

Cross-Validation of Data in SAXS and Cryo-EM

Bijan Afsari

Center for Imaging Science
Johns Hopkins University
Baltimore, Maryland 21218
Email: bijan@cis.jhu.edu

Jin Seob Kim

Department of Mechanical Engineering
Johns Hopkins University
Baltimore, Maryland 21218
Email: jkim115@jhu.edu

Gregory S. Chirikjian

Department of Mechanical Engineering
Johns Hopkins University
Baltimore, Maryland 21218
Email: grege@jhu.edu

Abstract—Cryo-Electron Microscopy (EM) and Small Angle X-ray Scattering (SAXS) are two different data acquisition modalities often used to glean information about the structure of large biomolecular complexes in their native states. A SAXS experiment is generally considered fast and easy but unveiling the structure at very low resolution, whereas a cryo-EM experiment needs more extensive preparation and post-acquisition computation to yield a 3D density map at higher resolution. In certain applications, one may need to verify if the data acquired in the SAXS and cryo-EM experiments correspond to the same structure (e.g., prior to reconstructing the 3D density map in EM). In this paper, a simple and fast method is proposed to verify the compatibility of the SAXS and EM experiments. The method is based on averaging the 2D correlation of EM images and the Abel transform of the SAXS data. The results are verified on simulations of conformational states of large biomolecular complexes.

I. INTRODUCTION

Cryo-Electron Microscopy (EM) [1] and Small Angle X-ray Scattering (SAXS) [2]–[4] are two popular -but very different- data acquisition modalities used to glean information about the structure of large biomolecular complexes in their native states. Both methods differ from crystallography in that no crystallization is needed (at the expense of lower resolution). The advantage is that crystallization in many cases is either a very lengthy process or may destroy biologically important features. A SAXS experiment is generally considered as fast and easy, but unveiling the structure at very low resolution, whereas a cryo-EM experiment needs more preparation and post-acquisition computation to yield a 3D density map at a higher resolution. The low resolution in SAXS can be roughly attributed the fact that, the SAXS data is the *spherical average* of the scattering pattern of the complex under study. In cryo-EM, similar to standard tomography, one obtains *numerous* 2D projections of a 3D complex, but in contrast to tomography the projections are at *random* (unknown) directions; moreover, the projections are extremely noisy. As a result, the 3D volume reconstruction in cryo-EM involves complicated post-processing and still the resolution is not very high.

In order to investigate the relationship between the function and the shape of a biomolecular complex, one may want to combine these two experimental data modalities (or fuse them). In such applications, one may first need to verify if the data acquired in the SAXS and cryo-EM experiments correspond to the same structure (e.g., prior to reconstructing the 3D density map). In this paper, we introduce a simple yet

effective method that enables fast verification of the compatibility of data collected in SAXS and Cryo-EM experiments. Roughly speaking, we relate the planar *correlations* of EM images to the SAXS data. To the best of our knowledge, our work is the first attempt in establishing such a relation and the derivations in Section III-A are new. The main benefit of our approach is that it enables the validation without the need for *aligning* and *classification* of the EM images or 3D reconstruction of the volume, which are complicated steps [1]. The *translation-invariance* property of the correlation function is the key enabling factor to achieve this. The combination and validation of SAXS and EM data has been appeared in the literature (e.g., [5], [6]). However, such methods are mostly based on image processing techniques or visual verification. Correlation functions in the context of EM data have been used before [7], but it became clear that for reconstruction of 3D density map the correlation function might not be adequate [8]. Here, we prove that the correlation functions of EM images after averaging can be related to the SAXS data via the Abel transform.

This paper is organized as follows: In Section II we review the mathematical modeling the EM and SAXS data. In Section III the relation between the SAXS and EM data is established and an algorithm for the validation is expressed. In Section IV we perform some simulations to support our approach, and Section V concludes the paper. The focus of this paper is deriving the basic mathematical methodology and a plausibility study with simulated data. Applications with actual SAXS and EM data will appear later.

II. THE EM AND SAXS DATA AND THEIR RELATION

A. The Cryo-EM Data

The reader is referred to [1] for detailed physical and mathematical analysis of cryo-EM modeling and reconstruction. Here, we proceed with a very basic mathematical model. We model the 3D atomic density of a large biomolecular complex of interest as a uniform density map $\chi(\mathbf{r})$, where $\chi : \mathbb{R}^3 \rightarrow \mathbb{R}_{\geq 0}$ and $\mathbf{r} = (x, y, z)^\top \in \mathbb{R}^3$ [1]. Although any uniform density function $\mathbb{R}^3 \rightarrow \mathbb{R}_{\geq 0}$ will work here, we specifically use $\chi(\mathbf{r})$ as the characteristic function, which is defined as [9], [10]

$$\chi(\mathbf{r}) \doteq \begin{cases} 1 & \text{if } \mathbf{r} \in B \\ 0 & \text{if } \mathbf{r} \notin B \end{cases}$$

where B denotes a biomolecular complex viewed as a solid body. With the characteristic function, a number of geometric quantities can be computed. For example, the volume of the complex body is computed as

$$V(B) = \int_{\mathbb{R}^3} \chi(\mathbf{r}) d\mathbf{r} = \int_B 1 d\mathbf{r}$$

where $d\mathbf{r} = dx dy dz$ is the usual integration measure for \mathbb{R}^3 .

In a cryo-EM experiment a frozen sample containing the complex is imaged. Inside the sample instances of the complex appear at *random positions* and *orientations*. This can be modeled by random rigid body motion $g = (R, \mathbf{t}) \in SE(3)$, where $R \in SO(3)$ is the rotation component and $\mathbf{t} \in \mathbb{R}^3$ is the translation component; $SE(3)$ and $SO(3)$ denote, respectively, the Lie groups of rigid body motions and rotations in \mathbb{R}^3 . A copy of the complex under random rotation and translation $g = (R, \mathbf{t}) \in SE(3)$ can be modeled as

$$\chi_g(\mathbf{r}) = \chi(g^{-1} \cdot \mathbf{r}), \quad (1)$$

where \cdot denotes the usual action of $SE(3)$ on \mathbb{R}^3 defined as $g \cdot \mathbf{r} = R\mathbf{r} + \mathbf{t}$. In the process of imaging, essentially every copy of the complex is imaged, and this is modeled by the projection along the z axis in a global frame:

$$(\chi_g(\mathbf{r}))^p = \chi_g^p(x, y) = \int_z \chi(g^{-1} \cdot \mathbf{r}) dz \quad (2)$$

Thus an EM image is the projection of a randomly translated and oriented copy of the complex. In reality, the EM images are highly noisy to the extent that signal-to-noise ratio (SNR) of 1/100 (i.e., the noise energy 100 times the signal energy) is quite common. Moreover, other effects such as the contrast transfer function of the microscope further deteriorate the images. Here, we ignore such effects. The standard reconstruction of the 3D volume $\chi(\mathbf{r})$ from the 2D noisy projections $\{\chi_g^p(x, y)\}_g$ is a complicated and lengthy process [1]. We show in Section III that certain information that relates to the SAXS experiment can be obtained from the EM images with relatively simple operations and low computational load.

B. The SAXS Data

Before progressing further, we have to mention that the SAXS and EM data are generated based on different physical principles and atomic properties. However, one expects that the information relevant to the geometry of a large complex would not be much affected by this difference. Thus we postulate that both SAXS and EM experiments yield information about the complex modeled by a uniform density χ .

Having this assumption in mind, we formulate the mathematics of SAXS data. The data collected in a SAXS experiment can be related to the (spatial) correlation function of $\chi(\mathbf{r})$ (also known as the Patterson function [4]), which is defined as

$$C_\chi(\mathbf{r}) = \chi(\mathbf{r}) * \chi(-\mathbf{r}) = \int_{\mathbb{R}^3} \chi(\mathbf{r}') \chi(\mathbf{r}' - \mathbf{r}) d\mathbf{r}', \quad (3)$$

where $*$ denotes the convolution operation. Thus the correlation function $C_\chi(\mathbf{r})$ is the convolution of $\chi(\mathbf{r})$ with its

reflection across the origin. The correlation function for a function defined on \mathbb{R} and \mathbb{R}^2 is defined similarly. In the frequency domain we have $\widehat{C}_\chi(\boldsymbol{\omega}) = |\widehat{\chi}(\boldsymbol{\omega})|^2$, where $\widehat{f}(\boldsymbol{\omega})$ denotes the Fourier transform of $f(\mathbf{r})$ and $\boldsymbol{\omega} \in \mathbb{R}^3$ is the spatial frequency vector. Let us write $\mathbf{r} = r\mathbf{u}$, where $\mathbf{u} \in \mathbb{S}^2$ (\mathbb{S}^2 being the unit sphere in \mathbb{R}^3) and $r = \|\mathbf{r}\| \in \mathbb{R}$. So we write $C_\chi(\mathbf{r}) = C_\chi(r\mathbf{u})$. Then the SAXS experiments gives a profile which (in spatial domain) can be expressed as (see [3] for details):

$$\gamma_\chi(r) = \int_{\mathbf{u} \in \mathbb{S}^2} C_\chi(r\mathbf{u}) d\mathbf{u}. \quad (4)$$

The meaning of this equation is that the SAXS data $\gamma_\chi(r)$ is the spherical average of the correlation function $C_\chi(\mathbf{r})$. Note that the pair distance distribution function $p_\chi(r)$, which is one of two key quantities in SAXS experiments, can be written as (see [4] for details)

$$p_\chi(r) = r^2 \gamma_\chi(r).$$

The source of equation (4) is that the copies of the complex $\chi(\mathbf{r})$ are randomly (uniformly) directed in the liquid sample. This is, in fact, related to our approach in relating the EM data and the SAXS data. Note that we make a distinction between *orientation*, which is coded by a rotation matrix $R \in SO(3)$, and *direction* which is coded by a direction vector $\mathbf{u} \in \mathbb{S}^2$.

III. RELATION BETWEEN SAXS AND EM DATA

A. Relating the SAXS and EM Data via the Abel Transform

Let us denote the translated version of $\chi(\mathbf{r})$ (i.e., $\chi(\mathbf{r} - \mathbf{r}_0)$) as $\chi_{\mathbf{r}_0}$, where \mathbf{r}_0 is the translation vector. We start by noting that $C_\chi(\mathbf{r}) = C_{\chi_{\mathbf{r}_0}}(\mathbf{r})$, i.e., the correlation function is invariant under translations. This holds both in the case of 1D or 2D correlations. Thus the planar (2D) correlation of an EM image $\chi_g^p(x, y)$ (see (2) for its definition) only depends on $R \in SO(3)$, the rotation part of $g \in SE(3)$. Therefore, we write

$$\begin{aligned} C_{\chi_g^p}(x, y) &= C_{\chi_R^p}(x, y) = \chi_R^p(x, y) * \chi_R^p(-x, -y) \\ &= \int_{x', y'} \chi_R^p(x', y') \chi_R^p(x' - x, y' - y) dx' dy' \end{aligned} \quad (5)$$

where χ_R^p is defined by setting $g = (R, \mathbf{0})$ in (2). We also note that projection operation commutes with convolution and correlation operations. This follows from the Fourier slice theorem [11]. Therefore, using the translation-invariance property and the commutativity property we can write:

$$(C_{\chi_g}(\mathbf{r}))^p = C_{\chi_R^p}(x, y) = \chi_R^p(x, y) * \chi_R^p(-x, -y), \quad (6)$$

where p denotes the projection operation as defined in (2). Next, we assume that the R component of g is *uniformly* distributed on $SO(3)$, i.e., the EM images are coming from uniformly oriented copies of the complex. This assumption is important in our derivations. However, we stress that it is known that, in practice, the uniformity assumption may be violated, as there are the so-called *preferred orientations* that most copies of the complex assume in the frozen sample [1, ch. 3]. The existence of preferred orientations is a quite complicated

phenomenon; thus here we retain the uniformity assumption and leave studying the effect of preferred orientations and the extent to which the uniformity assumption can be relaxed to a later work. Now, under the uniformity assumption, by averaging all the planar correlations across all orientations R we get the *circularly symmetric* function:

$$E_R\{C_{\chi_R^p}(x, y)\} = \gamma_{\chi^p}(\sqrt{x^2 + y^2}) = \gamma_{\chi^p}(r), \quad (7)$$

where, here, with some abuse of notation, $r = \sqrt{x^2 + y^2}$ and E_R denotes expectation (average) with respect to the random variable R . By the commutativity property we have

$$\gamma_{\chi^p}(r) = E_R\{C_{\chi_R^p}(x, y)\} = (E_R\{C_{\chi_R}(\mathbf{r})\})^p, \quad (8)$$

that is, $\gamma_{\chi^p}(r)$ is the average of the correlation of χ_R across all (random) orientations R of the complex. In the above, we also used the fact the mathematical expectation and projection operations commute.

The next step is to relate the above averaged planar correlation to the SAXS data $\gamma_\chi(r)$ in (4). Notice that $\gamma_\chi(r)$ is the average of C_{χ_R} across the coset $\mathbb{S}^2 = SO(3)/SO(2)$. But under the uniformity assumption $\gamma_\chi(r)$ is not different from the average of $C_{\chi_R}(\mathbf{r})$ over $SO(3)$, namely, $E_R\{C_{\chi_R}(\mathbf{r})\}$. Thus it follows that, under the uniformity assumption, the EM averaged correlation $\gamma_{\chi^p}(r)$ in (8) is equal to the projection of the SAXS data $\gamma_\chi(r)$, when viewed as a 3D *spherically symmetric* function. This projection can be expressed in terms of the Abel transform [12, Ch. 9]. The Abel transform of a one dimensional function $f : [0, \infty) \rightarrow \mathbb{R}$ is defined as

$$\text{Abel}(f)(r) \doteq 2 \int_r^\infty \frac{f(r')r'}{\sqrt{r'^2 - r^2}} dr'. \quad (9)$$

This relation is easy to prove using Pythagoras's theorem and a simple change of variable. From our discussion it follows that:

$$\gamma_{\chi^p}(r) = \text{Abel}(\gamma_\chi)(r) = 2 \int_r^\infty \frac{\gamma_\chi(r')r'}{\sqrt{r'^2 - r^2}} dr', \quad (10)$$

which means that the average of the correlations of EM images equals the Abel transform of the SAXS data. In practice, we expect this equality to hold up to a scale, due to the fact the sources for EM imaging and SAXS experiment have different amplitudes. Moreover, we are tacitly assuming that the SAXS and EM data are based on the same physical properties of the complex. In reality, however, the EM images are formed based on the scattering of electron beams by, primarily, the nuclei of the atoms, whereas the SAXS data is formed based on the diffraction of X-ray beams by electrons in the atoms.

B. Evaluating the Abel Transform: Accuracy Issues

Our approach requires evaluating the Abel transform of the SAXS data $\gamma_\chi(r)$. In the Abel transform, the integrand is singular at $r = r' \neq 0$. This can cause some error in evaluating the transform, especially when the SAXS data $\gamma_\chi(r)$ is known with only finite radial resolution or noise is high. Note that from (10) for $r = 0$, the singularity is removed, and $\text{Abel}(\gamma_\chi)(0) = 2 \int_0^\infty \gamma_\chi(r') dr'$.

In general, there are two approaches in evaluating singular integrals: *eliminating the singularity* with a change of variable and *ignoring the singularity* [13]. The first method, essentially requires knowing the integrand with infinite accuracy, which is impractical in our application. However, the method of ignoring-the-singularity is practical, and as its name suggests requires no special provision. The caveat, however, is that the presence of a singularity adversely affects the rate of convergence in terms of the integration step-size [14]. We elaborate on this issue (following [14]).

Let $f : (0, 1] \rightarrow \mathbb{R}$ be continuous on $(0, 1]$. Assume that f can be written as $f(r) = r^\alpha g(r)$ with $\alpha > -1$, $g(r)$ being a function whose derivative on $[0, 1]$ exists, is continuous and integrable in absolute value, i.e., $\int_0^1 |\frac{dg}{dr}| dr < \infty$. Notice that under these conditions the integral $\int_0^1 f(r) dr$ exists. Such a singularity is called an *algebraic singularity* [14]. In the case of the Abel transform we have $\alpha = -\frac{1}{2}$, as for most densities $C_\chi(\mathbf{r})$ and, hence, $\gamma_\chi(r)$ are smooth. Now, let us take the simple rectangle integration rule $I_n(f) = \frac{1}{n} \sum_{i=1}^n f(\frac{i}{n})$, the step-size being $h = \frac{1}{n}$. Note that, in this approximation, we are ignoring the singularity at $r = 0$ (e.g., instead of writing $\frac{1}{n} \sum_{i=0}^{n-1} f(\frac{i}{n})$, which obviously cannot be calculated). The approximation error is $E_n = |\int_0^1 f(r) dr - I_n(f)|$. It can be shown that $E_n \xrightarrow{n \rightarrow \infty} 0$ with the rate $(\frac{1}{n})^{1+\alpha}$ [14]. This means that for the Abel transform the rate of convergence will be $\frac{1}{\sqrt{n}}$. It is interesting also to mention that, even more complicated integration methods such as the midpoint, trapezoid, or quadrature rules have the same rate of convergence [14], although the actual error for a given step-size might be different. To put this rate of convergence in perspective, note that if f is continuous on $[0, 1]$ (no singularity, $\alpha = 0$), then the rate would be $\frac{1}{n}$. Also recall that in the case of a function with smooth second order derivative on $[0, 1]$, the convergence rate of the trapezoid and the midpoint rules improve to $\frac{1}{n^2}$. Finally, we stress that the approximation error always exists (due to the discretization in computing the integral), however, in this case, due to the singularity it is worse compared with the case of a nonsingular integrand.

Next, we perform a numerical experiment in the case of the SAXS data of a spherical object. Consider a spherical object of radius R and with uniform density, i.e.,

$$\chi_S(\mathbf{r}) = \chi_S(r) = \begin{cases} 1 & \text{if } \|\mathbf{r}\| (= r) \leq R, \\ 0 & \text{otherwise.} \end{cases} \quad (11)$$

For such a body the correlation function is given by

$$\begin{aligned} C_{\chi_S}(\mathbf{r}) &= C_{\chi_S}(r) \\ &= \begin{cases} \frac{1}{12} \pi (4R + r)(2R - r)^2 & \text{if } \|\mathbf{r}\| (= r) \leq 2R, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \quad (12)$$

This formula is derived using the formula for the intersection volume of two spheres. Due the spherical symmetry of C_{χ_S} , we have $\gamma_{\chi_S}(r) = C_{\chi_S}(r)$. We use the rectangle integration method with various step-sizes to evaluate the Abel transform of $C_{\chi_S}(r)$ with $R = 100$. Figure 1 shows the results for step-sizes $h = 1, 1/2, 1/4, 1/8, 1/16, 1/32, 1/64$. Also the exact

curve (calculated using Matlab's adaptive step-size integration method) is shown. Clearly the *slow* rate of improvement in the accuracy matches the theory explained above.

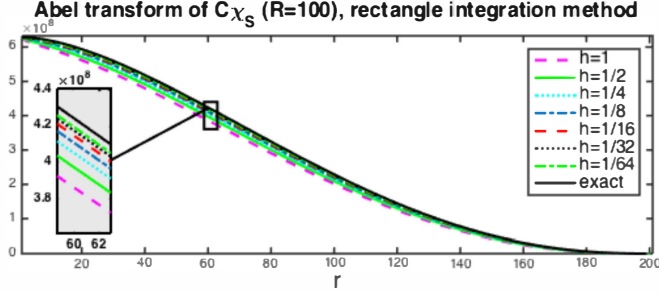


Fig. 1: The effect of integration step-size in evaluating the Abel transform of $C_{\chi_S}(r)$ (12).

In our applications, we cannot have very small step-sizes because usually the SAXS data is available on a finite grid and additionally the values at the grid nodes can be noisy. Notice that an interpretation of the above error analysis is that if we have $\gamma_\chi(r)$ at a resolution of $\frac{1}{n}$ we get its Abel transform with an error of order $\frac{1}{\sqrt{n}}$. However, this also means that any error in $\gamma_\chi(r)$ may be amplified in $\text{Abel}(\gamma_\chi)(r)$. Thus the net effect the singularity is that we have to accept some moderate error in evaluating the Abel transform of $\gamma(r)$ and hence in the matching between the EM profile $\gamma_{\chi^p}(r)$ and the SAXS profile $\text{Abel}(\gamma_\chi)(r)$ (equation (10)). This may hinder our ability to distinguish between complexes that have very similar SAXS data $\gamma_\chi(r)$.

C. Removing the Noise in EM Planar Correlations

The noise in an EM image can be extremely strong and we need to devise methods to mitigate its effect. Let us denote the noisy version of the image $\chi_R^p(x, y)$ by $\tilde{\chi}_R^p(x, y)$. We assume an additive noise model (which is a reasonable assumption in the EM imaging mechanism). Thus we can write

$$\tilde{\chi}_R^p(x, y) = \chi_R^p(x, y) + \nu(x, y), \quad (13)$$

where $\nu(x, y)$ is the noise. If we assume that the noise is spatially white with zero mean and variance σ_ν^2 , and further that it is uncorrelated with the image χ_R^p , then we can see that

$$C_{\tilde{\chi}_R^p}(x, y) = \begin{cases} C_{\chi_R^p}(0, 0) + \sigma_\nu^2 & \text{if } x = y = 0, \\ C_{\chi_R^p}(x, y) & \text{otherwise.} \end{cases} \quad (14)$$

We have derived this equation under the ergodicity assumption meaning that statistical and spatial averages are equal, which in practice holds to a good extent. Note that (14) means that if we know or can estimate the noise variance σ_ν^2 , then we can remove the effect of noise from the correlations $C_{\tilde{\chi}_R^p}$. In practice, we may be able to estimate σ_ν^2 from parts of the image, where most likely the actual projection image of the complex is not present (e.g., the corners of the image).

IV. EXPERIMENTS

An important problem in structural biology is to determine the conformational states of a large biomolecular complex (e.g., open vs. closed). Both SAXS and EM experiments can be performed for this purpose, and in some cases one may want to verify if the samples in the SAXS and EM experiments contain the complex at the same conformational state. We perform two simulated experiments on a synthetic model and one on a simulated complex with ligand-binding domains.

A. Conformational States: The Two-Body Model

We simulate different conformational states of a large biomolecular complex by modeling the complex with two solid ellipsoids of the same size and varying the (unknown) angle θ between the principal axes of the two ellipsoids. We denote the complex at conformation associated with angle θ by χ_θ . The goal is to decide whether the EM and SAXS experiments were conducted on the complex at the same conformation in the respective samples, and also to examine the angular resolution that can be achieved. Before proceeding further, we stress that although our mathematical models were in spatially continuous setting, in the experiments we move to a discrete spatial setting.

The data is generated as follows. The axes lengths of each ellipsoid are $a = 21.3$, $b = c = 6.7\text{\AA}$. A volume V of size $N \times N \times N$ ($N = 129$) voxels is generated that contains the complex χ_θ with conformational angle θ . The complex χ_θ itself is modeled by a solid (uniform) body as explained earlier in Section II. To simulate the EM experiment, the complex is rotated (inside V) around the center of V by a randomly (uniformly) generated rotation matrix R . Then a projection along the z -axis is computed, which results in an $N \times N$ image. The top two images in Figure 2 show random projections of conformational angles of $\theta = 30^\circ$ and $\theta = 45^\circ$. Next samples of zero mean white Gaussian noise of variance σ_ν^2 is added to the images as shown in the second row of Figure 2. In the figure, the signal-to-noise ratio (SNR) is $\frac{1}{10}$. For each image the SNR is defined as the average energy of the projection of the complex to the energy of the noise in that image; and σ_ν^2 is determined accordingly (the energy of noise is $N\sigma_\nu^2$). In this experiment $F = 100$ random rotation matrices are uniformly generated on $SO(3)$ and corresponding to each rotation matrix $K = 10$ noisy samples are generated. Thus a total of 1000 noisy images are generated. Next, the spatial correlation of each planar image is found. Notice that this results in a correlation image of size $2N - 1$. To simulate the noise-removal step described in Section III-C, a patch of size $W \times W$ pixels, where $W = 35$, at a corner of each image is selected and the variance of noise estimated in that patch. Then the variance of the noise is subtracted from the correlation at $(0, 0)$ (see (14)); and all such denoised correlations are averaged. The result is an almost circularly symmetric image. However, we further symmetrize it by circular averaging to get an approximation of $\gamma_{\chi_\theta}^p(r)$. The circular averaging is performed at rather high angular resolution. Since the images are discrete spatially, for

circular averaging one needs to perform spatial interpolation. We also used interpolation to increase the resolution in the radial variable r to $H = 0.5\text{\AA}$. This improves visualization and smoothness of the results. To simulate the SAXS data generation we simply find the 3D correlation of the volume V and perform a spherical averaging (combined with spatial interpolation) to get an approximation of $\gamma_{\chi_\theta}(r)$ (see (4)). We obtain $\gamma_{\chi_\theta}(r)$ with resolution of $H = 0.5\text{\AA}$. Finally, to estimate the SAXS profile $\text{Abel}(\gamma_{\chi_\theta})(r)$, we use the rectangle integration rule. The bottom left plot in Figure 2 shows the graphs of the estimated EM profiles $\gamma_{\chi_\theta}^p(r)$ and SAXS profiles $\text{Abel}(\gamma_{\chi_\theta})(r)$ for conformational angles of $\theta = 30^\circ$ and $\theta = 45^\circ$. It can be seen that the corresponding SAXS and EM profiles are similar for moderate to large values of r . However, for small values of r the SAXS profiles for both 30° and 45° are becoming closer to each other and deviate from the EM profiles. The fact that the SAXS profiles for small r are very similar is expected, and reflects the fact that SAXS data can give low-resolution information about a structure. In fact, theoretically $\gamma_{\chi_{30^\circ}}(0) = \gamma_{\chi_{45^\circ}}(0)$ and $\gamma_{\chi_{30^\circ}}(r) \approx \gamma_{\chi_{45^\circ}}(r)$ for small r since in both conformations the total volume is the same and the individual bodies are the same. Thus effectively $\text{Abel}(\gamma_{\chi_{30^\circ}})(r) \approx \text{Abel}(\gamma_{\chi_{45^\circ}})(r)$ for small r . The observed discrepancy between the EM and SAXS profiles can be attributed to both noise and the integration error in evaluating the Abel transform. In the bottom right plot of Figure 2 we have plotted the normalized version of the EM and SAXS profiles, where each profile integrates to 1. Specifically, we define the normalized profiles as

$$\gamma_{\chi_\theta}^{p,n}(r) = \frac{\gamma_{\chi_\theta}^p(r)}{\int_0^\infty \gamma_{\chi_\theta}^p(r) dr}, \quad \text{Abel}^n(\gamma_\chi) = \frac{\text{Abel}(\gamma_\chi)(r)}{\int_0^\infty \text{Abel}(\gamma_\chi)(r) dr}. \quad (15)$$

In practice this normalization is very useful, primarily because the EM and SAXS profiles are, at best, proportional to each other and may differ in their scaling significantly since they come from very different experimental sources.

Next we perform a similar experiment to estimate the smallest resolvable angle. For that we define a notion of distance between the SAXS and EM profiles. Based on the forgoing discussion, the difference at small r may be very large but with very little relation to the conformational angles. Thus we choose a value r_0 and we compare the profiles for $r \geq r_0$. We choose a simple L^2 -based distance as:

$$d(\gamma_{\chi_{\theta_1}}^{p,n}, \text{Abel}^n(\gamma_{\chi_{\theta_2}})) = \sqrt{\int_{r_0}^\infty |\gamma_{\chi_{\theta_1}}^{p,n}(r) - \text{Abel}^n(\gamma_{\chi_{\theta_2}})(r)|^2 dr}. \quad (16)$$

Of course, in practice a discretized version of this will be used. In this experiment, we chose the SNR much lower at $\frac{1}{100}$. At this extremely low SNR, no vestige of the (projection of the) complex is visible. We generate conformations with angles $\theta \in \{30^\circ, 35^\circ, \dots, 70^\circ\}$. Table I shows the entries of a distance matrix, where each entry is the distance defined in (16) between SAXS and EM profiles at respective indicated angles. All the distances are scaled by a fixed number for

ease of presentation. We have chosen $r_0 \approx 20\text{\AA}$ in (16). As it can be seen, in some cases (underlined) the EM and SAXS profiles of angles that differ by 5° can be closer than those of the actual (correct) angles. However, for differences larger than or equal to 10° such a confusion is not observed. Our experiments show that in a larger number of trials confusion for angles larger than equal to 10° is very rare. Thus, the angular resolution is roughly $\pm 5^\circ$ with very high certainty.

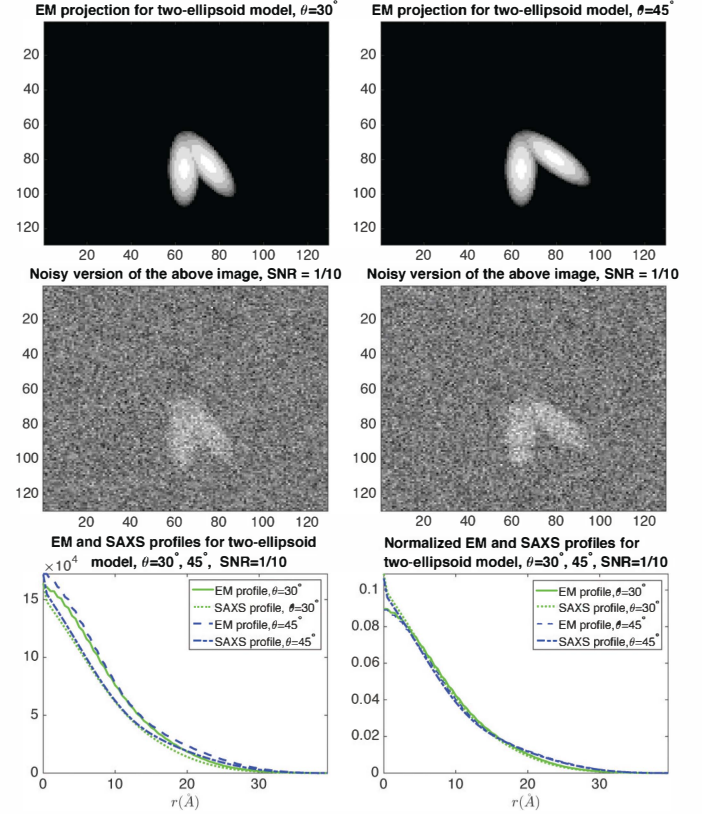


Fig. 2: The top row shows random EM projections of the two body complex χ_θ at two conformational angles $\theta = 30^\circ$ and $\theta = 45^\circ$ (see Section IV-A for more details). The middle row shows the same images contaminated by noise (SNR = 1/10). The bottom graphs show the estimated corresponding SAXS and EM profiles (the left graph shows un-normalized profiles and the right one shows the normalized profiles—see (15)).

B. Glutamate Receptor Ligand-Binding Domain Conformations

Here, we perform a rather similar experiment while we generate the data from Protein Data Bank (PDB) [15]. Specifically, we consider the ligand-binding domain (LBD) of glutamate receptor, which is known to play a crucial role in human brain activities such as memory formation and learning process [16]. We consider three conformational states of LBD: ‘apo’ or unliganded state (PDB entry: 1FTO), antagonist-bound state (PDB entry: 1FTL), and partial agonist-bound state (PDB entry: 1FTK). It is experimentally shown that the conformation of ‘apo’ state is more similar to the antagonist-bound state than

SAXS \ EM	30°	35°	40°	45°	50°	55°	60°	65°	70°
30°	<u>0.29</u>	0.52	<u>0.97</u>	0.76	<u>1.28</u>	<u>1.39</u>	1.66	<u>1.91</u>	<u>1.88</u>
35°	<u>0.21</u>	<u>0.28</u>	0.66	<u>0.49</u>	<u>0.99</u>	1.11	<u>1.38</u>	1.66	1.63
40°	<u>0.48</u>	<u>0.29</u>	<u>0.34</u>	<u>0.29</u>	<u>0.68</u>	0.83	<u>1.08</u>	1.40	1.37
45°	<u>0.82</u>	0.56	<u>0.19</u>	<u>0.37</u>	<u>0.37</u>	0.55	0.77	1.12	<u>1.09</u>
50°	1.11	0.83	0.43	0.61	<u>0.21</u>	0.34	<u>0.49</u>	0.85	0.83
55°	1.37	1.08	0.71	0.86	0.35	<u>0.30</u>	0.26	0.61	0.60
60°	<u>1.58</u>	<u>1.28</u>	<u>0.94</u>	1.07	0.57	0.42	<u>0.20</u>	0.41	0.41
65°	1.75	1.46	1.15	1.26	0.77	<u>0.58</u>	0.34	<u>0.27</u>	<u>0.28</u>
70°	<u>1.88</u>	<u>1.59</u>	1.31	1.41	<u>0.94</u>	0.74	0.50	<u>0.25</u>	<u>0.26</u>

TABLE I: A distance matrix for the two-body model experiment in Section IV-A. Each entry shows the distance (16) between the SAXS and EM profiles at the corresponding body-angles. All the distance values are normalized by a common factor for ease of presentation. The underlined entries show the cases where the distances between the SAXS and EM profiles of two different angles give a distance smaller than the correct angle.

to the agonist-bound one [17]. As we will see shortly, the same conclusion can be drawn from our simulation results. In our simulation, all atom coordinates were considered.

We simulate the effect of each atom by a $3\text{\AA} \times 3\text{\AA} \times 3\text{\AA}$ cube of constant intensity centered at each atomic coordinate present in the corresponding PDB file of each complex. The rest of simulation is as in the previous example with volume size $N = 129$, $\text{SNR}=1/100$, number of random rotations $F = 100$, number of images per orientation $K = 10$, and patch size $W = 35$ for denoising. In the first experiment we compare the SAXS and EM profiles of complexes 1FTL and 1FTK. The top panels in Figure 3 show the atomic configuration of these two complexes. The left bottom graph shows the unnormalized SAXS and EM profiles for both. The large jump at $r = 0$ is due to error in estimating the noise variance. As it can be seen the profiles of distinct conformations are easily separable despite the fact that the EM and SAXS of each complex differ slightly (due to noise and the singularity effect). Here, however, this difference for smaller values of r is significant, presumably because the total volume of 1FTL and 1FTK is substantially different. In the next experiment depicted in Figure 4 we compare the profiles of 1FTO and 1FTL. As it can be seen the two conformations are quite similar and their corresponding profiles also become indistinguishable, which conforms with the earlier mentioned fact that these two conformational states are close. Note that in the current simulations, we used uniformly distributed random rotations. In reality, however, it might be the case that there are preferred orientations of the molecules in the sample grid [1, ch. 3]. In order to test that, we tried non-uniform rotations, such as the case where there are finite fixed number of rotational directions. Our preliminary and limited investigation shows that results such as those in Fig. 3 and 4 were not influenced significantly by non-uniform rotations. In our future work, we will try to understand how these results will be affected due to the presence of preferred orientations and the violation of the non-uniformity assumption (see III-A and [1, ch. 3]), together with applying our method to larger and more complex macromolecules.

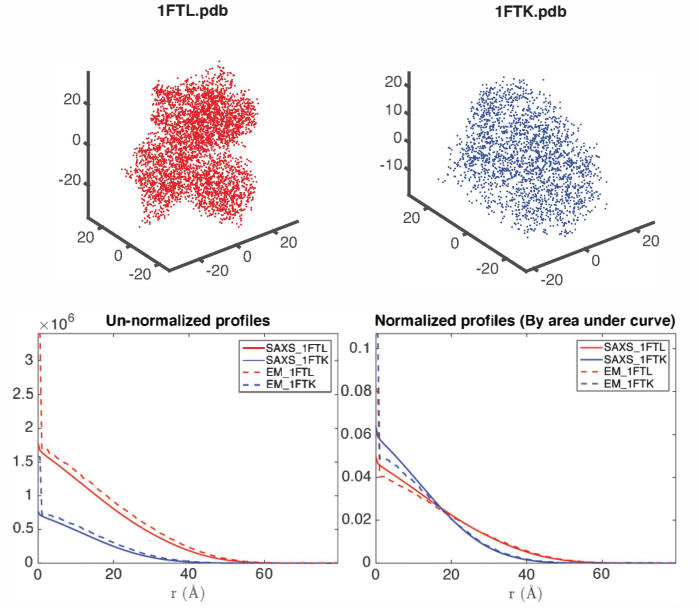


Fig. 3: Comparing SAXS and EM profiles ($\text{Abel}(\gamma_\chi)(r)$ and $\gamma_{\chi^p}(r)$, respectively) and their normalized versions for 1FTL and 1FTK, see Section IV-B for details.

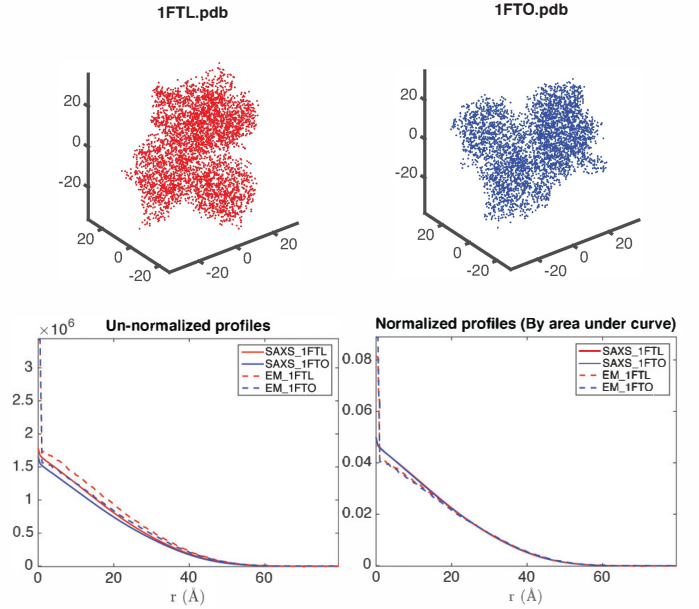


Fig. 4: Comparing SAXS and EM profiles ($\text{Abel}(\gamma_\chi)(r)$ and $\gamma_{\chi^p}(r)$, respectively) and their normalized versions for 1FTL and 1FTO, see Section IV-B for details.

V. CONCLUSIONS

We have presented a simple and fast approach to check the compatibility of SAXS and cryo-EM data. By noting that the SAXS data is indeed derived from the correlation function (i.e., the Patterson function, which is the self-convolution of the density function), we proposed a method based on the correlation functions both in SAXS and EM. In this paper,

we derived the basic mathematical methodology. In particular, we used the fact that averaging the correlation functions of EM images (prior to 3D reconstruction) is equivalent to the Abel transform of the SAXS data. Then we investigated the plausibility with simulated data and the actual PDB data. We also discussed limitations due to the singularity in the Abel transform in SAXS and the noise effect in EM images. An important question is the effect of preferred orientations in our results and simulations, which is the subject of our future research. The current study is a first step toward the fusion of two different experimental modalities, which in turn will enhance the understanding of the relationship between the function and the shape of large bimolecular complexes.

ACKNOWLEDGMENT

This work has been supported by the National Institute of General Medical Sciences of the NIH under award number R01GM113240. The authors thank the anonymous reviewers for insightful comments.

REFERENCES

- [1] J. Frank, "Three-dimensional electron microscopy of macromolecular assemblies: Visualization of biological molecules in their native," 2006.
- [2] D. I. Svergun and M. H. Koch, "Small-angle scattering studies of biological macromolecules in solution," *Reports on Progress in Physics*, vol. 66, no. 10, p. 1735, 2003.
- [3] L. Feigin, D. I. Svergun, and G. W. Taylor, *Structure analysis by small-angle X-ray and neutron scattering*. Springer, 1987.
- [4] D. I. Svergun, M. H. Koch, P. A. Timmins, and R. P. May, *Small Angle X-Ray and Neutron Scattering from Solutions of Biological Macromolecules*. Oxford University Press, 2013.
- [5] B. Vestergaard, S. Sanyal, M. Roessle, L. Mora, R. H. Buckingham, J. S. Kastrup, M. Gajhede, D. I. Svergun, and M. n. Ehrenberg, "The SAXS solution structure of RF1 differs from its crystal structure and is similar to its ribosome bound cryo-EM structure," *Molecular Cell*, vol. 20, no. 6, pp. 929–938, 2005.
- [6] P. Bron, E. Giudice, J.-P. Rolland, R. M. Buey, P. Barbier, J. F. Díaz, V. Peyrot, D. Thomas, and C. Garnier, "Apo-Hsp90 coexists in two open conformational states in solution," *Biology of the cell / under the auspices of the European Cell Biology Organization*, vol. 100, no. 7, pp. 413–425, 2008.
- [7] M. Schatz and M. Van Heel, "Invariant classification of molecular views in electron micrographs," *Ultramicroscopy*, vol. 32, no. 3, pp. 255–264, 1990.
- [8] M. van Heel, M. Schatz, and E. Orlova, "Correlation functions revisited," *Ultramicroscopy*, vol. 46, no. 1, pp. 307–316, 1992.
- [9] H. Dong and G. S. Chirikjian, "A computational model for data acquisition in SAXS," in *BDB '14 Proceeding of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics*, 2014, pp. 695–702.
- [10] H. Dong, J. S. Kim, and G. S. Chirikjian, "Computational analysis of SAXS data acquisition," *J. Comput. Biol.*, vol. 22, pp. 787–805, 2015.
- [11] F. Natterer, *The mathematics of computerized tomography*. Siam, 1986, vol. 32.
- [12] R. Bracewell, *Fourier analysis and imaging*. Springer Science & Business Media, 2003.
- [13] P. J. Davis and P. Rabinowitz, *Methods of numerical integration*. Academic Press, 1984.
- [14] M. El-Tom, "On ignoring the singularity in approximate integration," *SIAM Journal on Numerical Analysis*, vol. 8, no. 2, pp. 412–424, 1971.
- [15] H. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. Bhat, H. Weissig, I. Shindyalov, and P. Bourne, "The Protein Data Bank," *Nucleic Acids Res.*, vol. 28, no. 1, pp. 235–242, 2000.
- [16] R. Dingleline, K. Borges, D. Bowie, and S. F. Traynelis, "The glutamate receptor ion channels," *Pharmacol. Rev.*, vol. 51, pp. 7–61, 1999.
- [17] D. R. Madden, N. Armstrong, D. Svergun, J. Pérez, and P. Vachette, "Solution X-ray scattering evidence for agonist- and antagonist-induced modulation of cleft closure in a glutamate receptor ligand-binding domain," *J. Biol. Chem.*, vol. 280, pp. 23 637–23 642, 2005.