

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

Jeffrey J. Richards<sup>1‡</sup>

June 21, 2018

## 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 Lysozyme.

## 1 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 protein's structure including X-ray diffraction, NMR, and electron microscopy.<sup>1</sup>

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 especially for looking at protein complexes. 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.<sup>2</sup> 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.

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.<sup>3,4</sup> Proteins fall into this category due to their chemical nature. They are composed of sequences of amino acids each of which has a distinct contrast for both neutrons and X-rays. Therefore, the total scattering pattern from a protein contains information not just about its shape envelop (i.e., the region of the protein inaccessible to solvent molecules) but also correlations between regions of varying contrast within its internal volume.<sup>5,6</sup>

As a result of these characteristics, more advanced analysis techniques must be employed which expand beyond analytical model descriptions to gain insight into relevant structural features of proteins. These advanced techniques are particularly powerful when combined with computation methods such as molecular simulations.<sup>7</sup> NIST has been involved in this work through the CCP-SAS Project.<sup>8</sup>

## 2 Theory

### 2.1 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  $\lambda$  is the wavelength. The incident wave induces in a finite volume element,  $dV$ , a resonant dipole that emits waves in all directions.<sup>9</sup> 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  $|k_f| = 2\pi/\lambda$ .<sup>9</sup> Therefore, the scattering process is defined by the change in angle,  $\theta$ , of the scattering to incident radiation. This change is commonly expressed as scattering vector,  $Q$ , with magnitude  $Q = \frac{4\pi}{\lambda} \sin(\theta/2)$ . Elastic scattering provides the basis for a number of experimental techniques including small angle X-ray and neutron scattering (SAXS/SANS), X-ray Diffraction (XRD), and grazing incidence small angle neutron and X-ray scattering (GISANS/GISAXS)

Fluctuations in scattering intensity arise due to constructive and destructive interference of emerging waves from the sample. This is easily understood by considering two volume elements within a sample emitting waves in all directions as in Figure 1. One is centered at the origin,  $O$ , and another at a distance,  $|r|$ . 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

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

†\*The author acknowledges the help of Thomas Cleveland and Susan Krueger for their fruitful discussions and advice in preparing the proteins and performing the measurements presented in this work.

‡<sup>1</sup>J.J. Richards is an NRC Postdoctoral Fellow at the NIST Center for Neutron Scattering.

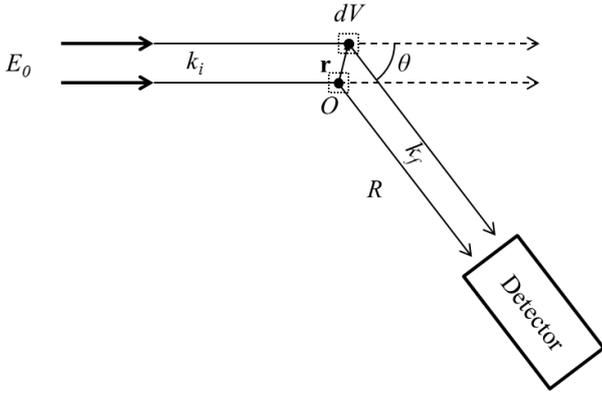


Figure 1: Schematic representation of a small angle scattering experiment

the finite distance separating the two volume elements, the outgoing waves are phase shifted by  $-Q \cdot r$  radians. This produces 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.<sup>5</sup>

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

While Equation 1 is completely general, it is not often practical to use for disordered scatterers. A more convenient description 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,  $\rho$  and the atomic positions are now 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,  $\rho$  is the scattering length density measured in a vacuum. To account for 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  $\rho_s$  is the scattering length density of the solvent.

## 2.2 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  $\phi_p$ , the volume fraction of scatters in the sample,  $V_{part}$ , the average volume of the scatterer, and the square contrast,  $\Delta\rho^2 = (\bar{\rho} - \rho_s)^2$ . The  $Q$ -dependence of the intensity originates from the particular

details of the scattering entity's 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  $S(Q)$ ,  $I(Q) \sim P(Q)S(Q)$ . Therefore, in general form, Equation 8 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  $\lim_{\phi_p \rightarrow 0} S(Q) \rightarrow 1$ .

With the form factor isolated from the structure factor, it is possible to analyze the scattering pattern to identify relevant pertinent structural features such as the size and shape of the particles. For a homogeneous particle (i.e. one composed of only one scattering length density) and in the absence of a structure factor, the Guinier approximation is often the starting point for structural analysis. In the small angle limit,  $Q \rightarrow 0$ ,  $\frac{\sin(Q \cdot r_{mn})}{Q \cdot r_{mn}} = 1 - Q^2 r^2 / 6 + Q^4 r^4 / 120 + \dots$  via the McLaurin series expansion. Truncating to the quadratic term, it can be shown that equation 2 reduces to:

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

This is the 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 and is generally valid under the condition of  $QR_g < 1$ .

## 2.3 The Stuhrmann Plot

For homogeneous particles, the Guinier approximation is a convenient place to start 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 scattering from heterogeneities within its internal volume.<sup>10</sup> Stuhrmann was able to decouple these approximations by assuming that the solvent penetration did not influence the structure via Equation 6,

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

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,  $\rho_s$  (eg., through isotopic substitution), the total scattering intensity for any inhomogeneous particle can be decomposed into separate terms as shown in reference,<sup>10</sup>

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

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. Though some progress can be made via spherical harmonic expansion. This is the case for many proteins. The complex interplay of the protein envelope with the distribution of amino acid residues complicates unique structural identification. Contrast variation in these cases can serve to separate those contributions.

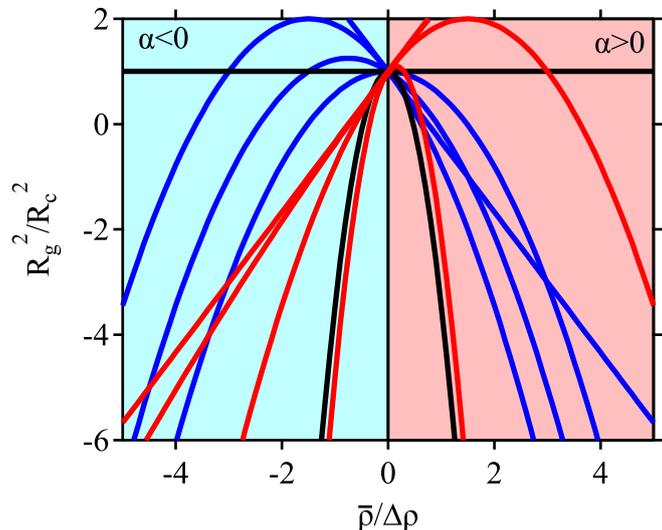


Figure 2: Stuhmann Plots for various  $\alpha$  and  $\beta$  values. The blue traces are for  $\alpha < 0$  and the red traces are for  $\alpha > 0$ . As  $\beta$  increases, the position of the peak maximum of the parabola shifts toward the ordinate. A homogeneous particle is described by the solid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

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

$$R_g^2 = R_c^2 + \alpha/\Delta\rho - \beta/\Delta\rho^2 \quad (8)$$

In Equation 8,  $R_c$  is the radius of gyration of the particle at infinite contrast (i.e. if it were homogeneous),  $\alpha$  describes the relative scattering length density distributed radially from the particles center of mass, and  $\beta$  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  $\alpha$  and  $\beta$ . 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 proteins by their average distribution of internal inhomogeneities. Representative Stuhmann plots are shown in Figure 2. Note that because of the negative sign in front of  $\beta$ , Stuhmann plots should always be concave down.

The general features of a Stuhmann plot are shown for a variety of different scenarios in Figure 2. It is particularly valuable to normalize both the x-axis and y-axis to the average values for comparison of different particles and proteins. The sign of  $\alpha$  determines the quadrant of the vertex of parabolic profile and the magnitude of  $\beta$  determines the latus rectum calculated as  $1/2\beta$ . In the normalized representation the vertex of the parabola can be located by  $(\frac{\alpha}{2\beta}, R_c^2 + \frac{\alpha^2}{2\beta})$ .

### 3 Experimental Planning

There are a number of things necessary to perform a good contrast variation experiment on proteins and other heterogeneous scatterers. We seek to generate a Stuhmann plot for Lysozyme. Lysozyme is a antimicrobial 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.

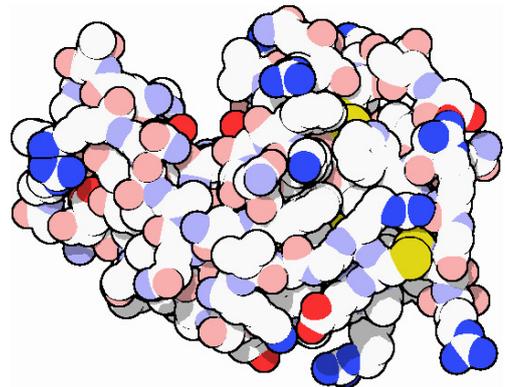


Figure 3: PDB rendering of Lysozyme (*image reproduced from <http://pdb101.rcsb.org/motm/9>*)

#### 3.1 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 and the incoherent background. 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.

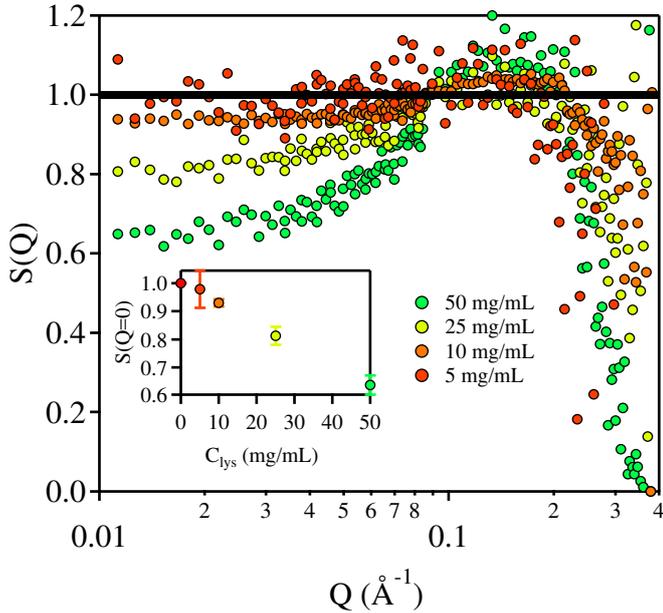


Figure 4: Static structure factor,  $S(Q)$  from hen egg white (HEW) Lysozyme measured at several concentrations.  $S(Q)$  is calculated through normalization to  $\lim_{\phi \rightarrow 0} I_s(Q)/\phi$ . The inset shows  $S(Q \rightarrow 0)$  vs concentration of protein in mg/mL. The structure factor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

### 3.2 Subtracting the Background

The second aspect of a contrast variation experiment 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 proteins is frequently performed for water based samples, this weak intensity is measured against the fact that the incoherent background increases with hydrogen content. In  $D_2O$ , this background is  $0.05 \text{ cm}^{-1}$ , whereas in  $H_2O$  it can reach  $1.1 \text{ cm}^{-1}$ . 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 protein sample in 90%  $D_2O$ .

### 3.3 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, contrast points should be selected for mixtures of  $D_2O/H_2O$  that span

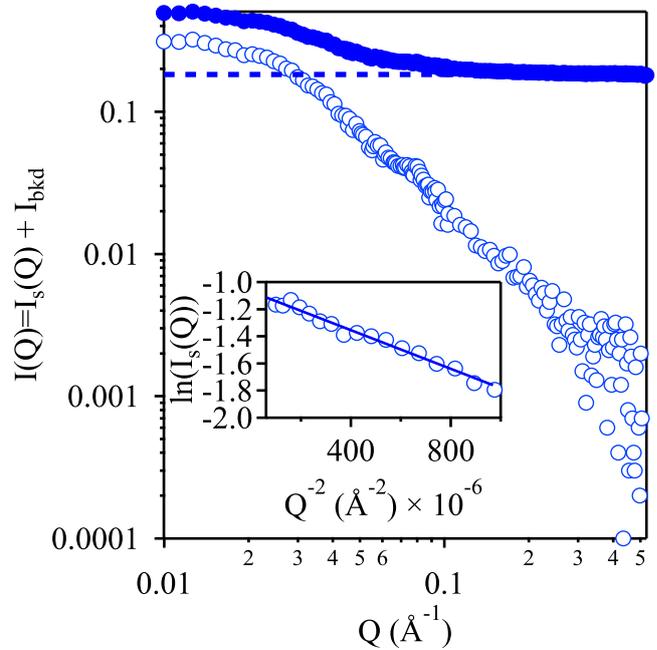


Figure 5:  $I(Q)$  vs  $Q$  for a protein dissolved in  $\rho_s = 5.7 \times 10^{-6} \text{ \AA}^{-2}$  (10 %  $H_2O$ ) The incoherent background,  $I_{bkd}$ , is shown as a horizontal dashed line and the background subtracted sample intensity,  $I_s(Q)$  is shown as well. Within the inset is a so-called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

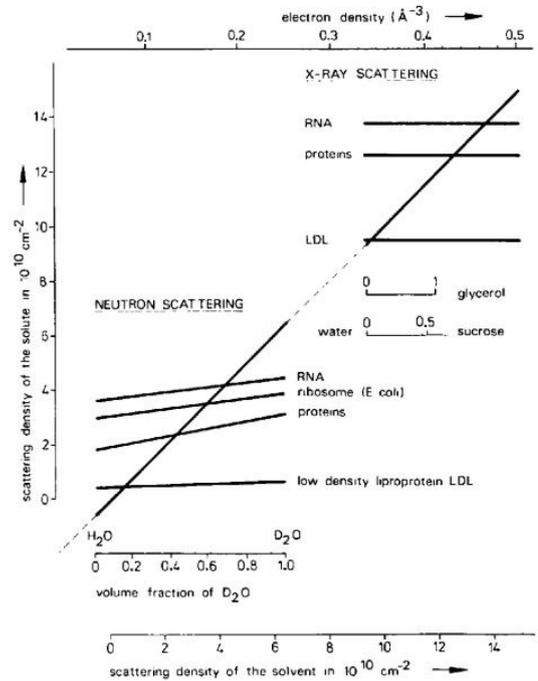


Figure 6: (Figure reproduced from<sup>11</sup>). Scattering densities of various solutes and solvents in X-ray and neutron scattering. The scattering density of solutes in  $D_2O/H_2O$  mixtures is slightly increasing due to H/D exchange. Note that a broader range of contrast can be achieved in neutron scattering.

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.

$$\rho_s = \sum_i \phi_i \rho_i \quad (9)$$

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.

### 3.4 Radius of Gyration

With the experiment appropriately planned, the samples can be prepared, loaded into the SANS instrument and scattering measurements can be performed. Once the scattering intensity is obtained at multiple detector configurations, the raw scattering intensity can be corrected for empty cell, blocked beam and plotted on absolute scale, a procedure which is performed by Steve Kline's Reduction Macros in IGOR Pro.<sup>6</sup> Once the scattering curves are in absolute scale, the procedure outlined in the above sections can then be applied. With the background subtracted and the data extrapolated to zero concentration, it is now possible to calculate the radius gyration of the samples at each contrast.

While there are many ways to perform this calculation, the classic method is through a linearization, where the  $I(Q)$  vs.  $Q$  plot is converted to a plot of  $\ln(I(Q))$  vs.  $Q^2$ . If data is collected at sufficiently low- $Q$  values, then this linearization will produce a straight line as shown in the inset in Figure 5. The origin of this linearization is in Equation 5 where the slope of this line is now  $-R_g^2/3$  and the intercept is  $\ln(I(0))$ . The validity of the Guinier approximation is maintained under the condition where  $QR_g < 1$ . This condition can be checked iteratively by selecting an initial  $Q$ -range, finding the slope, and reestablishing the  $Q$ -range until the results converge. The condition  $QR_g < 1.2$  can yield errors in excess of 20%.

## 4 Constructing the Stuhrmann Plot for Lysozyme

The construction of the Stuhrmann Plot requires a measure of the radius of gyration of the particle at infinite dilution measured in at least three different solvents. It then further requires that one identify the contrast match point. As discussed above, the intercept of the Guinier linearization yields  $I(0) = (\Delta\rho)^2 V_p$  for each contrast. From this estimate, a further linearization,  $\sqrt{I_0}$  vs.  $\rho_s$ , yields the contrast match point of the protein,  $\bar{\rho}$ .

Perform this linearization for every contrast. The results of this procedure are shown as  $R_g^2$  vs.  $1/\Delta\rho$  as shown in Figure 8. The experimentally determined contrast match point was found to be  $\bar{\rho} = 2.3 \times 10^{-6} \text{ \AA}^{-2}$  or 45%  $H_2O$ . Armed with this value, the Stuhrmann plot can be constructed as shown in Figure 8. The preliminary measurement shows excellent agreement with Stuhrmann's result for the envelope radius of gyration,  $R_c = 13.9$ . However, there are significant deviations in the apparent shapes of the curves with our preliminary measurement showing a much greater degree of eccentricity

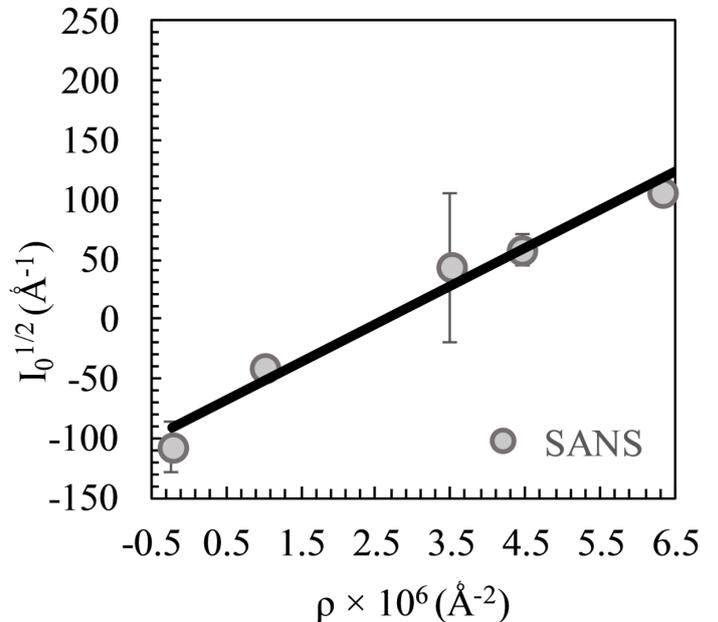


Figure 7: Contrast Match point determination,  $\sqrt{I_0}$  vs  $\rho_s$  measured for Lysozyme five different contrasts. The contrast match point is identified as the x-intercept.

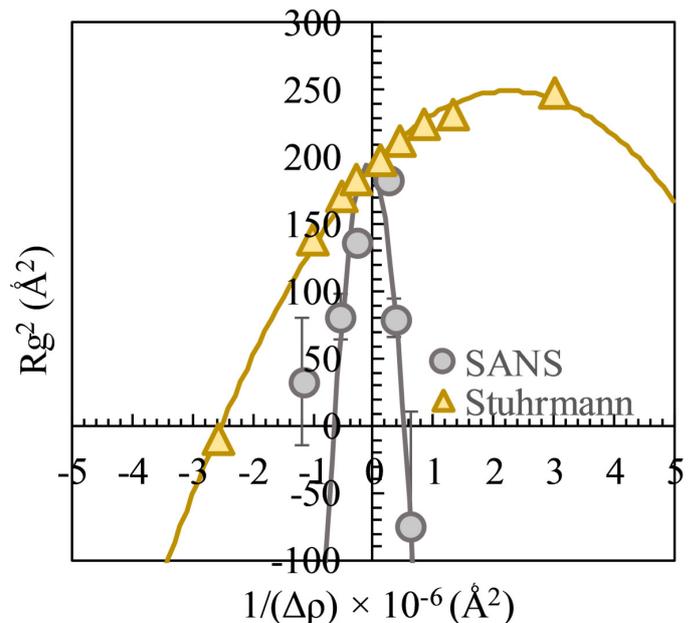


Figure 8: Normalized Stuhrmann plot,  $\frac{R_g^2}{R_c^2}$  vs.  $\frac{\bar{\rho}}{\Delta\rho}$ , measured for Lysozyme at five different contrasts. The yellow triangles are the values measured by Stuhrmann in 1976. Best fit to Equation 8 shown as solid lines.

as indicated by the more narrow parabolic shape of our recent measurement. These differences are also reflected in the fit parameters extracted from equation 8. The values are tabulated below

	SANS	Stuhrmann
$R_c$	13.9	13.9
$\alpha \times 10^6$	-0.76	4.92
$\beta \times 10^{12}$	56.0	1.09

## ACKNOWLEDGMENT

The author acknowledges the contributions of Paul Butler in the preparation of this document and many fruitful discussions. This work benefited from CCP-SAS software developed through a joint EPSRC (EP/K039121/1) and NSF (CHE-1265821) grant.

## 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] J. S. Pedersen, D. Posselt, and K. Mortensen. Analytical treatment of the resolution function for small-angle scattering. *Journal of Applied Crystallography*, 23(4):321–333, 8 1990.
- [4] 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.
- [5] O. Glatter and O. Kratky. *Small Angle X-Ray Scattering*. Academic Press Inc., 1982.
- [6] Steven R. Kline. Reduction and analysis of SANS and USANS data using IGOR Pro. *Journal of Applied Crystallography*, 39(6):895–900, 11 2006.
- [7] Steven C. Howell, Xiangyun Qiu, and Joseph E. Curtis. Monte Carlo simulation algorithm for B-DNA. *Journal of Computational Chemistry*, 37(29):2553–2563, 11 2016.
- [8] Paul D. Butler, Stephen Perkins, Joseph Curtis, and Steve King. CCP-SAS Project, 2018.
- [9] 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.
- [10] L A Feigin and D I Svergun. *Structure Analysis by Small-Angle X-Ray and Neutron Scattering*. Springer US, Boston, MA, 1987.
- [11] H. B. Stuhrmann and a. Miller. Small-angle scattering of biological structures. *Journal of Applied Crystallography*, 11(5):325–345, 10 1978.