama, F.: Flexible fitting of high-resolution x-ray structures into cryoelectron microscopy maps using biased molecular dynamics simulations. *Biophys. J.* **95**(12), 5692–5705 (2008)
15. Punjani, A., Fleet, D.: 3D flexible refinement: Structure and motion of flexible proteins from cryo-EM. *Microsc. Microanal.* **28**(S1), 1218–1218 (2022)
16. Punjani, A., Fleet, D.J.: 3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *J. Struct. Biol.* **213**(2), 107702 (2021)
17. Punjani, A., Rubinstein, J.L., Fleet, D.J., Brubaker, M.A.: cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**(3), 290–296 (2017)
18. Punjani, A., Zhang, H., Fleet, D.J.: Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nat. Methods* **17**(12), 1214–1221 (2020)
19. Ray, S., Lindsay, B.G.: The topography of multivariate normal mixtures. *Ann. Stat.* **33**(5), 2042–2065 (2005)
20. Scheres, S.H.: A bayesian view on cryo-EM structure determination. *J. Mol. Biol.* **415**(2), 406–418 (2012)
21. Scheres, S.H.: RELION: implementation of a bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**(3), 519–530 (2012)
22. Schrödinger LLC: The PyMOL molecular graphics system, version 1.8 (2015)
23. Sugiyama, M., Nakajima, S., Kashima, H., Buenau, P., Kawanabe, M.: Direct importance estimation with model selection and its application to covariate shift adaptation. In: Advances in Neural Information Processing Systems 20 (2007)
24. Tang, G., et al.: EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**(1), 38–46 (2007)
25. Wrapp, D., et al.: Cryo-em structure of the 2019-ncov spike in the prefusion conformation. *Science* **367**(6483), 1260–1263 (2020)
26. Wu, Z., Chen, E., Zhang, S., Ma, Y., Mao, Y.: Visualizing conformational space of functional biomolecular complexes by deep manifold learning. *Int. J. Mol. Sci.* **23**(16), 8872 (2022)
27. Yuan, J., Chen, K., Zhang, W., Chen, Z.: Structure of human chromatin-remodelling PBAF complex bound to a nucleosome. *Nature* **605**(7908), 166–171 (2022)
28. Zhong, E.D., Bepler, T., Berger, B., Davis, J.H.: CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nat. Methods* **18**(2), 176–185 (2021)
29. Zhong, E.D., Bepler, T., Davis, J.H., Berger, B.: Reconstructing continuous distributions of 3D protein structure from cryo-EM images. In: International Conference on Learning Representations, ICLR 2020 (2020)

30. Zhong, E.D., Lerer, A., Davis, J.H., Berger, B.: CryoDRGN2: Ab initio neural reconstruction of 3D protein structures from real cryo-EM images. In: International Conference on Computer Vision, ICCV 2021, pp. 4046–4055. IEEE (2021)
31. Zivanov, J., et al.: New tools for automated high-resolution cryo-EM structure determination in RELION-3. *elife* **7**, e42166 (2018)
32. Zong, B., et al.: Deep autoencoding gaussian mixture model for unsupervised anomaly detection. In: International Conference on Learning Representations, ICLR 2018 (2018)