JOURNAL OF
APPLIED
CRYSTALLOGRAPHY

Volume 54 (2021)

Supporting information for article:

Experimental noise in small-angle scattering can be assessed using the Bayesian indirect Fourier transformation

Andreas Haahr Larsen and Martin Cramer Pedersen

Supporting Information

SI.1. Examples of simulated data with $M = 50$ points

Figure S1 shows examples of simulated datasets with a detector binning the data into 50 bins. The data form the basis for Figures 2 and S3.

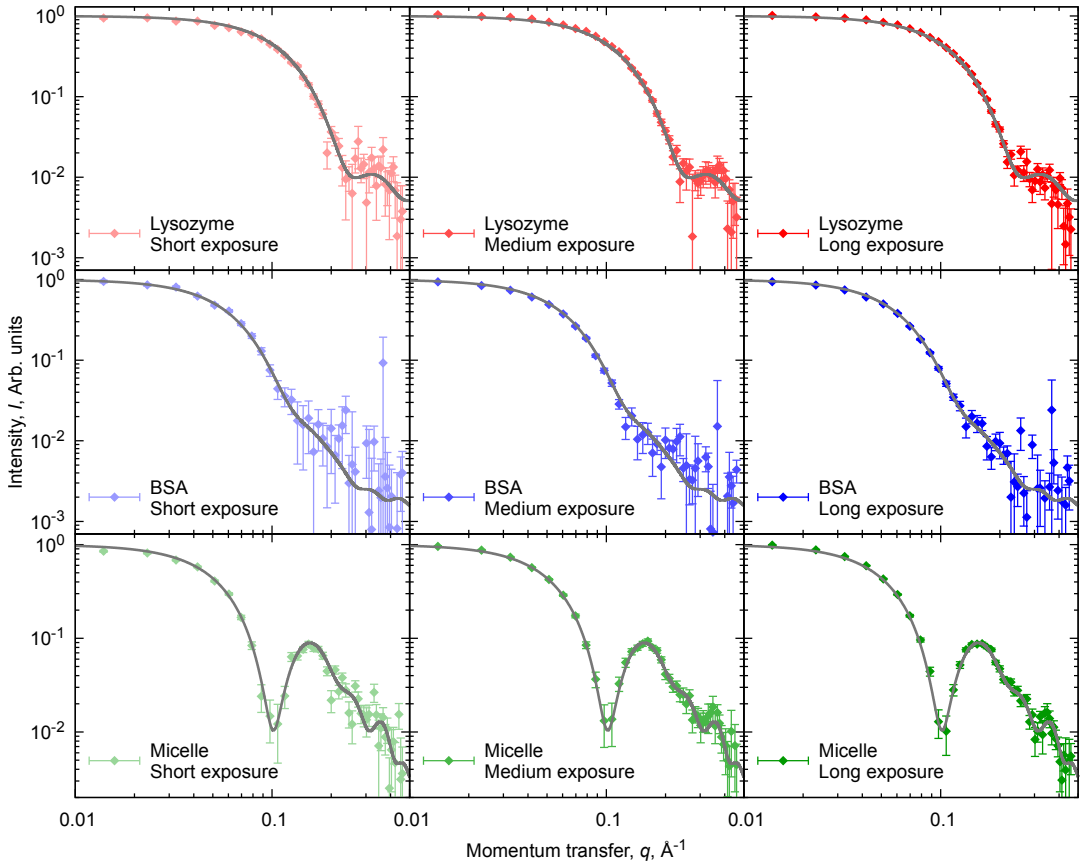

Fig. S1. Examples of simulated data with 50 datapoints.

SI.2. Full version of the correlation plots

The full version of the correlation plot in Figure 4 for the versions of our data with rescaled errors are shown in Figure S2.

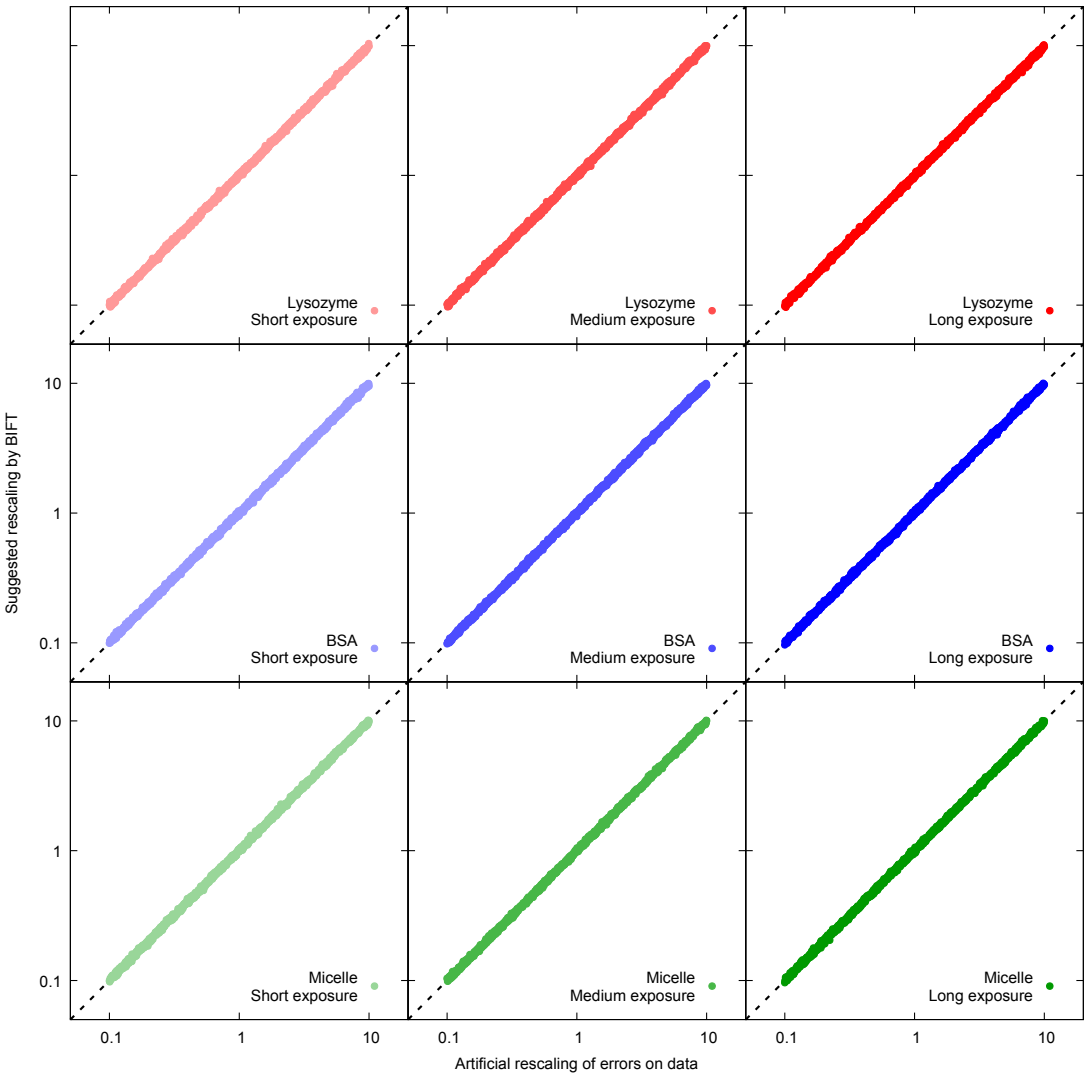

Fig. S2. Full version of Figure 4.

SI.3. Distributions of χ^2s for different choices of N_{DoF}

All distributions of χ^2 for our datasets with 50 datapoints for the BIFT algorithm and for the structural model with different choices for N_{DoF} are shown in Figure S3. Corresponding statistics can be found in Table S1.

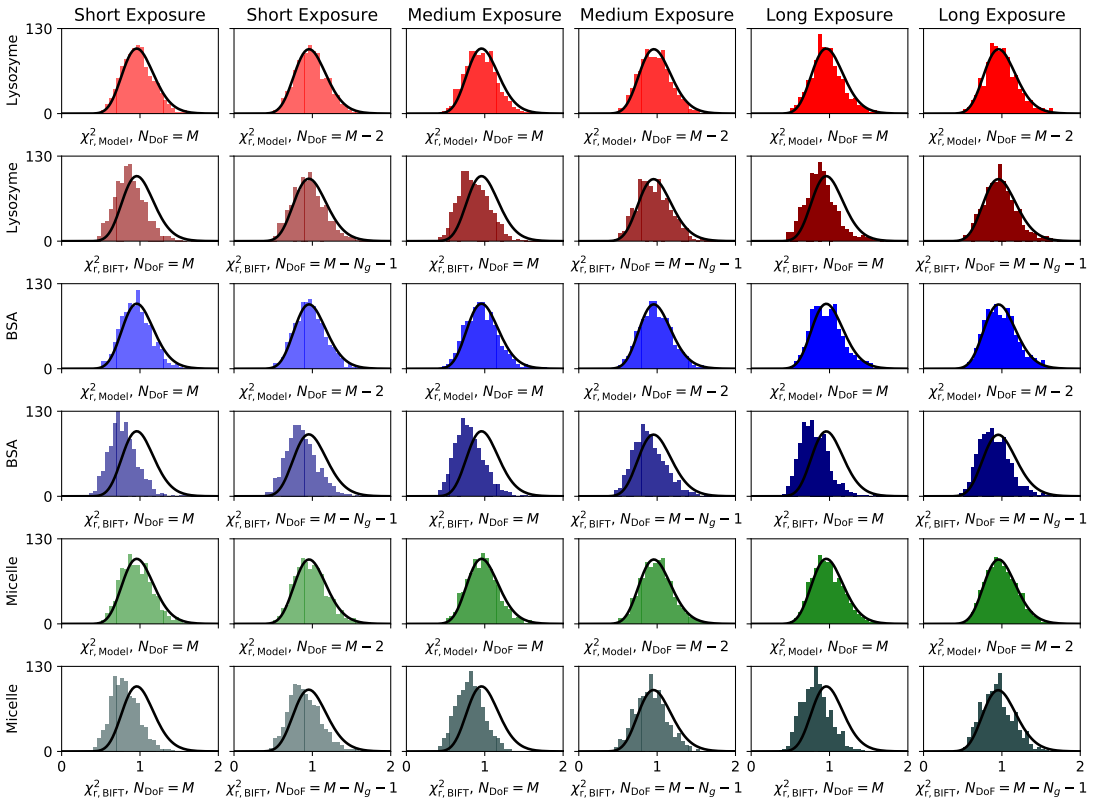

Fig. S3. Full version of Figure 2. Average values for each histogram given in Table S1.

Model	Exposure time	Choice of χ_r^2			
		$\frac{\chi^2_{\text{Model}}}{M}$	$\frac{\chi^2_{\text{Model}}}{M-2}$	$\frac{\chi^2_{\text{BIFT}}}{M}$	$\frac{\chi^2_{\text{BIFT}}}{M-N_g-1}$
Lysozyme	Short	0.95	0.99	0.87	0.95
-	Medium	0.93	0.97	0.86	0.95
-	Long	0.95	0.99	0.89	0.98
BSA	Short	0.93	0.97	0.78	0.87
-	Medium	0.94	0.98	0.80	0.90
-	Long	0.95	0.99	0.81	0.90
Micelle	Short	0.94	0.94	0.82	0.90
-	Medium	0.95	0.95	0.82	0.92
-	Long	0.95	0.95	0.83	0.94
Mean of average values		0.94	0.98	0.83	0.92
Standard deviation		0.01	0.01	0.03	0.03

Table S1. *Average of the* 1000 ${\chi_r}^2$ *values for each sample/exposure time. Histograms for the same values are displayed in Figure S3.* $M = 50$ *is the number of datapoints, and* N_g *is the number of "good parameters" from the BIFT algorithm.*

SI.4. Structural parameters for the presented SAS fits

The parameters refined from the fits in Figure 5 can be found in Table S2.

	SAXS only			SAXS and SANS			SANS only		
	Original	Rescaled Constant	Rescaled q-dep.	Original	Rescaled Constant	Rescaled q-dep.	Original	Rescaled Constant	Rescaled q-dep.
χ^2_r	1.4	7.2	1.4	1.8	1.3	1.4	0.08	0.94	0.94
Structural parameters									
ϵ	1.4 ± 0.2	1.4 ± 0.3	1.4 ± 0.3	1.4 ± 0.3	1.4 ± 0.5	1.3 ± 0.2	1.5 ± 2.0	1.5 ± 0.5	1.4 ± 0.6
A, Å^2	59.4 ± 1.3	59.5 ± 2.9	59.5 ± 2.9	59.4 ± 2.1	58.9 ± 2.7	59 ± 1	67 ± 222	70 ± 67	70 ± 64
N	149 ± 12	149 ± 27	149 ± 27	152 ± 10	153 ± 5	152 ± 4	139 ± 499	143 ± 205	143 ± 193
$R_{g,tag}$, Å	12 ± 7	12 ± 16	12 ± 16	10 ± 11	11 ± 17	9 ± 11	43 ± 1015	29 ± 339	30 ± 341
V_{MSP}, 10^3 Å^3	26.5 ± 0.6	26.5 ± 1.3	26.5 ± 1.3	26.6 ± 0.8	26.6 ± 1.0	26.7 ± 0.6	24 ± 175	19 ± 64	19 ± 42
V_{lip} Å^3	1006 ± 7	1006 ± 16	1006 ± 15	1005 ± 9	1005 ± 12	1004 ± 5	1143 ± 3513	1186 ± 1021	1191 ± 1187
Contrast-specific parameters									
σ_X, Å	4.1 ± 0.4	4.1 ± 0.9	4.1 ± 0.9	4.2 ± 0.6	4.3 ± 1.2	4.3 ± 0.6	–	–	—
B_X, 10^{-4} cm^{-1}	2 ± 1	2 ± 3	2 ± 3	2 ± 3	2 ± 6	2 ± 4	–	–	—
σ_N, Å	–	–	—	4.7 ± 2.9	4.8 ± 1.3	4.9 ± 1.1	5.0 ± 7.3	5.1 ± 6.9	5.1 ± 6.4
B_{100}, 10^{-4} cm^{-1}	–	–	—	9 ± 4	9 ± 1	85 ± 8	9 ± 3	9 ± 1	85 ± 7
B_0, 10^{-4} cm^{-1}	–	–	—	250 ± 188	252 ± 43	251 ± 32	249 ± 162	249 ± 38	249 ± 28

Table S2. *Parameters refined during the fitting of a phospholipid nanodisc model to data in Figure 5 before rescaling and after rescaling the errors using BIFT. The model is described in the literature (Skar-Gislinge* et al., *2010; Skar-Gislinge & Arleth, 2011). The nanodisc model was refined from, respectively, SAXS alone, SAXS and SANS together, or SANS data alone (including SANS samples with* 42% *D_2O and* 100% *D_2O in the buffer). ϵ: axis ratio of bilayer, A: area per lipid headgroup, N: number of lipids per nanodisc, $R_{g,tag}$: radius of gyration of histidine tag, V_{MSP}: volume of membrane scaffolding protein, V_{lip}: volume of phospholipid, DLPC, σ_X: SAXS Roughness, B_X: SAXS background, σ_{SANS}: SANS Roughness, B_{100}: SANS (*100% *D_2O) background, B_0: SANS (*0% *D_2O) background.*

SI.5. Details on the probability of $\chi^2_{\mathrm{r,BIFT}}$

We use the probability of a given value of $\chi^2_{\mathrm{r,BIFT}}$ to assess whether the experimental errors in a given dataset are appropriate. This is the probability for getting a particular $\chi^2_{r,\mathrm{BIFT}}$ or any value more extreme, which can be compute as:

$$P(\chi_r^2) = \begin{cases} 2\displaystyle\int_0^{\chi_r^2} \mathrm{d}\bar{\chi}_r^2\, p(\bar{\chi}_r^2) & \text{for } \chi_r^2 \leq \widetilde{\chi_r^2} \\ 2\displaystyle\int_{\chi_r^2}^{\infty} \mathrm{d}\bar{\chi}_r^2\, p(\bar{\chi}_r^2) & \text{for } \chi_r^2 > \widetilde{\chi_r^2} \end{cases} \tag{SI.1}$$

where $p(\bar{\chi}_r^2)$ is a probability density that depends on the variable $\bar{\chi}_r^2$. $\widetilde{\chi_r^2}$ is the median and can be approximated by:

$$\widetilde{\chi_r^2} \approx \left(1 - \frac{2}{9N_{\mathrm{DoF}}}\right)^3, \tag{SI.2}$$

where N_{DoF} is the degrees of freedom. The median is unity for large N_{DoF}. P is the two-tailed P-value for $\chi_{\mathrm{r}}{}^2{}_{,\mathrm{BIFT}}$ given the null hypothesis that the experimental errors in question are appropriate (Figure S4). The probability is unity when $\chi_r{}^2 = \widetilde{\chi}_r{}^2$, i.e. in most practical cases, when $\chi_r{}^2$ is unity, as all other values are more extreme.

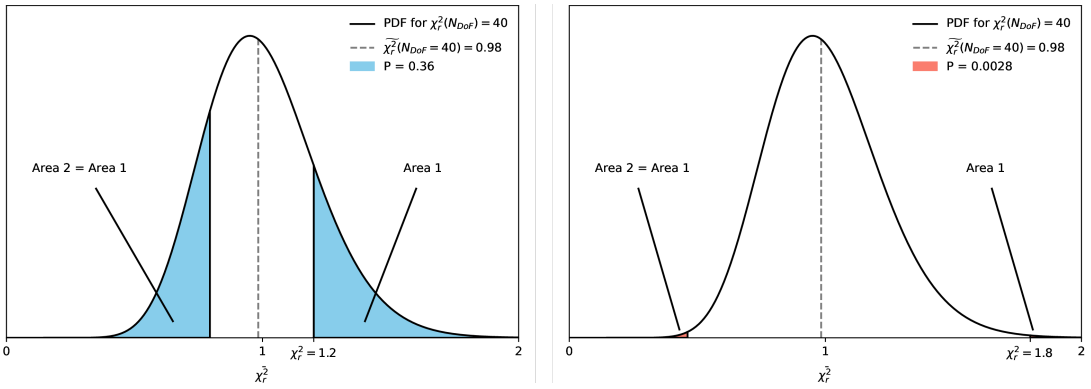

Fig. S4. Probability of a given $\chi_r{}^2$ value, illustrated for two different $\chi_r{}^2$ values and $N_{\mathrm{DoF}} = 40$. The probability equals the area under the graphs for all $\bar{\chi}_r^2 \geq \chi_r^2$ plus the same area from the left-side tail. On the left, $\chi_r^2 = 1.2$ gives $P = 0.36$ and is thus far above our suggested significance level of 0.003, so errors are probably correct. On the right, $\chi_r^2 = 1.8$ and $P = 0.0028$, i.e. below the significance level, so the errors are probably underestimated.

SI.6. Residual plots of BIFT fits

Figure S5 show the fits and residual plots of the data introduced in Figure 5. Of particular interest is the magnitude of the residuals in the various regions of the datasets, as this indicates the need for a q-dependent rescaling of the experimental errors, as done in panel C.

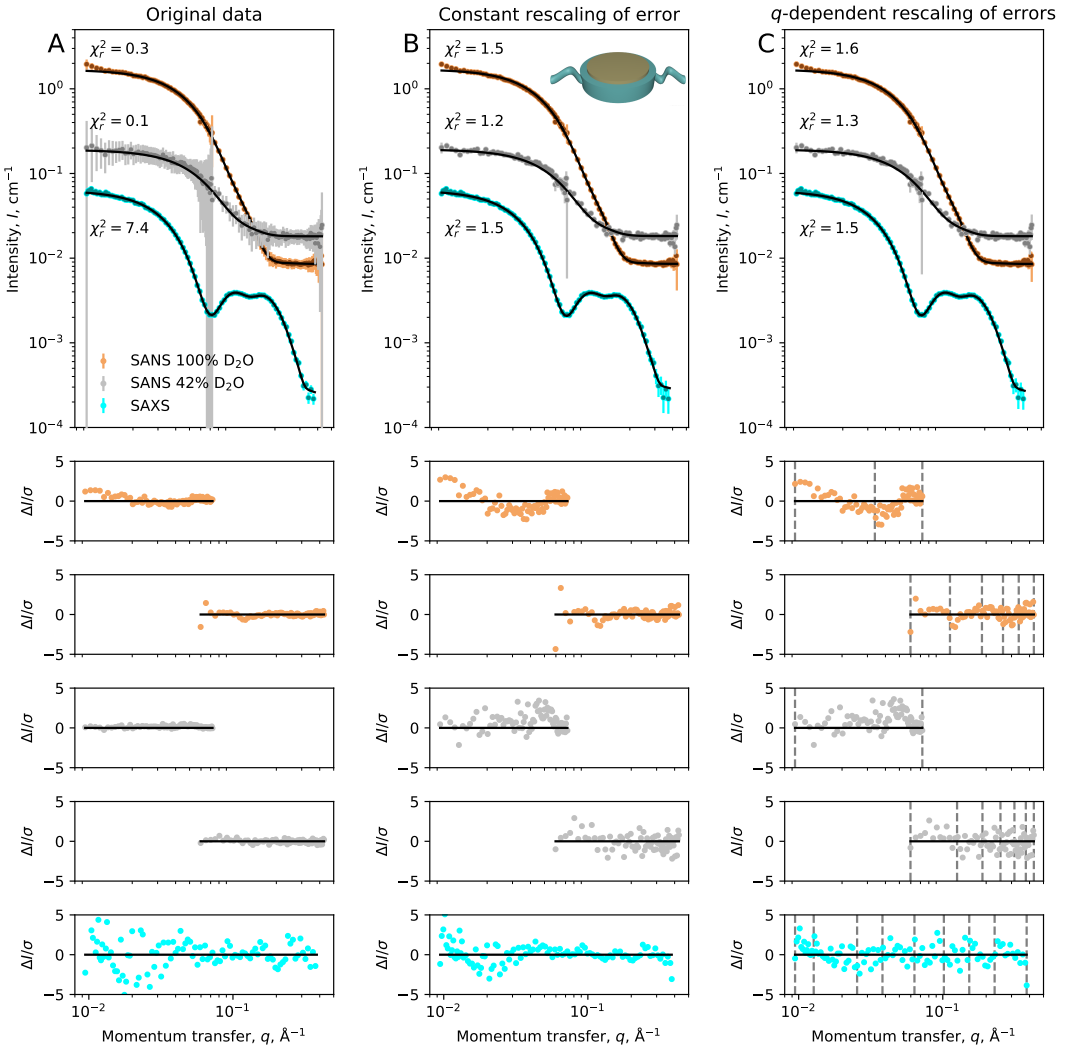

Fig. S5. Plots and normalized residuals for the BIFT fit of the example data (Figure 5), visualising q-dependency of the misestimated errors. The normalized residuals are expected to lie within a range of -3 and 3, assuming only statistical noise. ()A) Original data and model fits. (B) Data and fits after rescaling experimental errors with a constant. Inset shows the geomerical model: a protein/lipid nanodisc. (C) Data and fits after rescaling errors in with factors varying along q. Shannon bins (see main text) with minimum 10 datapoints in each bin are shown with vertical grey dashed lines in the residual plots.

SI.7. Simulations and tests of data with aggregation or hard-sphere potential tendencies

Examples of the data simulated for this part of the study are shown in Figures S6 and S7. The plots establishing the correlation between the output of the BIFT algorithm and the simulated noise levels can be found in Figure S8.

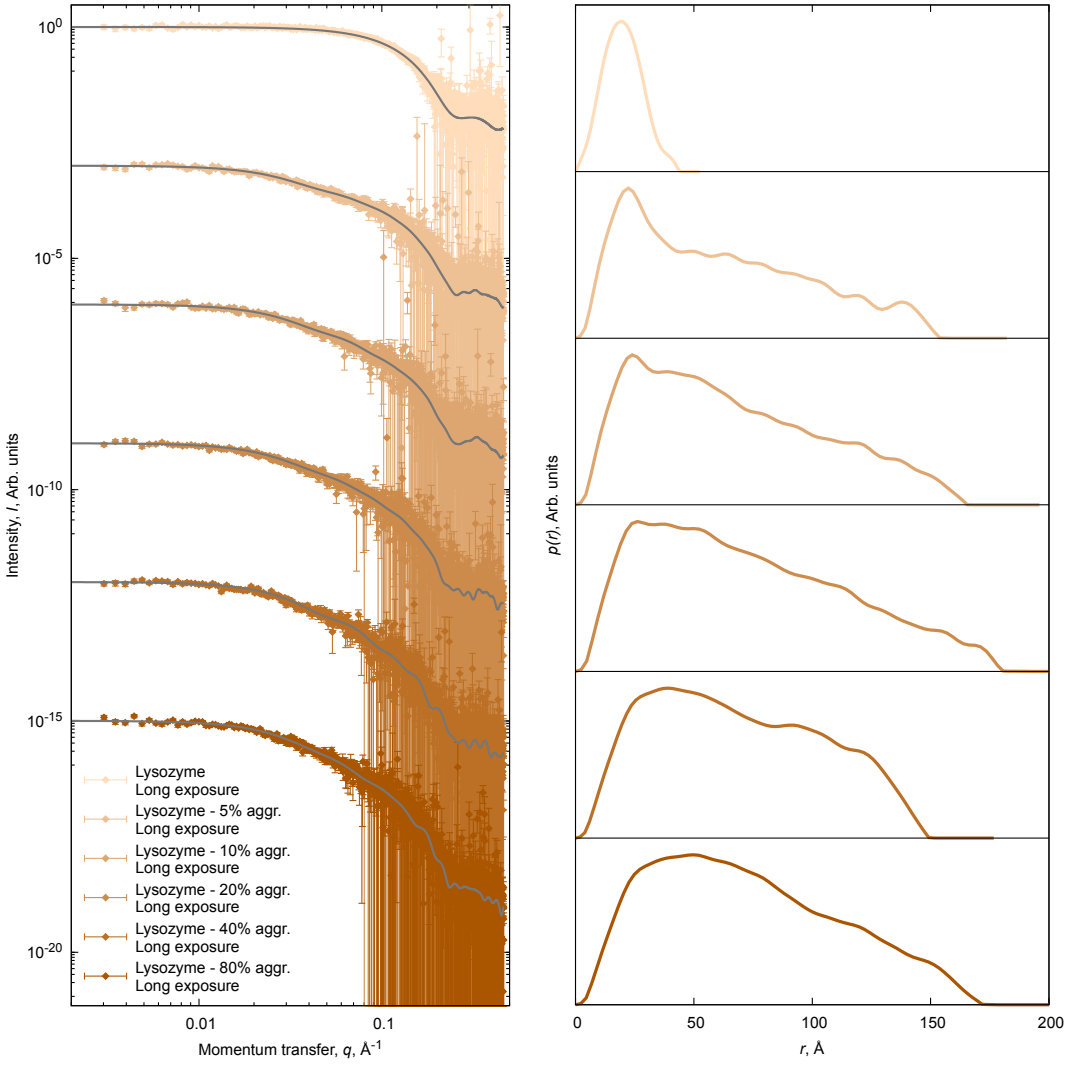

Fig. S6. Simulated data and corresponding $p(r)$ distributions for lysozyme with aggregation. The grey lines are the BIFT fits.

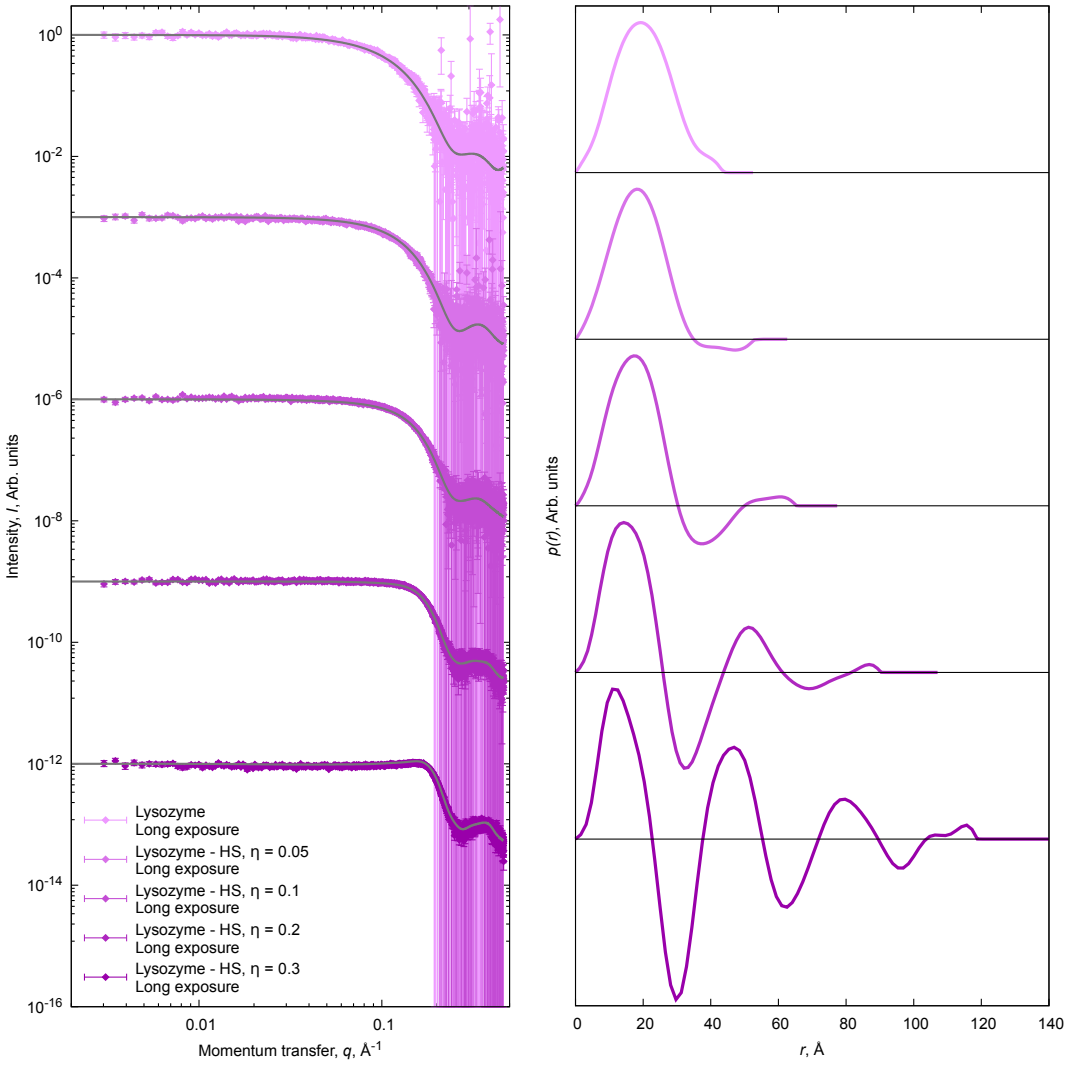

Fig. S7. Simulated data and corresponding $p(r)$ distributions for lysozyme with a hard-sphere structure factor. The grey lines are the BIFT fits.

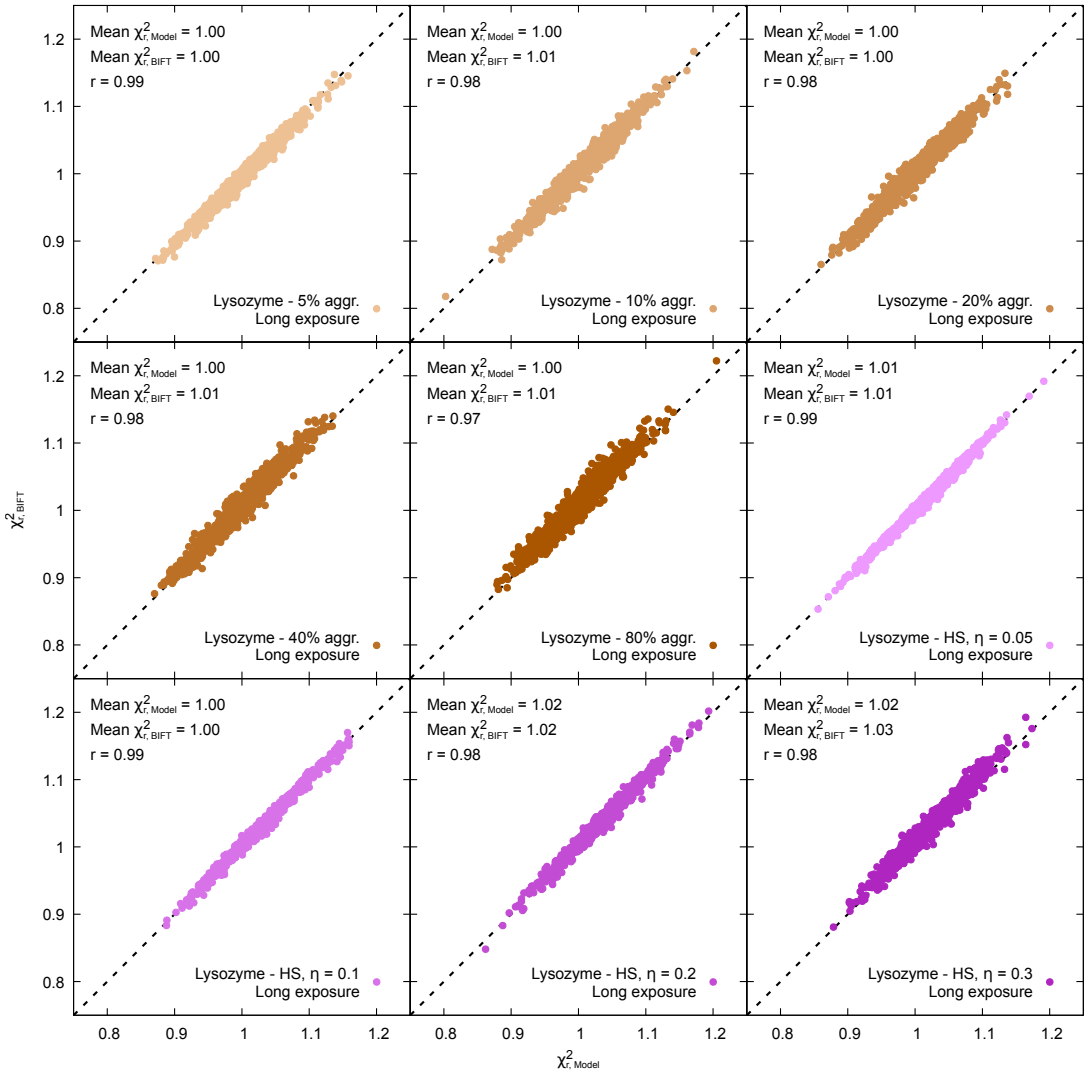

Fig. S8. Correlation plots for the data simulated with either aggregation or a contribution from a hard-sphere potential. As previously, Pearson coefficients, r, and means are printed in the individual plots.