

Investigation of molecular and mesoscale clusters in undersaturated glycine aqueous solutions

*Georgina Zimbitas**, *Anna Jawor-Baczynska*&*, *Maria Jazmin Vesga**, *Nadeem Javid**,
*Barry D. Moore***, *John Parkinson***, *Jan Sefcik*+++*

* Department of Chemical and Process Engineering, University of Strathclyde, 75
Montrose Street, Glasgow, G1 1XJ, UK

& Present address: AstraZeneca, Research and Development, Silk Road Business Park,
Charter Way, Macclesfield, SK10 2NA, UK

+ School of Chemistry and Biosciences, University of Bradford, Bradford, BD7 1DP, UK

** WestCHEM, Department of Pure and Applied Chemistry, University of Strathclyde,
295 Cathedral Street, Glasgow, G1 1XL, UK

+++ EPSRC Future Manufacturing Hub in Continuous Manufacturing and Advanced
Crystallisation, University of Strathclyde, 99 George Street, Glasgow, G1 1RD, UK

KEYWORDS

clustering, mesostructured liquid phase, glycine, scattering

Abstract

In this work we used DLS, NTA, SAXS and NMR to investigate populations, size distributions and structure of clusters in undersaturated aqueous solutions of glycine. Molecular and colloidal scale (mesoscale) clusters with radii around 0.3-0.5 nm and 100-150 nm, respectively, were observed using complementary experimental techniques. Molecular clusters are consistent with hydrated glycine dimers present in equilibrium with glycine monomers in aqueous solutions. Mesoscale clusters previously observed in supersaturated glycine solutions appear to be indefinitely stable in mutual equilibrium within mesostructured undersaturated solutions across all glycine concentrations investigated here, down to as low as 1 mg/g of water.

1. Introduction

Aqueous solutions of highly soluble molecules, such as those of small amino acids, are usually assumed to be essentially homogenous systems with some degree of local structuring due to specific interactions on the sub-nanometre scale (e.g. molecular clusters, hydration shells); these molecular structures (molecular clusters) usually do not exceed several solute molecules[1-3]. Such molecular clusters have been reported in both experiments and simulations in aqueous solutions of many organic and inorganic systems[4-10]. In addition to the small molecular clusters, larger colloidal scale (or mesoscale) clusters have also been reported and there have been of great interest lately due to potential roles played by mesoscale structures in crystal nucleation[2, 11-15]. Recently published reviews[7,9] emphasize that there are various molecular self-

assembly pathways in nucleating solutions and highlight the role of pre-nucleation structures in nucleation of a wide range of organic and inorganic systems.

The presence of sub-micron size domains with liquid-like properties has been widely reported in concentrated solutions of large organic molecules such as proteins[11-13, 16-27] as well as in those of smaller molecules[28-41]. Detailed studies of both super- and undersaturated aqueous solutions of NaCl, (NH₄)₂SO₄ and Na-citrate, using dynamic light scattering (DLS), revealed that such systems contain not only solvated ions (with radii below 1nm) but also larger structures with radii varying from 50nm to 500nm[5]. Using static and dynamic light scattering (SLS and DLS, respectively) further investigation of undersaturated aqueous solutions of common small organic molecules, such as various amino acids and amines[42], citric acid[5, 43, 44], glucose[44], and urea[43, 44], also revealed the presence of large-scale supermolecular structures with broad size distributions, within ranges of several hundred nanometers.

A detailed light scattering study shows that these structures can be characterized as close-to-spherical, discrete domains, which present higher solute density with respect to the lesser dense rest of the solution[44-46], with approximately 10³ to 10⁸ solute molecules found to be present in these structures. A study of under- and supersaturated aqueous solutions of DL-alanine revealed mesoscale domains, with radius sizes ranging from 50 nm to 300 nm, existed well below the solid-liquid equilibrium concentration (saturation limit)[10]. These liquid-like, solute-rich mesoscale domains are dispersed within the bulk solution and are not to be considered a separate phase, but instead the

liquid phase containing these clusters can be seen as a single liquid phase in thermodynamic equilibrium with the whole system.

The dissolution of glycine crystals in water leads to the formation of an optically clear solution which contains not only a stable population of small molecular clusters, but also contains glycine-rich, liquid-like mesoscale clusters, also referred to as nanodroplets, as observed using SAXS and DLS[10, 15]. It has been suggested that the presence of very large nanodroplets, of over 350 nm radius, is required to facilitate crystal nucleation in order to provide the critical mass of glycine essential for productive crystal nucleation[15]. It has been also observed using NTA that mesoscale species were present in undersaturated glycine solutions[17]. However, structure and composition of mesoscale clusters and phase behaviour of such mesostructured liquid phases is very little understood. There has been recently emerging theoretical understanding of thermodynamics of such phases and potential mechanisms of their formation through competing short-range and long-range interactions[47-50], but there is a lack of experimental data on phase behaviour and properties of mesostructured liquid phases. In this work, we used a range of complementary experimental techniques, including Dynamic Light Scattering (DLS), Nanoparticle Tracking Analysis (NTA), Nuclear Magnetic Resonance (NMR), and Small Angle X-ray Scattering (SAXS), to investigate mesostructured glycine solutions over a wide range of glycine concentrations to determine populations, sizes and structures of mesoscale clusters as a function of glycine concentration in aqueous solutions.

2. Experimental Methods

2.1 Materials

All chemicals and solvents were of laboratory reagent grade and used without further purification: Glycine puriss. $\geq 99.0\%$, NT (Fluka), 3-(trimethylsilyl)-2-propionic-d₄ sodium salt (TSP_d₄) puriss. $\geq 97.0\%$ (Sigma Aldrich), Deuterium Oxide isotopic purity 99.9% (Alfa Aesar). Deionized water was supplied from an in-house Millipore Water System, 18 M Ω /cm and was further filtered using either 0.1 μ m (Whatman Cat No 6784-1301) or 1 μ m (Whatman Cat. No. 6784-2510) PTFE syringe filters for the preparation of solutions. Concentrations are shown as mg/g and refer to mg of glycine present in g of water.

2.2 Solutions preparation

Aqueous solutions of glycine were prepared by introducing a weighed amount of solid glycine into a known volume of filtered (via PTFE syringe filter) deionized water present in a glass vial with a screw-on cap. Concentrations prepared ranged from 1 to 230 mg/g. A magnetic stirrer was placed in the glycine solution. The solution was then placed on a magnetic stirring plate, present within an incubator at a constant temperature of 55 °C. The solution was then stirred for 16 to 24 hrs. Prior to analysis samples were filtered using either 0.1 μ m or 1 μ m PTFE syringe filters, unless stated otherwise. All syringes, filters, cuvettes and tubes were preheated in the incubator at 55 °C, to avoid premature cooling of the solution during filtering and transfer. NTA and DLS sample cells were preheated to 50-55 °C before the solution was introduced into them and then left to cool to room temperature (18 – 25 °C), after which they were allowed to equilibrate at room temperature for at least 5min prior to data collection. The concentrations used were below

the solubility limit of glycine in water, 249.9 mg/g at 25 °C, in order to avoid glycine crystallisation.

2.3 Dynamic Light Scattering (DLS)

DLS measurements were carried out using an ALV/LSE- 5004 instrument, equipped with temperature control at a scattering angle of $\theta = 90^\circ$ and a laser light wavelength of $\lambda = 632.8$ nm. DLS is a well-established experimental technique for studying nanoscale particles in dispersions. By measuring the time-dependent fluctuations of scattered light intensity, arising from Brownian motion, average diffusion coefficients and corresponding hydrodynamic radii can be inferred. From the analysis of a measured autocorrelation function the average hydrodynamic radius (R_h) of clusters present was estimated using cumulant analysis[51]. For DLS measurements, original concentrated solutions were prepared as described above and solutions at lower concentrations were obtained via dilutions of the original solution using filtered Millipore water. Both diluted and original solutions were kept in the incubator at 55 °C, and all DLS measurements were performed between 24 hrs and 48 hrs hours after solution preparation. The scattered intensity measured by DLS is shown here in the form of the normalised count rate, NCR, (a.u.), calculated via multiplying the average count rate (kcps) by 10^6 and dividing by the recorded monitor diode value, corresponding to the incident beam intensity.

2.4 Nanoparticle Tracking Analysis (NTA)

Size distribution and number concentrations of mesoscale clusters were determined using a Nanosight LM10 instrument with a temperature control unit. Nanosight NTA 3

software was used to analyse videos and calculate the size and concentration of clusters. The camera settings of the instrument were set to the 'Autosettings' option, to allow the software to optimize the shutter and gain settings. The sample was introduced into the viewing unit and images of patterns of the laser light scattered by diffusing objects was captured by a CCD camera attached to a microscope. A video was recorded and processed, with each observed individual object 'tracked' by the nanoparticle tracking analysis (NTA) software. Each video was recorded for 60 seconds and the processing parameters of brightness and gain were optimized by the software. Particle size was calculated from the Brownian motion analysis, whereas the diffusion coefficient was calculated from the mean squared displacement of the particle tracked. The calculated diffusion coefficient was subsequently substituted into the Stokes-Einstein equation in order to obtain hydrodynamic radii of individual objects tracked. Estimation of the particle concentration was based on the particle count in the illuminated volume ($5.125 \times 10^4 \mu\text{m}^3$) calculated from the dimensions of the field of view (at a magnification of 20x). The total particle count refers to the summation of the particle count up to and including radius of 250 nm. It has been reported that particle concentration measurements are subject to a variation of up to 25 % between identical samples[52].

2.5 Nuclear Magnetic Resonance (NMR)

Measurements of glycine aqueous solutions were carried out within a Bruker 400 MHz, AV400 Instrument using 16 scans and a relaxation delay of 10 sec. The aqueous medium used consisted of a 7:3 H₂O:D₂O w/w ratio. All NMR experiments were conducted using a double coaxial tube system to allow absolute measurement of the glycine concentration:

the main sample tube was a Norell 400MHz, 5 mm diameter, 7" length NMR tube, whereas the internal reference tube was a Wilmad coaxial insert, stem L32 mm. This setup provided for a sample capacity of 530 μ l, and reference capacity of 60 μ l. 1mM of 3-(trimethylsilyl)-2-propionic-d4 sodium salt (TSP_d4) in deuterium oxide was used to provide the internal reference concentration. ^1H NMR Diffusion Ordered Spectroscopy (DOSY) measurements were performed at 27 °C (300 K) and results obtained were used for determination of diffusion coefficients of species within the glycine solutions.

2.6 Small Angle X-ray Scattering (SAXS)

Small angle x-ray scattering measurements were carried out at the SWING beamline at the SOLEIL synchrotron source in Gif-sur-Yvette, France. Aqueous solutions of glycine (ranging from 10 mg/g to 230 mg/g) were prepared at 55 °C and filtered using PTFE 0.1 μm syringe filters. The solutions were injected into a thermostated automatic sample changer and cooled and equilibrated for 5 min at 25 °C. A selected sample would then be automatically passed through a thermostated capillary for exposure to x-ray radiation and detection of scattered intensity patterns. The beam energy was 8keV and the sample-detector distance was 2.227 m, accessing a Q value range between 0.001 \AA^{-1} and 0.17 \AA^{-1} . Five measurements of a single sample were recorded with an exposure time of 1 s each and 2 s delay time, checked for repeatability and then averaged. The PCCD 170170 (AVIEX) 2D-detector was present in a vacuum chamber equipped with a pumping system to obtain a primary vacuum of 10^{-6} bar. The vacuum achieved allowed for lower angles to be reached and also reduced the probability of interference from scattering from air molecules. The same measurement procedure was used for pure water samples. The

integration and processing of the scattering data was carried out by using the software provided at the beamline.

3. Results and Discussion

Below we report and discuss results from Dynamic Light Scattering (DLS), Nanoparticle Tracking Analysis (NTA), Small Angle X-ray Scattering (SAXS), and Nuclear Magnetic Resonance with Diffusion Ordered Spectroscopy (NMR-DOSY) measurements of glycine aqueous solutions within a wide range of glycine concentrations (1 – 230 mg/g). These solutions were undersaturated with respect to solid glycine and showed presence of molecular scale as well as mesoscale clusters across all concentrations investigated, and we aimed to investigate populations, size and structures of clusters in these solutions.

3.1 DLS measurements

DLS measurement provide us with time averaged scattered intensities and intensity averaged autocorrelation functions, from which mean hydrodynamic radii (R_h) can be estimated using cumulant analysis[51]. The autocorrelation functions measured clearly showed two characteristic decay times (see supplementary info), corresponding to a bimodal distribution of species present in glycine aqueous solutions. For filtered solutions, characteristic times of the first decay appeared at times below 0.005 ms, corresponding to molecular clusters, while the second decay appeared at characteristic times over 1 ms, corresponding to mesoscale clusters.

In Figure 1 we show results from DLS measurements in terms of the mean hydrodynamic radius of mesoscale clusters (Fig 1a) and molecular clusters (Fig 1b), as well the

normalised scattered intensity (Fig 1c). Two sets of solutions were investigated: those which were prepared by filtering the original concentrated solution into the filtered water (referred to as filtered solutions), and those which were prepared by adding the original concentrated solution, without filtration, into the filtered water (referred to as unfiltered solutions).

The mean mesocluster radius is between 100 and 150 nm, with a slight upward trend at larger glycine concentrations above 150 mg/g (Fig 1a). At lower glycine concentrations below 50 mg/g there is a larger uncertainty in estimated values as signal to noise ratio decreases with decreasing concentrations. It can be also seen that unfiltered solutions seem to have somewhat larger mean radii, although these differences are close to marginal given accuracy of these estimates as indicated by error bars in Fig 1a. It is interesting to note that the mean size of mesoclusters varies very little over the whole range of concentrations (1 - 230 mg/g), similarly to what was previously observed in DL-alanine aqueous solutions[17].

The mean molecular cluster radius (Fig.1b) for unfiltered solutions varies between 0.3 and 0.5 nm and it increases with increasing glycine concentrations. For unfiltered solutions the mean sizes of molecular clusters seem to be higher but there are large uncertainties in estimated values for lower concentrations up to 100 mg/g but the difference from filtered solutions does not seem to be significant given errors bars in Fig. 1b. As a radius of the glycine molecule was reported to be 0.28 nm[53], the mean hydrodynamic radius of molecular clusters observed at lower glycine concentrations is in fact consistent with hydrated glycine monomers. As glycine concentration increases, the reversible formation of hydrogen-bonded glycine dimers is increasingly driven by the

mass action principle; it was estimated that about 10% of glycine is in the form of a dimer at the highest concentration considered here[54]. The hydrodynamic radius of the glycine dimer is larger than that of the glycine monomer, and in the mean hydrodynamic radius measured by DLS it is weighted proportionally to its scattering intensity which is proportional to the cluster mass squared[55, 56] which is 4 times that of the monomer, and therefore it is expected that the mean hydrodynamic radius of molecular clusters would be increasing with glycine concentration.

The normalised scattered intensity measured by DLS for filtered and unfiltered solutions is shown in Figure 1c. A horizontal dashed line shows the background scattered intensity value corresponding to pure water. It can be seen for filtered solutions the scattered intensity increases linearly with glycine concentration between 1 and 230 mg/g. This is different from what was observed in DL-alanine solutions where appeared to be a critical clustering concentration indicated by a sharp drop off in the scattered intensity between at concentrations between 10 and 40 mg/g[17]. However, in glycine solutions the scattered intensity gradually increases across the whole range of concentrations.

The scattered intensity of unfiltered solutions is significantly higher than that of filtered solutions, which means that there is higher number concentration of scattering objects and/or they are larger (and thus scattering more strongly) than in filtered solutions. As filter pore sizes (1 μm) are larger than mean mesoscale cluster diameters, filtration may remove few very large clusters or break them apart into smaller ones. It may also be that some clusters smaller than the nominal filter pore size become trapped by filters due to their distribution of pore sizes and thus cluster number concentration is reduced in filtered solutions.

Filtered solutions were also used to create another set of solutions to assess filtration effects: the filtered solutions were either transferred directly into DLS cells without further filtration or filtered into DLS cells using 1 μm PTFE filters (double filtered solutions – see supplementary info). Both filtered and double-filtered solutions exhibit similar behaviour, in terms of mean hydrodynamic radii of mesoclusters and molecular clusters, as well as the normalised scattered intensity.

This demonstrated that repeated filtration does not change either size or population of clusters in glycine solutions at a given concentration, although it cannot be ruled out that filtration may disrupt the equilibrium distribution of clusters which then quickly reform under these conditions.

3.2 NTA measurements

NTA measurements provide us with estimated number concentrations of detected objects and corresponding number-based distribution of hydrodynamic radii of individual objects. We note that range of the sizes that could be analysed using NTA is reported to be from around 20 nm up to 1 μm (depending on optical contrast material and solvent type), thus precluding the detection of molecular clusters.

Figure 2 shows the hydrodynamic radius distribution of mesoclusters for unfiltered and filtered glycine solutions of varying concentrations. From these distributions we can see that while the mode size is similar between 60 and 80 nm for all glycine concentrations, distributions become significantly broader with increasing glycine concentration, with the distribution tail extending towards 250 nm. A comparison of mean hydrodynamic radii of

mesoclusters measured by NTA and those measured by DLS is shown in Figure 4. While the NTA mean hydrodynamic radius is the number based average, the DLS mean hydrodynamic radius reports the intensity weighted average[55, 56], which is more sensitive toward larger objects which are weighted proportionally to their mass squared. The intensity weighted average is higher than the number based average unless all objects have the same size, which is not case here (Figure 2), and therefore the DLS mean hydrodynamic radius is expected to be higher than the NTA mean hydrodynamic radius as seen in Figure 3. Furthermore, small number of large objects would be hardly detected by NTA provides number-based size distributions based in tracking individual objects, while there may be a significant effect on DLS intensity and mean size. This may explain why there is a little difference between NTA measurements of filtered and unfiltered solutions compared to DLS measurements.

In Figure 4, the scattered intensity measured via DLS is compared to the total number concentration measured via NTA for filtered and unfiltered solutions. The total number concentration increases with glycine concentration, indicating that the scatterer population increases with concentration. The increase in scatterer population is further supported by the fact that for both filtered and unfiltered solutions of the same concentration, the overall particle counts are similar.

3.3 NMR measurements

^1H NMR-DOSY measurements provide us with values of diffusion coefficients for species detected by ^1H NMR. Measurements were performed with aqueous solutions of glycine at concentrations between 10.2 mg/g and 206 mg/g which correspond to undersaturated

conditions in both H₂O and D₂O [57]. The measured diffusion coefficients were similar to those measured in aqueous (H₂O) glycine solutions by Huang et al[54]. The mean hydrodynamic radii of glycine molecular clusters were calculated using the Stokes-Einstein equation and the diffusion coefficient values obtained from DOSY (Table 1), with the viscosity taken as that of a 7:3 H₂O:D₂O w/w mixture at 300 K ($\eta = 0.91051$ cp @ 300 K), with the D₂O viscosity ($\eta = 1.0447$ cp at 300 K) determined from values in Millero et al[58] and the H₂O viscosity $\eta = 1,0016$ cp at 300 K. From these calculations the hydrodynamic radius was found to vary from 0.241 to 0.345 nm (Table 1 and Figure 7). We note that the mean hydrodynamic radius obtained from NMR-DOSY measurements is a number-based average in case that multiple species contribute to the same signal, as the NMR signal intensity is proportional to the number of nuclei with the same chemical shift. For example, if a signal comes from both glycine monomer and dimer, it would be proportional to the total number of glycine molecules contained in all monomers and dimers per unit volume.

3.4 SAXS measurements

SAXS measurements provide us with angle dependent scattered intensities $I(Q)$, where the angular dependence is presented in terms of the scattering vector $Q = (4\pi/\lambda)\sin(\theta/2)$, where λ is the beam wavelength and θ is the scattering angle. By recording the elastic scattering of x-rays at low angles, SAXS allows for the acquisition of information about the size and structures of species within a sample that presents inhomogeneities in the nm range. $I(Q)$ data can be analysed to obtain the mean radius of gyration as well as structural information by fitting data with appropriate theoretical form factors.

SAXS measurements were performed with glycine aqueous solutions at concentrations ranging from 10 to 230 mg/g, filtered using 0.1µm PTFE syringe filters. SAXS results confirm the presence of small molecular clusters (high Q value) and large mesoscale clusters (low Q value) in undersaturated glycine solutions (Figure 5), consistent with previous observations from supersaturated glycine solutions [8,12]. This is also in line with results obtained via DLS discussed above.

Guinier analysis was performed in the Q range between 0.041-0.29 Å⁻¹ (for high Q measurements – see supplementary info). The linear plots are produced by using the following:

$$\ln[I(Q)] = \ln[I_0] - \frac{R_g^2}{3} Q^2 \quad \text{Equation 1}$$

From the linear fit the mean radius of gyration (R_g) for the molecular clusters can be estimated. In Figure 6 we show these radii together with intensities $I(Q)$ for $Q=0.05$ (approximating a low Q limit of the scattering intensity for molecular clusters scattering at high Q values) as a function of glycine concentration. The radii are similar to the value from a more concentrated glycine solution (3.6 M) where the mean radius of gyration was reported as 0.34 nm [12]. We note that for a glycine monomer and dimer, R_g is expected to be around 0.29-0.3 nm and 0.37-0.4 nm, respectively [12].

In Figure 7 we show the mean radii of gyration of molecular clusters obtained from SAXS compared with the mean hydrodynamic radii of molecular clusters obtained from DLS and the mean hydrodynamic radii obtained from NMR-DOSY. The radius of gyration is defined as the root mean square of distance from the centre of mass, and is thus expected to be lower than the hydrodynamic radius (for example for a sphere the ratio between the two is $(3/5)^{1/2} = 0.77$). In terms of averaging over multiple species present, the mean radius

of gyration is based on species contributions weighted by the species mass squared, similarly to the mean hydrodynamic radius[55, 56]. On the other hand, the mean hydrodynamic radius from NMR-DOSY is based on species contributions weighted by the number of atoms, i.e., the species mass. Therefore, if only one species is contributing, the mean hydrodynamic radius from DLS and NMR should be close to each other, as we see is the case at the low glycine concentration when glycine dimers are barely present. As glycine concentration increases and more glycine dimers are formed, the DLS values become higher than NMR-DOSY values as expected. However, results from SAXS are not consistent with DLS and NMR: while the radius of gyration should be lower than the hydrodynamic radius, it should also increase with increasing glycine concentration due to more dimers forming, but this is not the case. In fact, the radius of gyration from SAXS seems to be close to that expected for the monomer and the lowest glycine concentration it appears to be too high. We can conclude that the trend observed for the radii of gyration from SAXS is unexplained, although the magnitude is in the expected range.

By using experimentally measured size distribution of mesoclusters obtained from NTA, the SAXS intensity $I(Q)$ corresponding to various structural models was calculated for the entire ensemble of mesoclusters, for the full range of concentrations of analysed glycine solutions.

Figure 8 shows how the mesocluster scattering intensity was obtained from experimental $I(Q)$ data. The signal relating to mesoclusters (grey solid line) was obtained by subtracting the SAXS intensity at low Q values (signal from both mesoclusters and molecular clusters, experimental data, black solid line) from those at high Q values (signal from molecular clusters, grey dashed line). The specific mesocluster intensity signal is denoted here as

$I(Q)_o$ and defined as the intensity value for the best fit d_f at $Q(\text{\AA}^{-1}) = 1 \times 10^{-5}$ minus the intensity value of the high Q plateau region value (Table 2 and straight dashed line as seen in Figure 10). Results for calculated $I(Q)_o$ values for the various concentrations of aqueous glycine solutions measured via SAXS can be found in Table 2. Differences in the $I(Q)_o$ values obtained from high and low Q measurements are mostly due to the fitting of the extrapolation towards $Q(\text{\AA}^{-1}) = 1 \times 10^{-5}$, the trend, however, is the same in both cases, with the $I(Q)_o$ value decreasing with concentration. The exception here is the value at 230 mg/g, which was calculated using data from unfiltered NTA results, whereas the remaining data was calculated using NTA data from filtered solutions.

Figure 8 also shows an example of the calculated scattering intensity for both a smooth sphere and a mass fractal model, for a 100 mg/g glycine aqueous solution. The form factor for a smooth sphere with radius R is given by the following equation:

$$P(Q) = \left[3 \frac{(QR) - ((QR) \cos(QR))}{(QR)^3} \right]^2 \quad \text{Equation 2}$$

This model does not have any adjustable parameters and gave a poor fit to the SAXS data, as the slope in the power law region was much less -4, the value expected for smooth spheres. In order to better describe potential cluster structures, we considered an exponential form factor for a mass fractal with radius R and fractal dimension d_f given by[31]:

$$P(Q) = \frac{\sin[(d_f-1)\tan^{-1}(Q\xi)]}{(d_f-1)Q\xi(1+Q^2\xi^2)^{\frac{d_f-1}{2}}} \quad \text{Equation 3}$$

where ξ is the cut off, defined as:

$$\xi^2 = \frac{2R^2}{d_f(d_f+1)} \quad \text{Equation 4}$$

The individual cluster structure factors obtained via intensity-weighted averaging are used in the fractal fit, which was performed using the dimensionless average structure factor[59] employed for experimental data obtained via dynamic and static light scattering:

$$P(Q) = \frac{P(Q)_i N_i m_i^2}{\sum N_i m_i^2} \quad \text{Equation 5}$$

where $m_i = R_i^{d_f}$ is the mass of a cluster with radius R_i , $P(Q)_i$ is the corresponding form factor obtained from Eq. 3 and N_i is the cluster number concentration obtained from NTA. The only adjustable parameter in this model is the cluster mass fractal dimension d_f (cf. Figure 8).

Overall best fit results for SAXS data obtained for high and low Q data is shown in Table 2, resulting in estimated fractal dimensions between 2.2 and 2.5 for all glycine concentrations where suitable data were available. We note that SAXS data for the lowest glycine concentrations (10 mg/g) were too noisy within the low Q region and thus it was not possible to estimate any properties of mesoscale clusters from SAXS data at such low concentrations.

Fractal dimensions around 2.5 correspond to fairly compact domains with irregular surfaces and are typical for percolation clusters, while lower fractal dimensions around 2.2 indicate more open structures. This range of fractal dimension indicates that percolation and/or reversible aggregation or restructuring (Lazzari et al., 2016) may be responsible for formation of mesoscale clusters in undersaturated glycine solutions. We also note that higher fractal dimensions around 2.7, corresponding to more compact,

denser structures, were previously estimated for mesoscale clusters in supersaturated glycine solutions [4,15].

4. Conclusions

Our results show that undersaturated glycine aqueous solutions are mesostructured liquid phases where hydrated glycine monomer and dimers (molecular clusters) are present alongside mesoscale clusters with radii in colloidal domain (100-150 nm) across a wide range of glycine concentrations (1-230 mg/g).

Molecular clusters were detected by three complementary experimental methods: DLS, NMR-DOSY and SAXS. Mean hydrodynamic radii of molecular clusters are increasing with glycine concentration due to increasing fraction of glycine present in hydrogen-bonded glycine dimers, as shown by both DLS and NMR-DOSY measurements. Surprisingly, mean radii of gyration from SAXS were found to be independent of glycine concentration, and this discrepancy with the other two methods remains currently unexplained.

Mesoscale clusters were also detected by three separate experimental methods: DLS, NTA and SAXS. Mean hydrodynamic radii of mesoscale clusters were measured by both DLS and NTA, but NTA also provided their cluster size distribution and number concentrations. While mean hydrodynamic radii of mesoscale clusters varied very little with glycine concentrations, their size distributions extended significantly towards larger sizes with increasing glycine concentrations.

Using size distributions and concentrations from NTA measurements, and assuming the mesoscale clusters have mass fractal-like structure, we developed a structural model to

fit SAXS data with a single adjustable parameter: the cluster fractal dimension. Using this model we estimated fractal dimensions of mesoscale clusters across a range of glycine concentrations (50-230 mg/g).

In this work we demonstrated how multiple experimental techniques can be used in conjunction to provide quantitative insights into phase behaviour and properties of mesostructured liquid phases. Developing better experimental and theoretical understanding of complex clustering phenomena in solutions of small molecules is crucial for further development of rational design and control of self-assembly and nucleation processes.

Acknowledgments

This project was supported by EPSRC funding via the SynBIM project (Grant Reference EP/P006892/1) and by the Synchrotron SOLEIL.

The authors would like to declare there are no financial or non-financial competing interests with regards to the work presented in this paper.

References

1. Cerreta, M.K. and K.A. Berglund, *The structure of aqueous solutions of some dihydrogen orthophosphates by laser Raman spectroscopy*. Journal of Crystal Growth, 1987. **84**(4): p. 577-588.
2. Rusli, I.T., G.L. Schrader, and M.A. Larson, *Raman spectroscopic study of NaNO₃ solution system - solute clustering in supersaturated solutions*. Journal of Crystal Growth, 1989. **97**(2): p. 345-351.
3. Hassan, S.A., *Morphology of ion clusters in aqueous electrolytes*. Physical Review E, 2008. **77**(3): p. 031501.
4. Chattopadhyay, S., et al., *SAXS Study of the Nucleation of Glycine Crystals from a Supersaturated Solution*. Crystal Growth & Design, 2005. **5**(2): p. 523-527.
5. Georgalis, Y., A.M. Kierzek, and W. Saenger, *Cluster Formation in Aqueous Electrolyte Solutions Observed by Dynamic Light Scattering*. The Journal of Physical Chemistry B, 2000. **104**(15): p. 3405-3406.
6. Erdemir, D., et al., *Relationship between Self-Association of Glycine Molecules in Supersaturated Solutions and Solid State Outcome*. Physical Review Letters, 2007. **99**(11): p. 115702.
7. Gebauer, D., A. Volkel, and H. Colfen, *Stable prenucleation calcium carbonate clusters*. Science, 2008. **322**(5909): p. 1819-22.
8. Davey, R.J., S.L. Schroeder, and J.H. ter Horst, *Nucleation of organic crystals--a molecular perspective*. Angew Chem Int Ed Engl, 2013. **52**(8): p. 2166-79.
9. Jawor-Baczynska, A., B.D. Moore, and J. Sefcik, *Effect of mixing, concentration and temperature on the formation of mesostructured solutions and their role in the nucleation of DL-valine crystals*. Faraday Discuss, 2015. **179**: p. 141-54.
10. Jawor-Baczynska, A., et al., *Population and size distribution of solute-rich mesospecies within mesostructured aqueous amino acid solutions*. Faraday Discussions, 2013(167): p. 425-440.
11. Niimura, N., et al., *Aggregation in supersaturated lysozyme solution studied by time-resolved small angle neutron scattering*. Journal of Crystal Growth, 1995. **154**(1): p. 136-144.
12. Muschol, M. and F. Rosenberger, *Liquid-liquid phase separation in supersaturated lysozyme solutions and associated precipitate formation/crystallization*. The Journal of Chemical Physics, 1997. **107**(6): p. 1953-1962.
13. Haas, C. and J. Drenth, *Understanding protein crystallization on the basis of the phase diagram*. Journal of Crystal Growth, 1999. **196**(2): p. 388-394.
14. Gebauer, D., et al., *Pre-nucleation clusters as solute precursors in crystallisation*. Chemical Society Reviews, 2014. **43**(7): p. 2348-2371.
15. Jawor-Baczynska, A., J. Sefcik, and B.D. Moore, *250 nm Glycine-Rich Nanodroplets Are Formed on Dissolution of Glycine Crystals But Are Too Small To Provide Productive Nucleation Sites*. Crystal Growth & Design, 2013. **13**(2): p. 470-478.
16. Grouazel, S., et al., *BPTI liquid-liquid phase separation monitored by light and small angle X-ray scattering*. Acta Crystallographica Section D, 2002. **58**(10 Part 1): p. 1560-1563.

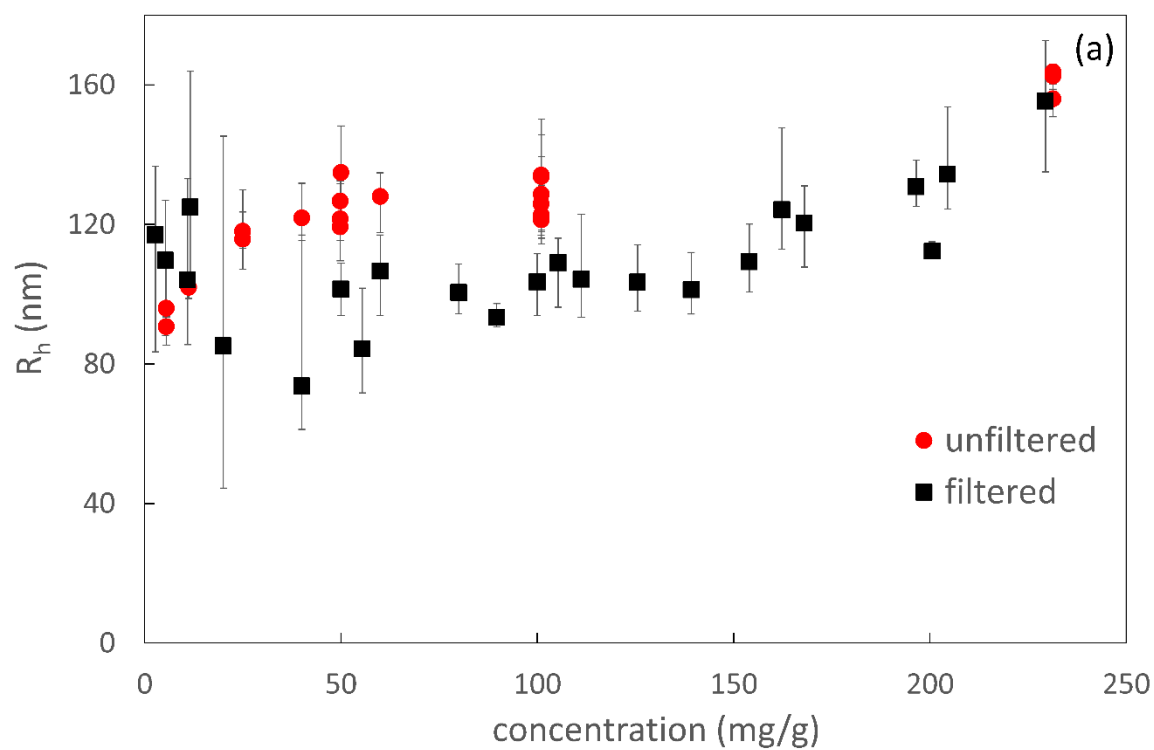
17. Onuma, K. and N. Kanzaki, *Multi-angle static and dynamic light scattering investigation of lysozyme association: From crystallization to liquid-liquid phase separation*. Journal of Crystal Growth, 2007. **304**(2): p. 452-459.
18. Gliko, O., et al., *A Metastable Prerequisite for the Growth of Lumazine Synthase Crystals*. Journal of the American Chemical Society, 2005. **127**(10): p. 3433-3438.
19. Gliko, O., et al., *Metastable Liquid Clusters in Super- and Undersaturated Protein Solutions*. The Journal of Physical Chemistry B, 2007. **111**(12): p. 3106-3114.
20. Taratuta, V.G., et al., *Liquid-liquid phase separation of aqueous lysozyme solutions: effects of pH and salt identity*. The Journal of Physical Chemistry, 1990. **94**(5): p. 2140-2144.
21. Ishimoto, C. and T. Tanaka, *Critical Behavior of a Binary Mixture of Protein and Salt Water*. Physical Review Letters, 1977. **39**(8): p. 474-477.
22. Galkin, O. and P.G. Vekilov, *Are Nucleation Kinetics of Protein Crystals Similar to Those of Liquid Droplets?* Journal of the American Chemical Society, 2000. **122**(1): p. 156-163.
23. Galkin, O. and P.G. Vekilov, *Control of protein crystal nucleation around the metastable liquid-liquid phase boundary*. Proceedings of the National Academy of Sciences, 2000. **97**(12): p. 6277-6281.
24. Galkin, O., et al., *Liquid-Liquid Separation in Solutions of Normal and Sickle Cell Hemoglobin*. Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(13): p. 8479-8483.
25. Galkin, O., et al., *Two-step mechanism of homogeneous nucleation of sickle cell hemoglobin polymers*. Biophys J, 2007. **93**(3): p. 902-13.
26. Yau, S.T., B.R. Thomas, and P.G. Vekilov, *Real time, in-situ, monitoring of apoferritin crystallization and defect formation with molecular resolution*. Journal of Crystal Growth, 2001. **232**(1): p. 188-194.
27. Filobelo, L.F., O. Galkin, and P.G. Vekilov, *Spinodal for the solution-to-crystal phase transformation*. The Journal of Chemical Physics, 2005. **123**(1): p. 014904.
28. Groen, H. and K.J. Roberts, *Nucleation, Growth, and Pseudo-Polymorphic Behavior of Citric Acid As Monitored in Situ by Attenuated Total Reflection Fourier Transform Infrared Spectroscopy*. The Journal of Physical Chemistry B, 2001. **105**(43): p. 10723-10730.
29. Bonnett, P.E., et al., *Solution crystallisation via a submerged liquid-liquid phase boundary: oiling out*. Chemical Communications, 2003(6): p. 698-699.
30. Heeley, E.L., et al., *Are metastable, precrystallisation, density-fluctuations a universal phenomena?* Faraday Discussions, 2003. **122**(0): p. 343-361.
31. Sorensen, T.J., et al., *Cluster Formation in Precrystalline Solutions*. Chemical Engineering & Technology, 2003. **26**(3): p. 307-312.
32. Veessler, S., et al., *Phase Transitions in Supersaturated Drug Solution*. Organic Process Research & Development, 2003. **7**(6): p. 983-989.
33. Veessler, S., et al., *Crystallization in the Presence of a Liquid-Liquid Phase Separation*. Organic Process Research & Development, 2006. **10**(4): p. 841-845.
34. Lafferrère, L., C. Hoff, and S. Veessler, *In Situ Monitoring of the Impact of Liquid-Liquid Phase Separation on Drug Crystallization by Seeding*. Crystal Growth & Design, 2004. **4**(6): p. 1175-1180.

35. Lafferrère, L., C. Hoff, and S. Veessler, *Study of liquid–liquid demixing from drug solution*. Journal of Crystal Growth, 2004. **269**(2): p. 550-557.
36. Deneau, E. and G. Steele, *An In-Line Study of Oiling Out and Crystallization*. Organic Process Research & Development, 2005. **9**(6): p. 943-950.
37. Waizumi, K. and T. Eguchi, *Novel Observation of Nucleation and Growth of Insulin Crystals via Liquid Droplets Generated by Liquid–Liquid Phase Separation*. Chemistry Letters, 2005. **34**(12): p. 1654-1655.
38. Kimura, M., *Characterization of the Dense Liquid Precursor in Homogeneous Crystal Nucleation Using Solution State Nuclear Magnetic Resonance Spectroscopy*. Crystal Growth & Design, 2006. **6**(4): p. 854-860.
39. Wang, X., J.M. Gillian, and D.J. Kirwan, *Quasi-Emulsion Precipitation of Pharmaceuticals. 1. Conditions for Formation and Crystal Nucleation and Growth Behavior*. Crystal Growth & Design, 2006. **6**(10): p. 2214-2227.
40. Kiesow, K., F. Ruether, and G. Sadowski, *Solubility, crystallization and oiling-out behavior of PEGDME: 1. Pure-solvent systems*. Fluid Phase Equilibria, 2010. **298**(2): p. 253-261.
41. Kiesow, K., F. Tumakaka, and G. Sadowski, *Experimental investigation and prediction of oiling out during crystallization process*. Journal of Crystal Growth, 2008. **310**(18): p. 4163-4168.
42. Hagemeyer, D., et al., *Direct experimental observation of the aggregation of [small alpha]-amino acids into 100-200 nm clusters in aqueous solution*. RSC Advances, 2012. **2**(11): p. 4690-4696.
43. Sedlák, M. and D. Rak, *Large-Scale Inhomogeneities in Solutions of Low Molar Mass Compounds and Mixtures of Liquids: Supramolecular Structures or Nanobubbles?* The Journal of Physical Chemistry B, 2013. **117**(8): p. 2495-2504.
44. Sedlák, M., *Large-Scale Supramolecular Structure in Solutions of Low Molar Mass Compounds and Mixtures of Liquids: I. Light Scattering Characterization*. The Journal of Physical Chemistry B, 2006. **110**(9): p. 4329-4338.
45. Sedlák, M., *Large-Scale Supramolecular Structure in Solutions of Low Molar Mass Compounds and Mixtures of Liquids: II. Kinetics of the Formation and Long-Time Stability*. The Journal of Physical Chemistry B, 2006. **110**(9): p. 4339-4345.
46. Sedlák, M., *Large-Scale Supramolecular Structure in Solutions of Low Molar Mass Compounds and Mixtures of Liquids. III. Correlation with Molecular Properties and Interactions*. The Journal of Physical Chemistry B, 2006. **110**(28): p. 13976-13984.
47. Sweatman, M.B. and L. Lue, *The giant SALR cluster fluid: a review*. Advanced Theory and Simulations, 2019.
48. Ferreiro-Rangel, C.A. and M. Sweatman, *Cluster formation in binary fluids with competing short-range and long-range interactions*. Molecular Physics, 2018. **116**(21-22): p. 3231-3244.
49. Sweatman, M.B. and L. Lue, *The cluster vapor to cluster solid transition*. J. Chem. Phys. 2016, **144**(17): 171102.
50. Sweatman, M.B., R. Fartaria, and L. Lue, *Cluster formation in fluids with competing short-range and long-range interactions*. The Journal of chemical physics, 2014. **140**(12): 124508.

51. Pusey, P.N., P. Lindner, and T. Zemb, *Dynamic Light Scattering*, in *Neutrons, X-rays and Light: Scattering Methods Applied to Soft Condensed Matter*. 2002, Elsevier Science B.V. p. 203.
52. *Analytical Software Operation Manual*. 2010, NanoSight LM 10 & NTA 2.1.
53. Orttung, W.H., *POLARIZABILITY AND APPARENT RADIUS OF GLYCINE FROM REFRACTIVE INDEX DATA*. *The Journal of Physical Chemistry*, 1963. **67**(5): p. 1102-1105.
54. Huang, J., T.C. Stringfellow, and L. Yu, *Glycine Exists Mainly as Monomers, Not Dimers, in Supersaturated Aqueous Solutions: Implications for Understanding Its Crystallization and Polymorphism*. *Journal of the American Chemical Society*, 2008. **130**(42): p. 13973-13980.
55. Sandkühler, P., J. Sefcik, and M. Morbidelli, *Scaling of the Kinetics of Slow Aggregation and Gel Formation for a Fluorinated Polymer Colloid*. *Langmuir*, 2005. **21**(5): p. 2062-2077.
56. Waldner, M.H., et al., *Initial growth kinetics and structure of colloidal aggregates in a turbulent coagulator*. *Powder technology*, 2005. **156**(2-3): p. 226-234.
57. Jelińska-Kazimierczuk, M. and J. Szydłowski, *Isotope effect on the solubility of amino acids in water*. *Journal of Solution Chemistry*, 1996. **25**(12): p. 1175-1184.
58. Millero, F.J., R. Dexter, and E. Hoff, *Density and viscosity of deuterium oxide solutions from 5-70.deg*. *Journal of Chemical & Engineering Data*, 1971. **16**(1): p. 85-87.
59. Lattuada, M., H. Wu, and M. Morbidelli, *A simple model for the structure of fractal aggregates*. *Journal of colloid and interface science*, 2003. **268**(1): p. 106-120.
60. ten Wolde, P.R. and D. Frenkel, *Enhancement of protein crystal nucleation by critical density fluctuations*. *Science*, 1997. **277**(5334): p. 1975-1978.
61. Anwar, J. and P.K. Boateng, *Computer simulation of crystallization from solution*. *Journal of the American Chemical Society*, 1998. **120**(37): p. 9600-9604.
62. Shore, J.D., D. Perchak, and Y. Shnidman, *Simulations of the nucleation of AgBr from solution*. *The Journal of Chemical Physics*, 2000. **113**(15): p. 6276-6284.

Figures (need to be in colour)

Figure 1. Hydrodynamic radius R_h (nm) of mesoclusters (a) and molecular clusters (b), and the total scatter intensity (c) versus the concentration of aqueous glycine solutions. Solutions were either filtered with 1 μm PTFE filters prior to DLS measurements (black \blacksquare), or not filtered at all (red \bullet).



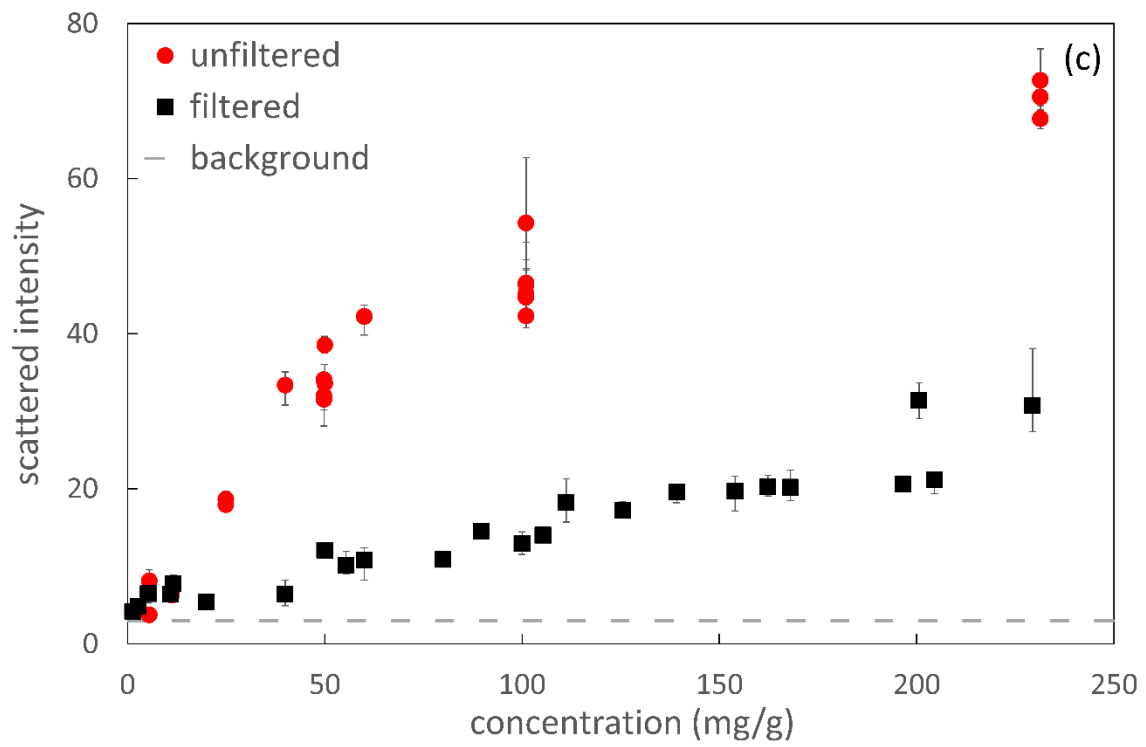
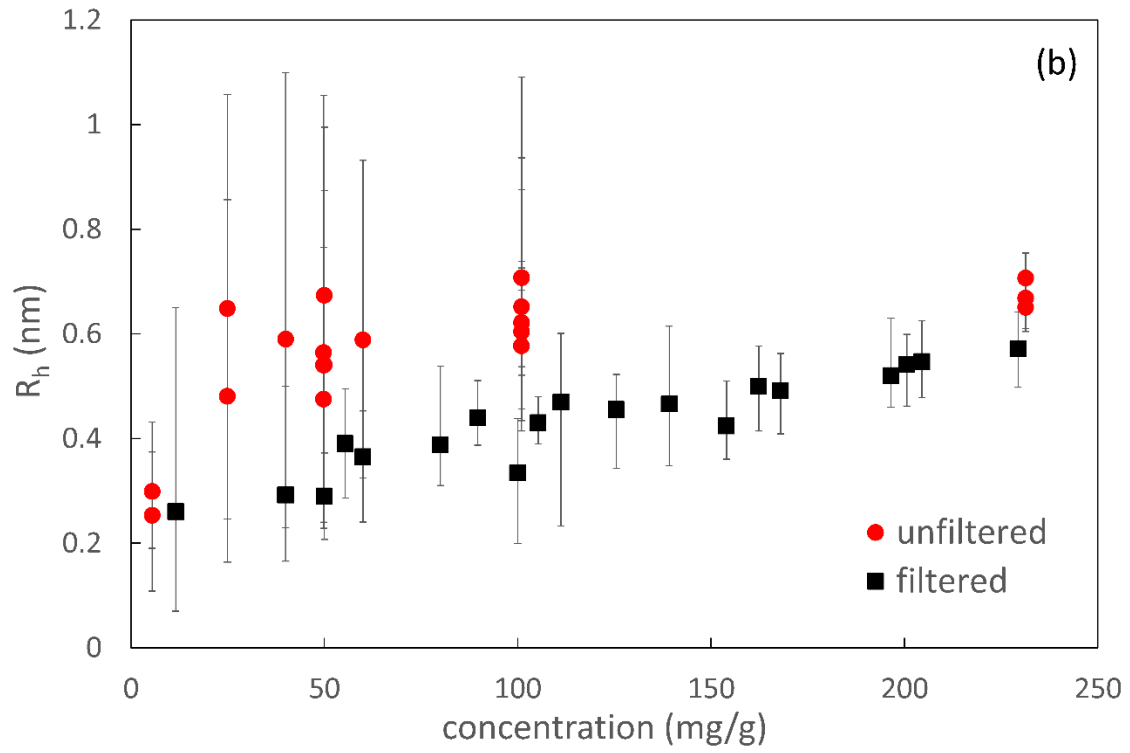


Figure 2. Number based hydrodynamic radius distributions of mesoscale clusters for selected unfiltered and filtered solutions from NTA measurements.

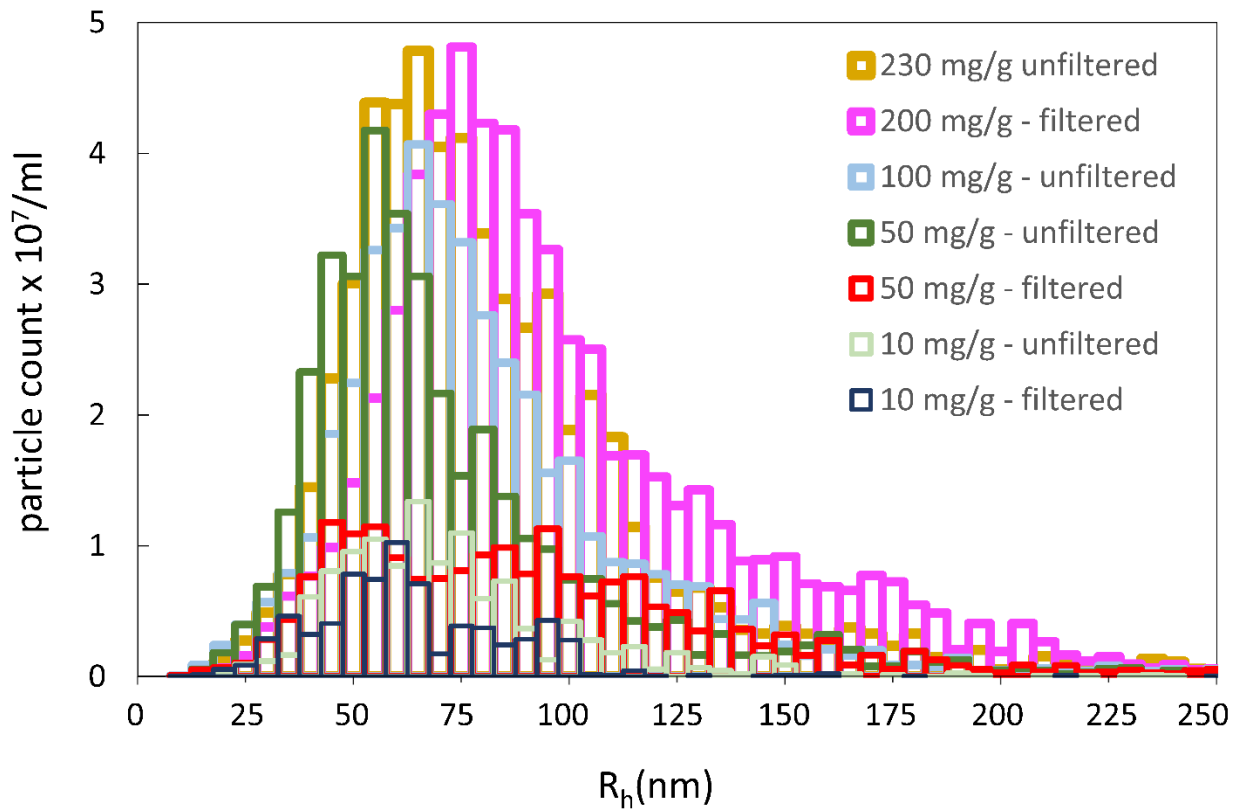


Figure 3. Mean hydrodynamic diameter vs aqueous glycine solution concentration, as measured via NTA (square, black), and DLS (circle, red). Block markers represent filtered solutions and open markers unfiltered solutions. NTA and DLS measurements here were performed on the same day.

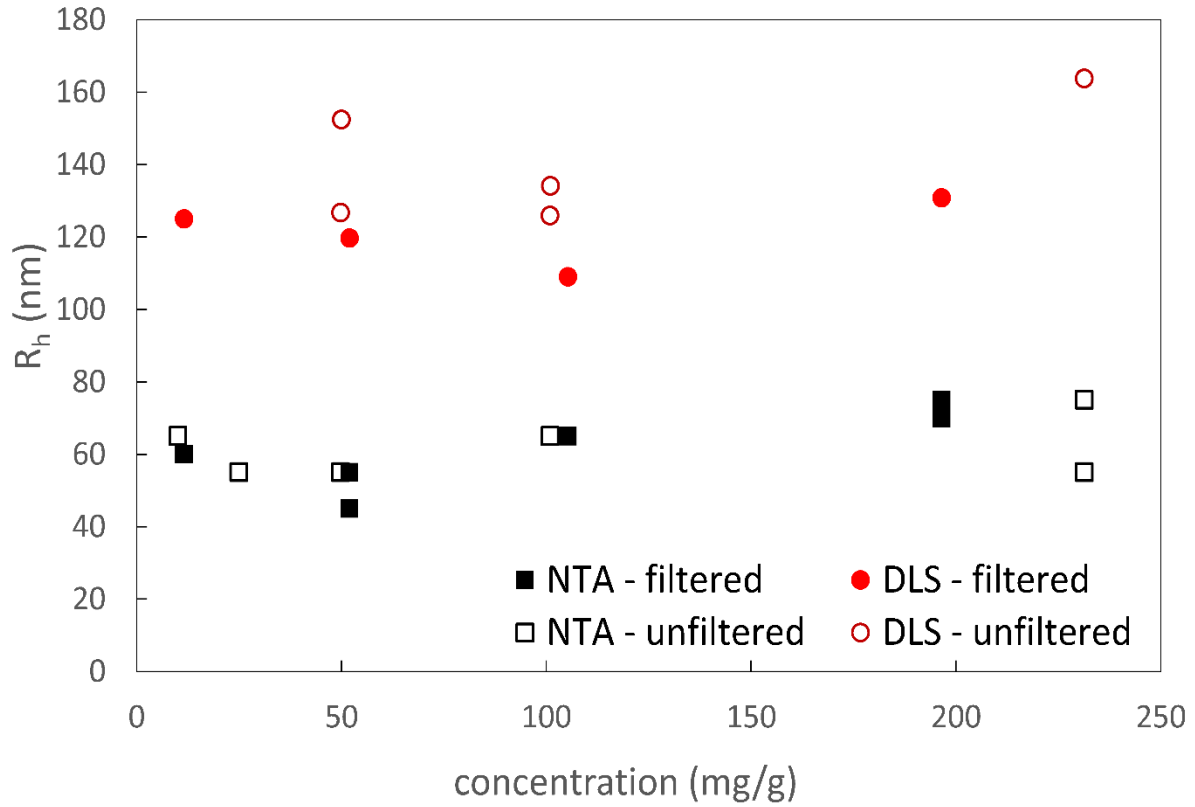


Figure 4. The total number concentration from NTA (red ●) with overall scattering intensity (normalised count rate) as obtained using DLS (black ■) for aqueous glycine solutions of different concentrations. DLS results are plotted as filtered (a) and non-filtered (b) samples.

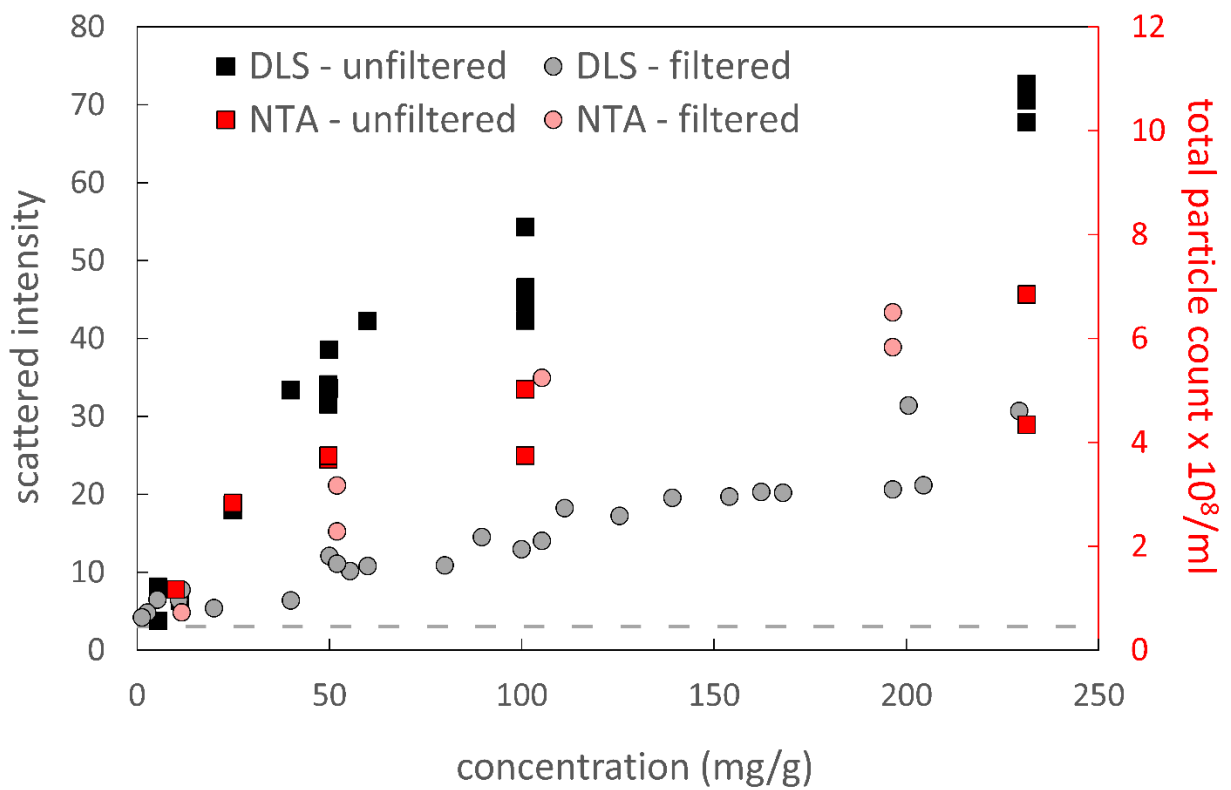


Figure 5. Variation of scattering intensity $I(Q)$ versus the momentum transfer $Q=4\pi/\lambda \sin(\theta/2)$ for various concentrations of filtered (0.1 μm PTFE) aqueous glycine solutions. (a) low Q measurement, (b) high Q measurements.

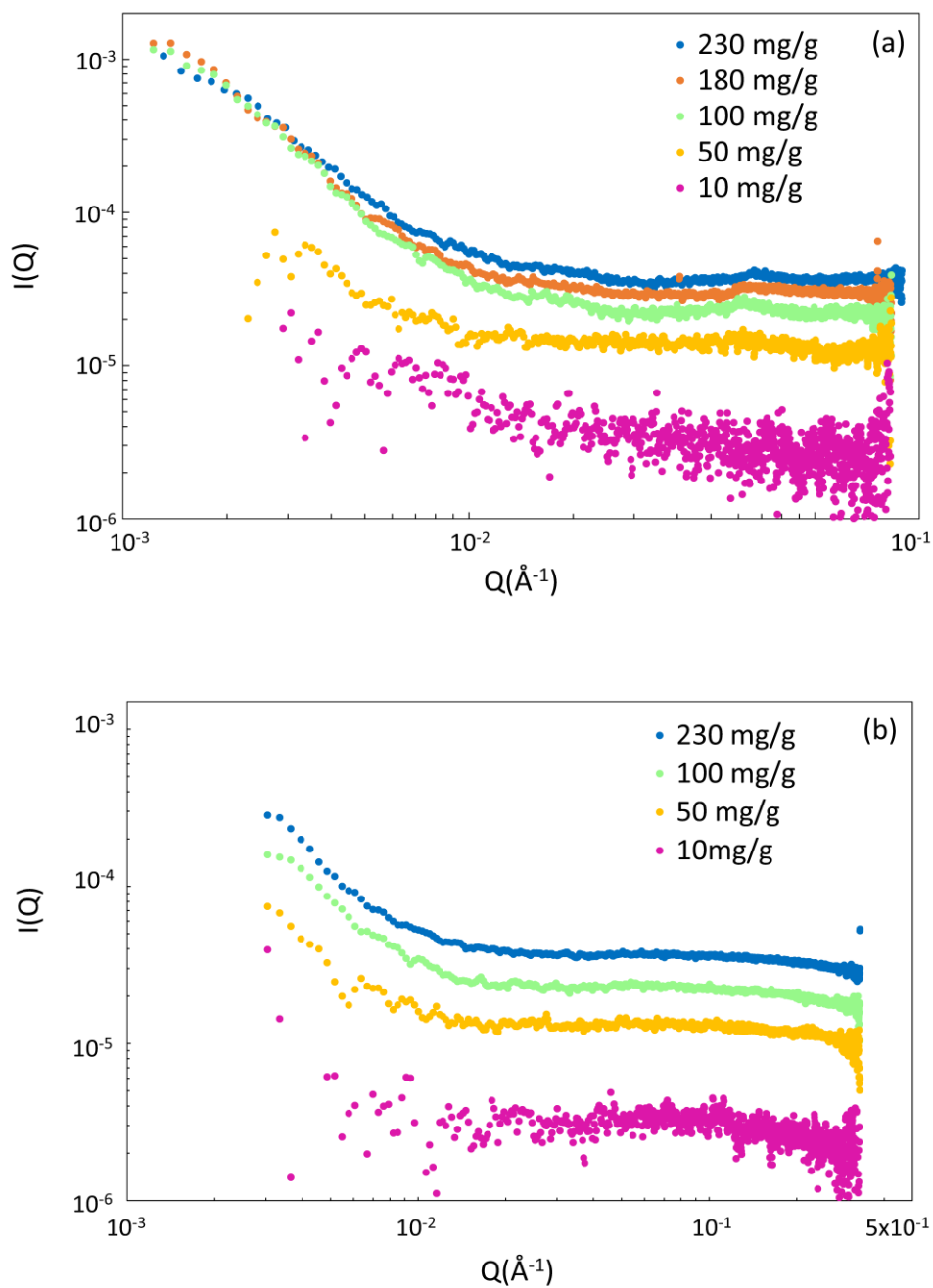


Figure 6. Variation in SAXS scattering intensity (black ■) and radius of gyration R_g (red ●) at $Q = 0.05 \text{ \AA}^{-1}$ (high Q measurements), over a concentration range of aqueous glycine solutions.

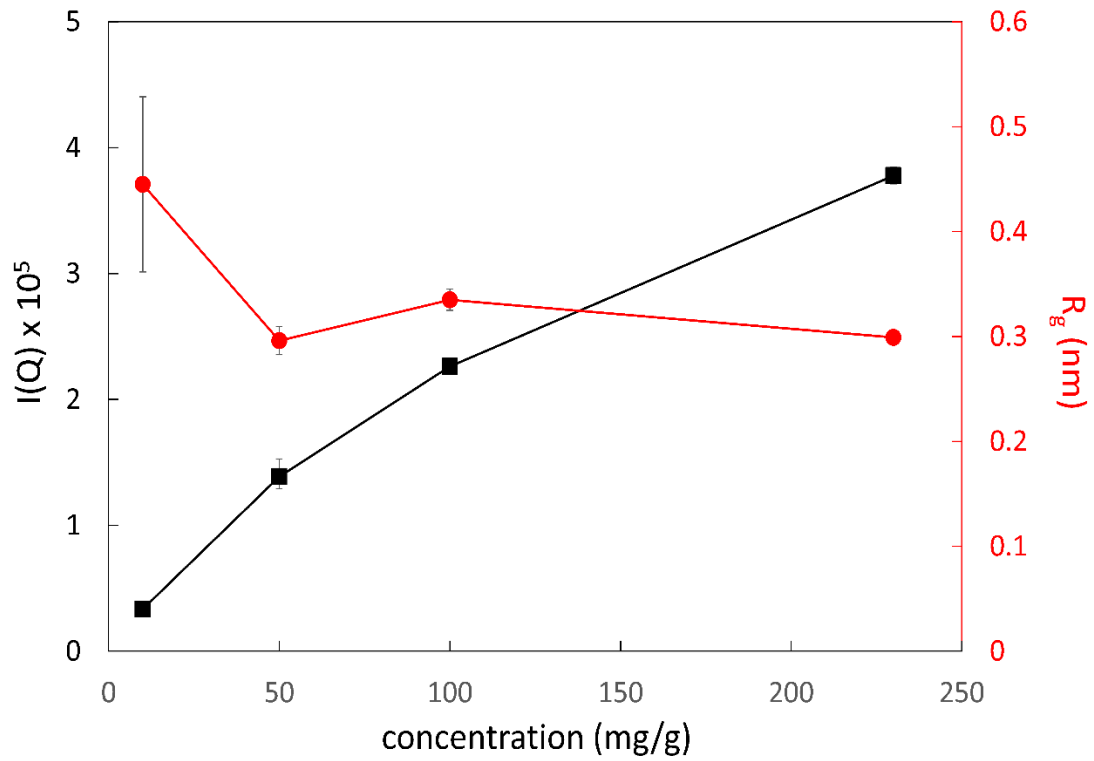


Figure 7. SAXS radius of gyration R_g (red ●) compared to molecular cluster hydrodynamic radius R_h from DLS (black ■) and NMR (grey ◆) data, for filtered solutions.

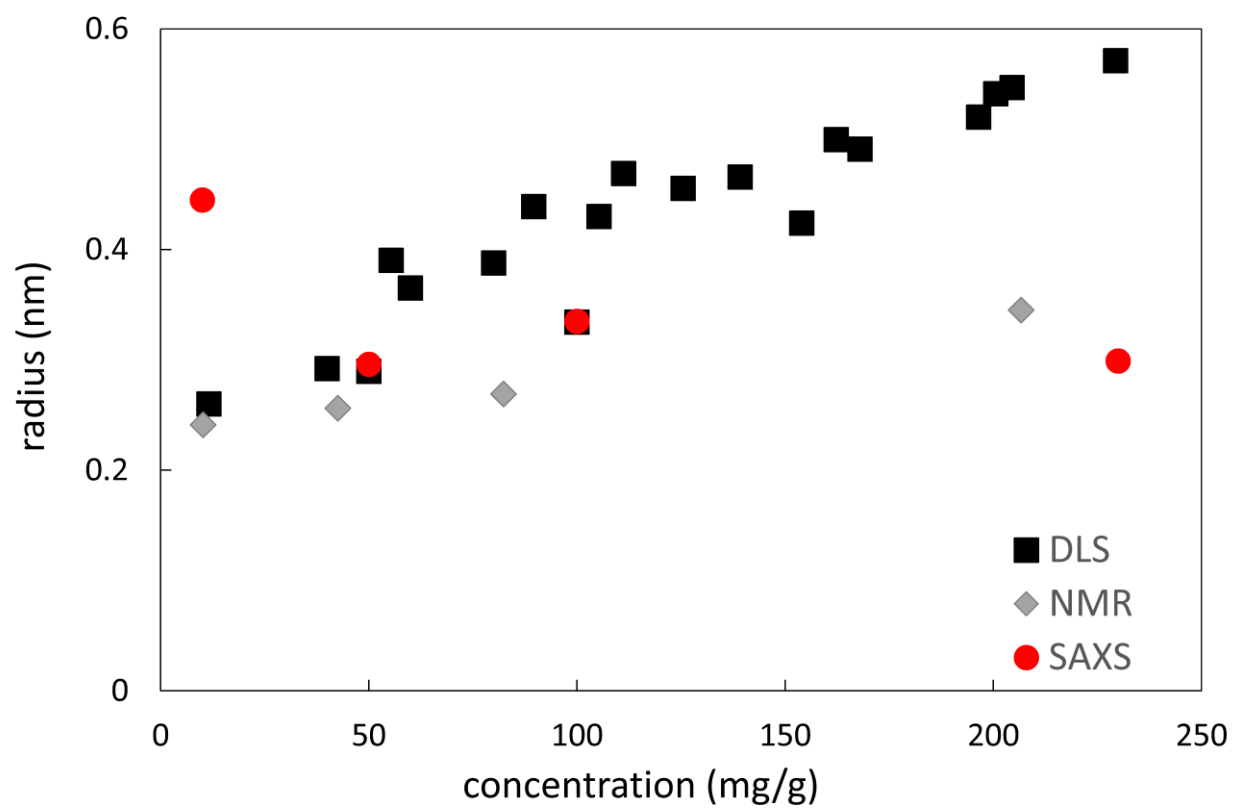
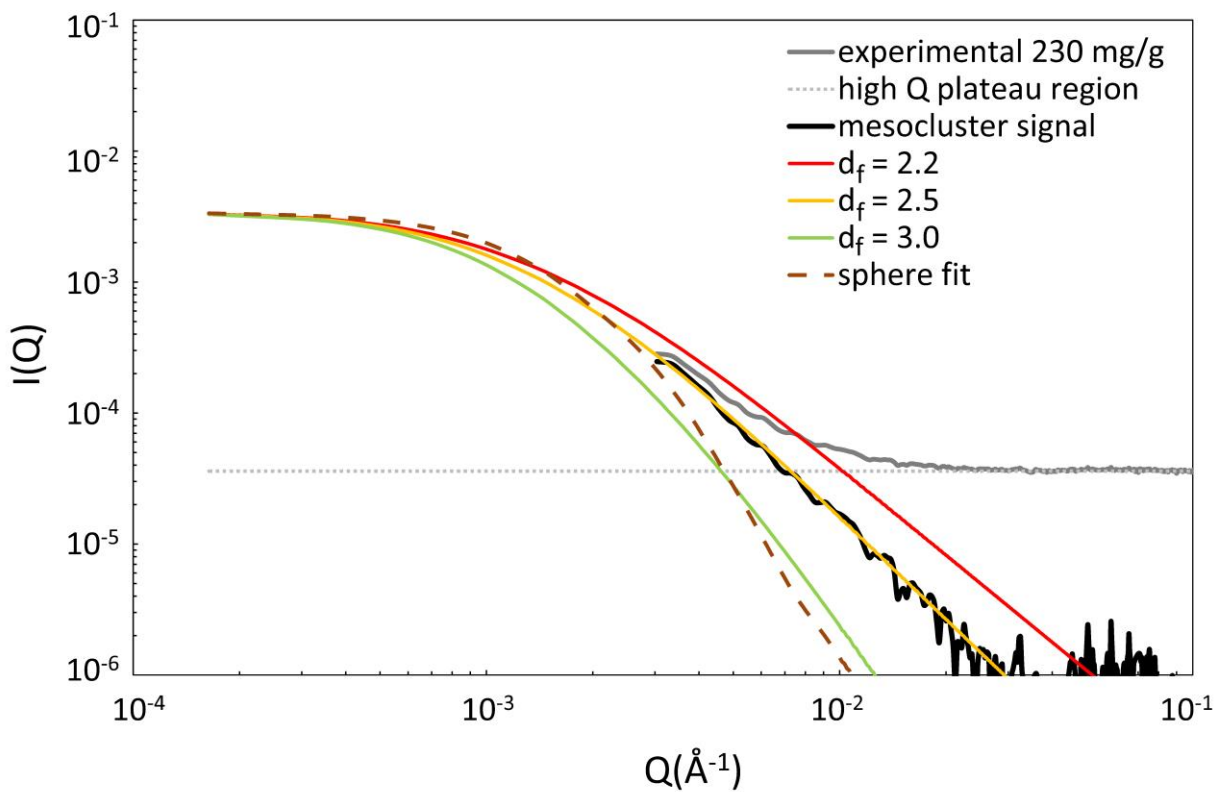


Figure 8. SAXS experimental data (grey solid line) for a 230 mg/g aqueous glycine solution, plotted with best fits using smooth sphere model (dashed brown line) and the mass fractal model for varying fractal dimension (coloured solid lines), with the best fit being obtained for $d_f = 2.5$. Included in the plot is the scattering from mesoscale clusters (black solid line), as determined after subtraction of the high Q plateau region (dashed grey line) from the low Q region.



Tables

Table 1: Diffusion coefficients from DOSY NMR and calculated hydrodynamic diameters for species present in aqueous glycine solutions of varying concentrations at 300 K.

concentration (mg/g)	diffusion coefficient $\times 10^{10}$ (m ² /s)	Hydrodynamic radius (nm)
206.7	7.0	0.345
82.4	8.96	0.269
42.6	9.42	0.256
10.2	10.0	0.241

Table 2: Fractal dimensions and $I(0)$ for molecular and mesoscale clusters estimated from High and Low Q SAXS data for varying concentrations of aqueous glycine solutions.

concentration (mg/g)	molecular cluster estimated $I(0) \times 10^5$	fractal dimension (d_f)		mesoscale cluster estimated $I(0) \times 10^3$	
		high Q data	low Q data	high Q data	low Q data
230	3.6	2.5	2.2	3.87	2.78
180	2.9	-	2.5	-	4.47
100	2.25	2.5	2.5	4.83	3.38
50	1.35	2.5	2.2	1.09	0.3

Supplementary Info

The autocorrelation data from DLS clearly shows two characteristic decay times (Figure S1), corresponding to two typical sizes of species present in the glycine aqueous solutions. For filtered solutions, the first decay, with a characteristic decay time around 0.0037 ms, corresponds to freely diffusing objects (molecular clusters). The second decay, appearing at approximately 2 ms, is indicative of the presence of larger, slower diffusing bodies (mesoclusters).

A Matlab script was prepared whereby exponential fits were performed on the autocorrelation curve and the subsequent radius of hydration for the two species present was obtained (Figure S2).

Figure S3 depicts a typical ^1H NMR spectrum for the aqueous glycine solutions used in this paper. $I(Q)_0$ values, along with the total number concentrations (from filtered solutions, NTA) and overall scattering intensity (from filtered solutions, DLS) are plotted in Figures S4-S5. $I(Q)_0$ initially begins to increase slowly with glycine concentration, showing a rapid increase after app. 50 mg/g. This coincides with the increase in radius seen for filtered solutions in DLS (Figure 2a), but also with changes occurring in the curves of the unfiltered samples in Figures S4 and S5.

Figure S1. Example of DLS autocorrelation function (black, solid line) and corresponding decay time distribution (red, dotted line) for a filtered 230 mg/g glycine aqueous solution at 19°C.

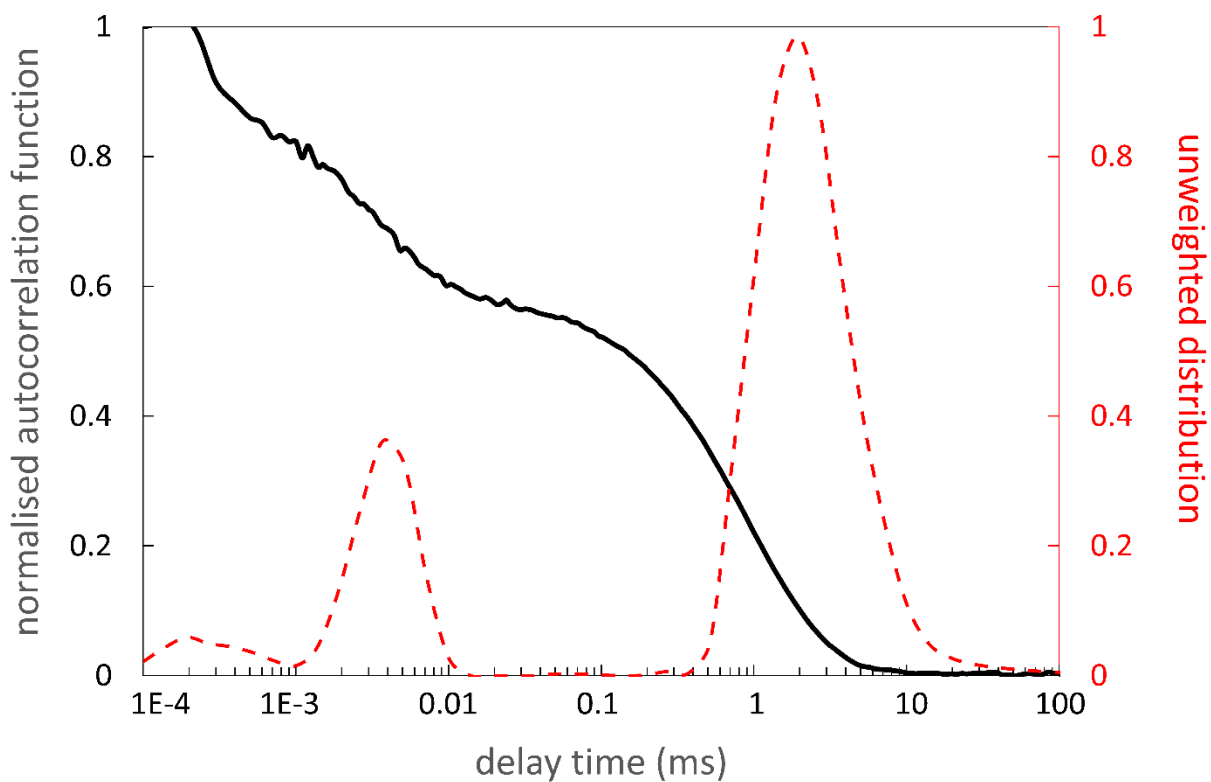


Figure S2. Example of 2-component exponential fit on an experimentally obtained DLS autocorrelation curve (black, solid line). The yellow and red lines are the fits on the experimental curve that correlate to the signals from molecular clusters and mesoclusters, respectively. The dashed blue line is the summation for the two fits, which reveals itself as being an extremely good overall fit to the experimental curve. The dashed grey line is an example of the water signal (millipore water, filtered with 0.2 μm PTFE filter). A filtered 230 mg/g glycine aqueous solution at 19 °C was used.

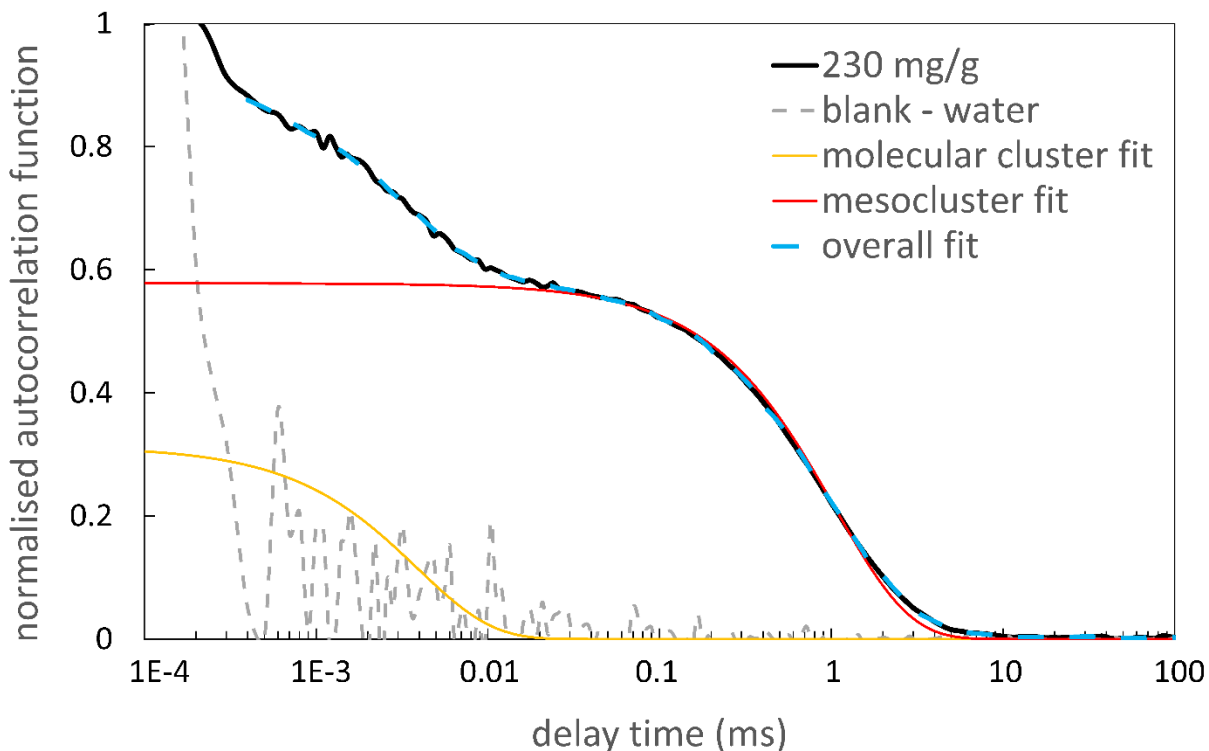
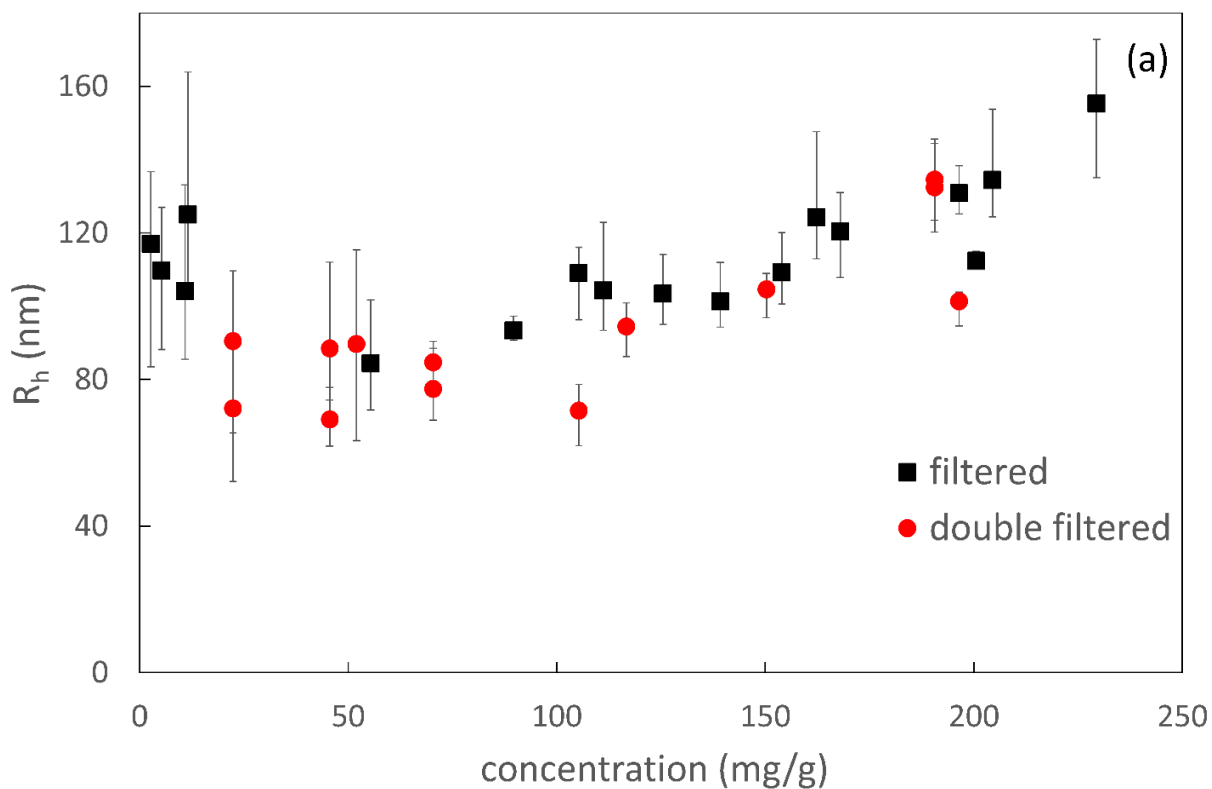


Figure S3. Hydrodynamic radius R_h (nm) of mesoclusters (a) and molecular clusters (b), and the total scatter intensity (c) versus the concentration of aqueous glycine solutions. All solutions were filtered with 1 μm PTFE filters prior to DLS measurements (black ■), and some solutions were filtered again (1 μm PTFE) when being placed into the DLS cells (red ●).



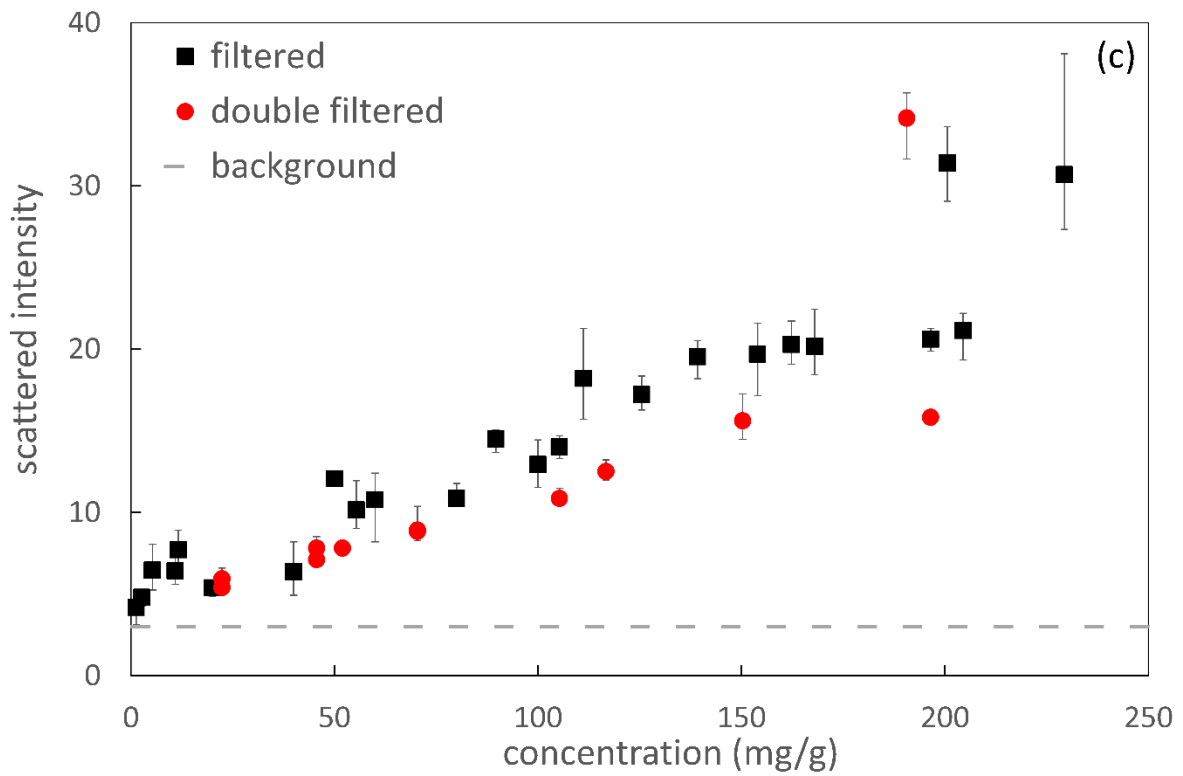
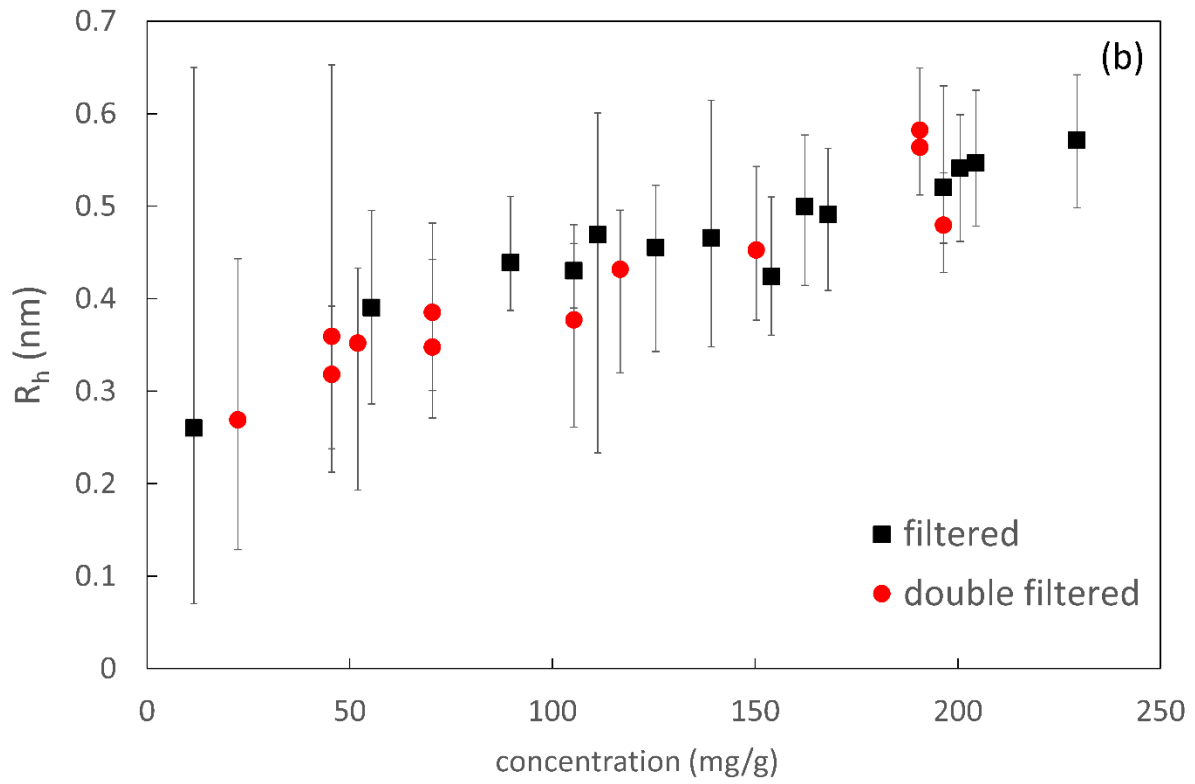


Figure S4. Typical ^1H NMR spectrum of glycine solution together with diffusion coefficient measurement (DOSY). Glycine concentration: 207 mg/g, temperature: 300 K, solvent: ratio $\text{D}_2\text{O}:\text{H}_2\text{O}$ of 3:7 w/w.

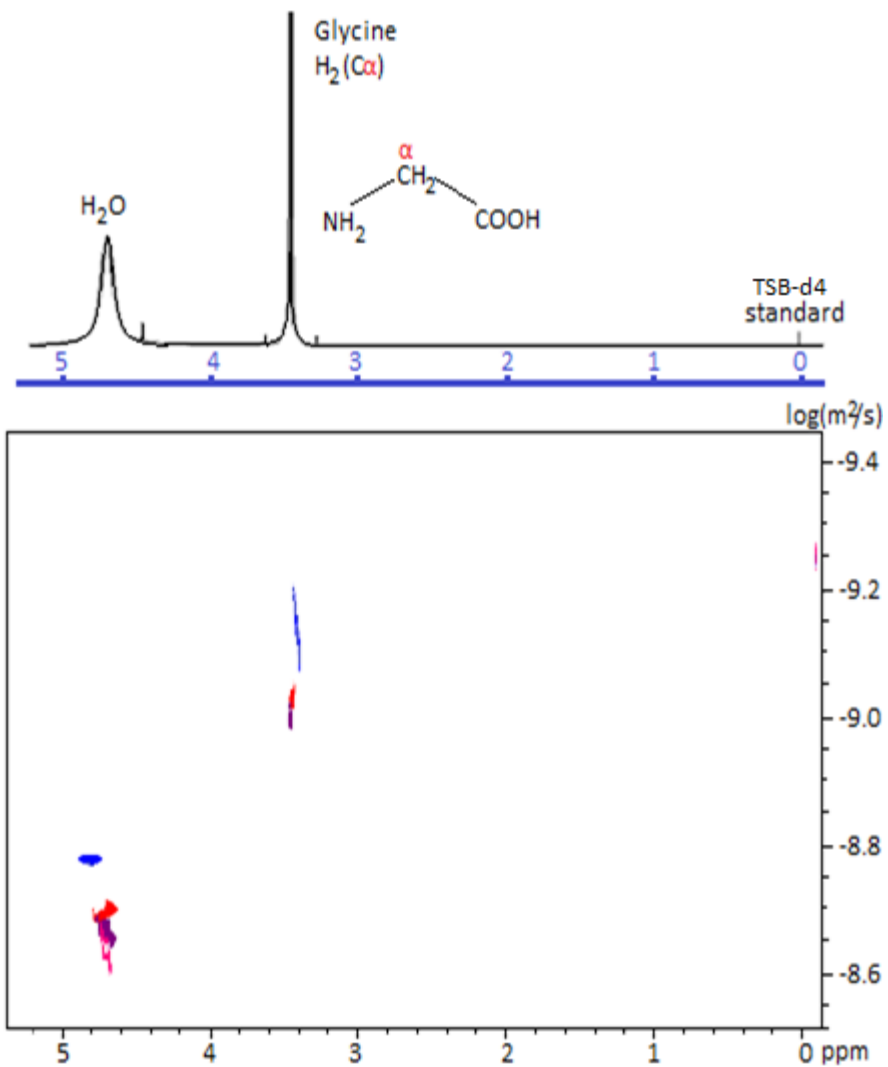


Figure S5. Guinier fits were performed on the scattering profile of aqueous glycine solutions (high Q measurements).

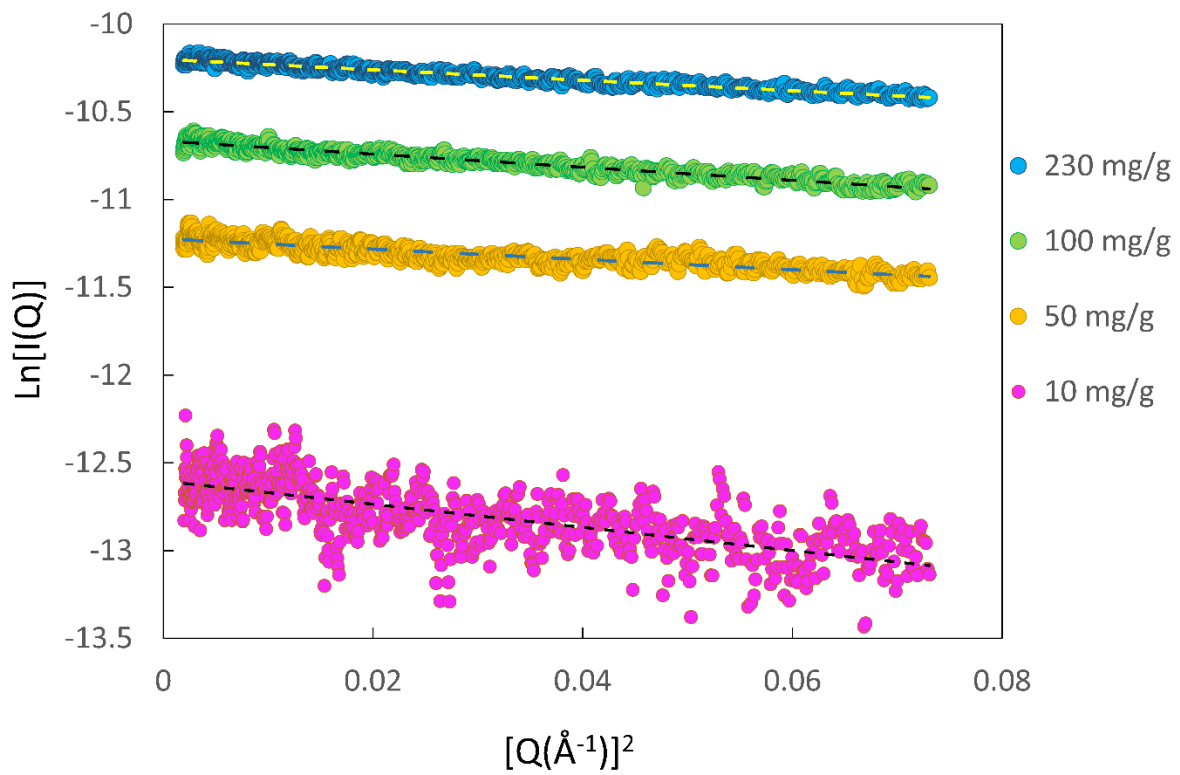


Figure S6. The measured scattering intensity for low $I(Q)_0$ from SAXS (black line) and the total number concentration from NTA for filtered (red solid line) and unfiltered (red dashed line) aqueous glycine solutions of different concentrations.

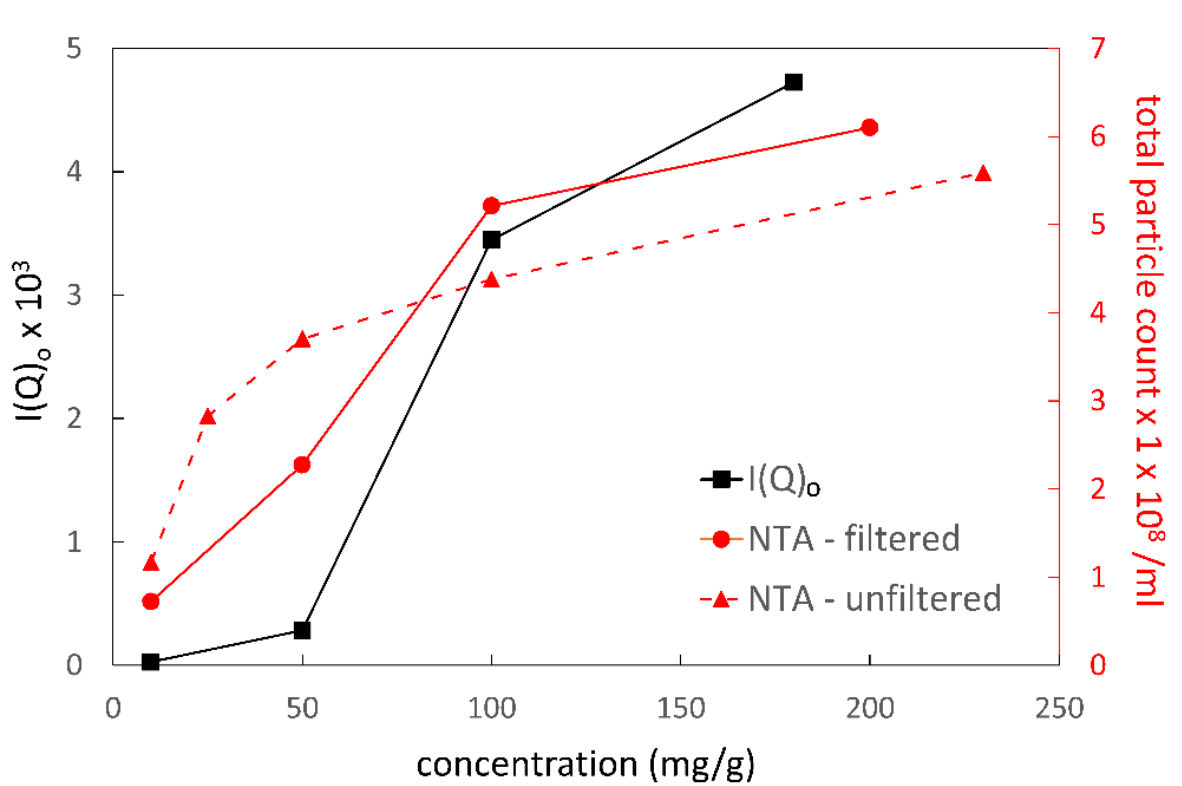


Figure S7. The measured scattering intensity for low $I(Q)_o$ from SAXS (black line) and the scattered intensity (NCR) from DLS for filtered (red solid line) and unfiltered (red dashed line) aqueous glycine solutions of different concentrations.

