Aggregation Analysis of Therapeutic Proteins, Part 2

Analytical Ultracentrifugation and Dynamic Light Scattering

Tsutomu Arakawa, John S. Philo, Daisuke Ejima, Kouhei Tsumoto, and Fumio Arisaka

Aggregation is a major problem for long-term storage stability of therapeutic proteins and their shipping and handling, as has been extensively reviewed (**1**). We discussed the mechanisms of protein aggregation at length in Part 1 of this review (**2**). As we also described there, although size exclusion chromatography (SEC) has generally been the routine method for detecting and quantifying protein aggregation (**3–5**), a major drawback of SEC is the loss of aggregates by their nonspecific binding to the column (**6**, **7**). Thus, column (matrix)-free techniques, such as analytical ultracentrifugation (AUC), field-flow fractionation (FFF), and dynamic light scattering (DLS), now find increasing application in aggregation analysis. It is also now fairly common for regulatory agencies to request aggregation analysis using one or more of these methods to cross-check the SEC analysis.

In this installment we review two of those technologies, AUC and DLS.

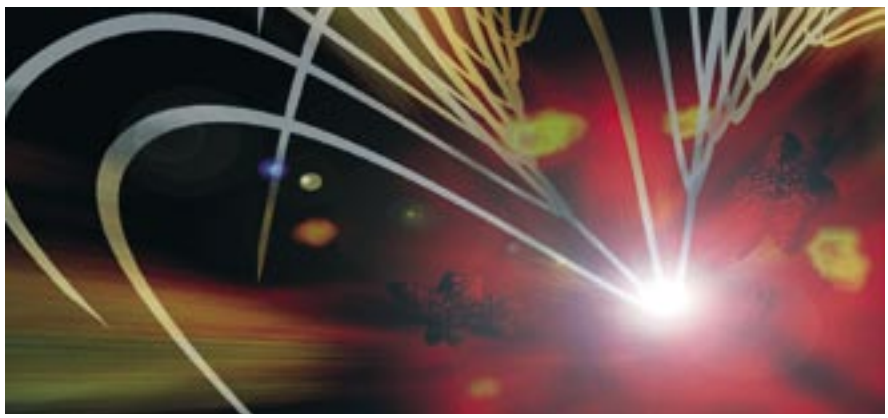
Product Focus: Proteins

Process Focus: Analysis and downstream processing

Who Should Read: QA/QC, process development, manufacturing, and analytical personnel

Keywords: Aggregation, analytical ultracentrifugation, sedimentation velocity, dynamic light scattering

Level: Intermediate



WWW.PHOTOS.COM

Because they are probably much less familiar than HPLC and electrophoresis, and because the raw data require significant mathematical analysis, we discuss these two and interpretation of their data in more depth here than in Part 1. And because biotechnology applications of FFF are relatively new and the technique is relatively unfamiliar, we shall devote Part 3 of this article to describing FFF principles, some of its hardware and operational variations, and example data. That final installment will conclude by summarizing the strengths and weaknesses of all five analytical methods we have covered.

Analytical Ultracentrifugation

An analytical ultracentrifuge is an instrument that monitors the concentration distribution of solutes within a centrifuge cell in real time while the rotor is spinning. The basic principles of analytical ultracentrifugation (AUC) can be illustrated by considering the sedimentation due to gravity of a thin solution of uniform sand particles within an optically-transparent cylinder (Figure 1). From a side view, an optical detector (your eye) can see that the distribution of sand particles is initially uniform at time zero. Sedimentation of the particles over time creates a boundary separating the sand-free upper layer from the sand-filled lower layer.

In AUC, a centrifugal force of up to about 250,000*g* is used to accelerate sedimentation of very small particles, such as proteins, while an optical detector scans the concentration of protein within an optically transparent cell (which generally uses sapphire windows) from the top to bottom — just as the eye scans the concentration of sand from the top to bottom in the above example. Because separation is based on molecular size, no gel structure or column matrix is needed to provide size fractionation; AUC is matrix-free. More specifically, solutes sediment and separate along the radial

direction according to their sedimentation coefficients. Simultaneously, the solutes will diffuse according to Fick's law depending on the diffusion coefficient of each component, causing the boundaries to spread. (Such diffusion is negligibly small for large particles such as sand, so the boundary is sharp in that example.)

Sedimentation Velocity: Two fundamental modes of measurement are *sedimentation velocity* (SV) and *sedimentation equilibrium*. Aggregation analysis of proteins almost exclusively uses sedimentation velocity because it causes physical separation of molecular species of different mass or shape. In sedimentation velocity, a rotor is spun at a high speed such that both monomeric and aggregated species sediment completely to the outside of the rotor (pellet) within a few hours. As protein molecules sediment, a moving boundary appears between two regions in the cell: one region where the solutes are already depleted and another where solutes are still present.

Sedimentation Equilibrium: For sedimentation equilibrium, on the other hand, the rotor speed is set relatively low, such that the solutes are gently pushed to the outside of the rotor, and no component forms a pellet. Under this condition, no true boundary forms, but the concentration of solute near the meniscus will gradually decrease while that at the bottom will increase due to the combined effects of sedimentation and diffusion. Eventually the system will reach equilibrium, and the concentration distribution within the cell becomes constant in time, forming an exponential or exponential-like concentration gradient. It can be readily shown that the slope of a graph of the logarithm of concentration with respect to r^2, where r denotes the radial distance from the center of revolution, will give the weight average molecular weight of the solutes, and that this equilibrium distribution is independent of the shape of the molecule.

Overall, sedimentation equilibrium is primarily a thermodynamic rather than a separation method. It is particularly useful for characterizing the strength and stoichiometry of reversible binding interactions (whether self-association or binding between different macromolecules). Of the two methods, sedimentation velocity is far more sensitive to solute heterogeneity. Because we are interested in aggregation, we focus here on sedimentation velocity, but sedimentation equilibrium also has important applications for protein pharmaceutical analysis (**8**).

SV Interpretation: In principle, the moving boundaries in a sedimentation velocity experiment contain a wealth of information, including heterogeneity or homogeneity of hydrodynamically distinct molecular species and the sedimentation coefficients and diffusion coefficients of the solutes (a function of both molecular weight and the shape of the solute molecule). If both the sedimentation coefficient and diffusion coefficient of a solute can be determined, it is possible to calculate its molecular weight based on the Svedberg equation (**9**, page 39).

In practice, however, it can be quite challenging to obtain information about heterogeneity (amount of material sedimenting at different rates) while simultaneously determining the diffusion coefficients of each component by sedimentation velocity. The difficulty lies in distinguishing broadening of the moving boundary due to heterogeneity (the presence of multiple components with similar sedimentation coefficients) from that due to diffusion. A number of data analysis methods recently developed to overcome this difficulty couple the power of modern desktop computers with the robust theory governing flow of material within a centrifuge cell. One such approach models the moving boundaries as a mixture of a fixed number of discrete components. The concentration, sedimentation coefficient, and diffusion coefficient of each component are determined by nonlinear least-squares fitting, as first implemented in the SVEDBERG analysis program (**10**) and now also implemented in several other programs (**11–14**).

Figure 1: Schematic showing sedimentation due to gravity of a thin solution of sand

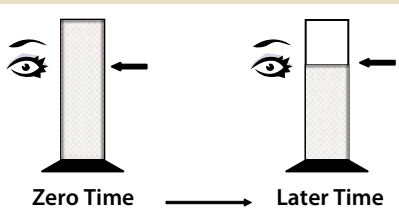
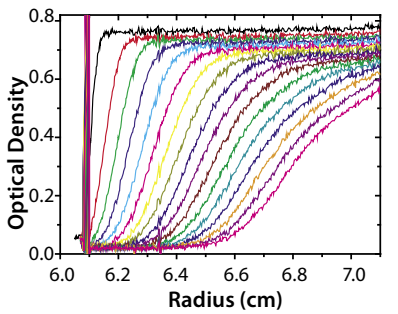


Figure 2: Sedimentation boundaries for a sample of EPO at 0.5 mg/mL in 20 mM citrate pH 7.0 after heating to 79 °C for 20 min; these data were recorded at 60,000 rpm and 20 °C using absorbance scans at 280 nm. Every third scan is shown.



For example, we used sedimentation velocity to examine the aggregation of erythropoietin (EPO) after it was heated to 79 °C for 20 min in 20 mM citrate, pH 7.0. Figure 2 shows the sedimentation pattern, in which the abscissa denotes the distance from the center of revolution in centimeters and the ordinate denotes the absorbance at 280 nanometers. Before such a run, it is important to make a scan at low speed, commonly at 3,000 rpm. The value of absorbance can then be compared with the total absorbance of the solute at the final run speed to confirm that no large aggregates have pelleted during acceleration. In this case, no large aggregates sedimented during acceleration to 60,000 rpm. At run speed the boundary moves from the meniscus (vertical spike at the left) toward the cell base (right). Simultaneously, the concentration gradient, as represented by the slope at the midpoint of the boundary between the baseline and the plateau region, decreases over time due to diffusion.

Although our sample actually contained a considerable fraction of aggregates, the aggregates did not produce distinct boundaries in

Table 1: Aggregate content determined by AUC and SEC

Sample	SV*	SEC (+ arginine)	SEC (– arginine)
MAb-a	5.7	2.6	2.5
MAb-a stressed	52.6	43.4	26.5
MAb-b	4.3	0.9	0.8
MAb-b stressed	45.8	42.6	38.5

*SV = Sedimentation velocity

Figure 2 because those boundaries are broadened by diffusion (although trained eyes may discern that the main boundary appears asymmetric). Nonetheless it is possible to quantify the extent of aggregation. Analysis of this sample as a discrete mixture of monomer and irreversible aggregates via the SVEDBERG program gave a best fit as 74 ± 2% monomer, 24 ± 2% dimer, 1.4 ± 0.5% tetramer, and 0.3 ± 0.2% octamer. This result qualitatively agrees with the native gel analysis of the same sample (**2**, referring to Figure 3B, lane 3, in Part 1 of this series, www.bioprocessintl.com). However, such resolution and quantitation are not possible by native gel analysis. After similar heat treatment in 20 mM Tris, the monomer content was much higher at 95.5%, indicating formation of much less aggregate, consistent with the apparent absence of aggregate bands in Figure 3B, lane 4 (Part 1).

One problem with this type of analysis is that it requires an assumption about the number of components present. Further, for a minor component present at low levels, the data generally do not contain sufficient information to uniquely determine its diffusion coefficient and therefore its molecular weight. Hence, in our EPO example the molecular weights of the "dimer," "tetramer," and "octamer" components were constrained to be two, four, and eight times the value for the monomer (and it is important to confirm that the returned sedimentation coefficients are actually consistent with such constraints, as they were in this case).

Therefore some alternative approaches to data analysis approaches are designed to enhance the resolution for minor components by modeling the influence of diffusion more generically without requiring explicit assumptions about the number of components present. In essence, such methods sacrifice the detailed information about the diffusion of each component (and hence the molecular weights) to obtain better information about the heterogeneity of sedimentation coefficients and better quantitation of the fraction of each species.

One such approach is an improved form of the van Holde and Weischet method (**15**). The one we illustrate here is the *c(s)* method developed by Peter Schuck at NIH and implemented in his program SEDFIT (**16**). This approach derives a high-resolution distribution of sedimentation coefficients

Figure 3: Sedimentation velocity analysis of ~0.5 mg/mL monoclonal antibody MAb-b at 50,000 rpm and 20 °C in 50 mM Tris-HCl, 0.1 M NaCl, pH 7.5; (a) raw sedimentation boundaries, showing every third scan; (b) sedimentation coefficient distribution derived from the data in (a).

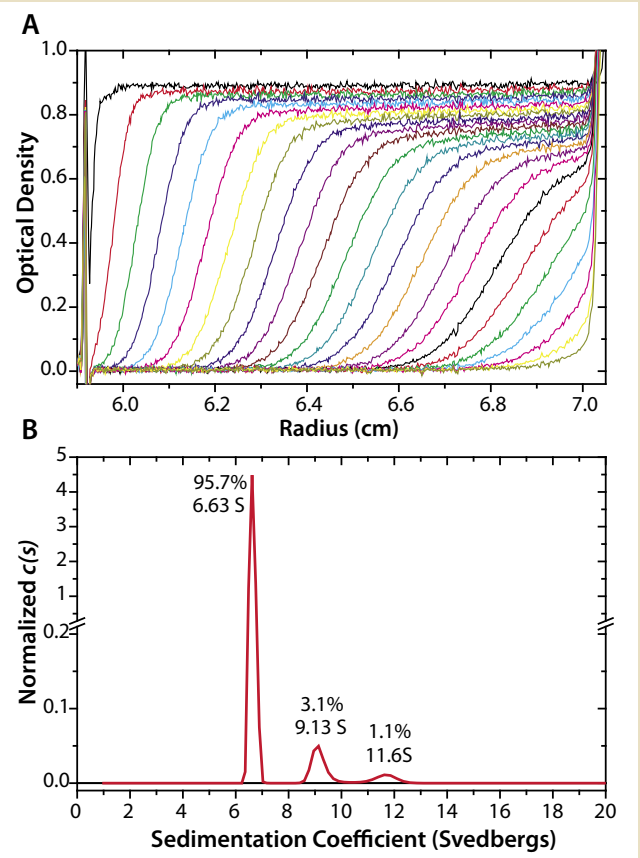
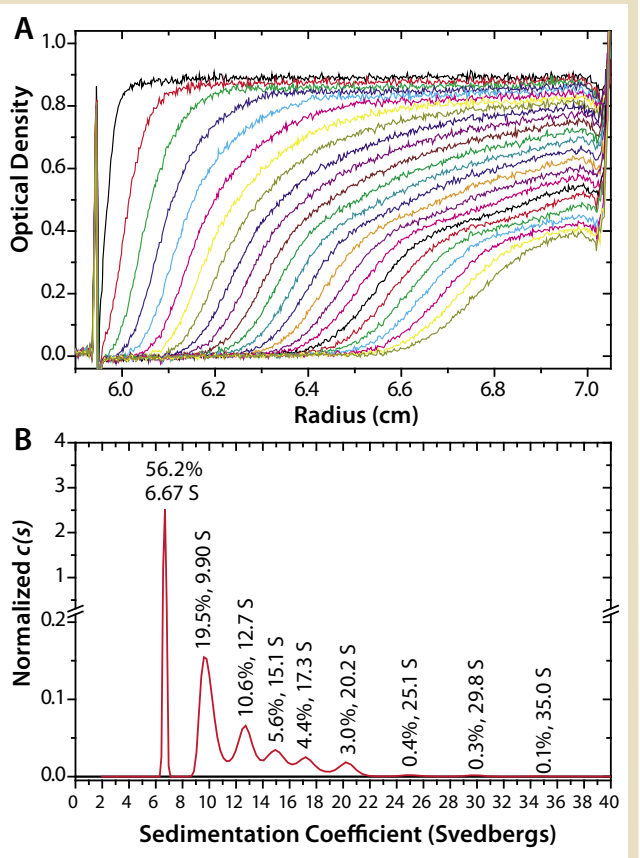


Figure 4: Sedimentation velocity analysis of monoclonal antibody MAb-b after heating; (a) raw sedimentation boundaries showing every third scan; (b) sedimentation coefficient distribution derived from the data in (a).



(concentration as a function of sedimentation coefficient) where each species produces a peak at its corresponding sedimentation coefficient (like a chromatogram). As in a chromatogram, the area under each peak gives the relative concentration of that species. One key strength of this method is that it does not require an analyst to explicitly assume how many such peaks are present.

Figure 3A shows the moving boundaries for monoclonal antibody MAb-b (one of two antibodies for which native gel and SEC data were presented in Part 1). Figure 3B shows the distribution of sedimentation coefficients as calculated from Figure 3A by the *c(s)* method in SEDFIT. The major molecular species, antibody monomer, has a sedimentation coefficient of 6.63 S (Svedbergs) and accounts for 95.7% of the total sedimenting solutes. The species at 9.13 S, which accounts for only 3.1% of the total, is most likely a dimer, and the peak at 11.6 S (1.1%) is probably trimer. Note that because sedimentation coefficients depend on molecular shape as well as mass, it is generally not possible to uniquely assign a stoichiometry to minor aggregate peaks such as these based solely on their observed sedimentation coefficients (which is one drawback to this approach).

As noted above, in principle if we could accurately measure the width of the boundary for each aggregate species, then we could measure its diffusion coefficient as well as its sedimentation coefficient and hence derive its mass. However, in practice for species present at levels of only a few percent, the signal–noise ratio is often not high enough, and the degree of separation is not good enough, to permit accurately measuring the diffusion coefficients.

Figure 4A shows sedimentation boundaries after the antibody is subjected to heat stress. Even to the untrained eye this sedimentation pattern is obviously quite different from that of Figure 3A, and several boundaries for individual aggregate species are at least partially resolved. The sedimentation coefficient distribution derived from these data (Figure 4B) indicates that the fraction of monomer (at 6.77 S in this sample) has decreased to only 56.2%, whereas at least eight peaks for aggregates are now detected. Note that the higher molecular weight of antibodies relative to EPO means that boundaries for antibodies are less broadened by diffusion, and the intrinsic resolution of the separation is better. This is an important reason why it is possible to mathematically resolve so many components and detect peaks present at levels of only a few tenths of a percent. The most prominent aggregate peaks at 9.90, 12.7, 15.1, 17.3, and 20.2 S fall very close to the positions expected (relative to monomer) for dimer, trimer, tetramer, pentamer, and hexamer based on model hydrodynamic calculations for compact oligomers of spheres (**17**). The

Figure 5: Sedimentation velocity analysis of ~0.5 mg/mL monoclonal antibody MAb-a (same conditions as in Figure 3); (a) raw sedimentation boundaries showing every third scan; (b) sedimentation coefficient distribution derived from the data in (a).

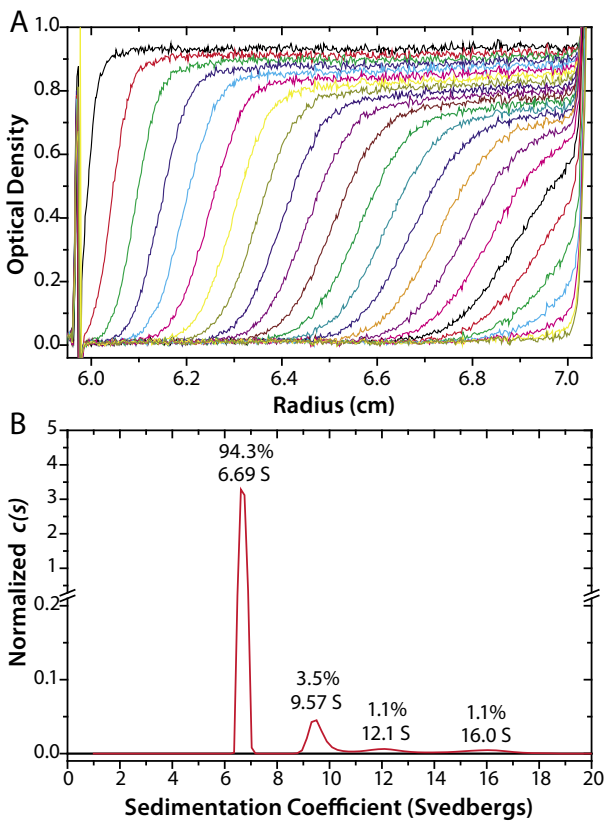
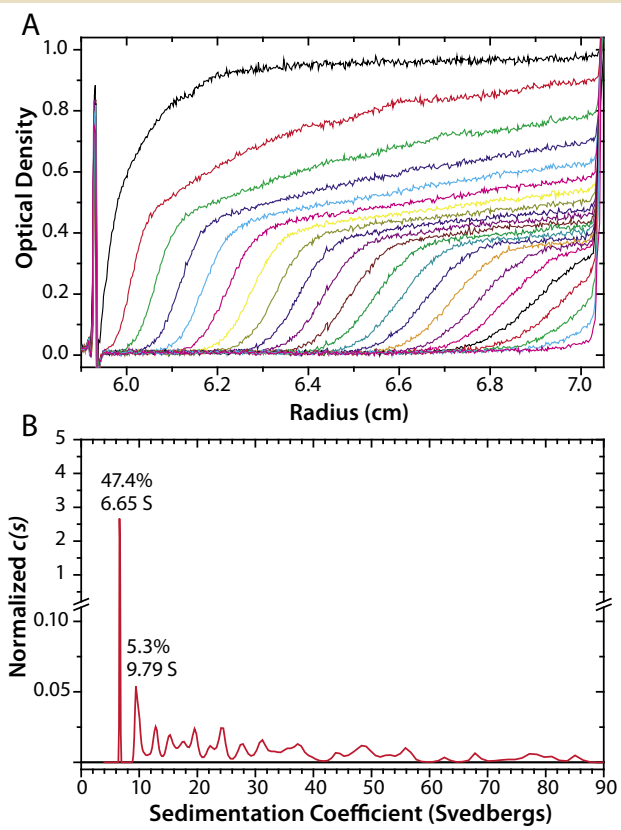


Figure 6: Sedimentation velocity analysis of monoclonal antibody MAb-a after heating; (a) raw sedimentation boundaries showing every third scan; (b) sedimentation coefficient distribution derived from the data in (a).



largest species detected (at 35 S) corresponds roughly to a dodecamer.

Figures 5 and 6 show analogous experiments on control and heat-stressed antibody MAb-a. This purified antibody starts off with a somewhat higher aggregate content of 5.7% (Figure 5). After heating, the sample becomes quite heterogeneous (Figure 6). The total aggregate content rises to 52.6%, distributed among very many poorly resolved peaks, including some very large species sedimenting at up to 85 S (roughly 45-mers). Interestingly the distributions of aggregates formed by the two antibodies are quite different from one another even though heat stress produces about 50% total aggregate for both. The aggregate content of 52.6% for heat-stressed MAb-a measured by sedimentation velocity is close to the estimate by SEC in the presence of 0.2 M arginine (Table 1).

The results for both antibodies illustrate another strength of sedimentation velocity: its ability to cover a fairly wide range of molecular weights in one analysis. Even when SEC does give full recovery of all aggregate species, typically all aggregates larger than trimer or tetramer coelute in a single peak at the exclusion limit (void volume) of the column. By contrast, such larger aggregates can be resolved and quantified by SV.

Another significant advantage of SV is that the analysis often can be conducted directly in the formulation buffer, whereas for SEC or FFF a different elution buffer is often required to achieve good sample recovery (as discussed in Part 1 for SEC). Any time solvent conditions change, the distribution of noncovalent aggregates can potentially change (**7**). Detergents and high levels of sucrose or other sugars can, however, produce significant interference in SV analysis.

Dynamic Light Scattering

Dynamic light scattering (DLS) is also known as *quasi-elastic light scattering* (QELS) or *photon correlation spectroscopy* (PCS). In this method, what is studied are fluctuations in the intensity of light scattered from a protein solution over time scales from ~0.1 µs to ~0.1 s are due to the Brownian motion of the molecules. Larger molecules move (diffuse) more slowly than smaller ones, and thus the time scale of the scattering fluctuations provides a measure of molecular size. The size parameter measured by DLS is the hydrodynamic size; by definition the hydrodynamic radius of a protein molecule is the radius of a spherical particle with the same diffusion coefficient as that protein molecule.

When used in batch mode, DLS is a matrix-free technique and hence is not subject to erroneous conclusions regarding aggregation due to nonspecific adsorption of the aggregates. DLS also often requires no dilution, thus preventing potential dissociation of reversible aggregates.

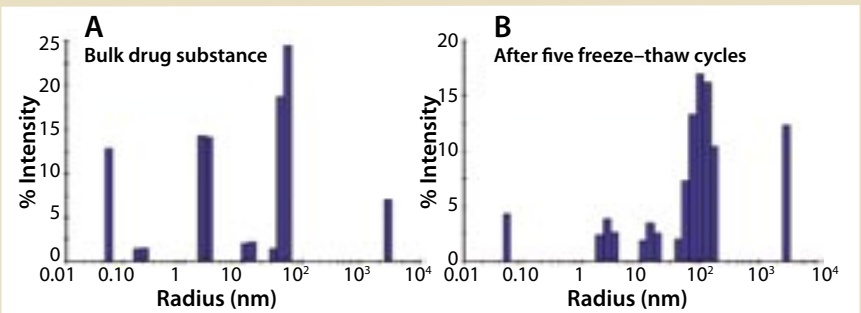
The other primary light scattering technique is *static* or *classical* light scattering. Here, the time-averaged scattering intensity at one or more scattering angles is measured, and that information can be used to directly measure the true weight-average molecular weight of a sample. A great strength of all light scattering techniques is that their sensitivity increases with increasing molecular weight (and under many circumstances it is directly proportional to molecular weight). That means that light scattering methods can be very sensitive for detecting small amounts of very large aggregates. Somewhat ironically though, this great sensitivity to large species is also a primary weakness of scattering techniques because the scattering from traces of dust or other large particulates can easily dominate over that caused by a protein.

Therefore, with light scattering methods it is often necessary to have some means of knowing that the signals really came from protein and not dust. For static scattering, that is usually accomplished by using the detector in flow mode, following separation by SEC or FFF (which presumably separates dust from the protein). In this application we directly measure the molecular weight for each peak as it elutes, and that value is independent of molecular shape or the actual elution position relative to molecular standards (**18**). This approach can be very useful for identifying the true molecular weight (stoichiometry) of protein aggregates (**19**, **20**), and it is one we recommend. However, it suffers from the same drawbacks as standard SEC and FFF methods: loss of aggregates due to nonspecific adsorption, and alterations of aggregate distribution caused by changes in concentration and/or use of an elution buffer that differs from the formulation buffer.

DLS also can be used in flow mode as an on-line detector following SEC, so do not discuss that application here (because DLS, when used on-line, is no longer a measurement of protein aggregates in the actual formulation).

For DLS, no physical separation is needed to distinguish the scattering due to dust or other large particles because the Brownian motion of such particles is much slower than that of a protein monomer. That is, in DLS we can achieve a mathematical separation of large from small components because they produce scattering fluctuations on different time scales. Relatively simple data analysis

Figure 7: Dynamic light scattering analysis of a ~20 kDa protein therapeutic; (a) distribution of scattering intensity for bulk drug substance at ~0.5 mg/mL; (b) distribution obtained after five cycles of freeze-thaw stress



algorithms for DLS give a single average hydrodynamic size for the sample and a measure of the homogeneity in size. It is often more informative, though, to use more complex algorithms that process the measured scattering fluctuations to yield a size distribution histogram.

For example Figure 7A shows the size distribution obtained for the bulk drug substance of a ~20 kDa recombinant protein. This distribution histogram shows the percentage of the scattering intensity that arises from species at different radii. Note that the individual bins in the histogram are fairly broad, with neighboring bins differing in radius by about 33%.

Given that the protein is quite pure and contains rather low levels of aggregates when analyzed by other techniques, it is at first rather shocking to see that DLS has resolved the sample into six peaks ranging in radius from 0.05 nm to 2.9 µm. What does this mean? In fact, the two peaks at smallest radii (0.05 and 0.19 nm) are due to scattering by water and salts, not to protein. The main protein peak at mean radius 2.34 nm, which represents 28% of the total scattering intensity. Based on globular protein standards, that radius implies a molecular weight of 25 kDa — significantly above the true monomer mass. That discrepancy is not due to aggregation, but rather, it arises because the monomer does not have a compact globular shape. This illustrates that DLS is not a good method for measuring absolute molecular weights because the hydrodynamic radius depends on molecular shape as well as molecular mass.

If the peak at 2.34 nm represents monomer, then do the peaks at mean radii of 12.7 nm, 51.1 nm, and 2.90 µm represent protein aggregates? Perhaps, but remember that this technique does not discriminate the chemical nature of particles producing the scattering, so it is possible that one or more of those peaks is actually due to trace particulate contaminants (nonprotein species).

If these three peaks do represent product aggregates, their radii imply molecular weights of approximately 1.3 MDa, 33 MDa, and 430 GDa. Thus they are very large aggregates indeed! Precisely because they are so large, they scatter light very strongly, and even though together they represent two-thirds of the total protein scattering intensity, they in fact are present in very low amounts on the basis of percentage by weight. The estimated weight fraction for the 12.7-nm peak (which is 5.0% of the total protein scattering intensity) is only 0.095%, whereas the 51.1-nm peak (53% of the intensity) represents only 0.017% by weight.

DLS is unfortunately unable to accurately quantify weight fractions. Repeated measurements of the same sample will often give values differing by a factor of two or more. Further, the algorithms for such calculations are considered proprietary by instrument manufacturers, and different instruments therefore give different values. For very large species such as the peak at 2.90 µm (8.4% of intensity), which are larger than the wavelength of the light, the scattering intensity will vary strongly with scattering angle and also on the detailed shape of the particle (due to internal reflections). Thus in our view it is not possible to reliably estimate the weight fraction for such large particles within even an order of magnitude, although they probably represent less than 0.01% by weight.

This sample therefore illustrates a number of the strengths and weaknesses of batch-mode DLS. DLS can cover an enormous range of size in one measurement (a radius range from 0.1 nm to 1 µm corresponds to a factor of ~10^{12} in molecular weight). It is also extremely sensitive to small amounts of large aggregates and can often detect species larger than 100 nm in radius when they are present only in the parts-per-million range by weight. On the other hand, quantitation of weight fractions is rather poor (and sometimes can't be done at all). Further, it can be hard to distinguish which species are product aggregates rather than particulate contaminants.

DLS is often more useful for comparing one sample to another (different lots, different formulations, and so on) rather than as an absolute measurement. With the recent advent of DLS instruments accepting samples in 96- or 384-well plates, it is now possible to compare fairly large numbers of samples. As an example of using DLS for a forced-degradation study, Figure 7B shows data for a drug substance after five cycles of freeze-thaw stress. The total scattering intensity of the stressed sample is more than four-fold higher than that of the starting material. As shown by the size distribution, the difference arises from an increase in scattering for all three of the putative aggregate peaks that we detected in the bulk. In this sample we see aggregate peaks at mean radii of 15.0 nm (8.3% of intensity, 0.55% by weight), 110 nm (69.5% of intensity, 0.028% by weight), and ≥2.90 µm (12.9% of intensity). Because the weight fraction of these species increases with stress, this proves they are assembling from smaller species, strongly suggesting they are indeed product aggregates rather than contaminants.

Figure 8: Hydrodynamic radius distribution for a recombinant protein at ~50 mg/mL indicates presence of large, reversible aggregates (not seen at low concentrations)

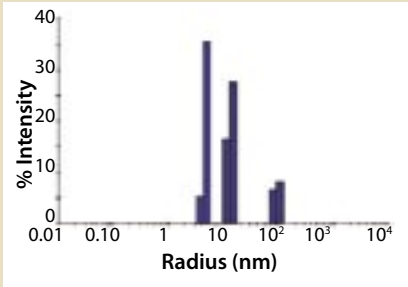
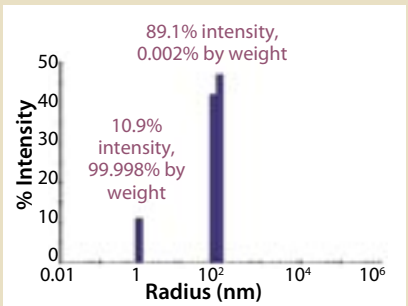


Figure 9: Hydrodynamic radius distribution of a fresh solution of a small peptide that eventually (weeks to months) forms visible particles; the particles at a ~80-nm radius are a precursor to the visible particles and diagnostic of "good" and "bad" manufacturing lots.



Why was the salt peak at 0.19 nm in the control sample not detected in the stressed sample? This is an example of a phenomenon we call "blinded by the light" — the salt peak is simply lost in the glare from the aggregates. Typically the analysis algorithms will suppress any peak that represents less than 1–2% of the total scattering intensity because such a peak might be false (due to noise in the data). In the control sample, the salt peak is only 2.9% of the total intensity. Thus, when the total intensity rises more than fourfold after stress, this peak falls back below the detection threshold. We aren't really interested in water or salt peaks, so in this case being "blinded by the light" doesn't really matter. However, for highly aggregated samples the main protein peak is often lost in a similar fashion, and thus DLS may fail to detect 99.9+% of the protein in a sample! Obviously if the main peak is not detected, you cannot make any estimate of weight fractions.

A final point to note from this example is that the radius for the main peak shifts upward from 2.34 to 2.75 nm, which is significant compared with the usual precision of 2–3%. This shift is almost certainly due to the creation of additional small oligomers (dimer, trimer, tetramer) by freeze–thaw stress. The resolution of DLS is very low (remember, the separation is purely mathematical); typically two species cannot be resolved as separate peaks unless their radii differ by a factor of ~2 or more (which corresponds to a factor of ~8 or more in molecular weight). Thus the "main peak" for any protein sample is likely to include contributions from any small oligomers that may be present, and consequently DLS is generally quite poor at detecting such species. In this case the upward shift in radius is due to the increased amount of unresolved small oligomers within that main peak, but many different combinations of oligomers could produce the same shift in radius. So there is no way to quantify this increase in aggregate content.

Measurements at High Concentrations: Today many protein pharmaceuticals, particularly antibodies and Fc-fusion proteins, are being formulated at concentrations above 10 mg/mL (and often considerably higher). It would be highly desirable to have a simple tool that tells us whether a protein remains monomeric or reversibly associates to form oligomers at such high concentrations (**21**). Although DLS can be used for concentrated protein solutions, at such high concentrations the interparticle and particle–solvent interactions can cause substantial distortions in apparent diffusion coefficients and hence the measured hydrodynamic size (**22**). In general, particle–particle interactions ("molecular crowding" effects) will cause the apparent size to grow at high concentrations, even if the protein remains 100% monomeric. All proteins are electrically charged, and at high concentrations their diffusion can couple to the rapid diffusion of small counter-ions, producing a decrease in the apparent size of the proteins. Such electrostatic effects can be very strong when the ionic strength of a solution is low and the protein electric charge is high, reducing the apparent size by factors of two, three, or more (and thus suggesting that the molecular weight has changed by an order of magnitude or more!).

Because concentration-dependent size distortions can be either positive or negative, it can be quite difficult to correctly interpret the absolute measured sizes at high concentrations and whether any change is due to reversible self-association. If however we see a qualitative change in size distribution measured by DLS, such as the appearance of new peaks as the concentration increases, that is a clear indication that the sample is no longer homogeneous at high concentration due to formation of aggregates. Figure 8 illustrates this approach. At low concentrations this protein shows only a single peak. However, in data taken at ~50 mg/mL, we see additional peaks due to formation of large (but reversible) aggregates. The measured radii of all of these components are significantly distorted from their true values at this high concentration, but the change in homogeneity is unambiguous.

Using DLS to Detect Precursors to Visible Particulates: Aggregation that forms visible particulates ("snow" or "floaters"), precipitates, or so-called "subvisible" particles with diameters from ~10 to 50 μm is a fairly common problem for protein pharmaceuticals. This can be especially troubling for several reasons:

- Just a single visible particle, or very low levels of subvisible particles, make a vial of product unusable.
- The particles may not appear until months after manufacturing.
- Often the levels of particles vary widely from one manufacturing lot to another.
- Particles typically represent a very low weight fraction of the total protein (often <0.1%).

These factors combine to make it very difficult to track down the cause of the problem and to be certain that corrective action will be effective.

In many cases this type of aggregation is an example of a nucleation-controlled or "seeded" reaction (**23**). This means that formation of smaller aggregates is thermodynamically unfavorable, but once an aggregate does reach a certain size (the so-called critical nucleus), its growth through association of additional monomers becomes very likely. The kinetics of such nucleation-controlled assembly processes exhibit a lag phase during which no large aggregates are formed. However, once the critical nuclei are formed (which may take days or months), assembly then accelerates rapidly.

Therefore what is needed is an assay that can detect those smaller precursors to the large particles and predict which product lots will eventually form large particulates. Because the concentration of critical nuclei usually remains at all times quite low (perhaps 0.01% or less), such an assay needs great sensitivity. As we have already seen, however, DLS is very good at detecting trace levels of very large aggregates, and that is exactly why we have found it to be an effective tool for tackling and solving this type of aggregation problem (**7**).

Figure 9 shows some DLS data for a small peptide that tends to form threadlike visible particles. The particles at a radius of ~80 nm represent a precursor to the visible particles (a critical nucleus). Even though these ~80-nm particles represent only ~0.002% of the sample by weight, that level is easily detected and is also high enough to trigger formation of visible particles within a day or two. Peptide lots initially containing much less of that ~80-nm aggregate do not form visible particles (or at least don't form them quickly enough to affect use of the product).

For a number of different proteins and peptides we have similarly identified particles with radii from roughly 20 to 400 nm that appear to be critical intermediates in formation of visible or subvisible particles and that correlate with "bad" versus "good" manufacturing lots. With such an assay in hand we can then try to find changes that will prevent formation of particles larger than 10 µm. A change in formulation might do this, but we have also had great success in examining samples from different stages in manufacturing, identifying a particular manufacturing step where the damage occurs that eventually leads to particulate formation, and altering that manufacturing step to permanently solve the aggregation problem.

A Final Note

Thus far we have been talking about cases where the critical nuclei consist of product aggregates (homogeneous nucleation). However, it is important to be aware that contaminant particles can also serve as nuclei or "seeds" for product aggregation. In one case silica particles shed by vials appear to have caused a particulate problem (**24**). Anecdotal reports implicate particles of silicones (used in manufacturing plastic tubing or as lubricants in prefilled syringes) and particles of vacuum pump oil from lyophilizers as heterogeneous nuclei for aggregation of other protein products. Because DLS can detect particles of any chemical type, it is also useful for tackling these types of problems.

References

1 Wang W. Protein Aggregation and Its Inhibition in Biopharmaceutics. *Int. J. Pharm.* 289(1-2) 2005: 1–30.

2 Arakawa T, et al. Aggregation Analysis of Therapeutic Proteins, Part 1: General Aspects and Techniques for Assessment. *BioProcess International* 4(10) 2006: 42–49.

3 Wang W. Instability, Stabilization, and Formulation of Liquid Protein Pharmaceuticals. *Int. J. Pharm.* 185(2) 1999: 129–188.

4 Jones AJS. Analysis of Polypeptides and Proteins. *Adv. Drug Deliv. Rev.* 10(1) 1993: 29–90.

5 Ahrer K, et al. Analysis of Aggregates of Human Immunoglobulin G Using Size-Exclusion Chromatography, Static and Dynamic Light Scattering. *J. Chromatogr. A* 1009(1–2) 2003: 89–96.

6 Stulik K, Pacakova V, Ticha M. Some Potentialities and Drawbacks of Contemporary Size-Exclusion Chromatography. *J. Biochem. Biophys. Methods* 56(1–3) 2003: 1–13.

7 Philo JS. Is Any Measurement Method Optimal for All Aggregate Sizes and Types? *AAPS J.* 8(3) 2006: E564–E571.

8 Philo JS. Analytical Ultracentrifugation. *Methods for Structural Analysis of Protein Pharmaceuticals.* AAPS Press: Arlington, VA, 2005; 379–412.

9 Svedberg T, Pedersen KO. *The Ultracentrifuge.* Clarendon Press: Oxford, 1940.

10 Philo JS. Measuring Sedimentation, Diffusion, and Molecular Weights of Small Molecules by Direct Fitting of Sedimentation Velocity Concentration Profiles. *Modern Analytical Ultracentrifugation.* Schuster TM, Laue TM, Eds. Birkhauser: Boston, 1994; 156–170.

11 Behlke J, Ristau O. Molecular Mass Determination by Sedimentation Velocity Experiments and Direct Fitting of the Concentration Profiles. *Biophys. J.* 72(1) 1997: 428–434.

12 Demeler B, Saber H. Determination of Molecular Parameters by Fitting Sedimentation Data to Finite-Element Solutions of the Lamm Equation. *Biophys. J.* 74(1) 1998: 444–454.

13 Schuck P, MacPhee CE, Howlett GJ. Determination of Sedimentation Coefficients for Small Peptides. *Biophys. J.* 74(1) 1998: 466–474.

14 Stafford WF, Sherwood PJ. Analysis of Heterologous Interacting Systems by Sedimentation Velocity: Curve Fitting Algorithms for Estimation of Sedimentation Coefficients, Equilibrium and Kinetic Constants. *Biophys. Chem.* 108(1–3) 2004: 231–243.

15 Demeler B, van Holde KE. Sedimentation Velocity Analysis of Highly Heterogeneous Systems. *Anal. Biochem.* 335(2) 2004: 279–288.

16 Schuck P. Size-Distribution Analysis of Macromolecules by Sedimentation Velocity Ultracentrifugation and Lamm Equation Modeling. *Biophys. J.* 78, 2000: 1606–1619.

17 Garcia de la Torre JG, Bloomfield VA. Hydrodynamic Properties of Complex, Rigid, Biological Macromolecules: Theory and Applications. *Q. Rev. Biophys.* 14(1) 1981: 81–139.

18 Arakawa T, Philo JS. Size-Exclusion Chromatography with On-Line Light-Scattering, Absorbance, and Refractive Index Detectors for Studying Proteins and Their Interactions. *Anal. Biochem.* 240(2) 1996: 155–166.

19 Endo Y, et al. Heat-Induced Aggregation of Recombinant Erythropoietin in the Intact and Deglycosylated States As Monitored by Gel Permeation Chromatography Combined With a Low-Angle Laser Light Scattering Technique. *J. Biochem.* (Tokyo) 112, 1992: 700–706.

20 Andya JD, Hsu CC, Shire SJ. Mechanisms of Aggregate Formation and Carbohydrate Excipient Stabilization of Lyophilized Humanized Monoclonal Antibody Formulations. *AAPS. PharmSci.* 5(2) 2003: E10.

21 Shire SJ, Shahrokh Z, Liu J. Challenges in the Development of High Protein Concentration Formulations. *J. Pharm. Sci.* 93(6) 2004: 1390–1402.

22 Andries C, Clauwaert J. Photon Correlation Spectroscopy and Light Scattering of Eye Lens Proteins at High Concentrations. *Biophys. J.* 47(5) 1985: 591–605.

23 Chi EY, et al. Physical Stability of Proteins in Aqueous Solution: Mechanism and Driving Forces in Nonnative Protein Aggregation. *Pharm. Res.* 20(9) 2003: 1325–1336.

24 Chi EY, et al. Heterogeneous Nucleation-Controlled Particulate Formation of Recombinant Human Platelet-Activating Factor Acetylhydrolase in Pharmaceutical Formulation. *J. Pharm. Sci.* 94(2) 2005: 256–274.

Corresponding author ***Tsutomu Arakawa*** *is president of Alliance Protein Laboratories, Thousand Oaks, CA 91360, tarakawa2@aol.com;* ***John S. Philo*** *is vice president and director of biophysical chemistry at Alliance Protein Laboratories;* ***Daisuke Ejima*** *is a chief chemist at AminoScience Laboratories, Ajinomoto Co., Inc., Kawasaki, Kanagawa 210-8681, Japan;* ***Kouhei Tsumoto*** *is an associate professor at the Department of Medical Genome Sciences, The University of Toyko, Kashiwa 277-8562, Japan; and* ***Fumio Arisaka*** *is an associate professor in the Department of Molecular Bioprocessing, Tokyo Institute of Technology, Yokohama 226-8501, Japan.*