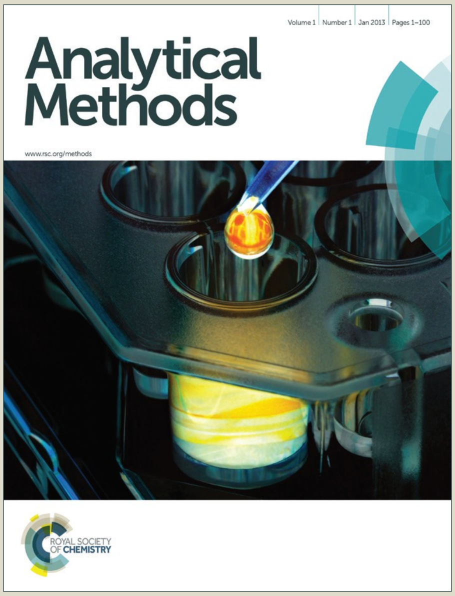
Analytical Methods

Accepted Manuscript



ROYAL SOCIETY
OF CHEMISTRY

www.rsc.org/methods

Methodologies for the Taylor Dispersion Analysis for mixtures, aggregates and the mitigation of buffer mismatch effects

Seyi Latunde-Dada*, Rachel Bott, Karl Hampton, Jenika Patel and Oksana Iryna Leszczyszyn

Malvern Instruments Ltd., Grovewood Road, Malvern, Worcestershire, WR14 1XZ, UK.

ABSTRACT: Taylor Dispersion Analysis (TDA) is a fast and simple method for determining hydrodynamic radii. For a mixture of solutes, current TDA methods lead to an average hydrodynamic radius for the different constituents of the mixture. In this paper, we first derive this average for the fitting method. Next, we present a method of deconvoluting a taylorgram from a mixture into its constituent taylorgrams so that the hydrodynamic radii and relative proportions of the individual components can be obtained. Using the differentials and integrals of the taylorgram, near-accurate initial estimates for the parameters of the constituent taylorgrams are obtained. These are used as seed parameters in least-squares fitting algorithms which find the optimum solutions for the constituent taylorgrams. Furthermore, the proximity of the seed parameters to the solutions is a measure of the confidence in the accuracy of the fits. The method is applied to two-, three- and four- component mixtures as well as aggregated samples with good agreement obtained between the results and the expected values. In addition, a fitting method which mitigates the effect of concentration mismatches between the eluent phase and the sample matrix phase is presented.

1. Introduction

Taylor dispersion analysis (TDA) is a fast and absolute method for determining the diffusion coefficients, and hence the hydrodynamic radii of molecules. The method, sometimes referred to as Taylor-Aris dispersion, was first described by Taylor in his classic paper[1]. In 1956, Aris developed the method further by accounting for the longitudinal diffusion of the molecules[2].

In the early years, this technique was applied to the determination of gaseous[3] and liquid diffusion coefficients[4-6]. With the use of fused silica microcapillaries, TDA regained interest and has been used to analyze amino acids, peptides, proteins, small molecules, macromolecules, nanoparticles and biosensors[7-24]. Since it is absolute, no calibration is required and the knowledge of the sample concentration is not required.

The diffusion coefficient of the injected solute can be deduced by fitting Taylor's solution to the concentration profile or taylorgram of the solute[25]. Alternatively, this can be achieved by calculating the moments of the profile[3-8, 12, 26, 27](moment method) or by measuring its height and area[28]. The analysis can either be carried out at a single detection point or at two spatially separated detection points. These methods are referred to as single detection TDA and double detection TDA respectively.

There have been studies dealing with the use of TDA for mixtures of solutes and, more specifically for aggregation analysis[16, 17, 21, 26, 29-33]. Most of these methods have relied on the use of the moment and area-height methods to calculate average diffusion coefficients for mixtures. These averages vary depending on the TDA method used. For example, if a mass-concentration-sensitive detector is used, the moment method gives a mass-weighted harmonic average diffusion coefficient[17] whilst the area-height method gives the square of the mass-weighted average value of the square roots of the individual diffusion coefficients[25, 27]. These averages are known as Taylor average diffusion coefficients which can be converted to Taylor average hydrodynamic radii. A recent method known as the cumulant method has been used[30] to compute a polydispersity index based on the ratio between the Taylor average from the moment method and the gamma average diffusion coefficient. Another method proposed recently[31] proposes the use of the Constrained Regularized Linear Inversion (CRLI) approach as a new data processing method to extract the probability density functions of diffusion coefficients from taylorgrams. This is a more rigorous approach which requires no *a priori* knowledge of the number of components.

The fitting method for TDA provides functions which closely match the taylorgram by adjusting the parameters of the functions. To do this effectively requires the input of seed parameters which act as initial estimates from which the fitting algorithm can find the optimum solutions. In general, the better the initial parameter estimates, i.e. the closer the initial estimates are to the optimal solutions, the greater is the probability of finding the optimal solutions. The more unconstrained or random an initial estimate is, the higher the risk of finding spurious local minima which lead to inaccurate or unphysical solutions.

There are currently a few methods that provide initial estimates for multi-component taylorgrams using information intrinsic to the distribution in question. Such a method could pave the way for more accurate, reproducible and robust deconvolution of the hydrodynamic radii and composition of taylorgrams generated by solute mixtures and make TDA a viable option for the characterization of solute mixtures and reversible aggregates in biopharmaceutical formulations without the need for separation. An example of such a method involved the use of the diffusion coefficients obtained from the cumulant method or from an equivalent log-normal distribution to constrain the range of possible diffusion coefficients for the CRLI approach[31].

In this paper, alternative methods for the determining the initial parameter estimates for multi-component taylorgrams are presented. These are obtained by differentiating and integrating the taylorgram and solving the resulting sets of linear equations. The estimates are then used as seeds for a least-squares fitting algorithm to obtain the individual diffusion coefficients and relative proportions of the components of a mixture. Furthermore, the closer the fitted parameters are to the initial estimates, the greater the degree of confidence in the accuracy of the results.

In addition, TDA requires that the concentrations of the eluent phase (run buffer) and the sample matrix phase (sample buffer) be matched. This, however, is not always achieved experimentally and usually results in a negative contribution to the taylorgrams which give biased results. By treating the mismatch in concentration as an independent contribution to the taylorgram, it is possible to isolate this contribution to enable the analysis of the solute. This fitting method is also presented in this paper.

The paper is arranged as follows. First, the least-squares fitting method as applied to a single solute component will be described and the expression for the Taylor-average hydrodynamic radius from applying this fit to a multi-component taylorgram without deconvolution is derived. Next, the methods of parameter estimation for multi-component mixtures will be presented as well as the method for mitigating the effects of mismatched buffer phases. These methods are then used to analyze a range of solute mixtures and aggregated samples. The results from the fits are compared to the expected values.

2. Theoretical Methods

2.1. The least-squares fitting method and the Taylor-average hydrodynamic radius

The theoretical concentration distribution C which arises at a time t when a solute made up of a single component type undergoes Taylor dispersion is given by:

$$C \propto C_0 \sqrt{\frac{t_r}{t}}\, e^{-\frac{u^2(t-t_r)^2}{4kt}} \qquad (1)$$

where C_0 is the initial solute concentration, t_r is known as the mean residence time, u is the mean flow speed of the carrier solution and k is the dispersion coefficient.

At large values of t, Equation 1 is approximated by a Gaussian distribution:

$$C = Ae^{-\frac{(t-t_r)^2}{2\sigma^2}} \qquad (2)$$

where A is the peak amplitude, the dispersion coefficient k is related to the standard deviation of the Gaussian σ by

$$k = \frac{u^2\sigma^2}{2t_r} \qquad (3)$$

Note that the standard deviations have units of time. The diffusion coefficient D and hydrodynamic radius R_h are related to k by:

$$D = \frac{{r_c}^2 u^2}{48k} \tag{4}$$

$$R_h = \frac{k_B T}{6\pi\eta D} \tag{5}$$

where r_c is the capillary radius, k_B is Boltzmann's constant, T is the temperature and η is the viscosity of the carrier solution.

The standard deviation σ can be determined by fitting Equation 2 to the taylorgram. Typically, a least-squares fitting method is employed and to ensure an accurate fit, the unknown parameters such as A, t_r and σ are initially estimated. The peak amplitude A is estimated as the maximum value of the taylorgram, the residence time t_r is estimated as the time at which this maximum occurs whilst σ is determined from the width of the taylorgram at half the maximum value. These initial parameter estimates serve as seeds from which the fitting algorithm can find the optimum solution.

Compared to the method for single-component taylorgrams, the estimation of the parameters is not as straightforward for multi-component mixtures because of the increased number of parameters to be estimated. However, in the following section, a method using data intrinsic to the taylorgram in question is presented. This relies on the differentiation and integration of the taylorgram to obtain relations between the parameters which can be solved using well-known mathematical techniques.

The fitted expression to taylorgrams in Equation 2 has three unknown parameters namely, the standard deviation σ, the residence time t_r and the amplitude A. To obtain the best fit Gaussian to a Taylorgram, these three parameters are varied until the sum of the residuals R^2 is minimized. For a mixture of solutes, the taylorgram takes on a different shape. Assuming, there are no interactions between the samples, the taylorgram can be assumed to be the sum of Gaussian distributions; one corresponding to each component of the mixture. Hence, the theoretical solution to the n-component dispersion problem is given by:

$$g = \sum_{i=1}^{n} A_i \,.\, e^{-0.5.\left(\frac{t-t_r}{\sigma_i}\right)^2} \tag{6}$$

If the single Gaussian distribution

$$A_F e^{-\frac{(t-t_r)^2}{2{\sigma_F}^2}} \tag{7}$$

is fitted to the multi-component taylorgram in Equation 6, the sum of residuals, R^2, is given by:

$$R^2 = \sum \left[A_F e^{-\frac{(t-t_r)^2}{2{\sigma_F}^2}} - \sum_{i=1}^{n} A_i \, e^{-\frac{(t-t_r)^2}{2{\sigma_i}^2}} \right]^2 \tag{8}$$

By converting this sum to an integral over time and solving for the minima, the following closed expressions are obtained for the fitted parameters:

$$\frac{A_F}{\sqrt{2}} - \sum_{i=1}^{n} \frac{A_i \sigma_i}{\sqrt{\sigma_i^2 + \sigma_F^2}} = 0 \tag{9}$$

$$\frac{A_F}{\sqrt{2}} - \sum_{i=1}^{n} \frac{2 A_i {\sigma_i}^3}{(\sigma_i^2 + \sigma_F^2)^{3/2}} = 0 \tag{10}$$

which are simultaneously satisfied by the parameters of the best-fit function to the taylorgram.

Eliminating A_F in the Equations 9 and 10 gives

$$\sum_{i=1}^{n} \frac{A_i \sigma_i (\sigma_F^2 - \sigma_i^2)}{(\sigma_i^2 + \sigma_F^2)^{3/2}} = 0 \tag{11}$$

This can be re-written in terms of hydrodynamic radii R_h^i which scales with σ_i^2 and mass-concentrations M_i which scales with $A_i\sigma_i$ to give:

$$\sum_{i=1}^{n} M_i \frac{\left(R_h^F - R_h^i\right)}{\left(R_h^F + R_h^i\right)^{3/2}} = 0 \tag{12}$$

where R_h^F is the Taylor average hydrodynamic radius obtained from the fitting method. As can be seen, obtaining an explicit expression for R_h^F is complicated. However, it may be observed that the inverse cube of the hydrodynamic radius is approximately weighted by the square of the mass concentration and hence for a mixture of solutes with equal mass-concentrations, the Taylor average hydrodynamic radius is biased toward the smaller components.

2.2. Parameter estimation for multi-component taylorgrams

In the long time limit of Taylor dispersion, the taylorgram obtained from a mixture of *n* non-interacting components can be approximated by:

$$g = \sum_{i=1}^{i=n} A_i \,.\, e^{-0.5.\left(\frac{t-t_r}{\sigma_i}\right)^2} \qquad (13)$$

where A_i and σ_i are the respective amplitudes and standard deviations for the *i*-th component, t_r is the residence time of the mixture at the observation point and *t* is the measurement time.

The residence time t_r is estimated as the time at which the maximum value of the taylorgram, *x* occurs. Likewise, the sum of the amplitudes A_i is estimated by *x*.

$$x = \sum_{i=1}^{i=n} A_i \qquad (14)$$

Figure 1 shows the second differential $\frac{d^2g}{dt^2}$ of a taylorgram with respect to time. This can be obtained by the difference method or by using the Savitzky-Golay filter[34].

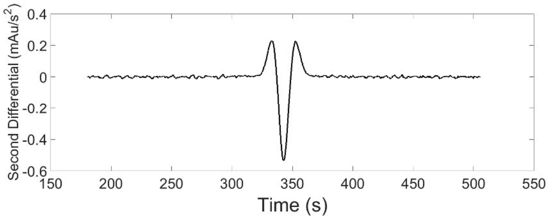


Figure 1. Second differential of a taylorgram.

Its value *y* at $t = t_r$ (or its maximum negative value) provides an estimate for the following sum:

$$y = -\frac{d^2g}{dt^2}\|_{t=t_r} = \sum_{i=1}^{i=n} \frac{A_i}{{\sigma_i}^2} \qquad (15)$$

Figure 2 shows the integral of a taylorgram over time.

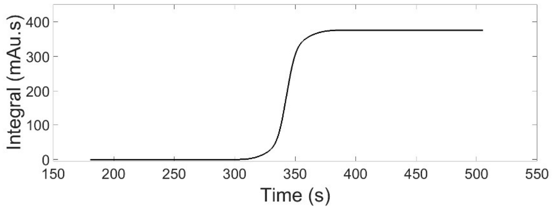


Figure 2. Integral of a taylorgram.

Its value *I* at $t = t_r$ provides an estimate *z* for the following sum:

$$z = \sqrt{\frac{2}{\pi}} I = u = \sum_{i=1}^{i=n} A_i \sigma_i \qquad (16)$$

Finally, Figure 3 shows the double integration of a taylorgram over time.

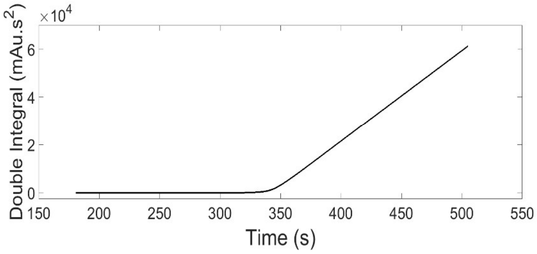


Figure 3. Double integral of a taylorgram.

Its value *u* at $t = t_r$ provides an estimate for the following sum:

$$u = \sum_{i=1}^{i=n} A_i \sigma_i^2 \qquad (17)$$

The four equations, Equations 14-17, provide simultaneous equations in the amplitudes A_i and standard deviations σ_i which can be solved for mixtures with differing numbers of components. In principle further differentials and integrations can be computed to establish further relations between these parameters but in this paper, the equations are limited to the four presented. The solutions to these equations can then be used as seed parameters for a least-squares fitting algorithm to deconvolve the mixture's taylorgram. It should be noted that even though these seed parameters are obtained by assuming the components have the same residence times, these are allowed to vary by the fitting algorithm. From the standard deviations of the deconvoluted taylorgrams, the hydrodynamic radii can be determined for the individual components. Furthermore, with knowledge of the extinction coefficient of each component, the proportion of each component in the mixture can be estimated by computing the area under each individual taylorgram.

There are three different cases (one with four sub-cases) to which the methods described above can be applied. These are:

1A) A two-component mixture of unrelated components e.g. two different solutes, a monomer and its aggregates.

1B) A two-component mixture where the size of one component is known e.g. a monomer of known hydrodynamic radius and its aggregates.

1C) A two- or three-component mixture where the components are related but the hydrodynamic radii and proportions of the components are not known e.g. a mixture of monomer, dimers and trimers.

1D) The individual hydrodynamic radii of all components in a two- three- or four-component mixture are known but their proportions within the mixture are not.

2) A three- and four- component mixture of unrelated components.

3) A mismatch in the concentrations of the sample matrix phase (sample buffer) and the eluent phase (run buffer) which results in a taylorgram which is the sum of two which differ in sign. The negative contribution comes from the lower concentration of the buffer in the injected sample. Typically, the positive contribution of the solute is dominant in the taylorgram. Such a scenario can arise as a result of evaporation, sample mishandling or when solute/buffer components change during storage.

The application of the method to each case will now be described in turn.

Case 1A: The two unrelated components have distributions with non-negligible amplitudes and unknown radii.

In this case, the four equations, Equations 14-17, reduce to:

$$x = A_1 + A_2$$

$$y = \frac{A_1}{{\sigma_1}^2} + \frac{A_2}{{\sigma_2}^2}$$

$$z = A_1\sigma_1 + A_2\sigma_2$$

$$u = A_1\sigma_1^2 + A_2\sigma_2^2 \tag{18}$$

Solving the four equations simultaneously gives:

$$\sigma_1 = 0.5\left(wk + \sqrt{(wk)^2 - 4k}\right)$$

$$\sigma_2 = \frac{k}{\sigma_1}$$

$$A_2 = \frac{z - \sigma_1 x}{\sigma_2 - \sigma_1}$$

$$A_1 = x - A_2 \tag{19}$$

where

$$w = \sqrt{\frac{yu - x^2}{ux - z^2}}$$

$$k = \frac{u}{zw - x} \tag{20}$$

Case 1B: The hydrodynamic radius of one of the two components is known *a priori*

An estimate of the standard deviation of the known component is obtained from its hydrodynamic radius by rearranging Equation 4. In doing so, one of the four unknown parameters has been determined *a priori* and any three of the four equations in Equation 18 may be used to obtain initial estimates for the remaining parameters. There are four such combinations of three equations which can be used.

Case 1C: Two or three components with known ratios between their hydrodynamic radii.

In this case, there is only one unknown standard deviation since the other one (or two in the case of three components) is related by a ratio known *a priori*. For the case of three components, there are three unknown amplitudes and one unknown standard deviation. Hence, with the corresponding four equations (cf. Equations 14 – 17, the four unknowns (A_1, A_2, A_3 and σ_1) can be solved for.

$$x = A_1 + A_2 + A_3$$

$$y = \frac{A_1}{{\sigma_1}^2} + \frac{A_2}{a{\sigma_1}^2} + \frac{A_3}{b{\sigma_1}^2}$$

$$z = A_1\sigma_1 + A_2\sqrt{a}\sigma_1 + A_3\sqrt{b}\sigma_1$$

$$u = A_1\sigma_1^2 + A_2 a\sigma_1^2 + A_3 b\sigma_1^2 \qquad (21)$$

where *a* and *b* are the known diffusion coefficient ratios of the second and third components to the first component respectively.

These simultaneous equations can be reduced to a single quartic equation in σ_1 which can be solved using the traditional methods. From this solution, estimates can then be made for A_1, A_2 and A_3 by substitution. A similar solution can be obtained for a mixture of two components by solving any three of the four corresponding equations.

Case 1D: Two, three or four components with known hydrodynamic radii.

In this case, since the hydrodynamic radii are known, the only unknowns are the amplitudes. Hence, for the general case, Equations 14 – 17 can be solved for the *n* (< 5) unknowns (A_n) by reducing them to a matrix equation. The following is the equation applicable to *n*=4 components.

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ {}^{1}/_{\sigma_1^2} & {}^{1}/_{\sigma_2^2} & {}^{1}/_{\sigma_3^2} & {}^{1}/_{\sigma_4^2} \\ \sigma_1 & \sigma_2 & \sigma_3 & \sigma_4 \\ \sigma_1^2 & \sigma_2^2 & \sigma_3^2 & \sigma_4^2 \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \\ A_3 \\ A_4 \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \\ u \end{bmatrix} \qquad (22)$$

where σ_1, σ_2, σ_3 and σ_4 are known *a priori* (determined from Equation (4). This can be solved by well-known matrix methods to obtain initial estimates for A_1, A_2, A_3 and A_4. For two- (and three-) component mixtures a 2 by 2 (and 3 by 3) matrix constructed from any two (and three) of the four equations may be used to obtain initial estimates.

Case 2: A three- and four- component mixture of unrelated components.

In these cases, where there is no *a priori* information about the composition of the mixtures, there are six and eight unknowns respectively which cannot be solved for explicitly using the four equations to obtain the initial parameter estimates. However, by assuming the taylorgram as an ensemble of two taylograms initially, it is possible to successively solve the four equations to obtain approximations to the initial estimates for all the parameters. This approximate method is outlined below.

(i) Assuming, the taylorgram is a mixture of two taylorgrams, use the method of Case 1A to solve for the two parameter pairs (A_1', σ_1') and (A_2', σ_2'). Let the fitted models from these parameters be g_1' and g_2' respectively.

(ii) For a three-component mixture, create another data set by subtracting the model with the smaller area from the taylorgram data i.e. if $\sum g_1'$ < $\sum g_2'$, create g_3'= taylorgram data – g_1'. For a four component mixtures, create two data sets: g_3'= taylorgram data – g_1' and g_4'= taylorgram data – g_2'.

(iii) For the three-component mixture, apply the method of Case 1A to data set g_3' and solve for two more parameter pairs (A_3', σ_3') and (A_4', σ_4'). The seed parameters for the three-component fit are then: (A_2', σ_2'), (A_3', σ_3') and (A_4', σ_4'). Likewise, for the four-component mixture, apply the method of Case 1A to data sets g_3' and g_4' to obtain four parameter pairs which would serve as seed parameters for the four-component fit. Note that there are other alternatives to deconvoluting the talyorgrams in this way at this stage which may be tailored accordingly to the type of size distribution expected.

The method described may be adjusted accordingly if one, two or three of the hydrodynamic radii are known *a priori*.

Case 3: The taylorgram is a sum of two which differ in sign in which the positive taylorgram is dominant e.g. due to a mismatch of concentrations between the sample and run buffers.

The effect of mismatches between the solute and run buffers on the resulting taylorgrams are illustrated in Figures 4 and 5. In both cases the concentration of solute buffer is lower than that of the run buffer, which manifests as a negative contribution to the concentration profile. In Figure 4, the hydrodynamic radius of the buffer molecules is smaller than for the solute and vice versa in Figure 5.

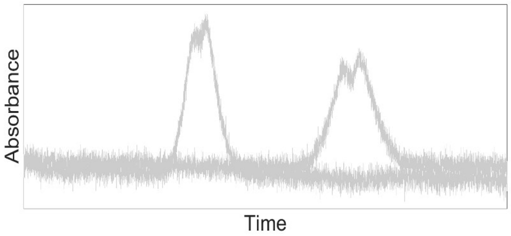


Figure 4. IgG in PBS with mismatched buffer phases (buffer molecules are smaller than the solute molecules).

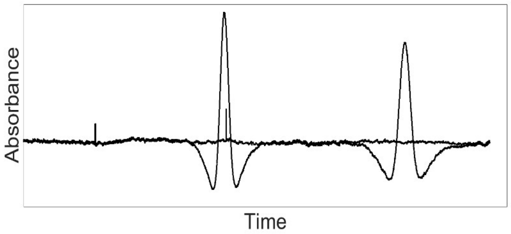


Figure 5. Caffeine in BSA with mismatched buffer phases (buffer molecules are larger than the solute molecules).

In this case, since a mixture of two unrelated components is being considered, the application of a two-component fit to the taylorgram is required so that the standard deviation of the sample can be extracted. As before, there are four unknown parameters, A_1, A_2, σ_1 and σ_2, and thus four equations are required to obtain initial estimates. These may be obtained as described earlier but in cases where the solute is the dominant component in the mixture, a simpler method may be implemented. The method for the case illustrated in Figure 5 is described as follows.

The first equation can be obtained from the value of the absorbance x at the dip in the trace in Fig. 4 i.e. at $t \sim t_r$.

$$x = A_1 - A_2 \tag{23}$$

The second and third equations are obtained from the maximum absolute value *y* of the differential of the profile. For a single component fit, this occurs at a time *t'* given by

$$t' = t_r \pm \sigma_1 \tag{24}$$

where the taylorgram has an absolute value y given by

$$y = \frac{A_1}{\sigma_1}\exp(-0.5) \tag{25}$$

where the subscript 1 corresponds to the solute component.

In Figure 6 is shown the second differential of the taylorgram. At the dip at the center of the taylorgrams shown in Figure 4, the second differential is non-zero and positive as shown in Figure 6 (whereas, it is negative for a single component trace). The value *z* at this time is given by:

$$z = \frac{A_1}{{\sigma_1}^2} - \frac{A_2}{{\sigma_2}^2} \tag{26}$$

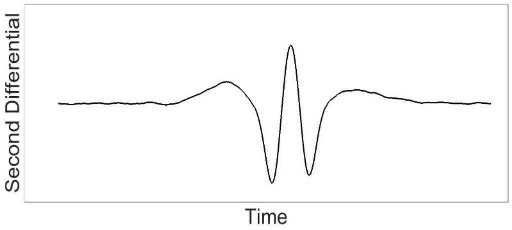


Figure 6. Second differential of a taylorgram with mismatched buffers (buffer molecules are smaller than the solute molecules).

Since both A_1 and σ_1 are exclusively determinable from Equations 24 and 25, it is straightforward to obtain A_2 and σ_2 from Equations 23 and 26 by direct substitution. Similar expressions can be obtained for the case where the hydrodynamic radius of the buffer is larger than for the solute.

3. Experimental Section

TDA measurements were undertaken with the Viscosizer 200 or Viscosizer TD instruments (Malvern Instruments Ltd., Worcestershire, UK) fitted with a standard uncoated capillary (ID 75 µm, OD 360 µm, Malvern Instruments Ltd., Worcestershire, UK) using a 214 nm wavelength filter. Delivery of narrow solute plugs was achieved by pressure-driven injection at 50 mbar for 12 s. To ensure accuracy and because of the increased signal strength at the first detection window, single-detection TDA at the first window will be used to compute the hydrodynamic radii and relative proportions of the solutes in the mixtures. The decreased signal strength at the second detection window introduces errors due to the influence of baseline noise particularly on wide and low-amplitude taylorgrams.

The two-, three- and four- component mixtures of solutes were prepared using the highest quality reagents (from Sigma Aldrich, Suffolk, UK unless otherwise stated) as follows:

3.1. Two-component mixtures: To test Cases 1A and 1B, three different mixtures were prepared from samples of 0.25 mg/ml caffeine (R_h ~ 0.3 nm[35]), 20mg/mL Bovine Serum Albumin (BSA, R_h ~ 3.8 nm[36, 37]), 20 mg/mL Myoglobin (R_h ~ 2.1 nm[38]) and 20 mg/ml Immunoglobin G (IgG, R_h ~ 5.8 nm[39]) prepared in Phosphate Buffered Saline (PBS pH 7.4) in equal volumes. The individual components were analyzed and the hydrodynamic radii were found to agree with the literature values. The mixtures comprised a) Caffeine and BSA, b) BSA and Myoglobin, c) Myoglobin and IgG. Three replicates of each mixture were analyzed at a run pressure of about 100-140 mbar at 20^oC.

3.2. Three-component mixtures: To test Case 1C and 2, a three-component mixture was prepared from samples of 1 mg/mL caffeine and two Nanospheres^{TM} 3000 Series Size Standards (R_h ~ 30 nm and 100 nm; Fisher Scientific, Leicestershire, UK) dissolved in 0.01M NaCl in equal volumes. Three replicates of the mixture were analyzed at a run pressure of 36 mbar at 25^oC.

3.3. Four-component mixtures: To test case 1D and 2, a four-component mixture was prepared from samples of 1 mg/mL caffeine (R_h ~ 0.3 nm) and three Nanospheres^{TM} 3000 Series Size Standards (R_h ~ 30 nm, 100 nm and 200 nm) dissolved in 0.01M NaCl in equal volumes. Three replicates of the mixture were analyzed at a run pressure of 18 mbar at 25^oC.

3.4. Samples with a mismatch between the solute and run buffers: To test Case 3, it was necessary to contrive a mismatch between the solute and run buffer. Samples of 1 mg/mL and 2 mg/mL Insulin (R_h ~ 2 nm, 20% (v/v) acetic acid, pH 1.9) were prepared in their respective buffer solutions and left exposed to air for 3-5 hours. The hydrodynamic radius of Insulin was determined *a priori* using the Viscosizer instrument. In both of these experiments, the buffer molecules are smaller than the solute molecules. Another sample comprising 1 mg/mL caffeine (0.3 nm) dissolved in a 5 mg/mL BSA/PBS solution was also prepared to illustrate the case where the buffer molecules are larger than the solute molecules. Three replicates of each sample were analyzed at a run pressure of 140 mbar at 20^oC.

3.5. Aggregated samples: Samples of 20 mg/mL BSA (R_h ~ 3.8 nm, PBS pH 7.4) and 3 mg/mL BSA (R_h ~ 5 nm, PBS/arginine) were heat-stressed at 65^oC to varying degrees to induce aggregation. Mixtures of stressed and non-stressed samples were combined to generate samples with 3, 4, 6, 9, 50, 75 and 100 % (v/v) aggregated samples. The hydrodynamic radii of unstressed BSA in the two different buffer phases were determined *a priori* using the Viscosizer instrument. These are different due to conformational changes of BSA in the different ionic environments. At least five replicates of each mixture was analyzed at a run pressure of 140 mbar at 20^oC. The two component model was applied to the data to extract the proportion of aggregated material in each sample.

4. Results and Discussion

4.1. Cases 1A and 1B: Two-component mixtures with unknown radii and where one of the radii is known a priori.

Figures 7-9 show three examples of the fits obtained to the two-component mixtures. The estimated hydrodynamic radii from the fits are presented in Table 1 for Cases 1A (where both radii are unknown) and 1B (where one of the radii is known). Note that for Case 1B, the radius of each component has been obtained by fixing the other component of the mixture and vice versa. Also shown for Case 1A are the seed hydrodynamic radii which are the values of the radii obtained from the initial estimates before the fits are applied. As can be seen, for all three mixtures, the estimated hydrodynamic radii obtained are in reasonable agreement with the expected values.

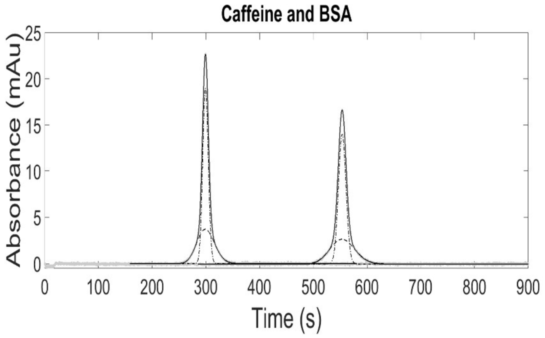


Figure 7. Fits to taylorgrams from a mixture of caffeine and BSA (Grey: taylorgram, Solid: full fit, Dot-dashed: smaller component, Dashed: larger component)

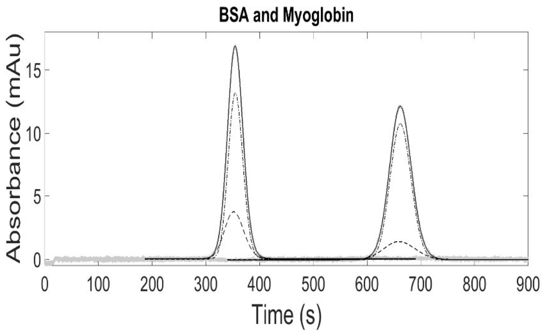


Figure 8. Fits to taylorgrams from a mixture of BSA and Myoglobin (Grey: taylorgram, Solid: full fit, Dot-dashed: smaller component, Dashed: larger component)

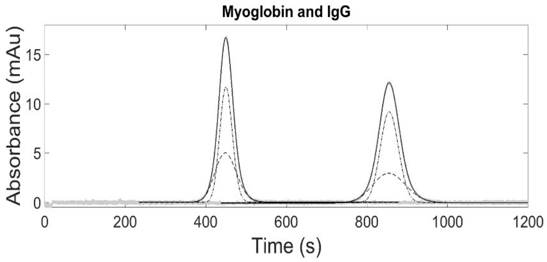


Figure 9. Fits to taylorgrams from a mixture of Myoglobin and IgG (Grey: taylorgram, Solid: full fit, Dot-dashed: smaller component, Dashed: larger component)

Table 1. Estimated hydrodynamic radii for two-component mixtures.

Mixture	Solute	Nom. R_h (nm)	Seed R_h (nm) 1A	Est. R_h (nm) 1A	Est. R_h (nm) 1B
Caff. + BSA	Caff.	0.3	0.38 +/-0.01	0.340 +/-0.002	0.346 +/- 0.001
	BSA	3.8	3.8 +/- 0.3	3.62 +/- 0.06	3.44 +/- 0.02
BSA + Myog.	BSA	3.8	3.05 +/-0.03	3.65 +/- 0.02	3.76 +/- 0.07
	Myog.	2.1	1.79 +/-0.06	2.08 +/- 0.02	2.11 +/- 0.11
Myog. + IgG	Myog.	2.1	2.3 +/- 0.4	2.08 +/- 0.08	2.07 +/- 0.01
	IgG	5.8	7 +/- 6	5.8 +/- 0.5	5.9 +/- 0.1

4.2. Case 1C: Three-component mixtures with the ratios of the radii known *a priori*.

Figure 10 shows the fits to taylorgrams obtained from a three-component mixture and the results for the hydrodynamic radii are presented in Table 3. These were obtained with *a priori* knowledge of the ratios of the hydrodynamic radii. Also shown in Table 3 are the seed values for the hydrodynamic radii obtained before the fits were applied. As can be seen, the estimated values for the hydrodynamic radii are in reasonable agreement with the nominal values.

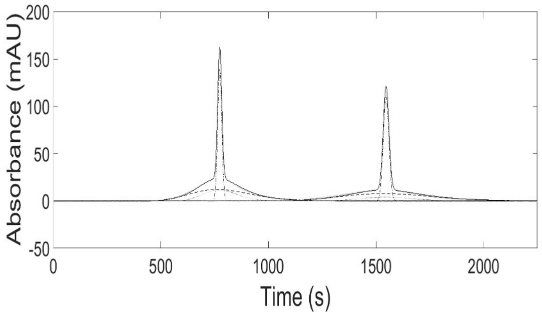


Figure 10. Fits to taylorgrams from three-component mixtures (Grey: taylorgram, Solid: full fit, Dot-dashed: Caffeine, Dashed: 30nm nanospheres, Dotted: 100nm nanospheres)

Table 2. Estimated hydrodynamic radii for three-component mixtures when the size ratios are known *a priori*.

Caffeine + 30 nm + 100 nm	Nominal R_h (nm)	Seed R_h (nm)	Est. R_h (nm)
Caffeine	0.3	0.37 +/- 0.02	0.36 +/- 0.02
30 nm nanospheres	30	37 +/- 3	27 +/- 2
100 nm nanospheres	100	125 +/- 16	102 +/- 6

4.3. Case 1D: Four-component mixtures with the four radii known *a priori.*

Figure 11 shows fits to taylorgrams obtained from the mixture of four components and the results for the areas beneath the taylorgrams are presented in Table 4. These were obtained with a priori knowledge of the four hydrodynamic radii. Also shown are the expected areas which were determined by analyzing each component independently. The seed values obtained before the fits were applied are also presented. As can be seen, the estimated areas are in reasonable agreement with the expected values.

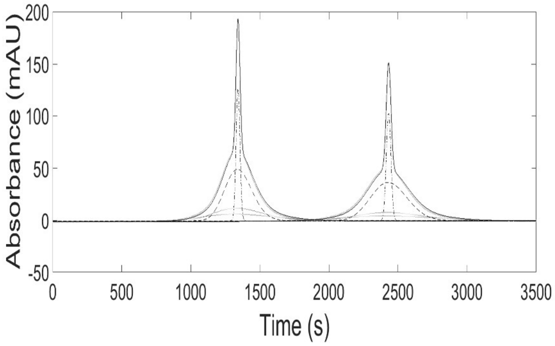


Figure 11. Fits to taylorgrams from four-component mixtures (Grey: taylorgram, Solid: full fit, Dot-dashed: Caffeine, Dashed: 30nm nanospheres, Dotted: 100nm (larger of the two profiles) and 200nm nanospheres)

Table 3. Estimated taylorgram areas for four-component mixtures with the radii known *a priori.*

Caffeine + 30 nm + 100 nm	Nominal R_h (nm)	Seed R_h (nm)	Est. R_h (nm)
Caffeine	0.3	0.427 +/- 0.004	0.35 +/- 0.01
30 nm nanospheres	30	38.5 +/- 0.3	32.3 +/- 0.9
100 nm nanospheres	100	133 +/- 3	107 +/- 3

4.4. Case 2: Three- and four-component mixtures of solutes with unknown radii *a priori*.

In Tables 4 and 5 are shown the results from the application of the fitting method for Case 2 (where the radii are unknown *a priori*) to the three- and four-component mixtures respectively. Note that since the four-component mixtures which were run at a pressure of 18 mbar, the hydrodynamic radius expected for caffeine is ~ 0.5 nm when Taylor's solution to the dispersion equation is used. The correct value can be obtained by using the Taylor-Aris solution which accounts for the additional contribution to the dispersion from longitudinal diffusion should be used.

As can be seen, the estimated radii are in agreement with the expected values. A closer look at the seed parameters reveals that while these were reasonably approximated for the three-component mixture, this was only partially the case for the four-component mixture where three of the radii (caffeine, 30 nm and 200 nm nanospheres) and one area (caffeine) are reasonably approximated.

As mentioned earlier, unlike for Cases 1A-1D, this is an approximate method for determining the seed parameters. However, its usefulness may be demonstrated in two ways. The first is that since the seeds are obtained directly from operations on the taylorgram data, they are physically feasible. This is because the initial seeds obtained this way already provide an approximate fit to the data although they may be inaccurate. The second benefit of this method is its potential applicability to taylorgrams with an arbitrary number of components such as polydisperse samples. By successively applying the method described above to such taylorgrams, it may be possible to obtain a measure of the sample's polydispersity from the range of seeds obtained for the parameters. This is evident from seeds obtained for the four-component mixture where the seed radii span the range of the expected radii. The application of this method to the quantification of polydispersity will be the subject of future work.

Table 4. Estimated hydrodynamic radii for three-component mixtures with no *a priori* information.

Caffeine + 30 nm + 100 nm + 200 nm	**Nominal Area (mAUs)**	**Seed Area (mAUs)**	**Est. Area (mAUs)**
Caffeine	4000	3521 +/- 75	4627 +/- 148
30 nm nanospheres	12000	19836 +/- 1058	12982 +/- 756
100 nm nanospheres	6500	9202 +/- 166	6057 +/-79
200 nm nanospheres	3500	3063 +/- 352	3769 +/- 535

Table 5. Estimated hydrodynamic radii and taylorgram areas for four-component mixtures with no *a priori* information.

Caffeine + 30 nm + 100 nm + 200 nm	**Nom. Area (mAUs)**	**Seed Area (mAUs)**	**Est. Area (mAUs)**	**Seed R_h (nm)**	**Est. R_h (nm)**
Caffeine (~ 0.5 nm)	4000	4947 +/- 95	4960 +/- 92	0.52 +/- 0.01	0.51 +/- 0.01
30 nm nano-spheres	12000	8103 +/- 111	13032 +/- 244	27.4 +/- 0.4	31 +/- 0.3
100 nm nano-spheres	6500	7866 +/- 386	6853 +/-800	47.3 +/- 0.3	99 +/- 12
200 nm nano-spheres	3500	7342 +/- 235	3145 +/- 965	171 +/- 2	223 +/- 16

4.5. Case 3: Solutes with a mismatch between the solute and run buffers

Figures 12-14 show fits to three taylorgrams showing a mismatch between the sample buffer and run buffer phases. The values for the hydrodynamic radii obtained from the fits are presented in Table 6. These are in reasonable agreement with the nominal values. Note that the radii of the buffer molecules are estimated from the standard deviations of the negative taylorgrams.

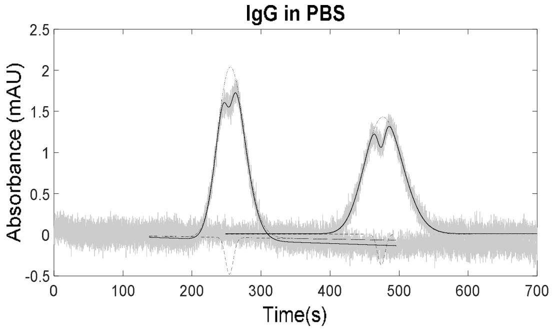


Figure 12. Fits to IgG in PBS with a mismatch in the buffer phases (Grey: taylorgram, Solid: full fit, Dot-dashed: solute, Dashed: mismatched buffer)

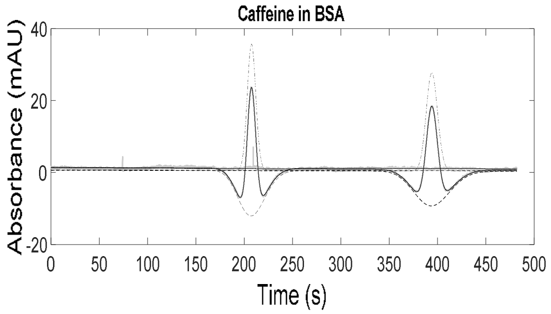


Figure 13. Fits to caffeine in BSA with a mismatch in the buffer phases (Grey: taylorgram, Solid: full fit, Dot-dashed: solute, Dashed: mismatched buffer)

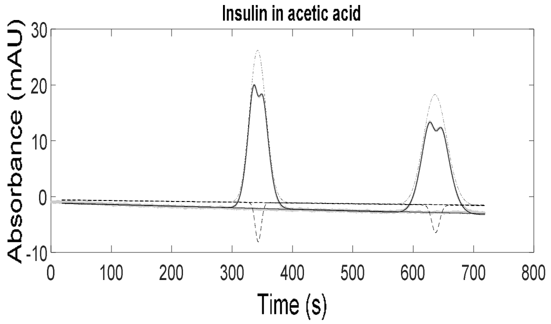


Figure 14. Fits to Insulin in acetic acid with a mismatch in the buffer phases (Grey: taylorgram, Solid: full fit, Dot-dashed: solute, Dashed: mismatched buffer)

Table 6. Estimated hydrodynamic radii from taylorgrams with mismatched buffer phases.

Sample	Nominal % aggregation	Estimated % aggregation
BSA in PBS	3	3.5 +/- 0.2
	4	3.9 +/- 0.3
	6	6.8 +/- 0.4
	9	7.7 +/- 0.7
BSA in Arginine	50	57.0 +/- 0.3
	75	81.8 +/- 0.2
	100	97 +/- 6

4.6. Aggregated samples

Figures 15 and 16 show examples of two-component fits to taylorgrams obtained from aggregated BSA samples. It will be noted from the taylorgrams that the bulk of the aggregates have similar residence times to the monomer indicating that they are not too large in size. The estimated levels of aggregation computed from the areas beneath the individual taylorgrams are compared to the expected values in Table 7 and can be seen to be in reasonable agreement. The slight discrepancies in the values obtained may be attributed to external factors such as additional aggregation occurring between the time the samples were prepared and when they were analyzed.

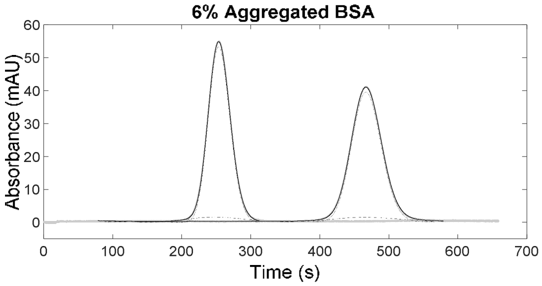


Figure 15. Two-component fits to taylorgrams from a 6% aggregated BSA sample (Grey: taylorgram, Solid: full fit, Dot-dashed: BSA monomer, Dashed: BSA aggregates)

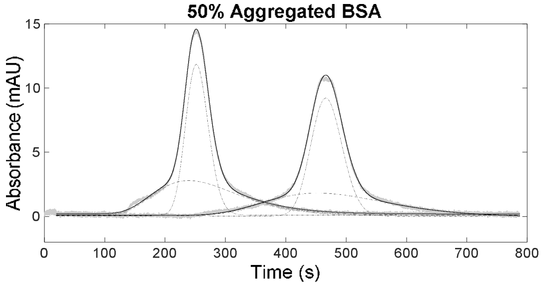


Figure 16. Two-component fits to taylorgrams from a 50% aggregated BSA sample (Grey: taylorgram, Solid: full fit, Dot-dashed: BSA monomer, Dashed: BSA aggregates)

Table 7. Estimated levels of aggregation for aggregated BSA samples

Sample	Solute and buffer	Nominal R_h (nm)	Estimated R_h (nm)
IgG in PBS	IgG	5.8	5.9 +/- 0.2
	PBS mismatch		0.21 +/- 0.04
Caffeine in BSA	Caffeine	0.3	0.339 +/- 0.003
	BSA mismatch	3.8	3.76 +/- 0.03
Insulin in acetic acid	Insulin	2.1	2.11 +/- 0.03
	Acetic acid mismatch		0.25 +/- 0.01

5. Conclusions

In this work, a method has been described for the determination of the initial estimates of parameters for the constituent taylorgrams of a multi-component mixture. These initial estimates can then be used in a least-squares fitting algorithm to obtain optimum solutions for the constituent taylorgrams from which the hydrodynamic radii and relative proportions of the components may be calculated. The method was applied successfully to two-, three- and four-component mixtures as well as aggregated samples. These methods could feasibly be applied to the determination of the onset of and quantification of polydispersity and aggregation in a sample and this will be the focus of further work.

ACKNOWLEDGEMENTS

The authors would like to thank Professor David Goodall (Paraytec and University of York) and Professor Robert Forbes (University of Bradford) for fruitful discussions.

AUTHOR INFORMATION

Corresponding Author

* Email: seyi.latunde-dada@malvern.com.

Author Contributions

‡These authors contributed equally.

REFERENCES

1. Taylor, G. *Proc. Roy. Soc. Lond. A* **1953**, *219*, 186.
2. Aris, R. Proc. Roy. Soc. Lond. A, **1956**, 235, 67.
3. Giddings, J. C.; Seager, S. L. *J.Chem.Phys.* **1960**, *33*, 1579-1580.
4. Grushka, E.; Kikta, E. J. *J. Phys. Chem.* **1974**, *78*, 2297-2301.
5. Ouano, A. C. *Ind. Eng. Chem. Fundam.* **1972**, *11*, 268-271.
6. Pratt, K. C.; Wakeham, W. A., *Proc. R. Soc., London A* **1974**, 393-406.
7. Bello, M. S.; Rezzonico, R.; Righetti, R. G. *Science* **1994**, *266*, 773-776.
8. Mes, E. P. C.; Kok, W. Th.; Poppe, H.; Tijssen, R. *J. Polym. Sci., Part B: Polym. Phys.* **1999**, *37*, 593-603.
9. Hulse, W. L.; Forbes, R. *Int. J. Pharm.* **2011**, *416*, 394.
10. Hulse, W. L.; Forbes, R. *Int. J. Pharm.* **2011**, *411*, 64.
11. Hawe, A.; Hulse, W. L.; Jiskoot, R.; Forbes, R. *Pharm. Res.* **2011**, *28*, 2302.
12. Ostergaard, J.; Hensen, J.; *Anal. Chem.* **2009**, *81*, 8644.
13. Sharma, U.; Gleason, N. J.; Carbeck, J. D. *Anal. Chem.*, **2005**, *77*, 3.
14. Niesner, R.; Heintz, A.; *J. Chem. Eng. Data* **2000**, *45*, 1121.
15. Cottet, H.; Martin, M.; Papillaud, A.; Souaid, E.; Collet, H.; Commeyras, A.; *Biomacromolecules* **2007**, *8*, 3235.
16. Cottet, H.; Biron J. P.; Cipelletti, L.; Matmour, M.; Martin, M. *Anal. Chem.* **2010** *82*, 1793.
17. Cottet, H.; Biron, J. P.; Martin, M. *Anal. Chem.* **2007** *79(23)*, 9066-9073.
18. Le Saux, T.; Cottet, H. *Anal. Chem.* **2008,** *80*, 1829.
19. D'Orlye, F.; Varenne, A.; Gareil, P. *J. Chromatogr. A* **2008**, *1204*, 226.
20. Franzen, U.; Vermehren, C.; Jensen, H.; Ostergaard, *J. Electrophoresis* **2011**, *32*, 738.
21. Boyle, W. A.; Buchholz, R. F.; Neal, J. A.; McCarthy, J. L. *J. Appl. Polym. Sci.* **1991**, *42*, 1969.
22. Quinn, J. G. *Anal. Biochem.* **2012,** *421*, 391-400.
23. Quinn, J. G. *Anal. Biochem.* **2012,** *421*, 401-410.
24. Chapman, A. J. S.; Goodall, D. *Chromatogr. Today* **2008**, *1*, 22-24.
25. Umecky, T.; Kuga, T.; Funazukuri, T. *J.Chem.Eng.Data* **2006**, *51*, 1705-1710.
26. Kelly, B.; Leaist D. G. *Phys. Chem. Chem. Phys.,* **2004,** *6*, 5523-5530.
27. Alizadeh, A.; De Castro, C. A.; Wakeham, W. A. *Int. J. Thermophys.* **1980,** *1*, 243-284.
28. Barooah, A.; Chen, S. H. *J. Polym. Sci., Part B: Polym. Phys.* **1985**, *23*, 2457-2468.
29. Chamieh, J.; Cottet, H. *J. Chromatogr. A* **2012,** *1241,* 123-127.
30. Cipelletti, L.; Biron, J.-P.; Martin, M.; Cottet, H. *Anal. Chem.* **2014**, *86*, 6471–6478.
31. Cipelletti, L.; Biron, J.-P.; Martin, M.; Cottet, H. *Anal. Chem.* **2015**, *87*, 8489–8496.
32. Price, W.E. *J. Chem. Soc. Faraday Trans.* **1988**, *1*, 2431–2439.
33. Steuer, W.; Jost, K.; Halasz, I. *Chromatographia* **1985**, *20*, 13–19.
34. Savitzky, A.; Golay, M. J. E. *Anal. Chem.* **1964**, *36*, 1627-1639.
35. Leaist, D.; Hui, L. *J. Phys. Chem.* 94 **1990**, *94,* 8741-8744.
36. Jachimska, B.; Wasilweska, M.; Adamczyk, Z. *Langumuir* **2008**, *24*, 6866-6872.
37. Brownsey, G. J.; Noel, T. R.; Parker, R.; Ring, S. R. *Biophy. J.* **2003** *85* 3943-3950.
38. Zaia, J.; Annan, R. S.; Biemann, K. *Rapid Commun. Mass Spectrom.* **1992** *6* 32-36.
39. Zhao, H.; Graf, O.; Milovic, N.; Luan, X.; Bluemel, M.; Smolny, M. *J. Pharm. Sci.* **2010** *99* 2279-2294.