Theory for sequence selection via phase separation and oligomerization

Ivar S. Haugerud, Giacomo Bartolucci, Dieter Braun, Christoph A. Weber

Angaben zur Veröffentlichung / Publication details:

Haugerud, Ivar S., Giacomo Bartolucci, Dieter Braun, and Christoph A. Weber. 2026. "Theory for sequence selection via phase separation and oligomerization." *Proceedings of the National Academy of Sciences* 123 (5): e2422829123. https://doi.org/10.1073/pnas.2422829123.

PNAS RESEARCH ARTICLE | PHYSICS
BIOPHYSICS AND COMPUTATIONAL BIOLOGY OPEN ACCESS

Theory for sequence selection via phase separation and oligomerization

Ivar S. Haugerud[a], Giacomo Bartolucci[a,b], Dieter Braun[c], and Christoph A. Weber[a,1]

Affiliations are included on p. 11.

Edited by Alexei A. Kornyshev, Imperial College London, London, United Kingdom; received November 15, 2024; accepted December 20, 2025, by Editorial Board Member Daan Frenkel

Nonequilibrium selection pressures were proposed for forming oligonucleotides with rich functionalities encoded in their sequences, such as catalysis. Since phase separation was shown to direct various chemical processes, we ask whether condensed phases can provide mechanisms for sequence selection. To answer this question, we use nonequilibrium thermodynamics and describe the reversible oligomerization of different monomers to sequences at nondilute conditions prone to phase separation. We find that as sequences form, their interactions can trigger phase separation, which in turn enriches some sequences while depleting others. Our main result is that phase separation creates a selection pressure leading to specific sequence patterns when fragmentation maintains the system away from equilibrium. When fragmentation is slow, alternating sequences that interact more cooperatively with their surroundings are preferred. When fragmentation is fast, sequences with longer repeating motifs capable of more specific interactions are selected instead. Our finding that out-of-equilibrium condensed phases can provide a selection mechanism highlights their potential as versatile hubs for the evolution of functional sequences, a question relevant to the molecular origin of life and de novo life.

phase separation | nonequilibrium thermodynamics | sequence selection | origins of life | oligomerization

Significance

An unresolved question is how specific oligonucleotide sequences such as DNA and RNA, which store genetic information and catalyze reactions, evolved from nucleotide building blocks without sophisticated biological machinery. Our research suggests that phase-separated condensates—naturally occurring mesoscopic clusters of oligonucleotides—may have created a unique environment that mediated selection pressures favoring certain sequences. Using nonequilibrium thermodynamics, we explore conditions under which phase separation and intersequence interactions generate such pressures. We find that nonequilibrium driving via sequence fragmentation determines whether more or less cooperatively interacting sequences are selected. Our findings open up promising avenues to study how nonequilibrium physics and phase separation can select specific, potentially functional sequences.

The ability to store genetic information and catalyze chemical reactions makes oligonucleotides, such as DNA and RNA, key building blocks for the molecular origin of life (1–4). Without any specialized biological molecules, such as protein-based enzymes on early Earth, the mechanisms that can give rise to the emergence of long oligomers with prebiotic functions remain elusive (5–9). A particularly aspired function is folding into a secondary configuration capable of replicating unfolded oligonucleotide sequences. However, a fundamental problem is the exponentially large number of different sequences, particularly for the minimal sequence length required for complex functionalities such as replication (10). This problem renders the need for selection mechanisms at the molecular origin of life in the early stages when sequences grow in length and start exploring their exponential sequence space. Various selection mechanisms were proposed based on the nonequilibrium conditions presumably acting on early Earth (11), such as biased replication (12, 13), salt and temperature gradients (14, 15), wet–dry (16–18) or freeze–thaw cycles (19), shear-driven replication (20, 21), accumulation at liquid–vapor interfaces (22) or mineral surfaces (23), template aided-ligation (24–26), and finally via phase-separated condensates (27).

The significance of phase-separated condensates in providing a selection mechanism at the molecular origin of life was suggested roughly a century ago (28–31). These ideas have been revisited due to advances in sequencing technologies capable of resolving large sequence distributions (32), complemented by observations of phase separation of oligonucleotide sequences under prebiotic conditions. In particular, it was shown that the attractive interactions among different oligomeric sequences can result in the formation of additional phases such as hydrogels (33, 34), liquid crystalline phases (35, 36), and coacervates (27, 37–39). Interestingly, the saturation concentrations above which oligonucleotides such as single-stranded DNA and RNA phase separate can be low, in the order of μM, due to the strong interaction strength ($5k_BT$) of complementary basis pairs (37, 40). Such low saturation concentrations make oligonucleotide phase separation a likely and robust scenario, even under varying physicochemical conditions at early Earth. Furthermore, nonequilibrium thermodynamics implies that the chemical processes such as

Author contributions: All authors designed research; I.S.H. and C.A.W. performed research and analyzed data; and all authors wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. A.A.K. is a guest editor invited by the Editorial Board.

[1]To whom correspondence may be addressed. Email: christoph.weber@uni-a.de.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2422829123/-/DCSupplemental.

Published January 28, 2026.

polymerization, ligation, and fragmentation of oligonucleotides are generically coupled with phase separation of oligonucleotides (39, 41, 42). This coupling gives rise to feedback between the evolving sequence distribution and their propensity to form condensed phases. In other words, when considering the condensed phases as phenotypes and their local sequence distributions as genotypes, thermodynamics provides a coupling between phenotype and genotype—a prerequisite for Darwinian evolution.

The coupling between chemical processes and phase separation is expected to mediate a selection mechanism that depends on the intersequence interactions. Indeed, phase separation was shown to enrich gradually specific sequences from a pool composed of a few designed sequences (27). However, it remains an open question whether the interaction among sequences and particularly phase separation, can direct the oligomerization of sequences. If phase separation can direct oligomerization, it could represent a physicochemical mechanism that reduces the occupations of the exponentially growing sequence space. The coupling between sequence distributions and phase separation becomes especially interesting in the presence of competing effects to aggregation, namely hydrolysis (26, 43). Phase separation creates a phase that is both oligomer-rich and solvent-poor (39), possibly serving as a protective environment against hydrolysis, steering the fragmentation of specific sequences.

In this work, we developed a theoretical framework using nonequilibrium thermodynamics to describe the oligomerization of sequences at nondilute conditions. Such conditions require accounting for the interactions among all sequences, including the possibility of phase coexistence, as illustrated in Fig. 1. Using our theoretical framework, we find that the interactions among sequences direct oligomerization toward sequences that phase separate from the solvent and form a condensed phase. We show that phase separation favors forming more cooperatively interacting sequences over specialized sequences that interact only weakly with most other sequences. To unravel the role of phase separation in sequence selection, we maintain the system out of equilibrium by fragmentation inspired by the hydrolysis of oligonucleotides. We find that slow fragmentation rates select for sequences with a high information content, while low-information sequences with highly correlated sequence motifs emerge for large enough fragmentation rates. Thus, phase separation can mediate a selection pressure for sequences through selective oligomerization.

Theory for Oligomerization at Phase Equilibrium

To describe the dynamics of sequences in phase-separated systems, we use the volume fraction, ϕ_i^α, where α denotes the phase. Here, the component label i includes monomers and oligomers summarized in the set σ. For two monomeric building blocks, denoted a and b, the union of the set of monomers $\sigma_m = \{a, b\}$ and sequences $\sigma_o = \{aa, ab, ba, bb, \ldots\}$ gives all nonsolvent components $\sigma = \sigma_m \cup \sigma_o$. The volume fraction of a component, $\phi_i^\alpha = \nu_i N_i^\alpha / V^\alpha$, quantifies the volume occupied by N_i^α molecules of type i with molecular volume ν_i, relative to the phase volume V^α. For simplicity, we consider an incompressible system where molecular volumes ν_i of each component are constants.

In this work, we build on a theoretical framework developed for nondilute chemical kinetics where, if phase separation occurs, the different phases α are at phase equilibrium (41). In this framework, the time evolution of the components volume fractions $\phi_i^\alpha(t)$ in each phase α is governed by oligomerization rates r_i^α, the diffusive exchange fluxes between phases j_i^α, and changes in the phase volumes V^α (see Fig. 1 for an illustration of the different processes):

$$\frac{\mathrm{d}}{\mathrm{d}t}\phi_i^\alpha = r_i^\alpha - j_i^\alpha - \frac{\phi_i^\alpha}{V^\alpha}\frac{\mathrm{d}}{\mathrm{d}t}V^\alpha, \qquad [1a]$$

with the component label $i \in \sigma$ denoting monomers and oligomer sequences. In an incompressible system, oligomerization conserves volume ($\sum_{i\in\sigma} r_i^\alpha = 0$), such that the phase volume $V^\alpha(t)$ changes in time only due to partitioning fluxes of solvent j_s^α, and partitioning fluxes of monomers and sequences j_i^α ($i \in \sigma$) through the phase boundary:

$$\frac{1}{V^\alpha}\frac{\mathrm{d}}{\mathrm{d}t}V^\alpha = -j_s^\alpha - \sum_{i\in\sigma} j_i^\alpha. \qquad [1b]$$

The oligomerization rate r_i^α of component i results from the gain $r^\alpha_{j+m\rightleftharpoons i}$ and loss $r^\alpha_{i+m\rightleftharpoons j}$ contributions related to the possible oligomerization pathways. For simplicity, we discuss monomer pick-up and release (Fig. 2*A*; see *SI Appendix*, section 3 for studies also including ligation), leading to

$$r_i^\alpha = \sum_{j\in\sigma}\sum_{m\in\sigma_m}\left(r^\alpha_{j+m\rightleftharpoons i} - r^\alpha_{i+m\rightleftharpoons j}\right). \qquad [1c]$$

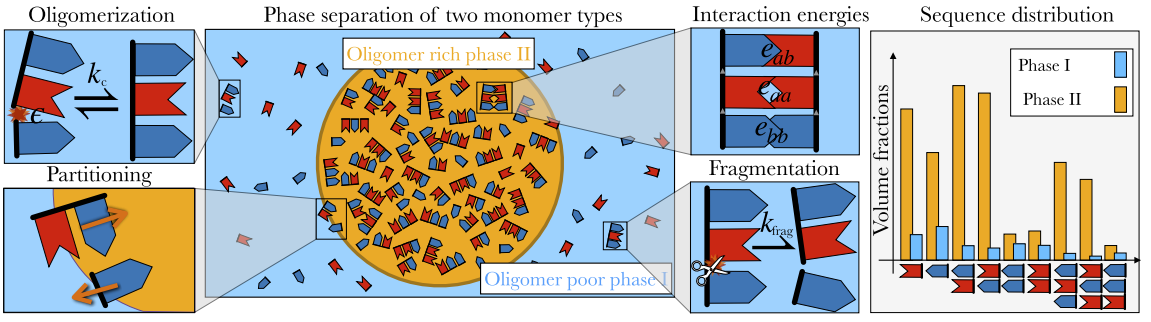

Fig. 1. Different monomers oligomerize to sequences that can fragment and interact leading to oligomer-rich and oligomer-poor phases. Two distinct monomeric units a and b reversibly oligomerize through monomer pick-up and release with a rate coefficient k_c and a free energy ϵ associated with each bond along the oligomer backbone. Sequence-dependent interactions between oligomers can lead to phase separation, resulting in the coexistence of an oligomer-rich and an oligomer-poor phase, mediating the partitioning of monomers and sequences. Sequences interact through nucleotide base specific interactions, with interaction parameters e_{ab}, e_{aa}, and e_{bb}. We also discuss the case when the system is maintained away from equilibrium by a fragmentation pathway occurring with a rate coefficient k_{frag}. Due to nondilute conditions, phase separation and reversible oligomerization are mutually coupled. This coupling gives rise to complex sequence distributions that differ between the two coexisting phases, I and II.

This oligomerization rate conserves the total volume (and mass) of monomers and monomers incorporated into sequences. In a nondilute system, the gain and loss contribution can be written as follows (44):

$$r^{\alpha}_{i+m\rightleftharpoons j} = k^{\alpha}_{imj}\left[\exp\left\{\frac{\mu^{\alpha}_i+\mu^{\alpha}_m}{k_BT}\right\}-\exp\left\{\frac{\mu^{\alpha}_j}{k_BT}\right\}\right], \qquad \textbf{[1d]}$$

where the terms in the square bracket denote the forward $r_{i+m\rightarrow j}$ and backward $r_{i+m\leftarrow j}$ reaction rates respectively, and μ^{α}_j denote the chemical potential of component j in phase α. In Eq. **1d**, forward and backward rates obey detailed balance of the rates (45, 46):

$$\frac{r^{\alpha}_{i+m\rightarrow j}}{r^{\alpha}_{i+m\leftarrow j}} = \exp\left\{\frac{\mu^{\alpha}_i+\mu^{\alpha}_m-\mu^{\alpha}_j}{k_BT}\right\}. \qquad \textbf{[1e]}$$

There is a single kinetic rate coefficient k^{α}_{imj} for each chemical reaction $i+m\rightleftharpoons j$ governing the relaxation toward chemical equilibrium, $\mu^{\alpha}_i+\mu^{\alpha}_m=\mu^{\alpha}_j$. For simplicity, we consider the kinetic rate coefficients constant, phase independent, and agnostic to the reactants, such that $k^{\alpha}_{imj}=k_c$.

The nondilute conditions require accounting for interactions among all components, the solvent, monomers, and oligomers. On a coarse-grained level, these interactions can be described by contributions to the free energy density $f(\{\phi_i\})$, which makes the chemical potential that characterizes the free energy cost to add component i to a mixture, $\mu_i(\{\phi_j\}_{j\in\sigma})=\nu_i\partial f/\partial\phi_i|_{\phi_{j\neq i}}$, dependent on all components $j\in\sigma$. The free energy density f contains mixing entropy contributions and enthalpic terms up to second order in each sequence volume fractions ϕ_i (mean-field); see, e.g., ref. 39. First-order enthalpic terms in f correspond to internal free energies that govern oligomerization equilibria in the dilute limit. In our model, the internal free energies of sequences depend on sequence length, and internal free energy scales are set by backbone bond energy ϵ (Fig. 1). We consider ϵ as sequence-independent and study values in the order of a few k_BT, consistent with biofilaments such as DNA and RNA (47). We note that for peptides, nearest neighbor interactions along the sequences are crucial, leading to a sequence-dependent internal free energy (48, 49). The interactions between sequences i and j are captured by second-order terms in the free energy and are of the form $e_{ij}\phi_i\phi_j$, where e_{ij} is the average interaction parameter. The quantity e_{ij} is calculated by performing a Boltzmann average over all possible combinations of nucleotide interactions (corresponding to base-pairs for DNA and RNA) between sequence i and j, as illustrated in Fig. 2*B*. Inspired by RNA and DNA, we investigate cases where nucleotides interact favorably with their opposite, i.e., $e_{aa}=e_{bb}=0$, and $e_{ab}<0$ (attractive due to Watson–Crick base pairing). For more information about the interactions, see *SI Appendix*, section 2.

Interactions between sequences can lead to phase separation once enough sequences have formed via oligomerization. If diffusion rates of sequences between phases are fast compared to oligomerization rates, phases are at phase equilibrium at each time of the oligomerization kinetics (see ref. 41 for a detailed discussion). For two coexisting phases $\alpha=\mathrm{I},\mathrm{II}$, chemical potentials μ^{α}_i and osmotic pressures $\Pi^{\alpha}=f^{\alpha}-\sum_{i\in\sigma}\phi^{\alpha}_i\mu_i/\nu_i$ are balanced: $\mu^{\mathrm{I}}_i=\mu^{\mathrm{II}}_i$ and $\Pi^{\mathrm{I}}=\Pi^{\mathrm{II}}$. In this work, we focus on the case of two coexisting phases and compare the results to the case where phase separation is suppressed (mixed case). To support this focus, we looked for multiphase equilibria for the parameters investigated in this work, finding exclusively two-phase coexistence (see *SI Appendix*, section 6 for further discussions). During the oligomerization kinetics in the presence of two phases at phase equilibrium, the oligomer-rich (II) and oligomer-poor (I) phases are constrained to the surface of a high-dimensional binodal manifold, where tie-lines connect the two coexisting phases; see Fig. 2*C*. The constraint of phase equilibrium during oligomerization is satisfied by solving $\partial_t\mu^{\mathrm{I}}_i=\partial_t\mu^{\mathrm{II}}_i$ and $\partial_t\Pi^{\mathrm{I}}=\partial_t\Pi^{\mathrm{II}}$, and self-consistently calculating the diffusive fluxes between the phases, j^{α}_i; for more details see *SI Appendix*, section 7. Note that the condition of phase coexistence implies the α-superscript of the chemical potentials in Eqs. **1**c, **1**d, and **1**e, can be omitted. Furthermore, as the reaction rate coefficient k_c is phase-independent, it implies an equal oligomerization rate r_i in the oligomer-poor and -rich phase throughout the chemical kinetics.

Since chemical potentials govern both the oligomerization rates (Eq. **1d**) as well as the diffusive fluxes between the coexisting phases (last paragraph), there is a coupling between oligomerization and phase separation. This coupling is a fundamental thermodynamic property of chemically reacting, nondilute mixtures (also away from equilibrium) (39, 41, 42). In our work, it is essential for the selective oligomerization of sequences due to their favorable interactions with their local, oligomer-rich or oligomer-poor environment.

To understand the role of phase separation in the oligomerization kinetics, we consider a mixed system as a reference. The mixed system is homogeneous, and the reaction kinetics follows

$$\mathrm{d}\phi_i/\mathrm{d}t = r_i \qquad \textbf{[2]}$$

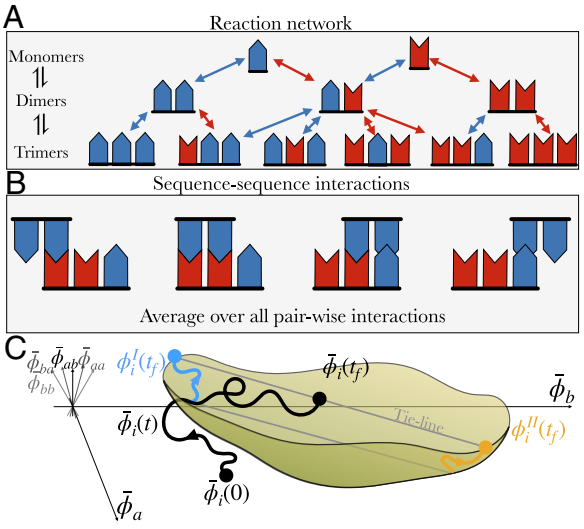

Fig. 2. Reaction network, sequence–sequence interactions, and oligomerization kinetics in the high-dimensional sequence space. (*A*) The reaction network for monomer pickup and release is illustrated for a subset of the sequence space (monomers, dimers, and trimers). (*B*) Interactions between sequences are determined by sliding over all possible pair-wise sequence configurations (*SI Appendix*, Eq. **S2**). We average over such configurations to calculate an interaction parameter for each sequence pair (*SI Appendix*, Eq. **S3**). (*C*) Oligomerization creates a trajectory in the high-dimensional sequence space. Initializing the system in a homogeneous state with an average volume fraction $\bar{\phi}_i(t=0)$, the oligomerization trajectory $\bar{\phi}_i(t)$ (black) may intersect with the binodal manifold (green surface), leading to two coexisting phases. The phases I and II exhibit different compositions of oligomers (oligomer-rich and oligomer-poor) that are connected by a tie-line. Phase compositions $\phi^{\mathrm{I/II}}_i(t)$ move along the binodal manifold (blue and orange) and relax to thermodynamic equilibrium or a nonequilibrium steady state at time t_f.

with the oligomerization rate given in Eqs. **1**c and **1**d. In simple terms, it is the same system as the phase-separated system, i.e., with the same physical parameters, but where phase coexistence is suppressed. We note that the mixed state has a higher free energy requiring nonequilibrium mixing to maintain the system away from phase equilibrium.

In the following, we numerically integrate Eq. **1** to evolve the volume fractions of monomers and sequences, $\phi_i^\alpha(t)$ with $i \in \sigma$ and $\alpha \in \{\mathrm{I}, \mathrm{II}\}$. The average volume fractions are found from a weighted sum over both phases,

$$\bar{\phi}_i = \left(V^{\mathrm{I}}\phi_i^{\mathrm{I}} + V^{\mathrm{II}}\phi_i^{\mathrm{II}}\right) V^{-1}, \qquad [3]$$

where V is the total system volume $V = V^{\mathrm{I}} + V^{\mathrm{II}}$. For all studies in this work, we initialize the system at $t = 0$ with an equal amount of a and b monomers and no other nonsolvent components. Specifically, $\bar{\phi}_a(t=0) = \bar{\phi}_b(t=0)$, and $\bar{\phi}_i(0) = 0\ \forall i \in \sigma_o$, which further sets the total sequences volume fraction $\bar{\phi}_{\mathrm{tot}} = \sum_{i\in\sigma}\bar{\phi}_i$ that is conserved during the oligomerization kinetics.

Sequence Evolution Toward Thermodynamic Equilibrium

Dimers, trimers, and longer oligomers emerge from a pool of monomers through oligomerization. As more and longer oligomers form, the decrease in mixing entropy and the stronger attractive interactions enhance the propensity to phase separate. When phase separation occurs at $t = t^*$ (Fig. 3*A*), an oligomer-rich phase (phase II) forms that coexists with an oligomer-poor phase (phase I). The two phases can differ by several orders in magnitude in their total oligomer volume fraction. The corresponding phase average $\bar{\phi}$ typically exceeds the one of the mixed system governed by Eq. **2**, where phase separation is suppressed.

The interactions among sequences in the presence of phase separation biases the oligomerization kinetics, leading to the enrichment of specific sequences. As examples, we show the evolution of two different 5-mers, an alternating sequence *ababa* (Fig. 3*B*) and a homo-5-mer *aaaaa* (Fig. 3*C*). Upon phase separation, the volume fractions in the two phases differ significantly between the two sequences. However, this difference is much more pronounced for the alternating sequence than for the homopolymeric sequence. This is because phase separation arises from attractive interactions between sequences, and alternating sequences tend to interact favorably with many others in the system (we later refer to these as cooperative), whereas homosequences interact strongly only with their homopolymeric complements. Most importantly, the sequence interactions also affect the oligomerization, biasing the formation toward such strongly interacting, phase-separation-prone sequences. Conceptionally, the mutual feedback between oligomerization and phase equilibrium arises from both processes being governed by the free energy (see next section discussing the system maintained away from equilibrium). This bias is reflected by the higher average volume fractions with $\bar{\phi}_{ababa}(t) \gg \bar{\phi}_{aaaa}(t)$ for late times t. It also explains why many more sequences of this type form throughout the entire system at thermodynamic equilibrium (black crosses in Fig. 3*D*).

To unravel the role of the mutual feedback between phase separation and oligomerization leading to the strong enrichment of specific sequences, we considered a mixed reference system (see dashed curves in Fig. 3 *A–C*). Physically, a mixed system is homogeneous, and phase separation is suppressed. Mathematically, this case corresponds to only considering the oligomerization

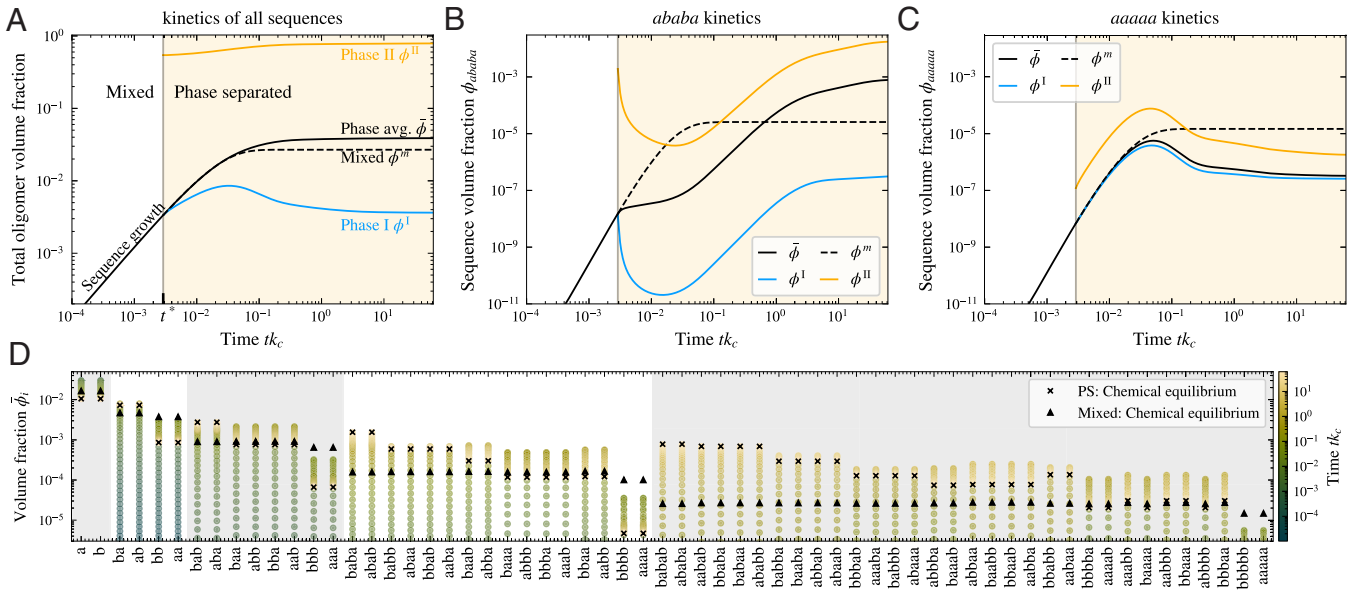

Fig. 3. Oligomerization kinetics triggers phase separation leading to the enrichment of specific sequences. (*A–C*) After initializing the system with the two monomers *a* and *b*, oligomer sequences form, increasing their volume fractions $\phi_i(t)$, causing the total oligomer volume fraction $\sum_{n=2}\bar{\Phi}(n)$ (*SI Appendix*, Eq. **S8**) to increase (*A*). Once sufficient oligomers have formed, the interactions among sequences trigger the formation of two coexisting, phase I (blue) and II (orange) with different volume fractions in each phase. After phase separation, different sequences undergo specific kinetics, leading to different phase compositions at thermodynamic equilibrium. Here, we display the dynamics of the specific oligomers $i =$ *ababa* (*B*), and $i =$ *aaaaa* (*C*). In the phase separation domain, we also show the average volume fraction (black, Eq. **3**). As a control, we compare it to the mixed system where phase separation is suppressed (dashed). (*D*) The complex time evolution of the sequence distribution ranging from monomers to five-mers is generally nonmonotonic, where time is indicated by the color bar. While the mixed reference case (triangle) shows an almost flat distribution for each sequence length at chemical equilibrium, there are large variations in the average volume fraction $\bar{\phi}_i(t)$ when the system can phase-separate (cross). Note that we show all sequences, ordered by their Hamming distance to the alternating sequence of equal length, and subordered by the number of *a*-*b*-neighbors.

kinetics in the nondilute mixture by solving Eq. **2**, such that the constraint of phase equilibrium is not fulfilled. In Fig. 3*A*, we see that the mixed system leads to a bit less total oligomer volume fraction $\sum_{n=2}\bar{\Phi}(n)$, where $\bar{\Phi}(n)$ is the phase-averaged total volume fraction of oligomers of length n (*SI Appendix*, Eq. **S8**), than the phase-separated system. In other words, phase separation only slightly enhances the net trend of monomers forming oligomers. For specific sequences, however, the time evolution toward chemical equilibrium is completely altered by phase separation. This effect can be seen by comparing the dashed (mixed) with blue (phase I) and orange (phase II) lines in Fig. 3 *B* and *C* for the two 5-mers, and when considering the full sequences distribution shown in Fig. 3*D*. The two sequences depicted in Fig. 3 *B* and *C* differ by less than a factor of two when mixed, while for the phase average, the same ratio is more than 10^3. Phase separation enhances *ababa* multiple orders of magnitude and suppresses *aaaaa* to the extent that it is more abundant for the mixed system. In Fig. 3*D*, the final chemical equilibrium volume fractions for each sequence are indicated by crosses, while the stationary values for the mixed state are displayed by triangles. We observe that the mixed case has only a little diversity between different sequences of the same length since it yields an almost flat sequence distribution. Note that the mixed system behaves completely differently from the phase-separated case, leading to specific sequences being strongly enriched while others are depleted.

A striking property of our theoretical framework is that it allows us to follow the kinetics of the full sequence distribution in time (Fig. 3*D*). Longer oligomers form at the expense of depleting the amount of initialized monomers. At early times, the kinetics is dominated by the entropic gain to distribute the monomer mass to larger oligomers. This fast entropy-driven growth at intermediate times causes the oligomer volume fractions to exceed its equilibrium value before relaxing to it due to sequence–sequence interactions (see, e.g., Fig. 3*C*). At later times, the enthalpic gains through interactions balance with the entropic costs of unequally distributed mass between sequences of the same length. In the mixed case, the system is not dense enough for the interactions to alter the chemical kinetics significantly, resulting in a simple monotonic increase in the volume fractions (dashed lines in Fig. 3 *B* and *C*). Eventually, thermodynamic equilibrium is reached where phase coexistence and chemical equilibrium are concomitantly satisfied, and the sequence distribution in each phase becomes stationary (black crosses in Fig. 3*D*). At thermodynamic equilibrium, the most and least abundant sequences among 5-mers differ in volume fraction by more than three orders of magnitude, as shown in Fig. 3 *B* and *C*. Phase separation thus strongly alters kinetics and the sequence distribution at thermodynamic equilibrium.

The key property of phase separation is that the coexisting phases act as local hubs with very different total oligomer mass and, thereby, very different intersequence interactions (Fig. 3*A*). These interactions affect oligomerization locally, enriching or depleting specific sequences. This hub-like effect is evident by comparing the five most abundant sequences in each phase (Fig. 4*A*). While the oligomer-poor (I) phase has no clear preference for any sequence (see also *SI Appendix*, Fig. S6*A*), the oligomer-rich (II) phase shows a clear preference toward alternating sequences. For the corresponding mixed system with a total volume fraction $\bar{\phi}_{\mathrm{tot}} = 0.10$ (Fig. 4*B*), the distribution is almost flat, similar to the oligomer-poor phase. Importantly, phase separation has a dual effect: it changes which sequences are enriched and depleted and also magnifies their depletion and enrichment. This can be seen by sequences in the oligomer-rich (II) phase (indicated by orangish color code) that do not appear in either the oligomer-poor (I) or the mixed case. Consistently, the oligomer-poor (I) phase and the mixed system at low total oligomer volume fraction enrich similar sequences such as *aabbabaa*.

The enrichment mediated by phase separation is most pronounced for small volumes of the oligomer-rich phase. This is apparent by comparing the equilibrium volume fraction averaged over both phases (Eq. **3**) with the mixed case where phase separation is suppressed, for varying the total volume fraction of monomers and oligomers $\bar{\phi}_{\mathrm{tot}}$; see Fig. 4*C*. Phase separation enriches the volume fraction of the alternating 8-mers, while depleting the homo-8-mers. The relative enrichment is largest for small volumes of the oligomer-rich phase. This behavior is due to the volume fractions of long sequences increasing exponentially with the total mass (*SI Appendix*, Eq. **S10**), while the phase average increases linearly with the composition of the oligomer-rich phase (Eq. **3**). As a result, there is a relative enrichment of

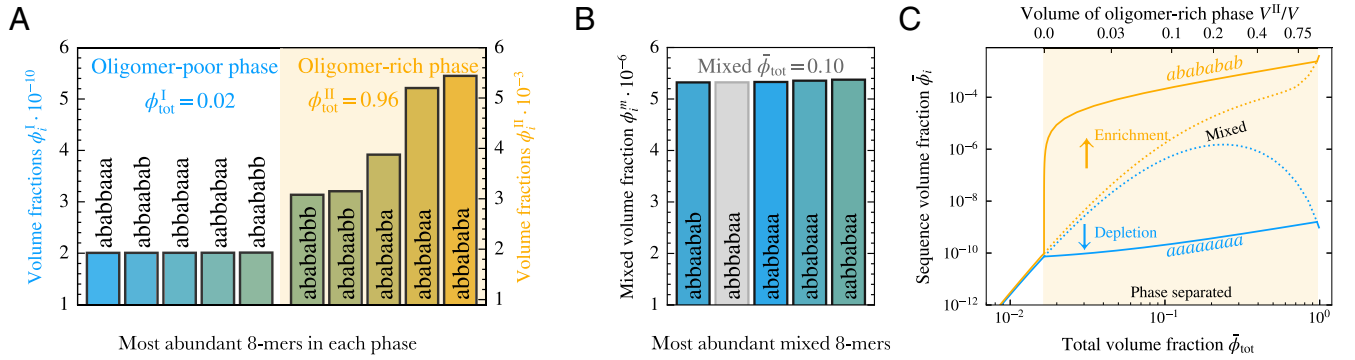

Fig. 4. Phase separation is key for the enrichment of alternating sequences and is most pronounced for small oligomer-rich phases. (*A*) The volume fraction of the five most abundant 8-mer sequences indicates that the oligomer-rich phase II biases oligomerization toward alternating sequences (within approx. one Hamming distance) by a compositional enhancement of around seven orders of magnitude relative to the oligomer-poor phase I. Moreover, in contrast to the oligomer-rich phase II, phase I also shows no significant compositional variations among sequences. (*B*) In the mixed system ($\bar{\phi}_{\mathrm{tot}} = 0.10$), the five most abundant sequences are similar to the oligomer-poor phase I and show only little compositional variations. Increasing $\bar{\phi}_{\mathrm{tot}}$ toward $\phi^{\mathrm{II}}_{\mathrm{tot}}$, the mixed phases consistently approach the composition of the oligomer-rich phase II (see *SI Appendix*, Fig. S1). (*C*) In the phase-separated regime ($\phi^{\mathrm{I}}_{\mathrm{tot}} < \bar{\phi}_{\mathrm{tot}} < \phi^{\mathrm{II}}_{\mathrm{tot}}$), the relative enrichment and depletion is largest for small volumes of the oligomer-rich phase II, V^{II} (see upper plot axis). While homopolymers (e.g., *aaaaaaaa*) get depleted at thermodynamic equilibrium, alternating sequences (e.g., *abababab*) become enriched by many orders of magnitude compared to the respective mixed cases (dotted).

specific sequences up to a 1,000-fold. For the mixed system, the two sequences depicted remain at approximately equal abundance until around $\bar{\phi}_{\mathrm{tot}} \simeq 0.1$, while phase separation is able to differentiate them at dilute conditions. Phase separation allows for the enrichment of specific sequences through the local creation of a dense environment at conditions only slightly above the saturation volume fraction, ($\phi^{\mathrm{I}}_{\mathrm{tot}} \lesssim \bar{\phi}_{\mathrm{tot}} \ll \phi^{\mathrm{II}}_{\mathrm{tot}}$), which would not be possible without phase separation. We note that weakly supersaturated conditions are common in experimental systems on oligonucleotide phase separation (50–53).

Sequence interactions collectively enrich similar sequence patterns in the respective phases. In the oligomer-poor phase, the oligomers are too dilute to affect each other's propensity through interactions, as seen in *SI Appendix*, section 4. In the oligomer-rich phase, the oligomer–oligomer interactions determine the sequence distribution, as seen in *SI Appendix*, Fig. S6*B*. Without interactions ($e_{ab} = 0$) or a too-weak interaction strength e_{ab}, the distribution remains approximately flat for each length, independent of the total volume fraction of monomers and sequences ϕ_{tot}. With stronger interactions (e_{ab} more negative), the volume fractions of the enriched/depleted sequences may enhance/deplete further (see Fig. 5*A* for 8-mers). However, a nonmonotonous behavior can be observed for some sequences when increasing the interaction strength e_{ab}. For example, *abaababbb* is enriched at $e_{ab} = -4k_BT$, but depleted at $-5k_BT$. Another example is *aababababaa* being the most abundant sequence at $e_{ab} = -3k_BT$, but that is far from the case at $-5k_BT$. The nonmonotonous dependence on the interaction strength e_{ab} originates from enrichment being a collective effect of the sequence distribution. For intermediate interaction strengths, the enrichment of *ab* is not strong enough for many repetitions of the *ab* pattern in the same sequence to minimize interaction energies, instead a combination with *aa* and *bb* minimizes the interaction energies. For stronger interaction, the energetic cost of *aa* and *bb* patterns depletes the previously enriched sequences. The effect is collective, as it is the entire sequence distribution that determines the optimal patterns.

Sequence interactions and phase separation mediate a strong bias toward sequences with almost alternating sequence patterns at thermodynamic equilibrium (Figs. 4*A* and 5*A*). This behavior is evident in the probability distribution for the proportion of *a-b* neighbors for 8-mers. The proportion of *a-b* neighbors takes the value of 0 for homo-oligomers and the value of 1 for a perfectly alternating oligomer (Fig. 5*B*, illustration left/right). Without interactions, the distribution is a binomial distribution symmetric around an *a-b* neighbor proportion of a half. With stronger interactions e_{ab}, the mean of the distribution shifts to a higher number of *a-b* neighbors, and the variance of the distribution decreases. The results support that oligomers close to homo-oligomers are statistically disfavored as they have fewer complementary strand motifs where sequences can attach. Due to the volume fractions decaying with length (*SI Appendix*, Eq. **S10**), the interactions with shorter sequences are the most relevant for enhancing motifs in longer sequences. Though alternating patterns are preferred, a perfect alternating pattern does not produce any complementary substrand for other nonperfectly alternating sequences and is, hence, not as enriched. The strongest enrichment occurs for approximately alternating sequences that deviate slightly from a perfect alternating pattern, allowing them to interact favorably with *aa* or *bb* segments of other sequences.

Phase separation and interactions among oligomers create long-range correlations along the oligomer sequence. The cor-

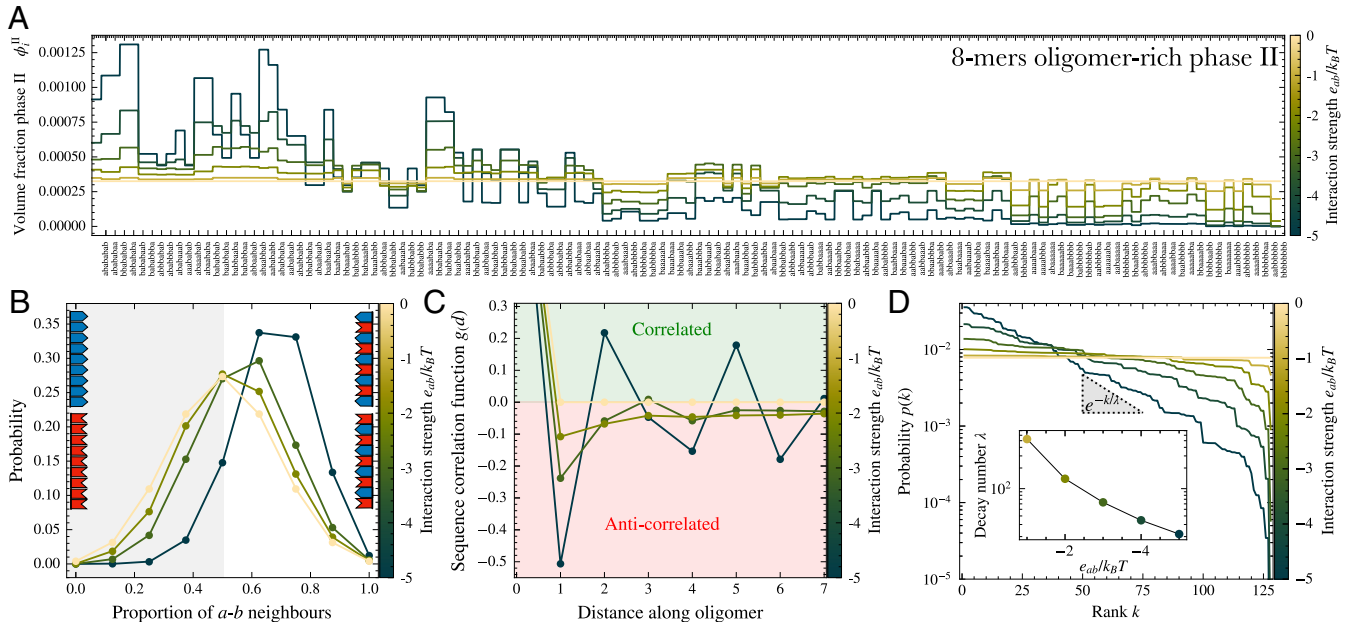

Fig. 5. Formation of alternating sequences requires phase separation and strong enough sequence interactions. (*A*) There are strong variations between the equilibrium volume fraction of sequence in the distributions of 8-mers in the oligomer-rich phase II. The enrichment or depletion of specific sequences originates from sequence–sequence interactions e_{ab} (values indicated by color code). (*B* and *C*) Alternating sequences require strong enough attractive interactions, e_{ab} This trend is reflected in (*B*), showing the *a-b* neighbor distribution getting more skewed toward more *a-b* neighbors, and (*C*) with the sequence correlation function $g(d)$ (Eq. **4**) developing a clear oscillatory behavior with a wave number of about two monomers (i.e., units of *ab*). (*D*) Ordering the sequences by abundance gives the rank distribution $p(k)$, which decays exponentially, $p \propto e^{-k/\lambda}$. The characteristic decay number λ decreases with more adhesive interactions e_{ab}. This behavior indicates that more monomers are shuffled to a smaller amount of sequences to enhance their abundance. In other words, variations in the sequence distribution enhance.

responding autocorrelation function

$$g(d) = \frac{\sum_{i\in\sigma} \delta_{n_i,l} \sum_{j=1}^{l} \phi_i \sum_{\pm} \begin{cases} 1 & \text{if } \Gamma_{i,j\pm d} = \Gamma_{i,j} \\ -1 & \text{if } \Gamma_{i,j\pm d} \neq \Gamma_{i,j} \end{cases}}{2l \sum_{i\in\sigma} \phi_i \delta_{n_i,l}}, \quad [4]$$

characterizes the occurrence of the same monomers along an oligomer of fixed length l. In the equation above, $\Gamma_{i,j}$ gives the monomer-type at position j of oligomer i. The autocorrelation function $g(d)$ shows pronounced oscillations for large enough interaction strength e_{ab} (Fig. 5*C*). Stronger interactions (more negative values of e_{ab}) give rise to stronger correlations, while no interactions ($e_{ab} = 0$) leave all sequences uncorrelated. The wave number of the correlation function $g(d)$ is approximately two, corresponding to the alternating pattern *ab*, which is statistically most common.

The stronger sequences interact (more negative e_{ab}), the more of the total monomer mass gets turned into enriched sequences. This trend can be quantified by the rank distribution $p(k)$ that orders the sequences by their abundance (rank k). We find that the rank distribution $p(k) \propto \exp(-k/\lambda)$ decays exponentially, with the characteristic rank scale λ decreases with the strength of interaction e_{ab}. Without interactions ($e_{ab} = 0$), the sequence distribution for a given length is flat, and so is the rank distribution $p(k)$ (yellow line shown in Fig. 5*D*). With increasing interaction strength, the more alternating sequences are enriched while most other sequences get more depleted. The characteristic rank scale λ thus decreases with interaction strength e_{ab} (*Inset* of Fig. 5*D*). As most sequences are depleted for large interaction strengths e_{ab}, stronger interactions reduce the size of the effectively occupied high-dimension sequence space.

The preference toward alternating sequences originates from the necessity of all sequences to interact with both *a*'s and *b*'s. The preference for alternating sequences results from the interaction strength e_{ab} and the volume fractions of all sequences being large enough, which both affect the distribution in an exponential fashion (*SI Appendix*, Eq. **S11**). While large sequences are dilute compared to short oligomers (dimers, trimers, ...), the dominant contributions in the exponential leading to the enrichment of alternating patterns in long oligomers (7-, 8-, 9-mers), stem from short oligomers (see *Inset* of *SI Appendix*, Fig. S6*B*). The origin of short alternating sequences arises from them interacting more favorably with their environment, meaning their respective phases, i.e., they are more cooperative. In Fig. 3*D* and *SI Appendix*, Fig. S6*B* for example, the dimer *ab* is much more abundant than both *aa* and *bb*, as it interacts favorably with both monomer types ($e_{aa,a} = 0$, $e_{aa,b} = e_{ab}$, while $e_{ab,a} = e_{ab,b} \approx e_{ab}$). But why are the most enriched sequences approximately one Hamming distance away from the perfectly alternating 8-mer *abababababab*, and not the perfectly alternating 8-mer itself? This is because a single nonalternating segment, *aa* or *bb*, enhances the average interactions with the diverse sequence environment, further boosting a sequence's cooperativity. The number of such nonalternating segments in the enriched sequences decreases as the interaction strength is increased (Fig. 5*A*).

Sequence Selection by Nonequilibrium Fragmentation

To decipher the effects of phase separation in sequence selection away from equilibrium, we break detailed balance of the rates between the fragmentation and fusion pathways $i \rightharpoonup j + |i-j|$; see *SI Appendix*, section 3 for further studies with nonequilibrium ligation leading to qualitative similar results. In general, both pathways rely on the break-up (formation) of bonds along the sequence backbone, where sequence i fragments into (forms from) the suboligomer/monomer j and $|i-j|$. An example is $aababab \rightharpoonup aa + bab$. Detailed balance of the rates is broken by omitting the thermal fusion (ligation) pathway in the following, considering it to be slow compared to fragmentation. Thus, the system is maintained away from equilibrium. The rate of the fragmentation pathway for sequence i in phase α is written as

$$h^{\alpha}_{i\rightharpoonup j+|i-j|} = k_{\text{frag}} \frac{\phi^{\alpha}_s \phi^{\alpha}_i}{n_i - 1} \delta_{j\in\mathcal{S}_i}, \quad [5]$$

where k_{frag} denotes the fragmentation rate coefficient that we choose, for simplicity, to be phase-independent. The fragmentation rate $h^{\alpha}_{i\rightharpoonup j+|i-j|}$ is proportional to the local volume fraction ϕ^{α}_i (α = I,II) of the fragmenting sequence. Each bond along the sequence backbone fragmentates equally, such that all possible length combinations of the two fragmentated subparts of i occur equally often. This is captured by the fragmentation rate being independent of j, as long as j is a continuous substring of i ($\delta_{j\in\mathcal{S}_i} = 1$), otherwise the rate is zero ($\delta_{j\notin\mathcal{S}_i} = 0$). The factor $(n_i - 1)^{-1}$, where n_i is the number of monomers of sequence i, ensures each bond in the system is equally likely to break. Inspired by the hydrolysis of biofilaments such as DNA and RNA, the fragmentation rate scales with the solvent volume fraction ϕ^{α}_s. In contrast to the studies discussed in the previous section, the system will not relax to thermodynamic equilibrium; rather, it settles in a nonequilibrium steady state, where generally there exists a net reaction r^{α}_i for all sequences in each phase α, which is balanced by the diffusive flux across the phase boundary. An emerging stationary sequence distribution is thus particularly distinct when the system is phase-separated from its mixed counterpart, where the oligomerization flux vanishes exactly at the nonequilibrium steady state.

Breaking detailed balance of the rates in the oligomerization kinetics gives rise to a selection pressure that strongly changes the sequence distribution compared to thermodynamic equilibrium. For the parameters studied, we always observe a nonequilibrium steady state with a stationary sequence distribution. For low fragmentation rate coefficients k_{frag}, alternating sequences are favored, and the sequence distribution remains almost unchanged compared to thermodynamic equilibrium, as shown in Fig. 6 and the yellow distribution in Fig. 7*A*. For intermediate values of k_{frag} (light green), the thermodynamic sequence bias is no longer evident, and the nonequilibrium sequence distribution at steady state is almost flat with approximately equal volume fractions for each sequence. For large fragmentation rate coefficients, we find that the thermodynamic bias of sequences observed for

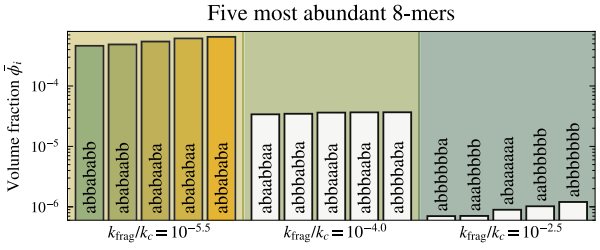

Fig. 6. Five most abundant sequences when phase separation is maintained away from equilibrium by fragmentation. With increasing fragmentation rate coefficient k_{frag}, the selection bias changes from alternating sequences to sequences with longer blocks of the same monomer type. For very large k_{frag}, these blocks approach the maximal sequence length, and the most abundant sequences deviate from homopolymers within 1 to 2 Hamming distances.

vanishing k_{frag} inverts, i.e., sequences that are most depleted at thermodynamic equilibrium become most abundant in the nonequilibrium steady state, and vice versa.

The inverted thermodynamic bias for large fragmentation rate coefficients k_{frag} originates from the thermodynamically unfavorable oligomers having the largest oligomerization rates. A sequence i is at its steady state when its net oligomerization flux r_i and its fragmentation flux h_i balance, that is $r_i = \sum_k h_{i\to k+|i-k|}$. The flux r_i scales exponentially with the chemical potential μ_j of sequences j that form i upon monomer-pickup $j + m = i$. Thus, sequences j that are more homopolymeric, increasing their chemical potential, balance the fragmentation rate h_i at larger values of ϕ_i. The result is, following the derivation in *SI Appendix*, section 8, a nonequilibrium steady state (NESS) sequence distribution in the oligomer-rich phase (II):

$$\phi_i^{\text{II,NESS}} \simeq \frac{k_c}{k_{\text{frag}}} \frac{\exp\left(\frac{\mu_m}{k_B T}\right)}{\frac{V^{\text{II}}}{V}\phi_s^{\text{II}}} \sum_{j\in\mathcal{R}_i} \exp\left(\frac{\mu_j}{k_B T}\right), \qquad [6]$$

where μ_m is the monomer chemical potential (*SI Appendix*, Eq. **S9**). We note that the effects of other sequences j on the steady state abundance of sequence i is described by the term $\sum_{j\in\mathcal{R}_i} \exp\left(\mu_j/(k_B T)\right)$. Thus, a sequence i is more abundant if the set of sequences that oligomerize into i ($j \in \mathcal{R}_i$) have unfavorable interactions with their environment, reflected in a larger value of their chemical potential μ_j. In other words, nonequilibrium fragmentation selects for sequences of lower cooperativity, that are homo-oligomer-like sequences. This selection mechanism is the opposite of the thermodynamic enrichment mechanism that favors sequences of higher cooperativity.

For increasing fragmentation rate coefficient k_{frag}, the most abundant sequences change from alternating oligomers to oligomers containing low complexity domains composed of the same monomeric unit. This trend is illustrated by the five most abundant sequences (Fig. 6) and by the probability distribution of *a*-*b*-neighbors (Fig. 7*B*). We see that the five most abundant sequences for weak nonequilibrium driving ($k_{\text{frag}}/k_c = 10^{-5.5}$) are alternating sequences with maximally two monomers of the same type neighboring each other. In this case, the *a*-*b*-neighbor distribution is skewed toward an *a*-*b*-proportion of 1, indicating that most *a*-monomers are neighbored by *b* monomers. Increasing k_{frag} to intermediate values enhances the length of domains composed of the same monomer type. For example, the five most abundant sequences contain sequences similar to *abbbaba*, containing both alternating and homo-oligomeric parts. The consequence is a $n_{a,b}$-distribution peaked around a half, but less spread than the flat distribution. When increasing the fragmentation rate coefficient k_{frag} further, domains composed of the same monomer type approach the max length, leading to almost homopolymeric sequences such as *abbbbbbb* or *abaaaaaa*. Consistently, the *a*-*b*-distribution is more skewed toward a small fraction of *a*-*b*-neighbors for increasing k_{frag}. To quantify the shift from alternating to sequences with more extended domains of the same monomer type, we use the correlation function defined in Eq. **4**. Fig. 7*C* confirms that faster fragmentation leads to an increase in the wave number of the correlation function, corresponding to longer correlated domains of the same monomer type.

For weak nonequilibrium driving the exponential decay of the rank distribution of 8-mers, $p(k) \simeq \exp\{(-k/\lambda)\}$ persists, as displayed in Fig. 7*D*. However, for intermediate fragmentation rates k_{frag}, the decay length λ increases to a maximum before decreasing. At the maximum, the sequence distribution is almost flat. A steeper rank distribution (smaller λ) indicates a stronger selection pressure, where only a small subset of the sequences are selected. For very large fragmentation rate coefficients k_{frag}, the rank distribution approximately follows a power-law $p(k) \propto k^{-\beta}$ with $\beta \simeq 3$.

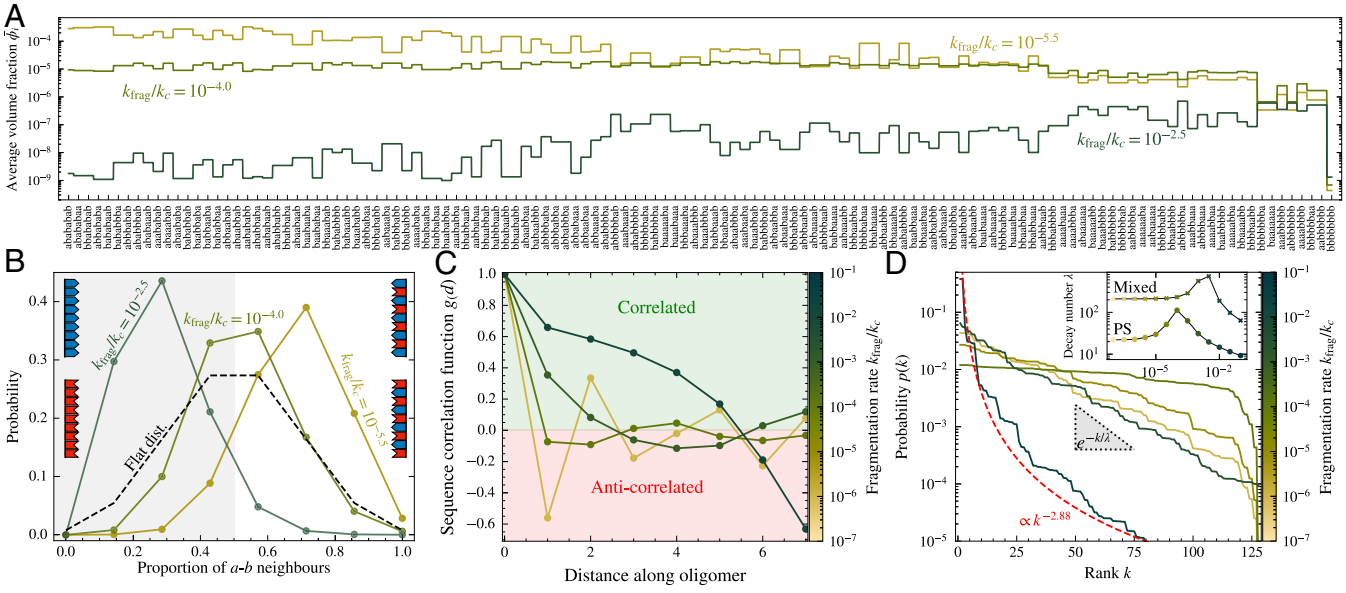

Fig. 7. Nonequilibrium fragmentation creates a selection pressure favoring sequences with longer blocks of the same monomer type. (*A* and *B*) With increasing fragmentation rate coefficient k_{frag}, the selection bias changes from alternating sequences to sequences with longer blocks of the same monomer type. For very large k_{frag}, these blocks approach the maximal sequence length (8mers in *A*), and the most abundant sequences deviate from homopolymers within 1 to 2 Hamming distances. (*B* and *C*) The selection trend for increasing k_{frag} from alternating sequences to sequences with larger blocks is supported by (*B*) showing a pronounced shift toward less *a*-*b*-neighbors and a pronounced increase of the wave number of the sequence correlation function $g(d)$ (see *C*). (*D*) The rank distribution $p(k) \propto \exp(-k/\lambda)$ decays exponentially for smaller fragmentation rate coefficients k_{frag}. The characteristic rank scale λ (*Inset*) has a nonmonotonous behavior and is maximal for intermediate values of k_{frag} because the sequence distribution is approximately flat (*A*). For the very large fragmentation rate coefficients k_{frag}, the rank distribution follows a power-law decay with $p(k) \propto k^{-\beta}$ illustrated by the exponent $\beta \simeq 2.88$ for $k_{\text{frag}}/k_c = 10^{-2.5}$.

Power-laws in rank distributions were reported for many complex systems such as city sizes, gene circuits, and self-replicating systems, and were shown as a requirement for open-ended evolution (54). Though the exponent is different from Zipf's law (55), future studies will scrutinize whether our phase-separated system with nonequilibrium oligomerization is complex enough to enable open-ended evolution (54). We note that without phase separation (mixed case), there is also a power-law behavior of the rank distribution for large k_{frag} with an exponent close to Zipf's law (*SI Appendix*, Fig. S3).

Maintaining oligomerization away from equilibrium gives rise to a strong selection mechanism, where only a small subset of all possible sequences accumulate most of the total oligomer volume fraction for that length. This effect is most pronounced for longer sequences. We characterize these observations by the quantity $\#_{90}(l)$ that is the smallest number of sequences corresponding to 90% volume fraction for a given length l (Fig. 8*A*):

$$\frac{\sum_{i\in\sigma}^{\#_{90}(l)} \phi_i \delta_{n_i,l}}{\sum_{i\in\sigma} \phi_i \delta_{n_i,l}} > 0.90\,. \qquad [7]$$

This quantity characterizes the effectively occupied sequence space. At thermodynamic equilibrium, and independent of the interaction strength e_{ab}, $\#_{90}(l)$ scales exponentially in sequence length l because longer length leads to exponentially more sequences (see Fig. 8*A* for low k_{frag}). Strikingly, the selection pressure, once driven away from equilibrium, gets so strong for fast enough fragmentation (large k_{frag}) such that $\#_{90}(l)$ scales even subexponentially (Fig. 8*A*). Though the sequence distribution is less diverse when maintaining oligomerization away from equilibrium, the selection pressure facilitates specific sequence motifs to win over the others. A relevant selection mechanism has to oppose the exponentially growing sequence space. In this light, our findings of subexponential scalings highlight the potential of phase separation with oligomerization maintained away from equilibrium in selecting specific oligomer sequences already during their oligomerization kinetics.

Selection mediated by phase separation can be further corroborated by introducing a cooperativity measure. The cooperativity Λ_i per sequence length n_i characterizes how well sequence i interacts with its environment:

$$\Lambda_i = \gamma_i^{-n_i}\left(\{\phi_j\}\right)\exp\left\{-\frac{\mu_i^0}{n_i\, k_B T}\right\}. \qquad [8]$$

Here, the chemical potentials have been decomposed in terms of the composition-independent reference chemical potential μ_i^0, and the composition-dependent activity coefficient γ_i, such that $\mu_i = \mu_i^0 + k_B T \log(\phi_i \gamma_i)$, with the expressions given in *SI Appendix*, Eq. **S14**. Chemical equilibrium enriches oligomers that interact well with sequences in their respective phase, corresponding to a larger average cooperativity (Fig. 8*B*). The longer an oligomer, the more cooperative segments exist at which other sequences can attach. We note that without phase separation (mixed case), sequences are relatively uncooperative and not well-adapted to their mixed environment. The increasing cooperativity trend with sequence length in the phase-separated case at thermodynamic equilibrium changes when maintaining oligomerization away from equilibrium. The cooperativity decreases as the selection pressure inverts the thermodynamic enrichment, leading to less alternating and more homopolymeric sequences. In other words, the faster the fragmentation, the more uncooperative sequences are selected. Cooperative sequences interact well with almost all other sequences. Selection of less cooperative sequences with longer repeating blocks of the same nucleotide can mediate more specific interactions, giving rise to secondary structure such as hairpins (57).

Finally, we ask whether the selected sequences store more or less information. We quantify the information content of each sequence S_i by the number of *a*-*b*-neighbors using Shannon's measure of information (58, 59):

$$S_i = -\sum_{j\in\sigma} p(j|n_{a,b}^{(i)}) \log_2\left(p(j|n_{a,b}^{(i)})\right), \qquad [9]$$

where $p(j|n_{a,b}^{(i)})$ is the probability of sequence j given the number of *a*-*b*-neighbors $n_{a,b}^{(i)}$ of sequence i, without prior knowledge about the sequence distribution. No prior knowledge corresponds to a flat sequence distribution. This means that each sequence of the given length with a specific $n_{a,b}^{(i)}$ value is equally likely. In this

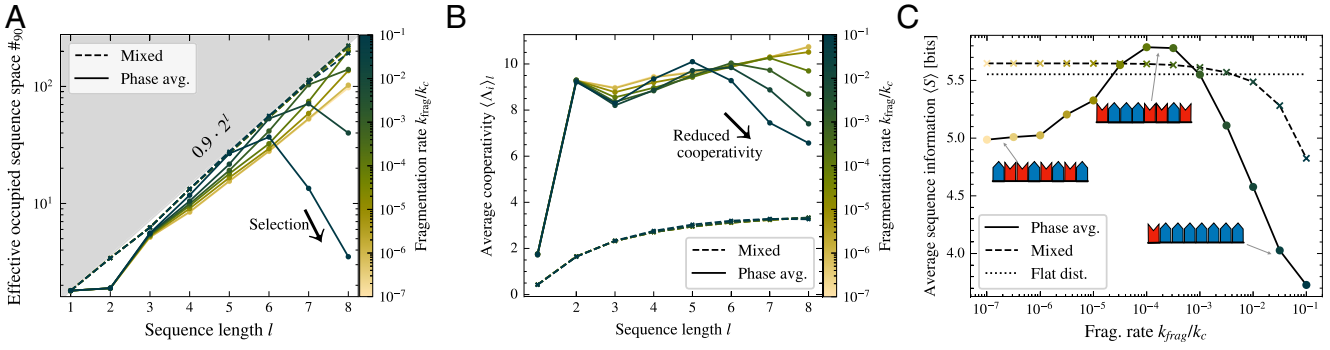

Fig. 8. Nonequilibrium fragmentation selects for less cooperative sequences of lower information. (*A*) The number of sequences $\#_{90}(l)$ occupying 90% of the oligomer volume of a given length l (Eq. **7**) deviates from the exponential scaling as observed at thermodynamic equilibrium (vanishing values of k_{frag}). For the large fragmentation rate coefficients k_{frag}, $\#_{90}(l)$ even decreases for long enough sequences, indicating a strong selection pressure selecting a subexponential subset of the possible sequence for a given length l. (*B*) The cooperativity characterizes how well sequences of each length interact with their environment (Eq. **8**). Thermodynamics promotes cooperativity (low k_{frag}) and increases with sequence length l as the sequences have more freedom to adapt their structure to suit their environment. Increasing the fragmentation rate coefficient k_{frag} changes the selection pressure to promote oligomers with uncooperative substrates, reducing the average cooperativity at large l. (*C*) The average sequence information over the distribution of 8-mers Eq. **9** has a local maximum with the fragmentation rate coefficient k_{frag}. In the limit of low and high fragmentation rates, the average information content of the sequence distribution decreases due to the selection of alternating sequences or sequences of larger homopolymeric blocks, respectively. Similar behavior is found for different measures of sequence information/complexity, such as the Kolmogorov complexity (54) and the Lempel-Ziv complexity (56).

case, Eq. **9** gives the maximal value of

$$S_i^* = -\log_2 \{p(i|n_{a,b}^{(i)})\}. \quad [10]$$

Thus, S_i^* quantifies the information gained in bits by knowing the exact sequence, with the only prior information being the number of *a-b* neighbors corresponding to sequence *i*. For example, there are only two sequences of length $l = 8$ that are homo-oligomers ($n_{a,b} = 0$). Thus, the probability of picking the correct homo-oligomer is one-half, yielding an information gain of 1 bit by finding out the actual sequence. Similarly, the sequence *abbbaaba* has a probability of 1.4% among all sequences of four *a-b*-neighbors, yielding an information gain of 6.1 bits. Thus, S_i^* assigns the same value of information to the sequences with the same value of $n_{a,b}$. The average information of a sequence distribution is calculated as $\langle S \rangle = \sum_{i\in\sigma} \delta_{n_i,l} \bar{\phi}_i S_i^* / \sum_{i\in\sigma} \delta_{n_i,l} \bar{\phi}_i$.

Phase-separated systems away from equilibrium change the information content of the selected sequences. Fig. 8*C* shows the average sequence information of 8-mers, $\langle S \rangle$, at the nonequilibrium steady state for increasing nonequilibrium fragmentation rate coefficient k_{frag}. For phase-separated systems, $\langle S \rangle$ has a nonmonotonous dependency on k_{frag}. Close to thermodynamic equilibrium (low k_{frag}), phase separation favors low-information sequences, i.e., alternating sequences. Increasing the fragmentation rate coefficient k_{frag} leads to an increase in the average sequence information, exceeding the average for a flat distribution of 8-mers. This is because there is an enrichment of sequences with 3 and 4 *a-b*-neighbors at the expense of homo-oligomers and alternating oligomers (Fig. 7*B*), which contain less information. Note that no such local maxima is observed without phase separation, where a monotonic decrease in average information content with fragmentation rate k_{frag} is observed. The same behavior is observed with alternative measures, such as Kolmogorov complexity (60), Lempel-Ziv (56), Lempel-Ziv-Welch (61), and, the information content of the distribution (58). Increasing k_{frag} further, in the case with phase separation, selects for longer blocks of the same monomer type, strongly decreasing the average information content. Such low-information sequences are interesting because they have the potential to show rich functionalities, such as catalytic activity or folding, and the extended blocks can mediate specific, lock-like interactions (57, 62–64).

Discussion

In our work, we propose a theoretical framework for the oligomerization kinetics of sequences at nondilute conditions. This framework builds on nonequilibrium thermodynamics and is used to explore the dynamics in high-dimensional sequence spaces. Using our framework, we show that the oligomerization kinetics triggers phase separation of sequences, which in turn creates a strong bias for the enrichment and depletion of specific sequences. A key finding is that, sequence distributions at thermodynamic equilibrium favors more cooperative sequences that interact well with their environment, while nonequilibrium fragmentation selects for low-information, less cooperative sequences.

We now discuss the two key prerequisites for the selection mechanism of our work, namely phase separation and nonequilibrium fragmentation, to highlight which experimental systems can be used to test our results at prebiotically consistent conditions. A relevant experimental setting is nucleotide mixtures composed of DNA or RNA sequences phase-separating into oligomer-rich and oligomer-poor phases. In general, when sequences are long and their concentrations are large enough, phase separation is favored since the attractive sequence interactions outcompete the mixing entropy. Moreover, salt concentrations must be high enough to screen the negative charges on the DNA or RNA backbone. Between DNA and RNA, attractive interactions occur through Watson–Crick base pairs; their sequence specificity is a fundamental prerequisite for the sequence-selection dynamics discussed in our work. The experimental conditions under which sequences can polymerize and ligate, and undergo phase transitions via sequence-specific interactions at prebiotically consistent conditions, are subtle and are discussed in detail below. The considered nonequilibrium driving force required for the selection dynamics discussed in this work is fragmentation. Nonequilibrium fragmentation does not need to be engineered experimentally. For RNA, it is an omnipresent property on reasonable experimental time-scales of hours to days (26, 43).

Phase transitions of DNA sequences toward a gel state had been known since the 1960s, when mechanically fragmented DNA the size of 200- to 400mers was upconcentrated beyond about 500 μM (65). In a prebiotic setting, this upconcentration could be achieved by drying (66), or by a temperature gradient using DNA with sticky ends (67). Moreover, defined multistrand constructs have been developed that exhibit a liquid–liquid phase transition by hybridization at much lower strand concentrations, and thus with the inclusion of a large amount of water (68–73). Recently, attempts have been made to extrapolate these experiments to short oligonucleotides or even single bases, leading to liquid DNA crystals (74, 75) or gel-like condensates (27, 33, 67). In ref. 74, the polymerization was triggered using EDC, as in enzyme-free studies on DNA replication (76). However, the requirement for such an activation agent significantly inhibited the formation of a phase transition via base-pairing.

Recently, polymerization from self-activation of 2′,3′-cyclic phosphate nucleotides in alkaline conditions was probed and shown to offer oligomeric RNA of mixed sequences (77–80). Especially the addition of amino acids was reported to have a profound effect on the base composition, leading to a much more balanced sequence distribution, attributed mostly to a general base catalytic opening of the 2′,3′-cyclic phosphate at the nucleotides by the amine group of the amino acids (81). In light of our theory, it is interesting to consider whether these results, obtained under drying conditions, could be affected by base-pairing interactions as expected at higher concentrations. At such higher concentrations, polymerization can occur in a dry state due to the need for molecular proximity. The crucial question is how this dry state is regulated by base pairing that occurs during drying. In many cases, an amorphous dry state was observed (81), indicating the possibility of maintaining base pairing upon drying.

In the future, it would be especially interesting to probe the regime of high concentration just before full drying, for example, by using a high relative humidity. Here, polymerization is expected to occur, albeit at a low rate, while hybridization can still mediate sequence-specific interactions. So far, wet–dry cycles have been optimized toward high yield at high pH of up to pH = 11 (80, 82). However, the hydrolysis of the 2′,3′-cyclic phosphate stops the reaction, and so far, schemes to recycle this activation group have not been implemented. A second setting could be pushing thermal traps to sufficiently high concentrations that phase transitions occur, and precipitation would allow for a dry-like polymerization. Similarly,

under freezing conditions, the upconcentration in water pockets (incorrectly dubbed eutectic ice) could lead to simultaneous polymerization and phase transitions. The identification of specific sequences that enhance slow ligation at neutral pH points in this direction (83), and similar findings are likely to be found for polymerization. Deep sequencing techniques, combined with HPLC and mass spectroscopy, can be used to determine whether more or less cooperative sequences were experimentally selected by calculating, for example, the proportion of *a-b* neighbors or the sequence correlation function (Fig. 7 *B* and *C*). For example, according to our theory, a histogram of the proportion of *a-b* neighbors being asymmetric with a maximum below 0.5 indicates the emergence of less cooperative sequences. We note that both quantities can also be computed experimentally if only a fraction of the sequence space can be sampled, thereby providing a means to test our theoretical predictions.

In summary, our theoretical work tackles the question of how oligonucleotide sequences are selected under varying nonequilibrium conditions—a question that is relevant for the molecular Origin of Life (84, 85), and for designing de novo life systems capable of undergoing Darwinian or Lamarckian evolution (86, 87). In particular, our finding of a subexponential effective occupation of the possible sequence space indicates a very strong selection pressure by simply considering nonequilibrium fragmentation. We expect that extending our theory to study other nonequilibrium driving forces, such as wet–dry, freeze–thaw, or nucleotide cycles (18, 88), may yield a richer selection outcome of sequences. When considering four or even more different monomers, we speculate that the observed trend to select sequences of low cooperativity and low information could give rise to exotic sequence patterns or motifs, which are the prerequisite for target-specific functionalities, such as catalysis (89, 90).

Materials and Methods

The free energy that determines both phase coexistence and reaction kinetics is the Flory–Huggins free energy density. The interaction parameters χ_{ij} are obtained by a Boltzmann average over all possible configurations in which two sequences can interact. A detailed discussion and all explicit expressions are provided in *SI Appendix*, section 2, together with the parameter values used (*SI Appendix*, section 5).

The diffusive fluxes j_i^{α} between the coexisting phases are computed such that the net compositional changes across all components in both phases produce identical shifts in the chemical potentials and in the osmotic pressure. This requirement constrains the dynamics to remain on the binodal (Fig. 2*C*). Our method extends the approach introduced in ref. 41 to multicomponent mixtures with species of different sizes. A full derivation is given in *SI Appendix*, section 7.

Data, Materials, and Software Availability. All other data are included in the article and/or *SI Appendix*. Computer codes on the numerical integration can be found here: https://git.rz.uni-augsburg.de/haugeriv/theory-for-sequence-selection-via-phase-separation-and-oligomerization (91).

ACKNOWLEDGMENTS. We are grateful for the insightful discussions with H. Vuijk, S. Gomez, M. Koul, and G. Granatelli. We are also grateful for the discussions with A. Schmid, F. Dänekamp, S. Wunnava, A. Serrao, and P. Schwintek on how to experimentally scrutinize our predictions. D.B. and C.A.W. thank for support by the German Research Foundation in the framework of the TRR 392: Molecular evolution in prebiotic environments (Project No. 521256690). C.A.W. acknowledges the European Research Council for the financial support under the European Union's Horizon 2020 research and innovation programme ("Fuelled Life," Grant Agreement No. 949021).

Author affiliations: [a]Faculty of Mathematics, Institute of Physics, Natural Sciences, and Engineering, University of Augsburg, Augsburg 86159, Germany; [b]Department of Physics, Universitat de Barcelona, Barcelona 08028, Spain; and [c]Systems Biophysics and Center of Nanoscience, Ludwig Maximilian University, Munich 80799, Germany

1. M. P. Robertson, G. F. Joyce, The origins of the RNA world. *Cold Spring Harb. Perspect. Biol.* **4**, a003608 (2012).
2. K. Kruger *et al*., Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell* **31**, 147–157 (1982).
3. M. J. Fedor, J. R. Williamson, The catalytic diversity of RNAs. *Nat. Rev. Mol. Cell Biol.* **6**, 399–412 (2005).
4. W. Gilbert, Origin of life: The RNA world. *Nature* **319**, 618 (1986).
5. K. R. Birikh, P. A. Heaton, F. Eckstein, The structure, function and application of the hammerhead ribozyme. *Eur. J. Biochem.* **245**, 1–16 (1997).
6. C. Deck, M. Jauker, C. Richert, Efficient enzyme-free copying of all four nucleobases templated by immobilized RNA. *Nat. Chem.* **3**, 603–608 (2011).
7. A. Mariani, D. A. Russell, T. Javelle, J. D. Sutherland, A light-releasable potentially prebiotic nucleotide activating agent. *J. Am. Chem. Soc.* **140**, 8657–8661 (2018).
8. T. Walton, W. Zhang, L. Li, C. P. Tam, J. W. Szostak, The mechanism of nonenzymatic template copying with imidazole-activated nucleotides. *Angew. Chem. (Int. Ed. Engl.)* **58**, 10812–10819 (2019).
9. S. Wunnava *et al*., Acid-catalyzed RNA-oligomerization from 3′,5′-cGMP. *Chemistry* **27**, 17581–17585 (2021).
10. M. J. McCall, P. Hendry, P. A. Jennings, Minimal sequence requirements for ribozyme activity. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 5710–5714 (1992).
11. A. Ianeselli *et al*., Physical non-equilibria for prebiotic nucleic acid chemistry. *Nat. Rev. Phys.* **5**, 185–195 (2023).
12. A. V. Tkachenko, S. Maslov, Spontaneous emergence of autocatalytic information-coding polymers. *J. Chem. Phys.* **143**, 045102 (2015).
13. A. V. Tkachenko, S. Maslov, Onset of natural selection in populations of autocatalytic heteropolymers. *J. Chem. Phys.* **149**, 134901 (2018).
14. C. B. Mast, S. Schink, U. Gerland, D. Braun, Escalation of polymerization in a thermal gradient. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 8030–8035 (2013).
15. T. Matreux, P. Aikkila, B. Scheu, D. Braun, C. B. Mast, Heat flows enrich prebiotic building blocks and enhance their reactivity. *Nature* **628**, 110–116 (2024).
16. H. M. Fares, A. E. Marras, J. M. Ting, M. V. Tirrell, C. D. Keating, Impact of wet-dry cycling on the phase behavior and compartmentalization properties of complex coacervates. *Nat. Commun.* **11**, 5423 (2020).
17. O. R. Maguire, I. B. A. Smokers, W. T. S. Huck, A physicochemical orthophosphate cycle via a kinetically stable thermodynamically activated intermediate enables mild prebiotic phosphorylations. *Nat. Commun.* **12**, 5517 (2021).
18. I. S. Haugerud, P. Jaiswal, C. A. Weber, Nonequilibrium wet-dry cycling acts as a catalyst for chemical reactions. *J. Phys. Chem. B* **128**, 1724–1736 (2024).
19. H. Mutschler, A. Wochner, P. Holliger, Freeze-thaw cycles as drivers of complex ribozyme assembly. *Nat. Chem.* **7**, 502–508 (2015).
20. M. J. Eleveld *et al*., Competitive exclusion among self-replicating molecules curtails the tendency of chemistry to diversify. *Nat. Chem.* **17**, 132–140 (2024).
21. K. Liu *et al*., Light-driven eco-evolutionary dynamics in a synthetic replicator system. *Nat. Chem.* **16**, 79–88 (2024).
22. M. Morasch *et al*., Heated gas bubbles enrich, crystallize, dry, phosphorylate and encapsulate prebiotic molecules. *Nat. Chem.* **11**, 779–788 (2019).
23. R. Mizuuchi *et al*., Mineral surfaces select for longer RNA molecules. *Chem. Commun.* **55**, 2090–2093 (2019).
24. P. W. Kudella, A. V. Tkachenko, A. Salditt, S. Maslov, D. Braun, Structured sequences emerge from random pool when replicated by templated ligation. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2018830118 (2021).
25. A. Calaça Serrão, F. T. Dänekamp, Z. Meggyesi, D. Braun, Replication elongates short DNA, reduces sequence bias and develops trimer structure. *Nucleic Acids Res.* **52**, 1290–1297 (2024).
26. A. Calaça Serrão *et al*., High-fidelity RNA copying via 2′,3′-cyclic phosphate ligation. *J. Am. Chem. Soc.* **146**, 8887–8894 (2024).
27. G. Bartolucci *et al*., Sequence self-selection by cyclic phase separation. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2218876120 (2023).
28. J. B. S. Haldane, The origin of life. *Ration. Annu.* **148**, 3–10 (1929).
29. A. I. Oparin, *The Origin of Life* (Dover Publications Inc., New York, NY, ed. 2, 1938).
30. S. W. Fox, The evolutionary significance of phase-separated microsystems. *Orig. Life* **7**, 49–68 (1976).
31. S. Tirard, JBS Haldane and the origin of life. *J. Genet.* **96**, 735–739 (2017).
32. T. Hu, N. Chitnis, D. Monos, A. Dinh, Next-generation sequencing technologies: An overview. *Hum. Immunol.* **82**, 801–811 (2021).
33. Y. Sato, T. Sakamoto, M. Takinoue, Sequence-based engineering of dynamic functions of micrometer-sized DNA droplets. *Sci. Adv.* **6**, eaba3471 (2020).
34. Z. Xing *et al*., Microrheology of DNA hydrogels. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 8137–8142 (2018).
35. M. Nakata *et al*., End-to-end stacking and liquid crystal condensation of 6- to 20-base pair DNA duplexes. *Science* **318**, 1276–1279 (2007).
36. G. Zanchetta, M. Nakata, M. Buscaglia, N. A. Clark, T. Bellini, Liquid crystal ordering of DNA and RNA oligomers with partially overlapping sequences. *J. Physics: Condens. Matter* **20**, 494214 (2008).
37. W. M. Aumiller, F. Pir Cakmak, B. W. Davis, C. D. Keating, RNA-based coacervates as a model for membraneless organelles: Formation, properties, and interfacial liposome assembly. *Langmuir* **32**, 10042–10053 (2016).
38. Bj. Jeon *et al*., Salt-dependent properties of a coacervate-like, self-assembled DNA liquid. *Soft Matter* **14**, 7009–7015 (2018).
39. G. Bartolucci, I. S. Haugerud, T. C. Michaels, C. A. Weber, The interplay between biomolecular assembly and phase separation. *eLife* **13**, RP93003 (2024).

40. D. M. Mitrea, R. W. Kriwacki, Phase separation in biology: Functional organization of a higher order. *Cell Commun. Signal.* **14**, 1 (2016).
41. J. Bauermann, S. Laha, P. M. McCall, F. Jülicher, C. A. Weber, Chemical kinetics and mass action in coexisting phases. *J. Am. Chem. Soc.* **144**, 19294–19304 (2022).
42. D. Zwicker, The intertwined physics of active chemical reactions and phase separation. *Curr. Opin. Colloid Interface Sci.* **61**, 101606 (2022).
43. T. Göppel, J. H. Rosenberger, B. Altaner, U. Gerland, Thermodynamic and kinetic sequence selection in enzyme-free polymer self-assembly inside a non-equilibrium RNA reactor. *Life* **12**, 567 (2022).
44. N. Van Kampen, Nonlinear irreversible processes. *Physica* **67**, 1–22 (1973).
45. F. Jülicher, A. Ajdari, J. Prost, Modeling molecular motors. *Rev. Mod. Phys.* **69**, 1269–1282 (1997).
46. C. A. Weber, D. Zwicker, F. Jülicher, C. F. Lee, Physics of active emulsions. *Rep. Prog. Phys.* **82**, 064601 (2019).
47. F. Kilchherr *et al*., Single-molecule dissection of stacking forces in DNA. *Science* **353**, aaf5508 (2016).
48. S. E. Toal *et al*., Randomizing the unfolded state of peptides (and proteins) by nearest neighbor interactions between unlike residues. *Chemistry* **21**, 5173–5192 (2015).
49. Y. Mu, Y. Q. Gao, Effects of hydrophobic and dipole-dipole interactions on the conformational transitions of a model polypeptide. *J. Chem. Phys.* **127**, 105102 (2007).
50. B. Dutagaci *et al*., Charge-driven condensation of RNA and proteins suggests broad role of phase separation in cytoplasmic environments. *eLife* **10**, e64004 (2021).
51. A. Jain, R. D. Vale, RNA phase transitions in repeat expansion disorders. *Nature* **546**, 243–247 (2017).
52. C. Roden, A. S. Gladfelter, RNA contributions to the form and function of biomolecular condensates. *Nat. Rev. Mol. Cell Biol.* **22**, 183–195 (2021).
53. D. Tauber, G. Tauber, R. Parker, Mechanisms and regulation of RNA condensation in RNP granule formation. *Trends Biochem. Sci.* **45**, 764–778 (2020).
54. B. Corominas-Murtra, L. F. Seoane, R. Solé, Zipf's law, unbounded complexity and open-ended evolution. *J. R. Soc. Interface* **15**, 20180395 (2018).
55. G. K. Zipf, *Selected Studies of the Principle of Relative Frequency in Language* (Harvard University Press, Cambridge, MA/London, England, 1932).
56. A. Lempel, J. Ziv, On the complexity of finite sequences. *IEEE Trans. Inf. Theory* **22**, 75–81 (1976).
57. A. Ciesiolka, M. Jazurek, K. Drazkowska, W. J. Krzyzosiak, Structural characteristics of simple RNA repeats associated with disease and their deleterious protein interactions. *Front. Cell. Neurosci.* **11**, 97 (2017).
58. C. E. Shannon, A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423 (1948).
59. A. Ben-Naim, D. Casadei, *Modern Thermodynamics* (World Scientific, Hackensack, NJ, 2017).
60. M. Li *et al*., *An Introduction to Kolmogorov Complexity and Its Applications* (Springer, 2008), vol. 3.
61. T. A. Welch, A technique for high-performance data compression. *Computer* **17**, 8–19 (1984).
62. M. Buiatti, C. Acquisti, G. Mersi, P. Bogani, M. Buiatti Jr., "The biological meanings of DNA correlations" in *Fractals in Biology and Medicine*, G. A. Losa, D. Merlini, T. F. Nonnenmacher, E. R. Weibel, Eds. (Springer, 2002), pp. 235–245.
63. M. Ackermann, L. Chao, DNA sequences shaped by selection for stability. *PLoS Genet.* **2**, e22 (2006).
64. A. Bacolla *et al*., Abundance and length of simple repeats in vertebrate genomes are determined by their structural properties. *Genome Res.* **18**, 1545–1553 (2008).
65. V. Luzzati, A. Nicolaieff, Etude par diffusion des rayons x aux petits angles des gels d'acide désoxyribonucléique et de nucléoprotéines:(note préliminaire). *J. Mol. Biol.* **1**, 127-IN5 (1959).
66. T. Bellini *et al*., Liquid crystal self-assembly of random-sequence DNA oligomers. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 1110–1115 (2012).
67. M. Morasch, D. Braun, C. B. Mast, Heat-flow-driven oligonucleotide gelation separates single-base differences. *Angew. Chem.* **128**, 6788–6791 (2016).
68. S. Biffi *et al*., Phase behavior and critical activated dynamics of limited-valence DNA nanostars. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 15633–15637 (2013).
69. S. Agarwal, D. Osmanovic, M. A. Klocke, E. Franco, The growth rate of DNA condensate droplets increases with the size of participating subunits. *ACS Nano* **16**, 11842–11851 (2022).
70. S. Wilken, A. Chaderjian, O. A. Saleh, Spatial organization of phase-separated DNA droplets. *Phys. Rev. X* **13**, 031014 (2023).
71. O. A. Saleh, S. Wilken, T. M. Squires, T. Liedl, Vacuole dynamics and popping-based motility in liquid droplets of DNA. *Nat. Commun.* **14**, 3574 (2023).
72. G. Fabrini *et al*., Co-transcriptional production of programmable RNA condensates and synthetic organelles. *Nat. Nanotechnol.* **19**, 1665–1673 (2024).
73. F. Giessler *et al*., Growth, dissolution and segregation of genetically encoded RNA droplets by ribozyme catalysis. bioRxiv [Preprint] (2025). https://doi.org/10.1101/2025.08.29.673008 (Accessed 19 November 2025).
74. M. Todisco *et al*., Nonenzymatic polymerization into long linear RNA templated by liquid crystal self-assembly. *ACS Nano* **12**, 9750–9762 (2018).
75. S. Marni, T. P. Fraccia, T. Bellini, Random-sequence DNA oligomers make liquid crystals: A case of collective ordering in a superdiverse environment. *ACS Nano* **18**, 34034–34043 (2024).
76. E. Edeleva *et al*., Continuous nonenzymatic cross-replication of DNA strands with in situ activated DNA oligonucleotides. *Chem. Sci.* **10**, 5807–5814 (2019).
77. A. V. Dass *et al*., RNA oligomerisation without added catalyst from 2′,3′-cyclic nucleotides by drying at air-water interfaces. *ChemSystemsChem* **5**, e202200026 (2023).
78. C. F. Dirscherl *et al*., A heated rock crack captures and polymerizes primordial DNA and RNA. *Phys. Chem. Chem. Phys.* **25**, 3375–3386 (2023).
79. E. Tekin *et al*., Prebiotic foam environments to oligomerize and accumulate RNA. *ChemBioChem* **23**, e202200423 (2022).
80. F. Caimi *et al*., High-yield prebiotic polymerization of 2′,3′-cyclic nucleotides under wet-dry cycling. *ACS Cent. Sci.* **11**, 1546–1557 (2025).
81. S. K. Rout *et al*., Amino acids catalyse RNA formation under ambient alkaline conditions. *Nat. Commun.* **16**, 5193 (2025).
82. X. Song, P. Simonis, D. Deamer, R. N. Zare, Wet-dry cycles cause nucleic acid monomers to polymerize into long chains. *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2412784121 (2024).
83. H. Mutschler *et al*., Random-sequence genetic oligomer pools display an innate potential for ligation and recombination. *eLife* **7**, e43022 (2018).
84. A. Ianeselli *et al*., Physical non-equilibria for prebiotic nucleic acid chemistry. *Nat. Rev. Phys.* **5**, 185–195 (2023).
85. I. Budin, J. W. Szostak, Expanding roles for diverse physical phenomena during the origin of life. *Annu. Rev. Biophys.* **39**, 245–263 (2010).
86. S. Otto, An approach to the de novo synthesis of life. *Acc. Chem. Res.* **55**, 145–155 (2022).
87. C. M. Kriebisch *et al*., A roadmap toward the synthesis of life. *Chem* **11**, 102399 (2025).
88. P. Jaiswal, I. S. Haugerud, H. D. Vuijk, C. A. Weber, Harvesting chemical power from cyclic environments. arXiv [Preprint] (2025). https://arxiv.org/abs/2510.10266 (Accessed 12 November 2025).
89. C. E. Weinberg, Z. Weinberg, C. Hammann, Novel ribozymes: Discovery, catalytic mechanisms, and the quest to understand biological function. *Nucleic Acids Res.* **47**, 9480–9494 (2019).
90. R. Micura, C. Höbartner, Fundamental studies of functional nucleic acids: Aptamers, riboswitches, ribozymes and DNAzymes. *Chem. Soc. Rev.* **49**, 7331–7353 (2020).
91. I. S. Haugerud, Theory for sequence selection via phase separation and oligomerization. GitLab. https://git.rz.uni-augsburg.de/haugeriv/theory-for-sequence-selection-via-phase-separation-and-oligomerization. Deposition 15 January 2026.