Size Refocusing on Small-Angle X-ray Scattering of Polydisperse Nanoparticles for Shape Determination

Authors

Siyu Wu[a], Xiaobing Zuo[b] and Yugang Sun[a]*

[a]Department of Chemistry, Temple University, 1901 North 13th Street, Philadelphia, Pennsylvania, 19122, United States

[b]X-ray Science Division, Argonne National Laboratory, 9700 South Cass Avenue, Bldg 433-E002, Lemont, IL, 60439, USA

Correspondence email: ygsun@temple.edu

Funding information National Science Foundation (award No. 2002960); U.S. Department of Energy, Office of Science (contract No. DE-AC02-06CH11357).

Synopsis A penalized iterative regression method using a gradient descent algorithm with regularization is presented to eliminate size-distribution-induced smearing effects in small-angle X-ray scattering (SAXS) profiles, enabling to retrieve the characteristic Bessel-type SAXS oscillations (including both peaks and valleys) of representative nanoparticles with varying shapes.

Abstract Small-angle X-ray scattering (SAXS) that records reciprocal-space signals with characteristic Bessel-type oscillations is a powerful technique for studying nanoparticles. However, the size polydispersity (or size distribution) of nanoparticles in ensemble samples smears the oscillational peaks and valleys in SAXS profiles, making it difficult to extract the accurate real-space information (e.g., three-dimensional geometry) of nanoparticles from the SAXS profiles. In this work, a method capable of eliminating the size-distribution-induced smearing effect from SAXS profiles by taking the known size distribution function into consideration has been developed. The method employs a penalized iterative regression to fit the pair distance distribution function (PDDF) derived from a SAXS profile, recovering the representative PDDF of the nanoparticles. The method has been evaluated with a series of nanoparticle systems with various shapes and size distributions, showing their PDDF profiles with high fidelity to the reference ideal PDDF profiles. Inverse Fourier transform of the recovered PDDF profiles gives the SAXS profiles presenting the characteristic Bessel-type oscillations, enabling reconstruction of the representative three-dimensional geometry of nanoparticles. This method will help use SAXS to image the synthesized colloidal nanoparticles where size polydispersity is inevitable.

Keywords: Small-angle X-ray scattering, polydispersity, pair distance distribution function, penalized iterative regression method.

1. Introduction

Small-angle scattering is a structure-probing technique that collects and analyzes scattered signals in the reciprocal space.(Li *et al.*, 2016) Unlike real-space techniques that directly capture microscopic images, translating the reciprocal-space scattering signals into real-space images requires complicated data processing, which is challenging and computationally expensive. As computational science rapidly advances, the complexity of data processing becomes an acceptable trade-off for easy experiment set-up, high-speed data collection, and high statistical representativeness associated with measuring many particles as an ensemble.(Kikhney & Svergun, 2015) For example, synchrotron small-angle X-ray scattering (SAXS) has been widely applied in the structural analysis of biomolecules. A number of bio-SAXS software packages have been developed in the past three decades for analyzing SAXS patterns, ranging from simple data-reduction (such as 2D image to 1D plot reduction, data averaging, and background subtraction) to complex modeling and fitting.(Ilavsky, 2012; Ilavsky & Jemian, 2009) In particular, developing 3D-shape reconstruction programs that aim to rebuild the geometry of particles has achieved high fidelity in recovering the shape of biomolecules from their SAXS profiles. Typical programs include dummy atom bead filling modeling (DAMMIN, DAMMIF)(Petoukhov *et al.*, 2007), iterative electron density retrieving (DENSS)(Grant, 2018), and deep learning methods (decodeSAXS)(He *et al.*, 2020).

SAXS is also considered a powerful tool for studying synthesized nanoparticles. Although biomolecules that have a well-defined chemical formula and folding pattern usually exhibit monodispersed size and geometric conformation in a solution, the synthesized colloidal nanoparticles exhibit (sometimes broad) distributions of size and shape.(Borchert *et al.*, 2005) Intensive efforts have been made to narrow down the size distribution and achieve shape uniformity by rigorously controlling the synthesis conditions.(Mantzaris, 2005) The dispersions with a size distribution of less than 15% are usually considered “monodisperse” nanoparticles. Such a small polydispersity can still cause problems for accurate SAXS data analysis. The theoretically sharp oscillations of SAXS signals of individual nanoparticles of different sizes overlap, resulting in a smeared SAXS profile without pronounced peaks and valleys.

The size distribution of an ensemble of particles can be retrieved from the SAXS profile when the particles exhibit the same simple geometry, for example, a mathematically ideal sphere. Many SAXS data analysis programs, including GNOM(Semenyuk & Svergun, 1991), SAXSfit(Ingham *et al.*, 2009), Irena(Ilavsky & Jemian, 2009), SASView(Doucet *et al.*, 2021) and FFSAS(Leng *et al.*, 2022) can extract the size distribution if the particle geometry is known. For example, Polte and co-workers used

in-situ time-resolved SAXS to study the growth mechanism of colloidal gold nanoparticles.(Polte *et al.*, 2010) A Schulz-Zimm distribution function fit the size distribution of gold nanoparticles while assuming their geometry to be a rigid sphere. Even in a more challenging situation where the nanoparticles are multimodally distributed (i.e., exhibiting two or more peaks in the size distribution statistics), the fitting can still provide high fidelity in recovering the size distribution function.(Thünemann *et al.*, 2009) The successful application of the size-distribution fitting algorithm can be attributed to the availability of empirical and theoretical models that describe the distribution functions, reducing the fitting process to only a few parameters such as the mean size and the standard deviation. However, when the geometry of the nanoparticles is the focus, the size distribution frustrates the nanoparticle geometry recovery as the oscillational SAXS signals for identifying the geometry are strongly smeared. The application of advanced shape-reconstruction programs in nanoparticle study is only limited to the highly monodisperse nanoparticles with <5% size distribution.(Sun *et al.*, 2017) With low monodispersity of nanoparticles, the geometry determined from the direct reconstruction programs using the smeared SAXS data is often misleading since the programs tend to take for granted that the smearing effect results from asymmetric geometry.(Ozerin *et al.*, 2006) Therefore, there is a demand for developing a method to deal with the size polydispersity of nanoparticles in order to achieve accurate shape from SAXS data.(Mittelbach & Glatter, 1998)

The problem of refocusing the SAXS profile of a polydisperse system can be analogous to the problem of deblurring a blurred image, for which various algorithms have been developed.(Bertero *et al.*, 2009) An image-deblurring problem can be simplified as solving the inverse problem of $f' = H \times f + n$, in which f' and f are the blurred image to be processed and the real image to be recovered, respectively. H is the blurring operator that indicates how the pixels in the real image diffused into the nearby ones, and n is the random noise added to the image. A typical deblurring procedure requires prior knowledge of H, which, in the SAXS size-refocusing problem, is the information about the size distribution. The size distribution of a nanoparticle system can be readily estimated using techniques such as electron microscopy and dynamic light scattering with distribution models involving only two parameters (mean size and standard deviation). The difference is that a blurred image is taken in the real space while a SAXS pattern is taken in the reciprocal space, which requires a proper transformation of the SAXS smearing problem to a real-space problem. In this work, we introduce a method to eliminate or reduce the size-distribution smearing effect in SAXS profile of a nanoparticle system with the same geometric shape. A penalized iterative regression method is used to fit the Fourier-transformed real-space SAXS profile with an estimated size-distribution function. A series of theoretical SAXS profiles of nanoparticle systems with different size distributions for various shapes have been generated as targets to train the method. The output of the estimated pair distance distribution functions (PDDF) of unisized model nanoparticles shows high reconstruction quality and fidelity to the idealized reference PDDF

profiles. The calculation relies on the pre-determined size distribution function but is robust even when the size distribution function is not accurately determined.

2. Methods

2.1. SAXS data calculation and processing procedure

SAXS intensity profiles of the perfect sphere particles were calculated using SASView, an open-source program designed for small-angle scattering modeling and data analysis. PDDF profiles of structures with various shapes of a given size were calculated using a self-developed Python script. In the typical calculation, a given structure was first approximated using a voxel model generated by filling a 64×64×64 grid. The autocorrelation of the filled grid was then calculated, resulting in a 127×127×127 matrix where each element corresponded to a pair distance ranging from 0 to $\sqrt{3}$ for a unit-size space. The discretized PDDF profile was obtained by grouping the elements in the autocorrelation matrix by their corresponding pair distances and calculating the sum. The PDDF profile could also add a size distribution using the forward Fourier transformation. Conversion from SAXS intensity profiles to PDDF profiles was conducted by GNOM, a program in the ATSAS package that uses indirect inverse Fourier transform and penalization methods.(Svergun, 1992) Conversion from PDDF profiles to SAXS intensity profiles was conducted by a self-developed Python script (pdf2iq). Cubic interpolation was performed to the discretized PDDF profiles before the conversion. 3D shape reconstruction was conducted by DAMMIF, an *ab initio* shape determination program that uses a single-phase dummy atom model, also available in the ATSAS package.(Petoukhov *et al.*, 2007) Unless specified otherwise, default modes and parameters were used in the mentioned program.

The size-refocusing fitting was performed with a self-developed Python program (SharPy). At the current stage of the development, SharPy can read the PDDF profiles generated by GNOM (files with *.out* extension) from SAXS data of a particle ensemble. The program reads the smeared PDDF profiles of the particle ensemble $P_{ens}(\boldsymbol{r})$, the type of the size distribution (default: log-normal), and the estimated mean size μ and standard deviation σ of the size distribution function. The objective of the program is to find target PDDF profile of the model particle, $\hat{P}_0$, that gives the lowest goal function. The search of $\hat{P}_0$ is conducted in a binned box with a dimension ($D_{\max}$) that is typically larger than the mean size μ of the particle. The oversized $D_{\max}$ ensures that the resultant PDDF curve can be included completely. The search is based on the Broyden–Fletcher–Goldfarb–Shanno algorithm (BFGS), a gradient descent algorithm that is called out using the *scipy.optimize* module in the scientific computation package SciPy.(Virtanen *et al.*, 2020) Optional functions are available to users to improve the performance of the computation potentially. For example, users can input an initial guess of $\hat{P}_0(\boldsymbol{r}/D_{\max})$ if they have *a priori* knowledge of the expected result. The default initial guess of $\hat{P}_0(\boldsymbol{r}/D_{\max})$ is a slightly randomized (e.g., 15%) Gaussian distribution close to the PDDF of a spherical

particle. In default mode, 51 points on the $\hat{P}_0(\boldsymbol{r}/D_{\max})$ need to be optimized, in which the first point (r=0) and the last point (r=D_{max}) are forced to be 0. After each iteration, the intensity of $\hat{P}_0(\boldsymbol{r}/D_{\max})$ is normalized to 1 by the maximum value. A typical optimization takes approximately ~100 iterations and costs approximately a few minutes on a typical commercial laptop. Future development of the program will include an interactive, user-friendly interface and additional functions that may be helpful for the SAXS analysis. All scripts and programs are available at https://github.com/wsy94/SharPy .

Current data processing flow includes: (1) computing the PDDF from the SAXS data of polydisperse nanoparticle sample using program GNOM and generating the GNOM output that contains the PDDF data; (2) program SharPy taking the GNOM output and size distribution parameters and generating the PDDF of the model particle; (3) program pdf2iq converting the PDDF of the model particle to the SAXS profile for further analysis.

2.2. SAXS of synthesized Ag nanoparticles

2.2.1. Chemicals

Silver nitrate ($AgNO_3$, 99.85%) was purchased from ACROS organics. Oleylamine ($C_{18}H_{35}NH_2$, 70%) was purchased from Sigma Aldrich. Acetone and hexane were purchased from Fisher Chemicals All chemicals were used as received without further purification.

2.2.2. Synthesis of Ag nanoparticles

The synthesis of Ag nanoparticles was carried out using the method described elsewhere with slight modification.(Peng *et al.*, 2010) A solution of 0.090 g of $AgNO_3$ (0.5 mmol) in 10 ml of oleylamine was prepared in a round-bottom flask. The flask was connected to a Schlenk line and protected with a nitrogen atmosphere while stirring at 600 rpm. The solution was degassed three times at 105 °C to remove moisture and oxygen, then heated to 190 °C at a ramping rate of ~10 °C/min. This temperature (190 °C) was maintained for 30 min to complete the reaction. The reaction solution was then rapidly cooled down by adding 20 ml of acetone. The Ag nanoparticles were collected by centrifugation at a rate of 2000 rpm for 10 min. The precipitate was re-dispersed in 5 ml of hexane. This centrifugation/re-dispersion process was repeated three times. The lastly collected precipitate was re-dispersed in 1:1 oleylamine: hexane to obtain a dispersion with concentration of approximately 5 mg Ag/ml

2.2.3. Characterizations of Ag nanoparticles

Transmission electron microscopy (TEM) images were taken using JEOL JEM-1400 microscope operated at 120 kV. SAXS measurements were conducted at the 12-ID-B beamline at the Advanced Photon Source (APS), Argonne National Laboratory (ANL). Colloidal Ag nanoparticle dispersions were loaded in quartz capillary tubes with diameters of ~1.5 mm. Each SAXS pattern was an average of at least 10 repeated measurements of a sample, with an exposure time of 1s for each measurement.

The SAXS profiles of colloidal Ag nanoparticles were background-subtracted against the SAXS profile of the 1:1 oleylamine: hexane solution.

3. Results and discussion

3.1. The size-distribution smearing effect

The reciprocal-space-based SAXS signal of a nanoparticle originates from the interference of the scattered X-rays by electrons in the particle, representing the Fourier transform of the three-dimensional (3D) structure of the nanoparticle in real space. The real-space PDDF ($P(r)$) of a particle represents the autocorrelation of the electron density map, ρ(r), of the particle:

$$P(r) = Vr^2\gamma_0(r) = V\int_{V_u}\rho(\vec{\boldsymbol{r}}+\vec{\boldsymbol{u}})\rho(\vec{\boldsymbol{u}})\mathrm{d}\,V_u \quad (1)$$

P(r) is a weighted histogram of electron pairs with a distance of *r* within the nanoparticle. For a single nanoparticle, the experimentally measured scattering intensity *I* (as a function of the scattering vector *q)* is related to *P(r)* by the Fourier transform:

$$I(q) = 4\pi\int_0^\infty P(r)\frac{\sin(qr)}{qr}\mathrm{d}r. \quad (2)$$

With the experimentally measured SAXS profile *I(q)*, the corresponding *P(r)* could be determined by using the inverse Fourier transform:

$$P(r) = \frac{1}{2\pi^2}\int_0^\infty I(q)qr\sin(qr)\,\mathrm{d}q. \quad (3)$$

In the SAXS measurement of a colloidal nanoparticle dispersion, the collected signal, $I_{ens}(q)$, of the nanoparticle ensemble is the sum of the SAXS signals of individual nanoparticles ($I_n(q)$) exposed to the X-ray beam:

$$I_{ens}(q) = \sum I_n(q). \quad (4)$$

If the nanoparticles have the same geometric shape (e.g., sphere) with only the difference in size, the SAXS signal of the nanoparticle ensemble can be expressed as

$$I_{ens}(q) = \int_0^\infty f(R)\cdot I_n(q,R)\mathrm{d}R, \quad (5)$$

where *f(R)* is the size distribution function that tells the frequency of particles with a size parameter of *R* (e.g., the radius of a sphere).

In this study, particles with the same geometric shape are defined as those that can be inter-converted by isotropic size scaling with respect to the center of the mass; for example, spherical particles with different radii have the same shape. For nanoparticles with the same shape but different sizes, the intensity *I* and scattering vector *q* scale as a function of the particle size parameter *R*. The SAXS intensity of a single nanoparticle, $I_n(q,R)$, is proportional to $V^2(R)$ (i.e., R^6). Therefore, larger nanoparticles contribute more significantly to the overall SAXS signals if they mix with smaller ones. The SAXS profile of an ensemble of spherical nanoparticles with non-uniform sizes represents a

volume-square-weighted average of the SAXS signal of individual particles. As the scattering vector q is in the reciprocal space, the q-axis of the scattering curve $I_n(q,R)$ inversely scales to the size parameter R: the curve shrinks towards the lower q region for large particles and stretches to the higher q region for small particles. If we assume the large and small particles have the same shape, the normalized scattering signal of a representative model particle can be described as $A(q')$ as a function of $q' = qR$. $A(q')$ is the pseudo form factor ($A(q')$ *vs* q') of the unit size particle for a given shape. For example, the theoretical SAXS signal of a spherical particle with a radius of R can be described as

$$I(q,R) = \left\{\frac{4\pi R^3}{3}\frac{3[\sin(qR)-qR\cos(qR)]}{(qR)^3}\right\}^2 = R^6 A(qR) = R^6 A(q'). \quad (6)$$

For an ensemble of spherical nanoparticles with a size distribution function $f(R)$, the total SAXS signal is

$$I_{ens}(q) = \int_0^\infty f(R)\cdot R^6 A(qR)\mathrm{d}R. \quad (7)$$

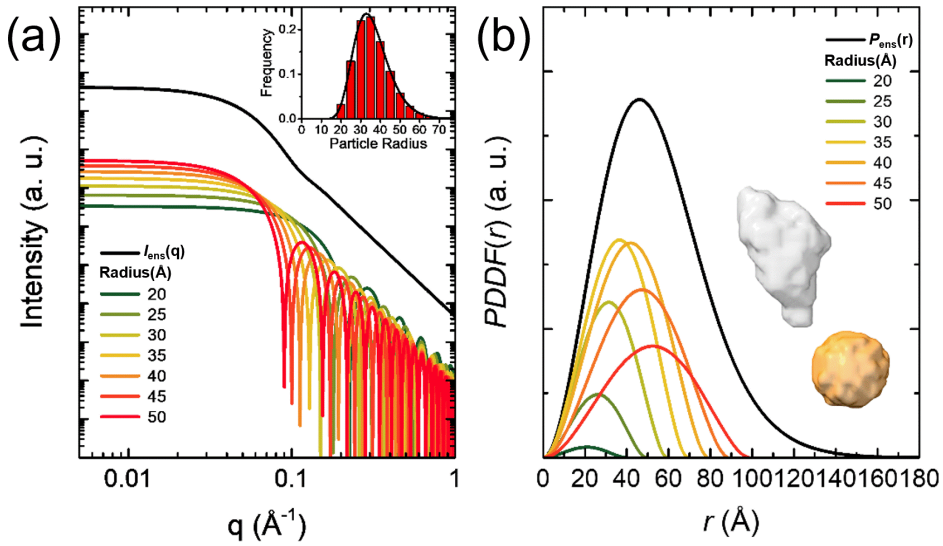

Figure 1 (a) Single particle SAXS profiles of spherical particles whose radii range from 20 Å to 50 Å (green to red curves) and the size-distribution smeared SAXS profile (black curve, offset for easy visualization) calculated using a μ = 35 Å, σ = 0.25 log-normal distribution (insert). (b) The corresponding PDDF profiles of spherical particles whose radii range from 20 Å to 50 Å (green to red curves), with their intensities normalized by the size-distribution populations (bars in inset, (a)) and the synthetic PDDF profile of the particle ensemble (black curve) considering the size-distribution effect. The models show the direct 3D shape reconstruction results of DAMMIF from the smeared PDDF profile of the particle ensemble (gray model) and the unsmeared PDDF profile of a single particle (orange model).

Figure 1(a) shows the smearing effect of size distribution using the computed SAXS profiles of spherical particles described by eq. (6). A log-normal size distribution function (Figure 1(a), insert) is applied to the particle ensemble:

$$f(R) = \frac{1}{R\sigma\sqrt{2\pi}} \exp\left(-\frac{(lnR - ln\mu)^2}{2\sigma^2}\right), \quad (8)$$

where μ is the median of particle radius (fixed as 35 Å), and σ is the distribution parameter (fixed as 0.25 or ~25% of the coefficient of variation). Figure 1(a) highlights the individual SAXS profiles of spherical particles with radii from 20Å to 50Å, which account for 95% of the particle populations in the ensemble. The oscillations in the SAXS profiles of individual nanoparticles with enlarging size shift towards the low q direction. Although the nanoparticles are perfectly spherical, the total oscillation in the ensemble SAXS profile does not show strong oscillation peaks and valleys. The reconstructed 3D geometry from this ensemble SAXS profile would be distorted and does not represent the particles in the ensemble (Figure 1b).

The PDDF calculated from the SAXS signal of a nanoparticle ensemble is expressed as:

$$P_{ens}(r) = \frac{1}{2\pi^2}\int_0^\infty I_{ens}(q)(q \cdot r)\sin(q \cdot r)\,\mathrm{d}q$$

$$= \frac{1}{2\pi^2}\int_0^\infty \int_0^\infty f(R) \cdot R^6 A(qR)(q \cdot r)\sin(q \cdot r)\mathrm{d}R\,\mathrm{d}q. \quad (9)$$

Due to the linear property in the Fourier transform, the linear combination of scattering signal [$I(q,R) = R^6A(qR)$] weighted by $f(R)$ in the reciprocal space also gives the linear combination of PDDF in the real space

$$P_{ens}(r) = \int_0^\infty f(R)\left[\frac{1}{2\pi^2}\int_0^\infty R^6 A(qR)(q \cdot r)\sin(q \cdot r)\,\mathrm{d}q\right]\mathrm{d}R$$

$$= \int_0^\infty f(R)\,P_{SP}(r,R)\,\mathrm{d}R, \quad (10)$$

where $P_{SP}(r,R)$ is the PDDF of a single particle with a size parameter R. Similar to the scattering intensity function $I(q)$ of the particles with the same shape that can be scaled by the size parameter R, their corresponding PDDFs can also be scaled. As inferred from eq. (10), the amplitude of $P_{SP}(r,R)$ is proportional to $V^2(R)$ (i.e., R^6).

Therefore, the PDDF of a particle with size parameter R can also be obtained from the stretching transformation from the PDDF of a unit-size particle with the same shape ($P_0(r')$):

$$P_{SP}(r,R) = R^6 P_0(r') = R^6 P_0(r/R), \quad (11)$$

in which $r' = r/R$ ($r' \in [0,1]$). Thus, $P_{ens}(r)$ can be reduced to a linear combination of a series of stretched $P_0(r')$ weighted by the size distribution:

$$P_{ens}(r) = \int_0^\infty f(R)\,R^6 P_0(r')\,\mathrm{d}R = \int_0^\infty f(R)\,R^6 P_0(r/R)\,\mathrm{d}R, \quad (12)$$

which has a similar description to $I_{ens}(q)$ in the reciprocal space expressed in eq. (7).

Figure 1(b) shows how the PDDF profiles of a series of spherical nanoparticles of varying sizes combine into a total PDDF profile, using the same log-normal size distribution described in eq. (8). In the log-normal distribution, the population of large particles gradually decreases. As a result, the ensembled total PDDF profile also features a diminishing tail, as does the size-distribution function. This tail is contributed by the large particles in the ensemble, even though they have a small population, since the PDDF intensity scales by R^6. Compared to the PDDF of an individual particle with a sharply defined D_{max} (the maximum pair distance within a particle), the tailing effect often leads to an overestimation of particle size and gives misleading geometry. In order to obtain reliable geometric information, the PDDF curve must be deconvoluted from the smeared profile before ab initio reconstruction. Directly using the PDDF profile of a particle ensemble with size distribution for DAMMIF shape reconstruction results in an asymmetric, distorted geometry (gray model, inset/Figure 1b) that offers no clue about the original particle shape (orange model, inset/Figure 1b).

3.2. A penalized regularization-based size-refocusing method on deconvoluting representative PDDF

Solving either the inverse problem described in eq. (7) in reciprocal space or eq. (12) in real space becomes necessary to extract the particle geometry information from SAXS measurements. Both equations are analogous to the image deblurring problem, with typical difficulties associated with inverse problems, such as having no unique solution and being highly sensitive to noise. Since there is no universal analytical equation to describe the scattering intensity $I(q)$ or the PDDF $P(r)$ for a particle with an arbitrary shape, these inverse problems cannot be reduced to fit an analytical equation with limited parameters. Therefore, the problem must be converted into a discrete function that can be solved numerically. Compared to the scattering intensity $I(q)$, which contains multiple oscillation waves with intensities following a power-law decay, the PDDF $P(r)$ is easier to solve in a discrete, numerical way. As the PDDF $P(r)$ is essentially a distribution function, it should have typical features such as continuous, positive, and converging to zero at the high and low limits. These features can be guidelines when conducting numerical fittings to exclude unreasonable solutions.

To find a possible numerical solution of $P_0(r/R)$, one can discretize the $P_0(r/R)$ into N segments (i.e., fitted $\hat{P}_0(\boldsymbol{r}/R)$ array) and try different value combinations until the fitted $\hat{P}_{ens}(\boldsymbol{r})$ gets close enough to the observed $P_{ens}(\boldsymbol{r})$. However, this approach will lead to an infinite number of solutions if no constraint is applied to $P_0(r/R)$. Additionally, this searching approach, if fully randomized (Monte-Carlo), will have a total computational complexity of d^N as the solution lies in an N-dimensional space. Therefore, the search for $P_0(r/R)$ must be conducted with a priori knowledge to tell what a good $P_0(r/R)$ is and in which direction one can find it. The qualification of a "good" $P_0(r/R)$ can be justified by using penalty functions, which has been discussed during the development of GNOM by

Svergun.(Svergun, 1992) Fundamentally, one can use Chi-squared (CHISQR) to describe how far the fitted $\hat{P}_{ens}(\boldsymbol{r})$ is from the observed one:

$$\text{CHISQR}=\frac{1}{N}\left(\frac{\hat{P}_{ens}(\boldsymbol{r})-P_{ens}(\boldsymbol{r})}{P_{ens}(\boldsymbol{r})}\right)^2, \quad (13)$$

where $\boldsymbol{r}$ is the array of spacing for numerical calculation. In this research, $\boldsymbol{r}$ is set as an evenly spaced array from r=0 to r=3μ (3-fold of the estimated median particle size, μ) at which the contribution of a particle to the ensembled $P_{ens}(r)$ is nearly vanished. Only adapting CHISQR as the only goal function to minimize does not give a fitted $\hat{P}_0(r/R)$ with high fidelity. The data points on the fitted $\hat{P}_0(r/R)$ are usually scattered and may have negative values. Those $\hat{P}_0(r/R)$ may mathematically give a good fit to the ensembled $P_{ens}(\boldsymbol{r})$, but are counterintuitive to our knowledge of the PDDF of real particles: the fitted $\hat{P}_0(r/R)$ is supposed to be smooth because the particles at the SAXS scale can be considered continuously-distributed matters. For a size-normalized model particle, where the maximum dimension of the fitting box is normalized to $D_{\max}$ (i.e., $r/D_{\max}$ spans from 0 to 1), the smoothness (SMOOTH) of the fitted array $\widehat{\boldsymbol{P}}_{\boldsymbol{0}}$ can be quantified as:

$$\text{SMOOTH}=\frac{1}{\pi}\left(\left\|\widehat{\boldsymbol{P}}'_0\right\|/\left\|\widehat{\boldsymbol{P}}_0\right\|\right), \quad (14)$$

where $\|\boldsymbol{A}\|$ denotes the norm of an array: $\|\boldsymbol{A}\|=\left(\sum_{i,j}\left|a_{i,j}\right|^2\right)^{1/2}$ and $\widehat{\boldsymbol{P}}'_0$ is the first derivative array of the function $\hat{P}_0(r/R)$. SMOOTH equals to ~1.1 for the PDDF of a solid sphere. The SMOOTH penalty term increases as the points on the $\widehat{\boldsymbol{P}}_{\boldsymbol{0}}$ array get scattered. Second, solid particles often have higher electron density than the surrounding environment in SAXS experiments. In such a case, the PDDF calculated from a background-subtracted SAXS profile should also have positive values. The positivity (POSITV) of an array can be quantified as:

$$\text{POSITV}=\frac{1}{2}\left(\frac{\left\|\widehat{\boldsymbol{P}}_0+\left|\widehat{\boldsymbol{P}}_0\right|\right\|}{\left\|\widehat{\boldsymbol{P}}_0\right\|}\right). \quad (15)$$

An array of all positive elements will have a POSITV value equal to 1 while an array of all negative elements has a value of 0. There are other penalty functions in the discussion of GNOM paper(Svergun, 1992) for quantifying the goodness of a fitted curve. To simplify the calculation, only SMOOTH and POSITV are adapted. The following results show that those two terms are enough to solve the current problem. A good fitting of $\hat{P}_0(r/R)$ should simultaneously have a CHISQR as low as possible, a reasonable SMOOTH, and a POSITV equal to (or close to) 1. Therefore, the overall goal function with the added penalty terms becomes:

$$\text{GOAL}=\text{CHISQR}+a1\,(\text{SMOOTH}-1.1)+a2\,(1-\text{POSITV}), \quad (16)$$

where a_1 and a_2 are the weighing factors of penalty terms that indicate how important those qualities should be considered during optimization. Empirical magnitudes of a_1 and a_2 are 10^{-3} and 10^4, respectively, which accept $\hat{P}_0(r/R)$ with SMOOTH from 1.1~4 but reject most $\hat{P}_0(r/R)$ with any negative values. With a goal function described in eq. (16), we used Broyden–Fletcher–Goldfarb–

Shanno algorithm (BFGS) to determine the gradient descent direction with adopting the approximated Hessian matrix.(Fletcher, 2000) The BFGS algorithm avoided the calculation of the inverse of large matrices, making it suitable for optimizing problems with a large number of variables. Such regularized fitting over eq. (12) will give the PDDFs at unit size, i.e., $\hat{P}_0(r/R)$ and at the representative median size R, i.e., $P(r,R)$.

The process of deconvoluting the SAXS and PDDF profiles of an individual particle from a particle ensemble is summarized in the flowchart in Figure 2. The penalized regularization-based size-refocusing method described above is implemented as a Python script named SharPy. Its contribution to the data processing procedure is emphasized in the flowchart in green. This Monte-Carlo-like process consists of two parts: guessing a potential result from the PDDF of the particle ensemble and the results from previous iterations and fitting the results of individual particle PDDF to that of the particle ensemble while evaluating the penalty terms. This guess-fit circle fills the gap in solving the inverse problem of deconvoluting the individual particle PDDF from the particle ensemble. This process makes it possible to sharpen the SAXS signals from the size-distribution smearing effect and eventually extract the geometry information of the polydisperse particle system.

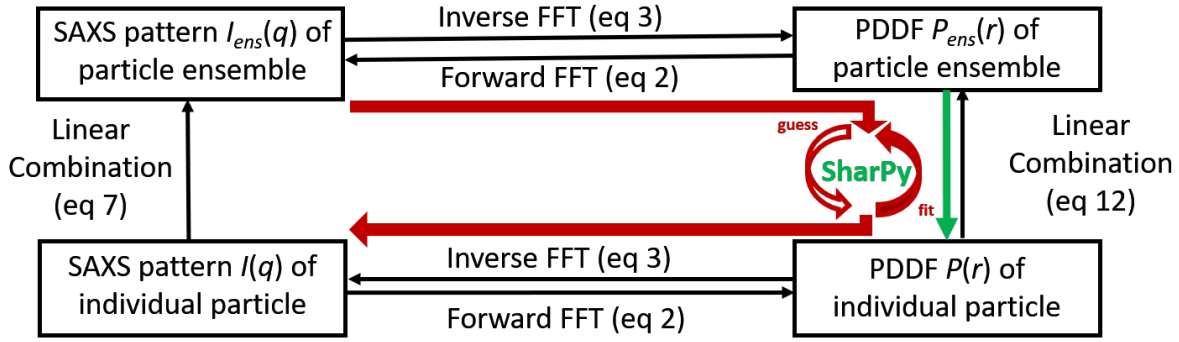

Figure 2 Flowchart describing the relationship of the SAXS and PDDF profiles of individual particle and particle ensemble. The red arrows highlight the process of deconvoluting the SAXS and PDDF profiles of an individual particle from a particle ensemble and the role of SharPy in this procedure is highlighted by green arrows.

Figure 3(a) shows the deconvoluted PDDF of a particle with the median size, i.e., $P_0(r,R)$ from the ensembled $P_{ens}(r)$ of spherical particles with a 25% log-normal size distribution and a median diameter of 70Å. The original ensembled $P_{ens}(r)$ (upper panel, black curve) has a vanishing tail caused by the size distribution, indicating an overestimation of the particle size. In contrast, the recovered PDDF of the particle at the median size (lower panel, purple dots) after size refocusing shows a clear cut-off D_{max} at 70 Å, which fits well with the theoretical PDDF calculated from an ideally spherical model (lower panel, gray curve). Figure 3(b) shows the SAXS scattering intensities converted from the corresponding PDDFs presented in Figure 3(a). Compared to the original one with the size distribution smearing effect (black curve), the recovered SAXS profile (purple curve) shows characteristic Bessel-type oscillations

of a symmetric sphere. The oscillations' position and spacing fit well with that calculated from the ideal model (gray curve). The recovered SAXS profile did not reach 100% fidelity to the original one, possibly because the potential of PDDF fitting function is relatively flat around the actual solution, and the Fourier transform further magnifies the discrepancy in resulting SAXS, especially in the log-log presentation. However, the deconvoluted PDDF is already accurate enough to quantitatively analyze the essential geometry parameters of particle ensembles.

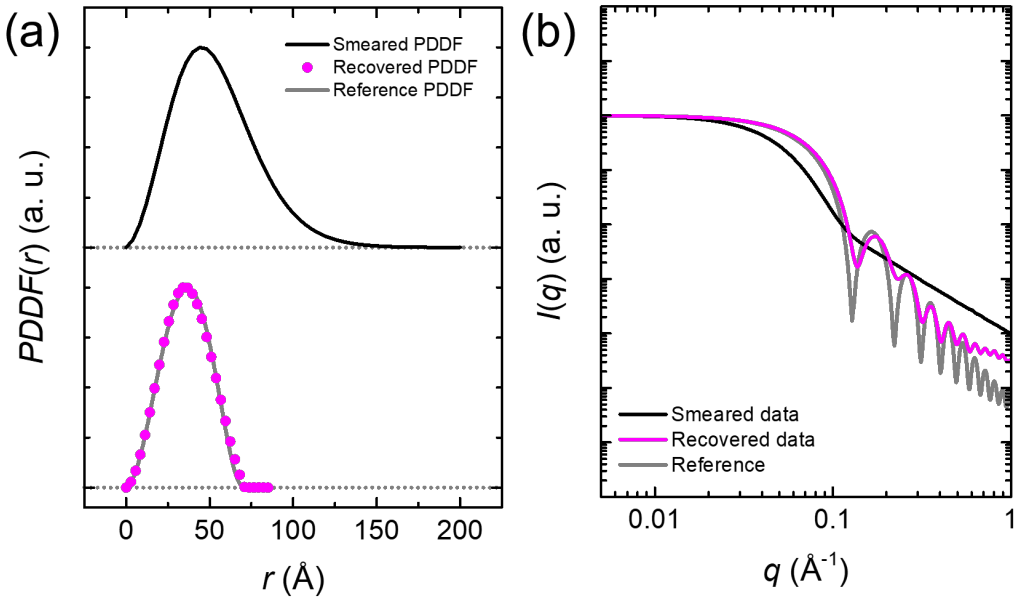

Figure 3 (a) The deconvoluted $\hat{P}_0(r)$ ($R = 35$Å) (pink dots) from the smeared $P_{ens}(r)$ (black curve) of a spherical particle ensemble with a median radius μ = 35 Å, $\sigma = 0.25$ log-normal distribution, which fits well with the reference PDDF profile computed from a mathematically ideal sphere model (gray curve). (b) The corresponding SAXS profiles calculated from the PDDF profiles in (a), with intensity normalized for comparison.

The roles of each penalty term are studied by optionally introducing them into the goal function. A slightly over-estimated $D_{\max}$ (80 Å) is chosen to ensure the search box includes the PDDF curve completely. As shown in Figure 4(a), adapting CHISQR as the only goal function to minimize does not result in a reasonable $\hat{P}_0(r/R)$ (gray plot). The data points on the fitted $\hat{P}_0(r/R)$ are scattered and turn negative where r > 70 Å. Introducing POSITV into the goal function removes the negative parts (green plot), but the main part of the curve is still scattered. The SAXS intensity profiles calculated from the fitted PDDF curves (Figure 4(b)) are far away from the reference presented in Figure 3(b) and even have negative values. Adding SMOOTH to the goal function only makes the fitted PDDF curve smooth, and the part where r > 70 Å is still negative (blue plot, Figure 4(a)). The difference between SAXS intensity profiles of adapting SMOOTH only and both POSITV and SMOOTH may not be notable (blue curve vs. purple curve, Figure 4(b)), but the negative parts still cause difficulties for the 3D shape-reconstruction programs. For example, a dummy atom modeling program like DAMMIF can get

confused by the negative part of the PDDF curve. The 3D structures reconstructed from the fitted PDDF have irregular shapes and fractal surfaces as the 3D reconstruction program tried to "fit" the scattered and negative parts in the PDDF (Figure 4(c)). It is not surprising that the fitted $\hat{P}_{ens}(\boldsymbol{r})$ and the corresponding $\hat{I}_{ens}(q)$ all fit well with the target $P_{ens}(\boldsymbol{r})$ and $I_{ens}(q)$, despite the addition of penalty terms. The high-quality fitting again emphasizes the importance of adding regulations for selecting the most reasonable solutions from many possible candidates.

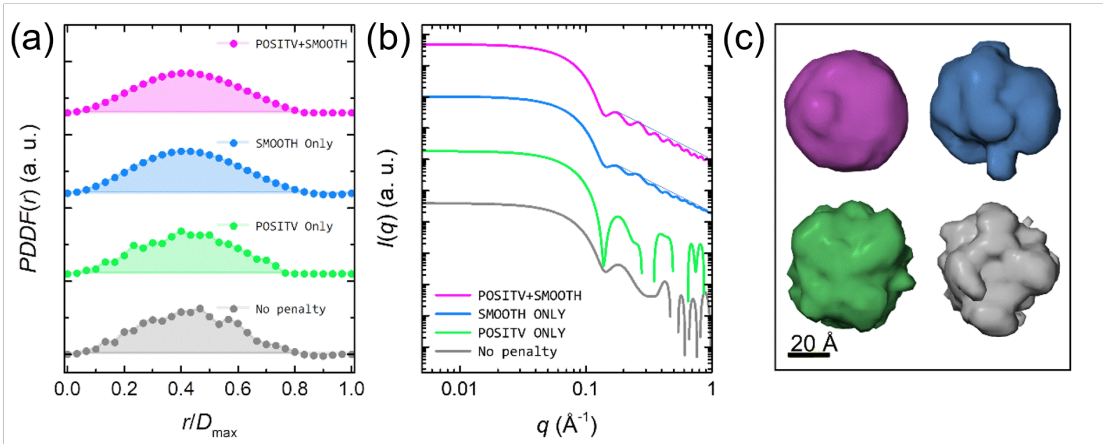

Figure 4 (a) The deconvoluted $\hat{P}_0(r/D_{\max})$ ($D_{\max} = 80$ Å) from the smeared $P_{ens}(r)$ of a spherical particle ensemble with a $\mu = 35$ Å, $\sigma = 0.25$ log-normal distribution, with penalty terms optionally added to the goal function: adding both POSITV and SMOOTH (pink), adding SMOOTH only (blue), adding POSITV only (green), and no penalty terms (gray). (b) Corresponding SAXS intensity profiles calculated from the PDDF profiles in (a), with intensities offset for easy visualization. (c) 3D reconstruction results from DAMMIF calculated using the SAXS intensity profiles in (b). The color coding is the same for (a-c).

Simulated SAXS profiles of particle systems with different shapes are generated to verify the generality of the developed approach. The well-described common particle shapes, such as dimers, plates, rods, and spheres, are reduced to voxel models rasterized by a 64×64×64 grid (more details see Methods section). Autocorrelation and spherical average are then applied to the voxel models to calculate the PDDF of the model particle according to eq. (1). A log-normal distribution with median half maximum dimension $\mu = 25$ Å, $\sigma = 0.25$ is applied to the PDDF to simulate the smeared $P_{ens}(\boldsymbol{r})$ and $I_{ens}(q)$. The fitting procedure was conducted in a search box larger than the maximum dimension of the particle ($D_{\max}$=80 Å), with only the σ values of the size distribution being provided. The tested models include:

(i) dimers consisting of a pair of identical spherical particles with varying spacings;
(ii) round plates with the same equatorial radius but with varying thicknesses;
(iii) rods with the same length but with varying equatorial radii;
(iv) hollow spheres with the same size but varying hollow interiors.

The maximum dimensions of the model particles are all normalized by $D_{\max}$. The PDDF recovered by the developed approach fits well with the PDDF directly calculated from the model (Figure 5), which shows an overall similarity >0.99 in most cases, quantified by the Pearson's correlation coefficient:

$$corr(A,B) = \frac{\sum(A_i-\bar{A})(B_i-\bar{B})}{\sqrt{\sum(A_i-\bar{A})^2\sum(B_i-\bar{B})^2}}, \quad (17)$$

where A and B are substituted by the arrays of fitted $\hat{P}_0(r/D_{\max})$ and reference $P_0(r/D_{\max})$, respectively. The distinct features of PDDF curves for particles with assorted shapes are accurately reconstructed, allowing for qualitative shape determination. A dimer structure can be identified by the presence of characteristic double peaks in the PDDF curve (Figure 5(a-c)). Other notable PDDF details include the broad and mostly symmetrical peak of a round plate (Figure 5(d-f)), the sharp apex and sloped peak tail of a cylindric rod (Figure 5(g-i)), and the triangular-shaped peak of a hollow shell (Figure 5(j-l)). The high accuracy also allows for quantitative analysis of the model particles. For example, it makes it possible to measure the distance between the two domains of a dimer (Figure 5(a-c)). When the size distribution is high, the two peaks in the PDDF curve of a dimer become overlapped and might be difficult to distinguish. However, data processing with the size-refocusing approach results in the two peaks that can be well separated with precise peak positions. The corresponding scattering intensity profiles $\hat{I}_0(q)$ also show similarity to the references $I_0(q)$ that are calculated from voxel models (Figure S1). All the fitted $\hat{I}_0(q)$ follow the Guinier law of $I\sim\exp\left(-\frac{1}{3}R_g^2q^2\right)$ in the low q range ($q < 1/R_g$, R_g representing radius of gyration). The power laws of the intensity decrease are also followed for different particles: $I\sim q^{-1}$ for 1D nanorods (Figure S1(d-f)) and $I\sim q^{-2}$ for 2D nanoplates (Figure S1(g-i)) in the corresponding q range (from ~ 0.1 to ~ 0.2Å^{-1}). As voxel models with fractal surfaces were used, the Porod law of $I\sim q^{-4}$ at high q range (q > 0.5 Å^{-1}) is not followed for a smooth particle. The data at high q pertain more to the details of high spatial resolution, such as the surface structure. It is unnecessary to include high q range scattering data as it requires high-precision modeling and high computational cost. Overfitting may also result in possible mismatching of the recovered $\hat{P}_0(r/D_{\max})$ to the reference $P_0(r/D_{\max})$. For example, the implementation of SMOOTH penalty term may cause a slight distortion on the apex of the PDDF of hollow shell structures. The possible overfitting is mitigated by adjusting the weighting factor of each penalty term (i.e., a_1 and a_2 in eq. (16)). To clarify, a smaller a_1 (1/10 to the regular value) was employed during the fitting of hollow shell models to reduce the distortion on the peak apex (Figure 5(j-l)). In practice, we suggest operators adjust those fitting parameters on a case-by-case basis to obtain good results.

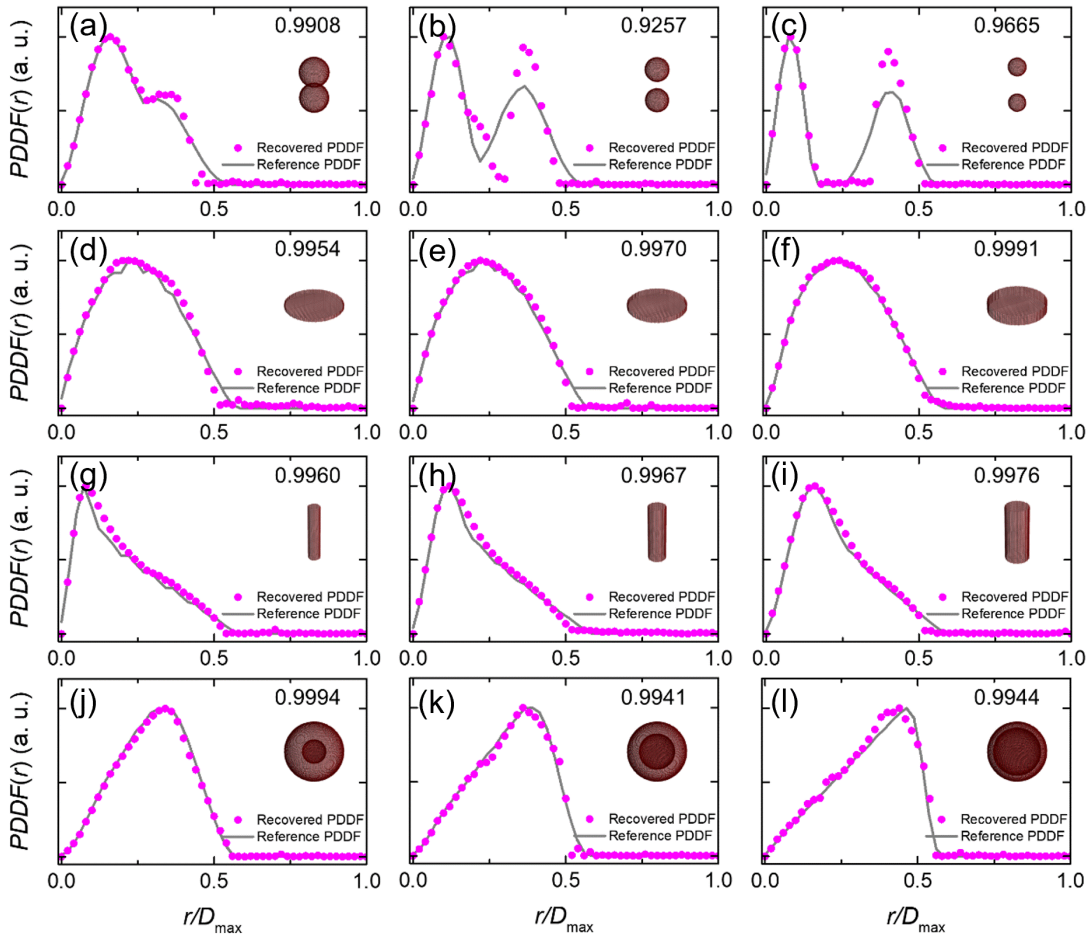

Figure 5 The recovered $\hat{P}_0(r/D_{\max})$ ($D_{\max} = 80$ Å, pink dots) from the smeared $P_{ens}(r)$ of particle ensembles with assorted shape: (a-c) dimers with increasing spacing, (d-f) circular nanoplates with increasing thickness, (g-i) nanorods with increasing equatorial radius while maintaining the same length, (j-l) nanoshells with enlarging hollow interiors while maintaining the same outer diameter. The size distribution was fixed to a log-normal distribution with a median size of 50 Å along the longest direction and a standard deviation of 0.25. The reference curves (gray curves) were computed from the 64×64×64 voxel models (insets). The numbers above the voxel models refer to the correlation coefficient between the recovered and reference PDDF curves.

The dynamics of the fitting process are presented in Figure 6, using one trial on the dimer case (Figure 5(a)) as an example. The fitting process starts with an initial guess of $\hat{P}_0(r/D_{\max})$, which is a perturbated (15%) Gaussian distribution curve. The purpose of introducing randomness is to explore the solution space and avoid potential local minima during optimization. After approximately 120 iterations, the process ends with a result showing the typical double-peak shape PDDF of a dimer structure (Figure 6(a)). Meanwhile, the size-distribution-smeared $\hat{P}_{ens}(r)$ gradually converges to the target $P_{ens}(r)$, which shows only a singular peak (Figure 6(b)). Figure S2 illustrates how the terms (i.e., CHISQR, SMOOTH, and POSITV) in the goal function evolve during the fitting process. As a potent

regulator, POSITV always sticks to 1, so any trial with negative values is rejected. At the beginning of the search, CHISQR quickly reduces to reach an overall fit for the target $P_{ens}(r)$ (the green stage in Figure 6(b)). The rapid reduction of CHISQR is at the expense of losing suitable SMOOTH. The temporary $\hat{P}_0(r/D_{\max})$ (the green stage in Figure 6(a)) is still noisy. After the initial increase, SMOOTH drops to approximately 2.5 while CHISQR continuously drops to 10^{-2} (the yellow stage). The refinement during this stage removes the spikes on the $\hat{P}_0(r/D_{\max})$ and pushes the $\hat{P}_{ens}(r)$ closer to the target $P_{ens}(r)$. According to the definition in eq. (14), a dimer's PDDF profile with 2 peaks and 1 deep has a SMOOTH value of ~3. At the late stage (orange to red), the peak on the $\hat{P}_0(r/D_{\max})$ splits into two peaks and the SMOOTH value raises back to ~3. The driving force at this stage is the continuous refinement of CHISQR. The combination of the gradient descent searching and the penalty terms makes the fitting process able to find the hidden peaks that are smeared in the ensembled $P_{ens}(\boldsymbol{r})$ due to the size distribution.

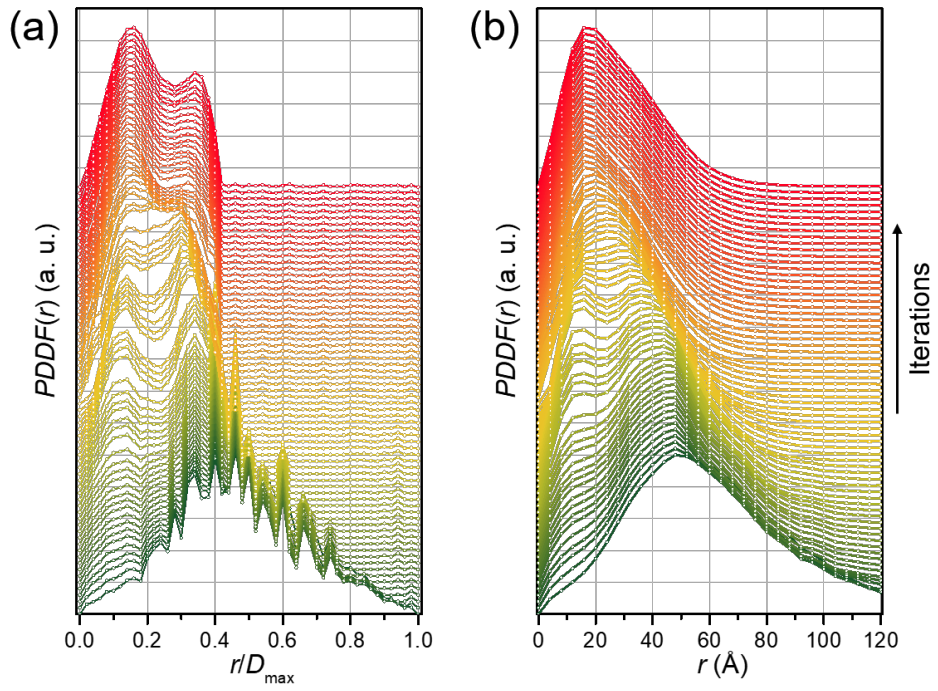

Figure 6 The evolution dynamics of (a) guessed PDDF of single particle $\hat{P}_0(r/D_{\max})$ and (b) fitted PDDF of particle ensemble $\hat{P}_{ens}(r)$ in the course of fitting the dimer models presented in Figure 5(a).

The size-refocusing approach, which is based on gradient descent searching with the regularization of penalty terms, performed well on the simulated dataset of particle ensembles with different geometries. However, as with any numerical fitting method used to solve an ill-posed inverse problem, there is a concern that the "good" performance may result from overfitting,(Hawkins, 2004) where a fitting that focuses too much on the trivial details of the input data. In an overfitting case, once those trivial details are covered, for example, by the experimental noise, the fitting will become unstable and generate unreliable results. We conducted experiments to determine if the method remains stable when facing

different interferences. According to eq. (12), possible interferences may come from either the observed $P_{ens}(r)$ or the pre-determined size-distribution function *f*(*R*). The dimer particle ensemble in Figure 5(a) was chosen as the test case, where the particles have a log-normal size distribution with a standard deviation of 0.25 (the gray curve). A 5% random perturbation was added to the input $P_{ens}(r)$ data to simulate the experimental noise generated from measurements and data handling processes. Despite this interference, the fitting process is still able to deconvolute the two-peak shaped $\hat{P}_0(r/D_{\max})$ from the $P_{ens}(r)$ of a dimer particle ensemble with a 25% log-normal size distribution (**Figure S3(a),** orange plot). The fitting result indicates that the overall $P_{ens}(r)$ curve is fitted instead of focusing on the trivial details from artificially generated data. The pre-determined size distribution function *f*(*R*) may also introduce errors to the fitting process. Typically, the size distribution of a particle ensemble grown from a solution can be described by using a log-normal size distribution (eq. (8)). The median particle size μ and the standard deviation σ need to be determined from the statistical analysis of particle size in order to fit with a log-normal size distribution. In this approach, we assume that other methods, like electron microscopy, determine *f(R)*. As a result, the fitting algorithm should be robust in case there are minor errors in determining the size distribution function *f*(*R*). The accuracy of the median size μ does not affect the fitting results because the fitting object is a size-normalized, unitless PDDF of the model particle. However, selecting a significantly underestimated μ should be avoided because it would result in the PDDF exceeding the limit. An overestimated μ does not cause any harm to the fitting since the margin is penalized to zero. The actual median size μ (or the maximum dimension) of the particle can still be determined from fitting even if the input median size μ is overestimated. As for the other input parameter σ, tests were conducted on the same dimer particle ensembles with σ=0.25 (~25%) log-normal distribution but giving a false *f*(*R*) with σ=0.20 (underestimated) or 0.30 (overestimated) to the fitting program (**Figure S3(a),** blue and purple plots). An overall peak shift is observed in both cases because the log-normal size distribution is an asymmetric distribution function: the portion of large particles increases as σ increases. It is important to point out that the SAXS measures the volume-squared average of properties of a particle ensemble. The underestimation or overestimation of the contribution of large particles leads to the averaged particle size shifting towards the opposite direction. In contrast, the geometry information is still well recovered. The influence of picking different types of distribution functions was also studied. While asymmetric distribution functions like log-normal and Schulz-Zimm are more accurate in describing low-quality particle ensembles, the symmetric Gaussian (or normal) distribution function is commonly used to describe the polydispersity of synthesized particles in many studies:

$$f(R) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(R-\mu)^2}{2\sigma^2}\right). \quad (18)$$

Robustness tests indicate that the fitting program can recover the characteristic two-peak feature of a dimer structure if the size distribution function used in the fitting process is substituted by a Gaussian

distribution function with the same μ and σ (**Figure S3(a)**, green plot). The corresponding size-refocused SAXS intensity profiles $I(q)$, converted from the PDDF, are compared in **Figure S3(b)**. The power law of $I \sim q^{-1}$ for a linearly placed dimer is followed in the corresponding q range (from ~ 0.1 to ~ 0.2Å^{-1}) in all cases. The size-refocused SAXS profiles are suitable for quantitative morphological analysis. For example, the distance between the center of two spheres of a dimer structure can be determined from the positions of the two peaks (P_1 and P_2) in its PDDF curve. For a dimer where the centers of two spheres are spaced twice as far as their radii, the PDDF peaks is spaced following $r(P_1)$: $r(P_2)$ = 0.5. The fitted PDDF with falsely given $f(R)$ has an $r(P_1)$: $r(P_1)$ ratio between 0.4 and 0.5, referring to dimer structures where the two domains slightly fused into each other, possibly forming bottleneck at the interface. A falsely given $f(R)$ can result in incomplete removal of the size-distribution smearing effect. It is unnecessary to strive for perfectionism in this issue since the size-distribution smearing effect causes a loss of information, making it impossible to refocus the SAXS profile perfectly. The more accurate we know about the size distribution, the higher quality we can get from the refocusing fitting. The results suggested that precise determination of particle size distribution function is essential even though our refocusing fitting is robust. The developed size refocusing approach can tolerate falsely given $f(R)$ and extract quantitative geometrical information from smeared data.

Real-world size distributions of nanoparticle ensembles are often complicated, displaying irregular curves such as multi-peaks. To assess the efficacy of our method in handling such cases, we present an example of refocusing the PDDF of spherical nanoparticle from an ensemble with a bimodal Gaussian size distribution. Specifically, we simulated a nanoparticle ensemble with a size distribution consisting of two Gaussian distributions (Figure S4(a)). One distribution had a mean diameter of 30 Å, while the other had a mean diameter of 90 Å. Both distributions exhibited a standard deviation of 15%. It is noteworthy that in SAXS measurements, the scattering intensity and PDDF are directly proportional to the volume squared of the particles. Consequently, the larger particle contributed 729 (3^6) times more to the scattering intensity and PDDF than the smaller particle. To ensure comparable contributions in the PDDF, we set their ratio based on the number of particles to 1:1000 (Figure S4(b). The size-refocusing algorithm also performed well in the case of the bimodal size distribution, as presented in Figure S4(c). The underlying deconvolution principle remains consistent regardless of the distribution function's appearance. The algorithm successfully retrieved the characteristic one-peak, symmetric feature of the spherical particle PDDF from the complex PDDF with bimodal size distributions.

We applied the size-refocusing fitting method to the experimental SAXS profiles of colloidal Ag nanoparticles. The Ag nanoparticles were synthesized from chemical reduction of $AgNO_3$ in oleylamine at elevated temperatures and characterized with transmission electron microscopy (TEM) (**Figure 7(a)**). The details of the synthesis, TEM characterization, and SAXS experiment are described in supporting materials. The synthesized Ag nanoparticles were spherical in shape, with an average radius of 7.9 nm. The size distribution, presented in Figure 7(b), indicates that the synthesized Ag nanoparticles have a

polydisperse distribution with a standard deviation of ~12%. The PDDF profile of the Ag nanoparticles was calculated from the measured SAXS profile. Both the measured SAXS profile and PDDF profiles were found to be smeared due to the polydispersity of the particles. We applied the size-refocusing method to the smeared PDDF profile using a 12% log-normal size distribution to obtain a more accurate representation of the spherical Ag nanoparticle geometry. The resulting recovered PDDF profile, shown in Figure 7(c), exhibits a more symmetric shape and a clear cut-off at D_{max} around 9 nm, which is close to the average diameter determined from the TEM measurement. The size-refocused SAXS profile was calculated from the recovered PDDF profile using the forward Fourier transform. The size-distribution smearing effect is removed from the SAXS profile, as shown in Figure 7(d). The size-refocused SAXS profile exhibits a clear oscillational structure, enabling more accurate determination of morphological parameters such as the radius of gyration.

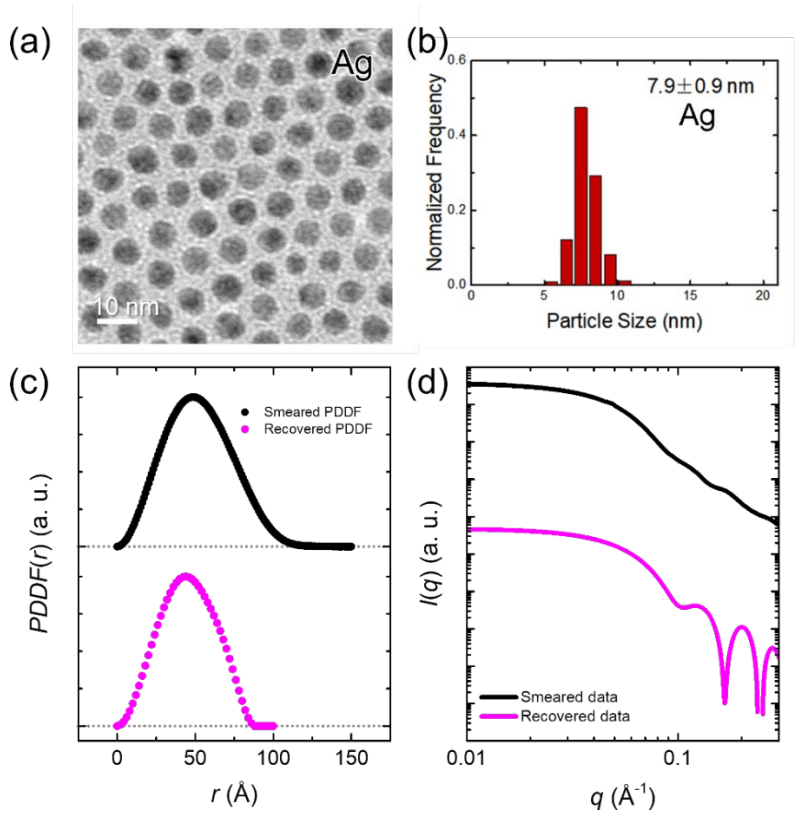

Figure 7 (a) TEM image of the synthesized Ag nanoparticles. (b) Statistic histogram of the diameter of Ag nanoparticles measured from TEM images. (c) PDDF profile and size-refocused PDDF profile of Ag nanoparticles using a 12% log-normal size distribution function. (d) SAXS profiles measured from the colloidal Ag nanoparticle dispersions and the size-refocused SAXS profile of Ag nanoparticles converted from the size-refocused PDDF, intensities vertically offset for visibility.

While our method has demonstrated promising results in addressing the challenges posed by size polydispersity in nanoparticles, it is important to acknowledge the limitations of this approach. First,

our focus on the nanoparticle polydispersity is their variations in size, and the PDDF of the particles in the ensemble should follow the scaling function in eq. (11). However, this requirement may not be applicable to complex nanoparticles composed of multiple domains, where polydispersity encompasses parameters beyond size distribution. For instance, nanoparticles with a core-shell structure exhibit polydispersity that involves variations in overall size, the core-shell ratio, and the symmetry of the core and shell. Our method will not work for such systems. Second, the current method also relies on the known particle size-distribution in the fitting process. However, in situations where techniques like TEM are not available, there is a need to extract both the size-distribution and accurate geometrical information solely from the SAXS profile. Although our method has shown effectiveness even when the size-distribution is not precisely determined, accurately determining the size distribution without any a priori knowledge of the particle geometry remains a challenge. Future research efforts could focus on developing functions to directly determine the size distribution from the SAXS profile, enabling simultaneous fitting of both the size distribution functions and the PDDF. This enhancement would significantly broaden the applicability of our size-refocusing algorithm and extend its potential in studying synthesized nanoparticles without the need of additional characterization techniques.

4. Conclusion

We have developed a practical fitting approach for reducing the size-distribution smearing effect in the PDDF profile of polydisperse particle system. The approach relies on gradient descent searching with regularization of penalty terms. The resulting PDDF of the size-refocused model particle is of high quality for shape reconstruction. While the size-refocusing procedure requires a pre-determined size distribution function, it is still robust even when the size-distribution function is not accurately determined. Size distributions of nanoparticle systems usually undermine the quality of SAXS data, rendering many such systems unsuitable for SAXS study. Preparing samples to narrow down polydispersity could be time-consuming and generate unnecessary waste. Separation techniques like size-exclusion chromatography and size-selective precipitation may not always work for various nanoparticle systems made of different materials. Our size-refocusing data fitting method provides a post-processing strategy for the polydispersity problem. The size-refocusing fitting will also benefit in-situ SAXS studies of nanoparticle growth and transformation, where polydispersity cannot be narrowed in real time. The size-refocusing procedure reveals detailed morphological information that benefits the application of advanced 3D reconstruction techniques on polydisperse nanoparticle systems. We expect that the size-refocusing fitting method will become an essential step in the SAXS data processing, helping researchers obtain clear and detailed 3D images from nanoparticles of interest.

Acknowledgements We express sincere thanks to Dr. Xiao-Min Lin (Center for Nanoscale Materials) for the support in the synthesis of Ag nanoparticles.

Funding information This work is supported by the National Science Foundation of United States under NSF award 2002960. SAXS data were collected at Beamline 12-ID-B of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science user facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357.

References

Bertero, M., Boccacci, P., Desiderà, G. & Vicidomini, G. (2009). *Inverse Problems* **25**, 123006.

Borchert, H., Shevchenko, E. V., Robert, A., Mekis, I., Kornowski, A., Grübel, G. & Weller, H. (2005). *Langmuir* **21**, 1931–1936.

Doucet, M., Cho, J. H., Alina, G., Attala, Z., Bakker, J., Bouwman, W., Butler, P., Campbell, K., Cooper-Benun, T., Durniak, C., Forster, L., Gonzalez, M., Heenan, R., Jackson, A., King, S., Kienzle, P., Krzywon, J., Murphy, R., Nielsen, T., O'Driscoll, L., Potrzebowski, W., Prescott, S., Ferraz Leal, R., Rozyczko, P., Snow, T. & Washington, A. (2021). SasView version 5.0.4 Zenodo.

Fletcher, R. (2000). Practical Methods of Optimization: Fletcher/Practical Methods of Optimization Chichester, West Sussex England: John Wiley & Sons, Ltd.

Grant, T. D. (2018). *Nat Methods* **15**, 191–193.

Hawkins, D. M. (2004). *J. Chem. Inf. Comput. Sci.* **44**, 1–12.

He, H., Liu, C. & Liu, H. (2020). *IScience* **23**, 100906.

Ilavsky, J. (2012). *J Appl Cryst* **45**, 324–328.

Ilavsky, J. & Jemian, P. R. (2009). *J Appl Cryst* **42**, 347–353.

Ingham, B., Li, H., Allen, E. L. & Toney, M. F. (2009).

Kikhney, A. G. & Svergun, D. I. (2015). *FEBS Letters* **589**, 2570–2577.

Leng, K., King, S., Snow, T., Rogers, S., Markvardsen, A., Maheswaran, S. & Thiyagalingam, J. (2022). *J Appl Cryst* **55**, 966–977.

Li, T., Senesi, A. J. & Lee, B. (2016). *Chem. Rev.* **116**, 11128–11180.

Mantzaris, N. V. (2005). *Chemical Engineering Science* **60**, 4749–4770.

Mittelbach, R. & Glatter, O. (1998). *Journal of Applied Crystallography* **31**, 600–608.

Ozerin, A. N., Muzafarov, A. M., Ozerina, L. A., Zavorotnyuk, D. S., Meshkov, I. B., Pavlova-Verevkina, O. B. & Beshenko, M. A. (2006). *Dokl Chem* **411**, 202–205.

Peng, S., McMahon, J. M., Schatz, G. C., Gray, S. K. & Sun, Y. (2010). *Proceedings of the National Academy of Sciences* **107**, 14530–14534.

Petoukhov, M. V., Konarev, P. V., Kikhney, A. G. & Svergun, D. I. (2007). *J Appl Cryst* **40**, s223–s228.

Polte, J., Ahner, T. T., Delissen, F., Sokolov, S., Emmerling, F., Thünemann, A. F. & Kraehnert, R. (2010). *J. Am. Chem. Soc.* **132**, 1296–1301.

Semenyuk, A. V. & Svergun, D. I. (1991). *Journal of Applied Crystallography* **24**, 537–540.

Sun, Y., Zuo, X., Sankaranarayanan, S. K. R. S., Peng, S., Narayanan, B. & Kamath, G. (2017). *Science* **356**, 303–307.

Svergun, D. I. (1992). *J Appl Cryst* **25**, 495–503.

Thünemann, A. F., Rolf, S., Knappe, P. & Weidner, S. (2009). *Anal. Chem.* **81**, 296–301.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F. & van Mulbregt, P. (2020). *Nat Methods* **17**, 261–272.

Supporting information

S1. Supplementary figures

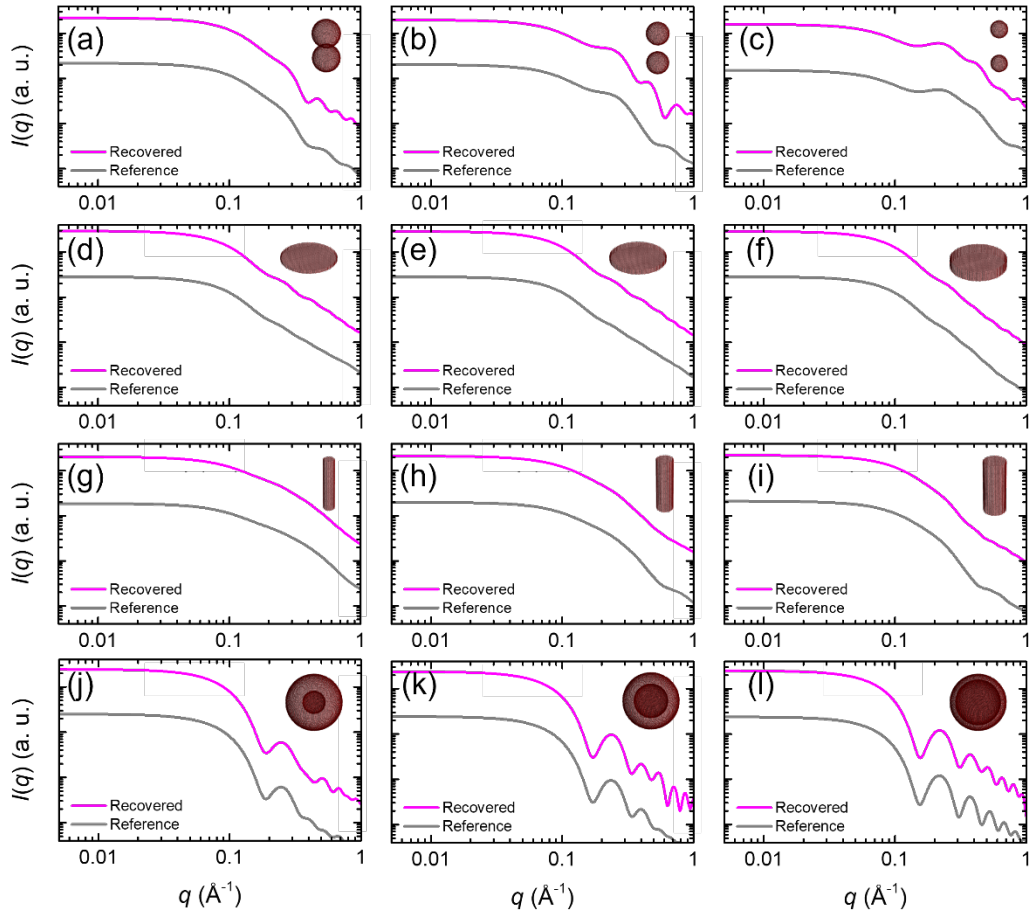

Figure S1 The corresponding SAXS intensity profiles calculated from the recovered PDDF profiles (pink curves) and reference PDDF profiles (gray curves) presented in Figure 5. The two SAXS curves in each panel were offset for easy visualization.

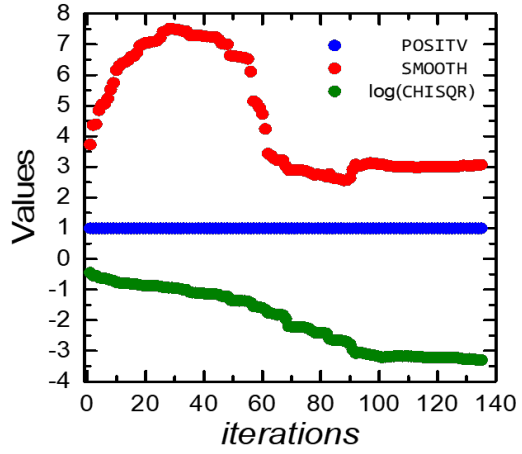

Figure S2 The evolution of values of each term in the goal function as a function of the number of iterations during the fitting procedure presented in Figure 6.

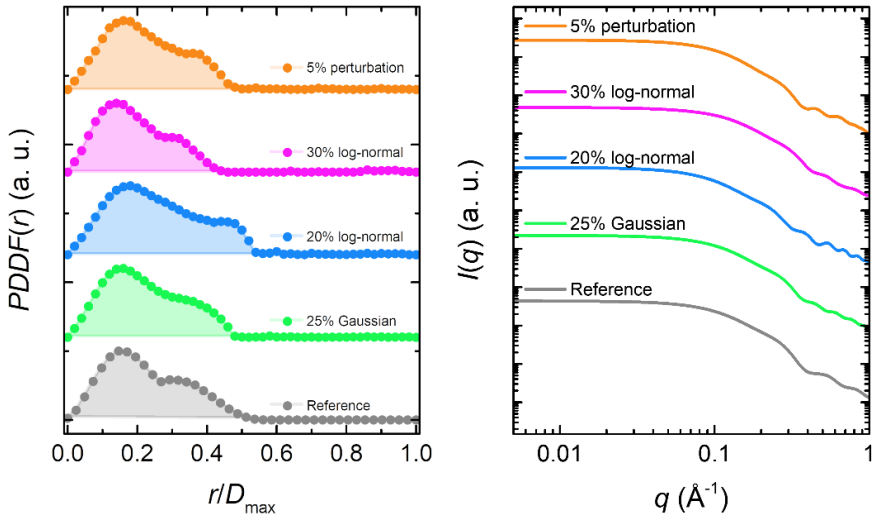

Figure S3 (a) The recovered $\hat{P}_0(r/D_{\max})$ ($D_{\max}$ = 80 Å) of dimer particle ensembles presented in Figure 5(a) at different falsely given conditions: orange - 5% perturbation (noise) added to $P_{ens}(r)$; pink - given a σ = 0.30 (~30%) log-normal size distribution; blue: given a σ = 0.20 (~20%) log-normal size distribution; green - given a σ = 0.25 (25%) Gaussian size distribution. Gray curve represents the reference curve calculated by the voxel model. (b) The corresponding SAXS intensity profiles calculated from the PDDF profiles presented in (a). The curves were offset for easy visualization.

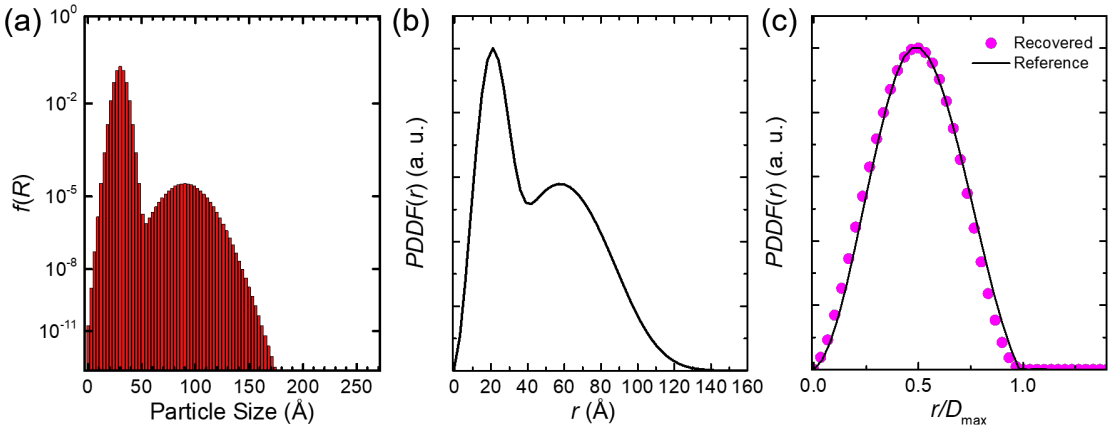

Figure S4 Refocusing the PDDF of spherical nanoparticle from a bimodal Gaussian size distribution. (a) Simulated size distribution consisting of two Gaussian distributions, with mean diameters of 30 Å and 90 Å, both exhibiting a standard deviation of 15%. The ratio of the respective particle numbers was set to 1:1000 to ensure comparable contributions of these two groups of nanoparticles to the PDDF. (b) The PDDF spherical nanoparticle ensembles with a bimodal Gaussian size distribution (c) The deconvoluted, size-refocused PDDF with the characteristic one-peak, symmetric feature of the spherical particle, compared to the reference.