Hydrodynamic techniques in the study of elastic macromolecules

Radost Waszkiewicz

University of Warsaw
Faculty of Physics

A dissertation submitted to the University of Warsaw
for the degree of Doctor of Philosophy
under the supervision of dr hab. Maciej Lisicki, prof. UW

Warsaw, 2024–03–25

Summary

Recent, increasingly accurate measurements of the properties of bio-relevant molecules challenge our understanding of macromolecules as having well-defined, fixed shapes that determine their function. The equilibrium distribution of their conformations is described by the Boltzmann distribution, which combines two terms: the influence of the temperature of the solvent and the potential energy of a given configuration, which in turn captures the elastic properties of the molecules. The behavior of elastic macromolecules is thus shaped by the competition of these two physical phenomena—whenever the valleys of the potential energy landscape are shallow in comparison to a typical thermal fluctuation, a wide variety of conformations are present; conversely, when they are deep, only small deviations from the energy-minimizing configurations can be observed.

The doctoral thesis presents a theoretical description of the conformational variability of elastic macromolecules and its effect on diffusion. The first part of the thesis provides an overview of theoretical fundamentals required for building coarse-grained models, which form the core of this work. It also delves into the theoretical underpinnings of the experimental methods used to validate simplifying assumptions made in the former. The second part of the thesis comprises a series of thematically linked publications and preprints in which we demonstrate methods for dealing with and taking advantage of both extremes of the elasticity spectrum.

Starting from molecules with very large persistence length compared to their size, we demonstrate how to model the approach of a very short DNA segment towards a nanopore and provide an analysis of the influence of wall interaction and hydrodynamic anisotropy in the nanopore capture process. By considering a rod-like molecule with uniformly distributed charge, we establish theoretical criteria for determining when and where the inclusion of wall corrections is necessary. Secondly, we investigate the impact of negative supercoiling and curvature on the hydrodynamic properties of 336 bp and 672 bp DNA minicircles. Utilizing linear elasticity theory and hydrodynamic calculations, we predict DNA shapes and diffusion coefficients. Our results show a favorable comparison with experimental data on diffusion and sedimentation coefficients obtained using analytical ultracentrifugation. For intermediate values of persistence length, we determine the range of lengths and g-forces under which sedimentation of a flexible, looped filament remains stable to buckling. Our analysis, based on linear elasticity theory combined with resistive force theory, yields a stability criterion reliant on a single dimensionless parameter.

In a more general case where both thermal fluctuations and elastic forces are significant, we propose a numerical approach. This approach is based on a stochastic differential equations integrator we developed, combined with hydrodynamic interactions based on the Rotne-Prager approximation. Additionally, we present a suite of Python packages designed to make small Brownian Dynamics simulations both fast to develop and fast to simulate, thanks to hardware acceleration.

We demonstrate that for intrinsically disordered proteins (IDPs), which represent the opposite extreme of the elasticity spectrum as compared to very stiff DNA, excluded volume interactions are the key factor determining the equilibrium conformational ensemble. We introduce the Globule-Linker model for generating conformations and combine it with the Minimum Dissipation Approximation to predict their hydrodynamic size. Using the comparison of the coarse-grained approach with the largest set of experimental values collected to date, we show that our first-principles approach outperforms phenomenological fits available in the literature.

Finally, we deal with the theoretical problem of equilibrium distributions of molecules with both very stiff degrees of freedom and comparatively free ones (such as the bond length and inter-bond angles, respectively, in molecular models). We identify the important details of the constraining potentials overlooked in earlier works and demonstrate a method of their computation, both in general and through specific examples.

Our results, which span a spectrum of possible elastohydrodynamic regimes, demonstrate the rich diversity of phenomena that arise from the competition of structural rigidity, viscous stresses and thermal fluctuations. In order to facilitate the analysis of such systems, we have created a number of open-source numerical tools, which have been published with documentation. The applicability of these tools was tested directly in relatively stiff biopolymers—the circular DNA—and soft structures of IDPs. Additionally, they provide an insight into the intermediate stiffness regimes, where thermal fluctuations compete with intramolecular interactions, and can be used for a better understanding of microscale elastohydrodynamic phenomena.

Streszczenie

Nowe, coraz dokładniejsze pomiary właściwości istotnych biologicznie cząsteczek podważają nasze rozumienie makromolekuł jako cząsteczek o dobrze określonych, stałych kształtach określających ich funkcję. Rozkład równowagowy ich konformacji opisuje rozkład Boltzmanna, który łączy dwa efekty: wpływ temperatury rozpuszczalnika oraz energię potencjalną danej konfiguracji, która z kolei opisuje właściwości elastyczne cząsteczki. Zachowanie elastycznych makromolekuł jest zatem określone przez współzawodnictwo dwóch zjawisk fizycznych: gdy studnie potencjału są płytkie w porównaniu z typowymi fluktuacjami termicznymi, występuje szeroka gama konformacji; z drugiej strony, gdy są one głębokie, można zaobserwować jedynie niewielkie odchylenia od konfiguracji minimalizujących energię makromolekuły.

W niniejszej rozprawie doktorskiej przedstawiono teoretyczny opis zmienności konformacyjnej elastycznych makromolekuł i jej wpływu na ich dyfuzję. Pierwsza część rozprawy zawiera przegląd podstaw teoretycznych wymaganych do budowy modeli gruboziarnistych, które stanowią istotę rozprawy oraz teoretyczne podstawy metod eksperymentalnych stosowanych do weryfikacji upraszczających założeń, wykorzystanych do konstrukcji modeli. Druga część rozprawy składa się z szeregu powiązanych tematycznie publikacji i preprintów, w których prezentujemy metody modelowania makromolekuł w szerokim zakresie ich sprężystości.

Zaczynając od cząsteczek o dalekim zasięgu korelacji elastycznych (tzn. o dużej długości persystencji, ang. persistence length) w porównaniu z ich rozmiarem, pokazujemy jak modelować zbliżanie się krótkiego fragmentu DNA do nanoporu. Analizujemy przy tym wpływ ścianek i anizotropii hydrodynamicznej makrocząstki w procesie wychwytywania cząstek przez nanopory. Określamy teoretyczne kryteria dla przypadków, w jakich konieczne jest uwzględnienie oddziaływania hydrodynamicznego cząsteczek ze ściankami, modelując kształt cząstki jako pręt z równomiernie rozłożonym ładunkiem elektrycznym. W kolejnych pracach badamy wpływ ujemnego trzeciorzędowego skręcenia (ang. supercoiling) i krzywizny DNA na właściwości hydrodynamiczne minipętli DNA o długości 336 i 672 par zasad. Wykorzystując liniową teorię elastyczności i modele hydrodynamiczne, przewidujemy kształty DNA i ich współczynniki dyfuzji w roztworze w zależności od stopnia skręcenia. Pomiary eksperymentalne współczynników dyfuzji i sedymentacji uzyskane za pomocą ultrawirowania analitycznego wykazują dobrą zgodność z naszymi przewidywaniami teoretycznymi. Dla pośrednich wartości zasięgu korelacji elastycznych wyznaczamy zakres długości i skalę sił zewnętrznych, przy którym sedymentacja elastycznej, cienkiej pętli pozostaje stabilna na wyboczenia. Nasza analiza, oparta na liniowej teorii elastyczności połączonej z teorią lokalnego oporu hydrodynamicznego, wyznacza kryterium stabilności zależne od pojedynczego bezwymiarowego parametru. W bardziej ogólnym przypadku, gdy istotne są zarówno fluktuacje termiczne, jak i siły sprężyste, prezentujemy podejście numeryczne. Opiera się ono na algorytmie całkowania stochastycznych równań różniczkowych, połączonym z modelem oddziaływań hydrodynamicznych opartym na przybliżeniu Rotne-Pragera. Powyższe metody zostały udostępnione jako zbiór paczek w Pythonie zaprojektowanych w celu szybkiego programowania i wykonywania symulacji dynamiki Brownowskiej.

W przypadku białek nieustrukturyzowanych (ang. intrinsically disordered proteins, IDP), które reprezentują przeciwny biegun spektrum elastyczności w porównaniu z bardzo sztywnym DNA i mogą być rozpatrywane jako bardzo wiotkie, determinantą ich konformacji w rozkładach równowagowych jest wykluczona objętość. Do modelowania białek IDP, proponujemy model konformacji GLM (ang. globule-linker model), który w połączeniu z hydrodynamicznym przybliżeniem minimalnej dyssypacji pozwala obliczyć wielkość hydrodynamiczną takich białek. Porównując nasze gruboziarniste podejście teoretyczne z największym jak dotąd zmierzonym zbiorem danych eksperymentalnych pokazujemy, że nasze podejście jest skuteczniejsze niż dopasowania fenomenologiczne dostępne w literaturze. W ostatnim artykule rozważamy teoretyczny problem rozkładów równowagowych cząsteczek o wielu stopniach swobody, z których część jest silnie związana (jak na przykład długości wiązań chemicznych). Pokazujemy, że istotne detale potencjałów realizujących więzy zostały przeoczone we wcześniejszych pracach oraz proponujemy właściwą, ścisłą matematycznie metodę ich uwzględnienia.

Nasze wyniki, obejmujące szeroki zakres korelacji elastycznych, pokazują różnorodność zjawisk wynikających z współwystępowania elastyczności, sił lepkich i fluktuacji termicznych. Aby ułatwić analizę takich układów, stworzyliśmy zbiór otwartych narzędzi numerycznych, które zostały opublikowane wraz z dokumentacją. Możliwość zastosowania powyższych narzędzi przetestowaliśmy bezpośrednio w stosunkowo sztywnym biopolimerze (minikółka z DNA) i w miękkich strukturach białek nieuporządkowanych (IDP). Przedstawione narzędzia dają wgląd w pośrednie reżimy sztywności, w których fluktuacje termiczne konkurują z oddziaływaniami wewnątrzcząsteczkowymi. Liczymy, że proponowane modele mogą posłużyć do lepszego zrozumienia zjawisk elastohydrodynamicznych w mikroskali.

Manuscripts comprising the Thesis

- Radost Waszkiewicz and Maciej Lisicki. "Hydrodynamic effects in the capture of rod-like molecules by a nanopore". In: *Journal of Physics: Condensed Matter* **33**. (2021), p. 104005. DOI: 10.1088/1361-648X/abd11b
- Radost Waszkiewicz, Piotr Szymczak, and Maciej Lisicki. "Stability of sedimenting flexible loops". In: *Journal of Fluid Mechanics* **919** (2021), A14. DOI: 10.1017/jfm.2021.350
- Radost Waszkiewicz, Maduni Ranasinghe, Jonathan M Fogg, Daniel J Catanese Jr, Maria L Ekiel-Jeżewska, Maciej Lisicki, Borries Demeler, Lynn Zechiedrich, and Piotr Szymczak. "DNA supercoiling-induced shapes alter minicircle hydrodynamic properties". In: *Nucleic Acids Research* **51**. (2023), pp. 4027–4042. DOI: 10.1093/nar/gkad183
- Radost Waszkiewicz, Maciej Bartczak, Kamil Kolasa, and Maciej Lisicki. "Pychastic: Precise Brownian dynamics using Taylor-Itō integrators in Python". In: *SciPost Physiscs Codebases* **11** (2023). DOI: 10.21468/SciPostPhysCodeb.11
- Radost Waszkiewicz, Agnieszka Michaś, Michał K. Białobrzewski, Barbara Klepka, Maja Cieplak-Rotowska, Zuzanna Staszałek, Bogdan Cichocki, Maciej Lisicki, Piotr Szymczak, and Anna Niedźwiecka. "Minimum dissipation approximation: A fast algorithm for the prediction of diffusive properties of intrinsically disordered proteins". In: *Journal of Physical Chemisty Letters (in review)* (2024). DOI: 10.1101/2024.02.05.578612
- Radost Waszkiewicz and Maciej Lisicki. "The trimer paradox: the effect of stiff constraints on equilibrium distributions in overdamped dynamics". In: *Physical Review E (in review)* (2024). DOI: 10.48550/arXiv.2403.13185

Funding

Research on the topic of the Thesis was supported by the grant *Dynamic deformations of elastic filaments in viscous flows*, financed by the National Centre of Science in Poland under the grant agreement to Maciej Lisicki no. 2018/31/D/ST3/02408

Contents

Introduction

1.1 Mesoscopic world of macromolecules

A fundamental technique for assessing the relative importance of different physical mechanisms within a theoretical framework, originating in the domain of fluid mechanics (cf. Reynolds, [7]), involves the consideration of dimensionless numbers. Here, we follow the approach of Naegele [8], where he motivates common approximations in colloidal physics by considering timescale ratios.

Colloidal hydrodynamics focuses on the properties of colloidal suspensions, which are mixtures of very small objects (with sizes of the order of 0.1 μm) and a solvent, typically water. At human-length scale, water is easily treated as incompressible (for example, by considering volume changes at typical pressures on the order of atmospheric pressure). However, this becomes less obvious at the microscopic length scale where the granularity of matter becomes significant. The relevant bulk density relaxation timescale τ_s at the length scales relevant to the colloidal object of size a can be estimated from the speed of sound in water, c, as $\tau_s = a/c \approx 6 \times 10^{-11}$s, which is much shorter than any experimentally relevant timescale, prompting us to use the incompressible Navier-Stokes equations.

For a Newtonian fluid (such as water) of density ρ and viscosity η, these equations describe the evolution of the velocity field $\boldsymbol{u}$ in response to the pressure gradient $\boldsymbol{\nabla} p$ and body forces $\boldsymbol{f}$ acting on the fluid as

$$\rho\left(\frac{\partial \boldsymbol{u}}{\partial t} + \boldsymbol{u}\cdot\boldsymbol{\nabla}\boldsymbol{u}\right) = -\boldsymbol{\nabla} p + \eta\Delta\boldsymbol{u} + \boldsymbol{f}, \tag{1.1}$$

together with the incompressible continuity equation

$$\nabla\cdot\boldsymbol{u} = 0. \tag{1.2}$$

The Navier-Stokes equation (1.1) is famously nonlinear and further simplifications are necessary for almost any problem of practical relevance. The relative importance of the nonlinear momentum advection term compared to the viscous dissipation term is measured by the Reynolds number Re [7]. In quiescent fluid, it can be estimated from the velocity of a colloidal particle V_p as

$$\mathrm{Re} = \frac{|\rho\boldsymbol{u}\cdot\boldsymbol{\nabla}\boldsymbol{u}|}{|\eta\Delta\boldsymbol{u}|} \sim \frac{\rho V_p L}{\eta}. \tag{1.3}$$

Taking $a = 100$ Å and $V_p \approx 1\mu$m/s we get $\mathrm{Re} \sim 10^{-7}$. Thus, we can disregard the nonlinear terms and arrive at the time-dependent Stokes equation

$$\rho\frac{\partial \boldsymbol{u}}{\partial t} = -\boldsymbol{\nabla} p + \eta\Delta\boldsymbol{u} + \boldsymbol{f}. \tag{1.4}$$

By taking the curl of both sides of this equation, we arrive at the vorticity diffusion equation, for the vorticity $\omega = \boldsymbol{\nabla}\times u$

$$\frac{\partial}{\partial t}\boldsymbol{\omega} = \frac{\eta}{\rho}\Delta\boldsymbol{\omega}, \tag{1.5}$$

which has a heat-kernel-type solution with characteristic momentum diffusion time of

$$\tau_\omega = \frac{\rho a^2}{\eta}. \tag{1.6}$$

It turns out that this timescale is of the same order as the Rayleigh particle velocity relaxation timescale τ_B (following the naming convention of van Kampen [9]) describing the exponential decay of

the velocity of a solid particle of radius a slowing down due to Stokes drag $F_{stokes} = 6\pi\eta a V_p = \zeta V_p$ as shown by

$$\tau_B = \frac{M}{\zeta} = \frac{2}{9}\left(\frac{\rho_p}{\rho}\right)\tau_\omega, \tag{1.7}$$

where ρ_p is the density of the colloidal particle, which is often neutrally buoyant.

Finally we can define the diffusive timescale τ_D as the time required for a particle to move distance comparable with its size while diffusing with coefficient of diffusion D

$$\tau_D = \frac{a^2}{D}. \tag{1.8}$$

Thus, for colloidal particles $\tau_D \gg \tau_\omega \sim \tau_B$ and on experimentally relevant time scales $\sim \tau_D$ we can neglect time dependent terms of the Stokes equation (this is not always the case; for example, in the motion of carpets of cilia or bacterial flagella the time dependent terms play a role [10]). Consequently, we arrive at the Stokes equations

$$\eta\Delta\boldsymbol{u} - \boldsymbol{\nabla}p + \boldsymbol{f} = 0, \tag{1.9}$$

$$\nabla \cdot \boldsymbol{u} = 0. \tag{1.10}$$

The Stokes equation (1.9) possesses important properties of linearity and instantaneous information propagation throughout the domain. This means that the velocity at any moment is fully determined by the fluid boundary conditions at the same time instance. These properties are vital for the construction of mobility matrices discussed in the next section.

1.2 Mobility matrices

An immediate corollary of the linearity of the Stokes equation is the existence of so-called mobility matrices – linear operators that provide particle velocities resulting from forces and torques acting within a given configuration of colloidal particles. Unfortunately, closed-form expressions for these relationships exist only in very simple geometries where flows can be easily computed, such as for a single sphere or ellipsoid in a bulk fluid [11]. Even for highly symmetric shapes, such as a torus [12, 13], solutions are, at best, expressed in terms of a series expansion. We can leverage the very simple solution for flow around a sphere and construct an approximate solution for flow around multiple spheres. In this manner, we can model many bodies of interest in a modular fashion by approximating them as a collection of small spheres, connected either rigidly or through appropriate potentials. This construction is facilitated by the method of mobility matrices, which we describe below.

Consider N spherical particles each of radius a located at $\boldsymbol{R}_i$ for $i \in (1 \dots N)$ moving with velocities $\boldsymbol{V}_i$ and angular velocities $\boldsymbol{\Omega}_i$ under the influcence of external forces $\boldsymbol{F}_i$ and torques $\boldsymbol{T}_i$. We seek the matrix connecting these velocties to these forces. For convenience this section uses supervector notation, here following convention of [8]. Let us denote a concatenated vector of particle velocities as $\boldsymbol{V} = (\boldsymbol{V}_1, \boldsymbol{V}_2, \dots, \boldsymbol{V}_N)^T$, angular velocities and $\boldsymbol{\Omega} = (\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \dots, \boldsymbol{\Omega}_N)^T$ and analogously for forces $\boldsymbol{F}_i$ and torques $\boldsymbol{T}_i$

From the linearity of the Stokes equation and the no-slip boundary conditions of the fluid velocity at the surface of the colloids we know that they obey a linear relationship. We can introduce mobility tensors $\boldsymbol{\mu}$ in the following fashion

$$\begin{pmatrix}\boldsymbol{V}\\ \boldsymbol{\Omega}\end{pmatrix} = -\begin{pmatrix}\boldsymbol{\mu}^{tt} & \boldsymbol{\mu}^{tr}\\ \boldsymbol{\mu}^{rt} & \boldsymbol{\mu}^{rr}\end{pmatrix} \cdot \begin{pmatrix}\boldsymbol{F}\\ \boldsymbol{T}\end{pmatrix}. \tag{1.11}$$

Often one is interested in a torque free case where only submatrix $\boldsymbol{\mu}^{tt}$ plays a role and is referred to as translational mobility matrix or grand translation mobility matrix (to emphasise that it captures particle-particle interactions).

For a point force $\boldsymbol{f}$ acting on the fluid Green's function of the Stokes equation is called the Oseen tensor $\mathbf{T}^0$ with fluid velocity satistying $\boldsymbol{u}(\boldsymbol{r}) = \mathbf{T}^0(\boldsymbol{r}).\boldsymbol{f}$. Said tensor is given by

$$\mathbf{T}^0(\boldsymbol{r}) = \frac{1}{8\pi\eta|\boldsymbol{r}|}\left(\mathbf{1} + \hat{\boldsymbol{r}}\hat{\boldsymbol{r}}\right). \tag{1.12}$$

For multiple spherical colloids suspended in a quiescent fluid, the mobility matrices depend on the relative positions of the spheres. In the lowest order approximation, it turns out that only pairwise

displacements $\boldsymbol{R}_{ij} = \boldsymbol{R}_i - \boldsymbol{R}_j$ are required to compute them. It may be tempting to simply evaluate the Oseen tensor at the centers of the spheres and use this as a mobility matrix; unfortunately, this approach can lead to a matrix that is not positive definite, which physically corresponds to negative dissipation of energy under some forcing. As a result, this approach is not just inaccurate but completely fruitless in the case of Brownian Dynamics where square roots of mobility matrices are required at each simulation step [8, 14]. Most methods for computing square roots (such as the Cholesky decomposition) are singular for singular matrices. Thus, even a very small error in the components of the mobility tensor can lead to arbitrarily large errors of the square root, unless positive definiteness can be guaranteed.

Perhaps the simplest positive-definite approximation can be constructed in the following manner. Introducing $R_{ij} = |\boldsymbol{R}_{ij}|$, we can determine the velocity of i^{th} sphere exactly from the surface tractions integral by combining Faxen's law [15] with the Oseen tensor

$$\boldsymbol{V_i} = -\mu_0^t \boldsymbol{F}_i + \sum_{j\neq i}^{N} \left(1 + \frac{a^2}{6}\Delta_i\right) \int_{S_j} \mathrm{d}S' \mathbf{T}^0(\boldsymbol{R}_i - \boldsymbol{r'}) \cdot \boldsymbol{f}^{(s)}(\boldsymbol{r'}). \tag{1.13}$$

A direct use of equation (1.13) is not very practical, since it requires solving for surface traction distribution. If we only take into account the average surface traction on each sphere, by taking $\boldsymbol{f}_j^{(s)}(\boldsymbol{r'}) = -\boldsymbol{F}_j/(4\pi a^2)$, one obtains a relationship of the form (1.11). Next, one can approximate the integrand by Taylor expanding it around the centres of spheres to second order (with the remainder term $\mathcal{O}((a/R_i)^4)$ because the integrals of all third-order terms vanish), and we get the lowest-order approximation in a for the mobility tensors

$$\boldsymbol{V}_i \approx -\mu_0^t \boldsymbol{F}_i - \sum_{j\neq i}^{N} (1 + \frac{a^2}{6}\Delta_i)(1 + \frac{a^2}{6}\Delta_j)\mathbf{T}^0(\boldsymbol{R}_i - \boldsymbol{R}_j) \cdot \boldsymbol{F_j} \tag{1.14}$$

$$\approx -\mu_0^t \boldsymbol{F}_i - \sum_{j\neq i}^{N} (1 + \frac{a^2}{3}\Delta_x)\mathbf{T}^0(\boldsymbol{x} = \boldsymbol{R}_{ij}) \cdot \boldsymbol{F}_j \tag{1.15}$$

because $\Delta_i \Delta_j \mathbf{T}^0(\boldsymbol{R}_{ij}) = 0$.

Evaluating the Laplacians and rearranging the explicit formulae for mobility tensors can be obtained

$$\boldsymbol{\mu}_{ii}^{tt,RP} = \mu_0^t \mathbf{1}, \tag{1.16}$$

$$\boldsymbol{\mu}_{ij}^{tt,RP} = \mu_0^t \left(\frac{3}{4}\left(\frac{a}{R_{ij}}\right)\left(\mathbf{1} + \hat{\boldsymbol{R}}_{ij}\hat{\boldsymbol{R}}_{\boldsymbol{ij}}\right) + \frac{1}{2}\left(\frac{a}{|R_{ij}|}\right)^3 \left(\mathbf{1} - 3\hat{\boldsymbol{R}}_{ij}\hat{\boldsymbol{R}}_{ij}\right)\right) \quad \text{for} \quad i \neq j. \tag{1.17}$$

The $\boldsymbol{\mu}_{ij}^{tt,RP}$ approximation is the well-known Rotne-Prager approximation [8, 16, 11]. It turns out that this approximation is always positive definite and thus a great candidate for Brownian Dynamics applications. Moreover, it turns out that the tensor divergence of $\boldsymbol{\mu}^{tt,RP}$ is equal to zero (which is not true for the exact solution $\boldsymbol{\mu}^{tt}$) eliminating another term—problematic for Brownian dynamics simulations—in a self-consistent manner[14]. A completely analogous procedure can be applied to the tr and rr parts of the mobility matrix. That and further improvements (such as differentiable continuation for overlaps and extension to collections of spheres of different sizes) are discussed by Zuk, Cichocki, and Szymczak [17]. In this work, we implemented the Rotne-Prager expressions in Python, as the `pygrpy` package.

1.3 Mathematical treatment of Brownian motion

Although the fundamentals of kinetic theory of gases had been established since the 1870s and Navier-Stokes equations since the 1840s, the formulation of the earliest fluctuation-dissipation-type equation, now known as the Stokes-Einstein relationship,

$$D = \frac{k_B T}{6\pi\eta a} \tag{1.18}$$

for a spherical particle of radius a in a fluid of viscosity η, first appeared in the contributions of Einstein [18] and Smoluchowski [19]. Their theoretical insights were experimentally verified by Perrin [20], culminating in his Nobel Prize in Physics in 1926. The conceptual challenge of merging two formerly distinct branches of physics was further complicated by the necessity to develop new mathematical tools

capable of handling random functions describing the position and velocity of Brownian particles. The physical heuristics devised to capture the probabilistic nature of temperature were later formalised into what is now known as stochastic calculus. The applications of stochastic calculus found utility in the economic context, where Merton and Scholes were awarded the Nobel Prize in Economics in 1997 for their development of the Black-Scholes-Merton formula for option pricing. However, the misuse of their results by the financial industry, coupled with inadequate regulation of the derivatives sector, led to the 2008 financial crisis [21].

We now turn to the technical details and outline the central complication of the stochastic calculus—second-order terms in the equivalent of the chain rule—using a physically motivated example. A more succinct description of this domain can also be found in Waszkiewicz et al. [4]. For simplicity, we will again focus on just a single colloidal sphere. In the case of isotropic diffusion in a quiescent fluid the problem is separable with each Cartesian direction being independent, in this simple case we can just consider scalar velocty and displacement. In more complex cases (single sphere near a wall, or more than one sphere, or any other shape) more care is needed because displecements in different directions are no longer independent and often also dependent on angular displacements which complicates the problem significantly. Following the notation of Öttinger [22], we will denote the time dependence in subscripts such that the (time-dependent) velocity of such a sphere is V_t. Historically, V_t was modelled by the Langevin equation

$$M\frac{\mathrm{d}V_t}{\mathrm{d}t} = -\zeta V_t + F_t^B \tag{1.19}$$

with F_t^B being a time dependent force arising from the bombardment of the colloid by the water molecules.

Langevin's heuristic solution of equation (1.19) ignored the unusual stochastic properties of F_t^B, to obtain V_t as a convolution of F_t^B with an appropriate Green's function. Such a convolution is a linear operator acting on (what we hope is) a Gaussian process and thus we should be able to compute variance of the velocity $\langle V_t^2\rangle$, with $\langle\cdot\rangle$ denoting the expected value by means of a double integral

$$\langle V_t^2\rangle = \frac{1}{M^2}\int_0^t \mathrm{d}t' \int_0^t \mathrm{d}t'' \exp\left(-\frac{\zeta}{M}(2t - t' - t'')\right)\langle F_{t'}^B F_{t''}^B\rangle. \tag{1.20}$$

Recall that the Brownian timescale is much longer than the fluid relaxation timescale $\tau_B \gg \tau_s$ and thus we postulate that the Brownian force has a singular correlation structure

$$\langle F_{t'}^B F_{t''}^B\rangle = \alpha_B\delta(t' - t''), \tag{1.21}$$

where δ deontes the Dirac delta functional.

Evaluating the integral (1.20) we get

$$\frac{1}{2}M\langle V_t^2\rangle = \frac{\alpha_B}{4\zeta}\left[1 - \exp\left(-2\frac{\zeta}{M}t\right)\right]. \tag{1.22}$$

By applying equipartition principle to the result (1.22) we arrive at the necessary amplitude of the Brownian fluctuations as

$$\alpha_B = 2k_BT\zeta. \tag{1.23}$$

This is a form of the fluctuation-dissipation theorem [23, 22].

Furthermore, we can heuristically go to the limit of high damping $M/\zeta \to 0$ and obtain an equation for the position of the particle

$$\frac{\mathrm{d}X_t}{\mathrm{d}t} = \frac{1}{\zeta}F_t^B. \tag{1.24}$$

A proper treatment of the equation (1.24) is possible with the help of stochastic differential equations (SDE). First, we define the standard Wiener process W_t as a Gaussian martingale with the following covariance structure

$$\langle W_{t_1}W_{t_2}\rangle = \int_0^{t_1}\mathrm{d}t' \int_0^{t_2}\mathrm{d}t''\delta(t' - t'') = \min(t_1, t_2) \tag{1.25}$$

Note that, this time the covariance is non-singular and the usual theory of Gaussian processes is directly applicable. Heuristically, we expect V_t to be described by the following integral

$$V_t = \frac{\sqrt{2k_bT\zeta}}{M}\int_0^t \exp\left(-\frac{\zeta}{M}(t - t')\right)\mathrm{d}W_{t'}. \tag{1.26}$$

Unfortunately, this integral cannot be performed path-wise because W_t has infinite variation in every interval, even though the issues of almost-surely nowhere differetniability of equation (1.24) are avoided. We need some generalisation of the usual integration to formalise this notion.

These integrals turn out to be formally tractable for a class of processes called non-anticipatory. Any such process (X_t, say) has the property, that for any time t, the future increments of the Wiener process $W_{t'} - W_t$ and past values of the process $X_{t''}$ ($t'' < t < t'$) are independent variables.

We begin the construction of the stochastic integral by considering non-anticipatory random step functions – processes that are constant on finite intervals (denoted with the indicator function $\mathbb{I}$) with step heights given by random variables $\widetilde{X}_i$ by

$$X_t = \sum_{j=1}^{n} \widetilde{X}_{j-1} \mathbb{I}(t \in [t_{j-1}, t_j]). \tag{1.27}$$

For such processes, the stochastic (Ito) integral is simply defined as

$$\int_0^{t_{\max}} X_t \mathrm{d}W_t = \sum_{j=1}^{n} \widetilde{X}_{j-1}(W_{t_j} - W_{t_{j-1}}). \tag{1.28}$$

There is a noticeable lack of symmetry in this expression – we evaluate the integrand at the left end of each interval. As a result, $\widetilde{X}_{j-1}$ and $W_{t_j} - W_{t_{j-1}}$ are independent variables because of the non-anticipatory nature of the process X_t. This 'direction of anticipation' asymmetry plays a central role in the differences between Ito and classical calculus (and is notably absent in Stratonovich calculus).

Thanks to the non-anticipation property, we can immediately conclude two very useful lemmas for random step functions (which are also true for general non-anticipating processes). First, Ito integrals are martingales

$$\left\langle \int_0^{t_{\max}} X_t \mathrm{d}W_t \right\rangle = 0. \tag{1.29}$$

Second, the variance of an Ito integral can be computed with a standard (non stochastic) integral

$$\left\langle \left(\int_0^{t_{\max}} X_t \mathrm{d}W_t \right)^2 \right\rangle = \int_0^{t_{\max}} \left\langle X_t^2 \right\rangle \mathrm{d}t, \tag{1.30}$$

a result dubbed Ito's lemma [24]. To complete the construction of Ito's integral, we need an appropriate limiting procedure whereby a sequence of approximating random-step-function processes is constructed and the original integral is the limit of approximating integrals. It turns out that the correct notion of the limit here is that of the mean square error, and that both the approximating processes and the integral itself converge in that sense.

To show explicitly that this notion of integration is really fundamentally distinct from the usual integration, consider a famous integral $\int_0^t W_{t'} \mathrm{d}W_{t'}$. We can easily construct a sequence of approximating step functions of W_t by a uniform discretisation of a given interval with the mesh size approaching to zero. By a straightforward calculation we obtain a surprising result

$$\int_0^t W_{t'} \mathrm{d}W_{t'} = \frac{1}{2} \left(W_t^2 - t \right). \tag{1.31}$$

The additional term $-t/2$ does not have a classical counterpart.

The fundamental advantage of using Ito's calculus over Langevin's heuristics (apart from being formally sound and the lemmas (1.29) and (1.30)) are the transformation rules of Ito's formula – the stochastic counterpart to chain rule.

Suppose that $\mathrm{d}X_t = A_t \mathrm{d}t + B_t \mathrm{d}W_t$ in the weak sense of Ito integral and $Y_t = f(X_t, t)$. Then we know that

$$\mathrm{d}Y_t = \left(\frac{\partial f}{\partial t} + \frac{\partial f}{\partial x} A_t + \frac{1}{2} \frac{\partial^2 f}{\partial x^2} B_t^2 \right) \mathrm{d}t + \frac{\partial f}{\partial x} B_t \mathrm{d}W_t. \tag{1.32}$$

This equation allows for solution of the previous integral (1.31)

$$\mathrm{d}(W_t)^2 = 2 W_t \mathrm{d}W_t + \mathrm{d}t, \tag{1.33}$$

but more importantly for the physics applications, it gives the change of coordinates rules which are vital when trying to take advantage of symmetries of studied systems.

Summarising, we can see that integral (1.31) is not only of academic interest. Equation (1.33) shows that this behavior exactly describes the square displacements, which play a central role in determining the diffusion coefficient. Given an SDE describing trajectories of Brownian motion, we can proceed by generating sample trajectories and attempt to estimate the diffusion coefficient from them. For a very simple case of a spherical colloid in a quiescent fluid, estimating diffusion directly from the mean square displacement works fine. However, for a molecule of arbitrary shape or an elastic one that changes shape while diffusing, the convergence of this method is impracticably slow. Better approaches to estimating the effective size are required, as outlined in the next section.

1.4 The hydrodynamic radius

Experiments such as Analytical Ultracentrifugation (AUC) observe macroscopic changes in the concentration field ϕ which evolves due to sedimentation and diffusion. In this Section, we discuss the notion of hydrodynamic radius which emerges naturally in the interpretation of experimental data, such as AUC but also in scattering techniques (FCS, DLS, SAXS), which we discuss later.

In the simplest case of isotropic diffusion, the concentration field evolves according to Fick's equation with macroscopic diffusion coefficient D

$$\frac{\partial}{\partial t}\phi = \nabla \cdot (D\nabla\phi). \tag{1.34}$$

In the case of dilute suspensions, the macroscopic D can be identified with the self-diffusion coefficient of a single colloid. For a suspension of microscopic solid spheres of size a the celebrated Stokes-Einstein relation (1.18) gives D in terms of viscosity and temperature, which we repeat below

$$D = \frac{k_B T}{6\pi\eta a}. \tag{1.35}$$

Ceteris paribus, the relationship (1.35) captures all relevant properties of the fluid in which colloids are suspended, which in biochemical context is referred to as the buffer. We can invert equation (1.35) to define hydrodynamic radius R_h for an arbitrary colloid as

$$R_h := \frac{k_B T}{6\pi\eta D}. \tag{1.36}$$

We define R_h as the size of a microscopic solid sphere with the same diffusion coefficient as the studied macromolecule. Since R_h is derived from the apparent diffusion coefficient D, it is influenced by surface effects such as hydration layers. On the other hand, in simple cases, R_h can be treated as a property of the molecule alone, disregarding colloid-solvent cohesion, and thus it can be computed in a more convenient way.

The diffusion coefficient can be in principle estimated directly from the sample trajectories generated from the SDEs knowing that (in 3 dimensions) the mean square of the displacement of a point on a molecule $\boldsymbol{R}(t) - \boldsymbol{R}(0)$ (for example one of its atoms) satisfies the asymptotic relationship

$$\left\langle \|\boldsymbol{R}(t) - \boldsymbol{R}(0)\|^2 \right\rangle \sim 6Dt, \qquad t \to \infty. \tag{1.37}$$

Many software implementations try to estimate D from short trajectories using the instantaneous diffusion coefficient D_s only defined by the limit

$$\left\langle \|\boldsymbol{R}(t) - \boldsymbol{R}(0)\|^2 \right\rangle \sim 6D_s t, \qquad t \to 0. \tag{1.38}$$

Considerations of the memory function show that $D_s > D$ for any tracking point [25], thus picking a tracking point with the minimal D_s reduces the estimation error (in some cases to zero).

For a rigid macromolecule, R_h can be calculated from the trace of the mobility matrix by transforming it to track the diffusion centre—a point which minimizes the short-time diffusion coefficient D_s [17]. If another reference point is chosen rotational motion contributes to the instantaneous diffusion coefficient leading to an overestimate of long term diffusion coefficient [26]. Note that the diffusion centre does not necessarily coincide with the centre of mass of the molecule.

For a rigid macromolecule modelled as a conglomerate of spherical beads, we can derive the hydrodynamic radius from its grand mobility matrix by considering the collective translational and rotational

motion of constituent beads around any given point [27]. Forces acting on individual beads can be computed by first deriving the grand friction matrix from grand mobility matrix, then imposing rigid body motion and then computing the total force and torque required for such motion, thus obtaining the conglomerate friction matrix. The conglomerate friction matrix is then inverted to obtain the conglomerate mobility matrix, which is then used to find the diffusion centre and, finally, conglomerate mobility matrix centred at that point as described in detail in Cichocki et al. [25].

An alternative method of computation of R_h is based on a heuristic observation that the trace of the Oseen tensor satisfies the Laplace equation and Monte Carlo methods of solution of Laplace's equation are then applied to solve a heat equation with constant temperature difference between the surface of the molecule and the ambient fluid. Heat flux is then used to estimate the effective size of such a molecule, as described by the authors of the Zeno package [28], which we used in Waszkiewicz et al. [3].

For an elastic macromolecule, there are more bounded degrees of freedom apart from the rotation. Deformation can also affect the instantaneous diffusion coefficient and a judicious choice of the tracking point can remove much of the spurious overestimation which comes from the approximation of long-time diffusion coefficient D with the instantaneous diffusion coefficient D_s.

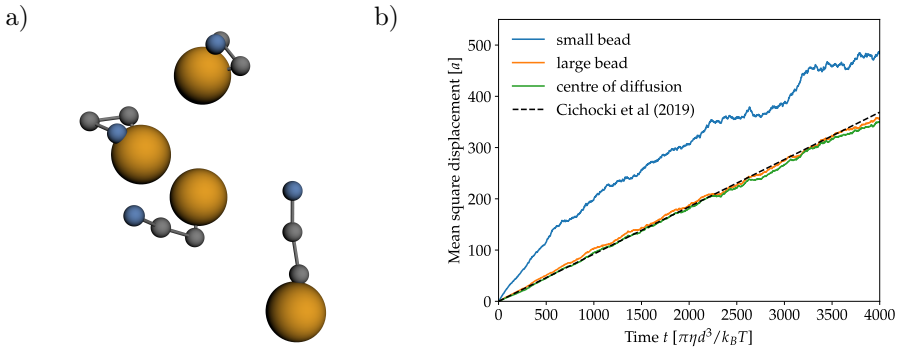


Figure 1.1: a) Representative configurations of an elastic molecule taken from a Brownian Dynamics simulation. b) Mean square displacement of three different tracked locations: the small bead at the end of the chain (blue), the large bead at the other end (orange) and the weighted average of all four beads compared with similar simulation from Cichocki et al. [25] (dashed line). After a long enough time the slope of the mean square displacement is the same regardless of the point tracked. Graphics CC BY 4.0[4]

We now outline the basic idea of estimation of R_h proposed by Cichocki et al. [25]. The crucial step is the choice of a tracking point of a particle. An intuitive first guess is to track 'the middle' or the geometric average of the constituent parts of the molecule, whereas a slightly more refined approach is to use the centre of mass – this is clearly not an optimal strategy, as shown by the exact result for the rigid molecule. Location, or more specifically, the weights in the weighted average, have to be derived from the hydrodynamic properties rather than from the hydrodynamically irrelevant mass. The need for selection of a correct tracking point is highlighted by 1.1, here the short term diffusion coefficient corresponds to the initial slope of the mean square displacement.

Treating the translational mobility $\boldsymbol{\mu}^{tt}$ as a tensor (with indices space, space, bead, bead) we can define an inverse matrix of per-particle ensemble-average traces of mobility b as

$$\mathbf{b}_{ij}\left(\left\langle \boldsymbol{\mu}^{tt}\right\rangle\right)_{jkll} = \boldsymbol{\delta}_{ik} \tag{1.39}$$

where $\langle\cdot\rangle$ denotes ensemble average. Then the hydrodynamic radius is estimated by

$$R_h \approx \frac{1}{2\pi\eta}\left(\sum_i\sum_j \mathbf{b}_{ij}\right)^{-1}. \tag{1.40}$$

We called this approach the minimum dissipation approximation (MDA) as explained in Waszkiewicz et al. [5]. A detailed derivation of the MDA method can be found in Cichocki et al. [25].

Clearly, the MDA method (and earlier, simpler and popular Kirkwood-Riseman method [29]) require estimating the ensemble averaged translational-grand-mobility matrix. Given the Rotne-Prager approximations of Eqs. (1.17), we require only the relative positions of the constituent elements of the

molecule and their effective sizes. These should be drawn from the equilibrium distribution given by the Boltzmannian

$$\mathrm{d}p \propto \exp\left(-\frac{U}{k_B T}\right) \mathrm{d}V. \tag{1.41}$$

There are two potential difficulties with computing this measure: determining the potential energy U and determining the volume element dV. We discuss both of them in greater detail in Waszkiewicz and Lisicki [6]. Both of these problems are more pronounced in coarse-grained models when very stiff springs are used as models of bonds, leading to superficially paradoxical results such as the 'trimer paradox' discussed therein.

The methods discussed above have broad applicability, but now we turn our attention to specific examples of molecules studied in this thesis to highlight their distinct physical properties. The stiffness of different macromolecules varies significantly, spanning many orders of magnitude, which in turn motivates the use of different modelling approaches in different cases. We will outline the methodology leading to these diverse approaches in the following sections.

1.5 Elastic macromolecules

Soft matter as a discipline of physics evolved out of 'colloid suspensions science' under the influence of two 'fuel sources' – bio-relevant measurements showing immediate applicability and a chase of observations of 'universal' validity, with universal being interpreted as insensitive to bio-chemical details of the molecules. A good example of intersection of these two currents are models of long polymeric chains – first elastic macromolecules to be successfully modelled. These include some synthetic plastics but also bio-molecules, such as the DNA or denatured proteins. A phenomenological approach to estimating the properties of coarse-grained models of such polymeric chains started with representing them as chains of N identical monomers, distributed randomly. A celebrated result of Rouse found the hydrodynamic radius $R_h \sim N^{1/2}$ for a Gaussian chain with N elements [30]. It was later improved by Zimm by the inclusion of excluded volume interactions [31] to yield $R_h \sim N^{\gamma}$ with $\gamma = 0.588$. For a more detailed overview the subject matter see Hermann and Gompper [16, chapter 3]. These scalings are valid under the assumption of small persistence length $P = EI/(k_B T)$, which captures the ratio of elastic forces (proportional to the bending rigidity EI, E being the Young's modulus and I the area moment of inerial) to Brownian forces (proportional to fluctuation energy $k_B T$).

In reality, there are more forces influencing the conformation of macromolecules. Their relative importance is easiest to compare using length scale ratios, which quantify the respective ranges of different interactions within the molecule. In our case, the length scales of interest include the already mentioned persistence length, the building block size (quantifying excluded volume interactions), and the Debye length (quantifying electrostatic interactions, scaling with the inverse square root of the ionic strength C_s of the buffer). Some values of these length scales, pertinent to the publications [3] and [5], are outlined in Table 1.1. The need for assessing the relative values for each problem is apparent – there is a factor of 160 difference in stiffness of protein linkers and DNA filaments.

Description	**Scaling**	**IDP [Å]**	**DNA [Å]**
Length	L	2000	1000
Persistence length	$P = EI/k_B T$	3	500
Building block		4	3
Debye length	$R_D \sim (C_s)^{-1/2}$	1	1
Applicable limit		hot	cold(?)

Table 1.1: Relevant length scales for the two studied problems. Length row represents typical values. Debye length computed for relevant (physiological) buffer conditions.

The first type of molecules investigated in this doctoral thesis are DNA minicircles, whose persistence length is comparable to the length of the very short DNA minicircles under investigation by the group of Professor Lynn Zechiedrich at Baylor College of Medicine, described in detail in [3]. Consequently, their conformation is primarily determined by the minimization of elastic forces and is relatively stable.

The second type of molecule investigated in this thesis is intrinsically disordered proteins (IDPs), analysed in [5]. Their persistence length is two orders of magnitude smaller and comparable to the

building block size (the C_α distance between consecutive amino acids along the chain). This characteristic leads to a large conformational variability in an equilibrium ensemble, making both measurement and modelling challenging.

Much of the theoretical enthusiasm in developing complex models of IDPs, with a multitude of interactions, is hampered by an insufficient quantity or quality of experimental calibration data necessary for the determination of molecular "material constants". In polymer science domains where data is plentiful, such as the area of globular proteins that can be readily measured by techniques like X-ray crystallography, significant numerical progress has been made using machine learning techniques, leading to celebrated examples like AlphaFold [32], which has achieved even popular science recognition.

Progress in the domain of IDPs is much more modest, even though they constitute a large proportion of proteins [33] and play an important role in multiple bio-relevant processes, such as the functioning of the COVID-19 virus [34] (to name just one example particularly resonant in 2024). In the opinion of the author this progress is (at least in part) determined by the amount of readily available experimental data; Compare over 200 thousand conformations available in the Protein Data Bank [35] with not even 50 high quality measurements of the diffusion coefficients of the IDPs we were able to extract from published works [5], 118 proteins in the Small Angle Scattering Database [36], or 462 proteins in the Protein Ensemble Database [37].

Simultaneously, a variety of biochemical insights into both folded and disordered protein behaviour invite phenomenological models with many covariates to explain the hydrodynamic size of these molecules. A specific example of such insight could be the relative rigidity of polyhistidine fragments and tags. Their presence (or, more simply, the share of histidines in the totality of amino acids forming a molecule) is sometimes used as an explanatory variable in the process of phenomenological modelling [38, 39].

As data concerning the hydrodynamic properties of IDPs is scarce, and a direct intervention in the protein sequence can be prohibitively expensive, since the synthesis of each new protein construct can take months, we encounter an issue more familiar to social scientists than physicists. The available information essentially forms an observational study where correlation is difficult to separate from causation[1]. Even if we perform all statistical analyses correctly and observe a statistically significant correlation between, for example, the presence of a his-tag and an increase in hydrodynamic size, how can we be sure that it is not simply a case of a confounding cause? One can easily imagine a scenario where compact proteins containing a his-tag are simply harder to synthesize, leading to a distorted sample (selection bias) and forming our phenomenological conclusion based on biased data. Correlations estimated this way can even have the opposite sign to the real causal effect, which is what we are really trying to estimate. A comprehensive and approachable overview of the discrepancies between causal and correlational findings in observational studies can be found in Cinelli, Forney, and Pearl [40]. For a more introductory exposition, we refer to Neal [41].

Until and unless we have a very large dataset of hydrodynamic properties of a bias-free, representative subset of IDPs, first-principles theoretical models should take precedence over phenomenological models. This is not just because of their precision, as demonstrated in Waszkiewicz et al. [5], but also because they correctly model the causal dependency between the covariates of interest and hydrodynamic size.

1.6 Diffusion coefficients in hot and cold limits

First-principles models of elastic molecules come in two flavors: atomistic and coarse-grained. Atomistic methods, as outlined by Karplus and Petsko [42], can be utilized, albeit with certain challenges. These methods either involve the explicit simulation of surrounding water molecules, demanding significant computational resources, or employ an implicit solvent scheme, both of which pose notable numerical difficulties [43]. Even if simulations are theoretically feasible, generating 10-100 millisecond long trajectories for the direct computation of long-time diffusion coefficients, using this method requires great computational power.

On the other hand, coarse-grained methods utilize larger units, typically amino acid residues for proteins, or base pairs for the DNA, as building blocks for conformation prediction schemes. These units are integrated with a separate hydrodynamic model to predict diffusion coefficients.

Addressing the problem directly, even within a coarse-grained perspective, still presents numerical challenges. Furthermore, the identification of a minimal model capable of reproducing experimentally

[1] This is in contrast to a typical physics experiment, akin to a randomized controlled trial in a social setting, where the experimenter is free to choose the value of the control variable and measure the outcome. In the example study of his-tag influence, we cannot simply randomly separate proteins into two groups and add or remove his-tags as required.

observed variations in diffusion coefficients holds an intrinsic value, offering a direct interpretation of observed variability through a restricted set of molecular mechanisms. In this context, two different minimal models, termed the "cold limit" and "hot limit," (inspired by the naming conventions such as [44, 45]) were employed to simulate the conformations of two classes of molecules: DNA minicircles and Intrinsically Disordered Proteins (IDPs).

A comparison of persistence length to the minicircle length and a visual examination of CryoEM figures from Irobalieva et al. [46], alongside equilibrium configurations of twisted beams as described by Coleman and Swigon [47] suggested that thermal fluctuations might exert only a minor influence on the overall behaviour of these molecules. Consequently, the problem was modelled in the "cold limit" (cf. Table 1.2), where thermal fluctuations were neglected, and only equilibrium configurations were computed, akin to operating at negligible absolute temperature. Subsequently, these configurations were treated as rigid bodies when determining the hydrodynamic radius, as detailed in Waszkiewicz et al. [3]. The influence of thermal fluctuations was further investigated using the `pychastic` package [4] which we developed post-publication.

The second type of molecules investigated in this study was the intrinsically disordered proteins (IDPs). These molecules exhibit much smaller stiffness, with a persistence length as small as 3 Å and an elastohydrodynamic length scale ($8\pi^3 EI/(L^2 g\rho)$) of approximately 2000 Å, comparable to the molecule's length. This characteristic further justifies our initial concern regarding the potential influence of buckling forces arising in sedimentation on the hydrodynamic sizes in AUC measurements. Fortunately, the group led by Prof. Anna Niedzwiecka at the Institute of Physics, Polish Academy of Sciences, who provided experimental data for our joint publication [5], employs fluorescence correlation spectroscopy (FCS) instead of AUC, which does not introduce large force gradients on the molecule. Given that the persistence length is comparable to the building block size, we decided to focus on the *hot limit* of the problem, where elastic forces are disregarded relative to thermal fluctuations. However, the excluded volume interactions do not vanish at high temperatures; thus, this approach can be treated as a limit $T \to \infty$ approximation, hence the name. More precisely, bending forces arising from the distribution of Ramachandran angles were neglected, while the forces governing bond length were taken to be infinitely strong. This approach requires care when handling bond-angle distributions, and some apparently paradoxical results can arise, as discussed in Waszkiewicz and Lisicki [6]. We show there that the intuitive distribution 'uniform on a sphere' indeed arises as a limit in the case of linear filaments (notably, a different distribution arises for molecules with loops as shown therein).

Manuscript	**Elasticity**	**Thermodynamics**	**Hydrodynamics**
DNA [3, 2]	✓	.	✓
IDP [5, 6]	.	✓	✓
Pychastic [4]	✓	✓	✓

Table 1.2: Three domains of interest for the study of elastic molecules. Very high and very low stiffness of DNA and IDPs respectively allowed us to use simplified models. The numerical package Pychastic provided an additional tool to explore intermediate regimes where all three factors have to be incorporated.

The two experimentally inspired problems, even though solvable to a satisfactory degree within the presented approximations, leave a desire to assess the size of the error introduced by these particular simplifications. These error estimates can be compared either to the precision of the experimental data or to the estimates of other approximation errors in the model, such as approximations of the hydrodynamic mobility tensors. This can guide the effort in further improvements to the numerical method; For example, should we first work on including thermal fluctuations to the conformations or rather improve hydrodynamic mobility tensor approximations?

When the mobility tensors are simply modeled with the Rotne-Prager approximation (such as the tensors of the `pygrpy` package), the errors introduced there are of the order of 2% [17]. Assessing errors introduced by approximate conformer generation, whether in the hot ($T \to \infty$) or cold ($T \to 0$) limit, necessitates simulating the ensemble at finite temperatures.

Various methods can be employed for this purpose, but a simulation grounded in physical principles holds particular appeal. To achieve good performance, the Brownian Dynamics method was selected. This approach relies on formulating the dynamical equation governing conformational changes in the form of an SDE.

Surprisingly, in 2022, the authors found that the only high-quality SDE integration package available

was `differentialequations.jl` in Julia [48]. Since Julia is still not very popular among physicists (or at least not inside the soft-matter community), this resulted in frequent re-implementation of SDE integration methods for each problem and often re-implementation of the hydrodynamic interaction tensors as well, possibly due to the fact that the Rotne-Prager-Yamakawa tensors most frequently used in our group were originally implemented in Fortran [17].

We addressed this issue by ensuring that the packages `pychastic` (for SDE integration) and `pygrpy` (for Rotne-Prager-Yamakawa tensors) are easily accessible, with single command installation and compatibility across various operating systems including Linux distributions, MacOS, and (as of 2023) even Windows 10 [49]. Additionally, substantial effort was invested in creating straightforward yet realistic examples in the documentation. This not only simplifies the learning curve but also aims to encourage a wider adoption of our integration package. For a comprehensive understanding of implementation and usage, we refer to the detailed information provided in Waszkiewicz et al. [4].

The specific implementation of `pychastic` further streamlines the development of similar models by eliminating the need for explicit force specification. Instead, users can now simply define the energy, and the forces can be derived programmatically using `jax.grad` without compromising the speed and precision. This feature proved particularly advantageous in simulating shear-relaxed dynamics of DNA minicircles in [3], where the energy depends on the non-local quantity of writhe (Wr).

The answer to the correct development direction (in the opinion of the author) lies in future experimental data with resolutions capable of probing inaccuracies of the force-fields used or showing clear dependencies on buffer conditions such as temperature or ionic strength. Such measurements are in progress, and early results provide insights into future DNA models [50].

1.7 Experimental techniques

For the author, the experimental challenges of studying biomacromolecules have been an eye-opening lesson in patience. The excellent work on biosynthesis by the groups at Baylor College of Medicine (Houston, Texas, USA) and the Institute of Physics of the Polish Academy of Sciences (Warsaw, Poland) should not be underestimated. These two collaborations provide an overview of many experimental challenges: from the success or failure of initial biosynthesis, obtaining correct concentrations, to purification from by-products of biosynthesis and the stability of the resultant molecular constructs, to practical obstacles such as replacing discontinued lab equipment and cross-border shipping.

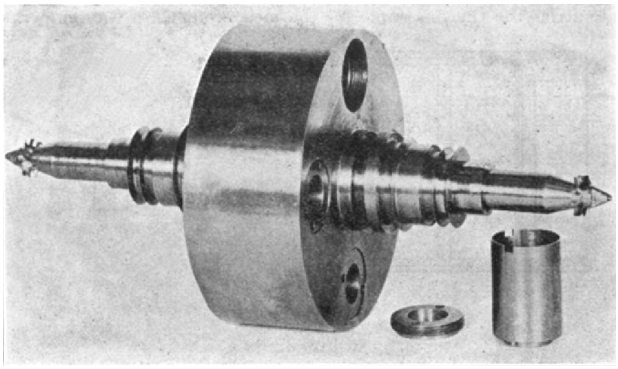
Even in the fortunate scenario, when one obtains a good sample of the molecule to study, the measurement itself is a demanding task. Both AUC and FCS methods rely on time-series analysis of an optical signal. In that sense, they are both indirect methods requiring model fitting within the experimental procedure. This complicates the analysis of direct measurement error (uncertainty of a single measurement), with the current Monte Carlo-based AUC analysis method providing no uncertainty estimates for samples of high purity. However, the single-measurement error is not the only, and not even the leading, source of uncertainty in the measurements of diffusion (and sedimentation) properties of the molecules. Since buffer conditions, concentration, and even the time from synthesis [51] affect the final outcome, these factors must be incorporated into the error analysis to arrive at a comprehensive confidence interval for the measurements. Unfortunately, some authors do not provide any error estimates [52, 53], and most of them do not discuss multiple sources of error.

The following Subsections describe the theoretical underpinnings of the experimental methods relevant to the publications included in this Thesis.

1.7.1 Analytical Ultracentrifugation

Analytical Ultracentrifugation (AUC) is one of the oldest techniques in the domain of colloidal science, having been developed by T. Svedberg in 1923, who received a Nobel prize in 1926 for his contributions to colloid chemistry. In AUC, a colloidal suspension is put inside a rapidly rotating centrifuge to increase the sedimentation rate of the molecules: urrent centrifuges spin with G-forces as large as 10'000G. Inside the centrifuge rotor, a small window allows for optical measurements of either absorption or transmission (in recent models involving multiple wavelengths) which change as a result of concentration variation.

Changes of concentration inside the container are governed by the Lamm equationthat describes the temporal variation of the local concentration field ϕ and that contains the divergence of two currents: diffusive – proportional to concentration gradient and sedimentation – proportional to the local outward

a) 

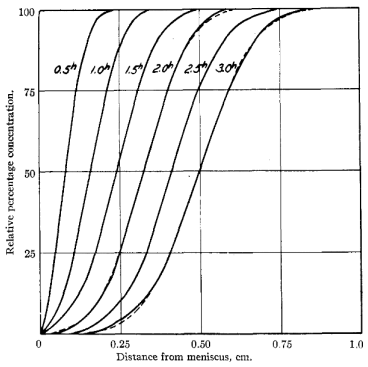
b) 

Figure 1.2: a) Rotor from an ultracentrifuge. b) Haemoglobin concentration profiles from an early AUC experiments with rotor spinning at 42'000rpm or 104'000 G. Profiles derived from photographs taken at 30 minute intervals are compared with Lamm equation predictions (dashed). Reprinted with permission from [54]. Copyright 1927 American Chemical Society.

acceleration

$$\frac{\partial \phi}{\partial c} = \nabla \cdot \left(D \nabla \phi + s\omega^2 \boldsymbol{R} \phi \right), \tag{1.42}$$

where D is the diffusion coefficient, and D is the sedimentation coefficient. Derivation of the values of s and D from experimental, time-dependent concentration profiles is done via a fitting procedure which requires efficient solution schemes for the PDE (1.42) [55, 56]. The finite element method combined with nonlinear least-squares techniques and Monte-Carlo based error estimation gives both values and uncertainties of the values of s and D that can be used to determine other molecular parameters, such as R_h or molecular mass.

1.7.2 Fluorescence Correlation Spectroscopy

The overview of the Fluorescence Correlation Spectroscopy (FCS) technique follows that of Tompson [57] and Gregor and Enderlein [58]. The FCS method, introduced about 50 years after invention of AUC, relies on the temporal analysis of fluorescence signal to determine the properties of the studied sample. These include the hydrodynamic size (relevant to this work), but it is also possible to determine adsorption or reaction kinetics using this method.

In the case of diffusion constant measurements, a very dilute sample of a studied molecule (typically marked with an added fluorophore) is illuminated by a laser inside a confocal microscope. An excited fluorophore then emits light back through the microscope but at a slightly shorter wavelength (due to the Stokes shift) which reaches the detector shielded from laser light with a dichroic mirror (cf Fig. 1.3).

The recorded fluorescence signal $F(t)$ is time dependent, since the number of molecules inside the laser beam changes in time due to diffusion. FCS experiments perform best when at each moment the expected number of excited molecules is close to one giving the largest relative fluctuations of the fluorescence signal.

The frequency of the fluctuations due to diffusion can be quantified (and thus used to derive the diffusion coefficient) by considering the signal autocorrelation function $G(\tau)$ given by

$$G(\tau) = \langle F(t) F(t-\tau) \rangle \tag{1.43}$$

with $\langle \cdot \rangle$ denoting time average. This function is composed of two terms – a constant background (in the case of non-interacting molecules) term due to photons coming from two different molecules, and a delay-dependent term quantifying the probability of detecting a photon from the same molecule again. For simplicity of this overview, we focus only on the last term.

Suppose that a position-dependent probability of excitation describing the shape of the excitation volume can be approximated by an axisymmetric Gaussian profile $U(\boldsymbol{r})$, given by

$$U(\boldsymbol{r}) = \kappa \exp\left(-\frac{2}{a^2}\left(x^2 + y^2\right) - \frac{2}{b^2} z^2 \right), \tag{1.44}$$

with $\boldsymbol{r} = [x, y, z]$ and κ is some overall constant.

If we now consider the Green's function $g(\boldsymbol{\rho}, \tau)$ of the diffusion problem given by

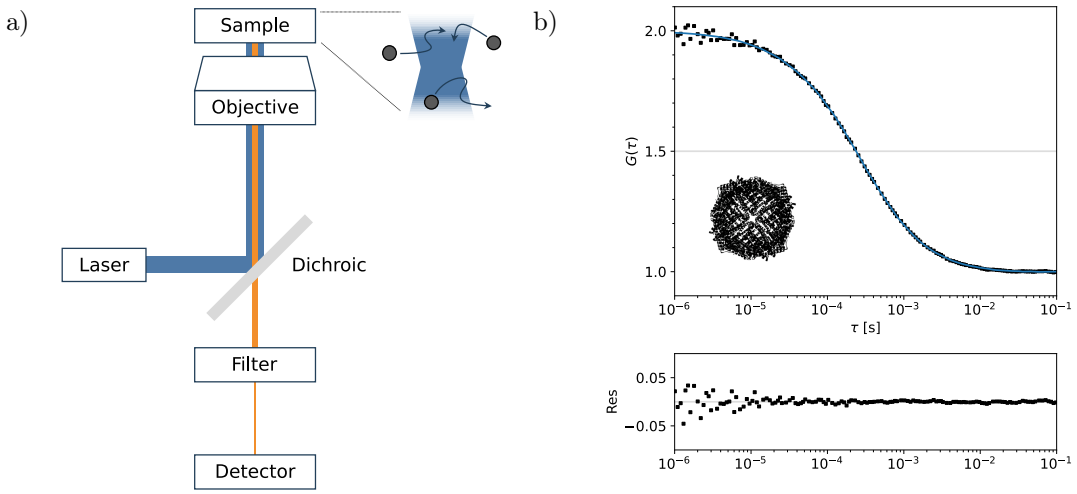


Figure 1.3: a) A typical FCS setup diagram. b) Fluorescence correlation spectroscopy (FCS) data with a fited a model for 3d diffusion (Waszkiewicz et al. [5] CC-BY-NC 4.0)

$$g(\boldsymbol{\rho}, \tau) = \frac{1}{(4\pi D\tau)^{3/2}} \exp\left(-\frac{|\rho|^2}{4D\tau}\right) \tag{1.45}$$

where $\boldsymbol{\rho}$ is the distance, and τ is the time, we can express the autocorrelation function according to

$$G(\tau) = \int\int U(\boldsymbol{r}+\boldsymbol{\rho}) g(\boldsymbol{\rho}, \tau) U(\boldsymbol{r}) \mathrm{d}\boldsymbol{r}\mathrm{d}\rho \tag{1.46}$$

$$= \frac{\pi^{3/2}}{8} \frac{a^2 b}{(1+4D\tau/a^2)\sqrt{1+4D\tau/b^2}}, \tag{1.47}$$

which expresses thae fact that the particle has travelled the distance $\boldsymbol{rho}$ during a time interval τ. Since the parameters a and b cannot be known a priori, one typically fits a simplified expression

$$G(\tau) = G(0)\left(\left(1+\frac{\tau}{\tau_D}\right)\left(1+\gamma\frac{\tau}{\tau_D}\right)^{1/2}\right)^{-1} + G(\infty), \tag{1.48}$$

where γ quantifies the aspect ratio of the excitation volume and τ_D characteristic time of the diffusion. We can obtain the diffusion coefficient D from the residence time by comparing a reference sample of known D and computing a ratio.

1.7.3 Small Angle X-ray Scattering

This section follows the notation of Hermann [59]. Another experimental technique is based on the analysis of the X-ray scattering on the colloidal particles as a function of the beams deflection vector $\boldsymbol{q}$ defined as a difference between incoming wave vector $\boldsymbol{k}_i$ and scattered wave vector $\boldsymbol{k}_s$.

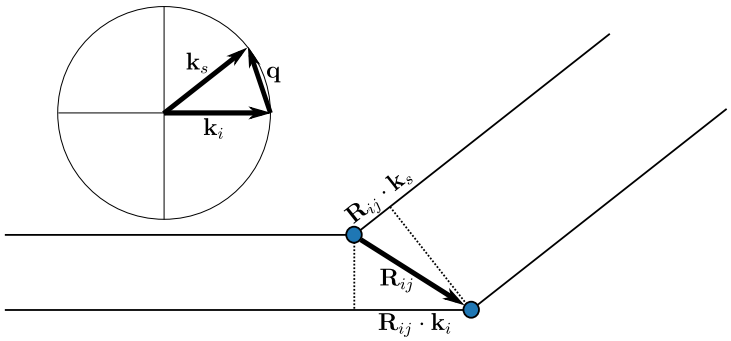


Figure 1.4: Phase shifts in Born approximation.

For colloidal particles, the Born approximation of single-event scattering (as show in Fig. 1.4) is a sufficient description to model experimental data. Within this approximation, the scattering pattern is computed as an interference pattern of the scattered signals from different scattering sites due to phase shift $\Delta\Phi$ resulting from the difference in the length of the optical path

$$\Delta\Phi = \boldsymbol{R}_{ij} \cdot (\boldsymbol{k}_s - \boldsymbol{k}_i) = \boldsymbol{R}_{ij} \cdot \boldsymbol{q}. \tag{1.49}$$

We can express the complex amplitude of the scattered wave $A(\boldsymbol{q})$ as a Fourier transform of the scattering intensity distribution $\rho(\boldsymbol{r})$ as

$$A(q) = A_0 \int \rho(\boldsymbol{r}) \exp(i\boldsymbol{q} \cdot \boldsymbol{r}) \mathrm{d}\boldsymbol{r}. \tag{1.50}$$

Since the observed intensity on the screen I is proportional to the squared absolute value of the wave amplitude $I(\boldsymbol{q}) \propto |A(\boldsymbol{q})|^2$, we can express $I(\boldsymbol{q})$ as a double integral

$$\begin{aligned} |A(\boldsymbol{q})|^2 &= AA^* && (1.51)\\ &= A_0^2 \int\int \rho^*(\boldsymbol{r}')\rho(\boldsymbol{r}'') \exp(i\boldsymbol{q} \cdot (\boldsymbol{r}' - \boldsymbol{r}'')) \mathrm{d}\boldsymbol{r}' \mathrm{d}\boldsymbol{r}''. && (1.52) \end{aligned}$$

If we assume that the scattering sites are point-like, we can express the scattering density distribution as a sum of Dirac distributions δ centred at the locations of scattering sites $\boldsymbol{s}_i$

$$\rho(\boldsymbol{r}) = \sum_i \rho_i \delta(\boldsymbol{r} - \boldsymbol{s}_i), \tag{1.53}$$

where ρ_i gives scattering intensity on site i. The integral (1.52) reduces then to a double sum

$$I(\boldsymbol{q}) = A_0^2 \sum_i \sum_j \rho_i \rho_j \exp(i\boldsymbol{q} \cdot (\boldsymbol{s}_i - \boldsymbol{s}_j)). \tag{1.54}$$

Inside the colloid, the distribution of orientations of the macromolecules with respect to the lab frame is invariant under action of rotations. For a vector $\boldsymbol{v} = \boldsymbol{s}_i - \boldsymbol{s}_j$ this simply implies $\boldsymbol{v}$ distributed uniformly on a sphere and its projection onto the vector $\boldsymbol{q}$ being distributed uniformly on an interval according to the orange slicing theorem[2] giving

$$\begin{aligned} \langle I(\boldsymbol{q}) \rangle &= A_0 \sum_i \sum_j \rho_i \rho_j \int \exp(iqs) \mathbb{I}(|s| < |\boldsymbol{s}_i - \boldsymbol{s}_j|) \mathrm{d}s && (1.55)\\ &= A_0 \sum_i \sum_j \rho_i \rho_j \operatorname{sinc}(|q||\boldsymbol{s}_i - \boldsymbol{s}_j|) && (1.56)\\ &= A_0 \sum_i \sum_j \rho_i \rho_j \operatorname{sinc}(|q||\boldsymbol{R}_{ij}|), && (1.57) \end{aligned}$$

where the sinc function is defined as $\operatorname{sinc} x = \sin x / x$. Thus, the prediction of the scattering intensity in a SAXS measurement of a colloidal suspension reduces to the computation of the distance matrix R_{ij} between the scattering sites, given the values of scattering intensities $\rho_i(q)$. In the case of elastic macromolecules, the scattering signal is simply averaged over an ensemble of possible molecular conformations. With the aid of tabulated values of scattering intensities of each amino acid obtained by Tong, Yang, and Lu [60], a Python package, `saxs_single_bead`, was developed. This package facilitates a simple computation of SAXS profiles in the one-bead-per-residue approximation. It serves as a complement to our ensemble generation schemes operating within this coarse-graining resolution.

[2] the orange slicing theorem: if you slice an orange into slices of equal thickness each slice has the same share of the orange peel

Software packages

Most of the source code utilized in the preparation of this thesis has been made available in public repositories on GitHub. Furthermore, considerable effort was invested in structuring the software into Python packages with structured documentation to facilitate ease of use, as detailed in Table 2.1.

The package `pychastic`, described in detail in Ref. [4], was initially conceived as a tool for generating equilibrium configurations and studying memory effects in the diffusion of macromolecules, an ongoing research project of the group. It is capable of simulating Stochastic Differential Equations, including the equations of Brownian Dynamics. Formulating the full equations of Brownian Dynamics requires the computation of mobility tensors, which can be evaluated using components of the `pygrpy` package. This package can also compute the hydrodynamic size of rigid molecules based on the same grand mobility matrices.

Another package, `glm_mda_diffusion`, relies on `pygrpy` and serves as the Pythonic counterpart to the publication by Waszkiewicz et al. [5]. It can compute the hydrodynamic size of Intrinsically Disordered Proteins (IDPs) using an annotated amino acid sequence from conformers generated by the `sarw-spheres` package. The `sarw-spheres` package generates self-avoiding random walks consisting of tangent spheres of variable size. Its recursive implementation allows for very fast sampling compared to naive sphere-by-sphere randomization.

Finally, the `saxs-single-bead` package provides methods for computing Small-Angle X-ray Scattering (SAXS) profiles from the locations of amino acids in both rigid cases and ensembles (for example, generated using the `sarw-spheres` package).

Package	**Description**	**GitHub**	**Docs**
`pychastic`	SDE solver	[61]	[62]
`pygrpy`	Rotne-Prager mobility tensors	[63]	[64]
`glm_mda_diffusion`	GLM+MDA diffusion calculator	[65]	
`sarw-spheres`	Globule-linker conformer generator	[66]	
`saxs-single-bead`	One site per amino-acid SAXS engine	[67]	[68]
`pywrithe`	Computing writhe of a curve	[69]	[70]

Table 2.1: Software packages developed in the course of preparation of this doctoral Thesis.

Main results of the Thesis

The main results of this Thesis are a series of six research papers, four of which already published in international academic journals. Four papers contain theoretical and numerical results related to the modelling of elastic macromolecules in various settings or the development of practical tools for modelling. The remaining two manuscripts contain results of experiments, along with theoretical models that aim to interpret and rationalise the measured quantities.

Paper I The publication, titled *Hydrodynamic Effects in the Capture of Rod-like Molecules by a Nanopore* [1], co-authored by Maciej Lisicki, provides an analysis of the influence of wall interaction and hydrodynamic anisotropy in the process of capture by a nanopore, relevant to short DNA fragments and a method of their sequencing. A theoretical consideration of a rod-like molecule with uniformly distributed charge provides simple scaling-based criteria for determining when and where inclusion of the wall corrections is required.

Paper II The article, titled *Stability of Sedimenting Flexible Loops* [2], co-authored by Piotr Szymczak and Maciej Lisicki, provides a linear stability analysis of elastic loops within the resistive-force theory framework coupled with elastic forces modeled with Euler-Bernoulli equation. We were able to establish a semi-analytic stability criterion and re-derive the dimensionless quantity governing the buckling instability for this and similar problems.

Paper III The publication, titled *DNA Supercoiling-induced Shapes Alter Minicircle Hydrodynamic Properties* [3], co-authored by Maduni Ranasinghe, Jonathan M. Fogg, Daniel J. Catanese Jr, Maria L. Ekiel-Jeżewska, Maciej Lisicki, Borries Demeler, Lynn Zechiedrich, and Piotr Szymczak, is a result of theoretical-experimental collaboration with the team based at Baylor College of Medicine and Rice University (JMF, DJC, and LZ; experts in biosynthesis, DNA properties, and sample preparation) and the team based at the University of Lethbridge (MR and BD; experts in AUC measurements) with a theoretical collaboration between the Institute of Fundamental Technology Research, Polish Academy of Sciences (MLEJ), and a team at the University of Warsaw (RW, PS, and ML). In this publication, we have determined the impact of negative supercoiling and curvature on the hydrodynamic properties of DNA by subjecting 336 bp and 672 bp DNA minicircles to analytical ultracentrifugation (AUC). We then utilised linear elasticity theory and hydrodynamic calculations to predict the DNA shapes and diffusion coefficients.

Paper IV The publication, titled *Pychastic: Precise Brownian Dynamics using Taylor-Ito Integrators in Python*, co-authored by Maciej Bartczak, Kamil Kolasa, and Maciej Lisicki, is a result of work on the implementation of efficient stochastic differential equations solvers capable of convenient treatment of Brownian Dynamics (BD) problems. By expressing BD equations as Ito integrals, we can leverage the classical methods of truncated Taylor-Ito integrators. As part of the documentation of the `pychastic` package, we show how to deal with common BD obstacles: calculations of the divergence of the mobility tensor in the diffusion equation and discontinuous trajectories encountered when working with dynamics on S^2 and $SO(3)$. With a vectorization-oriented implementation, we have achieved performance comparable to earlier implementations in lower-level programming languages.

Paper V The manuscript, titled *Minimum Dissipation Approximation: A Fast Algorithm for the Prediction of Diffusive Properties of Intrinsically Disordered Proteins*, co-authored by Agnieszka Michaś, Michał K. Białobrzewski, Barbara Klepka, Maja Cieplak-Rotowska, Zuzanna Staszałek, Bogdan Cichocki, Maciej Lisicki, Piotr Szymczak, and Anna Niedźwiecka, is the result of an experimental-theoretical collaboration between the team based at the Institute of Physics

of the Polish Academy of Sciences (AM, MKB, BK, MCR, ZS, and AN; experts in protein biosynthesis and FCS experiments) and the theory team at the University of Warsaw (RW, BC, ML, and PS). In our study, we demonstrate a fast numerical method combining simple conformational sampling and approximate hydrodynamic interactions to estimate the diffusion coefficients of intrinsically disordered proteins (IDPs), even in the presence of structured do- mains, with a precision surpassing the classical Kirkwood-Riseman approximation. With a new collection of diffusion coefficient measurements, we can quantitatively compare our predictions with multiple models present in the literature (such as power-laws and power-laws with sequence-dependent corrections).

Paper VI The manuscript, titled *The trimer paradox: the effect of stiff constraints on equilibrium distributions in overdamped dynamics*, co-authored by Maciej Lisicki, deals with the problem of very stiff constraints and limiting bond-angle distributions arising from those constraints in coarse-grained spring-bead models of macromolecules in the diffusive regime. It shows that even though the solution of the paradox was elucidated as early as 1984 by van Kampen, explicit (and correct) treatment of this limit was missing from well-known books such as those by Frenkel [71]. By a careful treatment of singular distributions, we show that a combination of metric properties of the constraining manifold and the Hessian of the constraining field are required for the correct determination of bond angles, and that 'uniform on a sphere' distributions for harmonic constraining potentials are not universal, with potentially large deviations for small cyclic molecules. These results establish theoretical foundations for the globule-linker model and should guide further work on the minimum dissipation approximation.

The aim of the presented work was twofold: to address the immediate needs of the experimental groups we have been collaborating with, but also to establish robust 'null hypothesis' models, which are easy enough to use and incorporate all of the fundamental interactions required for the modelling of diffusion, but not more. As such, deviations from these models can be used as quantitative indicators of a significant contribution of new physical phenomena (for example, electrostatic interactions or the formation of transient bridges between distant parts of the molecules).

3.1 Paper I: *Hydrodynamic effects in the capture of rod-like molecules by a nanopore*

PAPER I

"*Hydrodynamic effects in the capture of rod-like molecules by a nanopore*"

Radost Waszkiewicz and Maciej Lisicki

COMMENTARY

Translocation of biomolecules through nanopores lies at the heart of many biological processes, such as cell signalling. It is a significant element of modern nanotechnology and sequencing techniques which give access to the structure of DNA or RNA in small quantities. Thus, it enables low-cost genotyping and testing without the need to resort to PCR amplification or chemical labelling. In a relevant experimental setup for DNA analysis, the efficiency of sequencing hinges on two critical factors: the translation speed through the nanopore and the capture radius of the nanopore.

The complex process of translocation is now understood in great detail and has been studied extensively to assess the importance of multiple contributing factors. However, the process of approach and capture by a nanopore still poses a challenge. Existing capture models mostly neglect hydrodynamic interactions, which might play an important role when the DNA fragment is sufficiently close to the nanopore. This work aimed to fill this gap by discerning and examining the time scales associated with different modes of motion – rotational and translational – induced by different types of forces present in the system: Brownian fluctuations, electrostatic attraction to the pore, and hindering effects of hydrodynamic interaction with the wall. By the analysis of different regimes, we successfully identified the pore distance at which wall interaction terms exert a dominant influence and therefore should be accounted for in quantitative models.

To perform the analysis, we constructed a simple coarse-grained model of an anisotropic, rod-shaped particle, representing the elongated nature of a DNA filament of a length short compared to its persistence length. In this case, a molecule can be treated as a rigid body, for which an approximate form of the mobility tensor in the presence of a wall has been proposed [72]. In this work, we performed scaling analysis to identify the regimes where different types of forces are dominant. We calculated the trajectories of a model nanorod near a wall in Mathematica, incorporating the near-wall corrected mobility tensor, which takes into account both the anisotropy of the particle itself, and its coupling to the wall. We then formulated criteria as to where the inclusion of hydrodynamic interactions with the bounding surface could prove beneficial for quantitative modelling of the capture process.

In this study, the PhD candidate: co-developed the scaling analysis and the theoretical description, implemented the near-wall corrections to the mobility tensors and generated all numerical results and visualisations presented, participated in the analysis of results, prepared all figures, wrote the first draft and edited all subsequent versions.

IOP Publishing Journal of Physics: Condensed Matter
J. Phys.: Condens. Matter **33** (2021) 104005 (8pp) https://doi.org/10.1088/1361-648X/abd11b

Hydrodynamic effects in the capture of rod-like molecules by a nanopore

Radost Waszkiewicz and Maciej Lisicki*

Institute of Theoretical Physics, Faculty of Physics, University of Warsaw, Poland

E-mail: mklis@fuw.edu.pl and rwaszkiewicz@fuw.edu.pl

Received 30 September 2020, revised 30 November 2020
Accepted for publication 7 December 2020
Published 23 December 2020

Abstract

In the approach of biomolecules to a nanopore, it is essential to capture the effects of hydrodynamic anisotropy of the molecules and the near-wall hydrodynamic interactions which hinder their diffusion. We present a detailed theoretical analysis of the behaviour of a rod-like molecule attracted electrostatically by a charged nanopore. We first estimate the time scales corresponding to Brownian and electrostatic translations and reorientation. We find that Brownian motion becomes negligible at distances within the pore capture radius, and numerically determine the trajectories of the nano-rod in this region to explore the effects of anisotropic mobility. This allows us to determine the range of directions from the pore in which hydrodynamic interactions with the boundary shape the approach dynamics and need to be accounted for in detailed modelling.

Keywords: diffusion, nanopore, hydrodynamic interactions, hydrodynamic mobility

(Some figures may appear in colour only in the online journal)

1. Introduction

Nanopore sequencing is now a well-established technique for the determination of structure of biomolecules [1, 2], such as DNA [3], RNA [4], or proteins [5]. The molecules, which are typically slender filaments, are electrophoretically transported to the nanopore and then translocated through an orifice. The process of passage or threading, controlled by a combination of electric [6, 7], electrokinetic [8], entropic [9], osmotic [10, 11] and mechanical forces [12], is now well understood and explored.

However, the approach to the pore is described in less detail. Available models characterise the dynamics of the molecule by its diffusion coefficient D and electrophoretic mobility μ_e. This simplified approach has proved useful to establish the general properties of the system. In reference [13], Grosberg and Rabin determined the concentration of DNA near the pore using the Smoluchowski equation formalism. Qiao *et al* defined the capture radius of the pore [14], being the distance at which thermal fluctuations become comparable to the electrostatic potential energy, which bounds the region of attraction of the pore. In the following works, they extended this notion by introducing the orientational capture radius [15], being the range at which the electric field strongly orients the colloidal rods. These models, however, neglect the anisotropy of the particles, and of the hydrodynamic interactions with the wall which hinder diffusion at close distances. Hydrodynamic effects are known to alter the trajectories of close sedimenting particles [16, 17] by coupling to their inherent shape anisotropy and lead to a general slow-down of colloidal dynamics [18].

In this contribution, we fill this gap by formulating a detailed theoretical approach which accounts for anisotropic diffusivity of a model nano-rod both due to its non-isotropic shape and due to the particle-wall flow-mediated interactions. We first use this model to determine the time scales corresponding to the subsequent phases of motion of a nano-rod approaching a pore: purely Brownian motion far from the pore, electric field-induced translation and reorientation, and the wall influence region. Then we provide a quantitative insight by solving the equations of motion numerically for a collection of initial positions and orientation. We describe the resulting trajectories and mechanisms shaping the motion. This allows us to determine the wall influence region.

* Author to whom any correspondence should be addressed.

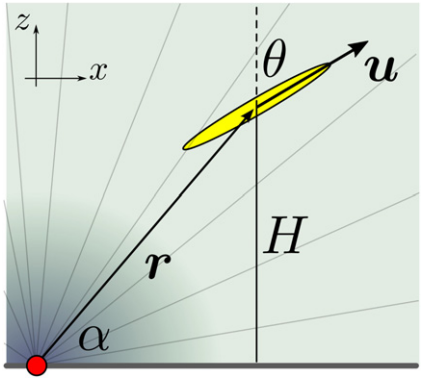


Figure 1. Sketch of a nanorod close to a nanopore denoted by a red dot. The pore generates a radial electric field, modelled by a point charge placed in its location at the origin. The configuration of the rod is given by its position $\boldsymbol{r}$ and orientation $\boldsymbol{u}$, which corresponds to an inclination angle θ. In addition, we denote by α the polar angle at which the rod is seen from the pore.

We focus on microparticles which can be modelled as stiff rods. This is appropriate for biomolecules of length L shorter than their persistence length L_p. Examples of such nano-rods include dsDNA shorter than $L_\mathrm{p} \approx 50$ nm (or ca 150 bp) or *fd*-viruses [19], with $L = 880$ nm and $L_\mathrm{p} = 2.8\,\mu$m, translocating through solid-state pores [20]. For longer molecules, the effects of elasticity and changing conformation can lead to coiling [21] or knotting [12, 22] and must be taken into account for proper modelling.

The paper is organised as follows. First, we describe the general model in section 2.1, specifying the form of electrostatic interactions in section 2.2 and hydrodynamic interactions in section 2.3. The following section 3 presents our results, divided into scaling insights in section 3.1, and a numerical analysis of the trajectories in section 3.2. We summarise our conclusions in section 4.

2. Model

2.1. Motion of the nanorod

We consider an ellipsoidal rod of length L with its centre located at a point $\boldsymbol{r}$, as sketched in figure 1. The aspect ratio of the rod is $p = 10$. The director of the rod is a unit vector $\boldsymbol{u}$. The nanopore is at the centre of a Cartesian lab coordinate system, with the z axis being normal to the wall, defined by the xy plane. Thus, the rod is at a distance $H = \boldsymbol{r} \cdot \boldsymbol{e}_z$ from the wall and its inclination angle θ is determined by $\cos\theta = \boldsymbol{u} \cdot \boldsymbol{e}_z$.

The rod is charged and moving in an electric field generated by the nanopore. The field exerts an electrostatic force, $\boldsymbol{F}_\mathrm{e}$ and torque, $\boldsymbol{T}_\mathrm{e}$, on the rod. On the other hand, the suspending fluid reacts to the motion of the particle by exerting a frictional force, $\boldsymbol{F}_\mathrm{h}$, and torque, $\boldsymbol{T}_\mathrm{h}$. Because the flow is characterised by a small Reynolds number, the lack of inertia yields the following equations of motion of the particle

$$\boldsymbol{F}_\mathrm{e} + \boldsymbol{F}_\mathrm{h} = \boldsymbol{0}, \tag{1}$$

$$\boldsymbol{T}_\mathrm{e} + \boldsymbol{T}_\mathrm{h} = \boldsymbol{0}, \tag{2}$$

which determine the translational and rotational velocity of the particle, $\boldsymbol{V}$ and $\boldsymbol{\Omega}$, respectively. These are then used to evaluate the trajectory and orientation according to

$$\frac{\partial \boldsymbol{r}}{\partial t} = \boldsymbol{V}, \tag{3}$$

$$\frac{\partial \boldsymbol{u}}{\partial t} = \boldsymbol{\Omega} \times \boldsymbol{u}. \tag{4}$$

2.2. Electrostatic interactions with a nanopore

Following previous works [15], we model the interaction between a pore and a rod using the Coulomb potential from a point source. We assume that in the capture process rods are uniformly charged with an effective electrophoretic charge Q [13]. We will measure the strength of the electric field by the capture radius λ_e: a distance where the potential energy $Q\Psi$ of the rod is comparable to the thermal fluctuations, so $Q\Psi(\lambda_\mathrm{e}) = k_\mathrm{B}T$, with k_B being the Boltzmann constant and T the temperature. In this setting, the electric field at a location $\boldsymbol{r}$ can be written as

$$\boldsymbol{E}(\boldsymbol{r}) = -\frac{k_\mathrm{B}T\lambda_\mathrm{e}}{Q}\frac{\hat{\boldsymbol{r}}}{r^2}, \tag{5}$$

where the hat denotes a unit vector. The potential energy of the particle in the electric field is given by an integral along the rod

$$\Psi_\mathrm{e} = \frac{k_\mathrm{B}T\lambda_\mathrm{e}}{L}\int_{-L/2}^{L/2}\frac{\mathrm{d}s}{r(s)}, \tag{6}$$

where $r(s)$ represents distance from the pore to the rod element s. From this we can obtain exact integral expressions for the force $\boldsymbol{F}_\mathrm{e}$ and torque $\boldsymbol{T}_\mathrm{e}$ acting on the rod by appropriate differentiation. Since in the remainder of the paper we will be interested in intermediate particle–pore distances, when $L/r \ll 1$, we expand them and to leading order we find

$$\boldsymbol{F}_\mathrm{e} = -k_\mathrm{B}T\lambda_\mathrm{e}\frac{\hat{\boldsymbol{r}}}{r^2} + \mathcal{O}((L/r)^2), \tag{7}$$

$$\boldsymbol{T}_\mathrm{e} = -\frac{k_\mathrm{B}T\lambda_\mathrm{e}L^2}{4}\frac{(\hat{\boldsymbol{r}}.\boldsymbol{u})(\hat{\boldsymbol{r}}\times\boldsymbol{u})}{r^3} + \mathcal{O}((L/r)^4). \tag{8}$$

The force is radially attracting the rod and falls off as r^{-2}. When rod is oriented at an angle to the direction towards the pore, uneven force distribution generates a torque which reorients the rod, forcing it to point towards the pore. This torque is proportional to the electric field gradient, thus it scales as r^{-3}.

2.3. Near-wall mobility of a rod-like particle

On the colloidal length and time scales, relevant for nanopore experiments, the flow field $\boldsymbol{v}(\boldsymbol{r})$ around a particle is described by the stationary Stokes equations [23]

$$\eta\nabla^2\boldsymbol{v}(\boldsymbol{r}) - \nabla p(\boldsymbol{r}) = -\boldsymbol{f}(\boldsymbol{r}), \qquad \nabla\cdot\boldsymbol{v}(\boldsymbol{r}) = 0, \tag{9}$$

where $\boldsymbol{f}(\boldsymbol{r})$ is the force density the particle exerts on the fluid, and $p(\boldsymbol{r})$ is the modified pressure field. If a particle is moving in a quiescent fluid, the frictional force and torque are linearly

related to its translational and angular velocities, $\boldsymbol{V}$ and $\boldsymbol{\Omega}$, via the friction (or resistance) tensor [23, 24]

$$\begin{pmatrix} \boldsymbol{F}_{\mathrm{h}} \\ \boldsymbol{T}_{\mathrm{h}} \end{pmatrix} = -\begin{pmatrix} \boldsymbol{\zeta}^{\mathrm{tt}} & \boldsymbol{\zeta}^{\mathrm{tr}} \\ \boldsymbol{\zeta}^{\mathrm{rt}} & \boldsymbol{\zeta}^{\mathrm{rr}} \end{pmatrix} \begin{pmatrix} \boldsymbol{V} \\ \boldsymbol{\Omega} \end{pmatrix}. \tag{10}$$

The indices tt and rr above denote the translational and rotational parts, respectively, while the tensors $\boldsymbol{\zeta}^{\mathrm{tr}}$ and $\boldsymbol{\zeta}^{\mathrm{rt}}$ describe the translation–rotation coupling.

If a particle is moving under the action of a known force and torque, a complementary problem is formulated using the mobility tensor $\boldsymbol{\mu}$ which is an inverse of the friction tensor

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}^{\mathrm{tt}} & \boldsymbol{\mu}^{\mathrm{tr}} \\ \boldsymbol{\mu}^{\mathrm{rt}} & \boldsymbol{\mu}^{\mathrm{rr}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\zeta}^{\mathrm{tt}} & \boldsymbol{\zeta}^{\mathrm{tr}} \\ \boldsymbol{\zeta}^{\mathrm{rt}} & \boldsymbol{\zeta}^{\mathrm{rr}} \end{pmatrix}^{-1} = \boldsymbol{\zeta}^{-1}. \tag{11}$$

Finally, the mobility tensor is related to the diffusion matrix $\boldsymbol{D}$ by the fluctuation–dissipation theorem

$$\boldsymbol{D} = k_{\mathrm{B}} T \boldsymbol{\mu}. \tag{12}$$

Thus, the diffusive properties of the particle are completely determined by its hydrodynamic mobility.

For an axisymmetric particle, the friction (and mobility) tensors have a high degree of symmetry. In a bulk system, the configuration of a particle is given by its axial unit vector $\boldsymbol{u}$, and the friction tensor can be explicitly written as

$$\boldsymbol{\zeta}^{\mathrm{tt}} = \zeta^{\mathrm{t}}_{\parallel} \boldsymbol{uu} + \zeta^{\mathrm{t}}_{\perp} (\mathbf{1} - \boldsymbol{uu}), \tag{13}$$

$$\boldsymbol{\zeta}^{\mathrm{rr}} = \zeta^{\mathrm{r}}_{\parallel} \boldsymbol{uu} + \zeta^{\mathrm{r}}_{\perp} (\mathbf{1} - \boldsymbol{uu}), \tag{14}$$

$$\boldsymbol{\zeta}^{\mathrm{tr}} = \boldsymbol{\zeta}^{\mathrm{rt}} = \mathbf{0}, \tag{15}$$

using only four coefficients. For ellipsoids, analytical formulae are available for the bulk diffusion tensor and are listed in appendix A. Otherwise, efficient schemes for the calculation of bulk hydrodynamics properties of macromolecules modelled as collections of beads, such as HYDRO++ [25, 26], GRPY [27], or HYDROMULTIPOLE [24, 28] are also available.

The presence of a confining boundary changes this situation, since the hydrodynamic tensors now depend both on the distance to the boundary, and on the orientation of the nano-rod with respect to the surface. The friction tensors of a near-wall particle, $\boldsymbol{\zeta}_w$, may be written as:

$$\boldsymbol{\zeta}_w = \boldsymbol{\zeta}_0 + \boldsymbol{\Delta\zeta}_w, \tag{16}$$

where $\boldsymbol{\zeta}_0$ is the bulk resistance tensor, and the last term is a wall-induced correction. An analytical leading-order approximation to $\boldsymbol{\Delta\zeta}_w$, with the expansion parameter L/H being the ratio of the size of the particle, L, to the wall–particle distance H, was derived previously by some of us [29]. Earlier works provide the components of the diffusion tensor for very slender filaments close to walls only for particular alignments [30, 31]. The treatment proposed in reference [29] allows for an efficient calculation of the diffusion tensor of a slender rod-like particle for moderate and large wall–particle distances. By inverting $\boldsymbol{\zeta}_w$ from equation (16), we arrive at a convenient approximation to the near-wall mobility $\boldsymbol{\mu}_w = \boldsymbol{\zeta}_w^{-1}$, which will serve as the starting point for present work. The correction terms have the following form

$$\boldsymbol{\Delta\zeta}^{\mathrm{tt}}_w = -\frac{\mathbf{A}_1}{2H} + \frac{\mathbf{A}_2}{(2H)^2} + \mathcal{O}(H^{-3}), \tag{17}$$

$$\boldsymbol{\Delta\zeta}^{\mathrm{tr}}_w = -\frac{\mathbf{B}}{(2H)^2} + \mathcal{O}(H^{-3}), \tag{18}$$

$$\boldsymbol{\Delta\zeta}^{\mathrm{rt}}_w = -\frac{\mathbf{B}^{\mathrm{T}}}{(2H)^2} + \mathcal{O}(H^{-3}), \tag{19}$$

$$\boldsymbol{\Delta\zeta}^{\mathrm{rr}}_w = -\frac{\mathbf{C}}{(2H)^3} + \mathcal{O}(H^{-4}). \tag{20}$$

The tensors $\mathbf{A}_{1,2}$, $\mathbf{B}$, $\mathbf{C}$ above are derived from the multipole expansion of the Blake tensor [32] (Green's function for the wall-bounded geometry) and they depend on the bulk components of the friction tensor of a rod-like particle and its orientation angle θ but not on the wall–particle distance. For completeness, we write them explicitly in appendix B.

Notably, there are other strategies for tackling the problem of near-wall mobility or various levels of accuracy, such as boundary integral equations [33] or finite element method simulations [34]. Approximate bead-model numerical schemes on a similar level of accuracy to the analytical correction above supplemented by lubrication treatment of close configurations have been introduced by Swan and Brady [35]. More accurate multipole expansion approaches have also been developed [36]. However, comparisons of the analytical correction presented above in equations (17)–(20) with accurate bead-model HYDROMULTIPOLE scheme [36] for a rod of aspect ratio $p = 10$ have shown the validity of the correction for distances up to $H/L \sim 1$, provided that the rod is far from touching the wall, in which case lubrication corrections become important. Thus, in this contribution we will use the approximate correction, bearing in mind that the analysed model has been developed for moderate particle-nanopore distances. Here, we restrict our attention to the semi-analytical scheme presented above, since the dynamics occur mostly in the range of wide separations of the particle and the wall.

3. Results

3.1. Scaling analysis

The length of the particle is L. For dsDNA the length of $L = 100$ bp corresponds to ca 34 nm. Another length scale in the problem is the electrostatic capture radius λ_{e}, which at room temperature is in the range of micrometers [15], so $\lambda_{\mathrm{e}}/L \gg 1$.

The interplay between different time scales in the problem determines the ranges in which different mechanisms of capture dominate. We list them all for convenience in table 1. Far enough from the pore, the dynamics are purely Brownian, and the relevant time scale is $\tau_{\mathrm{B}} \sim L^2/\langle D^{\mathrm{t}} \rangle \approx 6\pi\eta L^3/k_{\mathrm{B}}T$, where $\langle D^{\mathrm{t}} \rangle$ is the average diffusion coefficient of the nano-rod. Rotational Brownian motion occurs on the same time scale $1/\langle D^{\mathrm{r}} \rangle \sim 8\pi\eta L^3/k_{\mathrm{B}}T$. In the presence of an electrostatic force (7) driving the translational motion, the velocity scales as $V \sim$

Table 1. Time scales in the system.

Symbol	Scaling	Description
τ	$6\pi\eta L^4/k_\mathrm{B}T\lambda_\mathrm{e}$	Basic time scale
τ_B	$6\pi\eta L^3/k_\mathrm{B}T$	Brownian (diffusive) time
$\tau_\mathrm{e}^\mathrm{t}$	$6\pi\eta L^2r^2/k_\mathrm{B}T\lambda_\mathrm{e}$	Electrostatic translation time
$\tau_\mathrm{e}^\mathrm{r}$	$6\pi\eta Lr^3/k_\mathrm{B}T\lambda_\mathrm{e}$	Electrostatic rotation time
τ_w	$6\pi\eta H^2r^2/k_\mathrm{B}T\lambda_\mathrm{e}$	Wall-induced rotation time

$\langle D^\mathrm{t}\rangle\,\lambda_\mathrm{e}/r^2 \sim k_\mathrm{B}T\lambda_\mathrm{e}/6\pi\eta Lr^2$. Thus the relevant translational time scale is $\tau_\mathrm{e}^\mathrm{t} \sim L/V \approx 6\pi\eta L^2r^2/k_\mathrm{B}T\lambda_\mathrm{e}$. The electrostatic torque falls off quicker with distance, so the relevant rotation time scale derived from it becomes $\tau_\mathrm{e}^\mathrm{r} \approx 6\pi\eta L^2r^3/k_\mathrm{B}T\lambda_\mathrm{e}$. By comparing the Brownian and electrostatic rotational time scale, we recover the scaling for the orientational capture radius defined in reference [15]. We note that in the mobility (or diffusion) matrix $\boldsymbol{\mu}$, the translational elements scale as ηL, the coupling terms as ηL^2, and the rotational terms as ηL^3. This scaling changes close to the wall, where an additional length scale, the wall–particle distance H, comes into play. When the rod comes close to the wall, reorientation due to hydrodynamic interactions with the boundary becomes important, with the relevant time scale τ_w derived from the scaling form of the equation $\boldsymbol{\Omega} = \boldsymbol{\mu}^\mathrm{rt}\cdot\boldsymbol{F}_\mathrm{e} + \boldsymbol{\mu}^\mathrm{rr}\cdot\boldsymbol{T}_\mathrm{e}$. By comparing different time scales we determine four general regimes of motion, sketched in figure 2. The furthest region is Brownian, but closer to the pore when $\tau_\mathrm{B} \sim \tau_\mathrm{e}^\mathrm{t}$, translational motion is driven by electrostatic forces but rotational motion remains Brownian. Moving even closer, electrostatic torque dominates over thermal reorientation. The boundaries between these regions are determined by the radial distance from the pore. However, in the vicinity of the wall, the diffusivity and hence mobility are generally hindered by hydrodynamic interactions with the wall, which become the dominant driving mechanism. We thus conclude that for the analysis of the behaviour of a field-driven nanorod, since we focus on the dynamics in the range $L < r \ll \lambda_\mathrm{e}$, Brownian motion may be neglected, as it would only influence the initial orientations with which the rod would enter the area dominated by electrostatic interactions. It is essential, however, to retain the hydrodynamic anisotropy of the rod, as it influences both the translational and rotational motion in the presence of a strong electric field.

For quantitative calculations, we choose dimensionless units with the basic length L. In the presence of an electric field, the more appropriate time unit is $\tau = \tau_\mathrm{B}L/\lambda_\mathrm{e} = 6\pi\eta L^4/k_\mathrm{B}T\lambda_\mathrm{e}$. For dsDNA mentioned above, this time scale is of the order of 10^{-7} s. With this choice, we can write the force and torque acting on the rod as

$$\boldsymbol{F}_\mathrm{e} = -\frac{\hat{\boldsymbol{r}}}{r^2}, \qquad \boldsymbol{T}_\mathrm{e} = -\frac{(\hat{\boldsymbol{r}}\cdot\boldsymbol{u})(\hat{\boldsymbol{r}}\times\boldsymbol{u})}{r^3}. \tag{21}$$

3.2. Discussion of trajectories

Before focussing on the numerical solutions of the full set of equations, it is informative to consider a very simple case of an anisotropic particle subject to a central force without the wall influence and with no external torque. In this case, the particle's mobility tensor has the form as in equation (13), with two coefficients $\mu_\parallel$, $\mu_\perp$. For very slender rods, we have $\mu_\parallel \approx 2\mu_\perp$.

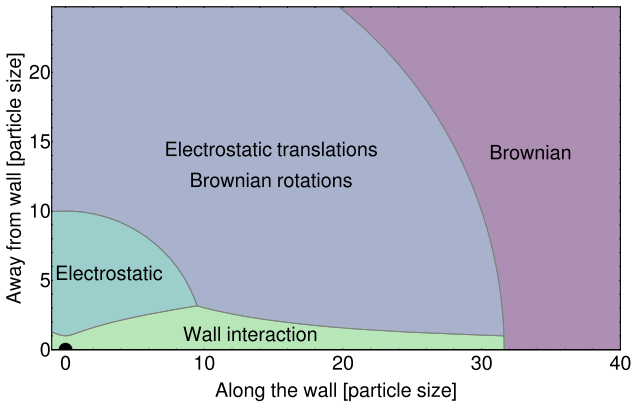


Figure 2. The division of near-pore space into regions coloured by dominant terms determining the dynamics of the rod. The pore is located at the origin, and the wall coincides with the bottom border of the graph. Closest to the surface, wall interaction terms are the most important. When the particle is moving further away from the wall, we expect concentric regions of electrostatically determined dynamics, electrostatic translations with Brownian rotations, and fully Brownian respectively. This corresponds to the intuition that electrostatic torque decays faster than force when moving away from the pore. The boundaries between subsequent regions are obtained by comparing the respective times scales of motion. Expressions used to determine the time scales are collected for convenience in table 1. For this calculation we assumed $\lambda_\mathrm{e} = 10^3L$.

We can describe such a situation by taking a coordinate system centred at the pore, and oriented with the principal axes of the body ($\parallel, \perp$). Within this parametrisation, the equations of motion of the centre of the rod ($r_\parallel, r_\perp$) are

$$\frac{\partial r_\parallel}{\partial t} = \mu_\parallel\frac{F(r)}{r}r_\parallel \tag{22}$$

$$\frac{\partial r_\perp}{\partial t} = \mu_\perp\frac{F(r)}{r}r_\perp. \tag{23}$$

By displaying the equations in this form, it is immediately clear that the function $F(r)$ only influences the time dependence and has no bearing on the trajectory. For bounded F, there is a hyperbolic fixed point at the origin with two important trajectories intersecting at it: $r_\parallel = 0$ and $r_\perp = 0$. The first corresponds to slower, sideways motion and the second to faster, axial motion. We conclude that a particle with different values of the drag coefficients along different axes will almost always approach the fixed point along the slowest axis, which in the case of a rod-like colloids means that the particles would most often approach the pore broadside. This conclusion holds sway whenever the torques acting on the rod are negligible and thus it concisely describes the initial dynamics when starting far enough from the pore. Approaching the slow rectilinear trajectory is accomplished by hydrodynamic 'gliding' effect where velocity stays at an angle to the force in the initial stages of motion. Moreover, we predict the existence of two types of trajectories—concave and convex—depending on whether rod initially points towards or away from the wall compared to the pore direction.

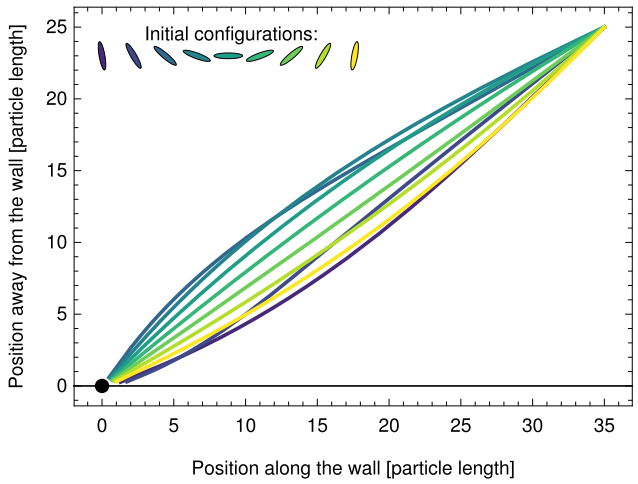


Figure 3. Trajectories of the particle centre depending on the initial orientation of the particle. Due to differences of drag coefficients in the direction along and across the particle, the velocity is never aligned with the drag force. This leads to gliding, either higher above the wall or closer to the wall depending on the orientation, importantly ruling out a straight path towards the pore.

To explore the dynamics driven by the interplay between electrostatic and hydrodynamic forces acting on the nanorod, we integrate the deterministic equations of motion numerically. As argued before, we neglect the influence of Brownian motion in the range of distances under consideration. The equations of motion for the system are

$$\begin{pmatrix} \boldsymbol{V} \\ \boldsymbol{\Omega} \end{pmatrix} = \boldsymbol{\mu}_w \begin{pmatrix} \boldsymbol{F}_\mathrm{e} \\ \boldsymbol{T}_\mathrm{e} \end{pmatrix}, \tag{24}$$

which together with equations (3) and (4) allow us to calculate the trajectories of nano-rods. We consider an ellipsoidal rod of aspect ratio $p = 10$. We integrated the equations of motion using `NDSolve` command of Mathematica 12.1 with `Method -> {"EquationSimplification" -> "Residual"}` option enabled to deal with algebraic-differential nature of the equations.

We present the resulting trajectories in figure 3, starting from a point $\boldsymbol{r}_0 = (35, 25)$ above the wall, thus seen at an angle $\alpha_0 \approx 2\pi/9$ from the nanopore. Rods starting at different orientations glide sideways with respect to the field direction due to their shape (and therefore also mobility) anisotropy. For the initial angle $\theta_0 < \alpha_0$, the trajectories are convex and they approach the vicinity of the pore from below the direction $\alpha_0 = \theta_0$. The initial inclination $\theta_0 > \alpha_0$ leads to a concave path gliding above the $\boldsymbol{r}_0$ direction. The shape of the trajectories is also dependent on the aspect ratio of the nanorod, which determines the parallel and perpendicular friction (mobility) coefficients. At large distances, the wall does not influence the observed dynamics. A closer look into the particles' orientations, sketched in figure 4 for four chosen initial orientations ($\theta_0 = k\pi/36$, with $k = 3, 11, 22, 34$), reveals a strong alignment with the field lines, predicted by earlier works neglecting hydrodynamic anisotropy [15]. The time dependence of the rods' orientation shows a systematic change, the rate of which increases when approaching the pore due to the

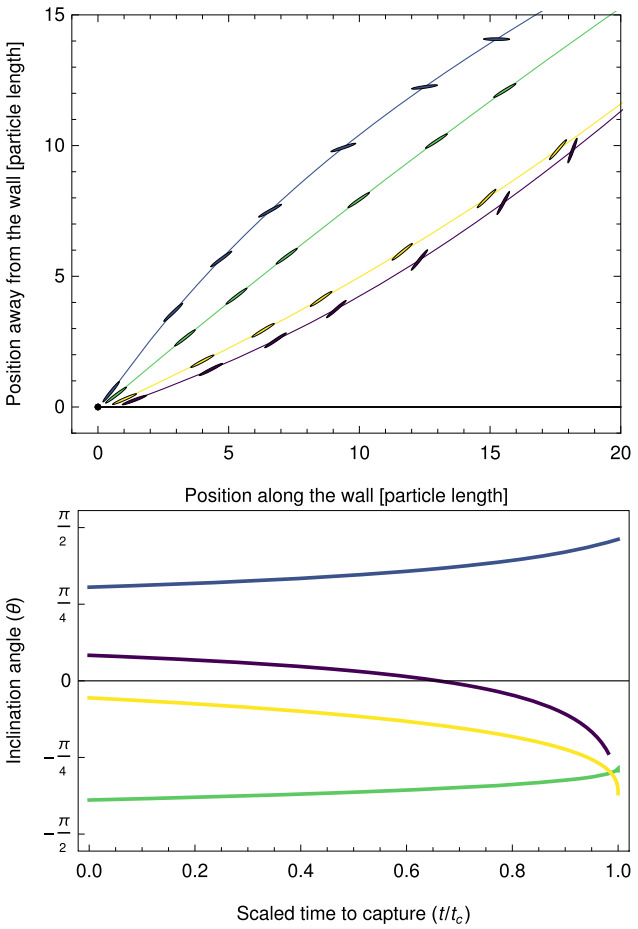


Figure 4. Top: four sample trajectories with the particle orientation sketched at equal translation intervals. The particles start from an initial position (35, 25) at four different initial orientations ($\theta_0 = k\pi/36$, with $k = 3, 11, 22, 34$), depending on which they follow convex or concave gliding trajectories. Bottom: inclination angles corresponding to the trajectories plotted above against normalized time. Reorientation rapidly accelerates in the later stages of motion because the initial offset due to the sideways glide close to the starting point has an increasingly larger effect on the approach angle.

increasingly strong attractive force and aligning torque. At the considered separations from the pore, we see no pronounced effect of the translation–rotation coupling due to the wall, which in this case is too weak to compete with field-driven motion. It would come into significance at near-touching configurations where we also expect the detailed geometry of the pore to matter.

To test how strong the alignment with the field lines is, in figure 5 we present the results of numerical simulations for a spectrum of initial orientation angles θ_0 for rods released far away from the pore at three representative values of the polar angle α_0. For each starting configuration, we determine the final orientation θ_f and plot the final angle to the local field line, $|\alpha_\mathrm{f} - \theta_\mathrm{f}|$, at the final position of the particle. For most initial values of the approach angle, conclusions are similar to those from models neglecting wall interaction terms—the final deviation angle from the field lines direction is typically of the order of $\pi/50$. For small angles α_0, however, the relationship

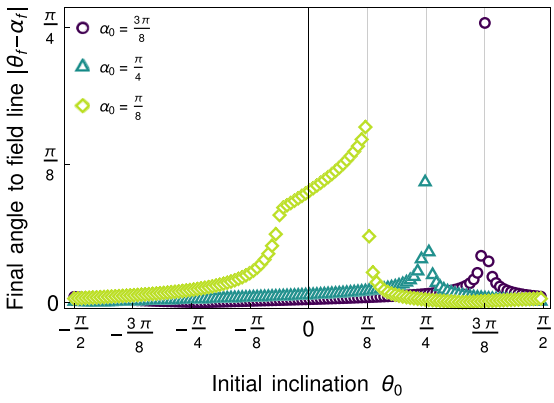


Figure 5. The final angle between the rod and the field line plotted against the initial inclination angle at different starting positions far away from the pore with different polar angles: $\alpha_0 = \pi/8$—close to the wall, $\alpha_0 = \pi/4$—intermediate, $\alpha_0 = 3\pi/8$—far from the wall. Trajectories with $\theta_0 < \alpha_0$ are convex and initially glide towards the wall, and trajectories with $\theta_0 > \alpha_0$ are concave and initially glide away from the wall. Importantly, when the two angles are similar, $\theta_0 \approx \alpha_0$, a straight trajectory is unstable because of drag anisotropy and leads to relatively large values of the final angle between the field lines and the rod, $|\theta - \alpha|$. Additionally, for small α_0 this relationship is asymmetric due to wall influence where the hydrodynamic torque from wall drag competes with electrostatic reorientation in the late stages of movement.

between the initial inclination and the final angle to field lines is asymmetric due to wall interactions, which increase the offset to the field lines. For example, we observe angles greater than $\pi/16$ in the region $(-\pi/8, \pi/8)$. When interpreting this result, it is notable that in 3D under uniform distribution of charged particles over the hemisphere of possible initial directions, small angles α_0 occur much more often (proportionally to $\cos\alpha_0$), meaning that such effect can hinder the capture of a substantial number of particles. The area of the region with $\alpha_0 < \pi/8$ accounts for nearly 40% of the considered hemisphere. Thus it remains important in a high percentage of capture events to properly resolve hydrodynamic interactions with the confining boundary.

Notably, we have chosen the initial orientations to lie within the xz plane. By solving the full, three-dimensional equations of motion (3), (4) and (24), we confirm that the qualitative characteristics of motion remain unchanged even when the initial orientation of the rod has an out-of-plane component. The reorientation in the region far away from the wall remains mostly Brownian, while in the near-pore region the strong alignment mechanism brings the dynamics to a plane to which we restrict our attention from the beginning. We thus do not see any changes as compared to the axisymmetric configurations analysed here.

4. Conclusions

We have presented an analysis of motion of a colloidal nanorod driven by the electric field of a nanopore in a viscous fluid with a particular focus on the inclusion of hydrodynamic interactions and a detailed analysis of the different regimes of motion. The nanopore is modelled as a point charge which attracts a uniformly charged rod-like particle. Basing on scaling arguments, we identified the various time scales of motion and demonstrated that far away from the pore the motion of the particle is purely Brownian but as soon as it reaches the electrostatic capture radius λ_e it is systematically attracted towards the pore. Its initial dynamics are then governed by an electrostatic force driving its translational motion, with the velocity resulting from the balance of this force and the fluid drag force. Since the latter is anisotropic, and depends on the orientation of the particle, the motion resembles sideways gliding towards the pore. At closer distances, the electrostatic torque becomes pronounced and strongly aligns the rods with the electric field lines, as reported previously [15]. However, earlier studies neglected the role of hydrodynamic interactions both on the level of the particle anisotropy, and the wall-induced increase of friction.

In this contribution, we have outlined a theoretical approach which takes into account both the anisotropy of the particle and the wall hindrance effect. Supported by scalings, we explored the trajectories at intermediate distances, when Brownian motion can be neglected, but the rod is far enough from the pore to disregard the field and flow effects of the pore geometry. At large and moderate distances, we have employed an approximate analytical scheme, in which the friction tensor of a colloid close to the wall can be decomposed into its bulk value, encoding the particle anisotropy, and a wall-induced correction. This allowed us to formulate a deterministic system of equations which can be solved for arbitrary initial position and orientation of the nano-rod.

For starting points at a large polar angle α, we find that gliding trajectories are governed by the shape anisotropy of the rod, and the wall plays no significant role. However, for smaller approach angles, hydrodynamic interactions with the wall significantly alter the angle at which the rod approaches the near-pore region. We have shown the extent of this region to be as large as $\pi/8$ which accounts for nearly 40% of spherical area in 3D, thus signifying the importance of wall effects in the proper modelling of dynamics in confined geometry.

Our work shows, basing on a scaling analysis, that it is justified to neglect the role of Brownian motion in the near-pore region. Previous works on a similar system included Brownian motion on the level of translational motion only, by imposing a constant diffusion coefficient of the particle [15]. However, in order to properly resolve the question of Brownian displacements and rotations, one should account for two facts: (a) the non-spherical shape of the rod which makes its diffusion anisotropic even in a bulk system, and (b) the presence of the wall which renders the anisotropic diffusion tensor of the particle a wall–particle distance-dependent and introduces translation–rotation coupling. Even then, we would expect the effect of Brownian motion to be pronounced outside the near-pore region which is dominated by electrostatic interactions. Including these effects would be an interesting direction of future research.

The analysis of trajectories close to the pore remains a separate issue, since its non-planar geometry has a significant influence on the trajectories, both by changing the structure of the electric field, which can no longer be modelled by a

point charge, and by the different character of hydrodynamic interactions, where the pore opening shapes the lubrication flow and friction landscape for the colloid. For an insight into these dynamics, detailed models of the pore structure should be implemented.

Acknowledgments

We thank Piotr Szymczak for helpful discussions. The work of RW and ML was supported by the National Science Centre of Poland Grant Sonata to ML No. 2018/31/D/ST3/02408.

Appendix A. Bulk friction tensors of an ellipsoid

The elements of the bulk friction tensor for an ellipsoidal colloid are known analytically [23]. For a prolate spheroid with a long axis $a = L/2$ and a short axis c, corresponding to an eccentricity $e = \sqrt{a^2 - c^2}/a$, the bulk friction coefficients in $\boldsymbol{\zeta}_0$ can be written as

$$\zeta_{\parallel}^{\mathrm{t}} = (6\pi\eta a)\frac{8}{3}e^3\left[-2e + \left(1+e^2\right)\ell\right]^{-1}, \tag{A.1}$$

$$\zeta_{\perp}^{\mathrm{t}} = (6\pi\eta a)\frac{16}{3}e^3\left[2e + \left(3e^2-1\right)\ell\right]^{-1}, \tag{A.2}$$

$$\zeta_{\parallel}^{\mathrm{r}} = (8\pi\eta a^3)\frac{4}{3}e^3(1-e^2)\left[2e - \left(1-e^2\right)\ell\right]^{-1}, \tag{A.3}$$

$$\zeta_{\perp}^{\mathrm{r}} = (8\pi\eta a^3)\frac{4}{3}e^3(2-e^2)\left[-2e + \left(1+e^2\right)\ell\right]^{-1}, \tag{A.4}$$

$$\zeta^{dr} = (\pi\eta a^3)\frac{16}{3}e^5\left[-2e + \left(1+e^2\right)\ell\right]^{-1}, \tag{A.5}$$

$$\ell = \log\left(\frac{1+e}{1-e}\right). \tag{A.6}$$

The last coefficient, ζ^{dr}, links the stresslet (symmetric dipole moment) on the spheroid with the rate-of-strain tensor of an external flow.

Appendix B. Wall correction terms

The correction terms are described in detail in reference [29]. It is most convenient to specify the components of the near-wall friction tensor in a body-fixed frame of reference. It is defined by a set of basis vectors $\{\boldsymbol{u}, \boldsymbol{u}_{\perp 1}, \boldsymbol{u}_{\perp 2}\}$, where $\boldsymbol{u}$ is the director along the long axis of the nano-rod, $\boldsymbol{u}_{\perp 1} = (\boldsymbol{e}_z \times \boldsymbol{u})/\left|\boldsymbol{e}_z \times \boldsymbol{u}\right|$ is parallel to the wall and perpendicular to the particle axis, and $\boldsymbol{u}_{\perp 2} = \boldsymbol{u}_{\perp 1} \times \boldsymbol{u}$ completes the orthonormal basis. We can write the tensors in equations (17)–(20) explicitly in the body-fixed frame RW. For convenience, we define the following shorthand notation

$$s \equiv \sin\theta, \qquad c \equiv \cos\theta. \tag{B.1}$$

For the translational part (17), we find the correction's angular dependence as

$$\mathbf{A}_1 = -\frac{3}{16\pi\eta}\begin{pmatrix} (\zeta_{\parallel}^{\mathrm{t}})^2(1+c^2) & 0 & -\zeta_{\parallel}^{\mathrm{t}}\zeta_{\perp}^{\mathrm{t}}sc \\ 0 & (\zeta_{\perp}^{\mathrm{t}})^2 & 0 \\ -\zeta_{\parallel}^{\mathrm{t}}\zeta_{\perp}^{\mathrm{t}}sc & 0 & (\zeta_{\perp}^{\mathrm{t}})^2(1+s^2) \end{pmatrix}, \tag{B.2}$$

$$\mathbf{A}_2 = \frac{9}{256\pi^2\eta^2}\begin{pmatrix} A_{\parallel} & 0 & A_{\parallel\perp} \\ 0 & A_{\perp 1} & 0 \\ A_{\parallel\perp} & 0 & A_{\perp 2} \end{pmatrix}. \tag{B.3}$$

with the coefficients

$$A_{\perp 1} = (\zeta_{\perp}^{\mathrm{t}})^3, \tag{B.4}$$

$$A_{\parallel} = (\zeta_{\parallel}^{\mathrm{t}})^3(1+c^2)^2 + \zeta_{\perp}^{\mathrm{t}}(\zeta_{\parallel}^{\mathrm{t}})^2 s^2c^2, \tag{B.5}$$

$$A_{\perp 2} = (\zeta_{\perp}^{\mathrm{t}})^3(1+s^2)^2 + (\zeta_{\perp}^{\mathrm{t}})^2\zeta_{\parallel}^{\mathrm{t}}s^2c^2, \tag{B.6}$$

$$A_{\parallel\perp} = -\zeta_{\parallel}^{\mathrm{t}}\zeta_{\perp}^{\mathrm{t}}[\zeta_{\parallel}^{\mathrm{t}}(1+c^2) + \zeta_{\perp}^{\mathrm{t}}(1+s^2)]sc. \tag{B.7}$$

The translation–rotation coupling part reads

$$\mathbf{B} = \frac{3\zeta^{dr}}{16\pi\eta}\begin{pmatrix} 0 & \zeta_{\parallel}^{\mathrm{t}}(1+c^2)s & 0 \\ 0 & 0 & \zeta_{\perp}^{\mathrm{t}}c \\ 0 & -\zeta_{\perp}^{\mathrm{t}}(1+s^2)c & 0 \end{pmatrix}. \tag{B.8}$$

Finally, the rotational part $\mathbf{C}$ in equation (20) has the form

$$\begin{aligned}\mathbf{C} = &-\frac{1}{16\pi\eta}\begin{pmatrix} (\zeta_{\parallel}^{\mathrm{r}})^2(5-3c^2) & 0 & 3\zeta_{\parallel}^{\mathrm{r}}\zeta_{\perp}^{\mathrm{r}}sc \\ 0 & 5(\zeta_{\perp}^{\mathrm{r}})^2 & 0 \\ 3\zeta_{\parallel}^{\mathrm{r}}\zeta_{\perp}^{\mathrm{r}}sc & 0 & (\zeta_{\perp}^{\mathrm{r}})^2(5-3s^2) \end{pmatrix} \\ &+ \frac{3\zeta^{dr}}{16\pi\eta}\begin{pmatrix} 0 & 0 & \zeta_{\parallel}^{\mathrm{r}}sc \\ 0 & -2\zeta_{\perp}^{\mathrm{r}}(1-2c^2) & 0 \\ \zeta_{\parallel}^{\mathrm{r}}sc & 0 & 2\zeta_{\perp}^{\mathrm{r}}c^2 \end{pmatrix} \\ &- \frac{3(\zeta^{dr})^2}{16\pi\eta}\begin{pmatrix} 0 & 0 & 0 \\ 0 & 3+c^2-c^4 & 0 \\ 0 & 0 & 1+2c^2 \end{pmatrix}. \end{aligned} \tag{B.9}$$

ORCID iDs

Radost Waszkiewicz https://orcid.org/0000-0002-0376-1708

Maciej Lisicki https://orcid.org/0000-0002-6976-0281

References

[1] Wanunu M 2012 Nanopores: a journey towards DNA sequencing *Phys. Life Rev.* **9** 125–58

[2] Branton D *et al* 2008 The potential and challenges of nanopore sequencing *Nat. Biotechnol.* **26** 1146–53

[3] Henrickson S E, Misakian M, Robertson B and Kasianowicz J J 2000 Driven DNA transport into an asymmetric nanometer-scale pore *Phys. Rev. Lett.* **85** 3057–60

[4] Kasianowicz J J, Brandin E, Branton D and Deamer D W 1996 Characterization of individual polynucleotide molecules using a membrane channel *Proc. Natl Acad. Sci.* **93** 13770–3

[5] Pradeep W, Hu R, Prasad B, Yamazaki H, Cressiot B, Zhao Q, Whitford P C and Wanunu M 2017 Nanopore-based measurements of protein size, fluctuations, and conformational changes *ACS Nano* **11** 5706–16

[6] Amit M, Nivon L and Branton D 2001 Voltage-driven DNA translocations through a nanopore *Phys. Rev. Lett.* **86** 3435–8
[7] Forrey C and Muthukumar M 2007 Langevin dynamics simulations of ds-DNA translocation through synthetic nanopores *J. Chem. Phys.* **127** 015102
[8] Ye A and Qian S 2011 Electrokinetic particle translocation through a nanopore *Phys. Chem. Chem. Phys.* **13** 4060–71
[9] Muthukumar M 2010 Theory of capture rate in polymer translocation *J. Chem. Phys.* **132** 195101
[10] Hatlo M M, Panja D and van Roij R 2011 Translocation of DNA molecules through nanopores with salt gradients: the role of osmotic flow *Phys. Rev. Lett.* **107** 068101
[11] Jeon B J and Muthukumar M 2014 Polymer capture by α-hemolysin pore upon salt concentration gradient *J. Chem. Phys.* **140** 015101
[12] Szymczak P 2014 Translocation of knotted proteins through a pore *Eur. Phys. J. Spec. Top.* **223** 1805–12
[13] Grosberg A Y and Rabin Y 2010 DNA capture into a nanopore: interplay of diffusion and electrohydrodynamics *J. Chem. Phys.* **133** 165102
[14] Qiao L, Ignacio M and Slater G W 2019 Voltage-driven translocation: defining a capture radius *J. Chem. Phys.* **151** 244902
[15] Qiao L and Slater G W 2020 Capture of rod-like molecules by a nanopore: defining an 'orientational capture radius' *J. Chem. Phys.* **152** 144902
[16] Russel W B, Hinch E J, Leal L G and Tieffenbruck G 1977 Rods falling near a vertical wall *J. Fluid Mech.* **83** 273
[17] Mitchell W H and Spagnolie S E 2015 Sedimentation of spheroidal bodies near walls in viscous fluids: glancing, reversing, tumbling and sliding *J. Fluid Mech.* **772** 600–29
[18] Happel J and Brenner H 1991 *Low Reynolds Numbers Hydrodynamics* (Dordrecht: Kluwer)
[19] Dogic Z and Fraden S 2001 Development of model colloidal liquid crystals and the kinetics of the isotropic–smectic transition *Phil. Trans. R. Soc.* A **359** 997–1015
[20] McMullen A, de Haan H W, Tang J X and Stein D 2014 Stiff filamentous virus translocations through solid-state nanopores *Nat. Commun.* **5** 4171
[21] de Haan H W and Slater G W 2013 Translocation of 'rod-coil' polymers: probing the structure of single molecules within nanopores *Phys. Rev. Lett.* **110** 048101
[22] Sharma R K, Agrawal I, Liang D, Doyle P S and Garaj S 2019 Complex DNA knots detected with a nanopore sensor *Nat. Commun.* **10** 4473
[23] Kim S and Karrila S J 1991 *Microhydrodynamics: Principles and Selected Applications* (Boston: Butterworth-Heinemann)
[24] Ekiel-Jeżewska M L and Wajnryb E 2009 Precise multipole method for calculating hydrodynamic interactions between spherical particles in the Stokes flow *Theoretical Methods for Micro Scale Viscous Flows* ed F Feuillebois and A Sellier (Kerala: TransWorld Research) pp 127–72
[25] García de la Torre J, del Rio Echenique G and Ortega A 2007 Improved calculation of rotational diffusion and intrinsic viscosity of bead models for macromolecules and nanoparticles *J. Phys. Chem.* B **111** 955–61
[26] Fernandes M X and García de la Torre J 2002 Brownian dynamics simulation of rigid particles of arbitrary shape in external fields *Biophys. J.* **83** 3039–48
[27] Zuk P J, Cichocki B and Szymczak P 2018 GRPY: an accurate bead method for calculation of hydrodynamic properties of rigid biomacromolecules *Biophys. J.* **115** 782–800
[28] Cichocki B, Felderhof B U, Hinsen K, Wajnryb E and Bławzdziewicz J 1994 Friction and mobility of many spheres in Stokes flow *J. Chem. Phys.* **100** 3780–90
[29] Lisicki M, Cichocki B and Wajnryb E 2016 Near-wall diffusion tensor of an axisymmetric colloidal particle *J. Chem. Phys.* **145** 034904
[30] Katz D F, Blake J R and Paveri-Fontana S L 1975 On the movement of slender bodies near plane boundaries at low Reynolds number *J. Fluid Mech.* **72** 529–40
[31] De Mestre N J and Russel W B 1975 Low-Reynolds-number translation of a slender cylinder near a plane wall *J. Eng. Math.* **9** 81–91
[32] Blake J R 1971 A note on the image system for a Stokeslet in a no-slip boundary *Math. Proc. Camb. Phil. Soc.* **70** 303–10
[33] Hsu R and Ganatos P 1989 The motion of a rigid body in viscous fluid bounded by a plane wall *J. Fluid Mech.* **207** 29–72
[34] De Corato M, Greco F, D'Avino G and Maffettone P L 2015 Hydrodynamics and Brownian motions of a spheroid near a rigid wall *J. Chem. Phys.* **142** 194901
[35] Swan J W and Brady J F 2007 Simulation of hydrodynamically interacting particles near a no-slip boundary *Phys. Fluids* **19** 113306
[36] Cichocki B, Jones R B, Kutteh R and Wajnryb E 2000 Friction and mobility for colloidal spheres in Stokes flow near a boundary: the multipole method and applications *J. Chem. Phys.* **112** 2548

3.2 Paper II: *Stability of sedimenting flexible loops*

Paper II

"*Stability of sedimenting flexible loops*"

Radost Waszkiewicz, Piotr Szymczak, and Maciej Lisicki

Commentary

Analytical Ultracentrifugation, a method for the determination of hydrodynamic radius and effective density of molecules, involves applying extremely high centrifugal forces to a colloidal suspension. This well-established technique has been successfully used to measure various colloidal particles, however, its application to flexible macromolecules raises concerns over the influence of large forces (or, more precisely, large force gradients) imposed on the molecule. In extreme cases the high values of compressional tension along the sedimenting particle can lead to buckling during sedimentation. To understand when this might happen, we focused on an idealised problem involving the sedimentation of a circular fibre, aiming to eliminate end-corrections from the buckling consideration.

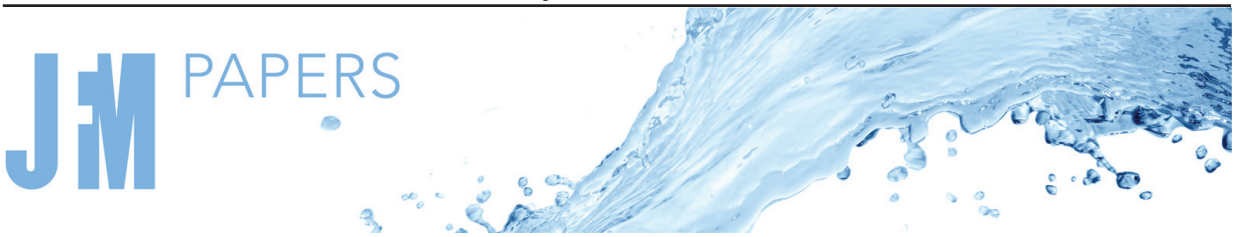
Earlier studies of sedimenting flexible rings using bead-model approach [73] showed complex dynamics with a wide variety of periodic orbits depending on setup parameters and initial conditions. Assessing which parts of this complexity are attributable to hydrodynamic interaction, which to bead-model discretisation and which are intrisic to similar setups was a secondary goal of this investigation.

To determine the cases where buckling may occur, we performed linear stability analysis within the Resistive Force Theory approximation under which hydrodynamic drag is a local quantity. Under this simplification the tension inside the loop at each moment can be computed from an ordinary differential equation.

We calculated the evolution of the conformation of a sedimenting loop using a custom numerical integrator based on a truncated Fourier series. These results were compared with a linear stability analysis of the initial circular configuration, thus providing a near-analytical stability boundary. We obtained good agreement between numerical and analytical approaches and favourable comparison with earlier work.

In this study, the PhD candidate: co-developed the theoretical description, conducted a stability analysis of the shape evolution equation of the fibre and computed the stability criterion using the proposed matrix method, generated all numerical results and visualisations presented, prepared all figures, wrote the first draft and edited all subsequent versions. Additionally they are a corresponding author.

J. Fluid Mech. (2021), *vol.* 919, A14, doi:10.1017/jfm.2021.350



Stability of sedimenting flexible loops

Radost Waszkiewicz[1],†, **Piotr Szymczak**[1] **and Maciej Lisicki**[1],†

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

(Received 8 July 2020; revised 26 March 2021; accepted 20 April 2021)

We study the behaviour of circular flexible loops sedimenting in a viscous fluid by numerical simulations and linear stability analysis. The numerical model involves a local slender-body theory approximation for the flow coupled to the Euler–Bernoulli elastic forces for an inextensible fibre. Starting from an inclined circle, we simulate the dynamics using truncated Fourier modes to observe three distinct regimes of motion: absolute stability, two- and three-dimensional dynamics, depending on the relative importance of the elastic and gravitational forces. We identify the governing parameter and develop a simple semi-analytic stability criterion, which we verify numerically. In all cases, sedimenting loops converge to stable, planar shape equilibria with one free parameter related to the initial conditions and material properties of the fibre.

Key words: slender-body theory, Stokesian dynamics, microfluidics

1. Overview

Biological processes are one of many inspirations of elastohydromechanics (Shelley & Ueda 2000; Schoeller *et al.* 2021). Slender biological objects emerge in multiple contexts, motivating detailed investigation. Starting from the sub-cellular level, examples include DNA and protein folding dynamics (Goldstein & Langer 1995), lipids usually forming cell walls assembling into long filaments (Rudolph, Ratna & Kahn 1991) or microtubules helping healing by contracting wounds (Ehrlich, Grislis & Hunt 1977). Another large area of interest is motility – cells moving inside a fluid environment or cells inducing motion of a fluid. An iconic example of such motion is sperm cells. On closer investigation it turns out that flexibility plays an important role in their motion, and the interplay between elasticity and viscous forces causes changes to the beating pattern in response to the changing environment (Cosentino Lagomarsino, Capuani & Lowe 2003; Fauci & Dillon 2006; Gaffney *et al.* 2011). Biological 'optimisation' for viscosity gradients can also be found in mucus transport inside the lungs where the correct length, stiffness and active deformation of cilia provide the necessary movement of multiple layers of

† Email addresses for correspondence: radost.waszkiewicz@gmail.com, mklis@fuw.edu.pl

fluid with varying viscosity, essential for healthy respiration (Fulford & Blake 1986). On larger length scales, bacterial complexes were observed joining into elongated structures exhibiting a complex dynamics because of elastohydrodynamic effects (Mendelson *et al.* 1995; Goldstein, Powers & Wiggins 1998).

Most of these examples are set in a microscale context, and thus the observed dynamics is dominated by the viscous interactions with the surrounding fluid (Lauga & Powers 2009).

Flexible fibres with free ends have been studied in multiple settings, including their sedimentation, both experimentally (Herzhaft & Guazzelli 1999) and numerically (Li *et al.* 2013). The free-end configuration was investigated first, because methods of producing slender filaments were already developed, and because it is of interest for both industrial applications and biological settings. Further, the one-dimensional structure provides a particularly elegant, treatable and successful way of modelling (Wiggins *et al.* 1998).

In the case of low Reynolds number flows in such settings, elastic elongated filaments can be modelled using various simulation methods, for example: the immersed boundary (IB) method (Peskin 2002), regularised singularity methods (Cortez, Fauci & Medovikov 2005), bead-spring models (Kuei *et al.* 2015; Słowicka, Wajnryb & Ekiel-Jeżewska 2015; Schoeller *et al.* 2021) and discrete and continuous variants of slender-body theory (SBT) (Tornberg & Shelley 2004; Saintillan, Darve & Shaqfeh 2005). The reduced dimensionality of the filament offers a computational advantage, which has been used in variants of the IB technique to study the whirling instability of spinning filaments (Lim & Peskin 2004) and hydrodynamic bundling of bacterial flagella (Lim & Peskin 2012). A combination of SBT and the regularised Stokeslet method has also been formulated by Cortez & Nicholas (2012) and profitably applied e.g. to explain the motion of flagella in dinoflagellates (Hguyen *et al.* 2011). See Nguyen, Cortez & Fauci (2014) for a review of this approach. On the other hand, in methods which treat the filament as fundamentally one-dimensional, such as SBT, one faces problems when the mesh along the filament is too fine (small in comparison with the reduced length scale), even when smoothing the integral kernels, as discussed by Tornberg & Shelley (2004). Finally, any numerical scheme for the time evolution of elastic filaments must address the stiffness of the equations caused by presence of high-order spatial derivatives in the equations of motion responsible for bending rigidity. Due to the very high rate at which disturbances of small wavelength are damped, the issue of stiffness becomes even more pronounced with finer mesh sizes. Practically all numerical works to date use an implicit integration scheme, while here we present a different approach.

We focus on a different configuration – looped filaments with no free ends. The dynamics of microscale loops in viscous flows was previously analysed in the context of growing smectic-A liquid crystals, which were modelled by Shelley & Ueda (2000) using SBT, and for circular filaments with a non-zero inherent twist and bend, explored using a variant of the IB method by Lim *et al.* (2008) in the context of over- and underwinding of DNA leading to dynamic transitions of shapes. Our work is motivated by the experimental work of Alizadehheidari *et al.* (2015) on circular DNA confined to nanofluidic channels (and in particular its breaking), and that of Koche *et al.* (2020) linking extrachromosomal circular DNA properties with neuroblastoma, and partly inspired by previous numerical work using bead-spring hydrodynamic models (Gruziel-Słomka *et al.* 2019). Electrophoretic and ultracentrifugation measurements of mobility pose questions about what constitutes a flexible regime, correct values of drag coefficients and the stress distribution along the filament.

The inclusion of elasticity is necessary to analyse the aforementioned systems. It was observed experimentally for red blood cells (Jay & Canham 1972) and numerically for

flexible chains of beads (Gruziel *et al.* 2018; Gruziel-Słomka *et al.* 2019) that high flexibility leads to a change in orientation (and sometimes shape) of sedimenting objects, affecting their sedimentation speed. Independently, in the case of linear filaments, Reichert & Stark (2005) observed that including elasticity can change the behaviour qualitatively when looking at bundling vs non-bundling flagella. Even for a fixed shape, a change in orientation can alter the sedimentation velocity by 25 %, as shown by Tchen (1954) and in further analytical solutions for sedimenting tori (Cox 1970; Majumdar & O'Neill 1977; Johnson & Wu 1979). Furthermore, as noted by Box *et al.* (2020) and Kodio, Goriely & Vella (2020), dynamical buckling can occur in similar settings, resulting in significant shape changes of the filaments. In bead-spring models, Gruziel-Słomka *et al.* (2019) observed the existence of an elasticity threshold beyond which flexible loops undergo significant changes in sedimentation dynamics. While stiff loops were seen to attain almost planar, oval shapes and sediment vertically or at an acute angle to gravity, depending on their stiffness, more flexible fibres exhibited a complex shape evolution. Our work aims to explore this stability threshold in slender-body dynamics, both analytically and in terms of numerical simulations.

In this work, we analyse the dynamics of slender elastic loops by linear stability analysis in a coupled elastohydrodynamic model, and by numerical simulations introducing a new method based on Fourier expansions. The mathematical elegance of the periodic boundary conditions allows us to simplify the theoretical analysis and gain an analytical insight into the stability question. Our results contribute to the explanation of horizontal sedimentation preference. We also derive explicit expressions for tension distribution along the filament, which comply with the work of Alizadehheidari *et al.* (2015) on DNA loops in microfluidic channels, where typical locations of ruptures correspond to the highest tension in our model.

2. Qualitative description

We focus on a thin, inextensible, looped elastic filament, settling in a viscous fluid under gravity. The filament has a length L and bending stiffness A. We consider the dynamics in the Stokesian regime of low Reynolds numbers, where the fluid drag forces are proportional to the local velocity of the filament. Solutions for the terminal velocity for a rigid loop were known even before the development of the SBT (Tchen 1954; Majumdar & O'Neill 1977) and were tested experimentally in some cases (Amarakoon *et al.* 1982). The distribution of shear forces from the fluid acting on toroidal particles in these solutions is not uniform and thus has to be balanced by an equal and opposite force from the particle. For loops that are not perfectly stiff, these forces may partially arise from elastic deformation.

The presence of elastic forces can give rise to a complex dynamics of the sedimenting loop. To understand the stability of sedimenting circular loops, we first consider a simpler case. A classical example of beam instability under external compression has a solution known since the mid-eighteenth century (Euler 1759). Buckling under internal forcing (a heavy column buckling under its own weight) was revisited later by Greenhill (1881). For a beam of length L under the action of an external force F, stability results from a competition of this forcing with the stiffness of the beam. Because the bending rigidity is quantified by the product $A = EI$, with E being the Young's modulus of the beam material, and I being its cross-sectional moment of inertia, the relevant dimensionless quantity is EI/FL^2, capturing the stiffness to external force ratio, and there exists a critical threshold value of this quantity above which the initial shape becomes unstable.

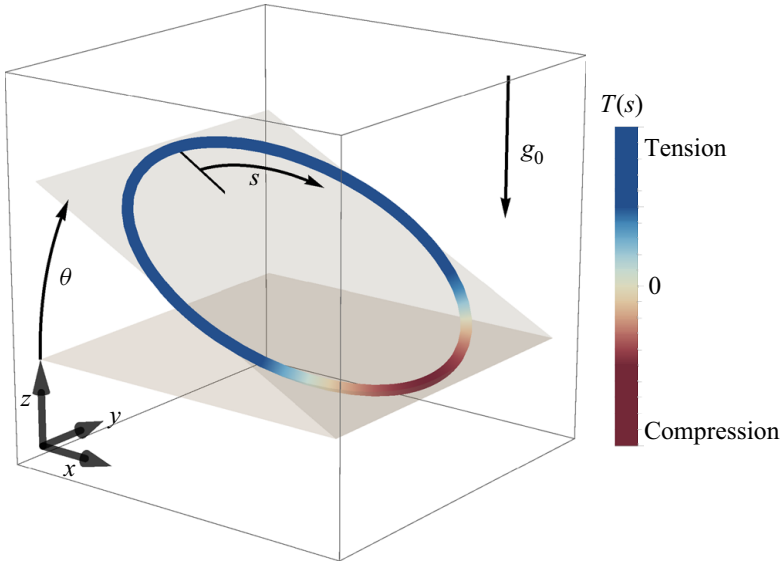


Figure 1. A three-dimensional sketch of the studied system. The direction of gravity $\boldsymbol{g_0}$, angle of inclination θ and the arc length variable s are highlighted. Additionally, the colour indicates the tension in the beam, either positive (aft or trailing side) or negative valued (fore or leading side). Such a tension distribution is possible when the dimensionless gravity-to-stiffness ratio and inclination angle are large enough to cause compression due to drag anisotropy.

We thus expect that a large enough compression is sufficient to destabilise an elastic filament. On the other hand, when we compute forces on a straight beam bent into the shape of a circle, we find that bending forces are balanced by the tension (negative compression, or stretching) in the beam of value $T = EI/R^2$, where R is the radius of the circle. Hence, in the absence of the fluid, we have two mechanisms which stabilise the shape against perturbations: the negative compression rate in the beam and the forces that resist bending. The presence of a fluidic environment introduces additional hydrodynamic forces that may be responsible for a local compression of the sedimenting loop.

Uniform drag would not lead to any compression, so even for a qualitative explanation of hydrodynamic buckling it is necessary to include the dependence of the drag force density on the location on the loop. The intuition that the drag is larger in the areas where the filament is perpendicular to the flow is exemplified by a known result of local SBT: that the ratio of the drag coefficients of a slender body in the directions parallel and perpendicular to the centreline of the body is $1/2$ (Guazzelli & Morris 2012, p. 79) up to $O(\epsilon)$, where $\epsilon \ll 1$ is the slenderness parameter of the body, or the filament aspect ratio. Consider now a sedimenting loop of circular shape, as in figure 1. The local gravitational force density is uniform on the circle, but we expect the drag forces to be higher at the top and the bottom in comparison with the sides, which results in the front of the circle being hydrodynamically compressed and its back being stretched. The compression can be destabilising, leading to a spontaneous shape change at the fore side of the loop – in sedimentation this is the lowest point of the loop – which can then lead to reorientation and further deformation. For stiff loops, the global balance of hydrodynamic, elastic and gravitational forces suggests that the hydrodynamic forces are proportional to the total weight of the loop only via a geometric, dimensionless factor. Because the flow field around stable configurations of flexible loops has to be identical to that around stiff loops of the same shape, we expect this scaling to hold for flexible loops as well. Recalling the stability threshold form to be EI/FL^2, and with the force scaling as mass times gravity, this reasoning finally hints at $EI/((\rho_L g L)L^2)$ being the dimensionless number governing this set-up, similarly to simpler buckling examples (here, ρ_L is the fibre linear density corrected for buoyancy).

3. Local slender-body equations

We model the fibre as a slender elastic beam of length L in a viscous fluid. To account for hydrodynamic interactions, we use a local SBT which is a far-field approximation of the Stokes flow due to an obstacle with a very small aspect ratio (Batchelor 1970; Cox 1970; Johnson 1980), allowing for shape parametrisation using the centreline position $\boldsymbol{x}(s)$, where $s \in [0, L]$ is the arc length.

The Stokes approximation is valid if two dimensionless constants are very small: the Reynolds and Stokes numbers, measuring the relative importance of viscous to advective and viscous to inertial terms in the Navier–Stokes equation, respectively. In this case, we neglect the inertial and time-dependent terms in the Navier–Stokes equations, and arrive at the Stokes equations describing the flow field $\boldsymbol{u}$ of an incompressible fluid with viscosity μ under external body force density $\boldsymbol{g}$

$$\mu \nabla^2 \boldsymbol{u} = \nabla p - \boldsymbol{g}, \tag{3.1}$$

$$\nabla \cdot \boldsymbol{u} = 0, \tag{3.2}$$

where p denotes pressure. These equations are linear and thus admit the Green's fundamental solution, also called the Stokeslet, which reads

$$\boldsymbol{u}_S(\boldsymbol{r}) = \frac{1}{8\pi\mu r}\left(\mathbb{I} + \frac{\boldsymbol{r}\boldsymbol{r}}{r^2}\right) \cdot \boldsymbol{f}, \tag{3.3}$$

where $\mathbb{I}$ denotes a unit tensor and $r = |\boldsymbol{r}|$. The Stokeslet is associated with a point force $\boldsymbol{f}$ acting on the fluid at the origin. Notably, its derivatives are also solutions to the Stokes equations. One of particular importance is the Stokes doublet, which has a dipolar flow character (Blake & Chwang 1974) and decays faster ($\sim 1/r^2$) than the Stokeslet solution.

The SBT solves for the flow around a slender object of radius r and length L, with a typical aspect ratio (slenderness parameter) $\epsilon = r/L \ll 1$, by approximating the force density on its surface by a distribution of Stokeslets and Stokes doublets along the centreline.

This is motivated by the idea that distributing the singularities should be sufficient to model the flow at distances large in comparison with the typical radius of the filament. Matching the 'inner' expansion of the flow field with the 'outer' flow produced by the body as a whole, and taking into account the boundary conditions on the surface of the rod, allows for expression of the centreline velocity of the filament, $\boldsymbol{u}(s)$, in terms of two linear operators $\boldsymbol{\Lambda}$ and $\boldsymbol{K}$ acting on the force density applied to the filament as

$$\boldsymbol{u}(s) = -\boldsymbol{\Lambda}[\boldsymbol{f}](s) - \boldsymbol{K}[\boldsymbol{f}](s). \tag{3.4}$$

The operators take the form

$$\boldsymbol{\Lambda}[\boldsymbol{f}](s) = \left[(c+1)\mathbb{I} + (c-3)\partial_s \boldsymbol{x}\, \partial_s \boldsymbol{x}\right] \cdot \boldsymbol{f}(s), \tag{3.5}$$

$$\boldsymbol{K}[\boldsymbol{f}](s) = \int_0^L \left(\frac{\mathbb{I} + \hat{\boldsymbol{R}}(s,s')\hat{\boldsymbol{R}}(s,s')}{|\boldsymbol{R}(s,s')|} \cdot \boldsymbol{f}(s') - \frac{\mathbb{I} + \partial_s \boldsymbol{x}\, \partial_s \boldsymbol{x}}{|s-s'|} \cdot \boldsymbol{f}(s) \right) \mathrm{d}s', \tag{3.6}$$

where $\boldsymbol{R}(s,s') = \boldsymbol{x}(s) - \boldsymbol{x}(s')$, $\hat{\boldsymbol{R}}(s,s') = \boldsymbol{R}(s,s')/|\boldsymbol{R}(s,s')|$ and $c = -2\log(\epsilon)$ is a function of the slenderness parameter. This method was initially developed by Batchelor (1970) and later improved by Cox (1970), Keller & Rubinow (1976) and Johnson (1980). The non-local contribution ($\boldsymbol{K}[\boldsymbol{f}]$ together with the c-independent part of $\boldsymbol{\Lambda}[\boldsymbol{f}]$) vanishes on comparison with the local one at a rate $o(1/\log(\epsilon))$. In this contribution, we shall take

advantage of this asymptotic behaviour by neglecting $\boldsymbol{K}[\boldsymbol{f}]$ entirely. For slender fibres, neglecting the non-local term is a great simplification towards an analytical treatment of the resulting equations and leads to the local SBT, also known as resistive-force theory (RFT) introduced by Gray & Hancock (1955), in which the local velocity (in a quiescent fluid) is related to the local hydrodynamic drag force on the filament, $\boldsymbol{f}_h$, by

$$\boldsymbol{u}(s) = -\frac{c}{8\pi\mu}\,(\mathbb{I} + \partial_s\boldsymbol{x}\,\partial_s\boldsymbol{x})\cdot\boldsymbol{f}_h(s). \tag{3.7}$$

More recently, RFT has been increasingly popular as a modelling technique in biological fluid dynamics and the analysis of the motion of slender filaments, leading, e.g. to a general qualitative agreement with experimental observations of deforming flagella (Lauga & Eloy 2013), or giving insights into the buckling (De Canio, Lauga & Goldstein 2017) and swirling instabilities (Stein *et al.* 2021) in the microtubule cytoskeleton. In fact, high-precision tracking of swimming sperm revealed that RFT can quantitatively predict the complex trajectory of a sperm cell (Friedrich *et al.* 2010). Even in the case when the slender filaments come close together, RFT has proven to be useful in predicting their bundling behaviour (Man, Koens & Lauga 2016). In Stokes flow, the forces acting on a suspended body balance out to zero. In our case, this involves elastic forces, gravity and hydrodynamic drag on the filament, so the no-net-force condition can be written as

$$\boldsymbol{f}_{el} + \boldsymbol{f}_g + \boldsymbol{f}_h = 0. \tag{3.8}$$

We model the elastic forces $\boldsymbol{f}_{el}$ according to the Euler–Bernoulli beam theory, which takes into account only the local curvature of the filament and the longitudinal tension (Tornberg & Shelley 2004; Euler, Fellmann & Mikhai 2016). The elastic force density is then given by

$$\boldsymbol{f}_{el} = EI\,\partial_{ssss}\boldsymbol{x} - \partial_s(T(s)\partial_s\boldsymbol{x}), \tag{3.9}$$

where EI is the flexural (bending) rigidity and $T(s)$ is the tension of the filament. The second term imposes a constant length of the filament with T acting as a Lagrange multiplier.

The tension is determined by the inextensibility equation $|\partial_s\boldsymbol{x}| = 1$, which can be rewritten as a condition on the filament velocity by taking the time derivative

$$0 = \tfrac{1}{2}\partial_t|\partial_s\boldsymbol{x}|^2 = \partial_{st}\boldsymbol{x}\cdot\partial_s\boldsymbol{x} = \partial_s\boldsymbol{u}\cdot\partial_s\boldsymbol{x}, \tag{3.10}$$

and noting that it is satisfied initially. One problem that arises due to this treatment is the lack of a correcting mechanism in cases when the length changes slightly due to numerical errors. We implement the solution proposed by Tornberg & Shelley (2004) by introducing a numerical stabilisation term, recasting (3.10) as

$$0 = \partial_s\boldsymbol{x}\cdot\partial_s\boldsymbol{u} - w(1 - \partial_s\boldsymbol{x}\cdot\partial_s\boldsymbol{x}), \tag{3.11}$$

with w controlling the absolute extension penalty.

The force balance condition, (3.8), governs the dynamics. The hydrodynamic force density on the filament is determined by the sum of the gravitational and elastic forces (as in (3.9)), which is then used to compute the velocity of the filament centreline via (3.7). Equation (3.10) closes the system by imposing the filament inextensibility. We now rescale these equations to arrive at a dimensionless system. Firstly, we choose the dimensionless arc length to have a domain $s \in [0, 2\pi]$ for additional convenience when expanding in Fourier series. This results in the characteristic length $L/2\pi$ scaling for $\boldsymbol{x}$.

Secondly, we introduce $L^3/(8\pi^3 EI)$ as the force scale. Finally, we rescale the time by choosing $2\mu L^3/(\pi EIc)$ as our velocity scale. This leads to equations in the form

$$\boldsymbol{f} = -\partial_{ssss}\boldsymbol{x} + (\partial_s \boldsymbol{x}\partial_s + \partial_{ss}\boldsymbol{x})T(s) - g_0\hat{\boldsymbol{e}}_z, \tag{3.12}$$

$$\boldsymbol{u} = \boldsymbol{f} + (\partial_s \boldsymbol{x} \cdot \boldsymbol{f})\partial_s \boldsymbol{x}, \tag{3.13}$$

$$0 = \partial_s \boldsymbol{x} \cdot \partial_s \boldsymbol{u} - w(1 - \partial_s \boldsymbol{x} \cdot \partial_s \boldsymbol{x}), \tag{3.14}$$

with a single dimensionless parameter $g_0 = L^3 g\rho_L/(8\pi^3 EI)$, analogous to (the inverse of) that used by Gruziel-Słomka *et al.* (2019) for the bead-spring model. Here, $\rho_L = \pi r^2(\rho_{beam} - \rho_{fluid})$ is the mass per unit length of the fibre corrected for buoyancy, with ρ_{beam} and ρ_{fluid} being the densities of the beam material and the fluid, respectively. We note that w is merely a numerical stabilisation constant having no influence on the solutions under exact evolution.

4. Linear stability analysis of the planar circle solution

We now use the equations of motion to study the sedimentation dynamics of looped filaments with the initial condition that they are perfectly circular and inclined at an angle θ_0 to the horizontal plane, as in figure 1. A rigid circle solution (including full hydrodynamics with non-local terms) was already known previously (Tchen 1954; Majumdar & O'Neill 1977; Johnson & Wu 1979) and serves as the starting point for our stability analysis.

Equations (3.12)–(3.14) admit a translating ($\partial_s \boldsymbol{u} = \boldsymbol{0}$) solution with a single parameter θ_0 as

$$\boldsymbol{x}_0(s, t) = [\sin s, \cos\theta_0 \cos s, \sin\theta_0 \cos s] + \boldsymbol{u}_0 t, \tag{4.1}$$

$$T_0(s) = 1 + \frac{g_0 \sin\theta_0}{3}\cos s, \tag{4.2}$$

$$\boldsymbol{u}_0 = \left[0, \frac{g_0 \sin 2\theta_0}{6}, \frac{g_0(7 - \cos 2\theta_0)}{6}\right]. \tag{4.3}$$

Note that, for $g_0(\sin\theta_0)/3 > 1$, there appears an area of negative tension (compression) in the beam, also the tension is largest on the aft side of the loop, explaining the observations of Alizadehheidari *et al.* (2015) that such flexible loops of DNA tend to break near fore or aft more frequently than in between – we propose that this effect is compounded by the higher curvature, as noted in the mentioned work.

Similarly to a sedimenting slender rod, this solution exhibits a lateral drift due to the friction anisotropy. The maximal settling angle (the angle between the sedimentation velocity and gravity) is $\gamma_{max} = \tan^{-1}(1/(4\sqrt{3})) \approx 8.2^\circ$ (as compared with $\gamma_{max} \approx 19.47^\circ$ for a thin rod (Guazzelli & Morris 2012, p. 83)). It is relatively small, because some parts of the circle contribute to the downwards force, while not contributing to the sideways force at any chosen angle.

We perturb the solution (4.3) by taking

$$\boldsymbol{x}(s) = \boldsymbol{x}_0 + \tilde{\boldsymbol{x}} + O(\tilde{\boldsymbol{x}}^2), \tag{4.4}$$

where a tilde over a symbol denotes the perturbation function. A general form of the perturbation turns out to be analytically intractable due to the complexity of the coupled equations for tension perturbation, so further simplifying assumptions are necessary. Here,

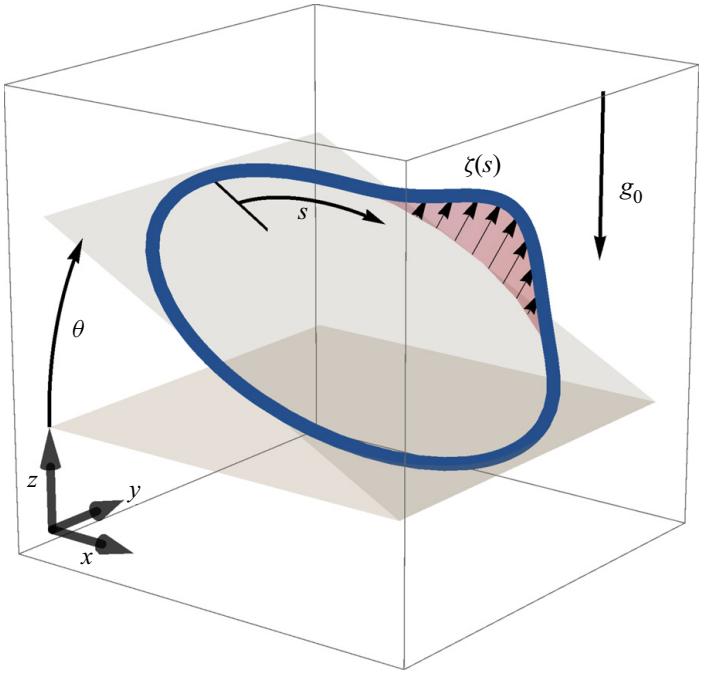


Figure 2. Diagram of the considered initial perturbation $\zeta(s)$. Presented linear stability analysis focuses only on the perturbation in the direction perpendicular to the plane in which the unperturbed solution lies.

we consider specific perturbations in the direction perpendicular to the circle's plane, as in figure 2, so that $\tilde{\boldsymbol{x}}$ is of the form

$$\tilde{\boldsymbol{x}} = \zeta(s)[0, -\sin\theta_0, \cos\theta_0]. \tag{4.5}$$

The presented method gives rise to two problems when trying to expand to in-plane perturbations. Firstly, taking a dot product with an in-plane vector instead of a normal vector leads to significantly more complex equations. The second complication is that in-plane perturbations are inherently two-dimensional and cannot be described by just a single scalar function. Choosing only specific normal perturbation is justified by the intuitive insight that comes from the tractable form of the resulting linear stability analysis problem.

Assuming that the associated perturbation of tension $\tilde{T}$ is $O(\zeta)$, we neglect quadratic- and higher-order terms in ζ. Then the force $\boldsymbol{f}$ in the perturbed system is of the form $\boldsymbol{f} = \boldsymbol{f}_0 + \tilde{\boldsymbol{f}}$ where

$$\boldsymbol{f}_0 = \frac{g_0 \sin\theta_0}{3}\left[\sin 2s,\ \cos 2s\cos\theta_0,\ \frac{3}{\sin\theta_0} + \cos 2s\cos\theta_0\right], \tag{4.6}$$

$$\tilde{\boldsymbol{f}} = -\partial_{ssss}\tilde{\boldsymbol{x}} + \partial_s T_0\,\partial_s\tilde{\boldsymbol{x}} + \partial_s\tilde{T}\,\partial_s\boldsymbol{x}_0 + \tilde{T}\,\partial_{ss}\boldsymbol{x}_0 + T_0\,\partial_{ss}\tilde{\boldsymbol{x}}. \tag{4.7}$$

Finally, we get a linear resulting perturbation to the velocity $\boldsymbol{u} = \boldsymbol{u}_0 + \tilde{\boldsymbol{u}}$ with

$$\tilde{\boldsymbol{u}} = \tilde{\boldsymbol{f}} + (\tilde{\boldsymbol{f}}\cdot\partial_s\boldsymbol{x}_0)\partial_s\boldsymbol{x}_0 + (\boldsymbol{f}_0\cdot\partial_s\tilde{\boldsymbol{x}})\partial_s\boldsymbol{x}_0 + (\boldsymbol{f}_0\cdot\partial_s\boldsymbol{x}_0)\partial_s\tilde{\boldsymbol{x}}. \tag{4.8}$$

This can be put into the inextensibility condition $\partial_s\boldsymbol{u}\cdot\partial_s\boldsymbol{x} = 0$. With $\partial_s\boldsymbol{u}_0 = \boldsymbol{0}$ and $\partial_s\tilde{\boldsymbol{u}}\cdot\partial\tilde{\boldsymbol{x}}$ being a second-order term, the inextensibility equation for the perturbed shape

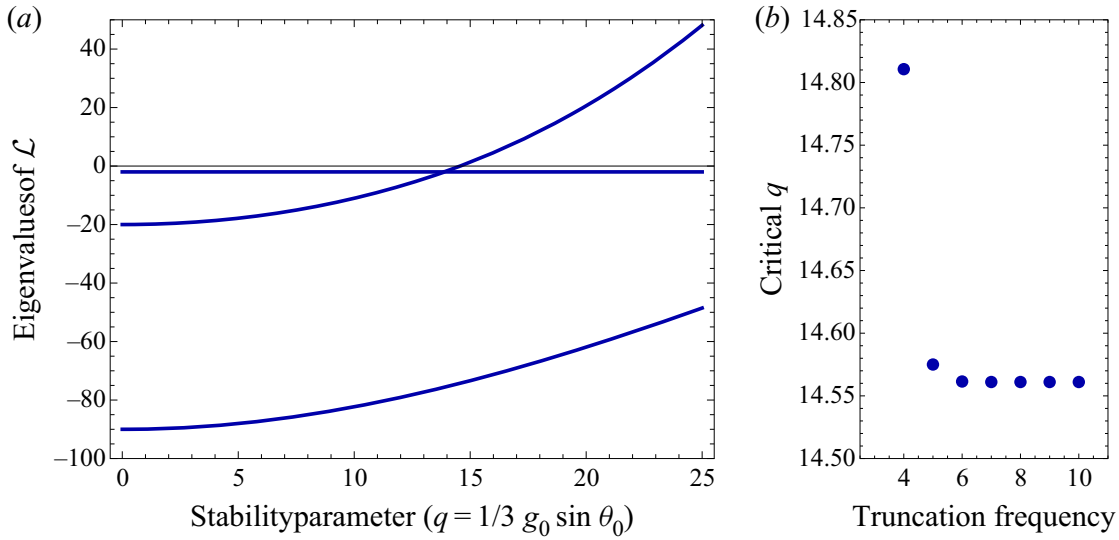


Figure 3. (*a*) Largest eigenvalues of the linear operator $\mathcal{L}$ depending on the stability parameter q (dark blue lines) describing the linearised stability problem. At q approximately 14.5 the largest eigenvalue crosses zero, which corresponds to the appearance of an unstable solution of the time-dependent equation. (*b*) The critical value of q computed for various values of the truncation frequency n. For $n < 4$ the behaviour of $\mathcal{L}$ is completely different, but for $n \geq 4$ the critical value of the stability parameter changes by a very small fraction. This is possible because the most unstable mode is dominated by low-frequency oscillations.

is simply

$$0 = \partial_s \tilde{\boldsymbol{u}} \cdot \partial_s \boldsymbol{x}_0. \tag{4.9}$$

It could be in principle solved for the tension perturbation. Instead, a more convenient way of proceeding is to note that

$$\partial_t \zeta = \tilde{\boldsymbol{u}} \cdot [0, -\sin\theta_0, \cos\theta_0] = \partial_{ss}\zeta + \frac{g_0 \sin\theta_0}{3}\left[(\partial_s\zeta)\sin s + (\partial_{ss}\zeta)\cos s\right] - \partial_{ssss}\zeta = \mathcal{L}[\zeta], \tag{4.10}$$

which has all the information needed to analyse the evolution of perturbation in the direction of the initial perturbation. The perturbation dynamics is now governed by a single parameter only

$$q = \tfrac{1}{3} g_0 \sin\theta_0. \tag{4.11}$$

Equation (4.10) can be rewritten as a diffusion-like equation of the form

$$\partial_t \zeta = -\partial_{ssss}\zeta + \partial_s(T_0 \partial_s \zeta). \tag{4.12}$$

This highlights the essential role of negative tension in the development of shape instability, which takes the role of the diffusion coefficient in (4.12) and the only other term is $\partial_{ssss}\zeta$, which has an additional stabilising effect. To determine the stability of the linear partial differential equation (4.12), we examine the eigenvalues of the linear operator $\mathcal{L}$ on the right-hand side. The value of the initial tension $T_0(s) = 1 + q\cos s$ gives rise to a simple analytical form of this operator

$$\mathcal{L} = -\partial_{ssss} + \partial_{ss} + q((\sin s)\,\partial_s + (\cos s)\,\partial_{ss}). \tag{4.13}$$

The periodicity of ζ can be enforced by analysing $\mathcal{L}$ action on $\{\sin ks, \cos ks\}$ basis on $L^2(S^1)$. We note that

$$\mathcal{L}[\sin ks] = -(k^2 + k^4)\sin ks + \tfrac{1}{2}kq(k+1)\sin(k-1)s + \tfrac{1}{2}kq(k-1)\sin(k+1)s, \tag{4.14}$$

and similarly for $\cos ks$. Therefore, to find the eigenvalues of $\mathcal{L}$, it is sufficient to consider linear combinations of the sine and cosine parts of the Fourier expansion separately, as $\mathcal{L}$ maps the span of either one to itself. Moreover, on each of the subspaces, the restricted maps are the same and thus have identical eigenvalues.

The operator $\mathcal{L}$ on span$[\sin(kx)]$ has the following matrix representation:

$$\mathcal{L} = \begin{bmatrix} -2 & -3q & & & & & \\ 0 & -20 & -6q & & & & \\ & -q & -90 & -10q & & & \\ & & -3q & -272 & & & \\ & & & -6q & \ddots & -qk(k+1)/2 & \\ & & & & & -k^2 - k^4 & \\ & & & & & -qk(k-1)/2 & \ddots \end{bmatrix}. \tag{4.15}$$

We can obtain approximations to $\mathcal{L}$ by truncating at a desired n. For a given n the condition that q is critical translates to $\mathcal{L}$ having one eigenvalue equal to zero, which can be expressed as $\det \mathcal{L} = 0$, which is a polynomial equation in q. Such equations have fast numerical solvers, allowing for computation of the critical value with high accuracy. We examined this for $n \in (1, 2, \ldots 60)$ to verify that the highest eigenvalue of $\mathcal{L}$ was determined with satisfactory precision – the convergence is extremely fast (at least exponential), as illustrated in figure 3. We find that the critical value of $q = g_0(\sin\theta_0)/3$ is 14.56105439107. Above this critical value the largest eigenvalue is positive, as illustrated.

5. Numerical method

In order to verify the predictions of the theoretical model and the simplified linear stability analysis, we solve the equations of motion numerically. Because all the functions characterising the elastic loop are periodic, we represent them in the form of (truncated) Fourier series. More precisely, an approximation to a function $f(s)$ is numerically represented by a complex valued $2n$-dimensional vector f_α, such that

$$f(s) \approx \sum_{\alpha=-n}^{n} f_\alpha \exp(\mathrm{i}\alpha s). \tag{5.1}$$

For a smooth f, such series converge exponentially. We simulate the equations of motion by computing truncated series approximations to the position $\boldsymbol{x}$, velocity $\boldsymbol{u}$, and tension T up to a fixed order n.

The governing equations can be coded as three affine maps (of type $\boldsymbol{x} \mapsto \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$) as follows:

$$T \xrightarrow[(1)]{\text{Euler–Bernoulli equation}} \boldsymbol{f} \xrightarrow[(2)]{\text{local SBT}} \boldsymbol{u} \xrightarrow[(3)]{\text{inextensibility equation}} \varepsilon. \tag{5.2}$$

These correspond to the Euler–Bernoulli theory (map (1) – (3.9)) allowing for computation of the force density given the tension T, local SBT mapping force density $\boldsymbol{f}$ to the local velocity u (map (2) – (3.7)) and the inextensibility equation mapping the local velocity to local length creation, the error term ε, which we try to minimise in the

simulation (map (3) – (3.11)). Maps (1), (2) are approximated by matrix equations of dimension corresponding to the truncation order n. This is chosen in order to keep $\boldsymbol{x}$ and $\boldsymbol{u}$ expanded to the same order. Nevertheless, it is essential to compute the inextensibility equation map (3) including higher-order ($2n$) terms. These three maps are combined to obtain a relationship between the coefficients of the Fourier expansion of the tension distribution T and the local filament length creation ε, which should be as close to zero as possible. Because there are more terms in the expansion of the local length creation than in the tension distribution, we can only attempt to make them as close to zero as possible. The combined affine map from the tension T to the error term ε therefore induces an overdetermined system of linear equations. These are solved for T by L^2 error minimisation, which is the same as solving the ordinary least squares (OLS) problem (Fourier basis is orthonormal) with an additional restriction that $T(s)$ is real valued. It might be tempting to simplify this procedure by using the same truncation order on the error terms as on the tension expansion, leading to an exact solution for T instead of an optimisation problem, but if we leave out the highly oscillatory terms in the error map (3), then the solver is oblivious to the filament length increase due to oscillations with frequencies higher than $n/2$ in the tension expansion, resulting in an exponential explosion of high-frequency vibrations. Such behaviour comes from the terms in the equations where two functions are multiplied (such as $\partial_s T\, \partial_s \boldsymbol{x}$); there two terms of a given wavenumber can combine to one term with double the wavenumber. Additional care needs to be taken to ensure that the trajectories remain real valued (the error can accumulate in the complex-valued $\boldsymbol{x}$ Fourier expansions). This was achieved by projecting the solution onto the allowed subspace at each time step.

Equations (5.2) for the tension T are solved at each evaluation of $\boldsymbol{u}$, and this value of tension is used to compute the velocity in an explicit integration scheme with a variable time step of the Runge–Kutta–Feldberg (fifth-order) method. This algorithm, however, is at best $O(n^3)$ in the truncation frequency (because the OLS minimiser is $O(n^3)$) and in reality even slower, as more degrees of freedom necessitate a decrease of the time step. For $n = 6$, our implementation was running at a speed of 40 dimensionless time units per hour for typical values of parameters on one thread of a typical 2.5 GHz processor; for $n = 8$, the speed decreased to 5 dimensionless time units per hour (giving a very crude estimate for complexity of approximately $O(n^7)$). This makes investigations of large values of n impractical. For our calculations, we choose $n = 6$. We discuss this choice further in § 6.1. Most of the simulations were run with the help of GNU Parallel software (Tange 2011).

6. Sedimentation modes

Motivated by the linear stability analysis performed above for circular loops, we now explore numerically the evolution of elastic rings starting at arbitrary inclination angles for a range of the elastohydrodynamic parameter g_0. We thus choose similar initial conditions to those used by Gruziel-Słomka *et al.* (2019). In the simulations, we observe three distinct sedimentation regimes (corresponding to terminal shapes) depending on the stiffness of the loop. They are demonstrated in figure 4 with top and in-plane views of the terminal sedimentation shapes. For very stiff loops, characterised by low values of g_0, marked in purple in the figure and referred to as stable, we observe stable sedimentation of a circular shape. The dynamics is then given by the translating solution of (4.3) – we see no change in the shape or the sedimentation angle. When the elasticity of the loop is increased, we observe a different terminal regime referred to as in-plane dynamics, where the shape of the loop evolves in a two-dimensional plane defined by the initial angle (cases marked in

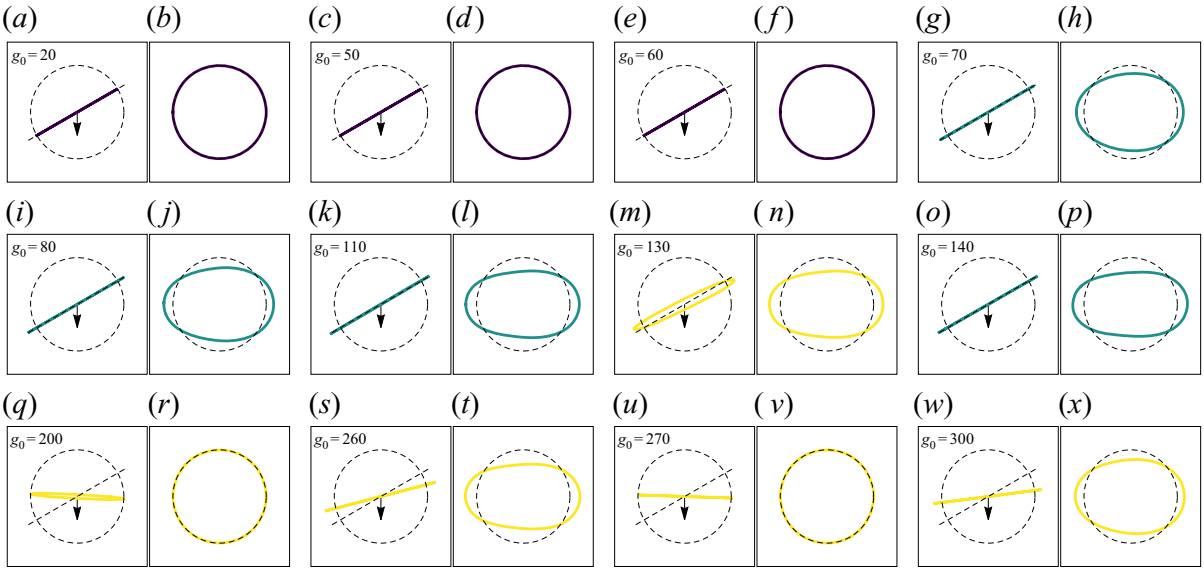


Figure 4. Selection of terminal shapes for varying values of forcing (inverse stiffness) g_0 with an initial angle $\theta_0 = 30^\circ$ and $n = 6$ coloured by trajectory type: stable (dark purple), in-plane dynamics (green-blue), three-dimensional dynamics (yellow). The left panels (a,c,e,g,i,k,m,o,q,s,u,w) show side views (sedimentation downwards) with a unit circle and initial tilt plane marked by dashed lines; black arrows indicate the direction of gravity. When the terminal shape aspect ratio is 1, regardless of the terminal inclination angle, the projection fits inside the unit circle. The right panels (b,d,f,h,j,l,n,p,r,t,v,x) show the shape within the final sedimentation plane aligned with the principal axes of the loop. Lowest points on the respective loops are located on the left side of panels (b,d,f,h,j,l,n,p,r,t,v,x). Dashed unit circle is plotted for reference. Note that for sufficiently large values of g_0 (highly flexible filaments) the terminal tilt angle changes erratically with initial conditions.

green–blue) – we only see changes of the shape, while the inclination angle of the loop remains unchanged. The circular state is then unstable and evolves into a prolate loop. When the stiffness is reduced further, for high values of g_0, we observe three-dimensional evolution (marked in yellow and referred to as three-dimensional dynamics) with a change in the angle of the terminal sedimentation plane with respect to the initial inclination. In the three-dimensional dynamics regime the loop leaves the initial plane essentially immediately – it does not go through dynamics similar to the in-plane regime. Deviations from a circular shape are necessary but not sufficient for a change in the inclination angle, as exemplified by the existence of the in-plane dynamics regime. For the three-dimensional dynamics regime in each simulation we observe in-plane and out-of-plane perturbations appearing spontaneously every time, and the inclination angle always changes.

In order to test the stability criterion derived in § 4, we plot the observed sedimentation modes depending on the dimensionless gravity g_0 and the initial sedimentation angle θ_0 in figure 5. Clearly, the stability criterion with $q \approx 14.5$ (solid line) divides the regions of absolute stability (with purple markers) and full three-dimensional dynamics (yellow markers), with the planar evolution states in between. All instances of stable behaviour are inside the predicted stability region. The initially unstable behaviour – yellow points to the right of the stability curve – involves a complex transient evolution that finally settles on a stable configuration at a smaller inclination angle. Even though the final shape in figure 4 is planar (prolate or circular), an example of the full shape transition in figure 6 shows significant bending with a complete deviation from the initial plane and the establishment of a new ellipsoidal shape in a different plane, essentially always at an angle smaller than the initial θ_0, and followed by a relaxation to the final shape. In most cases, the trajectory consists of three phases. Firstly, the loop folds in half starting with the fore side of the loop

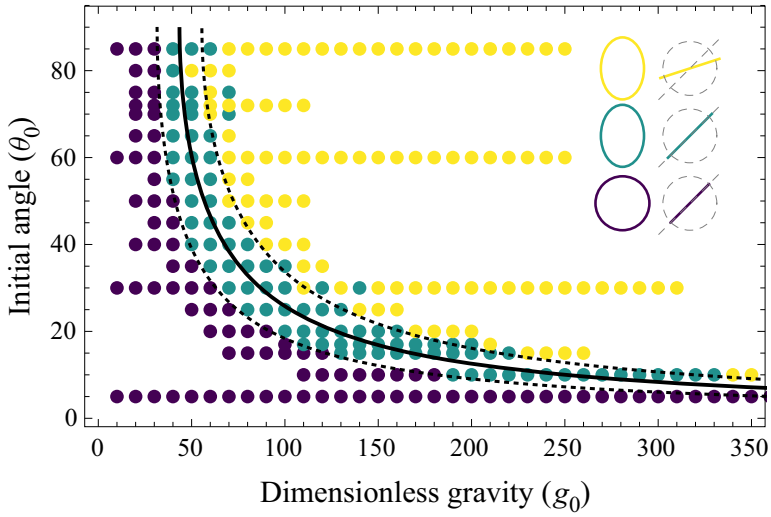


Figure 5. Sedimentation regimes showing stable translation (dark purple), in-plane dynamics (green-blue) and three-dimensional dynamics (yellow), observed in simulations for truncation frequency $n = 6$. Solid line represents the linear stability analysis prediction ($q \approx 14.5$); dashed lines are empirical in-plane dynamics boundaries ($q \approx 10.5$ and $q \approx 18.5$) presented as eye guides.

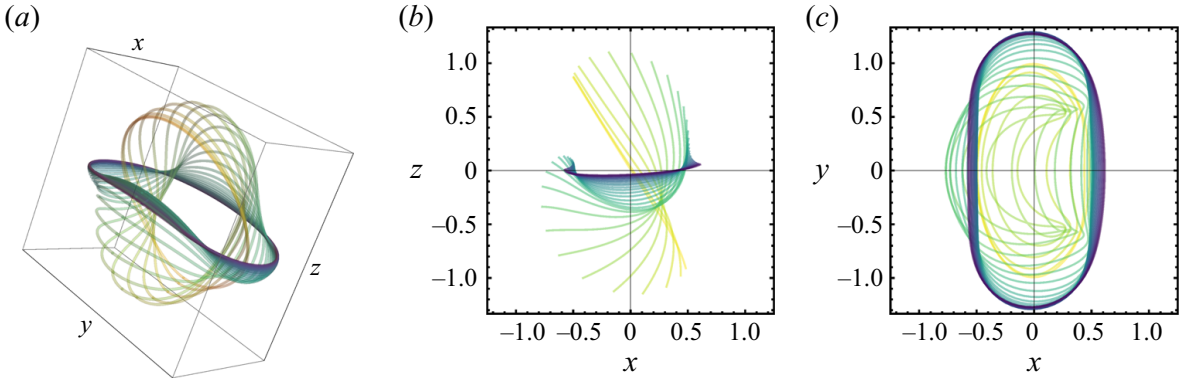


Figure 6. Sample of three-dimensional time evolution dynamics. Three-dimensional rendering (*a*), side view (*b*), top view (*c*). Images show shape evolution when starting in an unstable equilibrium. Snapshots are taken at regular time intervals with changing colour: initial light yellow to final dark purple. The loop starts at initial angle $\theta_0 = 60^\circ$ subject to a gravitational force of dimensionless value $g_0 = 150$. The fore side of the loop folds upwards leading to the formation of two lobes and an ellipsoidal shape (best visible in top view). At long times, the loop converges to a near-horizontal plane to finally (after a very slow dynamics) relax to a perfect circle (not shown in the figure).

falling faster than the centre of mass and immediately after being 'blown' backwards by the drag force. This is related to the loop deforming towards a more prolate shape. With the increasing fore–aft distance, and the centre of mass position remaining symmetrically in the middle of the loop, it is necessary for the fore side to move faster than the centre of mass during this stage. We regard this as the primary reason for the different behaviours of flexible rods and loops. In the initial stages of motion a loop extends in plane, elongating its long axis, while a rod retains its constant length. We note, however, that this effect lasts for a very short period of time, and is present only at the early stages of the evolution. Secondly, the two lobes formed by the fold relax towards the terminal plane. The dynamics then becomes very slow and the loop attains the terminal shape within the terminal plane. Notably, this behaviour is different from the case of a free-end filament, in which the

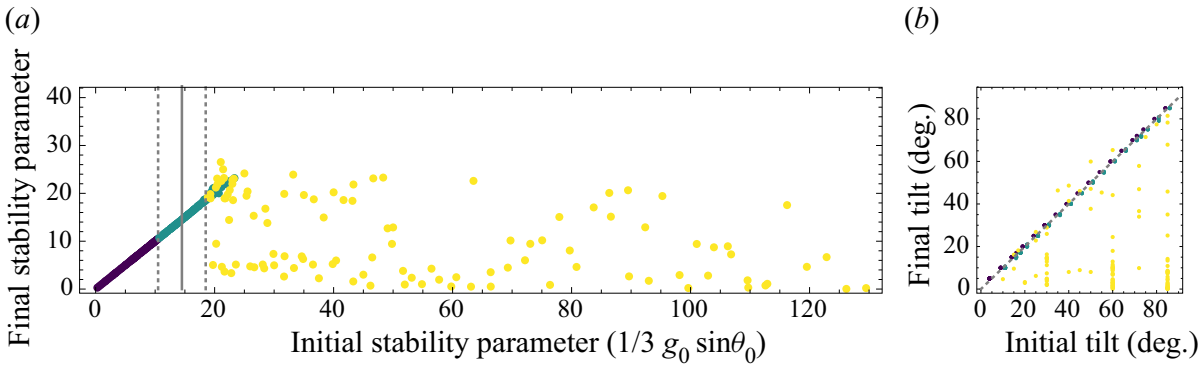


Figure 7. (*a*) A comparison between the initial and the final values of the stability parameter in all simulations. Linear stability analysis prediction (solid line at $q \approx 14.5$) is marked together with estimated in-plane regime boundaries of $q \approx 10.5$ and $q \approx 18.5$ (dashed lines). The points are coloured by the type of trajectory: stable (dark purple), in-plane dynamics (green-blue), three-dimensional dynamics (yellow). (*b*) A comparison between the initial and the final values of tilt angle. The stable (darkest) and the in-plane dynamics (medium) points were shifted 1° in (*a*,*b*) respectively, for clarity. The unordered scatter of the three-dimensional dynamics points (yellow) shows that the final tilt is difficult to predict, but in the vast majority of cases it is smaller than the initial tilt angle.

lowest part of the filament initially sediments slower than the centre of mass (Li *et al.* 2013), leading to a different shape evolution path.

Lastly, the green–blue points span across the stability boundary in figure 5. This type of behaviour is not taken into account by our simplistic linear stability analysis, because of the assumption of a normal direction of perturbation, while here the loop stays in the initial plane. Nevertheless, this type of dynamics is observed in the vicinity of the predicted stability boundary and can be regarded as an intermediate stage between complete instability and the complete lack of shape change. A vast majority of such trajectories are bounded by $10.5 < (g_0 \sin\theta_0)/3 < 18.5$. This region closely follows the stability curve in a wide range of the control parameter values and has boundaries that appear to have a similar functional form to the analytical predictions.

To further analyse the relationship between the initial and final sedimentation planes, in figure 7 we plot the initial stability parameter $q = g_0 \sin\theta_0/3$ vs the final parameter (calculated as $g_0 \sin\theta/3$) for all the cases investigated. The stability threshold is again marked with a solid line, together with the empirical strip of two-dimensional evolving shapes between the dashed lines. For stiff loops, at low values of q, the evolution does not affect the sedimentation angle, and thus we see the expected linear correlation, which persists for the semi-stable states which still remain in the initial plane. For unstable loops of high flexibility, we see no apparent correlation between the initial and the final sedimentation angles, as seen clearly in figure 7(*b*). Although the transition from stable to unstable dynamics is quite accurately grasped by our estimates, we note that the detailed evolution in this mode of motion is sensitive to the initial conditions, as discussed further.

In figure 8, we explore the deviation from a circular shape vs the final stability parameter. To this end, we plot the aspect ratio of the final shape for all our data, defined as the ratio of the highest to the middle eigenvalue of the spatial covariance matrix. The absolutely stable loops remain circular, and once we enter the unstable region, in the vicinity of the stability criterion we see a continuous increase of the aspect ratio. For unstable loops undergoing the full three-dimensional evolution, we do not see a systematic trend, but the final aspect ratios remain within the trend seen for stiffer loops, confirming that these configurations are globally stable solutions of the loop evolution equations. In addition, we see that the planar evolution is governed solely by the stability parameter $q = g_0 \sin\theta/3$, also in the

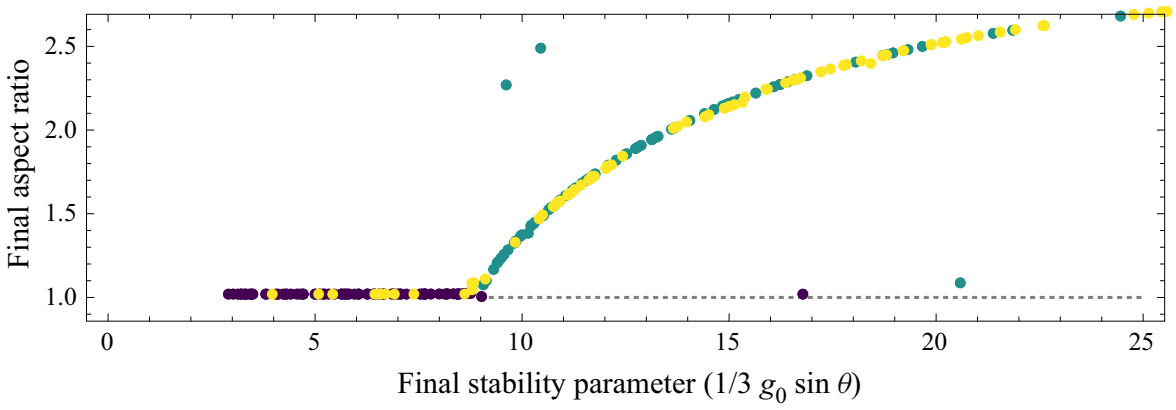


Figure 8. Aspect ratio of the terminal centreline shape plotted against the terminal value of stability parameter $g_0(\sin\theta)/3$. Large aspect ratios refer to highly elongated loops. A functional relationship confirms that the in-plane dynamics depends only on the stability parameter. The final shape is determined by the final sedimentation angle (and *vice versa*). Regime boundary between stable and the in-plane dynamics is clearly visible on this graph as the point where medium shade points reach the aspect ratio = 1 line. When the initial stability parameter is greater than $q^* \approx 9$, loops move from a stable to an in-plane dynamics regime. Note that for the stable and in-plane dynamics trajectories (darkest and medium points) the initial and the final tilt are the same, but for the three-dimensional dynamics the terminal tilt angle is smaller than the initial one (explaining lightest points with aspect ratios of unity). Regardless of the initial configuration and parameters, it is sufficient for the terminal stability parameter to be smaller than q^* to ensure the terminal configuration takes the shape of a circle. Dashed line shows the unstable branch of the solutions.

cases where the final value of θ is very small and the initial three-dimensional dynamics converges to a perfectly circular equilibrium shape, the same shape as in the case of stable sedimentation modes.

6.1. *Influence of truncation order*

All the presented summaries of simulations are results of a numerical scheme terminated at $t = 10$ of the dimensionless time, which corresponds to a sedimentation distance of the order of 10^3 in terms of loop radius for moderate values of stiffness, upon confirming no further shape evolution. This is long enough for all the simulated shapes to attain the final configuration. In most cases, all rapid changes in shape (each at time scales of approximately 0.5) are finished when t reaches approximately 3, then the terminal angle is selected. A much slower relaxation of shape within the terminal plane follows with the characteristic time of $t \approx 1$.

In the section above, we studied the stability of initial configurations and the attraction to a stable shape for different values of the initial angle θ_0 and the truncation orders $n = 6$. Numerical investigation of the eigenvalues of the truncated $\mathcal{L}$ operator shows that the stability boundary should be largely independent of the truncation order n. Even though the initial rate at which the instability develops will be independent of n (in an unstable equilibrium it is the numerical noise that initiates movement), we should expect quantitatively similar behaviour for the terminal motion of the loops.

In figure 9, we redraw the stability diagram, figure 5, for different truncation order of $n = 4$ (a) and $n = 8$ (b). Regardless of the initial angle and the truncation order, the stability region is correctly predicted by the simplified linear stability analysis. Moreover, when the loops are initially in an unstable position, they eventually reach the terminal angles, which are constrained by the stability region. However, the truncation order changes the exact behaviour predicted for a particular stiffness and initial sedimentation angle. Notably, for very small values of n the intermediate regime of in-plane dynamics cannot be observed.

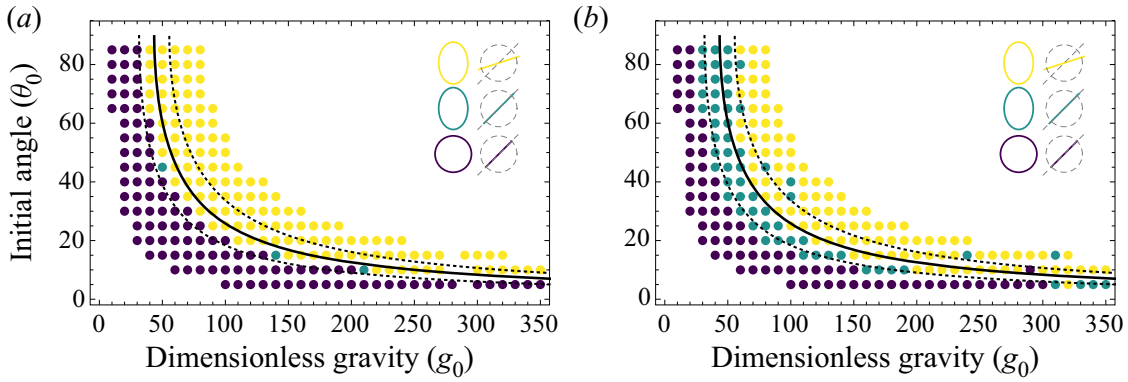


Figure 9. Comparison of regimes of sedimentation observed in simulation for truncation frequency $n = 4$ (*a*) and $n = 8$ (*b*). Solid line represents the boundary of the stable region as obtained from the linear stability analysis. Dashed lines are eye guides to facilitate the comparison of the two panels. Very small values of n do not exhibit the intermediate regime of in-plane dynamics. However, the proposed stability criteria seem to hold sway.

It is possible that this effect appears because the loop is unable to attain configuration close enough to the terminal shape of the in-plane dynamics, due to the small number of degrees of freedom. For a larger number of truncation modes, we see that the extent of the two-dimensional evolution region decreases, with more unstable states close to the stability criterion, but nevertheless the clear division between the modes remains in place, rendering the simplistic criterion a useful tool for assessing the loop stability.

Far from the stability boundary, the dynamics of the loop follows the qualitative conclusions derived from the analytical treatment, and the lower boundary (stable to in-plane dynamics) is weakly sensitive to the truncation frequency – additional simulations for $n = 10$ on a restricted set of initial angles confirm that a lower stability boundary shows no further change with inclusion of higher-order terms in expansions. We note, however, that the upper boundary (in-plane to three-dimensional dynamics) is a fragile one, and the selection of one attractor over the other is sensitive to the details of numerical implementation. The discussed discrepancies can be of three origins: (i) change in stability due to the truncation of high frequencies, (ii) change in perturbation power spectrum due to the change of dimensionality or (iii) change in perturbation due to the change in numerical stability of the OLS minimiser procedure. The presented analysis of truncation order in figure 3 gives us confidence that the high-frequency modes make negligible contributions to the stability problem due to the extremely high damping by elastic ∂_{ssss} terms. Therefore, our expectation is that effects (ii) and (iii) are the primary reasons for the observed differences between the smallest and largest values of the truncation order.

7. Conclusions

In this contribution, we modelled the behaviour of elastic loops sedimenting under gravity. To this end, we combined the local SBT with the Euler–Bernoulli beam theory to develop analytical insights into the dynamics and proposed a Fourier basis expansion method for effective numerical implementation. Our approach takes advantage of the periodicity of all the relevant functions in this setting, complementing our analytical treatment.

In simulations, when starting from an inclined circle, we identified three distinct regimes of motion, depending on the relative importance of gravity and loop stiffness, combined into a single dimensionless parameter g_0. For stiff loops, or low values of g_0, we see no effect of elasticity and the loops sediment as circular rings. When increasing the softness,

sedimenting loops remain oriented in the initial plane but attain elongated and slightly asymmetric shapes. For even softer filaments, the loops exhibit a transient instability, undergoing a three-dimensional shape evolution, where the fore edge of the loop is bent and leaves the initial plane, but the dynamics eventually settles on a planar shape at an angle different from the original. The said terminal angle is hard to predict and under fixed g_0 determines whether the final configuration will be circular or prolate. Small enough angles corresponding to terminal stability parameter $q = g_0 \sin\theta/3$ smaller than approximately $q^* = 9$ always result in an eventual circular shape.

To explain the transition between these regimes, we propose a simple theory based on linear stability analysis, with a further assumption that perturbation is taken in the direction perpendicular to the initial plane. This specific choice of perturbation leads to an analytical insight into the dynamics which we doubt to be possible with an arbitrary perturbation. To circumvent this difficulty, we additionally perform numerical simulations to provide a description for arbitrary initial conditions. Under the chosen simplification we identify the most important parameter controlling the motion as the initial stability parameter $q = g_0 \sin\theta_0/3$, and perform a near-analytic determination of the stability boundary at $q \approx 14.5$.

The results of numerical simulations are in satisfactory agreement with the simplistic approach of the linear stability analysis, thus confirming the validity of our approach for finding the stability threshold. Both the absolutely stable and the unstable regimes fit entirely within the domains predicted by the theory. The intermediate regime of planar shape evolution appears in close proximity of the stability boundary. We conclude that such a simplified linear stability analysis is a useful tool in both the three-dimensional dynamics case as well as for in-plane dynamics because it correctly predicts both scaling and approximate values of the stability parameter of regime transitions. For sufficiently stiff loops, we compare our numerical codes to the existing analytical results of Johnson & Wu (1979) and Tchen (1954) for stiff loops. We confirmed the expected agreement both qualitatively (translation without change of orientation) and quantitatively in terms of translation velocity asymptotics for very slender rods. Below the stability threshold, however, we see differences from the bead-spring model (Gruziel-Słomka *et al.* 2019), where sufficiently stiff sedimenting loops attained vertical or tilted oval shapes, in contrast to loops sedimenting without a change of orientation in our model. Increasing the flexibility leads to a deviation from the initial shape, resulting in an approach to a different equilibrium circle. Beyond the stability threshold, we also find that the details of intermittent evolution of more flexible fibres differ between the bead-spring results of Gruziel-Słomka *et al.* (2019) and slender-body models. This might be partly due to the lack of non-local terms in our resistive-force SBT and due to the different geometric details of both systems, i.e. a slender filament vs a chain of beads. In particular, we note that the stable circular configuration found in the RFT approximation is no longer a solution when the full hydrodynamics is included. However, the shape of the final tilt angle vs stiffness curve is similar in the bead-spring and RFT models.

The presented approach shows an attractive interpretation of the compression (negative tension) on the fore side of sedimenting objects as a negative diffusion coefficient in the governing equation of the linear stability analysis. This gives intuitive grounding to the experimental results such as those of Jay & Canham (1972) where red blood cells show a preference for horizontal sedimentation when their flexibility is increased, or Gruziel *et al.* (2018) where the preference for horizontal sedimentation was seen for knotted elastic fibres. It also provides support for the interpretation of experiments of the DNA loop

rupture dynamics (Alizadehheidari *et al.* 2015), in which the loops break in locations corresponding to the maximal tension in our description.

The conclusion that vertically oriented loop configurations are forbidden due to their instability is a general physical observation applicable in similar elastohydrodynamic settings. The presented results show that observations of instability from free-end simulations (Li *et al.* 2013) are applicable only to some extent: circular configuration gives rise to a tension offset which substantially improves stability in comparison with the free-end configuration.

We look forward to additional experimental verification of the conclusions of this paper, either in the microscale with biological fibres or in macroscopic experiments such as those with knotted bead chains of Gruziel *et al.* (2018).

Acknowledgements. We would like to thank M. Ekiel-Jeżewska, M. Gruziel-Słomka and B. Cichocki for discussions. Science Club Fenix is acknowledged for providing computational power for simulations.

Funding. The work of M.L. and R.W. was supported by the National Science Centre of Poland grant Sonata to M.L. no. 2018/31/D/ST3/02408. P.S. was supported by the National Science Centre (Poland) under research grant 2015/19/D/ST8/03199.

Declaration of interests. The authors report no conflict of interest.

Author ORCIDs.
Radost Waszkiewicz https://orcid.org/0000-0002-0376-1708;
Piotr Szymczak https://orcid.org/0000-0001-8940-7891;
Maciej Lisicki https://orcid.org/0000-0002-6976-0281.

References

Alizadehheidari, M., *et al.* 2015 Nanoconfined circular and linear DNA: equilibrium conformations and unfolding kinetics. *Macromolecules* **48** (3), 871–878.

Amarakoon, A.M.D., Hussey, R.G., Good, B.J. & Grimsal, E.G. 1982 Drag measurements for axisymmetric motion of a torus at low Reynolds number. *Phys. Fluids* **25** (9), 1495–1501.

Batchelor, G.K. 1970 Slender-body theory for particles of arbitrary cross-section in Stokes flow. *J. Fluid Mech.* **44**, 419–440.

Blake, J.R. & Chwang, A.T. 1974 Fundamental singularities of viscous flow. *J. Engng Maths* **8** (1), 23–29.

Box, F., Kodio, O., O'Kiely, D., Cantelli, V., Goriely, A. & Vella, D. 2020 Dynamic buckling of an elastic ring in a soap film. *Phys. Rev. Lett.* **124**, 198003.

Cortez, R., Fauci, L. & Medovikov, A. 2005 The method of regularized Stokeslets in three dimensions: analysis, validation, and application to helical swimming. *Phys. Fluids* **17** (3), 031504.

Cortez, R. & Nicholas, M. 2012 Slender body theory for stokes flows with regularized forces. *Commun. Appl. Maths Comput. Sci.* **7**, 33–62.

Cosentino Lagomarsino, M., Capuani, F. & Lowe, C.P. 2003 A simulation study of the dynamics of a driven filament in an Aristotelian fluid. *J. Theor. Biol.* **224** (2), 215–224.

Cox, R. 1970 The motion of long slender bodies in a viscous fluid part 1. General theory. *J. Fluid Mech.* **44** (4), 791–810.

De Canio, G., Lauga, E. & Goldstein, R.E. 2017 Spontaneous oscillations of elastic filaments induced by molecular motors. *J. R. Soc. Interface* **14** (136), 20170491.

Ehrlich, H.P., Grislis, G. & Hunt, T.K. 1977 Evidence for the involvement of microtubules in wound contraction. *Am. J. Surg.* **133** (6), 706–709.

Euler, L. 1759 *Sur la force des colonnes (concerning the strength of columns)*. Memoires de l'academie des sciences de Berlin.

Euler, L., Fellmann, E.A. & Mikhai, G.K. 2016 Letter 60, 12 December 1942. In *Leonhardi Euleri Commercium Epistolicum cum Daniele, Johanne II, Johanne III Bernoulli...* (ed. E.A. Fellmann & G.K. Mikhajlov, with A. Kleinert, M. Mattmüller, U. Monecke & A. Verdun), Leonhardi Euleri Opera Omnia, series IVA, vol. 3, part 1. Springer.

Fauci, L.J. & Dillon, R. 2006 Biofluidmechanics of reproduction. *Annu. Rev. Fluid Mech.* **38** (1), 371–394.

FRIEDRICH, B.M., RIEDEL-KRUSE, I.H., HOWARD, J. & JÜLICHER, F. 2010 High-precision tracking of sperm swimming fine structure provides strong test of resistive force theory. *J. Exp. Biol.* **213** (8), 1226–1234.

FULFORD, G.R. & BLAKE, J.R. 1986 Muco-ciliary transport in the lung. *J. Theor. Biol.* **121** (4), 381–402.

GAFFNEY, E.A., GADÊLHA, H., SMITH, D.J., BLAKE, J.R. & KIRKMAN-BROWN, J.C. 2011 Mammalian sperm motility: observation and theory. *Annu. Rev. Fluid Mech.* **43** (1), 501–528.

GOLDSTEIN, R.E. & LANGER, S.A. 1995 Nonlinear dynamics of stiff polymers. *Phys. Rev. Lett.* **75** (6), 1094–1097.

GOLDSTEIN, R.E., POWERS, T.R. & WIGGINS, C.H. 1998 Viscous nonlinear dynamics of twist and writhe. *Phys. Rev. Lett.* **80** (23), 5232–5235.

GRAY, J. & HANCOCK, G.J. 1955 The propulsion of sea-urchin spermatozoa. *J. Expl Biol.* **32** (4), 802–814.

GREENHILL, A.G. 1881 Determination of the greatest height consistent with stability that a vertical pole or mast can be made, and the greatest height to which a tree of given proportions can grow. *Proc. Camb. Phil. Soc.* **4**, 65–73.

GRUZIEL, M., THYAGARAJAN, K., DIETLER, G., STASIAK, A., EKIEL-JEŻEWSKA, M. & SZYMCZAK, P. 2018 Periodic motion of sedimenting flexible knots. *Phys. Rev. Lett.* **121**, 127801.

GRUZIEL-SŁOMKA, M., KONDRATIUK, P., SZYMCZAK, P. & EKIEL-JEŻEWSKA, M. 2019 Stokesian dynamics of sedimenting elastic rings. *Soft Matt.* **15**, 7262–7274.

GUAZZELLI, E. & MORRIS, J. 2012 *Physical Introduction to Suspension Dynamics*. Cambridge University Press.

HERZHAFT, B. & GUAZZELLI, É. 1999 Experimental study of the sedimentation of dilute and semi-dilute suspensions of fibres. *J. Fluid Mech.* **384**, 133–158.

HGUYEN, H., ORTIZ, R., CORTEZ, R. & FAUCI, L. 2011 The action of waving cylindrical rings in a viscous fluid. *J. Fluid Mech.* **671**, 574–586.

JAY, A.W. & CANHAM, P.B. 1972 Sedimentation of single human red blood cells, differences between normal and glutaraldehyde fixed cells. *J. Cell Physiol.* **80** (3), 367–372.

JOHNSON, R.E. 1980 An improved slender-body theory for Stokes flow. *J. Fluid Mech.* **99**, 411–431.

JOHNSON, R.E. & WU, T.Y. 1979 Hydromechanics of low-Reynolds-number flow. Part 5. Motion of a slender torus. *J. Fluid Mech.* **95** (2), 263–277.

KELLER, B. & RUBINOW, S. 1976 Slender-body theory for slow viscous flow. *J. Fluid Mech.* **75** (4), 705–714.

KOCHE, R.P., *et al.* 2020 Extrachromosomal circular DNA drives oncogenic genome remodeling in neuroblastoma. *Nat. Genet.* **52** (1), 29–34.

KODIO, O., GORIELY, A. & VELLA, D. 2020 Dynamic buckling of an elastic ring. *Phys. Rev. E* **101**, 053002.

KUEI, S., SŁOWICKA, A.M., EKIEL-JEŻEWSKA, M.L., WAJNRYB, E. & STONE, H.A. 2015 Dynamics and topology of a flexible chain: knots in steady shear flow. *New J. Phys.* **17** (5), 053009.

LAUGA, E. & ELOY, C. 2013 Shape of optimal active flagella. *J. Fluid Mech.* **730**, R1.

LAUGA, E. & POWERS, T.R. 2009 The hydrodynamics of swimming microorganisms. *Rep. Prog. Phys.* **72** (9), 096601.

LI, L., MANIKANTAN, H., SAINTILLAN, D. & SPAGNOLIE, S.E. 2013 The sedimentation of flexible filaments. *J. Fluid Mech.* **735**, 705–736.

LIM, S., FERENT, A., WANG, X.S. & PESKIN, C.S. 2008 Dynamics of a closed rod with twist and bend in fluid. *SIAM J. Sci. Comput.* **31** (1), 273–302.

LIM, S. & PESKIN, C.S. 2004 Simulations of the whirling instability by the immersed boundary method. *SIAM J. Sci. Comput.* **25** (6), 2066–2083.

LIM, S. & PESKIN, C.S. 2012 Fluid-mechanical interaction of flexible bacterial flagella by the immersed boundary method. *Phys. Rev.* E **85**, 036307.

MAJUMDAR, S.R. & O'NEILL, M.E. 1977 On axisymmetric stokes flow past a torus. *Z. Angew. Math. Phys.* **28** (4), 541–550.

MAN, Y., KOENS, L. & LAUGA, E. 2016 Hydrodynamic interactions between nearby slender filaments. *Eur. Lett.* **116** (2), 24002.

MENDELSON, N.H., THWAITES, J.J., KESSLER, J.O. & LI, C. 1995 Mechanics of bacterial macrofiber initiation. *J. Bacteriol.* **177** (24), 7060–7069.

NGUYEN, H., CORTEZ, R. & FAUCI, L. 2014 Computing flows around microorganisms: slender-body theory and beyond. *Am. Math. Mon.* **121** (9), 810–823.

PESKIN, C.S. 2002 The immersed boundary method. *Acta Numerica* **11**, 479–517.

REICHERT, M. & STARK, H. 2005 Synchronization of rotating helices by hydrodynamic interactions. *Eur. Phys. J.* E **17**, 493–500.

RUDOLPH, A.S, RATNA, B.R. & KAHN, B. 1991 Self-assembling phospholipid filaments. *Nature* **352** (6330), 52–55.

SAINTILLAN, D., DARVE, E. & SHAQFEH, E.S.G. 2005 A smooth particle-mesh Ewald algorithm for Stokes suspension simulations: the sedimentation of fibers. *Phys. Fluids* **17** (3), 033301.

SCHOELLER, S.F., TOWNSEND, A.K., WESTWOOD, T.A. & KEAVENY, E.E. 2021 Methods for suspensions of passive and active filaments. *J. Comp. Phys.* **424**, 109846.

SHELLEY, M.J. & UEDA, T. 2000 The Stokesian hydrodynamics of flexing, stretching filaments. *Physica* D **146** (1), 221–245.

STEIN, D.B., DE CANIO, G., LAUGA, E., SHELLEY, M.J. & GOLDSTEIN, R.E. 2021 Swirling instability of the microtubule cytoskeleton. *Phys. Rev. Lett.* **126**, 028103.

SŁOWICKA, A.M., WAJNRYB, E. & EKIEL-JEŻEWSKA, M.L. 2015 Dynamics of flexible fibers in shear flow. *J. Chem. Phys.* **143** (12), 124904.

TANGE, O. 2011 Gnu parallel – the command-line power tool; login. *USENIX Mag.* **36** (1), 42–47.

TCHEN, C.-M. 1954 Motion of small particles in skew shape suspended in a viscous liquid. *J. Appl. Phys.* **25** (4), 463–473.

TORNBERG, A.-K. & SHELLEY, M.J. 2004 Simulating the dynamics and interactions of flexible fibers in Stokes flows. *J. Comput. Phys.* **196**, 8–40.

WIGGINS, C.H., RIVELINE, D., OTT, A. & GOLDSTEIN, R.E. 1998 Trapping and wiggling: elastohydrodynamics of driven microfilaments. *Biophys. J.* **74** (2), 1043–1060.

3.3 Paper III: *DNA supercoiling-induced shapes alter minicircle hydrodynamic properties*

Paper III

"*DNA supercoiling-induced shapes alter minicircle hydrodynamic properties*"

Radost Waszkiewicz, Maduni Ranasinghe, Jonathan M Fogg, Daniel J Catanese Jr, Maria L Ekiel-Jeżewska, Maciej Lisicki, Borries Demeler, Lynn Zechiedrich, and Piotr Szymczak

Commentary

Bio-relevance of the DNA requires little introduction. Even though the basic structure of the DNA is well understood at least since the 1950s, understanding the secondary structure of the DNA filament remains an important but challenging task. Since the overall 3D conformation of the DNA can influence gene expression and is known to affect biological affinity, understanding the forces governing its elasticity is crucial. By using a selection of small DNA loops which differ only in the linking number, we were able to measure conformational changes of the DNA via its hydrodynamic properties in the diffusive measurements.

Earlier works such as that of Coleman and Swigon [47] established the correct energy density functional to model slender loops subject to torsional stress. Unfortunately, the original code giving piecewise analytical solutions for the problem is lost and thus re-implementation of the solver was required. That work used DNA loops as inspiration but was never directly tested against experimental results. On the other hand, a variety of software packages are capable of predicting hydrodynamic radius given the shape of a rigid molecule (Zeno, US-Somo, GRPY), but as far as we know, they were not yet used for prediction of hydrodynamic properties of DNA loops.

This study emerged from the collaborative efforts of three distinct groups. The first group, specialising in theoretical modelling, comprised the PhD candidate, Maciej Lisicki, and Piotr Szymczak from the Faculty of Physics, Univeristy of Warsaw and Maria L Ekiel-Jeżewska from the Institute of Fundamental Technological Research, Polish Academy of Sciences. The second group, responsible for the biosynthesis of DNA minicircles, included Jonathan M Fogg, Daniel J Catanese Jr, and Lynn Zechiedrich from the Department of Pharmacology and Chemical Biology at Baylor College of Medicine and Rice Univeristy. The third group, overseeing the AUC measurements, included Maduni Ranasinghe and Borries Demeler from the Department of Chemistry and Biochemistry at the University of Lethbridge.

The choice of DNA minicircles topoisomers as the subject of collaboration proved strategic. For the therory group, the shared hydration properties among different topoisomers and their loop structure, as opposed to rods, facilitated a more straightforward modelling process. The biosynthesis group leveraged their prior experience in preparing these molecules, coupled with an interest in understanding the physical origins of changes in the bio-activity of writhed DNA. Lastly, for the AUC group, the remarkable sample stability of DNA minicircles allowed for numerous repetition of experiments. Using the selected approach, we were able to calibrate the hydrodynamic thickness of the DNA filament against measurements of the configurations which adopt a toroidal shape and to use that value to predict the hydrodynamic radius of all topoisomers that were experimentally measured.

In this study, the PhD candidate: proposed and implemented a numerical method to obtain equilibrium configurations of the supercoiled DNA minicircles. They computed the hydrodynamic properties of these configurations using `GRPY`, `pygrpy`, and Zeno software (with only Zeno being included in the published manuscript); wrote the first draft and edited all subsequent versions of the manuscript, produced all graphs and visualisations incorporated in the manuscript.

Published online 27 March 2023

Nucleic Acids Research, 2023, Vol. 51, No. 8 ***4027–4042***
https://doi.org/10.1093/nar/gkad183

DNA supercoiling-induced shapes alter minicircle hydrodynamic properties

Radost Waszkiewicz [1], Maduni Ranasinghe [2], Jonathan M. Fogg [3], Daniel J. Catanese, Jr. [4], Maria L. Ekiel-Jeżewska [5,*], Maciej Lisicki [1,*], Borries Demeler [2,6,*], Lynn Zechiedrich [3,*] and Piotr Szymczak [1,*]

[1]Institute of Theoretical Physics, Faculty of Physics, University of Warsaw, Pasteura 5, 02-093 Warsaw, Poland, [2]University of Lethbridge, Dept. of Chemistry and Biochemistry, Alberta, T1K3M4, Canada, [3]Department of Molecular Virology and Microbiology, Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Department of Pharmacology and Chemical Biology, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA, [4]Department of Biosciences, Rice University, 6100 Main St., Houston, TX 77005-1827, USA, [5]Institute of Fundamental Technological Research, Polish Academy of Sciences, A. Pawińskiego 5B, 02-106 Warsaw, Poland and [6]University of Montana, Department of Chemistry and Biochemistry, Missoula, MT 59812, USA

Received November 17, 2022; Revised February 25, 2023; Editorial Decision February 27, 2023; Accepted February 28, 2023

ABSTRACT

DNA in cells is organized in negatively supercoiled loops. The resulting torsional and bending strain allows DNA to adopt a surprisingly wide variety of 3-D shapes. This interplay between negative supercoiling, looping, and shape influences how DNA is stored, replicated, transcribed, repaired, and likely every other aspect of DNA activity. To understand the consequences of negative supercoiling and curvature on the hydrodynamic properties of DNA, we submitted 336 bp and 672 bp DNA minicircles to analytical ultracentrifugation (AUC). We found that the diffusion coefficient, sedimentation coefficient, and the DNA hydrodynamic radius strongly depended on circularity, loop length, and degree of negative supercoiling. Because AUC cannot ascertain shape beyond degree of non-globularity, we applied linear elasticity theory to predict DNA shapes, and combined these with hydrodynamic calculations to interpret the AUC data, with reasonable agreement between theory and experiment. These complementary approaches, together with earlier electron cryotomography data, provide a framework for understanding and predicting the effects of supercoiling on the shape and hydrodynamic properties of DNA.

INTRODUCTION

Nearly seventy years after Rosalind Franklin's meticulous work that led to the first description of the structure of DNA (1), we are still working to understand how this remarkable molecule is organized, stored, activated, and segregated into daughter cells (2–9). It is becoming increasingly apparent that negative supercoiling (the underwinding of the DNA double helix) provides a secondary or 'hidden code' that contributes to the three-dimensional (3-D) organization of the genome (10) that can be used by cells as a 'molecular servomechanism' to detect and regulate gene expression (11).

We recently discovered that the degree of curvature, dictated by DNA loop length, additionally tunes supercoiling-mediated effects and promotes mechanical crosstalk to expose DNA bases at specific distant sites (12). Exposed DNA bases drastically increase DNA flexibility to change the 3-D structure of DNA, which, conversely, influences the location and frequency of the disruptions to base pairing (12). Therefore, 3-D shape and base exposure are manifestations of supercoiling and looping (13). These findings underscore how supercoiling-dependent conformational changes may allow DNA to be an active participant in its transactions (13,14).

We previously used minicircles of a few hundred base pairs and defined supercoiling to determine how supercoiling and looping modulates the 3-D structure of DNA (13). DNA loops in this length range are found in na-

*To whom correspondence should be addressed. Email: piotrek@fuw.edu.pl
Correspondence may also be addressed to Maria L. Ekiel-Jeżewska. Email: mekiel@ippt.pan.pl
Correspondence may also be addressed to Maciej Lisicki. Email: maciej.lisicki@fuw.edu.pl
Correspondence may also be addressed to Borries Demeler. Email: demeler@uleth.ca
Correspondence may also be addressed to Lynn Zechiedrich. Email: elz@bcm.edu

ture, e.g., wrapped around the nucleosome, and are involved in gene regulation (15–18). Human topoisomerase IIα relaxes supercoiled DNA minicircles (13), demonstrating that they are biologically active. Thus, minicircles are biologically relevant and mimic DNA loops naturally occurring in cells.

We previously determined how supercoiling modulates the shapes of minicircles using electron cryotomography (cryoET) (13). Although informative, these assays are laborious and only limited DNA sequences and buffer conditions have been explored (12,13,19). CryoET provides 3-D information on individual DNA minicircles of defined supercoiling (DNA topoisomers) (13), but the approach requires skill in the art, is time-consuming and the resulting structures are of low resolution. Increasingly powerful, atomic force microscopy (AFM) can show the helical repeat of DNA as well as areas of base pair disruption (19), but the sequence is unidentifiable and the technique requires that DNA is adsorbed onto a flat surface. Not only might this adsorption distort DNA conformations but it means that they are visualized as 2-D projections with limited 3-D information.

Computational modeling would be of great value in helping predict DNA negative supercoiling and looping behavior, but thus far fails to account for supercoiling- and looping-mediated site-specific base exposure or resulting conformational changes. For example, most efforts at understanding looping (20) or cyclizing of DNA (21) ignore supercoiling. Therefore, new modeling efforts are needed, including the parameters of supercoiling and degree of curvature (dictated by DNA loop length). Before this modeling can be improved, however, detailed parameters of supercoiled loops of DNA must be determined.

Whereas cryoET is impractical to use for multiple conditions, AUC and electrophoresis rapidly assess properties in solution, and can be used to test multiple conditions simultaneously. Toward the goal of understanding how DNA sequence and 3-D shape are affected by negative supercoiling and looping, here we combined state-of-the-art AUC (22) with mathematical modeling to determine hydrodynamic parameters of supercoiled DNA minicircles. We derived partial specific volume (PSV) and anisotropy, and measured the sedimentation and diffusion coefficient for minicircle DNA of different degrees of negative supercoiling and lengths. With these values, we determined the density of DNA. We discovered that DNA length and supercoiling strongly affect the sedimentation properties of minicircle DNA but have either no or only a minimal effect on the PSV.

We generalized the continuum elastic framework to accurately predict the previously observed DNA minicircle 3-D shapes (13). This generalization provides additional and complimentary information that will allow us to interpret supercoiling- and curvature-dependent DNA structural alterations. Emboldened by this accomplishment, we then combined the measured elastic and hydrodynamic properties of DNA minicircles using bead models and considering force and hydrodynamic effects to compute the hydrodynamic sedimentation and diffusion coefficients. These modeling results compared favorably to AUC measurements.

MATERIALS AND METHODS

Chemicals and reagents

MseI, Nb.BbvCI, Proteinase K, T4 DNA Ligase, low molecular weight DNA ladder, and 100 bp DNA ladder were purchased from New England Biolabs (Ipswich, MA, USA). Adenosine triphosphate (ATP), antifoam 204, dithiothreitol (DTT), ethidium bromide, and RNase A were purchased from Sigma-Aldrich (St. Louis, MO, USA). Acrylamide, ampicillin, chloroform, and sodium chloride were purchased from Fisher Scientific (Pittsburgh, PA, USA). All other chemicals were purchased from VWR International (West Chester, PA, USA).

Generation and purification of minicircle DNA

Plasmid pMC336 (13) was used to generate both the 336 bp and 672 bp minicircles via λ-integrase-mediated site-specific recombination as described (23). Double-length 672 bp minicircles contain two copies of the 336 bp minicircle sequence in tandem orientation and are generated by the recombination used to generate 336 bp minicircle DNA.

Generation of different DNA topologies

The 'supercoiled' samples are the 336 bp or 672 bp minicircle products of the purification process. These were analyzed without further manipulation. To make nicked DNA, the minicircles were nicked at a single site using the nicking endonuclease Nb.BbvCI according to the manufacturer's protocol. The 672 bp minicircle contains two copies of the BbvCI site and was thus nicked at both locations. Following nicking, the DNA was subsequently incubated at 80 °C for 20 minutes to inactivate Nb.BbvCI. Linear 336 bp was generated by incubating supercoiled 336 bp minicircle with MseI according to the manufacturer's protocol. The linearized DNA was subsequently incubated at 65 °C for 20 minutes to inactivate the enzyme. 'Relaxed' 336 bp minicircle DNA was generated by incubating the nicked minicircles with T4 DNA ligase in 50 mM Tris-Cl pH 7.5, 10 mM $MgCl_2$, 1 mM ATP, and 10 mM DTT overnight at room temperature. 'Hypernegatively supercoiled' 336 bp was generated in an identical manner as 'relaxed', except for the addition of ethidium bromide (6.5 μg/ml) to the ligation reaction. Ligations were subsequently extracted with butanol (to both reduce the volume and to remove the ethidium bromide), extracted with chloroform, then precipitated with ethanol. The nicked, linearized, and supercoiled minicircle samples were also subjected to butanol and chloroform extraction, and ethanol precipitation in a similar manner to both remove the enzymes and to ensure that any differences observed could not be attributed to differences in how the samples were made. Following ethanol precipitation, DNA was resuspended in 50 mM Tris-Cl pH 8.0, 150 mM NaCl, and 10 mM $CaCl_2$. DNA samples were subsequently subjected to multiple rounds of buffer exchange in the same buffer using an Amicon 0.5 ml centrifugal filter to ensure that buffer conditions were equal across all samples. DNA concentrations were determined using a Nanodrop spectrophotometer.

Geometry and topology of DNA minicircles

DNA supercoiling is defined by the linking number (Lk), the total number of times the two single DNA strands coil about one another (24). Lk is, thus, an integer number by construction if both strands are covalently closed. If one or both of the strands is not covalently closed, e.g., for nicked and linear DNA, Lk can adopt non-integer values. Another quantity determining the shape of DNA is the equilibrium helical repeat h defined as the number of base pairs between two locations where the backbones are aligned and is measured in base pairs per turn. The value of h is buffer dependent and is ∼10.42 bp/turn in 10 mM $CaCl_2$ (13). Using h, we can calculate the angle between terminal base pairs of a straight linear DNA segment of a given length L, which gives us the reference value $Lk_0 = L/h$. Because h can, in principle, take any value, Lk_0 is not restricted to integer values and usually has a fractional part. For relaxed 336 bp minicircles, we get $Lk_0 = 32.2$ while for 672 bp $Lk_0 = 64.4$. Therefore, the deviation from the most relaxed DNA structure is measured by the difference between Lk_0 and Lk denoted by $\Delta Lk = Lk - Lk_0$, which for the relaxed configuration of the 336 bp minicircle yields $\Delta Lk = -0.2$. ΔLk is typically scaled to the DNA length to give the superhelical density $\sigma = \Delta Lk/Lk_0$.

Because Lk is constrained to integer values, it is sufficient to report ΔLk rounded to the nearest integer to uniquely identify experimental configurations, as used in Ref. (13). For the relaxed 336 bp minicircle, we round −0.2 to 0. For simplicity, we follow this convention when reporting experimental values in this work. At the same time, we keep track of the fractional parts to accurately compute the elastic properties of the minicircles.

Gel electrophoresis

DNA samples were analyzed by electrophoresis through 5 % (for 336 bp and 672 bp samples) or 4 % (for 672 bp samples) polyacrylamide gels (acrylamide:bis-acrylamide = 29:1) in Tris-acetate buffer (pH 8.2) containing either 150 mM NaCl and 10 mM $CaCl_2$ (5 % gels) or 10 mM $CaCl_2$ (4 % gels) at 125 V (∼6 V/cm) for 8 hours. Buffer was continuously recirculated during electrophoresis. DNA samples were also analyzed by electrophoresis through 1.5 % and 3 % agarose gels (Seakem LE agarose, Lonza, Rockland, ME) in TAE (Tris-acetate + 1 mM EDTA) buffer at 100 V for 3 hours. Gels were subsequently stained with SYBR Gold (ThermoFisher Scientific, Waltham, MA), then visualized using a FOTO/ANALYST Investigator imaging system (Fotodyne, Hartland, WI, USA) with quantitation using ImageQuant TL, version 8.1 (GE Healthcare Life Sciences, Marlborough, MA, USA).

Analytical ultracentrifugation

Linearized, nicked, relaxed, supercoiled, and hypernegatively supercoiled 336 bp minicircles, and supercoiled and nicked 672 bp samples, were measured by sedimentation velocity using an AN50Ti rotor in a Beckman Coulter Optima AUC at the Canadian Center for Hydrodynamics at the University of Lethbridge in Alberta, Canada. For details of minicircles used, see Supplementary Table S1.

All samples were measured in 50 mM Tris-Cl pH 8.0, 150 mM NaCl, and 10 mM $CaCl_2$. 460 μl of each sample at an absorbance (A) of 0.6 at 260 nm were loaded into cells fitted with sapphire windows and 12 mm double channel epon charcoal centerpieces (Beckman Coulter, Indianapolis, IN, USA). Data were collected in intensity mode at 260 nm, and at 20 °C at five different rotor speeds of 10, 14, 25, 35, and 45 krpm. After data collection at each speed was completed, AUC cells were thoroughly shaken to redistribute the minicircle DNA uniformly. Depending on minicircle topology and length, at 10 krpm, pelleting occurred between 559–770 scans, requiring 50–70 hours. At 45 krpm, pelleting occurred after 88–159 scans, requiring 2–4 hours. The density and viscosity of the buffer, estimated with UltraScan, was 1.00682 g/ml and 1.02667 cP, respectively.

AUC data analysis

All data were analyzed with UltraScan-III, version 4.0 (6345) (22), using the UltraScan data acquisition module (25). UltraScan fits experimental data to finite element solutions of the Lamm equation, deriving distributions for sedimentation and diffusion coefficients (26,27). Optimization is achieved by parallel distributed data analysis, which was performed on the UltraScan Science Gateway using XSEDE resources (Expanse, Bridges 2, Stampede), and high-performance computing clusters at the University of Montana and University of Lethbridge. The optimization process proceeds through a series of model refinement steps, which employs the two-dimensional spectrum analysis (2DSA) (28). This refinement process removes systematic noise contributions contained in the raw data and obtains exact boundary conditions (the radial positions at the meniscus and the bottom of the cells) as described in (29). The final 2DSA refinement result is used to initialize a genetic algorithm analysis (GA) (30), which is followed by a Monte Carlo GA analysis (31). The total concentration determined from each speed between identical samples was also compared to ensure no material was lost due to aggregation or degradation, and samples were comparable across all speeds for a global analysis. The Monte Carlo GA results from identical samples and different speeds were combined to initialize a global GA analysis over all speeds. UltraScan supports simultaneous fitting to datasets from multiple experiments performed at different speeds. A global analysis benefits from the enhanced signal of the diffusion coefficient at low speeds and the improved sedimentation signal at higher speeds (32,33). This feature also enhances signal-to-noise ratios and improves the confidence limits for the determined hydrodynamic parameters. The global fitting algorithm in UltraScan is further explained in (22).

Hydrodynamic properties

Translational diffusion coefficient. The translational diffusion coefficient D is inversely proportional to the translational frictional coefficient f,

$$D = \frac{k_B T}{f}, \tag{1}$$

where k_B is the Boltzmann constant, and T the absolute temperature. In this work, diffusion occurs at very low concentrations of solute, which allows the analysis of transport coefficients in terms of single-particle properties only. For microscopic solid spheres of radius R suspended in a liquid of temperature T and viscosity η the Stokes-Einstein relationship reads

$$D = \frac{k_B T}{6\pi\eta R}. \tag{2}$$

This relationship can be generalized to non-spherical molecules by introducing the effective hydrodynamic radius R_h, defined as

$$R_h = \frac{k_B T}{6\pi\eta D}. \tag{3}$$

In aqueous solvents, macromolecules are typically hydrated, which adds to their apparent size and friction. The value of R_h derived from measured D includes these effects.

The hydrodynamic anisotropy of the DNA minicircles was characterized by the frictional ratio f/f_0, being the ratio of the measured frictional coefficient f and the frictional coefficient f_0 of a spherical particle of the same volume. The anisotropy equals 1.0 for a spherical molecule and exceeds 1.0 for non-spherical molecular shapes.

Sedimentation coefficient. The sedimentation coefficient s depends on the molar mass M, the translational frictional coefficient f, and the buoyancy of the particle, which is a function of its PSV, $\bar{v}$, and solvent density ρ,

$$s = \frac{M(1-\bar{v}\rho)}{Nf}, \tag{4}$$

where N is Avogadro's number. The Svedberg equation describes the ratio of the two parameters measured in a sedimentation velocity experiment, s and D, and provides a way to estimate the molar mass M, if the PSV is known:

$$\frac{s}{D} = \frac{M(1-\bar{v}\rho)}{Nk_B T}. \tag{5}$$

Apparent PSV. Eq. (5) considers a two-component system—an analyte with anhydrous molar mass M and a solvent with density ρ. However, our experimental solution also contains buffer components and ions that may be bound to the analytes. The degree of counterions bound to the analyte is dependent on solvent conditions and the ionic strength of the solvent, particularly for charged molecules (34). PSV is defined as the change in volume when one gram of analyte is added to the solvent, and is typically reported in units of ml/g. Because we do not know the precise amount of counterions bound to the analyte, we consider an apparent partial specific volume $\bar{v}'$, which can only be considered constant for a single solvent at a constant temperature and pressure. Rearranging the Svedberg equation allows the determination of the apparent PSV, provided the molar mass and the solvent density are known and the sedimentation and diffusion coefficients have been determined experimentally from a sedimentation velocity experiment

$$\bar{v}' = \frac{1}{\rho}\left(1 - \frac{Nsk_B T}{MD}\right). \tag{6}$$

In our case, the molar masses are 207.576 kDa for the 336 bp minicircle and 415.152 kDa for the 672 bp minicircle, as calculated from the sequence (12) using molbiotools.com/dnacalculator. UltraScan automatically estimates the solvent density and viscosity from the buffer composition, and adjusts the experimental $s_{T,B}$ and $D_{T,B}$ values to standard conditions (water at 20 °C) using the density and viscosity estimates from the buffer components, see ((35), p. 117)

$$s_{20,w} = s_{T,B}\frac{(1-\bar{v}'\rho)_{20,w}}{(1-\bar{v}'\rho)_{T,B}}\frac{\eta_{T,B}}{\eta_{20,w}}, \tag{7}$$

$$D_{20,w} = D_{T,B}\frac{T_{20}}{T}\frac{\eta_{T,B}}{\eta_{20,w}}, \tag{8}$$

where $s_{T,B}$ is the observed sedimentation coefficient at experimental conditions (temperature $T = 293.15$ K and buffer B). However, for the $s_{20,w}$ corrections, the partial specific volume of DNA at standard conditions is required, but it is not known to us and impossible for us to measure. While literature values are reported for NaDNA (0.54–0.55 ml/g) (36,37), topoisomers here were studied in 10 mM calcium, which has a higher binding affinity to DNA than Na (38). Hence, we report here the experimentally measured values of s and D for all topoisomers, and the apparent partial specific volume under experimental conditions, $\bar{v}'$, calculated by Eq. (6) and assuming a two component system.

Finding equilibrium shapes of loops

To model the shapes of DNA minicircles with a given Lk, a variant of the Kirchhoff beam theory for inextensible rods (39) was used, which describes the twisting and bending of a uniform elastic filament of constant steric thickness d_s, which was set to 20 Å in all computations. The helical repeat of the DNA yields a reference value of $Lk_0 = L/h$ for a given length L.

To model a DNA minicircle with a given Lk (and ΔLk), we used an elastic beam representation in which a (closed) beam is characterized by two constants: bending rigidity A and geometric torsional stiffness ω (describing the cross-sectional shape, equal to 2/3 for circular cross sections). The energy density has quadratic contributions from the residual excess twist density Ω and local curvature κ. The total energy is thus given by

$$E_{\text{rod}} = \frac{1}{2}\int A\left(\kappa^2 + \omega\Omega^2\right)\mathrm{d}s, \tag{9}$$

where Ω is computed from Lk and the shape of the filament centerline with the help of the Călugăreanu theorem (40)

$$Lk = Tw + Wr, \tag{10}$$

where twist (Tw) is defined as

$$Tw = Lk_0 + \frac{1}{2\pi}\int_0^L \Omega\,\mathrm{d}s, \tag{11}$$

and writhe (Wr) is defined in the standard way (41). Scaling by $k_B T$, the energy function can be made dimensionless, leaving the (width-to-length) aspect ratio d_s/L and ω as the

only parameters of the model. This approach was used by Coleman and Swigon (42) to categorize equilibrium shapes of looped filaments for a single aspect ratio $d_s/L = 8.2 \times 10^{-3}$ (corresponding to a DNA minicircle of length 718 bp and $d_s = 20$ Å), which constitutes a benchmark for our computations. Coleman and Swigon began by solving the problem of a free beam segment subject to boundary conditions at each end, and two beams in contact along a contact line. Such solution fragments can be glued together at contact points to form a looped solution, subject to gluing conditions that ensure the continuity of the first two derivatives and appropriate jump conditions to account for beam-beam steric forces. This approach uses the same expression for the beam energy but addresses the energy minimization in a different way—either by solving an ordinary differential equation subject to appropriate boundary conditions when no contact forces are needed or by direct minimization subject to no-overlap constraint when contact forces are present.

Determination of critical ΔLk

The stability of a computed minicircle shape depends on its ΔLk (42,43). For sufficiently small $|\Delta Lk|$, a flat circular configuration is the only equilibrium solution. Upon increasing $|\Delta Lk|$, at a thickness-dependent threshold value of critical (minimal) Lk_{crit}, a figure-8 solution becomes admissible and the flat circular and figure-8 shapes coexist. Supercoiling further, above the thickness-independent threshold of $Lk_{\max} = \sqrt{3}/\omega$, the flat circular shape is no longer a solution and only writhed configurations exist (43). For the prediction of Lk_{crit} for initial writhing of a minicircle, an approach based on solving an ordinary differential equation for centerline shape was used. The minimal value of $|\Delta Lk|$ required for writhing can be characterized by the existence of a configuration with a single contact point but with zero contact force. An ordinary differential equation was written for the beam centerline with minimum energy in a Cartesian parametrization subject to the boundary condition of a single contact point and no contact force and solved numerically using Mathematica, where boundary value problems are solved by the shooting method with conjugate gradient descent. For any given separation of the centerline at the self-contact location, one value of residual twist density was found. The relationship between the two was used to derive $Lk_{\text{crit}}(d_s/L)$.

Energy minimization of DNA minicircle shapes

Having determined the range of $|\Delta Lk|$ for which writhed configurations can be stable, the space of admissible configurations was examined and those configurations that minimized the elastic energy were investigated. Here, because of the presence of contact forces, a different numerical method was used. Representing the centerline shape with periodic cubic splines, direct energy minimization was performed over all possible shapes without self-intersections. Curves with 16 nodal points with enforced dihedral symmetry were subjected to a Monte Carlo minimization procedure. The bending energy was calculated directly from curvature using the adaptive Simpson's algorithm and taking advantage of the twice continuously-differentiable nature of the cubic splines. The precise estimation of Wr, required to compute Tw, was performed by approximating the curve by 200 linear segments and using an algorithm proposed by Levitt (44) to deal with the singularities of the Gauss formulation. Steric interaction was introduced by tracking self-intersections through a large number ($20\,L/d_s$) of sample points along the curve and a suitable steric energy penalty. Length constraint was imposed by computing the apparent length at each optimization step using the adaptive Simpson's method and by imposing an energy penalty for the deviation from the prescribed length. Multiple sets of different control parameters for numerical optimization were tested to ensure both fast convergence and satisfactory precision. All final computations were done with identical discretization and penalty characteristics. For final values of penalty parameters, Monte Carlo procedure parameters, and initial conditions, see data availability section.

Models for hydrodynamic radius

Solutions for Stokes flow around a slender toroidal object were developed by Johnson (45) that provide an asymptotic approximation in terms of slender-body theory. A fully analytical approach based on toroidal harmonics used by Goren and O'Neill (46) allows exact computations of all elements of the mobility matrix for a torus with an arbitrary aspect ratio. For a rigid, axially symmetric particle of a given length L and hydrodynamic thickness d_h, the mobility coefficients for translation along the symmetry axis and perpendicularly to it, $\mu_z(L, d_h)$ and $\mu_x(L, d_h)$, respectively, are sufficient to compute the hydrodynamic radius R_h by taking the inverse of the arithmetic mean

$$R_h = \frac{1}{6\pi\eta}\left(\frac{3}{2\mu_x(L, d_h) + \mu_z(L, d_h)}\right). \qquad (12)$$

To theoretically determine R_h for an arbitrarily shaped molecule, a rigid bead model of its structure was constructed and its hydrodynamic radius was calculated using the ZENO software package (47,48). In our case, the configuration of a minicircle was represented by 400 spherical and overlapping beads placed on the shape centerline, with diameters corresponding to the hydrodynamic thickness of the DNA molecule and the distance between overlapping beads summing up to the length of the molecule. This structure was then used to evaluate R_h for the composite particle. The diffusion coefficient at a given T and η is calculated from the definition of R_h in Eq. (2). The sedimentation coefficient is obtained from the Svedberg relation, eq. (5).

RESULTS

Rationale

The study of DNA supercoiling and curvature has benefited from multiple complementary theoretical and experimental approaches. The combination of multiple approaches takes

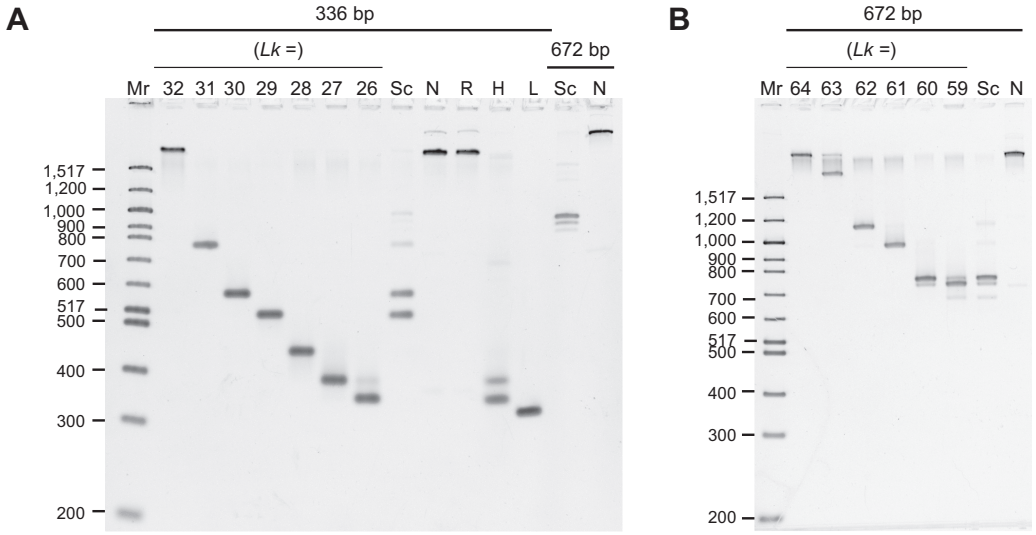


Figure 1. Electrophoretic mobility of minicircle DNA. (**A**) DNA samples were analyzed by polyacrylamide gel electrophoresis (5 % polyacrylamide) in 150 mM NaCl and 10 mM $CaCl_2$ (the same conditions used in analytical ultracentrifugation). Mr: 100 bp DNA ladder, lanes 2–8: 336 bp minicircle topoisomer markers (*Lk* as indicated), lanes 9–13: 336 bp minicircle DNA samples (Sc: 'supercoiled,' N: nicked, R: relaxed, H: 'hypernegatively supercoiled,' L: linear), lanes 14–15: 672 bp DNA samples (Sc: 'supercoiled,' N: nicked). (**B**) Determination of topoisomer identity in 672 bp samples. DNA samples were analyzed by electrophoresis on a 4 % polyacrylamide gel in the presence of 10 mM $CaCl_2$. Mr: 100 bp DNA ladder, lanes 2–8: 672 bp minicircle topoisomer markers (*Lk* as indicated), lanes 9–10: 672 bp DNA samples (Sc: 'supercoiled' as isolated from the bacteria, N: nicked).

advantage of the knowledge gained from each approach while helping shore up their individual limitations.

We previously (13) used 336 and 672 bp minicircles to study the effect of supercoiling and looping on DNA structure and have extensive 3-D structural data. AUC requires more material than cryoET (13), AFM (19), or other biochemical and biophysical analyses (12). A wide range of supercoiling was explored by testing relaxed, nicked, supercoiled, and hyper-negatively supercoiled minicircle samples. Supercoiled minicircle DNA was obtained from bacterial cells and, therefore, approximates the steady-state level of supercoiling in bacteria. Hypernegatively supercoiled samples have been further manipulated to increase the level of supercoiling and allowed us to determine whether AUC can distinguish different topoisomers from each other. To determine the effect of circularity, linearized DNA samples were also analyzed.

We characterized the minicircle samples by gel electrophoresis, which allows the topoisomer distribution of each sample to be precisely determined. Polyacrylamide gel electrophoresis effectively separates minicircle topoisomers and provides some insight into the conformational differences, although the theory underlying the differential migration is not fully understood.

We then applied advanced theoretical modeling to see whether it can explain the previously observed 3-D conformations of these minicircles (13). It was reasonably successful, and these advanced mathematical models could then be used to help analyze and interpret AUC data. Algorithms used elsewhere to interpret AUC data approximate shape as a sphere to a line, but because of the diversity of shapes of DNA, in this paper a theoretical model was constructed to determine specific shapes and use them to evaluate diffusion and sedimentation coefficients.

Electrophoretic characterization of DNA minicircles

DNA minicircles were analyzed by polyacrylamide gel electrophoresis. Both the helical repeat and conformation of DNA are sensitive to solution conditions (49,50). Here we used the same conditions used in analytical ultracentrifugation experiments (150 mM NaCl and 10 mM $CaCl_2$).

Supercoiled topoisomers migrated much more rapidly on the polyacrylamide gel than relaxed topoisomers (Figure 1). In comparison, the different topologies had relatively similar mobilities on an agarose gel (Supplementary Figure S1). This difference in migration on polyacrylamide gels can be at least partially explained by the relative compactness of supercoiled minicircle conformations (13). The nicked and relaxed topoisomers had near-identical migration, suggesting that the single-strand break in the nicked minicircle does not significantly affect the global conformation. The lack of difference is explained by the number of helical turns in the 336 bp minicircle studied being close to a perfect integer value of 32 (under these conditions), resulting in the base pairs flanking the nick site being in close rotational alignment, allowing for favorable base stacking across the nick (12). We previously showed that when the rotational alignment is out of phase (i.e., when the number of helical turns deviates from a perfect integer value), the effect of a nick on polyacrylamide gel migration is much more pronounced (12).

The topoisomer distribution for the samples taken through to AUC analyzes was measured from quantification of digital images of the fluorescently stained gels using image analysis software. The 'supercoiled' 336 bp sample contained primarily $\Delta Lk = -3$ (48 %), $\Delta Lk = -2$ (41 %), and $\Delta Lk = -1$ (7 %) topoisomers. The sample also con-

Table 1. Apparent partial specific volume for DNA minicircle topoisomers in the buffer

	Partial specific volume[a] in mg/l	
Sample	336 bp	672 bp
Linear	0.479	ND
Relaxed	0.470	ND
Nicked	0.469	0.495
Supercoiled	0.488	0.494
Hypernegatively supercoiled	0.479	ND
Average value[b]	0.482 ± 0.011	

[a]Determined by global sedimentation velocity analysis using the known molar masses. [b]For all the 336 bp and 672 bp minicircle DNA species. ND, not determined.

tained trace amounts of nicked 336 bp (1 %) and supercoiled 672 bp (3 %) minicircle DNA. The topoisomer distribution obtained (mean $\sigma \sim -0.08$) reflects the supercoiling level in the bacterial strain used to generate the minicircles. The 'hypernegatively supercoiled' sample contained primarily $\Delta Lk = -6$ (61 %) and $\Delta Lk = -5$ (33 %) topoisomers, with trace amounts of nicked 336 bp (4 %) and supercoiled 672 bp (3 %) minicircle DNA. This sample (with mean $\sigma \sim -0.15$) is representative of the very high levels of dynamic supercoiling generated transiently during transcription.

The supercoiled 672 bp sample contained primarily $\Delta Lk = -4$ (63 %), $\Delta Lk = -5$ (24 %), $\Delta Lk = -6$ (6 %) topoisomers, and trace amounts of nicked 672 bp (2 %), $\Delta Lk = -2$ (4 %) and $\Delta Lk = -3$ (2 %) topoisomers. The identity of the topoisomers present in the supercoiled 672 bp sample was determined on a separate gel with 672 bp topoisomer markers. The topoisomer distributions of each sample are compiled in Supplementary Table S1.

Analytical ultracentrifugation of DNA minicircles

Apparent PSVs obtained from a global multispeed genetic algorithm-Monte Carlo analysis for each minicircle sample are summarized in Table 1. The derived PSV values did not show any apparent pattern that would indicate a dependence of the PSV on topoisomer conformation, and resulted in a near constant value of 0.482 ± 0.011 ml/g over all tested minicircles (see Table 1). The average PSV for 336 bp minicircles was 0.477 ml/g while the average PSV for 672 bp minicircles was 0.494 ml/g.

Using the determined average PSV value of 0.482 ml/g, a frictional ratio f/f_0 was derived from the sedimentation and diffusion coefficients obtained in the global analysis. Plots of the frictional ratio as a function of sedimentation coefficient and the diffusion coefficient as a function of the sedimentation coefficient are shown in Figure 2. DNA topology had a significant effect on sedimentation and diffusion coefficients. In contrast to polyacrylamide gel electrophoresis, for which linear migrates fastest (Figure 1), circular molecules (nicked, relaxed, supercoiled, and hypernegatively supercoiled) all sedimented faster than linear.

AUC was additionally able to differentiate relaxed, supercoiled, and hypernegatively supercoiled samples. The supercoiled or hypernegatively supercoiled samples containing a mixture of topoisomers behaved as single species in AUC. This result can be rationalized by noting that the topoisomers in experiments differ in Lk by 1 only. While the difference between nicked and supercoiled minicircles is clearly seen, it may be difficult to resolve by AUC minicircle topoisomers differing by only one or a few Lk.

Predicted shapes of DNA minicircles

Before employing our mathematical models to predict the effect of supercoiling and curvature on the hydrodynamic properties of DNA, we needed to first test how well these models predicted known equilibrium 3-D minicircle shapes. We adopted the strategy of building a coarse-grained representation of the equilibrium shapes of the minicircles obtained using our energy minimization codes. These models are reduced representations of macromolecules still capable of retaining key physical aspects (51). This approach has been highly successful in calculations of biomolecule properties in solution (52). Having found the shapes of DNA minicircles, we calculated the hydrodynamic radius for each conformation. The hydrodynamic radius was used to calculate the sedimentation and diffusion coefficients. We also tested our models against the 3-D structures of these minicircles, previously observed experimentally (13). Our aim was to develop a practical predictive theoretical framework to determine the measured transport coefficients.

Electrostatic screening and Brownian contributions. DNA molecules have substantial charge, which can be exploited, e.g., in electrophoretic measurements of DNA of different lengths. A qualitative comparison of the elastic and electrostatic forces is possible by considering scaling arguments. The Debye-Hückel equation is a well-established model of electrostatic interaction in a buffer containing counterions (53). In this approach, the interaction potential decays exponentially with separation due to screening. The decay rate is quantified by a characteristic distance, the Debye length (R_D). Comparing R_D with the typical distances between base pairs gives a crude estimation of the influence of electrostatic forces. We estimated R_D for our setup to be 1.45 Å from the ionic strength of 230 mM using an ionic strength-based estimate $R_D^{-1} = \kappa_0\sqrt{C_s}$, where C_s is the molar salt concentration in moles per liter and $\kappa_0 = 0.329$ $Å^{-1}l^{1/2}mol^{-1/2}$, as reported by Lim *et al.* (54). Thus the Debye length is much smaller than an average distance between different segments of the DNA molecule. Notably, this value of R_D was also much lower than that of earlier work, such as 30 Å in Ref. (55) for different buffer conditions.

The persistence length P of polyelectrolytes is the sum of two contributions,

$$P = P_{\text{in}} + P_{\text{el}}, \tag{13}$$

where P_{in} is an intrinsic persistence length due to the rigidity of the backbone, and P_{el} is an electrostatic persistence length, which accounts for buffer-dependent repulsion between neighboring ionic sites (56). The latter can be related to the Debye length as $P_{\text{el}} = R_D^2(4l_B)^{-1}$, with l_B being the Bjerrum length, according to the Odijk-Skolnick-Fixman

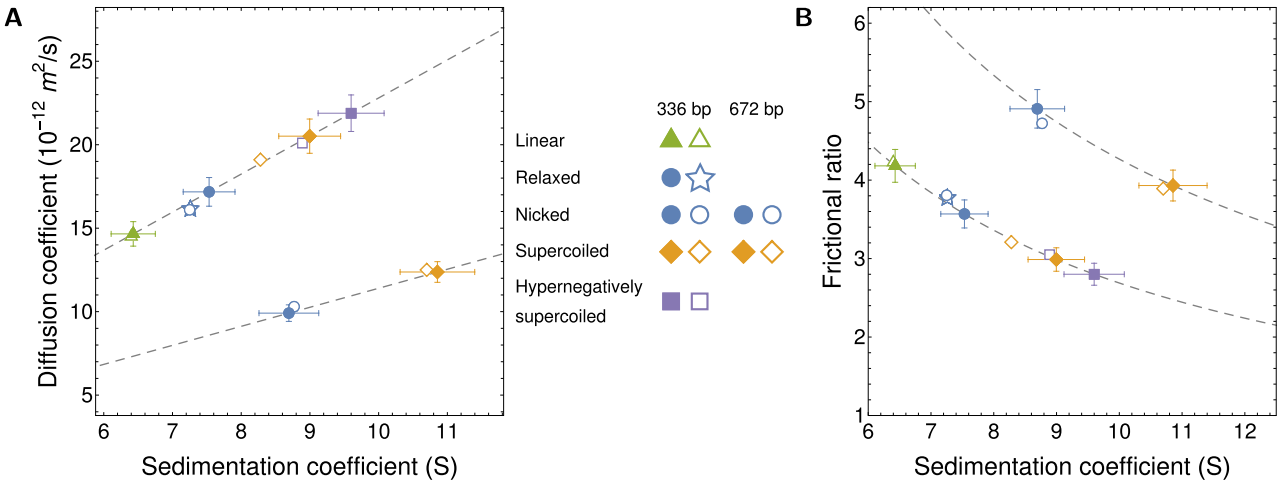


Figure 2. Measured and predicted diffusion and sedimentation coefficients for DNA minicircles. AUC measurements using global Monte Carlo-Genetic Algorithm analysis are marked as empty symbols. Theoretical predictions are presented with filled symbols. (**A**) Diffusion coefficient as a function of the sedimentation coefficient for topoisomers of 336 bp (upper branch) and 672 bp (lower branch) minicircle DNA. Dashed lines represent the constant mean value of PSV determined from AUC experiments upon which theoretical predictions of the sedimentation coefficient are based. Experimental and theoretical data for 336 bp and 672 bp relaxed and nicked minicircles overlay almost completely and thus are impossible to discern in the plots. (**B**) Frictional ratio as a function of the sedimentation coefficient. Sedimentation coefficients s are measured in svedberg units (S), with 1 S = 10^{-13} s.

theory (57–59). For linear or relaxed double-stranded DNA at room temperature, the persistence length in 0.1 M NaCl is approximately 500 Å (150 bp) (60). Based on persistence length, one can define effective bending and twisting energies for a circular shape as

$$E_{\text{bend}} = 4\pi^2 k_B T \frac{P}{L} = 4\pi^2 \frac{A_{\text{eff}}}{L} \quad \text{and} \tag{14}$$

$$E_{\text{twist}} = 6\pi^2 k_B T (\Delta Lk)^2 \frac{P}{L} = 6\pi^2 (\Delta Lk)^2 \frac{A_{\text{eff}}}{L}, \tag{15}$$

which, via eq. (13), include both elastic and electrostatic contributions. Here, $A_{\text{eff}} = k_B T P$ is the effective bending rigidity of the DNA. The remaining long-ranged electrostatic contribution E_{lr} can be estimated as the interaction energy between N equal charges q at a typical distance comparable to the radius of the loop, $L/2\pi$, which amounts to

$$E_{\text{lr}} \approx \frac{q^2 N}{2\epsilon L} \exp\left(-\frac{L}{2\pi R_D}\right), \tag{16}$$

where ϵ is the dielectric permittivity of the buffer. Note that these estimates do not take into account short-range electrostatic interactions between distant parts of the DNA that may come close together when supercoiled. For a circular loop with $|\Delta Lk| = 2$ and equivalent length of 336 bp, we found $E_{\text{bend}} \approx 17\, k_B T$, $E_{\text{twist}} \approx 105\, k_B T$, and a negligible value of E_{lr}. The effective bending and twisting energies considerably exceed the typical energy of thermal fluctuations, $k_B T$, which is somewhat surprising given that the minicircles are longer than P. This result may be a consequence of the relatively high energy stored in the form of bending energy E_{bend}, which renders the circular configuration less prone to Brownian shape disturbances than a torsional-stress-free linear DNA. We therefore use the stiff beam approximation to describe minicircle shapes.

Critical ΔLk for writhing. The shape of a DNA minicircle depends primarily on $|\Delta Lk|$ and the width to length ratio d_s/L. To predict 336 bp and 672 bp minicircle conformations for $\Delta Lk = 0$ to $|\Delta Lk| = 5$, we developed a theoretical framework that assumed that DNA can be modeled by a continuous elastic beam of ΔLk with the steric thickness d_s and stiffness A as determined by the inter-phosphate distance and persistence length, respectively. The previous theoretical study of Coleman and Swigon (42) focused on a circularized 718 bp DNA fragment and determined the structure of possible stable configurations, describing the contact diagrams in great detail. Our results were consistent with their findings for a particular thickness to length ratio ($d_s/L = 0.0082$) but were applied to an arbitrary aspect ratio.

In Figure 3A we present the shapes obtained by our minimization procedure. Close to $\Delta Lk = 0$, circular configurations were stable. At a critical value of $|\Delta Lk|$, a figure-8 shape became energetically favorable, and the loop writhed to relax excess twist. For higher $|\Delta Lk|$, the number of self-contacts increased, leading to a more writhed configuration. In this process, regions of high and low curvature emerged along the loop, as sketched in Figure 3B, with higher curvature present for higher values of $|\Delta Lk|$, as expected intuitively. For open configurations, the curvature was nearly constant and equal to $2\pi/L$. The position along the loop changed from 0 to 1 and was measured from the point of contact or (in the case of multiple contacts) from the center of symmetry. With increasing $|\Delta Lk|$, regions of increased curvature appear far away from self-contact points.

For larger $|\Delta Lk|$, increased curvature was present even at points close to the contact line. The variation of curvature along the loop centerline, shown in Figure 3C, might hint at sites of potential base pair instability or other configurational changes. The effect of localized curvature regions would be further compounded by the transmission

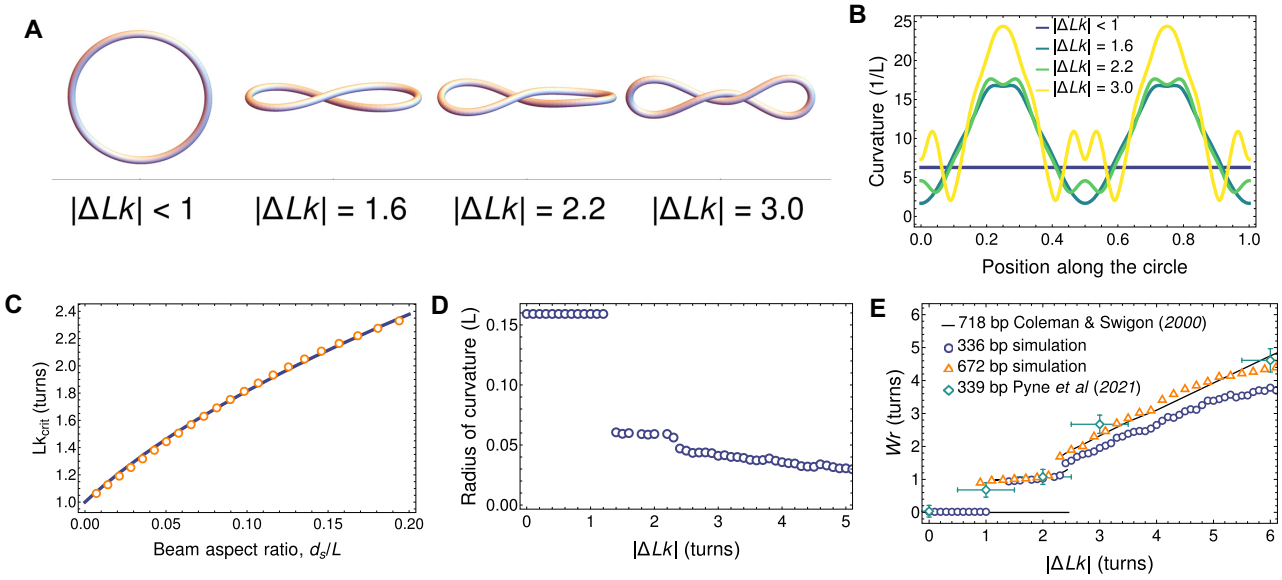


Figure 3. Elastic equilibrium shapes of model DNA minicircles. **(A)** Energy minimizing shapes of 336 bp minicircle of various ΔLk (with $d_s/L = 0.018$). For small $|\Delta Lk|$ values (<1.6), loops adopt open circle configurations. For intermediate values ($\Delta Lk = 1.6$ or 2.2), a single point of polymer contact is observed; for larger values of ΔLk (>2.2), continuous contact is observed. **(B)** Curvature distribution along the supercoiled loop centerline in shapes corresponding to panel (**A**). The position along the loop is measured from the point of contact (or the center of symmetry in multiply touching configurations). **(C)** Lk_{crit} above which a writhed configuration can be stable, plotted as a function of the beam aspect ratio (steric diameter to length, d_s/L). Solid line shows the approximation of eq. (17). **(D)** Minimal radius of curvature along the loop as a function of ΔLk for the 336 bp minicircle. This radius decreases monotonically with ΔLk, leading to increasing bending stresses. **(E)** Writhe of energy-minimizing shapes as a function of the aspect ratios $d_s/L = 0.018$, 0.0090, and 0.0082 for 336 bp, 672 bp, and 718 bp DNA minicircles respectively. For small values of $|\Delta Lk|$, only flat (open circle) configurations are permitted but above Lk_{crit}, one or more twists is relaxed by writhing. For large values of $|\Delta Lk|$, around 90 % of torsional energy is relaxed by shape deformation. Our results are shown next to the continuum model predictions of Coleman and Swigon (42) and atomistic MD simulations of Pyne *et al.* (19).

of mechanical stress along the DNA backbone to promote DNA kinking and base pair disruption at distant sites (13,19,61,62) as reported by Fogg *et al.* (12).

The value of Lk_{crit} for the open circle-figure-8 transition depends on the aspect ratio d_s/L. In Figure 3C, we show the dependence of Lk_{crit} on the aspect ratio numerically, showing that for thicker beams a transition to writhed configurations required an increased negative supercoiling. A very thin filament can be stably writhed for almost any value of $|\Delta Lk|$ greater than 1. However, for beams with a larger thickness, more torsional stress was required to stabilize writhed configurations and prevent them from unwrithing to a circular, open configuration.

For increased values of d_s/L, we also empirically found a convenient approximate expression for

$$Lk_{crit} = \sqrt{1 + 23.1\frac{d_s}{L}}, \qquad (17)$$

which we show in Figure 3C. To further characterize the writhed shapes, in Figure 3D we plotted the minimal radius of curvature of a twisted beam as a function of $|\Delta Lk|$. For low values of $|\Delta Lk|$, the constant value reflected the purely circular equilibrium. Above Lk_{crit}, the loop became writhed, with an increased curvature at the apices. The monotonic decrease of radius of curvature with $|\Delta Lk|$ showed that increasingly supercoiled beams tended to have tighter bends and, therefore, stored larger bending energy. Finally, in Figure 3E, we present the writhe of the resulting configuration, calculated with our method as a function of $|\Delta Lk|$. The values were calculated for both lengths–336 and 672 bp. For comparison, we also plotted the results of computations of Coleman and Swigon (42) for $L = 718$ bp as a solid line, and results of base pair-resolution MD simulations of Pyne *et al.* (19) for a 339 bp minicircle, confirming the observed trend. The total Wr as a function of $|\Delta Lk|$ seemed to be weakly dependent on the minicircle length. Writhing, therefore, emerged as a universal mechanism of stress release for twisted loops.

Shape and stability of supercoiled configurations. Although purely elastic considerations suggest a plethora of possible writhed configurations, we mostly observed the simple minicircle conformations in the cryoET measurements of Irobalieva *et al.* (13), rather than the more intricate shapes. The latter have higher energies and are thus less frequently realized (42). For a given ΔLk and at low temperatures, the shapes associated with higher energies were unfavorable compared to the ground state solution determined by our energy minimization procedure.

Equilibrium shapes measured in cryoET experiments in Ref. (13) were divided into eight groups termed open circle, open figure-8, figure-8, racquet, handcuffs, needle, rod, and other. For a quantitative comparison, we reduced the complexity by distinguishing only between contact-free (corresponding to open minicircle configurations) and self-touching (corresponding to writhed minicircle configurations) solutions. Using the elastic beam framework, we de-

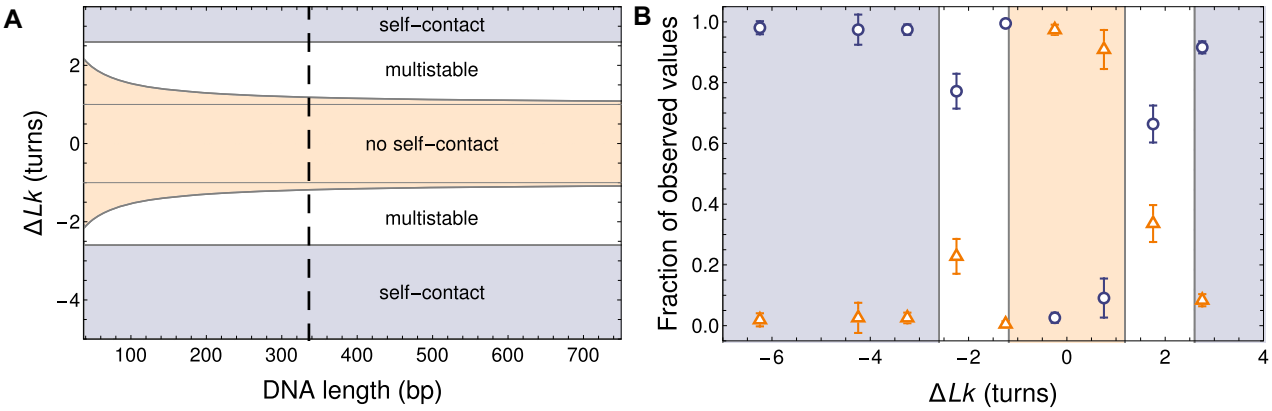


Figure 4. Regimes of shape stability for model DNA minicircles. (**A**) Phase diagram of different regimes of stability of supercoiled DNA as a function of ΔLk and DNA length. The phase space is divided into three regions: only open circular configurations permissible (tan-shaded area), both circular and writhed configurations permissible (multistability; white area), and only writhed configurations permissible (blue-shaded area). The dashed line marks the 336 bp minicircles, for which experimental shape data are available. (**B**) Plot of the relative occurrence rate of configurations with (blue circles) and without (orange triangles) points of self contact, based on cryoET shapes of 336 bp minicircles measured by Irobalieva *et al.* (13) and overlayed on our theoretical stability predictions with the same color coding as in panel (**A**). Error bars represent 2 standard deviations.

termined the regions of stability of these solutions in terms of ΔLk for a given DNA loop length. We present them in Figure 4A. The orange-shaded region close to $\Delta Lk = 0$ favors an open circle as the stable configuration. For undertwisted configurations, we identified a region of multistability, where we found both open circular and figure-8 shapes. Finally, when the inherent ΔLk was large enough, marked by the blue shading in the figure, the self-touching shapes became the only stable energy minimum. The stable circular shape region narrowed down from $|\Delta Lk| < 1.1$ for 336 bp minicircles to $|\Delta Lk| < 1.2$ for 672 bp and 718 bp minicircles.

For the particular length of $L = 336$ bp, in Figure 4B we compared the theoretical predictions to the cryoET data (13). We measured the fractions of open and self-touching configurations in the population of minicircle shapes. For ΔLk close to zero, we saw that most of the shapes were open loops, with only a fraction of about 10 % showing self-contact. This situation changed with increased ΔLk. In the predicted region of multistability, we saw that the fractions of writhed configurations increased, but there was still a pronounced population of open circular shapes that vanished almost completely when increasing $|\Delta Lk|$. This observation is in excellent agreement with our prediction for the new stability region at $|\Delta Lk| > 2.6$. In particular, the most relaxed state (closest to $\Delta Lk = 0$) was slightly undertwisted ($\Delta Lk < 0$). For configurations with $\Delta Lk = -0.2$ and $\Delta Lk = 0.8$, only open configurations were predicted while for $\Delta Lk = -2.2$, -1.2 and 1.8, we observed large conformational variability corresponding to two solutions in the uniform elasticity model. For $\Delta Lk = -4.2$, -3.2, and 2.8, open configurations were no longer permitted by the uniform elasticity theory and these open configurations were largely absent in the prior cryoET measurements for these topoisomers. Interestingly, an outlier at $\Delta Lk = -1.2$ was observed in the cryoET data, where we saw a surprising lack of open circular configurations, which could not be simply explained by our coarse-grained model. One potential explanation for this puzzling observation has been postulated that involves a coupling of limited base pair disruption with writhing (12).

Predictions of hydrodynamic radius of DNA minicircles with different ΔLk

Non-writhed configurations. Configurations with small values of ΔLk adopt toroidal conformations; for those configurations, we have found a convenient expression for the hydrodynamic radius, consistent with asymptotic solutions by Johnson and Wu (45) and Johnson (63)

$$R_h = \left(\frac{L}{2\pi} + \frac{d_h}{2}\right) \frac{72\pi\left(\frac{1}{16x^2}+1\right)x}{25x + 6(11x+8)\log 8x + 16}, \quad (18)$$

with $x = L/\pi d_h$. Here, log denotes the natural logarithm. Notably, Eq. (18) agrees with the numerical results of Goren and O'Neill (46) even for non-slender tori. For more slender tori, when $L/d_h > 30$, a simpler expression can be fitted without loss of accuracy, given by

$$R_h = L\left(\frac{11}{6}\log\left(\frac{2L}{d_h}\right) + 1.13\right)^{-1} \quad (19)$$

as obtained by Adamczyk *et al.* (64) from the numerical results of the bead model shape approximations. Qualitatively, R_h is of the same order of magnitude as the experimental DNA lengths and thus it approximately scales with the mass of the DNA molecule. This scaling for approximating DNA is in contrast with globular models used for some proteins that scale with the cubic root of their mass instead.

Writhed configurations. Configurations with intermediate values of $|\Delta Lk|$ adopted either toroidal or writhed shapes. To generate hydrodynamic predictions for DNA minicircles with larger values of $|\Delta Lk|$, we combined elastic energy minimization with hydrodynamic bead models. To this end, we used equilibrium shapes obtained for each $|\Delta Lk|$ within the elastic beam model described before, and produced their

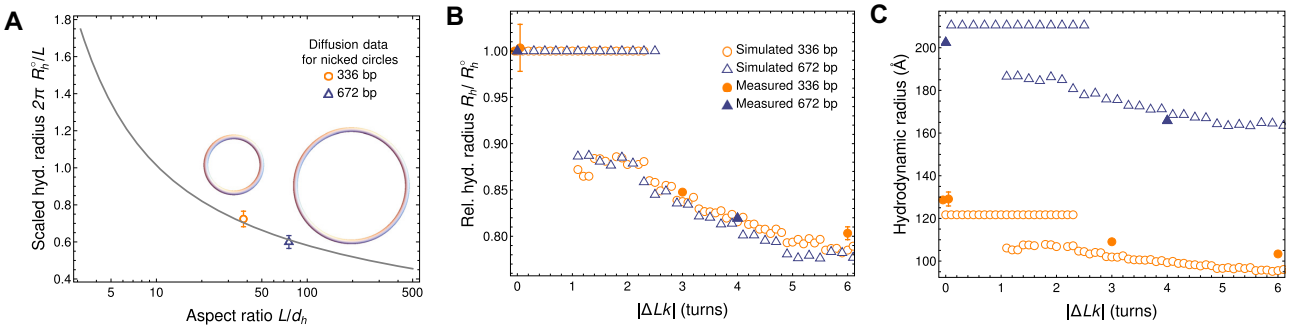


Figure 5. Hydrodynamic radius of DNA minicircles. (**A**) Hydrodynamic (hyd.) radius R_h° of open-circular DNA, scaled by the geometric radius $L/2\pi$ of a torus, plotted for a range of DNA aspect ratios L/d_h. Comparing the ZENO results for toroidal particles (solid line) to diffusion measurements of minicircles with 336 bp and 672 bp yields the fitted (common) hydrodynamic thickness $d_h = 29.4$ Å. We note that ZENO approximation yields high-precision results for toroidal particles (48). This value was used in all subsequent computations. Circular sketches representing molecules preserve both the relative scale and thickness. Note the logarithmic scale on the horizontal axis. (**B**) R_h of minicircle shapes, relative (rel.) to the hydrodynamic radius R_h° of the relaxed, open-circular shape for a range of $|\Delta Lk|$. We present results of simulations (open circles and triangles) along with experimental data for 336 bp (filled diamonds) and 672 bp (filled squares) minicircles. For 336 bp (nicked), 336 bp (relaxed), and 672 bp (nicked), the simulations show a region of constant R_h where the shape is independent of $|\Delta Lk|$. For supercoiled 336 bp containing a mixture of $\Delta Lk = -1$, -2 and -3, the R_h is about 15 % smaller than that of an open circle. For larger values of $|\Delta Lk|$, the theoretical approach seems to correctly grasp R_h of the resulting highly compact conformers. **(C)** Absolute values of the R_h for minicircles from (**B**), with the same symbols.

representation as a collection of 400 rigidly attached and overlapping spherical beads arranged such that the length and thickness of the DNA molecule were left unchanged. The bead model was then used to calculate the hydrodynamic mobility of the conglomerate within the ZENO package (47). We estimated the error of our predictions of the hydrodynamic radius to be about 5 % by comparing the results of bead-model calculations to known analytical solutions for highly symmetric shapes of model molecules.

Our modeling approach reduced the problem of finding diffusion and sedimentation coefficients to the computation of the hydrodynamic radius R_h. Once calculated theoretically, R_h was used to predict the diffusion coefficient from the Stokes-Einstein relation, Eq. (2), provided that the viscosity of the environment η is known. The presented results were adjusted for the buffer viscosity. The prediction of R_h for DNA of any shape required additionally the knowledge of the hydrodynamic thickness of the loop d_h. To determine the effective thickness of the molecules (related to the existence of a hydration layer), we assumed that, as a local property, it does not depend on the shape. Then, we used the measured AUC data for nicked and relaxed ($\Delta Lk = 0$) 336 bp and 672 bp minicircles, which we assumed to be toroidal. Knowing the dependence of the DNA R_h on the molecule aspect ratio from the ZENO software tool, we compared it to the AUC results for the R_h of the 336 bp and 672 bp DNA minicircles, and we fitted the same value of d_h to both minicircles to reproduce their experimental R_h. The dependence of R_h on the aspect ratio is presented in Figure 5A, together with the two measured values. Fitting the theoretical curve to these two data points yielded the hydrodynamic diameter of $d_h = 29.4$ Å (corresponding to $d_h/L = 2.6 \times 10^{-2}$ for the 336 bp minicircles and $d_h/L = 1.3 \times 10^{-2}$ for the 672 bp minicircles). We used this value for all subsequent calculations. We note, however, that the estimation of d_h from diffusion measurements cannot be precise due to the logarithmic dependence of hydrodynamic parameters on this value, so the fitted value should be treated as more approximate than the number of digits provided.

The value of excess thickness over inter-phosphate distance was significantly larger than that reported by Fernandes *et al.* (65) (22.8 Å)—without the details of the solvent ionic strength; however, as argued by Penkova *et al.* (66), the hydration shell can be as thick as 16 Å, corresponding to diameters as large as 40 Å. This value is sensitive not only to the ionic strength of the solvent, but also to the details of ion composition (67). Moreover, small deviations from the toroidal shape of the nicked and relaxed DNA shapes are expected, as caused by Brownian motion.

Figure 5B demonstrates that the difference of shapes with different $|\Delta Lk|$ is significant and similar in theory and experiments. We present therein the hydrodynamic radius for writhed configurations calculated for the two investigated DNA lengths with the same thickness ($d_h = 29.4$ Å). The plot shows the radius relative to that of a toroidal particle of the same d_h, denoted by R_h°. We normalized theoretical results by calculations at $\Delta Lk = 0$, while experimental data are rescaled by respective results for a nicked/relaxed configuration of a minicircle of a given length. Open circles and triangles in the graph mark theoretical results. For the 336 bp minicircle, two experimental values were available, namely that of a nicked and relaxed configuration. In these cases, we rescaled the experimental data by the mean of the two radii. The values of the relative radii are equal to unity for $\Delta Lk = 0$ by definition, but even the non-zero ΔLk simulations predicted a region of stable circular configurations with unchanged R_h.

Increasing $|\Delta Lk|$ led to the emergence of highly writhed configurations, which tended to be more compact and therefore had a smaller R_h than the open circular configurations seen at low $|\Delta Lk|$. We predict that topological writhing could reduce R_h by about 15 % in experimentally relevant conditions as shown in Figure 5B.

We additionally plotted available experimental results for relaxed and supercoiled minicircles, which compared fa-

Table 2. Comparison of predicted and measured diffusion and sedimentation coefficients in the buffer

Sample	$\|\Delta Lk\|$	D ($\mu m^2/s$)		s (S)	
		predicted	measured	predicted	measured
336 linear	not applicable	14.7 ± 0.7	14.5	6.4 ± 0.3	6.40
336 relaxed	0	17.2 ± 0.9	16.2	7.5 ± 0.4	7.26
336 nicked	0	17.2 ± 0.9	16.1 ± 0.4	7.5 ± 0.4	7.25 ± 0.02
336 supercoiled[a]	**3**, 2, 1	20.5 ± 1.0	19.0	9.0 ± 0.5	8.28
336 hypernegatively supercoiled[b]	**6**, 5	21.9 ± 1.1	20.1 ± 0.2	9.6 ± 0.5	8.89
672 nicked	0	9.9 ± 0.5	10.3	8.7 ± 0.4	8.77
672 supercoiled[c]	**4**, 5, 6	12.4 ± 0.6	12.5	10.9 ± 0.5	10.72

[a]Predictions done for the dominant species in the mix ($\Delta Lk = -3$) highlighted in bold. [b]Predictions done for the dominant species in the mix ($\Delta Lk = -6$). [c]Predictions done for the dominant species in the mix ($\Delta Lk = -4$). Where experimental errors are not shown, the confidence limits from the global Monte Carlo analysis of the AUC data were exactly 0. We roughly estimated the errors of theoretical values as 5 %. Sedimentation coefficients are measured in svedbergs, with 1 S = 10^{-13} s. Experimental confidence limits are reported only when the last digit of the result would be affected.

vorably with theoretical predictions. The theory seems to correctly determine R_h obtained from AUC measurements, which opens an efficient route to calculate the hydrodynamic transport coefficients also for highly writhed conformations. In Figure 5C we also presented the same results in absolute terms. Because the fitted value of hydrodynamic thickness in Figure 5A lies between the estimates based on 336 bp and 672 bp only, we see the value of R_h for $\Delta Lk = 0$ to be slightly overestimated by the simulation for 672 bp and underestimated for 336 bp. We emphasize that this difference is an effect of the thickness fitting procedure. This systematic difference can be caused by slightly non-toroidal DNA shapes caused by Brownian motion.

Our AUC data, together with the theoretical modeling shown in Figure 5, yield a length-invariant observation—the ratio of R_h of open circular (nicked or relaxed) minicircle topoisomers to the R_h of compact (supercoiled) topoisomers is approximately 5:4. This value holds true for both 336 bp and 672 bp minicircles, even though the length L is longer than the persistence length P for 672 bp minicircles, with $L \approx 4P$. This length-invariant ratio of 5:4 holds true as long as thermal effects are negligible. This theoretical result is particularly robust because it is independent of the viscosity of the buffer or the minicircle length. We attributed this robustness to the logarithmic dependence of the hydrodynamic models on the thickness of slender bodies. Additionally, the ratio of R_h of 336 bp linear to R_h of 336 bp nicked is approximately 7:6.

These results can also be viewed in absolute rather than relative terms (Figure 3), where the theoretical predictions for the values of the sedimentation and diffusion coefficients are displayed together with the experimental results. We obtained a good agreement for the larger of the two minicircles (672 bp) and for linear, relaxed, and nicked samples of the smaller minicircle (336 bp). For highly writhed configurations of the smaller minicircle, however, theoretical predictions suggest a more compact conformation than observed. This result could be attributed to the limits of applicability of the linear elasticity theory to the very tight bend radii necessary for energy minimization in these configurations, or to a rough estimate of shapes and width of nicked and relaxed minicircles.

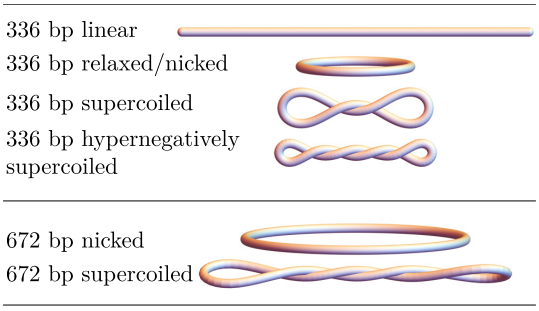


Figure 6. Sketches of model shapes in a given minicircle configuration used for hydrodynamic simulations with ΔLk specified in the caption of Table 2. The sketches have realistic aspect ratios (d_h/L) and preserve the relative size.

To test the effectiveness of our modeling approach, in Table 2 we compared the measured and predicted values of D and s. To compute D, data from nicked samples were used as a calibration of the only fitting parameter: the effective hydrodynamic diameter of the DNA molecule. With the use of d_h, all the D values were purely theoretical predictions based on the DNA length and the value determined in separate experiments.

Because supercoiled and hyper-supercoiled samples contain mixtures of minicircles with different ΔLk, theoretical predictions are given for the most common species in a mixture. These configurations are sketched in Figure 6. To calculate s, we used the approximately constant value of the PSV determined from the AUC experiments. In all cases, the deviation between theory and experiment was below 8 %. The simplified hydroelastic model provides practical estimates of the experimentally accessible quantities, and thus may be used to discern populations of minicircles differing in ΔLk.

Our modeling experiments reveal that the change of ΔLk by one turn (for example from $\Delta Lk = -4$ to $\Delta Lk = -5$) changes the frictional ratio by ~2 %—a value similar to theory-experiment deviations and similar to the noise level coming from the Monte Carlo optimizations of energy minimizing shapes. Such small gradients could be the reason why AUC measurements cannot distinguish topoisomers that differ in ΔLk by only a single turn, especially if the

concentrations of the different topoisomers in the mix differ. The samples analyzed by AUC contained a mixture of topoisomers that behaved as a single species because the major species in each sample were within one ΔLk of each other.

DISCUSSION

Methods that measure the hydrodynamic radius R_h, such as gel electrophoresis and diffusion-sedimentation AUC experiments, can be qualitatively described as sorting molecules by their size. Unlike for gel electrophoresis, however, when assessing diffusion via AUC, size contributions can be separated from viscosity, temperature, and PSV by the appropriate scaling, which yields R_h.

In electrophoresis, molecules squeeze their way through pores in the gel matrix. One of the major differences between agarose and polyacrylamide is the average pore size, with agarose gels typically having larger sized pores (68). In typical electrophoresis experiments, the primary determinant of the migration speed is the molecular weight and charge of the molecule. Here, the DNA molecule of a given length has a fixed charge and molecular weight, but the speed of migration through the matrix of small pores like the polyacrylamide gel varies greatly when the shape of the molecule is changed (although charge, and thus cations and hydration shell may be altered with negative supercoiling (Randall 2006; Randall 2009)). Comparing gel results to the AUC data allows us to confidently say that knowing the R_h is insufficient to predict its electrophoretic mobility as the gel electrophoresis and AUC separated the minicircle topoisomers differently. Supercoiling influences minicircle electrophoretic mobility much more than hydrodynamic mobility (as measured by AUC); the same change in supercoiling increases electrophoretic mobility by 400 % while increasing mobility only by 20 %. Differences between the methods can be even more dramatic. The linearized 336 bp minicircle showed 600 % increase of electrophoretic mobility, while at the same time (in the AUC measurements) having a 14 % *smaller* hydrodynamic mobility as compared to nicked or relaxed form.

Introducing circularization presents a significant obstacle to extending quantitative electrophoresis methods, such as one proposed by Ziraldo *et al.* (69). Calibrating gel electrophoresis measurements with a ladder of relaxed or nicked circular DNA of different lengths would be insufficient because the contribution of the degree of supercoiling strongly affects the apparent DNA mass derived from that method (at least by a factor of two). These difficulties are further compounded by the dependence on the absolute value of applied electrophoretic-mediated force, as reported by Iubini *et al.* (70). For large electrophoretic-mediated forces, linear DNA is expected to migrate faster than the circular form, which is opposite to what occurred in AUC, while for small forces, the circular form migrates faster than the linear, which is the same as in AUC. One possible explanation for the differences between electrophoresis and AUC is that as the DNA is pulled through the pores of the gel matrix, it has to change its shape (this would be most significant for the $\Delta Lk = 0$ samples) whereas there is no sieving in AUC. AUC, therefore, better reflects the solution properties of DNA, while gel electrophoresis offers higher resolution for separation by means we do not yet fully understand.

Diffusion-sedimentation measurements provided by AUC allow accurate hydrodynamic modeling. To elucidate the shapes and properties of DNA minicircles, we proposed a coarse-grained model that represents the DNA as an elongated, uniform elastic beam. By minimizing the elastic energy of a beam with a given superhelical density, we were able to find equilibrium shapes of the minicircles. We note that our coarse-grained models of minicircles are oblivious to their sequence and do not exploit information on sequence-dependent elastic properties. However, the uniform beam model of DNA elasticity presented here predicts shapes of DNA minicircles (shown in Figures 3A and 6) that compare favorably with the direct observations of 336 bp minicircles (13) (shown in Figure 4B).

Our theoretical model predicts a very weak dependence of the shapes of DNA minicircles with a few hundred base pairs on the aspect ratio d_s/L, as shown in Figure 4A. According to the model, the change in ΔLk, alone, gives sufficient information to describe basic features of the minicircle configuration: $\Delta Lk = 0$ is an open circle; $|\Delta Lk| = \pm 1$ is at the transition between open circle and writhed conformations; $|\Delta Lk| = \pm 2$ is multi-stable; and $|\Delta Lk| > 3$ always exhibits self-contact (compare with the corresponding shapes in Figures 3A and 6). This result means that no matter whether the minicircle was 336 or 672 bp, the loss of three helical turns was enough to disallow the open circle conformation as an energy-minimal solution. These results suggest that it is the absolute value of ΔLk and not the superhelical density $\sigma = \Delta Lk/Lk_0$ that governs the conformational landscape of small DNA minicircles (with the length having a small influence via the d_s/L ratio). This finding provides an important input to future models of circularized polymers.

Incorporating torsional interactions is more difficult than just bending interactions and is often neglected when constructing models of circularized molecules (71–74). Because the torsional forces play a role even when ΔLk is close to zero, care must be taken when generalizing such models for the context of supercoiled DNA. We have shown here that regardless of the length of the DNA molecule the torsional forces are of the same order of magnitude as pure bending forces.

While both bending and shearing forces remain comparable in magnitude even for longer minicircles, the relative importance of thermal fluctuations increases proportionally to the length. For the short minicircles considered here, thermal effects are not strong enough to push them out of their global energy minimum corresponding to the presented shapes. The results of Coleman and Swigon (42), however, show that mechanically stable branched plectonemic configurations exist that correspond to local energy minima of higher elastic energy. These high energy states can be accessed provided that the forces from thermal motion exceed the elastic forces, which is the case for minicircles of lengths much longer than the persistence length. For the minicircles studied here, we assumed that the minimal elastic energy shape was rigidly maintained throughout an experiment. To predict hydrodynamic properties of

longer DNA circles, prone to conformational flexibility due to Brownian effects, one has to use different modeling techniques, e.g., such as those inspired by models of intrinsically disordered proteins (75).

The elastic-energy minimizing centerline shapes were found for minicircles with different ΔLk with hydrodynamic thickness determined by calibration based on AUC experimental data. These shapes were then used to construct hydrodynamic models to compute R_h, which, in turn, was used to compare the theoretical diffusion and sedimentation coefficients with the AUC measurement. The comparison (Table 2) shows general agreement, which is satisfactory given the simplicity of the underlying coarse-grained model. These results confirm the predictive capabilities of uniform elasticity models, combined with a hydrodynamic calculations package (ZENO) to interpret and guide AUC measurements. We expect that the proposed modeling strategy could be beneficially employed to similar problems in the dynamics of DNA and perhaps extended further to account for sequence-specific effects and modified intramolecular interactions. One way forward would be to introduce local, sequence-dependent elasticity by allowing the bending stiffness of the elastic beam model to vary along the centerline of the minicircle. This extension should aid the studies of localized kinking, seen in earlier experiments (61), and the formation of bubbles in DNA minicircles under negative supercoiling (12).

This work represents significant progress in understanding and modeling the sedimentation of a biological molecule with a complex and dynamic conformation. Sedimentation of roughly spherical molecules (e.g., many proteins) is fairly well understood, but linear and minicircle DNA do not adopt a spherical conformation. Using 336 bp minicircles as a model system we are now able, for the first time, to fully test the theoretical models. We anticipate that this testing will allow us to further improve these models and expand the use of AUC to include a larger repertoire of important and complex biological molecules. We have successfully applied two stage modeling (combination of energy minimization to find the shape then applying hydrodynamic modeling for rigid configurations) to find hydrodynamic properties of a DNA molecule. Our approach developed here yielding the prediction of shape and hydrodynamic properties is applicable to other complex biomolecules.

CONCLUSION

The work presented here is a step toward understanding how supercoiling and curvature affect DNA shape and hydrodynamic properties, which in turn affect important DNA activities. The DNA solvation shell and counterions as well as DNA shape all affect DNA structure to influence how the DNA code is protected, accessed, modified, and activated. It will take a combination of approaches to fully understand this remarkable molecule.

DATA AVAILABILITY

The UltraScan software used to analyze the AUC data is open source and freely available from Github repository. The AUC data itself is available upon request from the UltraScan LIMS server at the Canadian Center for Hydrodynamics.

The algorithm for finding minimal energy shapes, initial conditions, and final configurations can be found in the Zenodo repository 7501673 (76).

Lk-Wr pairs computed from energy minimal shapes as well as hydrodynamic radii computed using the ZENO software can be found in the Zenodo repository Zenodo repository 7501675 (77).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The Authors thank Prof. De Witt Sumners and Dr. David Swigon for helpful remarks.

FUNDING

The work of RW and ML was supported by the National Science Center of Poland grant Sonata to ML no. 2018/31/D/ST3/02408. This work was also supported by the Canada 150 Research Chairs program (C150-2017-00015, BD), the National Institutes of Health (1R01GM120600, BD) and the Canadian Natural Science and Engineering Research Council (DG-RGPIN-2019-05637, BD). The Canadian Center for Hydrodynamics is funded by the Canada Foundation for Innovation (CFI-37589, BD). UltraScan supercomputer calculations were supported through NSF/XSEDE grant TG-MCB070039N (BD), and University of Texas grant TG457201 (BD). MLEJ was supported in part by National Science Center of Poland under grant UMO-2018/31/B/ST8/03640. The authors acknowledge funding (to LZ) from National Institutes of Health grant R35 GM141793 and National Science Foundation grant MCB 2107527.

Conflict of interest statement. Daniel J. Catanese, Jr., Jonathan M. Fogg, and Lynn Zechiedrich are co-inventors on issued and pending patents covering the supercoiled minicircle technology and uses and furthermore hold equity stake in Twister Biotech, Inc.

REFERENCES

1. Franklin,R.E. and Gosling,R.G. (1953) Molecular configuration in sodium thymonucleate. *Nature*, **171**, 740–741.
2. Lieberman-Aiden,E., van Berkum,N.L., Williams,L., Imakaev,M., Ragoczy,T., Telling,A., Amit,I., Lajoie,B.R., Sabo,P.J., Dorschner,M.O. *et al.* (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, **326**, 289–293.
3. Dorman,C.J. (2013) Genome architecture and global gene regulation in bacteria: making progress towards a unified model?. *Nat. Rev. Microbiol.*, **11**, 349–355.
4. Fisher,J.K., Bourniquel,A., Witz,G., Weiner,B., Prentiss,M. and Kleckner,N. (2013) Four-dimensional imaging of E. coli nucleoid organization and dynamics in living cells. *Cell*, **153**, 882–895.
5. Kleckner,N., Fisher,J.K., Stouf,M., White,M.A., Bates,D. and Witz,G. (2014) The bacterial nucleoid: nature, dynamics and sister segregation. *Curr. Opin. Microbiol.*, **22**, 127–137.
6. Dorman,C.J. (2019) DNA supercoiling and transcription in bacteria: a two-way street. *BMC Mol. Cell Biol.*, **20**, 26.

7. Baranello,L., Levens,D., Gupta,A. and Kouzine,F. (2012) The importance of being supercoiled: How DNA mechanics regulate dynamic processes. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, **1819**, 632–638.
8. Magnan,D. and Bates,D. (2015) Regulation of DNA replication initiation by chromosome structure. *J. Bacteriol.*, **197**, 3370–3377.
9. Drew,H.R., Weeks,J.R. and Travers,A.A. (1985) Negative supercoiling induces spontaneous unwinding of a bacterial promoter. *EMBO J.*, **4**, 1025–1032.
10. Kim,S.H., Ganji,M., Kim,E., van der Torre,J., Abbondanzieri,E. and Dekker,C. (2018) DNA sequence encodes the position of DNA supercoils. *eLife*, **7**, e36557.
11. Levens,D. (2008) How the c-myc promoter works and why it sometimes does not. *J. Natl. Cancer Ins. Monograp.*, **39**, 41–43.
12. Fogg,J.M., Judge,A.K., Stricker,E., Chan,H.L. and Zechiedrich,L. (2021) Supercoiling and looping promote DNA base accessibility and coordination among distant sites. *Nat. Commun.*, **12**, 5683.
13. Irobalieva,R.N., Fogg,J.M., Catanese,D.J., Sutthibutpong,T., Chen,M., Barker,A.K., Ludtke,S.J., Harris,S.A., Schmid,M.F., Chiu,W. *et al.* (2015) Structural diversity of supercoiled DNA. *Nat. Commun.*, **6**, 8440.
14. Fogg,J.M., Randall,G.L., Pettitt,B.M., Sumners,D. W.L., Harris,S.A. and Zechiedrich,L. (2012) Bullied no more: when and how DNA shoves proteins around. *Quart. Rev. Biophys.*, **45**, 257–299.
15. Bond,L.M., Peters,J.P., Becker,N.A., Kahn,J.D. and Maher,L.J. (2010) Gene repression by minimal lac loops in vivo. *Nucleic Acids Res.*, **38**, 8072–8082.
16. Czapla,L., Grosner,M.A., Swigon,D. and Olson,W.K. (2013) Interplay of protein and DNA structure revealed in simulations of the lac operon. *PLoS ONE*, **8**, e56548.
17. Garcia,H.G., Grayson,P., Han,L., Inamdar,M., Kondev,J., Nelson,P.C., Phillips,R., Widom,J. and Wiggins,P.A. (2007) Biological consequences of tightly bent DNA: The other life of a macromolecular celebrity. *Biopolymers*, **85**, 115–130.
18. Luger,K., Mäder,A.W., Richmond,R.K., Sargent,D.F. and Richmond,T.J. (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, **389**, 251–260.
19. Pyne,A.L.B., Noy,A., Main,K.H.S., Velasco-Berrelleza,V., Piperakis,M.M., Mitchenall,L.A., Cugliandolo,F.M., Beton,J.G., Stevenson,C. E.M., Hoogenboom,B.W. *et al.* (2021) Base-pair resolution analysis of the effect of supercoiling on DNA flexibility and major groove recognition by triplex-forming oligonucleotides. *Nat. Commun.*, **12**, 1053.
20. Basu,A., Bobrovnikov,D.G., Qureshi,Z., Kayikcioglu,T., Ngo,T.T.M., Ranjan,A., Eustermann,S., Cieza,B., Morgan,M.T., Hejna,M. *et al.* (2021) Measuring DNA mechanics on the genome scale. *Nature*, **589**, 462–467.
21. Li,K., Carroll,M., Vafabakhsh,R., Wang,X.A. and Wang,J.-P. (2022) DNAcycP: a deep learning tool for DNA cyclizability prediction. *Nucleic Acids Res.*, **50**, 3142–3154.
22. Demeler,B. and Gorbet,G.E. (2016) Analytical ultracentrifugation data analysis with UltraScan-III. In: Uchiyama,S., Arisaka,F., Stafford,W.F. and Laue,T. (eds.) *Anal. Ultracentr*. SpringerJapan, Tokyo, pp. 119–143.
23. Fogg,J.M., Kolmakova,N., Rees,I., Magonov,S., Hansma,H., Perona,J.J. and Zechiedrich,E.L. (2006) Exploring writhe in supercoiled minicircle DNA. *J. Phys. Cond. Matt.*, **18**, S145–S159.
24. Fuller,F.B. (1978) Decomposition of the linking number of a closed ribbon: A problem from molecular biology. *Proc. Natl. Acad. Sci.*, **75**, 3557–3561.
25. Savelyev,A., Gorbet,G.E., Henrickson,A. and Demeler,B. (2020) Moving analytical ultracentrifugation software to a good manufacturing practices (GMP) environment. *PLOS Comput. Biol.*, **16**, e1007942.
26. Cao,W. and Demeler,B. (2005) Modeling analytical ultracentrifugation experiments with an adaptive space-time finite element solution of the Lamm equation. *Biophys. J.*, **89**, 1589–1602.
27. Cao,W. and Demeler,B. (2008) Modeling analytical ultracentrifugation experiments with an adaptive space-time finite element solution for multicomponent reacting systems. *Biophys. J.*, **95**, 54–65.
28. Brookes,E., Cao,W. and Demeler,B. (2010) A two-dimensional spectrum analysis for sedimentation velocity experiments of mixtures with heterogeneity in molecular weight and shape. *Eur. Biophys. J.*, **39**, 405–414.
29. Demeler,B. (2010) Methods for the design and analysis of sedimentation velocity and sedimentation equilibrium experiments with proteins. *Curr. Protoc. Protein Sci.*, **60**, 7.13.1–7.13.24.
30. Brookes,E.H. and Demeler,B. (2007) Parsimonious regularization using genetic algorithms applied to the analysis of analytical ultracentrifugation experiments. In: *Proceedings of the 9th annual conference on genetic and evolutionary computation - GECCO '07, London*. pp. 361–368.
31. Demeler,B. and Brookes,E. (2008) Monte Carlo analysis of sedimentation experiments. *Coll. Pol. Sci.*, **286**, 129–137.
32. Williams,T.L., Gorbet,G.E. and Demeler,B. (2018) Multi-speed sedimentation velocity simulations with UltraScan-III. *Eur. Biophys. J.*, **47**, 815–823.
33. Gorbet,G.E., Mohapatra,S. and Demeler,B. (2018) Multi-speed sedimentation velocity implementation in UltraScan-III. *Eur. Biophys. J.*, **47**, 825–835.
34. Eisenberg,H. (2003) Modern analytical ultracentrifugation in protein science: Look forward, not back. *Prot. Sci.*, **12**, 2647–2649.
35. Van Holde,K.E. (1985) Physical Biochemistry, Prentice-Hall, Englewood Cliffs.
36. Durchschlag,H. (1986) Specific volumes of biological macromolecules and some other molecules of biological interest. In: Hinz,H.-J. (ed.) *Thermodynamic data for Biochemistry and Biotechnology*. Springer Verlag, Berlin, Heidelberg.
37. Durchschlag,H. (1989) Determination of the partial specific volume of conjugated proteins. *Coll. Pol. Sci.*, **267**, 1139–1150.
38. Korolev,N., Lyubartsev,A.P., Rupprecht,A. and Nordenskiöld,L. (1999) Competitive Binding of Mg^{2+}, Ca^{2+}, Na^{+}, and K^{+} Ions to DNA in Oriented DNA Fibers: Experimental and Monte Carlo Simulation Results. *Biophys. J.*, **77**, 2736–2749.
39. Landau,L.D. and Lifshitz,E.M. (2009) *Theory of elasticity, Elsevier, Butterworth-Heinemann*. Amsterdam
40. Călugăreanu,G. (1959) L'intégrale de Gauss et l'Analyse des nœuds tridimensionnels. *Revue Roumaine de Mathématiques Pures et Appliquées*, **4**, 5–20.
41. Bates,A.D. and Maxwell,A. (2005) *DNA topology*. Oxford University Press, Oxford
42. Coleman,B.D. and Swigon,D. (2000) Theory of supercoiled elastic rings with self-contact and its application to DNA plasmids. *J. Elas.*, **60**, 173–221.
43. Zajac,E.E. (1962) Stability of two planar loop elasticas. *J. Appl. Mech.*, **29**, 136–142.
44. Levitt,M. (1983) Protein folding by restrained energy minimization and molecular dynamics. *J. Mol. Biol.*, **170**, 723–764.
45. Johnson,R.E. and Wu,T.Y. (1979) Hydromechanics of low-Reynolds-number flow. Part 5. Motion of a slender torus. *J. Fluid Mech.*, **95**, 263–277.
46. Goren,S.L. and O'Neill,M.E. (1980) Asymmetric creeping motion of an open torus. *J. Fluid Mech.*, **101**, 97–110.
47. Juba,D., Audus,D.J., Mascagni,M., Douglas,J.F. and Keyrouz,W. (2017) ZENO: Software for calculating hydrodynamic, electrical, and shape properties of polymer and particle suspensions. *J. Res. Natl. Ins. Stand. Tech.*, **122**, 20.
48. Hubbard,J.B. and Douglas,J.F. (1993) Hydrodynamic friction of arbitrarily shaped Brownian particles. *Phys. Rev. E*, **47**, R2983–R2986.
49. Rybenkov,V.V., Vologodskii,A.V. and Cozzarelli,N.R. (1997) The effect of ionic conditions on DNA helical repeat, effective diameter and free energy of supercoiling. *Nucleic Acids Res.*, **25**, 1412–1418.
50. Xu,Y.C. and Bremer,H. (1997) Winding of the DNA helix by divalent metal ions. *Nucleic Acids Res.*, **25**, 4067–4071.
51. Pak,A.J. and Voth,G.A. (2018) Advances in coarse-grained modeling of macromolecular complexes. *Cur. Opin. Struct. Biol.*, **52**, 119–126.
52. de la Torre,J.G. and Bloomfield,V.A. (1981) Hydrodynamic properties of complex, rigid, biological macromolecules: theory and applications. *Quart. Rev. Biophys.*, **14**, 81–139.
53. Hückel,E. and Debye,P. (1923) The theory of electrolytes: I. Lowering of freezing point and related phenomena. *Physikalische Zeitschrift*, **24**, 185–206.
54. Lim,S., Kim,Y. and Swigon,D. (2011) Dynamics of an electrostatically charged elastic rod in fluid. *Proc. Roy. Soc. A: Math., Phys. Eng. Sci.*, **467**, 569–590.

55. Robertson,R.M., Laib,S. and Smith,D.E. (2006) Diffusion of isolated DNA molecules: Dependence on length and topology. *Proc. Natl. Acad. Sci.*, **103**, 7310–7314.
56. Mantelli,S., Muller,P., Harlepp,S. and Maaloum,M. (2011) Conformational analysis and estimation of the persistence length of DNA using atomic force microscopy in solution. *Soft Matt.*, **7**, 3412.
57. Odijk,T. (1977) Polyelectrolytes near the rod limit. *J. Poly. Sci.: Poly. Phys. Edn.*, **15**, 477–483.
58. Skolnick,J. and Fixman,M. (1977) Electrostatic persistence length of a wormlike polyelectrolyte. *Macromolecules*, **10**, 944–948.
59. Manning,G.S. (2001) Counterion condensation on a helical charge lattice. *Macromolecules*, **34**, 4650–4655.
60. Manning,G.S. (2006) The persistence length of DNA is reached from the persistence length of its null isomer through an internal electrostatic stretching force. *Biophys. J.*, **91**, 3607–3616.
61. Lionberger,T.A., Demurtas,D., Witz,G., Dorier,J., Lillian,T., Meyhöfer,E. and Stasiak,A. (2011) Cooperative kinking at distant sites in mechanically stressed DNA. *Nucleic Acids Res.*, **39**, 9820–9832.
62. Wang,Q., Irobalieva,R.N., Chiu,W., Schmid,M.F., Fogg,J.M., Zechiedrich,L. and Pettitt,B.M. (2017) Influence of DNA sequence on the structure of minicircles under torsional stress. *Nucleic Acids Res.*, **45**, 7633–7642.
63. Johnson,R.E. (1980) An improved slender-body theory for Stokes flow. *J. Fluid Mech.*, **99**, 411–431.
64. Adamczyk,Z., Sadlej,K., Wajnryb,E., Ekiel-Jeżewska,M.L. and Warszyński,P. (2010) Hydrodynamic radii and diffusion coefficients of particle aggregates derived from the bead model. *J. Colloid Interface Sci.*, **347**, 192–201.
65. Fernandes,M.X. (2002) Calculation of hydrodynamic properties of small nucleic acids from their atomic structure. *Nucleic Acids Res.*, **30**, 1782–1788.
66. Penkova,N.A., Sharapov,M.G. and Penkov,N.V. (2021) Hydration shells of DNA from the point of view of terahertz time-domain spectroscopy. *Int. J. Mol. Sci.*, **22**, 11089.
67. Laage,D., Elsaesser,T. and Hynes,J.T. (2017) Water dynamics in the hydration shells of biomolecules. *Chem. Rev.*, **117**, 10694–10725.
68. Stellwagen,N.C. (2009) Electrophoresis of DNA in agarose gels, polyacrylamide gels and in free solution. *Electrophoresis*, **30**, S188–S195.
69. Ziraldo,R., Shoura,M.J., Fire,A.Z. and Levene,S.D. (2019) Deconvolution of nucleic-acid length distributions: a gel electrophoresis analysis tool and applications. *Nucleic Acids Res.*, **47**, e92–e92.
70. Iubini,S., Orlandini,E., Michieletto,D. and Baiesi,M. (2018) Topological sieving of rings according to their rigidity. *ACS Macro Lett.*, **7**, 1408–1412.
71. Waszkiewicz,R., Szymczak,P. and Lisicki,M. (2021) Stability of sedimenting flexible loops. *J. Fluid Mech.*, **919**, A14.
72. Gruziel,M., Thyagarajan,K., Dietler,G., Stasiak,A., Ekiel-Jeżewska,M.L. and Szymczak,P. (2018) Periodic motion of sedimenting flexible knots. *Phys. Rev. Lett.*, **121**, 127801.
73. Gruziel-Słomka,M., Kondratiuk,P., Szymczak,P. and Ekiel-Jeżewska,M.L. (2019) Stokesian dynamics of sedimenting elastic rings. *Soft Matter*, **15**, 7262–7274.
74. Gruziel-Słomka,M., Kondratiuk,P., Szymczak,P. and Ekiel-Jeżewska,M.L. (2022) Correction: Stokesian dynamics of sedimenting elastic rings. *Soft Matter*, **18**, 4811–4811.
75. Cichocki,B., Rubin,M., Niedzwiecka,A. and Szymczak,P. (2019) Diffusion coefficients of elastic macromolecules. *J. Fluid Mech.*, **878**, R3.
76. Waszkiewicz,R. (2023) RadostW/twisted-loop: Manuscript version. Zenodo, https://doi.org/10.5281/zenodo.7501673.
77. Waszkiewicz,R. (2023) RadostW/twisted-dna-results: Manuscript version. Zenodo, https://doi.org/10.5281/zenodo.7501675.

3.3.1 Thermal Effects on the Shapes of DNA Minicircles

The model presented in this publication [3] has a limited assessment of the influence of thermal fluctuations on the shape and hydrodynamics of DNA loops. This was caused by lack of readily available tools for Brownian Dynamics simulation under non-local force fields—such as the energy functional depending on writhe, which is a generically non-local quantity.

Using the later developed `pychastic` package or Ref. [4], we conducted simulations to explore the shapes of DNA minicircles at finite (room) temperatures and computed their apparent diffusion coefficient using the Minimum Dissipation method derived by Cichocki et al. [25].

In these simulations, we assumed that torsional stresses in the DNA are relaxed. Consequently, the total elastic energy E_{el} of such a minicircle is given by

$$E_{el} = \frac{1}{2}\int_0^L EI\left(\kappa^2 + \omega\left(\frac{2\pi(Lk - Wr)}{L}\right)^2\right) dl, \tag{3.1}$$

with the notation explained in Ref. [3]. Both κ and Wr can be computed from the shape of the centreline alone (or a suitable discretisation thereof), thus reducing the dimensionality of the problem, as compared to an approach where both the position and orientation of filament segments are simulated. It is worth noting that simulating rotational dynamics, even of a single body, is a complex task, as outlined in Waszkiewicz et al. [4]. Given that Wr is a nonlocal quantity, the computation of forces from E_{el} can be cumbersome. Fortunately, this task can be accomplished with the assistance of `jax.grad`, which is compatible with the SDE solver `pychastic`.

We thus get a further insight into the intermediate region of linking numbers ($1 < Lk < 2.5$) where the DNA minicircles have multiple stable configurations at absolute zero temperature. We discussed the zero temperature limit in Waszkiewicz et al. [3].

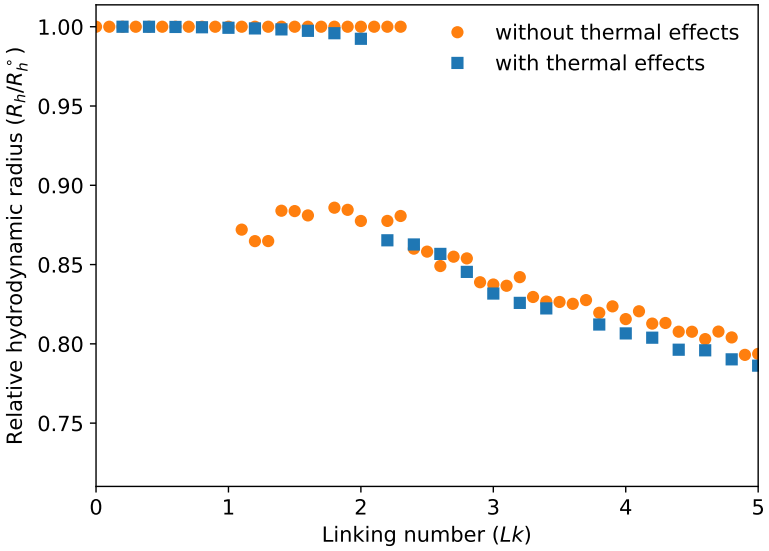


Figure 3.1: Hydrodynamic radius relative to $Lk = 0$ configuration of the 336bp minicircle as a function of the linking number.

It turns out that the effect of thermal fluctuations on the overall hydrodynamic radius outside of the intermediate region are negligible, as shown in Fig. 3.1. Moreover, the range of values of Lk for which the equilibrium ensemble contains both open configurations and configurations with self-contact is very small—less than 0.2 turns. This contrasts starkly with CryoEM data of Irobalieva et al. [46], where a multitude of conformations were observed for a wide range of Lk values. One possible explanation for this discrepancy lies in the temperature dependence of the geometric properties of DNA, as hinted at by the results of Ranasinghe et al. [50].

The example above highlights our motivations for the development of the `pychastic` package which is described in detail in the next publication.

3.4 Paper IV: *Pychastic: Precise Brownian dynamics using Taylor-Itō integrators in Python*

PAPER IV

"*Pychastic: Precise Brownian dynamics using Taylor-Itō integrators in Python*"

Radost Waszkiewicz, Maciej Bartczak, Kamil Kolasa, and Maciej Lisicki

COMMENTARY

This paper focuses on the novel implementation of numerical integration schemes for Brownian Dynamics simulations. The major theoretical limitation in our earlier studies revolved around an incomplete, approximate treatment of Brownian motion. In Ref. [1], we simplified the dynamics to a 2D case due to Mathematica's inability to handle full $SO(3)$ dynamics, as outlined in the publication below. Similarly, in Ref. [3], the challenges arose from the nature of the Writhe quantity, leading to a non-local force field that made assessing the effects of thermal fluctuations on the shape of minicircles exceptionally challenging.

Addressing the treatment of Brownian motion through stochastic differential equations presented a well-posed mathematical problem, with a range of solutions described for example in Ref. [74]. However, surprisingly, there was no readily available, easy-to-use software solution, such as the package described in the publication below.

We took advantage of the modern Python features such as introspection and metaprogramming, which allowed efficient automatic differentiaion of functions combined with the package `jax` for just-in-time compilation of Python code to machine code for performance. The resultant package, `pychastic`, is now available to instal via `pip` or directly from GitHub [61].

In this study, the PhD candidate: conceptualised the problem, led a small programming team, which included the author, Maciej Bartczak, and Kamil Kolasa, co-developed a Python package complete with documentation, examples, and tests, created test cases and debugging tools that facilitated the correction of multiple typos found in existing literature (outlined in the publication). He co-designed illustrative examples showcasing concrete applications of the tools developed during the Thesis preparation. Additionally, he wrote the first draft and edited all subsequent versions of the manuscript.

Pychastic: Precise Brownian dynamics using Taylor-Itō integrators in Python

Radost Waszkiewicz*, Maciej Bartczak, Kamil Kolasa and Maciej Lisicki

Institute of Theoretical Physics, Faculty of Physics, University of Warsaw
L. Pasteura 5, 02-093 Warsaw, Poland

* radost.waszkiewicz@fuw.edu.pl

Abstract

In the last decade, Python-powered physics simulations ecosystem has been growing steadily, allowing greater interoperability, and becoming an important tool in numerical exploration of physical phenomena, particularly in soft matter systems. Driven by the need for fast and precise numerical integration in colloidal dynamics, here we formulate the problem of Brownian Dynamics (BD) in a mathematically consistent formalism of the Itō calculus, and develop a Python package to assist numerical computations. We show that, thanks to the automatic differentiation packages, the classical truncated Taylor-Itō integrators can be implemented without the burden of computing the derivatives of the coefficient functions beforehand. Furthermore, we show how to circumvent the difficulties of BD simulations such as calculations of the divergence of the mobility tensor in the diffusion equation and discontinuous trajectories encountered when working with dynamics on S^2 and $SO(3)$. The resulting Python package, Pychastic, is capable of performing BD simulations including hydrodynamic interactions at speeds comparable to dedicated implementations in lower-level programming languages, but with a much simpler end-user interface.

Received 12-09-2022
Accepted 30-01-2023
Published 03-04-2023
doi:10.21468/SciPostPhysCodeb.11

Check for updates

This publication is part of a bundle: Please cite both the article and the release you used.

DOI	Type
doi:10.21468/SciPostPhysCodeb.11	Article
doi:10.21468/SciPostPhysCodeb.11-r0.2	Codebase release

Contents

1 Introduction

The dynamics of soft and colloidal matter systems is of importance for numerous technological and industrial processes, such as food products, pastes, creams, and gels [1]. Another important example are biological systems which involve aqueous suspensions of colloids, macromolecules, polymers, and cells. The diversity of the constituent elements, together with the tunability of their direct interactions, and the presence of hydrodynamic interactions (HI) mediated by the suspending fluid [2], gives rise to a multitude of complex dynamic phenomena, which can only be explored using appropriate numerical techniques [3,4].

The choice of a suitable method derives from the characteristic time scales of the investigated dynamics [5]. In the context of macromolecules, when short-time effects are of interest, Molecular Dynamics (MD) simulations are a popular choice [6]. Short times may be comparable to the solvent relaxation time scale, which for a fluid of kinematic viscosity ν and speed of sound c scales as ν/c^2. For water, this time is of the order of 10^{-13} s. The idea behind MD simulations is to solve Newtonian equations of motion for atoms or molecules, which are a set of second-order ODEs with effective interaction potentials. In this case, HI can be resolved using either explicit solvent methods that resolve the molecular structure of the solvent or more approximate implicit solvent models. However, the typical time scales of colloidal motion are much longer; the velocity relaxation time for colloidal particles of size a and density comparable to that of the fluid is approximately a^2/ν, amounting to ca. 10^{-8} s for a 100 nm colloid. On coarser time scales, the colloidal velocity relaxes multiple times, and the Rayleigh description in terms of velocity becomes irrelevant [5,7]. Instead, the description in terms of the position of a Brownian particle becomes possible on diffusive time scales, during which a particle diffuses over a distance of its own radius. This time scale is $a^2/D \sim \eta a^3/k_B T$, where η is the dynamic viscosity of the fluid, k_B is the Boltzmann constant, and T is the temperature. For the aforementioned particle, this time scale is ca. 10^{-3} s, and corresponds to the minimal time resolution of typical light scattering and microscopy experiments. Thus, the description of the dynamics on such coarse scales is made solely using the position of the particle. The clear separation of discussed time scales, combined with the numerical stiffness of Newton's equations, renders MD simulations not applicable to explore diffusive dynamics.

Instead, BD simulations present a convenient alternative, building on models suitable for

diffusive time scales, and involving accordingly long time steps for numerical computations. Since the velocity of particles varies very rapidly on such time scales, the position becomes a random variable whose properties are dictated by the fluctuation-dissipation theorem [7].

The development of a suitable theoretical description to explain the Brownian motion and its subsequent experimental verification at the beginning of the 20th century were a groundbreaking step that confirmed the atomic structure of matter [8]. Early works by pioneers such as Einstein [9], Smoluchowski [10], Fokker [11], and Planck [12], led to the description of diffusive processes through the underlying probability density functions (PDFs) for the positions of the particles and their temporal evolution, rather than representations of individual trajectories. The realisation that the path of a Brownian particle resembles a nowhere differentiable function paved the route for a different approach, now called stochastic calculus, with the classical works of Kolmogorov [13], Wiener, and finally Itō [14]. With these tools, the problem of Brownian motion can be recast in the form of stochastic differential equations (SDEs), which offer a practical route for numerical simulations.

In this contribution, we present our approach to the implementation of such Brownian dynamics. Our goal was to show how modern programming techniques available in Python: reflective and functional programming paradigms allow one to develop concise implementations of physically relevant problems. These constitute relatively recent developments in computer modelling, and we follow this trend to facilitate faster modelling of diffusive phenomena.

First, in Sec. 2 we review three popular ways of describing the Brownian motion using the Langevin equation, the Fokker-Planck equation, and Itō equation. We then provide an overview of available numerical integration packages for SDEs in Sec. 3.1. Next, we present our new package, Pychastic, in Sec. 3, along with three examples (in Sec. 4) of problems in which the packages greatly facilitate numerical computations. We conclude the work in Sec. 5.

2 Three vantage points: Langevin, Fokker-Planck, and Itō

Perhaps the most popular approach to modelling the trajectories of Brownian particles dates back to Langevin's work of 1908 [15]. The Langevin equation can be rationalised as an extension of Newtonian mechanics to include the effects of fluctuations by adding a stochastic force, F_n, acting on a particle. We write it below for a particle with a constant friction coefficient ζ. The average value of the force is zero, while its covariance at temperature T is given by the fluctuation-dissipation theorem [7], $\langle F_n(t)F_n(t')\rangle = 2k_BT\zeta\delta(t-t')$, with k_B being the Boltzmann constant. The equation of motion of the particle in the presence of a deterministic force F reads then

$$m\ddot{x} = -\zeta\dot{x} + F(x,t) + F_n(t)\,. \tag{1}$$

For the case of a spherical particle of radius a in Stokes flow the friction coefficient is $6\pi\eta a$, η being the dynamic viscosity. We note that even though $x(t)$ is not differentiable once, and certainly not differentiable twice at any point even in the usual distributional sense (because the Lebesgue integral requires finite variation and realisations of the Wiener process have infinite variation), Eq. (1) can be given a proper interpretation by transforming it into an integral form. However, the above equation proved practical for numerical calculations, e.g. using an integration scheme analogous to the Euler-Maruyama method.

An alternative description via the Fokker-Planck equation circumvents the difficulties of interpreting the dynamics on a single trajectory in the Langevin equation by focussing on a probabilistic description. On the Brownian time scale, the position of a particle is now treated as a random variable, and the probability density function (PDF) $P(x,t)$ of finding the particle

at a location x at a time t evolves according to the partial differential equation

$$\frac{\partial P(x,t)}{\partial t} = \zeta^{-1}\frac{\partial}{\partial x}\left(F(x,t)P(x,t)\right) + D\frac{\partial^2 P(x,t)}{\partial x^2}\,, \tag{2}$$

with the diffusion coefficient defined by $D = k_B T/\zeta$. The problem is now well-posed mathematically, and the tools of mathematical analysis of PDEs can be employed to study the evolution of the underlying probability distribution, e.g. by examining its moments. However, questions involving individual trajectories become less straightforward: problems involving first passage time have to be dealt with with a careful treatment of the boundary conditions (e.g. substances vanishing in chemical reactions). Moreover, from the point of view of numerical simulations, this approach becomes impractical for large systems. Using finite-difference methods, for N particles in a 3D simulation box of size L and mesh size Δx, we typically require $(L/\Delta x)^{3N}$ points to track the PDF, which may become prohibitively large.

The mathematical difficulties of the Langevin equation are absent in the proper treatment of SDEs in the Itō formalism. If now dX denotes the position increment of a particle in a time interval dt, we can write it as

$$dX = F dt + \sqrt{2D}dW\,, \tag{3}$$

where $F dt$ denotes the systematic drift of the particle and $\sqrt{2D}dW$ denotes the stochastic (diffusive) component of motion (provided that D is constant and the metric tensor is constant as well). This equation is meant in the distributional sense with respect to the Itō integral, that is

$$X(T) = \int_0^T dX = \int_0^T F(X,t)dt + \int_0^T \sqrt{2D}dW\,. \tag{4}$$

Here, all integrals are taken in the Itō sense. Such a trajectory-focused formulation effectively deals with the ill-defined derivatives in the Langevin equation. Using Itō's lemma imposes formal rules of transformation of the coefficients, freeing us from a canonical coordinate description. Finally, the estimation of observable quantities such as expected values, correlation functions or equilibrium distributions of low-dimensional projections of evolving variables can be recovered using a Monte Carlo approach, which for M simulations converges as $\sqrt{M}$, regardless of dimensionality. The constant of proportionality of this convergence is controlled only by the variance of the variable of interest, which is often independent of the dimensionality of the equation. For example, in the problem of diffusion of a single particle in a semidilute suspension, the variance of the mean squared displacement of the tracer is weakly dependent on the number of particles in the simulation volume.

3 Pychastic: description of the package

3.1 Available numerical integration packages

Only a few SDE integration packages have been made available in recent years. Two notable examples are `DifferentialEquations.jl` for Julia and `ItoProcess` being a part of Mathematica. Now, we present `Pychastic` for Python, which takes advantage of the popularity of this language. Our package source code is available on Github, up-to-date documentation on ReadTheDocs and ready to install via `pip` via Python Package Index. As a preliminary comparison, we note that `DifferentialEquations.jl` has the largest variety of integrators (for example, many options for stiff equations), while Mathematica's `ItoProcess` and `Pychastic` contain essentially the same algorithms. However, the biggest drawback of `ItoProcess` is the lack of step post-processing. When working with SDEs defined on manifolds whose universal cover is not $\mathbb{R}^n$, such as a sphere S^2 or the space of rigid rotations $SO(3)$, any parameterisation

Table 1: Available SDE integration packages.

Package	`DifferentialEquations.jl`	`ItoProcess`	`Pychastic`
Language	Julia	Mathematica	Python
License	MIT	proprietary	MIT
Codebase	open	closed	open
scalar SDEs	yes	yes	yes
vector SDEs	yes	yes	yes
strong convergence	up to order 1.5	up to order 1.5	up to order 1.5
weak convergence	up to order 2.0	up to order 2.0	up to order 2.0
supports events	yes	no	yes

of $\mathbb{R}^n$ will contain singularities, thus an integrator which cannot handle discontinuous paths cannot reproduce, e.g. 3-dimensional rotational dynamics [16]. Table 1 contains a synthetic comparison of the three mentioned packages.

3.2 Implementation details

`Pychastic` contains implementations of three numerical integration schemes based on the Taylor-Itō expansion. They are the schemes of strong order 1/2, 1, and 3/2 which in the package are referred to as Euler, Milstein, and Wagner-Platen schemes, respectively.

The basis for these integration schemes is the Taylor-Itō expansion [18] which generalises the deterministic Taylor expansion to SDEs. We write it here for a one-dimensional problem driven by a Wiener process W

$$dX = a(X)dt + b(X)dW\,. \tag{5}$$

Following the notation of Kloeden & Platen [18], the Taylor-Itō expansion up to strong order 3/2 for a scalar function X has the form

$$\begin{aligned} X(T) =& X(0) + a\Delta + b\Delta W \\ &+ \frac{1}{2}bb'\left((\Delta W)^2 - \Delta\right) \\ &+ a'b\Delta Z + \frac{1}{2}\left(aa' + \frac{1}{2}b^2a''\right)\Delta^2 \\ &+ (ab' + \frac{1}{2}b^2b'')(\Delta W\Delta - \Delta Z) + \frac{1}{2}b\left(bb'' + (b')^2\right)\left(\frac{1}{3}(\Delta W)^2 - \Delta\right)\Delta W\,, \end{aligned} \tag{6}$$

where

$$\Delta = \int_0^T dt = T\,, \tag{7}$$

$$\Delta W = \int_0^T dW = W(T) - W(0)\,, \tag{8}$$

$$\Delta Z = \int_0^T \int_0^s dW(s)\,ds\,. \tag{9}$$

The stochastic Euler scheme is based on the first line of this expansion, with three terms only. The Milstein scheme includes the next term, namely the second line of Eq. (6). Finally, the Wagner-Platen scheme includes all the terms mentioned above. We note from this expansion

that even the simplest schemes require correct computation of coefficients multiplying principal Wiener integrals (Δ, ΔW, ΔZ, and others in the multidimensional case) and sampling from correct distributions corresponding to these integrals. Importantly, an analogous expansion for vector quantities is considerably more complex, and other Wiener integrals arise, as detailed in [18]. For brevity, we do not write this expansion here explicitly, but we have implemented the vector Taylor-Itō expansion in `Pychastic` to enable simulations of both scalar and vector processes.

Although expressions for these coefficient functions can be written explicitly in principle, in `Pychastic` we take advantage of functional programming tools, which results in a greatly simplified implementation. First, by introducing the $\mathcal{L}^0$ and $\mathcal{L}^j$ operators (again using the Kloeden & Platen notation), we can express all coefficient functions by repeated application of the $\mathcal{L}$ operators to the a and b functions. Using the `jax.grad` functionality, this is implemented directly as

```
def tensordot1(a, b):
    return jax.numpy.tensordot(a, b, axes=1)

def tensordot2(a, b):
    return jax.numpy.tensordot(a, b, axes=2)

# Taylor-Ito expansion operators
def L_t_operator(f,problem):
    @wraps(f)
    def wrapped(x):
        b_val = problem.b(x)
        val = tensordot1(jax.jacobian(f)(x), problem.a(x)) + 0.5 * tensordot2(
            jax.hessian(f)(x), tensordot1(b_val, b_val.T)
        )
        return val[:,jnp.newaxis,...] #indexing convention [spatial, time, ...
    = noiseterms/time]

    return wrapped

def L_w_operator(f,problem):
    @wraps(f)
    def wrapped(x):
        val =  tensordot1(jax.jacobian(f)(x), problem.b(x))[:,jnp.newaxis,...]
        return jnp.swapaxes(val,1,-1)[:,...,0] # indexing convention [spatial,
    noiseterms, ... = noiseterms/time]

    return wrapped
```

Second, different integration methods – Euler, Milstein, and Wagner-Platen – were implemented following the book by Kloeden & Platen [18]. These methods make use of samples of the principal Wiener integrals listed above. Unfortunately, the text contains typographic errors, which were found by examining the mean, variance and covariance properties of fundamental multiple Itō and Stratonovich integrals using a testing suite that is part of the `Pychastic` package. The errors we found were corrected appropriately and are listed in the Appendix A.

Finally, using the `jax.lax.scan` functionality of the package `jax`, we can generate many trajectories in parallel by programmatically vectorising the code representing a and b provided by the user.

4 Examples of usage

In the following, we show a few examples of usage that take advantage of various functionalities of `Pychastic`.

4.1 First passage problems, polar random walk

In nearly all physical applications of Brownian Dynamics, the simulated properties of the systems are observed via their moments rather than individual realisations of the underlying stochastic process. In fact, in many situations with random forcing, the paths of the process are only a model of reality and cannot be directly observed, as opposed to their statistical effect. For a smooth quantity g observed through its expected value $\mathbb{E}$ in thermodynamic equilibrium, it is the weak convergence rate that controls the error of $\mathbb{E}(g(X_t))$ at fixed t. With this in mind, it is tempting to dismiss methods with a high strong order of convergence as impractical.

A natural candidate for a counterexample are first passage times, where the answer requires a more subtle reasoning. Since the first passage problem is *path-dependent*, it would seem that strong convergence is important. However, the theorem regarding the first passage times established by Whitt [17] shows that the weak convergence of the process approximations $X_n \to X$ implies a weak convergence of the first passage times $T(X_n) \to T(X)$. In conclusion, for all popular physical applications, the weak order of convergence is the important one. In consequence, the Milstein scheme is never a good choice, since it is equivalent to the Euler scheme in terms of weak order of convergence, but is more computationally intensive and harder to implement. We illustrate the surprising result of [17] by numerical simulation in a familiar setting.

A simple analytically solvable case of the first passage problem is a two-dimensional diffusion process in the (X, Y) plane with constant drift velocity $\boldsymbol{\nu}$ and constant diffusion coefficient σ^2 in a system with an absorbing barrier at $Y = Y_\mathrm{b} = 2$. The problem admits an exact solution, in which the probability density function for the hitting time $t_\mathrm{hit} = \mathrm{argmin}_t(Y_t > Y_\mathrm{b})$ is given by

$$p_{t_\mathrm{hit}}(\tau) = \frac{Y_\mathrm{b}}{\sqrt{2\pi\sigma^2\tau^3}} \exp\left(-\frac{(Y_\mathrm{b} - \nu_y\tau)^2}{2\sigma^2\tau}\right), \tag{10}$$

and so $\mathbb{E}[t_\mathrm{hit}] = 1/2$.

Numerically, we calculate the sample of the first passage times and locations for $\boldsymbol{\nu} = (\nu_x, \nu_y) = (0, 4)$ with stopping condition $Y = Y_\mathrm{b} = 2$ and initial condition $(X, Y) = (2, 0)$. The Itō equations for the two-dimensional diffusion driven by Wiener processes W and W' with drift velocities $[\nu_x, \nu_y]$ take the form of

$$dX = \quad \nu_x dt \quad + \sigma dW_t\,, \tag{11}$$
$$dY = \quad \nu_y dt \quad + \sigma dW'_t\,, \tag{12}$$

where σ is a parameter that describes the strength of the noise compared to the drift. We first transform this problem into a different set of coordinates, for which we choose polar coordinates. While we retain the analytical solution, we acquire non-linear terms in the governing equations, which will help us study the convergence issues which arise when noise and drift terms are generally dependent on the position. Eq. (12) transformed into polar coordinates (r, ϕ) is expressed as

$$dr = \quad \left(\tfrac{1}{2r} + \nu_y \sin\phi + \nu_x \cos\phi\right) dt \quad + \sigma\cos\phi\, dW_t + \sigma \sin\phi\, dW'_t\,, \tag{13}$$
$$d\phi = \quad \tfrac{1}{r}\left(\nu_y \cos\phi + \nu_x \sin\phi\right) dt \quad - \sigma\frac{\sin\phi}{r} dW_t + \sigma\frac{\cos\phi}{r} dW'_t\,. \tag{14}$$

The transformed equations are nonlinear, coupled, with nondiagonal, noncommutative noise terms. Vector equations in the package `Pychastic` are defined by providing both the drift vector `a` and the noise matrix `b` as callable to the `SDEProblem` constructor. Denoting the configuration by $\boldsymbol{q} = (r, \phi)$, we write

```
import jaq.numpy as jnp

y_drift = 4.0
sigma = 1
y_barrier = 2

def drift(q):
    return jnp.array(
        [1 / (2 * q[0]) + y_drift * jnp.sin(q[1]), y_drift * jnp.cos(q[1]) / q[0]]
    )

def noise(q):
    return sigma * jnp.array(
        [[jnp.cos(q[1]), jnp.sin(q[1])], [-jnp.sin(q[1]) / q[0], jnp.cos(q[1]) / q[0]]]
    )

problem = pychastic.sde_problem.SDEProblem(
    a=drift,
    b=noise,
    x0=jnp.array([2.0, 0.0]),
    tmax=2.0,
)
```

The `jax.numpy` package is a functional cousin of `numpy` [19], which is the fundamental package for scientific computing with Python [20]. The functional focus of `jax.numpy` enables automatic differentiation, and thus facilitates the use of the high-order method without the need for numerical differentiation, which can be imprecise and computationally costly. We obtain the trajectories using the `solve` method of `SDESolver`.

```
solver = pychastic.sde_solver.SDESolver(dt=2**(-5), scheme='milstein')
solution = solver.solve_many(problem,n_samples,seed=0)
```

`Pychastic` supports simulating many trajectories simultaneously with the `solve_many` method. Our strategy allows for concurrency in computation of random variables and equation coefficients thanks to the `jax` package we use as back-end (which in turn relies on vectorised mathematical operations provided by the `XLA` supporting architecture of most modern processors). The `jax` package is a fusion of three capabilities: `numpy`-like API for array-based computing, functional transformations (such as vectorisation, parallelisation and automatic differentiation), and modular back-end, allowing developers to test their work with just CPU and deploy the same code later on a GPU (or TPU) capable hardware. Furthermore, we avoid notoriously slow `Python` loops by using `jax.lax.scan` routines, taking advantage of just-in-time compilation and asynchronous dispatch, avoiding the problem of global interpreter lock, which cripples imperatively coded Python programmes. A more detailed guide on the advantages (and common pifalls) of `jax` package can be found in its documentation on ReadTheDocs.

Importantly, even though the Milstein and Wagner-Platen schemes require the values of spatial derivatives of noise and drift terms, we did not have to provide them explicitly. They were calculated using an automatic differentiation procedure from the coefficient functions provided in the `SDEProblem` constructor.

We compare the results obtained with different solvers and time steps in Fig. 1. Sample

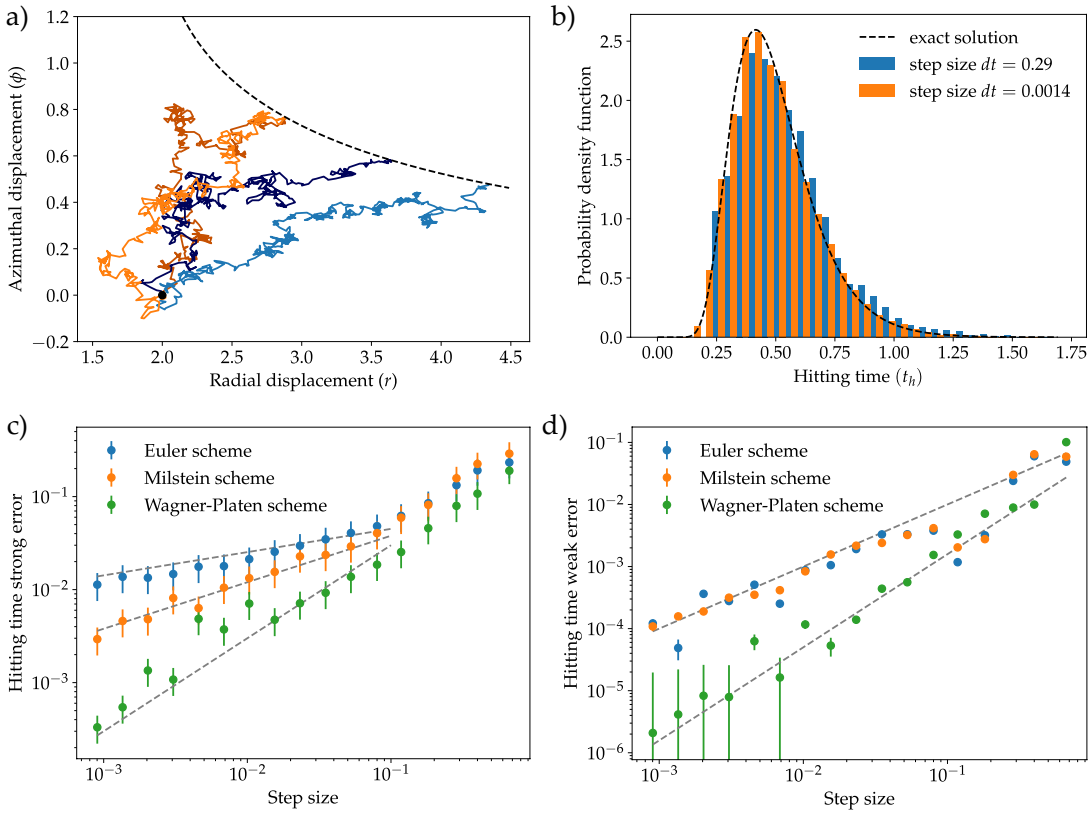


Figure 1: **a)** Four sample trajectories generated from Eq. (14) using the Euler algorithm together with the barrier $r\sin\phi = y = 2$. **b)** Distribution of hitting times for two different step sizes in the Euler algorithm together with the analytic solution of the problem, given by Eq. (10). Large step sizes typically overestimate the hitting time. **c)** Strong (path-wise) error of the hitting time for different step sizes and solvers together with power-law eye guides: $dt^{1/4}$, $dt^{1/2}$, dt. **d)** Error of the expected value of the first passage time together with eye-guides: $dt^{1/2}$, dt. Contrary to the strong error case, here Euler and Milstein algorithms coincide by Whitt's theorem.

trajectories starting from the point $(r,\phi) = (2,0)$ are shown in Fig. 1a. The barrier at $y = 2$ becomes a curve in polar parameterisation. In Fig. 1 we present the resulting distribution of hitting times measured from an ensemble of $N = 1000$ trajectories with two different time-step sizes and compared to the exact solutions. Choosing a too large step size can lead to an overestimation of the typical hitting time. Having implemented three stochastic integration algorithms, we compare the strong error of estimation of the hitting time in Fig. 1c. It is a measure of the numerical error per trajectory, defined as $\langle|\tilde{\tau}-\tau|\rangle$, where $\tilde{\tau}$ is the numerical estimate of hitting time and τ is actual hitting time that can be computed using the exact solution and a particular realisation of the Wiener process. The average is taken over an ensemble of realisations of the Wiener process. For different schemes, the strong error scales differently with the step size – from the linear dependence on $dt^{1/4}$ for the Euler scheme, through $dt^{1/2}$ for the Milstein algorithm, to dt for Wagner-Platen. We note that these exponents are different (smaller) than for strong convergence at a fixed time [18]. Similarly, in Fig. 1d we present the weak error of estimation of the average value of the hitting time. The weak error is defined as $|\langle\tilde{\tau}\rangle - \langle\tau\rangle|$, and reflects the difference between the ensemble average value of hitting time estimate and the true ensemble average hitting time (computed directly from the theoreti-

cal distribution). This quantity highlights the lack of difference between Euler and Milstein schemes, in agreement with Whitt's theorem [17] stating the dependence of the expected first passage time on the weak convergence rate of the scheme only. The power laws for convergence – $dt^{1/4}$ for Euler and Milstein algorithms and dt for the Wagner-Platen scheme – again have smaller exponents than for those for weak convergence at a fixed time.

4.2 Rotational Brownian motion, `step_post_processing` function

To highlight the ease-of-use features of the `Pychastic` package, we present the problem of rotational Brownian diffusion simulations. General BD simulations that involve rigid bodies require the application of finite rotations to diffusing objects. The resulting algorithms are widely applied to study problems in physics and biology. While resolving translational motion is straightforward, e.g. with the standard Ermak-McCammon (Euler) algorithm [21], the rotational part is more involved to simulate, because the domain of rotational motion is $SO(3)$ and this has to be taken into account when solving the equations of motion. One difficulty lies in the commonly used rotational coordinate systems, such as Euler angles [22], which contain strong singularities around the polar orientations. When curvilinear coordinates are used to describe the motion, the metric tensor gives rise to nontrivial additional terms in the equations of motion, the so-called metric or drift terms, which are frequently overlooked in rotational BD algorithms. For a summary, see Ref. [23].

To overcome these limitations, we reimplement the problem following the rotation-vector-based formulation of Evensen *et al.* [24]. We describe the angular position of the particle by the angle of rotation Φ around a unit vector $\boldsymbol{\delta}$ collinear with the axis of rotation. Generalised coordinates are encoded in the vector $\boldsymbol{q} = \Phi\boldsymbol{\delta}$. The rotational mobility matrix of the particle in the body-fixed frame is $\boldsymbol{\mu}_{\text{body}}$. We transform it into the laboratory frame by

$$\boldsymbol{\mu} = \boldsymbol{\Omega}^T \cdot \boldsymbol{\mu}_{\text{body}} \cdot \boldsymbol{\Omega}\,, \tag{15}$$

where $\boldsymbol{\Omega}$ is the relevant rotation matrix. We also define a velocity transformation matrix $\boldsymbol{\Xi}$ to further introduce the transformed mobility $\widehat{\boldsymbol{\mu}}$ by

$$\widehat{\boldsymbol{\mu}} = \boldsymbol{\Xi} \cdot \boldsymbol{\mu} \cdot \boldsymbol{\Xi}^T\,. \tag{16}$$

Now we can write the Itō SDE corresponding to the evolution of the generalised coordinates as

$$d\boldsymbol{q} = \widehat{\boldsymbol{\mu}} \cdot \left(\frac{\partial}{\partial \boldsymbol{q}} \log V\right) dt + k_B T \left(\frac{\partial}{\partial \boldsymbol{q}} \cdot \widehat{\boldsymbol{\mu}}\right) dt + \sqrt{2k_B T}\,\widehat{\boldsymbol{\mu}}^{1/2} \cdot\, d\boldsymbol{W}\,. \tag{17}$$

Here, V is the density of the volume element ($SO(3)$ Haar measure) with respect to the Lebesgue measure on $\mathbb{R}^3$. In the literature $\partial \log V / \partial \boldsymbol{q}$ is often called the metric force $\boldsymbol{F}^{(m)}$. Explicit expressions for $\boldsymbol{\Xi}$, $\boldsymbol{\Omega}$ and $\log V$ are given in Appendix B.

Since the evolution equation depends on the divergence of a product of two orientation-dependent matrices, writing Eq. (17) explicitly is quite cumbersome. Earlier works of [24] and [26] avoid this difficulty by approximating the gradients by sampling the transformed mobility matrix at nearby locations in the phase space. We can address this problematic term directly thanks to the automatic differentiation capabilities of the `jax` package.

```
def metric_force(q):
    # Metric force, providing the Boltzmann distribution in equilibrium.
    phi = jnp.sqrt(jnp.sum(q ** 2))
    scale = jax.lax.cond( # Taylor expansion for the polar orientations.
        phi < 0.01,
        lambda t: -t / 6.0,
        lambda t: jnp.sin(t) / (1.0 - jnp.cos(t)) - 2.0 / t,
```

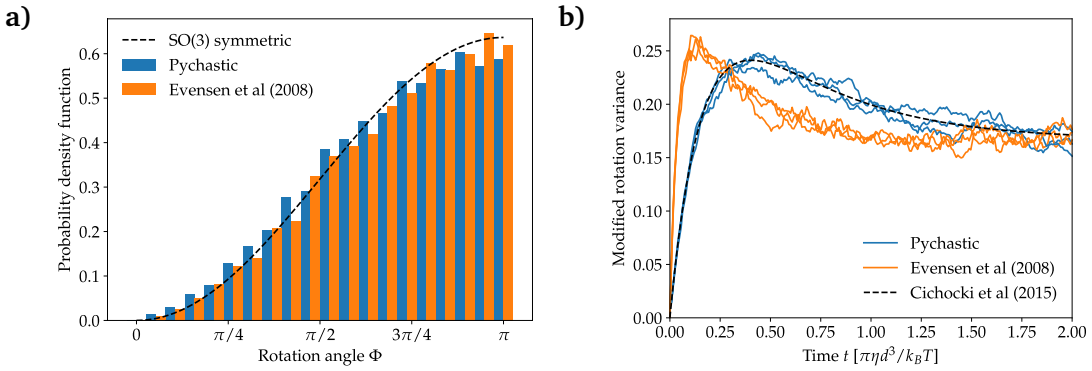


Figure 2: **a)** Distribution of rotation angles Φ after a long time in $N = 10^4$ simulations compared with the $SO(3)$ symmetric distribution $P^{eq}(\Phi)$ given by Eq. (18). **b)** Comparison of the time-dependent correlation structure arising from equation (17) for a spherical particle. Dashed lines show exact predictions from Ref. [25] and Eq. (19). Solid, colored lines show average values of 10^3 simulations (blue line corresponds to equation (17) and orange trajectories correspond to Ref. [24] containing a typo in their Eq. 12). Note that even though equilibrium distributions coincide, time dependent correlation structure is different.

```
        phi,
    )
    return jax.lax.cond(
        phi > 0.0, lambda: (q / phi) * scale, lambda: jnp.array([0.0, 0.0, 0.0
    ])
    )

def t_mobility(q):
    # Mobility matrix transformed to rotation vector coordinates.
    Return transformation_matrix(q) @ mobility @ (transformation_matrix(q).T)

def drift(q):
    # jax.jacobian has the differentiation index last (like mu_ij d_k) so
    divergence is contraction of the first and last axis.
    return (
    t_mobility(q) @ metric_force(q)
    + jnp.einsum("iji->j", jax.jacobian(t_mobility)(q))
    )
```

First, including the divergence of the mobility matrix in the new coordinate system is as simple as adding a correct contraction of `jax.jacobian` of the mobility tensor. To obtain this, we use the convenient `jnp.einsum` tool. This function, similarly to its analogue `np.einsum`, takes in a tensor, represented in the memory by an array of arrays of arrays etc. with the number of levels corresponding to the tensor rank, and performs contraction described symbolically by the index notation. In our case `"iji->j"` means 'contract the first index with the last index of a rank-3 tensor to produce a vector.'

Second, the domain of the equation is $SO(3)$, which is not covered by $\mathbb{R}^3$ (since $\pi_1(SO(3)) = \mathbb{Z}_2$ [16]). As a result, some trajectories, continuous in $SO(3)$, will necessarily be discontinuous in the $\mathbb{R}^3$ parameterisation. We adopt the method proposed by Ref. [24], where after each step of integration we project on the principal value of the rotation angle. In `pychastic`, we achieve this using the `step_post_processing` capability.

```python
def canonicalize_coordinates(q):
    phi = jnp.sqrt(jnp.sum(q ** 2))
    max_phi = jnp.pi
    canonical_phi = jnp.fmod(phi + max_phi, 2.0 * max_phi) - max_phi
    return jax.lax.cond(
        phi > max_phi,
        lambda canonical_phi, phi, q: (canonical_phi / phi) * q,
        lambda canonical_phi, phi, q: q,
        canonical_phi,
        phi,
        q,
    )
solver = pychastic.sde_solver.SDESolver(dt=0.01)
trajectories = solver.solve_many(
    problem,
    step_post_processing=canonicalize_coordinates,
    n_trajectories=1000
)
```

The simulations were validated by checking the equilibrium properties and the time-dependent correlation structure. The equilibrium distribution $P^{eq}(\Phi)$ for the rotation angle Φ is given by

$$P^{eq}(\Phi) = \frac{1-\cos\Phi}{\pi}\,. \tag{18}$$

To characterise the time evolution, we use the components of the modified rotation vector components $\Delta u_k(t) = -\frac{1}{2}\epsilon_{ijk}\Omega_{ij}(t)$, for which exact theoretical predictions were derived by Cichocki *et al.* [25] for an arbitrarily shaped molecule. For a spherical particle, the correlation functions are given by

$$\langle \Delta u_k(t)\Delta u_l(t)\rangle_0 = \left[\frac{1}{6} - \frac{5}{12}e^{-6D_r t} + \frac{1}{4}e^{-2D_r t}\right]\delta^K_{kl}\,, \tag{19}$$

where δ^K is the Kronecker delta. The diffusion coefficient for a sphere is $D_r = k_B T/\pi\eta d^3$, where d is the diameter of the particle.

In Fig. 2, we present a comparison between the results obtained with an algorithm based on Eq. (17) (blue lines), and a similar method based on equations (5-12) from Ref. [24] (orange). Fig. 2a shows the equilibrium distribution of the rotation angle Φ, together with the analytical prediction of Eq. (18). We note that although the equilibrium distribution predicted by both algorithms shows exact agreement with the theoretical predictions, the transformation matrix $\boldsymbol{\Xi}$ in Eq. (12) of Ref. [24] contains an error. We corrected this typographical error and present the proper formulation of the equations of motion matrices in the Appendix B. The discrepancy is visible in the time-dependent correlation function in Fig. 2b. Although at long times the system tends to the same equilibrium solution for both approaches, we see the agreement of our algorithm with the theoretical predictions of Eq. (19) (dashed line) and the deviation of the orange lines. This discrepancy highlights the need for better test cases that reliably test all properties of the simulated equation, and not solely the equilibrium distribution.

As mentioned above, the algorithm based on Ref. [24] is singularity-free, contrary to approaches based on Euler angles, which contain strong singularities around the polar orientations. However, it should be noted that the implementation of the code may still face some limitations, such as the computational inability to calculate $\sin\Phi/\Phi$ for $\Phi \to 0$. For these numerically restricted cases, we implemented the Taylor expansion of $\boldsymbol{\Xi}$, $\boldsymbol{\Omega}$ and $\log V$, based on Ref. [26]. However, we also report some typos in this publication. Therefore, in Appendix B we provide correct formulations of the Taylor-expanded terms.

4.3 Bead models with hydrodynamic interactions, `pygrpy` package integration

Hydrodynamically interacting beads (with and without springs) have been used successfully in modelling the properties of elastic macromolecules [21,27–31]. Some questions about elastic macromolecules can be answered by computing the equilibrium ensemble of conformations (with methods such as Markov Chain Monte Carlo [32]). Simulations involving hydrodynamic interactions can, on the other hand, provide access to the dynamics and answer questions, e.g., about time scales of conformational change [33], mechanisms of protein association [34], pore translocation [35–37], near-wall hindered diffusion [38], and many other dynamical processes.

The starting point is often the Hamiltonian, which describes intramolecular interactions between the constitutive subunits of the molecule, modelled by a collection of beads. The potential energy landscape can then be used to compute the interaction forces. Thanks to the `jax autograd` capabilities, the forces arising from many-body mechanical interactions can be automatically calculated from the potential energy of the system.

The computation of hydrodynamic interactions, encoded in the mobility tensors of the respective macromolecules is a more involved task. Numerous methods are available, varying in scope, degree of precision, and complexity. For a review of popular methods applied to macromolecules, see Ref. [39]. For completeness, we supplemented `pychastic` with the package `pygrpy` for the calculation of grand mobility tensors in the Rotne-Prager-Yamakawa (RPY) approximation [40, 41], generalised to beads of different sizes [42]. The procedure is a Python port of the `GRPY` package [39]. The Python package `pygrpy` simplifies the implementation of similar bead-spring simulations, both stochastic and deterministic. Furthermore, `pygrpy` is compatible with the `jax` functional paradigm and allows automatic differentiation and vectorisation.

The RPY approximation is by far the most popular method of accounting for hydrodynamic interaction in numerical models of soft matter systems [43]. The mobility tensors calculated in this way preserve positive definiteness and are divergence-free, significantly simplifying the BD algorithm [21]. In essence, RPY is a far-field approximation that includes all terms that decay slower than the inverse third power of the interparticle distances, but is less accurate at smaller distances. When the particles come close together, it is necessary to include higher-order terms of the multipole expansion [44] and lubrication corrections [45]. However, as shown by Żuk *et al.* [42], the RPY approximation can be generalised to overlapping particles, and thus it can also be used to model complex-shaped particles as conglomerates of rigidly glued overlapping spheres, in particular to calculate hydrodynamic properties of biological macromolecules, as in the GRPY method [39].

To highlight the ease of use and interoperability of `pychastic` and `pygrpy`, we implement a benchmark problem proposed by Cichocki et al. [46]. It concerns the diffusion coefficient of an elastic "macromolecule" composed of 4 beads of radii $r_i \in \{3, 1, 1, 1\}$ joined into a string, as shown in Fig. 3a. Here, the length scale is the radius of a small bead a. The neighbouring beads interact directly via harmonic potentials, with the equilibrium distance $d_i = 4$ and the spring constant $k = 5.5\ k_B T/a$, and indirectly through hydrodynamic interactions.

We determine the diffusion coefficient D by tracking the mean square displacement $\langle(\boldsymbol{x}(t) - \boldsymbol{x}(0))^2\rangle$ (MSD) of a point $\boldsymbol{x}$ of the molecule. If the observation time is long enough, MSD $\approx 6Dt$. However, for shorter times, the coefficient of proportionality in the MSD(t) curve is different and depends on the choice of the reference point on the molecule. Although biophysical experiments typically measure the long-time diffusion coefficient, numerical simulations have easier access to short-time diffusivity. To minimise the difference, one should choose a particular point, called the centre of diffusion, which can be constructed as a weighted average of the positions of beads with weights determined from the hydrodynamic mobilities of an ensemble of equilibrium configurations of the elastic molecule [46].

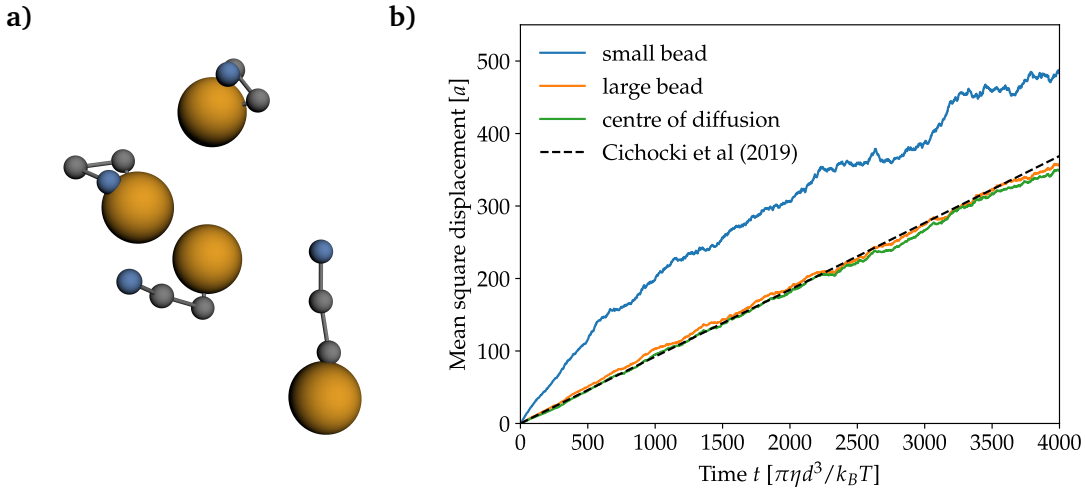


Figure 3: **a)** Representative configurations of four co-diffusing beads connected with harmonic springs at time (bottom to top) $t/\tau_d = 2, 2000, 4000, 6000$. **b)** Mean square displacement of three different tracked locations: the small bead at the end of the chain (blue), the large bead at the other end (orange) and the weighted average of all four beads with weights corresponding to the effective centre of diffusion, defined according to Ref. [46]. The dashed line corresponds to numerical results reported in [46] based on an extremely long, single-trajectory simulation.

We use this system to test our algorithm. In Fig. 3b, we show the temporal evolution of the MSD when different reference points on the elastic molecule are chosen: the small terminal bead, the large bead, and the centre of diffusion. The time is scaled by $\tau_d = \pi\eta d^3/k_BT$. We computed hydrodynamic interactions using the package `pygrpy`, which allows computations of mobility tensors for macromolecules composed of unequally sized and possibly overlapping spherical beads in the Rotne-Prager-Yamakawa approximation. The dashed line is the numerical result of [46] where the diffusion coefficient was estimated using a single very long BD trajectory. Indeed, we see that tracking the centre of diffusion provides a good estimate of the long-time behaviour of the MSD. When one chooses to track one of the smaller beads or one of the beads further away from the middle of the molecule, precision decreases primarily because rotational diffusion plays a bigger role in the motion of such tracers. The strength of our algorithm is that we use many trajectories instead of a single long one, thus avoiding the need for a more complex method of calculating the diffusion coefficient, involving recursive subdivisions of the simulation interval described by Frenkel and Smit [47].

```
radii = jnp.array([3.0,1.0,1.0,1.0]) # sizes of spheres used
n_beads = len(radii)
equilibrium_dist = 4.0
spring_constant = 5.5

def pot_ene(x): # potential energy
    locations = jnp.reshape(x,(n_beads,3))
    distance_ab = jnp.sqrt(jnp.sum((locations[0] - locations[1])**2))
    distance_bc = jnp.sqrt(jnp.sum((locations[1] - locations[2])**2))
    distance_cd = jnp.sqrt(jnp.sum((locations[2] - locations[3])**2))
    ene = 0.5*spring_constant*jnp.sum(
    ([distance_ab,distance_bc,distance_ca] - equilibrium_dist)**2
    )
    return ene
```

```
15
16 def drift(x):
17      locations = jnp.reshape(x,(n_beads,3))
18      mu = pygrpy.jax_grpy_tensors.muTT(locations,radii)
19      force = -jax.grad(pot_ene)(x)
20      return jnp.matmul(mu,force)
21
22 def noise(x):
23      locations = jnp.reshape(x,(n_beads,3))
24      mu = pygrpy.jax_grpy_tensors.muTT(locations,radii)
25      return jnp.sqrt(2)*jnp.linalg.cholesky(mu)
26
27 problem = pychastic.sde_problem.SDEProblem(
28       drift,
29       noise,
30       x0 = jnp.reshape(jnp.array([
31       [-2.,0.,0.],
32       [2.,0.,0.],
33       [6.,0.,0.],
34       [10.,0.,0.]
35       ]),(3*n_beads,)),
36       tmax = 10000.0)
```

5 Conclusion

We developed a novel Python package `pychastic` dedicated to efficient numerical solutions of SDEs. The package implements the classical truncated Taylor-Itō integrators up to strong order $\mathcal{O}(dt^{3/2})$ providing a precise treatment of both weak (e.g. equilibrium distributions, diffusion coefficients) and strong problems. We included a set of simple test cases that provide exact reference points for testing future stochastic integration algorithms. The analysis of three-dimensional rotational Brownian motion benchmarks is particularly important because it encompasses many of the difficult aspects of SDE approaches to Brownian dynamics: divergence terms in the evolution equation, handling discontinuous trajectories (unavoidable in the case of $SO(3)$), and spurious agreements when testing only the equilibrium distribution. We hope that `pychastic` will ease future studies of Brownian dynamics problems, especially problems that involve hydrodynamic interactions. The project is open source, and we hope to encourage collaboration and its further development.

Acknowledgements

The authors thank Piotr Szymczak for his insightful feedback.

Funding information The work of ML, RW, and MB was supported by the National Science Centre of Poland (FundRef DOI: http://dx.doi.org/10.13039/501100004281) grant Sonata to ML no. 2018/31/D/ST3/02408.

A Appendix: Typos in integration schemes in Ref. [18]

In the book: *Numerical solution of stochastic differential equations* by Peter E. Kloeden and Eckhard Platen in equation 5.8.11 we have found the following typos:

1. In the definition of $D^p_{j1,j2,j3}$: the fourth summand should read $\zeta_{j1,r}\eta_{j3,l+r}$ instead of $j1$.
2. In the definition of b_j: there is a $1/\pi$ factor missing before sum.
3. In the definition of $C^p_{j1,j2}$: second term should be $+1/r\eta_{j1,r}\eta_{j2,l}$ instead of $-l/r$.

B Appendix: Rotational Brownian motion

In equation (17) we recall the equations of rotational Brownian motion, formulated in Ref. [24]. Angular orientation is described by the vector $\boldsymbol{a} = (a_1, a_2, a_3) = \Phi\boldsymbol{\delta}$, where Φ is the angle of rotation around a unit vector $\boldsymbol{\delta}$, which corresponds to the axis of rotation. Because the expressions in Ref. [24] contain typos, here we present the proper expressions for coordinate transformation and rotation matrices $\boldsymbol{\Xi}$ and $\boldsymbol{\Omega}$:

$$\boldsymbol{\Xi} = \left(\frac{1}{\Phi^2} - \frac{\sin\Phi}{2\Phi(1-\cos\Phi)}\right)\begin{pmatrix} a_1a_1 & a_1a_2 & a_1a_3 \\ a_2a_1 & a_2a_2 & a_2a_3 \\ a_3a_1 & a_3a_2 & a_3a_3 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} \frac{\Phi\sin\Phi}{1-\cos\Phi} & -a_3 & a_2 \\ a_3 & \frac{\Phi\sin\Phi}{1-\cos\Phi} & -a_1 \\ -a_2 & a_1 & \frac{\Phi\sin\Phi}{1-\cos\Phi} \end{pmatrix}, \tag{20}$$

$$\boldsymbol{\Omega} = \frac{1}{\Phi^2}\begin{pmatrix} \Phi^2\cos\Phi & -a_3\Phi\sin\Phi & a_2\Phi\sin\Phi \\ a_3\Phi\sin\Phi & \Phi^2\cos\Phi & -a_1\Phi\sin\Phi \\ -a_2\Phi\sin\Phi & a_1\Phi\sin\Phi & \Phi^2\cos\Phi \end{pmatrix} + \frac{1-\cos\Phi}{\Phi^2}\begin{pmatrix} a_1a_1 & a_1a_2 & a_1a_3 \\ a_2a_1 & a_2a_2 & a_2a_3 \\ a_3a_1 & a_3a_2 & a_3a_3 \end{pmatrix}. \tag{21}$$

The matrix $\boldsymbol{\Xi}$ transforms the velocities from the Cartesian coordinate system to the one described by $\boldsymbol{a}$. The matrix $\boldsymbol{\Omega}$ is a simple rotation matrix. These two matrices combined allow transformation of the body-fixed mobility matrix $\boldsymbol{\mu}_{\text{body}}$ to the lab-fixed, $\boldsymbol{a}$-described mobility matrix, given as $\widehat{\boldsymbol{\mu}} = \boldsymbol{\Xi}\cdot\boldsymbol{\Omega}^T\cdot\boldsymbol{\mu}_{\text{body}}\cdot\boldsymbol{\Omega}\cdot\boldsymbol{\Xi}^T$.

The term $\left(\frac{\partial}{\partial\boldsymbol{q}}\log V\right)$ from Eq. (17) can be associated with the metric force $\boldsymbol{F}^{(m)}$, which guarantees the Boltzmann distribution in equilibrium and is given as

$$\left(\frac{\partial}{\partial\boldsymbol{q}}\log V\right) = \boldsymbol{F}^{(m)} = -k_BT\left(\frac{\sin(\Phi)}{1-\cos(\Phi)} - \frac{2}{\Phi}\right)\boldsymbol{\delta}\,. \tag{22}$$

Due to the numerical limitations, it is sometimes inevitable to perform the Taylor expansion of the above quantities, as presented in Ref. [26]. However, we also found typos in this case. Therefore, we provide the correctly expanded matrices $\boldsymbol{\Xi}$ and $\boldsymbol{\Omega}$ for $\Phi \to 0$,

$$\boldsymbol{\Xi} = \frac{1}{12}\begin{pmatrix} a_1a_1 & a_1a_2 & a_1a_3 \\ a_2a_1 & a_2a_2 & a_2a_3 \\ a_3a_1 & a_3a_2 & a_3a_3 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 2 & -a_3 & a_2 \\ a_3 & 2 & -a_1 \\ -a_2 & a_1 & 2 \end{pmatrix} + O(\Phi^2), \tag{23}$$

$$\boldsymbol{\Omega} = \begin{pmatrix} 1 & -a_3 & a_2 \\ a_3 & 1 & -a_1 \\ -a_2 & a_1 & 1 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} a_1a_1 & a_1a_2 & a_1a_3 \\ a_2a_1 & a_2a_2 & a_2a_3 \\ a_3a_1 & a_3a_2 & a_3a_3 \end{pmatrix} + O(\Phi^2). \tag{24}$$

The Taylor-expanded metric force around $\Phi \to 0$ becomes

$$\boldsymbol{F}^{(m)} = k_BT\frac{\Phi}{6}\boldsymbol{\delta} + O(\Phi^3). \tag{25}$$

References

[1] P.-G. De Gennes, *Soft matter*, Science **256**, 495 (1992), doi:10.1126/science.256.5056.495.

[2] S. Kim and S. J. Karrila, *Microhydrodynamics: Principles and selected applications*, Butterworth-Heinemann, Oxford, UK, ISBN 9780750691734 (1991), doi:10.1016/C2013-0-04644-0.

[3] M. Praprotnik, L. D. Site and K. Kremer, *Multiscale simulation of soft matter: From scale bridging to adaptive resolution*, Annu. Rev. Phys. Chem. **59**, 545 (2008), doi:10.1146/annurev.physchem.59.032607.093707 .

[4] U. D. Schiller, T. Krüger and O. Henrich, *Mesoscopic modelling and simulation of soft matter*, Soft Matter **14**, 9 (2018), doi:10.1039/C7SM01711A.

[5] J. K. G. Dhont, *An introduction to dynamics of colloids*, Elsevier, Amsterdam, Netherlands, ISBN 9780080535074 (1996).

[6] M. Karplus and G. A. Petsko, *Molecular dynamics simulations in biology*, Nature **347**, 631 (1990), doi:10.1038/347631a0.

[7] N. van Kampen, *Stochastic processes in physics and chemistry*, Elsevier, Amsterdam, Netherlands, ISBN 9780080475363 (2007), doi:10.1016/B978-0-444-52965-7.X5000-4.

[8] R. M. Mazo, *Brownian motion*, Oxford University Press, Oxford, UK, ISBN 9780199556441 (2008), doi:10.1093/acprof:oso/9780199556441.001.0001.

[9] A. Einstein, *Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen*, Ann. Phys. **322**, 549 (1905), doi:10.1002/andp.19053220806.

[10] M. von Smoluchowski, *Zur kinetischen theorie der Brownschen molekularbewegung und der suspensionen*, Ann. Phys. **326**, 756 (1906), doi:10.1002/andp.19063261405.

[11] A. D. Fokker, *Die mittlere energie rotierender elektrischer dipole im strahlungsfeld*, Ann. Phys. **348**, 810 (1914), doi:10.1002/andp.19143480507.

[12] M. Planck, *Sitzungsberichte der Königlich Preussischen akademie der wissenschaften zu Berlin*, Deutsche Akademie der Wissenschaften zu Berlin, Berlin, Germany (1917).

[13] A. Kolmogoroff, *Über die analytischen methoden in der wahrscheinlichkeitsrechnung*, Math. Ann. **104**, 415 (1931), doi:10.1007/BF01457949.

[14] K. Itô, *On stochastic differential equations*, American Mathematical Society, Providence, Rhode Island, USA, ISBN 9780821812044 (1951).

[15] P. Langevin, *Sur la théorie du mouvement brownien*, C. R. Acad. Sci. **146**, 530 (1908).

[16] B. C. Hall, *Lie groups, Lie algebras, and representations: An elementary introduction*, Springer, Cham, Switzerland, ISBN 9783319134673 (2015), doi:10.1007/978-3-319-13467-3.

[17] W. Whitt, *Weak convergence of first passage time processes*, J. Appl. Prob. **8**, 417 (1971), doi:10.2307/3211913.

[18] P. Kloeden and E. Platen, *Numerical solution of stochastic differential equations*, Springer, Berlin, Heidelberg, ISBN 9783662126165 (1992), doi:10.1007/978-3-662-12616-5.

[19] C. R. Harris et al., *NumPy*, https://numpy.org/.

[20] C. R. Harris et al., *Array programming with NumPy*, Nature **585**, 357 (2020), doi:10.1038/s41586-020-2649-2.

[21] D. L. Ermak and J. A. McCammon, *Brownian dynamics with hydrodynamic interactions*, J. Chem. Phys. **69**, 1352 (1978), doi:10.1063/1.436761.

[22] H. Goldstein, *Classical mechanics*, Pearson Education, London, UK, ISBN 9788177582833 (2002).

[23] I. M. Ilie, W. J. Briels and W. K. den Otter, *An elementary singularity-free rotational Brownian dynamics algorithm for anisotropic particles*, J. Chem. Phys. **142**, 114103 (2015), doi:10.1063/1.4914322.

[24] T. R. Evensen, S. N. Naess and A. Elgsaeter, *Brownian dynamics simulations of rotational diffusion using the Cartesian components of the rotation vector as generalized coordinates*, Macromol. Theory Simul. **17**, 403 (2008), doi:10.1002/mats.200800031.

[25] B. Cichocki, M. L. Ekiel-Jeżewska and E. Wajnryb, *Brownian motion of a particle with arbitrary shape*, J. Chem. Phys. **142**, 214902 (2015), doi:10.1063/1.4921729.

[26] I. M. Ilie, W. K. den Otter and W. J. Briels, *Rotational Brownian dynamics simulations of clathrin cage formation*, J. Chem. Phys. **141**, 065101 (2014), doi:10.1063/1.4891306.

[27] R. M. Jendrejack, J. J. de Pablo and M. D. Graham, *Stochastic simulations of DNA in flow: Dynamics and the effects of hydrodynamic interactions*, J. Chem. Phys. **116**, 7752 (2002), doi:10.1063/1.1466831.

[28] R. G. Larson, T. T. Perkins, D. E. Smith and S. Chu, *Hydrodynamics of a DNA molecule in a flow field*, Phys. Rev. E **55**, 1794 (1997), doi:10.1103/PhysRevE.55.1794.

[29] A. Cressman, Y. Togashi, A. S. Mikhailov and R. Kapral, *Mesoscale modeling of molecular machines: Cyclic dynamics and hydrodynamical fluctuations*, Phys. Rev. E **77**, 050901 (2008), doi:10.1103/PhysRevE.77.050901.

[30] C.-C. Hsieh, S. Jain and R. G. Larson, *Brownian dynamics simulations with stiff finitely extensible nonlinear elastic-Fraenkel springs as approximations to rods in bead-rod models*, J. Chem. Phys. **124**, 044911 (2006), doi:10.1063/1.2161210.

[31] J. Huang and T. Schlick, *Macroscopic modeling and simulations of supercoiled DNA with bound proteins*, J. Chem. Phys. **117**, 8573 (2002), doi:10.1063/1.1511506.

[32] B. A. Berg, *Markov chain Monte Carlo simulations and their statistical analysis*, World Scientific, Singapore, ISBN 9789812389350 (2004), doi:10.1142/5602.

[33] S. He and H. A. Scheraga, *Macromolecular conformational dynamics in torsional angle space*, J. Chem. Phys. **108**, 271 (1998), doi:10.1063/1.475378.

[34] S. H. Northrup, J. A. Luton, J. O. Boles and J. C. L. Reynolds, *Brownian dynamics simulation of protein association*, J. Comput.-Aided Mol. Des. **1**, 291 (1988), doi:10.1007/BF01677278.

[35] T. Yamanouchi and A. Satoh, *Improvement of trapping performance of magnetic particles by magnetic multi-poles via Brownian dynamics simulations of magnetic rod-like particles in a Hagen-Poiseuille flow*, Mol. Phys. **120**, e2067503 (2022), doi:10.1080/00268976.2022.2067503.

[36] R. Waszkiewicz and M. Lisicki, *Hydrodynamic effects in the capture of rod-like molecules by a nanopore*, J. Phys.: Condens. Matter **33**, 104005 (2020), doi:10.1088/1361-648x/abd11b.

[37] L. Qiao and G. W. Slater, *Capture and translocation of a rod-like molecule by a nanopore: Orientation, charge distribution and hydrodynamics*, Phys. Chem. Chem. Phys. **24**, 6444 (2022), doi:10.1039/D2CP00313A.

[38] S. A. Rogers, M. Lisicki, B. Cichocki, J. K. G. Dhont and P. R. Lang, *Rotational diffusion of spherical colloids close to a wall*, Phys. Rev. Lett. **109**, 098305 (2012), doi:10.1103/PhysRevLett.109.098305.

[39] P. J. Zuk, B. Cichocki and P. Szymczak, *GRPY: An accurate bead method for calculation of hydrodynamic properties of rigid biomacromolecules*, Biophys. J. **115**, 782 (2018), doi:10.1016/j.bpj.2018.07.015.

[40] J. Rotne and S. Prager, *Variational treatment of hydrodynamic interaction in polymers*, J. Chem. Phys. **50**, 4831 (1969), doi:10.1063/1.1670977.

[41] H. Yamakawa, *Transport properties of polymer chains in dilute solution: Hydrodynamic interaction*, J. Chem. Phys. **53**, 436 (1970), doi:10.1063/1.1673799.

[42] P. J. Zuk, E. Wajnryb, K. A. Mizerski and P. Szymczak, *Rotne-Prager-Yamakawa approximation for different-sized particles in application to macromolecular bead models*, J. Fluid Mech. **741**, R5 (2014), doi:10.1017/jfm.2013.668.

[43] G. Nägele, *Brownian dynamics simulations*, Forschungszentrum Jülich, Jülich, Germany, ISBN 9783893364305 (2006).

[44] B. Cichocki, R. B. Jones, R. Kutteh and E. Wajnryb, *Friction and mobility for colloidal spheres in Stokes flow near a boundary: The multipole method and applications*, J. Chem. Phys. **112**, 2548 (2000), doi:10.1063/1.480894.

[45] M. L. Ekiel-Jeżewska and E. Wajnryb, *Precise multipole method for calculating hydrodynamic interactions between spherical particles in the stokes flow*, Transworld Research Network, Trivandrum, India, ISBN 9788178954004 (2009).

[46] B. Cichocki, M. Rubin, A. Niedzwiecka and P. Szymczak, *Diffusion coefficients of elastic macromolecules*, J. Fluid Mech. **878**, R3 (2019), doi:10.1017/jfm.2019.652.

[47] D. Frenkel and B. Smit, *Understanding molecular simulation: From algorithms to applications*, Academic Press, Cambridge, UK, ISBN 9780122673511 (2002), doi:10.1016/B978-0-12-267351-1.X5000-7.

3.5 Paper V: *Minimum dissipation approximation: A fast algorithm for the prediction of diffusive properties of intrinsically disordered proteins*

PAPER V

"*Minimum dissipation approximation: A fast algorithm for the prediction of diffusive properties of intrinsically disordered proteins*"

Radost Waszkiewicz, Agnieszka Michaś, Michał K. Białobrzewski, Barbara Klepka, Maja Cieplak-Rotowska, Zuzanna Staszałek, Bogdan Cichocki, Maciej Lisicki, Piotr Szymczak, and Anna Niedźwiecka

COMMENTARY

Intrinsically Disordered Proteins (IDPs) constitute a large class of bio-relevant elastic macromolecules. They typically consist of approximately rigid folded domains connected with flexible linkers. This type of molecules comprises up to a third of proteins in eukaryotes, yet it remains understudied because it cannot be easily analysed using crystallographic techniques due to their ever changing conformations. Determining the relative importance of different physical processes that determine the conformations of IDPs and the role of flexible linkers in interactions between their domains remains an open problem.

To the best knowledge of the authors, this is the first publication comparing a first-principles approach (in contrast to many phenomenological attempts such as Ref. [38, 75]) to prediction of hydrodynamic size of IDPs with experimental data. Earlier approaches to the prediction of the protein hydrodynamic radius (for example Ref. [76]) generally focus on techniques which approximate the protein as a rigid body. From a modelling perspective, intrinsically disordered proteins differ significantly from, e.g. short strands DNA we focused on before. While IDPs consist of globular fragments connected with linkers, where the former are essentially rigid, and the latter are almost ideally flexible. Such proteins are ideal subjects for the application of the Minimum Dissipation Approximation [25], specifically designed for the fast computation of diffusion coefficients for molecules that exhibit substantial conformational variability and possess parts of different sizes.

This study emerged from the collaborative efforts of two distinct groups. The first group, specialising in theoretical modelling, included the PhD candidate, Bogdan Cichocki, Maciej Lisicki and Piotr Szymczak from the Faculty of Physics, University of Warsaw. The second group that focused on synthesis and FCS measurements of IDPs, comprised Agnieszka Michaś, Michał K. Białobrzewski, Barbara P. Klepka, Maja K. Cieplak-Rotowska, Zuzanna Staszałek, and Anna Niedźwiecka from the Institute of Physics of the Polish Academy of Sciences.

We were able to obtain a favourable comparison with experimental data on the level of hydrodynamic radius R_h, with a substantially better agreement than all prior phenomenological models selected for comparison, and noticeably better agreement than a simple power law fit. The resulting code was packaged into an easy to use python library and command line tool called `glm_mda_diffusion` ready to install via `pip` and available directly on GitHub.

In this study the PhD candidate: co-conceptualized the modelling approach, investigated the available conformer generation schemes, implemented the Python port `pygrpy` of the generalised Rotne-Prager mobility tensors (originally implemented in Fortran [17]), implemented the globule-linker conformer generation scheme, performed numerical calculations and statistical analysis to assess deviations between theory and experiment. Additionally, they wrote the first draft and edited all subsequent versions of the manuscript.

Hydrodynamic Radii of Intrinsically Disordered Proteins: Fast Prediction by Minimum Dissipation Approximation and Experimental Validation

Radost Waszkiewicz,[†,§] Agnieszka Michaś,[‡,§] Michał K. Białobrzewski,[‡] Barbara P. Klepka,[‡] Maja K. Cieplak-Rotowska,[‡,¶] Zuzanna Staszałek,[‡] Bogdan Cichocki,[†] Maciej Lisicki,[†] Piotr Szymczak,[*,†] and Anna Niedzwiecka[*,‡]

†*Institute of Theoretical Physics, Faculty of Physics, University of Warsaw, L. Pasteura 5, 02-093 Warsaw, Poland*

‡*Institute of Physics, Polish Academy of Sciences, Aleja Lotnikow 32/46, PL-02668 Warsaw, Poland*

¶*present address: IMol Polish Academy of Sciences, Flisa 6, PL-02247 Warsaw, Poland*

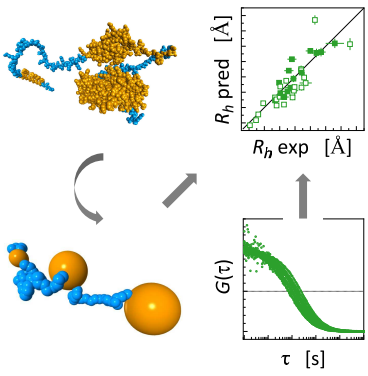
§*These authors have contributed equally to this work and share first authorship*

E-mail: piotrek@fuw.edu.pl; annan@ifpan.edu.pl

Abstract

The diffusion coefficients of globular and fully unfolded proteins can be predicted with high accuracy solely from their mass or chain length. However, this approach fails for intrinsically disordered proteins (IDPs) containing structural domains. We propose a rapid predictive methodology for estimating the diffusion coefficients of IDPs. The methodology uses accelerated conformational sampling based on self-avoiding random walks and includes hydrodynamic interactions between coarse-grained protein subunits, modeled using the generalized Rotne-Prager-Yamakawa approximation. To estimate the hydrodynamic radius, we rely on the minimum dissipation approximation recently introduced by Cichocki *et al*. Using a large set of experimentally measured hydrodynamic radii of IDPs over a wide range of chain lengths and domain contributions, we demonstrate that our predictions are more accurate than the Kirkwood approximation and phenomenological approaches. Our technique may prove valuable in predicting the hydrodynamic properties of both fully unstructured and multidomain disordered proteins.

TOC Graphic



Keywords

intrinsically disordered protein, hydrodynamic radius, diffusion coefficient, coarse-grained model, minimum dissipation approximation

Intrinsically disordered proteins (IDPs) constitute an extensive class of biological macromolecules, and their role in the homeostasis of a living cell has been increasingly recognized in recent decades.[1,2] The frequency of long intrinsically disordered regions (IDRs) in proteins differs significantly between the kingdoms of life, ranging from 2 % in archaea to 33 % in eukaryotes.[3] The IDP molecules display different degrees of structural disorder. Their chains can encompass either several folded globular domains, or supersecondary structures connected by flexible linkers, sparse secondary structural elements, or can be completely natively unstructured. Disordered proteins exhibit a notable characteristic – the absence of a stable, well-defined relative spatial arrangement of their fragments. Instead, their equilibrium properties can be described through a broad set of rapidly inter-converting conformers, posing a challenge for analysis, particularly in the context of long chains.[4]

The average geometric properties of IDPs, including their shape and size, are determined by the equilibrium ensemble of conformational states. This equilibrium state is intricately influenced by environmental conditions,[5] such as temperature,[6] ionic strength[7,8], osmolality,[9] crowding[10], post-translational modifications[11], and the presence of specific molecular binding partners[12]. The formation of transient or more stable non-covalent complexes introduces another non-trivial dependence of the IDP equilibrium geometry on environmental factors.

Because the shape and availability of the binding sites necessary for the interaction of IDP with ligands, other proteins, and nucleic acids are greatly influenced by the environment, IDPs often act as higher-order regulators in key cellular processes such as gene expression[11,13] signaling[2,14], or extracellular biomineralization[15]. It is the different conformations of these flexible proteins, which enable IDPs to perform their multiple functions[1]. In particular, it is worth emphasizing the important roles of IDPs in health and disease, *e.g.*, the role of the p53 protein as a tumour suppressor[16], mutations of which are often responsible for human cancers, the function of 4E-BPs in the inhibition of eukaryotic translation initiation[11,17–19], the significance of GW182 protein in the recruitment of the multi-protein machinery necessary for microRNA-mediated gene

silencing[20–22], or the importance of Tau, FUS, and α-synuclein proteins in neurodegenerative diseases[23,24]. Since the elastic properties of these biomolecules are responsible for the proper functioning of IDPs in the cellular context, *i.e.* for the association of complexes and the formation of biomolecular condensates via liquid-liquid phase separation such as, e.g., RNA-processing membraneless organelles[25,26] much attention is paid to the hydrodynamic properties of IDPs. Experimental techniques, such as analytical ultracentrifugation (AUC), size exclusion chromatography (SEC), pulsed-field gradient nuclear magnetic resonance (PFG–NMR), dynamic light scattering (DLS), and fluorescence correlation spectroscopy (FCS), offer insights into hydrodynamic parameters (as reviewed in[27]). However, due to the distinct limitations of each experimental approach, ongoing research aims to devise phenomenological methods for calculating the hydrodynamic radius (R_h). These methods may involve deriving R_h from the radius of gyration determined by small-angle X-ray scattering (SAXS)[28,29] or exploiting the conformational backbone propensity of IDPs.[30,31]

Simultaneously, significant effort is being invested in developing numerical models that extract the characteristics of IDPs from conformational ensembles obtained using molecular dynamics (MD) simulations or energy minimization algorithms[32–41]. However, the molecular flexibility of IDPs introduces substantial complexities when determining their hydrodynamic properties. Two main issues here are the large number of degrees of freedom and the long timescales of relaxation of the internal coordinates of the molecules. These factors prohibit direct calculation of the experimentally relevant long-time diffusion coefficient from either molecular or Brownian dynamics trajectories. One popular approximation that circumvents this difficulty is to assume that the macromolecule is 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 an equilibrium ensemble[42–45]. Nevertheless, the validity and accuracy of this approximation remain uncertain. Additionally, the generation of conformational ensembles can be a bottleneck for long chains (beyond approximately 300 amino acid residues) because it requires time-consuming MD simulations and/or the construction of new databases of

short peptide conformations.

There is, therefore, a strong need to develop a numerically efficient solution that would enable reliable calculation of the long-time diffusion coefficient of any long-chain IDP, such as one with 1000 amino acid residues, solely based on its sequence information.

In this study, we introduce a new theoretical approach to both generating conformational ensembles of IDPs and calculating their hydrodynamic properties. This method enables a swift estimation of the diffusion coefficient for long IDPs in a matter of minutes, with superior accuracy compared to existing methods. This assertion is substantiated through rigorous testing of the model on a diverse set of experimental results obtained for 43 proteins. The dataset includes both literature data and R_h values measured for a set of new IDP constructs using FCS under mild conditions (see Supporting Information).

We present our results in terms of the hydrodynamic radius of a molecule, R_h. This radius represents the size of a solid sphere that possesses the same translational diffusion coefficient, D, as the given molecule under identical buffer conditions. Therefore, $R_h = k_B T/6\pi\eta D$, where T is the temperature and η is the viscosity.

An important observation by Fixman[46,47] is that the diffusion coefficient of a flexible macromolecule is time-dependent, with well-defined short- and long-time limits. The disparity between the two is attributed to the effects associated with relaxation of the internal coordinates of the molecule, as well as rotation of the macromolecule as a whole.[46,48,49] The positivity of the dissipation rate in the system implies that the long-time diffusion coefficient (D_l) is always smaller than the short-time diffusivity (D_s).[47] The focus of theoretical approaches should be the determination of the former quantity, as it is the one measured in experiments utilizing techniques like FCS, AUC, or DLS. Unfortunately, the calculation of D_l is significantly more challenging than that of D_s because it involves the computation of time-dependent quantities, such as the memory function, which describes the relaxation effects. An additional point to keep in mind is that the value of the short-time diffusion coefficient depends on the choice of the point that one tracks.[49–52] In contrast, the long-time diffusivity is independent of the choice of reference point.[53]

The methods for predicting the diffusion coefficient can be broadly split into three

categories: atomistic, phenomenological, and coarse-grained. For small proteins, high-resolution, atomistic MD methods can be used,[54] but they require either simulating the surrounding water molecules explicitly, which is very computationally intensive, or an implicit solvent scheme. In the case of implicit solvent methods, addressing hydrodynamic interactions between distant parts of the molecule[55–58] and thermalization[59] pose significant challenges. Additionally, even for the smallest proteins, it is prohibitively difficult to obtain statistically meaningful data over the 10-100 millisecond scale, which would enable the direct computation of the long-time diffusion coefficient.

The other extreme are phenomenological models that predict R_h from the number of residues N and possibly other parameters, such as total charge or amino acid composition. Theoretical considerations of Rouse, who modelled a protein as a Gaussian chain[60] gave foundation to the power law relationship $R_h \sim N^{1/2}$. The classical Rouse model employs random displacements between the monomers. If we assume complete independence of displacements between each consecutive pair of monomers, the central limit theorem dictates that as N approaches infinity, the squared end-to-end distance should conform to a scaled $\chi^2(3)$ distribution. Consequently, the dimensions of such an idealized chain are expected to scale with $\sqrt{N}$. Later work of Zimm included the effect of excluded volume[61], which resulted in the scaling $R_h \sim N^{\gamma}$ with $\gamma = 0.588$.

Phenomenological size–length relationships that include other variables involve a number of fitting parameters. As a result, their range of applicability outside of the fitting dataset is difficult to assess. An alternative phenomenological approach proposed by Pesce *et al.*[29] employs the radius of gyration obtained from SAXS experiments to estimate R_h. This is substantiated by the observation that within the Kirkwood-Riseman approximation[62] R_h and R_g share the same scaling relationship with N as long as the pair-displacement distribution converges under appropriate scaling to a Gaussian for large N.

Finally, coarse-grained models, like our method, employ larger units (typically one or two per amino acid residue) as building blocks for the structure prediction scheme, along with approximate interaction potentials between subunits, to simulate the equilibrium

ensemble of configurations for a given molecule. These configurations are then combined with an approximation of the hydrodynamic properties to compute the diffusion coefficient. Essentially, the computation of the latter for elastic macromolecules addresses two interconnected challenges: predicting the conformations of molecules based on available biochemical data and then using these conformations to predict hydrodynamic properties.

The different exponents in the power-law relationships of Rouse[60] and Zimm[61] demonstrate that even the most basic method for approximating configurations must take into account excluded volume interactions.

A software that can accommodate excluded volume interactions for a disordered chain is Flexible Meccano (FM).[34] In addition to volume exclusion, it considers the distribution of Ramachandran angles determined from crystallographic protein structures when sampling conformations. However, FM treats the entire chain as unstructured, so it cannot be used to model proteins that possess both globular and unfolded segments, which are in fact much more common than fully unstructured chains. Unfortunately, FM has a closed license that precludes necessary modifications to accommodate folded regions of proteins.

The complex angle distributions used by FM are crucial when computing NMR parameters that are sensitive to short-range details of the pair-distribution function, such as residual dipolar couplings, paramagnetic relaxation enhancement, or J-coupling. However, upon closer examination, the pair-distance distribution generated by FM and a simpler model presented in this paper, globule-linker model (GLM; described below), become virtually identical for amino acids separated by more than 15 residues along the chain.

The highly localized differences between structures at small sequential distances have a minimal influence on the estimations of R_h. It is important to recall that for amino acid residues separated by a distance r, the dipolar coupling decays as r^{-3}, while the decay rate of hydrodynamic interactions (HI) is only r^{-1}. Therefore, HI are long-range and less sensitive to near-neighbor distributions, with contributions to the diffusion coefficient of near-neighbors and far-neighbors being $O(N)$ and $O(N^{2-\gamma}) = O(N^{1.4})$, respectively.

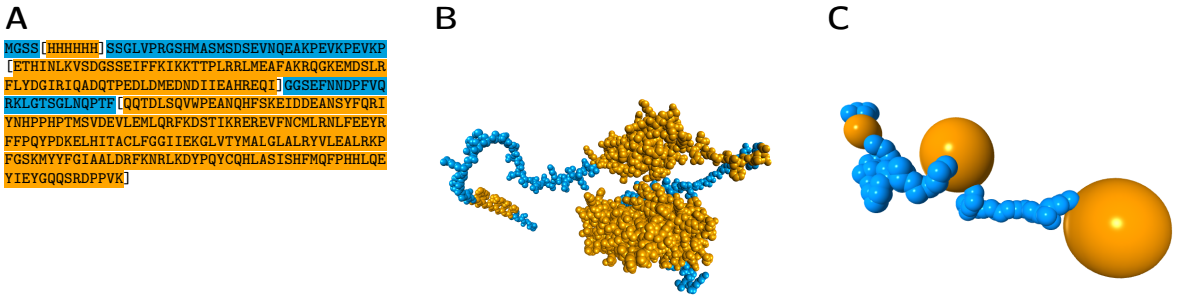


Figure 1. Construction of the coarse-grained globule-linker model (GLM) for an illustratory IDP, H_6-SUMO-CNOT1(800-999), containing three ordered domains of different sizes (no. 28 in Table S1). **A)** Sequence with highlighted ordered (orange) and disordered (blue) segments, and domain boundaries marked by square brackets. **B)** Representative full atom conformation generated by AlphaFold2 (for visualisation purposes only,[63,64] beads with van der Waals radii, hydrogen atoms omitted for clarity); ordered clusters (orange) form dense blobs connected with linkers (blue). **C)** Visualisation of a representative configuration generated using the GLM method where beads are displayed with their hydrodynamic radii.

Guided by these considerations, we have implemented the simplest extension of Zimm's chain — the globule-linker model (GLM), designed to comprehensively represent IDPs containing globular domains connected by unstructured fragments. In the model (Figure 1 A-C), we represented the protein as an assembly of spheres of different sizes. Disordered segments of length N were modeled as chains of N identical spheres, each with a diameter equal to the C_α-C_α distance, while structured domains were represented by single spheres, each of them described by their mass (m). The size of a sphere representing a structured domain was computed as $R_h = (3m/4\pi\rho_{\text{globular}})^{1/3} + a_{\text{hydration}}$ with $\rho_{\text{globular}} = 0.52$ Da / Å^3,[65] and the single layer hydration shell taken to be $a_{\text{hydration}} = 3$ Å thick. Within this approach, information about domain boundaries in the protein sequence is sufficient to construct an appropriate bead approximation of the IDP. The identification of protein sequence fragments to be treated as ordered regions and mimicked by larger beads in the GLM model was done using Disopred3.[66] The fragment was assumed to be ordered if the disorder probability P was less than 50 % for at least three subsequent amino acid residues, including loops linking such fragments but not exceeding 14 residues.[67]

Using a recursive approach, it is possible to generate GLM conformations with a time complexity of $O(N^{1+\gamma})$, which provides a satisfactory ensemble for the largest of the proteins considered here in under a minute using only a personal computer (a single thread

at 1.8GHz). The speed of the recursive approach should be contrasted with an iterated one where steps are simply added one by one, and intersecting chains are discarded. This easier-to-implement method is characterized by a time complexity of $O(\exp(N))$, which becomes prohibitively slow for chains with $N > 20$.

We have transformed the sampled conformations into a hydrodynamic model by increasing bead sizes in the disordered fragments to $R_{\text{disordered}} = 4.2$ Å, corresponding to the median value for all aminoacids.[68] In the resulting hydrodynamic model of linkers, neighbouring beads show substantial overlaps, requiring a careful treatment of the mobility matrices (see[69] for details). Note that the value of $R_{\text{disordered}}$ has only a minor impact on the final results, since the hydrodynamic radius of long slender filaments depends logarithmically on their thickness.[70–73]

To compute R_h from the estimated ensembles we have implemented two algorithms: the Kirkwood formula, and minimum dissipation approximation (MDA) method of Cichocki *et al*.[53] Within the first approach,[74] the hydrodynamic radius of a macromolecule is approximated by

$$\frac{1}{R_h^{\mathrm{K}}} = \frac{1}{N^2} \sum_{i=1}^{N} \left(\frac{1}{a_i} + \sum_{j=1, j\neq i}^{N} \left\langle \frac{1}{r_{ij}} \right\rangle \right), \tag{1}$$

where N is the total number of beads in the IDP model, a_i is the hydrodynamic radius of bead i, $r_{ij} = |\mathbf{r}_j - \mathbf{r}_i|$ is the distance between beads i and j, and $\langle \cdot \rangle$ denotes the average over the equilibrium ensemble. One can show that this corresponds to the ensemble-averaged short-time diffusion coefficient of the geometric center of the macromolecule, $\mathbf{r}_c = N^{-1} \sum_{i=1}^{N} \mathbf{r}_i$. Note that the geometric center fluctuates as the shape of the molecule evolves and does not correspond to any fixed position within it. A simplified form of the Kirkwood formula is often used[39,75,76]

$$\frac{1}{R_h^{\mathrm{K}}} \approx \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} \left\langle \frac{1}{r_{ij}} \right\rangle, \tag{2}$$

where the single-bead terms $1/a_i$ are dropped, as their contribution becomes negligible in the large N limit. This is the form that we will also use in the present work.

A better estimate of R_h, corresponding to the long-time diffusion coefficient, requires

a more in-depth description of the hydrodynamic interactions between the beads. To this end, one introduces the mobility matrix $\boldsymbol{\mu}$,[48] which links the velocities of the beads with the forces acting on them, according to

$$\mathbf{U}_i = \sum_j \boldsymbol{\mu}_{ij} \mathbf{F}_j, \tag{3}$$

where $\mathbf{U}_i$ is the velocity of bead i whereas $\mathbf{F}_j$ is the force with which bead j acts on the fluid. Based on the mobility matrix, one defines a matrix $\mathbf{A}$ indexed by the bead labels $(i,\, j)$, $A_{ij} = 2\pi\eta \mathrm{Tr} \left\langle \boldsymbol{\mu}_{ij} \right\rangle$ and its inverse $\mathbf{B} = \mathbf{A}^{-1}$. One can then construct the MDA[53] for R_h as

$$R_h^{\mathrm{MDA}} = \sum_{i,j} B_{ij}. \tag{4}$$

Note that the above formula is general, and can be used for different models of hydrodynamic interactions - both simple models (*e.g.* Oseen or Rotne–Prager far-field approximation[77]) or in more sophisticated approaches, like the multipole expansion method[78,79]. In this work, we use the generalized Rotne-Prager approximation for the calculation of the mobility matrix, as described in[80–82]. This approximation is now also available as a Python package, `pygrpy`.[83] For non-overlapping beads, the elements of the matrix $\mathbf{A}$ have then a particularly simple form: $A_{ij} = \langle 1/r_{ij} \rangle$ for $i \neq j$, and $A_{ii} = 1/a_i$. The formulas for overlapping beads can be found in the Supporting Information.

The MDA corresponds to the calculation of the short-time diffusion coefficient of the diffusion center of a molecule,[52] which is a point inside the molecule where D_s is minimal. The position of the diffusion center is $\mathbf{r}_d = \sum_{i=1}^{N} x_i \mathbf{r}_i$, with the weights given by $x_i = \sum_j B_{ij} / \sum_{k,j} B_{kj}$. Since D_s is always larger than its long-time counterpart, D_l, MDA provides the best estimation for the long-time diffusion coefficient out of all methods that utilize D_s for this purpose. The MDA turns out to be more robust when dealing with large differences in the sizes of beads used to model constituent parts of the macromolecule, because in such cases the equal weights of the geometric center of the macromolecule differ significantly from the optimal weights of the diffusion center.

We combined each method of generating conformers with each method of computing

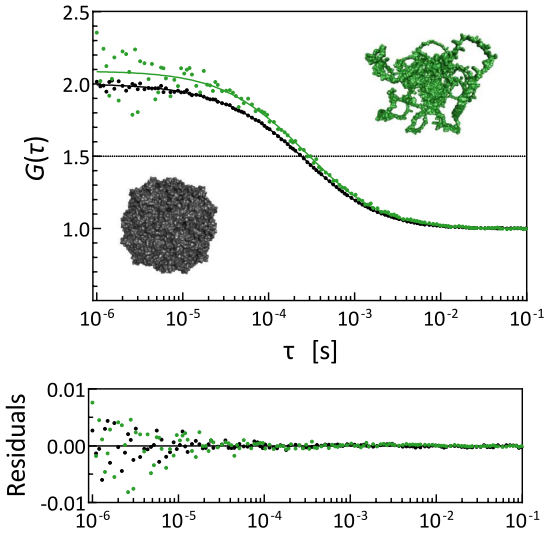


Figure 2. Examples of normalized FCS autocorrelation curves with raw fitting residuals for an intrinsically disordered H_6-SUMO-GW182SD-mCherry ($N = 809$, $R_h = 66 \pm 6$ Å) (green) in comparison with apoferritin ($N = 4200$, $R_h = 58 \pm 3$ Å) (black). Crystal structure of apoferritin (pdb id. code `2w0o`[84]) and putative conformation of H_6-SUMO-GW182SD-mCherry predicted by AlphaFold[63] are shown for illustration purposes, preserving relative sizes of solvent accessible surfaces of atoms.

R_h, which resulted in four different theoretical approaches, the predictions of which (Table S2) were then compared with experimental data. For this purpose, we have obtained 15 new IDP constructs covering a wide range of chain lenghts, folded domain contents and charge states, and determined their R_h using FCS (Figure 2 and S2, S3, S4, S5, S6; for further experimental details see Supporting Information).

The experimental benchmark set (Table S1) was thus composed of both the new FCS measurements, and R_h selected from the literature based on the following criteria: the proteins had sequences could be unambiguously identified in the literature or in the UniProtKB database, were measured at well defined, mild conditions (temperature of 20-26 °C, buffer of pH 7-8, ionic strength corresponding to 75-300 mM NaCl), and their hydrodynamic radii were determined directly from appropriate experiments without conversions from other experimental quantities, such as R_g.[29,85–102] This is, to our best knowledge, the largest benchmark set encompassing experimental R_h for 38 IDPs and 6 globular model proteins, measured at comparable conditions.

The results of tests performed for our four theoretical approaches against the bench-

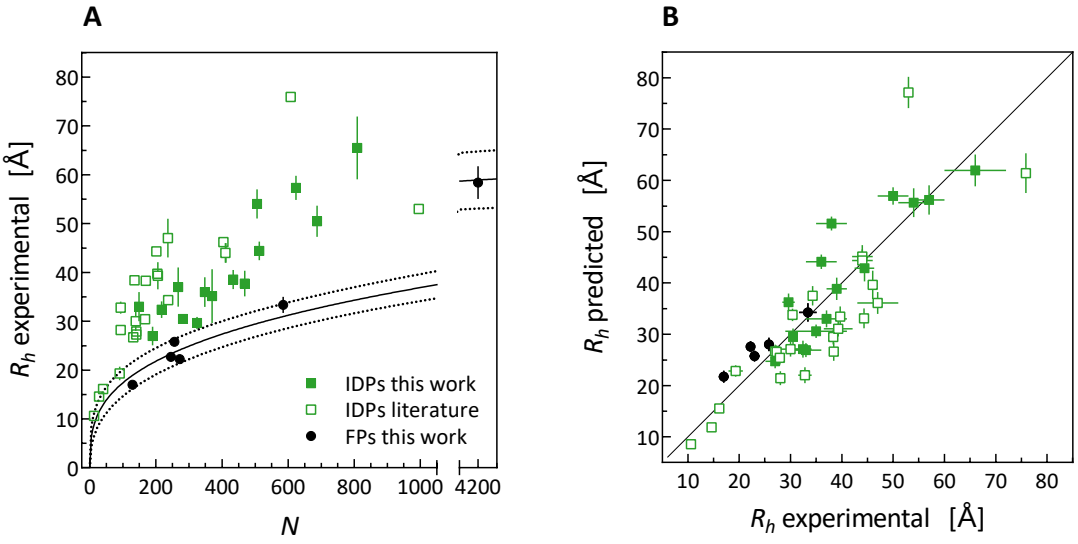


Figure 3. **A)** Experimental R_h values plotted against number of amino acid residues in the protein chain, N, and power-law curve fitted to R_h of folded proteins (FP) together with 95% confidence band. **B)** Direct comparison of the predicted vs. measured R_h values for all of the proteins modeled using the MDA+GLM approach.

Table 1. Comparison of error statistics of various models.

model	n_{fp}	RMSD[Å]	RMSRD[%]	R^2	R^2_{adj}	Q_3^{AE}[Å]	Q_3^{RE}[%]
MDA + GLM	0	7.09	18.15	0.71	0.71	6.80	22.51
MDA + GLM(ND)	0	9.48	28.02	0.48	0.48	11.88	29.31
KR + GLM	0	12.82	34.69	0.05	0.05	17.59	42.95
KR + GLM(ND)	0	9.25	27.31	0.50	0.50	11.11	29.44
random coil	1	9.60	27.71	0.47	0.45	10.28	33.69
power law	2	8.46	24.80	0.59	0.56	9.63	26.08
power law (lit.[103])	2	12.01	36.94	0.16	0.12	14.37	39.51
PPII-based (lit.[30])	3	17.25	49.09	−0.72	−0.86	20.62	59.54
sequence-based (lit.[103])	7	22.90	50.78	−2.05	−2.66	19.59	58.32

n_{fp} – number of fitting parameters, ND – no domain information, Q_3 – 3$^{\mathrm{rd}}$ quartile

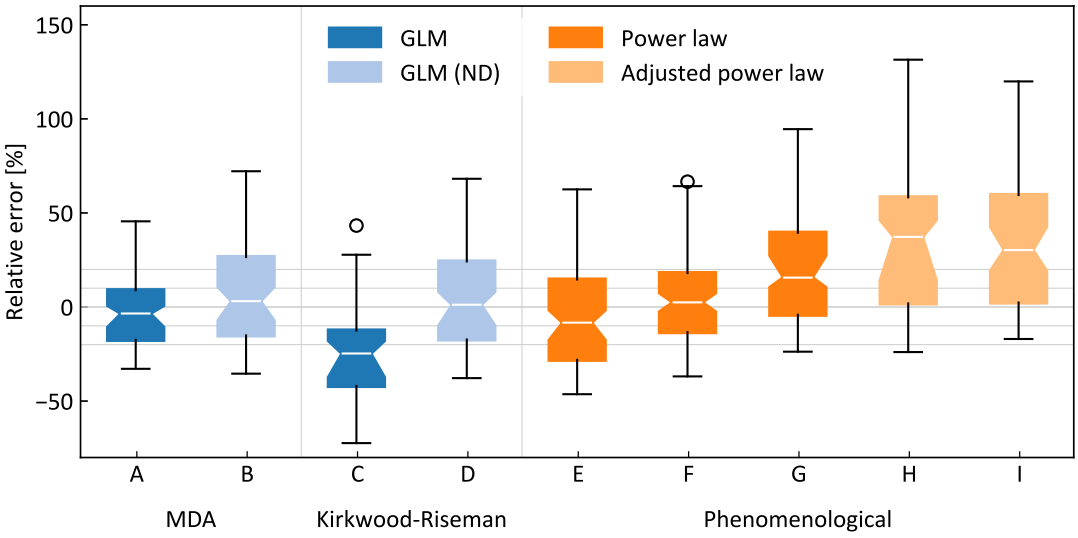


Figure 4. Comparison of different methods of estimation of R_h. Boxes show interquartile range with median confidence bands marked by notches. MDA with GLM ensemble generation (**A**) performs best on the IDP benchmark set with standard errors of 18.15 % and 7.09 Å (compared to 24.80 % and 8.46 Å for a simple power law). Methods based on the Kirkwood-Riseman R_h estimation (**C,D**) typically underestimate hydrodynamic size of the molecule. Power law fits with one free parameter (**E**) and two free parameters (**F**) evaluated using leave-one-out cross validation are compared with the formerly reported power law[103] (**G**) and a sequence-based model[103] (**H**) which takes into account total charge of the molecule, and a model based on polyproline II structure propensities[30] (**I**). Methods with no knowledge about the presence of domains in the IDP (ND; **B,C**) significantly overestimate the hydrodynamic size of the molecule. Domain data can be incorporated into our ensemble generation engine leading to more accurate estimates of R_h (**A**). Note that experimental uncertainty also contributes to the errors presented above and in Table 1.

mark set are gathered in Table 1, and Figure 4 shows a visual comparison of the deviations between theory and experiment. Additionally, we provide power law fits [104,105] and power law fits of Marsh *et al.* [103] for comparison of the prediction accuracy (Table S3).

We compare the accuracy of the previous and new model under six metrics (Table 1): the square root of mean square deviation (RMSD), square root of the mean square relative deviation (RMSRD), Pearson's coefficient (R^2), Pearson's coefficient adjusted for fitting parameters (R^2_{adj}), 3$^{\text{rd}}$ quartile of the absolute error Q_3^{AE}, and 3$^{\text{rd}}$ quartile of the relative error Q_{75}^{RE}. Whenever a fitting procedure is required, we use leave-one-out cross-validation to compute error metrics. We have chosen to test the relative deviations as well in order to reduce the undue weight given to the new, very long sequences in our dataset. Similarly, outlier-robust metrics of the 3$^{\text{rd}}$ quartile were included to reduce the impact of a single sequence misprediction on the final comparisons. In all evaluation metrics, the MDA+GLM approach performs the best. Surprisingly, it is the only model that performs better than the power law baseline in any of the evaluation metrics.

Although explicit intramolecular interactions of the amino acid residues are neglected in MDA+GLM approach, the main cause of discrepancies between the experimental and predicted R_h values (Figure S8) appeared to be the intrinsic properties of individual experimental methods, which suffer from typical errors or limitations and are usually not taken into account when reporting the final experimental results. PGF-NMR measurements are the most unambiguous and accurate, but their effective application is limited to smaller proteins (up to 200-300 amino acid residues long) at high concentrations. It is worth noting that the agreement of values of R_h predicted by MDA+GLM with the PGF-NMR results is excellent (Figure S8 C). FCS is the only method that addresses the self-diffusion of molecules at the limit of low concentrations. Raw FCS measurements can be refined to exclude possible oligomerization or aggregation during the experiment based on the count-rates. However, it is impossible to avoid proteolytic instability of proteins and, consequently, the appearance of impurities with a lower molar mass, which may potentially result in apparently lower values of R_h (Figure S8 B). On the other hand, SEC is the easiest approach to remove lower mass impurities, but it involves diffusion of

molecules at higher concentrations through a medium with pores of a specific shape under the influence of pressure. An additional common disadvantage is calibration based on R_h of standard proteins determined in various conditions and the lack of appropriate propagation of the calibration experimental uncertainty. Consequently, SEC measurements can be highly scattered (Figure S8 D). The largest outlier in our analysis concerns R_h determined using SEC for fesselin without providing experimental uncertainty (Id. 43, Tables S1, S2, Figure S7, S8 D). The DLS method is the most prone to overestimating experimental values (Figure S8 E), since the presence of even a small number of aggregates with a larger molar mass generates a huge contribution to the intensity of scattered light. Finally, AUC yields sedimentation coefficients, and their interpretation in terms of exact values of R_h requires some assumptions that are not obvious for IDPs, such as *e.g.* partial specific protein volume.[106] The second largest outlier in our set is the OMM-64 protein (Id. 39, Tables S1, S2, Figure S7, Figure S8 F) with the R_h value determined using AUC, which is very close to the power function curve for completely denatured proteins.[107]

In conclusion, we have presented a simple, first principles model for the prediction of R_h without any fitting parameters and achieved favourable comparison with a large benchmark set. Moreover, due to the relative simplicity of the model, all of the calculations for a given protein can be performed in about a minute on a typical laptop, which is contrasted with MD-based conformer generation methods that require supercomputers and take many days. Furthermore, to our surprise, the GLM-MDA approach demonstrates satisfactory convergence even with ensemble sizes as small as 40 conformers.

Our benchmark set, in which the previously known IDPs were complemented by a set of newly obtained proteins, constitutes a significant step forward in predicting hydrodynamic properties of IDPs. It includes a higher conformational variety, with a stronger emphasis on multidomain proteins, longer chains, and a much wider range of charge states compared to the reference sets used previously.[30,103] This diversity allows for more reliable testing of theoretical models.

Further developments of the MDA+GLM model are needed to take into account the

dependence of R_h on the environmental conditions[6–8] and the formation of complexes. However, our results clearly demonstrate that the relatively simple globule–linker model for conformational ensemble construction, in combination with the minimum dissipation approximation, can serve as the starting point for developing further phenomenological corrections. These improvements could incorporate factors such as amino acid sequence composition, residue charge, and counterion binding. When using MDA+GLM, all excluded volume effects should be correctly accounted for, with any further deviations hinting at interesting physical and chemical properties of the molecules.

Acknowledgement

The work of AM, MKB, BPK, MKC-R, ZS, and AN was supported by the National Science Centre of Poland Sonata-Bis grant no. UMO-2016/22/E/NZ1/00656 to AN. The work of RW and ML was supported by the National Science Centre of Poland Sonata grant no. 2018/31/D/ST3/02408 to ML. The authors thank Prof. Nahum Sonenberg and Dr. Marc Fabian for sharing plasmids for some protein constructs and Dr. Joanna Żuberek and Dr. Mateusz Kogut for helpful discussion. The research was performed in the NanoFun laboratories co-financed by ERDF within the POIG.02.02.00-00-025/09 Program.

Supporting Information Available

`Tables S1 S5 Figures S1-S8 Materials and Methods.pdf` containing experimental R_h , Figures S1, S2, S3, S4, S5, S6, S7, S8, and additional methodological details. `Tables S2-S4.xlsx` containing theoretical R_h from GLM, phenomenological R_h, and protein sequences with marked domains.

The programs referenced in this article can be found on GitHub. For reader convenience, an API and a command-line Python utility, `glm_mda_diffusion`, have been provided.[108] This package builds upon previous ones, namely `pychastic` and `sarw_spheres`, both of which are also accessible on GitHub.

References

(1) Oldfield, C. J.; Dunker, A. K. Intrinsically disordered proteins and intrinsically disordered protein regions. *Annual Review of Biochemistry* **2014**, *83*, 553–584.

(2) Wright, P. E.; Dyson, H. J. Intrinsically disordered proteins in cellular signalling and regulation. *Nature Reviews Molecular Cell Biology* **2015**, *16*, 18–29.

(3) Ward, J. J.; Sodhi, J. S.; McGuffin, L. J.; Buxton, B. F.; Jones, D. T. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *Journal of Molecular Biology* **2004**, *337*, 635–645.

(4) Shea, J.-E.; Best, R. B.; Mittal, J. Physics-based computational and theoretical approaches to intrinsically disordered proteins. *Current Opinion in Structural Biology* **2021**, *67*, 219–225.

(5) Uversky, V. N. Intrinsically disordered proteins and their environment: effects of strong denaturants, temperature, pH, counter ions, membranes, binding partners, osmolytes, and macromolecular crowding. *The Protein Journal* **2009**, *28*, 305–325.

(6) Langridge, T. D.; Tarver, M. J.; Whitten, S. T. Temperature effects on the hydrodynamic radius of the intrinsically disordered N-terminal region of the p53 protein. *Proteins: Structure, Function, and Bioinformatics* **2014**, *82*, 668–678.

(7) Müller-Späth, S.; Soranno, A.; Hirschfeld, V.; Hofmann, H.; Rüegger, S.; Reymond, L.; Nettels, D.; Schuler, B. Charge interactions can dominate the dimensions of intrinsically disordered proteins. *Proceedings of the National Academy of Sciences* **2010**, *107*, 14609–14614.

(8) Wohl, S.; Jakubowski, M.; Zheng, W. Salt-dependent conformational changes of intrinsically disordered proteins. *The Journal of Physical Chemistry Letters* **2021**, *12*, 6684–6691.

(9) Moses, D.; Yu, F.; Ginell, G. M.; Shamoon, N. M.; Koenig, P. S.; Holehouse, A. S.; Sukenik, S. Revealing the hidden sensitivity of intrinsically disordered proteins to

their chemical environment. *The Journal of Physical Chemistry Letters* **2020**, *11*, 10131–10136.

(10) Wang, Y.; Benton, L. A.; Singh, V.; Pielak, G. J. Disordered protein diffusion under crowded conditions. *The Journal of Physical Chemistry Letters* **2012**, *3*, 2703–2706.

(11) Bah, A.; Vernon, R. M.; Siddiqui, Z.; Krzeminski, M.; Muhandiram, R.; Zhao, C.; Sonenberg, N.; Kay, L. E.; Forman-Kay, J. D. Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. *Nature* **2015**, *519*, 106–109.

(12) Vancraenenbroeck, R.; Harel, Y. S.; Zheng, W.; Hofmann, H. Polymer effects modulate binding affinities in disordered proteins. *Proceedings of the National Academy of Sciences* **2019**, *116*, 19506–19512.

(13) Borgia, A.; Borgia, M. B.; Bugge, K.; Kissling, V. M.; Heidarsson, P. O.; Fernandes, C. B.; Sottini, A.; Soranno, A.; Buholzer, K. J.; Nettels, D. et al. Extreme disorder in an ultrahigh-affinity protein complex. *Nature* **2018**, *555*, 61–66.

(14) Seiffert, P.; Bugge, K.; Nygaard, M.; Haxholm, G. W.; Martinsen, J. H.; Pedersen, M. N.; Arleth, L.; Boomsma, W.; Kragelund, B. B. Orchestration of signaling by structural disorder in class 1 cytokine receptors. *Cell Communication and Signaling* **2020**, *18*, 1–30.

(15) Evans, J. S. The biomineralization proteome: protein complexity for a complex bioceramic assembly process. *Proteomics* **2019**, *19*, 1900036.

(16) Krois, A. S.; Dyson, H. J.; Wright, P. E. Long-range regulation of p53 DNA binding by its intrinsically disordered N-terminal transactivation domain. *Proceedings of the National Academy of Sciences* **2018**, *115*, E11302–E11310.

(17) Fletcher, C. M.; McGuire, A. M.; Gingras, A.-C.; Li, H.; Matsuo, H.; Sonenberg, N.; Wagner, G. 4E binding proteins inhibit the translation factor eIF4E without folded structure. *Biochemistry* **1998**, *37*, 9–15.

(18) Fletcher, C. M.; Wagner, G. The interaction of eIF4E with 4E-BP1 is an induced fit to a completely disordered protein. *Protein Science* **1998**, *7*, 1639–1642.

(19) Gingras, A.-C.; Raught, B.; Gygi, S. P.; Niedzwiecka, A.; Miron, M.; Burley, S. K.; Polakiewicz, R. D.; Wyslouch-Cieszynska, A.; Aebersold, R.; Sonenberg, N. Hierarchical phosphorylation of the translation inhibitor 4E-BP1. *Genes & Development* **2001**, *15*, 2852–2864.

(20) Sheu-Gruttadauria, J.; MacRae, I. J. Phase transitions in the assembly and function of human miRISC. *Cell* **2018**, *173*, 946–957.

(21) Cieplak-Rotowska, M. K.; Tarnowski, K.; Rubin, M.; Fabian, M. R.; Sonenberg, N.; Dadlez, M.; Niedzwiecka, A. Structural dynamics of the GW182 silencing domain including its RNA recognition motif (RRM) revealed by hydrogen-deuterium exchange mass spectrometry. *Journal of The American Society for Mass Spectrometry* **2018**, *29*, 158–173.

(22) Raisch, T.; Valkov, E. Regulation of the multisubunit CCR4-NOT deadenylase in the initiation of mRNA degradation. *Current Opinion in Structural Biology* **2022**, *77*, 102460.

(23) Louros, N.; Schymkowitz, J.; Rousseau, F. Mechanisms and pathology of protein misfolding and aggregation. *Nature Reviews Molecular Cell Biology* **2023**, 1–22.

(24) Chakraborty, P.; Zweckstetter, M. Role of aberrant phase separation in pathological protein aggregation. *Current Opinion in Structural Biology* **2023**, *82*, 102678.

(25) Banani, S. F.; Rice, A. M.; Peeples, W. B.; Lin, Y.; Jain, S.; Parker, R.; Rosen, M. K. Compositional control of phase-separated cellular bodies. *Cell* **2016**, *166*, 651–663.

(26) Forman-Kay, J. D.; Ditlev, J. A.; Nosella, M. L.; Lee, H. O. What are the distinguishing features and size requirements of biomolecular condensates and their implications for RNA-containing condensates? *RNA* **2022**, *28*, 36–47.

(27) Białobrzewski, M. K.; Klepka, B. P.; Michaś, A.; Cieplak-Rotowska, M. K.; Staszałek, Z.; Niedźwiecka, A. Diversity of hydrodynamic radii of intrinsically disordered proteins. *European Biophysics Journal* **2023**, *52*, 607–618.

(28) Nygaard, M.; Kragelund, B. B.; Papaleo, E.; Lindorff-Larsen, K. An efficient method for estimating the hydrodynamic radius of disordered protein conformations. *Biophysical Journal* **2017**, *113*, 550–557.

(29) Pesce, F.; Newcombe, E. A.; Seiffert, P.; Tranchant, E. E.; Olsen, J. G.; Grace, C. R.; Kragelund, B. B.; Lindorff-Larsen, K. Assessment of models for calculating the hydrodynamic radius of intrinsically disordered proteins. *Biophysical Journal* **2023**, *122*, 310–321.

(30) Tomasso, M. E.; Tarver, M. J.; Devarajan, D.; Whitten, S. T. Hydrodynamic radii of intrinsically disordered proteins determined from experimental polyproline II propensities. *PLoS Computational Biology* **2016**, *12*, e1004686.

(31) English, L. R.; Tilton, E. C.; Ricard, B. J.; Whitten, S. T. Intrinsic α helix propensities compact hydrodynamic radii in intrinsically disordered proteins. *Proteins: Structure, Function, and Bioinformatics* **2017**, *85*, 296–311.

(32) Mao, A. H.; Crick, S. L.; Vitalis, A.; Chicoine, C. L.; Pappu, R. V. Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proceedings of the National Academy of Sciences* **2010**, *107*, 8183–8188.

(33) Różycki, B.; Kim, Y. C.; Hummer, G. SAXS ensemble refinement of ESCRT-III CHMP3 conformational transitions. *Structure* **2011**, *19*, 109–116.

(34) Ozenne, V.; Bauer, F.; Salmon, L.; Huang, J.-r.; Jensen, M. R.; Segard, S.; Bernadó, P.; Charavay, C.; Blackledge, M. Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. *Bioinformatics* **2012**, *28*, 1463–1470.

(35) Mittal, J.; Yoo, T. H.; Georgiou, G.; Truskett, T. M. Structural ensemble of an intrinsically disordered polypeptide. *The Journal of Physical Chemistry B* **2013**, *117*, 118–124.

(36) Mittal, A.; Holehouse, A. S.; Cohan, M. C.; Pappu, R. V. Sequence-to-conformation relationships of disordered regions tethered to folded domains of proteins. *Journal of Molecular Biology* **2018**, *430*, 2403–2421.

(37) Das, P.; Matysiak, S.; Mittal, J. Looking at the disordered proteins through the computational microscope. *ACS Central Science* **2018**, *4*, 534–542.

(38) Estaña, A.; Sibille, N.; Delaforge, E.; Vaisset, M.; Cortés, J.; Bernadó, P. Realistic ensemble models of intrinsically disordered proteins using a structure-encoding coil database. *Structure* **2019**, *27*, 381–391.

(39) Baul, U.; Chakraborty, D.; Mugnai, M. L.; Straub, J. E.; Thirumalai, D. Sequence effects on size, shape, and structural heterogeneity in intrinsically disordered proteins. *The Journal of Physical Chemistry B* **2019**, *123*, 3462–3474.

(40) de la Torre, J. G.; Cifre, J. H. Hydrodynamic properties of biomacromolecules and macromolecular complexes: concepts and methods. A tutorial mini-review. *Journal of Molecular Biology* **2020**, *432*, 2930–2948.

(41) Gomes, G.-N. W.; Krzeminski, M.; Namini, A.; Martin, E. W.; Mittag, T.; Head-Gordon, T.; Forman-Kay, J. D.; Gradinaru, C. C. Conformational ensembles of an intrinsically disordered protein consistent with NMR, SAXS, and single-molecule FRET. *Journal of the American Chemical Society* **2020**, *142*, 15697–15710.

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

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

(44) Schmidt, R. R.; Cifre, J. H.; de la Torre, J. G. Translational diffusion coefficients of macromolecules. *Eururopean Physical Journal E* **2012**, *35*, 130.

(45) de la Torre, J. G. *Analytical Ultracentrifugation. Instrumentation, Software, and Applications*; Springer, 2016; pp 195–217.

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

(47) Fixman, M. Variational bounds for polymer transport coefficients. *Journal of Chemical Physics* **1983**, *78*, 1588–1593.

(48) Happel, J.; Brenner, H. *Low Reynolds Number Hydrodynamics*; Noordhoff, Leyden, 1973.

(49) Cichocki, B.; Ekiel-Jeżewska, M. L.; Wajnryb, E. Communication: Translational Brownian motion for particles of arbitrary shape. *Journal of Chemical Physics* **2012**, *136*.

(50) Wegener, W. A. Bead models of segmentally flexible macromolecules. *Journal of Chemical Physics* **1982**, *76*, 6425–6430.

(51) Harvey, S. C.; Mellado, P.; García de la Torre, J. Hydrodynamic resistance and diffusion coefficients of segmentally flexible macromolecules with two subunits. *Journal of Chemical Physics* **1983**, *78*, 2081–2090.

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

(53) Cichocki, B.; Rubin, M.; Niedzwiecka, A.; Szymczak, P. Diffusion coefficients of elastic macromolecules. *Journal of Fluid Mechanics* **2019**, *878*, R3.

(54) Karplus, M.; Petsko, G. A. Molecular dynamics simulations in biology. *Nature* **1990**, *347*, 631–639.

(55) Doi, M.; Edwards, S. F. *The Theory of Polymer Dynamics*; Oxford University Press, 1988; Vol. 73.

(56) Prakash, J. R. The kinetic theory of dilute solutions of flexible polymers: Hydrodynamic interaction. *Rheology Series* **1999**, *8*, 467–517.

(57) Szymczak, P.; Cieplak, M. Hydrodynamic effects in proteins. *Journal of Physics: Condensed Matter* **2011**, *23*, 033102.

(58) Skolnick, J. Perspective: On the importance of hydrodynamic interactions in the subcellular dynamics of macromolecules. *The Journal of Chemical Physics* **2016**, *145*.

(59) Frenkel, D.; Smit, B. *Understanding molecular simulation: from algorithms to applications*; Elsevier, 2001; Vol. 1.

(60) Rouse, P. E. A Theory of the Linear Viscoelastic Properties of Dilute Solutions of Coiling Polymers. *The Journal of Chemical Physics* **1953**, *21*, 1272–1280.

(61) Zimm, B. H. Dynamics of Polymer Molecules in Dilute Solution: Viscoelasticity, Flow Birefringence and Dielectric Loss. *Journal of Chemical Physics* **1956**, *24*, 269–278.

(62) Kirkwood, J. G.; Riseman, J. The intrinsic viscosities and diffusion constants of flexible macromolecules in solution. *The Journal of Chemical Physics* **1948**, *16*, 565–573.

(63) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589.

(64) Ruff, K. M.; Pappu, R. V. AlphaFold and implications for intrinsically disordered proteins. *Journal of Molecular Biology* **2021**, *433*, 167208.

(65) Murphy, L. R.; Matubayasi, N.; Payne, V. A.; Levy, R. M. Protein hydration and unfolding–insights from experimental partial specific volumes and unfolded protein models. *Folding and Design* **1998**, *3*, 105–118.

(66) Jones, D. T.; Cozzetto, D. DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics* **2015**, *31*, 857–863.

(67) Choi, Y.; Agarwal, S.; Deane, C. M. How long is a piece of loop? *PeerJ* **2013**, *1*, e1.

(68) Longsworth, L. Diffusion measurements, at 25°C, of aqueous solutions of amino acids, peptides and sugars. *Journal of the American Chemical Society* **1953**, *75*, 5705–5709.

(69) Zuk, P.; Wajnryb, E.; Mizerski, K.; Szymczak, P. Rotne–Prager–Yamakawa approximation for different-sized particles in application to macromolecular bead models. *Journal of Fluid Mechanics* **2014**, *741*, R5.

(70) Cox, R. The motion of long slender bodies in a viscous fluid Part 1. General theory. *Journal of Fluid Mechanics* **1970**, *44*, 791–810.

(71) Johnson, R. E.; Wu, T. Y. Hydromechanics of low-Reynolds-number flow. Part 5. Motion of a slender torus. *Journal of Fluid Mechanics* **1979**, *95*, 263–277.

(72) Majumdar, S. R.; O'Neill, M. E. On axisymmetric stokes flow past a torus. *Zeitschrift für Angewandte Mathematik und Physik ZAMP* **1977**, *28*, 541–550.

(73) Waszkiewicz, R.; Szymczak, P.; Lisicki, M. Stability of sedimenting flexible loops. *Journal of Fluid Mechanics* **2021**, *919*, A14.

(74) Kirkwood, J. G. The general theory of irreversible processes in solutions of macromolecules. *Journal of Polymer Science* **1954**, *12*, 1–14.

(75) Liu, B.; Dünweg, B. Translational diffusion of polymer chains with excluded volume and hydrodynamic interactions by Brownian dynamics simulation. *Journal of Chemical Physics* **2003**, *118*, 8061–8072.

(76) Vovk, A.; Zilman, A. Effects of Sequence Composition, Patterning and Hydrodynamics on the Conformation and Dynamics of Intrinsically Disordered Proteins. *International Journal of Molecular Sciences* **2023**, *24*, 1444.

(77) Kim, S.; Karrila, S. J. *Microhydrodynamics: Principles and Selected Applications*; Butterworth-Heinemann: London, 1991.

(78) Mazur, P.; van Saarloos, W. Many-sphere hydrodynamic interactions and mobilities in a suspension. *Physica A* **1982**, *115*, 21–57.

(79) Cichocki, B.; Felderhof, B. U.; Hinsen, K.; Wajnryb, E.; Bławzdziewicz, J. Friction and mobility of many spheres in Stokes flow. *Journal of Chemical Physics* **1994**, *100*, 3780–3790.

(80) Rotne, J.; Prager, S. Variational treatment of hydrodynamic interaction in polymers. *Journal of Chemical Physics* **1969**, *50*, 4831–4837.

(81) Yamakawa, H. Transport properties of polymer chains in dilute solution: Hydrodynamic interaction. *Jorunal of Chemical Physics* **1970**, *53*, 436–443.

(82) Zuk, P.; Cichocki, B.; Szymczak, P. GRPY: An Accurate Bead Method for Calculation of Hydrodynamic Properties of Rigid Biomacromolecules. *Biophysical Journal* **2018**, *115*, 782–800.

(83) Waszkiewicz, R.; Bartczak, M.; Kolasa, K.; Lisicki, M. Pychastic: Precise Brownian dynamics using Taylor-Itō integrators in Python. *SciPost Physics Codebases* **2023**, 11.

(84) de Val, N.; Declercq, J.-P.; Lim, C. K.; Crichton, R. R. Structural analysis of haemin demetallation by L-chain apoferritins. *Journal of Inorganic Biochemistry* **2012**, *112*, 77–84.

(85) McCubbin, W. D.; Kay, C. M.; Lane, B. G. Hydrodynamic and optical properties of the wheat germ Em protein. *Canadian Journal of Biochemistry and Cell Biology* **1985**, *63*, 803–811.

(86) Guez, V.; Nair, S.; Chaffotte, A.; Bedouelle, H. The anticodon-binding domain of tyrosyl-tRNA synthetase: state of folding and origin of the crystallographic disorder. *Biochemistry* **2000**, *39*, 1739–1747.

(87) Bouvier, M.; Stafford, W. F. Probing the three-dimensional structure of human calreticulin. *Biochemistry* **2000**, *39*, 14950–14959.

(88) Danielsson, J.; Jarvet, J.; Damberg, P.; Gräslund, A. Translational diffusion measured by PFG-NMR on full length and fragments of the Alzheimer Aβ (1–40) peptide. Determination of hydrodynamic radii of random coil peptides of varying length. *Magnetic Resonance in Chemistry* **2002**, *40*, S89–S97.

(89) Karlin, D.; Longhi, S.; Receveur, V.; Canard, B. The N-terminal domain of the phosphoprotein of morbilliviruses belongs to the natively unfolded class of proteins. *Virology* **2002**, *296*, 251–262.

(90) Longhi, S.; Receveur-Bréchot, V.; Karlin, D.; Johansson, K.; Darbon, H.; Bhella, D.; Yeo, R.; Finet, S.; Canard, B. The C-terminal domain of the measles virus nucleoprotein is intrinsically disordered and folds upon binding to the C-terminal moiety of the phosphoprotein. *Journal of Biological Chemistry* **2003**, *278*, 18638–18648.

(91) Sánchez-Puig, N.; Veprintsev, D. B.; Fersht, A. R. Human full-length Securin is a natively unfolded protein. *Protein Science* **2005**, *14*, 1410–1418.

(92) Sánchez-Puig, N.; Veprintsev, D. B.; Fersht, A. R. Binding of natively unfolded HIF-1α ODD domain to p53. *Molecular Cell* **2005**, *17*, 11–21.

(93) Khaymina, S. S.; Kenney, J. M.; Schroeter, M. M.; Chalovich, J. M. Fesselin is a natively unfolded protein. *Journal of Proteome Research* **2007**, *6*, 3648–3654.

(94) Zhang, X.; Perugini, M. A.; Yao, S.; Adda, C. G.; Murphy, V. J.; Low, A.; Anders, R. F.; Norton, R. S. Solution conformation, backbone dynamics and lipid interactions of the intrinsically unstructured malaria surface protein MSP2. *Journal of Molecular Biology* **2008**, *379*, 105–121.

(95) Mittag, T.; Orlicky, S.; Choy, W.-Y.; Tang, X.; Lin, H.; Sicheri, F.; Kay, L. E.; Tyers, M.; Forman-Kay, J. D. Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor. *Proceedings of the National Academy of Sciences* **2008**, *105*, 17772–17777.

(96) Paz, A.; Zeev-Ben-Mordehai, T.; Lundqvist, M.; Sherman, E.; Mylonas, E.; Weiner, L.; Haran, G.; Svergun, D. I.; Mulder, F. A.; Sussman, J. L. et al. Biophysical characterization of the unstructured cytoplasmic domain of the human neuronal adhesion protein neuroligin 3. *Biophysical Journal* **2008**, *95*, 1928–1944.

(97) Habchi, J.; Mamelli, L.; Darbon, H.; Longhi, S. Structural disorder within Henipavirus nucleoprotein and phosphoprotein: from predictions to experimental assessment. *PLoS One* **2010**, *5*, e11684.

(98) Choi, U. B.; McCann, J. J.; Weninger, K. R.; Bowen, M. E. Beyond the random coil: stochastic conformational switching in intrinsically disordered proteins. *Structure* **2011**, *19*, 566–576.

(99) Perez, R. B.; Tischer, A.; Auton, M.; Whitten, S. T. Alanine and proline content modulate global sensitivity to discrete perturbations in disordered proteins. *Proteins: Structure, Function, and Bioinformatics* **2014**, *82*, 3373–3384.

(100) Yarawsky, A. E.; English, L. R.; Whitten, S. T.; Herr, A. B. The proline/glycine-rich region of the biofilm adhesion protein Aap forms an extended stalk that resists compaction. *Journal of Molecular Biology* **2017**, *429*, 261–279.

(101) Poznar, M.; Hołubowicz, R.; Wojtas, M.; Gapiński, J.; Banachowicz, E.; Patkowski, A.; Ożyhar, A.; Dobryszycki, P. Structural properties of the intrinsically disordered, multiple calcium ion-binding otolith matrix macromolecule-64 (OMM-64). *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* **2017**, *1865*, 1358–1371.

(102) Więch, A.; Rowińska-Żyrek, M.; Wątły, J.; Czarnota, A.; Hołubowicz, R.; Szewczuk, Z.; Ożyhar, A.; Orłowski, M. The intrinsically disordered C-terminal F

domain of the ecdysteroid receptor from Aedes aegypti exhibits metal ion-binding ability. *The Journal of Steroid Biochemistry and Molecular Biology* **2019**, *186*, 42–55.

(103) Marsh, J. A.; Forman-Kay, J. D. Sequence determinants of compaction in intrinsically disordered proteins. *Biophysical Journal* **2010**, *98*, 2383–2390.

(104) Le Guillou, J.; Zinn-Justin, J. Critical exponents for the n-vector model in three dimensions from field theory. *Physical Review Letters* **1977**, *39*, 95.

(105) Ding, F.; Jha, R. K.; Dokholyan, N. V. Scaling behavior and structure of denatured proteins. *Structure* **2005**, *13*, 1047–1054.

(106) Philo, J. S. SEDNTERP: a calculation and database utility to aid interpretation of analytical ultracentrifugation and light scattering data. *European Biophysics Journal* **2023**, 1–34.

(107) Wilkins, D. K.; Grimshaw, S. B.; Receveur, V.; Dobson, C. M.; Jones, J. A.; Smith, L. J. Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry* **1999**, *38*, 16424–16431.

(108) Waszkiewicz, R. GitHub: glm-mda-diffusion. https://github.com/RadostW/glm-mda-diffusion, 2023; Accessed: 2023-12-25.

3.5.1 Comparing Conformations to SAXS Data

As detailed in Waszkiewicz et al. [5], computation of the diffusion coefficient for a given IDP requires one to generate samples from the equilibrium ensemble. To this end, we proposed to use the Globule-Linker Model, in which statistical variability was represented with a self-avoiding random walk. To assess the quality of the conformer generation method independently of hydrodynamic modelling, we utilised available SAXS data, not included in the final manuscript. One example of such a comparison is illustrated in Figure 3.2, where experimental data from Sicorello et al. [77] is compared with our implementation of the Globule-Linker Model (computed with `sarw-spheres` code; blue solid line) combined with our implementation of the one-site-per-amino-acid scattering model (computed with `saxs-single-bead` code; orange line). This model is based on form factors computed by Tong, Yang, and Lu [60].

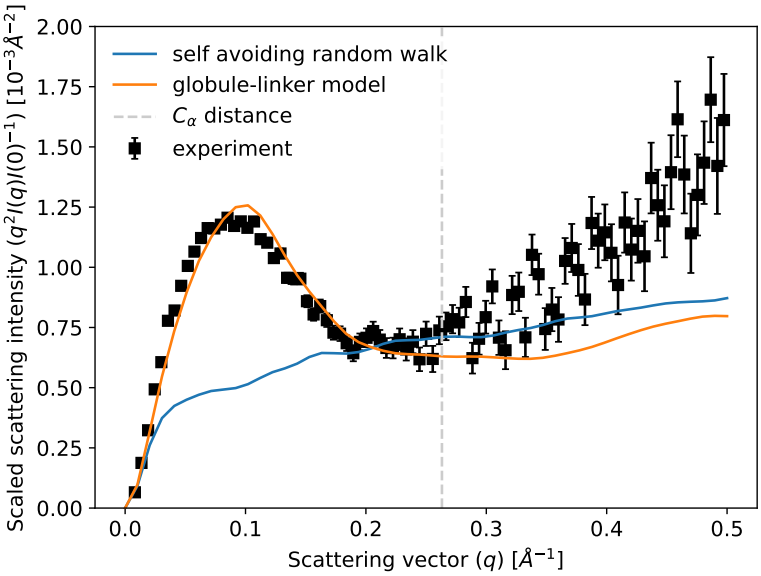


Figure 3.2: A Kratky plot of scaled scattering intensity, along with a comparison of our theoretical predictions generated using `saxs_single_bead`, is presented against experimental data for ataxin-3 (pdb id `1yzb`), as reported by Sicorello et al. [77]. Predictions based on conformers generated using the globule-linker model (orange line) demonstrate good agreement for scattering vectors corresponding to intramolecular distances larger than the C_α distance, which represents the coarse-graining resolution. An even simpler conformer generation scheme based on self-avoiding random walk, which completely disregards domain information (blue line), is capable of capturing only the initial slope of the scattering curve.

The conformations used to predict the SAXS curve in Figure 3.2 were generated using the Globule-Linker Model engine, with the globules replaced by crystallographic data retrieved from the Protein Data Bank (PDB)[35]. The excellent agreement of the orange curve in the range of scattering vectors corresponding to features larger than the C_α distance (small values of q) is a result of the combination of adequate modelling of both the rigid and flexible parts.

The treatment of the distribution in the Globule-Linker engine relies on the heuristic assumption that bond vectors are uniformly distributed on a sphere (barring excluded volume interactions). The theoretical basis of this approach relies on the treatment of the stiff spring limit in which bonds between the constituent parts of a molecule are treated as arbitrarily rigid springs. We examine this limit in greater detail to better inform the treatment and simulation of freely jointed chains.

The author would like to thank Dr. Bartosz Różycki for his guidance regarding SAXS experimental data.

3.6 Paper VI: *The trimer paradox: the effect of stiff constraints on equilibrium distributions in overdamped dynamics*

PAPER VI

"*The trimer paradox: the effect of stiff constraints on equilibrium distributions in overdamped dynamics*"

Radost Waszkiewicz and Maciej Lisicki

COMMENTARY

Feely jointed chains form the basis of many soft matter models and the determination of their statistical properties were amongst the earliest victories of the field, as demonstrated by the predictions of end-to-end distances of Rouse and Zimm. Moreover, a freely jointed chain is a fundamental component of globule-linker conformer generation scheme which used together with the Minimum Dissipation Approximation (MDA, [25]), yielded predictions for the hydrodynamic radii of IDPs in [5]. On the other hand, MDA assumes loose binding with short time diffusion along the bonds contributing to the short time diffusion coefficient. In previous works, we computed hydrodynamic radii only in two extreme cases: very loosely bound components (MDA) or completely rigid conglomerates (GRPY [17]), with the intermediate regime not treated by either. A first step towards designing an extension of coarse-grained molecules with rigid bond constraints but capable of reorientation is to understand the equilibrium distributions in such cases.

The problem of a freely jointed trimer—three beads connected by flexible springs—in the limit of infinitely stiff bonds was first approached by Fixman [78]. They identified some, but not all, of the complications of the stiff limit. Further refinements of this work, outlined the correct treatment but only in a low-dimensional setting and, to our best knowledge, no general treatment for larger systems was presented. Moreover, no attention was given to the computational feasibility of the techniques outlined.

In order to fill this gap, in the following we have rephrased the stiff bonds limit in a mathematically rigorous way. By examining the action of the limiting distribution on a smooth compactly supported trial function, all the terms contributing to the limiting distribution arise naturally.

By examining a general parametrisation which separates soft from hard coordinates, we were able to provide an efficient way of computing the confinement shape factor. We have thus bridged the gap in previous works by providing an explicit treatment of cases with many degrees of freedom, and thus with many constrains. By considering a cyclic tetramer, we have shown that the shape-factor corrections can be arbitrarily large, even for molecular bonds modelled with identical harmonic springs.

In this study the PhD candidate: proposed the research question of the manuscript, performed the calculation of the limiting distribution, performed the numerical simulations used in illustrations, wrote the first draft and edited all subsequent versions of the manuscript.

The trimer paradox: the effect of stiff constraints on equilibrium distributions in overdamped dynamics

Radost Waszkiewicz[1, *] and Maciej Lisicki[1, †]

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

(Dated: March 21, 2024)

We reconsider the classical problem of a freely joined chain of Brownian particles connected by elastic springs and study its conformational probability distribution function in the overdamped regime in the limit of infinite stiffness of constraints. We show that the well-known solution by Fixman is missing a shape-related term, later alluded to but not computed by Helfand. In our approach, the shape term, also termed zero-point energy, arises explicitly from a careful treatment of the distributional limit. We present a computationally feasible method of calculation of the shape term and demonstrate its validity in a couple of examples.

I. INTRODUCTION

Molecular Dynamics and Brownian Dynamics simulations are now standard tools for detailed modelling of a plethora of molecular and mesoscopic systems, where structural complexity poses a challenge to theoretical calculations. Upon the introduction of a suitable coarse-graining scheme to represent the composition of a given molecule, its conformational space is prescribed by specifying intramolecular interactions, such as chemical bonds or electrostatic forces, between the subunits. The nature of these interactions endows molecular models with flexibility, which then in turn affects their conformational variability, as well as their macroscopic and statistical properties, such as rheology, diffusivity, and thermodynamics, and also biological function.

A popular idealisation for a wide variety of systems is purely mechanical, where the molecule is represented by beads connected with springs, an idea which probably originated from Kramers [1], who built on concepts of Kuhn, and termed it the pearl-necklace model. Variations of this idea are now the cornerstone of polymer physics [2], but have also been used to describe other classical many-body systems. Physically, the introduction of intramolecular interactions leads to constrained dynamics with multiple time scales. Typically, the characteristic time scales associated with internal vibrations of molecular bonds are much shorter than those of translational and rotational motion or bond angle dynamics. The disparity of time scale of relaxation of hard (or stiff) degrees of freedom (such as bond lengths) in comparison to soft degrees of freedom (such as bond angles) is a source of stiffness in the problem and hinders fast numerical simulation of such systems [3].

In many applications, the dynamics of the hard degrees of freedom are of secondary importance, and one method of circumventing this difficulty is to treat them as fully constrained (i.e. rigid), thus eliminating the short timescale, and enabling faster calculation. This is the basis of many algorithms, such as SHAKE [4] and its further extensions. During the development of such rigid models, a disconnect was discovered between the equilibrium bond angle distributions in the rigid simulations and the simulations with bonds modelled as very stiff springs.

The simplest example in which the complexity of the constrained dynamics can be appreciated is the case of a flexible trimer: a hypothetical molecule containing 3 subunits and 2 harmonic bonds, depicted in Fig. 2a. The angle between the two bonds is denoted by ψ. In the classical textbook [3], Frenkel and Smit present an apparently paradoxical result: the marginal distribution of the bond angle in the case of rigid bonds $\mathrm{d}p_{\mathrm{rigid}}$ and the limiting distribution of spring-like bonds where the spring stiffness k is taken to infinity $\mathrm{d}p_{k\to\infty}$ do not coincide. More precisely, they report

$$\mathrm{d}p_{k\to\infty} = N_1 \sin\psi \, \mathrm{d}\psi \tag{1}$$

$$\mathrm{d}p_{\mathrm{rigid}} = N_2 \sin\psi \sqrt{1 - \frac{\cos^2\psi}{4}} \, \mathrm{d}\psi \tag{2}$$

for appropriate normalisation constants N_1 and N_2. Qualitatively, distribution (1) corresponds to a spherically uniform distribution of the second bond direction when the coordinates are aligned with the first bond, while in Eq. (2) it seems that an additional force resists bond alignment. In an early and widely cited work, Fixman [5] presented a derivation to argue that the ratio of the two probability densities, Eqs. (1) and (2), can be computed with the knowledge of the constraining surfaces alone and does not require knowledge of the shape of the confining potential. This presupposes that both of these distributions are well defined whenever the constraint is given, which we argue is not the case.

In this contribution, we present a clear mathematical procedure leading to the aforementioned limits in the overdamped regime when inertial effects can be neglected. We show that the distribution given by Eq. (1) is not completely determined by fixing the bond lengths. To correctly predict the limiting distribution, knowledge about the nature of confinement is needed and its structure survives in the final expression. Furthermore, we

* radost.waszkiewicz@gmail.com
† mklis@fuw.edu.pl

demonstrate that Fixman's expressions for the stiff limit miss a term, which is important even in the case of harmonic springs.

II. STIFF VS. RIGID CONSTRAINTS

The first calculations of the equilibrium properties of flexible polymers can be traced to the papers by Kramers [1], Gō and Sheraga [6] and Fixman [5]. Results of the latter paper were used essentially without changes in works such as [7–10]. Early papers focused on the treatment of constraints in the Lagrangian and Hamiltonian pictures, both classical and quantum-mechanical, by defining the soft and stiff degrees of freedom and integrating out generalised coordinates corresponding to the hard variables. The limiting distribution is then obtained for infinitely stiff constraints when stiff variables are eliminated.

Fixman [5] presented a purely classical calculation of the limiting distribution, correctly writing the action of the distribution on a test function but erroneously assuming that the constraining potential depends only on the hard coordinates. As a result, details of the confining potential cannot appear in his final expressions, which contain only the projected phase-space volume element terms. Gō and Sheraga [6] treated the problem quantum-mechanically and correctly identified two contributions to the equilibrium probability density of the soft coordinates: the projected volume elements and the zero-point energy of the vibrational motion of the hard coordinates. However, they then stated that the zero-energy contribution in typical molecular systems varies only a little when changing the soft coordinates and may be altogether neglected. We show in the following that this is not necessarily the case in overdamped dynamics. By focusing on a concrete classical two-dimensional example with one constraint, Helfand [11] correctly identified the rigid-rod type distribution as a uniform distribution on the constraining manifold and computed the stiff-spring limit in a similar simple case. Unfortunately, these results are not expressed in a form applicable to higher-dimensional examples with multiple constraints.

Due to this lack of generality, later works such as that of Hinch [8], cite Helfand as a reference to an erroneous claim that the bond angle distribution becomes non-uniform in the case of rigid-rod system. The pitfall lies in the assumption that the distribution should be uniform under an action of a rotation matrix on a single bond, presumably by analogy with a dimer. However, there is a distinct difference between these two cases; In the case of a dimer, the true physical symmetries of the equations of motion – rigid body rotation of the whole system and translations – act transitively on the space of all possible configurations, thus the equilibrium distribution can be determined from symmetry arguments. More generally, for a rigid body the rotation group SO(3) can be treated as a topological manifold, resulting in a natural notion of uniformity, with the uniform distribution given by the Haar measure. For the trimer, however, we have configurations that are meaningfully distinct. There is no symmetry that changes the bond angle ψ. While the equilibrium distribution for 6 out of the 7 degrees of freedom can be determined from the six symmetries of translation and rotation, the relative weight of configurations with different ψ has to refer to the physical problem, not just geometric considerations.

Hinch [8] approached the problem differently, by manipulating the Langevin equations, which are notoriously difficult to handle, and erroneously assumed that details of the confining potential are of little importance to the equilibrium distribution of the constrained configurations. Van Kampen and Lodder [12] noted that the approach of Helfand [11] is generally applicable to constrained systems and comment that the discrepancy between the stiff-spring and rigid-rod distributions in the case of the trimer molecule is due to the fact that *the width of this gully is not the same everywhere* [12]. However, they did not quantify the effect of this width in a higher-dimensional case. As a result, the knowledge of the influence of the shape of the constraining potential appears to have been lost as multiple works (for example, [3, 13, 14]) mention only the volume element densities and omit the *zero point energy* or, equivalently, the potential *shape* terms from their descriptions of constrained dynamics, widely using the result of Fixman [5] as reference.

In the following, we show a way to compute the limiting distribution by integrating out the hard degrees of freedom in the overdamped limit. We explicitly find the metric contribution arising from the transformation of coordinates and the shape term that represents the details of the constraining potential and which remains imprinted on the configurational distribution when the hard coordinates are taken to be infinitely stiff.

III. THE STIFF SPRING LIMIT

We present a derivation of the limiting distribution in the following setting, sketched in Fig. 1. Consider a molecule (or a polymer) with a handful of subunits (atoms, beads, or monomers) and N degrees of freedom in a heat bath at constant temperature T, with $\beta = 1/k_B T$. The conformation of the molecule is described by an N-dimensional vector $\boldsymbol{q}$. We parameterise conformations by Cartesian coordinates of monomers q_i with $i \in 1 \dots N$. The potential energy of the molecule has two components: the conformation-dependent energy $U(\boldsymbol{q})$ and the confining (springs) potential $k^2 W(\boldsymbol{q})$ where k is a large parameter describing the spring stiffness. Note that k here would correspond to the square root of a harmonic spring constant. The probability density p_k of the equilibrium distribution of the system configuration with respect to the Lebesgue measure $\mathrm{d}\boldsymbol{q}$ is given

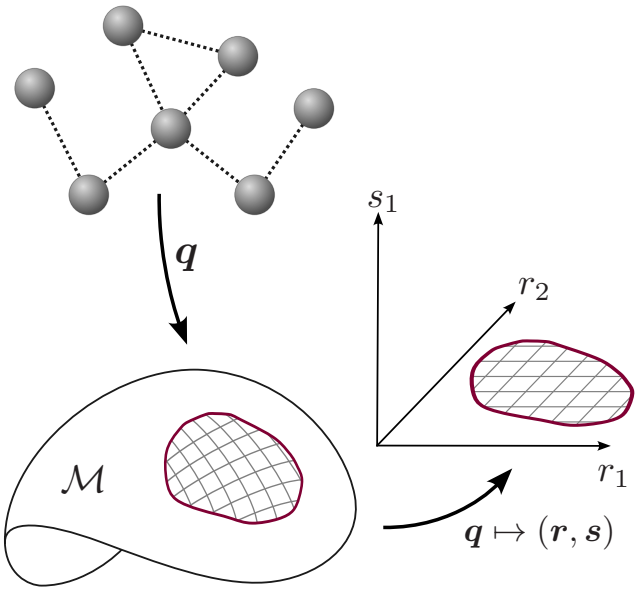


FIG. 1. A flexible polymer with N degrees of freedom is modelled as beads connected with springs. Conformations of the polymer are described by a vector $\boldsymbol{q}$. The potential energy of the confining springs, $k^2 W(\boldsymbol{q})$, attains a minimum on a submanifold $\mathcal{M}$. We define an orthonormal system of coordinates on a compact neighbourhood of $\mathcal{M}$, enclosed by the red solid line, by introducing soft degrees of freedom $\boldsymbol{r}$ on $\mathcal{M}$ and hard degrees of freedom $\boldsymbol{s}$, which are normal to $\mathcal{M}$ and correspond to the confining springs. The hard degrees of freedom will be integrated out to find the limiting distribution for stiff springs.

by the Boltzmann distribution

$$p_k = N(k) \exp\left\{-\beta\left[U(\boldsymbol{q}) + k^2 W(\boldsymbol{q})\right]\right\}, \tag{3}$$

with an appropriate normalising function $N(k)$. Suppose now that the spring potential is non-negative, $W(\boldsymbol{q}) \geq 0$, and attains a minimum for a configuration lying on a smooth submanifold $\mathcal{M}$ of dimension $M < N$, so that $W(\boldsymbol{q}) = 0 \iff \boldsymbol{q} \in \mathcal{M}$. To determine the weak limit of p_k, we consider integrals I_k for a compactly supported smooth test function $\phi(\boldsymbol{q})$ given by

$$I_k = \int p_k(\boldsymbol{q}) \phi(\boldsymbol{q}) \mathrm{d}\boldsymbol{q}. \tag{4}$$

Since $\mathcal{M}$ is smooth, we can use the tubular embedding theorem to define new orthonormal coordinates $\boldsymbol{q} \mapsto (\boldsymbol{r}, \boldsymbol{s}) = (r_1, r_2, \dots, r_M, s_1, s_2, \dots)$ in the vicinity of $\mathcal{M}$, where $\boldsymbol{q} \in \mathcal{M} \iff \boldsymbol{s} = 0$. In other words, coordinates r_i represent the *soft* degrees of freedom and s_i represent the *hard* directions. Since the new coordinates are orthonormal we know that the determinant of the transformation $(q_i) \mapsto (r_i, s_j)$ equals 1. Additionally, I_k is finite by compactness of the support of ϕ and continuity of p_k, so by Fubini's theorem we can replace the volume integral by an iterated one. Then, I_k is given by

$$I_k = \iint p_k(\boldsymbol{r}, \boldsymbol{s}) \phi(\boldsymbol{r}, \boldsymbol{s}) \mathrm{d}\boldsymbol{s} \mathrm{d}\boldsymbol{r} \tag{5}$$

We proceed by expanding p_k in a Taylor series with respect to $\boldsymbol{s} = (s_1, s_2, \dots)$ to second order as $W(\boldsymbol{r}, \boldsymbol{s}) = \boldsymbol{s}^T \mathbf{H}(\boldsymbol{r}) \boldsymbol{s}$, where $\mathbf{H}(\boldsymbol{r})$ is the Hessian of W, evaluated at $\boldsymbol{s} = 0$. Note that the linear term vanishes, since at $\boldsymbol{s} = 0$ the potential W attains a minimum. We can thus write

$$I_k = \iint N(k) \exp\left\{-\beta U(\boldsymbol{r}, \boldsymbol{s}) - \beta k^2 \left[\boldsymbol{s}^T \mathbf{H}(\boldsymbol{r}) \boldsymbol{s} + o(||\boldsymbol{s}||^2)\right]\right\} \phi(\boldsymbol{r}, \boldsymbol{s}) \mathrm{d}\boldsymbol{s} \mathrm{d}\boldsymbol{r} \tag{6}$$

We now make a substitution $k\boldsymbol{s} = \boldsymbol{t}$ to arrive at

$$I_k = \iint \frac{N(k)}{k} \exp\left\{-\beta U(\boldsymbol{r}, k^{-1}\boldsymbol{t}) - \beta\left[\boldsymbol{t}^T \mathbf{H}(\boldsymbol{r}) \boldsymbol{t} + ||\boldsymbol{t}||^2 \frac{o(k^{-2}||\boldsymbol{t}||^2)}{k^{-2}||\boldsymbol{t}||^2}\right]\right\} \phi(\boldsymbol{r}, k^{-1}\boldsymbol{t}) \mathrm{d}\boldsymbol{t} \mathrm{d}\boldsymbol{r} \tag{7}$$

Whenever W increases sufficiently fast with the distance from the constraining manifold, we may take a limit inside the integral by the dominated convergence theorem to arrive at the limiting value $I = \lim_{k\to\infty} I_k$ given by

$$I = \iint \left(\lim_{k\to\infty} \frac{N(k)}{k}\right) \exp\left[-\beta U(\boldsymbol{r}, \boldsymbol{0}) - \beta\, \boldsymbol{t}^T \mathbf{H}(\boldsymbol{r}) \boldsymbol{t}\right] \phi(\boldsymbol{r}, \boldsymbol{0}) \mathrm{d}\boldsymbol{t} \mathrm{d}\boldsymbol{r} \tag{8}$$

and, since $\mathbf{H}$ is full rank, perform the integral over $\boldsymbol{t}$ to arrive at

$$I = L \int |\mathbf{H}(\boldsymbol{r})|^{-1/2} \exp\left[-\beta U(\boldsymbol{r}, \boldsymbol{0})\right] \phi(\boldsymbol{r}, \boldsymbol{0}) \mathrm{d}\boldsymbol{r} \tag{9}$$

with the constant

$$L = \lim_{k\to\infty} \frac{N(k)\sqrt{(2\pi)^{N-M}}}{k}, \tag{10}$$

being independent of the soft coordinates.

In a particular physical setting, finding the orthonormal coordinates $(\boldsymbol{r}, \boldsymbol{s})$ may be prohibitively difficult, but we can relax this strict requirement by considering another parametrisation $\boldsymbol{w}$ such that $\boldsymbol{q} = \boldsymbol{\zeta}^*(\boldsymbol{w})$, and that still separates *soft* from *hard* degrees of freedom by

$$\boldsymbol{q} \in \mathcal{M} \iff \forall i > M : w_i = \xi_i, \tag{11}$$

for a set of constants ξ_i. This parametrisation is not necessarily orthonormal. For convenience, we define a map $\boldsymbol{\zeta} : \mathbb{R}^M \to \mathbb{R}^N$ such that $\boldsymbol{\zeta}(w_1, w_2, \ldots, w_M) = \boldsymbol{\zeta}^*(w_1, w_2, \ldots, w_M, \xi_{M+1}, \xi_{M+2}, \ldots, \xi_N)$ which takes values of the *soft* coordinates and returns points on $\mathcal{M}$. We now change variables under the integral from $\boldsymbol{r}$ to $\boldsymbol{w}$ to arrive at

$$I = L \int \frac{\left|\mathbf{J}^T\mathbf{J}\right|^{1/2}}{|\mathbf{H}(\boldsymbol{\zeta})|^{1/2}} \exp\left[-\beta U(\boldsymbol{\zeta})\right] \phi(\boldsymbol{\zeta}) \mathrm{d}w_1 \ldots \mathrm{d}w_M, \tag{12}$$

where $\mathbf{J}$ is the Jacobian of $\boldsymbol{\zeta}(\boldsymbol{w})$. Note that $\boldsymbol{\zeta} : \mathbb{R}^M \to \mathbb{R}^N$, so the Jacobian is not a square matrix. We can express the above using the Dirac-δ distributions. Taking

$$\delta(\boldsymbol{w}' - \boldsymbol{\xi}') = \prod_{i=M+1}^{N} \delta(w_i - \xi_i), \tag{13}$$

the limiting integral is recast as

$$I = L \int \frac{\left|\mathbf{J}^T\mathbf{J}\right|^{1/2}}{|\mathbf{H}|^{1/2}} \exp\left(-\beta U\right) \phi(\boldsymbol{\zeta}^*(\boldsymbol{w})) \delta(\boldsymbol{w}' - \boldsymbol{\xi}') \, \mathrm{d}\boldsymbol{w} \tag{14}$$

for an appropriate constant L, and since ϕ is arbitrary we arrive at $p_\infty = \lim_{k\to\infty} p_k$ such that

$$\mathrm{d}p_\infty = L \frac{|\mathbf{J}^T\mathbf{J}|^{1/2}}{|\mathbf{H}|^{1/2}} \exp\left[-\beta U(\boldsymbol{\zeta}^*(\boldsymbol{w}))\right] \delta(\boldsymbol{w}' - \boldsymbol{\xi}') \, \mathrm{d}\boldsymbol{w}, \tag{15}$$

in the weak sense. We have thus arrived at the central result of this manuscript.

The limiting distribution comprises two important factors. First, the metric term $|\mathbf{J}^T\mathbf{J}|$ describing the projection of the surface element of $\mathcal{M}$ onto $\mathbb{R}^M$ in the parametrisation $\boldsymbol{\zeta}(\boldsymbol{w})$; This term was correctly computed by Fixman [5]. Second, the *shape* term $|\mathbf{H}|$, also called the zero-point energy in the quantum-mechanical setting, which was often missing in the derivations.

A. Computation of the shape term

A careful choice of parametrisation allows for computation of the metric term with relative ease. However, the Hessian term $|\mathbf{H}|$ may be more difficult to compute. In this Section, we present a feasible approach to its computation.

Let $\mathbf{B}$ be the full Hessian of W at a point $\boldsymbol{q}$,

$$\mathbf{B}_{ij} = \frac{\partial^2}{\partial q_i \partial q_j} W(\boldsymbol{q}). \tag{16}$$

Since the eigenvectors of $\mathbf{B}$ are orthogonal and zero eigenvectors lie inside the tangent space $T\mathcal{M}$, we can compute $|\mathbf{H}|$ from the product of the non-zero eigenvalues of $\mathbf{B}$. We can use the knowledge of the tangent space of $\mathcal{M}$ to help us compute the eigenvalues of $\mathbf{B}$. If our confining function is of the form $W(\boldsymbol{q}) = \sum_i P_i^2$, we can find vectors that lie in the normal space $\mathcal{M}^\perp$ [15, proposition 8.15] by computing

$$\mathbf{A}_{ji} = \frac{\partial}{\partial q_j} P_i \tag{17}$$

The vectors $\{\mathbf{A}_{j1}, \ldots, \mathbf{A}_{j,N-M}\}$ are not pairwise orthogonal but are orthogonal to $\mathcal{M}$, and thus can form a basis of the normal space at $\boldsymbol{q}$. We can write the eigenvalue problem using an arbitrary vector b_i as

$$\lambda \mathbf{A}_{ki} b_i = \mathbf{B}_{kj} \mathbf{A}_{ji} b_i. \tag{18}$$

This is a system of N equations with $N - M$ unknowns. We can eliminate redundant equations by contracting each side with $\mathbf{A}_{ij}$ to finally get

$$\lambda(\mathbf{A}^T\mathbf{A})\boldsymbol{b} = (\mathbf{A}^T\mathbf{B}\mathbf{A})\boldsymbol{b} \tag{19}$$

and thus the product of eigenvalues in this problem is simply

$$|\mathbf{H}| = \frac{|\mathbf{A}^T\mathbf{B}\mathbf{A}|}{|\mathbf{A}^T\mathbf{A}|}. \tag{20}$$

We see that there are two terms in this expression. The denominator $|\mathbf{A}^T\mathbf{A}|$ measures the angles between the gradients of the constraining functions, while the numerator $|\mathbf{A}^T\mathbf{B}\mathbf{A}|$ measures the shape and strength of the confining field in these directions.

IV. CALCULATION FOR THE TRIMER PROBLEM

In the aforementioned case of a trimer, depicted in Fig. 2(a), we have 3 beads located at $\boldsymbol{r}_a, \boldsymbol{r}_b$ and $\boldsymbol{r}_c$ with bond extensions P_a, P_c given by

$$P_a = |\boldsymbol{r}_a - \boldsymbol{r}_b| - l_0 \tag{21}$$
$$P_c = |\boldsymbol{r}_c - \boldsymbol{r}_b| - l_0, \tag{22}$$

l_0 being the equilibrium bond length, and the confining potential given by

$$W(\boldsymbol{q}) = P_a^2 + P_c^2. \tag{23}$$

We now pick the parametrisation $\boldsymbol{q} = \boldsymbol{\zeta}^*(\boldsymbol{w})$ with $\boldsymbol{w} = [x, y, z, \alpha, \beta, \gamma, \psi, l_a, l_c]^T$. Cartesian coordinates (x, y, z) describe the position of the central bead (b), (α, β, γ) are the Euler angles that encode the direction of bead a seen from b, and ψ is the bond angle between the edges ab and bc. The positions of beads are thus given by

$$\boldsymbol{r}_a = [x, y, z]^T + l_a \mathbf{E}_{\alpha\beta\gamma}[1, 0, 0]^T, \tag{24}$$
$$\boldsymbol{r}_b = [x, y, z]^T, \tag{25}$$
$$\boldsymbol{r}_c = [x, y, z]^T + l_c \mathbf{E}_{\alpha\beta\gamma}[\cos\psi, \sin\psi, 0]^T. \tag{26}$$

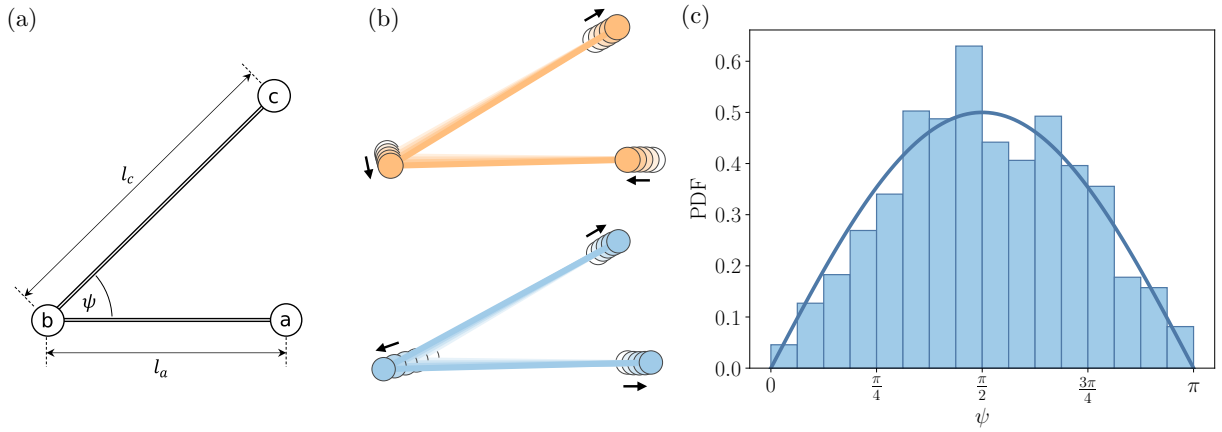


FIG. 2. (a) Parametrisation used for the trimer uses the position of the central particle (b), bond angle ψ and bond lengths l_a and l_b, with identical equilibrium spring lengths l_0. Rigid-body rotation around the particle b is omitted for clarity. (b) Motion along the symmetric (top) and antisymmetric eigenvectors of the hessian $\mathbf{H}$ of the harmonic confinement function. Note that the central bead must move for the overall motion to be orthogonal to the constraining manifold $\mathcal{M}$. (c) Equilibrium probability density function (PDF) for the bond angle ψ from the Langevin numerical simulation. Solid line is the theoretical prediction of Eq. (32).

We defined the rotation matrix $\mathbf{E}_{\alpha\beta\gamma}$ in terms of the Euler angles in the Appendix. In addition, we note that $P_a = l_a - l_0$ and $P_c = l_c - l_0$. We compute the metric term as

$$|\mathbf{J}^T\mathbf{J}| = \frac{1}{2} l_0^8 \sin^2\beta \sin^2\psi\,(7-\cos 2\psi)\,, \tag{27}$$

from the matrix $\mathbf{J}$ listed in the Appendix. Next, we compute the shape term using

$$\mathbf{A}^T\mathbf{A} = \begin{bmatrix} 2 & \cos\psi \\ \cos\psi & 2 \end{bmatrix}, \tag{28}$$

$$\mathbf{A}^T\mathbf{B}\mathbf{A} = \begin{bmatrix} 9+\cos 2\psi & 8\cos\psi \\ 8\cos\psi & 9+\cos 2\psi \end{bmatrix}, \tag{29}$$

thus finding

$$|\mathbf{H}| = \frac{|\mathbf{A}^T\mathbf{B}\mathbf{A}|}{|\mathbf{A}^T\mathbf{A}|} = 2(7-\cos 2\psi). \tag{30}$$

We therefore have

$$\frac{|\mathbf{J}^T\mathbf{J}|^{1/2}}{|\mathbf{H}|^{1/2}} = \frac{l_0^4}{2}\sin\beta\sin\psi, \tag{31}$$

and for the final distribution, we find

$$\mathrm{d}p_\infty = N\sin\beta\sin\psi\;\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z\,\mathrm{d}\alpha\,\mathrm{d}\beta\,\mathrm{d}\gamma\,\mathrm{d}\psi, \tag{32}$$

where N is a normalisation constant. Therefore, the marginal density of the bond angle is proportional to $\sin\psi$ as expected. However, this is simply not the case of a uniform distribution—that distribution corresponds to $|\mathbf{H}| = \text{const}$. This can be examined in greater detail by looking at the eigenvectors of $\mathbf{H}$, visualised in Fig. 2(b). There are two eigenvectors which correspond to symmetric and antisymmetric motion of the terminal beads. Counterintuitively, the vectors in $\mathcal{M}^\perp$ change the orientation of at least one bond when the bond length changes.

The simplification lies in the fortuitous cancellation of the $(7-\cos 2\psi)$ term, rather than its absence. This is best highlighted by an example where no such cancellation occurs, which we present in the next Section.

We additionally corroborate this result with a Brownian Dynamics simulation in Python using the package `pychastic` [16]. Choosing k_BT as the energy scale, l_0 as the distance scale and l_0^2/D, as the timescale, with D being the diffusion coefficient, the corresponding stochastic BD equation takes the form

$$\mathrm{d}\boldsymbol{r}_i = -\boldsymbol{\nabla}(U + k^2 W)\mathrm{d}t + \sqrt{2}\,\mathrm{d}\boldsymbol{\mathcal{B}}_i, \tag{33}$$

with $\boldsymbol{\mathcal{B}}_i$ denoting the standard Wiener process. We have performed $n_{\mathrm{traj}} = 4000$ simulations with $k = 35$ (in units of k_BT/l_0) for a time up to $t_{max} = 10$ with time step $\delta t = 10^{-5}$. The simulation results in Fig. 2(c) complement the prediction of Eq. (32). We use the same code for a less studied example of a tetramer in the following.

V. CYCLIC TETRAMER

A more complex, and yet fully tractable analytically, is the case of a cyclic tetramer, a molecule of 4 beads joined into a quadrilateral. For the sake of brevity, we discuss the dynamics in two dimensions. The calculation in 3D is completely analogous but involves much longer intermediate expressions due to the additional rotational degrees of freedom and a dihedral angle. We pick the parametrisation where the 4 beads are located at $\boldsymbol{r}_a, \boldsymbol{r}_b, \boldsymbol{r}_c$ and $\boldsymbol{r}_d$,

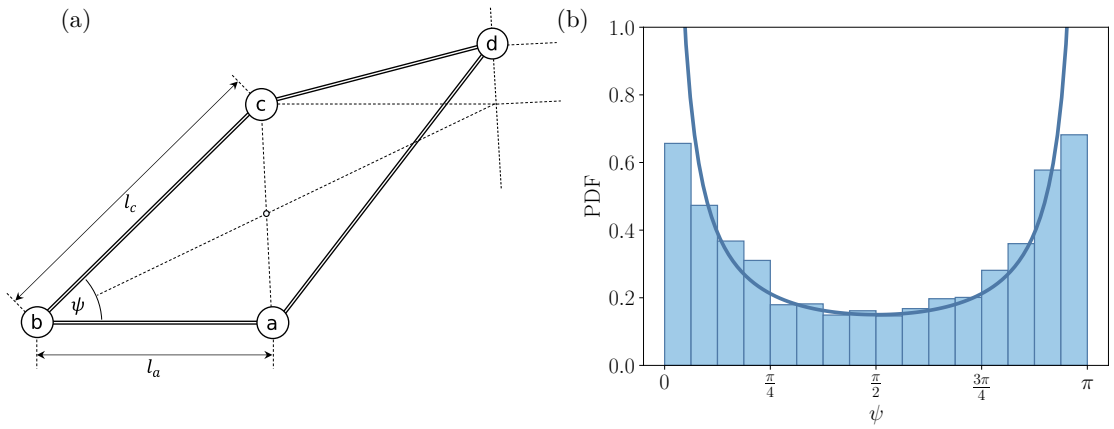


FIG. 3. (a) The shape of a cyclic tetramer is specified with the bead b taken as reference point. The directions of sections ab and bc span a rhombus, which is used to define the position of bead d. (b) Equilibrium probability density function (PDF) for the bond angle ψ from the numerical simulation. Solid line is the theoretical prediction of Eq. (47).

with their positions given by

$$\begin{aligned} \boldsymbol{r}_a &= [x,y]^T + l_a \mathbf{E}_\alpha [1,0]^T, && (34)\\ \boldsymbol{r}_b &= [x,y]^T, && (35)\\ \boldsymbol{r}_c &= [x,y]^T + l_c \mathbf{E}_\alpha [\cos\psi, \sin\psi]^T, && (36)\\ \boldsymbol{r}_d &= [x,y]^T + \mathbf{E}_\alpha \left((1+\delta)\, \boldsymbol{v}_\delta + \epsilon\, \boldsymbol{v}_\epsilon \right), && (37)\\ & && (38) \end{aligned}$$

where $\mathbf{E}_\alpha$ is a rotation matrix defined in the Appendix, and the vectors $\boldsymbol{v}_{\delta,\epsilon}$, written as

$$\begin{aligned} \boldsymbol{v}_\delta &= \frac{[l_a + l_c \cos\psi, l_c \sin\psi]^T}{2}, && (39)\\ \boldsymbol{v}_\epsilon &= \frac{[l_a - l_c \cos\psi, -\sin\psi]^T}{2}, && (40) \end{aligned}$$

point along the diagonals of the rhombus spanned by the edges ab and bc. Finally, ϵ and δ encode the position of the bead located at $\boldsymbol{r}_d$. The four constraining surfaces are given by

$$\begin{aligned} P_a &= |\boldsymbol{r}_a - \boldsymbol{r}_b| - l_0 && (41)\\ P_b &= |\boldsymbol{r}_b - \boldsymbol{r}_c| - l_0 && (42)\\ P_c &= |\boldsymbol{r}_c - \boldsymbol{r}_d| - l_0 && (43)\\ P_d &= |\boldsymbol{r}_d - \boldsymbol{r}_a| - l_0 && (44) \end{aligned}$$

and $\mathcal{M}$ is given by $l_a = l_c = l_0, \delta = 1, \epsilon = 0$.

The metric term is simply constant with $|\mathbf{J}^T\mathbf{J}| = 16 l_0^4$, while the shape term is non-trivial. Denoting $\cos\psi = c_\psi$ for brevity, the shape factors are given by

$$\mathbf{A}^T\mathbf{A} = \begin{bmatrix} 2 & c_\psi & 0 & -c_\psi \\ c_\psi & 2 & -c_\psi & 0 \\ 0 & -c_\psi & 2 & c_\psi \\ -c_\psi & 0 & c_\psi & 2 \end{bmatrix}, \quad (45)$$

$$\mathbf{A}^T\mathbf{B}\mathbf{A} = 4 \begin{bmatrix} c_\psi^2+2 & 2c_\psi & -c_\psi^2 & -2c_\psi \\ 2c_\psi & c_\psi^2+2 & -2c_\psi & -c_\psi^2 \\ -c_\psi^2 & -2c_\psi & c_\psi^2+2 & 2c_\psi \\ -2c_\psi & -c_\psi^2 & 2c_\psi & c_\psi^2+2 \end{bmatrix}, \quad (46)$$

yielding $|\mathbf{H}| = 256 \sin^2\psi$. As a result, we get that

$$\mathrm{d}p_\infty \propto \frac{1}{\sin\psi} \mathrm{d}\psi, \quad (47)$$

which is not non-normalizable near 0 and π. Physically, as we approach the stiff spring limit, the tetramer spends more and more time in the folded state, and unless some repulsive potential is added, the molecule in the stiff spring limit looks predominantly like a trimer, alternating between beads (a, c) and (b, d) coinciding in space.

This surprising result is also seen in the numerical simulation. In this case, a strong repulsive potential was added to prevent bead overlaps for distances less than $0.05 l_0$. The final bond angle distribution shown in Figure 3(b) coincides well with $0.15/\sin\psi$.

VI. CONCLUSIONS

In this work, we have provided a procedure by which one can find the equilibrium probability probability distribution function of the configurations of a bead-spring model of a polymer in the overdamped regime in the limit of infinite stiffness of the bonds. In particular, we have demonstrated that the shape of the confining spring potential persists in the limiting distribution. By rephrasing the problem as a mathematically rigorous limit, we have shown that the classical expression given by Fixman [5], later reproduced in numerous works and books, for example in Ref. [3], is missing a critical term describing the shape of the confining potential, $|\mathbf{H}|$.

This shows that the reasoning presented by Frenkel & Smit [3] cannot be universally applied. First, the dynamics and the distribution of a polymer connected with (stiff or soft) springs is independent of the mass of the monomers in the overdamped regime, and thus the limiting distribution cannot depend on the mass of the

monomers. Second, putting aside the question of existence of the rigid rod distribution (or the lack thereof), the supposition that the stiff spring limit should always be the same, regardless of the details of the confining potential is clearly not true, and harmonic potentials will lead to a different outcome than springs that realise uniform confinement around the constraining manifold, as previously remarked, e.g., by van Kampen & Lodder [12], although without a general mathematical formulation. The difference can be arbitrarily large, as shown in the tetramer example. Frenkel & Smit remark in their book that for the bond length constraints of the type most often used in Molecular Dynamics simulations, the effect of hard constraints on the distribution functions seems to be relatively small [3, Chapter 15.1.1]. We believe this statement to be misleading. In the discussed case of the trimer molecule, the effect might indeed be small, but we have shown that for the tetramer system, which is still a relatively simple setup, the effect is pronounced. To provide a practical insight, we propose a straightforward technique for the analysis of the limiting distributions once the confining potential is specified in detail.

ACKNOWLEDGMENTS

The Authors would like to thank Bartłomiej Lewandowski for his insightful comments regarding convergence of distributions. The work of ML and RW was supported by the National Science Centre of Poland (FundRef DOI: http://dx.doi.org/10.13039/501100004281) grant Sonata to ML no. 2018/31/D/ST3/02408.

Appendix A: Auxiliary matrices used in calculations

Using the notation $s_\alpha = \sin\alpha$, $c_\alpha = \cos\alpha$ we can compactly express the matrices used. The 2D rotation matrix is given by

$$\mathbf{E}_\alpha = \begin{bmatrix} c_\alpha & -s_\alpha \\ s_\alpha & c_\alpha \end{bmatrix}. \tag{A1}$$

The 3D rotation matrix is given by

$$\mathbf{E}_{\alpha\beta\gamma} = \begin{bmatrix} c_\alpha c_\beta c_\gamma - s_\alpha s_\gamma & -c_\alpha c_\beta s_\gamma - c_\gamma s_\alpha & c_\alpha s_\beta \\ c_\beta c_\gamma s_\alpha + c_\alpha s_\gamma & c_\alpha c_\gamma - c_\beta s_\alpha s_\gamma & s_\alpha s_\beta \\ -c_\gamma s_\beta & s_\beta s_\gamma & c_\beta \end{bmatrix}. \tag{A2}$$

For our paremetrisation of the trimer, the Jacobian $\mathbf{J}$ can be expressed as a block matrix with 3×4 blocks $\mathbf{Q}_i$

$$\mathbf{J} = \begin{bmatrix} \mathbf{1} & \mathbf{Q}_1 \\ \mathbf{1} & \mathbf{Q}_2 \\ \mathbf{1} & \mathbf{Q}_3 \end{bmatrix}. \tag{A3}$$

Writing $\gamma + \psi = \theta$ for brevity, the blocks $\mathbf{Q}_i$ are given by

$$\mathbf{Q}_1^T = l_0 \begin{bmatrix} -s_\alpha c_\beta c_\gamma - c_\alpha s_\gamma & c_\alpha c_\beta c_\gamma - s_\alpha s_\gamma & 0 \\ -c_\alpha s_\beta c_\gamma & -s_\alpha s_\beta c_\gamma & -c_\beta c_\gamma \\ -s_\alpha c_\gamma - c_\alpha c_\beta s_\gamma & c_\alpha c_\gamma - s_\alpha c_\beta s_\gamma & s_\beta s_\gamma \\ 0 & 0 & 0 \end{bmatrix} \tag{A4}$$

$$\mathbf{Q}_2^T = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{A5}$$

$$\mathbf{Q}_3^T = l_0 \begin{bmatrix} -s_\alpha c_\beta c_\theta - c_\alpha s_\theta & c_\alpha c_\beta c_\theta - s_\alpha s_\theta & 0 \\ -c_\alpha s_\beta c_\theta & -s_\alpha s_\beta c_\theta & -c_\beta c_\theta \\ -s_\alpha c_\theta - c_\alpha c_\beta s_\theta & c_\alpha c_\theta - s_\alpha c_\beta s_\theta & s_\beta s_\theta \\ -s_\alpha c_\theta - c_\alpha c_\beta s_\theta & c_\alpha c_\theta - s_\alpha c_\beta s_\theta & s_\beta s_\theta \end{bmatrix} \tag{A6}$$

[1] H. A. Kramers, The behavior of macromolecules in inhomogeneous flow, The Journal of Chemical Physics **14**, 415 (1946).
[2] P. Flory, *Principles of Polymer Chemistry*, Baker lectures 1948 (Cornell University Press, 1953).
[3] D. Frenkel and B. Smit, *Understanding molecular simulation: from algorithms to applications* (Elsevier, 2023).
[4] J.-P. Ryckaert, G. Ciccotti, and H. J. Berendsen, Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, Journal of Computational Physics **23**, 327 (1977).
[5] M. Fixman, Classical statistical mechanics of constraints: a theorem and application to polymers., Proceedings of the National Academy of Sciences of the United States of America **71**, 3050 (1974).
[6] N. Gō and H. A. Scheraga, Analysis of the Contribution of Internal Vibrations to the Statistical Weights of Equilibrium Conformations of Macromolecules, The Journal of Chemical Physics **51**, 4751 (1969).
[7] M. R. Pear and J. H. Weiner, Brownian dynamics study of a polymer chain of linked rigid bodies, The Journal of Chemical Physics **71**, 212 (1979).
[8] E. Hinch, Brownian motion with stiff bonds and rigid constraints, Journal of Fluid Mechanics **271**, 219 (1994).
[9] W. K. den Otter and W. J. Briels, The calculation of free-energy differences by constrained molecular-dynamics simulations, The Journal of Chemical Physics **109**, 4139 (1998).
[10] T. W. Liu, Flexible polymer chain dynamics and rheolog-

ical properties in steady flows, The Journal of Chemical Physics **90**, 5826 (1989).
[11] E. Helfand, Flexible vs rigid constraints in statistical mechanics, The Journal of Chemical Physics **71**, 5000 (1979).
[12] N. van Kampen and J. Lodder, Constraints, American Journal of Physics **52**, 419 (1984).
[13] A. Vitalis and R. V. Pappu, Methods for monte carlo simulations of biomacromolecules, Annual reports in computational chemistry **5**, 49 (2009).
[14] M. P. Allen, Introduction to molecular dynamics simulation, in *Computational soft matter: from synthetic polymers to proteins*, Vol. 23, edited by N. Attig, K. Binder, H. Grubmüller, and K. Kremer (NIC Series Jülich, Germany, 2004) pp. 1–28.
[15] J. M. Lee, *Introduction to Smooth Manifolds* (Springer New York, 2003).
[16] R. Waszkiewicz, M. Bartczak, K. Kolasa, and M. Lisicki, Pychastic: Precise brownian dynamics using Taylor-Itō integrators in Python, SciPost Physics Codebases , 011 (2023).

Conclusions

The objectives of this Thesis were twofold: (i) to identify the theoretical approaches capable of modelling experimentally relevant macromolecules, and (ii) to develop a modular system capable of accommodating various coarse-graining approximations and experimental techniques. The results demonstrate the feasibility of such a unified approach and its ability to predict macromolecular diffusion coefficients with an accuracy that enables comparison with experiments. Throughout the preparation of this PhD Thesis, we examined numerous macromolecular scenarios, encompassing a wide range of elasticity, from near-rigid DNA fragments and minicircles, through elastic loops subject to an external force and molecules of intermediate elasticity in heat bath, to highly elastic linkers in bead models of Intrinsically Disordered Proteins (IDPs), towards in a general treatment of molecules that combines both extremes in the study of the stiff springs limit. The study yielded several results, the most important of which are the following:

1. Scaling-based analysis of the approach of rod-like molecules to a nanopore, incorporating hydrodynamic interactions with the wall and Brownian reorientations [1].

2. Linear stability analysis and numerical investigation of the buckling instability induced by hydrodynamic drag during the sedimentation of flexible loops, modelled with Euler-Bernoulli elasticity and resistive force theory [2].

3. The development of a coarse-grained modeling approach to computing the conformations of supercoiled DNA minicircles by minimization of their elastic energy under for different values of linking number [3].

4. Computation of hydrodynamic radii, diffusion coefficients and sedimentation coefficients for the computed models of DNA minicricle topoisomers [3].

5. Development and implementation of a Python package integrating stochastic differential equations of Brownian Dynamics [4].

6. Development of the Globule Linker Model for expedited sampling of IDP conformations, combined with the Minimum Dissipation Approximation for computing their hydrodynamic size [5].

7. Analytic determination of the limit of the equilibirum distributions for a general elastic macromolecule with stiff and soft degrees of freedom [6].

The key practical achievements of this Thesis can be summarised as follows:

- Successful prediction of hydrodynamic radii for DNA minicircles at different degrees of supercoiling and given values of linking number.

- Successful prediction of hydrodynamic radii for 38 different intrinsically disordered proteins—the largest benchmark set to date.

- Provision of user-friendly, well-documented, and publicly available Python implementations for all proposed methods (without compromising the prediction speed).

Consequently, the presented methods can serve as a numerically feasible null-hypothesis model in future investigations by various experimental groups, with significant deviations from its predictions indicating potential new physical phenomena. Moreover, we anticipate that the soft matter physics community will leverage the software developed in this Thesis in various ways, either as-is for predicting diffusion coefficients of similar molecules or by extending its capabilities through its modular design and utilizing its components independently.

Other research activity

- Radost Waszkiewicz, John Burnham Shaw, Maciej Lisicki, and Piotr Szymczak. "Goldilocks fluctuations: dynamic constraints on loop formation in scale-free transport networks". In: *Physical Review Letters* (2024). DOI: 10.1103/PhysRevLett.132.137401

  ABSTRACT: Adaptive transport networks are known to contain loops when subject to hydrodynamic fluctuations. However, fluctuations are no guarantee that a loop will form, as shown by loop-free networks driven by oscillating flows. We provide a complete stability analysis of the dynamical behaviour of any loop formed by fluctuating flows. We find a threshold for loop stability that involves an interplay of geometric constraints and hydrodynamic forcing mapped to constant and fluctuating components. Loops require fluctuation in the relative size of the flux between nodes, not just a temporal variation in the flux at a given node. Hence, there is both a minimum and a maximum amount of fluctuation relative to the constant-flux component where loops are supported.

- Jan Turczynowicz, Radost Waszkiewicz, and Łukasz Gładczuk. "Preventing sinking of a disk by leveraging the Parametric Hydraulic Jump phenomenon". In: *arXiv preprint* (2023). DOI: 10.48550/arXiv.2312.13099

  ABSTRACT: Although it is commonly expected that a metal disk placed on the surface of water will sink, our investigation has revealed a surprising phenomenon: a vertical jet directed onto the disk from above can allow it to remain afloat. This result defies intuition, as one would assume that the force of the jet's impact would cause the disk to sink.We have discovered that this phenomenon occurs as a result of water displacement from the top of the disk caused by the impacting jet, operating through a mechanism similar to a classical hydraulic jump. This displacement generates a difference in hydrostatic pressures, resulting in an upward buoyancy force capable of balancing the force of gravity. In contrast to the classical case, here the jump radius is fixed by the geometric parameters of a disk, a phenomenon we refer to as the parametric hydraulic jump. To further explore this effect, we have presented a theoretical model based on scaling laws, which provides the conditions required for the disk to float. The dimensionless scaling constant's value was determined through an independent experiment. Finally, we conducted experiments on the disk's floating and sinking, which showed a good match with the proposed theory.

- Radost Waszkiewicz and Honorata Bogusz. "The Impact of Parenthood on Labour Market Outcomes of Women and Men in Poland". In: *arXiv preprint* (2024). DOI: 10.48550/arXiv.2306.12924

  ABSTRACT: We examine the gender gap in income in Poland in relation to parenthood status, employing the placebo event history method adapted to low-resolution data (Polish Generations and Gender Survey). Our analysis reveals anticipatory behavior in both women and men who expect to become parents. We observe a decrease of approximately 20 percent in mothers' income post-birth. In contrast, the income of fathers surpasses that of non-fathers both pre- and post-birth, suggesting that the fatherhood child premium may be primarily driven by selection. We note an increase (decrease) in hours worked for fathers (mothers). Finally, we compare the gender gaps in income and wages between women and men in the sample with those in a counterfactual scenario where the entire population is childless. Our findings indicate no statistically significant gender gaps in the counterfactual scenario, leading us to conclude that parenthood drives the gender gaps in income and wages in Poland.

Bibliography

[1] Radost Waszkiewicz and Maciej Lisicki. "Hydrodynamic effects in the capture of rod-like molecules by a nanopore". In: *Journal of Physics: Condensed Matter* **33**. (2021), p. 104005. DOI: 10.1088/1361-648X/abd11b.

[2] Radost Waszkiewicz, Piotr Szymczak, and Maciej Lisicki. "Stability of sedimenting flexible loops". In: *Journal of Fluid Mechanics* **919** (2021), A14. DOI: 10.1017/jfm.2021.350.

[3] Radost Waszkiewicz, Maduni Ranasinghe, Jonathan M Fogg, Daniel J Catanese Jr, Maria L Ekiel-Jeżewska, Maciej Lisicki, Borries Demeler, Lynn Zechiedrich, and Piotr Szymczak. "DNA supercoiling-induced shapes alter minicircle hydrodynamic properties". In: *Nucleic Acids Research* **51**. (2023), pp. 4027–4042. DOI: 10.1093/nar/gkad183.

[4] Radost Waszkiewicz, Maciej Bartczak, Kamil Kolasa, and Maciej Lisicki. "Pychastic: Precise Brownian dynamics using Taylor-Itō integrators in Python". In: *SciPost Physiscs Codebases* **11** (2023). DOI: 10.21468/SciPostPhysCodeb.11.

[5] Radost Waszkiewicz, Agnieszka Michaś, Michał K. Białobrzewski, Barbara Klepka, Maja Cieplak-Rotowska, Zuzanna Staszałek, Bogdan Cichocki, Maciej Lisicki, Piotr Szymczak, and Anna Niedźwiecka. "Minimum dissipation approximation: A fast algorithm for the prediction of diffusive properties of intrinsically disordered proteins". In: *Journal of Physical Chemisty Letters (in review)* (2024). DOI: 10.1101/2024.02.05.578612.

[6] Radost Waszkiewicz and Maciej Lisicki. "The trimer paradox: the effect of stiff constraints on equilibrium distributions in overdamped dynamics". In: *Physical Review E (in review)* (2024). DOI: 10.48550/arXiv.2403.13185.

[7] Osborne Reynolds. "An experimental investigation of the circumstances which determine whether the motion of water shall be direct or sinuous, and of the law of resistance in parallel channels". In: *Philosophical Transactions of the Royal Society of London* **174** (1883), pp. 935–982. DOI: 10.1098/rstl.1883.0029.

[8] Gerhard Nägele. "Colloidal hydrodynamics". In: *Physics of complex colloids*. IOS Press, 2013, pp. 507–601. DOI: 10.3254/978-1-61499-278-3-507.

[9] N.G. van Kampen. *Stochastic Processes in Physics and Chemistry*. North-Holland Personal Library. Elsevier Science, 2011. ISBN: 9780080475363.

[10] Da Wei, Parviz G Dehnavi, Marie-Eve Aubin-Tam, and Daniel Tam. "Measurements of the unsteady flow field around beating cilia". In: *Journal of Fluid Mechanics* **915** (2021), A70. DOI: 10.1017/jfm.2021.149.

[11] S. Kim and S.J. Karrila. *Microhydrodynamics: Principles and Selected Applications*. Butterworth-Heinemann series in chemical engineering. Elsevier Science, 2013.

[12] Simon L. Goren and Michael E. O'Neill. "Asymmetric creeping motion of an open torus". In: *Journal of Fluid Mechanics* **101**. (1980), pp. 97–110. DOI: 10.1017/S0022112080001553.

[13] S. R. Majumdar and M. E. O'Neill. "On axisymmetric stokes flow past a torus". In: *Zeitschrift für angewandte Mathematik und Physik ZAMP* **28**. (1977), pp. 541–550. DOI: 10.1007/BF01601334.

[14] Donald L Ermak and J Andrew McCammon. "Brownian dynamics with hydrodynamic interactions". In: *The Journal of Chemical Physics* **69**. (1978), pp. 1352–1360. DOI: 10.1063/1.436761.

[15] Hilding Faxén. "Der Widerstand gegen die Bewegung einer starren Kugel in einer zähen Flüssigkeit, die zwischen zwei parallelen ebenen Wänden eingeschlossen ist". In: *Annalen der Physik* **373**. (1922), pp. 89–119. DOI: 10.1002/andp.19223731003.

[16] J. K. G. Hermann and G. Gompper. "Introduction". In: *Soft Matter: From Synthetic to Biological Materials*. Ed. by Dhont J., G. Gompper, and G. Nagele. Schriften des Forschungszentrums Julich, 2008, p. I.

[17] Pawel J Zuk, Bogdan Cichocki, and Piotr Szymczak. "GRPY: an accurate bead method for calculation of hydrodynamic properties of rigid biomacromolecules". In: *Biophysical Journal* **115**. (2018), pp. 782–800. DOI: 10.1016/j.bpj.2018.07.015.

[18] A. Einstein. "Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen". In: *Annalen der Physik* **322**. (1905), pp. 549–560. DOI: 10.1002/andp.19053220806.

[19] M. von Smoluchowski. "Zur kinetischen Theorie der Brownschen Molekularbewegung und der Suspensionen". In: *Annalen der Physik* **326**. (1906), pp. 756–780. DOI: 10.1002/andp.19063261405.

[20] J Perrin. "Movement brownien et realite molec". In: *Annales de Chimie et de Physique* **18** (1909), pp. 1–114.

[21] Financial Crisis Inquiry Commission. *Conclusions of the Financial Crisis Inquiry Commission*. Stanford University. 2010. URL: http://fcic-static.law.stanford.edu/cdn_media/fcic-reports/fcic_final_report_conclusions.pdf.

[22] H.C. Öttinger. *Stochastic Processes in Polymeric Fluids: Tools and Examples for Developing Simulation Algorithms*. Springer Berlin Heidelberg, 2012. ISBN: 9783642582905.

[23] NG van Kampen and JJ Lodder. "Constraints". In: *American Journal of Physics* **52**. (1984), pp. 419–424. DOI: 10.1119/1.13647.

[24] Kiyosi Itô. "On a formula concerning stochastic differentials". In: *Nagoya Mathematical Journal* **3** (1951), pp. 55–65. DOI: 10.1017/S0027763000012216.

[25] Bogdan Cichocki, Marcin Rubin, Anna Niedzwiecka, and Piotr Szymczak. "Diffusion coefficients of elastic macromolecules". In: *Journal of Fluid Mechanics* **878** (2019), R3. DOI: 10.1017/jfm.2019.652.

[26] Bogdan Cichocki, Maria L Ekiel-Jeżewska, and Eligiusz Wajnryb. "Brownian motion of a particle with arbitrary shape". In: *The Journal of Chemical Physics* **142**. (2015). DOI: 10.1063/1.4921729.

[27] Maria L. Ekiel-Jeżewska and Eligiusz Wajnryb. "Precise multipole method for calculating hydrodynamic interactions between spherical particles in the Stokes flow". In: *Methods Micro Scale Viscous Flows* **661**. (2009), pp. 127–172.

[28] Derek Juba, Debra J. Audus, Michael Mascagni, Jack F. Douglas, and Walid Keyrouz. "ZENO: Software for calculating hydrodynamic, electrical, and shape properties of polymer and particle suspensions". In: *Journal of Research of National Institute of Standards and Technology* **122**. (2017). DOI: 10.6028/jres.122.020.

[29] John G Kirkwood and Jacob Riseman. "The intrinsic viscosities and diffusion constants of flexible macromolecules in solution". In: *The Journal of Chemical Physics* **16**. (1948), pp. 565–573. DOI: 10.1063/1.1746947.

[30] Prince E. Rouse. "A Theory of the Linear Viscoelastic Properties of Dilute Solutions of Coiling Polymers". In: *The Journal of Chemical Physics* **21**. (1953), pp. 1272–1280. DOI: 10.1063/1.1699180.

[31] Bruno H. Zimm. "Dynamics of Polymer Molecules in Dilute Solution: Viscoelasticity, Flow Birefringence and Dielectric Loss". In: *Journal of Chemical Physics* **24**. (1956), pp. 269–278. DOI: 10.1063/1.1742462.

[32] John Jumper et al. "Highly accurate protein structure prediction with AlphaFold". In: *Nature* **596**. (2021), pp. 583–589. DOI: 10.1038/s41586-021-03819-2.

[33] Jonathan J Ward, Jaspreet S Sodhi, Liam J McGuffin, Bernard F Buxton, and David T Jones. "Prediction and functional analysis of native disorder in proteins from the three kingdoms of life". In: *Journal of Molecular Biology* **337**. (2004), pp. 635–645. DOI: 10.1016/j.jmb.2004.02.002.

[34] Bartosz Różycki and Evzen Boura. "Conformational ensemble of the full-length SARS-CoV-2 nucleocapsid (N) protein based on molecular simulations and SAXS data". In: *Biophysical Chemistry* **288** (2022), p. 106843. DOI: 10.1016/j.bpc.2022.106843.

[35] *Protein Data Bank*. https://www.rcsb.org/. Accessed: 2023-11-02.

[36] *Small Angle Scattering Biological Data Bank*. https://www.sasbdb.org/search/. Accessed: 2023-11-02.

[37] *The Protein Ensemble Database*. https://proteinensemble.org/. Accessed: 2023-11-02.

[38] Maria E Tomasso, Micheal J Tarver, Deepa Devarajan, and Steven T Whitten. "Hydrodynamic radii of intrinsically disordered proteins determined from experimental polyproline II propensities". In: *PLoS Computational Biology* **12**. (2016), e1004686. DOI: 10.1371/journal.pcbi.1004686.

[39] Joseph A Marsh and Julie D Forman-Kay. "Sequence determinants of compaction in intrinsically disordered proteins". In: *Biophysical Journal* **98**. (2010), pp. 2383–2390. DOI: 10.1016/j.bpj.2010.02.006.

[40] Carlos Cinelli, Andrew Forney, and Judea Pearl. "A crash course in good and bad controls". In: *Sociological Methods & Research* (2022), p. 00491241221099552. DOI: 10.1177/00491241221099552.

[41] Brady Neal. *Introduction to Causal Inference from a Machine Learning Perspective*. https://www.bradyneal.com/Introduction_to_Causal_Inference-Dec17_2020-Neal.pdf. 2020.

[42] Martin Karplus and Gregory A. Petsko. "Molecular dynamics simulations in biology". In: *Nature* **347**. (Oct. 1990), pp. 631–639. DOI: 10.1038/347631a0.

[43] Daan Frenkel and Berend Smit. *Understanding molecular simulation: from algorithms to applications*. Vol. 1. Elsevier, 2001.

[44] Charles Murphy, Antoine Allard, Edward Laurence, Guillaume St-Onge, and Louis J Dubé. "Geometric evolution of complex networks with degree correlations". In: *Physical Review E* **97**. (2018), p. 032309. DOI: 10.1103/PhysRevE.97.032309.

[45] J Bleibel, A Dominguez, Martin Oettel, and S Dietrich. "Capillary attraction induced collapse of colloidal monolayers at fluid interfaces". In: *Soft Matter* **10**. (2014), pp. 4091–4109. DOI: 10.1039/C3SM53070A.

[46] Rossitza N. Irobalieva et al. "Structural diversity of supercoiled DNA". In: *Nature Communications* **6**. (2015), p. 8440. ISSN: 2041-1723. DOI: 10.1038/ncomms9440.

[47] Bernard D. Coleman and David Swigon. "Theory of Supercoiled Elastic Rings with Self-Contact and Its Application to DNA Plasmids". In: *Journal of Elasticity* **60**. (2000), pp. 173–221. DOI: 10.1023/A:1010911113919.

[48] Christopher Rackauckas and Qing Nie. "DifferentialEquations.jl–a performant and feature-rich ecosystem for solving differential equations in Julia". In: *Journal of Open Research Software* **5**. (2017). DOI: 10.5334/jors.151.

[49] *A community supported Windows build for jax*. https://github.com/cloudhan/jax-windows-builder. Accessed: 2023-11-02.

[50] Maduni Ranasinghe, Jonathan M Fogg, Daniel J Catanese Jr, Lynn Zechiedrich, and Borries Demeler. "Suitability of double-stranded DNA as a molecular standard for the validation of analytical ultracentrifugation instruments". In: *European Biophysics Journal* **52**. (2023), pp. 267–280. DOI: 10.1007/s00249-023-01671-y.

[51] Suman Nag, Bidyut Sarkar, Arkarup Banerjee, Bankanidhi Sahoo, KAS Varun, and Sudipta Maiti. "The nature of the amyloid-β monomer and the monomer-oligomer equilibrium". In: *Biophysical Journal* **100**. (2011), 202a. DOI: 10.1074/jbc.M110.199885.

[52] Monika Poznar, Rafał Hołubowicz, Magdalena Wojtas, Jacek Gapiński, Ewa Banachowicz, Adam Patkowski, Andrzej Ożyhar, and Piotr Dobryszycki. "Structural properties of the intrinsically disordered, multiple calcium ion-binding otolith matrix macromolecule-64 (OMM-64)". In: *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* **1865**. (2017), pp. 1358–1371. DOI: 10.1016/j.bbapap.2017.08.019.

[53] Svetlana S Khaymina, John M Kenney, Mechthild M Schroeter, and Joseph M Chalovich. "Fesselin is a natively unfolded protein". In: *Journal of Proteome Research* **6**. (2007), pp. 3648–3654. DOI: 10.1021/pr070237v.

[54] The Svedberg and JB Nichols. "The application of the oil turbine type of ultracentrifuge to the study of the stability region of carbon monoxide-hemoglobin". In: *Journal of the American Chemical Society* **49**. (1927), pp. 2920–2934. DOI: 10.1021/ja01410a046.

[55] Borries Demeler and Gary E. Gorbet. "Analytical Ultracentrifugation Data Analysis with UltraScan-III". In: *Analytical Ultracentrifugation*. Ed. by Susumu Uchiyama, Fumio Arisaka, Walter F. Stafford, and Tom Laue. Tokyo: Springer Japan, 2016, pp. 119–143. ISBN: 978-4-431-55983-2. DOI: 10.1007/978-4-431-55985-6_8.

[56] Weiming Cao and Borries Demeler. "Modeling Analytical Ultracentrifugation Experiments with an Adaptive Space-Time Finite Element Solution of the Lamm Equation". In: *Biophysical Journal* **89**. (2005), pp. 1589–1602. DOI: 10.1529/biophysj.105.061135.

[57] Nancy Tompson. "Chapter 6 FCS". In: *Fluorescence Spectroscopy*. Ed. by Joseph R. Lakowicz. Vol. 1. Kluwer Academic Publishers, 2002, pp. 337–378.

[58] I. Gregor and J. Enderlein. "Fluorescence Correlation Spectroscopy". In: *Soft Matter: From Synthetic to Biological Materials*. Ed. by Dhont J., G. Gompper, and G. Nagele. Schriften des Forschungszentrums Julich, 2008, A3.

[59] R. P. Hermann. "Scattering of Neutrons and Photons". In: *Soft Matter: From Synthetic to Biological Materials*. Ed. by Dhont J., G. Gompper, and G. Nagele. Schriften des Forschungszentrums Julich, 2008, A1.

[60] Dudu Tong, Sichun Yang, and Lanyuan Lu. "Accurate optimization of amino acid form factors for computing small-angle X-ray scattering intensity of atomistic protein structures". In: *Journal of Applied Crystallography* **49**. (2016), pp. 1148–1161. DOI: 10.1107/S1600576716007962.

[61] *Pychastic is a stochastic differential equations integrator written entirely in python.* https://github.com/radostw/stochastic. Accessed: 2023-11-02.

[62] *Pychastic documentation.* https://pychastic.readthedocs.io/en/latest/. Accessed: 2023-11-02.

[63] *Python port of Generalized Rotne Prager Yamakawa hydrodynamic tensors.* https://github.com/RadostW/PyGRPY/. Accessed: 2023-11-02.

[64] *PyGRPY documentation.* https://pygrpy.readthedocs.io/en/latest/. Accessed: 2023-11-02.

[65] *Globule-Linker-Model, Minimum-Dissipation-Approximation diffusion coefficient calculator.* https://github.com/radostw/glm-mda-diffusion/. Accessed: 2024-02-29.

[66] *Genrate self avoiding random walks (SARW) for spheres of given sizes.* https://github.com/RadostW/sarw_spheres. Accessed: 2023-11-02.

[67] *All Python package to compute small angle X-ray scattering (SAXS) profiles in one-bead-per-residue approximation with numpy.* https://github.com/RadostW/saxs_single_bead. Accessed: 2023-11-02.

[68] *saxs-single-bead documentation.* https://saxs-single-bead.readthedocs.io/en/latest/. Accessed: 2023-11-02.

[69] *Compute writhe of closed curve with numpy and jax.numpy.* https://github.com/RadostW/PyWrithe. Accessed: 2023-11-02.

[70] *PyWrithe documentation.* https://pywrithe.readthedocs.io/en/latest/. Accessed: 2023-11-02.

[71] Daan Frenkel and Berend Smit. "Chapter 15 - Tackling Time-Scale Problems". In: *Understanding Molecular Simulation (Second Edition)*. Ed. by Daan Frenkel and Berend Smit. Second Edition. San Diego: Academic Press, 2002, pp. 409–429. ISBN: 978-0-12-267351-1. DOI: 10.1016/B978-012267351-1/50017-1.

[72] Maciej Lisicki, Bogdan Cichocki, and Eligiusz Wajnryb. "Near-wall diffusion tensor of an axisymmetric colloidal particle". In: *The Journal of Chemical Physics* **145**. (2016). DOI: 10.1063/1.4958727.

[73] Magdalena Gruziel, Krishnan Thyagarajan, Giovanni Dietler, Andrzej Stasiak, Maria L Ekiel-Jeżewska, and Piotr Szymczak. "Periodic motion of sedimenting flexible knots". In: *Physical Review Letters* **121**. (2018), p. 127801. DOI: 10.1103/PhysRevLett.121.127801.

[74] Peter E Kloeden and Eckhard Platen. *Stochastic differential equations*. Springer, 1992.

[75] Julie D Forman-Kay, Jonathon A Ditlev, Michael L Nosella, and Hyun O Lee. "What are the distinguishing features and size requirements of biomolecular condensates and their implications for RNA-containing condensates?" In: *RNA* **28**. (2022), pp. 36–47. DOI: 10.1261/rna.079026.121.

[76] Patrick J Fleming and Karen G Fleming. "HullRad: Fast calculations of folded and disordered protein and nucleic acid hydrodynamic properties". In: *Biophysical Journal* **114**. (2018), pp. 856–869. DOI: 10.1016/j.bpj.2018.01.002.

[77] Alessandro Sicorello, Bartosz Różycki, Petr V Konarev, Dmitri I Svergun, and Annalisa Pastore. "Capturing the conformational ensemble of the mixed folded polyglutamine protein ataxin-3". In: *Structure* **29**. (2021), pp. 70–81. DOI: 10.1016/j.str.2020.09.010.

[78] M Fixman. "Classical statistical mechanics of constraints: a theorem and application to polymers." In: *Proceedings of the National Academy of Sciences of the United States of America* **71** (1974), pp. 3050–3. DOI: 10.1073/pnas.71.8.3050.

[79] Radost Waszkiewicz, John Burnham Shaw, Maciej Lisicki, and Piotr Szymczak. "Goldilocks fluctuations: dynamic constraints on loop formation in scale-free transport networks". In: *Physical Review Letters* (2024). DOI: 10.1103/PhysRevLett.132.137401.

[80] Jan Turczynowicz, Radost Waszkiewicz, and Łukasz Gładczuk. "Preventing sinking of a disk by leveraging the Parametric Hydraulic Jump phenomenon". In: *arXiv preprint* (2023). DOI: 10.48550/arXiv.2312.13099.

[81] Radost Waszkiewicz and Honorata Bogusz. "The Impact of Parenthood on Labour Market Outcomes of Women and Men in Poland". In: *arXiv preprint* (2024). DOI: 10.48550/arXiv.2306.12924.