The gamma distribution model for pulsed-field gradient NMR studies of molecular-weight distributions of polymers

Magnus Röding [a,*], Diana Bernin [b], Jenny Jonasson [a], Aila Särkkä [a], Daniel Topgaard [c], Mats Rudemo [a], Magnus Nydén [b,d]

[a] Department of Mathematical Statistics, Chalmers University of Technology and Gothenburg University, Gothenburg, Sweden
[b] Applied Surface Chemistry, Department of Chemical and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden
[c] Division of Physical Chemistry, Center for Chemistry and Chemical Engineering, Lund University, Sweden
[d] Ian Wark Research Institute, University of South Australia, Mawson Lakes Campus, Adelaide, South Australia 5095, Australia

ARTICLE INFO

ABSTRACT

Self-diffusion in polymer solutions studied with pulsed-field gradient nuclear magnetic resonance (PFG NMR) is typically based either on a single self-diffusion coefficient, or a log-normal distribution of self-diffusion coefficients, or in some cases mixtures of these. Experimental data on polyethylene glycol (PEG) solutions and simulations were used to compare a model based on a gamma distribution of self-diffusion coefficients to more established models such as the single exponential, the stretched exponential, and the log-normal distribution model with regard to performance and consistency. Even though the gamma distribution is very similar to the log-normal distribution, its NMR signal attenuation can be written in a closed form and therefore opens up for increased computational speed. Estimates of the mean self-diffusion coefficient, the spread, and the polydispersity index that were obtained using the gamma model were in excellent agreement with estimates obtained using the log-normal model. Furthermore, we demonstrate that the gamma distribution is by far superior to the log-normal, and comparable to the two other models, in terms of computational speed. This effect is particularly striking for multi-component signal attenuation. Additionally, the gamma distribution as well as the log-normal distribution incorporates explicitly a physically plausible model for polydispersity and spread, in contrast to the single exponential and the stretched exponential. Therefore, the gamma distribution model should be preferred in many experimental situations.

1. Introduction

Pulsed-field gradient nuclear magnetic resonance (PFG NMR) is a powerful method to evaluate translational motion such as diffusion or flow [1,2]. If the sample studied is sufficiently monodisperse, the (mean) self-diffusion coefficient can be obtained by fitting a single exponential function to the observed signal attenuation [3]. Also multi-component systems can easily be studied because of their chemical shift resolution. However, when the self-diffusion of for example a polymer is characterized by a molecular weight distribution and a corresponding distribution of self-diffusion coefficients, we obtain a more complex signal attenuation [4–7]. The probability distribution of self-diffusion coefficients may be extracted by using an inverse Laplace transform, typically using the CONTIN framework, where no specific functional form or shape assumptions are imposed, but the solution is regularized for smoothness [8–11]. Even though being notoriously difficult and very noise-sensitive,

this approach has been widely applied in various techniques, e.g. NMR relaxation and diffusion measurements, and dynamic light scattering [12,13]. An alternative approach is to assume a specific but flexible functional form, using e.g. the so-called stretched exponential, cumulant expansions, or a log-normal distribution model [14,15,5]. Many functional forms yield more or less identical fits and results in terms of the first two moments of a distribution (mean and variance) [16]. Thus, it is possible to choose the specific functional form based on simplicity and computational convenience.

In this paper, we propose to use a model based on a gamma distribution of self-diffusion coefficients. We discuss and compare different models for evaluating the NMR signal attenuation obtained from a PFG NMR experiment of polymers in water. In particular, we compare the computational speed and performance of the gamma model to the single exponential, the stretched exponential, and the log-normal models, all of which are already accepted and spread within the community. Although the gamma distribution has been mentioned in the literature as a model for a distribution of self-diffusion coefficients [15], its performance and suitability for

* Corresponding author.
E-mail address: roding@chalmers.se (M. Röding).

the analysis of PFG NMR data appears not to have been thoroughly evaluated. To do this is the purpose of this work.

We first describe the basic theory and put the gamma distribution model in context. Then, we evaluate the single exponential, the stretched exponential, the log-normal, and the gamma model on experimental data sets of two polyethylene glycol (PEG) solutions with different polydispersity and also on several simulated data sets. The models are compared in terms of estimated mean self-diffusion coefficient, spread and polydispersity index. The performance of the gamma distribution model is compared to all other models, and in particular to its most obvious competitor, the log-normal distribution model.

2. Theory

For a single self-diffusion coefficient, Stejskal and Tanner [17] showed that the echo decay is exponential (the so called *single exponential model*),

$$I(k) = I_0 \exp(-k\langle D \rangle), \tag{1}$$

where I_0 is the signal intensity before decay (for $k = 0$), $\langle D \rangle$ is the mean self-diffusion coefficient (actually, it is the only one, but we stick to the $\langle D \rangle$ notation for consistency), and

$$k = (\gamma g \delta)^2 \left(\Delta - \frac{\delta}{3} \right). \tag{2}$$

Here, γ is the proton magnetogyric ratio ($\gamma = 2.6752 \times 10^8$ rad T^{-1} s^{-1}), g is the gradient strength, δ is the gradient pulse duration, and Δ is the time lapse between the leading edges of the gradient pulses. In a plot of $\log I(k)$ vs. k, a single self-diffusion coefficient will manifest itself by a linear decay (straight line). Generally, the echo attenuation shows non-linear behavior for molecules with a large degree of polydispersity. The obvious interpretation is that a single self-diffusion coefficient cannot accurately describe the features of the system. Perhaps the simplest way of dealing with this problem is to use the *stretched exponential* model,

$$I(k) = I_0 \exp(-(kD_{app})^\beta), \tag{3}$$

a phenomenological relationship which is able to express polydispersity to some extent through the 'stretch' parameter β [18,19], and some attempt have been made to interpret this in terms of polydispersity [20]. However, the relation between spread and the value of the beta parameter is complicated [20]. However, the stretched exponential model does not correspond to an actual *distribution* of self-diffusion coefficients. Accordingly, there is no expression relating D_{app}, the *apparent* self-diffusion coefficient, to the true, mean self-diffusion coefficient $\langle D \rangle$; however, as was pointed out by Callaghan (personal communication) it can be shown that

$$\left\langle \frac{1}{D} \right\rangle = \frac{\frac{1}{\beta} \Gamma\left(\frac{1}{\beta} \right)}{D_{app}}, \tag{4}$$

where Γ is the gamma function. However, since $\langle 1/D \rangle^{-1} \neq \langle D \rangle$, any estimate of the mean self-diffusion coefficient will be weighted toward the slowly diffusing molecules. Therefore, a more physically adequate approach is to start by assuming that the self-diffusion coefficient D follows a probability distribution $P(D)$, which yields that the attenuation is an integral (weighted sum) over different exponential decays,

$$I(k) = I_0 \int_0^\infty P(D) \exp(-kD) dD. \tag{5}$$

In general, the functional form of $P(D)$ is unknown [6]. However, a very common assumption is that the self-diffusion coefficients are log-normally distributed [21]. The log-normal distribution has probability density

$$P_L(D; \mu, \sigma_L) = \frac{1}{D\sqrt{2\pi\sigma_L^2}} \exp\left(-\frac{(\log D - \mu)^2}{2\sigma_L^2} \right), \tag{6}$$

with $\mu = \log(\langle D \rangle) - \sigma_L^2/2$ where $\langle D \rangle$ is the mean of D, and $\sigma_L^2 = \log(1 + CV^2)$. The spread, or coefficient of variation (CV), is defined as CV = (standard deviation/mean) $\times$ 100%. However, the log-normal distribution does not yield an analytically tractable integral in Eq. (5). We suggest an alternative for the distribution $P(D)$ that does yield a tractable integral, namely the gamma distribution,

$$P_G(D; \kappa, \theta) = D^{\kappa-1} \frac{\exp(-D/\theta)}{\Gamma(\kappa)\theta^\kappa}, \tag{7}$$

where Γ is the gamma function, κ is the 'shape' parameter, and θ is the 'scale' parameter. Replacing $P(D)$ in Eq. (5) with P_G gives the echo decay model

$$I(k) = I_0(1 + k\theta)^{-\kappa}, \tag{8}$$

which can be written by using mean self-diffusion coefficient $\langle D \rangle$ and standard deviation σ_G as

$$I(k) = I_0\left(1 + k\sigma_G^2/\langle D \rangle\right)^{-\langle D \rangle^2/\sigma_G^2} \tag{9}$$

(see A for details regarding the gamma distribution model). This latter expression is not only more transparent to the practitioner, but was also found to be more numerically well-behaved. We will be interested in the spread CV and the polydispersity index, defined as

$$PDI = \frac{M_w}{M_n}, \tag{10}$$

where M_w and M_n are the weight-average and the number-average molecular weights, respectively [22]. The polydispersity index is per se a measure of the width of the *molecular-weight distribution*, whereas with PFG NMR we estimate the *self-diffusion coefficient distribution*. The self-diffusion coefficient D can be related to the molecular weight M by

$$D = KM^{-\alpha}, \tag{11}$$

with K and α being scaling parameters. From the parameters of the self-diffusion coefficient distribution, the polydispersity index can be computed as

$$PDI = \exp\left(\frac{\sigma_L^2}{\alpha^2} \right) \tag{12}$$

for the log-normal model and by

$$PDI = \left(1 + \frac{\sigma_G^2}{\langle D \rangle^2} \right)^{1/\alpha^2} \tag{13}$$

for the gamma model (see B for details about the polydispersity index calculations). In this paper we use $\alpha = 0.525$, the value previously measured for dilute PEG in water [6].

For multiple components which cannot be resolved due to their chemical shift, it is assumed that the signal attenuation is a weighted sum of several attenuations of any of the above types.

Fitting the models to data is performed by the standard non-linear least squares method [23]. We minimize the sum

$$S = \sum_n (I_{obs}(k_n) - I(k_n))^2, \tag{14}$$

where $I_{obs}(k_n)$ is the normalized signal intensity for $k = k_n$, yielding least squares (or equivalently, if the noise truly is Gaussian and independent, maximum likelihood [24,25]) estimates of the parameters for the chosen model.