New biological entities (NBEs, therapeutic proteins such as interferons or antibodies) are much more complex than new chemical entities (NCEs), the classic “chemical” active ingredients. First, they are much larger. The average molecular weight of antibodies is ~150,000 g/mol. Second, most NBEs contain three-dimensional structural elements — with the protein secondary and tertiary structure being the most prominent, but quaternary structures are also known for some. The 3D structures are essential for correct bioactivity (1), but they are not rigid, “frozen” structures. Most proteins show a certain structural flexibility, which enables their correct molecular interactions (e.g., of an antibody with its antigen or receptor). The extreme conformational states of that flexibility go from the native, correctly folded conformation to a completely denatured state, in which the protein adopts a more or less random conformation. Various interim conformational states can also be adopted.

The large size and complex structure of proteins make them prone to instabilities caused by chemical and biological degradations. In addition, a colloidal (physical) instability is often encountered, which can manifest as protein particle formation, aggregation, association, precipitation, and/or adsorption to materials used in, for example, medical devices, primary packaging, and tubing during filling. Colloidal instability and protein aggregate formation present a major problem for long-term storage stability, shipping, and handling (2). Aggregates are defined as protein assemblies of higher molecular order formed by unfolded (denatured) and/or partially unfolded monomers. By contrast, protein associates are built up of native monomers (Figure 1) and can be redissolved to yield those native monomers again (1). The term protein particles refers to all protein-containing assemblies, independent of the nature and structure of the protein. Figure 1 represents schematically the formation of protein particles. Protein associates are formed by physical association of native protein monomers, whereas aggregates are made of irreversibly or partially denatured monomers. Protein association — the first step in the formation of protein associates — can be studied using the osmotic second virial coefficient (3, 4). It depends strongly on the solution conditions of a formulation.

The size of protein particles varies from dimers to extremely large...
multimer units. As described below, they can be induced by the presence of nonproteinaceous (extrinsic) particles that act as nucleation sites. The reasons protein particles form in solution are manifold. Induction of physical or thermal stresses, for example, may change the secondary/tertiary structures and lead to denaturation or partial denaturation, followed by formation of particles. Excipients and other ingredients also strongly affect the colloidal stability of a liquid protein formulation — especially so-called “highly concentrated” liquid formulations (HCLFs, with protein concentration ~40–200 mg/mL) requested for subcutaneous application of antibodies, for example. So colloidal stability is a critical issue to consider.

Induction of aggregates may also be due to the presence of extrinsic particles. Chi et al. showed aggregate formation with recombinant human platelet-activating factor acetylhydrolase (rhPAF-AH) in the presence of silica particles (5). Others presented the appearance of protein aggregates induced by silicone oil (6, 7). Recently, it was shown that stainless steel particles generated during a filling process using a piston pump can induce IgG particle formation (8). In recent years, the presence of protein particles has been intensively discussed because of their immunogenic potential (9).

A number of techniques are used to characterize protein particles: high-performance size-exclusion chromatography (HP-SEC), light-obscuration microscopy, nephelometry, analytical ultracentrifugation, and field-flow fractionation (FFF). Each has strength and limitations. Because of technical impacts on the samples, correlation of results obtained with different techniques is complicated and in some cases improper (10, 11). Additionally, protein particles can cover an extremely broad size range, from a few nanometers up to a few micrometers the latter formed by millions of antibody monomer units. With this particle size range up to six orders of magnitude, no available technique can correctly analyze samples containing such a multitude of sizes. Techniques are especially scarce for subvisual particles. Among the techniques available for particle characterization, photon correlation spectroscopy (PCS) offers the advantage of detecting particles in a broad dynamic range from a few nanometers to a few micrometers (11–13).

**Materials and Methods**

We used proteins (lysozyme, bovine serum albumin, and concanavaline) purchased from Sigma Aldrich in Germany (www.sigmaaldrich.com/germany.html) and humanized monoclonal IgG1 antibodies produced by mammalian cell culture technology. Protein purity was determined using SEC and sodium-dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE, 14). The monomer content of each protein sample was >99% (based on HP-SEC). IgG concentration in sample solutions was
determined by ultraviolet (UV) measurement at 279 nm (14).

Latex standard samples came from Thermo Fisher Scientific Inc. in Waltham, MA (www.thermo.com). Photon correlation spectroscopy experiments were performed on a Zetasizer Nano ZS analyzer from Malvern Instruments in the United Kingdom (www.malvern.com) using a QS 3-mm quartz cuvette from Hellma GmbH & Co. KG in Müllheim, Germany (www.hellma-worldwide.com) with a filling volume of 75 µL. All measurements were performed at 20 °C if not otherwise specified.

**Basic Principles**

PCS is also known as dynamic light scattering (DLS) or quasielastic light scattering (QELS) (15–17). It is mainly used to measure the hydrodynamic size (Figure 2), polydispersity, and size distribution of dissolved or suspended particles (18, 19). PCS measures the light (in modern instruments laser light) scattered from particles dissolved or dispersed in solution.

Particles in solution undergo Brownian motion, which depends on physical properties such as temperature, viscosity, and so on. This causes fluctuations of scattered light intensity to be observed. Particle motion changes the phase of scattered light, so constructive and destructive interferences among light scattered from different particles are observed (20). Phase change depends on the motion of particles in relation to the wavelength (λ) of scattered light (21). For a particle moving at distances comparable to λ, a significant phase change is observed.

That fluctuation is generally studied in the time range of ~10^-7–10^-1 second. Because the size, shape, and mass of the scattering particles determine their motion and velocity (larger ones moving slower), a defined correlation function is obtained. So PCS data represents the intensity of the scattered light, measured by a photocurrent as a function of time (22). The resulting time correlation curve (Figure 3) contains information regarding the diffusion properties of particles within an analyzed sample.

Equation 1 describes the correlation function, G(τ), τ = delay time.

\[ G(\tau) = \int I(t) \cdot I(t + \tau) \, dt \]  

**Equation 1**

PCS therefore measures and correlates single scattered photons (intensity, I). The correlation procedure is a statistical method for measuring the degree of nonrandomness in an apparently random data set (23). Autocorrelation is the convolution of the intensity signal as a function of time with itself (Equation 1).

Experimentally, the intensity fluctuation of scattered light is measured quantitatively by recording the arrival times (at t and τ in Equation 1) of individual scattered photons and computing the autocorrelation function of those times over ~1 µs to 100 ms as described above (11). The rate of the intensity fluctuation is directly linked to the diffusion coefficient (D) of the particles. That in turn is a direct measurement of the hydrodynamic diameter D_H according to Equation 2 (the Stokes-Einstein relation), in which k_B represents the Boltzmann constant, T is the absolute temperature, η is the solution viscosity, and D_H is the hydrodynamic diameter. The term 3πD_Hη is the so-called friction coefficient (24).

\[ D_H = \frac{k_B T}{3\pi\eta D_H} \]  

**Equation 2**

By definition, PCS measures a particle diameter/radius of a hypothetical hard sphere that diffuses with the same speed as the particle under investigation. In reality, most particles (especially macromolecules) are not perfect spheres; they usually are nonspherical, more or less hydrated, and they might contain bound ions. So the size determined by PCS — the hydrodynamic diameter/radius (also known as Stokes’s radius) — is an apparent size containing contributions from dynamic hydrated and solvated particles (25). The size in PCS is reduced to “one” number: the diameter/radius of a spherical particle — although, for example, the elliptical size and shape of lysozyme is better represented by two size numbers in two dimensions derived from its crystallographic structure (Figure 2). Crystallographic data represent the size of the molecule as a crystal, however, with more or less “frozen” dynamics, and that can be different from the protein’s size and shape in solution (26, 27).
**Measurement and Data Evaluation**

In PCS, particle size distribution is derived by applying a mathematical procedure called **deconvolution** upon the experimental obtained intensity autocorrelation function (28). In most cases, deconvolution is performed using a nonnegatively least square (NNLS) algorithm. The main complexity is that deconvolution of the intensity autocorrelation curve to an intensity distribution is an ill-defined problem (29).

Thus, to determine the distributions of diffusion coefficients for particle populations under investigation, the diffusion coefficients are calculated in PCS by applying a multiexponential fit algorithm to the experimental correlation curve. The main question at this point is which algorithm best represents and determines the analyzed size distribution. In addition to the diffusion coefficients, the correlation curve contains noise information, which renders both fit and data interpretation more complicated (25).

No generally valid algorithm can be applied for all samples. The correct choice of algorithm depends on different factors: e.g., the type of samples analyzed and the working size range of the experimental set-up. The level of noise in the experimentally obtained correlogram is one of the most important aspects to consider. However, taking these aspects into account, PCS enables generation of extremely fruitful information that cannot be derived using other techniques (11, 30).

The algorithms mostly used are, for example, the CONTIN algorithm developed by Steven Provencher (31) or the regularization algorithm written by Maria Ivanova (30) or the general purpose (multiple narrow mode) algorithm used by certain PCS users. A main difference among these algorithms comes from the fact that each was especially developed for specific sample properties. For example, the regularization algorithm was optimized for dust-free small-particle samples such as pure proteins and small micelles. Another difference lies in the smooth parameter (sometimes called the *alpha parameter*), which is used to “smooth” the autocorrelation function before analysis. It focuses on the estimated noise level in a data set, and an understimation of the noise level can lead to “ghost peaks” in the distribution (12, 25).

Thus, it is crucial to understand the algorithm used, to be aware of the application and limitations of each, and to test whether a given algorithm is suitable for the intended application (28). For more details, please consult the literature (12, 25, 28, 31).

To better understand the influence of various algorithms for deconvolution of the measured intensity-autocorrelation function of a sample, we analyzed a ternary mixture comprising 30-nm, 80-nm, and 1,000-nm particles at a molar ratio of 185,000:3,850:1 in water using three different algorithms (general purpose, CONTIN, and multiple narrow mode). Figure 3a represents the correlogram of the sample, and Figure 3b shows the corresponding intensity particle size distributions using different algorithms. Clearly, the number and sharpness of resolved peaks differ because the smooth (alpha) parameter is different for those three algorithms:

- CONTIN (alpha parameter varies, 0.1 used as a standard)
- general purpose (alpha parameter = 0.01)
- multiple narrow mode (alpha parameter = 0.001).

Because baseline resolution is achieved, the calculated sizes (peak positions) are independent of the alpha parameter, as Figure 3 shows. Only the apparent width of the peaks is altered. An inappropriate conclusion from this example would be to use the multiple narrow mode algorithm for all samples. But because its alpha parameter is quite low, that algorithm can lead to calculation of ghost peaks resulting from noise in a data set. They cannot be directly identified and would lead to wrongly interpreted data because an additional particle population would be “observed” that is not based on the presence of real particles.

Behind the NNLS algorithm for extracting diffusion coefficient information respective to size information from an experimental correlogram, the cumulant approach is often used (18, 19, 33). From cumulant analysis, the mean or z-average size of a particle distribution from a PCS data set is calculated by moment analysis of a linear measured correlogram (24, 33). By this means, the first cumulant is used to calculate the intensity-weighted “z-average” mean size, and the second cumulant is used to calculate a parameter defined as the polydispersity index (PDI) (34). So the cumulant analysis assumes a single particle population and represents the distribution as simple and Gaussian-like, with the z-average being the mean value and the PDI describing the relative variance of a hypothetical Gaussian (34).

The NNLS algorithm uses the majority of the exponential decay correlogram for data fitting, but the cumulant analysis uses only its initial portion. An advantage to using the
cumulant analysis is the yield of a more repeatable mean value. However, this analysis is not suitable for polydisperse samples. The z-average of the sample in Figure 3 is 110 nm with a PDI of 0.45. Figure 3 represents the intensity distribution. Other distributions such as volume or number distribution are also common (Figure 4). They are calculated from the intensity distribution using either the Mie or Rayleigh-Gans-Debye (RGD) theory (12, 20, 21).

Figure 4 illustrates the investigation of a ternary mixture composed of 30-, 80-, and 1,000-nm particles at a molar ratio of 185,000:3,850:1. As the figure shows, there is a large difference in the results and interpretation of data — partly due to the relationship between the scattering intensity of the particles, which is (according to Rayleigh) proportional to the square of their molecular weight: \( I \propto R^6 \), with \( R \) as the particle radius \( (R \times \lambda / 10) \) (35). The number of large particles (1,000 nm) is extremely low in our sample. Using intensity as well as volume distribution, however, that particle population is detected (it would not be using the number distribution). Thus, a very small number of large particles dominate the PCS-measured intensity distribution, whereas a very large number of small particles (30 nm) are barely represented.

This example shows that large particles (e.g., external dust contamination) can falsify results, so we recommend performing PCS measurements under “clean” conditions (e.g., using a laminar-flow cabinet). But this weakness can also be seen as a strength of PCS. It is highly sensitive for the detection of large particles. When a totally unknown sample is investigated, analysis of the three distributions might be helpful. However, although transformation of the experimentally yielded intensity distribution to volume and number distributions seems to be easy, it is important to remember the assumptions that have to be accepted for this transformation:

- that all particles are spherical
- that all have a homogenous and equivalent density
- that their optical properties are known (refractive index)
- that the intensity distribution is without error (33, 36).

In some studies, mass distributions are also presented. But the terms mass and volume are synonymous with regard to size distribution.

**Sample and Technical Details:**
Samples can be analyzed by PCS without pretreatment (usually no separation procedure is required), depending to some extent on the overall particle size distribution in a sample. In some cases, it might be helpful for data interpretation to “classify” those particles using ultracentrifugation, for example. In such cases, the impact of sample pretreatment has to be carefully evaluated. Another aspect to consider is that the diffusion coefficients of all particles composing a sample are within the resulting correlation curve. Because PCS data are derived from Brownian motion, there is some degree of uncertainty inherent to PCS measurements. So the theoretical limit of PCS baseline resolution for a polydisperse sample is 1.7 times the particle size (12). Experience shows, however, that the limit is more likely in the range of 3 to 5 times the particle size. A sample with 10-nm and 30-nm particles can be resolved into two single peaks, but the presence of 20-nm particles will turn that result into one broad peak because the three “theoretical” peaks at 10, 20, and 30 nm cannot be baseline resolved. Thus, the presence of dimers in a monomeric antibody solution broadens the peak. For such samples, the polydispersity index is a valuable parameter for characterization.

Modern PCS instruments use compact laser diodes, high-end fiber optics, and very sensitive avalanche photo diodes (APDs) as detectors (12). Scattered light is usually measured at a single angle of 90°, but newer applications often use larger scattering angles (e.g., 155° or 173°) called backscattering to reduce multiple scattering (a phenomenon especially observed with highly scattering samples) because it can lead to misleading results.

The required PCS sample volume is quite low (75 µL using the Hellma QS 3-mm cuvette), and samples can be recovered. Some PCS systems use 96- to 1,356-well microplate formats for screening large numbers of samples (as in protein crystallography). Depending on the application of interest, it can be advantageous in some cases to pretreat a sample to remove extremely large particles. They can be removed by filtration or centrifugation. Data acquisition is fast (within a few minutes), and data are usually taken in triplicate and then averaged. Sample temperature and viscosity must be known for absolute calculation of size distribution. Viscosity is an important parameter, particularly for analyzing highly concentrated protein formulations.

**APPLICATION, RESULTS, DISCUSSIONS**
PCS is used for characterizing the hydrodynamic diameter of particles in solution. Figure 5 represents the intensity distribution of four different proteins. The single protein solution shows one peak, corresponding to one
distribution, which in each case is related to the monomer unit. Corresponding hydrodynamic diameters for freshly prepared proteins in those different solutions are as follows: concanavaline A 8 nm, antibody 11 nm, bovine serum albumin 6–7 nm, and lysozyme 3.6 nm.  

Mechanical Means: For an unfiltered BSA solution, three populations can be detected (using multiple narrow mode) located at 6, 18, and 80 nm (Figure 6). Applying a physical stress (stirring) to the solution induces the appearance of a fourth peak at ~500 nm, showing the formation of large protein particles. Although the fourth peak seems to “grow” with respect to the others, the intensity distribution shown overestimates the presence of large particles because of the relation \( I \propto R^6 \), illustrating the extremely high sensitivity of this method for large particles.

A freshly prepared antibody solution of 99.9% of monomers as determined by HP-SEC shows in PCS analysis one peak with a hydrodynamic diameter of 11 nm (Figure 7). Shaking the sample for 20 minutes induced the formation of large particles (0.7 µm), but other protein aggregates were not detected. Figure 7a shows the corresponding autocorrelation function. Changes were obvious by visual inspection of the “raw” data.

Interfering Ingredients: Protein aggregates can also be generated by the presence of certain excipients. Figure 8 shows an example of a HCLF formulation (50-mg/mL protein concentration) at pH 6.1, but with different buffer salts. Protein aggregates are formed in citrate buffer and grow rapidly over time. When such particles become much too large, they sediment and thus cannot be detected by PCS.  

PCS can be used to investigate samples for protein concentrations up to 200 mg/mL. For these samples, it is crucial to measure real viscosity and use that for characterizing the size distribution. Figure 9 represents the intensity distribution of the monomer peak for an antibody at various protein concentrations (10–145 mg/mL). Clearly, the peak shifts to larger sizes as a function of protein concentration, indicating protein–protein interactions and associations. Particle–particle excluded volume effects usually cause the apparent hydrodynamic diameter to increase with concentration, even when protein molecules remain entirely monomeric (11).

That experiment shows the formation of antibody associates. PCS measurement of highly concentrated samples can be difficult because of multiple scattering. So for investigating HCLF solutions, the optical path must be sufficiently short to prevent it. The phenomenon can also be prevented simply by choosing a scattering volume that is very close to the inner wall of the cuvette. The ways for light to move through the solution are thus shortened, so most photons are scattered only once (12). Another approach presented by Lämmle uses photon cross-correlation to enable measurement of very highly concentrated protein solutions (37).

Dilution of the 145-mg/mL sample (Figure 9) back to 10 mg/mL led to the appearance of a single peak located at 10–11 nm, as was found for a protein sample at the same concentration before the concentration step. No particle population is observed in the size range ≤5 µm (data not shown).

Pros and Cons

Experimentally derived PCS data are represented by the intensity autocorrelation curve, which contains all information regarding the diffusion coefficient distribution of an ensemble collection of particles in a sample solution. Diffusion coefficients are obtained by a mathematical procedure called deconvolution, which is applied to the intensity autocorrelation function. This procedure is ill defined. Several deconvolution algorithms are available for different applications. From the diffusion coefficient, the hydrodynamic diameter/radius is calculated.

One drawback to PCS is that it allows only qualitative information, making absolute quantification infeasible. Furthermore, this is a nonspecific detection method (particles are detected without being identified). For reliable results, some sample-specific parameters (e.g., dynamic viscosity) have to be known. And results can be affected by different PCS algorithms, which must be well understood before the right one can be chosen. Because of PCS’s high sensitivity for large particles, dust contaminants strongly affect measurement. So samples should be prepared under a laminar flow cabinet. When investigating protein samples, users must be aware of the presence of excipients (especially detergents) that may self-associate and thus form larger particles. Up to the critical micelle concentration of a detergent, micelles are formed. Absorption of incident laser light, which can occur when measuring colored samples, is another drawback. Even the choice of cuvette can strongly influence the quality of size determinations. Sample temperatures should also be controlled during measurement.

The main advantages of using PCS come from its noninvasive and nondestructive nature for measuring the size distribution of a broad range of particles (from 1–2 nm to ~5–7 µm) and allowing sample reuse. So it can analyze samples containing broad distributions of particle species. Short measurement times and high sensitivities to larger particles underline the usefulness of this technique for investigating protein samples. The sample volume can be as low as a few microliters. Unlike SEC, for example, PCS can even be used to measure HCLFs (up to 200 mg/mL), which is not feasible with most other techniques. So HCLF formulations can be characterized at their original concentration. But correct data interpretation requires strong knowledge about the data evaluation software chosen and the use of intelligent experimental design approaches.

Acknowledgments

We thank Stefan Bassarab and Uli Nienhaus for helpful and constructive discussions.


Corresponding author Dr. rer. nat. Patrick Garidel is associate director of pharmaceutical development at Boehringer Ingelheim Pharma GMB & Co. KG, D-88397 Biberach an der Riss, Germany; 49-7351-545556, fax 49-7351-835556; patrick.garidel@boehringer-ingelheim.com; www.boehringer-ingelheim.com. Fabian Kebbel is a biochemistry masters student at the University of Ulm (www.uni-ulm.de).