


Runaway transition in irreversible polymer condensation with cyclization

Maria Panoukidou ^{1,*} Simon Weir,^{1,*} Valerio Sorichetti ^{2,3} Yair Gutierrez Fosado ¹
Martin Lenz ^{3,4} and Davide Michieletto ^{1,5,†}

¹*School of Physics and Astronomy, University of Edinburgh, Peter Guthrie Tait Road, Edinburgh, EH9 3FD, United Kingdom*

²*Institute of Science and Technology Austria, 3400 Klosterneuburg, Austria*

³*Laboratoire de Physique Théorique et Modèles Statistiques (LPTMS), CNRS, Université Paris-Saclay, F-91405 Orsay, France*

⁴*PMMH, CNRS, ESPCI Paris, PSL University, Sorbonne Université, Université de Paris, F-75005, Paris, France*

⁵*MRC Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh EH4 2XU, United Kingdom*



(Received 24 October 2022; revised 10 January 2024; accepted 21 February 2024; published 20 May 2024)

The process of polymer condensation, i.e., the formation of bonds between reactive end groups, is ubiquitous in both industry and biology. Here we study generic systems undergoing polymer condensation in competition with cyclization. Using a generalized Smoluchowski theory, molecular dynamics simulations and experiments with DNA and ATP-consuming T4 ligase, we find that this system displays a transition, from a ring-dominated regime with finite-length chains at infinite time to a linear-polymers-dominated one with chains that keep growing in time. Finally, we show that fluids prepared close to the transition may have widely different compositions and rheology at large condensation times.

DOI: [10.1103/PhysRevResearch.6.023189](https://doi.org/10.1103/PhysRevResearch.6.023189)

I. INTRODUCTION

Linear polymer condensation is the process by which two polymeric end groups react to form a bond. Beyond its relevance to industry [1], and biotechnology [2], it underpins the biophysics of DNA repair and cloning [3]. In the absence of loop formation, polymer condensation will yield linear chains with average length $\langle l \rangle = 1/(1-p)$, where p is the extent of the condensation reaction [1,4]. However, looping, or cyclization, is expected to be favorable in certain conditions [5–7]. Several theories on reversible polymer condensation and experiments have, over the last decades, attempted to reach a consensus on whether the polymers in such systems will all eventually convert into rings or whether there always will be a linear population at a large timescale [8–15]. Despite this, the polymer physics and chemistry communities have not yet reached a consensus [15–17]. Additionally, there is little literature on irreversible polymer condensation, which we also refer to as “ligation,” henceforth in analogy with the biological process of connecting DNA segments by the enzyme ligase.

Here we study irreversible linear polymer condensation using a combination of theory, simulations, and experiments. First, we show that irreversible polymer condensation is well captured by a modified Smoluchowski coagulation equation [18,19] with an additional sink term that captures ring formation. By spanning a range of monomer concentrations c , we discover that above a critical $c^\dagger \simeq 0.1c^*$ there is a

“runaway” transition characterized by a population of chains that permanently escape cyclization. Here $c^* = l_0/(4/3\pi R_g^3)$ denotes the overlap concentration of polymers, with l_0 and R_g the initial polymer length and radius of gyration, respectively. This transition separates a regime ($c < c^\dagger$) in which all the chains are converted into rings at infinite time, from one ($c > c^\dagger$) in which the length of the linear chains diverges in time. The consequence of this runaway transition is that systems prepared close to c^\dagger and driven out-of-equilibrium by irreversible condensation will display markedly different architectural and rheological features at large enough times.

Our work differs from classic and also more recent papers on polymer condensation and cyclization [8,11,11,16,20] because it deals with irreversible condensation while implementing subdiffusive search and cyclization in a Smoluchowski framework and because it suggests through theory, simulations, and experiments that a runaway transition is expected beyond a critical concentration. We also argue that DNA is particularly suitable to test these theories, as we can readily visualise the products of ligation reactions by gel electrophoresis and distinguish linear and circular forms by treating the samples with exonuclease, as described below. We conclude our paper by discussing the implications of our findings in the design of soft materials and DNA cloning.

II. METHODS

A. Molecular dynamics simulations

We model a 6500-bp-long linear DNA molecule as a bead-spring polymer made of $l_0 = 174$ beads. The total number of polymer chains is $N_c = 200$. The polymers are modeled via the Kremer-Grest model [21]. Each bead has a diameter $\sigma = 13$ nm (or ~ 38 bp), modeled as a truncated and shifted Lennard-Jones (LJ) potential (Weeks-Chandler-

*These authors contributed equally to this work.

†Corresponding author: davide.michieletto@ed.ac.uk

Andersen, WCA),

$$U_{\text{LJ}}(r) = 4\epsilon[(\sigma/r)^{12} - (\sigma/r)^6 + 1/4], \quad (1)$$

for $r < r_c = 2^{1/6}\sigma$ and 0 otherwise. Here r represents the distance between beads, and $\epsilon = 1.0$ (in LJ units) parametrizes the strength of the potential. The diameter of the bead, σ , defines the length units in our system. Consecutive beads are connected through a permanent finite extensible nonlinear elastic (FENE) bond,

$$U_{\text{FENE}}(r) = -0.5KR_0^2 \log[1 - (r/R_0)^2], \quad (2)$$

with $K = 30\epsilon/\sigma^2$ and $R_0 = 1.5\sigma$, which is summed to a Weeks-Chandler-Andersen potential to yield an equilibrium bond length of around 0.9σ . The bending stiffness of the polymer is controlled by a Kratky-Porod interaction,

$$U_b(r) = \frac{k_B T l_p}{\sigma} (1 - \cos \theta), \quad (3)$$

which constrains the angle (θ), defined by the two tangent vectors connecting three consecutive beads along the polymer. Here, $l_p = 4\sigma = 150$ bp is the persistence length of DNA. We note that as $l_0 \gg l_p$, we are always in the flexible chain regime. The solvent is simulated implicitly using a Langevin thermostat so that the time evolution of our system is governed by the stochastic partial differential equations,

$$m\dot{\mathbf{r}} = -\zeta\dot{\mathbf{r}} - \nabla U + \sqrt{2k_B T \zeta} \delta, \quad (4)$$

where \mathbf{r} is the position of a particle, ζ its friction, m its mass, U the sum of the interaction potentials discussed above, and δ white noise with unit variance. The diffusion timescale is $\tau_B = \zeta\sigma^2/k_B T$. The integration of the Langevin equation is done with a velocity-Verlet algorithm using a time step $\Delta t = 0.01\tau_B$ in LAMMPS [22].

Various monomer densities were considered, ranging from $10^{-2}c^*$ to $1c^*$, where $c^* = 0.012\sigma^{-3}$ is the monomer concentration at which the polymers start to overlap. The overlap concentration c^* was measured by computing the radius of gyration R_g of the polymers in equilibrium at infinite dilution. All the systems were equilibrated for a sufficient amount of time to ensure that the polymer chains have moved at least a distance equal to R_g .

After the equilibration step, 40 replicas of production runs were started for each number density considered. The ligation is performed stochastically and is attempted every $t_l = \tau_B$ between two end beads that are closer than $R_c = 1.1\sigma$ using the `fix bond/create` LAMMPS command. The choice of the time in between ligation attempts, t_l , was made so that it was much shorter than the relaxation time of the chains; in this way the condensation process is diffusion limited. The distance threshold R_c was chosen so that the new bond created is a FENE with cutoff 1.5σ and to avoid unstable simulations. The probability of successful ligation (i.e., bond formation) is set to $p_l = 0.1$. This value was chosen to avoid ‘‘granularity’’ in the stochastic condensation reaction. If this parameter was set to 1, all the ends that can react would do so in a single time step introducing granular events in our simulations. Setting $p_l < 1$ introduces some randomness that simply maps to a smaller average condensation rate. We have tested slightly different choices of these parameters, and we found that the main results and qualitative behavior of our results are not

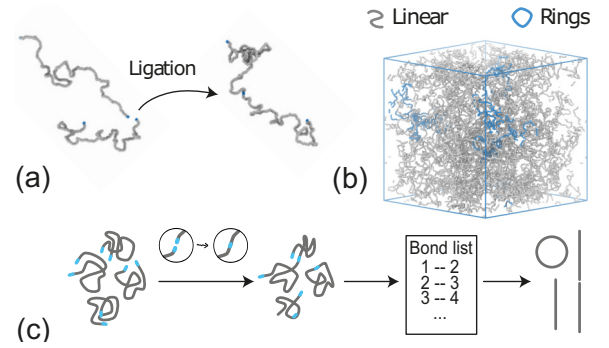


FIG. 1. Molecular dynamics simulations of irreversible condensation. (a) Sketch of a condensation (also referred to as ‘‘ligation’’) event in a molecular dynamics (MD) simulation. (b) Snapshot of the simulation box with rings colored in blue and linear chains in gray. (c) Simulation workflow: we track the bond list to reconstruct the length and topology of the polymers and how these evolve in time.

affected. In particular, we have tested that the reactions remain diffusion limited, even with our choice of p_l . A schematic representation of the simulation process is shown in Fig. 1.

Once ligated, the bond formed between the polymers is irreversible and cannot be broken, therefore accounting for the formation of a covalent bond between the DNA fragments. During the ligation process, snapshots of the system are taken every 10^6 time steps on both the three-dimensional coordinates of the beads and the bond list at those time steps. From the bond list we can, later on, reconstruct the topology of the individual polymers, i.e., if fused with others to form linear chains or if circularized.

For the topology reconstruction, the trajectories and bond lists were analyzed using our PYTHON code. The description of the algorithm can be found in Appendix A.

B. The DSMC algorithm

The modified Smoluchowski equation proposed here [see Eq. (5) below] can only be solved analytically for certain forms of the condensation rate $k_1(i, j)$ and of the cyclization rate $k_0(l)$. As our molecular dynamics simulations are practically limited to systems of hundreds of chains, to characterize the behavior of larger systems we solve the Smoluchowski equation numerically employing the direct-simulation Monte Carlo (DSMC) algorithm [23–26]. DSMC is a powerful stochastic method to solve differential equations such as Eq. (5) and which samples the correct ligation kinetics in the limit of large system sizes. The algorithm employed here is similar to that described in Ref. [26], with the difference that here we do not include fragmentation, but instead, we include ring formation. The description of the algorithm also follows Ref. [26]. The starting point for the Monte Carlo algorithm is an array \mathbf{m} of length N_c , each element i of which contains a number m_i which represents the mass/length of the chain i :

$$\mathbf{m} = (m_1, m_2, \dots, m_{N_c}).$$

A value of 0 corresponds to the absence of a certain chain. Moreover, to satisfy mass conservation, we ensure that $\sum_{i=1}^{N_c} m_i = N_c$ is true at any time during the simulation. Here, N_c denotes the total number of polymer chains. We will also

consider an analogous array \mathbf{r} of length N_c (initially empty), where we save the masses of the rings.

For an initial monodisperse condition, we set $\mathbf{m}_0 = (1, 1, \dots, 1)$. After the array \mathbf{m} is initialized, we run the DSMC simulation, which consists of repeating a Monte Carlo step a large number of times (described in detail in Appendix B). The execution is terminated when the system has reached a state in which there is a single linear chain and several nonreactive rings, where the only possible reaction is the cyclization of the remaining linear chain.

C. Experiments

1. Ligation reactions with DNA

We perform irreversible condensation on linear DNA using T4 ligase New England Biolabs (NEB). This enzyme consumes ATP to form a covalent bond between two proximal and complementary double-stranded DNA ends. More specifically, we perform irreversible condensation on a monodisperse solution of linear, $l_0 = 6500$ -bp-long plasmid (referred to as “1288” plasmid here) that is converted into a linear form by using a restriction enzyme (XhoI). This linearization step is checked on gel electrophoresis. The equilibrium radius of gyration of this linear DNA molecule is about $R_g \simeq l_p \sqrt{l_0/3l_p} \simeq 0.2 \mu\text{m}$ (in agreement with diffusion data from Ref. [27]). This yields an overlap concentration $c^* = 3l_0M_w/(4N_A\pi R_g^3) \simeq 0.2 \mu\text{g}/\mu\text{l}$ with $M_w = 650 \text{ g/mol}$ the molecular weight of a DNA base pair and N_A the Avogadro number. For the low-DNA-concentration experiments, we set the sample at $0.01c^*$, i.e., $c = 2 \mu\text{g}/\text{ml}$. To perform ligation we use T4 ligase (NEB, M0202L, 1U corresponds to 0.5 ng or 0.00735 pmol of protein according to Ref. [28]) and work at a 1x T4 ligase reaction buffer concentration, which contains 1 mM ATP. To classify the topology of the DNA under ligation, we perform time-resolved gel electrophoresis. We prepare a master solution of DNA at the desired concentration, 1x ligation buffer and 2 U/ μl T4 ligase.

After adding T4 ligase, we draw aliquots at time intervals and heat-inactivate the reaction by heating the aliquot at 65 °C for 15 min. We then split the aliquot and treat one of the two subaliquots using exonuclease (RecBCD, Lucigen), an enzyme that digests linear, but not circular, DNA. Finally, we treat all aliquots with Nb.BbvCI nickase (NEB, R0631L) to relax the supercoiled population [29]. The resulting aliquots are run on a gel: we load 20 ng of DNA from each aliquot onto a 1% agarose gel prepared using 1x TAE buffer. A standard λ DNA - HindIII digest (NEB, N3012S) marker is also loaded. The gel is run at $\sim 2.5 \text{ V/cm}$ for 5 h and poststained with SybrGold (ThermoFisher) for 30 min. A Syngene G-box and GENESYS software is used to image the gels.

The combination of nickase (relaxing the DNA supercoiling) and exonuclease (fully digesting linear DNA molecules) allowed the topology of the DNA in each band to be unambiguously identified. Further, the λ DNA - HindIII digest marker confirmed the bands were of the correct size for monomer and dimer lengths. Here the terms “monomer” and “dimer” refer to a single DNA molecule and two molecules ligated, respectively. To extract the relative amount of molecules in each lane, we compute, using ImageJ, the intensity of each lane and account for the fact that the band

with dimers has chains that are twice as long. We then normalize against the sum of the three bands to obtain the relative fraction of chains in each population.

2. Microrheology

The viscosity of the systems is measured using particle tracking microrheology. Solutions are made by mixing 8 μl of 1288 linearized plasmid at different concentrations to a final concentration in the range 2 ng/ μl –500 ng/ μl , with 1 μl of 40 U/ μl T4 ligase and 1 μl of T4 ligase reaction buffer. Control solutions are prepared at the same time and in the same manner, substituting additional Tris-EDTA buffer for the T4 ligase. The samples are kept at room temperature on a roller for several days. The samples are then spiked with $a = 800 \text{ nm}$ PVP (Polyvinylpyrrolidone)-coated polystyrene beads, pipetted and sealed onto a slide and imaged using an inverted microscope. We take a 30-min movie, and we analyze the movies using a particle tracking algorithm (trackpy [30]) and extract the trajectories and mean-squared displacements (MSD) of the tracers $\langle \Delta r^2(t) \rangle = \langle [\mathbf{r}(t + \tau) - \mathbf{r}(t)]^2 \rangle$. Diffusion coefficients are extracted by fitting to the MSDs via $\text{MSD} = 2Dt$. The viscosity is obtained using the Stokes-Einstein relation [31], $\eta = k_B T / (3\pi Da)$.

III. RESULTS

A. Smoluchowski equation with cyclization

In this result section, we first propose a modified Smoluchowski equation [18,19] describing polymers undergoing irreversible condensation (ligation) and cyclization. Linear polymers undergo irreversible ligation with rate $k_1(i, j)$, with i, j the polymerization indexes of the reactants and cyclization with rate $k_0(q)$. The concentrations of linear polymers of polymerization index q at time t , $n_q(t)$, and of rings, $n_q^r(t)$, are thus governed by the following equations:

$$\dot{n}_q(t) = \frac{1}{2} \sum_{ij:i+j=q} k_1(i, j)n_i(t)n_j(t) - n_q(t) \sum_{i=1}^{\infty} k_1(q, i)n_i(t) - k_0(q)n_q(t) \quad (5a)$$

$$\dot{n}_q^r(t) = k_0(q)n_q(t). \quad (5b)$$

Once a linear chain undergoes cyclization, it becomes a ring and cannot undergo ligation anymore, as the reactions are assumed to be irreversible. The kinetics is also constrained by the requirement that the total mass is conserved:

$$\sum_{q=1}^{\infty} q[n_q(t) + n_q^r(t)] = M/V = n \quad \forall t, \quad (6)$$

where M is the total number of monomers and V is the system’s volume. Assuming that the reaction takes place on a timescale larger than the Rouse relaxation time, the length dependence of the annealing rate is [32,33]

$$k_1(i, j) = \tilde{\kappa}_1(D_i + D_j)(R_i + R_j) = \kappa_1(i^{-\alpha} + j^{-\alpha})(i^\nu + j^\nu), \quad (7)$$

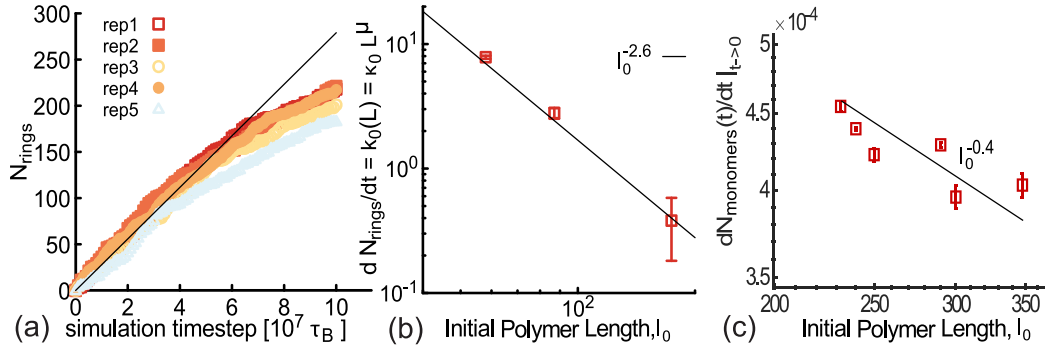


FIG. 2. Validation of cyclization and annealing rates. (a) Number of rings N_{rings} as a function of time for systems initialized to $N = 400$ chains with $l_0 = 87$ monomers each (five independent replicas). The black line represents the numerical derivative of the average $\langle N_{\text{rings}} \rangle$ in the limit $t \rightarrow 0$. (b) The numerical derivative dN_{rings}/dt displays a power-law decay with initial polymer length and with exponent -2.6 , close to the value $-4\nu = 2.4$ for $\nu = 0.588$ (as predicted, see text Sec. III A). (c) Value of $k_1(L)$ obtained from fitting the analytical solution of Eq. (9b), neglecting the second term, in the limit $t \rightarrow 0$. The exponent of the power law is close to the predicted value $\nu - \alpha$ (see text Sec. III A) with $\nu = 0.588$ and $\alpha = 1$.

where $l = il_0$ is the length of a polymer with a degree of polymerization i and l_0 is the initial polymer length, so that the chain's radius of gyration is $R_i = l_0 i^\nu$. In Eq. (7), $\tilde{\kappa}_1$ is a dimensionless constant and κ_1 is a constant that depends on temperature and the viscous friction of the solvent ζ . For example, in the Rouse model [34], $D_i = k_B T / (\zeta l_0 i)$ ($\alpha = 1$) and thus $\kappa_1 = \tilde{\kappa}_1 k_B T / \zeta$.

This condensation rate captures the diffusion-controlled search process [32,33]. The cyclization rate is taken to be $k_0(q) = \kappa_0 q^\mu$, where $\mu = -4\nu$. Note that this is different from the classic Shimada-Yamakawa theory [20,35], which would predict $\mu = -3\nu$ at lengths larger than l_p because we (i) are out-of-equilibrium and (ii) account for the subdiffusion of the polymer end within the volume of the coil.

In equilibrium, the looping probability of a chain is given by the Shimada-Yamakawa formula [20,36]. For $l \gg l_p$ the looping probability of a polymer decays as $P(l) \sim l^\mu$ with $\mu = -3\nu$. This looping probability also holds for an irreversible, nonequilibrium scenario if the process is reaction limited. This is because the chain ends would have the time to explore many conformations and to diffuse the whole volume of the chain, $V \sim l^{3\nu}$, before reacting (as it would happen in equilibrium). In a diffusion-limited, irreversible ligation process, one should instead compute the time it takes for an end to diffuse over a certain distance ξ . The dynamics of the end is described by the Rouse model [34] so that $\xi = b[k_B T t / (\zeta b^2)]^{1/4}$, where b is the size of a Kuhn monomer. Then, setting $\xi = R$ (the size of the polymer coil) one obtains $(R/b)^4 = k_B T t / (\zeta b^2)$, which implies $k_0 \sim t^{-1} \sim R^{-4} \sim l^{-4\nu}$. So considering $\mu = -4\nu$ effectively takes into account the fact that the chain ends are performing a sub-diffusive search process within the polymer coil, as expected for Rouse dynamics.

We have verified that the rate of cyclization scales as the length of the chain to the power -4ν by measuring the rate at which rings are produced for different lengths of the linear chains [Figs. 2(a) and 2(b)]. We have done this by changing the initial length l_0 and by running short simulations, in turn assuming that the system has had no time to create dimers, trimers, etc. and by measuring the number of rings formed. We have observed that the rate of ring formation at early times

$\dot{N}_{\text{rings}} \sim l_0^{-2.6}$, which is close to the expected $\mu = -4\nu = 2.4$ with $\nu = 0.588$. Thus, both theory and simulations suggest that the *diffusion-limited, irreversible* looping probability of a polymer scales with its length as $l^{-4\nu}$.

To validate the functional form used for the condensation rate $k_1(i, j)$ [Eq. (7)], we solve the Smoluchowski equation in the limit of small concentration and short times, where only monomer, dimer, and monomer ring populations are assumed to be present (see next section). In Fig. 2(c) we plot the condensation rate k_1 as a function of different initial polymer lengths obtained by fitting the analytical solution of Eq. (9b) (see below) to the monomer chain population, omitting the second term since no rings were present in these conditions and at early times. From this quantity we fit a power law $l^{\nu-\alpha}$ with $\nu = 0.588$ and find $\alpha \simeq 1$ yields a good fit to the simulated data. This validates de Gennes' hypothesis for the functional form of the condensation rate [Eq. (7)] and our choices for ν and α .

In our experiments and simulations, we typically track the mean length of the polymers as a function of time,

$$\langle l(t) \rangle = l_0 \frac{\sum_{i=1}^{\infty} i [n_i(t) + n_i^r(t)]}{\sum_{i=1}^{\infty} [n_i(t) + n_i^r(t)]}, \quad (8)$$

and we fit this observable with the numerical solution of the full Smoluchowski equation, Eqs. (5a) and (5b). This is practically implemented in a MATLAB code. The numerical evaluation of the system is iterated to find the best free parameters κ_0 and κ_1 that fit the mean length $\langle l(t) \rangle$ obtained from simulations or experiments. The fit is done using the nonlinear least squares MATLAB function `lsqcurvefit`. The rate of ring formation κ_0 and the rate of linear chains formation κ_1 are extracted from this fit by considering 40 independent replicas, allowing us to obtain the error on the fit parameters. For experiments, we typically average over three independent replicas. The numerical and fitting algorithms are described in detail in Appendix C.

1. Time dependence of the mean length: Dilute regime

At short times and in the dilute regime, we can assume that the formation of rings and short n -mers is more favorable.

This assumption is valid in the experiments whenever only linear monomers, dimers, and monomer rings are visible in the gel electrophoresis after ligation. In more dense solutions the presence of rings consisting of more than two monomer chains will be present and is observed in our simulations. Under very dilute conditions, we can thus assume that only monomers, dimers, and monomer rings are present. Denoting the number density of monomer rings, linear monomers, and dimers as n_1^r , n_1 , and n_2 , respectively, the Smoluchowski equations describing the system take the form

$$\frac{dn_1^r(t)}{dt} = k_0(1)n_1(t) \quad (9a)$$

$$\frac{dn_1(t)}{dt} = -k_1(1,1)n_1^2(t) - k_0(1)n_1(t) \quad (9b)$$

$$\frac{dn_2(t)}{dt} = \frac{1}{2}k_1(1,1)n_1^2(t). \quad (9c)$$

We solve Eq. (9b) neglecting the second term as $n_1^2 \ll 1$ in the infinite dilution limit:

$$n_1(t) = n_1(0)e^{-k_0(1)t}. \quad (10)$$

The concentration of monomer rings is thus

$$\frac{dn_1^r(t)}{dt} = k_0n_1(0)e^{-k_0(1)t}, \quad (11)$$

which yields

$$n_1^r(t) = n_1(0)[1 - e^{-k_0(1)t}]. \quad (12)$$

Substituting in Eq. (9c), we get

$$\frac{dn_2(t)}{dt} = \frac{1}{2}k_1(1,1)n_1^2(t) = \frac{1}{2}k_1(1,1)[n_1(0)e^{-k_0(1)t}]^2, \quad (13)$$

from which one obtains

$$n_2(t) = \frac{k_1(1,1)}{4k_0(1)}n_1^2(0)[1 - e^{-2k_0(1)t}]. \quad (14)$$

Assuming these three are the only contributions to the system, the mean length is then given by the following relation:

$$\begin{aligned} \langle l(t) \rangle &= \frac{l_0n_1(t) + l_0n_1^r(t) + 2l_0n_2(t)}{n_1(t) + n_1^r(t) + n_2(t)} \\ &= l_0 \frac{n_1(t) + n_1^r(t) + 2n_2(t)}{n_1(t) + n_1^r(t) + n_2(t)}. \end{aligned} \quad (15)$$

In denser solutions, where the population is more polydisperse, the Smoluchowski equation cannot be solved analytically and we refer to the next section for a scaling prediction and to Sec. III C 1 for a perturbative approach in the limit of small cyclization rate.

As mentioned above, we validate de Gennes' equation for the condensation rate [Eq. (7)] by running short simulations at very high dilution. We then fitted the change in number of ring monomers with the closed solutions [Eq. (12)] for different values of initial polymer length l_0 . Similarly, we fit the solution of Eq. (9b) without the ring term to the population of monomers. From these data we validate the scaling of the rates $k_0(l_0) = \kappa_0 l_0^{-\mu}$ and $k_1(l_0, l_0) = \kappa_1 l_0^{\alpha-\nu}$ as a function of length l_0 [Figs. 2(b) and 2(c)].

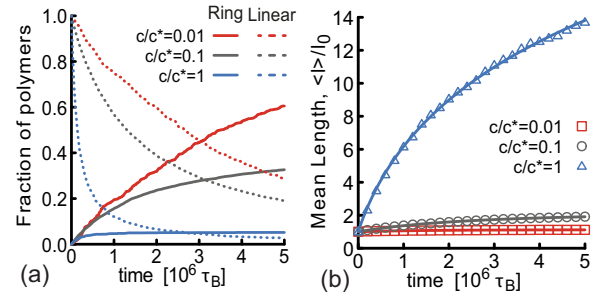


FIG. 3. Results from MD simulations. (a) Number fraction of linear (dotted line) and ring (solid line) polymers during the condensation process for different concentrations c (averaged across 40 independent replicas). (b) Number averaged polymer length for the simulations in (a) as a function of time (symbols), and fitted (solid line) with a numerical solution of Eq. (5) (see Sec. III A for details).

2. Time dependence of the mean length: Concentrated regime

Here we give scaling arguments for the solution of the Smoluchowski equation in the concentrated limit, with the assumption that ring formation is negligible. At the mean-field level, we can make the simplifying assumption that the system can be described by a *single characteristic length scale* l [37]. Under this assumption, the annealing rate scales as

$$k_1(l) \sim DR \sim l^{\nu-\alpha}. \quad (16)$$

The total polymer density n thus follows $\dot{n} = -k_1(l)n^2$, so that from the dimensional analysis the time evolution of the characteristic length is [38,39]

$$l(t) \sim t^{1/(1+\alpha-\nu)} \sim t^{1/(1-\lambda)} \equiv t^\gamma, \quad (17)$$

with $\lambda = \nu - \alpha$. For Rouse dynamics, one has $\alpha = 1$, whereas $\alpha = 2$ for reptation [40]. The Flory exponent has value $\nu = 1/2$ for ideal chains and $\nu = 0.588$ for self-avoiding chains [40]. Assuming concentrations above overlap but still far from the melt concentration (for which one would have ideal chain statistics and $\alpha = 1/2$), we can assume $\nu = 0.588$, so that $\gamma \simeq 0.7$ if the system is unentangled and $\gamma \simeq 0.4$ in the presence of entanglement. We note, however, that using Eq. (16) in the presence of entanglements is only valid for times longer than the reptation time $\tau_R \sim l^3$ [41].

B. Linear DNA condensation

1. Simulations

We first simulate linear condensation using molecular dynamics. As detailed in the Methods section, we simulate polymers with $N = 174$ beads of size $\sigma \sim 38$ bp and persistence length $l_p = 4\sigma = 150$ bp. These polymers are thus designed to coarse-grain 6.5-kb-long DNA plasmids, which will be employed in experiments (see next section). During the simulation, we take snapshots of the system and record the list of bonds to reconstruct the topology of the polymers (see Fig. 1). Over the simulation time, the number of initial linear chains decreases due to the formation of (i) longer linear polymers or (ii) circular chains [Fig. 3(a)]. Additionally, lower monomer concentrations c promote the formation of more rings at large times and a slower decrease of the linear species. We also note that (i) the number fraction of

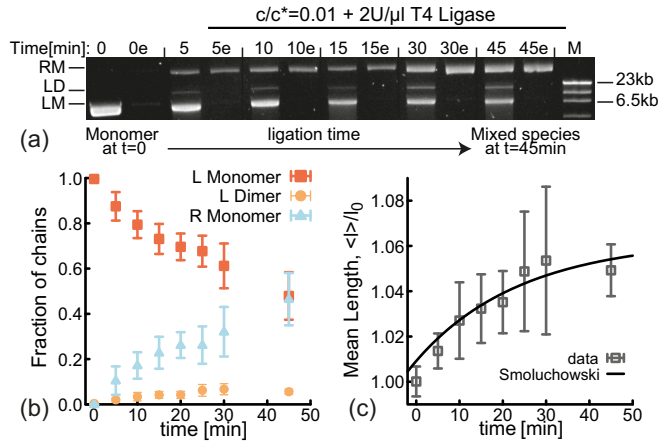


FIG. 4. Experiments of DNA condensation at low concentration. (a) Time-resolved gel electrophoresis during ligation of a 2 ng/ μl ($c/c^* = 0.01$) solution of 6.5-kb linearized plasmid. The lanes marked with “e” are treated with exonuclease to remove linear DNA; “RM” indicates ring monomers, “LD” linear dimers, and “LM” linear monomers. The term monomer here refers to a single DNA molecule. The last column is a λ -HindIII marker as a reference of DNA fragment lengths. The numbers at the top represent the minutes for which the solution was incubated with ligase enzyme. (b) Number fraction of polymers in ring and linear (monomer and dimer) topologies obtained from image analysis of the gel in (a). (c) Number averaged length calculated from three independent gels at 2 ng/ μl (symbols) and associated fit (solid line) using Eq. (5) numerically solved as described in Sec. III A.

rings converges to a finite value at large time, and that (ii) while the number of linear chains appears to go to zero, their mean length increases [Fig. 3(b)]. Accordingly, the (number) average length of polymers grows more quickly for larger c [Fig. 3(b)]. Thus, we conclude that loop formation competes with the growth of the chains and that cyclization is dominant in dilute systems. Interestingly, the curves of the mean length $\langle l(t) \rangle$ can be fitted extremely well by the numerical solution of the Smoluchowski equation, Eq. (5) [Fig. 3(b)].

2. Experiments

As described in the Methods section, we can perform DNA condensation using solutions of linearized DNA plasmids, mixed with ATP and DNA T4 ligase. We then perform a time-resolved experiment, where we draw aliquots from a master reaction at given time points from the addition of the T4 ligase. By running the aliquots on agarose gels we can visualize and compute the fraction of molecules in the linear and ring, monomeric, dimeric, etc. states. Figure 4(a) reports a picture of one such gel, displaying a single band of monomeric linear DNA (as it disappears after exonuclease treatment) at $t = 0$, evolving into three bands, one of which is exonuclease resistant (a monomer ring) at larger times. In Fig. 4(b) we plot the relative abundance of these populations, from which we obtain the number average molecular length $\langle l(t) \rangle$ [Fig. 4(c)].

3. Dimensionless topological parameter

Since we initialize our simulations and experiments below entanglement conditions, we fix $\alpha = 1$ as expected for Rouse

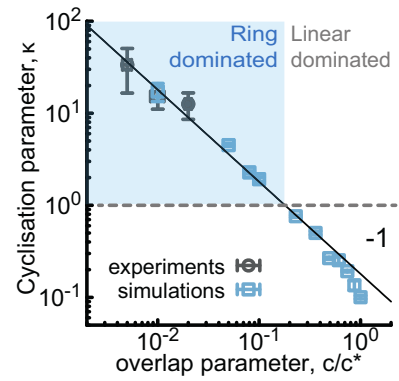


FIG. 5. The dimensionless topological parameter. Dimensionless topological (also “cyclization”) parameter $\kappa = 2\kappa_0/(n_0\kappa_1)$ as a function of c/c^* . This is obtained by fitting simulations and experimental curves $\langle l(t) \rangle$ with the numerical solution of the Smoluchowski equation with κ_0 and κ_1 as free parameters. There are no other free parameters in our model. The scaling $\kappa \sim (c/c^*)^{-1}$ is consistent with κ_0 and κ_1 being independent of concentration, and with this assumption breaking down near $c/c^* \simeq 1$ where three-body interactions become important.

dynamics and $\nu = 0.588$ as expected for self-avoiding polymers [34] (we verified these exponents through direct MD simulations in Figs. 2(a)–2(c)). In general, the Smoluchowski coagulation equation [Eq. (5)] is then solved numerically to fit the data of mean length versus time, $\langle l(t) \rangle$, obtained in simulations and experiments via the free parameters κ_1 and κ_0 . A key number in our system is the ratio of the rates at which polymers are condensed, κ_1 , and the one at which rings are formed, κ_0 . We thus define a dimensionless “topological parameter” $\kappa \equiv 2\kappa_0/(n_0\kappa_1)$, where n_0 is the number density of monomeric chains of length l_0 at the start of the simulation or experiment.

Albeit related to the classic j factor employed in DNA looping [6,20], our topological parameter is more naturally interpreted as the number of rings formed for every two linear chains that are fused together. Intuitively, this number determines the final topological composition of the system. At $\kappa \gg 1$, we expect the final state of the system to be dominated by rings, while for $\kappa \ll 1$, we expect it to be dominated by linear chains. Importantly, since $k_0 \sim \langle l(t) \rangle^{-4\nu}$, the probability of ring formation decreases in time as the average length of the linear chains increases. Accordingly, and even though our system has a ring-only irreversible absorbing state, we conjecture that the strongly decreasing looping probability may effectively yield a very long time transient in which the system is dominated by entangled linear chains with circular contaminants (see below for more simulations on this).

Importantly, we expect the Smoluchowski equation to be valid only in the limit in which three-body interactions are negligible, and the values of κ_0 and κ_1 should be independent of concentration only when c is small enough. By plotting $\kappa \equiv 2\kappa_0/(n_0\kappa_1)$ as a function of c/c^* (where c^* is computed at the beginning of the simulation or experiment), we show that κ scales as $n_0^{-1} \sim (c/c^*)^{-1}$ in both simulations and experiments until $c \simeq c^*$ where it starts to deviate (Fig. 5); this confirms that the Smoluchowski approximation is valid

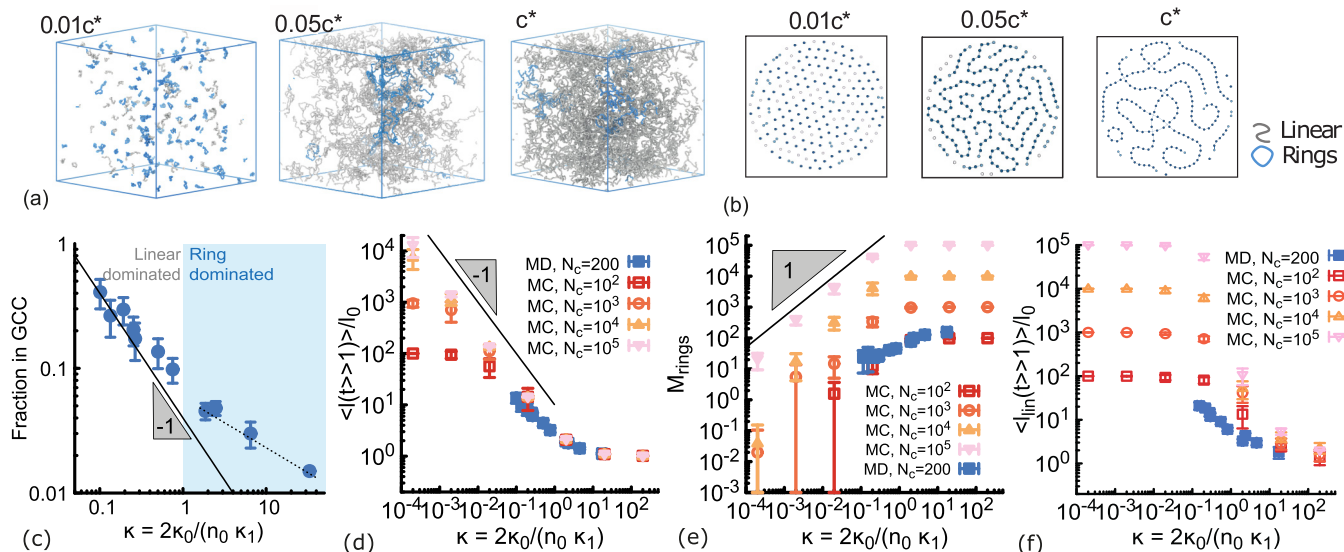


FIG. 6. Runaway transition. (a) Snapshots of MD simulations with initially $N_c = 200$ chains: the blue polymers represent rings of any length, gray polymers linear chains. (b) In the corresponding graph representations, each circle represents a single chain and its color represents the number of chains connected to it: 0 (gray), 1 (cyan), or 2 (blue). For example: a linear monomer chain is gray, a ring monomer is blue, a linear dimer has two cyan nodes, and a dimer ring has two blue nodes. (c) Fraction of monomers in the giant connected component of the system as a function of κ . (d) Average polymer length (including both linear and ring) at a large time as a function of κ . (e) Average mass of rings $M_{\text{rings}} = N_{\text{rings}} \times \langle l_{\text{rings}}(t \gg 1) \rangle$ at large simulation time. Note that some simulations may not display any rings, thus bringing the average M_{rings} below 1. (f) Average mass of linear chains at large simulation time as a function of κ . MC = Monte Carlo; MD = molecular dynamics.

in this range of concentrations. Importantly, in Fig. 5 we also identify the crossover value $\kappa = 1$ (at which the initial cyclization rate is larger than the dimerization rate) around $c/c^* \simeq 0.1 - 0.2$. We note that the agreement between simulations and experiments is excellent for small c/c^* . However, quantitative analysis of gel electrophoresis images at larger c/c^* is challenging due to the poor separation of multimeric bands.

C. Runaway transition

The results in Fig. 3 suggest that at large c/c^* the chains tend to grow longer and cyclization is suppressed; at the same time, the density of reactive ends and the speed of spatial exploration of the chains becomes smaller, thus suppressing dimerization. Due to this kinetic competition, we ask whether the system can truly display a “runaway” phase, defined as a regime where at least one chain permanently escapes cyclization and its length diverges in time. One way to address this question is to look at the number of chains that belong to the longest chain in the system and how this quantity changes in time. By using a graph representation of our simulations [see Figs. 6(a) and 6(b)], we can compute the fraction of chains (nodes) that belongs to the giant connected component (GCC), i.e., the largest cluster of connected monomer chains [Fig. 6(b)]. In Fig. 6(a) one can visually appreciate that at large reaction time, rings (blue) are abundant at low c/c^* , while linear chains (gray) are more abundant at large c/c^* . These systems display a qualitatively different graph topology [Fig. 6(b)]. At small c/c^* (large κ) the network of monomers is mostly disconnected; accordingly, even when the fraction of unreacted bonds goes to 0 at $t \rightarrow \infty$, the average length of the polymers does not diverge. On the contrary, at larger c/c^* we

observe only a few rings and some very long chains that are connecting most of the nodes in the system. Overall, the graph appears much more connected and approaching percolation, i.e., where most of the nodes belong to the GCC, whose size grows with the size of the system [see Fig. 6(c)].

1. Calculation of mass converted into rings at infinite time

Although our simulations support the notion that small values of κ will result in linear chains of increasing lengths and vanishing cyclization rate, they are fundamentally limited to finite-size systems where the cyclization rate of the largest chain never rigorously goes to 0. To estimate the amount of mass that is converted into rings at long times, we do a perturbative calculation valid in the limit of small κ . We start from the continuum Smoluchowski equation:

$$\frac{dn_l(t)}{dt} = \frac{1}{2} \int_0^l K(y, l-y) n_y(t) n_{l-y}(t) dy - \int_0^\infty K(y, l) n_l(t) n_y(t) dy - \frac{1}{2} \kappa l^\mu n_l(t). \quad (18)$$

We define $K \equiv k_1/\kappa_1$, which is thus a scaling function such that $K(ai, al) \sim a^\lambda k_1(i, l)$, where $\lambda = \nu - \alpha$ [42]. We now treat κ_0 perturbatively, starting with $\kappa_0 = 0$. In this case there is no mass lost into rings and we can thus write a conservation law:

$$\int_0^\infty l n_l(t) dl = 1 \quad \forall t. \quad (19)$$

Even for κ nonzero, we assume the loss of mass to cyclization remains finite and of order κ . We will check the self-consistency of this assumption below. Using the mass conservation and Eq. (18), we can write the following scaling

relations: $l^2 n = 1$, $nt^{-1} = l^{1+\lambda} n^2$. We therefore obtain

$$l \sim t^{1/(1-\lambda)}, \quad (20)$$

which is the same as Eq. (17). Note that we must have $\lambda < 1$ for the average length of polymers to increase over time. We can also write the density distribution as

$$n \sim t^{-2/(1-\lambda)}, \quad (21)$$

which in the limit of long times or large lengths may be written as

$$n_l(t) \sim t^{-2/(1-\lambda)} \mathcal{G}\left(\frac{l}{t^{1/(1-\lambda)}}\right), \quad (22)$$

where \mathcal{G} is a scaling function that only depends on the ratio $l/t^{1/(1-\lambda)}$.

We now introduce the ring length distribution $n_l^r(t)$ and its evolution equation as

$$\frac{dn_l^r(t)}{dt} = 2\kappa_0 l^\mu n_l(t). \quad (23)$$

Since at time $t = 0$ there are no rings, we can then write

$$n_l^r(t \rightarrow \infty) = 2\kappa_0 l^\mu \int_0^\infty n_l(t) dt. \quad (24)$$

We can plug in the result we obtained for the distribution of length of linear chains Eq. (22) to yield

$$\begin{aligned} n_l^r(t \rightarrow \infty) &= 2\kappa_0 l^\mu \int_0^\infty t^{-2/(1-\lambda)} \mathcal{G}\left(\frac{l}{t^{1/(1-\lambda)}}\right) dt \\ &= 2\kappa_0 l^\mu (1-\lambda) l^{-(1+\lambda)} \int_0^\infty x^{-\lambda} \mathcal{G}(x) dx, \end{aligned} \quad (25)$$

where we defined $x = l/t^{1/(1-\lambda)}$. Thus, the number density of polymers that are converted into rings over infinite time is

$$n_l^{r\infty} = 2\kappa_0 (1-\lambda) l^{\mu-1-\lambda} \int_0^\infty x^\lambda \mathcal{G}(x) dx. \quad (26)$$

Since $\lambda < 1$ and assuming $\mathcal{G}(x) = \mathcal{O}(1)$ when $x \rightarrow 0$, the integral converges at 0. For convergence of this integral at ∞ , we also require that the scaling function decays faster than $x^{\lambda-1}$.

Assuming this functional form for the distribution of ring lengths at infinite time, we now compute the total average mass transformed into rings at infinite time as

$$\begin{aligned} M_{\text{rings}}^\infty &= \int_1^\infty \ln_l^{r\infty}(t) dl \\ &= 2\kappa_0 (1-\lambda) \int_0^\infty x^{-\lambda} \mathcal{G}(x) dx \int_1^\infty l^{\mu-\lambda} dl. \end{aligned} \quad (27)$$

The convergence of this integral requires that $\lambda - \mu > 1$, and in this case we get

$$M_{\text{rings}}^\infty = 2\kappa_0 \frac{1-\lambda}{\lambda-\mu-1} \int_0^\infty x^{-\lambda} \mathcal{G}(x) dx. \quad (28)$$

From this equation we see that the fraction of mass in rings at infinite time $M_{\text{rings}}^\infty/M_0$ converges to a finite value proportional to κ_0 (and hence < 1 at small κ_0).

With this calculation we have thus shown that at small enough but nonzero κ_0 , the fraction of mass turning into rings

is finite if $\lambda - \mu = \nu - \alpha + 4\nu > 1$ or $\nu > \alpha/5$, which is valid for any type of polymer in the nonentangled ($\alpha = 1$) regime. This implies that in this regime we expect the cyclization probability to decay fast enough and cannot prevent the runaway of the $M_0 - M_{\text{rings}}^\infty$ mass into linear chains that keep growing in time.

Consistently with this, in both molecular dynamics (MD) and Monte Carlo (MC) simulations, we never observe the formation of rings larger than ten initial monomers. As shown here using asymptotic theory, the mass fraction of linear polymers goes to a finite limit at $t \rightarrow \infty$ in a thermodynamic system. We find that the key condition to ensure the existence of a runaway transition is that the cyclization rate $k_0 = \kappa_0 l^\mu$ decays strongly enough. More specifically, we require the exponent μ to be $\mu = -4\nu < -4\alpha/5$ or $\nu > \alpha/5$. This condition is always met in the Rouse unentangled ($\alpha = 1$) regime, provided that the polymers are not fully collapsed ($\nu = 1/3$). This argument establishes the existence of a runaway transition in the limit of large time and at large enough concentrations c/c^* .

2. Direct-simulation Monte Carlo simulations of irreversible condensation

To formally address the existence of a true runaway transition in the thermodynamic limit, we compute the fraction of monomers belonging to linear species in systems of increasing size. To perform this calculation, we employ direct-simulation Monte Carlo [23–26] to solve the Smoluchowski equation in systems with up to 10^5 chains. We run the DSMC code until it has reacted all ends apart from two and compute the average length of the linear population of chains, $\langle l_{\text{lin}}(t \gg 1) \rangle$. As shown in Fig. 6(c), our MD simulations show that at $\kappa = 1$ the GCC displays a change in scaling, growing as $GCC \sim \kappa^{-1} \sim c/c^*$ as $\kappa \rightarrow 0$, suggesting that a qualitative change in behavior takes place at around $\kappa \simeq 1$. In Fig. 6(d) we also plot the number of averaged chain lengths at an arbitrarily large time when the DSMC code has evolved the system as long as possible and has generated only a single linear chain. Figure 6(d) suggests that the linear-dominated regime ($\kappa < 1$) displays an average polymer length at a large reaction time that scales as $\langle l(t \gg 1) \rangle/l_0 \sim \kappa^{-1} \sim c/c^*$. Additionally, the fraction of mass “lost” in forming rings grows as $M_{\text{rings}} \sim \kappa$ and is thus negligible for small enough κ [Fig. 6(e)]. Finally, as shown in Fig. 6(f), the mean length of the linear chains $\langle l_{\text{lin}}(t \gg 1) \rangle$ displays a plateau for $\kappa \lesssim 1$, which grows with the system size, strongly indicating a true runaway transition at the critical value $\kappa \simeq 1$ or $c/c^* \simeq 0.1 - 0.2$.

D. Dynamics and rheology

To test the consequences of the runaway transition on the dynamics and rheology of the system, we perform microrheology experiments and compute dynamics in MD simulations. DNA microrheology is well established, and the effects of DNA concentration, length, and topology on microrheology have been studied in the past [43–49]. Here, we perform microrheology by tracking 800-nm PVP-coated polystyrene beads added in a solution of DNA that has been treated with either 40U T4 ligase for a week (and thus to full extent of reaction) or with buffer for a week (control) at different initial

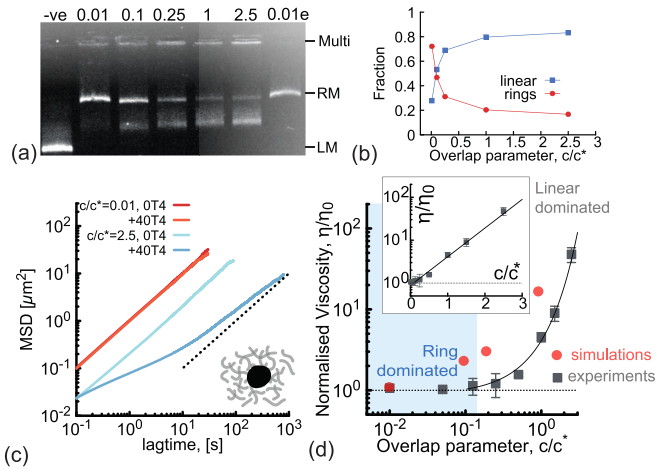


FIG. 7. Rheological consequences of the runaway transition. (a) Time-resolved gel electrophoresis during ligation of $c/c^* = 0.01, 0.1, 0.25, 1,$ and 2.5 solution of 6.5 kb linearized plasmid. The lanes marked with “e” are treated with exonuclease to remove linear DNA; “RM” indicates ring monomers, “LM” linear monomers, and “Multi” various lengths of linear structures. The numbers at the top represent the concentrations of DNA. (b) Fraction of linear and ring molecules of any size as a function of overlap parameter c/c^* . (c) Mean-squared displacements (MSDs) of 800 -nm PVP-coated polystyrene beads diffusing in solutions of DNA. We compare the MSDs in solutions treated for 1 week either with $40U$ T4 ligase or buffer (control). (d) Viscosity of the ligated solutions η , normalized by the viscosity of the control η_0 vs η/η_0 of the ligated solutions in simulations. The inset shows the same plot in a log-linear scale to highlight the exponential increase.

concentrations. We ran a small aliquot of the samples in a gel and observed that indeed at $c/c^* \simeq 0.1$ the fraction of linear chains overcomes the rings at large times [Figs. 7(a) and 7(b)].

At low concentrations, our microrheology shows that the MSD of the tracer particles is unaffected by DNA ligation [Fig. 7(c)]. On the contrary, for $c/c^* \geq 0.1$, we find that the MSDs of the tracers in the ligated systems are much slower and display a stronger subdiffusive behavior than the control [Fig. 7(c)]. From the MSD we extract the large time diffusion coefficient D of the tracers and the effective viscosity of the sample via the Stokes-Einstein equation [31]. The plot of the normalized viscosity [Fig. 7(d)] suggests that a dynamical transition takes place at around $c/c^* = 0.1 - 0.2$ (or $\kappa \simeq 1 - 2$), which matches the structural runaway transition seen before (Fig. 6). After the transition the viscosity increases exponentially with the concentration [see inset of Fig. 7(d)]. This suggests a relaxation process dominated by end-retraction [34], possibly due to the threading of very long linear chains through small rings [50–54] or pseudo-knotted parts of their own extremely long contour [55,56]. We note that, especially at large c/c^* , the ligated solution is extremely elastic, and the passive tracers do not display a freely diffusive behavior even after a lag time of 10 min. We thus argue that the reported η/η_0 may be lower bounds at large c/c^* , which would render the transition even more dramatic. All this implies that, intriguingly, near the transition $c/c^* \simeq 0.1$, systems prepared at similar concentrations may display extremely different rheology at large condensation times. To

further support the existence of a qualitative change in the dynamics, we compute the values of viscosity obtained in MD simulations through the diffusion coefficient of the center of mass of chains that have been ligated for long time at different initial concentrations [see red circles in Fig. 7(d)]. One can appreciate that our simulations also suggest a qualitative difference in dynamics for $c/c^* \geq 0.1$, albeit the transition appears less dramatic than in experiments; we argue that this may be due to finite-size effects present in MD simulations.

IV. CONCLUSION

We have studied a system of linear polymers undergoing irreversible condensation in competition with cyclization. We have shown that the key adimensional parameter controlling growth kinetics is $\kappa = 2\kappa_0/(n_0\kappa_1)$, naturally interpreted as the number of rings formed for any one dimerization. At large concentrations (or $\kappa < 1$) dimerization is kinetically favored and drives the growth of linear chains. While growth disfavors cyclization, it also reduces the number of available reactive ends and the annealing rate of the chains [see Eq. (7)], disfavoring further growth. Despite this, we discover that the net result of this kinetic competition is a runaway transition for $\kappa < 1$ if the cyclization rate decays strongly enough with polymer length, i.e. m with $\nu > \alpha/5$, with ν the metric exponent (typically $1/2$ for random walks and 0.588 for self-avoiding walks) and α the dynamics exponent (typically 1 for Rouse and 2 for reptative dynamics). In these conditions, the fraction of monomers transformed into rings is finite, thus leaving the rest of the monomers available to form a permanently growing linear chain which then drives a runaway reaction.

We also discover that the runaway transition has deep consequences on the rheology and triggers an exponential increase for $\kappa < 1$ (or $c/c^* > 0.1$). Our results suggest that it may be possible to tune the final topological composition of ligated systems by judiciously choosing c/c^* . For instance, the most likely regime to form large rings and ring-linear blends [51,52] is near the transition $c/c^* \simeq 0.1$. Mixing polymer families with different reactive ends further enhances the designability, as it introduces different c^* for each family. Our results can be used to optimize the conditions for DNA engineering, e.g., transfection vectors [2] ought to be ligated at $c/c^* < 0.1$, whereas synthetic chromosome assemblies [57] at large c/c^* . Finally, it may be possible to couple dissipative DNA breakage reactions [48,58,59] with ATP-consuming ligation to create dense solutions of self-sustained topologically active viscoelastic fluids, which would be an interesting active fluid to investigate in the future.

Source codes are available, see Ref. [60].

ACKNOWLEDGMENTS

D.M. acknowledges the support of the Royal Society via a University Research Fellowship. This project has received support from the European Research Council (ERC) under the European Union’s Horizon 2020 Research and Innovation Program (Grant Agreement No. 947918 to D.M. and No. 677532 to M.L.). The authors acknowledge insightful

discussions with Daan Noordermeer and Antonio Valdes, who also kindly gifted us with the 1288 plasmid.

APPENDIX A: TOPOLOGY RECONSTRUCTION ALGORITHM

We will refer to the topology reconstruction code from now on as TR. The code takes as input the instantaneous trajectory and bond list from LAMMPS and checks for newly formed linear and ring chains. The output of the PYTHON code is a file containing the number and length of linear chains that have formed in a given simulation time step. Similar files are produced for the ring chains. These files are then used to calculate the average length and the number of linear/ring chain figures.

The starting point of the TR algorithm is an array \mathbf{b} of size $N_b \times 2$; each row $b_i = (id_1, id_2)$ represents the IDs of atoms that are bonded within the system:

$$\mathbf{b} = \begin{bmatrix} id_1 & id_2 \\ id_3 & id_4 \\ \vdots & \vdots \\ id_{N_b-1} & id_{N_b} \end{bmatrix}.$$

Since not all particles are linked together, some do not appear in the array \mathbf{b} . To avoid operations with large sparse arrays, the matrix \mathbf{b} is mapped to $\tilde{\mathbf{b}}$ that contains only indexes from 1 to the maximum number of atoms connected M : $\mathbf{b} \rightarrow \tilde{\mathbf{b}}$.

The next step of the TR algorithm is to create a connectivity matrix \mathbf{C} based on the list $\tilde{\mathbf{b}}$. Each row of \mathbf{C} represents an atom index and consists of three components $\mathbf{C}(id_i, :) = (id_{i-1}, id_{i+1}, flag)$. Since in our case a particle can be linked with two more particles, the first two components of each row id_{i-1}, id_{i+1} represent the connections of particle id_i (note that id_{i-1} and id_{i+1} are not necessarily consecutive in one dimension but can be any other particle bonded to particle i). The third component, $flag$, takes only the values 0, 1 and accounts for the particles id_i that already belong to a polymer chain. The flag column of \mathbf{C} is initialized to zeros. During the reading process of the connectivity matrix, the algorithm switches the flags to one of the particles that are already considered to belong in a chain. Rings are extracted in the same manner, and a ring is found if the current atom index is the same as the starting atom index. This reading process outputs N_c arrays that have different lengths, and each of them contains the particle (mapped) ids that are connected in a polymer chain. The final step of the TR algorithm is to map the atom indexed back to the original ones: $M^{-1} : \tilde{\mathbf{b}} \rightarrow \mathbf{b}$. This algorithm is very generic and can be applied also in cases where the atoms are not initially in polymers as in our case but rather individual atoms that can connect during the simulation.

APPENDIX B: DESCRIPTION OF THE MONTE CARLO STEP

In this paragraph we will describe the single Monte Carlo (MC) step, which is repeated a large number of times during the numerical resolution of Eq. (5) performed using the DSMC algorithm. With reference to Eq. (5), we define $n_f \equiv V \sum_{i=1}^{\infty} n_i$ (total number of chains). Before the start of the simulation, we give an estimate of the maximum annealing

rate k_{\max} and of the maximum cyclization rate k_0^{\max} . The exactness of the algorithm does not depend on this initial choice; however, choosing values that are too far from the actual maximum rates can lead to a reduced efficiency [23].

During every MC step, we either attempt to perform a ligation reaction (with probability p) or a cyclization one (with probability $1 - p$). The value of p is calculated initially and then updated during the simulation in such a way that the average number densities $n(l)$ satisfy (5). At the beginning of each MC step, p is evaluated as

$$p^{-1} = 1 + \frac{2Nk_o^{\max}}{(n_f - 1)nk_{\max}}. \quad (\text{B1})$$

We will show below that this choice also guarantees that the simulation samples the correct number of cyclization and ligation events per unit volume and unit time as required by Eq. (5).

We define a *waiting time* variable that is set to zero at the beginning of the simulation. After each reaction, a waiting time increment is generated. These increments are also chosen to guarantee the correct number of ligation and fragmentation reactions per unit of time and volume, as detailed below. We can now describe the MC step, during which the following actions are performed:

(1) We evaluate the probability of annealing p according to Eq. (B1). The explicit form of p , Eq. (B1), will be discussed in detail below.

(2) We pick a random number $0 \leq r \leq 1$ from a uniform distribution. If $r \leq p$, we attempt a ligation event:

(1) We pick a pair of elements of the array \mathbf{m} , denoted α, β at random. Since there are $n_f(n_f - 1)$ ordered pairs of chains, the probability of picking a specific pair is $[n_f(n_f - 1)]^{-1}$. Let the length associated with these elements be $m_\alpha = i$ and $m_\beta = j$.

(2) We evaluate the ligation rate $k_1(i, j)$ for the two chains. If $k_1(i, j) > k_{\max}$, we set $k_1^{\max} = k_1(i, j)$ and return to (1). Otherwise, we continue.

(3) We pick another random number r' and perform the ligation if $r' \leq k_1(i, j)/k_1^{\max}$. If ligation is unsuccessful, we return to (1). Otherwise, we continue.

(4) We increment the waiting time by $\Delta t_{i,j}^{\text{lig}} = \frac{2AN}{n_f(n_f-1)nk_{ij}}$. Here A is a parameter, the only condition of which is that it must be between 0 and 1, as we discuss in more detail below.

(5) After incrementing the waiting time, we update \mathbf{f} by setting $m_\alpha = 0$ and $m_\beta = i + j$.

(3) If $r > p$, we attempt a cyclization event:

(1) We pick a chain γ at random with probability n_f^{-1} . Let $m_\gamma = l$.

(2) We evaluate the cyclization rate $k_0(l)$. If $k_0(l) > k_0^{\max}$, set $k_0^{\max} = k_0(l)$ and return to (1). Otherwise, we continue.

(3) We extract another random number $0 \leq r' \leq 1$ from a uniform distribution and perform cyclization if $r' \leq k_0(l)/k_0^{\max}$. If cyclization is unsuccessful, we return to (1). Otherwise, we continue.

(4) We increment the waiting time by $\Delta t_l^{\text{cyc}} = \frac{1-A}{n_f k_0(l)}$, with A defined above in step (2).

(5) We record the value of l in \mathbf{r} and set $m_\gamma = 0$.

We now prove that the definitions of p [Eq. (B1)], the waiting time increments $\Delta t_{i,j}^{\text{lig}}$ (for ligation) and $\Delta t_{i,k-i}^{\text{cyc}}$ (for cyclization) give several ligation and cyclization events per unit time, which is consistent with the Smoluchowski equation, Eq. (5). Over a single MC step, the mean number of ligation events involving the ordered pair of filaments (α, β) is

$$\langle \#L_{\alpha,\beta} \rangle \equiv \frac{p}{n_f(n_f - 1)} \frac{k_1(m_\alpha, m_\beta)}{k_1^{\text{max}}}. \quad (\text{B2})$$

We note that in the algorithm we consider (m_α, m_β) as an ordered pair, and thus in (B2) we consider the reaction $(i, j) \rightarrow l$ as distinct from $(j, i) \rightarrow l$. The mean number of ligation events involving *any two chains* with lengths i, j can be obtained by multiplying the above quantity by $2(1 - \delta_{ij}/2)V^2n_in_j$. The factor $2(1 - \delta_{ij}/2)$ takes into account the fact that, as mentioned above, for $i \neq j$, there are two ways to perform the ligation, whereas for $i = j$ there is only one. The factor $V^2n_in_j$ is the product of the volume fractions of filaments of lengths i and j . We thus have

$$2V^2n_in_j \left(1 - \frac{\delta_{ij}}{2}\right) \times \frac{p}{n_f(n_f - 1)} \frac{k_1(i, j)}{k_1^{\text{max}}} \\ = Vk_1(i, j)n_in_j \left(1 - \frac{\delta_{ij}}{2}\right) \Delta t, \quad (\text{B3})$$

where we have equated the mean number of ligation events involving any two chains with lengths i, j to the value required by the Smoluchowski equation. Recalling that $n = N/V$, we thus find

$$\Delta t = \frac{2pN}{n_f(n_f - 1)nk_1^{\text{max}}}. \quad (\text{B4})$$

Equation (B4) relates the time interval Δt to the probability of ligation. We will now obtain a second equality involving p and Δt , which will allow us to prove that the expression Eq. (B1) for p guarantees the correct number of ligation and cyclization events per unit time.

The mean number of cyclization events involving chains γ is

$$\langle \#C_\gamma \rangle \equiv \frac{(1-p)k_0(m_\gamma)}{n_f k_0^{\text{max}}}. \quad (\text{B5})$$

To obtain the mean number of cyclizations of a generic l -mer we need to multiply this quantity by Vn_l , *i.e.*, the volume fraction of filaments of length l . Equating this quantity to the expected number of rings formed in a time interval Δt we obtain

$$Vn_l \times \frac{(1-p)k_0(l)}{n_f k_0^{\text{max}}} = k_0(l)n_l V \Delta t, \quad (\text{B6})$$

and hence

$$\Delta t \equiv \frac{1-p}{n_f k_0^{\text{max}}}. \quad (\text{B7})$$

By equating the two expressions for Δt , Eqs. (B4) and (B7), we find Eq. (B1). We have thus proven that the latter is the correct expression of p , which gives the correct number of cyclization and ligation events per unit time and unit volume, as required by the Smoluchowski equation.

Finally, we will prove below that the constants A and $1 - A$ introduced when calculating the waiting time increments are consistent with Eqs. (B4) and (B7). To show this, it is sufficient to observe that the total time increment during an MC step is

$$\Delta t = \sum_{0 \leq \alpha < \beta \leq n_f - 1} \langle \#L_{\alpha,\beta} \rangle \Delta t_{m_\alpha, m_\beta}^{\text{lig}} + \sum_{i=1}^{m_\gamma - 1} \langle \#C_\gamma \rangle \Delta t_{m_\gamma}^{\text{cyc}} \\ = \sum_{0 \leq \alpha < \beta \leq n_f - 1} \left[\frac{pk_{m_\alpha, m_\beta}}{n_f(n_f - 1)k_{\text{max}}} \right] \left[\frac{2AN}{n_f(n_f - 1)nk_{m_\alpha, m_\beta}} \right] \\ + \sum_{i=1}^{m_\gamma - 1} \left[\frac{(1-p)k_o(l)}{nk_o^{\text{max}}} \right] \frac{1-A}{n_f k_o(l)} \\ = \frac{2ApN}{n_f(n_f - 1)nk_{\text{max}}} + \frac{(1-A)(1-p)}{n_f k_o^{\text{max}}}. \quad (\text{B8})$$

One can see that this equality is consistent with Eqs. (B4) and (B7). We note that the algorithm samples on average the correct kinetics independently of the value of A , as long as $0 \leq A \leq 1$. Here we take $A = 1$, meaning that the waiting time increment is calculated only after a successful ligation reaction but not after a successful cyclization reaction.

APPENDIX C: NUMERICAL INTEGRATION OF MODIFIED SMOLUCHOWSKI

Solving the Smoluchowski equation to fit the data from MD simulations consists of two main parts:

(1) We create an objective function for the *lsqcurvefit* (called *Obj_smoluchowski*) that takes as input the array of initial coefficient guess $K_0 = (\kappa_1, \kappa_0)$ and the time data array *xdata*. It returns the average length as a function of time, array *ydata*. In the objective function:

(1) An array $\mathbf{L} = \mathbf{n} \cdot l_0$ is initialized where $n = \{1, 2, \dots, N_c = 200\}$ and $l_0 = 174$. This represents the set of lengths that can be found in the system (recall that we initialize our MD simulations with 200 chains of 174 beads each). Also, the arrays with the number density of linear and ring chains are initialized as follows, $\mathbf{n}_0 = (N_c/vol, 0, \dots, 0)_{1 \times N_c}$ and $\mathbf{n}_{r_0} = (0, \dots, 0)_{1 \times N_c}$, since initially all the molecules are linear chains. Here, *vol* denotes the volume of the simulation box.

(2) for $t = \{1$ to simulation final step time} do call $(\mathbf{n}_{\text{L}_{\text{new}}}, \mathbf{n}_{\text{R}_{\text{new}}}) = \text{exEuler_smoluchowski}(\mathbf{n}_{\text{L}}, \mathbf{n}_{\text{R}}, K)$ function (see point 2 below)

(3) update arrays $\mathbf{n}_{\text{L}} = \mathbf{n}_{\text{L}_{\text{new}}}$ and $\mathbf{n}_{\text{R}} = \mathbf{n}_{\text{R}_{\text{new}}}$. Calculate the total average length $\mathbf{l}_{\text{total}}$ as

$$\mathbf{l}_{\text{total}}(\mathbf{t}) = \frac{\mathbf{n}_{\text{L}_{\text{new}}} \cdot \mathbf{L} + \mathbf{n}_{\text{R}_{\text{new}}} \cdot \mathbf{L}}{\sum_i^{N_m} n_{\text{L}_{\text{new}}}^i + \sum_i^{N_m} n_{\text{R}_{\text{new}}}^i}$$

(4) *exit for loop* and parse $\mathbf{l}_{\text{total}}(\mathbf{t})$ to *ydata*.

(2) The *exEuler_smoluchowski* function takes as input the initial number densities of linear and ring chains and the reaction rates $\mathbf{n}_{\text{L}}, \mathbf{n}_{\text{R}}, K = (\kappa_1, \kappa_0)$. Based on the given rates K , it outputs the final number density arrays $\mathbf{n}_{\text{L}_{\text{new}}}, \mathbf{n}_{\text{R}_{\text{new}}}$, after the reactions have taken place. When this function is called, the number densities of linear and ring chains of each population

are updated according to Eq. (5). The monomer, dimer, and so on populations are increased according to the first two terms of Eq. (5), while the number of them that is converted into rings is subtracted by the $\mathbf{n}_{L_{\text{new}}}$ and added to the $\mathbf{n}_{R_{\text{new}}}$ array.

In the first two terms of Eq. (5a), the rate $k_1(i, j)$ is not a scalar quantity but rather a matrix that follows the relation

Eq. (7). The extracted coefficient against which the fitting is optimized is the scalar κ_1 . Similarly, for the sink term of Eqs. (5a) and (5b), the equation $k_0(l) = \kappa_0 l^{-4\nu}$ is used and the fitting coefficient exported is the scalar κ_0 .

The coefficients K are updated iteratively by the *lsqcurvefit* algorithm to best fit the data. Once the optimum values are obtained, the algorithm terminates.

-
- [1] M. Rubinstein and H. R. Colby, *Polymer Physics* (Oxford University Press, Oxford, UK, 2003).
- [2] R. T. Oliynyk and G. M. Church, Efficient modification and preparation of circular DNA for expression in cell culture, *Commun. Biol.* **5**, 1393 (2022).
- [3] B. Alberts, A. Johnson, J. Lewis, D. Morgan, and M. Raff, *Molecular Biology of the Cell* (Taylor & Francis, London, 2014), p. 1464.
- [4] P. J. Flory, Molecular size distribution in linear condensation polymers, *J. Am. Chem. Soc.* **58**, 1877 (1936).
- [5] M. E. Cates and S. J. Candau, Ring-driven shear thickening in wormlike micelles? *Europhys. Lett.* **55**, 887 (2001).
- [6] R. Vafabakhsh and Taekjip Ha, Extreme bendability of DNA Less than 100 base pairs long revealed by single-molecule cyclization, *Science* **337**, 1097 (2012).
- [7] H. Zhou, J. Woo, A. M. Cok, M. Wang, B. D. Olsen, and J. A. Johnson, Counting primary loops in polymer gels, *Proc. Natl. Acad. Sci. USA* **109**, 19119 (2012).
- [8] H. Jacobson and W. H. Stockmayer, Intramolecular reaction in systems, *J. Chem. Phys.* **18**, 1600 (1950).
- [9] P. J. Flory and J. A. Semlyen, Macrocyclization equilibrium constants and the statistical configuration of poly(dimethylsiloxane) chains, *J. Am. Chem. Soc.* **88**, 3209 (1966).
- [10] K. Suematsu and T. Okamoto, Theory of ring formation in irreversible system: AB model, *Colloid Polym. Sci.* **270**, 421 (1992).
- [11] C. C. Chen and E. E. Dormidontova, Ring-chain equilibrium in reversibly associated polymer solutions: Monte Carlo simulations, *Macromolecules* **37**, 3905 (2004).
- [12] G. Ercolani and D. Stefano, Ring-expanding polymerization by reversible ring fusion. A fascinating process driven by entropy, *J. Phys. Chem. B* **112**, 4662 (2008).
- [13] C. Madeleine-Perdrillat, F. Delor-Jestin, and P. De Sainte Claire, The polysiloxane cyclization equilibrium constant: A theoretical focus on small and intermediate size rings, *J. Phys. Chem. B* **118**, 330 (2014).
- [14] S. Di Stefano and L. Mandolini, The canonical behavior of the entropic component of thermodynamic effective molarity. An attempt at unifying covalent and noncovalent cyclizations, *Phys. Chem. Chem. Phys.* **21**, 955 (2019).
- [15] H. R. Kricheldorf, S. M. Weidner, and F. Scheliga, Synthesis of cyclic polymers and flaws of the Jacobson-Stockmayer theory, *Polym. Chem.* **11**, 2595 (2020).
- [16] M. Lang and K. S. Kumar, Simple and general approach for reversible condensation polymerization with cyclization, *Macromolecules* **54**, 7021 (2021).
- [17] H. R. Kricheldorf, S. M. Weidner, and J. Falkenhagen, The role of transesterifications in reversible polycondensations and a reinvestigation of the Jacobson-Beckmann-Stockmayer experiments, *Polym. Chem.* **13**, 1177 (2022).
- [18] M. v. Smoluchowski, Versuch einer mathematischen theorie der koagulationskinetik kolloider lösungen, *Z. Phys. Chem.* **92U**, 129 (1918).
- [19] R. M. Ziff, Kinetics of polymerization, *J. Stat. Phys.* **23**, 241 (1980).
- [20] J. Shimada and H. Yamakawa, Ring-closure probabilities for twisted wormlike chains. Application to DNA, *Macromolecules* **17**, 689 (1984).
- [21] K. Kremer and G. S. Grest, Dynamics of entangled linear polymer melts: A molecular-dynamics simulation, *J. Chem. Phys.* **92**, 5057 (1990).
- [22] S. Plimpton, Fast parallel algorithms for short-range molecular dynamics, *J. Comput. Phys.* **117**, 1 (1995).
- [23] A. L. Garcia, C. Van Den Broeck, M. Aertsens, and R. Serneels, A Monte Carlo simulation of coagulation, *Physica A* **143**, 535 (1987).
- [24] K. Liffman, A direct simulation Monte Carlo method for cluster coagulation, *J. Comput. Phys.* **100**, 116 (1992).
- [25] F. E. Kruis, A. Maisels, and H. Fissan, Direct simulation Monte Carlo method for particle coagulation and aggregation, *AIChE J.* **46**, 1735 (2000).
- [26] Q. D. Tran, V. Sorichetti, G. Pehau-Arnaudet, M. Lenz, and C. Leduc, Fragmentation and entanglement limit vimentin intermediate filament assembly, *Phys. Rev. X* **13**, 011014 (2023).
- [27] R. M. Robertson, S. Laib, and D. E. Smith, Diffusion of isolated DNA molecules: Dependence on length and topology, *Proc. Natl. Acad. Sci. USA* **103**, 7310 (2006).
- [28] W. H. Taylor and P. J. Hagerman, Application of the method of phage T4 DNA ligase-catalyzed ring-closure to the study of DNA structure: II. NaCl-dependence of DNA flexibility and helical repeat, *J. Mol. Biol.* **212**, 363 (1990).
- [29] A. Bates and A. Maxwell, *DNA Topology* (Oxford University Press, Oxford, UK, 2005).
- [30] J. C. Crocker, M. T. Valentine, E. R. Weeks, T. Gisler, P. D. Kaplan, A. G. Yodh, and D. A. Weitz, Two-point microrheology of inhomogeneous soft materials, *Phys. Rev. Lett.* **85**, 888 (2000).
- [31] J.-P. Hansen and I. R. McDonald, *Theory of Simple Liquids: With Applications to Soft Matter* (Academic Press, New York, 2013).
- [32] P. G. De Gennes, Kinetics of diffusion-controlled processes in dense polymer systems. I. Nonentangled regimes, *J. Chem. Phys.* **76**, 3316 (1982).
- [33] A. Y. Grosberg, P. G. Khalatur, and A. R. Khokhlov, Polymeric coils with excluded volume in dilute solution: The invalidity of the model of impenetrable spheres and the influence of excluded volume on the rates of diffusion-controlled

- intermacromolecular reactions, *Die Makromol. Chem. Rapid Commun.* **3**, 709 (1982).
- [34] M. Doi and S. Edwards, *The Theory of Polymer Dynamics* (Oxford University Press, Oxford, UK, 1988).
- [35] A. Rosa, N. B. Becker, and R. Everaers, Looping probabilities in model interphase chromosomes, *Biophys. J.* **98**, 2410 (2010).
- [36] A. Rosa and R. Everaers, Structure and dynamics of interphase chromosomes, *PLoS Comput. Biol.* **4**, e1000153 (2008).
- [37] V. Sorichetti and M. Lenz, Transverse fluctuations control the assembly of semiflexible filaments, *Phys. Rev. Lett.* **131**, 228401 (2023).
- [38] P. G. J. van Dongen and M. H. Ernst, Kinetics of reversible polymerization, *J. Stat. Phys.* **37**, 301 (1984).
- [39] P. G. J. van Dongen and M. H. Ernst, Dynamic scaling in the kinetics of clustering, *Phys. Rev. Lett.* **54**, 1396 (1985).
- [40] M. Doi and S. Edwards, *The Theory of Polymer Dynamics* (Oxford University Press, Oxford, UK, 1988).
- [41] P. G. De Gennes, Kinetics of diffusion-controlled processes in dense polymer systems. II. Effects of entanglements, *J. Chem. Phys.* **76**, 3322 (1982).
- [42] P. Meakin and M. H. Ernst, Scaling in aggregation with breakup simulations and mean-field theory, *Phys. Rev. Lett.* **60**, 2503 (1988).
- [43] T. G. Mason, K. Ganesan, J. H. van Zanten, D. Wirtz, and S. C. Kuo, Particle tracking microrheology of complex fluids, *Phys. Rev. Lett.* **79**, 3282 (1997).
- [44] X. Zhu, B. Kundukad, and J. R. Van Der Maarel, Viscoelasticity of entangled λ -phage DNA solutions, *J. Chem. Phys.* **129**, 1 (2008).
- [45] B. A. Krajina, C. Tropini, A. Zhu, P. Digiacomo, J. L. Sonnenburg, S. C. Heilshorn, and A. J. Spakowitz, Dynamic light scattering microrheology reveals multiscale viscoelasticity of polymer gels and precious biological materials, *ACS Central Sci.* **3**, 1294 (2017).
- [46] M. Tanoguchi and Y. Murayama, Length dependence of viscoelasticity of entangled-DNA solution with and without external stress, *AIP Adv.* **8**, 105218 (2018).
- [47] J. Smrek, J. Garamella, R. Robertson-Anderson, and D. Michieletto, Topological tuning of DNA mobility in entangled solutions of supercoiled plasmids, *Sci. Adv.* **7**, 1 (2021).
- [48] D. Michieletto, P. Neill, S. Weir, D. Evans, N. Crist, V. A. Martinez, and R. M. Robertson-Anderson, Topological digestion drives time-varying rheology of entangled DNA fluids, *Nat. Commun.* **13**, 4389 (2022).
- [49] Y. A. G. Fosado, J. Howard, S. Weir, A. Noy, M. C. Leake, and D. Michieletto, Fluidification of entanglements by a DNA bending protein, *Phys. Rev. Lett.* **130**, 058203 (2023).
- [50] J. Roovers, Viscoelastic properties of polybutadiene rings, *Macromolecules* **21**, 1517 (1988).
- [51] M. Kapnistos, M. Lang, D. Vlassopoulos, W. Pyckhout-Hintzen, D. Richter, D. Cho, T. Chang, and M. Rubinstein, Unexpected power-law stress relaxation of entangled ring polymers, *Nat. Mater.* **7**, 997 (2008).
- [52] J. D. Halverson, G. S. Grest, A. Y. Grosberg, and K. Kremer, Rheology of ring polymer melts: From linear contaminants to ring-linear blends, *Phys. Rev. Lett.* **108**, 038301 (2012).
- [53] Y. Zhou, C. D. Young, M. Lee, S. Banik, D. Kong, G. B. McKenna, R. M. Robertson-Anderson, C. E. Sing, and C. M. Schroeder, Dynamics and rheology of ring-linear blend semidilute solutions in extensional flow: Single molecule experiments, *J. Rheol.* **65**, 729 (2021).
- [54] D. Parisi, J. Ahn, T. Chang, D. Vlassopoulos, and M. Rubinstein, Stress relaxation in symmetric ring-linear polymer blends at low ring fractions, *Macromolecules* **53**, 1685 (2020).
- [55] D. Michieletto, D. Marenduzzo, E. Orlandini, G. P. Alexander, and M. S. Turner, Dynamics of self-threading ring polymers in a gel, *Soft Matter* **10**, 5936 (2014).
- [56] B. W. Soh, A. R. Klotz, R. M. Robertson-Anderson, and P. S. Doyle, Long-lived self-entanglements in ring polymers, *Phys. Rev. Lett.* **123**, 048002 (2019).
- [57] N. Annaluru, H. Muller, L. A. Mitchell, S. Ramalingam, G. Stracquadanio, S. M. Richardson, J. S. Dymond, Z. Kuang, L. Z. Scheifele, E. M. Cooper, Y. Cai, K. Zeller, N. Agmon, and J. S. Han, Designer eukaryotic chromosome, *Science* **344**, 55 (2014).
- [58] E. Del Grosso, E. Franco, L. J. Prins, and F. Ricci, Dissipative DNA nanotechnology, *Nat. Chem.* **14**, 600 (2022).
- [59] L. Heinen and A. Walther, Programmable dynamic steady states in ATP-driven nonequilibrium DNA systems, *Sci. Adv.* **5**, 32 (2019).
- [60] Our source code is available at <https://git.ecdf.ed.ac.uk/taplab/dna-ligation.git>.