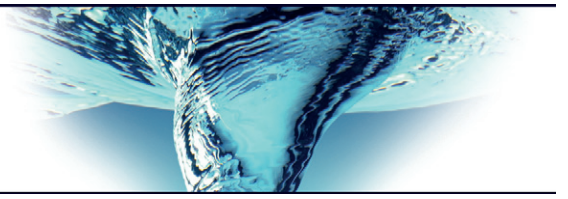
J. Fluid Mech. (2019), *vol.* 878, R3, doi:10.1017/jfm.2019.652



JFM RAPIDS
journals.cambridge.org/rapids

Diffusion coefficients of elastic macromolecules

Bogdan Cichocki[1], Marcin Rubin[1], Anna Niedzwiecka[2] and Piotr Szymczak[1,†]

[1]Institute of Theoretical Physics, Faculty of Physics, University of Warsaw, Pasteura 5, 02-093 Warsaw, Poland

[2]Institute of Physics, Polish Academy of Sciences, Aleja Lotnikow 32/46, PL-02668 Warsaw, Poland

(Received 23 June 2019; revised 6 August 2019; accepted 6 August 2019)

In elastic macromolecules, the value of the short-time diffusion coefficient depends on the choice of the point the displacement of which is tracked. On the other hand, the experimentally more relevant long-time diffusion coefficient is independent of the reference point, but its estimation usually requires computationally expensive Brownian dynamics simulations. Here we show how to obtain a precise estimate of the long-time diffusion coefficient of elastic macromolecules in a fast and robust manner, without invoking Brownian dynamics.

Key words: colloids

1. Introduction

Precise estimation of the diffusion constant of biomolecules is important for quantitative analysis of transport in living cells. It is also crucial for the proper interpretation of biophysical experiments, such as fluorescence correlation spectroscopy, ultracentrifugation or dynamic light scattering. If the internal structure of the macromolecule is rigid, then its transport properties can be efficiently calculated by a variety of methods, including bead modelling (Bloomfield, Dalton & Van Holde 1967; de la Torre & Bloomfield 1978; Byron 2008; Zuk, Cichocki & Szymczak 2018), boundary element methods (Allison 1999; Aragon 2004) or path integral techniques (Kang, Mansfield & Douglas 2004; Mansfield & Douglas 2008; Juba *et al.* 2017).

However, it becomes increasingly clear that the structure of most biomolecules is flexible and fluctuating rather than rigid. Even well-folded proteins undergo slow, large-scale movements of subunits, which are referred to as protein breathing (Makowski *et al.* 2008). In many cases, one finds flexible linkers or loops connecting domains, allowing for inter-domain hinge motions or flap motions (Jacobs, Kuhn & Thorpe 2002; Thorpe *et al.* 2005). For example, HIV-1 protease has two molecular

† Email address for correspondence: piotrek@fuw.edu.pl

flaps which move a distance of up to 7 Å when the enzyme becomes associated with a substrate (Ishima *et al.* 1999). At the extreme end of the flexibility spectrum are proteins with disordered sequences that fail to form a permanent tertiary structure but can adopt a variety of transient conformations. These are referred to as intrinsically disordered proteins (IDPs) (Dyson & Wright 2005; Oldfield & Dunker 2014). It is estimated that 30 %–50 % of eukaryotic proteins contain at least one long disordered region. These proteins participate in important regulatory functions in the cell, including transcription, translation and cell signalling (Galea *et al.* 2008; Oldfield *et al.* 2008; Dosztányi, Mészáros & Simon 2009). Several IDPs have been shown to be associated with various diseases such as cancer and neurodegenerative diseases (Iakoucheva *et al.* 2002; Uversky, Oldfield & Dunker 2008). The importance of IDPs in cellular processes calls for rethinking of the classical structure–function paradigm (Wright & Dyson 1999; Berlow, Dyson & Wright 2018), that protein function depends on a fixed three-dimensional structure.

Flexibility of the molecules brings considerable complications into the calculation of their hydrodynamic properties. Instead of the 6 degrees of freedom of a rigid molecule, we now need to deal with $6N$ degrees of freedom, with N standing for the number of rigid subunits. To make the problem tractable, a number of approximations have been adopted over the years. Perhaps the simplest was introduced by Kirkwood & Riseman (1948) in the context of polymer solutions. Here, the polymer is assumed to be rigidly frozen in one of a large number of possible conformations. Transport properties are then calculated by treating the molecule as a rigid body, and the results are averaged over the equilibrium ensemble. Such a rigid body approximation was later taken up by Zimm (1980, 1982) and de la Torre and co-workers (Schmidt, Cifre & de la Torre 2012; de la Torre 2016), and incorporated into the HYDRO set of programs (de la Torre 2016); however, the validity of this approximation and its accuracy remain unclear (Schmidt *et al.* 2012).

To facilitate the analytical treatment of polymer dynamics, Kirkwood & Riseman (1948) introduced an additional 'pre-averaging' approximation in which the chain conformation-dependent hydrodynamic disturbance, induced by the motion of the polymer, is replaced by its equilibrium average over all chain conformations. However, while pre-averaged Kirkwood–Riseman theory was successful in explaining many dynamical properties of polymers, a number of studies (Zimm 1980; de La Torre, Jimenez & Freire 1982) indicated that the errors introduced by this approach can be considerable.

An important observation, due to Fixman (1981, 1983), is that the diffusion coefficient in a flexible macromolecule will also be time-dependent, with well-defined short- and long-time limits. The difference between the two is due to memory effects related to the relaxation of the internal coordinates of the molecule. Due to the positiveness of the memory function, the long-time diffusion coefficient can be shown to be always smaller than the short-time diffusivity (Fixman 1983). These effects have been further studied by Liu & Dünweg (2003) using the Zwanzig–Mori projection operator technique.

From another angle, Wegener (1982) and Harvey, Mellado & García de la Torre (1983) have considered relatively simple molecules with one flexible joint, which was taken into account by the introduction of appropriate generalized coordinates. One problem that they encountered was that the diffusion coefficient turned out to be dependent on the choice of the origin. In an insightful paper, Wegener (1985) proposed that the appropriate reference point (which he called centre of diffusion, C_D) should correspond to the minimal short-time diffusion coefficient, but he was

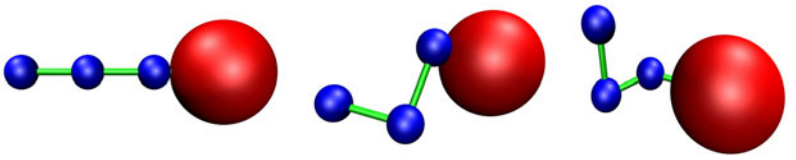


FIGURE 1. Snapshots of the trajectory of an elastic molecule undergoing Brownian motion.

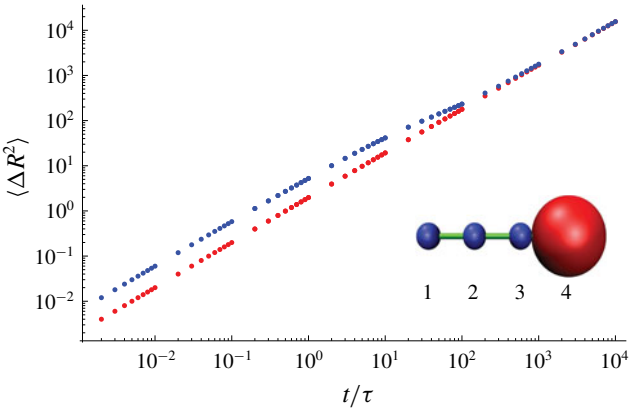


FIGURE 2. Mean square displacement (MSD) of a given point on a molecule versus time. The molecule consists of four beads connected with springs, as shown in the inset. The blue points correspond to the MSD of bead no. 1, whereas the red points correspond to that of bead no. 4.

not able to find an explicit formula for its position. He also hypothesized that the minimal short-time diffusion coefficient is equal to the long-time diffusion coefficient, accessible by experimental measurements.

Below, we show how to find the centre of diffusion for a rather broad class of elastic macromolecules, which can be represented as a collection of N beads interacting with a general intramolecular potential. We also demonstrate that the short-time diffusion coefficient calculated at the centre of diffusion is not equal to the long-time diffusion coefficient (as claimed by Wegener), but the difference between the two is small and can be calculated at a modest computational expense.

Before we turn to a formal description of the method, let us first briefly present the main problems one encounters when analysing the diffusion of flexible molecules, using a simple example. Consider a small linear molecule, which consists of four beads – three of radius a, and the fourth of radius $3a$ (figure 1). The beads are connected with harmonic springs of an equilibrium length $l_0 = 4a$ and spring constant $k = 5.5kT/a$. The hydrodynamic interactions between them are described within the Rotne–Prager approximation (Rotne & Prager 1969; Yamakawa 1970; Zuk *et al.* 2014). The evolution is calculated using the Brownian dynamics (BD) algorithm (Ermak & McCammon 1978). In figure 2, we show the mean square displacement of the centres of the first and fourth bead. The distances are measured in a, whereas the time unit is $\tau = a^2/6D_0$, where D_0 is the diffusion coefficient of a single bead of radius a.

There are several observations to be made based on the data in figure 2. First, for a given point P of the molecule, the mean square displacement $\Delta R^2(t)$ is linear in time only at very short and very long times. The corresponding definitions of the short-time and long-time translational diffusion coefficient read:

$$\langle (\boldsymbol{R}(t) - \boldsymbol{R}(0))^2 \rangle \sim 6D_s(P)t, \quad t \to 0 \tag{1.1}$$

and

$$\langle (\boldsymbol{R}(t) - \boldsymbol{R}(0))^2 \rangle \sim 6D_l\, t, \quad t \to \infty. \tag{1.2}$$

In the above, $\langle \cdots \rangle$ denotes the equilibrium average.

Second, the value of the short-time diffusion coefficient depends on the choice of the point that we track. Conversely, the long-time diffusivity is independent of the choice of the reference point. Finally, the value of the long-time diffusion coefficient is always smaller than that of its short-time counterpart. For the data in figure 2, $D_l = 0.2898 \pm 0.0002$ (in the units of D_0), whereas the short-time diffusion coefficients are $D_s(1) = 1$ (for the small bead) and $D_s(4) = 1/3$ (for the large bead).

In experiments, the long-time diffusion coefficient is usually measured, due to the time scales involved in either fluorescence correlation spectroscopy, ultracentrifugation or dynamic light scattering. However, a direct assessment of D_l in the BD simulations is very hard, if not impossible, for complex biomolecules. The main reason for this is a relatively slow convergence of the value of D_l with the time interval over which the data is collected, due to the presence of relatively long relaxation times, related to the relaxation of the internal degrees of freedom as well as the rotation of the macromolecule as a whole. The acquirement of statistically meaningful BD data on these time scales is computationally expensive, mostly due to the necessity of generating a normally distributed random vector with a covariance determined by the N-particle diffusion tensor. To give an example, the generation of figure 2 requires three days of computation time on a Intel Xeon E5-2670 workstation at 2.50 GHz. The bead models of real biological macromolecules are much more complex, comprising hundreds of beads. Since the computational complexity of simulating the BD trajectory increases with the number of beads (N) as N^3, the calculation becomes prohibitively expensive to carry out.

2. Short- and long-time diffusion coefficient

Turning to a formal description, let us introduce a general model of a macromolecule as a collection of beads of different sizes (possibly overlapping), suspended in a viscous fluid and interacting with potential forces. We assume that the Reynolds number is small and thus the motion is an overdamped one. On a mesoscopic level (Van Kampen 1992), the state of such a system is then described by positions of the centres of the beads $\boldsymbol{X} = (\boldsymbol{R}_1, \boldsymbol{R}_2, \ldots, \boldsymbol{R}_N)$. An important step in the analysis is splitting $\boldsymbol{X}$ into the internal and external coordinates. The internal coordinates describe simply the relative positions of the beads with respect to each other

$$\boldsymbol{X}_{int} = (\boldsymbol{R}_{12}, \boldsymbol{R}_{23}, \ldots, \boldsymbol{R}_{N-1,N}), \tag{2.1}$$

with $\boldsymbol{R}_{ij} = \boldsymbol{R}_i - \boldsymbol{R}_j$. The external coordinate, on the other hand, describes the position of one specific point P of the macromolecule in the laboratory frame. This position will be given by a vector $\boldsymbol{R}$, which is a linear combination of the positions of the beads

$$\boldsymbol{R} = x_1\boldsymbol{R}_1 + x_2\boldsymbol{R}_2 + \cdots + x_N\boldsymbol{R}_N, \tag{2.2}$$

with the non-negative coefficients x_i summing up to one (that is, $\sum_{i=1}^{N} x_i = 1$). The configuration vector can thus be written as $\boldsymbol{X} = (\boldsymbol{R}, \boldsymbol{X}_{int})$. The vector of the coefficients x_i will be denoted as $\boldsymbol{x}$:

$$\boldsymbol{x} = \{x_1, x_2, \ldots, x_N\}. \tag{2.3}$$

Note that for identical beads, the choice $x_1 = x_2 = \cdots x_N = 1/N$ corresponds to the position of the centre of mass of the system.

The potential energy of particle interactions $U(\boldsymbol{X}_{int})$ is only a function of the internal coordinates, $\boldsymbol{X}_{int}$. Similarly, the N-particle mobility matrix $\boldsymbol{\mu}$ describing the hydrodynamic interactions between the beads is also a function of $\boldsymbol{X}_{int}$ only. The mobility matrix links the velocities of the particles with the forces acting on them, according to

$$\boldsymbol{U}_i = \sum_j \boldsymbol{\mu}_{ij} \boldsymbol{F}_j, \tag{2.4}$$

where $\boldsymbol{U}_i$ is the velocity of bead i and $\boldsymbol{F}_j$ is the force with which particle j act on the fluid. Note that the mobility matrix needs to be positive-definite for all particle configurations (Happel & Brenner 1973).

Let $P(\boldsymbol{X}, t)$ be the probability density of finding the system in configuration $\boldsymbol{X}$ at time t, with the normalization $\int \mathrm{d}\boldsymbol{X}\, P(\boldsymbol{X}, t) = 1$, where $\mathrm{d}\boldsymbol{X} = \mathrm{d}\boldsymbol{R}\, \mathrm{d}\boldsymbol{X}_{int}$ and $\mathrm{d}\boldsymbol{X}_{int} = \mathrm{d}\boldsymbol{R}_{12} \ldots \mathrm{d}\boldsymbol{R}_{1N}$. At equilibrium, the distribution becomes uniform in $\boldsymbol{R}$ and Boltzmannian in the internal coordinates – that is, $P^{eq}(\boldsymbol{X}) = (1/V) P^{eq}(\boldsymbol{X}_{int})$ with $P^{eq}(\boldsymbol{X}_{int}) \sim \mathrm{e}^{-\beta U(\boldsymbol{X}_{int})}$. The evolution of a Brownian system is governed by the generalized Smoluchowski equation

$$\partial_t P(\boldsymbol{X}, t) = \mathcal{L}(\boldsymbol{X}) P(\boldsymbol{X}, t), \tag{2.5}$$

where

$$\mathcal{L}(\boldsymbol{X}) = \sum_{i,j=1}^{N} \frac{\partial}{\partial \boldsymbol{R}_i} \cdot \boldsymbol{D}_{ij}(\boldsymbol{X}_{int}) \cdot \left(\frac{\partial}{\partial \boldsymbol{R}_j} + \beta \frac{\partial U(\boldsymbol{X}_{int})}{\partial \boldsymbol{R}_j} \right) \tag{2.6}$$

is the evolution operator and

$$\boldsymbol{D}_{ij}(\boldsymbol{X}_{int}) = k_B T \boldsymbol{\mu}_{ij}(\boldsymbol{X}_{int}) \tag{2.7}$$

is the diffusion matrix. Note that $\partial/\partial \boldsymbol{R}_i = x_i (\partial/\partial \boldsymbol{R}) + \partial/\partial \boldsymbol{R}_{i,int}$, where the second term on the right-hand side denotes the derivative with respect to $\boldsymbol{R}_i$ through the dependence on $\boldsymbol{X}_{int}$.

In the following, we will characterize the motion of the point $\boldsymbol{R}$ (2.2) by introducing the time- and origin-dependent diffusion coefficient

$$D(t, \boldsymbol{x}) = \frac{1}{6} \frac{\mathrm{d}}{\mathrm{d}t} \langle [\boldsymbol{R}(t) - \boldsymbol{R}(0)]^2 \rangle, \tag{2.8}$$

where $\langle \cdots \rangle$ denotes an equilibrium average (that is, $\langle A \rangle = \int A(\boldsymbol{X}) P_{eq}(\boldsymbol{X})\, \mathrm{d}\boldsymbol{X}$) and $\boldsymbol{x}$ is the weight vector defined in (2.3). Note that here we adopt the convention that $P_{eq}(\boldsymbol{X})$ is always placed at the right-hand side of the integrand, which is important if A is a differential operator.

Making use of the stationarity of the stochastic process governing the motion of the molecule,

$$D(t, \boldsymbol{x}) = -\frac{1}{3} \frac{\mathrm{d}}{\mathrm{d}t} \langle \boldsymbol{R}(t) \cdot \boldsymbol{R}(0) \rangle = -\frac{1}{3} \frac{\mathrm{d}}{\mathrm{d}t} \mathrm{Tr} \langle \boldsymbol{R} \mathrm{e}^{\mathcal{L}(\boldsymbol{X})t} \boldsymbol{R} \rangle, \tag{2.9}$$

which allows us to express $D(t)$ in the following form (Akcasu 1982; Fixman 1983; Cichocki & Hinsen 1992; Liu & Dünweg 2003):

$$D(t,\boldsymbol{x}) = D_s(\boldsymbol{x}) - \int_0^t M(\tau,\boldsymbol{x})\,\mathrm{d}\tau, \tag{2.10}$$

where

$$D_s(\boldsymbol{x}) \equiv D_s(\boldsymbol{x}) = -\frac{1}{3}\mathrm{Tr}\langle \boldsymbol{R}\mathcal{L}(\boldsymbol{X})\boldsymbol{R}\rangle = \frac{k_BT}{3}\sum_{i,j=1}^{N} x_i x_j \langle \mathrm{Tr}\,\boldsymbol{\mu}_{ij}\rangle \tag{2.11}$$

is the short-time diffusion coefficient and the memory function $M(t)$ is given by

$$M(t,\boldsymbol{x}) = \tfrac{1}{3}\mathrm{Tr}\langle \boldsymbol{V}\mathrm{e}^{\mathcal{L}_{int}(\boldsymbol{X}_{int})t}\boldsymbol{V}\rangle. \tag{2.12}$$

Here, $\mathcal{L}_{int}(\boldsymbol{X}_{int})$ is the Smoluchowski operator $\mathcal{L}(\boldsymbol{X})$, in which the derivatives with respect to $\boldsymbol{R}_i$ were replaced by the derivatives with respect to $\boldsymbol{R}_{i,int}$. The particle flux, $\boldsymbol{V}$, is given by

$$\boldsymbol{V} = \mathcal{D}(\boldsymbol{X})\boldsymbol{R} = \sum_{i,j=1}^{N} x_j \left\{ \left[\frac{\partial}{\partial \boldsymbol{R}_{i,int}} \cdot \boldsymbol{D}_{ij}(\boldsymbol{X}_{int}) \right] - \beta \frac{\partial U(\boldsymbol{X}_{int})}{\partial \boldsymbol{R}_{i,int}} \cdot \boldsymbol{D}_{ij}(\boldsymbol{X}_{int}) \right\}, \tag{2.13}$$

where we have introduced the operator $\mathcal{D}(\boldsymbol{X})$ adjoint to $\mathcal{L}(\boldsymbol{X})$, given by $\mathcal{L}(\boldsymbol{X})P^{eq}(\boldsymbol{X})\cdots = P^{eq}(\boldsymbol{X})\mathcal{D}(\boldsymbol{X})$.

In biophysical experiments, a long-time diffusion coefficient is usually measured, $D_l = D(t=+\infty)$, related to displacements much larger than the size of the molecule. Note that, for a bound system, the long-time diffusion coefficient does not depend on the choice of the tracked point $\boldsymbol{R}$. Indeed, the difference of diffusion coefficients for two different choices of $\boldsymbol{x}$ follows

$$D(t,\boldsymbol{x}) - D(t,\boldsymbol{x}') = -\frac{1}{3}\frac{\mathrm{d}}{\mathrm{d}t}\langle [\boldsymbol{R}(t) - \boldsymbol{R}'(t)] \cdot [\boldsymbol{R}(0) + \boldsymbol{R}'(0)]\rangle, \tag{2.14}$$

where we have used (2.8) together with the condition of detailed balance. For a bound system, the vector $\boldsymbol{R}(t) - \boldsymbol{R}'(t)$ is of a limited length; hence, $D(t,\boldsymbol{x}) - D(t,\boldsymbol{x}') \to 0$ as $t \to \infty$. In our case this statement means that D_l becomes independent of the choice of the coefficients $\boldsymbol{x}$. On the other hand, both D_s and $M(\tau)$ do depend on the choice of the reference point. We then obtain

$$D_l = D_s(\boldsymbol{x}) - \int_0^{+\infty} M(\tau;\boldsymbol{x})\,\mathrm{d}\tau. \tag{2.15}$$

The direct computation of the long-time diffusion coefficient by means of BD simulation is computationally expensive (Liu & Dünweg 2003; Schmidt *et al.* 2012). However, a relatively simple estimate of D_l can be obtained by the analysis of (2.15). First, note that the right-hand side of the above relation is the difference between two non-negative quantities. This is a consequence of the fact that the Smoluchowski equation (2.5) governing the evolution of the system describes overdamped dynamics with detailed balance (Van Kampen 1992). Consequently, the lower the value of $D_s(\boldsymbol{x})$, the closer it is to the long-time diffusion coefficient. The upper bound for D_l can then be obtained by taking the minimum of the short-time diffusion coefficient $D_s(\boldsymbol{x})$ with respect to $\boldsymbol{x}$.

The formal procedure is the following. We introduce a matrix $\boldsymbol{A}$ indexed by the particle labels (i, j) as

$$A_{ij} = \tfrac{1}{3}\mathrm{Tr}\langle \boldsymbol{\mu}_{ij}\rangle. \tag{2.16}$$

We can then express $D_s(\boldsymbol{x})$ as

$$D_s(\boldsymbol{x}) = k_B T(\boldsymbol{x}^{\mathrm{T}} \cdot \boldsymbol{A} \cdot \boldsymbol{x}), \tag{2.17}$$

where the superscript T stands for the transposition. Next, we solve the variational problem of finding the minimum of $D_s(\boldsymbol{x})$ with respect to $\boldsymbol{x}$. The normalization condition for $\boldsymbol{x}$ can be expressed as $\boldsymbol{x}^{\mathrm{T}} \cdot \boldsymbol{i} = 1$, where $\boldsymbol{i}$ is a column vector of ones. The variational problem for $D_s(\boldsymbol{x})$ with respect to x_i, $i = 1, 2, \ldots, N$, is then

$$\delta\{\boldsymbol{x}^{\mathrm{T}} \cdot \boldsymbol{A} \cdot \boldsymbol{x} - \lambda(\boldsymbol{x}^{\mathrm{T}} \cdot \boldsymbol{i} - 1)\} = 0. \tag{2.18}$$

From the above, we get the extremum condition $2\boldsymbol{A} \cdot \boldsymbol{x}_{min} - \lambda \boldsymbol{i} = 0$. This corresponds to a minimum, due to the positive definiteness of the matrix $\boldsymbol{A}$. Explicitly,

$$\boldsymbol{x}_{min} = \frac{\lambda}{2}\boldsymbol{A}^{-1} \cdot \boldsymbol{i}, \tag{2.19}$$

where the coefficient λ can be obtained from the normalization condition, yielding $\lambda/2 = (\boldsymbol{i}^{\mathrm{T}} \cdot \boldsymbol{A}^{-1} \cdot \boldsymbol{i})^{-1}$, so that

$$\boldsymbol{x}_{min} = \frac{\boldsymbol{A}^{-1} \cdot \boldsymbol{i}}{\boldsymbol{i}^{\mathrm{T}} \cdot \boldsymbol{A}^{-1} \cdot \boldsymbol{i}} = \frac{\sum_j (\boldsymbol{A}^{-1})_{ij}}{\sum_{i,j} (\boldsymbol{A}^{-1})_{ij}}. \tag{2.20}$$

Finally, we get the minimum value of the short-time diffusion coefficient as

$$D_s(\boldsymbol{x}_{min}) = k_B T \boldsymbol{x}_{min}^{\mathrm{T}} \cdot \boldsymbol{A} \cdot \boldsymbol{x}_{min} = k_B T \frac{1}{\boldsymbol{i}^{\mathrm{T}} \cdot \boldsymbol{A}^{-1} \cdot \boldsymbol{i}} = k_B T \frac{1}{\sum_{i,j} (\boldsymbol{A}^{-1})_{ij}}. \tag{2.21}$$

3. Discussion

Returning to the example of the chain molecule of figure 2, for the Rotne–Prager model of hydrodynamic interactions, (2.21) gives the diffusion constant of $D_s(\boldsymbol{x}_{min}) = 0.2919$, which is less than 0.7 % off the long-time diffusion coefficient obtained in the BD simulations. On the other hand, the Kirkwood formula, which corresponds to calculating the diffusion coefficient in the geometrical centre of a macromolecule (that is, for $x_i = 1/N$), gives

$$D_K = \frac{k_B T}{N^2} \sum_{i,j} (\boldsymbol{A})_{ij} = 0.375, \tag{3.1}$$

which significantly overestimates the value of the long-time diffusion coefficient. In fact, D_K is even larger than the single-body diffusion coefficient of the largest bead (that is, for $x_i = \delta_{4i}$), which is $D_4 = k_B T A_{44} = 0.33(3)$. A slightly better estimate can be obtained by taking the centre of mass of the macromolecule, with $x_i \sim a_i^3$, where a_i

is the radius of the ith bead. In this case $D_{CM} = 0.312$, which is still, however, about 8 % off the value obtained from the BD. Another possibility is to take the weight proportional to the friction coefficients of individual beads (that is, $x_i \sim a_i$). This leads to $D = 0.309$, about 7 % off the BD value.

Note that (2.21) itself is exact, as no approximations have been made in the derivation. One can use it for different models of hydrodynamic interactions – both in simple models (for example, Oseen or Rotne–Prager far-field approximation (Kim & Karrila 1991)) or in more sophisticated approaches, like the multipole expansion method (Mazur & van Saarloos 1982; Felderhof 1988; Cichocki *et al.* 1994). The interaction potential can also be arbitrary, provided that it keeps the system bound. Importantly, the estimation of the long-time diffusion coefficient using (2.21) is not intensive computationally, since it does not require BD simulations. All that is needed are the equilibrium averages $\boldsymbol{A}_{ij}$ (2.16), which can be obtained, for example, by Monte Carlo sampling (Binder 1995).

In polymer physics, pre-averaging approximation is popular (Kirkwood & Riseman 1948; Zimm 1956; Dubois-Violette & De Gennes 1967; Yamakawa 1971; Bird *et al.* 1987; Doi & Edwards 1988; Prakash 1999), in which the diffusion matrix $\boldsymbol{D}_{ij}$ is replaced in the evolution equation by its equilibrium average. This leads to a significant simplification of the dynamics. The particle flux (2.13) is then simply

$$\boldsymbol{V}(\boldsymbol{x}) = -\beta \sum_{i,j=1}^{N} \boldsymbol{F}_i \boldsymbol{A}_{ij} x_j, \tag{3.2}$$

where we have used the fact that $\langle \boldsymbol{D}_{ij} \rangle = \boldsymbol{A}_{ij} \boldsymbol{I}$. In particular, if $\boldsymbol{x}$ corresponds to the diffusion centre, then the flux vanishes, since $\boldsymbol{V}(\boldsymbol{x}_{min}) \sim \sum \boldsymbol{F}_i = 0$. Hence, within the pre-averaging approximation, the memory function at the centre of diffusion vanishes; consequently, the long-time diffusion coefficient at this point is equal to its short-time counterpart. In fact, as shown in Zimm (1956), Yamakawa (1971), Bird *et al.* (1987), Öttinger (1987, 1996) with the pre-averaged dynamics, the motion of the centre of resistance separates out from that of the internal configuration and becomes a Markov process. In this literature, $\boldsymbol{x}_{min}$ is usually called 'centre of resistance', since it also has the property that the net torque of hydrodynamic forces around this point vanishes during a uniform translation of the molecule relative to the fluid (Zimm 1956; Yamakawa 1971). However, these properties of $\boldsymbol{x}_{min}$ hold only within the pre-averaging approximation (Bird *et al.* 1987; Öttinger 1987). To avoid confusion regarding the properties of $\boldsymbol{x}_{min}$, we have decided to call it 'centre of diffusion', following Wegener (1985), who coined that term for a point inside the molecule where the diffusion coefficient is minimal.

If we go beyond the pre-averaging approximation, the memory function no longer vanishes, and we expect a small difference between the long-time diffusion coefficient and its short-time counterpart, calculated at $\boldsymbol{x}_{min}$. This can be estimated by a direct calculation of the memory function given by (2.12). For our chain molecule, the memory function is shown in figure 3. We present there $M(t)$ for three different points of the molecule – one of the small beads (no. 1 in the inset of figure 2), the large bead (no. 4) and the centre of diffusion. The two former points exhibit exponential tails with the same characteristic decay time, connected with the rotation of the system as a whole. At the same time, the dynamics of the centre of diffusion shows a much smaller memory contribution. Even more importantly, the decay of $M(t)$ is now much faster – the long relaxation time connected with the slowest decaying mode is not present. This means that the calculation of memory correction

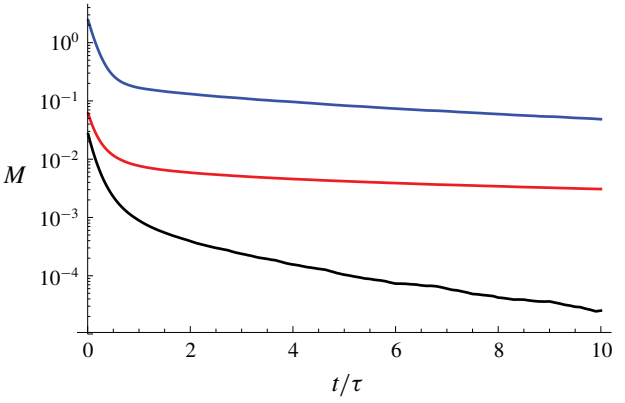


FIGURE 3. The memory function $M(t)$ given by (2.12) for three different points on the molecule – bead no. 1 (blue), bead no. 4 (red) and the centre of diffusion (black).

to the diffusion coefficient at $\boldsymbol{x}_{min}$ does not require the generation of long Brownian trajectories. Already the integration up to $T=5$ gives $\int_0^{\mathrm{T}} M(t)\,\mathrm{d}t = 0.002$, which differs just by 5 % from the integral up to $T=100$. Note that the memory correction itself is just 0.7 % of the long-time diffusivity, when calculated at the centre of diffusion. Contrastingly, the memory contribution to D_l calculated at bead no. 1 corresponds to as much as 250 % of the final value, whereas that calculated at bead no. 4 is approximately 15 %.

Finally, let us briefly recall the case of a rigid macromolecule. In such a case, the memory function can be calculated analytically (Cichocki, Ekiel-Jezewska & Wajnryb 2012; Cichocki, Ekiel-Jeżewska & Wajnryb 2015) and it can be shown that there exists a reference point for which the memory function vanishes; thus the mean square displacement of this point is linear over the entire time range.

In summary, we have presented a simple and accurate method of estimating the experimentally relevant long-time diffusivity of elastic macromolecules, based on the minimization of the short-time diffusion coefficient with respect to the tracked point. The method becomes exact within the pre-averaging approximation. In the case considered here, the long-time diffusivity is overestimated slightly (within 1 %) using this method. The exact value of this difference can be estimated by calculating the memory function, $M(t)$, at the centre of diffusion, where the short-time diffusivity is minimal. This calculation, although involving dynamical simulations, is relatively inexpensive, since $M(t)$ decays at this point on a much faster time scale than at any other point of the macromolecule. The computational simplicity of the method is crucial for the efficient estimation of diffusion coefficients of large and complex elastic macromolecules, the bead models of which involve hundreds of components.

Naturally, the accuracy of the method will depend on the system studied, including the number of beads, their sizes and interbead potentials. It seems, however, that elimination of the slowest decaying mode in the memory function at $\boldsymbol{x}_{min}$ will in general produce a high-accuracy estimate of D_l. The full assessment of the accuracy of the method will only be possible if the relaxation times corresponding to the Smoluchowski dynamics (2.5) are estimated, which should be the subject of further investigations.

Acknowledgement

This work was partially supported by grant no. UMO-2016/22/E/NZ1/00656 from the National Science Center, Poland.

References

AKCASU, A. Z. 1982 Comments on the diffusion coefficient and first cumulant. *Macromolecules* **15** (5), 1321–1324.

ALLISON, S. A. 1999 Low Reynolds number transport properties of axisymmetric particles employing stick and slip boundary conditions. *Macromolecules* **32** (16), 5304–5312.

ARAGON, S. 2004 A precise boundary element method for macromolecular transport properties. *J. Comput. Chem.* **25** (9), 1191–1205.

BERLOW, R. B., DYSON, H. J. & WRIGHT, P. E. 2018 Expanding the paradigm: intrinsically disordered proteins and allosteric regulation. *J. Molecular Biol.* **430** (16), 2309–2320.

BINDER, K. 1995 *Monte Carlo and Molecular Dynamics Simulations in Polymer Science*. Oxford University Press.

BIRD, R., HASSAGER, O., ARMSTRONG, R. & CURTISS, C. 1987 Kinetic theory. In *Dynamics of Polymeric Liquids*, vol. 2. John Wiley.

BLOOMFIELD, V., DALTON, W. & VAN HOLDE, K. 1967 Frictional coefficients of multisubunit structures. I. Theory. *Biopolymers* **5** (2), 135–148.

BYRON, O. 2008 Hydrodynamic modeling: the solution conformation of macromolecules and their complexes. *Method. Cell Biol.* **84**, 327–373.

CICHOCKI, B., EKIEL-JEZEWSKA, M. & WAJNRYB, E. 2012 Intrinsic viscosity for Brownian particles of arbitrary shape. *J. Phys.: Conf. Ser.* **392**, 012004.

CICHOCKI, B., EKIEL-JEŻEWSKA, M. L. & WAJNRYB, E. 2015 Brownian motion of a particle with arbitrary shape. *J. Chem. Phys.* **142** (21), 214902.

CICHOCKI, B., FELDERHOF, B. U., HINSEN, K., WAJNRYB, E. & BŁAWZDZIEWICZ, J. 1994 Friction and mobility of many spheres in Stokes flow. *J. Chem. Phys.* **100**, 3780–3790.

CICHOCKI, B. & HINSEN, K. 1992 Dynamic computer simulation of concentrated hard sphere suspensions. II. Re-analysis of mean square displacement data. *Physica* A **187** (1–2), 133–144.

DOI, M. & EDWARDS, S. F. 1988 *The Theory of Polymer Dynamics*, vol. 73. Oxford University Press.

DOSZTÁNYI, Z., MÉSZÁROS, B. & SIMON, I. 2009 Bioinformatical approaches to characterize intrinsically disordered/unstructured proteins. *Brief. Bioinform.* **11** (2), 225–243.

DUBOIS-VIOLETTE, E. & DE GENNES, P.-G. 1967 Quasi-elastic scattering by dilute, ideal, polymer solutions. II. Effects of hydrodynamic interactions. *Physics* **3** (4), 181–198.

DYSON, H. J. & WRIGHT, P. E. 2005 Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **6** (3), 197–208.

ERMAK, D. L. & MCCAMMON, J. A. 1978 Brownian dynamics with hydrodynamic interactions. *J. Chem. Phys.* **69**, 1352–1360.

FELDERHOF, B. U. 1988 Many-body hydrodynamic interactions in suspensions. *Physica* A **151**, 1–16.

FIXMAN, M. 1981 Inclusion of hydrodynamic interaction in polymer dynamical simulations. *Macromolecules* **14** (6), 1710–1717.

FIXMAN, M. 1983 Variational bounds for polymer transport coefficients. *J. Chem. Phys.* **78** (3), 1588–1593.

GALEA, C. A., WANG, Y., SIVAKOLUNDU, S. G. & KRIWACKI, R. W. 2008 Regulation of cell division by intrinsically unstructured proteins: intrinsic flexibility, modularity, and signaling conduits. *Biochemistry* **47** (29), 7598–7609.

HAPPEL, J. & BRENNER, H. 1973 *Low Reynolds Number Hydrodynamics*. Noordhoff.

HARVEY, S. C., MELLADO, P. & GARCÍA DE LA TORRE, J. 1983 Hydrodynamic resistance and diffusion coefficients of segmentally flexible macromolecules with two subunits. *J. Chem. Phys.* **78** (4), 2081–2090.

IAKOUCHEVA, L. M., BROWN, C. J., LAWSON, J. D., OBRADOVIĆ, Z. & DUNKER, A. K. 2002 Intrinsic disorder in cell-signaling and cancer-associated proteins. *J. Molecular Biol.* **323** (3), 573–584.

ISHIMA, R., FREEDBERG, D. I., WANG, Y.-X., LOUIS, J. M. & TORCHIA, D. A. 1999 Flap opening and dimer-interface flexibility in the free and inhibitor-bound HIV protease, and their implications for function. *Structure* **7** (9), 1047–1055.

JACOBS, D. J., KUHN, L. A. & THORPE, M. F. 2002 Flexible and rigid regions in proteins. In *Rigidity Theory and Applications*, pp. 357–384. Springer.

JUBA, D., AUDUS, D. J., MASCAGNI, M., DOUGLAS, J. F. & KEYROUZ, W. 2017 Zeno: software for calculating hydrodynamic, electrical, and shape properties of polymer and particle suspensions. *J. Res. Natl Inst. Stand. Technol.* **122**, 20.

KANG, E.-H., MANSFIELD, M. L. & DOUGLAS, J. F. 2004 Numerical path integration technique for the calculation of transport properties of proteins. *Phys. Rev.* E **69** (3), 031918.

KIM, S. & KARRILA, S. J. 1991 *Microhydrodynamics: Principles and Selected Applications*. Butterworth-Heinemann.

KIRKWOOD, J. G. & RISEMAN, J. 1948 The intrinsic viscosities and diffusion constants of flexible macromolecules in solution. *J. Chem. Phys.* **16**, 565–573.

DE LA TORRE, J. G., JIMENEZ, A. & FREIRE, J. J. 1982 Monte Carlo calculation of hydrodynamic properties of freely jointed, freely rotating, and real polymethylene chains. *Macromolecules* **15** (1), 148–154.

LIU, B. & DÜNWEG, B. 2003 Translational diffusion of polymer chains with excluded volume and hydrodynamic interactions by Brownian dynamics simulation. *J. Chem. Phys.* **118** (17), 8061–8072.

MAKOWSKI, L., RODI, D. J., MANDAVA, S., MINH, D. D., GORE, D. B. & FISCHETTI, R. F. 2008 Molecular crowding inhibits intramolecular breathing motions in proteins. *J. Molecular Biol.* **375** (2), 529–546.

MANSFIELD, M. L. & DOUGLAS, J. F. 2008 Improved path integration method for estimating the intrinsic viscosity of arbitrarily shaped particles. *Phys. Rev.* E **78** (4), 046712.

MAZUR, P. & VAN SAARLOOS, W. 1982 Many-sphere hydrodynamic interactions and mobilities in a suspension. *Physica* A **115**, 21–57.

OLDFIELD, C. J. & DUNKER, A. K. 2014 Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu. Rev. Biochem.* **83**, 553–584.

OLDFIELD, C. J., MENG, J., YANG, J. Y., YANG, M. Q., UVERSKY, V. N. & DUNKER, A. K. 2008 Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. *BMC Genomics* **9** (1), S1.

ÖTTINGER, H. C. 1987 Translational diffusivity from the Zimm model. *J. Chem. Phys.* **87** (5), 3156–3165.

ÖTTINGER, H. C. 1996 *Stochastic Processes in Polymeric Fluids: Tools and Examples for Developing Simulation Algorithms*. Springer.

PRAKASH, J. R. 1999 The kinetic theory of dilute solutions of flexible polymers: hydrodynamic interaction. In *Rheology Series*, vol. 8, pp. 467–517. Elsevier.

ROTNE, J. & PRAGER, S. 1969 Variational treatment of hydrodynamic interaction in polymers. *J. Chem. Phys.* **50**, 4831–4837.

SCHMIDT, R. R., CIFRE, J. H. & DE LA TORRE, J. G. 2012 Translational diffusion coefficients of macromolecules. *Eur. Phys. J.* E **35** (12), 130.

THORPE, M. F., CHUBYNSKY, M., HESPENHEIDE, B., MENOR, S., JACOBS, D. J., KUHN, L. A., ZAVODSZKY, M. I., LEI, M., RADER, A. & WHITELEY, W. 2005 Flexibility in biomolecules. In *Current Topics In Physics: In Honor of Sir Roger J. Elliott*, pp. 97–112. World Scientific.

DE LA TORRE, J. G. 2016 The HYDRO software suite for the prediction of solution properties of rigid and flexible macromolecules and nanoparticles. In *Analytical Ultracentrifugation. Instrumentation, Software, and Applications*, pp. 195–217. Springer.

DE LA TORRE, J. G. & BLOOMFIELD, V. A. 1978 Hydrodynamic properties of macromolecular complexes. IV. Intrinsic viscosity theory, with applications to once-broken rods and multisubunit proteins. *Biopolymers* **17**, 1605–1627.

UVERSKY, V. N., OLDFIELD, C. J. & DUNKER, A. K. 2008 Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu. Rev. Biophys.* **37**, 215–246.

VAN KAMPEN, N. G. 1992 *Stochastic Processes in Physics and Chemistry*. Elsevier.

WEGENER, W. A. 1982 Bead models of segmentally flexible macromolecules. *J. Chem. Phys.* **76** (12), 6425–6430.

WEGENER, W. A. 1985 Center of diffusion of flexible macromolecules. *Macromolecules* **18** (12), 2522–2530.

WRIGHT, P. E. & DYSON, H. J. 1999 Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Molecular Biol.* **293** (2), 321–331.

YAMAKAWA, H. 1970 Transport properties of polymer chains in dilute solution: hydrodynamic interaction. *J. Chem. Phys.* **53**, 436–443.

YAMAKAWA, H. 1971 *Modern Theory of Polymer Solutions*. Harper & Row.

ZIMM, B. H. 1956 Dynamics of polymer molecules in dilute solution: viscoelasticity, flow birefringence and dielectric loss. *J. Chem. Phys.* **24** (2), 269–278.

ZIMM, B. H. 1980 Chain molecule hydrodynamics by the Monte-Carlo method and the validity of the Kirkwood–Riseman approximation. *Macromolecules* **13** (3), 592–602.

ZIMM, B. H. 1982 Sedimentation of asymmetric elastic dumbbells and the rigid-body approximation in the hydrodynamics of chains. *Macromolecules* **15** (2), 520–525.

ZUK, P. J., CICHOCKI, B. & SZYMCZAK, P. 2018 GRPY: an accurate bead method for calculation of hydrodynamic properties of rigid biomacromolecules. *Biophys. J.* **115** (5), 782–800.

ZUK, P. J., WAJNRYB, E., MIZERSKI, K. A. & SZYMCZAK, P. 2014 Rotne–Prager–Yamakawa approximation for different-sized particles in application to macromolecular bead models. *J. Fluid Mech.* **741**, R5.