

RESEARCH ARTICLE

Moment-based metrics for molecules computable from cryogenic electron microscopy images

Andy Zhang¹ , Oscar Mickelin¹ , Joe Kileel² , Eric J. Verbeke¹ , Nicholas F. Marshall³ ,
Marc Aurèle Gilles¹  and Amit Singer^{1,4} 

¹Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ, USA

²Department of Mathematics and Oden Institute, University of Texas at Austin, Austin, TX, USA

³Department of Mathematics, Oregon State University, Corvallis, OR, USA

⁴Department of Mathematics, Princeton University, Princeton, NJ, USA

Corresponding author: Andy Zhang; Email: az8940@princeton.edu

Received: 15 November 2023; **Revised:** 27 January 2024; **Accepted:** 30 January 2024

Keywords: alignment-free metric; Kam's method; method of moments; protein structural similarity; rotationally invariant distance; structural search

Abstract

Single-particle cryogenic electron microscopy (cryo-EM) is an imaging technique capable of recovering the high-resolution three-dimensional (3D) structure of biological macromolecules from many noisy and randomly oriented projection images. One notable approach to 3D reconstruction, known as Kam's method, relies on the moments of the two-dimensional (2D) images. Inspired by Kam's method, we introduce a rotationally invariant metric between two molecular structures, which does not require 3D alignment. Further, we introduce a metric between a stack of projection images and a molecular structure, which is invariant to rotations and reflections and does not require performing 3D reconstruction. Additionally, the latter metric does not assume a uniform distribution of viewing angles. We demonstrate the uses of the new metrics on synthetic and experimental datasets, highlighting their ability to measure structural similarity.

Impact Statement

Single-particle cryogenic electron microscopy (cryo-EM) is a popular method to obtain three-dimensional (3D) reconstructions of biological molecules from noisy two-dimensional (2D) tomographic projection images. Many iterative techniques for this reconstruction require initializations sufficiently close to the unknown structure to obtain high-quality reconstructions. To help select an initialization from a database of known structures, this paper introduces a metric to compare the similarity of known 3D structures with a stack of noisy 2D tomographic projection images of an unknown structure. We show that this metric distinguishes differing structures and present an efficient method to compute it, notably without performing 3D reconstruction.

1. Introduction

Single-particle cryogenic electron microscopy (cryo-EM) enables high-resolution reconstruction of three-dimensional (3D) biological macromolecules from a large collection of noisy and randomly oriented projection images. Kam's method⁽¹⁾ is one of the earliest methods proposed for homogeneous reconstruction in cryo-EM. It is a statistical method-of-moments approach applied to the cryo-EM reconstruction problem, where the computation of low-order statistics of two-dimensional (2D) images allows for the estimation of 3D structure through solving a polynomial system. Kam's method has helped push the

theoretical understanding of the reconstruction process – under certain conditions, it is a provable algorithm and provides bounds for the estimated structure’s quality in terms of the noise level and the number of images.^(2–8) Kam’s method also enjoys other remarkable properties:

1. It bypasses the need for angular assignment, typically a large computational burden in competing methods.
2. It is a streaming algorithm and is thus theoretically much faster than iterative methods.
3. It can – in theory – break the detection limit of the minimal size of proteins that can be observed in cryo-EM.⁽⁹⁾

While theoretically attractive, Kam’s method has not been able to yield high-resolution reconstructions as yet. One direction that is currently being explored to resolve this issue is the development of better priors, for instance, based on the sparsity of the solution as in the study by Bendory et al.⁽⁷⁾ Separately, there has been considerable, continued interest in utilizing the vast amount of solved structures stored in the Protein Data Bank (PDB)⁽¹⁰⁾ to improve cryo-EM reconstructions.

Leveraging the PDB as a prior, we propose a method to match either projection images or molecular volumes to a database of previously solved structures (Section 3). We use this procedure as a rotationally and reflectionally invariant metric that can be directly computed from raw image datasets without needing a 3D reconstruction process. Importantly, our metric neither relies on prior knowledge of rotations nor assumes a uniform rotational distribution, making it applicable to typical datasets.

To demonstrate the efficacy of our metric, we compare it to existing methods and show empirically that it achieves similar performance to alignment-based metrics. As an application, we use our metric to compute a low-dimensional embedding of a subset of the PDB into the Euclidean plane, visually showcasing how structurally similar proteins are embedded near each other (Section 4.2). Further, we apply the version of the metric that can be directly computed from stacks of 2D images and show that it gives an efficient methodology to identify the nearest neighbors in a database to a given set of experimental moments on synthetic and real datasets (Sections 4.3 and 4.4).

2. Background

This section presents the mathematical preliminaries needed to define our metric. Let $\Phi: \mathbb{R}^3 \rightarrow \mathbb{R}$ be the electrostatic potential of a molecule and $\hat{\Phi}: \mathbb{R}^3 \rightarrow \mathbb{C}$ be its Fourier transform, which we define by

$$\hat{\Phi}(\zeta) = \int_{\mathbb{R}^3} \Phi(x) e^{-i\zeta \cdot x} dx.$$

A single projection image is given by

$$I = h * P R \Phi + \varepsilon,$$

and its Fourier transform is

$$\hat{I} = H \cdot S R \hat{\Phi} + \hat{\varepsilon},$$

where P is the projection operator, S is the slicing operator, h is a point spread function, H is the contrast transfer function (CTF), ε is noise, and $R \in \text{SO}(3)$ is a rotation operator. We assume that the Fourier transform $\hat{\Phi}$ can be expanded in a spherical harmonic expansion:

$$\hat{\Phi}(r, \theta, \phi) = \sum_{\ell=0}^L \sum_{m=-\ell}^{\ell} A_{\ell,m}(r) Y_{\ell}^m(\theta, \phi), \quad (1)$$

where (r, θ, ϕ) are spherical coordinates, and Y_{ℓ}^m denotes the complex spherical harmonic:

$$Y_{\ell}^m(\theta, \phi) := \left(\frac{(\ell-m)!(2\ell+1)}{4\pi(\ell+m)!} \right)^{1/2} e^{im\phi} P_{\ell}^m(\cos \theta),$$

where P_ℓ^m are the associated Legendre polynomials, $A_{\ell,m}(r)$ are r -dependent coefficients, and L is a bandlimit parameter. See Eq. 14.30.1 in⁽¹¹⁾ for the definitions of Y_ℓ^m and P_ℓ^m .

Let $\rho : SO(3) \rightarrow \mathbb{R}$ be the probability density function of the rotational distribution, which without loss of generality is invariant to in-plane rotations and reflections. (Note that by augmenting the dataset with in-plane rotations and reflections of all 2D images, one can always reduce to the case of such an invariant distribution ρ ; for example, see the study by Ponce and Singer⁽¹²⁾). More formally, ρ is a function on $SO(3)/O(2) \simeq \mathbb{S}^2/\{\pm 1\}$, which is formed by identifying antipodal points on the sphere $\mathbb{S}^2$.⁽¹³⁾ Thus, we model the density as a function $\rho : SO(3) \rightarrow \mathbb{R}$ with an even-degree spherical harmonic expansion:

$$\rho(R) = \sum_{\ell=0}^P \sum_{m=-2\ell}^{2\ell} B_{2\ell,m} \overline{Y_{2\ell}^m}(\theta(R), \varphi(R)), \quad (2)$$

where $(\theta(R), \varphi(R))$ represents the third column of the rotation matrix given by R in spherical coordinates, and $P \in \mathbb{Z}_{\geq 0}$ is a bandlimit parameter (see Section 4.2 in the study by Sharon et al.⁽⁸⁾). The analytical first and second moments $\mathbf{m}_1$ and $\mathbf{m}_2$ of the Fourier-transformed projection images with respect to ρ are

$$\begin{aligned} \mathbf{m}_1(x, y) &= \int (R \cdot \hat{\Phi})(x, y, 0) \rho(R) dR, \quad \text{and} \\ \mathbf{m}_2((x, y), (x', y')) &= \int (R \cdot \hat{\Phi})(x, y, 0) (R \cdot \hat{\Phi})(x', y', 0) \rho(R) dR, \end{aligned} \quad (3)$$

where dR denotes integration with respect to the Haar measure on $SO(3)$. It will be convenient to write (x, y) and (x', y') in terms of polar coordinates (r, ϕ) and (r', ϕ') , respectively. In Appendix A.1, we show in (12) and (13) that the first moment only depends on r , that is, $\mathbf{m}_1 = \mathbf{m}_1(r) : \mathbb{R}_{\geq 0} \rightarrow \mathbb{C}$, and that the second moment only depends on r, r' and $\Delta\phi = \phi - \phi'$, that is, $\mathbf{m}_2 = \mathbf{m}_2(r, r', \Delta\phi) : \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \times [-2\pi, 2\pi] \rightarrow \mathbb{C}$. We write $\mathbf{m}_1 = \mathbf{m}_1(\hat{\Phi}, \rho)$ and $\mathbf{m}_2 = \mathbf{m}_2(\hat{\Phi}, \rho)$ to denote the first and second moments defined by $\hat{\Phi}$ and ρ when discussing multiple structures. The basis of Kam's method is that the moments in (3) can be estimated from images and related to expansion coefficients for the potential $\hat{\Phi}$; see Appendix A for explicit formulas.

3. Definition of Kam's metrics

We now use metrics between the moments in (3) to define similarity between proteins and stacks of images. A first function is used to measure the similarity of two known structures by the moments of their potential as defined in (3). The second is used to measure the similarity between a known structure and the unknown structure observed in a dataset of images.

Crucially, the metrics can be computed without performing 3D alignment of the structures, reducing their computational costs compared to other approaches. Moreover, one of the metrics can be directly computed from noisy and CTF-affected projection images. This enables a nearest neighbor search among known structures to determine an initialization for the 3D reconstruction pipeline, especially in the expectation–maximization procedure.^(14,15)

3.1. Kam's volume metric d_{vKam}

Here, we introduce the first of Kam's metrics, which measures the similarity of two 3D structures. We use this to perform dimensionality reduction to visualize the relationship between structures from a subset of the PDB.

In detail, given two 3D structures Φ_1 and Φ_2 , we define the distance between them through their first and second moments $\mathbf{m}_1$ and $\mathbf{m}_2$ under a uniform distribution of viewing directions, which we denote by $\rho = \rho_u$. We will derive the explicit equations for the uniform case in (17) and (18). We then measure the resulting weighted deviation of the first and second moments by

$$\|\mathbf{m}_2(\hat{\Phi}_1, \rho_u) - \mathbf{m}_2(\hat{\Phi}_2, \rho_u)\|_{L^2(\mathbb{R}^2 \times \mathbb{R}^2)}^2 + \lambda \|\mathbf{m}_1(\hat{\Phi}_1, \rho_u) - \mathbf{m}_1(\hat{\Phi}_2, \rho_u)\|_{L^2(\mathbb{R}^2)}^2, \quad (4)$$

where $\lambda \geq 0$ is a hyperparameter that we set to 1 for all experiments. The moments will be represented on a discretized voxel grid, and we therefore replace the continuous norms with discrete norms. More specifically, we will represent the second moment using a grid $r, r' \in \{r_1, \dots, r_{\lfloor N/2 \rfloor}\}$ and $\Delta\phi : = \phi - \phi' \in \{\Delta\phi_1, \dots, \Delta\phi_{2N}\}$, where N is the number of pixels of one side of the discretized volume. We define the grid points $r_k = \delta k/N$, $\Delta\phi_j = 2\pi j/N - 2\pi$, for $k = 1, \dots, \lfloor N/2 \rfloor$, and $j = 1, \dots, 2N$, where δ is the side length of the volume grid in angstroms. We then use the following two approximations to the continuous norms above

$$\|\mathbf{M}\|_{w_2}^2 := \sum_{\substack{j=1, \dots, 2N \\ k_1=1, \dots, \lfloor N/2 \rfloor \\ k_2=1, \dots, \lfloor N/2 \rfloor}} |\mathbf{M}(\Delta\phi_j, r_{k_1}, r_{k_2})|^2 r_{k_1} r_{k_2}, \quad \|\mathbf{N}\|_{w_1}^2 := \sum_{k=1, \dots, \lfloor N/2 \rfloor} |\mathbf{N}(r_k)|^2 r_k. \quad (5)$$

With these norms, we define the metric comparing two sets of moments of two 3D structures by

$$d_{\text{vKam}}(\Phi_1, \Phi_2) := \left(\|\mathbf{m}_2(\widehat{\Phi}_1, \rho_u) - \mathbf{m}_2(\widehat{\Phi}_2, \rho_u)\|_{w_2}^2 + \lambda \|\mathbf{m}_1(\widehat{\Phi}_1, \rho_u) - \mathbf{m}_1(\widehat{\Phi}_2, \rho_u)\|_{w_1}^2 \right)^{1/2}, \quad (6)$$

This distance is rotationally invariant since for any rotation R , we have $R \cdot \widehat{\Phi} = R \cdot \widehat{\Phi}$ and the moments $\mathbf{m}_1$ and $\mathbf{m}_2$ in (2) satisfy

$$\mathbf{m}_i(R \cdot \widehat{\Phi}, \rho) = \mathbf{m}_i(\widehat{\Phi}, R^T \cdot \rho), \quad (7)$$

as can be seen through a change of variables in (3). When $\rho = \rho_u$ is uniform, clearly $R^T \cdot \rho = \rho$, which therefore shows rotational invariance of the cost function in (4), up to the discretization of the volume grid. Note that this bypasses the need for an alignment step. We detail the procedure for computing $\mathbf{m}_1$, $\mathbf{m}_2$ and therefore d_{vKam} in Appendix A.1. Under certain conditions, it has been demonstrated that the second moment of the image collection identifies the 3D structure uniquely^(2-4,6,7) or up to a finite list of candidate structures.⁽⁸⁾ In Section 4.2, we show that our metric is alike other similarity scores but remarkably does not rely on alignment.

3.2. Kam's image metric d_{iKam}

We now introduce a metric between the empirical moments computed from a set of experimental projection images and the moments computed from the atomic coordinates of a known structure that compares images to the known structure. We detail the procedure for computing these moments in Appendix A.1.

If the distribution of poses in the experimental dataset would be known to be uniform, the empirical moments could directly be substituted for $\mathbf{m}_1$ and $\mathbf{m}_2$ in (6) and the metric could be defined as the deviation between the moments of the two structures. In practice, however, the distribution of angles is not uniform and is unknown. Since the moments are functions of this distribution, it must therefore be inferred.

We will show in (12) and (13) that $\mathbf{m}_1$ and $\mathbf{m}_2$ depend linearly on the expansion coefficients $B_{p,u}$ of the distribution of viewing directions. The optimization problem minimizing the discrepancy between the moments of the two structures is, therefore, a linear least-squares problem in $B_{p,u}$. It follows from Table 3 of⁽⁸⁾ that this linear least-squares is Zariski-generically full-rank (although not necessarily well-conditioned) for various small bandlimits L and P . Solving this optimization problem efficiently eliminates the unknown rotational distribution. We then define the metric between the moments of the structure Φ and the experimental moments $\tilde{\mathbf{m}}_1, \tilde{\mathbf{m}}_2$ by

$$\min_{\rho \in \mathcal{P}} \|\tilde{\mathbf{m}}_2 - \mathbf{m}_2(\widehat{\Phi}, \rho)\|_{L^2(\mathbb{R}^2 \times \mathbb{R}^2)}^2 + \lambda \|\tilde{\mathbf{m}}_1 - \mathbf{m}_1(\widehat{\Phi}, \rho)\|_{L^2(\mathbb{R}^2)}^2, \quad (8)$$

where $\lambda \geq 0$ is a hyperparameter which we set to 1 for all experiments and

$$\mathcal{P} = \left\{ \rho(R) : \text{SO}(3) \rightarrow \mathbb{R} \text{ s.t. } \rho(R) = \sum_{\ell=0}^P \sum_{m=-2\ell}^{2\ell} B_{2\ell,m} \overline{Y_{2\ell}^m}(\theta(R), \varphi(R)), \text{ and } \int_{\text{SO}(3)} \rho(R) dR = 1 \right\}, \quad (9)$$

is the set of admissible distributions of viewing directions that are invariant to global reflections and in-plane rotations, where $(\theta(R), \varphi(R))$ is as in (2). To simplify the optimization problem and lead to faster algorithms, note that we do not impose positivity of the distributions $\rho \in \mathcal{P}$, though this could be enforced, for instance, by imposing linear constraints $\rho(R_i) \geq 0$ for a suitable choice of $R_i \in \text{SO}(3)$. Moreover, the constraint $\int_{\text{SO}(3)} \rho(R) dR = 1$ is equivalent to imposing $B_{0,0} = 1$,⁽⁸⁾ which can be achieved by removing $B_{0,0}$ from the set of optimization variables and fixing its value to 1. The values of the bandlimit parameters L, P and the hyperparameter λ used in our numerical experiments are given in [Appendix B.1](#).

Just as in the previous section, we replace the continuous norms in (8) by discrete norms to define the metric between empirical moments and the moments from a 3D structure as

$$d_{\text{iKam}}((\tilde{\mathbf{m}}_1, \tilde{\mathbf{m}}_2), \Phi) := \left(\min_{\rho \in \mathcal{P}} \|\tilde{\mathbf{m}}_2 - \mathbf{m}_2(\widehat{\Phi}, \rho)\|_{w_2}^2 + \lambda \|\tilde{\mathbf{m}}_1 - \mathbf{m}_1(\widehat{\Phi}, \rho)\|_{w_1}^2 \right)^{1/2}. \quad (10)$$

The cost function in (8) is rotationally invariant, in that it does not depend on the orientation of Φ , since (7) implies that

$$\begin{aligned} & \min_{\rho \in \mathcal{P}} \|\tilde{\mathbf{m}}_2 - \mathbf{m}_2(R \cdot \widehat{\Phi}, \rho)\|_{L^2(\mathbb{R}^2 \times \mathbb{R}^2)}^2 + \lambda \|\tilde{\mathbf{m}}_1 - \mathbf{m}_1(R \cdot \widehat{\Phi}, \rho)\|_{L^2(\mathbb{R}^2)}^2 \\ &= \min_{\rho \in \mathcal{P}} \|\tilde{\mathbf{m}}_2 - \mathbf{m}_2(\widehat{\Phi}, R^T \cdot \rho)\|_{L^2(\mathbb{R}^2 \times \mathbb{R}^2)}^2 + \lambda \|\tilde{\mathbf{m}}_1 - \mathbf{m}_1(\widehat{\Phi}, R^T \cdot \rho)\|_{L^2(\mathbb{R}^2)}^2 \\ &= \min_{\rho \in \mathcal{P}} \|\tilde{\mathbf{m}}_2 - \mathbf{m}_2(\widehat{\Phi}, \rho)\|_{L^2(\mathbb{R}^2 \times \mathbb{R}^2)}^2 + \lambda \|\tilde{\mathbf{m}}_1 - \mathbf{m}_1(\widehat{\Phi}, \rho)\|_{L^2(\mathbb{R}^2)}^2. \end{aligned} \quad (11)$$

where the last equality follows because $R^T \cdot \rho$ lies in $\mathcal{P}$, since rotating a viewing angle distribution over $\text{SO}(3)$ results in another viewing angle distribution over $\text{SO}(3)$.

At the cost of solving the small linear system detailed in [Appendix A.3](#), our method allows for the comparison between a stack of images and a resolved structure, without performing a 3D reconstruction. Furthermore, we precompute the least-squares matrices necessary for optimization, after which the distance function can be calculated in real time. With sufficient storage and precomputation, the procedure is scalable to the entirety of the PDB.

In particular, d_{iKam} can be used in an efficient scheme to match a stack of synthetic images to the potentials of nearby PDB structures. By selecting a subset of the PDB database, one can efficiently compute $d_{\text{iKam}}((\tilde{\mathbf{m}}_1, \tilde{\mathbf{m}}_2), \Phi)$ for each Φ in the subset and find the nearest neighbors. The method for processing image moments in practice is detailed in [Appendix A.4](#), and the computational complexity of the metric is derived in [Appendix A.5](#).

4. Results

4.1. Existing measures of structural similarity

There are several existing methods for reporting structural similarity between two known volumes. We list two approaches based on computing alignment and Zernike moments. We compare both d_{iKam} and d_{vKam} to these approaches in the experiments in the following subsections. Note that the following existing metrics are limited to measuring similarity between two structures and cannot compare images to structures, whereas d_{iKam} can.

1. *Euclidean alignment*: A classical approach for comparing the similarity of two structures is to sample the volumes on a 3D grid and calculate the Euclidean distance between pairs over rotations and translations. However, this method is expensive to compute since optimization over $\text{SO}(3)$ is required to align the structures. Accelerated methods for computing these alignments by maximizing the correlation between two volume maps over rotations and translations have been

implemented in various programs, for example, via gradient ascent in Chimera.⁽¹⁶⁾ Further acceleration can be achieved by calculating volumetric correlations by expanding the volumes in a well-chosen basis and applying dimensionality reduction⁽¹⁷⁾ or by maximizing the correlation between common lines in projection images generated from the volumes.⁽¹⁸⁾ Similar alignment methods, such as those described in the study by Bartesaghi et al. and Xu et al.^(19,20), are also used in electron tomography for sub-volume similarity. In this paper, we use a Bayesian optimization algorithm to minimize an Euclidean loss function, as described in the study by Singer and Yang,⁽²¹⁾ to compute the alignment and minimum distance between two volumes.

2. *Zernike moments*: Another metric for structural similarity is to expand the molecule's structure in Zernike polynomials and compute a metric from the Zernike expansion coefficients, as described in the study by Guzenko et al.⁽²²⁾, which is used by the PDB for structural similarity search.

4.2. Applying d_{vKam} to a PDB subset

To test the ability of d_{vKam} to discern the similarity between 3D structures, we first generate a database using 1420 structures downloaded from the PDB.⁽¹⁰⁾ The subset chosen here was selected by filtering for human proteins with an experimental structure at resolution between 1 and 3 Å and a molecular weight between 150 and 250 kDa. We use this subset because it encompasses a diverse range of shapes and symmetries as well as many homologous structures. Additionally, the weight range reflects a smaller and more challenging protein size for a typical cryo-EM experiment.⁽²³⁾ In the future, a larger database containing the entire PDB can also be generated.

Using our database, we first generate a discretized potential for each structure as described in [Appendix A.2](#). The first and second moments of each structure can then be computed using (3). We then compute d_{vKam} in (6) pairwise for all structures in the database.

To compare the performance of d_{vKam} against existing metrics, we calculate pairwise scores using d_{vKam} , Euclidean alignment, and the Zernike metric. We then plot the returned scores against each other and calculate a ranking similarity using normalized discounted cumulative gain⁽²⁴⁾ (NDCG). We use this metric since it is a popular method to quantify the similarity between sets of rankings; its calculation is given in [Appendix A.6](#).

In [Figure 1](#), we report the NDCG scores between pairs of metrics. All NDCG scores are close to 1, indicating strong agreement among the three different metrics on which structures are most similar. However, the alignment metric and $\log(d_{\text{vKam}})$ share the highest average NDCG score. To verify the statistical significance of this agreement, we report a t-test by selecting 10 different subsets, showing that the NDCG score between d_{vKam} and the alignment metric is statistically significantly higher (with a p -value $p \approx 8 \times 10^{-9}$) than the NDCG score between the alignment metric and the Zernike metric. We thus conclude that d_{vKam} provides a fast and accurate alternative for the alignment metric.

Although it is the most interpretable metric, Euclidean alignment is computationally expensive to execute for all pairs of structures in a database. To achieve a manageable runtime for alignment, we calculate pairwise Euclidean alignment distances for a subset of the database of size 100. Pairwise alignment on this subset took 8 hours on a 2.6 GHz Intel Skylake Central Processing Unit (CPU) running AVX-512 using 16 physical cores and 80 GB random-access memory (RAM). To do pairwise alignment via Bayesian optimization for the entire database of 1420 structures would require 46 days of computation, whereas using d_{vKam} (including precomputation) to calculate pairwise distances between all 1420 structures in the database requires 3 minutes on the same hardware. Despite containing an alignment component, the Zernike metric is also fast, taking 3 minutes to compute pairwise distances for the entire database.

After observing high agreement between d_{vKam} and the other metrics, we compute a 2D embedding of the similarity between structures in our database using t-distributed stochastic neighbor embedding (t-SNE)⁽²⁵⁾ (see [Figure 2](#)). Analogous t-SNE plots for the alignment metric and Zernike metric are reported in [Appendix B.2](#). We find that d_{vKam} provides interpretable results in identifying similar molecules from their moments without the need for alignment. In particular, we observe that both

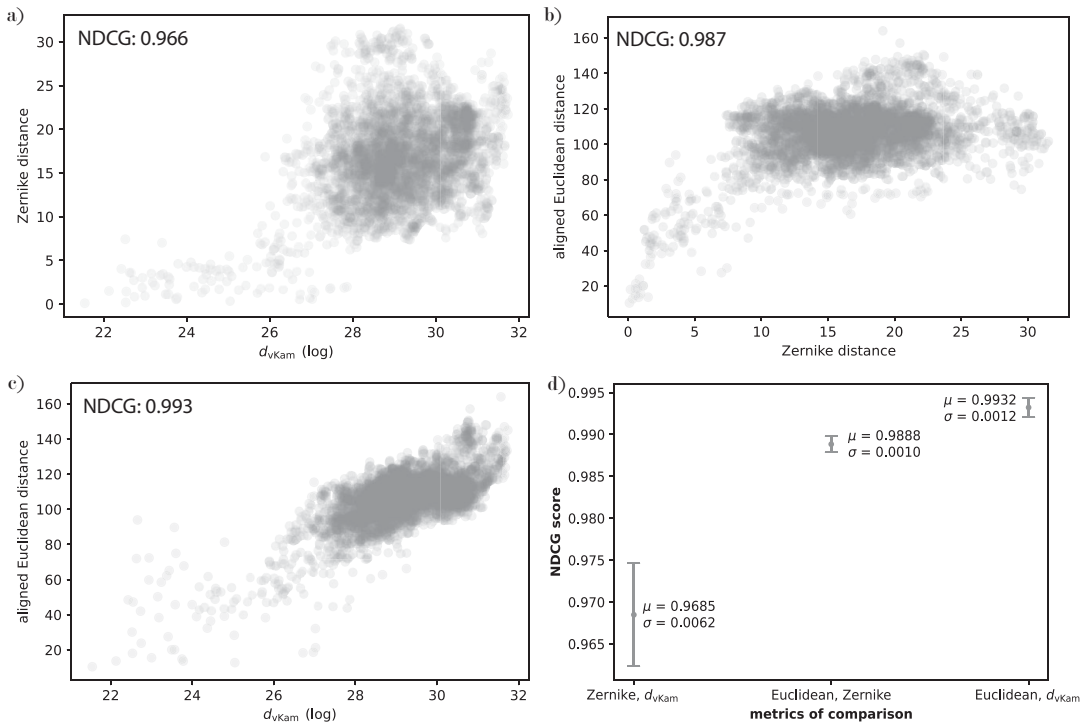


Figure 1. Comparison between d_{vKam} , the Zernike metric, and Euclidean alignment. a-c) A random size 100 subset of the database is selected. Then, pairwise similarity metrics are calculated and plotted, where each point represents a pair of structures. The NDCG score is calculated using the metric on the y-axis as the predicted metric, and the metric on the x-axis as the true metric. d) The procedure is repeated with 10 randomly selected size 100 subsets, and the mean (μ) and standard deviation (σ) of the NDCG scores are calculated. The error bars and points visualize $\mu \pm \sigma$.

homologous (i.e., structures with similar sequences) and similar-shaped structures are shown to be clustered together.

4.3. Database search using d_{iKam} with synthetic cryo-EM data

We next demonstrate the ability of d_{iKam} to accurately find a match for the moments computed from projection images to a database of analytical moments computed from the atomic coordinates of known structures. To test our metric, we use the same dataset as the previous section, selecting the protein structure of a Mas-related G-protein-coupled receptor (available as entry PDB-7VV3⁽²⁶⁾) from our database described in Section 4.2. We use this entry because our database includes several similarly shaped yet nonidentical structures, on which we examine our metric's performance.

We generate a synthetic cryo-EM dataset as illustrated in Figure 3: We take 25000 clean projection images from a nonuniform distribution over $SO(3)$ at viewing angles given by a mixture of three von Mises–Fisher distributions.⁽²⁷⁾ To simulate cryo-EM data, the images are then corrupted with one of 100 unique radial CTFs, after which we add white noise with a signal-to-noise ratio (SNR) of 0.1. We define the SNR by taking the signal as the average squared intensity over each pixel in all the clean images and setting the noise variance to the appropriate ratio of the signal. These simulated images are generated using the ASPIRE software package⁽²⁸⁾ and have parameters consistent with many experimental datasets.

We then compute the moments of the simulated images as will be shown in (12) and (13) and compare them to the database of moments using the image-to-volume metric described in (8). We also report the

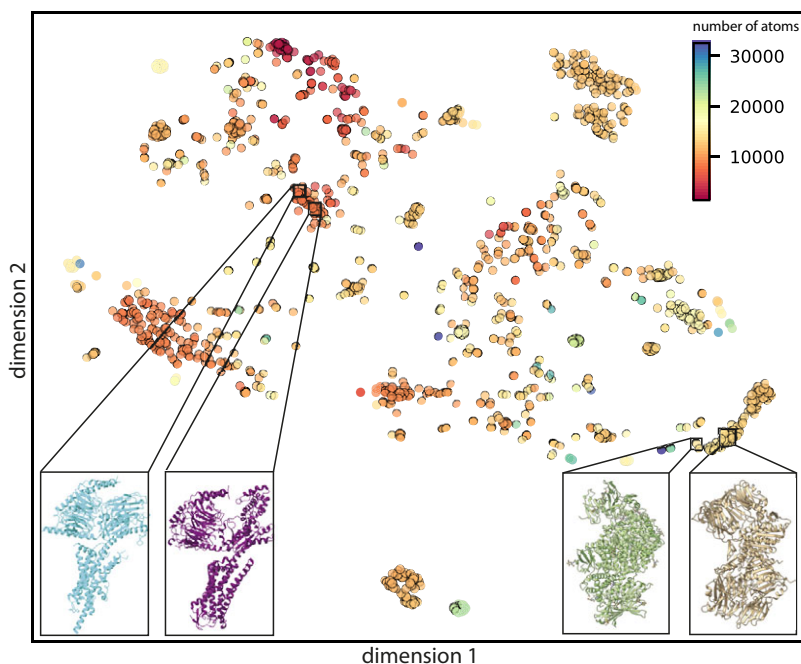


Figure 2. 2D embedding of protein structures based on their similarity using d_{vKam} . The analytical moments of 1420 proteins were computed and compared using (6), and t-SNE was applied for visualization. Each node represents a single structure and is colored by the number of atoms. Distinct clusters containing homologous or similarly shaped structures suggest that d_{vKam} provides interpretable results.

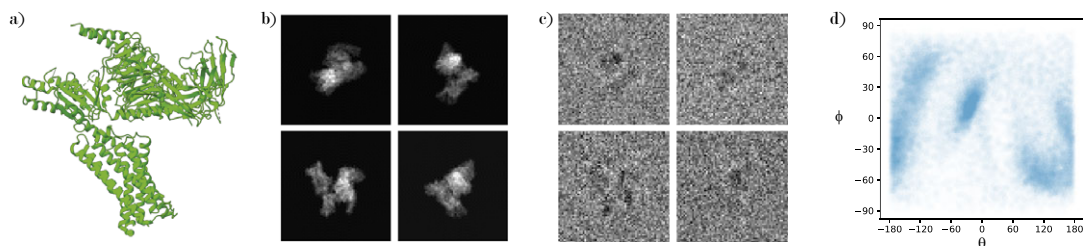


Figure 3. Visualization of the generation of simulated images. (a) Protein structure of PDB-7VV3. (b) Clean projection images from PDB-7VV3 generated with a nonuniform viewing angle distribution. (c) Projection images corrupted with a CTF and white noise with $\text{SNR}=0.1$. (d) Distribution of nonuniform viewing angles.

effect of varying the number of images on the metric's performance in Appendix B.3. Using our metric, we can rank the similarity of the image's moments to our database as shown in Figure 4. We show that the most similar score (*i.e.*, the smallest value in image Kam's metric) corresponds to the ground truth structure used to generate the images. Furthermore, based on our results, the next top 116 structures correspond to structures with similar volumes and sequences. These results demonstrate that we are able to compare directly between noisy, CTF-corrupted images and known structures. This approach could be especially valuable if there is no known model for initialization in 3D reconstruction or if the molecule generating the images is unknown.⁽²⁹⁾

We report alignment scores between molecules in our database to PDB entry 7VV3, compare these to our metric's scores, and plot the results in Figure 5. Most notably, when the protein structure becomes less

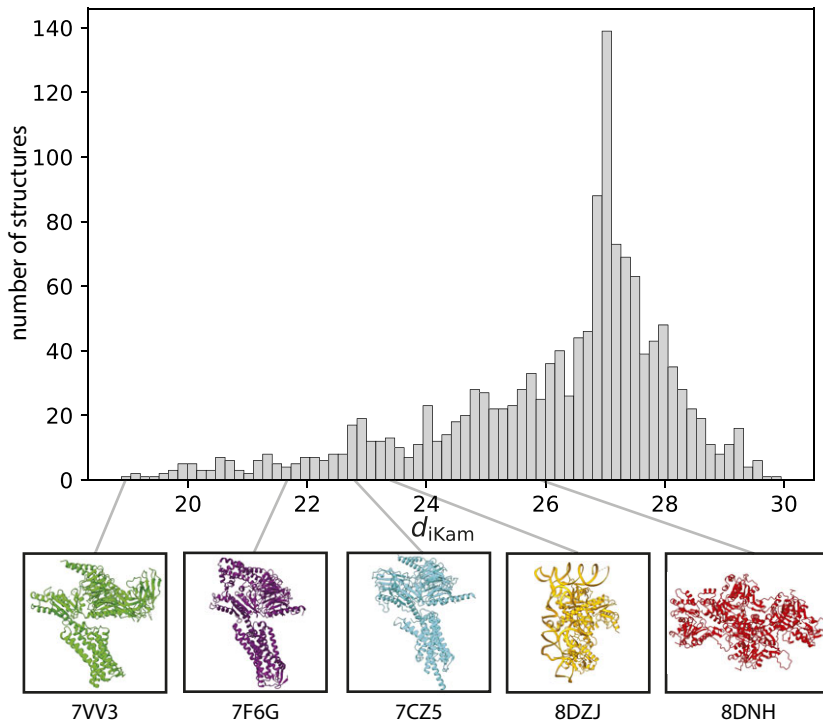


Figure 4. Histogram ranking of dissimilarities computed using d_{iKam} on simulated noisy projection images generated from PDB-7VV3.

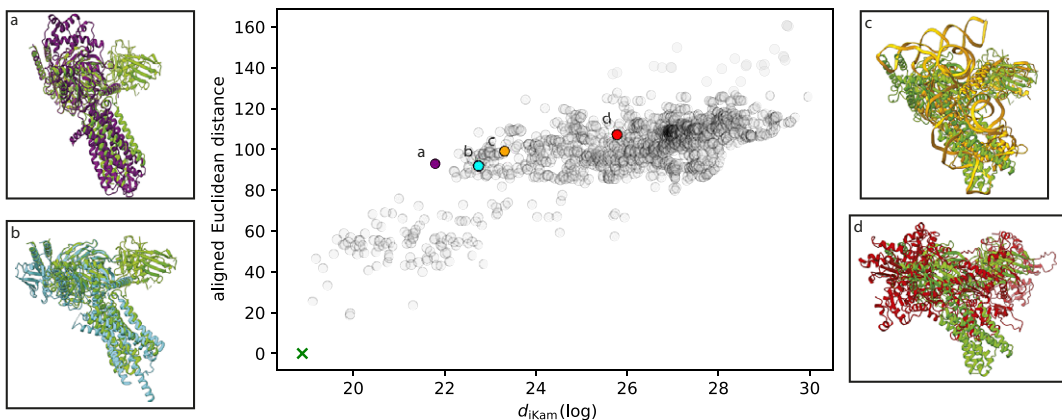


Figure 5. Comparison between the rankings given by d_{iKam} (computed from simulated images) and the minimum Euclidean distance after alignment (computed from volumes). The structures shown are superimposed with the ground truth after alignment in panels (a)–(d). The points on the graph that correspond to these structures are colored and labeled. The ground truth corresponds to the green cross in the lower left.

Note that the Euclidean alignment metric shows stagnation whereas Kam's metric does not.

similar to the ground truth (7VV3), the alignment metric begins to lose discriminative power. Figure 5 shows structures with varying degrees of dissimilarity as having the same score (~ 100). In contrast, our metric retains discriminative power, ranking structures with similar sequences/functions before structures with similar shapes.

Alignment via Bayesian optimization between one structure and the 1420 structures in the database took 95 minutes using the hardware described in Section 4.2. Aside from the computational cost, the interpretation of the optimal rotation returned by alignment becomes unclear when comparing two structures that are not volumetrically similar. On the other hand, our metric does not return an alignment between two structures, which could render it less useful when an explicit alignment must be computed. Without this alignment, it may become harder to visually compare their volumes.

It is computationally costly to generate and perform moment estimation on synthetic images for every molecule in the database. As such, to compare the performance of our metric against the Zernike metric, we select from our database a random subset of 100 structures. For each structure, we repeat the process we perform on PDB-7VV3: First, we generate a nonuniform distribution over $\mathbb{S}^2$ as a mixture of three von Mises–Fisher distributions with random means, weights, and covariance matrices. We then generate 25000 images, corrupt with SNR = 0.1 and radial CTFs, compute the moments, and search across the database.

For every structure, we recover the ground truth as one of the first six lowest-scoring molecules. Moreover, 88 of the 100 tests recovered the ground truth as the lowest-scoring molecule. To evaluate how well the metrics agree on structural similarity, we compute the size of the intersection between the top ten structures returned by our metric and those returned by the Zernike metric. As shown in Figure 6, we find that the metrics agree on two to three structures, and a large number of structures agree only on the ground truth structure. When they occur, disagreements between the metrics are likely due to the presence of near-identical molecules in the database.

4.4. Toward matching experimental datasets by d_{iKam}

While our simulated result shows success in matching a synthetic cryo-EM dataset to PDB structures, many experimental cryo-EM datasets are corrupted by a large number of unmodeled effects that we have not considered. Among the real-data effects are scattering potential's corruption by a solvent effect,⁽³⁰⁾ the B-factor,⁽³¹⁾ a global scaling ambiguity, imperfect centering, junk particles, non-radial CTF, and imperfect noise model. Our simulation falls short on these counts.

In a first step toward applying d_{iKam} to real experimental datasets, we compare the moments of a stack of images deposited in the Electron Microscopy Public Image ARchive (EMPIAR)⁽³²⁾ to the moments of its preprocessed 3D reconstructions given by the program CryoSPARC⁽³³⁾. We select the dataset EMPIAR-10076,⁽³⁴⁾ a heterogeneous dataset containing five major structures. The dataset is well

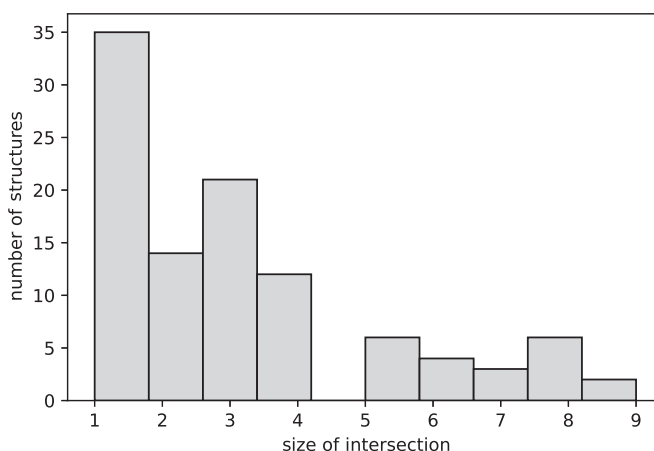


Figure 6. Comparison between d_{iKam} , computed from simulated images, and the Zernike metric, computed from volumes. Here, we repeat simulated experiments 100 times. Then, the size of the intersection of the top ten structures returned by d_{iKam} and the Zernike metric is plotted as a histogram.

characterized, and each image in the dataset has been classified into one of the five major states⁽³⁴⁾ or “junk” particles, which we discard. We use the classification to generate five separate datasets, allowing us to compute five different moments, one for each of the major states. This test case allows us to examine our metric matching on a real dataset, while bypassing some of the issues associated with comparing datasets and volumes obtained in different experimental conditions.

We downsample the image stack to 64×64 , center using the deposited shift, and mask the images with a circular binary mask of radius 0.8 times half the side length of the image. We then estimate the moments for each structure and compare them to moments computed analytically from preprocessed volume reconstructions of the five major structures, as well as two other distinctly shaped ribosomes from the Electron Microscopy Data Bank⁽³⁵⁾ (EMDB), EMD-8457 and EMD-2660, used as a baseline. Scaling issues between the moment computed from the images and the moment computed from the volume are resolved by examining the diagonal entries of the second moments. Specifically, we find a multiplicative scaling factor that best matches the diagonal of the image-computed second moment and those of the volume-computed second moment under a uniform distribution with respect to the l^2 norm.

As shown in Figure 7, it is observed that Kam’s metric recovers the ground truth structure at the lowest distance for the experimental images corresponding to structure 001. We note that the scores for molecules 001 and 002, as well as molecules 003 and 004, are almost identical in value. Also, we find that the analytical moments are closer to each other than to the experimentally determined moments. Finally, the metric reports the baseline structures, which are very different in shape and size, at the largest distances.

In Figure 8, we plot the distances between the five reconstructions (or in the case of d_{iKam} , their experimental images) and the seven candidate structures given by both of our metrics. The exact values for d_{iKam} are given in Appendix B.4. There is also scaling ambiguity in d_{vKam} since our reconstructions are preprocessed; hence, we use the same approach as above: We scale each candidate structure’s moment by a multiplicative scaling factor that best matches the candidate structure’s diagonal entries of the second moment with those of the ground truth structure. Analyzing the trends in each row, we observe that the

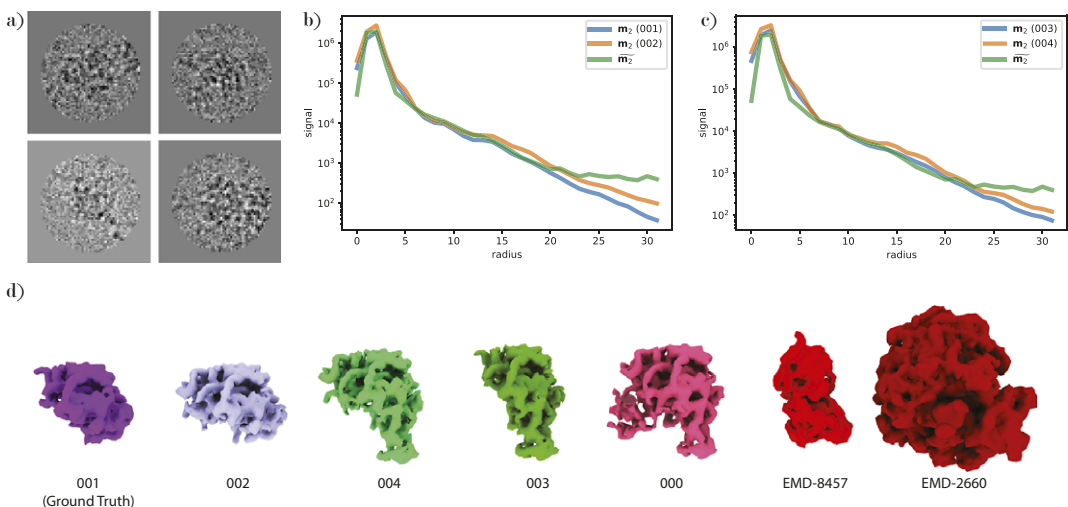


Figure 7. d_{iKam} visualization and ranking results for experimental data corresponding to structure 001 (a) Experimental images from EMPIAR-10076 corresponding to structure 001 downsampled to 64×64 pixels, centered, and with binary mask applied. (b) Comparison between diagonal entries of the second moment computed from the reconstructed volumes 001 and 002 and the moment estimated from experimental images corresponding to structure 001. (c) Comparison between diagonal entries of the second moment computed from the reconstructed volumes 003 and 004 and the moment estimated from experimental images corresponding to structure 001. (d) The five reconstructions (000–004) and two baseline structures (EMD-8457 and EMD-2600) ranked using d_{iKam} , ordered from left to right.

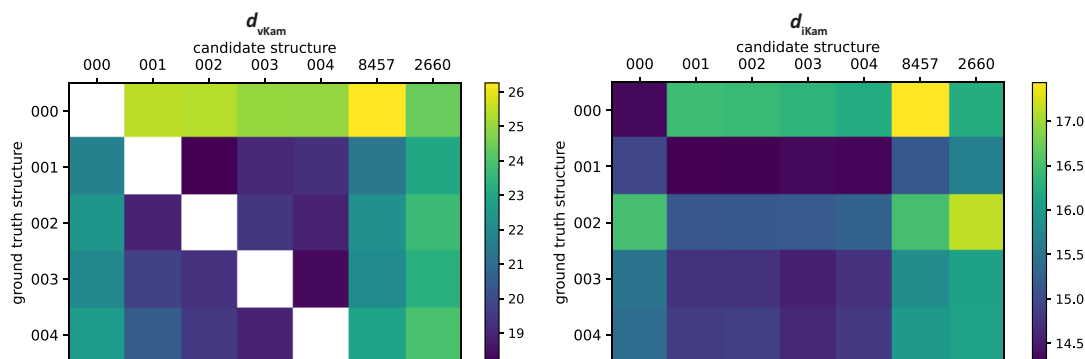


Figure 8. Visualization of $\log(d_{vKam})$ and $\log(d_{iKam})$ values on the seven candidate structures. Here, EMD-8457 and EMD-2660 are listed as 8457 and 2660 for brevity. Note that there are five ground truth structures but seven candidate structures since EMD-2660 and EMD-8457 are baseline structures for which there are no images in the experimental dataset.

metrics seem to agree on the general ranking of the molecules. While the structures 001, 002 and 003, 004 are very similar, d_{vKam} shows that the metric distinguishes between them given accurate moment estimation, whereas d_{iKam} loses some discriminative power. However, when it comes to distinct molecules such as EMD-8457 and EMD-2660, both metrics agree on their rankings.

5. Limitations and future work

d_{iKam} currently falls short of being directly applicable to experimental datasets. As stated in Section 4.4, there are several unmodeled effects not considered in this work that could lead to unexpected results for real data. The net effect of ignoring these experimental considerations is to bias our moment estimator, which may explain the inability of d_{iKam} to detect the smaller differences between structures 001 and 002, as well as 003 and 004. Developing an estimator that is robust to outliers (such as junk particles) could help alleviate this.

While we address a few of these parameters, we do so with prior knowledge. For example, the shifts used to center images are a byproduct of the reconstruction process. In future work, we aim to develop methods to correct these effects directly from the raw images. Likewise, here we have controlled for experiment-specific artifacts by using images and structures resolved from the same experiment, whereas in the future we wish to compare across all structures. Furthermore, in the future we seek to compare moments computed from real data directly to the PDB, by appropriately correcting for the discrepancies between PDB and reconstructed structures.

Even with our current mitigations, issues such as the B-factor and inaccuracies in the noise model remain completely unmodeled. Further studies will be required to investigate which of these omissions is important and which can safely be made. Then, our method could be modified to account for the important effects.

6. Discussion

We introduced structural similarity metrics for proteins based on moments, inspired by the moment computation in Kam's method. d_{vKam} compares known 3D structures according to the difference between the moments of their potentials. We showed that the metric accurately captures similarity according to the rotationally aligned Euclidean metric, an interpretable but expensive-to-compute molecular similarity metric. Therefore, d_{vKam} allows for the efficient comparison of large number of known structures. A potential application is to improve the similarity search presently in the PDB, which uses the Zernike metric – a fast but less principled metric that involves learning weights and which our results suggest performs worse than ours.

A second metric, termed d_{iKam} , allows for the computation of a similarity score between an unknown structure present in a large cryo-EM dataset and a solved structure. The computation of this metric does not require a 3D reconstruction process for the image stack and therefore is very efficient. We showed on

simulated projection images that our method could discriminate between even very similar proteins with reasonably sized datasets. If it were to work on experimental datasets, d_{iKam} could become a versatile tool for 3D reconstruction. Typical reconstruction algorithms used in practice are only locally optimal and thus require good initialization, which d_{iKam} could provide by returning the homologous structures present in the PDB. By extending the database to the entirety of the PDB and including structure predictions, both solved and predicted structures could be quickly compared against.

Beyond its application to experiments, d_{iKam} demonstrates that Kam's method is a feasible strategy for high-resolution reconstruction. Recent works have improved the viability of Kam's method by using sparsity⁽⁷⁾ or neural network⁽³⁶⁾ priors; likewise, the search over the PDB using Kam's metric can be interpreted as simply running Kam's method under a very strong prior, where only a finite number of structures appear with nonzero probability. Our results suggest that, if one could formulate a tractable prior over the manifold of proteins, Kam's method could yield high-resolution reconstructions.

Acknowledgments. J.K. and M.A.G. thank Bronson Zhou for helpful conversations.

Data availability statement. Replication code can be found at <https://github.com/aszhang107/moment-based-metrics/>.

Author contribution. All authors conceived of the project and designed the algorithms. A.Z. wrote the software and performed the experiments. All authors wrote the manuscript and approved its submission.

Funding statement. Part of this research was performed while authors J.K., E.J.V., N.F.M., M.A.G., and A.S. were visiting the long program on Computational Microscopy at the Institute for Pure and Applied Mathematics, which is supported by NSF DMS 1925919. A.Z., O.M., E.J.V., M.A.G., and A.S. are supported in part by AFOSR FA9550-20-1-0266, the Simons Foundation Math +X Investigator Award, NSF DMS 2009753, and NIH/NIGMS R01GM136780-01. J.K. is supported in part by NSF DMS 2309782, NSF CISE-IIS 2312746, and start-up grants from the College of Natural Science and Oden Institute at the University of Texas at Austin. N.F.M. is supported in part by a start-up grant from Oregon State University.

Competing interest. The authors declare none.

References

1. Kam Z (1980) The reconstruction of structure from electron micrographs of randomly oriented particles. *Journal of Theoretical Biology* **82**(1), 15–39.
2. Bhamre T, Zhang T and Singer A (2017) Anisotropic twicing for single particle reconstruction using autocorrelation analysis. Preprint, [arXiv:1704.07969](https://arxiv.org/abs/1704.07969).
3. Bhamre T, Zhang T and Singer A (2015) Orthogonal matrix retrieval in cryo-electron microscopy. In *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*. Brooklyn, NY, USA: IEEE, pp. 1048–1052.
4. Levin E, Bendory T, Boumal N, Kileel J and Singer A (2018) 3D ab initio modeling in cryo-EM by autocorrelation analysis. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. Washington, DC, USA: IEEE, pp. 1569–1573.
5. Bandeira AS, Blum-Smith B, Kileel J, Niles-Weed J, Perry A and Wein AS (2023) Estimation under group actions: Recovering orbits from invariants. *Applied and Computational Harmonic Analysis*, **66**, 236–319.
6. Huang S, Zehni M, Dokmanić I and Zhao Z (2023) Orthogonal matrix retrieval with spatial consensus for 3D unknown view tomography. *SIAM Journal on Imaging Sciences* **16**(3), 1398–1439.
7. Bendory T, Khoo Y, Kileel J, Mickelin O and Singer A (2023) Autocorrelation analysis for cryo-EM with sparsity constraints: improved sample complexity and projection-based algorithms. *Proceedings of the National Academy of Sciences* **120**(18), e2216507120.
8. Sharon N, Kileel J, Khoo Y, Landa B and Singer A (2020) Method of moments for 3-D single particle ab initio modeling with non-uniform distribution of viewing angles. *Inverse Problems* **36**(4), 044003.
9. Bendory T, Hadi I and Sharon N (2022) Compactification of the rigid motions group in image processing. *SIAM Journal on Imaging Sciences* **15**(3), 1041–1078.
10. Berman HM (2000) The protein data bank. *Nucleic Acids Research* **28**(1):235–242.
11. NIST Digital Library of Mathematical Functions. Release 1.1.11 of 2023-09-15. In Olver FWJ, Olde Daalhuis AB, Lozier DW, Schneider BI, Boisvert RF, Clark CW, Miller BR, Saunders BV, Cohl HS and McClain MA (eds.). Available at <https://dlmf.nist.gov/>.
12. Ponce C and Singer A (2011) Computing steerable principal components of a large set of images and their rotations. *IEEE Transactions on Image Processing* **20**(11), 3051–3062.
13. Bredon GE (1993) *Topology and geometry* (Vol. **139**). Springer New York, NY: Springer Science & Business Media.
14. Scheres SHW (2012) RELION: implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology* **180**(3), 519–530.

15. Scheres SHW (2012) A Bayesian view on cryo-EM structure determination. *Journal of Molecular Biology* **415**(2), 406–418.
16. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH and Ferrin TEF (2020) UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science* **30**(1), 70–82.
17. Rangan AV (2022) Radial recombination for rigid rotational alignment of images and volumes. *Inverse Problems* **39**(1), 015003.
18. Harpaz Y and Shkolnisky Y (2023) Three-dimensional alignment of density maps in cryo-electron microscopy. *Biological Imaging* **3**, e8.
19. Bartesaghi A, Sprechmann P, Liu J, Randall G, Sapiro G and Subramaniam S (2008) Classification and 3d averaging with missing wedge correction in biological electron tomography. *Journal of Structural Biology* **162**(3):436–450.
20. Xu M, Beck M and Alber F (2012) High-throughput subtomogram alignment and classification by fourier space constrained fast volumetric matching. *Journal of Structural Biology* **178**(2), 152–164.
21. Singer A and Yang R (2023) Alignment of density maps in Wasserstein distance. Preprint, [arXiv:2305.12310](https://arxiv.org/abs/2305.12310).
22. Guzenko D, Burley SK and Duarte JM (2020) Real time structural search of the Protein Data Bank. *PLoS Computational Biology* **16**(7), e1007970.
23. Cianfrocco MA and Kellogg EH (2020) What could go wrong? A practical guide to single-particle cryo-EM: From biochemistry to atomic models. *Journal of Chemical Information and Modeling* **60**(5), 2458–2469.
24. Järvelin K and Kekäläinen J (2002) Cumulated gain-based evaluation of IR techniques. *ACM Transactions on Information Systems (TOIS)* **20**(4), 422–446.
25. Van der Maaten L and Hinton G (2008) Visualizing data using t-SNE. *Journal of Machine Learning Research* **9**(86), 2579–2605.
26. Yang F, Guo L, Li Y, Wang G, Wang J, Zhang C, Fang GX, Chen X, Liu L, Yan X and Liu Q (2021) Structure, function and pharmacology of human itch receptor complexes. *Nature* **600**(7887), 164–169.
27. Watson GS (1982) Distributions on the circle and sphere. *Journal of Applied Probability* **19**(A), 265–280.
28. Wright G, Andén J, Bansal V, Xia J, Langfield C, Carmichael J, Brook R, Shi Y, Heimowitz A, Pragier G, Sason I, Moscovich A, Shkolnisky Y and Singer A (2023) ASPIRE – Algorithms for single particle reconstruction software package.
29. Verbeke EJ, Mallam AL, Drew K, Marcotte EM and Taylor DW (2018) Classification of single particles from human cell extract reveals distinct structures. *Cell Reports*, **24**(1), 259–268.e3.
30. Shang Z and Sigworth FJ (2012) Hydration-layer models for cryo-em image simulation. *Journal of Structural Biology* **180**(1), 10–16.
31. Rosenthal PB and Henderson R (2003) Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *Journal of Molecular Biology* **333**(4), 721–745.
32. Iudin A, Korir PK, Somasundharam S, Weyand S, Cattavittello C, Fonseca N, Salih O, Kleywegt GJ and Patwardhan A (2022) EMPIAR: The electron microscopy public image archive. *Nucleic Acids Research* **51**(D1), D1503–D1511.
33. Punjani A, Rubinstein JL, Fleet DJ and Brubaker MA (2017) cryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods* **14**(3), 290–296.
34. Davis JH, Tan YZ, Carragher B, Potter CS, Lyumkis D and Williamson JR (2016) Modular assembly of the bacterial large ribosomal subunit. *Cell* **167**(6), 1610–1622.e15.
35. Turner J and The wwPDB Consortium (2023) EMDB - The Electron Microscopy Data Bank.
36. Khoo Y, Paul S and Sharon N (2023) Deep neural-network prior for orbit recovery from method of moments. *Journal of Computational and Applied Mathematics*, **444**, 115782.
37. Biedenharn LC and Louck JD (1981) *Angular Momentum in Quantum Physics: Theory and Application*, Volume 8 of Encyclopedia of Mathematics and Its Applications. Reading, MA: Addison-Wesley Publishing Co..
38. Barnett AH, Magland J and Klinteberg L (2019) A parallel nonuniform fast Fourier transform library based on an “exponential of semicircle” kernel. *SIAM Journal on Scientific Computing* **41**(5), C479–C504.
39. Barnett AH (2021) Aliasing error of the $\exp(\beta\sqrt{1-z^2})$ kernel in the nonuniform fast Fourier transform. *Applied and Computational Harmonic Analysis* **51**, 1–16.
40. Peng L-M, Ren G, Dudarev S and Whelan M (1996) Robust parameterization of elastic and absorptive electron atomic scattering factors. *Acta Crystallographica Section A, Foundations of Crystallography* **52**, 257–276.
41. Singer A (2021) Wilson statistics: Derivation, generalization and applications to electron cryomicroscopy. *Acta Crystallographica Section A, Foundations and Advances* **77**(Pt 5), 472–479.
42. Driscoll JR and Healy DM (1994) Computing Fourier transforms and convolutions on the 2-sphere. *Advances in Applied Mathematics* **15**(2), 202–250.
43. Wieczorek MA and Meschede M (2018) SHTools: Tools for working with spherical harmonics. *Geochemistry, Geophysics, Geosystems* **19**(8), 2574–2592.
44. Marshall NF, Mickelin O, Shi Y and Singer A (2023) Fast principal component analysis for cryo-electron microscopy images. *Biological Imaging* **3**, e2.
45. Marshall NF, Mickelin O and Singer A (2023) Fast expansion into harmonics on the disk: A steerable basis with fast radial convolutions. *SIAM Journal on Scientific Computing* **45**(5), A2431–A2457.

Appendix

A. Methodology

In this section, we describe the computational details of the method.

A.1. Moment derivation

Prior work⁽⁸⁾ has shown that the analytical first and second moments of cryo-EM images generated by $\hat{\Phi}$ and ρ equal

$$\mathbf{m}_1(\hat{\Phi}, \rho) = \sum_{\ell, m} B_{\ell, -m} A_{\ell, m}(r) N_{\ell}^0 \frac{1}{2\ell+1} (-1)^m, \quad (12)$$

where the sum ranges over (ℓ, m) such that $0 \leq \ell \leq \min(L, P)$, ℓ is even, and $-\ell \leq m \leq \ell$, and

$$\mathbf{m}_2(\hat{\Phi}, \rho) = \sum_n e^{in(\phi' - \phi)} \sum_{\ell, m, \ell', m', \ell'', m''} A_{\ell, m}(r) A_{\ell', m'}(r') N_{\ell}^m N_{\ell'}^{-n} B_{\ell'', -m-m'} C_{\ell''}(\ell, \ell', m, m', n, -n) \frac{(-1)^{m+m'}}{2\ell''+1}, \quad (13)$$

where

$$N_{\ell}^n := \begin{cases} \sqrt{\frac{2\ell+1}{4\pi}} \frac{(\ell-n)!}{(\ell+n)!} \frac{(-1)^n}{2^{\ell} \ell!} (\ell+n)! \binom{\ell}{(\ell+n)/2} (-1)^{(\ell-n)/2} & \text{if } n \equiv \ell \pmod{2} \\ 0 & \text{if } n \not\equiv \ell \pmod{2} \end{cases} \quad (14)$$

is an explicitly calculated constant,

$$C_{\ell''}(\ell, \ell', m, m', n, n') := C(\ell, m; \ell', m' | \ell'', m+m') C(\ell, n; \ell', n' | \ell'', n+n') \quad (15)$$

is a product of Clebsch–Gordan coefficients,⁽³⁷⁾ and the sum ranges over those indices $n, \ell, m, \ell', m', \ell''$ that satisfy

$$\begin{aligned} -L \leq n \leq L, \quad |n| \leq \ell \leq L, \quad |n| \leq \ell' \leq L, \quad -\ell \leq m \leq \ell, \quad -\ell' \leq m' \leq \ell', \\ \ell \equiv \ell' \equiv n \pmod{2}, \quad \max(|\ell - \ell'|, |m + m'|) \leq \ell'' \leq \min(\ell + \ell', P). \end{aligned} \quad (16)$$

See Sections 2.3.1 and 2.3.2 in,⁽⁸⁾ respectively, for the derivations of (12) and (13). In the case of the uniform density on $\text{SO}(3)$, we note that $N_0^0 = \frac{1}{\sqrt{4\pi}}$ so (12) and (13) simplify to the following:

$$\mathbf{m}_1(r) = \sqrt{\frac{1}{4\pi}} A_{0,0}(r), \quad (17)$$

$$\begin{aligned} \mathbf{m}_2(\Delta\phi, r, r') &= \frac{1}{4\pi} \sum_{\ell, m} A_{\ell, m}(r) A_{\ell, -m}(r') P_{\ell}(\cos(\Delta\phi + \pi)) \\ &= \frac{1}{4\pi} \sum_{\ell, m} A_{\ell, m}(r) \overline{A_{\ell, m}(r')} P_{\ell}(\cos(\Delta\phi)), \end{aligned} \quad (18)$$

where P_{ℓ} is the Legendre polynomial of degree ℓ and $\bar{z}$ denotes the complex conjugate of a complex number $z \in \mathbb{C}$. The simplification of (12) to (17) is immediate, whereas the simplification of (13) to (18) uses the sum rule for spherical harmonics; see (10) in the study by Kam.⁽¹⁾

A.2. Uniform case

This section details the method to compute d_{vKam} . Our algorithm takes as input a PDB identifier (a list of atomic coordinates), on which we center the atomic positions by subtracting the molecule's center of mass. Then, we use the three-dimensional nonuniform fast Fourier transform (NUFFT)^(38,39) to compute the discrete Fourier transform evaluated on a grid in spherical coordinates, that is, to compute

$$a_{kjl} = \sum_{i=1}^q \hat{f}_i(r_k \alpha_{ji}) e^{ix_i r_k \alpha_{ji}}, \quad \text{where } \alpha_{ji} = (\sin \theta_j \cos \varphi_l, \sin \theta_j \sin \varphi_l, \cos \theta_j), \quad (19)$$

where x_i denotes the coordinates of the i^{th} atom from the PDB identifier and q is the total number of atoms. The function $\hat{f}_i$ is the Fourier transform of the scattering potential of the i^{th} atom as reported in the study by Peng et al. and Singer.^(40,41) In real space, this corresponds to convolving a Gaussian mixture with a delta function, in other words, adding a Gaussian blob around the atom coordinate. Here, $r_k = \frac{k\delta}{N}$, $\theta_j = \frac{j\pi}{N}$, and $\varphi_l = \frac{2\pi l}{N}$ for $k=0, \dots, N/2$ and $j, l=0, \dots, N-1$ and δ is the side length of the volume grid in angstroms.

Lastly, we apply the spherical harmonic transform to a_{kjl} defined on the spherical coordinate grid $(r_k, \theta_j, \varphi_l)$ in (19) using SHTools.^(42,43) This gives us coefficients

$$A_{\ell, m}(r_k) = \sum_{\substack{0 \leq \ell \leq L \\ 0 \leq k \leq n-1}} a_{k, j, l} \overline{Y_{\ell}^m(\alpha_{ji})}.$$

Let ρ_u denote the uniform density on the sphere. In the discrete case, we sample each image as a 2D polar grid at $N/2$ radial points r and N angular points ϕ , where N is the number of pixels of one side of the projection images. In (12), the first moment $\mathbf{m}_1$ is indexed by r and is thus an $N/2$ length vector. Note that in (5), $\Delta\phi_j = 2\pi j/N - 2\pi$ for $j = 1, \dots, 2N$, but since $e^{in\Delta\phi}$ in (13) is $2\pi/n$ periodic, we have that $e^{in(2\pi-\Delta\phi)} = e^{-in\Delta\phi}$, and hence, $\mathbf{m}_2(r, r', \Delta\phi_j) = \mathbf{m}_2(r, r', \Delta\phi_{j+N})$. Thus, $\Delta\phi_j$ for $j = 1, \dots, N$ is redundant and we consider only $\Delta\phi_j$ for $j = N+1, \dots, 2N$, which enumerates $[0, 2\pi]$. Thus, $\mathbf{m}_2$ is a three-dimensional tensor of size $N \times N/2 \times N/2$, since there are N values for $\Delta\phi$ and $N/2$ values each for r and r' , where $\Delta\phi$ are points uniformly spaced between 0 and 2π and r and r' are N uniformly spaced points between 0 and δ . Equations (17) and (18) give

$$\mathbf{m}_1(k) = A_{0,0}(r_k), \quad (20)$$

$$\begin{aligned} \mathbf{m}_2(j, k_1, k_2) &= \frac{1}{4\pi} \sum_{\ell, m} A_{\ell, m}(r_{k_1}) A_{\ell, -m}(r_{k_2}) P_\ell(\cos(\Delta\phi_j + \pi)) \\ &= \frac{1}{4\pi} \sum_{\ell, m} A_{\ell, m}(r_{k_1}) \overline{A_{\ell, m}(r_{k_2})} P_\ell(\cos(\Delta\phi_j)). \end{aligned} \quad (21)$$

We then compute the metric given in (10). To better approximate the L_2 norm in the continuous case, we scale the difference of each entry $\mathbf{m}_2(j, k_1, k_2)$ by $\sqrt{r_{k_1} r_{k_2}}$ so that the squared norm is scaled by $r_{k_1} r_{k_2}$. More precisely, we define weighted ℓ^2 norms $\|\cdot\|_{w_1}$ and $\|\cdot\|_{w_2}$ on $\mathbb{R}^{N/2}$ and $\mathbb{R}^{N \times N/2 \times N/2}$, as described in (5). Let Φ and Φ' be two different molecules, and $\mathbf{m}_1, \mathbf{m}_2$ and $\mathbf{m}_1', \mathbf{m}_2'$ be the first and second moment tensors, respectively, from two different molecules. We define the distance between the moments as in (6).

A.3. Least squares for the nonuniform case

This section describes the process for generating and solving the least-squares system for B , the matrix encoding the viewing angle distribution. We use the following convention for the vectorization operator $\text{vec}(\cdot)$: If $\mathcal{M} \in \mathbb{C}^{i \times j}$, $\text{vec}(\mathcal{M})$ returns a vector of dimension ij obtained by stacking the columns of $\mathcal{M}$, that is,

$$\text{vec}\left(\begin{bmatrix} a_{1,1} & \cdots & a_{1,j} \\ \vdots & \ddots & \vdots \\ a_{i,1} & \cdots & a_{i,j} \end{bmatrix}\right) := [a_{1,1} \ \cdots \ a_{i,1} \ a_{1,2} \ \cdots \ a_{i,j}]^T. \quad (22)$$

The first moment is linear in B as shown in (12), so fitting a viewing distribution to observed moments can be solved through a least-squares problem. We detail this procedure in Algorithm 1.

Algorithm 1: Computation of least-squares matrix V for $\mathbf{m}_1$.

```

1 initialize  $V[i = 1, \dots, N/2] [(p = 1, \dots, P; m = -p, \dots, p)] \leftarrow 0$ .
2 for  $i = 1, \dots, N/2$  do.
3   for  $p = 1, \dots, P$  do.
4     for  $m = -p, \dots, p$  do.
5        $V[i] [(p; -m)] \leftarrow A_{\ell, m}(r_i) N_0^0 \frac{(-1)^m}{2\ell+1}$ 
6     end.
7   end.
8 end.
9 return  $V$ .
```

Algorithm 2: Computation of least-squares matrix $\mathcal{U}_{\ell, \ell'}^n$ for $\mathcal{B}_{\ell, \ell'}^n$.

```

1 Initialize  $\mathcal{U}_{\ell, \ell'}^n [i = 1, \dots, (2\ell+1)(2\ell'+1)] [(\ell'' = 1, \dots, P; m = -\ell'', \dots, \ell'')] \leftarrow 0$ .
2 for  $i = 1, \dots, (2\ell+1)(2\ell'+1)$  do.
3   for  $m = -\ell, \dots, \ell$  do.
4     for  $m' = -\ell', \dots, \ell'$  do.
5       for  $\ell'' = \max(|\ell - \ell'|, |m + m'|), \dots, \min(\ell' + \ell, P)$  do.
6          $\mathcal{U}_{\ell, \ell'}^n [i] [(\ell'', -m - m')] \leftarrow \mathcal{U}_{\ell, \ell'}^n [i] [(\ell'', -m - m')] + C_{\ell''}(\ell, \ell', m, m', n, -n) \frac{(-1)^{m+m'}}{2\ell''+1}$ 
7       end.
8     end.
9   end.
10 end.
11 return  $\mathcal{U}_{\ell, \ell'}^n$ .
```

For the second moment, we rewrite (13) more compactly:

$$\mathbf{m}_2(\Delta\phi) = \sum_n e^{in\Delta\phi} \sum_{\ell, \ell'} \mathcal{A}_\ell \mathcal{B}_{\ell, \ell'}^n \mathcal{A}_{\ell'}^*, \quad (23)$$

where

$$\left(\mathcal{B}_{\ell, \ell'}^n \right)_{m, m'} = \sum_{\ell''} N_\ell^n N_{\ell'}^{-n} B_{\ell'', -m-m'} C_{\ell''}(\ell, \ell', m, m', n, -n) \frac{(-1)^{m+m'}}{2\ell''+1} \quad (24)$$

is a matrix of size $(2\ell+1) \times (2\ell'+1)$ indexed by $m_1 = -\ell \dots \ell$ and $m_2 = -\ell' \dots \ell'$, and $(\mathcal{A}_\ell)_{m,r} = A_{\ell,m}(r)$ is a matrix of spherical harmonic coefficients indexed by m, r . Here, the sum ranges are detailed in (16). Since $\mathcal{B}$ is linear in B , we use Algorithm 2 to construct many linear systems $\mathcal{U}_{\ell, \ell'}^n$ such that

$$\mathcal{U}_{\ell, \ell'}^n \text{vec}(B) = \text{vec} \left(\mathcal{B}_{\ell, \ell'}^n \right),$$

Using the Kronecker product, (23) can be written as

$$\text{vec}(\mathbf{m}_2(\Delta\phi)) = \sum_n e^{in\Delta\phi} \sum_{\ell, \ell'} (\mathcal{A}_{\ell'} \otimes \mathcal{A}_\ell) \mathcal{U}_{\ell, \ell'}^n \text{vec}(B).$$

This, too, is linear in B :

$$\text{vec}(\mathbf{m}_2(\Delta\phi)) = U(\Delta\phi) \text{vec}(B), \quad \text{where} \quad U(\Delta\phi) = \sum_n e^{in\Delta\phi} \sum_{\ell, \ell'} (\mathcal{A}_{\ell'} \otimes \mathcal{A}_\ell) \mathcal{U}_{\ell, \ell'}^n. \quad (25)$$

By vertically appending V and copies of $U(\Delta\phi)$ for all values of $\Delta\phi$ in Section 3.1, we obtain the least-squares formulation

$$\min_x \|Ax - b\|,$$

where

$$A = \begin{pmatrix} V \\ U(\Delta\phi_1) \\ \vdots \\ U(\Delta\phi_n) \end{pmatrix}, \quad x = \text{vec}(B), \quad \text{and} \quad b = \begin{pmatrix} \text{vec}(\mathbf{m}_1) \\ \text{vec}(\mathbf{m}_2) \end{pmatrix}. \quad (26)$$

To solve this, we perform QR decomposition $A = QR$ and then solve the normal equations

$$A^* Ax = R^* Rx = A^* b = R^* Q^* b,$$

that is, we solve $Rx = Q^* b$. Since R is a square upper triangular matrix, we solve this using back substitution.

A.4. Change of bases for moment comparison

We compute moments from images using the fast method⁽⁴⁴⁾ that produces the moments expanded in the Fourier Bessel basis. Thus, a change of bases is required for moment comparison. The Fourier Bessel basis has several nice properties that make it advantageous to use when computing the moment from images; it is orthonormal, frequency-ordered, steerable, provides fast radial convolutions, and has a fast transform.⁽⁴⁵⁾ The Fourier Bessel basis functions can be written in polar coordinates (r, θ) as

$$\psi_{n,k}(r, \theta) = c_{n,k} J_n(\lambda_{nk} r) e^{in\theta}, \quad (27)$$

where J_n is a Bessel function of the first kind of order n , and λ_{nk} is the k -th smallest positive zero of J_n , and $c_{n,k}$ is a normalization constant.

We create a change of basis matrix $(B)_{(r, \theta), (n, k)} = \psi_{n,k}(r, \theta)$ by sampling on a Cartesian grid $(x, y) \in \{r_i\} \times \{r_i\}$ with the $\{r_i\}$ grid defined as in Section 3.1, where (r, θ) are the grid points (x, y) in polar coordinates. This yields the moments in real space

$$\begin{aligned} \mathbf{m}_1(\Phi, \rho)(x, y) &= B \mathbf{m}_1(\Phi, \rho)(n, k), \\ \mathbf{m}_2(\Phi, \rho)(x_1, y_1, x_2, y_2) &= B \mathbf{m}_2(\Phi, \rho)(n_1, k_1, n_2, k_2) B^*. \end{aligned}$$

Now, we compute the NUFFT to convert the moments into radially sampled polar coordinates in Fourier space as in (19). In practice, we do this for $\mathbf{m}_2$ by taking each row (which is indexed by (x_1, y_1)), reshaping it into an image, and applying the transform. We then apply the same process to the columns indexed by (x_2, y_2) .

A.5. Computational complexity

In the following, L is the molecule bandlimit, and see (1); P is the distribution bandlimit, and see (2); M is the number of projection images; and N is the image side length of the $N \times N$ pixel images. We assume that $P \leq L$.

There are three main steps for calculating the least-squares matrix for each structure in our database. We first calculate the least-squares matrices $\mathcal{U}_{\ell, \ell'}^n$ for $\mathcal{B}$ as described in alg:BLSalg. This needs only to be done once and does not need to be recomputed for each molecule. Calculating this matrix takes $O(PL^5)$ time and uses $O(PL^5)$ space. For the calculation of the least-squares matrix

itself, we precompute $(A_{\ell'} \otimes A_{\ell}) \mathcal{U}_{\ell, \ell'}^n$, for ℓ, ℓ', n as described in (25). These intermediate steps take $O(P^2 L^2 N^2 + L^3 N^2)$ time and use $O(L^3 N^2)$ space for forming the Kronecker product and subsequent matrix multiplication. Finally, the construction of the least-squares matrix A in (26) takes $O(L^3 N^3)$ time for the scalar multiplication of a matrix for each n, ℓ, ℓ' , and the least squares use $O(P^2 N^3)$ space. As such, the total computational complexity for calculating a least-squares matrix is $O(P^2 L^5 + P^2 L^2 N^2 + L^3 N^3)$ time and $O(P^2 N^3 + L^3 N^2 + PL^5)$ space.

The computation of moments from noisy projection images in the Fourier–Bessel basis takes $O(MN^3 + N^4)$ time and uses $O(MN^2 + N^3)$ space. To convert this to polar coordinates in Fourier space, we must first evaluate the moments in the Fourier–Bessel basis. This takes $O(N^2 \log N)$ time for each expansion, and we require $2N^2$ such expansions (see app:basis). Hence, in total, this step takes $O(N^4 \log N)$ time and uses $O(MN^2 + N^4)$ space. Converting into Fourier space using the NUFFT takes $O(N^4 \log N)$ time and uses $O(N^4 \log N)$ space, and we do this $2N^2$ times for a total of $O(N^6 \log N)$ time and space complexity. Storing the final moment uses $O(N^3)$ space (since the resulting matrix from the NUFFT is block circulant). Overall, computing moments from images and converting them to polar coordinates in Fourier space take $O(N^6 \log N + MN^3)$ time and use $O(MN^2 + N^6 \log N)$ space.

A.6. NDCG score

The NDCG⁽²⁴⁾ is calculated by taking the discounted cumulative gain (DCG) and normalizing by ideal discounted cumulative gain (IDCG):

$$\begin{aligned} DCG &= \sum_i \frac{\text{true score of item } i}{\log(i+1)}, \\ IDC G &= \sum_i \frac{i^{\text{th}} \text{ highest true score}}{\log(i+1)}, \\ NDCG &= \frac{DCG}{IDCG}, \end{aligned} \tag{28}$$

where i is enumerated in the order induced by the predicted scores.

The NDCG puts weight on scores that are agreed to be high by both metrics. However, our metric and the metrics we compare to are dissimilarity scores, so we prefer weight on scores that are considered low by both metrics. To remedy this, we use the reverse of the order enumerated by the predicted scores. For the true scores, we first normalize the scores to the range $[0, 1]$ and then take the exponential e^{-s} for each true score s .

B. Additional Results

B.1. Parameter selection

In the experiments, we set the bandlimit parameters to $P=6$ and $L=25$. Note that this value of P is comparable to previous work as described in the study by Sharon et al.⁽⁸⁾, whereas the higher value of L allows for a more accurate representation of the molecule in spherical harmonics. Furthermore, the hyperparameter λ was set to be 1. As shown in tab:lambda below, varying λ does not greatly impact the performance of the metric.

Table B1. Effect of the value of the hyperparameter λ on the ranking induced by $d_{i\text{Kam}}$

λ	A_1	A_2	A_3	A_4	A_5
1e-2	7VV3	7VUZ	7TRK	7TRP	7VDM
e-1	7VV3	7VUZ	7TRK	7TRP	7VDM
1	7VV3	7VUZ	7TRK	7TRP	7VDM
e1	7VV3	7VUZ	7TRK	7TRP	7VDM
e2	7VV3	7VUZ	7TRK	7TRP	7VDM
e3	7Y15	6 K41	7VDM	7VUZ	7E33

Note: A_i denotes the structure with the i th lowest value of $d_{i\text{Kam}}$. In each row, the entry shaded green indicates the ground truth structure.

B.2. Additional t-SNE plots

This appendix includes t-SNE visualizations of the Zernike metric on our database and the alignment metric on a subset of our database of size 100. The alignment metric is restricted to a subset of size 100 since calculating pairwise distances for a 1420 is computationally taxing; see Section 4.2. Visually, the Zernike t-SNE seems to have fewer distinct clusters than the t-SNE plot generated using d_{vKam} and also groups molecules with different numbers of atoms together. It seems possible that Zernike metric is less discriminative, although this may also be an artifact of t-SNE’s dimensionality reduction.

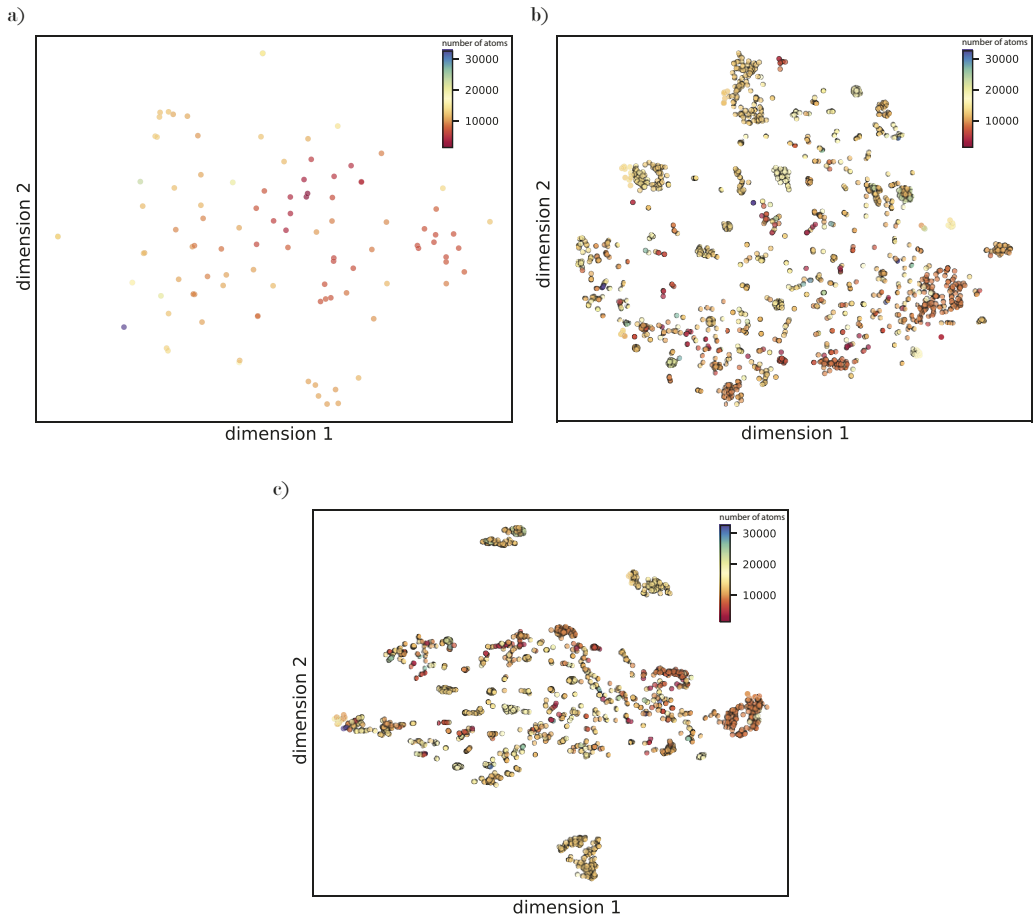


Figure B1. Additional *t*-SNE plots. (a) *t*-SNE plot of pairwise Euclidean alignment distances on a subset of size 100. (b) *t*-SNE plot of pairwise Zernike distances on the entire database. (c) *t*-SNE plot of pairwise d_{vKam} distances on the entire database.

B.3. Robustness to number of images

Here, we examine the robustness of our metric to inaccuracies of moment estimation. Specifically, we vary the number of noisy synthetic projection images that the metric has access to and record the highest-ranking structures.

Table B2. Effect of the number of projection images used for moment estimation on the ranking induced by d_{iKam}

Number of images	A_1	A_2	A_3	A_4	A_5	Relative M_1 error (%)	Relative M_2 error (%)
500	7VDM	7VUZ	7Y15	7E33	7VV3	1.49	8.23
1000	7VUZ	7VDM	7VV3	7TRK	7EJ8	0.76	6.22
2500	7VUZ	7VDM	7VV3	7TRK	7TRP	0.43	4.02
5000	7VV3	7TRK	7VUZ	7TRP	7VDM	0.27	2.95
10000	7VV3	7VUZ	7TRK	7TRP	7VDM	0.25	1.89
25000	7VV3	7VUZ	7TRK	7VDM	7TRP	0.19	1.37
50000	7VV3	7VUZ	7TRK	7VDM	7TRP	0.14	1.15

Note: A_i denotes the structure with the i th lowest value of d_{iKam} . In each row, the entry shaded green indicates the ground truth structure. The relative error in each moment is between the moment estimated from noisy projection images and the moment calculated from their clean counterparts.

B.4. Additional experimental results

Table B3 reports the metric’s rankings using experimental images corresponding to the five structures resolved from EMPIAR-10076.

Table B3. Performance of d_{iKam} on structures 001, 002, 003, 004, and 005 of EMPIAR-10076

Number of images	A_1	A_2	A_3	A_4	A_5	A_6	A_7
2018	000 (14.56)	004 (16.22)	EMD-2660 (16.23)	003 (16.32)	002 (16.42)	001 (16.45)	EMD-8457 (17.43)
12650	001 (14.28)	002 (14.29)	004 (14.34)	003 (14.35)	000 (14.95)	EMD-8457 (15.15)	EMD-2660 (15.64)
26104	001 (15.13)	002 (15.16)	004 (15.18)	003 (15.29)	000 (16.50)	EMD-8457 (16.51)	EMD-2660 (17.14)
26138	003 (14.56)	004 (14.73)	001 (14.74)	002 (14.74)	000 (15.50)	EMD-8457 (15.83)	EMD-2660 (16.08)
36561	003 (14.62)	004 (14.80)	001 (14.84)	002 (14.88)	000 (15.40)	EMD-8457 (16.00)	EMD-2660 (16.09)

Note: A_i denotes the structure with the i th lowest value of d_{iKam} . The value of $\log(d_{iKam})$ is reported next to each structure in parenthesis. In each row, the entry shaded green indicates the ground truth structure.