Small Angle Neutron Scattering studies of Lysozyme solutions

Susan Krueger[1] and Susana Teixeira[1,2]

[1]NIST Center for Neutron Research, National Inst. Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899.
[2]Dep. Chem. Biomolecular Engineering, U. Delaware, 150 Academy Str., Newark, DE 19716.

Abstract

Biomolecules such as proteins or nucleic acids are polymers that sustain life, with unique structures and broad scale dynamics. Inherent to their complexity, and ability to function in a variety of physicochemical environments, is a sensitivity that often renders biomolecules challenging to study. Typical samples are (i) isolated in very limited amounts, (ii) stable for short periods of time and (iii) prone to aggregation or polydispersity. Small Angle Neutron Scattering studies of biomolecules (BioSANS) require careful planning. A well-known model protein will be used to provide an example of common pitfalls - even for proteins generally considered structurally robust such as lysozyme - while highlighting the tremendous potential of BioSANS. Students will simulate SANS curves to compare to experimental data collected for lysozyme solutions, where subtle variations in ionic strength were introduced. During the BioSANS course students will assess data collection strategies, learn how contrast variation can be used, reduce SANS data collected at the NCNR and fit it using form and structure factor models.

1. Introduction

The structure of lysozyme was the first ever solved for an enzyme [1]. Lysozymes are ubiquitous in mammalian tissues, and are present in phagocyte-like cells in nonmammalian organisms as well. There are three dominant families of lysozymes with different structure and activities: chicken-type (c-type), goose-type (g-type) and invertebrate-type (i-type). Generally, both c- and g-lysozymes are basic proteins due to their high isoelectric point (pI) values. c-Lysozyme is a single chain of 129 amino acids (14.3 kDa) that folds into a globular structure containing 4 α-helices and 5 β-strands, stabilized by 4 disulfide bridges. The enzyme is able to lyse peptidoglycan in the bacterial cell wall (see Figure 1), as well as act synergistically with other antimicrobial polypeptides.

Hen egg-white c-lysozyme (HEWL) is extensively used in food due to its abundant availability and properties. The electrostatic charge distribution of HEWL (calculated pI=9.28 [2]), with 9 negatively charged and 17 positively charged residues, imparts important binding affinities (see for example [3, 4]). There is extensive ongoing research towards improving the inhibitory effects of HEWL, relevant to its antimicrobial applications, namely through the formation of protein gels or complexes with other compounds (for a review see for example [5]). Further developments require systematic studies under different conditions environments (pressure, pH, salts, etc).

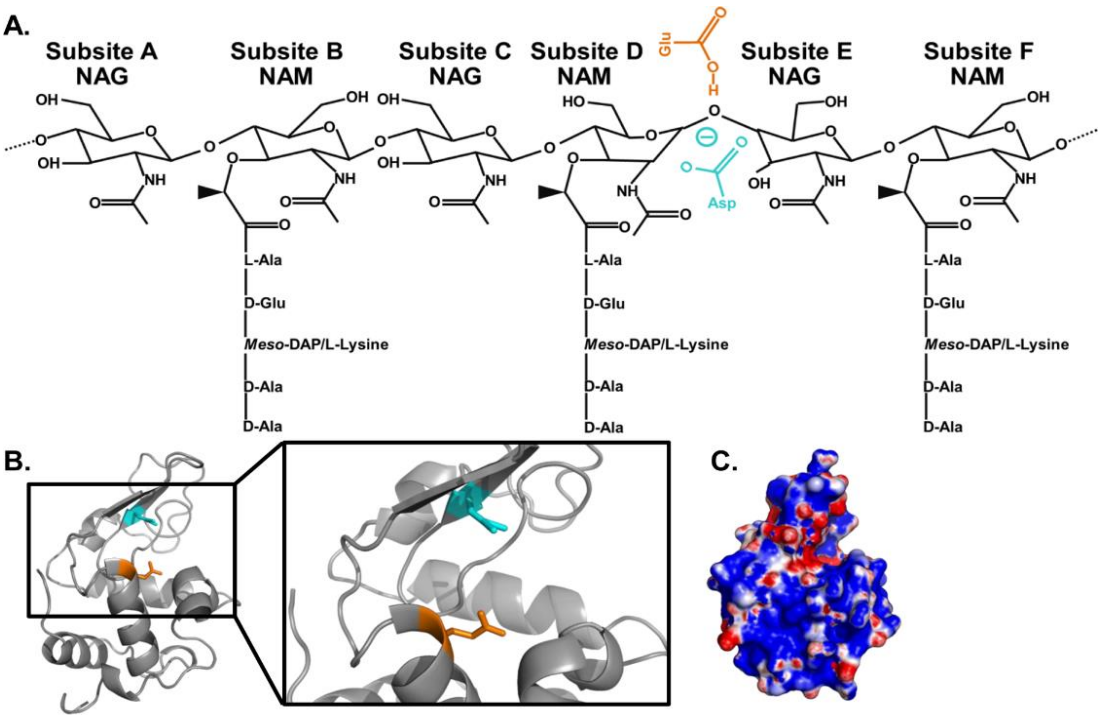

Figure 1. The enzymology and charge distribution of human lysozyme (PDB ID 1REX). **(A)** The active site accommodates up to 6 consecutive sugars annotated A-F. Lysozyme hydrolyzes the β-1,4 glycosidic bond between the subsite D and the subsite E. **(B)** Ribbon model of human lysozyme highlighting the essential active site residues in blue and orange. **(C)** Electrostatic potential map of human lysozyme, contoured at ± 5kT/e; blue, positive; red, negative; white, hydrophobic). Abbreviations: NAG, *N*-acetylglucosamine; NAM, *N*-acetylmuramic acid; PDB, Protein Data Bank. Image source: Ragland and Criss (2017) [6].

1.1. General SANS concepts

The solution environment sets the stage for protein stability and to the nature of intermolecular interactions present. Phase transitions are often caused by a combination of short-range attraction and long-range repulsive interactions, such as liquid-liquid phase transitions, where the solution separates into a protein-rich and a protein-deficient liquid phase [7]. Neutron scattering is particularly powerful to probe interactions because it covers a broad range of sizes: from ≈20Å to ≈20µm. The nondestructive nature of neutrons (i.e. no radiation damage) also facilitates in-situ measurements while various parameters are changed, e.g. ionic strength or temperature, or ligands are added. Furthermore, the penetrating power of neutron radiation allows for studies of samples contained in a variety of holders, such as pressure cells [8, 9] or chromatography columns [10], and a range of concentrations from a dilute to crowded regime [11]. SANS measures the elastic differential scattering cross section per solid angle (dΣ/dΩ) as a function of the magnitude of the momentum transfer (q):

$$\frac{d\Sigma}{d\Omega}(q) = n\,(\rho_{Protein} - \rho_{Solvent})^2 V_{protein}^2 P(q) S(q) + B \qquad (1)$$

where *n* is the number density of protein molecules, *ρ* is the neutron scattering length density (SLD), *P*(*q*) is the square of the form factor *F*(*q*), *S*(*q*) is the structure factor and B is the incoherent neutron

scattering background of the data. The momentum transfer depends on the scattering angle (2θ) and is defined as:

$$q = (4\pi/\lambda)\ sin\ \theta \qquad (2)$$

The form factor *F*(*q*) is affected by intraparticle interactions and describes the size and shape of the individual protein molecules. The structure factor *S*(*q*) is determined by interparticle interactions and contains information on the range and type of interactions. In dilute conditions, where the protein-protein interactions can be considered negligible, *S*(*q*) ≈ 1. For globular proteins, particularly when an atomic structural model is incomplete or unavailable, it is often possible to approximate the form factor with an ellipsoidal shape defined by equatorial (*a*) and polar radii (*b = c*):

$$P(q) = \int_0^1 F^2(q,\mu)\, d\mu \qquad (3)$$

$$F(q,\mu) = \frac{3(sin\ x\ - x\ cos\ x)}{x^3} \qquad (4)$$

$$x = q\sqrt{a^2\mu^2 + b^2(1-\mu^2)} \qquad (5)$$

$$V_{protein} = (4\pi/3)\, ab^2 \qquad (6)$$

where μ is the cosine of the angle between the directions of a and q. For an isotropic system, where interactions are uniform across the sample solution and do not depend on the orientation of the molecule, the structure factor can be written as:

$$S(q) = 1 + 4\pi n \int [g(r) - 1] \frac{\sin qr}{qr} r^2 dr \qquad (7)$$

where *g*(*r*) is the radial distribution function, governed by the interaction potential. *g*(*r*) is the probability of finding a protein molecule at a distance *r* from a reference particle centered at the origin.

1.2. Concentrated solutions of biological molecules

Solutions of biological molecules are often polydisperse and intermolecular interactions are affected by anisotropy in the solution. To account for these effects, equation (1) can be rewritten as:

$$\frac{d\Sigma}{d\Omega}(q) = n\,\Delta\rho^2 V_{protein}^2 P(q) S'(q) + B \qquad (8)$$

where the scattering contrast, i.e. the difference between the SLD of the protein and the SLD of the solvent, is written as $\Delta\rho$. Equation 8 describes the so-called β-approximation [12]: the structure factor is replaced by an apparent or effective structure factor *S'*(*q*) :

$$S'(q) = 1 + \beta(q)[S(q) - 1] \qquad (9)$$

$$\beta(q) = \frac{|\langle F(q)\rangle|^2}{\langle |F(q)|^2\rangle} \qquad (10)$$

where β is a q-dependent factor that suppresses oscillations of $S(q)$ due to the presence of polydispersity and molecular shape effects. For monodisperse spheres, $\beta = 1$ and equation 1 provides a good fit to the measured intensities. At high concentrations, $0 < \beta < 1$ and equation 8 should be used instead.

1.3. Dilute solutions of biological molecules

For dilute solutions, $\beta = 0$ and $S'(q) = S(q) \approx 1$. When structure factor effects on the scattered intensities can be neglected, model-independent parameters can be obtained from SANS data for the individual proteins in solution, such as the forward scattering intensity $I(0)$. The intensity at $q = 0$ is defined as:

$$I(0) = n\,\Delta\rho^2 V_{protein}^2 \qquad (11)$$

where n, the number density of protein molecules defined in equation 1, can be written in terms of the protein concentration, c (g cm^{-3}):

$$I(0) = \frac{cNA}{M_w} \qquad (12)$$

where N_A is Avogadro's number and M_w is the molar mass of the protein. If the SANS data are on an absolute scale (cm^{-1}), equations 11 and 12 can be used to relate $I(0)$ from the data to M_w, from which the presence of a monomer, dimer, or multimer may be inferable. Size information can also be extracted directly from SANS data without prior structural information, namely the radius of gyration, R_g , which provides an indication of the overall size of the molecule. The R_g can be calculated directly from the scattering curve using the **Guinier approximation** [13]:

$$I(q) = I(0)\, exp(-q^2 \frac{R_g^2}{3}) \qquad (13)$$

Equation 13 is valid on the low-q portion of the data to obtain R_g and $I(0)$, where $qR_g \leq 1.2$ for globular proteins. A second method that can be used to obtain R_g and $I(0)$ is to calculate the **distance distribution function**, $P(r)$ [14]. This method makes use of the full q-range of the SANS data and is typically obtained using an indirect Fourier transformation of the intensities [15-17]:

$$I(q) = 4\pi V \int_0^{Dmax} \frac{sin\,(qr)}{qr} P(r)\, dr \qquad (14)$$

where D_{max} is the maximum dimension beyond which $P(r) = 0$.

Although interatomic distances cannot be measured with SANS, since this information is smeared by the tumbling of molecules in solution, the folding state of a polymer such as a protein can be assessed by plotting $q^2I(q)$ as a function of q: a **Kratky plot** [14, 18]. Taking equation 13 and multiplying both sides by $(qR_g)^2/I(0)$:

$$\frac{I(q)}{I(0)}(qR_g)^2 = (qR_g)^2 \, exp\left(-q^2\frac{R_g^2}{3}\right) \qquad (15)$$

Dividing $I(q)$ by $I(0)$ normalizes the scattering profiles by mass and concentration (equations 11-12). Multiplying by $(qR_g)^2$ renders the plot dimensionless, so that a semi-quantitative assessment can be made for the flexibility and degree of unfolding of the molecules present in solution, provided that an accurate incoherent background subtraction has been carried out. It can be shown, by taking the derivative of the function in equation 15, that the first maximum of the function should be 1.104 at $(qR_g) = \sqrt{3}$ (see Figure 2).

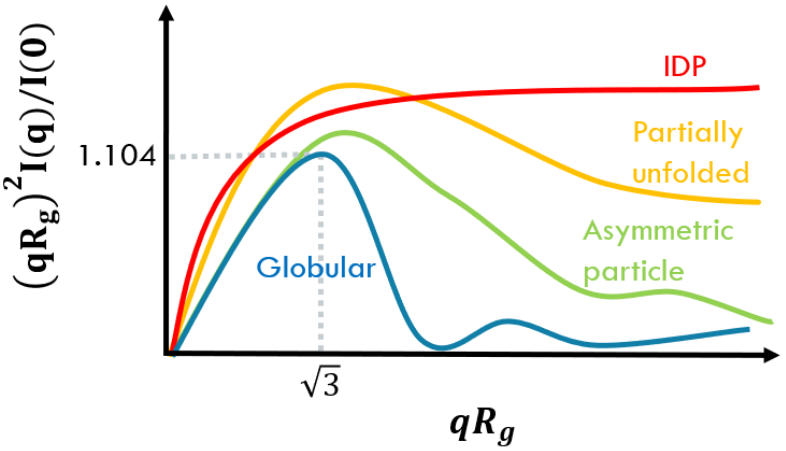

Figure 2. Schematic representation of a dimensionless Kratky plot for different types of proteins. A globular compact protein (blue) should have a Gaussian peak with an intensity of 1.104, regardless of its size, composition and concentration. Deviations from this behavior indicate a change in the flexibility or asymmetry. A hyperbolic plateau, as shown in red for an intrinsically disordered protein (IDP) is indicative of flexibility, while a deviation of the peak position from $\sqrt{3}$ (green) is characteristic of an asymmetry in the molecular shape.

1.4. Contrast variation and deuteration

As shown in equation 8, SANS intensities are a function of the contrast between the molecule of interest and the surrounding environment. Figure 3 shows a plot of the SLDs of common biological macromolecules, where the black line for water is indicative of the SLD expected for aqueous buffers. The SLD increases with increasing volume fractions of D_2O for almost all types of biomolecules, due to the exchange of labile hydrogen atoms against deuterium from the solvent.

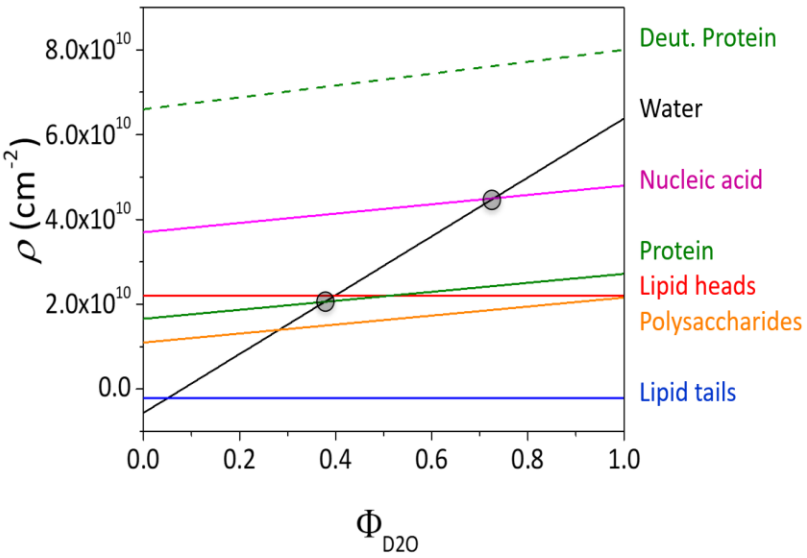

Figure 3. Typical scattering length densities (SLDs) of common biological macromolecules as a function of the volume fraction of D_2O in the solution. The transparent circles highlight **match points**: where scattering from the aqueous solvent and protein, or from the aqueous solvent and a nucleic acid, are effectively indistinguishable and contrast is zero.

For proteins, isotopic exchange occurs between amide nitrogen atoms of the protein backbone and the solvent. Other labile hydrogen atoms include those bound to O, N or S. Isotopic exchange is governed by pH, temperature and the solvent accessibility of the atoms within the protein structure. The sensitivity of neutron scattering techniques to differences between scattering from different nuclei, such as in H and D isotopes, enables SANS to distinguish between multiple types of molecules. Furthermore, by varying the percentage of D_2O present in solution, the contrast can be tuned to enhance or *match out* parts of the sample. For example, in a protein/DNA complex, contributions to scattering from the protein can be nominally reduced to zero, by dissolving the complex in a solution of ≈ 40% D_2O (see figure 3 match points). In the range 70-80% D_2O, on the other hand, the scattering profile of a solution of a protein/DNA complex would be determined by contributions from the protein only. It should be emphasized that contrast variation does not require isolating different components to study their structure: SANS probes the structure of individual components directly within a macromolecular complex [19]. This is a significant advantage over SAXS techniques, which are mostly insensitive to scattering from light isotope such as H or D: X-ray contrast variation requires more invasive sample preparation protocols, for example by varying salt concentration.

For single proteins in solution, using a solution with a large volume fraction of D_2O has the advantage of significantly reducing the incoherent neutron scattering background, which is largely dominated by contributions from the H atoms in the solvent. Proteins can also be selectively deuterated (a domain, for example), to highlight a particular structural feature of interest [20]. Protein-protein complexes can be resolved if one of the proteins is deuterated (see the difference between the SLD of a protein and its deuterated counterpart, shown in Figure 3).

2. Methods and exercises

2.1. Experiment planning: estimating contrast, minimum concentration and q-range of interest

SASSIE-web is an online program suite, freely available, to create atomistic models of molecular systems, calculate their scattering profiles and compare model profiles to experimental data [21, 22]. Using the Contrast Calculator [22] module in SASSIE-web, calculate the neutron SLD of HEWL (in buffers with D_2O fractions between 0 and 1), $\Delta\rho$ and $I(0)$ values, taking isotopic exchange into account. Using the SasCalc [23] module in SASSIE-web, calculate a model SANS curve – $I(q)$ as a function of q – from the provided PDB model structure for HEWL (modified from PDB ID code 193L, to include hydrogen atoms and for compatibility with the atom naming conventions used by the software). Compare the model curve to the SANS data that do not contain a structure factor, $S(q)$.

2.2. Sample and experimental data information

HEWL was obtained from Sigma (L-4919) and used without further purification. Two stock solutions of 150 mg/mL lysozyme were prepared in (A) D_2O and (B) 150 mM NaCl and D_2O. Dilutions of the stock solutions produced two series of samples in the absence (A) and presence (B) of salt, respectively, as described in Table I. All samples were degassed prior to data collection. Sample concentrations were confirmed by measuring UV absorption at 280 nm (theoretical extinction coefficient 38940 M^{-1} cm^{-1}). For higher concentration samples (45, 75 and 100 mg/mL), data were collected using 1 mm-thick cuvettes, while for lower concentration samples (1, 2.5, 5 and 15 mg/mL) the samples were loaded in 2 mm-thick cuvettes.

SANS data were collected for all samples at 25°C, using the NG7SANS 30m instrument of the NIST Center for Neutron Research [24] and a neutron wavelength of 6 Å ($\Delta\lambda/\lambda$ = 12.5%). Three sample-to-detector distances were used to cover a q-range of 0.003 – 0.55 $Å^{-1}$. Scattered intensities were recorded in a 640 mm x 640 mm ^{3}He position-sensitive proportional counter detector, with a 5.08 mm x 5.08 mm resolution.

2.3. Data reduction

Using IGOR Pro with the SANS macro routines developed at the NCNR [25], calculate the transmissions for all samples measured. Use the provided detector mask file to discard non-reliable data recorded on the detector. Build a reduction protocol for all instrument configurations and sample thicknesses used: normalize the raw counts to a common monitor count, and correct for empty cell counts, ambient room background counts, and nonuniform detector response. The data should be reduced to an absolute scale by normalizing the scattered intensity to the incident beam flux. For each sample, sort the data from the 3 instrument configurations into a single file, where the

radially averaged intensity, $I(q)$, can be plotted to show the scattering curve. Finally, subtract the incoherent neutron scattering background from the data.

Table I. Sample formulations. The sample pH was measured before degassing.

Expected [Lysozyme] mg/mL		Measured [Lysozyme] mg/mL (A_{280})	Measured pD
Buffer A	0	0	6.8
Series A D_2O solution	1.0	1.0	5.8
	2.5	2.4	4.7
	5.0	5.0	4.9
	15.0	15.1	4.5
	45.0	42.0	4.5
	75.0	74.6	4.5
	100.0	101.0	4.5
Buffer B	0	0	5.3
Series B 150 mM NaCl in D_2O	1.0	1.0	5.1
	2.5	2.5	4.6
	5.0	5.1	4.2
	15.0	15.4	4.3
	45.0	43.8	4.3
	75.0	72.7	4.3
	100.0	99.8	4.3

2.4. Data analysis (dilute solutions)

All data analysis will be conducted using SasView, a freely-available small-angle scattering analysis software package [26]. Load the reduced and background-subtracted data for the dilute samples into SasView. For this exercise, the dilute samples are assumed to be those that do not obviously show an interaction peak due to contribution from $S(q)$: 1.0 mg/mL data from Series A and 1.0 mg/mL, 2.5 mg/mL and 5.0 mg/mL data from Series B. Calculate R_g in two ways: 1) using the Guinier linear fit to the low q region of the curve (linear form of equation 13) and 2) using the Guinier shape-independent model in SasView. Determine $P(r)$ for these data using the inversion option in SasView**.**

2.5. Data fitting

Load the reduced background-subtracted data into SasView. Fit the data to a model using equations 8 - 10 with an ellipsoidal form factor $P(q)$ and a Hayter_MSA structure factor, $S(q)$. First, fit the 1.0 mg/mL data from both Series A and Series B where $S(q) \approx 1$. This will provide the parameters for the ellipsoid. Then, fit the 100.0 mg/mL data from both Series A and Series B, with the corresponding ellipsoid parameters fixed, to obtain $S(q)$. Plot the fitted $S(q)$ curves for each sample on the same graph to compare the effects of concentration and the presence of salt in the buffer. Fit the rest of the data. For each of the 2 series A and B, plot the fitted $S(q)$ curves on the same graph to compare the effects of concentration.

3. Goals of the BioSANS course

At the end of the BioSANS course, students will be able to plan a SANS experiment for writing a strong beamtime proposal and optimizing data collection strategies. Students will learn how to simulate SANS data from a protein sequence, an all-atom model, or a shape approximation to the structure. Sample preparation, including formulation conditions, sample thickness and sample environment will be discussed. The students will learn how to reduce experimental SANS data, collected prior to the school, on an absolute scale. Model-independent parameters from the experimental data will be compared with the simulated parameters from an all-atom model of HEWL; the students will use the theoretical background to discuss differences. The advantages and limitations of the dilute and concentrated solution regimes will be identified. The course emphasizes a typical BioSANS experiment, where the models available for fitting are not necessarily the best description of the solution structure under a certain sample environment. A discussion on experimental strategies will include the potential need for isotope labelling and the use of complementary techniques to support reliable data fitting.

Acknowledgements

We are very grateful to Brian Paul (U. Delaware), Jordan Berger (U. Delaware), Julia Greenfield (NCNR) and Paul Butler (NCNR) for helpful discussions and feedback. This work benefited from the use of the SasView application, originally developed under NSF award DMR-0520547. SasView contains code developed with funding from the European Union's Horizon 2020 research and innovation programme under the SINE2020 project, grant agreement No 654000. This work also benefitted from CCP-SAS software (SASSIE-web) developed through a joint EPSRC (EP/K039121/1) and NSF (CHE-1265821) grant.

Disclaimer

Certain commercial equipment, instruments, suppliers are identified to foster understanding. This does not imply recommendation or endorsement by the National Institute of Standards and

Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

REFERENCES

[1] L.N. Johnson, The early history of lysozyme, Nature Structural Biology 5(11) (1998) 942-944.

[2] M.R. Wilkins, E. Gasteiger, A. Bairoch, J.C. Sanchez, K.L. Williams, R.D. Appel, D.F. Hochstrasser, Protein identification and analysis tools in the ExPASy server, Methods Mol Biol 112 (1999) 531-52.

[3] A. Bijelic, C. Molitor, S.G. Mauracher, R. Al-Oweini, U. Kortz, A. Rompel, Hen egg-white lysozyme crystallisation: protein stacking and structure stability enhanced by a Tellurium(VI)-centred polyoxotungstate, Chembiochem 16(2) (2015) 233-41.

[4] M. Zaman, H.A. Safdari, A.N. Khan, S.M. Zakariya, S. Nusrat, T.I. Chandel, R.H. Khan, Interaction of anticancer drug pinostrobin with lysozyme: a biophysical and molecular docking approach, J Biomol Struct Dyn 37(16) (2019) 4338-4344.

[5] T. Wu, Q. Jiang, D. Wu, Y. Hu, S. Chen, T. Ding, X. Ye, D. Liu, J. Chen, What is new in lysozyme research and its application in food industry? A review, Food Chemistry 274 (2019) 698-709.

[6] S.A. Ragland, A.K. Criss, From bacterial killing to immune modulation: Recent insights into the functions of lysozyme, PLoS Pathog 13(9) (2017) e1006512.

[7] V.K. Aswal, S. Chodankar, J. Kohlbrecher, R. Vavrin, A.G. Wagh, Small-angle neutron scattering study of protein unfolding and refolding, Physical Review E 80(1) (2009) 011924.

[8] S.C.M. Teixeira, High-pressure small-angle neutron scattering for food studies, Current Opinion in Colloid & Interface Science 42 (2019) 99-109.

[9] S.C.M. Teixeira, J.B. Leao, C. Gagnon, M.A. McHugh, High pressure cell for Bio-SANS studies under sub-zero temperatures or heat denaturing conditions, Journal of Neutron Research 20(1-2) (2018) 11-21.

[10] M. Papachristodoulou, J. Doutch, H.S.B. Leung, A. Church, T. Charleston, L.A. Clifton, P.D. Butler, C.J. Roberts, D.G. Bracewell, In situ neutron scattering of antibody adsorption during protein A chromatography, Journal of Chromatography A 1617 (2020) 460842.

[11] J.E. Curtis, H. Nanda, S. Khodadadi, M. Cicerone, H.J. Lee, A. McAuley, S. Krueger, Small-angle neutron scattering study of protein crowding in liquid and solid phases: lysozyme in aqueous solution, frozen solution, and carbohydrate powders, J Phys Chem B 116(32) (2012) 9653-67.

[12] M. Kotlarchyk, S.H. Chen, Analysis of small angle neutron scattering spectra from polydisperse interacting colloids, The Journal of Chemical Physics 79(5) (1983) 2461-2469.

[13] G.F. A. Guinier, Small-angle scattering of X-rays, Wiley, University of Michigan, 1955.

[14] E. Walenta, Small angle x-ray scattering. Von O. GLATTER und O. KRATKY. London: Academic Press Inc. Ltd. 1982. ISBN 0-12-286280-5. X, 515 Seiten, geb. £ 43,60; US $ 81.00, Acta Polymerica 36(5) (1985) 296-296.

[15] O. Glatter, A new method for the evaluation of small-angle scattering data, Journal of Applied Crystallography 10(5) (1977) 415-421.

[16] P.B. Moore, Small-angle scattering. Information content and error analysis, Journal of Applied Crystallography 13(2) (1980) 168-175.

[17] A.V. Semenyuk, D.I. Svergun, GNOM - a program package for small-angle scattering data processing, Journal of Applied Crystallography 24(5) (1991) 537-540.

[18] V.M. Burger, D.J. Arenas, C.M. Stultz, A Structure-free Method for Quantifying Conformational Flexibility in proteins, Scientific Reports 6(1) (2016) 29040.

[19] S. Krueger, Designing and Performing Biological Solution Small-Angle Neutron Scattering Contrast Variation Experiments on Multi-component Assemblies, Adv Exp Med Biol 1009 (2017) 65-85.

[20] P.T. Reddy, R.G. Brinson, J.T. Hoopes, C. McClung, N. Ke, L. Kashi, M. Berkmen, Z. Kelman, Platform development for expression and purification of stable isotope labeled monoclonal antibodies in Escherichia coli, mAbs 10(7) (2018) 992-1002.

[21] J.E. Curtis, S. Raghunandan, H. Nanda, S. Krueger, SASSIE: A program to study intrinsically disordered biological molecules and macromolecular ensembles using experimental scattering restraints, Computer Physics Communications 183(2) (2012) 382-389.

[22] J. Curtis, Sassie-web: https://sassie-web.chem.utk.edu/sassie2/.

[23] M.C. Watson, J.E. Curtis, Rapid and accurate calculation of small-angle scattering profiles using the golden ratio, Journal of Applied Crystallography 46(4) (2013) 1171-1177.

[24] J.G.B. C. J. Glinka, B. Hammouda, S. Krueger, J.J. Moyer, W. J. Orts, The 30 m Small-Angle Neutron Scattering Instruments at the National Institute of Standards and Technology, J. Appl. Cryst. 31 (1998) 430-435.

[25] S.R. Kline, Reduction and Analysis of SANS and USANS data using Igor Pro, J. Appl. Crystallography 39(6) (2006) 895-900.

[26] Sasview for Small Angle Scattering Analysis - a SAS Community project. https://www.sasview.org/.