Molecular weight measurement of protein monomers in presence of others populations using Light Scattering measurements

Key Words: *Average molecular weight, Debye plot, Protein, Static Light Scattering (SLS), monomers, oligomers, polydispersity.*

Abstract
This note presents the basic concepts of average molecular weight determination based on light scattering measurements as well as some results obtained on standard Proteins using the "VASCO Kin" Instrument and its "NanoKin" software. We show, in particular, how one can use the Nanokin algorithms to deal with the presence of several populations in samples.

Introduction

Absolute molecular weight (Mw) measurement is of primary importance for anyone who deals with macromolecules such as proteins or polymers. Today, there are many known ways to estimate or determine averaged molecular weight such as Gel Permeation Chromatography (GPC), Membrane Osmometry, Sedimentation Velocity/Diffusion, and optical method such Static Light Scattering (SLS). Actually, the intensity of light scattered by a specie depends directly on its overall polarizability, which is itself, related to its molecular weight (or its size). The instrument "VASCO Kin" allows to measure the average molecular weight using such an optical method and features a dedicated calculation tool included in its software "NanoKin".
A common issue encountered with light scattering method is related to the presence of multiple populations in a sample and, in particular, in the case of protein monomers characterization, their coexistence with aggregates or oligomers having larger molecular weights. Actually, even in very low amount, the high level of scattered light intensities from these species can interfere significantly with the monomers signal. This phenomenon systematically induces some bias in the molecular weight result calculation.
Hereafter, we present the light scattering Mw measurement principle illustrated with Bovine Serum Albumin (BSA) and Lysozyme measurements and we show how the Mw calculation tool of "Vasco Kin" can improve the characterization of such a multi-populated samples using the Dynamic Light Scattering (DLS) sizes distribution.

SLS Molecular weight measurements: principle and theory

Static Light Scattering is a well-known characterization technique used in chemistry and macromolecules research which principles has been described many decades ago[1,2]. It is based on the fact that a portion of a laser beam focused into a sample is scattered in all directions by the macromolecules or particles dispersed in the liquid (see figure 1). Yet, this scattered light intensity depends on one hand, the sample properties, such as the molecular weight of solubilized species, their refractive index contrast, and their concentration for instance, and on the other hand, the optical set-up characteristic, such as the observation angle, the distance between the detector and the sample cell, the quantum efficiency of the detector, the scattering volume, *etc*.

[1] *P. Debeye (1944), "light scattering in solutions". J Appl Phys. 15 (4) :338*
[2] *B.H. Zimm (1945). "Molecular Theory of the scattering of lights in fluids". J. Chem . Phys 13 (4) :141*

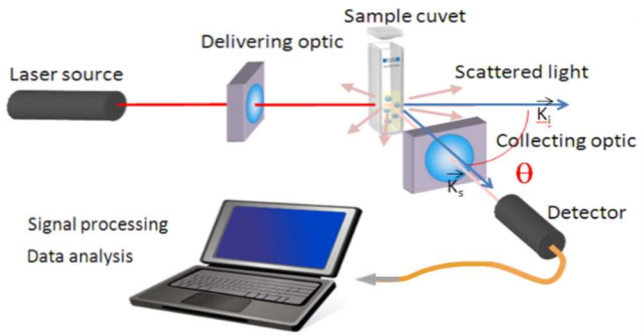

Figure 1: General set up of a light scattering experiment.

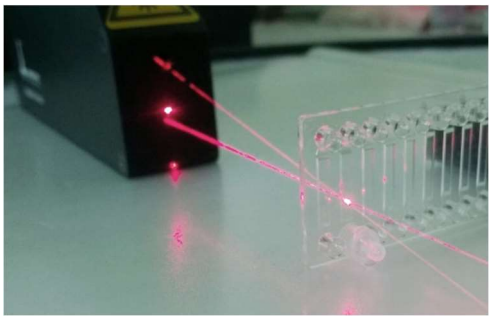

Figure 2: in-situ optical head of "VASCO Kin".

Then, it has been shown that the Rayleigh ratio **R(θ)** of a sample, related to the ratio of the scattered intensity to the incident intensity for a given scattering angle θ, can be approximated by the Zimm/Rayleigh equation:

$$\frac{KC}{R(\theta)} = \frac{1}{M_W\, P(\theta)} + 2\, A_2 C + \cdots \qquad (1)$$

Where $\mathbf{M_w}$ is the (weight averaged) molecular weight and **C** the macromolecules concentration.
K is called the optical contrast constant and is defined as:

$$K = \frac{4\pi^2\, n_0^2\, (\frac{dn}{dC})^2}{N_A\, \lambda^4} \qquad (2)$$

With $\boldsymbol{n_0}$ the refractive index of the solvent, $\boldsymbol{\lambda}$ the laser wavelength, $\boldsymbol{\frac{dn}{dC}}$ the differential refractive index increment of the solution upon the macromolecules concentration and $\boldsymbol{N_A}$ is the Avogadro number (6.023 x 10^{23}).
P(θ) is the corrective form factor that we have to consider with large particles/molecules with respect to the laser wavelength (> $\lambda/20$). When characterizing smaller objects like proteins or macromolecules, this factor can be approximated by 1 (Rayleigh domain). Note that it is precisely in this operating regime that we should use the "VASCO kin". Actually, addressing the Mw estimation of larger objects implies to determine P(θ) through the light intensity measurement at many angles (typically 6 or more) or to calculate it theoretically knowing the structure of the objects (Gaussian coil, rigid, sphere, rod..). Since we are addressing proteins with particle size smaller than 30 nm, we will assume the form factor equal to 1 in the next part of this note.
Finally, $\mathbf{A_2}$ is the second virial coefficient, a corrective factor for non-ideal solution related to the macromolecules-solvent interactions. The value and the sign of this coefficient is directly proportional to the affinity of the macromolecules with the solvent[3].

The Rayleigh ratio **R(θ)** has to be determined experimentally. In principle, it can be calculated by measuring the ratio of scattered light intensity on the incident light intensity at given angle and distance r (and normalized by unit of scattering volume and unit of scattering solid angle). In practical, these intensities of light are very difficult to measure due to multiple instrumental factors. In order to circumvent their tedious determination, one

[3] *For A_2 >0, Macromolecules have a good affinity for the solvent and the colloid should be rather stable (and polymers form large/spread random coils). For A_2 <0, the macromolecules have a better affinity for themselves than for the solvent and then tends to aggregate (and polymers form dense random coils). When A_2 =0, the strength of macromolecules-solvent interaction is the same magnitude as that of macromolecules-macromolecules interaction (Theta solvent).*

measures intensity ratios with respect to a reference sample, typically toluene for example. Knowing the theoretical Rayleigh ratio of toluene $\mathbf{R_{tol}}$ we can then write the following equation:

$$R(\theta) = \frac{I_A(\theta) - I_0(\theta)}{I_{tol}(\theta)} R_{tol} \left(\frac{n}{n_{tol}}\right)^2 N(\theta, n) \quad (3)$$

Where $\boldsymbol{I_A(\theta)}$ is the intensity scattered by the sample at the angle θ, $\boldsymbol{I_0(\theta)}$ and $\boldsymbol{I_{tol}(\theta)}$ the intensity scattered by the solvent only and the toluene in the same conditions (respectively the background and calibration measurement). $\boldsymbol{n}$ is r the refractive index of the sample and $\boldsymbol{n_{tol}}$ the one of toluene. The $\left(\frac{n}{n_{tol}}\right)^2$ term is involved in this equation to correct the solid angle normalization of scattered light between toluene and the sample. $\boldsymbol{N(\theta, n)}$ is another correction factor related to the change of scattering volume between toluene and the sample and highly depends on the optical set-up geometrical configuration (square or round cuvette, *etc.*).

Debye plot: analysis and interpretation

Equation (1) shows that the Rayleigh ratio has, at first approximation, a linear dependence with concentration C. Then, if one plots KC/R(θ) vs C at a constant scattering angle θ, we should in principle obtain a line having a slope equal to 2.$\mathbf{A_2}$ and an intercept at C= 0 equal to $\frac{1}{M_W}$ (assuming P(θ)= 1) as shown on figure 3. This representation is called the Debye plot.

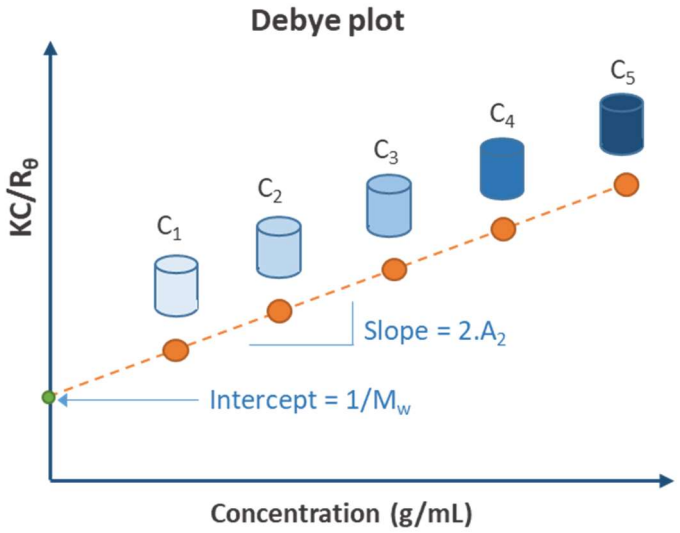

Figure 3: Debye plot. Values of KC/R(θ) are plotted as a function of sample concentration. The slope of the corresponding linear regression is equal to 2xA_2 while the intercept at C=0 is equal to 1/M_w.

Basically, the Mw measurement principle consists in measuring the average sample scattered intensity at different concentrations. Such a light intensity measurement has also to be performed for the solvent alone (background measurement) and for toluene or another reference material (calibration measurement). Then, for each concentration, the corresponding ratio KC/R(θ) is calculated using (2) and (3) and plots on a graph versus the concentration C. A linear regression fit allows to determine the intercept and the slope of the straight line from which Mw and A2 are deduced.

Examples of Mw measurement using the VASCO Kin

Bovine serum Albumin (BSA)

BSA is a small protein which is commonly used as a standard. According to literature, ideal monomeric albumin has a molecular weight of 66.5 kDa determined by mass spectrometry[4] but higher values, up to 69 kDa, have also been reported probably due to the co-presence of dimers or higher oligomers in solution[5]. BSA shows a natural tendency to dimerize depending on the buffer properties (ionic strength, pH…)[6].

The BSA used in this study is provided by Sigma-Aldrich under lyophilized powder form (≥ 96%, ref: A6003). This powder has been dissolved at different concentrations between 1 and 10 mg/mL in a 0.2µm-filtrated 0.1M KCL buffer. Measurements of both hydrodynamic diameters (DLS) and average scattered intensity (SLS) have been performed using "Vasco Kin" with its *in-situ* optical head in optical glass cuvette at 25°C (see figure 2).
Alongside with the samples measurements, the characterization of the solvent background, and filtered toluene were carried out in order to calibrate the instrument.
For this experiment, we used a dn/dc of 0.190 mL/g[7] and a $R_{tol}(\theta)$ toluene of 1.402E-5 cm-1[8].

Molecular weight calculation processes the average intensity of light scattered by the studied species in solution. However, some samples may contain several populations which all participate to the overall amount of light scattered by the sample and, then, affect significantly the result. This effect is well illustrated for the case of BSA sample as shown below.
Actually, high resolution DLS size characterization using the "Vasco Kin" shows that at least two populations are detected in these BSA solutions (see figure 4): a population having a hydrodynamic diameter close to 7nm, likely to contain mainly BSA monomers, and a larger population around 20nm, either oligomers, aggregates or impurities, present in low amount.

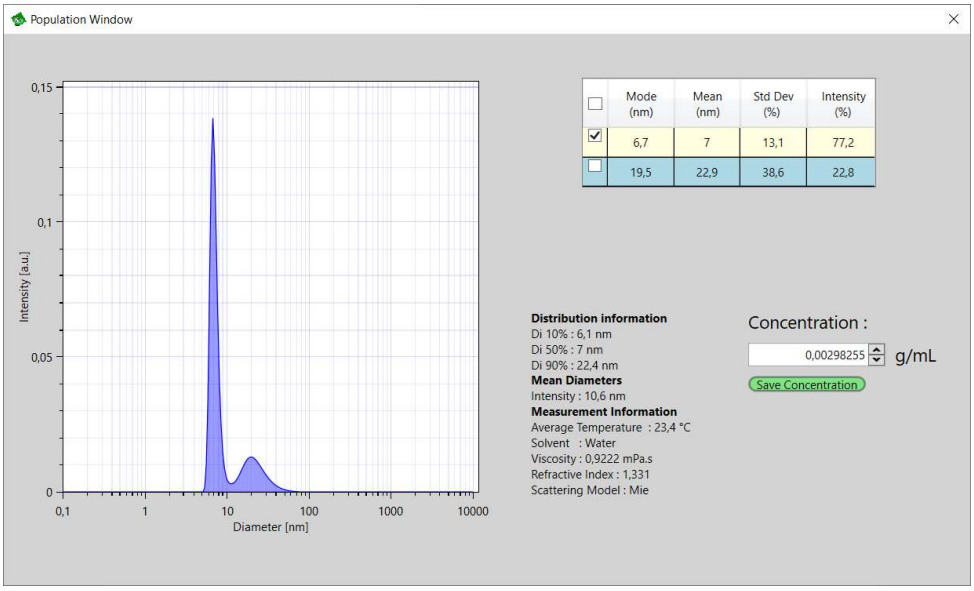

Figure 4: "NanoKin" software panel showing the DLS sizes measurement of BSA in 0.1M KCl buffer at 0,00298g/mL. This window also allows to select on which population is processed the Mw calculation.

[4] *Hirayama, K., et al., Rapid confirmation and revision of the primary structure of bovine serum albumin by ESIMS and Frit-FAB LC/MS, BBRC, 173(2), 639 (1990).*
[5] *Scatchard, G., et al., Preparation and properties of serum and plasma proteins; osmotic equilibria in solutions of serum albumin and sodium chloride J. Amer. Chem. Soc. 68, 26.0 (1946).*
[6] *D.C. Carter, J.X. Ho, Structure of serum albumin, Adv. Protein Chem.45 (1994) 153–203.*
[7] *A. Theisen and al, Refractive Increment Index Data-Book, Nottingham University Press.*
[8] *W. Kaye and al. Low-angle laser light scattering: Rayleigh factors and depolarization ratios. Appl. Opt., 13:1934–1937, 1974. 27, 96.*

Then, if the whole light intensity scattered by this sample is considered for the molecular weight calculation i.e. coming from the two populations, the software draws the corresponding Debye plot (see figure 5) and calculates a Mw value of **83.6 kDA** which is far from the expected values for BSA monomers.

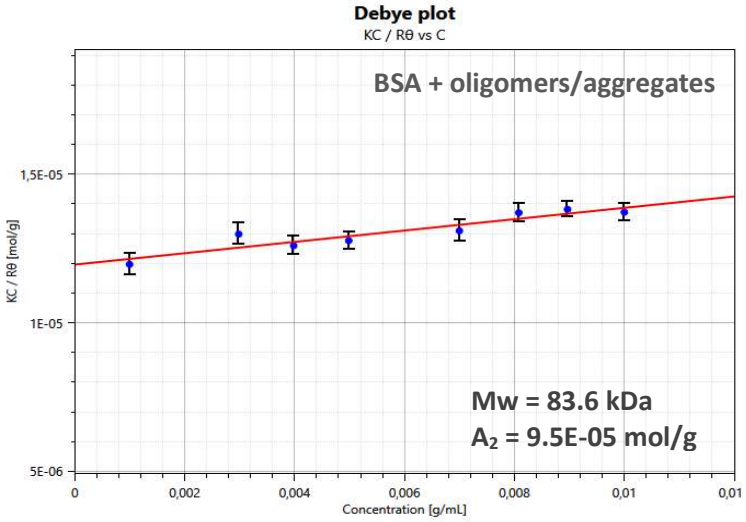

Figure 5: "Raw" Debye plot of BSA in 0.1M KCl in the case the whole sample scattering intensity is processed. The values of KC/R(θ) are plotted as a function of sample concentration (blue dots). The red line show the corresponding linear regression.

Yet, the "NanoKin" software allows to select the population of 7nm diameter to apply the processing to its sole scattering intensity. In this case, the R_θ calculation at a given concentration is weighted by the amplitude of the corresponding population determined from the DLS intensity size distribution (close to 77% in this example).
Note that this operation is only possible if the DLS inversion algorithm is sufficiently accurate to separate both populations. This can be especially tricky when dealing with small monomers in presence of their oligomers due to the very small gap between population average sizes. Actually, corresponding distributions proposed by DLS are easily overlapping in these cases, preventing a correct evaluation of their respective intensity.
To ensure a good data processing, the Nanokin software relies on the SBL (for Sparse Bayesian Learning) inversion algorithm which provides sizes distributions with a high level of resolution, as shown in this example with BSA.

By using this functionality, the software draws the corresponding Debye plot (see figure 6) and calculates the Mw value at **65.7 kDA** with a virial coefficient of **3.0E-4 mol/g**. As seen above, this result is close to the theoretical expected value and consistent with published values.

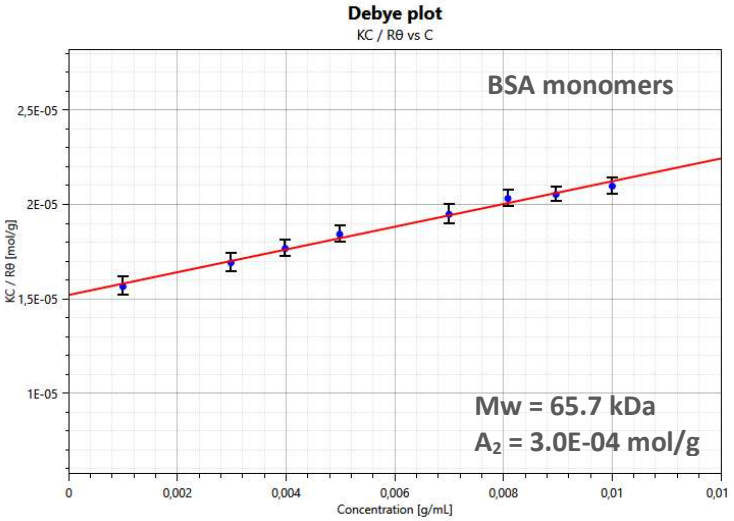

Figure 6: "Corrected" Debye plot of BSA in 0.1M KCl in the case only the intensity scattered by monomers is processed. The values of KC/R(θ) are plotted as a function of sample concentration (blue dots). The red line show the corresponding linear regression.

For each sample, the standard deviation of the average scattered light intensity measurement is displayed on the graph by an error bar (between 2 and 3% in this case). The overall error estimations is determined at 2.6 kDa for the molecular weight and 5E-5 mol/g for the virial coefficient. Note that this calculation is performed using the method of maximum slops (as define in ISO17025) and mainly involves the results obtained at minimum and

maximum concentrations. Then, the accuracy of these two measurements is critical for such an overall error estimation.

	Mw (kDa)	ΔMw (kDa)	**A_2 (mol/g)**	ΔA_2 (mol/g)
BSA + oligomers/aggregates	83.6	3.2	9.5E-05	3.7E-05
BSA monomers	**65.7**	2.6	**3.0E-04**	5.00E-05

Table 1: Molecular weights (M_w) and virial coefficients (A_2) of BSA calculated from the measurements shown in figure 5. "BSA + oligomers/aggregates" corresponds to the result using the whole scattered intensity while BSA monomers corresponds to the result after the intensity correction using the DLS distribution. ΔMw and ΔA_2 are the overall measurement errors estimated by the method of maximum slops.

Lysozymes

The procedure described above for BSA molecular weight measurement have been performed for Lysozyme proteins from chicken egg white, having a theoretical molecular weight of 14.3 kDa (ref 2970, from Sigma-Aldrich).
This powder has been dissolved at different concentrations between 0.8 and 10 mg/mL in a 0.2µm-filtrated 0.1M KCL buffer. Measurements of both hydrodynamic diameters (DLS) and average scattered intensity (SLS) have been performed using "Vasco Kin" with its *in-situ* optical head in optical glass cuvette at 25°C (see figure 2).
Alongside with the samples measurements, the characterization of the solvent background, and filtered toluene were carried out in order to calibrate the instrument.
The results have been calculated using a dn/dc of 0.183 mL/g[9] for the Lysozyme and a $R_{tol}(\theta)$ toluene of 1.402E-5 cm-1[10].
DLS size characterizations of these Lysozyme solutions show (see on figure 7) a population having a hydrodynamic diameter close to 3.9 nm, likely to contain mainly Lysozyme monomers, and a very large population of aggregates or impurities present in low amount, and having sizes from 20 nm to few microns.

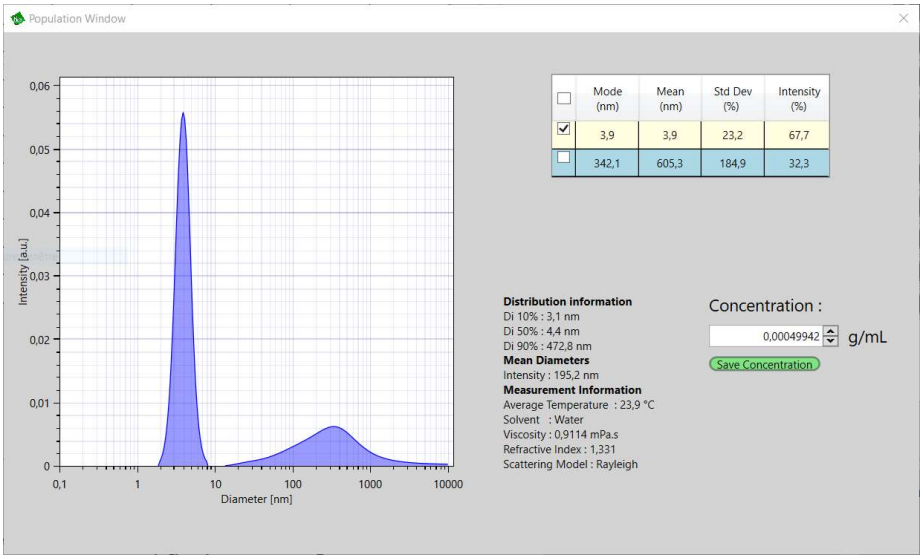

Figure 7: "NanoKin" software panel showing the DLS sizes measurement of Lysozyme in 0.1M KCl buffer at 0,000499g/mL. This window also allows to select on which population is processed the Mw calculation.

[9] *Narayanan, J., & Liu, X. Y. (2003). Protein Interactions in Undersaturated and Supersaturated Solutions: A Study Using Light and X-Ray Scattering. Biophysical Journal, 84(1), 523–532.*
[10] *W. Kaye and al. Low-angle laser light scattering: Rayleigh factors and depolarization ratios. Appl. Opt., 13:1934–1937, 1974. 27, 96.*

As explained with the BSA case, we use the Nanokin software functionality to apply the molecular weight calculation on the monomer population only. The corresponding Debye plot is shown on figure 8 while the resulting molecular weights and virial coefficients are reported in the table 2 below.
The calculated Mw value of 14.9 kDA is in good agreement with the theoretical molecular weight of 14.3kDa while the overall error estimated at 1.4 kDa for this analysis remains acceptable (<10%). The measurement accuracy could be improved, here, by purifying the monomeric component of samples and especially removing the larger inhomogeneities detected in corresponding sizes distributions.

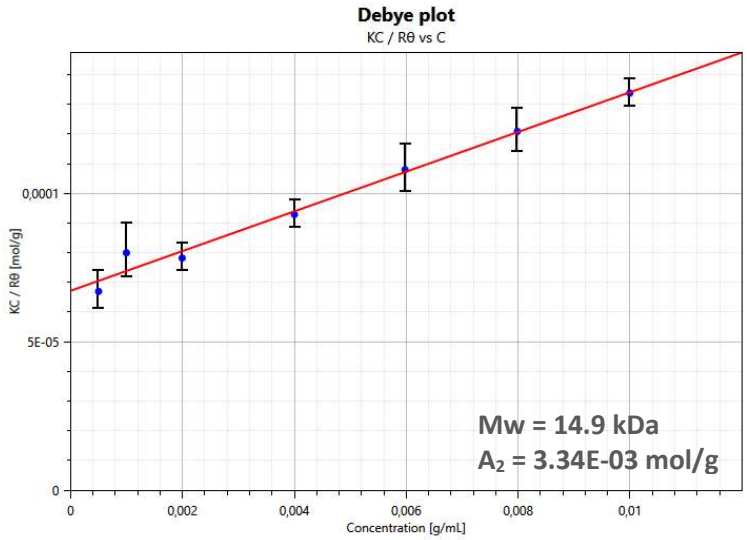

Figure 8: "Corrected" Debye plot of Lysozzyme in 0.1M KCl in the case only the intensity scattered by monomers is processed. The values of KC/R(θ) are plotted as a function of sample concentration (blue dots). The red line show the corresponding linear regression.

	Mw (kDa)	ΔMw (kDa)	**A_2 (mol/g)**	ΔA_2 (mol/g)
Lysozyme monomers	**14.9**	1.4	**3.34E-03**	5.75E-04

Table 2: Molecular weights (M_w) and virial coefficients (A_2) of Lysozyme calculated from the measurements shown in figure 8. ΔMw and ΔA_2 are the overall measurement errors estimated by the method of maximum slops.

Conclusion

The results shown in this note show that the molecular weight of BSA and Lysozyme have been successfully measured using the Vasco kin and its dedicated software tool. These measurements are in good agreement with the expected values for the monomer molecular weights and were achievable within the range of 10% maximum error estimation despite of the presence of large inhomogeneities (especially in the case of lysozymes).
We also observe that a multi-populated sample can be correctly characterized by using the corresponding DLS size distribution to focus the algorithm processing on the population of interest (the population of monomers here). This advanced functionality featured in the" Nanokin" software appears to be critical in such an optical measurement technic very sensitive to the presence of large aggregates or inhomogeneities even if, indeed, quality and purity of the protein solutions remain a key point of Mw measurement accuracy.