Contrast Variation Small Angle Neutron Scattering-Identifying the Unique Fingerprints for the Structure of Proteins

Jeffrey J. Richards[1]

***Abstract*— This summer school module will use contrast variation SANS measurements to determine the contrast match point, contrast dependent radius of gyration, and the basic functions for two model protein in solution.**

I. INTRODUCTION

Proteins are a fundamental component of life. Understanding their function in living systems is critical to developing strategies to treat human disease and reveal the fundamental nature of biophysical processes. The function of a protein is directly tied to its structure. Decades of effort have therefore been dedicated to measuring the structure of proteins in their native environments. There are many methods to determine a protien's structure including X-ray diffraction, NMR, and electron microscopy.[1]

While the preferred structural determination technique is often electron microscopy, it has several limitations for proteins. First, the small length scales involved, the lack of phase contrast, and the projection of a three-dimensional object onto a two-dimensional image complicate the analysis of the distribution of components within a protein volume. Second, to obtain a representative average over the many billions of particles in a typical formulation, this complex analysis must be carried out over thousands of particles which is often impractical and the protein is subject to damage by the beam. Finally, while electron microscopy is constantly improving the ability to resolve structures in environmental cells, these capabilities do not yet rival bulk characterization techniques.

Neutron Scattering is well suited to the challenge of determining the pertinent structural features of proteins.[2] Neutron Scattering methods probe intrinsically orientationally averaged structure over many particles ($N_{part} \sim 10^{18}$) for typical sample volumes). Further, the phase contrast can be changed readily through isotopic substitution without influencing the chemical identity of the species involved, and multiple length scales can be probed simultaneously. Neutron scattering is also inherently an in-situ technique, so the structure can be probed without being disturbed. The challenge with neutron scattering techniques compared with imaging is the interpretation of the scattering profile resulting from the measurement.

*This document is for the teaching and instruction of the NCNR Summer School

*The author acknowledges the help of Thomas Cleveland and Susan Kreller for their fruitful discussions and advice in the measurements presented in this work.

[1]J.J. Richards is an NRC Postdoctoral Fellow at the NIST Center for Neutron Scattering.

Whereas this challenge has been elucidated for relatively simple scattering objects that have homogeneous distributions of scattering length density (SLD), the structural interrogation of inhomogeneous particles is significantly more challenging. Much of the work to date using SANS has focused on particles with an axisymmetric distribution of SLD.[3] There are a great number of examples of particles that fall into this category including surfactant micelles, nanoparticles decorated with polymer brushes, and vesicles.[4] The scattering profiles for these particles are frequently treated using infinitely thin shells with unique contrast terms constructed such that they reproduce the radial variation of scattering length density.[5] The overall structure can then be described by the form factor of each shell multiplied by the phase factor that accounts for their relative separation.[6] Key to the unique refinement of structure in this case is the measurement at multiple solvent contrasts. This is achieved frequently using mixtures of deuterated and hydrogenated solvent.

While this simple analytical approach is often successful when applied to axisymmetric particles, it is more challenging when the heterogeneities in SLD are eccentric or those with domains whose center of mass does not correspond to the particles center of mass (e.g. egg-yolk particles or Janus particles).[7]–[9] For such particles, limited analytical treatments exist to adaquately describe their scattering profiles, and thus model independent approaches have been primarily used. However, computation methods such as Monte Carlo simulation are quite successful.[10]

II. THEORY

A. Small Angle Scattering

When radiation is incident on a sample, the momentum of the incident wave vector can be expressed as $|k_i| = 2\pi/\lambda$, where λ is the wavelength. The incident wave induces in a finite volume element, δV, within a material a resonant dipole that emits waves in all directions.[ref] For elastic scattering processes, there is little energy transfer from the incident radiation to the sample ($E_i = E_s$) and the fraction of scattered radiation is small relative to the portion that transmits. Under these conditions, the momentum of the wave-vector emanating from the sample is $|k_f| = 2\pi/\lambda$.[11] For scattering measurements, it is typical to measure the scattering intensity as a function of the scattering angle, θ, which is commonly expressed as scattering vector, Q, where $Q = \frac{4\pi}{\lambda} sin(\theta/2)$. Elastic scattering provides the basis for a number of experimental techniques including small angle X-ray and neutron scattering (SAXS/SANS), X-ray Diffraction

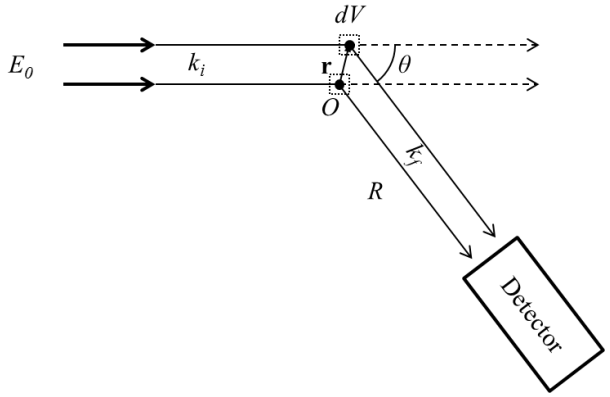

Fig. 1. Schematic representation of a small angle scattering experiment

(XRD), and grazing incidence small angle neutron and X-ray scattering (GISANS/GISAXS)

In order to understand this process in more detail, consider two differential volume elements within a sample, δV. Under the influence of an incident wave, scattered waves will emanate from each element in all directions. One is centered at the origin, O, and another at a distance, $|r|$, as shown in Figure 1. The scattered intensity, I_s, is measured as a function of angle on a detector at a distance R from the sample, where $I_s(Q,t) = |E(Q,t)|^2$, the squared amplitude of scattered radiation. The scattered waves produced from each scattering element propagate with magnitude, k_s, and due to the finite distance between elements, are phase shifted by $-Q \cdot r$ radians. This leads to interference that results in characteristic fluctuations in scattered intensity measured at the detector as a function of Q. For structural interrogation, the time-averaged intensity that results from a measurement is the Fourier transform of the combined contributions from all of the isolated scattering elements with scattering length, b, within the sample as shown in 1.[4]

$$\langle I_s(Q)\rangle = \frac{E_0^2}{R^2}\langle \sum_{m=1}^{N}\sum_{n=1}^{N} b_m(Q)b_n(Q)e^{-jQ\cdot(r_m - r_n)}\rangle \quad (1)$$

While Equation 1 is completely general, it is not often practical to use. A more convenient descriptions can be used which assumes the scattering originates from N identical particles. In this way, the collective ensemble behavior is what is measured so now Equation 1 can be reformulated to Equation 2, which is known as the Debye Equation. In this equation, the scattering lengths are replaced with scattering length densities, ρ and the spatial positions now are represented by vectors.

$$I_s(Q) = \int\int \rho(r_m)\rho(r_n)\frac{sin(Q\cdot r_{mn})}{Q\cdot r_{mn}}dr_m dr_n \quad (2)$$

In equation 2, ρ is the scattering length density measured in a vacuum. To account for solvent solvent, the contrast is introduced via equation 3.

$$\Delta\rho = V^{-1}\int_V (\rho(r) - \rho_s)dV = \bar{\rho} - \rho_s, \quad (3)$$

where $\bar{\rho}$ is the volume weighted average scattering length density of the object and ρ_s is the scattering length density of the solvent.

B. Scattering from Heterogeneous Particles

Within the context of a small angle scattering experiment, the scattered intensity collected on a detector will represent the ensemble average over its entire illuminated volume. In fact, regardless of shape or distribution of density fluctuations within the sample, $I(Q)$ will be proportional to ϕ_p, the volume fraction of scatters in the sample, V_{part}, the average volume of the scatterer, and $\Delta\rho^2 = (\bar{\rho} - \rho_s)$. The Q-dependence of the intensity originates from the particular details of the scattering entitys shape and the distribution of those domains within the sample. This Q-dependent component is commonly expressed as the product of the form factor, P(Q), and the structure factor S(Q), $I(Q) = P(Q)S(Q)$. Therefore, in general form, 4 approximates the scattering from a collection of identical objects.

$$I(Q) = \phi V_p \Delta\rho^2 P(Q)S(Q), \quad (4)$$

The form factor, $P(Q)$ is a function that is normalized such that $P(Q = 0) = 1$, and it contains all the details about the shape and internal density distribution of the particles. The structure factor encodes both dynamic and static interactions arising between particles. Decoupling the structure factor from the form factor is a critical component of the measurement of any sample of unknown structure. The easiest method to determine the relative contribution of $S(Q)$ to the scattering measurement is a dilution experiment where a concentration series is prepared and $I(Q)$ measured using a fixed contrast at each concentration. By dividing $\frac{I(Q)}{\phi_p}$, the role of structure factor can be identified by the condition $\lim_{\phi_p\to 0} S(Q) \to 1$.

For a homogeneous particle (i.e. one composed of only one scattering length density) and in the absence of a structure factor, in the small angle limit, $Q \to 0$, $\frac{sin(Q\cdot r_{mn})}{Q\cdot r_{mn}} = 1 - Q^2r^2/6 + Q^4r^4/120 + \ldots$ via the Mclaurin series expansion. Truncating to the quadratic term, it can be shown that

$$\lim_{Q\to 0} P(Q) = e^{-\frac{(QR_g)^2}{3}}. \quad (5)$$

This is Guinier approximation and R_g is known as the radius of gyration, which for a homogeneous particle is the root mean square distance of the mass of an object from its center of mass.

C. The Stuhrmann Plot

[12]

For homogeneous particles, the Guinier approximation is a convenient place to start for the analysis of scattering profiles. However, when a particle is composed of more than one component, it has less utility. In these cases, the shape function of the particle is convoluted with the

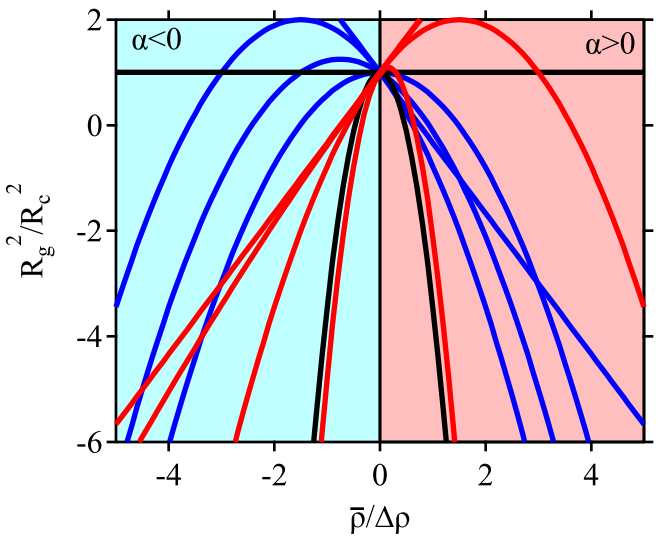

Fig. 2. Stuhrmann Plots for various α and β values. The blue traces are for $\alpha < 0$ and the red traces are for $\alpha > 0$. As β increases, the position of the peak maximum of the parabola shifts toward the ordinate. A homogeneous particle is described by the solid black line with a slope of 0.

scattering from heterogeneities within it. Stuhrmann was able to decouple these approximations by assuming that the solvent penetration did not influence the structure via 6,

$$\phi_p(r) = \begin{cases} 1, & \text{r is inside particle} \\ 0, & \text{r is outside particle} \end{cases} \tag{6}$$

Though it appears a trivial assumption, it can be shown that so defined, by measuring a suspension at a fixed concentration in solvents of varying scattering length density, ρ_s (eg., through isotopic substitution), the total scattering intensity for any inhomogeneous particle can be decomposed into separate terms as shown in Equation 7.

$$I(Q) = I_0 + 2\rho_s I_{01}(Q) + \rho_s^2 I_1(Q) \tag{7}$$

This equation implies a quadratic dependence of the total scattering intensity on solvent contrast at each Q-value. Further, it implies that any contrast variation experiment requires the measurement of at least three solvent contrasts in order to uniquely determine the scattering basic functions, $I_0(Q)$, $I_{01}(Q)$, and $I_1(Q)$. However, once these are determined any contrast can be predicted simply by applying Equation 7. Additionally, these basic functions can be fit independently to reconstruct the contributions to the measured curve, where $I_0(Q)$ is the shape function of the inhomogeneities, $I_1(Q)$ is the envelope scattering and $I_{01}(Q)$ is the cross-term.

Despite the power of this description, it often proves difficult to identify adequate analytical models for the basic functions. This is the case for many proteins. The complex interplay of the protein envelope with the peptide distribution complicates unique structural identification. Contrast variation in these cases can serve to separate those contributions.

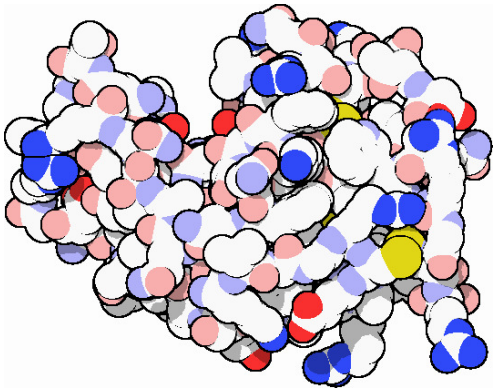

Fig. 3. PDB rendering of Lysozyme *(image reproduced from http://pdb101.rcsb.org/motm/9)*

In the absence of an adequate analytical description or simulation with which to compare the basic functions, Stuhrmann extended the Guinier approximation to account for the contrast dependence of the radius of gyration extracted from a fit to a scattering curve. Stuhrmann's equation, Equation 8, describes this dependence as a quadric dependence on the inverse contrast, $1/\Delta\rho$.

$$R_g^2 = R_c^2 + \alpha/\Delta\rho - \beta/\Delta\rho^2 \tag{8}$$

In Equation 8, R_c is the radius of gyration of the particle at infinite contrast (i.e. if it were homogeneous), α describes the relative scattering length density distributed radially from the particles center of mass, and β is a measure of the distance of the center of mass of the particle to the center of scattering length density of its heterogeneous components. In effect, the overall shape of the particle is described by R_c and the distribution of inhomogeneities by α and β. The result of this analysis showed that a plot of R_g^2 as a function of $1/\Delta\rho$ is a unique fingerprint that can be used to identify materials by their average distribution of internal inhomegeneities. Representative Stuhrmann plots are shown in Figure 2. Note that because of the negative sign in front of β, Stuhrmann plots should always be concave down.

III. Experimental Planning

There are a number of things necessary to perform a good contrast variation experiment on proteins and other heterogeneous scatterers. We will be performing these calculations before we go in the lab to prepare the samples. We will perform these calculations for the protein lysozyme. Lysozyme is a antimicrobrial enzyme produced by animals. It catalyzes the hydrolysis of glycosidic bonds in peptidoglycans. This hydrolysis reaction is mediated by Lysozyme's unique shape as shown in Figure 3. The C-shape of Lysozyme not only makes it an interesting protein for its biological function, but also the scattering patterns derived from it are feature rich due to the hydrophobic amino acid residues that reside in its core.

A. Concentration of Protein

First, a concentration must be identified where the structure factor, S(Q), is absent. This concentration ideally is in the dilute limit, but must be balanced against the need for coherent scattering intensity. This concentration can be found using the procedure outlined in the theory section by identifying the concentration when $S(Q) \rightarrow 1$. We performed this experiment for Lysozyme for concentrations spanning 5-50 mg/mL. By extrapolating the scattering curves to infinite dilution, P(Q) was recovered. Each concentration was then normalized using the curve at infinite dilution and from that $S(Q)$ was determined and is shown in Figure 4. The inset shows the intercept of the structure factor at $Q = 0$ versus concentration and from this plot it is clear that $S(Q) = 1$ for concentrations smaller than 5 mg/mL at this ionic strength. In order to obtain adequate statistics, therefore, we need to measure the samples at the largest concentration where the structure factor is minimized or absent.

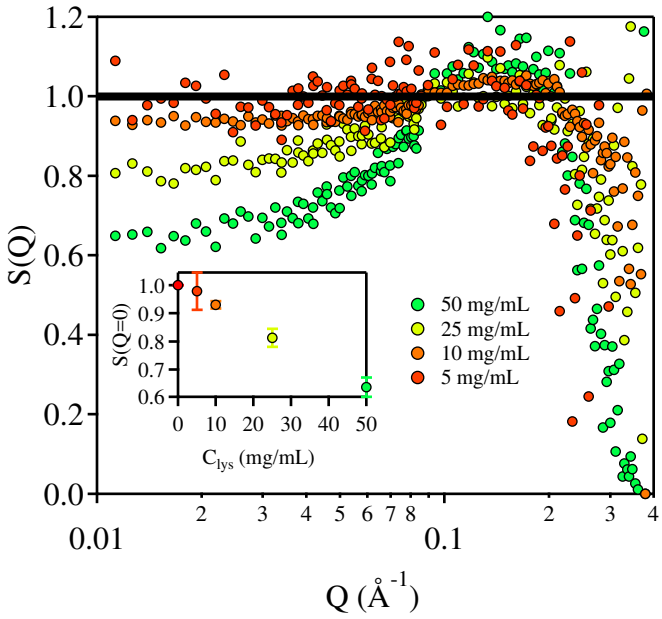

Fig. 4. Static Structure Factor, $S(Q)$ from HEW Lysozyme measured at several concentrations. $S(Q)$ is calculated through normalization to $lim_{\phi \rightarrow 0} I_s(Q)/\phi$. The inset shows $S(Q \rightarrow 0)$ vs concentration of protein in mg/mL. The structure factor approaches 1 at 5 mg/mL.

B. Subtracting the Background

The second aspect of a contrast variation experiments that makes them so challenging is that one needs to measure scattering curves near the match point in order to see the dependence of R_g on $\Delta\rho$. This implies via Equation 4 that the scattering intensity will be low and in the case of measuring at the match point $I(Q \rightarrow 0) = 0$. Because contrast variation for protiens is frequently performed for water based samples, this weak intensity is measured against the fact that the incoherent background increases with hydrogen

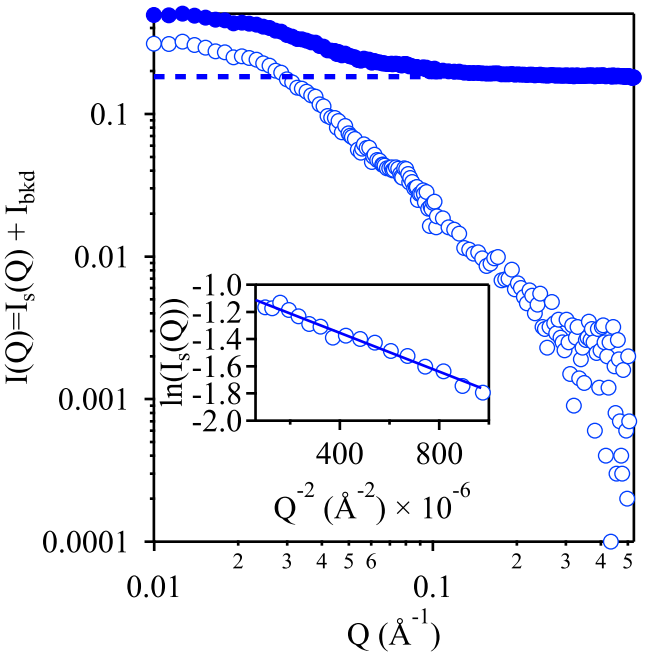

Fig. 5. $I(Q)$ vs Q for the NIST MAB dissolved in $\rho_s = 5.7 \times 10^{-6}$Å^{-2} (10 % H_2O) The incoherent background, I_{bkd}, is shown as a horizontal dashed line and the background subtracted sample intensity, $I_s(Q)$ is shown as well. Within the inset is a so-called Guinier plot which linearizes the data. From the slope on a Guinier plot, the R_g can be determined directly

content. In D_2O, this background is 0.05 cm^{-1}, whereas in H_2O it can reach 1.5 cm^{-1}. Given the volume of many proteins, the scattering will be dominated by the incoherent background at most contrasts. The incoherent scattering, I_{inc} is Q-independent and therefore can be simply subtracted directly from the measured $I(Q)$ curve to recover $I_s(Q)$ as shown in Figure 5 for a 5 mg/mL monoclonal antibody in 90% D_2O.

C. Selection of Contrasts and Number of Points

Third, the choice of contrast is also important in order to construct a Stuhrmann plot. We frequently have little *a priori* knowledge of the structure but can often estimate the protein's composition and density. Using these two parameters, it is possible to estimate $\bar{\rho}$. From this estimate, contrasts points should be selected for mixtures of D_2O/H_2O that span a broad range of contrasts so that the shape of the Stuhrmann plot is resolvable. A useful formula for calculating the solvent scattering length density is Equation 9.

$$\rho_s = \sum_i \phi_i \rho_i \tag{9}$$

The minimum number of contrasts that should be measured in order to determine parameters underlying the Stuhrmann plot is 3. However, it is good practice to perform the measurement using many more with at least 3 contrasts on both sides of the match point.

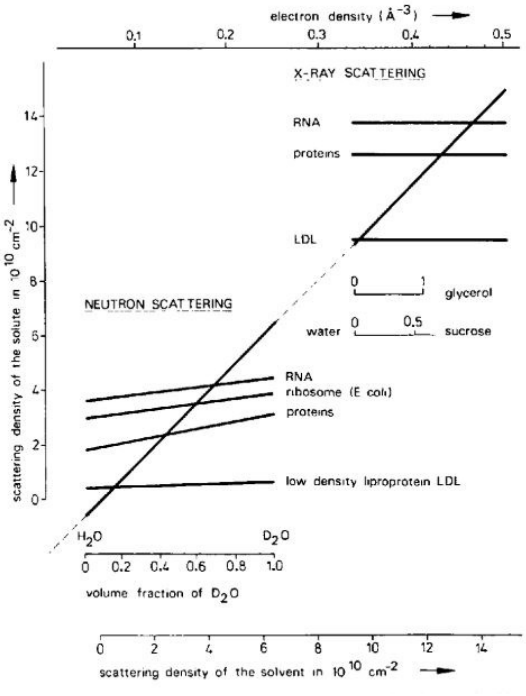

Fig. 6. (Figure reproduced from[13]). Scattering densities of various solutes and solvents in X-ray and neutron scattering. The scattering density of solutes in D_2O/H_2O mixtures is slightly increasing due to H/D exchange. Note that a broader range of contrast can be achieved in neutron scattering.

IV. Example

With the experiment appropriately planned, the samples can be prepared and measured. The raw scattering intensity is then corrected for empty cell, blocked beam and plotted on absolute scale. The incoherent background is then subtracted and this allows for the unique determination of I_0. This is shown in Figure 5 where the corrected scattering curve and background subtracted scattering curve are shown vs Q.

Linearizing the data by plotting $\ln(I_s(Q))$ vs. Q^2 allows for the graphical determination of R_g^2 via the $slope = -R_g^2/3$. The intercept on this plot yields I_0. Performing this linearization for every contrast measured yields R_g and I_0 at every $\Delta\rho$. From this estimate, a further linearization, $\sqrt{I_0}$ vs. ρ_s, yields the average contrast $\bar{\rho}$, which for the NIST MAB is 2.3×10^{-6}Å^{-2} or $(58.5\%H_2O)$. Armed with this value, the Stuhrmann plot can be constructed as shown in Figure 2.

V. Conclusion

Appendix

Appendixes should appear before the acknowledgment.

Acknowledgment

The author acknowledges the contributions of Paul Butler in the preparation of this document and many fruitful discussions.

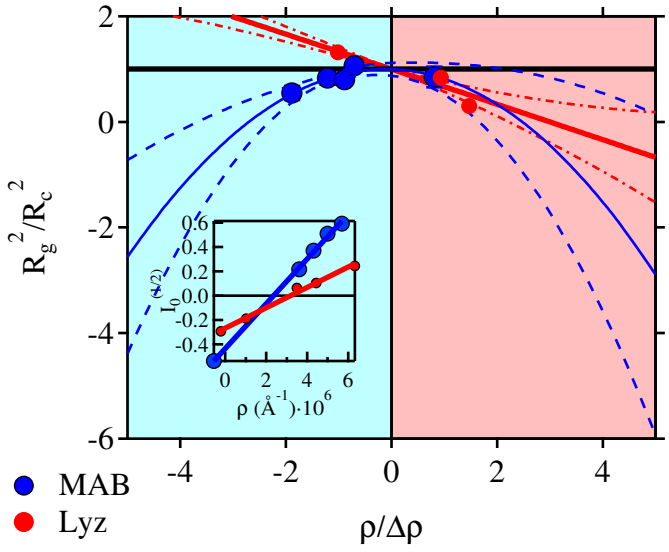

Fig. 7. Normalized Stuhrmann plot, $\frac{R_g^2}{R_c^2}$ vs. $\frac{\bar{\rho}}{\Delta\rho}$, measured for the NIST MAB and Lysozyme at five different contrasts. Best fit to Equation 8 shown as solid lines and 67 % confidence interval as dashed lines. The inset shows the$\sqrt{I_0}$ vs. ρ_s. The solid line shows x-intercept which signifies the contrast match point, $\bar{\rho}$

References

[1] J.?M. Wiencek. New Strategies for Protein Crystal Growth. *Annual Review of Biomedical Engineering*, 1(1):505–534, 8 1999.

[2] H B Stuhrmann. Neutron Small-Angle Scattering of Biological Macromolecules in Solution. *Journal of App*, 7(1):173–178, 1974.

[3] Jan Skov Pedersen. Analysis of small-angle scattering data from colloids and polymer solutions: modeling and least-squares fitting. *Advances in Colloid and Interface Science*, 70:171–210, 7 1997.

[4] O. Glatter and O. Kratky. *Small Angle X-Ray Scattering*. Academic Press Inc., 1982.

[5] Steven R. Kline. Reduction and analysis of SANS and USANS data using IGOR Pro. *Journal of Applied Crystallography*, 39(6):895–900, 11 2006.

[6] Mathieu; Doucet, Jae Hie; Cho, Gervaise; Alina, Jurrian; Bakker, Wim; Bouwman, Paul; Butler, Kieran; Campbell, Miguel; Gonzales, Richard; Heenan, Andrew; Jackson, Pavol; Juhas, Stephen; King, Paul; Kienzle, Jeff; Krzywon, Anders; Markvardsen, Torben; Nielsen, Lewis; O'Driscoll, Wojciech Potrzebowski, Ricardo; Ferraz Leal, Tobias; Richter, Piotr; Rozycko, and Adam Washington. SasView, 2017.

[7] H. Kaya. Scattering behaviour of Janus particles. *Applied Physics A: Materials Science and Processing*, 74(SUPPL.I):s507–s509, 12 2002.

[8] Tobias Fütterer, Gerrit A. Vliegenthart, and Peter R. Lang. Particle scattering factor of Janus micelles. *Macromolecules*, 37(22):8407–8413, 2004.

[9] Jeffrey J Richards, Curtis L Whittle, Guozheng Shao, and Lilo D Pozzo. Correlating structure and photocurrent for composite semiconducting nanoparticles with contrast variation small-angle neutron scattering and photoconductive atomic force microscopy. *ACS nano*, 8(5):4313–24, 5 2014.

[10] S J Henderson. Monte Carlo modeling of small-angle scattering data from non-interacting homogeneous and heterogeneous particles in solution. *Biophysical journal*, 70(4):1618–27, 4 1996.

[11] P Lindner and Th. Zemb. *Neutron, X-rays and Light. Scattering Methods Applied to Soft Condensed Matter*, volume 5. North-Holland, Amsterdam, 1st edition, 2002.

[12] L A Feigin and D I Svergun. *Structure Analysis by Small-Angle X-Ray and Neutron Scattering*. Springer US, Boston, MA, 1987.

[13] H. B. Stuhrmann and a. Miller. Small-angle scattering of biological structures. *Journal of Applied Crystallography*, 11(5):325–345, 10 1978.