University of Ljubljana Faculty of Mathematics and Physics



Department of Physics

Seminar - 4th year

Dynamic light scattering and application to proteins in solutions

Author: Dejan Arzenšek Advisor: prof. dr. Rudolf Podgornik Co-Advisor: dr. Drago Kuzman

Ljubljana, May 19, 2010

Abstract

Dynamic light scattering (DLS) measures time-dependent fluctuations in the scattering intensity arising from particles undergoing random Brownian motion. Diffusion coefficient and particle size information can be obtained from the analysis of these fluctuations. More specifically, the method provides the ability to measure size characteristics of proteins in a liquid medium. Proteins consist of polypeptide chains that are sensitive to a wide range of parameters such as temperature and chemical environment. Preparation method, storage conditions and/or buffer choice can all influence the size and quality of proteins in a sample. This seminar describes the theory of method and presents examples how light scattering may be used as a method of characterizing the size of proteins in solutions.

Contents

INTRODUCTION

1	INT	TRODUCTION	1
2	DY	NAMIC LIGHT SCATTERING THEORY	2
	2.1	SCATTERED LIGHT	2
	2.2	SCATTERING MODELS	5
		2.2.1 RAYLEIGH SCATTERING FOR SPHERICAL MOLECULES	6
	2.3	FLUCTATIONS AND AUTOCORRELATION FUNCTIONS	7
	2.4	FROM BROWNIAN MOTION TO THE HYDRODYNAMIC RADIUS	9
		2.4.1 PROTEIN HYDRODYNAMIC SIZES	10
3	MU	ULTIPLE SIZE DISTRIBUTION - MODALITY	11
	3.1	POLYDISPERSITY	11
4	DATA ANALYSIS IN DYNAMIC LIGHT SCATTERING		
	4.1	CALCULATION OF THE SIZE DISTRIBUTION	12
	4.2	OBTAINING SIZE INFORMATION FROM THE CORRELATION FUNCTION	12
5	AP	PLICATION TO PROTEINS	14
	5.1	TRACING PROTEIN AGGREGATES	14
	5.2	PROTEIN THERMAL STABILITY	15
	5.3	PROTEIN ASSEMBLY AND AGGREGATION STUDIES	15
		5.3.1 CONCENTRATED PROTEIN SAMPLES	16
		5.3.2 DYNAMIC DEBYE PLOT	17
6	CO	NCLUSION	18
R	References		

1 INTRODUCTION

Dynamic light scattering (DLS) is an important experimental technique in science and industry. It is also known as *Photon Correlation Spectroscopy* (PCS). The acronym PCS is only one of several different names that have been used historically for this technique. The first name given to the technique was quasi-elastic light scattering (QELS) because, when photons are scattered by mobile particles, the process is quasi-elastic. QELS measurements yield information on the dynamics of the scatterer, which gave rise to the acronym DLS (dynamic light scattering) [1].

Photon correlation spectroscopy has become a powerful light-scattering technique for studying the properties of suspensions and solutions of colloids, biological solutions, macromolecules and polymers, that is absolute, non-invasive and non-destructive. Technique is also useful for measuring the speed, for example for microorganisms that float in the solution, or to analyze flow in fluids. Shining a monochromatic light beam, such as a laser, onto a solution with particles in Brownian motion causes a Doppler Shift when the light hits the moving particle, changing the wavelength (typically red light at 633 nm or near-infrared at 830 nm for biomolecular applications [1].) of the incoming light. This change is related to the size of the particle (figure 1). It is possible to compute the sphere size distribution and give a description of the particle's motion in the medium, measuring the diffusion coefficient of the particle and using the autocorrelation function.

This technique is one of the most popular methods used to determine the size of particles. It is applicable in range from about 0.001 to several microns, which is difficult to achieve with

other techniques. This is because the dimensions are too small for optical spectroscopy and to large for electron microscopy. Problems are also part of the nature of the substances, that are usually liquids or gels. This may considerably change the properties. Also the investigation of materials by X-ray light is not suitable, since the solution is usually too rare for yielding sufficiently accurate results. If we make them thicker, their properties may be quite different [1].

Commercial "particle sizing" systems mostly operate at only one angle (90°) and use He-Ne red light laser (wavelength 633 nm). He-Ne lasers are cheap and compact, but weaker. The NIR lasers are used for biomolecular samples, because we need more power.

Today, DLS is recognized as a standard instrument widely used in industry, such as a biopharmaceutical industry. Knowledge on size, shape and morphology of the particles is of great importance in developing pharmaceutical formulations of proteins. These parameters namely influence the properties of proteins, their solubility, distribution in the body and indirectly also the technology of their production.



2 DYNAMIC LIGHT SCATTERING THEORY

2.1 SCATTERED LIGHT

 to

A tipical set-up for the scattering experiment consist of a laser beam illuminating a sample and a detector set up at scattering angle θ measuring the intensity $I(\theta, t)$ of the scattered light. Typically, the incident and scattered beams are shaped by apertures, slits, or by optics such as lenses. Usually, the incident beam is vertically polarized as the detector moves in a horizontal plane and by this is "catching" the strongest signal. The region of the sample which is illuminated by both incident and outgoing beam, and "seen" by the detector is the "scattering volume" [2]. Figure 2 shows a scheme of a typical light-scattering set-up, where the incident light is a plane electromagnetic wave:

$$\mathbf{E}_{\mathbf{i}}(\mathbf{r},t) = \hat{\mathbf{n}}_{\mathbf{i}} E_0 \exp[\mathbf{i}(\mathbf{k}_{\mathbf{i}} \cdot \mathbf{r} - \omega_i t)], \qquad (1)$$

with polarization $\hat{\mathbf{n}}_{\mathbf{i}}$ perpendicular to the scattering plane, amplitude E_0 . $\mathbf{k}_{\mathbf{i}}$ is wave vector of the incident light, having magnitude $|\mathbf{k}_i| = k_i = \frac{2\pi n}{\lambda_i}$ where λ_i is the wavelength of incident light in vacuo (ω_i is its angular frequency) and n is the refractive index of the scattering medium. $\mathbf{E}_{\mathbf{i}}(\mathbf{r},t)$ is the electric field at the point in space \mathbf{r} at time t.



Figure 2: **a)** Scheme of a typical lightscattering experiment, **b)** expanded view of the scattering volume, showing a volume element dV at position **r** from the origin O [2].

When the molecules in the illuminated volume are subjected to this incident electric field, their constituent charges experience a force. They are accelerated and consequently radiate light. The scattered electric field is the superposition of the scattered fields from subregions of the illuminated volume. If the subregions are optically identical (have the same dielectric constant), there will be no scattered light in other than the forward direction. If the subregions are optically different (have different dielectric constants), then the amplitudes of the light scattered are no longer identical.

Thus in *semimacroscopic view*, originally introduced by Einstein, light scattering is a result of local fluctations in the dielectric constant of the medimum [3]. We know from kinetic theory that molecules are constantly translating and rotating (i.e. Brownian motion), so that the instantaneous dielectric constant of a given subregion (which depends on the positions and orientations of the molecules) will fluctate and thus give rise to light scattering.

The Maxwell's equations for a nonconducting, nonmagnetic, nonabsorbing medium¹ with average dielectric constant ε_0 (and refractive index $n = \sqrt{\varepsilon_0}$), may be used to obtain the basic equation for the scattered field [3]. A medium has in the linear approximation², a *local* dielectric constant:

$$\underline{\underline{\varepsilon}}(\mathbf{r},t) = \varepsilon_0 \underline{\underline{I}} + \underline{\underline{\delta}} \underline{\underline{\varepsilon}}(\mathbf{r},t), \tag{2}$$

where $\underline{\delta \varepsilon}(\mathbf{r}, t)$ is the dielectric constant fluctation tensor and \underline{I} is the second-rank unit tensor. We considered that the dielectric contant has no time dependence due to the movement.

To obtain the component of the scattered electric field $E_S(R, t)$ at a large distance R from the scattering volume V, we have to consider the Maxwell's equations [3]. The scattered filed has polarization $\hat{\mathbf{n}}_{\mathbf{f}}$, propagation vector $\mathbf{k}_{\mathbf{f}}$ ($|\mathbf{k}_{\mathbf{f}}| = k_f = \frac{2\pi n}{\lambda_f}$; λ_f the wavelength of scattered light in vacuo), and frequency ω_f . Here we define the scattering vector \mathbf{q} in terms of the scattering geometry as $\mathbf{q} = \mathbf{k}_i - \mathbf{k}_f$. The angle θ between vectors \mathbf{k}_i and \mathbf{k}_f is called the scattering angle (figure 2). It is usually the case that the wavelength of the incident light is changed very little in the scattering process so that $|\mathbf{k}_i| = |\mathbf{k}_f|$.

¹Many of the applications of light scattering are to ionic solutions, which are conducting media. However since the ions are massive the charge density will vary on a much slower time scale than that specified by the laser frequency ($\sim 10^{14}Hz$). Thus the medium may be considered to be nonconducting as far as this derivation is concerned. Since we consider the medium to be nonabsorbing, we are restricted to incident wave lengths which are not resonant with any molecular transitions of the scattering medium.

 $^{^{2}}$ Note that in general the scattered field is much lower in amplitude than the incident field.

The dielectric fluctation can be expressed in terms of the spatial Furier transform:

$$\underline{\delta\varepsilon}(\mathbf{q},t) = \int_{V} d^{3}r \exp(\mathbf{i}\mathbf{q}\cdot\mathbf{r})\underline{\delta\varepsilon}(\mathbf{r},t), \qquad (3)$$

where $\delta \varepsilon_{if}(\mathbf{q}, t) \equiv \mathbf{\hat{n}_f} \cdot \underline{\delta \varepsilon}(\mathbf{q}, t) \cdot \mathbf{\hat{n}_i}$, is the component of the dielectric constant fluctation tensor (*structural tensor*) along the initial and final polarization directions. The component of the scattered electric field is:

$$E_S(\mathbf{q},t) = \frac{-k_f^2 E_0}{4\pi R \varepsilon_0} \exp(\mathbf{i}k_f R - \omega_i t) \delta \varepsilon_{if}(\mathbf{q},t).$$
(4)

This equation can be recognized as the formula describing the radiation due to an oscillating point dipole. So the incident electric field induces a dipole moment of strength proportional to Fourier component of dielectric fluctation tensor at arbitrary \mathbf{q} .

Equation (4) is written for the case when we know dielectric fluctation tensor throughout the solution. But for further discussion of this seminar, we have to consider N discrete scattering objects (particles like molecules or aggregates) suspended in a liquid in the scattering volume V, whose centers of mass at time t are described by position vector $\mathbf{r}_{\mathbf{j}}(t)$ (microscopic view). Therefore the dielectric fluctation tensor for a particle j, can be written (by integrating equation (4) over the particle):

$$\delta \varepsilon_{if}^{j}(\mathbf{q},t) = b_{j}(\mathbf{q},t)e^{\mathbf{i}(\mathbf{q}\cdot\mathbf{r}_{j}(t))},\tag{5}$$

where $b_j(\mathbf{q}, t)$ is the "scattering length" (component of structural tensor) of particle j and is also defined in terms of the spatial Furier transform as $b_j(\mathbf{q}, t) = \int_{V_{particle}} d^3r' \delta \varepsilon_{if}^j(\mathbf{r}', t) e^{\mathbf{i}(\mathbf{q}\cdot\mathbf{r}')}$. $\underline{b_j}(\mathbf{q}, t)$ varies in time because the particle rotates and vibrates, while the phase factor $e^{\mathbf{i}(\mathbf{q}\cdot\mathbf{r}_j(t))}$, \overline{var} is in time because the particle translates. In a fluid, if the particles are electronically weakly coupled and we ignore multiple scattering, equation (4) may be written simply as a summation:

$$E_S(\mathbf{q},t) = \frac{-k^2 E_0}{4\pi R\varepsilon_0} \exp(\mathbf{i}kR - \omega t) \sum_{j=1}^N b_j(\mathbf{q},t) e^{\mathbf{i}(\mathbf{q}\cdot\mathbf{r}_j(t))}.$$
(6)

Here is ω_i written as ω and k_f as k, because the scattering is ellastic. In equation (6), the multiple scattering is neglected because we have omitted higher order terms in $\delta \varepsilon_{if}^j$.

It is the scattering intensity that is measured directly, rather than the electric field. Intensities and fields are related by $I_S(\mathbf{q},t) \equiv |E_S(\mathbf{q},t)|^2$. The scattered field takes on the form of a random diffraction pattern, often reffered to as a "speckle pattern", which consist of dark and bright regions of light (constructive and destructive interference in the far field charge).

Each speckle, of intensity $|E_S(\mathbf{q},t)|^2$ (figure 3), constitutes a single Fourier component of a density fluctation of length scale $(\frac{2\pi}{q})$. In sufficiently dilute solutions, the particles so rarely encounter each other that we can assume their positions to be statistically independent (their behaviours are uncorre-



Figure 3: The scattered light falling on the detector [1].

lated) and were monodispersity is assumed³. But there has been growing interest in interactions

³A collection of objects are called **monodisperse**, or **monosized**, if they have the same size, and shape when discussing particles, and the same mass, when discussing polymers. A sample of objects that have an inconsistent size, shape and mass distribution are called **polydisperse**.

since concentrated particle solutions provide a challenging *many-body* problem. The many body problem is also of considerable practical importance, because many processes in *industry* (pharmaceutical, food, detergents, paints, etc.) involve concentrated solutions.

For identical particles we can write average intensity as:

$$I_{S}(\mathbf{q},t) = \frac{k^{4}|E_{0}|}{16\pi^{2}R^{2}\varepsilon_{0}^{2}} \langle N \rangle \langle b^{*}(\mathbf{q},0)b(\mathbf{q},t) \rangle \langle e^{\mathbf{i}\mathbf{q}\cdot\mathbf{\Delta r}(t)} \rangle.$$
(7)

We introduce $\langle e^{\mathbf{i}\mathbf{q}\cdot\mathbf{\Delta r}(t)}\rangle$ as the 'dynamic structure factor' which depends on translation of the particles where $\mathbf{\Delta r}(t) = \mathbf{r}(t) - \mathbf{r}(0)$ is the displacement of particle in time t. $\langle N \rangle$ is the average number of particles in volume V. The term $\langle b^*(\mathbf{q}, 0)b(\mathbf{q}, t)\rangle$ depends on rotation of particles.

2.2 SCATTERING MODELS

To interpret light scattering experiments, we begin with a discussion of light scattering theories. Classical light scattering theory was derived by Lord Rayleigh and is now called Rayleigh theory. Rayleigh developed theory for particles much smaller than the wavelength of light (tipically we take size less than $\lambda/10$ or around 60 nm for He-Ne laser as this criterion) and they have arbitrary forms. Many biomolecules will violate this criterion. So, the formal light scattering theory may be categorized in terms of two teoretical frameworks. One is the theory of *Rayleigh scattering*, the second is the teory of *Mie scattering* (after Gustav Mie) that encompasses the general spherical scattering solution without a particular bound on particle size. Rayleigh scattering theory is generally preferred if applicable, due to the complexity of the Mie scattering formulation. Sometimes is more desirable either the volume or number distribution, than the intensity distribution. Both can be calculated from the intensity distribution using Mie theory, assuming spherical particles and that there is no error in the intensity distribution.

When particles become larger than around 60nm the scattering changes from being isotropic, i.e. equal intensities in all directions (in the scattering plane), to a distortion in the forward scattering direction (figure 4). When the size of the particles becomes roughly equivalent to the wavelength of the illuminated light then we observe that the scattering becomes a complex function with maxima and minima with respect to angle: Mie theory solves exactly the Maxwell's equations linking the interaction of electromagnetic radiation with matter and predicts the observed maxima and minima when $d \sim \lambda$ (d is a diameter of a particle).



Figure 4: Comparison between scattering models. For particles size **larger** than a wavelength, Mie scattering predominates. Mie scattering is **not strongly wavelength dependent** (this is why clouds appear white) [4].

2.2.1 RAYLEIGH SCATTERING FOR SPHERICAL MOLECULES

In this model the main point is that the particle being so much smaller than the wavelength all regions within the particle are subject to a similar electric field so that the scattered waves produced by dipole oscillations of the electrons bound within the particle are in phase. The dipoles induced are also parallel with the plane of polarisation of the light (by definition), so if the scattered light is viewed in a plane at right angles to this plane of polarisation the scattered light will have no angular distribution and is equal in all directions (isotropic). All particle-sizing techniques have an inherent problem in describing the size of non-spherical particles. The sphere is the only shape that can be described by unique number. We measure some property of our particle and assume that this refers to a sphere, hence deriving our one unique number (the diameter of this sphere) to describe our particle. This ensures that we do not have to describe our 3-D particles with three or more numbers.

For a homogeneous spherical particle with radius R', term $\langle b^*(\mathbf{q}, 0)b(\mathbf{q}, t)\rangle$ in equation (7) may be calculated analytically. In solution the particles have the same tensor $\underline{\delta\varepsilon} = \delta\varepsilon \underline{I}$ and we have equality $\mathbf{\hat{n}_f} \cdot \underline{\delta\varepsilon}(\mathbf{q}, t) \cdot \mathbf{\hat{n}_i} = \delta\varepsilon(\mathbf{n_f} \cdot \mathbf{n_i})$. Our term becomes (also called a 'form factor $(P(\mathbf{q}))$ ' of the spherical particle) [2]:

$$\langle b^*(\mathbf{q},0)b(\mathbf{q},t)\rangle = \delta\varepsilon^2 (\mathbf{n_f} \cdot \mathbf{n_i})^2 (\int_{V_{particle}} e^{-\mathbf{iq}\cdot\mathbf{r}'} dV' \int_{V_{particle}} e^{\mathbf{iq}\cdot\mathbf{r}'} dV') \propto \\ \propto (\int_0^{2\pi} \int_{-1}^1 \int_0^{R'} e^{\pm iqr'\cos\theta'} r'^2 dr' d(\cos\theta') d\varphi')^2 = (\frac{4\pi}{a^3} [-qR'\cos(qR') + \sin(qR')])^2$$
(8)

In equation (7), we note the following dependence $I \propto \lambda^{-4}$. This indicates, for instance, that blue light is scattered more than red light.

In the Rayleigh limit $(qR' \ll 1)$, we could approximate term (8). So the intensity in Rayleigh scattering is finally:

$$I_{S}(\mathbf{q},t) = \frac{k^{4}|E_{0}|}{R^{2}\varepsilon_{0}^{2}} \langle N \rangle \delta\varepsilon^{2} (\mathbf{n_{f}} \cdot \mathbf{n_{i}})^{2} R^{\prime 6} \langle e^{\mathbf{i}\mathbf{q} \cdot \boldsymbol{\Delta}\mathbf{r}(t)} \rangle.$$
(9)

Notice the factor d^6 (*d* is the particle diameter). The d^6 factor means, that it is difficult with DLS to measure, say, a mixture of 1000 nm and 10 nm particles beacuse the contribution of the total light scattered by the small particles will be extremely small. A very simple way of describing the difference is to consider a sample that contains only two sizes of particles (5nm and 50nm) but with equal numbers of each size particle (figure 5).

The first graph at right shows the result as a number distribution. As expected the two peaks are of the same size (1:1) as there are equal number of particles. The second graph shows the result as a volume distribution. The area of the peak for the 50nm particles is 1000 times larger the peak for the 5nm (1:1000 ratio). This is because the volume of a 50nm particle is 1000 times larger that the 5nm particle (volume of a sphere is equal to $4/3\pi r^3$). The third graph



Figure 5: Two populations of spherical particles of diameters 5 nm and 50 nm present in equal numbers [5].

shows the result as an intensity distribution. The area of the peak for the 50nm particles is now 1,000,000 times larger than the peak for the 5nm (1:1000000 ratio). This is because large

particles scatter much more light than small particles, the intensity of scattering of a particle is proportional to the sixth power of its diameter (from Rayleigh's approximation). It is worth repeating that the basic distribution obtained from a DLS measurement is intensity - all other distributions are generated from this.

2.3 FLUCTATIONS AND AUTOCORRELATION FUNCTIONS

Pattern of scattered light comprises a grainy random diffraction, or 'speckle', pattern (figure 6). At some points in the far field the phases of the light scattered by the individual particles are such that the individual fields interfere largely constructively to give a large intensity; at other points destructive interference leads to a small intensity.



The intensity $I(\mathbf{q}, t)$ scattered to a point in the far field fluctuates randomly in time. Clearly, information on the motions of the particles is encoded in this random signal - at the simplest level, the faster the particles move, the more rapidly the intensity fluctuates.

When the light is scattered, negligible frequency shift occurs. Relative frequency shift $(\frac{\Delta\omega}{\omega})$ due to particle motion, on which the light is scattered (Doppler shift), is between 10^{-15} and maximum 10^{-16} . The particles will vary on a much slower time scale than that specified by the laser frequency ($\sim 10^{14}Hz$). Therefore, this effect is negligible. The rate at which these intensity fluctuations occur will depend on the size of the particles. Figure 7 schematically illustrates typical intensity fluctuations arising from a dispersion of large particles and a dispersion of small particles. The small particles cause the intensity to fluctuate more rapidly than the large ones.

It is possible to directly measure the spectrum of frequencies contained in the intensity fluctuations arising from the Brownian motion of particles. But the best way is to extract usable information from the fluctating intensity by constructing its time correlation function, defined as [2]:



Figure 7: Typical intensity fluctations for large and small particles [5].

$$G^{(2)}(q,\tau) = \langle I_S(\mathbf{q},0)I_S(\mathbf{q},\tau)\rangle \equiv \lim_{T \to \infty} \frac{1}{T} \int_0^T dt I_S(\mathbf{q},t)I_S(\mathbf{q},t+\tau)$$
(10)

As can be seen from its definition, this quantity effectively compares the signal $I_S(\mathbf{q}, t)$ with a delayed version $I_S(\mathbf{q}, t + \tau)$ of itself for all starting times t and for a range of delay times τ . At zero delay time equation (10) reduces to perfect correlation:

$$\lim_{\sigma \to 0} \langle I_S(\mathbf{q}, 0) I_S(\mathbf{q}, \tau) \rangle = \langle I_S^2(\mathbf{q}) \rangle.$$
(11)

For delay time much greater than the typical fluctuation time T_C of the intensity, fluctuations are uncorrelated so that the average can be separated:

$$\lim_{\tau \to \infty} \langle I_S(\mathbf{q}, 0) I_S(\mathbf{q}, \tau) \rangle = \langle I_S(\mathbf{q}, 0) \rangle \langle I_S(\mathbf{q}, \tau) \rangle = \langle I_S(\mathbf{q}) \rangle^2.$$
(12)

Thus, the intensity correlation function decays from the mean-square intensity at small delay times to the square of the mean at long times (the characteristic time T_C of this decay is a measure of the typical fluctuation time of the intensity). For a large number of monodisperse particles in Brownian motion, the correlation function is an exponential decaying function of the correlation time delay τ . The normalized time correlation function of the scattered intensity is defined by [2]:

$$g^{(2)}(q,\tau) \equiv \frac{\langle I_S(\mathbf{q},0)I_S(\mathbf{q},\tau)\rangle}{\langle I_S(\mathbf{q})\rangle^2}.$$
(13)

The normalized time correlation function of the scattered field is defined by:

$$g^{(1)}(q,\tau) \equiv \frac{\langle E_S^*(\mathbf{q},0)E_S(\mathbf{q},\tau)\rangle}{\langle I_S(\mathbf{q})\rangle}.$$
(14)

Scattered field E_S can be regarded as a sum of independent random variables $(E_S = \sum_n E_S^n)$, where E_S^n is the scattered field of the *n*-th subregion of the scattering volume. The central limit theorem implies that E_S , must be distributed according to a Gaussian distribution. A Gaussian distribution is completely characterized by its first and second moments (average and standard devation). Circumstances in which this assumption may be invalid, are in case of critical fluids (very long correlation lengths) and in very dilute solutions. On this assumption, two time correlation functions are connected via the 'Siegert relation' [3]:

$$g^{(2)}(q,\tau) = A(1+\beta[g^{(1)}(q,\tau)]^2).$$
(15)

Here β is a factor (intercept of the correlation function) which represents the degree of spatial coherence of the scattered light over the detector and is determined by the ratio of the detector area to the area of one speckle ($0 < \beta < 1$). So, the coherence factor β measures the degree of coherence. A is a baseline. In ideal example the detector aperture is usually chosen to accept about one speckle and β is taken as 1, such as baseline A. If the particles are large the signal will be changing slowly and the correlation will persist for a long time (figure 8). If the particles are small and moving rapidly then correlation will reduce more quickly.



Figure 8: Typical correlograms from a sample. As can be seen, the rate of decay for the correlation function is related to particle size as the rate of decay is much faster for small particles than it is for large [1].

Viewing the correlogram from a measurement can give a lot of information about the sample. The time at which the correlation starts to significantly decay is an indication of the mean size of the sample. The steeper the line, the more monodisperse the sample is. Conversely, the more extended the decay becomes, the greater the sample polydispersity.

2.4 FROM BROWNIAN MOTION TO THE HYDRODYNAMIC RADIUS

Brownian motion is the random motion of particles due to the bombardment by the particles suspended in within a liquid. The larger the particle, the slower the Brownian motion will be. Smaller particles are "kicked" further by the solvent molecules and move more rapidly. An accurately known temperature is necessary for DLS because knowledge of the viscosity is required (because the viscosity of a liquid is related to its temperature). The temperature also needs to be stable, otherwise convection currents in the sample will cause non-random movements that will ruin the correct interpretation of size. The velocity of the Brownian motion is defined by a property known as the **translational** ('free-particle') **diffusion coefficient** in infinitely-dilute solutions and is usually given as the symbol D_0 .

In previous section we saw, that the correlation function is an exponential decaying with time delay τ and $g^{(1)}(\tau)$ is the sum of all the exponential decays contained in the correlation function. It is frequently called the 'measured intermediate scattering function' $F(q,\tau) = \langle e^{i\mathbf{q}\cdot\boldsymbol{\Delta}\mathbf{r}(\tau)} \rangle$ as the dynamic structure factor from equation (7). According to the central limit theorem, the probability for a particle displacement should be the Gaussian distribution function. So we get [3]

$$F(q,\tau) = e^{\frac{-q^2 \langle \Delta \mathbf{r}^2(\tau) \rangle}{6}},\tag{16}$$

where $\langle \Delta \mathbf{r}^2(\tau) \rangle$ is the mean-square displacement of the particle in time τ and q is the scattering vector. For a diffusing particle we have a relation [3]:

$$\langle \Delta \mathbf{r}^2(\tau) \rangle = 6D_0 \tau, \tag{17}$$

and finally

$$F(q,\tau) = e^{-q^2 D_0 \tau}.$$
(18)

This relation is valid for a dilute solutions of identical non-interacting spheres. According to the Einstein relation the translational self-difusion coefficient is [3]:

$$D_0 = \frac{k_B T}{\zeta},\tag{19}$$

where is k_B Boltzmann's constant, T is the temperature and ζ is the friction constant. For spherical particles we have the Stokes approximation $\zeta = 6\pi\eta R_H$, where η is the viscosity of the solvent and R_H is the particles hydrodynamic radius.

Finally, the size of a particle is calculated from the translational diffusion coefficient by using the *Stokes-Einstein equation*:

$$R_H = \frac{k_B T}{6\pi\eta D_0}.$$
(20)

Note that the radius that is measured in DLS is a value that refers to how a particle diffuses within a fluid so it is referred to as a hydrodynamic diameter. The radius that is obtained by this technique is the radius of a sphere that has the same translational diffusion coefficient as the particle. Factors that affect the diffusion speed of particles are discussed in the following sections.

On figure 8 we saw that for the monodisperse dilute solution, $g^{(1)}(\tau)$ is presented by a single exponential as follows: $g^{(1)}(\tau) = e^{-\Gamma\tau}$. Now, we can obtain for correlation function:

$$g^{(2)}(\tau) = A(1 + \beta [e^{-2\Gamma\tau}]), \qquad (21)$$

where Γ is the decay rate (the inverse of the correlation time). The decay rate is:

$$\Gamma = \tau_c^{-1} = q^2 D_0. \tag{22}$$

The hydrodynamic diameter ($\mathbf{R}_{\mathbf{H}}$), that is being reported, is the diameter or the radius of the hard sphere⁴ that diffuses at the same speed as the particle or molecule being measured. The translational diffusion coefficient will depend not only on the size of the particle "core", but also on any surface structure, as well as the concentration and type of ions in the medium. The ions in the medium and the total ionic concentration can affect the particle diffusion speed by changing the thickness of the electric double layer called the Debye length (K^{-1}). Any change to the surface of a particle that affects the diffusion speed will correspondingly change the apparent size of the particle. The nature of the surface and the polymer, as well as the ionic concentration of the medium can affect the polymer conformation, which in turn can change the apparent size by several nanometres.

2.4.1 PROTEIN HYDRODYNAMIC SIZES

The hydrodynamic diameter of a non-spherical particle is the diameter of a sphere that has the same translational diffusion speed as the particle. If the shape of a particle changes in a way that affects the diffusion speed, then the hydrodynamic size will change. For example, small changes in the length of a rod-shaped particle will directly affect the size, whereas changes in the rod's diameter, which will hardly affect the diffusion speed, will be difficult to detect. The conformation of proteins are usually dependent on the exact nature of the dispersing medium. As conformational changes will usually affect the diffusion speed. Factors that influence the protein hydrodynamic sizes are the molecular weight of the molecule, the shape of conformation of the molecule and also whether the protein is in his native or folded state. In the case of anistropic objects, we could correct electric field autocorrelaton function. But, on average the proteins are in spherical shape, due to conformation changes and also aggregates form spherical like objects. On figure 9 we can see typical protein hydrodynamic sizes.



Figure 9: Typical protein hydrodynamic sizes. Insulin has the globular structure (spheroproteins comprising "globe"-like proteins). The size that we will get from the technique would be very similar to that of the molecule itself. In case of Immunoglobulin, the size we get is the radius of the sphere which has the same average diffusion speed as that molecule has. For Lysozyme the size is smaller as the dimension of the molecule [6].

⁴Hard spheres are widely used as model particles.

3 MULTIPLE SIZE DISTRIBUTION - MODALITY

For samples with a multiple size distribution, $g^{(1)}(\tau)$ generalizes to a sum of exponential functions:

$$g^{(1)}(\tau) = \sum_{n=1}^{M} A_n \exp(-\Gamma_n \tau).$$
 (23)

Here, the coefficients A_n represent the intensity-weighted contributions to the overal decay rate of particles having different diffusion coefficients $(D_{0_n} = \Gamma_n/q^2)$. This multiexponential dynamic process could be caused by mixtures of particles or polymers of different sizes or by unimodal distribution of particles of finite width (quite small or large).

3.1 POLYDISPERSITY

Polydispersity (**PdI**) in the area of light scattering is used to describe the width of the particle size distribution. In the light scattering area, the term polydispersity is derived from the polydispersity index, a parameter calculated from a Cumulants analysis of the DLS measured intensity autocorrelation function. In the Cumulants analysis [7], a single particle size is assumed and a single exponential fit is applied to the autocorrelation function. About this we will discus in next section.

If one were to assume a single size population following a Gaussian distribution, then the polydispersity index would be related to the standard deviation (σ) of the hypothetical Gaussian distribution in the fashion shown below [8].

$$PdI = \frac{\sigma^2}{Z_D^2},\tag{24}$$



Figure 10: Monodisperse and polydisperse size distribution [6].

where Z_D is Z average size, the intensity weighted mean hydrodynamic size (cumulants mean) of the ensemble collection of particles. It is important to emphasize, that Z average is the cumulants "hydrodynamic radius" value reported by most DLS analysis software and may represent the average of several species. Using the above expression, polydispersity can then be defined in the following terms:

Polydispersity Index (PdI) = Relative variance, Polydispersity (Pd) = Standard deviation (σ) or width (also known as the absolute polydispersity) or % Polydispersity (%Pd) = Coefficient of variation = $(PdI)^{1/2} \times 100$ (also called the relative polydispersity).

Of the three terms defined above, the coefficient of variation or % Pd is one of the more often used parameters in the area of protein analysis. The fitting of the measured autocorrelation function is an ill-posed problem. Which means that there are multiple distributions that can be transformed from the correlogram, depending upon how much noise is assumed. Because of the inherent uncertainty arising from this problem, DLS derived size distributions will always have a small degree of polydispersity, even if the sample consists of a very monodisperse analyte. As a rule of thumb, samples with %Pd <~ 20%, are considered to be monodisperse (figure 10).

4 DATA ANALYSIS IN DYNAMIC LIGHT SCATTERING

4.1 CALCULATION OF THE SIZE DISTRIBUTION

The particle size distribution from DLS is derived from a deconvolution of the measured intensity autocorrelation function of the sample. Generally, this deconvolution is accomplished using a non-negatively constrained least squares (NNLS) fitting algorithm, common examples being CONTIN, Regularization, and the General Purpose and Multiple Narrow Mode algorithms [9].

The fitting function for $g^{(1)}$ consists of a summation of single exponential functions (equation (23)) which is a constructed as a grid of exponentials with decay rate Γ_n . The weighting factor (A_n) is proportional to the intensity correlation function value, i.e. correlation function values at small times have a higher weight than those at large times. The main objective of the data inversion consists of finding the appropriate distribution of exponential decay functions which best describe the measured field correlation function. A spherical particle with radius R_n will produce a correlation function with decay rate Γ_n according to the following expression [9]:

$$\Gamma_n = (D_n q^2) = (k_B T / 6\pi \eta R_n) ((2\pi n / \lambda_i) \sin(\theta / 2))^2.$$
(25)

The normalized display of A_n vs. R_n (or A_n vs. diameter) is the intensity particle size distribution. The average sizes of the single distribution peaks are the intensity weighted averages, and are obtained directly from the size histogram using the following expression [9]:

$$\langle R \rangle = \frac{\sum_{n} A_{n} R_{n}}{\sum_{n} A_{n}}.$$
(26)

The peak width or standard deviation (σ), indicative of the distribution in the peak, is also obtained directly from the histogram [9]:

$$\sigma = \frac{\sqrt{\langle R^2 \rangle - \langle R \rangle^2}}{\langle R \rangle} \quad \text{where} \quad \langle R^2 \rangle = \frac{\sum_n A_n R_n^2}{\sum_n A_n}. \tag{27}$$

4.2 OBTAINING SIZE INFORMATION FROM THE CORRELATION FUNCTION

Size is obtained from the correction function by using various algorithms. There are two approaches that can be taken. A fit of a single exponential to the correlation function (cumulant fit) to obtain the average hydrodynamious radius - R_H (Z-average diameter) and a estimate of the width of the distribution (Polydispersity index) or fit of a multiple exponential to the correlation function to obtain the distribution of particle sizes (described in previous section) (figure 11). The size distribution obtained is a plot of the relative intensity of light scattered by particles in various size classes and is therefore known as an intensity size distribution. If the distribution by intensity is a single fairly smooth peak, then there is little point in doing the conversion to a volume distribution using the Mie theory. If the optical parameters are correct, this will just provide a slightly different shaped peak. However, if the plot shows a substantial



Figure 11: Obtaining DLS data from cumulant analysis on correlogram [6].

tail, or more than one peak, then Mie theory can make use of the input parameter of sample refractive index to convert the intensity distribution to a volume distribution.

In the Cumulants approach, the exponential fitting expression is expanded as a power series expansion around mean decay rate (with additional term $\mu_2 \tau^2$) to account for polydispersity or peak broadening effects, as shown below [10]:

$$g^{(2)}(\tau) = A(1 + \beta(\exp(-2\Gamma\tau + \mu_2\tau^2))).$$
(28)

The expression is then linearized and the data fit to the form shown below, where the D subscript notation is used to indicate diameter. The first Cumulant or moment (a_1) is used to calculate the intensity weighted Z average mean size and the second moment (a_2) is used to calculate a parameter defined as the polydispersity index (PdI) [10].

$$y(\tau) = \frac{1}{2} \ln[g^{(2)}(\tau) - A] = \frac{1}{2} \ln[A\beta \exp(-2\Gamma\tau + \mu_2\tau^2)] \approx \frac{1}{2} \ln[A\beta] - \langle\Gamma\rangle\tau + \frac{\mu_2}{2}\tau^2 = a_0 - a_1\tau + a_2\tau^2$$
$$Z_D = \frac{1}{a_1} \frac{K_B T}{3\pi\eta} [\frac{4\pi n}{\lambda_i} \sin(\frac{\theta}{2})]^2 \qquad PdI = \frac{2a_2}{a_1^2} \qquad \beta = \frac{\exp(2a_0)}{A}.$$
(29)

It can be shown that for the assumption of a single peak Gaussian size distribution the Z average size corresponds to the mean, with the square root of the polydispersity index corresponding to the relative standard deviation of that distribution. Note however, that obtaining a Z average and PdI does not mean that the distribution is Gaussian, although smaller PdI values are a good indication that a Gaussian might be a reasonable approximation of the real size distribution.

5 APPLICATION TO PROTEINS

Proteins (also known as polypeptides) are organic compounds made of amino acids arranged in a linear chain and folded into varios forms (e.g. globular form). The amino acids in a polymer are joined together by the peptide bonds between the carboxyl and amino groups of adjacent amino acid residues. Proteins consist of polypeptide chains that are sensitive to a wide range of parameters such as temperature and chemical environment. Preparation method, storage conditions and/or buffer⁵ choice can all influence the size and quality of proteins in a sample.

Protein stability is a particularly relevant issue today in the pharmaceutical field and will continue to gain more importance as the number of therapeutic protein products in development increases. However, if a therapeutic protein cannot be stabilized adequately, its benefits to human health will not be realized. Achieving this goal is particularly difficult because proteins are only marginally stable and the efficacy of protein drugs can be compromised by instability. From production to administration, various factors such as pH, shear and thermal stress can compromise the stability of therapeutic proteins. One major aspect of protein instability is self-association leading to aggregation⁶. Aggregates can reduce the efficacy of protein drugs and can lead to immunological reactions and toxicity.

During the development of a protein formulation, a combination of appropriate analytical methods must be used to detect subtle changes in the state of a protein to ensure the efficacy and safety. This seminar review the DLS applications for biophysical characterization of proteins and detection of aggregates.

5.1 TRACING PROTEIN AGGREGATES

The size of aggregates can be determined by DLS; however in the case of a multimodal distribution (e.g. presence of monomers and aggregates), the results are often biased towards the larger species. Scattering intensity is proportional to molecular radius of power six (figure 12), making the technique ideal for indentifying the presence of trace amounts of aggregate. The fact is that once we start to see the presence of aggregates, which may be there in very low mass population. Simply having a small amount of tetramer in presence of monomer will cause the significant change in the mean hydrodynamic size and also in the percentage polydispersity. But in volume (or mass) distribution, we may not see the presen



Figure 12: Scatterd light intesity versus radius. Size relating to scattered light intensity is very dependent on this relationship [6].

mass) distribution, we may not see the presence of aggregates or oligomers (figure 13).

⁵A buffer solution is an aqueous solution consisting of a mixture of a weak acid and its conjugate base or a weak base and its conjugate acid. Buffer solutions are used as a means of keeping pH at a nearly constant value in a wide variety of chemical applications.

⁶Particle aggregation is direct mutual attraction between particles (atoms or molecules) via van der Waals forces or chemical bonding. When there are collisions between particles in fluid, there are chances that particles will attach to each other and become larger particle. There are 3 major physical mechanisms to form aggregate: Brownian motion, fluid shear and differential settling.



Figure 13: Intensity and volume size distributions for two different protein size populations. The intensity distribution is closest to the raw data. The volume distributions corresponds to results derived from methods where concentrations are measured (UV-vis) [11].

5.2 PROTEIN THERMAL STABILITY

Thermal stability can be assessed by the measurement of the size and scattering intensity as a function of temperature. A protein 'melts' under the influence of heat when the molecules denature – leading to massive aggregation. This transition is visible in light scattering studies as a significant increase in both size and scattering intensity. The marked point where both the size and the intensity start to increase significantly is called the melting point (T_M) .



Figure 14: Melting point determination of the protein sample $(T_M = 56^{\circ}C)$ [12].

Melting point is representative of the stability of the protein under the current solution conditions. The protein that has a larger melting point is more thermal stable (figure 14).

5.3 PROTEIN ASSEMBLY AND AGGREGATION STUDIES

We do not only to know whether are the aggregates present, but possibly get the informations about resolving between aggregates and oligomers in our sample. And finally, how highly concentrated protein solutions influence on measurements and the influence of sample concentration on the results obtained (dynamic Debye plots). One of the problems with DLS is that we don't have ability to resolve monomer from oligomer and is very difficult to resove between oligomers and aggregates. The best that DLS can do in term of resolution is about double resize. So for example if we have a population in the sizes of 5 nm in order to baseline result another peak it would have to be 10 nm. Someone can think that if we take a double molecular weight, therefore we go from monomer to dimmer, we double the size. That is not the case. If we have information about conformation of the molecules. We can model what the size increase it will be. For the globular proteins double resize is the approximately the six time of the molecular weight. In another words, if we can baseline resolve monomer and hexsamer, we couldn't baseline resolve monomer, trimer, heksamer. We can use DLS in order to get the information about the presence of another ligaments in there. For example we taken Lysozyme and BSA. And we have the assumption that the BSA is the tetramer, because the molecular weight is four time of that monomer and molecular BSA is the roughly four time than that of Lysozyme. So we can simulate the presence of tetramer and monomer together and see what influence of various amounts of tetramers have on Z-average sizes we get and the mean hydrodynamic sizes we get and also percentage polydispersity. This is presented on picture 15. On this picture we see,



Figure 15: Simulation of the presence of tetramer and monomer together [6].

that when we get pure Lysozyme, we get a very low PdI and subsequently very low % Pd. As we start to doubt into the Lysozyme known masses of BSA, we can immediately begin to see the % Pd increasing. In this particular example, the mean size is not so sensitive, but % Pd we can certainly see that increase. In 0.9 % of BSA we can se that Z-average and also % Pd now increase quite significantly. From this is apparent that we can detect very small amounts of tetramer in the presence of monomer. In intensity distribution when we start with Lysozyme, we get narrow peak and BSA peak is large (pink second peak). In mass distribution we can see that very small amount of mass causes quite significant changes on sizes we were obtaining.

5.3.1 CONCENTRATED PROTEIN SAMPLES

At such high concentrations when the molecules are close to each other, the effects are no longer just simple contribution from electroviscous effects or structural alterations but nonspecific interactions arising as a result of polarity of surface residues, net charge, dipoles, and multipole moments in macromolecules come into play. The sum of these interactions contributes significantly to the free energy of the system. Nonspecific pair-wise interactions are weaker at larger distances and can be either repulsive (steric, electrostatic) or attractive (electrostatic, hydrophobic). At high concentrations ($\approx 100 \text{mg/ml}$), these interactions would increase the nonideal behavior in addition to the contributions from the excluded volume effect.

Measuring with DLS in high concentrated solutions may spoil measured correlation function. This is caused by effect called *multiple scattering*. It is very common that scattering centers are grouped together, and in those cases the radiation may scatter many times. The main difference between the effects of single and multiple scattering is that single scattering can usually be treated as a random phenomenon and multiple scattering is usually more deterministic. Because the location of a single scattering center is not usually well known relative to the path of the radiation, the outcome, which tends to depend strongly on the exact incoming trajectory, appears random to an observer. With multiple scattering, the randomness of the interaction tends to be averaged out by the large number of scattering events, so that the final path of the

radiation appears to be a deterministic distribution of intensity. We could see effects of protein concentration on much lower concentrations as 100 mg/ml. Particular in buffers which contain low salt, we could see these affects quite remarkably. So the question is, what is considered as high protein concentration. On figure 16 is shown a classic example:

In this particular example, we have the evidences that suggest the presence of salt shading intermolecular interactions. The molecule can be very charge indeed and when we increase the protein concentration thereby reduce intermolecular distances between these protein molecules, we are going to cause some nature repulsion between them. They will be desperately trying to get away from each other. And that will increase the diffusion speed. And because of the diffusion speed increases, the apparent size decreases. So the apparent size of the molecule is much smaller in the presence of dialysed water. When we put salt in there and thereby shade the repulsive forces present, the diffusion speed is



Figure 16: The effect of salt on measured protein size. The green curve shows protein measured in dylaised water, the red one in 500 mM NaCl solution [6].

slown down and therefore the apparent size now increases.

5.3.2 DYNAMIC DEBYE PLOT

If there is a concentration dependence in the results obtained, a dynamic Debye plot should be constructed. It is recommended that if we are making DLS measurement on any sample. we should check that the results we get are independent on concentration. In our case we have seen the dependence on concentration which can be minimized in the presence of salt. But as we go up with concentration the apparent size become smaller. In this situation it is recommended that we plot as known the dynamic Debye plot. We make measurements of the diffusion coefficient at several concentrations of the sample. And we plot the diffusion coefficient in the dependence of concentration and than extrapolate to zero con-



Figure 17: Dynamic Debye plot for protein Lysozyme at quite low concentration. [6].

centration to determine the diffusion coefficient in the lime of infinite dilution. In figure 17, we have a example at quite low concentration. In 10 mM NaCl is quite the significant change in the diffusion coefficient. It is increasing when concentration goes up. We believe that this is due to increasing repulsion because the intermolecule distances are now decreased. The presence of salt minimized that electrostatic repulsion effect and therefore diffusion coefficient are much more linear. In both of this cases if we extrapolate to the limite, we find that we get virtually the same diffusion coefficient. And this if we put it in Stokes-Einstein equation gives us same sizes. So, if we study the concentrated dependence, we will see changes in viscosity due to changes in Brownian motion and changes in intermolecular interactions. Brownian motion strongly de-

pends on intermolecular interactions. In example: repulsive interactions induce faster motion, attractive interactions induce slower. Therefore, the dynamic Debye plot provides information about the molecular interactions in a given solution.

6 CONCLUSION

DLS is rapid, non-invasive technique for determination of protein size. As any other particle sizing technique, DLS has advantages and disadvantages and it is particularly important to use it strictly within the framework of physical laws, if meaningful results are to be obtained. In DLS, scattering intensity fluctations are monitored and then correlated. The intensity fluctations are a consequence of particle motion, and then measured properly in the correlation analysis of the distribution of diffusion coefficients. The size is then calculated using the Stokes-Einstein equation. This technique provides rapid access to size information for the characterization of proteins. It is very sensitive to the presence of aggregates. This is extremely important considerations for the large majority of pharmaceutical and biomolecular applications. The behavior of proteins in solution is controlled by several interactions. At high concentrations, additional specific attractive interactions resulting from charge fluctuations contribute to self-association.

References

- [1] Zetasizer APS User Manual. December 2008. Manual on-http://www.malvern.com.
- [2] Anna Kozina. Cristallization Kinetics and Viscoelastic Properties of Colloid Binary Mixtures With Depletion Attraction. PhD thesis, 2009.
- [3] B.J. Berne R. Pecora. Dynamic Light Scattering: With Application to Chemistry, Biology and Physics. Dover Publications, third edition, 2000.
- [4] J.H. Wen. Dynamic light scattering: Principles, measurements, and applications. Course on-http://http://www.che.ccu.edu.tw/~rheology/DLS/outline3.htm, 9.4.2010.
- [5] Dynamic light scattering: An introduction in 30 minutes. Technical note (MRK656-01) on-http://http://www.malvern.com, 9.4.2010.
- [6] Protein interactions investigated with dynamic light scattering. 2009. Webinar on-http: //http://www.malvern.com.
- [7] D.E. Koppel. Analysis of macromolecular polydispersity in intensity correlation spectroscopy: The method of cumulants. *The Journal of Chemical Physics*, 57(11), December 1972.
- [8] What does polydispersity mean? Faq on-http://http://www.malvern.com, 9.4.2010.
- [9] How is the distribution calculated in dynamic light scattering? Faq on-http://http://www.malvern.com, 9.4.2010.
- [10] What is the z average? Faq on-http://http://www.malvern.com, 9.4.2010.
- [11] Effect of angle on resolving particle size mixtures using dynamic light scattering. Technical note (MRK1136-01) on-http://http://www.malvern.com, 9.4.2010.
- [12] L. Zhou J. Chen. Therapeutic protein characterization using light scattering techniques. Technical note (MRK704-01) on-http://http://www.malvern.com, 9.4.2010.