Article

On the Growth Rate of Non-Enzymatic Molecular Replicators

Harold Fellermann [1,2,⋆] **and Steen Rasmussen** [1,3]

[1] Center for Fundamental Living Technology, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark

[2] ICREA-Complex Systems Lab, Universitat Pompeu Fabra (GRIB), Dr Aiguader 80, 08003 Barcelona, Spain

[3] Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA

[⋆] Author to whom correspondence should be addressed; E-Mail: harold@ifk.sdu.dk.

Abstract: It is well known that non-enzymatic template directed molecular replicators $X + nO \xrightarrow{k} 2X$ exhibit parabolic growth $d[X]/dt \propto k[X]^{1/2}$. Here, we analyze the dependence of the effective replication rate constant k on hybridization energies, temperature, strand length, and sequence composition. First we derive analytical criteria for the replication rate k based on simple thermodynamic arguments. Second we present a Brownian dynamics model for oligonucleotides that allows us to simulate their diffusion and hybridization behavior. The simulation is used to generate and analyze the effect of strand length, temperature, and to some extent sequence composition, on the hybridization rates and the resulting optimal overall rate constant k. Combining the two approaches allows us to semi-analytically depict a replication rate landscape for template directed replicators. The results indicate a clear replication advantage for longer strands at lower temperatures in the regime where the ligation rate is rate limiting. Further the results indicate the existence of an optimal replication rate at the boundary between the two regimes where the ligation rate and the dehybridization rates are rate limiting.
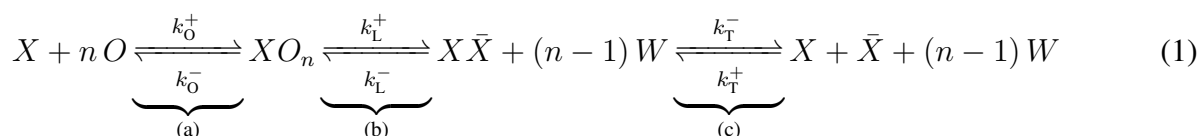
1. Introduction

Optimizing the yield of non-enzymatically self-replicating biopolymers is of great interest for many basic science and application areas. Clearly, early organisms could not emerge with a fully developed enzymatic gene replication machinery, so it is plausible that the first organisms had to rely on non-enzymatic replication [1–3]. Most bottom-up protocell models also rely on non-enzymatic biopolymer replication [4–9], which is also true for a variety of prospective molecular computing and manufacturing applications (for example, see [10,11]). Common for all of these research areas is the interest to obtain an optimal replication yield in the absence of modern enzymes. Depending on the details the biopolymer can be deoxyribonucleic acid (DNA), ribonucleic acid (RNA), peptide nucleic acid (PNA), *etc*. In the following we'll refer to them as XNA. In this study we investigate the detailed replication kinetics of well defined and relatively short XNA oligomers of interest to laboratory assembly of protocells as well as molecular computing and manufacturing applications.

Conceptually, XNA replication proceeds in three basic steps: (a) association, or *hybridization*, of n nucleotide monomers or oligomers with a single stranded, complementary template; (b) formation of covalent bonds in a condensation reaction, called *polymerization* in case of monomer condensation and *ligation* in case of oligomers; and finally (c) dissociation, or *dehybridization*, of the newly formed complementary strand:

$$X + n\,O \underbrace{\overset{k_{\mathrm{O}}^{+}}{\underset{k_{\mathrm{O}}^{-}}{\rightleftharpoons}}}_{(a)} XO_n \underbrace{\overset{k_{\mathrm{L}}^{+}}{\underset{k_{\mathrm{L}}^{-}}{\rightleftharpoons}}}_{(b)} X\bar{X} + (n-1)\,W \underbrace{\overset{k_{\mathrm{T}}^{-}}{\underset{k_{\mathrm{T}}^{+}}{\rightleftharpoons}}}_{(c)} X + \bar{X} + (n-1)\,W \tag{1}$$

Here, X and $\bar{X}$ denote the template strand and its complement, O denotes monomers/oligomers, W is the leaving group of the condensation reaction, and

$$K_i = \frac{k_i^{+}}{k_i^{-}} = e^{-\Delta G_i / k_{\mathrm{B}} T} \quad i \in \{\mathrm{O}, \mathrm{L}, \mathrm{T}\} \tag{2}$$

are the equilibrium constants of the three reactions. Note that the left hand transition of reaction scheme (1) is an abbreviation of a multi-step process that accounts for all n individual oligomer hybridizations and dehybridizations, which is only partly captured by the net reaction process.

The covalent condensation reaction is entirely activation limited. For nucleotide monophosphates, the leaving group corresponds to water which (due to its high concentration in aqueous solution) pushes the equilibrium to the hydrolyzed state. Product yields are significantly increased when using activated nucleotides, such as nucleotide triphosphate or imidazole.

Due to its complex reaction mechanism, non-enzymatic XNA replication from monomers suffers from various complications, namely inefficient extension of sequences containing consecutive adenine and thymine nucleotides [12,13], as well as side reactions such as partial replication and random strand elongation [14]. Consequently, experiments with activated nucleotides typically show little yield in aqueous solution, although results can be enhanced by employing surfaces (e.g., clay) or up-concentration in water-ice [2,15].
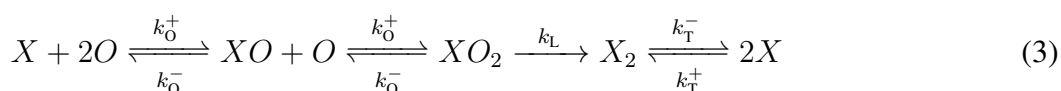
Replication from short activated oligomers, on the other hand, does produce high yields for both RNA and DNA [16–20] (and references therein). In particular, this observation has lead to the development of

minimal replicator systems, in which ligation of two oligomers is sufficient to form the complementary replica (see Figure 1). One of the reasons why these systems outperform replicators that draw from monomers is that the above side reactions expectedly occur, if at all, only to a negligible extent.

Figure 1. Minimal template directed replicator: two complementary oligomers hybridize to a template strand (upper part). An irreversible ligation reaction transforms the oligomers into the complementary copy of the template. The newly obtained double strand can dehybridize (lower part) thus allowing for iteration of the process. We assume that ligation is rate limiting, which implies that hybridization and dehybridization are in local equilibrium.
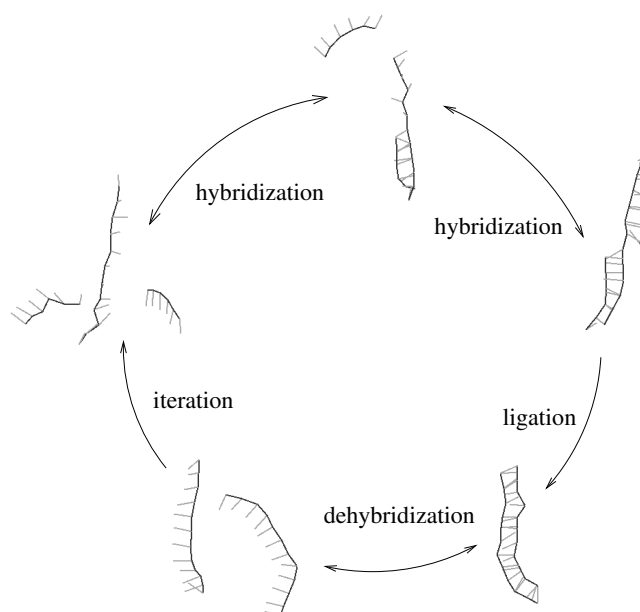
Neglecting both the production of waste as well as the hydrolysis of the ligation product, but explicitly taking into account the individual oligomer associations, minimal replicator systems (here for the case of a self-complementary template) can be written as

$$X + 2O \xrightleftharpoons[k_{\mathrm{O}}^-]{k_{\mathrm{O}}^+} XO + O \xrightleftharpoons[k_{\mathrm{O}}^-]{k_{\mathrm{O}}^+} XO_2 \xrightarrow{k_{\mathrm{L}}} X_2 \xrightleftharpoons[k_{\mathrm{T}}^+]{k_{\mathrm{T}}^-} 2X \tag{3}$$

where we have introduced the shorthand notation $k_{\mathrm{L}} = k_{\mathrm{L}}^+$.

In this article, we first develop a theoretical expression for the template directed replication rate of minimal replicator systems as a function of strand length and temperature. This analytical model provides transparent physical relations for how temperature, strand length and composition impact the overall replication rate. We then present a 3D, implicit solvent, constrained Brownian Dynamics model for short nucleotide strands, *i.e.*, strands with negligible secondary structures. The model does not attempt to be (quantitatively) predictive. In particular, we do not attempt to calibrate interaction parameters to experimental data, which prevents any sequence prediction. On the contrary, it is our aim to demonstrate that many of the replication properties of oligonucleotides arise from rather general statistical physics. The simulation is used to measure diffusion coefficients, effective reaction radii, and hybridization rates and their dependence on temperature, strand length, and, to some extent, sequence

information. This allows us to qualitatively obtain equilibrium constants K_O and K_T and to qualitatively sketch the effective replication rate k as a function of strand length and temperature. Our analysis focuses on minimal replicator systems in the context of chemical replication experiments as employed in protocell research and manufacturing applications, where the researcher controls reactant concentrations as well as most experimental parameters. However, we also discuss the impact of our findings in the context of origin of life research, were possible side reactions cannot be neglected.

2. Parabolic Growth and Replication Rate

Following and extending the derivation of [21,22], we assume that ligation is the rate limiting step. This translates into the following conditions for the rate constants:

$$k_L[XO_2] \ll k_O^+[X][O] \qquad k_L[XO_2] \ll k_O^+[XO][O] \qquad k_L[XO_2] \ll k_T^+[X]^2$$
$$k_L[XO_2] \ll k_O^-[XO] \qquad k_L[XO_2] \ll k_O^-[XO_2] \qquad k_L[XO_2] \ll k_T^-[X_2] \tag{4}$$

One can then assume a steady state of the hybridization/dehybridization reactions and express the total template concentration $[X]_{total} = [X] + [XO] + [XO_2] + 2[X_2]$ in terms of equilibrium constants as

$$[X]_{total} = \left(1 + K_O[O] + K_O^2[O]^2\right)[X] + 2K_T[X]^2 \tag{5}$$

When solved for $[X]$, this gives

$$[X] = \frac{1}{4K_T}\sqrt{8K_T[X]_{total} + (1 + K_O[O] + K_O^2[O]^2)^2} - \frac{1 + K_O[O] + K_O^2[O]^2}{4K_T} \tag{6}$$

Template directed replication typically suffers from product inhibition, where most templates are in double strand configuration, *i.e.*, $K_T[X]_{total} \gg 1$. Over the course of the reaction, this is tantamount of saying that $\sqrt{8K_T[X]_{total}} \gg 1 + K_O[O] + K_O^2[O]^2$. This allows us to approximate

$$\sqrt{8K_T[X]_{total} + (1 + K_O[O] + K_O^2[O]^2)^2} = \sqrt{8K_T[X]_{total}} + \sqrt{(1 + K_O[O] + K_O^2[O]^2)^2} + \mathcal{O}([X]_{total})$$

and simplify (6) to

$$[X] = \sqrt{\frac{[X]_{total}}{2K_T}} + \mathcal{O}([X]_{total}) \tag{7}$$

This is a lower bound of the single strand concentration, which is approached in the limit of vanishing oligomer concentration. By combining (3) and (7), we get

$$\frac{d[X]_{total}}{dt} = k_L[XO_2] = k_L K_O^2[O^2][X] \approx k\,[O]^2\sqrt{[X]_{total}} \tag{8}$$

with

$$k = k_L \frac{K_O^2}{\sqrt{2K_T}} \tag{9}$$

This well-established parabolic growth law is known to qualitatively alter evolutionary dynamics of XNA based minimal replicators and to promote coexistence of replicators rather than selection of the fittest [23–25].

Consequently, several strategies have been designed to overcome product inhibition in order to reestablish Darwinian evolution and survival of *only* the fittest [20,22,26,27]. Most of these approaches

hinge on a mechanism to lower the hybridization tendency of the product to the template. In this article, however, we accept parabolic growth and instead focus on the effective growth rate.

The key observation of Equation (9) is that, due to the steady state assumption, the overall growth rate is independent of the individual association and dissociation rates k_i^+, k_i^-, but only depends on the equilibrium constants K_O and K_T. Expressed in free energy changes, Equation (9) becomes

$$k = \exp\left[\log A + \left(\frac{1}{2}\Delta G_T - 2\Delta G_O - \Delta G_L^{\ddagger}\right)/k_B T\right] \tag{10}$$

where A and $\Delta G_L^{\ddagger}$ are the pre-exponential factor and activation energy of the ligation reaction, respectively, and we have used the Arrhenius equation

$$k_L = A e^{-\Delta G_L^{\ddagger}/k_B T} \tag{11}$$

We further observe that any potential optimum of (9) must obey

$$2k_L{}' K_T K_O = \frac{1}{2} k_L K_O K_T{}' - 4k_L K_T K_O' \tag{12}$$

where the prime indicates derivation with respect to any variable. Note that derivatives of k_L, K_T, and K_O can be taken with respect to parameters such as temperature and template length, whereas the notion of a derivative in sequence space is ill-defined. Therefore, Equation (12) can only give us partial information about an optimal growth rate.

It is well-known that the equilibrium constants K_O and K_T depend on various parameters such as temperature, salt concentration, strand length, and sequence information—all being relevant control parameters when designing replication experiments or delimiting origin of life conditions [28,29]. Furthermore, the two rates are interdependent as one expects K_T to rise with increasing K_O.

Qualitatively, the free energy of XNA hybridization obeys a form given by

$$\Delta G(N, T) = N\Delta G_{\text{base}} + \Delta G_{\text{init}} = N(\Delta H_{\text{base}} - T\Delta S_{\text{base}}) + \Delta H_{\text{init}} - T\Delta S_{\text{init}} \tag{13}$$

where N signifies the strand length, ΔG_{base} is the (maximal) energy change per base, ΔG_{init} is the initiation energy and $\Delta H_{\text{base}}, \Delta S_{\text{base}}$ are negative, whereas $\Delta H_{\text{init}}, \Delta S_{\text{init}}$ are positive. The right hand side of the equation expresses a saturation in the free energy per base as a function of the strand length; the free energy gain for each base pairing asymptotically becomes constant for long strands [30]. Inserting (13) into (10) and separating out the rate constant for the ligation reaction k_L, we obtain:

$$\frac{K_O^2}{\sqrt{2K_T}} \propto \exp\left[\left(\frac{1}{2}\Delta G(N, T) - 2\Delta G(N/2, T)\right)/k_B T\right]$$
$$\propto \exp\left[-\left(\frac{1}{2}N\Delta G_{\text{base}} + \frac{3}{2}\Delta G_{\text{init}}\right)/k_B T\right] \tag{14}$$

which, when differentiated for T, yields a positive dependence on temperature, iff

$$\frac{d}{dT}\frac{K_O^2}{\sqrt{2K_T}} > 0 \quad \Longleftrightarrow \quad N < -\frac{3\Delta H_{\text{init}}}{\Delta H_{\text{base}}}$$

Since $\Delta H_{\text{init}} > 0$, and $\Delta H_{\text{base}} \le 0$, this critical strand length is truly positive. It might surprise that $K_O^2/\sqrt{2K_T}$ can increase with decreasing temperature—the regime where templates are primarily

inhibited by the product. The results become understandable when considering that oligomers, with their lower hybridization rate, barely associate with the template if the temperature is raised.

Reintroducing the ligation reaction, this relation gets refined to

$$k = k_{\mathrm{L}} \frac{K_{\mathrm{O}}^2}{\sqrt{2K_{\mathrm{T}}}} = \exp\left[\log A - \left(\frac{1}{2}N\Delta G_{\mathrm{base}} + \frac{3}{2}\Delta G_{\mathrm{init}} + \Delta G_{\mathrm{L}}^{\ddagger}\right) / k_{\mathrm{B}}T\right] \tag{15}$$

with the critical strand length

$$\frac{dk}{d\mathrm{T}} > 0 \quad\Longleftrightarrow\quad N < N^* = -\frac{3\Delta H_{\mathrm{init}} + 2\Delta H_{\mathrm{L}}^{\ddagger}}{\Delta H_{\mathrm{base}}} \tag{16}$$

In words: we can identify a critical strand length N^* above which the overall replication rate k increases with decreasing temperature. This critical strand length is determined by the hybridization enthalpies $\Delta H_{\mathrm{base}}, \Delta H_{\mathrm{init}}$, and activation enthalpy change $\Delta H_{\mathrm{L}}^{\ddagger}$ of ligation.

Figure 2 depicts the graph of the replication rate landscape (15) that clearly identifies the optimum of Equation (12) as a saddle point. The corresponding temperature T^* where k changes its scaling with respect to strand length is—independent of the ligation reaction—given by

$$\frac{dk}{d\mathrm{N}} > 0 \quad\Longleftrightarrow\quad T < T^* = \frac{\Delta H_{\mathrm{base}}}{\Delta S_{\mathrm{base}}} \tag{17}$$

Figure 2. Effective replication rate k (given by Equation 15) as a function of strand length and temperature. For strands below a critical length N^* (here 10) the rate increases with temperature, for strands longer than N^*, the replication rate grows with decreasing temperature. The value of N^* is determined through Equation (16). Note the saddle point of the surface where T^* and N^* intersect (Equation 12) ($\Delta H_{\mathrm{base}} = -1.5k_{\mathrm{B}}T'$, $\Delta S_{\mathrm{base}} = -1k_{\mathrm{B}}$, $\Delta H_{\mathrm{init}} = 0.50k_{\mathrm{B}}T'$, $\Delta S_{\mathrm{init}} = 1.25k_{\mathrm{B}}$, $\Delta H_{\mathrm{L}}^{\ddagger} = 5.25k_{\mathrm{B}}T'$, $A = 10^3$).
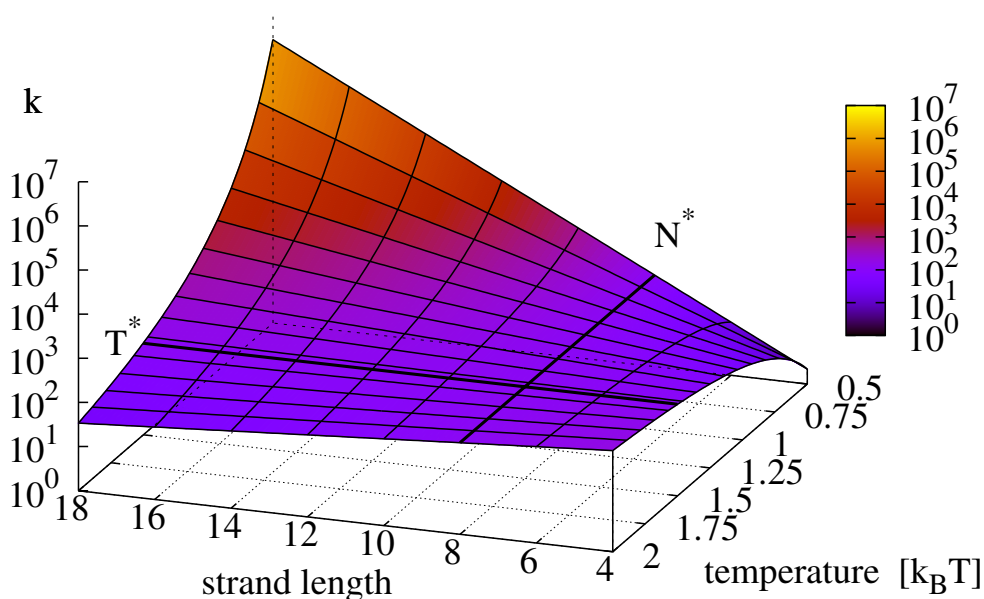
Can we obtain a higher replication rate by using non-symmetric oligomers? The rational behind this strategy is to increase the binding affinity of one oligomer to maybe decrease product inhibition. A simple refinement of Equation (14) allows us to capture this approach with our model:

$$\frac{K_{O_1} K_{O_2}}{\sqrt{2K_T}} \propto \exp\left[\left(\frac{1}{2}\Delta G(N,T) - \Delta G(N_1,T) - \Delta G(N_2,T)\right)/k_B T\right]$$

$$\propto \exp\left[\left(\left(\frac{1}{2}N - N_1 - N_2\right)\Delta G_{\text{base}} - \frac{3}{2}\Delta G_{\text{init}}\right)/k_B T\right] \tag{18}$$

where $N_1 + N_2 = N$ denote the lengths of oligomer strands O_1 and O_2. Thus, according to our simple thermodynamic considerations, non-symmetric variants of the replication process do not show more yield than the corresponding symmetric system: the binding affinity gained for the long oligomer strand is paid to hybridize the short oligomer strand.

Figure 2 seemingly implies that replication rates grow beyond any limit for long templates, which is unphysical. To resolve this inconsistency, it is important to remember that our findings are only valid in the regime where ligation is rate limiting. For very long XNA strands, however, double strands are so stable that dehybridization of the ligation product is expected to become the rate limiting step. Independent of the exact shape of the growth law, the dominant factor of the effective growth rate is given by

$$k_T^- = k^+ \exp\left[(N\Delta G_{\text{base}} + \Delta G_{\text{init}})/k_B T\right] \tag{19}$$

where k^+ summarizes both pathways of either product rehybridization or hybridization of oligomers followed by ligation. As k^+ is composed of hybridization (*i.e.*, diffusion plus orientational alignment) and ligation events, it varies only slightly with sequence length when compared to dehybridization rates for the case of large N. Therefore, the effective replication rate will be governed by the scaling

$$k \propto \exp(N\Delta G_{\text{base}}/k_B T) \tag{20}$$

with the limit

$$\lim_{N\to\infty} k = 0$$

since $\Delta G_{\text{base}} < 0$. As a consequence, we expect a full non-equilibrium study of the replication process to show a proper maximum in the replication rate as a function of strand length.

3. Spatially Resolved Replicator Model

Spatially resolved template-directed replicators have been previously simulated in the Artificial Life community using two-dimensional cellular automata and continuous virtual physics [14,31,32]. The model we present here is conceptually similar to, but simpler than other coarse-grained DNA models, e.g., [33–35]. Compared to our earlier work on hybridization and ligation [36], the model presented here is less computationally expensive while simultaneously being broader in its range of application.

We model nucleic acid strands as chains of hard spheres that are connected by rigid bonds. Each sphere has mass m, radius r, position and velocity $(\mathbf{x}_i, \mathbf{v}_i) \in \mathbb{R}^{3\times 3}$, as well as moment of inertia θ, orientation and angular momentum $(\boldsymbol{\omega}_i, \mathbf{L}_i) \in \mathbb{S}^2 \times \mathbb{R}^3$ representing the spatial orientation of the respective nucleotide. Further, each sphere has a type $t_i \in \{A, B, C, D\}$, and we define A and B (C and

D) to be complementary. The model is implicit in the sense that solvent molecules are not represented explicitly, but only through their effect on the nucleotide strands. We model the (translational and rotational) motion of each sphere by a *Langevin equation*

$$\dot{\mathbf{x}}_i = \mathbf{v}_i \tag{21}$$

$$m\,\dot{\mathbf{v}}_i = -\boldsymbol{\nabla}U_i(\mathbf{x},\boldsymbol{\omega}) - \gamma\mathbf{v}_i + \boldsymbol{\xi}_i \tag{22}$$

$$\dot{\boldsymbol{\omega}}_i = \frac{1}{\theta}\,\mathbf{L}_i \times \boldsymbol{\omega}_i \tag{23}$$

$$\dot{\mathbf{L}}_i = -\boldsymbol{\nabla}\hat{U}_i(\mathbf{x},\boldsymbol{\omega}) - \gamma\frac{\mathbf{L}_i}{\theta} + \hat{\boldsymbol{\xi}}_i \tag{24}$$

Here, γ is the friction coefficient, and $\boldsymbol{\xi}, \hat{\boldsymbol{\xi}}$ are zero mean random variables accounting for thermal fluctuations. Together, friction and thermal noise act as a thermostat: they equilibrate the kinetic energy with an external heat bath whose temperature is given by the following *fluctuation-dissipation-theorem* [37]:

$$\langle \boldsymbol{\xi}_i(t); \boldsymbol{\xi}_j(t') \rangle = 2\gamma k_{\mathrm{B}}T\,\delta_{ij}\delta(t-t') \tag{25}$$

$$\left\langle \hat{\boldsymbol{\xi}}_i(t); \hat{\boldsymbol{\xi}}_j(t') \right\rangle = 2\frac{\gamma}{\theta}k_{\mathrm{B}}T\,\delta_{ij}\delta(t-t') \tag{26}$$

Hence, a temperature change directly translates into a change of the Brownian noise amplitude. We use the moment of inertia for solid spheres $\theta = \frac{2}{5}mr^2$—noting that one could, in principle, use moment of inertia tensors to reflect the geometry of the individual nucleobases.

Equations (21)–(24) are solved under the constraints

$$|\mathbf{x}_i - \mathbf{x}_j| = r_{\mathrm{bond}} \quad \text{if } i, j \text{ bonded} \tag{27}$$

$$|\mathbf{x}_i - \mathbf{x}_j| = 2r \quad \text{if } |\mathbf{x}_i - \mathbf{x}_j| < 2r \text{ (and } i, j \text{ not bonded)} \tag{28}$$

$$|\boldsymbol{\omega}_i| = 1 \tag{29}$$

to account for rigid bonds (27) and hard spheres (28). By setting $r_{\mathrm{bond}} < 2r$, we can assert that strands do not penetrate each other. We define the following angles (see Figure 3):

$$\cos\theta_i = \left\langle \frac{\mathbf{x}_j - \mathbf{x}_i}{r_{\mathrm{bond}}} \cdot \frac{\mathbf{x}_k - \mathbf{x}_i}{r_{\mathrm{bond}}} \right\rangle \quad i, j \text{ and } i, k \text{ bonded}$$

$$\cos\phi_{ij} = \left\langle \frac{\mathbf{x}_j - \mathbf{x}_i}{r_{\mathrm{bond}}} \cdot \boldsymbol{\omega}_i \right\rangle \quad i, j \text{ bonded}$$

$$\cos\omega_{ij} = \left\langle \boldsymbol{\omega}_i \cdot \boldsymbol{\omega}_j \right\rangle \quad i, j \text{ bonded}$$

$$\cos\psi_{ij} = \left\langle \frac{\mathbf{x}_j - \mathbf{x}_i}{|\mathbf{x}_i - \mathbf{x}_j|} \cdot \boldsymbol{\omega}_i \right\rangle \quad i, j \text{ not bonded}$$

As much of the molecular geometry is already determined through the constraints, the innermolecular potentials U and $\hat{U}$ only need to account for strand stiffness (30), base orientation (31), and π-stacking (32). We set:

$$U_i^{\mathrm{bend}} = a_{\mathrm{bend}}\left(\frac{\theta_i}{\pi} - 1\right)^2 \quad \text{if } i \text{ not terminal, otherwise } 0 \tag{30}$$

$$\hat{U}_{ij}^{\mathrm{ortho}} = \hat{a}_{\mathrm{ortho}}\left(\frac{\phi_{ij}}{\pi} - \frac{1}{2}\right)^2 \tag{31}$$

$$\hat{U}_{ij}^{\mathrm{parallel}} = \hat{a}_{\mathrm{parallel}}\left(\frac{\omega_{ij}}{\pi} - 0\right)^2 \tag{32}$$

The minimum energy state of these definitions are stretched out nucleotide strands with orientations perpendicular to the strand and parallel to each other.

Figure 3. Geometry of the nucleotide strands. The figure shows the angles that define inner- and intermolecular interactions for one nucleobase (shaded in grey).
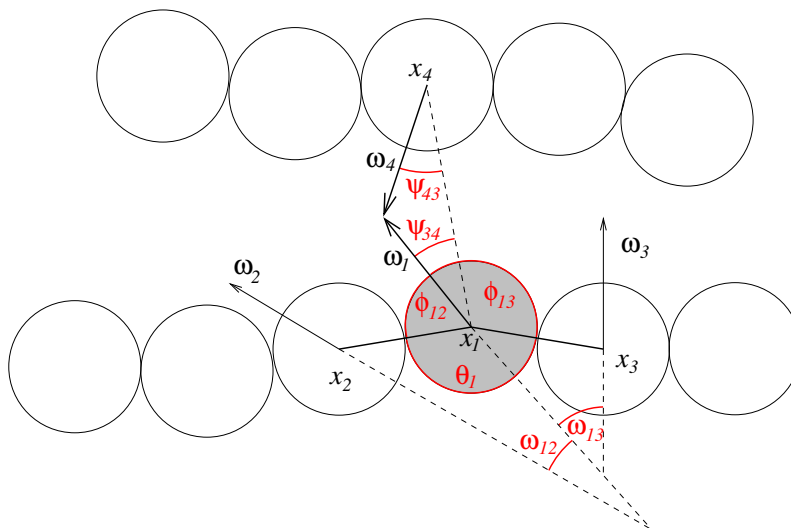
In addition, we define the following intermolecular potentials between non-bonded complementary nucleobases i and j:

$$U_{ij}^{\text{hybrid}} = -a_{\text{hybrid}}\, d\left(|\mathbf{x}_i - \mathbf{x}_j|\right) \cos\psi_{ij} \quad \text{if } |\mathbf{x}_i - \mathbf{x}_j| < r_{\text{c}} \tag{33}$$

$$\hat{U}_{ij}^{\text{hybrid}} = -\hat{a}_{\text{hybrid}}\, d\left(|\mathbf{x}_i - \mathbf{x}_j|\right)\left(\frac{\psi_{ij}}{\pi} - 1\right)^2 \quad \text{if } |\mathbf{x}_i - \mathbf{x}_j| < r_{\text{c}} \tag{34}$$

The shift and weighing function

$$d(r_{ij}) = \frac{1}{2}\left[\cos\left(\frac{r_{ij} - 2r}{r_{\text{c}} - 2r}\pi\right) + 1\right]$$

asserts that the potentials take on a minimum at particle contact and level out to zero at the force cutoff radius r_{c}. Equation (33) allows for a nucleobase i to attract its complement j along the direction of $\boldsymbol{\omega}_i$, while (34) orients $\boldsymbol{\omega}_i$ toward the complement.

Taking the above potentials together, we define

$$U_i(\mathbf{x}, \boldsymbol{\omega}) = U_i^{\text{bend}}(\mathbf{x}) + \sum_{\substack{i,j \\ \text{non-bonded} \\ \text{complementary}}} U_{ij}^{\text{hybrid}}(\mathbf{x}, \boldsymbol{\omega}) \tag{35}$$

$$\hat{U}_i(\mathbf{x}, \boldsymbol{\omega}) = \sum_{\substack{i,j \\ \text{bonded}}} \left(\hat{U}_{ij}^{\text{ortho}}(\mathbf{x}, \boldsymbol{\omega}) + \hat{U}_{ij}^{\text{parallel}}(\boldsymbol{\omega})\right) + \sum_{\substack{i,j \\ \text{non-bonded} \\ \text{complementary}}} \hat{U}_{ij}^{\text{hybrid}}(\mathbf{x}, \boldsymbol{\omega}) \tag{36}$$

Equations (21) to (29) are numerically integrated using a Velocity Verlet algorithm that, in each iteration, first computes unconstrained coordinates which are afterwards corrected with a SHAKE algorithm to satisfy the constraints [38,39]. Typical system configurations are shown in Figure 1.

4. Simulation Results

In the subsequent analyses, we will employ reduced units, *i.e.*, $m = 1$, $r_{\mathrm{c}} = 1$, and $k_{\mathrm{B}}T_0 = 1$ define the units of mass, length, and energy. From this, the natural unit of time follows as

$$\tau = r\sqrt{m/k_{\mathrm{B}}T_0}$$

The parameters r and r_{bond} are chosen to prevent crossing of strands ($r_{\mathrm{bond}} < 2r$). The ratio r_{bond}/r determines the double strand geometry which is modeled more sparse than in actual nucleic acid strands in order to compensate for the relatively shallow potentials of the coarse-grained model. The ratio r/r_{c} determines the distance at which complementary bases "feel each other" and has been set to two times the bead diameter.

Assuming a reference temperature $T_0 \equiv 37\ ^{\circ}C$, we set a_{bend} to loosely match the persistence length of about $p = 40$ Å for single stranded DNA [29,40]:

$$a_{\mathrm{bend}} = \frac{p}{2r_{\mathrm{bond}}}k_{\mathrm{B}}T = \frac{40\text{Å}}{2 \cdot 4.3\text{Å}}k_{\mathrm{B}}T = 4.65\ k_{\mathrm{B}}T \approx 5\ k_{\mathrm{B}}T$$

The parameters $\hat{a}_{\mathrm{ortho}}$, $\hat{a}_{\mathrm{parallel}}$ of the potential functions are chosen in order to promote stacking of single strands for temperatures up to at least $3k_{\mathrm{B}}T_0$ [29]. The parameter a_{hybrid} is on the order of magnitude of base pair interactions averaged from the interaction parameters in nearest-neighbor models [41]:

$$a_{\mathrm{hybrid}} = -\frac{1}{2}\langle \Delta G_{ij} \rangle = 1.13\ k_{\mathrm{B}}T \approx 1\ k_{\mathrm{B}}T$$

We point out that our model utilizes a high value $\hat{a}_{\mathrm{hybrid}}$ in order to promote fast hybridization. A list of all model parameters (unless otherwise noted) is given in Table 1.

Table 1. Model parameters in reduced units (unless otherwise noted).

Parameter	Value	Comment	Equations
m	1	particle mass	(22)–(24)
γ	3	friction coefficient	(22), (24)
$k_{\mathrm{B}}T_0$	1	equilibrium temperature	(25), (26)
Δt	0.05	numerical time step	
r	0.25	particle radius	(28)
r_{bond}	0.45	bond length	(27)
r_{c}	1	force cutoff radius	(33)–(34)
a_{bend}	5	strand stiffness	(30)
$\hat{a}_{\mathrm{ortho}}$	2.5	angular stretching	(31)
$\hat{a}_{\mathrm{parallel}}$	1	angular alignment	(32)
$\hat{a}_{\mathrm{hybrid}}$	10	angular hybridization	(34)
a_{hybrid}	1	complementary attraction	(33)

Note that it is not mandatory to relate one bead of the model to one physical nucleotide. Instead, each bead could also represent a short XNA subsequence (e.g., 2–4 nucleotides). While this would result in a

closer match of the ratio r_{bond}/r, the amplitudes of the potential functions would need to be adapted to reflect the changed representation.

4.1. Diffusion

In dilute solution, DNA diffusion depends primarily on temperature and strand length, as opposed to primary or secondary structure. In the limit of low Reynolds numbers, the diffusion coefficient of a sphere is given by the *Einstein-Stokes equation*

$$D = \frac{k_B T}{6\pi\eta r} \tag{37}$$

where η is the viscosity of the medium and r the radius of the sphere. For the Rouse chains implemented by our model, theory predicts a linear scaling of the effective Stokes radius with polymer length [42].

In order to compare our model polymer diffusion to (37), we perform simulations of single homopolymers (e.g., poly-C) and determine the diffusion coefficient from its measured mean square displacement

$$D = \frac{1}{6} \frac{|\mathbf{x}(\Delta t) - \mathbf{x}(0)|^2}{\Delta t}$$

Figure 4 shows results for strands of lengths $N = 1, \ldots, 10$ and temperatures $k_B T = 1, \ldots, 3$.

Figure 4. Diffusion coefficients measured for different strand lengths and temperatures (symbols) fitted to the prediction of the Einstein-Stokes relation (solid lines). For each parameter pair, 40 simulation runs over 1000τ have been averaged.
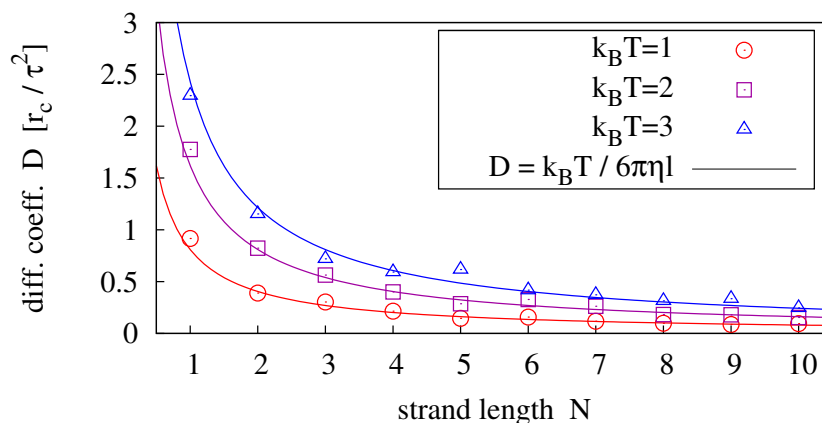
For the general scaling relation $r \propto N^\nu$ we obtain the most likely exponent from data fitting via ν and η as 1.06. By setting $\nu = 1$, and equivalently $r \propto N$, we obtain the best agreement between measurement and theory (by fitting via η only) for $\eta = 0.061 k_B T \tau^2 / r_c^2$ (see solid lines in Figure 4).

4.2. Radius of Gyration

Again in dilute solution, the radius of gyration

$$R_g^2 = \frac{1}{N-1} \sum_{i=1}^{N} |\mathbf{x}_i - \mathbf{x}_{mean}|^2$$

with $\mathbf{x}_{\mathrm{mean}}$ being the center of gravity of the chain, is expected to depend on chain length and temperature (or equally the backbone stiffness a_{bend}). As opposed to diffusion, we do expect the radius of gyration to change with sequence information, as it affects the secondary structure of the molecule. For homopolymers, we expect R_{g} to be well described by the Flory mean field model [42]

$$R_{\mathrm{g}} \propto (N-1)^{\nu}$$

We perform simulations of single homopolymers and self-complementary nucleotide strands and determine the radius of gyration. Figure 5 shows results for strands of lengths $N = 4, \ldots, 16$ and various backbone stiffness values. It is found that the Flory model is a good prediction, not only for homopolymers, but also for self-complementary strands. Expectedly, the radius of gyration is smaller for self-complementary strands. For $a_{\mathrm{bend}} = 0.2$, we find the radius of gyration of self-complementary strands to be slightly longer than the radius of gyration of a homopolymer with half the length—implying that the self-complementary strand is almost always in a hairpin configuration. For stronger backbone stiffness values, the effect is reduced.

Figure 5. Radius of gyration measured for different strand lengths and bending potentials (symbols) fitted to the prediction of the Flory mean field theory (solid lines). For each parameter pair, 40 simulation runs over 400τ have been averaged. The upper panel shows results for homopolymers (e.g., poly-C), the lower panel compares those to radii of self-complementary strands. The plots also show the boundaries for maximally stretched chains ($\nu = 1$—upper dotted line) and the expectation value of an ideal chain ($\nu = 3/5$—lower dotted line).
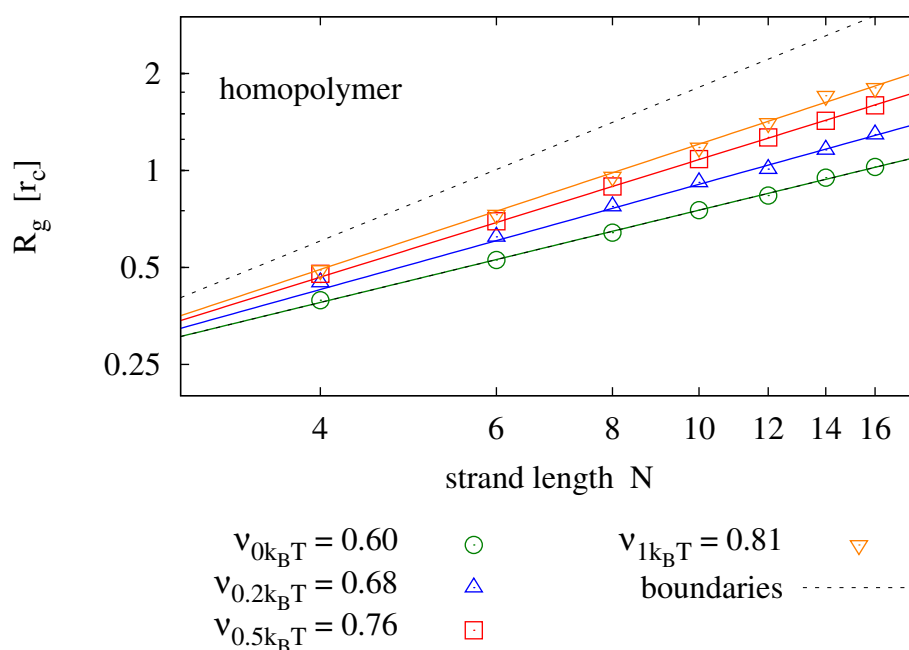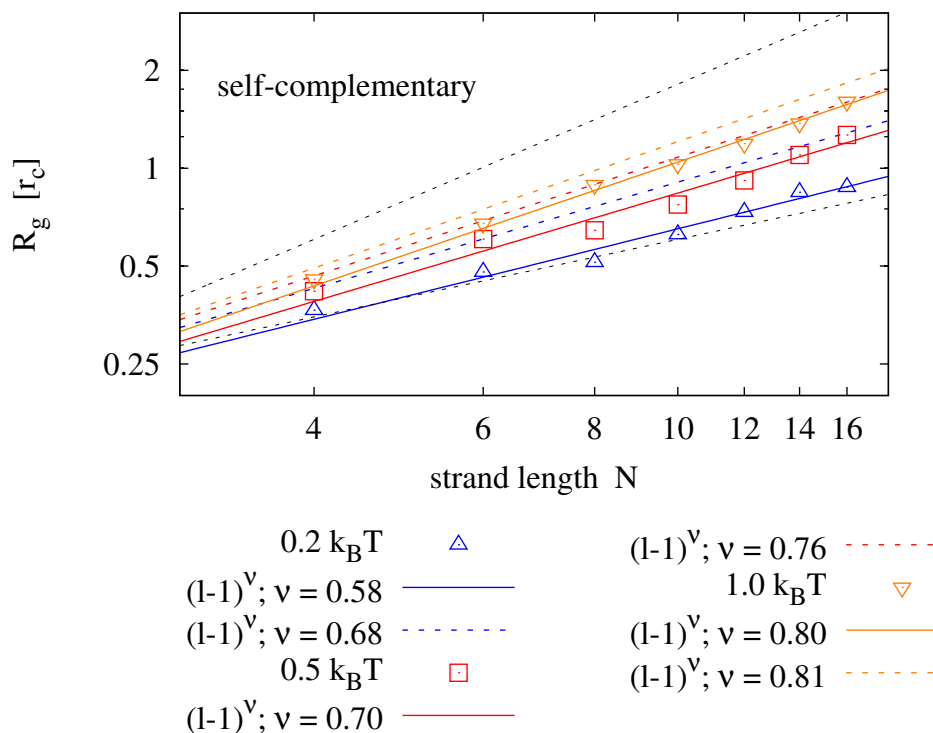
Figure 5. *Cont.*

4.3. Melting Behavior

We analyze the melting behaviour $[X]_2 \rightleftharpoons 2[X]$ of complementary nucleotide strands as a function of temperature for various strand lengths and sequences. We consider a base to be hybridized if there is a complementary base of another strand within a maximal distance of r_c. Denoting the fraction of hybridized nucleobases with $0 \leq \chi \leq 1$, we can compare the melting curves to the theoretical prediction

$$\chi(T) = \left(1 + e^{\frac{\Delta H - T \Delta S}{RT}}\right)^{-1} \tag{38}$$

where $\Delta H, \Delta S$ are constants depending on template length, sequence, and concentration.

Figure 6 shows melting curves for 18 different sequences and fits (via $\Delta H_i, \Delta S_i$) to the theoretical prediction, where each panel analyzes sequences that are subsequences of a common master sequence denoted in each panel. The graphs clearly show how the average hybridization increases with strand length for each master sequence. Inlays, where present, emphasize that the inverse of the melting point, at which $\chi(T_m) = 0.5$, scales linearly with the inverse of the strand length.

Comparing the individual panels to each other, we find that melting temperatures for strands of equal length are higher for sequences with identical adjacent bases. In fact, the melting behavior is dominated by the presence of identical adjacent bases: adding a single nucleobase to a strand that consists otherwise only of identical pairs (*i.e.*, moving from length 4 to 6 and from length 8 to 10 in panel two) has no significant impact on the observed melting temperature. We assume that this behavior is due to the fact that dehybridized nucleotides of a partly molten strand find more potential binding partners to enforce the stability of the partly molten strand, thereby promoting recombination of products.

Figure 6. Systems of size 10^3 are initialized with two complementary strands of length N. The sequence information is taken from the N central nucleotides of the master sequence denoted in each panel (e.g., $N = 6$ implies sequence CABACD in the first panel). Each system is simulated over 50000τ, and the average fraction χ of hybridized nucleobases is determined. Error bars show the average and standard deviation of 40 measurements. Solid lines show the theoretical prediction $\chi(T) = \left(1 + e^{\frac{\Delta H - T\Delta S}{RT}}\right)^{-1}$ fitted individually to each data set via ΔS and ΔH. Melting temperatures T_{m} are obtained from the relation $\chi(T_{\mathrm{m}}) = 0.5$, and their scaling as a function of strand length is depicted in the inlays for the cases where enough melting points had been observed.
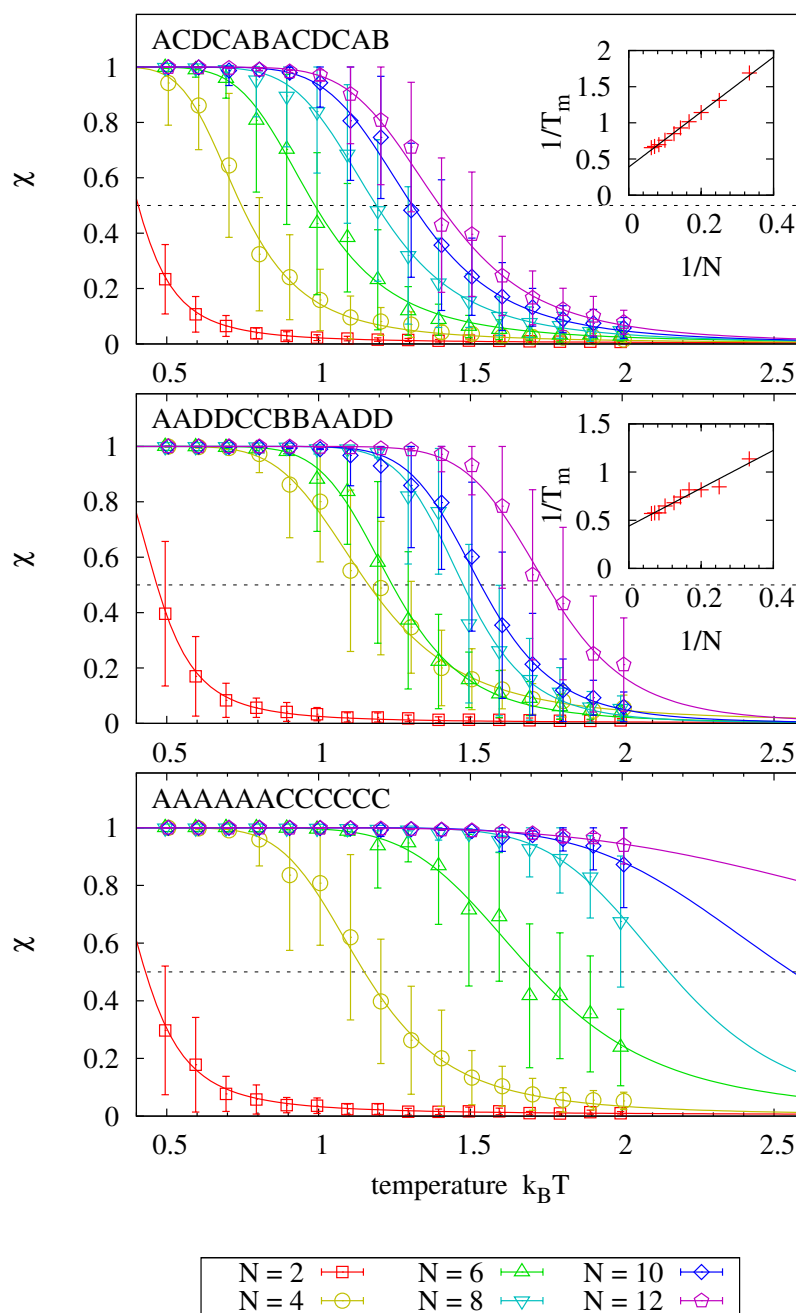
Experimentally obtained nearest neighbor interactions show indeed that there is a 40%–70% increase in the hybridization energy for AA and TT bases as compared to AT and TA pairs. However, adjacent CC/GG bases lower the melting energy by 15%–20% when compared to CG/GC pairs [41]. We suspect that this difference is rooted in a difference in stacking energies which is not sufficiently resolved in our parametrization model.

Up to now, we have analyzed hybridization of two complementary strands of equal length. How is the stability of the hybridization complex affected if one of the strands is replaced by two oligomers of half the length? We analyze the master sequence CBACBADDDDDD. Its left half is similar to the first sequence of Figure 6 with respect to non-identical neighboring bases. The right half has been chosen for its strong hybridization tendency. We run experiments as before and measure the hybridization of the left oligomer. By comparing its equilibrium rate to the one for two templates of half the length, we can determine how the dangling right hand side affects the equilibrium rate (e.g., we compare the hybridization of a 4-mer to an 8-mer template to the hybridization of two 4-mers.) Figure 7 shows that the hybridization fractions $\chi_O(N)$ and $\chi_T(N/2)$ are comparable for the analyzed sequence. We expect, however, that χ_O decreases when the two oligomers have more interaction possibilities than in the selected master sequence.

Figure 7. Melting curves for an oligomer that hybridizes to the left hand side of the master sequence in the presence of the right hand side oligomer. Data is obtained with the procedure described in Figure 6. For the analyzed master sequence, the results are comparable to those of two complementary strands of length $N/2$ (dotted lines).
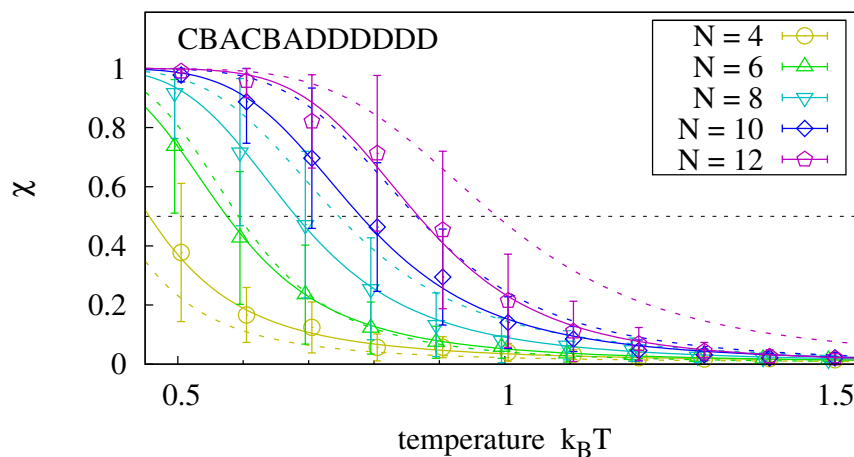
4.4. Effective Replication Rate

We can roughly equate

$$\chi \equiv \frac{2[X_2]}{[X]_{\text{total}}}, \qquad 1 - \chi \equiv \frac{[X]}{[X]_{\text{total}}}$$

and obtain an estimate for the equilibrium constant

$$K_T = \frac{[X_2]}{[X]^2} \equiv \frac{\chi}{2(1-\chi)^2} \frac{1}{[X]_{\text{total}}} \tag{39}$$

from the measurements. This equation has to be taken with some caution because the measured hybridization times reflect a non-trivial relation between diffusing reactants and rehybridization of partly molten complexes—both scaling differently with concentration. To truly obtain K, one is advised to repeat the simulations with varying concentrations, *i.e.*, box size. By means of Equation (39), we convert the melting data from Section 4.3 to obtain hybridization energy changes

$$\Delta G_{\mathrm{T}} = -k_{\mathrm{B}}T \log K_{\mathrm{T}}$$

$$N\Delta G_{\mathrm{base}} + \Delta H_{\mathrm{init}} - T\Delta S_{\mathrm{init}} = -k_{\mathrm{B}}T \log \frac{\chi}{2(1-\chi)^2} + k_{\mathrm{B}}T \log[X]_{\mathrm{total}}$$

In the latter equation, $-k_{\mathrm{B}} \log[X]_{\mathrm{total}}$ denotes the translational entropy for a box of size $[X]_{\mathrm{total}}^{-1}$, while ΔS_{init} accounts for the configurational entropy of a single strand. We combine both entropy terms, $\Delta S' = \Delta S_{\mathrm{init}} + k_{\mathrm{B}} \log[X]_{\mathrm{total}}^{-1}$, and fit

$$N\Delta G_{\mathrm{base}} + \Delta H_{\mathrm{init}} - T\Delta S' = -k_{\mathrm{B}}T \log \frac{\chi}{2(1-\chi)^2}$$

which allows for better comparison to the melting temperature plots, as $\Delta G_{\mathrm{T}} = 0 \Longleftrightarrow \chi = 1/2$.

Determining hybridization energies is difficult because hybridization is very stable at low temperatures, particularly for long XNA strands. Dehybridization then becomes a rare event, which requires unfeasibly long simulation runs in order to sample equilibrium distributions. Consequently, data for low temperatures and long strands is overshadowed by noise and we have excluded such data from the analysis. For the regime that is accessible to simulation, Figure 8 shows the measured hybridization energies fitted to the theoretical model of Equation (13)—see figure caption for details. The data follows the linear trend of the model and can recover the proper temperature scaling. However, we also observe deviations from the analytical prediction for $T > 1.9$. The plot confirms that hybridization energy changes are close to zero at the melting temperature of each double strand. For the simulation, where $[X]_{\mathrm{total}} = 0.001$, we obtain $\Delta S_{\mathrm{init}} = 1.33$, which confirms that ΔG_{init} is primarily entropic.

Figure 8. Hybridization energy changes ΔG_{T} obtained from the measurements of Section 4.3, sequence ACDCABACDCABACDCABAC (symbols), fitted to the analytical model of Equation (13) via the parameters $\Delta H_{\mathrm{base}} = -1.81$, $\Delta S = -0.756$, $\Delta H_{\mathrm{init}} = 0.470$, and $\Delta S' = -5.58$. Since $[X]_{\mathrm{total}} = 0.001$, we can estimate $\Delta S_{\mathrm{init}} = 1.33$.
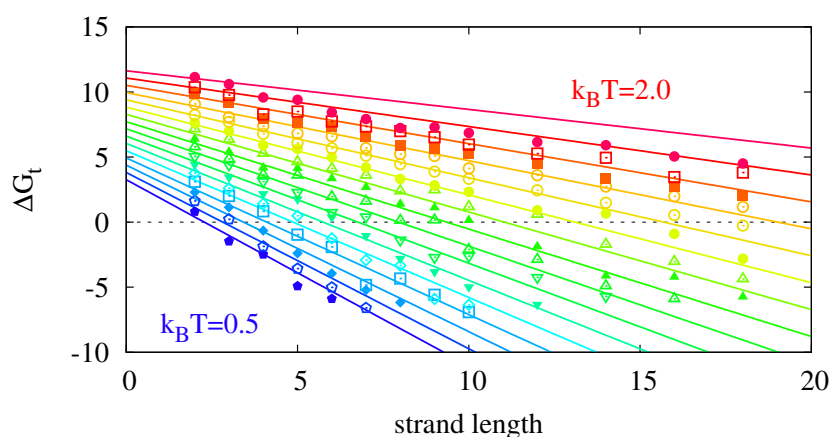
Unfortunately, the removal of noisy simulation results implies that we do not have measurements for the regime where the analytical model predicts the most features. To nevertheless obtain estimates for these points, we perform the same simulations as before but start with a perfectly hybridized complex. For short strands, the difference in initial conditions is negligible, as hybridization and dehybridization equilibrate within the simulated time span. For long strands / strands at low temperatures, the sampled χ values progressively overestimate the equilibrium time fraction of hybridization.

The rational behind these dehybridization measurements is the following: for long strands and low temperatures dehybridization of the ligation product becomes the rate limiting step and we are in the regime of Equation (20). Here, the effective replication rate is primarily governed by the rate of product dehybridization which in turn gives us an upper bound for the replication rate. By combining the results of the two scenarios, we implicitly relate the simulated time span with an assumed ligation rate.

Plugging the measured constants K_T and K_O into Equation (9), we obtain a replication rate landscape for minimal replicators which is depicted in Figure 9. The colored surface shows the effective equilibrium constant $K_O^2/\sqrt{2K_T}$, obtained from the original measurements. The grey surface shows the results from the dehybridization experiments. Finally, the replication rate landscape obtained via Equation (13) is shown as a mesh. For the analyzed master sequence and range of observation, the effective oligomer complex concentration $K_O^2/\sqrt{2K_T}$ varies over 13 orders of magnitude with highest rates for long strands ($N \geq 8$) and low temperatures ($k_B T \leq 0.8$).

Figure 9. Effective equilibrium constant $K_O^2/\sqrt{2K_T}$, obtained from the measurements of Figure 8 (colored) compared to the theoretical prediction of Equation (13) (mesh). Data shaded in gray is extrapolated from dehybridization experiments.
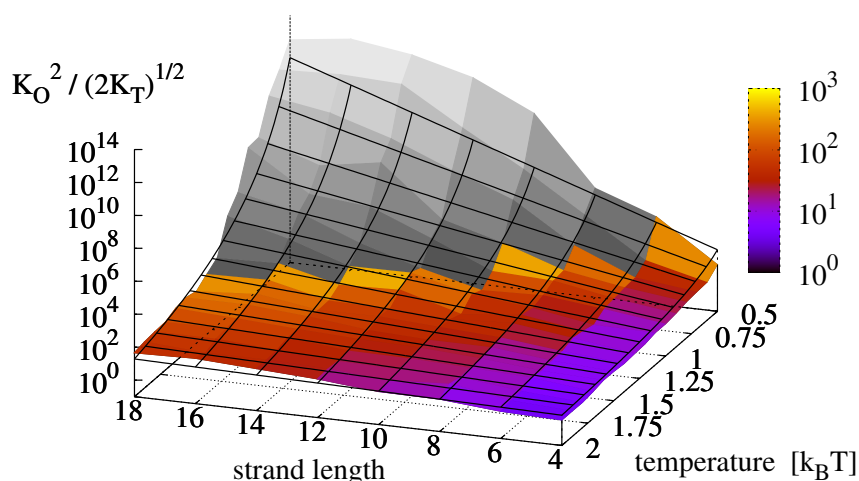

Contrary to the analytical derivations, the numerical results of the dehybridization experiments indicate a saturation and possibly a decrease of the replication rate for long strands at low temperatures, thereby supporting our hypothesis that the effective rate indeed possesses an optimum when dehybridization and ligation occur on comparable time scales.

The numerical simulations do not incorporate the ligation reaction. To include its temperature dependence, we superpose the Arrhenius Equation (11) with parameters as in Figure 2 onto Figure 9 and obtain the replication rate landscape shown in Figure 10. The resulting figure features a critical

strand length $N^* \approx 8$ at which the temperature scaling inverts. With the parameters obtained from the data fits, the critical temperature $T^* \approx 2.37\, T_0$ lies outside the analyzed area.

Figure 10. Final replication rate k as a function of template length and temperature. The figure is produced by superposing the data from Figure 9 with the Arrhenius equation for the ligation reaction following Equation (15) with $A = 10^3, \Delta H_{\mathrm{L}}^{\ddagger} = 6.52 k_{\mathrm{B}} T'$. For this parametrization, the critical strand length N^* above which the temperature dependence of the reaction inverts is 8.
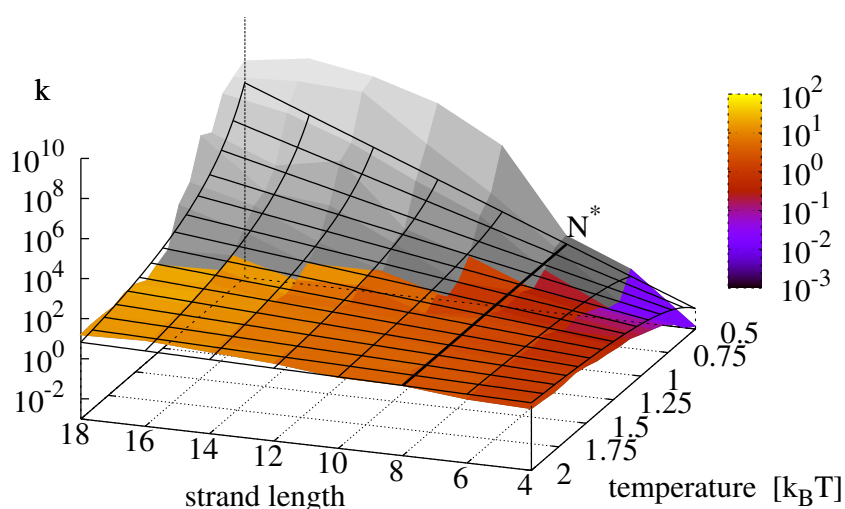
5. Discussion

The common strategy to increase the yield in template directed replication experiments is to increase the concentration of oligomers. This is certainly viable, and the fact that the growth rate k is proportional to the square of the oligomer concentration encourages this approach. Our investigation, however, indicates that oligomer concentration can be outweighed drastically by factors such as temperature, template length, as well as sequence information, which all influence the replication rate at least exponentially and thus over many orders of magnitude. These findings are consistent for the simple analytical expressions (Figure 2), for the simulations (Figure 8) as well as for the combined analysis (Figures 9 and 10).

Perhaps contrary to intuition, we find the highest growth rates for long replicators and low temperatures. This finding can be explained by the fact that the effective growth rate of minimal replicators features a critical strand length N^* at which the temperature dependence of the overall replication reaction inverts: below N^* the replication rate is dominated by the ligation reaction and its positive temperature scaling, whereas above N^*, the negative temperature scaling of the hybridization reactions becomes dominant, recall Figure 10.

We observe that hybridization rates are highly sequence dependent. In particular, our spatially resolved simulations reveal that adjacent identical nucleobases can drastically stabilize the hybridization complex. We expect that the overall replication process is primarily sequence specific near to the ligation sites, as it is known that mismatches near the ligation site effect the ligation the most [29].

We also find that there is no difference in the replication rates of symmetric versus asymmetric replicators. We see that from equation 18, where only the sum of the oligomer lengths appears: while the longer oligomer of an asymmetric replicator has a high binding affinity to the template and therefore promotes the formation of a hybridization complex, the short oligomer has a smaller binding affinity, such that the total asymmetric hybridization complex is as stable as its symmetric counterpart.

We emphasize that our approach hinges on the assumption that ligation is the rate limiting step of the replication reaction. Due to the temperature scaling of the diffusion, hybridization, and ligation processes, our approach breaks down for very low temperatures or very long template strands. As discussed in Section 2, Equation (20), for long strands and low temperatures the dehybridization of the templates becomes the rate limiting step. Exactly what happens in the transition region between these two limits requires a more detailed non-equilibrium analysis and is outside the scope of this investigation. The grey shaded area in Figures 9 and 10 depicts the expected landscape for the replication rate as we approach this transition zone from the regime where ligation is rate limiting, and it is clearly seen how the replication rate levels off as temperature decreases and the sequence length increases. In any event, we would expect the existence of a true optimal temperature for a given strand length, and equally a true optimal strand length for a given temperature, such that replication rates are maximized.

In the context of origin of life research, we cannot expect the presence of "clean" oligomer systems, as the ones we have used in current investigations. However, our findings clearly indicate that lower temperatures and longer strands have favorable replication rates under restricted conditions such as ligation limiting rate as well as single ligation events. Lower temperatures provide a qualitative mechanism for the preferred replication of longer information molecules, which may have implications for the "Snowball Earth" hypothesis.

Most importantly, in the context of minimal replicator experiments and applications, e.g., in protocell as well as molecular computing and fabrication research, our findings suggest a qualitative recipe for optimizing replication yields, as they relate experimentally accessible data such as melting temperatures and ligation rate to critical strand length (Equation 16) and temperature (Equation 17).

Acknowledgements

References and Notes

1. Gilbert, W. The RNA world. *Nature* **1986**, *319*, 618.
2. Monnard, P.A. The dawn of the RNA world: RNA Polymerization from monoribonucleotides under prebiotically plausible conditions. In *Prebiotic Evolution and Astrobiology*; Wong, J.T.F., Lazcano, A., Eds.; Landes Bioscience: Austin, TX, USA, 2008.

3. Cleaves, J.H. Prebiotic chemistry, the premordial replicator and modern protocells. In *Protocells: Bridging Nonliving and Living Matter*; Rasmussen, S., Bedau, M., Chen, L., Deamer, D., Krakauer, D., Packard, N., Stadler, P., Eds.; MIT Press: Cambridge, MA, USA, 2009; p. 583.

4. Rasmussen, S.; Chen, L.; Deamer, D.; Krakauer, D.C.; Packard, N.H.; Stadler, P.F.; Bedau, M.A. Transitions from nonliving to living matter. *Science* **2004**, *303*, 963–965.

5. Rasmussen, S.; Chen, L.; Stadler, B.M.R.; Stadler, P.F. Proto-organism kinetics: Evolutionary dynamics of lipid aggregates with genes and metabolism. *Orig. Life Evol. Biosph.* **2004**, *34*, 171–180.

6. Rasmussen, S.; Bailey, J.; Boncella, J.; Chen, L.; Collis, G.; Colgate, S.; DeClue, M.; Fellermann, H.; Goranovic, G.; Jiang, Y.; *et al.* Assembly of a minimal protocell. In *Protocells: Bridging Nonliving and Living Matter*; Rasmussen, S., Bedau, M., Chen, L., Deamer, D., Krakauer, D., Packard, N., Stadler, P., Eds.; MIT Press: Cambridge, MA, USA, 2008; pp. 125–156.

7. Szostack, W.; Bartel, D.P.; Luisi, P.L. Synthesizing life. *Nature* **2001**, *409*, 387–390.

8. Mansy, S.S.; Schrum, J.P.; Krishnamurthy, M.; Tobé, S.; Treco, D.A.; Szostak, J.W. Template-directed synthesis of a genetic polymer in a model protocell. *Nature* **2008**, *454*, 122–125.

9. Hanczyc, M. Steps towards creating a synthetic protocell. In *Protocells: Bridging Nonliving and Living Matter*; Rasmussen, S., Bedau, M., Chen, L., Deamer, D., Krakauer, D., Packard, N., Stadler, P., Eds.; MIT Press: Cambridge, MA, USA, 2009; p. 107.

10. The European Commission funded projects MatchIT. Available online: http://www.fp7-matchit.eu/ (access on 19 October 2011)

11. ECCell. Available online: http://homepage.ruhr-uni-bochum.de/john.mccaskill/ECCell/ (access on 19 October 2011).

12. Wu, T.; Orgel, L.E. Nonenzymic template-directed synthesis on oligodeoxycytidylate sequences in hairpin oligonucleotides. *J. Am. Chem. Soc.* **1992**, *114*, 317–322.

13. Wu, T.; Orgel, L. Nonenzymatic template-directed synthesis on hairpin oligonucleotides. 3. Incorporation of adenosine and uridine residues. *J. Am. Chem. Soc.* **1992**, *114*, 7963–7969.

14. Fernando, C.; Kiedrowski, G.v.; Szathmáry, E. A stochastic model of nonenzymatic nucleic acid replication: "Elongators" sequester replicators. *J. Mol. Evol.* **2007**, *64*, 572–585.

15. Monnard, P.A.; Dörr, M.; Löffler, P. Possible role of ice in the synthesis of polymeric compounds. Presented at the 38th COSPAR Scientific Assembly, Bremen, Germany, 15–18 July 2010.

16. Kiedrowski, G.V. A Self-replicating hexadeoxynucleotide. *Angew. Chem. Int. Ed.* **1986**, *25*, 932–935.

17. Sievers, D.; Kiedrowski, G.V. Self-replication of complementary nucleotide-based oligomers. *Nature* **1994**, *369*, 221–224.

18. Bag, B.G.; Kiedrowski, G.V. Templates, autocatalysis and molecular replication. *Pure App. Chem.* **1996**, *68*.

19. Joyce, G.F. Non-enzyme template-directed synthesis of RNA copolymers. *Orig. Life Evol. Biosph.* **1984**, *14*, 613–620.

20. Lincoln, T.A.; Joyce, G.F. Self-sustained replication of an RNA enzyme. *Science* **2009**, *323*, 1229–1232.

21. Wills, P.; Kauffman, S.; Stadler, B.; Stadler, P. Selection dynamics in autocatalytic systems: Templates replicating through binary ligation. *Bull. Math. Biol.* **1998**, *60*, 1073–1098.

22. Rocheleau, T.; Rasmussen, S.; Nielson, P.E.; Jacobi, M.N.; Ziock, H. Emergence of protocellular growth laws. *Philos. Trans. R. Soc. B* **2007**, *362*, 1841–1845.

23. Száthmary, E.; Gladkih, I. Sub-exponential growth and coexistence of non-enzymatically replicating templates. *J. Theor. Biol.* **1989**, *138*, 55–58.

24. Kiedrowski, G.v.; Wlotzka, B.; Helbing, J.; Matzen, M.; Jordan, S. Parabolic growth of a self-replicating hexadeoxynucleotide bearing a 3'-5'-phosphoamidate linkage. *Angew. Chem. Int. Ed.* **1991**, *30*, 423–426.

25. In particular, it has been shown that under parabolic growth conditions, competing replicators X_i grow when sufficiently rare:

$$[X_i] < \left(\frac{k_i}{k_{\text{base}}} \frac{\sum_j [X_j]}{\sum_j [X_j]^{1/2}} \right)^2 \implies \frac{d}{dt}[X_i] > 0$$

The equation captures the connection between the growth rate k_i and its selective pressure, such that replicator species with a high growth rate are also assigned a high evolutionary fitness. See [23] for the derivation.

26. Luther, A.; Brandsch, R.; Kiedrowski, G.V. Surface-promoted replication and exponential amplifcation of DNA analogues. *Nature* **1998**, *396*, 245–248.

27. Zhang, D.Y.; Yurke, B. A DNA superstructure-based replicator without product inhibition. *Nat. Comput.* **2006**, *5*, 183–202.

28. Owczarzy, R.; Vallone, P.M.; Gallo, F.J.; Paner, T.M.; Lane, M.J.; Benight, A.S. Predicting sequence-dependent melting stability of short duplex DNA oligomers. *Biopolymers* **1998**, *44*, 217–239.

29. Bloomfield, V.A.; Crothers, D.M.; Tinoco, I. *Nucleic Acids*; University Science Books: Sausalitos, CA, USA, 2000.

30. Poland, D.; Scheraga, H.A. Occurrence of a phase transition in nucleic acid models. *J. Chem. Phys.* **1966**, *45*, 1456–1463.

31. Hutton, T.J. Evolvable self-replicating molecules in an artificial chemistry. *Artif. Life* **2002**, *8*, 341–356.

32. Smith, A.; Turney, P.; Ewaschuk, R. Self-replicating machines in continuous space with virtual physics. *Artif. Life* **2003**, *9*, 21–40.

33. Klenin, K.; Merlitz, H.; Langowski, J. A Brownian Dynamics program for the simulation of linear and circular DNA and other wormlike chain polyelectrolytes. *Biophys. J.* **1998**, *74*, 780–788.

34. Tepper, H.L.; Voth, G.A. A coarse-grained model for double-helix molecules in solution: Spontaneous helix formation and equilibrium properties. *J. Chem. Phys.* **2005**, *122*, 124906.

35. Drukker, K.; Schatz, G.C. A model for simulating dynamics of DNA denaturation. *J. Chem. Phys. B* **2000**, *104*, 6108–6111.

36. Fellermann, H.; Rasmussen, S.; Ziock, H.J.; Solé, R. Life-cycle of a minimal protocell: A dissipative particle dynamics (DPD) study. *Artif. Life* **2007**, *13*, 319–345.

37. Kubo, R. The fluctuation-dissipation theorem. *Rep. Prog. Phys.* **1966**, *29*, 255.

38. Ryckaert, J.P.; Ciccotti, G.; Berendsen, H.J.C. Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n-Alkanes. *J. Comp. Phys.* **1977**, *23*, 327.

39. Note that our approach would not work in the absence of a thermostat: to describe rotational motion properly, one would need to define orientations and angular momenta in a local reference frame that moves with the extended object to which the oriented point particle belongs. In this manner, rotational motion of the extended object gets propagated down to the angular momenta of the particles it consists of (A QSHAKE algorithm would in addition be needed to properly conserve angular momenta in the constraints). While this approach is computationally significantly more cumbersome, we expect the result to be similar for the above model, in which rotation of extended objects is propagated down to its constituting particles through angular potentials and an overdamped thermostat.

40. Tinland, B.; Pluen, A.; Sturm, J.; Weill, G. Peristence length of single stranded DNA. *Macromolecules* **1997**, *30*, 5763–5765.

41. SantaLucia, J., Jr.; Allawi, H.T.; Seneviratne, A. Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry* **1996**, *35*, 3555–3562.

42. Teraoka, I. *Polymer Solutions—An Introduction to Physical Properties*; Wiley Interscience: New York, NY, USA, 2002.