Manifold-based transformation of probability distributions: Application to the inverse problem of reconstructing distributions from experimental data

Tomotaka Oroguchi [1,2,*] Rintaro Inoue [3] and Masaaki Sugiyama[3]

[1]*Department of Physics, Faculty of Science and Technology, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama 223-8522, Japan*

[2]*RIKEN SPring-8 Center, 1-1-1 Kouto, Sayo-cho, Sayo-gun, Hyogo 679-5148, Japan*

[3]*Institute for Integrated Radiation and Nuclear Science, Kyoto University, Kumatori, Sennan-gun, Osaka 590-0494, Japan*

(Received 11 October 2024; accepted 15 May 2026; published 15 June 2026)

Information geometry is a mathematical framework that elucidates the manifold structure of the probability distribution space (**p** space), providing a systematic approach to transforming probability distributions (PDs). In this study, we utilized information geometry to address the inverse problems associated with reconstructing PDs from experimental data. Our initial finding is that the Kullback-Leibler divergence, often considered nonmetric owing to its asymmetry, can serve as a valid metric under specific geometric conditions on the manifold. Based on this finding, we formulated the manifold-based gradient descent (MBGD) method, which was employed to visualize the internal structures—represented as PDs—of two types of systems: those with static constituent elements and those with dynamic state transitions. Through the application of MBGD, we successfully reconstructed the underlying PDs for both types of systems, outperforming the standard gradient descent methods that neglect the manifold structure of **p** space. Therefore, the present results demonstrate the essentiality of accounting for the manifold structure of **p** space in the inverse problems of reconstructing PDs. The ability of MBGD to accurately reconstruct PDs for systems with dynamic state transitions underscores its potential to provide valuable physical insights by visualizing internal structures.

DOI: 10.1103/vkck-1pcm

I. INTRODUCTION

Information geometry [1] is a framework that reveals the manifold structure inherent in the probability distribution space (**p** space). According to Chentsov's invariance theorem [2], the Fisher information metric is the unique Riemannian metric on **p** space [3,4]. Nagaoka and Amari further demonstrated that this metric endows **p** space with a dually affine manifold structure [5,6]. Naturally, this manifold structure results in the Kullback-Leibler (KL) divergence [7], serving as a canonical (albeit asymmetric) divergence between probability distributions (PDs) [8,9]. Since these fundamental mathematical principles define how PDs evolve or should be transformed, i.e., how they move on **p** space, information geometry is now recognized as a fundamental tool not only in information science [10,11] but also in various branches of physics [12], including phase transitions [13–16], complexity [17–19], nonequilibrium statistical mechanics [20–22], and quantum mechanics [23–27].

Despite the invaluable utility of information geometry, its application to inverse problems involving the reconstruction of PDs from experimental data remains largely unexplored. First, we briefly describe the physical definition of the inverse problem. The direct problem is defined as the forward mapping that converts a source PD, which characterizes the internal structure of the system, into observables [28,29]. Then, the inverse problem is defined as the reverse of the direct problem; it aims to reconstruct the hidden internal distribution from the experimental data of the observables [Fig. 1(a)]. In these problems, the forward mapping A is a functional of the PD expressed as

$$A: \int K(m,\ p(\mathbf{x}))d\mathbf{x} = y(m), \tag{1}$$

where $y(m)$, $p(\mathbf{x})$, and $K(m,\ p(\mathbf{x}))$ denote the observable, the PD, and the kernel, respectively. The variables m and $\mathbf{x}$ represent the coordinates in the data space of observables and in the internal distribution space, respectively. The inverse problem holds significant implications across various physical domains [30–50]. Systems in these applications can be broadly categorized into two types: (1) those whose constituent elements remain static and (2) those whose elements undergo state transitions. Systems falling into the former category are of particular interest in engineering applications, such as shape characterization of synthesized nanoparticles [30,31]. From a physics perspective, type (2) systems pose a greater challenge and hold more significance. In these systems, reconstructing PDs requires an understanding of the physical properties of the system elements, while the resulting outcomes contribute to refining this understanding. Such applications are in high demand in soft matter physics [32–43].

*Contact author: oroguchi@phys.keio.ac.jp

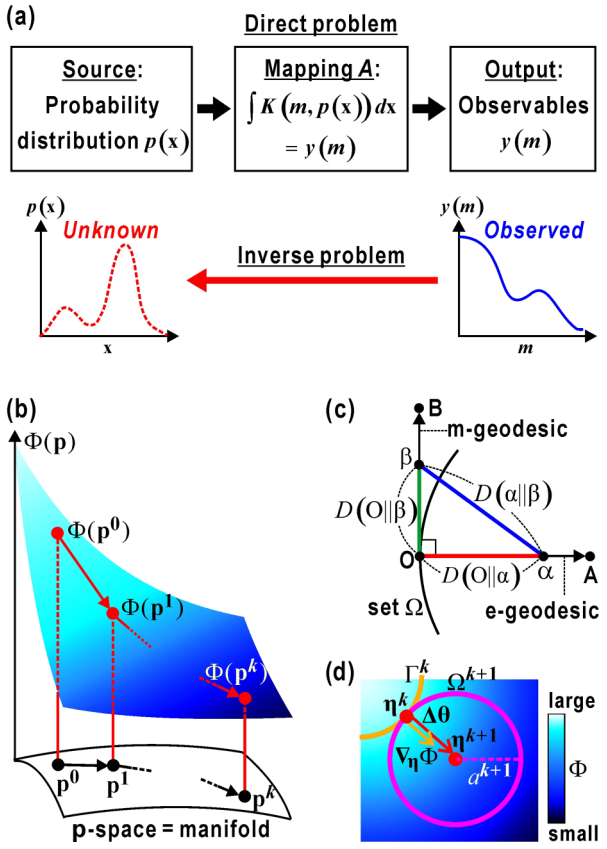

FIG. 1. Schematic of the inverse problem reconstructing of the probability distribution **p** from experimental data. (a) Schematic of the inverse problem in physics. (b) Iterative transformation of **p** from $\mathbf{p}^0$ under an objective functional $\Phi(\mathbf{p})$. (c) Geometry of the transformation path from point O to A in the **p** space. (d) Geometry of the gradient descent in the **p** space.

The inverse problem of reconstructing a PD is formulated as follows: Assuming that the distribution space **x** can be divided into n discrete states, $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$, within each of which the value of kernel K is almost constant, the PD can be represented by a vector $\mathbf{p} = (p_1, \ldots, p_n)$. This discretization corresponds to the projection of a single PD model to a point in **p** space [Fig. 1(b)]. Similarly, the observables can be also represented as a vector $\mathbf{Y} = (y(1), \ldots, y(M))$, with M data points. Let $\mathrm{Y}(\mathbf{p}) = (y(\mathbf{p}, 1), \ldots, y(\mathbf{p}, M))$ be the calculated observables through the forward mapping of **p**, and let $\mathbf{Y}^{\mathrm{EXP}} = (y^{\mathrm{EXP}}(1), \ldots, y^{\mathrm{EXP}}(M))$ represent the experimental data. We define an objective functional as $\Phi(\mathbf{Y}(\mathbf{p})||\mathbf{Y}^{\mathrm{EXP}})$, which evaluates the discrepancy between the model and the experimental data [hereafter, designated as $\Phi(\mathbf{p})$]. The goal of the inverse problem is to find the optimal **p** by minimizing $\Phi(\mathbf{p})$ [Fig. 1(b)]. The key task in this inverse problem lies in the transformation of **p** during minimization, which must adhere to two constraints: the normalization condition of PD, that is, $\sum_{i=1}^{n} p_i = 1$, and the non-negativity of the probability values. These constraints render the minimization process unstable, making the direct reconstruction of PDs from experimental data using the objective functional $\Phi(\mathbf{p})$ a difficult task. Therefore, various domain-specific algorithms have been developed for different applications. However, most approaches require parameter tuning by domain specialists. To achieve wide applicability, a universal approach based on a mathematical understanding of **p** space is essential.

Based on insights from information geometry, the constraints for PD, which mathematically correspond to Chentsov's invariance, transform **p** space into a manifold. Therefore, if transformations of PDs during the minimization process can be conducted to follow this manifold [Fig. 1(b)], the constraints will be automatically satisfied. Such manifold-based transformations would stabilize the minimization process and thereby enable the reconstruction of true PDs. However, the exact form of such transformations remains unclear, as defining a physical quantity to act as a distance on **p** space poses a challenge. In this study, we found that the KL divergence, typically considered nonmetric owing to its asymmetry, can satisfy the properties of a distance under specific geometric conditions. Such condition corresponds to an infinitesimal displacement along the exponential geodesic (e-geodesic) connecting two PDs [Fig. 1(c)]. This finding suggests that in the vicinity of a point on **p** space, the manifold can be locally approximated as a Euclidean space. This approximate Euclidean space allows for the formulation of the gradient descent on **p** space [Fig. 1(d)] as follows:

$$\Delta \ln \mathbf{p} = -\tau \nabla_{\mathbf{p}} \Phi(\mathbf{p}) + C, \tag{2}$$

where $\ln \mathbf{p} = (\ln p_1, \ldots, \ln p_n)$, $C = -\ln(\sum_{i=1}^{n} p_i \exp(-\tau \partial \Phi / \partial p_i))$, and τ denotes the step size. Unlike standard gradient descent methods, the left-hand side of Eq. (2) does not correspond to the Euclidean displacement $\Delta \mathbf{p}$, but rather to the displacement along e-geodesic, $\Delta \ln \mathbf{p}$. This signifies how PDs should be transformed on the manifold. Hereafter, we refer to Eq. (2) as a manifold-based gradient descent (MBGD) method.

To validate our findings on the KL divergence and the MBGD method built upon it, we conducted reconstruction simulations using model systems. When developing algorithms for inverse problems, it is necessary to evaluate not only their effectiveness in minimizing the objective functional but also their accuracy in the reconstruction of the underlying distribution that generates the experimental data [28,29,31,37]. However, in practice, the true distribution is not directly observable in general, making it challenging to assess the reconstruction accuracy using real data. Consequently, validation studies for algorithms often employ simulations using pseudoexperimental data generated from a synthesized pseudotrue model system. In this study, we used a particle size distribution (PSD) in solution as the model system, which corresponds to a type (1) system. We performed simulations in which the PSDs were reconstructed from small-angle x-ray scattering (SAXS) data [Fig. 3(a)]. This model system allowed for a comparison between the standard gradient descent and MBGD in terms of convergence behavior and accuracy in reconstructing the true distribution. The outcomes of our study unequivocally support our theoretical findings on the KL divergence, underscoring the significance of incorporating the manifold structure of **p** space in solving inverse problems. Furthermore, we delved into a method for assessing the inherent ill-posedness in inverse problems.

Based on the results obtained from the PSD model system, we applied the MBGD method to a type (2) system in which

the constituent elements undergo state transitions. Protein conformational ensembles were selected as an example of such soft matter systems. Most proteins adopt multiple conformational states, and the equilibrium between these states, that is, the conformational ensemble, is crucial for their biological functions [51–57]. While molecular dynamics (MD) simulations are commonly utilized to investigate these ensembles [58], the accuracy of MD is limited by the approximate nature of the molecular force fields [59]. Consequently, inverse problem approaches that refine MD-derived ensembles using experimental data are gaining importance [35,36,60]. The results presented in this study demonstrate that applying MBGD to such inverse problems provides a powerful framework for visualizing the components of soft matter systems.

II. THEORY

A. Proof of the geometric condition under which KL-divergence serves as a distance

Here, we define two nearby points O and A in the $\mathbf{p}$ space (or on the manifold), and consider the transformation from $\mathbf{p}^{\mathrm{O}}$ to $\mathbf{p}^{\mathrm{A}}$ [Fig. 1(c)]. The superscripts represent the points on the manifold hereinafter. Our primary objective here is to identify a local coordinate system and transformation path that satisfy the geometric conditions of the Euclidean space.

(1) An affine coordinate system can be established.

(2) For any point α on the path, the distance from point O, $d(\mathrm{O}\|\alpha)$, can be defined, satisfying the following conditions:

(a) Positivity and symmetry: $d(\mathrm{O}\|\alpha) = d(\alpha\|\mathrm{O}) \geqslant 0$.

(b) Pythagorean theorem: $d(\mathrm{O}\|\alpha) + d(\mathrm{O}\|\beta) = d(\alpha\|\beta)$.

(c) Triangle inequality: $\sqrt{d(\mathrm{O}\|\alpha)} + \sqrt{d(\mathrm{O}\|\beta)} \geqslant \sqrt{d(\alpha\|\beta)}$.

The information geometry reveals that the manifold structure of $\mathbf{p}$ space is a dually flat space [5,6], characterized by the Fisher information metric [3,4]:

$$g_n\left(\frac{\partial}{\partial\mu_j}, \frac{\partial}{\partial\mu_k}\right) = \sum_{i=1}^{n} p_i \left(\frac{\partial}{\partial\mu_j}\ln p_i\right)\left(\frac{\partial}{\partial\mu_k}\ln p_i\right). \quad (3)$$

where $\partial/\partial\mu_j$ denotes the tangent vector regarding the local coordinate system $\boldsymbol{\mu}$. In the dually flat manifold, mutually dual affine coordinate systems exist, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{n-1})$ and $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_{n-1})$, expressed as follows:

$$\theta_i = \ln(p_i/p_n) \quad \text{and} \quad \eta_i = p_i. \quad (4)$$

$\boldsymbol{\theta}$ and $\boldsymbol{\eta}$ are called exponential-connection and mixture-connection coordinate systems, respectively. In these coordinate systems, the Fisher information metric becomes

$$g_n\left(\frac{\partial}{\partial\theta_i}, \frac{\partial}{\partial\eta_j}\right) = \delta_{ij}. \quad (5)$$

Since only $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$ are dual affine coordinate systems on the manifold of $\mathbf{p}$ space, the use of either of them is necessary for the Euclidean space condition (1). Here, $\boldsymbol{\eta}$ is selected as a coordinate system, resulting in the introduction of the KL divergence [7–9], which is defined as follows:

$$D(\mathrm{O}\|\alpha) = \sum_{i=1}^{n} p_i^{\mathrm{O}} \ln \frac{p_i^{\mathrm{O}}}{p_i^{\alpha}}, \quad (6)$$

for the two points O and α. D is non-negative and quantifies the difference between the two distributions $\mathbf{p}^{\mathrm{O}}$ and $\mathbf{p}^{\mathrm{A}}$.

Here, we assume that point α lies on the e-geodesic [61] connecting points O and A [Fig. 1(c)]. Since e-geodesic is an affine line with respect to $\boldsymbol{\theta}$, the point α and tangent vector $\boldsymbol{v}$ along this line can be expressed as

$$\theta_i^{\alpha} = \theta_i^{\mathrm{O}} + s\left(\theta_i^{\mathrm{A}} - \theta_i^{\mathrm{O}}\right)(0 \leqslant s \leqslant 1) \quad \text{and}$$
$$\boldsymbol{v} = \sum_{i=1}^{n-1} \left(\theta_i^{\mathrm{A}} - \theta_i^{\mathrm{O}}\right)\frac{\partial}{\partial\theta_i}, \quad (7)$$

respectively. For the Euclidean space condition (1), the transformation along an e-geodesic is essential. Furthermore, we consider the displacement from O to A, $\Delta\mathbf{p}^{\mathrm{A}} = \mathbf{p}^{\mathrm{A}} - \mathbf{p}^{\mathrm{O}}$, to be infinitesimal, enabling the disregard of third- or higher-order terms. Consequently, the KL divergence $D(\mathrm{O}\|\alpha)$ approximately functions as a distance metric for any point α on the e-geodesic between O and A. This point is demonstrated below by proving distance condition (2).

Since $\sum_{i=1}^{n} \Delta p_i^{\alpha} = 0$ owing to the normalization condition, the second-order approximation of the Taylor expansion of the KL divergence (refer to Appendix A) becomes

$$D(\mathrm{O}\|\alpha) \simeq D(\alpha\|\mathrm{O}) \simeq \sum_{i=1}^{n} \frac{\left(\Delta p_i^{\alpha}\right)^2}{2p_i^{\mathrm{O}}} > 0, \quad (8)$$

thereby satisfying condition (a) [62]. Furthermore, a point β is introduced, located on the geodesic mixture [9] (m-geodesic) connecting points O and B [Fig. 1(c)]. The m-geodesic is an affine line with respect to $\boldsymbol{\eta}$. This indicates that point β and tangent vector $\boldsymbol{\omega}$ along this line can be expressed as

$$\eta_i^{\beta} = \eta_i^{\mathrm{O}} + t\left(\eta_i^{\mathrm{B}} - \eta_i^{\mathrm{O}}\right)(0 \leqslant t \leqslant 1) \quad \text{and}$$
$$\boldsymbol{\omega} = \sum_{i=1}^{n-1} \left(\eta_j^{\mathrm{B}} - \eta_j^{\mathrm{O}}\right)\partial/\partial\eta_j, \quad (9)$$

respectively. When the e- and m-geodesics are orthogonal at point O, the inner product of the tangent vectors $\boldsymbol{v}$ and $\boldsymbol{\omega}$, $g_n(\boldsymbol{v}, \boldsymbol{\omega})$, becomes 0. Therefore, by substituting Eqs. (3)–(9) into the aforementioned orthogonal relation, the Pythagorean theorem (b) can be obtained [61]:

$$D(\alpha\|\beta) = D(\mathrm{O}\|\alpha) + D(\mathrm{O}\|\beta) \quad (10)$$

(Appendix B). When the e- and m-geodesics are not orthogonal, the sum of the square roots of the KL divergence becomes

$$\sqrt{D(\mathrm{O}\|\alpha)} + \sqrt{D(\mathrm{O}\|\beta)} \simeq \sqrt{\sum_{i=1}^{n} \frac{\left(\Delta p_i^{\alpha}\right)^2}{2p_i^{\mathrm{O}}}} + \sqrt{\sum_{i=1}^{n} \frac{\left(\Delta p_i^{\beta}\right)^2}{2p_i^{\mathrm{O}}}}$$
$$\geqslant \sqrt{\sum_{i=1}^{n} \frac{\left(\Delta p_i^{\alpha} - \Delta p_i^{\beta}\right)^2}{2p_i^{\mathrm{O}}}}$$
$$\simeq \sqrt{D(\alpha\|\beta)}, \quad (11)$$

satisfying triangle inequality (c). Based on Eqs. (8), (10), and (11), any infinitesimal change along the e-geodesic connecting two points satisfies Euclidean distance conditions (2).

Next, the reverse relationship is observed. When the Euclidean space conditions are satisfied on the line connecting

points O and A, the line becomes e-geodesic. Consider a set of points, denoted as set Ω, whose KL divergences from point A have the same value with $D(\mathrm{O}\|\mathrm{A})$. Since the Euclidean nature is maintained between points O and A, with point O being a member of set Ω, both the set and point A hold Euclidean properties. In Euclidean space, a set of points equidistant from the center forms a hypersphere, with the normal vector on the surface pointing toward the center. Analogously, set Ω also forms a hypersphere, with any normal vector on the surface pointing toward the center, point A. Therefore, the ith component of the line connecting points O and A becomes

$$\{-\nabla_{\boldsymbol{\eta}} D(\mathrm{O}\|\mathrm{A})\}_i = -\frac{\partial D(\mathrm{O}\|\mathrm{A})}{\partial \eta_i} = \theta_i^{\mathrm{A}} - \theta_i^{\mathrm{O}}. \tag{12}$$

A comparison between Eqs. (12) and (7) reveals that the line is e-geodesic.

B. Formulation of MBGD

We formulated the gradient descent method in $\mathbf{p}$ space (MBGD) based on the geometric condition under which the KL divergence serves as a metric. In this formulation, we assume that the geometry of the gradient descent path in a general Euclidean space (Appendix C) holds true between two points on the approximate Euclidean path. We utilize the $\boldsymbol{\eta}$ coordinate system to describe the $\mathbf{p}$ space, with $\boldsymbol{\eta}^k$ and $\boldsymbol{\eta}^{k+1}$ representing the model PDs after the kth and $(k+1)$th transformations through the gradient descent, respectively [Fig. 1(d)]. Furthermore, let Γ^k be the contour of $\Phi(\boldsymbol{\eta})$ passing through $\boldsymbol{\eta}^k$ and let Ω^{k+1} be the hypersphere centered at $\boldsymbol{\eta}^{k+1}$ with radius a^{k+1}, a parameter that needs to be determined. As demonstrated in the gradient descent path within the general Euclidean space (refer to Appendix C), the relationship between two points on the path can be described by the geometric connection between contour Γ^k and hypersphere Ω^{k+1}, which are tangents at point $\boldsymbol{\eta}^k$. In other words, $\boldsymbol{\eta}^k$ represents the extremum at Ω^{k+1} [Fig. 1(d)]. Analogously, the gradient descent for $\mathbf{p}$ space can be determined by identifying the extremum of the Lagrangian expressed as follows:

$$L(\boldsymbol{\eta}^k) = \Phi(\boldsymbol{\eta}^k) + (a^{k+1} - D(k+1\|k))/\tau. \tag{13}$$

The extremum can be determined by solving $\nabla_{\boldsymbol{\eta}^k} L(\boldsymbol{\eta}^k) = 0$ using Eq. (12), resulting in

$$\Delta\boldsymbol{\theta} = -\tau \nabla_{\boldsymbol{\eta}} \Phi. \tag{14}$$

One advantage of Eq. (14) is its simplicity in determining the normalization constant. This can be shown by rewriting Eq. (14) in the form of Eq. (2). From Eq. (4), each component of Eq. (14) becomes

$$\Delta \ln p_i + \tau \frac{\partial \Phi}{\partial p_i} = \Delta \ln p_n + \tau \frac{\partial \Phi}{\partial p_n}. \tag{15}$$

Equation (15) is valid for all $1 \leqslant i \leqslant n-1$. Therefore, the right-hand side can be regarded as the normalization constant C. Consequently, the normalization condition following the transformation establishes the formulation for C, which simplifies Eq. (15) to Eq. (2).

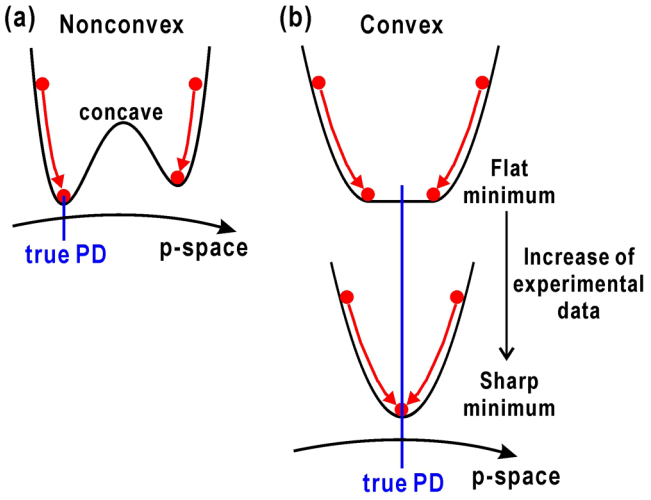

FIG. 2. Schematic of the relation between the uniqueness of the minimum and shape of the objective functional. (a) Nonconvex functional. (b) Convex functional: flat (upper) and sharp (lower) minima.

C. Ill-posedness of the objective functional

In addition to the difficulties associated with the treatment of PDs as reconstruction targets, a typical challenge in inverse problems is their ill-posed nature. This issue arises when the uniqueness of the minimizer of the objective functional is compromised. This uniqueness can be assessed by examining the shape of the objective functional [63–65], whether nonconvex or convex, as shown in Fig. 2. In the case of a nonconvex objective functional, multiple minima may exist, making it challenging to determine the minimum corresponding to the true PD [Fig. 2(a)]. This situation results in an ill-posed inverse problem, where the solution to which minimization converges depends on the initialization point in $\mathbf{p}$ space, that is, the initial PD model.

In contrast, a convex functional, characterized by a positive definite or semidefinite Hessian, has a unique minimum where the true PD may be located [Fig. 2(b)]. In this case, regardless of the initial model utilized, the minimization process is likely to converge to this unique minimum. However, the accuracy of the resulting optimal PD may still be influenced by the shape of the minimum, whether flat or sharp (strictly convex). In the case of a flat minimum [upper panel in Fig. 2(b)], the true PD would be surrounded by other solutions with the same objective functional value, making it difficult to determine which one corresponds to the true PD. In contrast, in the case of a strictly sharp minimum [lower panel in Fig. 2(b)], where the Hessian is positive definite, the minimization process converges to a unique PD. Therefore, the sharpness of the minimum in a convex functional can serve as an indicator of the degree of ill-posedness of the inverse problem. In practice, this sharpness is typically characterized by the number of nonzero eigenvalues of the Hessian [65].

Here, we show that the objective functional $\Phi(\mathbf{p})$ is convex when the experimental data are given as expectation values of observables with respect to the probability distribution. Various types of experimental data can be classified under this case; in particular, many instances where the data are derived from spectroscopic or scattering measurements of soft-matter

systems fall into this category [32–43]. Let $\mathbf{Y}_i = (y_i(1), \ldots, y_i(M))$ represent the observables for the ith state. Then, the mapping A in Eq. (1) takes the form of a Fredholm integral equation of the first kind, expressed as

$$A: \sum_{i=1}^{n} p_i y_i(m) = y(\mathbf{p}, m). \tag{16}$$

In this case, the corresponding inverse problem becomes linear. Since experimental errors are typically Gaussian, the negative log-likelihood of a model $\mathbf{p}$ given experimental data $\mathbf{Y}^{\mathrm{EXP}}$ reduces to χ^2 functions:

$$\chi^2 = \frac{1}{M} \sum_{m=1}^{M} \frac{\left(\sum_{i=1}^{n} p_i y_i(m) - y^{\mathrm{EXP}}(m)\right)^2}{(\sigma^{\mathrm{EXP}}(m))^2}, \tag{17}$$

where $\sigma^{\mathrm{EXP}}(j)$ denotes the error of the jth observable. This function is widely utilized to evaluate the disparity between models and experimental data [66,67]. Utilizing χ^2 as the objective functional $\Phi(\mathbf{p})$ results in a Hessian expressed as follows:

$$(\mathbf{H})_{ij} = \frac{\partial^2 \Phi(\mathbf{p})}{\partial p_i \partial p_j} = \frac{1}{2M} \sum_{m=1}^{M} \frac{y_i(m) y_j(m)}{\{\sigma^{\mathrm{EXP}}(m)\}^2}. \tag{18}$$

Equation (18) shows that the Hessian takes the form of a Gram matrix, which is positive definite, or at least positive semidefinite, at any point in $\mathbf{p}$ space. Therefore, the convexity of the objective functional is ensured in the present case. Then, the remaining issue concerns the sharpness of the minimum of the objective functional. To address this point, in subsequent reconstruction simulations using the model systems, we investigated the change in sharpness with decreasing reconstruction accuracy. The sharpness of the objective functional was assessed through the number of nonzero eigenvalues, Λ_i $(i = 1, \ldots, n)$, of its Hessian.

III. RESULTS

When developing algorithms for inverse problems, it is essential to evaluate their accuracy in reconstructing the underlying distribution that generates the experimental data [28,29,31,37]. To validate our findings on the KL divergence and MBGD method based on it, we conducted simulations using pseudoexperimental data generated from a synthesized pseudotrue system. As such a pseudotrue system, we employed a PSD in solution [Fig. 3(a)], which corresponds to a type (1) system (Sec. II A). Reconstructing PSDs is a well-known inverse problem [30,31], and SAXS data have been proven to be effective for such problems through studies utilizing Monte Carlo simulation approaches [31]. Furthermore, we investigated the relationship between the ill-posedness of the inverse problem and sharpness of the objective functional using the PSD system (Sec. II B).

Then, as an application example of MBGD to a type (2) system—where constituent elements undergo state transitions—we addressed the inverse problem of reconstructing protein conformational ensembles from SAXS data. Building on the simulation results from the PSD system, we initially assessed the feasibility of the approach using a pseudotrue ensemble generated by coarse-grained MD (CGMD) simulations [68] (Sec. II C). Finally, we applied the MBGD method to the experimental SAXS data for *Streptacidiphilus jiangxiensis* glucosamine kinase (SjGlcNK) [69], which undergoes significant conformational changes upon substrate binding, utilizing both all-atom MD (AAMD) and CGMD simulations (Sec. II D).

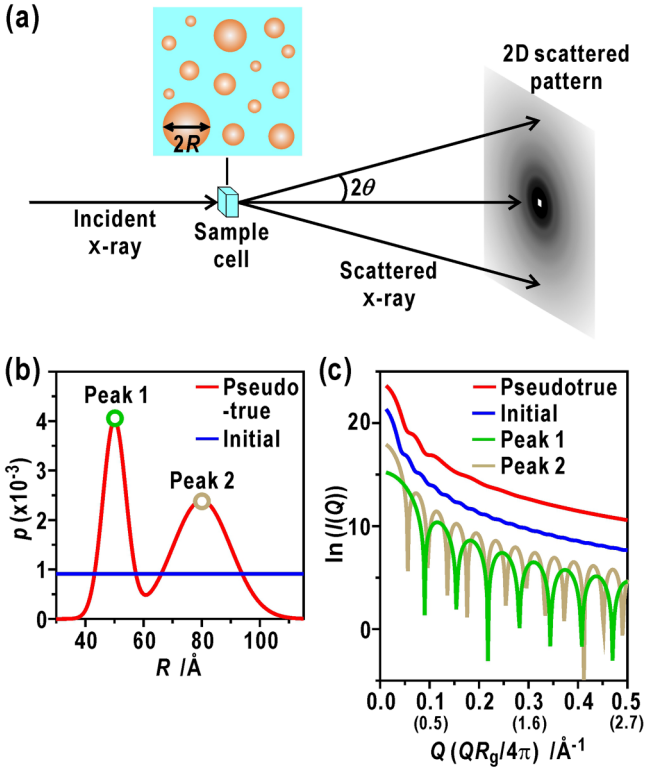

FIG. 3. Setup of the simulations that reconstruct particle size distribution from pseudoexperimental SAXS data. (a) Schematic of the SAXS measurement for PSD in solution [30,31]. (b) Pseudotrue (red) and initial uniform (blue) PSDs. The particles at the peaks in the small and large R areas are referred to as Peak 1 (green) and Peak 2 (brown). (c) SAXS data for the pseudotrue PSD, initial PSD, Peaks 1 and 2.

A. Reconstruction simulation of particle size distribution from SAXS data

First, we outline the setup for the PSD reconstruction simulations. In these simulations, we assumed a dilute solution in which interparticle interference could be neglected. For a pseudotrue PSD, $p^{\mathrm{true}}(R)$, which represents a PD over the particle radius R, we utilized a model comprising a mixture of two Gaussian functions [Fig. 3(b)]. While this Gaussian mixture model may not be realistic as a PSD [31], it allows for a more straightforward interpretation of the validation simulation results [Figs. 4(e), 5(a), and 5(d)]. The $p^{\mathrm{true}}(R)$ was discretized by grouping particles within each 0.2 Å bin along R into a single-size state ($n = 1100$ states). The SAXS data $I_i(Q)$ for the ith particle size R_i were calculated using the scattering function of a sphere, $I_i(Q) = [\Delta\rho V j_1(QR_i)/QR_i\,]^2$ ($Q = 4\pi \sin\theta/\lambda$, where λ and 2θ represent the wavelength of the incident X-ray beam and scattering angle, respectively) [Fig. 3(c)]. V is the volume of a particle. $\Delta\rho$ represents the

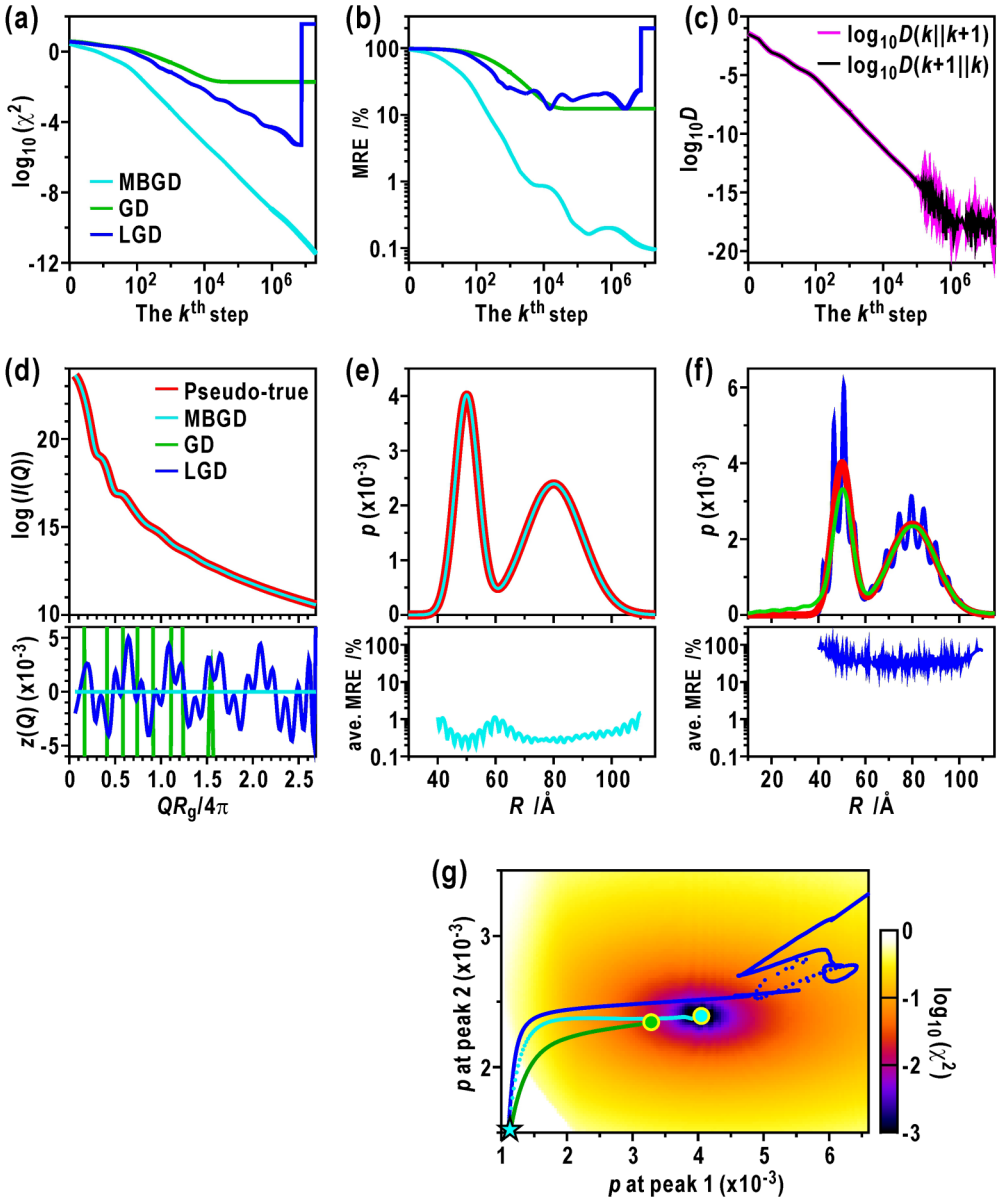

FIG. 4. Results of the simulations that reconstruct PSD using MBGD (cyan), GD (green), and LGD (blue). The minimization processes monitored through (a) χ^2 and (b) MRE. (c) Progress of forward (pink) and backward (black) KL divergences, $D(k||k+1)$ and $D(k+1||k)$, respectively, during the process by MBGD. (d) Comparisons of the pseudoexperimental SAXS data (red) and those obtained using MBGD, GD, and LGD. The residuals between the pseudoexperimental and calculated data were evaluated using $z(Q) = \{I^{\mathrm{CALC}}(Q) - I^{\mathrm{EXP}}(Q)\}/\sigma(Q)$ (lower). (e) Comparison of the pseudotrue PSD and those reconstructed using MBGD. The lower panel shows the binwise MRE averaged over 1000 reconstruction calculations performed with different initial models. (g) Comparison of the pseudotrue PSD and those reconstructed using GD and LGD. The lower panel shows the binwise MRE averaged over 1000 LGD calculations. Results for GD are not shown in the lower panel, as the GD calculations yielded identical reconstructions regardless of the choice of the initial model. (g) Minimization processes projected on the landscape of the objective functional in a two-dimensional space, where the coordinates correspond to the PSD values at Peaks 1 and 2. The star and circle represent the initial and final points of the process in this space, respectively. The color represents the χ^2 value (right bar).

electron density contrast between the particle and solvent, and was set as 1 for simplicity. Subsequently, the pseudoexperimental SAXS data, $I^{\mathrm{EXP}}(Q)$, were determined as expectation values with respect to $p^{\mathrm{true}}(R)$, and thus expressed as $I^{\mathrm{EXP}}(Q) = \sum_{i=1}^{n} p^{\mathrm{true}}(R_i) I_i(Q)$. Furthermore, the SAXS data calculated from $p(R)$ during the minimization process are expressed as $I^{\mathrm{CALC}}(Q) = \sum_{i=1}^{n} p(R_i) I_i(Q)$. For theoretical simplicity, noise was not introduced into $I^{\mathrm{EXP}}(Q)$, resulting in a complete minimization yielding an χ^2 value of 0 [Eq. (17)]. In the subsequent data plots, dependencies on both Q and $QR_g/4\pi$ are represented, where R_g represents the radius of gyration of the particle estimated from $I^{\mathrm{EXP}}(Q)$ using Guinier approximation [70], allowing for discussions independent of the particle size. For the simulations, we utilized SAXS data up to $Q = 0.5\,\text{Å}^{-1}$ ($QR_g/4\pi = 2.69$ and $M = 245$), which aligns with the typical upper limit of Q accessible in standard

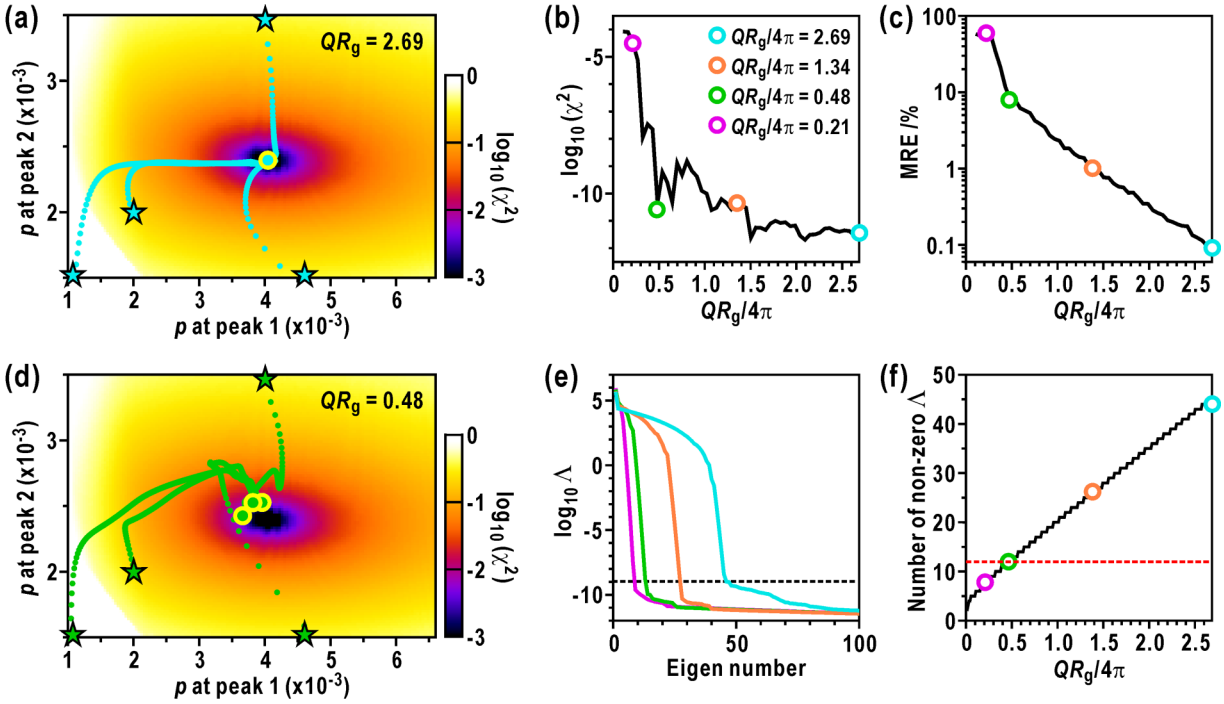

FIG. 5. Dependencies of the minimization and reconstruction accuracies on the amount of pseudoexperimental SAXS data. The amount of the data was reduced by decreasing the upper limit of QR_g. (a) Dependencies of the minimization processes on the initial PSD models when the SAXS data up to $QR_g = 2.69$ were utilized. These processes are projected on the landscape of the objective functional as well as in Fig. 4(e). The star and circle represent the initial and final points in these processes, respectively. (b) Dependencies of the minimized χ^2 value on the amount of the pseudoexperimental SAXS data. (c) Dependencies of the reconstruction accuracy monitored by MRE on the amount of the data. (d) Dependencies of the minimization processes on initial PSD models, when the SAXS data up to $QR_g = 0.48$ were utilized. (e) Distributions of the eigenvalues of the Hessian of the objective functional. The upper limits of QR_g are 2.69 (cyan), 1.34 (orange), 0.48 (green), and 0.21 (pink). The black dashed line represents the lower bound for the nonzero eigenvalues. (f) Dependencies of number of nonzero eigenvalues on the amount of data. The red dashed line represents the limit, below which the reconstruction accuracy becomes worse than 10%.

SAXS experiments. We adopted a uniform distribution as the initial PSD $p^0(R)$. In addition, to examine the dependence of the reconstruction on the initial model, we performed calculations using 1000 randomly generated Gaussian mixture models as the initial PSDs (Supplementary Method S5 and Fig. S1(a) of the Supplemental Material [71]). The accuracy of $p(R)$ was assessed using the model recovery error (MRE) [37], which is expressed as $\text{MRE} = \sum_{i=1}^{n} |p(R_i) - p^{\text{true}}(R_i)|$. The complete reproduction of $p^{\text{true}}(R)$ yielded an MRE value of 0.

In the reconstruction simulation of the PSD system initiated from the uniform distribution, our MBGD method successfully minimized the objective function and accurately reconstructed the pseudotrue PSD, showcasing its theoretical validity (cyan curves in Fig. 4). The excellent ability of the MBGD to minimize the objective functional was validated by the fact that the χ^2 value approached zero asymptotically [Figs. 4(a) and 4(d)]. In terms of accurate reconstruction, the MRE converged to within 0.1% [Fig. 4(b)], showing nearly complete reproduction of the pseudotrue PSD by MBGD [the upper panel in Fig. 4(e)]. The calculations initiated from the different Gaussian mixture models likewise achieved a comparable level of reconstruction accuracy [the lower panel in Fig. 4(e)], indicating negligible uncertainty in the reconstructions. From a practical perspective, MBGD is computationally efficient, requiring only a few seconds to reach 10 000 steps, during which the MRE falls below 1%. These results demonstrate the effectiveness of our manifold-based approach in the inverse problem of reconstructing PDs.

The evolution of the KL divergence during the minimization process also supports our theoretical finding that the KL divergence serves as a metric when the transformation of the PD is an infinitesimal displacement along an exponential geodesic. The equivalence between the forward and backward KL divergences $D(k||k+1)$ and $D(k+1||k)$, respectively, was maintained throughout this process [Fig. 4(c)]. Furthermore, both of the KL divergences were equal to the second-order approximation $\sum_{i=1}^{n} (\Delta p_i^{\alpha})^2 / 2 p_i^{\text{O}}$ (Fig. S2(a) in the Supplemental Material [71]), confirming the validity of the approximation in Eq. (8). To intentionally violate this geometric condition, we increased the step size τ, and beyond a certain threshold, the equivalence deteriorated and the minimization process failed (Figs. S2(b)–S2(d) in the Supplemental Material [71]). Furthermore, for values of τ below threshold, Eq. (8) remains valid. These outcomes further validate our findings regarding KL divergence. From a practical perspective, the upper bound for τ calculated from the aforementioned geometric condition (Note S1 in the Supplemental Material [71]) provides a good estimate of the actual upper bound (Figs. S2(b)–S2(d) in the Supplemental Material [71]). Moreover, when τ exceeded this upper bound, the calculation diverged within the first 100 minimization

steps (Figs. S2(e)–S2(f) in the Supplemental Material [71]). Therefore, the appropriateness of a chosen value of τ can be assessed with negligible computational cost. Consequently, by performing a few short test calculations in the vicinity of the estimated upper bound, one can select an optimal value of τ that avoids divergence without unnecessarily slowing down the MBGD computation due to an excessively small step size (Note S1 in the Supplemental Material [71]).

To further assess the indispensability of our manifold-based approach, we performed reconstruction simulations using conventional gradient descent, assuming a Euclidean geometry in either $\mathbf{p}$ or $\ln \mathbf{p}$ spaces. These two formulations are expressed as follows:

$$\Delta \mathbf{p} \propto -\tau \nabla_{\mathbf{p}} \Phi(\mathbf{p}), \tag{19}$$

and

$$\Delta \ln \mathbf{p} \propto -\tau \nabla_{\ln \mathbf{p}} \Phi(\mathbf{p}). \tag{20}$$

Hereafter, we refer to Eqs. (19) and (20) as the the gradient descent (GD) and log-weight gradient descent (LGD) methods, respectively. In these methods, we adopted the Lagrange multiplier method to satisfy the constraints imposed on $\mathbf{p}$, as well as those of other studies [35,36]. Although both GD and LGD reduced the χ^2 value [Fig. 4(a)], they failed to accurately reconstruct the pseudotrue PSD [Figs. 4(b) and 4(f)]. This failure was independent of the choice of the initial model, and the reconstructed PSDs exhibited substantial variability, indicating uncertainty in the reconstructions [the lower panel in Fig. 4(f)]. Since the pseudoexperimental data here includes no noises, a correct method, such as MBGD, would be expected to achieve complete reproduction. Furthermore, even in their ability to minimize the χ^2 values, both methods are inferior to MBGD. These behaviors of GD and LGD were independent of step size τ.

To ascertain the reason for the failure of Euclidean-based approaches, such as GD and LGD, in achieving accurate reconstruction, we examined how these three methods differ in their minimization process on the landscape of the objective functional [Fig. 4(g)]. To facilitate this analysis, we represented the landscape in a two-dimensional space, with the two coordinates corresponding to the PSD values at Peaks 1 and 2 [Fig. 3(b)]. The landscape was visualized by scanning the parameters of the mixture Gaussian model. As shown in Fig. 4(g), the minimization process in the MBGD successfully converged to the global minimum of the landscape. In contrast, the GD process became stuck before reaching the well containing the global minimum. In the LGD, the process initially approached the well but eventually diverged in an unintended direction. From a mathematical perspective, the gradient descent was formulated assuming that the objective function behaves like a potential function, generating a gradient field that directs the descent toward a minimum [72]. Therefore, the observed behaviors of the GD and LGD processes on the landscape demonstrate that by disregarding the manifold structure of the $\mathbf{p}$ space, these methods compromise the potential-function nature of the objective functional. In contrast, only the MBGD method, which properly considers the manifold structure, maintains this property. Consequently, the incorporation of manifold structure of $\mathbf{p}$ space is essential for inverse problems of reconstructing PDs from experimental data.

B. Evaluation of the ill-posedness of the inverse problem

As shown theoretically in Sec. II C, the objective functional becomes convex when the experimental data are presented as expectation values of observables with respect to the PD, such as $I^{\mathrm{EXP}}(Q)$. The convexity of the objective functional ensures a unique minimum, leading to results that are independent of the initial PD model [Fig. 2(b)]. Indeed, in the reconstruction simulations, utilizing a sufficient amount of SAXS data up to $QR_{\mathrm{g}}/4\pi = 2.69$ ($Q = 0.5$ Å^{-1}), the MBGD calculations consistently converged to the true PSD regardless of the choice of the initial PSD models [Fig. 5(a)]. This outcome validates the theoretical consideration regarding convexity. As discussed in Sec. II C, the ill-posedness in the present inverse problem can be reframed as a concern about the sharpness of the minimum of the objective functional.

Subsequently, we explored how the sharpness of the minimum was affected by reducing the amount of information contained in the pseudoexperimental SAXS data (the amount of SAXS data). The results, demonstrating the independence of minimizations from the initial PSD, indicate that the objective functional is sufficiently sharp to yield a unique and well-defined minimum when using SAXS data up to $QR_{\mathrm{g}}/4\pi = 2.69$. This implies that the SAXS data within this data range contain sufficient information for a unique PSD reconstruction. However, as the upper limit of $QR_{\mathrm{g}}/4\pi$ was gradually decreased, thereby reducing the amount of SAXS data utilized, both the minimization and reconstruction performance deteriorated accordingly [Figs. 5(b) and 5(c), respectively]. Then, the reconstruction accuracy deteriorated rapidly; the MRE exceeded 10%, when the upper limit fell below $QR_{\mathrm{g}}/4\pi = 0.48$. Notably, this limit approximately corresponds to the position of the first dip in the scattering function of the particle corresponding to Peak 1 [Fig. 3(c)], indicating that at least the first oscillation must be observed as the spacing between the dips is inversely proportional to the particle size. These findings suggest that reducing the amount of information contained in the experimental data flattens the minimum of the objective, thereby increasing the ill-posedness of the problem. When utilizing SAXS data up to $QR_{\mathrm{g}}/4\pi = 0.48$, the minimization process was stuck near the edge of the well containing the pseudotrue PSD, and the reconstruction results depended on the initial PSD models [Fig. 5(d)]. Nevertheless, the reconstruction accuracy remained below 10%, which was still better than that achieved by GD or LGD using SAXS data up to $QR_{\mathrm{g}}/4\pi = 2.69$.

Next, to quantitatively evaluate the ill-posedness of the reconstruction, we examined the changes in the eigenvalue distribution of the Hessian with the reduction in the amount of SAXS data [Fig. 5(e)]. A clear correlation is observed between the number of nonzero eigenvalues and ill-posedness [Figs. 5(c) and 5(f)]. When using SAXS data up to $QR_{\mathrm{g}}/4\pi = 2.69$, which provided a strictly sharp minimum of the objective functional, 44 nonzero eigenvalues were observed. This number decreased as the amount of SAXS data decreased, falling below 10 for $QR_{\mathrm{g}}/4\pi \leqslant 0.38$. In this regime, the reconstruction failed with an accuracy worse than MRE = 20%.

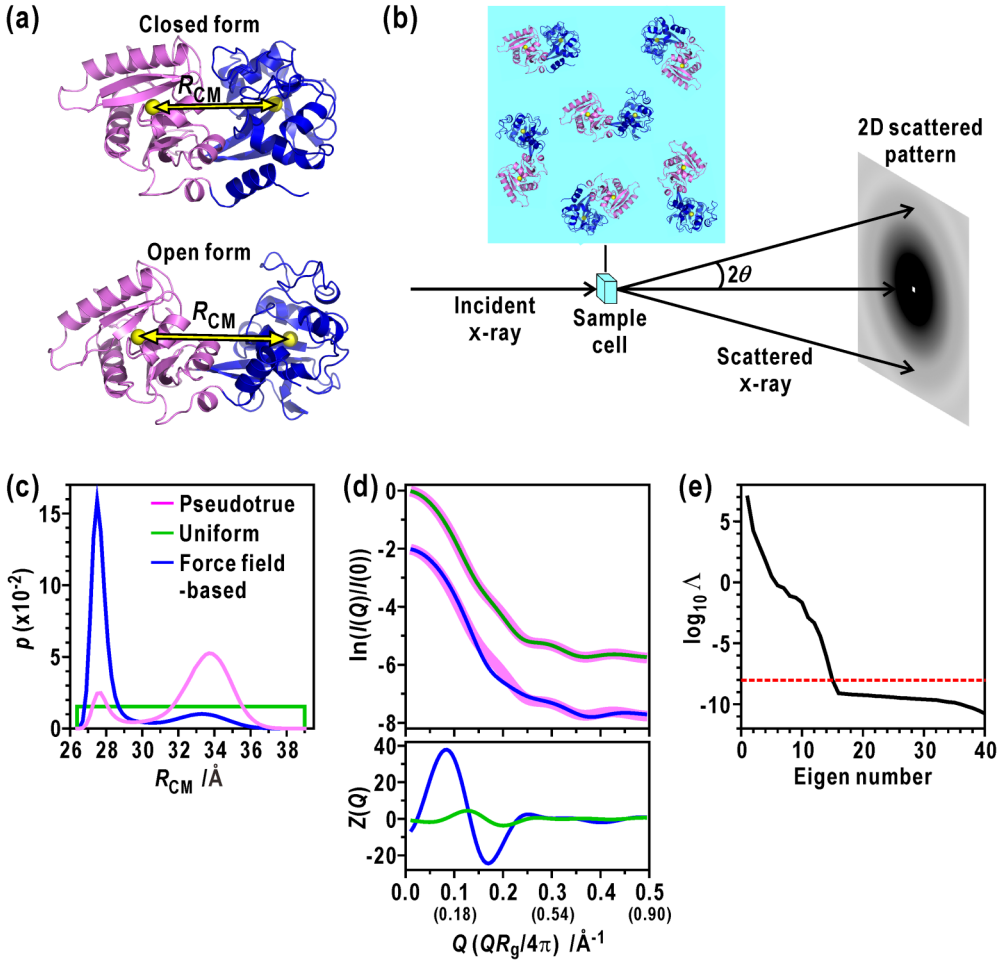

FIG. 6. Setup of the simulations that reconstruct protein conformational ensemble from pseudoexperimental SAXS data. (a) Transferrin as a model system. The crystal structures of the holo and apo forms are presented in the upper and lower panels, respectively (PDB IDs: 1A8E and 1BP5, respectively). The domain distance R_{CM} represents the distance between the mass centers of the two domains. (b) Schematic of the SAXS measurement for protein conformational ensembles in solution [75,76]. (c) Protein conformational ensembles can be represented as PDs on R_{CM} in the case of transferrin, showing the pseudotrue (pink), uniform (green), and force-field-based (blue) PDs. (d) Comparisons of the SAXS data (upper panel) among these PD models. The residuals were evaluated using $z(Q)$ (lower). (e) Distribution of the eigenvalues of the Hessian of the objective functional. The red dashed line represents the lower bound for the nonzero eigenvalues.

Consequently, the number of nonzero eigenvalues can serve as a valuable indicator for assessing the ill-posedness, even in inverse problems of PD reconstruction.

C. Reconstruction simulation of protein conformational ensembles from SAXS data

Next, we conducted simulations in which protein conformational ensembles were reconstructed from the pseudoexperimental SAXS data. As a model system, we utilized transferrin comprising two domains [73,74] [Fig. 6(a)]. Through AAMD and CGMD simulations (see Supplementary Methods S1 and S2, respectively, in the Supplemental Material [71]), we observed that the primary motion of transferrin is shown to be an open-close movement between the domains (domain motions) (Note S2 in the Supplemental Material [71]). In addition, the SAXS data of transferrin predominantly depended on the aforementioned open-close motion, allowing for the simplification the objective functional (Notes S3 and S4 in the Supplemental Material [71]). The red dashed line represents the lower bound for the nonzero eigenvalues of the ensemble as a one-dimensional PD based on the distance between the domains, R_{CM} [Fig. 6(a)], as well as the PSD system. To discretize a PD, $p(R_{CM})$, we grouped conformations within each bin of 0.2 Å along R_{CM} into a single conformational state, as substantial variations in SAXS data were observed among states of this size (Note S4 in the Supplemental Material [71]). The resultant number of states became 65. Hereafter, we employed multiple Gō CGMD simulations (Methods S2 in the Supplemental Material [71]) to generate model ensembles for the reconstruction simulations, for the following two reasons. First, since the structural information provided by SAXS data is of low resolution [75,76], the scattering profiles computed from CGMD ensembles are in good agreement with both experimental SAXS data and those calculated from AAMD simulations [77]. Second, CGMD is capable of generating multiple ensembles with distinct population distributions, making it particularly suitable for the present reconstruction simulations.

As a pseudotrue PD in reconstruction simulations, we generated a model PD with the primary populations located

at open conformational states [Fig. 6(c)]. Two types of PD models were prepared as initial PDs for the reconstruction simulations: nonphysical PDs and force-field-based PDs generated by CGMD. As nonphysical PDs, we employed a uniform distribution [green curve in Fig. 6(c)] and 1000 randomly generated Gaussian mixture models (Method S5 and Fig. S1(b) in the Supplemental Material [71]). As force-field-based PDs, we generated 100 distributions using CGMD with the same physical parameters as those used for the pseudotrue PD. To obtain PDs with different population characteristics, we varied the parameters of the multiple Gō model (Method S5 and Fig. S1(c) in the Supplemental Material [71]). As a representative force-field-based PD, a distribution predominantly populated in the closed conformational states is shown in Fig. 6(c). The uniform PD assumes a lack of prior knowledge regarding the physical principles governing protein motions, whereas the force-field-based PD utilizes the same physical parameters as the pseudotrue PD but with different populations. Despite the limitations of molecular force-field accuracy, protein motions can be computationally reproduced to some extent using MD simulations [58]. The use of the force-field-based PD as the initial ensemble virtually mimics this situation.

In the subsequent reconstruction simulations, the SAXS data calculated from the pseudotrue PD (Method S3 in the Supplemental Material [71]) were utilized as the pseudoexperimental data, with the χ^2 function serving as the objective functional. The experimental errors necessary for the calculation of χ^2 were determined based on the analyses on the SAXS biological data bank [78] (Method S4 in the Supplemental Material [71]). Because experimental SAXS data $I^{\mathrm{EXP}}(Q)$ represent the rotational and conformational averages of single-molecule x-ray scattering [Fig. 6(b)], they can be expressed as expected values [55,75,76]. This renders the objective functional convex. Significant differences in the SAXS data were observed among the pseudotrue, uniform, and force-field-based PDs [Fig. 6(d)], underscoring the effectiveness of SAXS in ensemble reconstruction. Notably, the number of nonzero eigenvalues of the Hessian of the χ^2 function reached 14 when utilizing data up to $QR_{\mathrm{g}}/4\pi = 0.90$ ($Q = 0.5\,\text{Å}^{-1}$) [Fig. 6(e)]. Based on the results of the reconstruction simulations using the PSD system (Fig. 5), while this number is deemed adequate for reconstructing the PD over R_{CM}, it may fall short of achieving complete reconstruction.

Consistent with the PSD system, the MBGD calculation applied to the conformational ensemble of transferrin succeeded in both reducing the χ^2 value [Figs. 7(a) and 7(c)] and reconstructing the pseudotrue PD [Figs. 7(b) and 7(d)] when employing the pseudoexperimental SAXS data up to $QR_{\mathrm{g}}/4\pi = 0.90$. We also conducted reconstruction simulations for other proteins exhibiting domain motions, namely, Mycobacterium tuberculosis EPSP synthase (*MtEPSPS*) [79] and guanylate kinase (GK) [80,81] (Notes S5 and S6, respectively, in the Supplemental Material [71]). For these proteins, the MBGD calculation consistently succeeded in reconstructing the pseudotrue PDs. Furthermore, as in the PSD system, the equivalence between the forward and backward KL divergences was maintained throughout the minimization processes for all these systems, thereby reaffirming the validity of the approximation in Eq. (8) (Note S7 in the Supplemental Material [71]).

However, the calculation results demonstrated a dependence on the initial PDs, likely stemming from the limited number of 14 nonzero eigenvalues. Initiating the MBGD calculations from nonphysical PDs yielded reconstructed PDs with an MRE of approximately 5% and variability in the range of 1%–30%, which is considered satisfactory [Fig. 7(d)]. In contrast, when initiated from force-field-based PDs, the calculations led to near-complete reconstructions with negligible uncertainty, achieving MRE values below 1% and variability in the range of 1%–5% [Fig. 7(e)]. In line with its superior reconstruction accuracy, the latter calculation yielded SAXS data that closely matched the pseudoexperimental data [Fig. 7(c)]. The reconstruction simulations for other proteins provided consistent results: superiority of the force-field-based PD as the initial input (Notes S5 and S6 in the Supplemental Material [71]). These results indicate that the current inverse problem of reconstructing the PDs of protein domain motion from SAXS data is inherently ill-posed. Nevertheless, for the one-dimensional domain motions of transferrin and *Mt*EPSPS, the MBGD calculations initiated from a uniform PD succeeded in achieving reconstruction accuracies exceeding 10% (Fig. 7(d) and Fig. S7(e) in the Supplemental Material [71]). Moreover, utilizing a force-field-based PD as the initial input can lead to near-complete reconstruction.

Next, we investigated the robustness of the reconstruction accuracy with respect to the flatness of the objective functional. Through reconstruction simulations with reduced amounts of pseudoexperimental SAXS data of transferrin (upper panel in Fig. 8), we assessed the number of nonzero eigenvalues of the Hessian of the objective functional (lower panel in Fig. 8) necessary to achieve reconstruction accuracy surpassing MRE = 10%. Remarkably, the force-field-based PD model significantly improved the robustness of the reconstruction accuracy. When the uniform PD was used as the initial input, the required number of nonzero eigenvalues was nine. In contrast, when the force-field-based PD was used, the required number was only four, which was significantly smaller than that required for the PSD system. Consistent results were also obtained for *Mt*EPSPS that undergoes one-dimensional domain motions: the required number of nonzero eigenvalues for the uniform and force-field-based PDs as the initial inputs were 10 and 3, respectively (Fig. S7 in the Supplemental Material [71]).

However, the required number depends on the complexity of protein motion, as exemplified by the case of GK, which undergoes two-dimensional domain motion (Note S6 in the Supplemental Material [71]). While the MBGD calculation initiated from a force-field-based PD succeeded in achieving reconstruction accuracy exceeding MRE = 10%, the calculation initiated from a uniform PD failed to achieve such reconstruction (Figs. S9 and S11 in the Supplemental Material [71]), even though the number of nonzero eigenvalues for the SAXS data was 12 (Fig. S10(e) in the Supplemental Material [71]). These results indicate that the information content contained in SAXS data alone is insufficient for visualization of two-dimensional domain motions in the absence of a

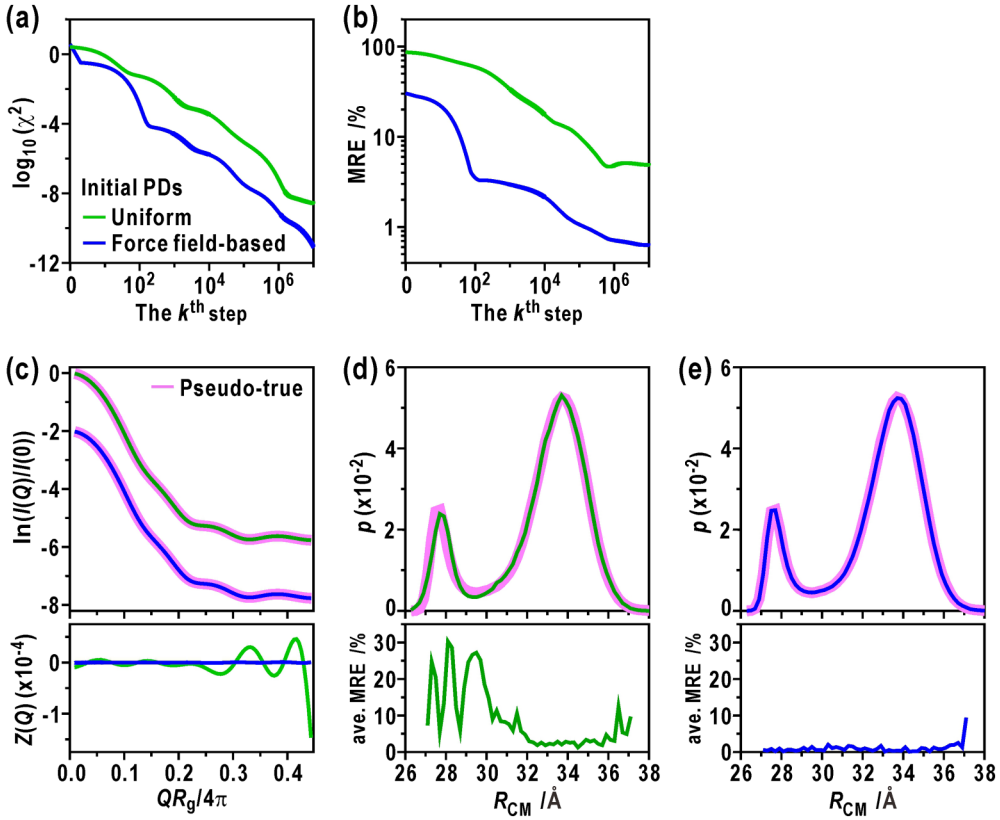

FIG. 7. Results of the simulations reconstructing conformational ensemble of transferrin from its pseudoexperimental SAXS data using the MBGD method. The green and blue represent the results for MBGD, which were initiated from the uniform and force-field-based PDs. Minimization processes monitored through (a) χ^2 and (b) MRE. (c) Comparisons of the pseudoexperimental SAXS data (red) and those obtained using MBGD. Residuals between the pseudoexperimental and calculated data evaluated using $z(Q)$ (lower). (d) Comparison of the pseudotrue PD and those reconstructed using MBGD calculation initiated from nonphysical PD models. The lower panel shows the binwise MRE averaged over 1000 reconstruction calculations performed with different initial models. (e) Comparison of the pseudotrue PD and those reconstructed using MBGD calculation initiated from force-field-based PD models. The lower panel shows the binwise MRE averaged over 100 reconstruction calculations performed with different initial models.

force-field-based PD. One possible approach to address such insufficient number of eigenvalues is to combine multiple sets of experimental data that provide complementary information about the targe PD. To examine the effectiveness of this approach, we performed an MBGD calculation using pseudoexperimental double electron-electron resonance (DEER) data, which provide an information on the distance between a pair of amino-acid residues, in addition to the SAXS data (Note S8 in the Supplemental Material [71]). While DEER data for a single pair of amino-acid residues alone provide little information on two-dimensional domain motions, integrating them with SAXS data results in a drastic increase in the number of nonzero eigenvalues, from 12 to 85 (Fig. S15(c) in the Supplemental Material [71]). This enhancement reflects the fact that DEER data encode information on domain motions that is distinct from that captured by SAXS (Fig. S15(d) in the Supplemental Material [71]), demonstrating that the increase in information content due to additional experimental data is synergistic rather than additive. As a consequence of the increased information content, even the MBGD calculation initiated from the uniform PD achieved reconstruction accuracy exceeding 10% (Fig. S16 in the Supplemental Material [71]). These results also indicate that, for a two-dimensional PD, an experimental information content corresponding to on the order of 80 nonzero eigenvalues is sufficient to render the minimum of the objective functional sharp.

D. Application of MBGD to experimental SAXS data of SjGlcNK

Finally, we applied the MBGD method to actual experimental SAXS data to elucidate the conformational ensembles. Prior to the application, we examined the effect of experimental noise on MBGD reconstruction using pseudoexperimental SAXS data for transferrin (Note S9 in the Supplemental Material [71]). The procedure employed in the application here is based on the results of that assessment. The target sample was SjGlcNK, an enzyme that catalyzes the phosphorylation of D-glucosamine (GlcN) [69]. SjGlcNK is composed of the N-terminal cap, intermediate, and C-lobe domains [Fig. 9(a)]. ATP binds to the active site located in the cleft between the intermediate and C-lobe domains, whereas GlcN binds to the C-lobe domain. X-ray crystallographic analyses revealed that the binding of ATP and GlcN induced domain-closure motion in SjGlcNK [upper panel in Fig. 9(b)], indicating significant

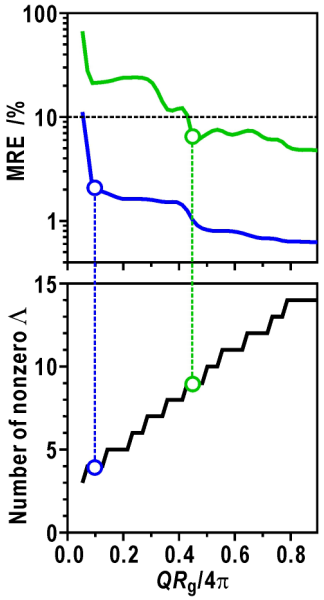

FIG. 8. Dependencies of the reconstruction accuracy monitored using MRE (upper) and the number of nonzero eigenvalues (lower) on the amount of data. The white-filled circles represent the upper bound of the data range. Narrowing the range further results in an MRE exceeding 10% (black dashed line).

conformational flexibility. We conducted 2-μs AAMD simulations to explore the conformational space of SjGlcNK. Principal component analysis (PCA) [82] revealed that the first and second principal components (PCs) corresponded to the open-close and twisting motions of the C-lobe domain relative to the intermediate domain, respectively, in both the AAMD and CGMD trajectories [Fig. 9(b)]. The twisting motion of the second PC was not observed in the crystallographic analyses [69]. The PMF map indicated that the AAMD simulations explored a wider range of conformational space compared with that of the crystal structures [Fig. 9(d)]. We first conducted the MBGD calculation using the AAMD-derived ensemble as the initial input. However, this approach resulted in minimization failure and a reconstructed ensemble with density discontinuities at its edges (Note S10 and Fig. S20 in the Supplemental Material [71]). Such discontinuities indicate insufficient conformational sampling by MD simulations (Note S9 and Fig. S18 in the Supplemental Material [71]). Therefore, to broaden the sampling range, we conducted CGMD simulations using 36 representative structures extracted from the AAMD ensemble as reference conformational states (AA + CGMD) [Fig. 9(e)]. The MBGD calculation utilizing this AA + CGMD-derived ensemble as the initial input successfully reproduced the experimental data [Fig. 9(c)] and reconstructed an ensemble with no discontinuities [Fig. 9(f)], indicating that the AA + CGMD simulations provided sufficiently broad sampling. Furthermore, the number of nonzero eigenvalues of the Hessian of the objective functional was nine. Based on the results of the reconstruction simulations presented in the previous section, this number suggests a reconstruction accuracy that surpasses MRE = 10%. It is noteworthy that the experimental SAXS data could not be explained without the twisting motion identified by AAMD. This result indicates that the realistic motions provided by AAMD are essential for accurate ensemble reconstruction.

The ensemble reconstructed by MBGD revealed that 71.2% of the population was distributed around the closed conformations observed in the crystal structure of the holo form, even in the absence of substrates. This indicates a population shift mechanism [83] for substrate recognition by SjGlcNK. Furthermore, a substantial population was observed at conformations where the active-site cleft was more open than that in the crystal structures of the apo form, referred as a "dropped-jar" conformation hereafter [Figs. 9(i) and 9(j)]. To validate the populations of these dropped-jar conformations, we conducted the MBGD calculation starting from an ensemble that excluded populations of these conformations. The results showed a reconstructed ensemble with discontinuities (Note S10 and Fig. S22 in the Supplemental Material [71]), demonstrating that the experimental data could not be considered without dropped-jar conformations. Notably, the closed and dropped-jar conformations were separated by a free-energy barrier not observed in the AAMD-derived ensemble, highlighting the need for improvements in the force field. Nevertheless, AAMD remains indispensable, as it can sample regions of the conformational space that are inaccessible by crystallographic analyses or CGMD alone.

The subsequent question concerns the reasons for the functional necessity of the dropped-jar conformations. A closer look at the conformational changes in the active-site cleft [Figs. 9(k)–9(n)] can provide some clues. The cleft of the substrate-bound structure was so narrow that the catalytic residue D118 and two magnesium ions were sandwiched between the phosphate groups of the bound ATP [Fig. 9(k)]. However, this sandwichlike structure was maintained in the half-closed crystal conformation of the apo form [Fig. 9(l)]. Consequently, the binding of ATP to the clefts of these crystal structure conformations would likely be difficult owing to the prevention of ATP access by the sandwichlike structure. In contrast, the clefts in the dropped-jar conformations were sufficiently wide to allow ATP access [Figs. 9(i) and 9(j)]. Based on these observations, we hypothesized that the dropped-jar conformations are crucial in facilitating ATP access to the active site of SjGlcNK.

In summary, the application of MBGD to the experimental SAXS data of SjGlcNK allows for the successful reconstruction of the conformational ensembles of this enzyme. The reconstruction results revealed a population-shift mechanism for substrate binding in this enzyme. Furthermore, through the reconstructed ensemble, we identified populations of ATP-accessible conformations that could not be detected through crystallographic analyses or MD simulations alone.

IV. DISCUSSION AND CONCLUSION

In this study, we applied the information geometry framework to inverse problems for reconstructing PDs from experimental data. First, we theoretically demonstrated that the KL divergence between two PDs can serve as a distance when the transformation between them corresponds to an infinitesimal displacement along an e-geodesic on the **p**-space manifold. Based on these findings, we formulated the

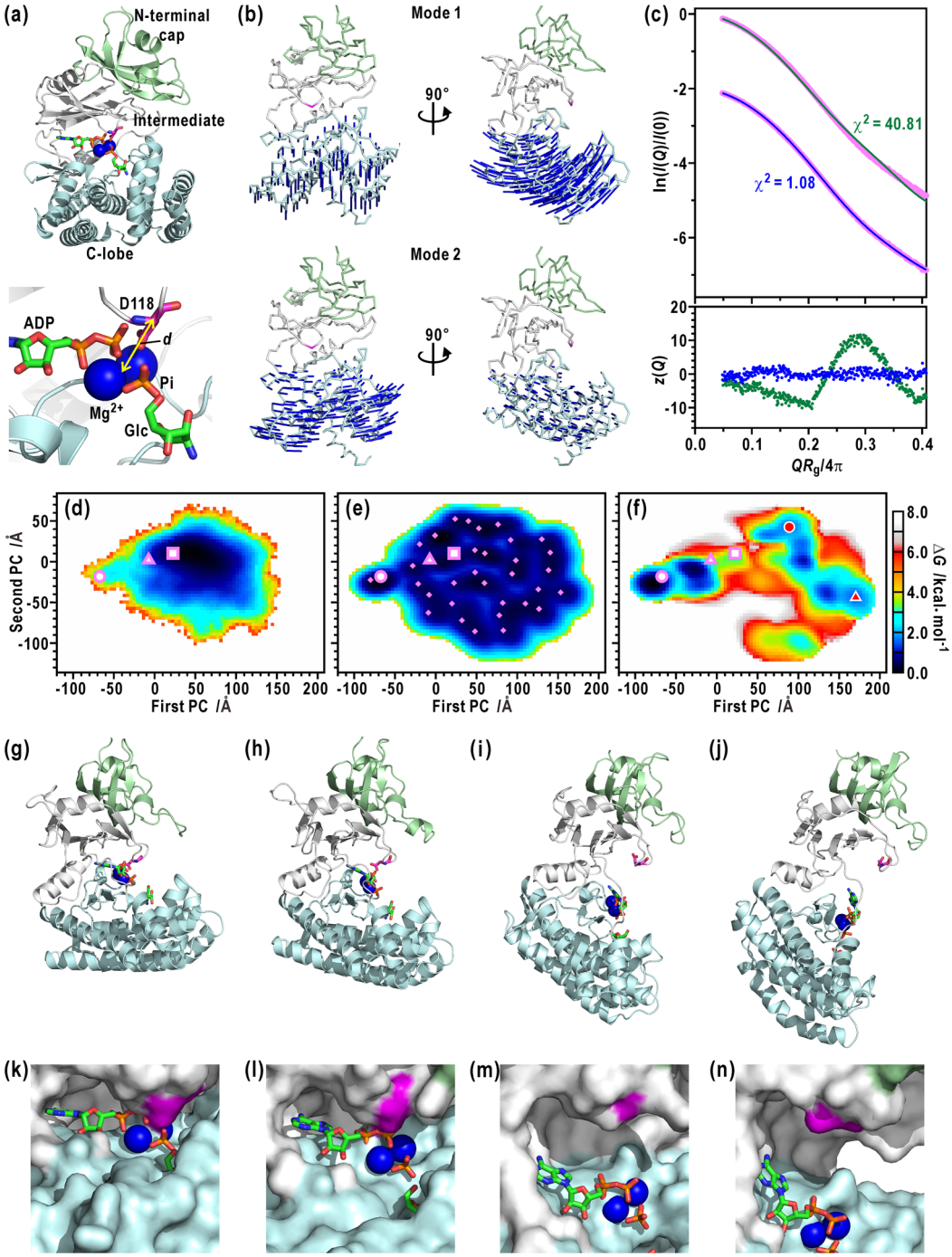

FIG. 9. Application of the MBGD method to experimental SAXS data of SjGlcNK under substrate-free conditions. (a) Crystal structure of the holo form of SjGlcNK (PDB ID: 6HWL); the entire structure (upper panel) and active-site cleft (lower panel). The domain structures are shown using different structures. (b) Domain motions of SjGlcNK identified using the PCA of the AAMD trajectory: the first PC mode (upper panel) and second PC mode (lower panel). The left and right panels represent the front and side views, respectively. The direction of C_α displacements along these PC modes is represented using blue arrows. (c) Comparison of the experimental (pink) and calculated SAXS data. The green and blue curves represent the data calculated from the AA + CGMD and reconstructed ensembles, respectively. (d)–(f) Ensembles projected on the first and second PC modes: the AAMD-, AA + CGMD-, and MBGD-reconstructed ensembles, respectively. The ensembles are visualized using a map plotting the potential of the mean force (PMF). The crystal structure of the holo form is represented by the pink circle, and those of the two apo forms are presented by the pink triangle and square. The structures utilized for the input of the AA + CGMD simulations are represented by the small pink diamonds. The two "dropped-jar" conformations are represented by the red circle and triangle. (g)–(j) Side views of the entire structure of the conformations shown in the PMF map [Fig. 9(f)]: crystal structure of the holo form [pink circle in Fig. 9(f)], the crystal structure of the apo form (pink triangle), dropped-jar conformation 1 (red circle), and dropped-jar conformation 2 (red triangle), respectively. (k)–(n) Closer views of the active-site cleft in the aforementioned conformations. The order is consistent with that of the entire structure. The D118 residue is represented in purple.

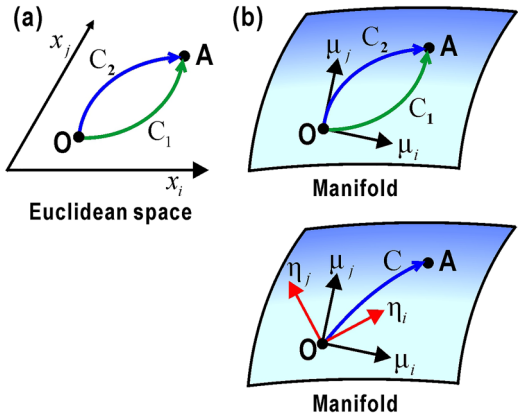

FIG. 10. Schematics of the line integration of gradient vector in Euclidean space [panel (a)] and manifold [panel (b)].

MBGD. The reconstruction simulations of the PSD system revealed that, unlike standard gradient descent methods, the MBGD method maintains the potential-function nature of the objective functional. This unique property enables MBGD to successfully achieve both the minimization of the objective functional and the accurate reconstruction of the true PD.

Observations of the GD and LGD processes in the landscape revealed that these standard gradient descent methods compromised the potential-function nature of the objective functional, unlike the MBGD method [Fig. 4(g)]. To provide a theoretical explanation for this compromise, we consider the following setup. Two fixed points, O and A, exist in either the Euclidean space or a manifold, and the gradient field $\mathbf{F} = -\nabla\Phi$ of the objective functional Φ is integrated along two arbitrary paths, C_1 or C_2 (Fig. 10). In the case of the Euclidean space $\mathbf{x}$ [Fig. 10(a)], the potential-function property of Φ implies that the integral is independent of the integration path:

$$-\int_{C_1} \mathbf{F}\cdot d\mathbf{x} = -\int_{C_2} \mathbf{F}\cdot d\mathbf{x} = \Phi \tag{21}$$

In contrast, on manifolds [Fig. 10(b)], owing to the curvature and torsion inherent in space [84], the integral of the gradient field $\mathbf{F}$ expressed in Eq. (21) depends on the choice of path and local coordinate systems, $\boldsymbol{\mu}$ and $\boldsymbol{\eta}$, such that

$$-\int_{C_1} \mathbf{F}\cdot d\boldsymbol{\mu} \neq -\int_{C_2} \mathbf{F}\cdot d\boldsymbol{\mu} \neq \Phi \quad \text{or} -\int_{C} \mathbf{F}\cdot d\boldsymbol{\mu} \neq -\int_{C} \mathbf{F}\cdot d\boldsymbol{\zeta} \neq \Phi. \tag{22}$$

Based on Eq. (22), neglecting the manifold structure of space compromises the potential-function nature of Φ. In the case of a general Riemannian manifold $\boldsymbol{\zeta}$ with the metric tensor $\mathbf{G}(\boldsymbol{\zeta})$, the gradient can be calibrated as $\mathbf{G}^{-1}(\boldsymbol{\zeta})\nabla_{\boldsymbol{\zeta}}\Phi(\boldsymbol{\zeta})$ [85], allowing for the recovery of the equality in Eq. (21). However, implementing this correction in $\mathbf{p}$ space remains challenging owing to the unknown normalization constant [86]. The MBGD method circumvents this challenge by incorporating the manifold structure with no gradient correction.

In this study, we employed an approach that transforms the discretized form of PD through displacements on the manifold of the $\mathbf{p}$ space. Alternatively, the PD can be expressed in an analytical form, where the transformation of its shape is realized by tuning the parameters of the function. The state-of-the-art analytical forms of such approaches are the generative models employed in machine learning [63,87,88]. Since the generative models have not yet been applied to inverse problems reconstructing internal distributions from experimental data, such an extension would be worth investigating in future studies. In such extension, it would be necessary to select properly the type of generative models and its parameters, whereas the present approach, which directly optimizes the probability values, is simpler and more straightforward. Our simple approach based on a discretized representation of the PD is also advantageous from a computational cost. The system used in our reconstruction simulations spans numbers of discrete states ranging from 60 to 2000; nevertheless, with an appropriate choice of the step size τ, reconstruction accuracies surpassing MRE $= 10\%$ were achieved within a few seconds for all systems (Table S1 in the Supplemental Material [71]).

The present applications of MBGD focus on the linear inverse problems, in which the experimental data are given as expectation values of observables. Scattering and spectroscopic data obtained from soft-matter systems, such as protein conformational ensembles, fall into this category of linear inverse problems [32–43]. In contrast, nonlinear inverse problems arise in experimental data associated with imaging the internal structure of a single object [89–91]. In nonlinear inverse problems, the Hessian of the objective functional $\Phi(\mathbf{p})$ does not take the form of a Gram matrix; as a consequence, the objective functional can become nonconvex, allowing multiple minima to exist. In such cases, locating the global minimum requires broad sampling of the objective functional over $\mathbf{p}$ space. A representative approach for such sampling is the Monte Carlo method. However, even during such stochastic sampling procedures, the PD must satisfy the normalizing and non-negativity constraints. Therefore, information geometry is expected to be useful for sampling PDs under these constraints. For example, Eq. (14) in the MBGD formulation indicates that such Monte Carlo sampling can be naturally carried out in the dual affine coordinate system.

As an application of MBGD to type (2) systems, whose constituent elements undergo state transitions, we conducted reconstructions of protein conformational ensembles from experimental SAXS data through the MBGD method. Our simulations have shown that accurate reconstruction with accuracy surpassing MRE $= 10\%$ is achievable when a physics-based PD is used as the initial input to MBGD. These findings underscore the significance of incorporating *a priori* physical knowledge of system components in reconstructions. The effectiveness of physics-based PDs has also been demonstrated in other studies [35,36,60]. In the reconstruction of protein conformational ensembles, MD-derived ensembles have been used as initial models and have succeeded to some extent in explaining experimental data, even without the use of manifold-based techniques.

Furthermore, the reconstructed PD can contribute to refining the physical understanding of the system by comparing it with the physics-based PD model utilized as the initial input. In the case of SjGlcNK, the energy barrier between the open and closed conformational states observed in the

reconstructed ensemble was not present in the AAMD-derived ensemble, meaning that the interdomain interactions through water molecules between domains were not properly captured by the current molecular force field [92,93]. In future studies, accumulating such discrepancies between physics-based model and reconstructed ensembles will be beneficial for the development of more accurate molecular force fields. Nevertheless, AAMD is indispensable for obtaining physics-based conformational ensemble required for the reconstruction of the true ensembles as the initial input.

The analyses conducted in this study on the ill-posedness of inverse problems suggest that the number of nonzero eigenvalues of the Hessian of the objective functional serves as a useful indicator of ill-posedness. For example, the reconstruction simulations for the PSD system and for protein domain-motion systems with a single degree of freedom indicate that, for a one-dimensional PD, the objective functional becomes nearly strictly convex when the number of nonzero eigenvalues is on the order of 10–20 (Figs. 5 and 7). Because SAXS data contain an experimental information content of this magnitude, MBGD calculations can reconstruct such PDs from SAXS data alone with accuracies better than 10% MRE, largely independent of the choice of the initial PD model, indicating low uncertainty in the reconstructions. In contrast, for protein domain motions with two degrees of freedom, an information content corresponding to about ten nonzero eigenvalues is insufficient to render the objective functional strictly convex, leading to a strong dependence of the reconstruction on the initial PD. Nevertheless, the MBGD method remains effective for identifying the most probable PD model when insufficient experimental data is combined with a physics-based PD model. Furthermore, by incorporating additional experimental data that provide information complementary to SAXS, such as DEER data, accurate reconstruction can be obtained even when starting from a uniform PD (Note S8 in the Supplemental Material [71]). A quantitative assessment of how such integration of multiple experimental data improves reconstruction accuracy and reduces uncertainty remains an important subject for future study.

In practice, the PDs in many physical systems are more complex, requiring higher-dimensional descriptions. Such examples for biomolecular conformational ensembles include multidomain proteins composed of more than two domains [94,95], intrinsically disordered proteins (IDPs) [96,97], and nuclear biomolecules [98,99]. For such complex systems, the experimental information content contained in SAXS data alone is insufficient to achieve accurate reconstruction. An effective strategy for addressing insufficient information content is to integrate different types of experimental data. As illustrated by the example in which SAXS data were integrated with DEER data, integrating experimental data that encode distinct structural information leads to a synergistic increase in information content (Note S8 in the Supplemental Material [71]). Moreover, the reconstruction simulations using such integrated data demonstrate that, by leveraging this increased experimental information content, the MBGD method can accurately reconstruct PDs even as the dimensionality of the system increases.

For such data integration, a variety of experimental techniques are available for studying biomolecular conformational ensembles, including solution neutron scattering [56,100,101], nuclear magnetic resonance [102–105], DEER spectroscopy [37,38], and cryo-electron microscopy [106,107]. Indeed, the integration of these diverse data sources yields significant benefits for the structural studies of IDPs [108–110]. Furthermore, as exemplified by the recent establishment of PDB-IHM [111], the importance of integrative approaches that combine diverse experimental and computational methods has become widely recognized in structural studies of complex biomolecules. However, even within the PDB-IHM, applications of such integrative approaches to conformational ensembles remain relatively rare. Integrative methods such as MBGD in this study are therefore important, as they provide a framework that enables such applications. In such applications, it will be necessary to assess which combinations of experimental data can yield a synergistic increase in information content, through reconstruction simulations such as those performed in this study. However, the primary objective here is to provide a theoretical demonstration that true PDs can be reconstructed from experimental data within the framework of information geometry. Systematic validation aimed at expanding such applications therefore remains an important subject for future studies.

In summary, this study demonstrated that the linear inverse problem of reconstructing PDs from experimental data is fundamentally solvable when the manifold structure of $\mathbf{p}$ space is properly considered. The application of MBGD to the actual experimental SAXS data of SjGlcNK enabled the visualization of functionally important conformational ensemble that cannot be assessed by crystallographic analyses or MD simulations alone, thereby demonstrating the practical utility of MBGD for reconstruction problems. PDs in type (2) systems, particularly in soft matter [32–43], are inherently difficult to observe directly, resulting in limited accumulation of visualization data. This limitation also poses challenges for the direct application of machine-learning approaches. In such contexts, inverse-problem-based methodologies, such as those presented in this study, provide an essential framework for visualizing PDs from experimental data. The collection of visualization data acquired through these approaches is expected to enhance our comprehension of physical systems and, in turn, facilitate the effective application of machine-learning techniques. We anticipate that manifold-based techniques, including MBGD, will be crucial in extending the applicability of inverse-problem approaches for visualizing PDs in type (2) systems beyond protein conformational ensembles.

ACKNOWLEDGMENT

This study was supported by Grants-in-Aid for Scientific Research (KAKENHI) from the Japan Society for the Promotion of Science (JSPS) under Grants No. 26104535, No. 26800227, No. 17K19209, No. 17H04854, and No. 21K03489 (to T.O.), and No. 18H05229 (to T.O., R.I., and M.S.).

DATA AVAILABILITY

The data that support the findings of this article are openly available [112].

APPENDIX A: TAYLOR EXPANSION OF THE KL DIVERGENCE

The Taylor expansions of the forward and backward KL divergences from point O to point α in $\mathbf{p}$ space, with respect to the displacement $\Delta\mathbf{p}^\alpha$, can be written as follows:

$$D(\mathrm{O}\|\alpha) = \sum_{i=1}^{n} \Delta p_i^\alpha + \sum_{i=1}^{n}\sum_{j=2}^{\infty} \frac{(-1)^j}{j} \frac{\left(\Delta p_i^\alpha\right)^j}{\left(p_i^{\mathrm{O}}\right)^{j-1}} \tag{A1}$$

and

$$D(\alpha\|\mathrm{O}) = \sum_{i=1}^{n} \Delta p_i^\alpha + \sum_{i=1}^{n}\sum_{j=2}^{\infty} \frac{(-1)^j}{(j-1)j} \frac{\left(\Delta p_i^\alpha\right)^j}{\left(p_i^{\mathrm{O}}\right)^{j-1}}, \tag{A2}$$

respectively. Therefore, assuming $\Delta\mathbf{p}^\alpha$ to be infinitesimal, the second-order approximation of the above expressions leads to Eq. (8).

APPENDIX B: PROOF OF PYTHAGOREAN THEOREM FOR POINTS ON E- AND M-GEODESICS

Here, we demonstrate that the points on e- and m-geodesics satisfy the Pythagorean theorem, as indicated in Eq. (10) when these geodesics are orthogonal. The proof utilizes points, O, α, and β, as referenced in the main text. By substituting Eqs. (5), (7), and (9) into the Fisher information metric, Eq. (3), we obtain

$$\sum_{i=1}^{n-1} \left(\theta_i^\alpha - \theta_i^{\mathrm{O}}\right)\left(\eta_i^\beta - \eta_i^{\mathrm{O}}\right) = 0. \tag{B1}$$

By substituting Eq. (4), the left-hand side of Eq. (A1) can be expressed using KL divergence as follows:

$$\begin{aligned}
&\sum_{i=1}^{n-1} \left(\theta_i^\alpha - \theta_i^{\mathrm{O}}\right)\left(\eta_i^\beta - \eta_i^{\mathrm{O}}\right) \\
&= \left(\sum_{i=1}^{n-1} \theta_i^\alpha \eta_i^\beta + \ln p_n^\alpha\right) - \left(\sum_{i=1}^{n-1} \theta_i^\alpha \eta_i^{\mathrm{O}} + \ln p_n^\alpha\right) \\
&\quad - \left(\sum_{i=1}^{n-1} \theta_i^{\mathrm{O}} \eta_i^\beta + \ln p_n^{\mathrm{O}}\right) + \left(\sum_{i=1}^{n-1} \theta_i^{\mathrm{O}} \eta_i^{\mathrm{O}} + \ln p_n^{\mathrm{O}}\right) \\
&= \left\{\sum_{i=1}^{n-1} p_i^\beta \ln p_i^\alpha + \left(1 - \sum_{i=1}^{n-1} p_i^\beta\right) \ln p_n^\alpha\right\} \\
&\quad - \left\{\sum_{i=1}^{n-1} p_i^{\mathrm{O}} \ln p_i^\alpha + \left(1 - \sum_{i=1}^{n-1} p_i^{\mathrm{O}}\right) \ln p_n^\alpha\right\} \\
&\quad - \left\{\sum_{i=1}^{n-1} p_i^\beta \ln p_i^{\mathrm{O}} + \left(1 - \sum_{i=1}^{n-1} p_i^\beta\right) \ln p_n^{\mathrm{O}}\right\} \\
&\quad + \left\{\sum_{i=1}^{n-1} p_i^{\mathrm{O}} \ln p_i^{\mathrm{O}} + \left(1 - \sum_{i=1}^{n-1} p_i^{\mathrm{O}}\right) \ln p_n^{\mathrm{O}}\right\} \\
&= \sum_{i=1}^{n} p_i^\beta \ln p_i^\alpha - \sum_{i=1}^{n} p_i^{\mathrm{O}} \ln p_i^\alpha \\
&\quad - \sum_{i=1}^{n} p_i^\beta \ln p_i^{\mathrm{O}} + \sum_{i=1}^{n} p_i^{\mathrm{O}} \ln p_i^{\mathrm{O}} \\
&= -\sum_{i=1}^{n} p_i^\beta \ln \frac{p_i^\beta}{p_i^\alpha} + \sum_{i=1}^{n} p_i^{\mathrm{O}} \ln \frac{p_i^{\mathrm{O}}}{p_i^\alpha} + \sum_{i=1}^{n} p_i^\beta \ln \frac{p_i^\beta}{p_i^{\mathrm{O}}} \\
&= -D(\beta\|\alpha) + D(\mathrm{O}\|\alpha) + D(\beta\|\mathrm{O}).
\end{aligned} \tag{B2}$$

Therefore, the Pythagorean theorem in Eq. (9) is obtained by substituting Eqs. (B2) and (8) into Eq. (B1).

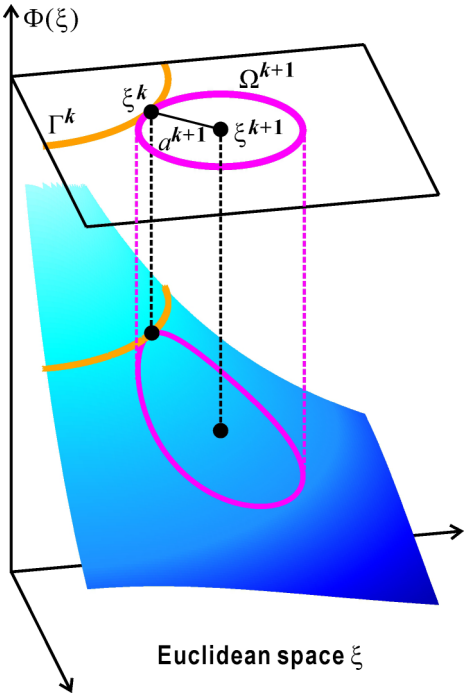

FIG. 11. Geometry of gradient descent path in general Euclidean space $\boldsymbol{\xi}$.

APPENDIX C: GEOMETRICS OF GRADIENT DESCENT PATH IN GENERAL EUCLIDEAN SPACE

This Appendix delves into the geometric properties of the gradient descent path in general Euclidean space $\boldsymbol{\xi}$. To accomplish this, we analyze the geometric relationship between two points $\boldsymbol{\xi}^k$ and $\boldsymbol{\xi}^{k+1}$ on the path, representing the positions after the kth and $(k+1)$th transformations through gradient descent, respectively (Fig. 11). The gradient descent formulation can be expressed using the Euclidean distance between these points, $d(k||k+1)$, as follows:

$$\nabla_{\boldsymbol{\xi}^k}(\Phi(\boldsymbol{\xi}^k) - d(k+1\|k)/\tau\ ) = 0. \tag{C1}$$

Mathematically, Eq. (B1) is equivalent to determining the extremum of a Lagrangian of the form

$$L = \Phi(\boldsymbol{\xi}^k) + (a^{k+1} - d(k+1\|k))/\tau, \tag{C2}$$

where a^{k+1} represents a parameter to be determined. In the general Lagrange multiplier method, $\boldsymbol{\xi}^{k+1}$ and a^{k+1} are known, whereas $\boldsymbol{\xi}^k$ and $1/\tau$ denote the position and multiplier to

be determined, respectively. In this scenario, by determining the extremum of the Lagrangian, Eq. (C2), is equivalent to determining the extremum of Φ ($\boldsymbol{\xi}$) on the hypersphere Ω^{k+1} defined by the constraint $a^{k+1} - d(k+1\|k) = 0$ (Fig. 11). However, in the gradient descent, the relationship between the variables is reversed: $\boldsymbol{\xi}^k$ and $1/\tau$ are known, whereas $\boldsymbol{\xi}^{k+1}$ and a^{k+1} represent the position and parameter to be determined, respectively. Therefore, when Γ^k is defined as the contour of $\Phi(\boldsymbol{\xi})$ that passes through $\boldsymbol{\xi}^k$, the task of determining the extremum of the Lagrangian, Eq. (C2), is equivalent to identifying the hypersphere Ω^{k+1} that is tangential to Γ^k at point $\boldsymbol{\xi}^k$ (Fig. 11). This represents the geometric relationship that must be maintained between the two points $\boldsymbol{\xi}^k$ and $\boldsymbol{\xi}^{k+1}$ along the gradient descent path. We assume that this geometric relationship holds true for the two points $\boldsymbol{\eta}^k$ and $\boldsymbol{\eta}^{k+1}$ along the gradient path in the $\mathbf{p}$ space, particularly when these points demonstrate an approximately Euclidean relationship. This assumption is expressed using Eq. (13).

[1] S. -I. Amari and H. Nagaoka, *Methods of Information Geometry* (Oxford University Press, New York, 2000), Vol. 191.

[2] N. N. Chentsov, Statistical decision rules and optimal inference, *Translations of Mathematical Monographs* (American Mathematical Society, Providence, USA, 1982), Vol. 53.

[3] R. A. Fisher, Theory of statistical estimation, Math. Proc. Camb. Philos. Soc. **22**, 700 (1925).

[4] C. R. Rao, Information and accuracy attainable in the estimation of statistical parameters, Bull. Calcutta. Math. Soc. **37**, 81 (1945).

[5] H. Nagaoka and S.-I. Amari, Differential geometry of smooth families of probability distribution *s*, Technical Report METR 82-7, Department of Mathematical Engineering and Instrumentation Physics, University of Tokyo, 1982.

[6] S.-I. Amari, *Differential-Geometric Methods in Statistics* (Springer, Berlin, 1985), Vol. 28.

[7] S. Kullback and R. A. Leibler, On information and sufficiency, Ann. Math. Stat. **22**, 79 (1951).

[8] S. Eguchi, Second order efficiency of minimum contrast estimators in a curved exponential family, Ann. Stat. **11**, 793 (1983).

[9] S. Eguchi, A characterization of second order efficiency in a curved exponential family, Ann. Inst. Stat. Math. **36**, 199 (1984).

[10] S. -I. Amari, *Information Geometry and Its Applications, Applied Mathematical Sciences* (Springer, Tokyo, 2016), Vol. 194.

[11] F. Nielsen, *Progress in Information Geometry: Theory and Applications* (Springer, Switzerland, 2021).

[12] D. Felice, C. Cafaro, and S. Mancini, Information geometric methods for complexity, Chaos **28**, 032101 (2018).

[13] H. Janyszek and R. Mrugala, Riemannian geometry and the thermodynamics of model magnetic systems, Phys. Rev. A **39**, 6515 (1989).

[14] G. Ruppeiner, Riemannian geometry in thermodynamic fluctuation, Rev. Mold. Phys. **67**, 605 (1995).

[15] W. Janke, D. A. Johnston, and R. P. K. C. Malmini, Information geometry of the Ising model on planar random graphs, Phys. Rev. E **66**, 056119 (2002).

[16] D. C. Brondy and A. Ritz, Information geometry of finite Ising models, J. Geometry Phys. **47**, 207 (2003).

[17] D. Felice, S. Mancini, and M. Pettini, Quantifying networks complexity from information geometry viewpoint, J. Math. Phys. **55**, 043505 (2014).

[18] R. Franzosi, D. Felice, S. Mancini, and P. Pettini, A geometric entropy detecting the Erdös-Renyi phase transition, Eur. Phys. Lett. **111**, 20001 (2015).

[19] N. Ay, Information geometry on complexity and stochastic interaction, Entropy **17**, 2432 (2015).

[20] U. Seifert, Stochastic thermodynamics, fluctuation theorems and molecular machines, Rep. Prog. Phys. **75**, 126001 (2012).

[21] S. Ito, Stochastic thermodynamic interpretation of information geometry, Phys. Rev. Lett. **121**, 030605 (2018).

[22] S. Ito, Stochastic time evolution, information geometry, and the Cramér-Rao bound, Phys. Rev. X **10**, 021056 (2020).

[23] A. Uhlmann, Density operators as an arena for differential geometry, Rep. Math. Phys. **33**, 253 (1993).

[24] D. Petz and C. Sudar, Geometries of quantum states, J. Math. Phys. **37**, 2662 (1996).

[25] P. Zanardi, P. Giorda, and M. Cozzini, Information-theoretic differential geometry of quantum phase transitions, Phys. Rev. Lett. **99**, 100603 (2007).

[26] P. Kumar, S. Mahapatra, P. Phukon, and T. Sarkar, Geodesics in information geometry: Classical and quantum phase transitions, Phys. Rev. E **86**, 051117 (2012).

[27] A. Fujiwara, Dually flat structures induced from monotone metrics on a two-level quantum state space, Eur. Phys. J. Plus **135**, 860 (2020).

[28] P. C. Hansen, *Discrete Inverse Problems: Insight and Algorithms* (SIAM, Philadelphia, 2010).

[29] M. Bertero, P. Boccacci, and C. De Mol, *Introduction to Inverse Problems in Imaging* (CRC Press, Boca Raton, FL, 2022).

[30] G. Tóth, Simultaneous Monte Carlo determination of particle size distribution and pair-correlation function of spherical colloids from a diffraction experiment, Langmuir **15**, 6718 (1999).

[31] S. Martelli and P. E. D. Nunzio, Particle size distribution of nanospheres by Monte Carlo fitting of small angle X-ray scattering curves, Part. Part. Syst. Charact. **19**, 247 (2002).

[32] J. Pikunic, C. Clinard, N. Cohaut, K. E. Gubbins, J. M. Guet, R. J. -M. Pellenq, I. Rannou, and J. N. Rouzaud, Structural modeling of porous carbons: Constrained reverse Monte Carlo method, Langmuir, **19**, 8565 (2003).

[33] T. Sato, M. Sugiyama, K. Itoh, K. Mori, T. Fukunaga, M. Misawa, T. Otomo, and S. Takata, Structural difference between liquidlike and gaslike phases in supercritical fluid, Phys. Rev. E **78**, 051503 (2008).

[34] H. Hagita, R. L. McGreevy, T. Arai, M. Inui, K. Matsuda, and K. Tamura, Frist example of multi-scale reverse Monte

Carlo modeling for small-angle scattering experimental data using reverse mapping from coarse-grained particles to atoms, J. Phys.: Condens. Matter **22**, 404215 (2010).
[35] J. Köfinger, L. S. Stelzl, K. Reuter, C. Allande, K. Reichel, and G. Hummer, Efficient ensemble refinement by reweighting, J. Chem. Theory Comput. **15**, 3390 (2019).
[36] S. Bottaro, T. Bengsten, and K. Lindorff-Larsen, Integrating molecular simulation and experimental data: A Bayesian/maximum entropy reweighting approach, Methods Mol. Biol. **2112**, 219 (2020).
[37] T. H. Edwards and S. Stoll, Optimal Tikhonov regularization for DEER spectroscopy, J. Mag. Res. **288**, 58 (2018).
[38] E. J. Hustedt, R. A. Stein, and H. S. Mchaourab, Protein functional dynamics from the rigorous global analysis of DEER data: Conditions, components, and conformations, J. Gen. Physiol. **153**, e201711954 (2021).
[39] M. Cordova, M. Balodis, B. S. de Almeida, M. Ceriotti, and L. Emsley, Bayesian probabilistic assignment of chemical shifts in organic solids, Sci. Adv. **7**, eabk2341 (2021).
[40] R. Grau-Crespo, S. Hamad, S. R. G. Balestra, R. Issa, T. D. Sparks, A. Femandes, B. L. Griffiths, R. F. Moran, D. Mackay, and S. E. Ashbrook, Capturing local compositional fluctuations in NMR modeling of solid solutions, Chem. Sci. **16**, 19357 (2025).
[41] C.-J. Feng, B. Dhayalan, and A. Tokmakoff, Refinement of peptide conformational ensembles by 2D IR spectroscopy: Application to Ala-Ala-Ala, Biophys. J. **114**, 2820 (2018).
[42] P. N. Akbar and R. Blümel, Inverse infrared spectral deconvolution for quantitative analysis of polymer mixtures in scattering media, Anal. Chem. **97**, 23080 (2025).
[43] A. L. Wilson, C. Outeiral, S. E. Dowd, A. J. Doig, P. L. A. Popelier, J. P. Waltho, and A. Almond, Deconvolution of conformational exchange from Raman spectra of aqueous RNA nucleosides, Comm. Chem. **3**, 56 (2020).
[44] A. S. Mishchenko, H. Matsui, and T. Hasegawa, Distribution of localized states from fine analysis of electron spin resonance spectra of organic semiconductors: Physical meaning and methodology, Phys. Rev. B **85**, 085211 (2012).
[45] S. Tokuda, S. Souma, K. Segawa, T. Takahashi, Y. Ando, T. Nakanishi, and T. Sato, Unveiling quasiparticle dynamics of topological insulators through Bayesian modelling, Comm. Phys. **4**, 170 (2021).
[46] Z. Hradil, Quantum-state estimation, Phys. Rev. A **55**, R1561 (1997).
[47] D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White, Phys. Rev. A **64**, 052312 (2001).
[48] R. Blume-Kohout, Optimal, reliable estimation of quantum states, New J. Phys. **12**, 043034 (2010).
[49] J. Suzuki, Y. Yang, and M. Hayashi, Quantum state estimation with nuisance parameters, J. Phys. A Math. Theor. **53**, 453001 (2020).
[50] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev, Phase retrieval with application to optical imaging, IEEE Signal Process. Mag. **32**, 87 (2015).
[51] K. Teilum, J. G. Olsen, and B. B. Kragelund, Functional aspects of protein flexibility, Cell. Mol. Life Sci. **66**, 2231 (2009).
[52] K. Teilum, J. G. Olsen, and B. B. Kragelund, Protein stability, flexibility and function, Biochim. Biophys. Acta **1814**, 969 (2011).
[53] M. D. Miller and G. N. Philips, Jr., Moving beyond static snapshots: Protein dynamics and the Protein Data Bank, J. Biol. Chem. **296**, 100749 (2021).
[54] R. Nussinov, Y. Liu, W. Zhang, and H. Jang, Protein conformational ensembles in function: Rules and mechanics, RSC Chem. Biol. **4**, 850 (2023).
[55] T. Oroguchi, H. Hashimoto, T. Shimizu, M. Sato, and M. Ikeguchi, Intrinsic dynamics of restriction endonuclease EcoO109I studied by molecular dynamics simulations and X-ray scattering data analysis, Biophys. J. **96**, 2808 (2009).
[56] R. Inoue, T. Oroguchi, T. Oda, B. Farago, A. Martel, L. Porcar, M. Sato, and M. Sugiyama, Internal dynamics of multi-domain protein as revealed by an optimized neutron spin echo measurement and all-atom molecular dynamics simulation, Phys. Rev. Res. **5**, 043154 (2023).
[57] T. Oroguchi and M. Nakasako, Changes in hydration structure are necessary for collective motions of a multi-domain protein, Sci. Rep. **6**, 26302 (2016).
[58] M. Karplus and J. A. McCammon, Molecular dynamics simulations of biomolecules, Nat. Struct. Biol. **9**, 646 (2002).
[59] S. Bottaro and K. Lindorff-Larsen, Biophysical experiments and biomolecular simulations: A perfect match? Science **361**, 355 (2018).
[60] M. Bonomi, G. T. Heller, C. Camilloni, and M. Vendruscolo, Principles of protein structural ensemble determination, Curr. Opin. Struct. Biol. **42**, 106 (2017).
[61] S. -I. Amari, Differential geometry of curved exponential families–Curvature and information loss, Ann. Stat. **10**, 357 (1982).
[62] S. L. Braunstein and C. M. Caves, Statistical distance and the geometry of quantum states, Phys. Rev. Lett. **72**, 3439 (1994).
[63] K. P. Murphy, *Probabilistic Machine Learning: An Introduction* (The MIT Press, Cambridge, MA, 2022).
[64] D. P. Bertsekas, *Convex Optimization Algorithms* (Athena Scientific, Belmont, MA, 2015).
[65] S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, Cambridge, 2004).
[66] C. Witkov and K. Zengel, *Chi-Squared Data Analysis and Model Testing for Beginners*, (Oxford University Press, Oxford, 2019).
[67] D. I. Svergun, C. Barberato, and M. H. J. Koch, CRYSOL–A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates, J. Appl. Cryst. **28**, 768 (1995).
[68] H. Kenzaki, N. Koga, N. Hori, Kanada R, W. Li, K. Okazaki, X. Yao, and S. Takada, CafeMol: A coarse-grained biomolecular simulator for simulating proteins at work, J. Chem. Theory Comput. **7**, 1979 (2011).
[69] J. A. Manso, D. Nunes-Costa, S. Macedo-Ribeiro, N. Empadinhas, and P. J. B. Pereira, Molecular fingerprints for a novel enzyme family in *Actinobacteria* with glucosamine kinase activity, mBio **10**, e00239-19 (2019).
[70] A. Guinier, G. Fourmet, C. B. Walker, and K. L. Yodowitch, *Small-angle Scattering of X-rays* (John Wiley and Sons, New York, 1955).
[71] See Supplemental Material at http://link.aps.org/supplemental/10.1103/vkck-1pcm for Method S1: All-atom MD simulations; Method S2: Coarse-grained MD simulations; Method S3: Calculation of SAXS data from CGMD simulations; Method S4: Q-dependent experimental errors of

SAXS data; Method S5: Preparation for initial distribution models for reconstruction simulations; Note S1: Dependence of MBGD results for the particle size distribution system on step size τ; Note S2: Analyses on conformational motions of transferrin; Note S3: Determination of projection coordinates and state resolution for representing probability distribution; Note S4: Dependencies of the calculated SAXS data of transferrin on conformational changes; Note S5: Reconstruction simulation of conformational ensembles of *Mt*EPSPS from SAXS data; Note S6: Reconstruction simulation of conformational ensembles of guanylate kinase from SAXS data; Note S7: Dependence of MBGD results for the protein conformational-ensemble systems on the step size τ; Note S8: Reconstruction simulation of conformational ensembles of guanylate kinase with increased amount of experimental data; Note S9: Reconstruction simulation of conformational ensembles of transferrin from pseudoexperimental data including noise; Note S10: Details in application of MBGD to experimental SAXS data of SjGlcNK; and Table S1, which also includes Refs. [113–118].

[72] H. B. Curry, The method of steepest descent for non-linear minimization problems, Quart. Appl. Math. **2**, 258 (1944).

[73] R. T. A. MacGillivray, S. A. Moore, J. Chen, B. F. Anderson, H. Baker, Y. Luo, M. Bewley, C. A. Smith, M. E. P. Murphy, Y. Wang, A. B. Mason, R. C. Woodworth, G. D. Brayer, and E. N. Baker, Two high-resolution crystal structures of the recombinant N-lobe of human transferrin reveal a structural change implicated in iron release, Biochemistry **37**, 7919 (1998).

[74] P. D. Jeffrey, M. C. Bewley, R. T. A. MacGillivray, A. B. Mason, R. C. Woodworth, and E. N. Baker, Ligand-induced conformational change in transferrins: Crystal structure of open form of the N-terminal half-molecule of human Transferrin, Biochemistry **37**, 13978 (1998).

[75] H. D. T. Mertens and D. I. Svergun, Structural characterization of proteins and complexes using small-angle X-ray solution scattering, J. Struct. Biol. **172**, 128 (2010).

[76] C. A. Brosey and J. A. Tainer, Evolving SAXS versatility: Solution X-ray scattering for macromolecular architecture, functional landscapes, and integrative structural biology, Curr. Opin. Cell Biol. **58**, 197 (2019).

[77] T. Ekimoto, Y. Kokabu, T. Oroguchi, and M. Ikeguchi, Combination of coarse-grained molecular dynamics simulations and small-angle X-ray scattering experiments, Biophys. Physicobiol. **16**, 377 (2019).

[78] E. Valentini, A. G. Kikhney, G. Previtali, C. M. Jeffries, and D. I. Svergun, SASBDB, a repository for biological small-angle scattering data, Nucleic Acids Res. **43**, D357 (2015).

[79] M. R. Marques, A. Vaso, J. R. Neto, M. A. Fossey, J. S. Oliveira, L. A. Basso, D. S. dos Santos, W. F. de Azevedo, Jr., and M. S. Palma, Dynamics of glyphosate-induced conformational changes of *Mycobacterium Tuberculosis* 5-enolpyruvylshikimate-3-phosphate Synthase (EC 2.5.1.19) determined by hydrogen-deuterium exchange and electrospray mass spectrometry, Biochemistry **47**, 7509 (2008).

[80] J. Blaszczyk, Y. Li, Y. Honggao, and X. Ji, Crystal structure of unligated guanylate kinase from yeast reveals GMP-induced conformational changes, J. Mol. Biol. **307**, 247 (2001).

[81] T. Stehele and G. E. Schulz, Refined structure of the complex between guanylate kinase and its substrate GMP at 2.0 Å resolution, J. Mol. Biol. **224**, 1127 (1992).

[82] A. Kitao, F. Hirata, and N. Gö, The effects of solvent on the conformation and the collective motions of protein: Normal mode analysis and molecular dynamics simulations of melittin in water and in vacuum, Chem. Phys. **158**, 447 (1991).

[83] D. D. Boehr, R. Nussinov, and P. E. Wright, The role of dynamic conformational ensembles in biomolecular recognition, Nat. Chem. Biol. **5**, 789 (2009).

[84] W. Rudin, *Principles of Mathematical Analysis* (McGraw-Hill, New York, 1976).

[85] S.-I. Amari, Natural gradient works efficiently in learning, Neural Comput. **10**, 251 (1998).

[86] Y. Chen, D. Z. Huang, J. Huang, S. Reich, and A. M. Stuart, Sampling via gradient flows in the space of probability measures, Math. Comp. (2026).

[87] P. Izmailov, P. Kirichenko, M. Finzi, and A. G. Wilson, Semi-supervised learning with normalizing flows, Proc. Mach. Learn. Res. **119**, 4615 (2020).

[88] G. Papamakarios, E. Nalisnick, D. J. Rezende, S. Mohamed, and B. Lakshminarayanan, Normalizing flows for probabilistic modeling and inference, J. Mach. Learn Res. **22**, 1 (2021).

[89] J. Miao, T. Ishikawa, I. K. Robinson, and M. M. Murnane, Beyond crystallography: Diffractive imaging using coherent x-ray light sources, Science **348**, 530 (2015).

[90] L. Borcea, Electrical impedance tomography, Inverse Prob. **18**, R99 (2002).

[91] Y. Hoshi and Y. Yamada, Overview of diffuse optical tomography and its clinical applications, J. Biomed. Opt. **21**, 091312 (2016).

[92] L. Wang, T. J. Martinez, and V. S. Pande, Building force fields: An automatic, systematic and reproducible approach, J. Phys. Chem. Lett. **5**, 1885 (2014).

[93] L. Wang, K. A. McKierman, J. Gomes, K. A. Beauchamp, T. Head-Gordon, J. E. Rice, W. C. Swope, T. J. Martínez, and V. S. Pande, Building a more predictive protein force field: A systematic and reproducible route to AMBER-FB15, J. Phys. Chem. B **121**, 4023 (2017).

[94] C. Vogel, M. Bashton, N. D. Kerrison, C. Chothia, and A. Teichmann, Structure, function and evolution of multidomain proteins, Curr. Opin. Cell Biol. **14**, 208 (2004).

[95] D. Ekman, A. K. Björklund, J. Frey-Skött, and A. Elofsson, Multi-domain proteins in the three kingdoms of life: Orphan domains and other unassigned regions, J. Mol. Biol. **348**, 231 (2005).

[96] V. N. Uversky, Unusual biophysics of intrinsically disordered proteins, Biochim. Biophys. Acta **1834**, 932 (2013).

[97] C. J. Oldfield and K. Dunker, Intrinsically disordered proteins and intrinsically disordered protein regions, Annu. Rev. Biochem. **83**, 553 (2014).

[98] Y. Qi and B. Zhang, Predicting three-dimensional genome organization with chromatin states, PLoS Comput. Biol. **15**, e1007024 (2019).

[99] H. Tjong, W. Li, R. Kalhor, C. Dai, S. Hao, K. Gong, Y. Zhou, H. Li, X. J. Zhou, M. A. Le Gros, C. A. Larabell, L. Chen, and F. Alber, Population-based 3D genome structure analysis reveals driving forces in spatial genome organization, Proc. Natl. Acad. Sci. USA **113**, E1663 (2016).

[100] M. Sonntag, P. K. A. Jagtap, B. Simon, M.–S. Appavou, A. Geerlof, R. Stehle, F. Gable, J. Hennig, and M. Sattler, Segmental, domain-selective perdeuteration and small-angle

neutron scattering for structural analysis of multi-domain proteins, Angew. Chem. Int. Ed. **56**, 9322 (2017).

[101] H. Nakagawa, T. Saio, M. Nagao, R. Inoue, M. Sugiyama, A. Ajito, T. Tominaga, and Y. Kawakita, Conformational dynamics of a multidomain protein by neutron scattering and computational analysis, Biophys. J. **120**, 3341 (2021).

[102] E. Z. Eisenmesser, O. Millet, W. Labeikovsky, D. M. Korzhnev, M. Wolf-Watz, D. A. Bosco, J. J. Skalicky, L. E. Kay, and D. Kern, Intrinsic dynamics of an enzyme underlies catalysis, Nature (London) **438**, 117 (2005).

[103] T. R. Alderson and L. E. Kay, NMR spectroscopy captures the essential role of dynamics in regulating biomolecular function, Cell **184**, 577 (2021).

[104] A. G. Palmer, III, Enzyme dynamics from NMR spectroscopy, Acc. Chem. Res. **48**, 457 (2015).

[105] L. Yu and R. Brüschweiler, Quantitative prediction of ensemble dynamics, shapes and contact propensities of intrinsically disordered proteins, PLoS Comput. Biol. **18**, e1010036 (2022).

[106] A. Dashti, G. Mashayekhi, M. Shekhar, D. B. Hail, S. Salah, P. Schwander, A. des Georges, A. Singharoy, J. Frank, and A. Ourmazd, Retrieving functional pathways of biomolecules from single-particle snapshots, Nat. Comm. **11**, 4734 (2020).

[107] M. Oide, T. Kato, T. Oroguchi, and M. Nakasako, Energy landscape of domain motion in glutamate dehydrogenase deduced from cryo-electron microscopy, FEBS J. **287**, 3472 (2020).

[108] N. Sibille and P. Bernadó, Structural characterization of intrinsically disordered proteins by the combined use of NMR and SAXS, Biochem. Soc. Trans. **40**, 955 (2012).

[109] A. G. Kikhney and D. I. Svergun, A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins, FEBS Lett. **589**, 2570 (2015).

[110] M. Chan-Yao-Chong, D. Durand, and T. Ha-Duong, Molecular dynamics simulations combined with nuclear magnetic resonance and/or small-angle X-ray scattering data for characterizing intrinsically disordered protein conformational ensembles, J. Chem. Inf. Model. **59**, 1743 (2019).

[111] B. Vallat, *et al.*, IHMCIF: An extension of the PDBx/mmCIF data standard for integrative structure determination methods, J. Mol. Biol. **436**, 168546 (2024).

[112] T. Oroguchi, R. Inoue, and M. Sugiyama, GitHub (2025), https://github.com/TOroguchi/MBGD.

[113] R. Salomon-Ferrer, D. A. Case, and R. C. Walker, An overview of the Amber biomolecular simulation package, WIREs Comput. Mol. Sci. **3**, 198 (2013).

[114] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, Smooth particle mesh Ewald method, J. Chem. Phys. **103**, 8577 (1995).

[115] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of *n*-alkanes, J. Comput. Phys. **23**, 327 (1977).

[116] G. P. Bourenkov, G. S. Kachalova, N. Strizhov, M. Bruning, A. Vagin, and H. D. Bartunik, *Mycobacterium Tuberculosis* epsp synthase in unliganded state, 2006, https://doi.org/10.2210/pdb2BJB/pdb.

[117] G. S. Kachalova, G. P. Burenkow, N. I. Strizhov, M. G. Brunning, and H. D. Bartunik, *Mycobacterium Tuberculosis* epsp synthase in complex with s3p, 2008, https://doi.org/10.2210/pdb2O0D/pdb.

[118] P. Chen and J. S. Hub, Interpretation of solution X-ray scattering by explicit-solvent molecular dynamics, Biophys. J. **108**, 2573 (2015).