7.S: Practical (wet-lab) considerations for small-angle scattering experiments.

Cy M Jeffries and Jill Trewhella (2011) Small-Angle Scattering: for structural molecular biologists, in *Quantitative biology: From molecular to cellular systems* **SUPPLEMENTARY INFORMATION** (M. Wall, ed.), Taylor and Francis Group, London – to be published.

7.S.1 Sample preparation.

It is perhaps useful to note what the characteristics of a 'good' sample for SAS analysis are. Using proteins as a case study, this section will discuss the requirements for preparing SAS-quality samples.

7.S.1.1. Gel electrophoresis - sample quality and assessment.

SDS-polyacrylamide gel electrophoresis is an excellent tool to assess protein sample purity and to monitor protein stability over time or to assess the effects of radiation exposure (*e.g.*, degradation due to bond breakage by X-ray induced free radical production (1).) Furthermore, the protein of interest can be excised from a gel and the amino acid sequence checked using matrix-assisted laser desorption/ionization (MALDI) time-of-flight mass spectrometry (peptide mass fingerprinting.) Combined, gel electrophoresis and mass spectrometry data offer a reassurance that a sample is pure, intact and has the correct sequence. Sample purity is important because scattering profiles represent the time and ensemble-averaged scattering from randomly oriented particles in solution, *i.e.*, *any* particle in solution and not just the particle of interest. As the scattered intensity is proportional to the square of the molecular volume, large molecular weight contaminants, even in relatively small amounts, are problematic. If a 'purity scale' could be devised for comparing and contrasting different methods used in structural and molecular biology, it might look something like this:

Enzyme kinetics << **NMR spectroscopy** << **Crystallography** <<<< **SAS**
~ 50 -80% <90% <95% **95-100%**

The scale reflects that NMR signals depend linearly on concentration and the signals of large molecular weight contaminants become broadened in the NMR spectrum and that crystallization, by its very nature, removes contaminating species. In contrast, doubling the molecular weight of a protein in a SAS experiment (effectively doubling the volume) quadruples its contribution to the scattering intensity (eq. **8**.) This dependency can cause issues if high molecular weight contaminants are present in a sample. For example, if a SAXS experiment was performed on a 97% w/v pure 14 kDa protein target contaminated with 3% w/v of a 140 kDa protein, that 3% contaminant would increase the overall scattered intensity by ~27% making it difficult, if not impossible, extract accurate structural parameters from the data. It is important to use gel electrophoresis (Figure S1) to ensure that proteins, especially those of a higher molecular weight relative to the target of interest, are not present in the sample (slight traces of smaller contaminants can be tolerated.)

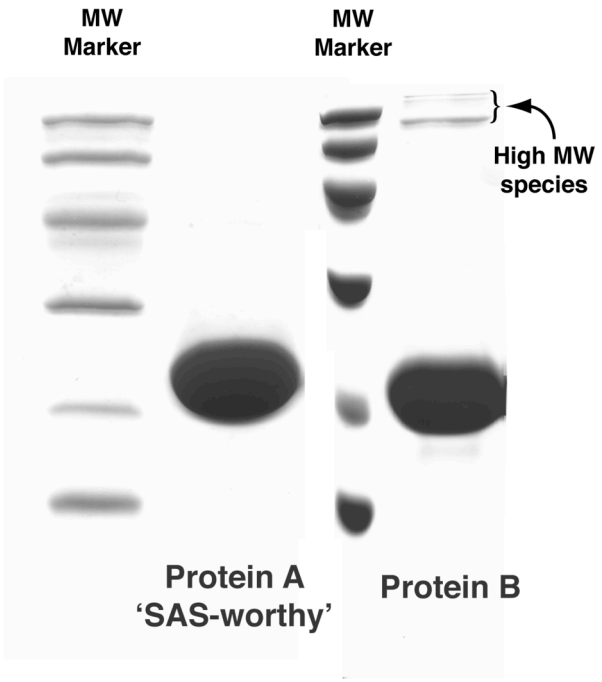

Figure S1. *Is it SAS-worthy?* Polyacrylamide gel electrophoresis is an excellent tool for assessing the purity of protein samples. The above gel shows two different protein samples under denaturing conditions. The sample on the right is pure enough for SAS experiments whereas the sample on the left has small traces of high molecular weight species that will significantly affect the scattering.

7.S.2. Solvent Matching.

As mentioned in 7.5, SAS is a subtractive technique — the solvent or blank measurement *is as important* as the sample measurement; get the solvent subtraction wrong and you can end up, for example, with unstable transforms between reciprocal (q)-space and real (r)-space or uncertain $I(0)$ analyses. Buffer or solvent matching is important for all biomacromolecular SAS investigations, but is especially so for small-angle neutron scattering that relies heavily on contrast variation (that can be further complicated by high incoherent scattering contributions from ^{1}H dominated systems (2).) It is generally optimal to perform the sample and solvent measurements for approximately the same time period so as to have similar counting statistics in

the high-*q* regime. Error estimates in scattering experiments are generally based on Poisson counting statistics (*i.e.* proportional to $\sqrt{N}$ where *N* is the number of scattering events counted.) As a result, the sample counting statistics will dominate at the smallest angles, but at the higher angles the counting statistics between the sample and solvent will be equally important.

Dialysis is an efficient means to perform solvent exchange as well as to screen multiple solution conditions in parallel to identify those for which a protein or complex is stable and monodisperse in solution. An alternative is to perform buffer exchange using size exclusion chromatography (SEC), where peak fractions containing a protein of interest and a protein-free fraction are collected from the column as sample and solvent, respectively. The advantage of SEC is that it is usually incorporated into purification procedures and can be a potential time saver by avoiding a dialysis step. The disadvantage of SEC is that only one buffer can be screened at any time and highly concentrated protein must be loaded onto the resins to take into account dilution effects (in most instances SAXS experiments are performed on solutions containing between ~2 and 10 $mg.mL^{-1}$ of the biomolecular species of interest.) In the case of complexes care must be taken that the subunits do not disassociate especially for moderately interacting systems (complexes with $K_d > 1$ μM are particularly difficult to isolate and may also be difficult to characterize using SAS unless the concentration is sufficiently high to ensure full complexation.) Here are three circumstances to consider:

i. *My protein should be in buffer X - I'll just make up fresh buffer X and use that as my blank.* This approach is not advised. It is almost impossible to know what buffer your protein is actually in unless exchanged. Weighing out sufficiently exact quantities of materials to make up new buffers is not generally posible.

ii. *I'll concentrate my sample using a membrane technique, like a spin column, then do SAXS on the sample and the membrane flow-through as a blank.* This method should be avoided if practical to do so. Protein concentrators can aggregate samples at the membrane/sample interface and small molecule concentrations can be different in the flow through and in the retained sample *e.g.*, there can be small molecule retention in the sample concentrate and thus the flow-through will no longer be solvent matched with the sample.

iii. *I have an expensive cofactor to add to my sample, I can't afford to do dialysis or SEC with that present!* Make up a concentrated stock solution of your expensive cofactor in a small volume of post dialysis or SEC buffer (that has gone through the column) and carefully add equivalent volumes to the buffer exchanged sample and solvent to the desired concentration prior to analysis.

7.S.3. Preparing for a neutron scattering experiment.

7.S.3.1. Isotopic labeling of proteins with deuterium.

There are a number of ways to deuterate a protein using recombinant protein expression techniques. On can use rich deuterated media based on algal-extract lyophilysates or minimal media made up in D_2O and if needed, deuterated sugars (3). Some research institutions even offer biodeuteration facilities (*e.g.*, The National Biodeuteration Facility at the Australian Nuclear Science and Technology Organisation, Lucas Heights, Australia or the Institut Laue-Langevin, Grenoble, France, and at the Oak Ridge National Laboratory, in Tennessee, USA.) Here we provide a brief outline for ^{2}H-labeling recombinant proteins in minimal media using *Escherichia coli*.

D_2O has different physical properties to regular light water that generally slows bacterial

growth and stresses the organism. Therefore it is important to use a strain of *Escherichia coli* that can survive and express protein in media with high levels of D_2O present (e.g., *E. coli* Bl21.) As D_2O is expensive it is best to initially assess the expression of the protein of interest in small volumes of media while screening for optimal pH, optimal antibiotic concentration and the addition of trace elements, etc. It is also important to adapt bacteria to their new deuterium rich environments (4). When adapting bacterial populations to D_2O minimal media, a standard protocol can include several small overnight cultures that have incrementally increased concentrations of D_2O until the final desired D_2O concentration in the media is reached (*e.g.*, night 1, ^1H_2O-LB rich media; night 2, 50% v/v ^1H_2O-LB/D_2O media; night 3, 70% v/v D_2O minimal media and; night 4 'the final culture' at the required percentage D_2O.) Leiting et al. (1998) (3) have determined the relationship between the percentage v/v D_2O used in minimal media preparations and the average incorporation of non-exchangeable deuterium in the final protein (Figure S2) that can be checked using a comparative analysis of mass peak shifts derived from peptide mass finger printing.

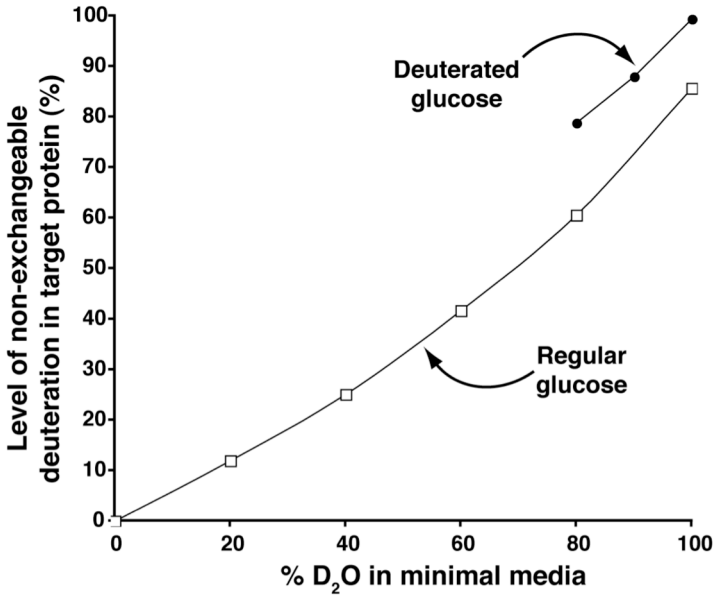

Figure S2. *Predictable labeling of proteins with deuterium*. By growing *E. coli* in D_2O-based minimal media it is possible to control the average incorporation of non-exchaneable deuterium into expressed proteins. Choosing the final level of deuteration in a protein will depend on the type of SANS experiment and the calculated match points of the protein and protein complex under investigation (adapted from Leiting et al. 1998 (3).)

After testing to ensure that the protein over expresses in small deuterated cultures (it may take 3–6 times longer to observe over expressed protein compared to regular media) it is often necessary to perform a medium scale purification (using regular ^1H buffers) to ensure that:

1) The deuterated protein is soluble.

2) The deuterated protein forms a complex with the desired non-deuterated partner.

3) The entire complex is soluble and remains monodisperse in high concentrations of D_2O containing buffers (where SANS experiments will be performed.)

The last point is very important. The 'D-bonds' in D_2O are stronger than regular hydrogen bonds and can affect protein solubility. SAXS or dynamic light scattering are excellent methods for determining whether a complex undergoes D_2O induced aggregation. The advantage of

SAXS is that conditions for the SANS experiment can be optimized, including buffer conditions, concentration series and *I*(0) analyses, using the smallest amounts of material. Due to the significant cost of D_2O, performing SAXS on a small scale is a good idea, especially if you have ready access to a laboratory based SAXS instrument. A full SANS contrast variation experiment requires the production of significantly more material than for a SAXS investigation — anywhere between 50–100 times as much (1.5–3 ml at ~5–10 $mg.ml^{-1}$ compared to 10–120 μl at ~2–10 $mg.ml^{-1}$ for SAXS.)

7.S.3.2 The neutron trip.

'Neutron trips' can be stressful affairs because one has to apply for relatively rare neutron beam time; the time is scheduled far in advance and the schedule is unforgiving. If your samples are not any good, you generally cannot reschedule without a new application and if you need a few more hours of beam time to finish an experiment, the next experimenter generally cannot be delayed. The sample sizes are large and deuteration is expensive, so a lot is at stake. The more preparation that goes into delivering quality samples to the beam line will reduce the stress and increase the success rate. Having a thorough experimental plan, good preliminary SAXS data, and an understanding of the logistics of sample preparation are the keys to success.

Before arrival.

In a SANS experiment, just as for SAXS, the magnitude of the scattering intensities will depend on the concentration and molecular weights of the components in solution as well as the collection time: the higher the concentration or molecular weight the better the counting statistics. However neutron fluxes are orders of magnitude less than X-ray sources so there is a

requirement for a SANS investigation to produce and use significantly larger samples that undergo longer exposure times (~15 min–4 hrs) compared to SAXS (1 s–1 hr). As well as characterizing samples as much as possible prior to arrival (using SAXS or DLS), it is worthwhile to calculate (*e.g.*, using MULCh (5)) all of the match points of the components and the complex using the experimentally determined level of deuteration on the relevant component. These calculations can be invaluable in the pre-planning of experimental designs in terms of the number of contrast experiments which can be completed within the timeframe allocated for the experiment. For example, as the SANS signal strength will be proportional to the volume fraction of components in a protein-protein complex, the level of deuteration on one of the components and complex concentration it might be necessary to supplement a full contrast series with stand-alone experiments to improve the counting statistics for certain contrast points (Figure S3).

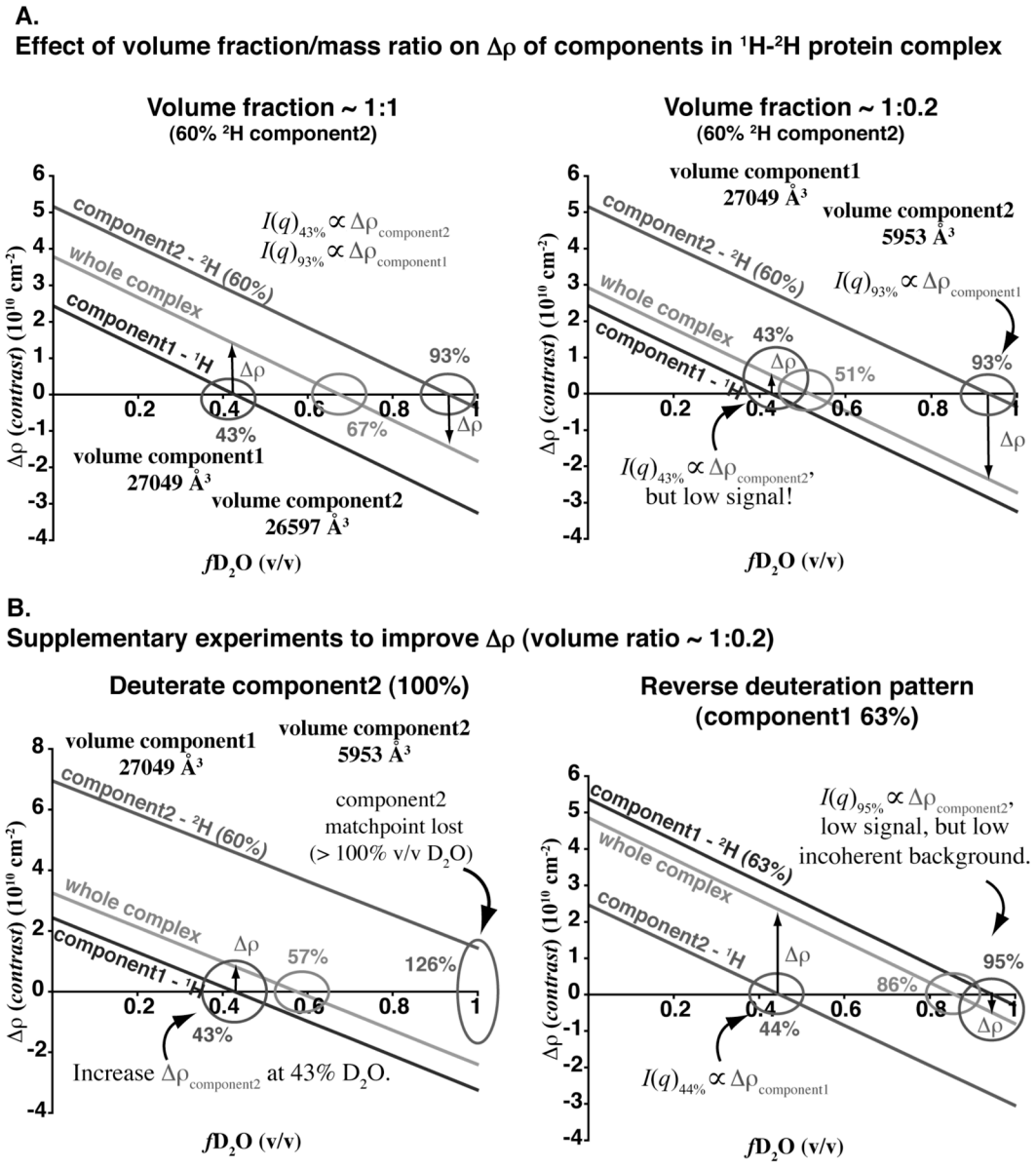

Figure S3. *SANS match point MULCh calculations - affect of volume fraction.* Using a 1:1 protein complex as an example, the *Contrast* module of MULCh (5) can be used to calculate the expected contrasts (Δρ) *vs* fraction of D_2O in the solvent (*f*D2O) taking into account the volume fractions (mass ratio) and the level of deuteration on one of the components in the complex (that can be optimized in terms of improving match point separation by altering ^2H growth conditions — Figure S2). Complexes that have components with radically different volume fractions can affect the contrast and hence the strength of the SANS intensities (**A**). Therefore, it might be necessary to supplement a full contrast variation experiment with 'stand alone' experiments to improve the statistics in the scattering signal at certain *f*D2O *e.g.*, fully deuterating smaller components to increase Δρ or reverse the component deuteration pattern (**B**.)

Upon arrival.

When you arrive at your favorite nuclear facility there are a couple of practical things to consider. Consistency in sample handling is important to reduce unwanted and unquantifiable errors during data acquisition. For example, forming a complex as a singular 'master stock' to the desired concentration for an experiment and apportioning the stock into smaller individual aliquots for dialysis against each of the buffers in the contrast series means that each sample derived from the master stock will be identical. Further, if sample dialysis is performed in sealable airtight bags to prevent atmospheric water contamination then, in theory, the only thing that should change in each sample is the contrast with the solvent. This approach should produce more consistent results as compared to making up multiple small samples derived from different protein preparations that can be affected by differences in protein concentration, mixing errors and deuteration levels. Care should also be taken when making up the contrast variation buffers, specifically in regard to pH. The pD of a solution is related to the pH via the following relationship:

$$\mathrm{pD} = \mathrm{pH} + 0.4.$$

If the H_2O and D_2O buffers are prepared at various H_2O/D_2O ratios, is it very difficult to pH/pD adjust them to the correct value *after* they have been mixed. However, by carefully weighing out the exact same amounts of buffer components in 100% v/v H_2O and in 100% v/v D_2O then adjusting the pH or pD to the correct value *before* they are mixed will help maintain pH/pD consistency across the contrast series. It is also pertinent to record the concentration of each sample post-dialysis to account for any subtle changes in concentration caused by H_2O/D_2O exchange. It may also be necessary, depending on the sample, to slowly raise the level of % v/v

D_2O in the solvent via multiple dialyses in ever increasing D_2O increments as opposed to dialyzing samples from H_2O into high D_2O buffers directly. After the dialysis step(s) is(are) completed, the samples and buffers (from the dialyzates) should be loaded into dry sample cells that are free of any external water condensation (which can cause problems with cold samples) before being loaded into the SANS instrument. It is generally also a good idea to degas samples as much as possible before loading them into the cells by gentle sonnication: the formation of tiny bubbles, due to subtle changes in temperature during extended neutron exposure times, will scatter neutrons strongly. Once you have measured your data, and performed all the important checks before you know it you will have a model of your complex!

Supplementary References

1. Kuwamoto, S., S. Akiyama, and T. Fujisawa. 2004. Radiation damage to a protein solution, detected by synchrotron X-ray small-angle scattering: dose-related considerations and suppression by cryoprotectants. J. Synchrotron Radiat. 11:462-468.
2. Rubinson, K. A., C. Stanley, and S. Krueger. 2008. Small-angle neutron scattering and the errors in protein structures that arise from uncorrected background and intermolecular interactions. J. Appl. Cryst. 41:456-465.
3. Leiting, B., F. Marsilio, and J. F. O'Connell. 1998. Predictable deuteration of recombinant proteins expressed in Escherichia coli. Anal. Biochem. 265:351-355.
4. Meilleur, F., K. L. Weiss, and D. A. Myles. 2009. Deuterium labeling for neutron structure-function-dynamics analysis. Methods Mol. Biol. 544:281-292.
5. Whitten, A. E., S. Z. Cai, and T. J. 2008. MULCh: modules for the analysis of small-angle neutron contrast variation data from biomolecular assemblies. J. Appl. Cryst. 41:222-226.