Transient States During Oligonucleotides Hybridization: Insights from Annealing Kinetics of Mismatches and Bulges

Marco Todisco[1], Dian Ding[2,3] and Jack W. Szostak[1*]
[1] Howard Hughes Medical Institute, Department of Chemistry, The University of Chicago, Chicago, Illinois 60637, USA.
[2] Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, USA.
[3] Department of Molecular Biology and Center for Computational and Integrative Biology, Massachusetts General Hospital, 185 Cambridge Street, Boston, Massachusetts 02114, USA.
[*] To whom correspondence should be addressed. Email: jwszostak@uchicago.edu

ABSTRACT
Oligonucleotides hybridization is crucial in various biological, prebiotic and nanotechnological processes, including gene regulation, non-enzymatic primer extension and DNA nanodevices assembly. Although extensive research has focused on the thermodynamics and kinetics of nucleic acids hybridization, the behavior of complex mixtures and the outcome of competition for target binding remain less understood. In this study, we investigate the impact of defects, such as mismatches and bulges, in a 12bp DNA or RNA duplex on its hybridization (k_{on}) and dehybridization (k_{off}) kinetics. Our findings reveal that these defects have a small effect on the hybridization kinetics, while significantly affecting the dehybridization kinetics in a position-dependent manner, with k_{off} changing by 6 orders of magnitudes while k_{on} changes by only one. Building upon this observation, we explore a competition scenario with multiple oligonucleotides, and observe transient low specificity for probe hybridization to its complementary target in solution. We characterize these long-lived metastable states and their evolution toward equilibrium, highlighting how mispaired duplexes can serve as substrates for prebiotically relevant chemical reactions. Our results suggest that transient low accuracy states may spontaneously emerge within all large enough complex systems comprising multiple strands, with potential negative repercussions on gene regulation within the realm of modern biology and the prebiotic preservation of genetic information.

INTRODUCTION
Precise target recognition in nucleic acids hybridization is important for gene regulation (1), PCR (2), *in-situ* hybridization (3, 4), proximity ligation assay (5), non-enzymatic template copying (6), assembly of nanodevices (7, 8) and supramolecular structures (9, 10). Failure in nucleic acid target recognition can lead to misregulation of biological processes (11, 12), amplification of erroneous sequences, and loss of information. Nucleic acid thermodynamics has been extensively studied to predict the hybridization energy for a given set of paired oligonucleotides, leading to the nearest neighbor (NN) model (13–15). The NN model is a reliable and widely used resource in designing the optimal sequences for a set of oligonucleotides in order to minimize mispairings and cross-interactions. Over the years the model has been expanded to include the description of a wide range of possible pairing configurations of multiple strands, involving dangling ends, coaxial stacking, loops, bulges, and mismatches. Systematic studies on internal mismatches and bulges have found them to be destabilizing both in a sequence context-dependent and position-dependent manner, with a more negative effect at internal locations (16–18). One of the major limitations of the NN model is its inherent restriction to the equilibrium description of the system, while neglecting the out-of-equilibrium dynamics, which can behave dramatically differently (19–22). Specifically, duplexes that are not expected at equilibrium may be initially present when multiple strands are mixed, with abundances that depend on the relative hybridization rates of the various competing species. The available literature on the effect of mismatches on hybridization rates is mostly limited to measurements of surface-immobilized oligonucleotides (23–25), showing a position-dependent and sequence context-dependent effect that usually does not affect the hybridization kinetics (k_{on}) by more than one order of magnitude. To the best of our knowledge, the only systematic work on the effect of mismatches measured in solution was performed by Cisse and colleagues (26). They found that hybridization rates of short fluorescently labeled DNA or RNA oligonucleotides enclosed in 200 nm diameter vesicles slows down by up to two orders of magnitude when mismatches are introduced. The authors proposed an empirical rule to explain their results, suggesting that seven contiguous base pairs are required for a fast hybridization of short DNA or RNA oligonucleotides, leading to the formulation of the so-called “rule of seven” (26).

Comparable hybridization rates are a requisite to create metastable states that are significantly different from the final equilibrium state (20). A direct consequence of the dramatic effect of mismatches in hybridization rates as determined by Cisse and colleagues would be that almost no mismatched duplexes are expected in solution at any time if these can be outcompeted by perfectly paired duplexes. If this is true, the equilibrium description is always well-suited for such systems and cross-interactions are negligible.
In this work, we present a systematic study on the effect of single mismatches, double mismatches, and bulges (collectively referred to as "defects") on the hybridization and dehybridization kinetics of short DNA or RNA duplexes in solution. To evaluate the kinetic effects of these defects, we used a combination of fluorescence-based techniques and exploited the emission of the adenine analogue 2-aminopurine. Finally, we tested the null hypothesis of equivalence between the equilibrium and out-of-equilibrium distribution of defected and perfectly paired duplexes in competing mixtures, and we studied their behavior in the presence of chemically reactive species. These results reveal the existence of a transient low accuracy state in oligonucleotide hybridization in complex mixtures: a metastable state in which otherwise absent defected duplexes can behave as substrates for subsequent reactions. This understanding enhances our ability to predict and manipulate nucleic acid hybridization processes.

MATERIALS AND METHODS

General

All DNA oligonucleotides were acquired from IDT with standard desalting and used as received. RNA oligonucleotides were synthesized in-house. Phosphoramidites were acquired from Chemgenes (Wilmington, MA), controlled pore glass (CPG) columns and all reagents for oligonucleotide synthesis and purification were acquired from Glen Research (Sterling, MA), while reagents used for cleavage and deprotection were acquired from Sigma-Aldrich (St. Louis, MO). All kinetics and competition experiments were performed in 1 M NaCl, 10 mM Tris-HCl buffer pH 8.0 unless otherwise specified. Primer extension and ligation reactions were performed in 100 mM $MgCl_2$, 200 mM Tris-HCl buffer pH 8.0. Buffer mixtures were prepared using Tris (1M), sodium chloride (5M) and magnesium chloride (1M) stock solutions from Invitrogen and pH-adjusted using HCl from Sigma-Aldrich. All fluorescence spectroscopy measurements were performed by tracking the fluorescent emission of 2-aminopurine (2Ap), a fluorescent base analogue of adenine whose quantum yield is reduced when in a double helix (27).

Solid phase synthesis of RNA oligonucleotides

RNA oligonucleotides were synthesized on an Expedite 8909 through conventional solid-phase synthesis and leaving the final 5′-DMT group on. Oligonucleotides synthesized this way were cleaved from the solid support using AMA (1.2 ml, 1:1 mixture of 40% w/w methylamine aqueous solution and 28% w/w NH_4OH aqueous solution) for 30 minutes. After cleavage, the oligonucleotides were incubated at 65°C for 20 minutes for deprotection. The deprotected oligonucleotides were dried for 30 minutes at 40°C in a speed-vac and then lyophilized overnight. The RNA powder was resuspended in DMSO (115 µl), TEA (60 µl) and TEA·3HF (75 µl), then incubated at 65°C for 2.5 hours to remove the TBDMS protecting groups. The resulting oligonucleotides were purified and deprotected from the 5′-DMT groups using GlenPak columns (Glen Research). Oligonucleotide concentrations were determined spectrophotometrically using a NanoDrop 2000 (Thermo Scientific) using the extinction coefficients computed through IDT oligo analyzer (28).

Synthesis of reactive species

The C*C imidazolium-bridged dimer and the 5′-2-aminoimidazole-activated pCGCA (*CGCA) were synthesized as previously described (29). To synthesize the imidazolium-bridged C*C dimer, cytidine monophosphate (1 equivalent) was mixed with 2-aminoimidazole (2AI) (0.45 equivalents) in dry dimethyl sulfoxide (DMSO). To this mixture, triethylamine (TEA) (20 equivalents), triphenylphosphine (TPP) (10 equivalents), and 2,2-dipyridyldisulfide (DPDS) (10 equivalents) were added. After a 30-minute incubation, the resulting product was precipitated using acetone with sodium perchlorate. The precipitant was then washed twice with a mixture of acetone and diethyl ether (1:1, v/v) and dried under vacuum. Subsequently, the dry pellet was resuspended in 20 mM triethylamine-bicarbonate buffer (TEAB), pH 8.0, and purified using reverse-phase flash chromatography on a 50 g C18Aq column over 12 column volumes (CVs) of 0-10% acetonitrile in 2 mM TEAB buffer (pH 8.0).
For the activated tetranucleotide *CGCA, the 5′-monophosphate oligonucleoside was activated by using 2AI (40 equivalents), TPP (40 equivalents), and DPDS (40 equivalents) in DMSO and TEA for 6 hours, as

described above. The resulting mixture was purified using preparative scale HPLC on a C18 reverse phase column over 27 CVs of 0-10% acetonitrile in 25 mM TEAB buffer (pH 7.5).

Stopped-flow measurements of hybridization kinetics

Hybridization constants for pairs of interacting oligonucleotides were measured using a Jasco FP-8500 Spectrofluorometer equipped with the SFS-852T Stopped-Flow accessory. All experiments were performed by acquiring the time-dependent emission of a 2Ap-containing strand while adding increasing quantities of complementary oligonucleotide at concentrations typically ranging from ≈ 0.5 μM to ≈ 5 μM. To analyze the data, the fluorescent traces were normalized for the signal acquired in a control experiment performed with the fluorescent oligonucleotide mixed with buffer only. To extract bimolecular kinetic constants, every measurement was independently fit in MATLAB as a bimolecular reaction, modeled using a set of differential equations solved with the variable-step solver *ODE15s* (see Supplementary Data 1 and Supplementary Figure 1A for more details).

Analytical model for hybridization kinetics

In this work the annealing process is modeled following the nucleation-zippering framework (30, 31), thus assuming that the hybridization process is controlled by the formation of a nucleus composed of a few base-pairs between the two oligonucleotides, followed by the rapid zippering of the two strands. To capture the effect of defects on this process, we used a variation of the simple analytical model that has recently been derived and employed (20, 32) for the study of nucleation-limited hybridization in short DNA and RNA oligonucleotides and accounting for the role of nucleation energy, nucleation rate and zippering rate in the overall hybridization kinetics.

$$k_{on} = k_{nuc} \sum_{i=1}^{L_A-n+1} \sum_{j=1}^{L_B-n+1} \frac{1}{1+e^{-\gamma+\Delta G_{i,j}/RT}} \qquad \text{(Eq. 1)}$$

with k_{nuc} and γ being respectively the rates at which any nucleus of size n is formed and the nucleus free energy required to successfully zipper up half of the times that the hybridization is initiated. The double sum is used to account for all nucleation sites of size n over the oligonucleotide A of length L_A and the oligonucleotide B of length L_B. The nucleation free energy $\Delta G_{i,j}$ for non-complementary segments was fixed as equal to infinity, while for complementary segments it was computed using Nupack 4.0 at 25°C and 1M NaCl, so that the model appropriately scales the probability of detaching after nucleation according to the nucleus binding strength. It is noteworthy that adding the length correction originally proposed in the implementation by Hertel and colleagues (32) reduces the model accuracy for our data, so it has been removed here.

Numerical simulations for hybridization kinetics

To explicitly account for the individual steps in the zippering process, we employed a double-ended zippering model that treats hybridization as a stochastic process over the Markov-chain of in-register bound states among two strands, similarly to that shown by Schaeffer, Flamm and Mensen (33–35). The propensity for every transition between two states differing by the addition or removal of one base-pair at the ends of the zippering duplex was computed using the Metropolis method (36). Simulations were performed by fixing the unimolecular rate for the formation of a new base-pair at the rate of 4.2×10^8 s^{-1} (33, 37). The hybridization rate at 1 μM of strand concentration was determined from the mean first passage time to complete zippering or detachment as implemented in the "first-step" method of Multistrand (20, 33).

Measurement of oligonucleotides pairing thermodynamics

The dissociation constant at 25°C (K_D, see Table 1 for list of names and definitions) for all oligonucleotides studied in this work was determined by measuring the interaction of a probe sequence with a 2-aminopurine-containing target. The extent of binding was determined by normalizing to the fluorescent emission of 2-aminopurine, which is maximal when the target is in a single stranded-state and minimal when it is duplexed.

For weakly binding oligonucleotides ($K_D \gtrsim 10^{-6}$ M), the K_D was measured by direct titration of the probe oligonucleotide. For tightly binding oligonucleotides, the binding constant at 25°C could not be directly assessed, and it had to be extrapolated by studying the melting of several solutions at different

oligonucleotides concentrations measured with a Jasco FP-8500 Spectrofluorometer equipped with an ETC-815 Peltier temperature-controlled cell holder. For each pair of oligonucleotides, we acquired the fluorescent emission at 370 nm while exciting at 305 nm and increasing the temperature with a gradient of 3°C/min. Every trace was independently fitted to obtain the K_D values at 25°C, as described previously (20). Typical results are depicted in Supplementary Figure 1B. From these values, the mean and standard deviation of the dissociation constants for each duplex were determined from a lognormal distribution. With this approach we found a modest 2Ap destabilization in DNA of +0.9 kcal/mol at room temperature, comparable with literature data (38), and a statistically non-significant destabilization in RNA at room temperature, in agreement with our recent assessment (20).

Measurement of competition between oligonucleotides

To study the competition in mixtures of oligonucleotides, we pre-mixed in buffer our target strand (T) bearing 2-aminopurine and a competitor sequence bearing a defect (C) at twice their final desired concentrations in buffer. To this, an equal volume of the complementary probe sequence (P) in buffer at twice its final desired concentration was added to initiate the reaction.

To measure the extent of hybridization of T:P, we recorded the time-dependent fluorescent signal (F_{mix}) of these mixtures. The signal was normalized to yield the fraction of bound target as follows:

$$\frac{[T:P]}{[T]_{tot}} = \frac{F_{mix} - F_{T:P}}{F_T - F_{T:P}} \qquad \text{(Eq. 2)}$$

Where F_{mix} is the measured fluorescent signal coming from the mixture, $F_{T:P}$ is the signal coming from a positive control solution containing pre-annealed target and probe oligonucleotides and F_T is the signal coming from the negative control solution with the target sequence only. To accurately determine the bound fraction at equilibrium, we annealed the competing mixtures and relative controls at 90°C for 3 minutes and repeated the measurements. Considering that the K_D of T:P in this study is significantly lower than the concentrations of the oligonucleotides employed here, it follows that all unhybridized target strands in solution must be effectively outcompeted by the competitor strand. Consequently, it has to be true that $[T:P]/([T:P]+[C:P]) = [T:P]/[T]_{tot}$. This study allowed us to independently confirm the previously extrapolated 2Ap destabilization in DNA or RNA at room temperature.

Kinetics of non-enzymatic primer extension and ligation in competing mixtures

In order for the reactions to work, we had to switch to a different buffer containing 200 mM Tris pH 8.0 and 100 mM $MgCl_2$, that does not to significantly alter the overall equilibration process (20).

All primer extension and ligation reactions were performed at 10-µL scale, with 1 µM template (probe sequence), 1 µM perfectly matched primer (target sequence), 1 µM mismatched primer (competitor sequence), 200 mM Tris (pH 8.0), and 100 mM $MgCl_2$. For each reaction, the template was first mixed with Tris 8.0 and $MgCl_2$ at the bottom of the tube. Then, the two different primers were loaded separately on the lid or wall of the tube. The stock solution of C*C/*CGCA was prepared freshly and added into the tube, followed by immediate spin down and mixing to initiate the reaction.

To sample the reaction at different time points, 0.5 µL of the reaction mixture was transferred to 25 µL of a quenching buffer composed by 25 mM EDTA, 1X TBE, and 2 µM RNA strands complementary to the template in formamide to dissociate the dye-labeled primers from the template. To analyze the reactions products, these were resolved by 20% (19:1) denaturing PAGE and their images in the green and red channels were acquired using a Typhoon 9410 scanner to detect both FAM-labeled and Cy5-labeled oligos. The reaction yield over time was quantified using the ImageQuant TL software.

RESULTS

Sequence design

To characterize the effect of defects on the behavior of DNA and RNA, we designed a probe of sequence 5′-CACGCATCACCA-3′ and systematically edited its complementary sequence to introduce mismatches or bulges. Every substitution was picked to minimize secondary structures and self-pairings as predicted by Nupack (15).

All target sequences and substitutions used for kinetic and thermodynamic characterization are listed in Supplementary Table 1, with every sequence name referencing the nature and the position of the defects, either mismatches (m) or bulges (b), along the double helix. For example, the sequence b_4 introduces a bulge in position 4 of the probe, and m_3 introduces an internal mismatch in position 3 of the probe.

Whenever an RNA sequence is used in this work, all deoxyribonucleotides are substituted by their respective ribonucleotides, and every deoxythymidine is substituted by a ribouridine. Sequences used to study the kinetics of non-enzymatic primer extension and ligation in competing mixtures are listed in Supplementary Table 2.

Effect of defects on short oligonucleotide hybridization

When two complementary target (T) and probe (P) oligonucleotides are mixed in solution, they quickly hybridize with a second order rate constant k_{on} to produce a double helix T:P. In this work we investigated the effect of mismatches and bulges on hybridization by systematically altering the position and nature of defects, yielding the results shown in Figure 1A. We found that all k_{on} values determined in this work do not vary by more than one order of magnitude with respect to the control sequence, a perfectly pairing 12 bp-long duplex (t12). Moreover, most defects reduce the rate of formation of the duplex, with hybridization rates for short DNA oligonucleotide affected the most by double mismatches, followed by single mismatches and then bulges.

By introducing defects in different positions along the duplex, in this study we have continuously reduced the length of the longest continuous stretch of base pairs down to 4. Interestingly, by doing so we could not measure any clear discontinuity in the measured hybridization rates, suggesting that “the rule of seven” as formulated by Cisse and colleagues (26) does not apply to our system.

In the nucleation-zippering framework, the hybridization is treated as a process controlled by the formation of a nucleus composed of a few base-pairs between the two oligonucleotides, followed by the rapid zippering of the two strands (30, 31). The nucleation rate is generally regarded as the limiting step, with a nucleus forming only every ~10^3 collisions (20, 33) and the total hybridization rate being the result of multiple hybridization pathways that can be initiated at different nucleation sites.

In this framework, a net reduction in the total number of available nucleation sites deriving from the introduction of defects would inevitably reduce the total hybridization rates (20, 32). By systematically computing the number of available nucleation sites as a function of their lengths, we found that the total number of continuous 3bp stretches is well correlated with the hybridization rates measured here (Pearson coefficient = 0.58, Figure 1B), analogously to what was recently shown by Hertel and colleagues for perfectly pairing DNA oligonucleotides (32).

Even though the hybridization rates for perfectly pairing oligonucleotides and oligonucleotides with bulges are especially well correlated with the total number of available nucleation sites, the variation in the hybridization rate for mismatched oligonucleotides is not well captured. We hypothesized that this discrepancy could be due to a reduced zippering rate for mismatched oligonucleotides, which must propagate their zippering across non-canonical base pairings to successfully form a stable double helix, consistent with literature on the folding energy landscape of hairpins with mismatches (39, 40) and a more recent assessment on short DNA duplexes containing abasic sites (41).

To capture this effect, we employed a simple analytical model as described in the Materials and Methods. The model was fit to our data using a single shared nucleation rate and a defect-dependent γ, greatly improving (Pearson coefficient = 0.76, Figure 1C) the model performance and finding a best fitting nucleation rate equal to 2.1×10^6 $M^{-1}s^{-1}$. With this specific choice of parameters, we find that γ is equal to +1.7 kcal/mol for perfectly paired duplexes, +2.0 kcal/mol for bulged duplexes and -5.8 kcal/mol for mismatched sequences, reflecting a significantly slower zippering rate for mismatches but not for bulges. It is important to emphasize how this simplified model does not explicitly consider each individual step of the zippering process. To obtain a more physically grounded insight in the phenomenon, we implemented a simple numerical simulation as described in the Materials and Methods, explicitly accounting for the whole zippering trajectory starting from any possible nucleus composed by a single base-pair between in-register complementary nucleotides. Assuming that the presence of mismatches in the sequence only affects the process of zippering but not the rate of formation of the initial nucleus, we found a best fitting nucleation rate equal to 2.0×10^6 M^{-1} s^{-1}, and an overall best fitting energy barrier to overcome internal mismatches in the hybridization process equal to 2.49 kcal/mol. In the context of this model, mismatches reduce the hybridization kinetics by causing a drop in the success rate when the zippering starts from a position adjacent to a defect and by dramatically increasing the time required to complete the hybridization. According to both approaches, bulges do not reduce the zippering rate, pointing to a fundamental role of coaxial stacking in stabilizing the growing duplex during the hybridization process.

To extend our findings to short RNA oligonucleotides, we studied the ribonucleic counterpart of a subset of the sequences here discussed. In this case, defects also cause the hybridization rate to vary within one

order of magnitude from the control, with the impact of the different defects comparable to what was measured in DNA. As a biologically relevant test case, we measure the effect of a single mismatch in the 8 bp long miR125 seed sequence in 200mM NaCl, and found that the presence of the defect does not significantly affect the hybridization rate (Supplementary Data 2 and Supplementary Figures 2 and 3). This result extends our findings to shorter sequences and lower ionic strength, in direct contrast to what previously seen in literature (26), and suggests the possibility of generalizing the conclusions of a minor role of defects in the hybridization kinetics as drawn from our study on 12 bp-long duplexes.

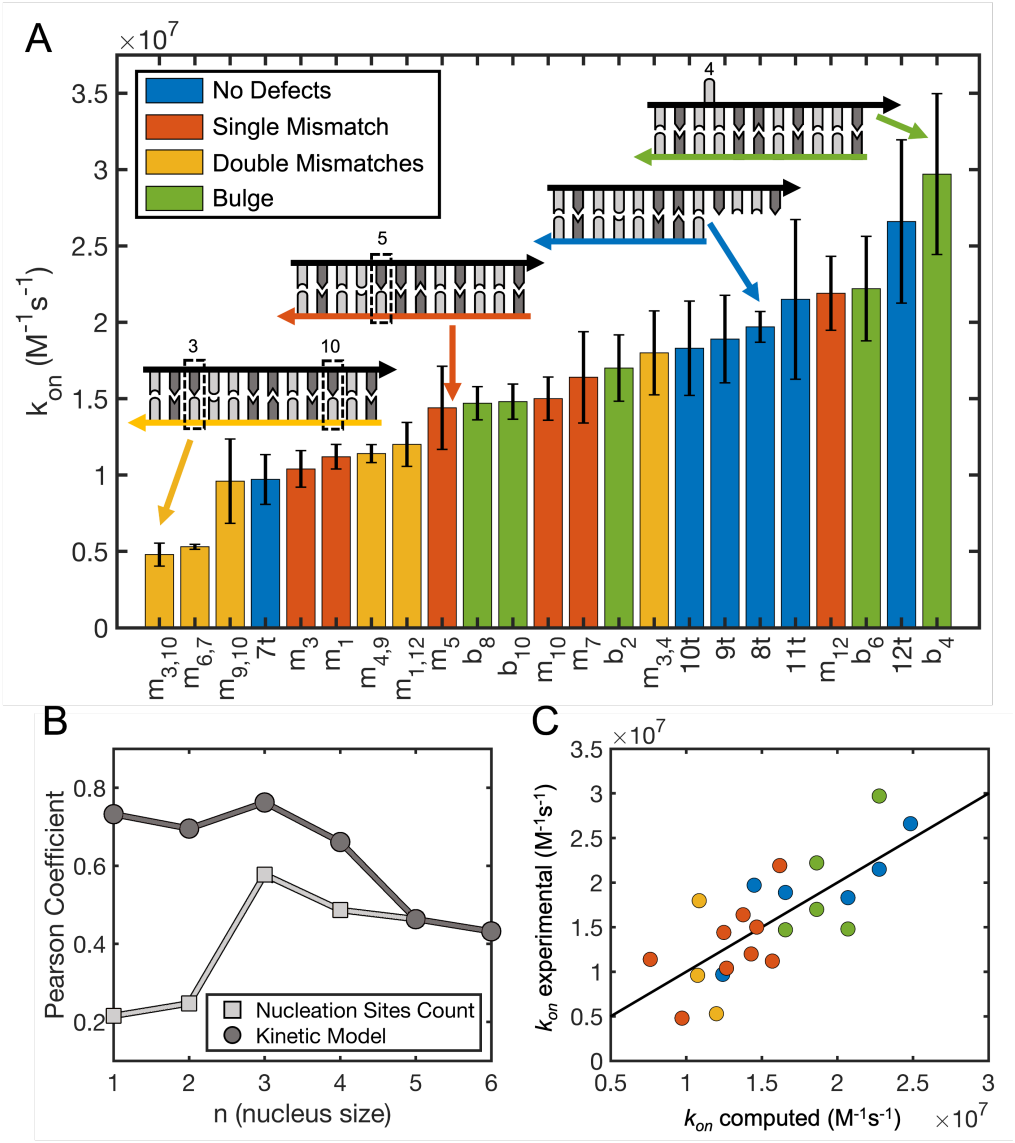

Figure 1. Hybridization kinetics for short DNA oligonucleotides with defects. (**A**) Hybridization kinetics constants for short DNA oligonucleotides with and without defects; either single mismatches, double mismatches, or bulges. Data have been sorted from lowest to highest hybridization rate to facilitate comparisons with the control (12t), which forms a perfectly pairing 12 bp-long duplex. The sketches show the general sequence designs, with the probe sequence (in black) fixed while changing the pairing oligonucleotides. (**B**) Correlation between experimentally determined hybridization rates and (orange) the total number of nucleation sites or (red) rates predicted by the kinetic model from Eq. 1 constructed with nucleation sites of different sizes. (**C**) Hybridization rates computed with the kinetic model described in Eq. 1 using n = 3 against experimentally determined rates.

Effect of defects on short oligonucleotides off-rates

The eventual separation of two hybridized strands occurs after a characteristic time, which is governed by the first-order rate constant k_{off}. To determine the k_{off} for our duplexes, we employed the relationship between the dissociation (K_D), hybridization and dehybridization rate constants, so that under the assumption of a two-state system it holds true that:

$$k_{off} = K_D \cdot k_{on} \qquad \text{(Eq. 3)}$$

We determined K_D at room temperature for our duplexes through a series of melting and binding experiments. Given K_D and k_{on}, we could calculate k_{off}, thus estimating the rate at which the duplexes dissociate and the strands separate.

The values of k_{on} and k_{off} for DNA determined in this work are shown in Figure 2 on a logarithmic scale to better emphasize the relative contribution of forward and reverse reactions to defects-mediated destabilization. Our data show that the main overall contribution of defects destabilization is the dehybridization constant of DNA. For all defects studied here, we find that the destabilization is position-dependent and consistent with NN predictions (see Supplementary Data 3 and Supplementary Figure 3). We confirmed the prior observation in literature (20, 41–43) that the logarithm of the dehybridization rates is well-correlated with the free energy of pairing of the two oligonucleotides both in DNA and RNA (R^2 = 0.97, Supplementary Figure 2D), even when defects are present. By treating dissociation as a thermally activated process (42–44), we can compute an average reduction in the activation energy caused by the defects in DNA as equal to -5.0 kcal/mol, -3.8 kcal/mol, and -2.6 kcal/mol for double mismatches, bulges, and single mismatches respectively, following a trend that is different from the one observed for k_{on}, where bulges do not significantly affect the process.

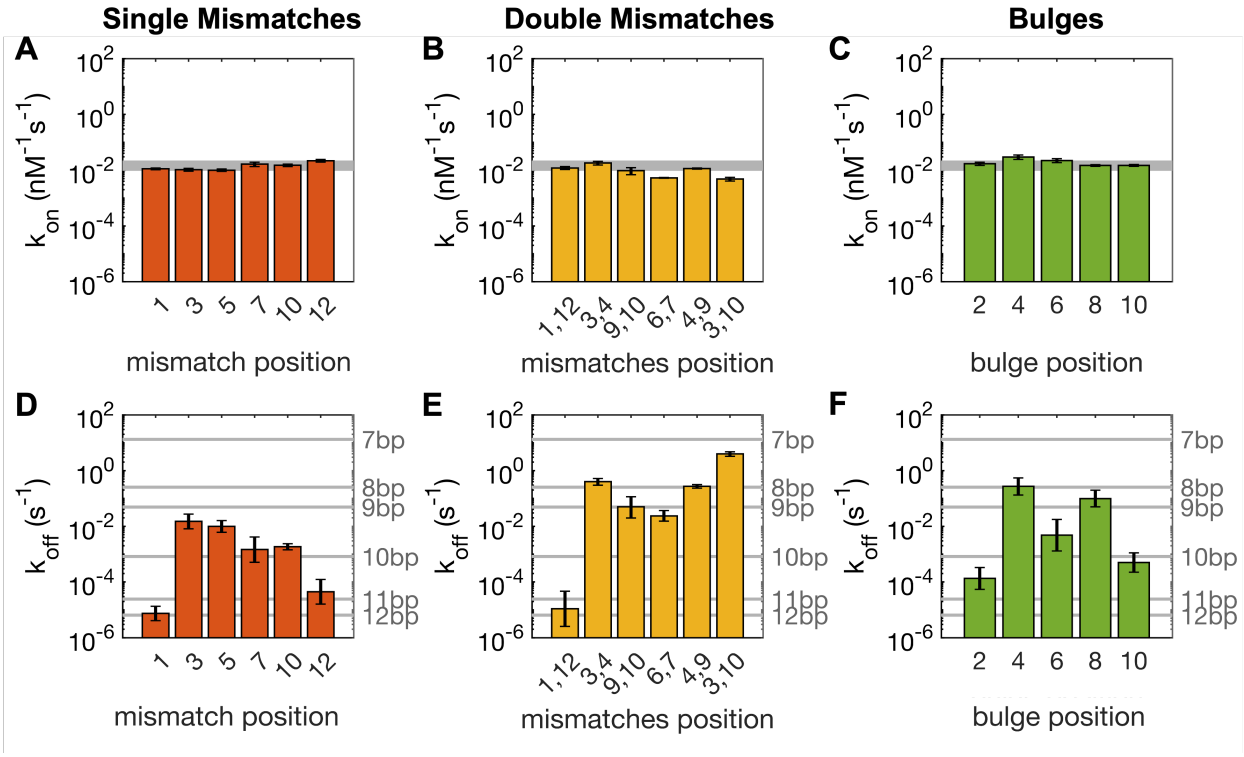

Figure 2. Kinetics for hybridization and dehybridization in short DNA oligonucleotides with defects. (**A, B, C**) Hybridization kinetics constants for oligonucleotides with defects, either single mismatches (**A**), double mismatches (**B**), or bulges (**C**). The grey shaded bands show k_{on} for control oligonucleotides of lengths varying from 7bp to 12bp, yielding values varying by roughly a factor of two. Data are shown on a logarithmic scale to directly compare the relative contributions of k_{on} and k_{off} to K_D. (**D, E, F**) Dehybridization kinetics constants for short DNA oligonucleotides with defects, either single mismatches (**D**), double mismatches (**E**), or bulges (**F**). Horizontal lines show k_{off} for a series of perfectly paired control duplexes.

Measurements of the binding accuracy of short oligonucleotides

We next extended our investigation to the direct competition between these defected oligonucleotides for binding to a probe in the presence of its perfectly complementary target.

When we add a target oligonucleotide (T) designed to bind its complementary oligonucleotide probe (P) in a complex mixture in the presence of another competing strand sharing a partial complementarity (C) to the probe, we expect a certain amount of both T:P and C:P duplexes to form (Figure 3A). In this scenario, every C:P duplex will be an off-target species. To assess the extent of target binding at any time, we monitored the emission of a 2Ap nucleotide within the target sequence as outlined in the Materials and Methods section. In order to quantify the interference caused by the competitor, we introduced a parameter termed binding accuracy (θ), defined as the ratio of correctly bound probe to the total bound probe:

$$\theta = \frac{[T:P]}{[T:P] + [C:P]} \qquad \text{(Eq. 4)}$$

Drawing upon our understanding of the behavior of complex mixtures of competing oligonucleotides (20), we expect rapid formation of both C:P and T:P duplexes in ~1 second. If the dehybridization rates of the duplexes are significantly smaller than their initial rates of formation, the relative amount of the two duplexes formed upon mixing of the three strands will depend uniquely on their relative hybridization rates (Figure 3B). In this early stage, an initial binding accuracy (θ_{in}) could be conveniently measured, typically assessed after approximately 2 seconds from hand mixing. Following the initial binding event and the partitioning of the probe oligonucleotide, the system progressively evolves and converges to an equilibrium state characterized by a stable binding accuracy, defined here as θ_{eq}.

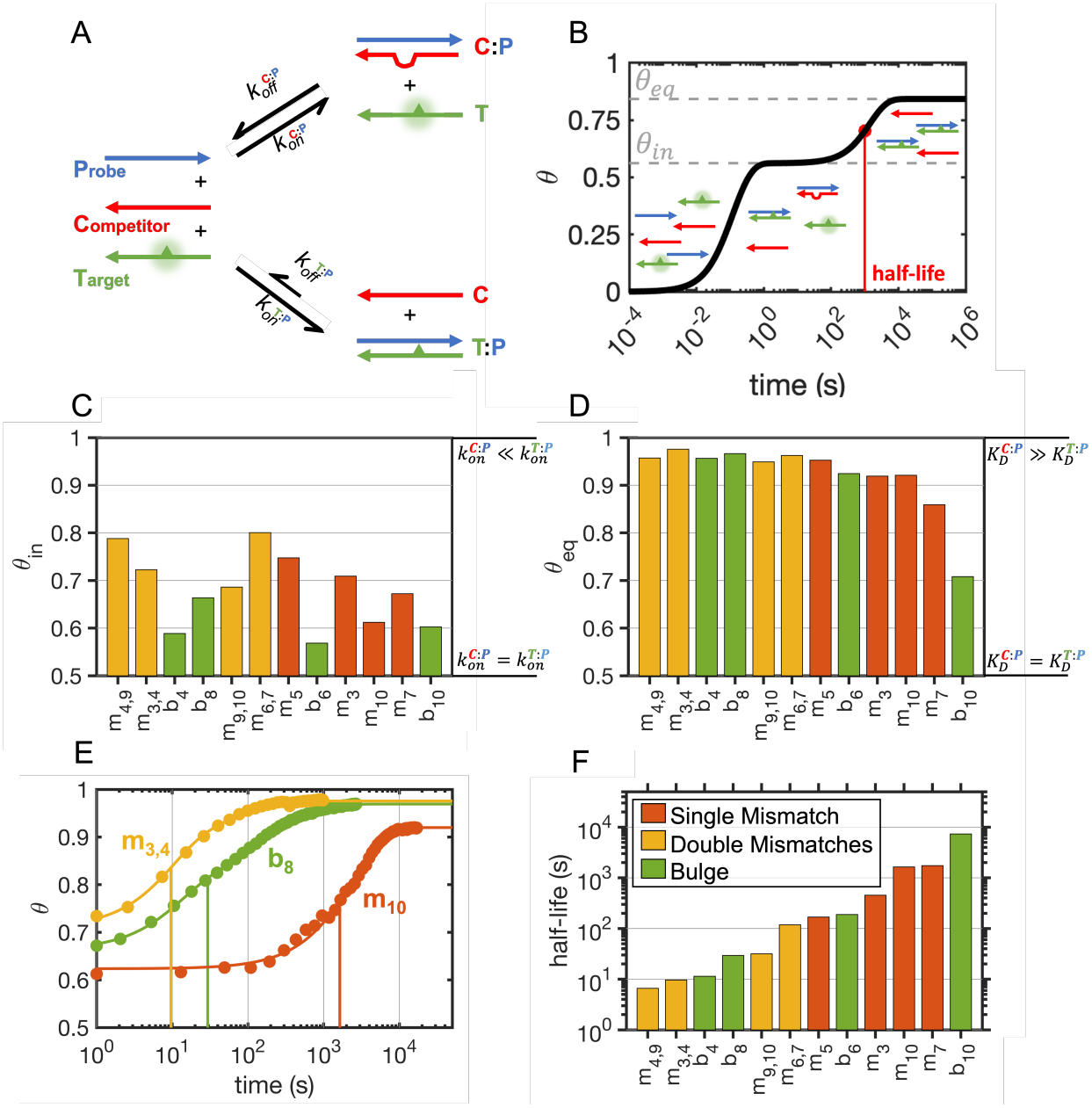

Figure 3. Competition experiment for DNA probe hybridization. (**A**) Reaction scheme for the competition experiment between equimolar amounts of probe sequence (P), perfectly pairing target (T) and defected competitor (C). Target strands bear a 2Ap nucleotide, sketched as a triangle, to keep track of their bound fraction over time. (**B**) Sketch of the time course of a typical competition experiment: the three strands are mixed and hybridize to produce a certain amount of perfect and defected duplex in ≈ 1 second, so that the system temporarily exhibits the initial binding accuracy θ_{in}. After a certain amount of time depending on k_{on}, k_{off} and strand displacement rates, the metastable state converges to the energy minimum that is associated with its equilibrium binding accuracy θ_{eq}. (**C**) Initial binding accuracy for competition experiments in case studies where $\theta_{in} < \theta_{eq}$. These values are expected to be equal to 1.0 for $k_{on}^{C:P} << k_{on}^{T:P}$ or 0.5 for $k_{on}^{C:P} = k_{on}^{T:P}$. (**D**) Binding accuracy at equilibrium for competition experiments in case studies where $\theta_{in} < \theta_{eq}$. These values are expected to be equal to 1.0 for $K_D^{C:P} >> K_D^{T:P}$ or 0.5 for $K_D^{C:P} = K_D^{T:P}$. (**E**) Equilibration traces measured in competition experiments. The continuous lines have been obtained by interpolating/extrapolating the data. Vertical lines highlight half-lives for the shown samples. (**F**) Measured equilibration timescale, defined as the time required so that 50% of the transient state is equilibrated (see

panel B). The transient low-accuracy state can last for a time ranging from a few seconds up to hours. All experiments were performed using 200 nM of each strand in 10 mM Tris-HCl pH 8.0 and 1 M NaCl.

In most cases studied here using a 2Ap-containing oligonucleotide 12t as the target sequence, we found that the initial binding accuracy (Figure 3C) differs from the equilibrium binding accuracy (figure 3D), with $\theta_{in} < \theta_{eq}$, meaning that as the system reaches equilibrium, we could detect an increase in hybridization accuracy (θ). The only exceptions to this were those cases where the presence of 2Ap caused a predictable increase in K_D comparable to the defects in the competitor strand, namely m_1, m_{12}, $m_{1,12}$ where we could measure a decrease in θ as a function of time, and b_2, where $\theta_{in} = \theta_{eq}$.
Having established the presence of transient low-accuracy states in a competitive scenario, our investigation shifted to understanding the duration of such states. Specifically, we aimed to determine whether they are short-lived or long-lived, in the sense that they persist for a timescale comparable to relevant processes requiring high specificity. To this end, we conducted a comprehensive study examining the evolution of the 2Ap signal in those specific conditions where the binding accuracy increases over time.
Equilibration within a multi-strand mixture is a complex process that may not adhere to a single exponential decay (20), making it challenging to define a simple equilibration rate (Figure 3E). Consequently, we chose to define the timescale for equilibration as the duration required for 50% of the transiently bound defected duplexes to disappear (Figure 3B,E). We found that the timescales associated with the equilibration process vary by several orders of magnitude, with sequences bearing double mismatches taking only a few seconds to equilibrate in the fastest cases, while sequences with single mismatches or bulges taking up to hours in the slowest cases (Figure 3F).

Predictions of short oligonucleotides binding accuracy
Having established that defected duplexes can be transiently present for a relative long time as the system equilibrates, we wanted to know whether their abundance could be predicted. Taking into account the modest destabilizing factor coming from 2Ap (+0.9 kcal/mol), we could calculate the exact analytical solution for the expected θ_{eq} (45) by knowing the total concentration of the three strands and the binding constants (K_D) for each duplex. Moreover, we could conveniently approximate predictions for the expected θ_{in} as follows:

$$\theta_{in} \approx \frac{k_{on}^{T:P}}{k_{on}^{T:P} + k_{on}^{C:P}} \qquad \text{(Eq. 5)}$$

It is important to highlight here that Equation 5 reduces to $\theta_{in} = 1.0$ in the case where $k_{on}^{C:P} << k_{on}^{T:P}$, so that similar hybridization rates are a critical requirement to have $\theta_{in} \neq \theta_{eq}$ in those cases when the target is designed to outcompete the competitor at equilibrium.
We found that both the measured initial ($R^2 = 0.60$, RMSE = 8.8%) and equilibrium ($R^2 = 0.89$, RMSE = 6.5%) binding accuracies are well predicted by our estimates as depicted in Figure 4A,B.
Importantly, θ_{in} and θ_{eq} are poorly correlated ($R^2 = 0.19$, RMSE = 32%) as depicted in Figure 4C, emphasizing how thermodynamic features and knowledge of the equilibrium state of the system lose their value when we enter the out-of-equilibrium regime.

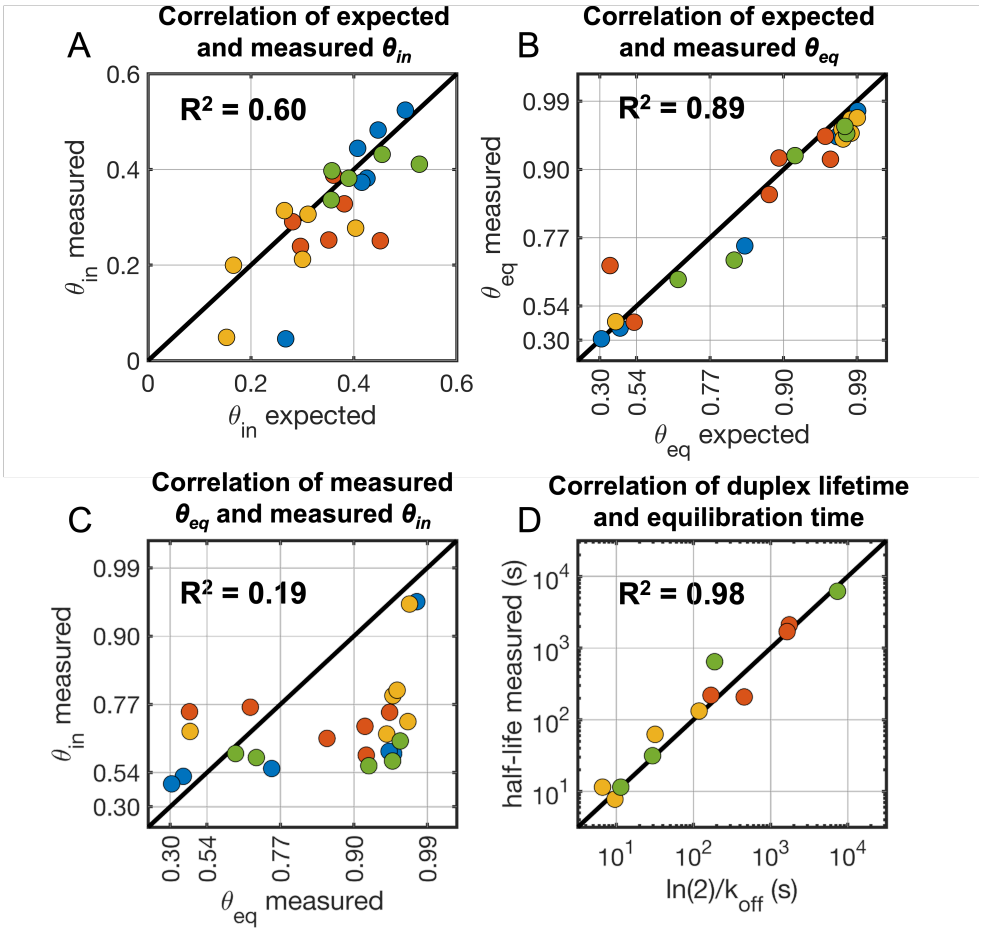

Figure 4. Predictions for competition experiments with DNA. (**A**) Initial binding accuracy (θ_{in}) for the competitive mixtures compared to values computed from Eq. 5 are in good agreement, showing that hybridization rates can be used to predict the initial partition of duplexes. (**B**) Binding accuracy at equilibrium is well predicted by a simple binding model based on experimentally determined nucleic acids dissociation constants. Axes have been stretched to better visualize the cluster of points close to 1.0 (**C**) Partition at equilibrium and after initial mixing are not well-correlated, showing the importance of expanding our current models for complex mixtures of oligonucleotides. Red dashed line shows best fitting line. Axes have been stretched to better visualize the cluster of points close to 1.0. (**D**) Comparison of the measured equilibration timescale and the lifetime of the defected duplex as computed from the dehybridization rate of the competitor oligonucleotide. All experiments were performed with 200nM of each strand in 10 mM Tris-HCl pH 8.0 and 1 M NaCl.

Finally, we observed a strong relationship (R^2 = 0.98) between the dehybridization rates of the defected duplexes and the duration of the equilibration process, a result pointing to a negligible role of strand displacement in the evolution of the system at these concentrations. This outcome is a natural consequence of the intentional absence of toeholds incorporated in the C:P duplexes, resulting in very slow strand displacement (46) so that the overall process is controlled by the dissociation and reshuffling of the strands, which is a relatively faster process in the defected DNA duplexes studied here. Interestingly, we found that this is not the case for competition among RNA strands (Supplementary Data 4 and Supplementary Figure 5) studied in this work: their much slower dehybridization rate compared to DNA (20, 42) would result in years-long transient states that do not match experimental results. This discrepancy can be understood in the context of an equilibration limited by strand displacement: given the lower boundary for the strand displacement rate as equal to ≈ 1 $M^{-1}s^{-1}$ in the absence of toeholds for non-mismatched sequences (20, 46), we can reasonably predict an upper limit of a few hours for the equilibration in a competing scenario with 1 μM of oligonucleotides, so that dehybridization and reshuffling

in RNA is effectively outcompeted by strand displacement. Still, the stronger pairing between RNA strands makes the equilibration up to 30 times slower than the DNA counterpart in the cases explored here, making the out-of-equilibrium C:P duplexes abundant and longer-lasting species that can be potential substrates for chemical and enzymatic reactions.

Off-target reactions in competing RNA mixtures

To test whether the defected duplexes present when the system is out-of-equilibrium can behave as substrates for subsequent reactions, we studied a mixture with the RNA $m_{4,9}$ sequence in the presence of the chemically reactive imidazolium-bridged C*C dinucleotides (47) and the *CGCA tetranucleotides bearing a 2-aminoimidazole moiety on the 5′ terminus (48).

The imidazolium-bridged C*C dinucleotide is known to chemically extend a primer on a template one nucleotide at a time over a typical timescale of minutes at saturating concentrations (29), and over a timescale of hours at lower concentrations. The activated tetranucleotide is known to ligate to the 3' end of a primer over a timescale of days at saturating concentrations (48). These prebiotically plausible reactions are considered essential intermediate steps for the origin of life from the RNA World, and can also serve as chemical models to test the out-of-equilibrium RNA hybridization.

To monitor the extent of chemical primer extension and ligation, the competing oligonucleotide C (bearing the double mismatch) was labelled with a 5′-Cy5 fluorescent dye, while the target oligonucleotide T was labelled with a 5′-FAM fluorescent dye. Denaturing polyacrylamide gels were imaged in the green and red channels to quantify the reacted target strand and the reacted competitor strand over time (Figure 5B,C and Supplementary Figure 6).

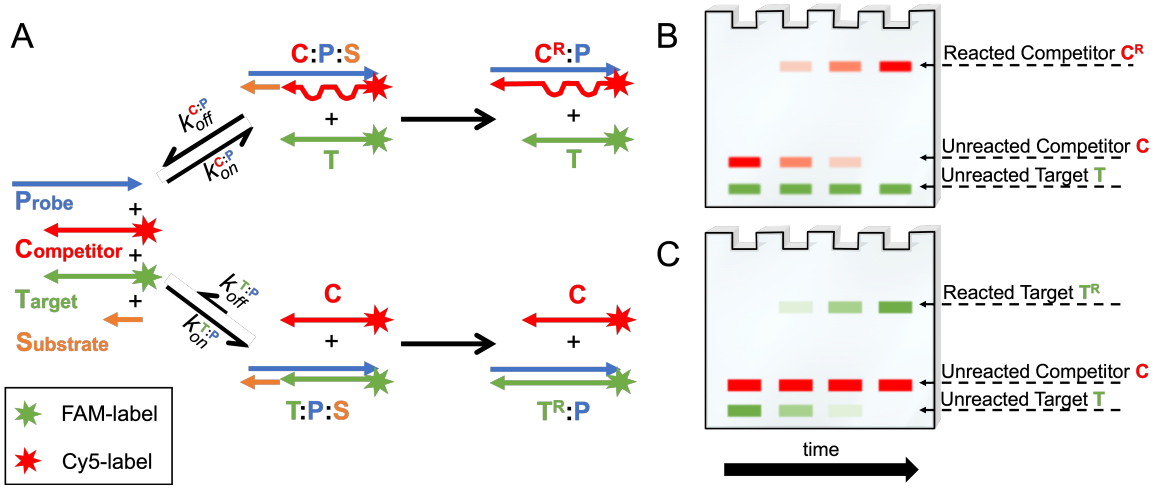

Figure 5. Experimental design for chemical reactions in a competition experiment. (**A**) Reaction scheme for the competition experiment between equimolar amounts of probe sequence (P), perfectly paired target (T) and defected competitor (C) in the presence of chemically reactive species (S). Both C:P and T:P duplexes can react with the reactants to form novel species C^R and T^R. For the two reactions to take place, the probe oligonucleotide presents a 5′ overhang to allow either imidazolium-bridged C*C dinucleotides binding and templated primer extension (overhang: 5′-GGGG-3′) or tetranucleotide binding and templated ligation (overhang: 5′-UGCG-3′) as depicted in the sketch. (**B**, **C**) Sketches of denaturing PAGE to resolve reaction products in a competition experiment. Both competitor, target sequences and relative reaction products can be tracked at the same time with the distinct fluorescent emissions of their conjugated fluorophores. The relative abundance of the reacted species can be quantified. In the examples presented here, we show (**B**) the hypothetical outcome of a reaction where θ is always equal to zero, so that only the defected C:P duplex is formed at any time, and (**C**) the hypothetical outcome of a reaction where θ is always equal to one, so that only the perfect T:P duplex is formed at any time. An example of a real PAGE overlay is shown in Supplementary Figure 6.

The equilibrium binding accuracy for the mixture with $m_{4,9}$ is predicted to be equal to 1.00, a value in agreement with our experimentally determined value of 0.97±0.02. Under a conventional description that only account for the equilibrium state, virtually no C:P complex would be available as a substrate for reactions. In contrast to the prediction made by only considering the equilibrium state, our results show

that even in the presence of a double mismatch, the competing oligonucleotide can in fact be a substrate for both non-enzymatic primer extension (Figure 6A,B) and ligation (Figure 6C) during the equilibration phase, with a strikingly large yield as high as ≈ 38%. This behavior is in agreement with an equilibration-limited reaction, so that once the metastable defected duplexes are removed from the pool during the system equilibration, no more C:P duplexes should be available as substrates for non-enzymatic reactions.

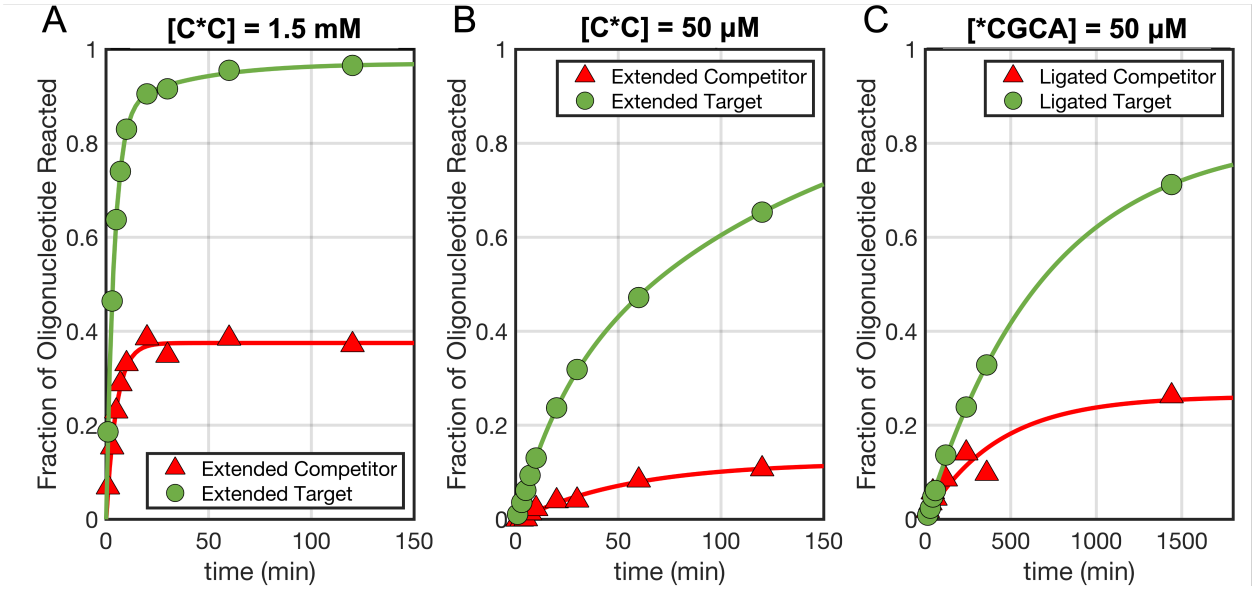

Figure 6. Chemical reactions in a competition scenario among RNA strands. (**A, B**) Primer extension reactions at different C*C substrate concentrations and (**C**) ligation reactions occurring on perfect (Target) and defected (Probe) duplexes. All continuous lines were obtained by interpolating/extrapolating the data. All reactions have been performed at room temperature in 200mM Tris-HCl pH 8.0 and 100mM $MgCl_2$.

Name	Definition
k_{on}	Hybridization rate of two oligonucleotides
k_{off}	Dehybridization rate of two oligonucleotides
K_D	Dissociation constant of a duplex
R^2	Coefficient of determination for predicted vs measured values
RMSE	Root mean squared error for predicted ($\hat{y}$) vs measured values (y). The value is given as a percentage, and it is computed over a dataset of size *m* as $\mathrm{RMSE}=\frac{100}{m}\sum_{i=1}^{m}\sqrt[2]{(\hat{y}_i-y_i)^2}/y_i$

Table 1. Table of names.

DISCUSSION

In this work we have systematically studied the effect of single mismatches, double mismatches, and bulges on the hybridization and dehybridization kinetics of short DNA and RNA duplexes. We have found that the defects studied here either have a negligible or negative effect on hybridization rate, with double mismatches affecting it the most, with a maximum ≈ 6x slow-down, and bulges affecting it the least. We found that k_{on} for perfectly paired sequences is well correlated with the ΔG of hybridization (R^2 = 0.76), analogously to what previously observed for short RNA oligonucleotides (20). This relationship does not hold for sequences bearing defects (R^2 = 0.08), showing how these increase the complexity of the hybridization phenomenon in a non-trivial way. To capture the role of defects, we employed a simple nucleation-limited model and a kinetic Monte Carlo model to fit the experimentally determined hybridization rates, finding that the hybridization slow-down can be explained in terms of loss of nucleation sites and added energy penalty for zippering the two strands. We found that our results can be qualitatively extended from DNA to RNA, and even to lower ionic strength and shorter sequences.

Our simple study on the combinatorics of nucleation sites suggests that 3 contiguous base-pairs are necessary and sufficient to make an effective nucleation site (Figure 1B) even in the context of defected sequences. No clear discontinuity in the measured hybridization rates was observed when reducing the length of the longest continuous stretch of base pairs, suggesting that “the rule of seven” as formulated by Cisse and colleagues (26) may be specific to vesicle-encapsulated oligonucleotides, and emphasizing the need for an orthogonal and more general study on defected nucleic acids kinetics in solution.
Our rate analysis shows that bulges have a minor effect – if any – on the zippering rate in DNA, suggesting that the continuous coaxial stacking prevents the energy barrier present with internal mismatches. On the contrary, their effect in dehybridization is negative and well-captured by the NN model of nucleic acids thermodynamics. While this observation points to an asymmetric role of defects in pairing and unpairing, it promotes the thermodynamic model for nucleic acids to a predictive tool for dehybridization kinetics for defected duplexes, simplifying future studies on oligonucleotides kinetics.
As a consequence of the similar hybridization rates for perfectly complementary and defected duplexes, we discovered the existence of a transient state in competing mixtures, where a considerable amount of defected duplexes could be measured. Our results on competing oligonucleotides mixtures corroborate the hypothesis that even in systems where defected duplexes are not expected at equilibrium, off-target binding is relevant when the system is out-of-equilibrium, over time scales that can be longer than prebiotically relevant chemical reactions involving RNA (29, 49) or biological phenomena (50). In these mixtures, we have shown how it is possible to efficiently process the unexpected metastable duplexes with non-enzymatic reactions, opening the question on the impact of such phenomenon in biological or prebiotic contexts (6) that are far from equilibrium.
To quantify the likelihood of observing out-of-equilibrium mispairings in prebiotic or biological mixtures of oligonucleotides, it is possible to derive a simple analytical relationship (Supplementary Eq. 1 and 2) between the length of a given oligonucleotides and its probability to bind with defects in a mixture of random sequences (Supplementary Data 5 and Supplementary Figure 7).
If we consider the primordial environment to host the first prebiotic reactions as being out-of-equilibrium (22) and subject to a continuous influx of energy required for replication and maintenance of information, the oligonucleotides wherein contained would have undergone a continuous change of the base-paired configurations while being cyclically re-equilibrated (22). Template copying is necessary for the purpose of maintaining genetic information and it cannot prescind from primer-template binding followed by primer extension. If a primer is mispaired in the wrong position over a template and polymerized, this could lead to the production of a novel sequence that was not originally encoded (Figure 7).

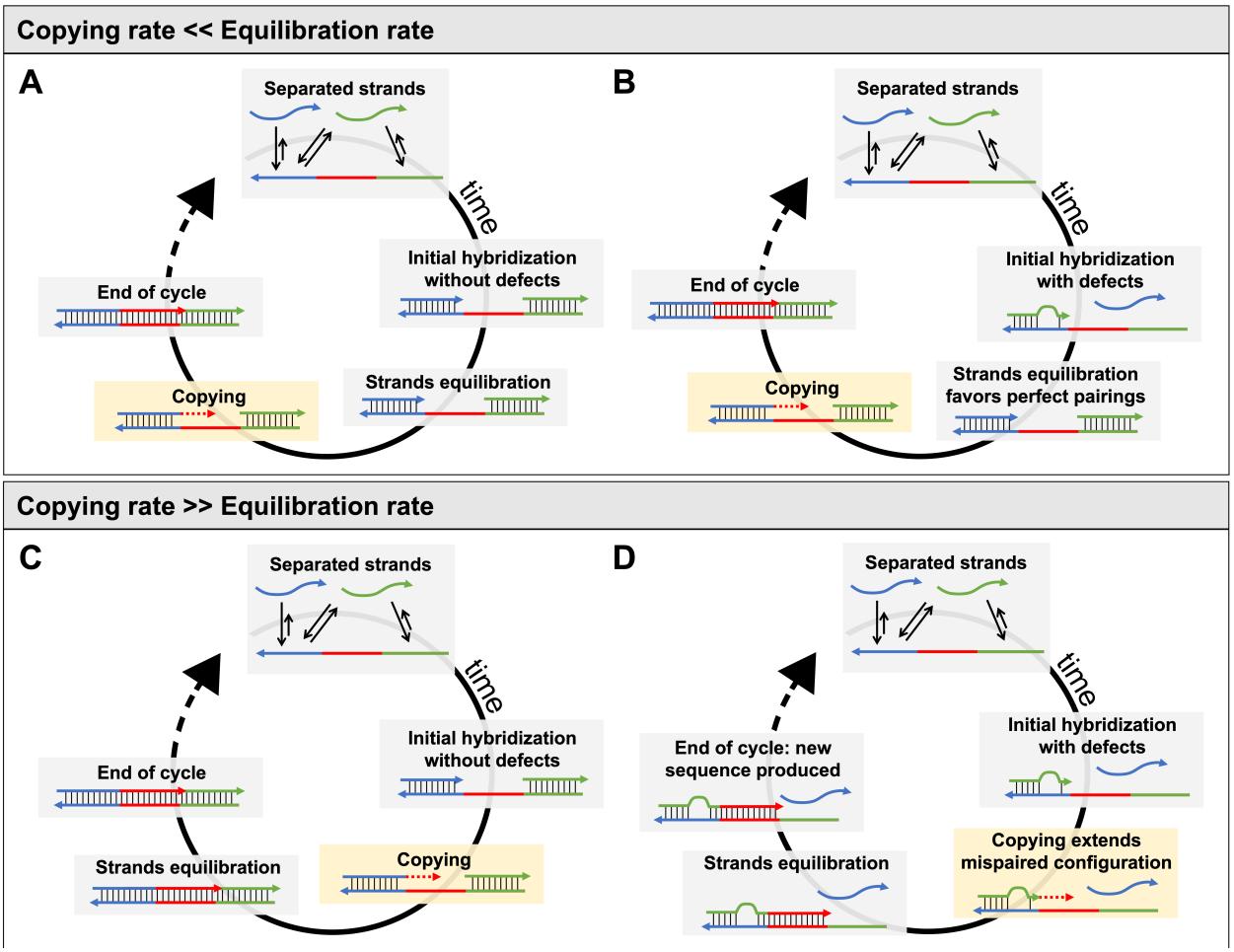

Figure 7. Copying of a template in a complex mixture. In a mixture of multiple RNA strands, these can separate because of chemical-physical changes such as dilution, thermal shock, reduction in ionic strength or change in pH (22, 51). As conditions favorable for pairing are re-established, the short oligonucleotides quickly hybridize. Depending on the chemical-physical conditions of the system, it may be possible that (**A, B**) the equilibration rate is much faster than the copying rate, so that regardless of the initial configuration of the duplexes, only the correctly paired primer will be elongated, and the information encoded in the template can be properly copied. (**C, D**) If the copying rate is faster than the equilibration rate, depending on the initial configuration it may be possible to either (**C**) properly copy the template or (**D**) to extend the mispaired oligonucleotide, creating a novel sequence that was not previously encoded in the template. Copying step occurring at different times is highlighted in yellow.

Even a relatively small 11 nt long primer has a significant chance of finding an alternative binding site with two defects over a genome of 100nt in length (Supplementary Figure 7), a typical size for prebiotically relevant ribozymes (52). We propose here that this physical mechanism may cause loss of information even with a chemistry allowing for perfect copying fidelity, and thus increase the apparent replication error rate experimentally determined (53). As we have shown in this study (Figure 6A,B), this loss of fidelity due to incorrect duplexing would be exacerbated by increasing the primer extension rate, suggesting the existence of an upper limit for a reliable copying rate, that is intimately linked to the equilibration rate of the system (Figure 7).
In a biological context, given the generally accepted view that most of the human genome is transcribed (54), and given a total sequence length of 10^9 nt, we can predict a significant chance ($p > 0.05$) of observing a random 20 nt oligonucleotide (a typical size for miRNA, ASO and siRNA) hybridizing with one or more defects. In a biological system, we can therefore predict that the transient low accuracy here discussed can exacerbate the toxicity of off-target hybridization and dysregulation of gene expression (12, 55, 56) by allowing the formation of undesired duplexes that would never be expected at equilibrium and

that can potentially be substrates for enzymatic activity. Our simple considerations point to the fact that over large genomes there should be plenty of defected duplexes formed by chance in long-lasting out-of-equilibrium states. This prediction raises questions on how biological systems can prevent such condition, possibly protecting long RNA transcripts through folding or thanks to binding proteins (57) that would prevent catastrophic mispairings.
Through this comprehensive analysis, we gained valuable insights into the kinetic and out-of-equilibrium properties of the oligonucleotide mixtures, shedding light on the intricate dynamics of their binding interactions and providing tools for future studies. Our findings show the importance of appropriately account for cross-interactions and pairing with mismatches and bulges when handling multiple strands in a mixture when out of equilibrium, and we foresee that these will help improve algorithms to simulate mixtures (21), design complex nucleic acids structures and predict the efficacy of gene and antisense therapies (58).

ACKNOWLEDGEMENTS
We would like to thank Aleksandar Radakovic, Brennan Ashwood, Daniel Duzdevich and Filip Boskovic for comments on the manuscript.

FUNDING
National Science Foundation [CHE-2104708 to J.W.S.]; Simons Foundation [290363 to J.W.S.]; Sloan Foundation; Moore Foundation. J.W.S. is an investigator of the Howard Hughes Medical Institute.

CONFLICT OF INTEREST
None declared.

REFERENCES
1. Bartel,D.P. (2009) MicroRNAs: Target Recognition and Regulatory Functions. *Cell*, **136**, 215–233.

2. Cha,R.S. and Thilly,W.G. (1993) Specificity, efficiency, and fidelity of PCR. *PCR Methods Appl*, **3**, S18-29.

3. Gall,J.G. and Pardue,M.L. (1969) Formation and detection of RNA-DNA hybrid molecules in cytological preparations. *Proc Natl Acad Sci U S A*, **63**, 378–83.

4. Pardue,M.L. and Gall,J.G. (1969) Molecular hybridization of radioactive DNA to the DNA of cytological preparations. *Proc Natl Acad Sci U S A*, **64**, 600–4.

5. Fredriksson,S., Gullberg,M., Jarvius,J., Olsson,C., Pietras,K., Gústafsdóttir,S.M., Östman,A. and Landegren,U. (2002) Protein detection using proximity-dependent DNA ligation assays. *Nature Biotechnology*, **20**, 473–477.

6. Zhou,L., Ding,D. and Szostak,J.W. (2021) The virtual circular genome model for primordial RNA replication. *RNA*, **27**, 1–11.

7. Saminathan,A., Devany,J., Veetil,A.T., Suresh,B., Pillai,K.S., Schwake,M. and Krishnan,Y. (2021) A DNA-based voltmeter for organelles. *Nat. Nanotechnol.*, **16**, 96–103.

8. DeLuca,M., Shi,Z., Castro,C.E. and Arya,G. (2020) Dynamic DNA nanotechnology: toward functional nanoscale devices. *Nanoscale Horiz.*, **5**, 182–201.

9. Hong,F., Zhang,F., Liu,Y. and Yan,H. (2017) DNA Origami: Scaffolds for Creating Higher Order Structures. *Chem. Rev.*, **117**, 12584–12640.

10. Biffi,S., Cerbino,R., Bomboi,F., Paraboschi,E.M., Asselta,R., Sciortino,F. and Bellini,T. (2013) Phase behavior and critical activated dynamics of limited-valence DNA nanostars. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 15633–15637.

11. Jackson,A.L., Bartz,S.R., Schelter,J., Kobayashi,S.V., Burchard,J., Mao,M., Li,B., Cavet,G. and Linsley,P.S. (2003) Expression profiling reveals off-target gene regulation by RNAi. *Nat Biotechnol*, **21**, 635–7.

12. Fedorov,Y., Anderson,E.M., Birmingham,A., Reynolds,A., Karpilow,J., Robinson,K., Leake,D., Marshall,W.S. and Khvorova,A. (2006) Off-target effects by siRNA can induce toxic phenotype. *Rna*, **12**, 1188–96.

13. SantaLucia,J. and Hicks,D. (2004) The Thermodynamics of DNA Structural Motifs. *Annu. Rev. Biophys. Biomol. Struct.*, **33**, 415–440.

14. Turner,D.H. and Mathews,D.H. (2010) NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. *Nucleic Acids Research*, **38**, D280–D282.

15. Zadeh,J.N., Steenberg,C.D., Bois,J.S., Wolfe,B.R., Pierce,M.B., Khan,A.R., Dirks,R.M. and Pierce,N.A. (2011) NUPACK: Analysis and design of nucleic acid systems. *J. Comput. Chem.*, **32**, 170–173.

16. Kierzek,R., Burkard,M.E. and Turner,D.H. (1999) Thermodynamics of Single Mismatches in RNA Duplexes. *Biochemistry*, **38**, 14214–14223.

17. Peyret,N., Seneviratne,P.A., Allawi,H.T. and SantaLucia,J. (1999) Nearest-Neighbor Thermodynamics and NMR of DNA Sequences with Internal A·A, C·C, G·G, and T·T Mismatches. *Biochemistry*, **38**, 3468–3477.

18. Naiser,T., Kayser,J., Mai,T., Michel,W. and Ott,A. (2008) Position dependent mismatch discrimination on DNA microarrays – experiments and model. *BMC Bioinformatics*, **9**, 509.

19. Rosenberger,J.H., Göppel,T., Kudella,P.W., Braun,D., Gerland,U. and Altaner,B. (2021) Self-Assembly of Informational Polymers by Templated Ligation. *Phys. Rev. X*, **11**, 031055.

20. Todisco,M. and Szostak,J.W. (2022) Hybridization kinetics of out-of-equilibrium mixtures of short RNA oligonucleotides. *Nucleic Acids Research*, **50**, 9647–9662.

21. Chamanian,P. and Higgs,P.G. (2022) Computer simulations of Template-Directed RNA Synthesis driven by temperature cycling in diverse sequence mixtures. *PLoS Comput Biol*, **18**, e1010458.

22. Ianeselli,A., Salditt,A., Mast,C., Ercolano,B., Kufner,C.L., Scheu,B. and Braun,D. (2023) Physical non-equilibria for prebiotic nucleic acid chemistry. *Nat Rev Phys*, 10.1038/s42254-022-00550-3.

23. Gotoh,M. (1995) A New Approach to Determine the Effect of Mismatches on Kinetic Parameters in DNA Hybridization Using an Optical Biosensor. *DNA Research*, **2**, 285–293.

24. Tawa,K. (2004) Mismatching base-pair dependence of the kinetics of DNA-DNA hybridization studied by surface plasmon fluorescence spectroscopy. *Nucleic Acids Research*, **32**, 2372–2377.

25. Nava,G., Ceccarello,E., Giavazzi,F., Salina,M., Damin,F., Chiari,M., Buscaglia,M., Bellini,T. and Zanchetta,G. (2016) Label-free detection of DNA single-base mismatches using a simple reflectance-based optical technique. *Physical Chemistry Chemical Physics*, **18**, 13395–13402.

26. Cisse,I.I., Kim,H. and Ha,T. (2012) A rule of seven in Watson-Crick base-pairing of mismatched sequences. *Nat Struct Mol Biol*, **19**, 623–627.

27. Rachofsky,E.L., Osman,R. and Ross,J.B.A. (2001) Probing Structure and Dynamics of DNA with 2-Aminopurine: Effects of Local Environment on Fluorescence. *Biochemistry*, **40**, 946–956.

28. Owczarzy,R., Tataurov,A.V., Wu,Y., Manthey,J.A., McQuisten,K.A., Almabrazi,H.G., Pedersen,K.F., Lin,Y., Garretson,J., McEntaggart,N.O., *et al.* (2008) IDT SciTools: a suite for analysis and design of nucleic acid oligomers. *Nucleic Acids Research*, **36**, W163–W169.

29. Ding,D., Zhou,L., Giurgiu,C. and Szostak,J.W. (2022) Kinetic explanations for the sequence biases observed in the nonenzymatic copying of RNA templates. *Nucleic Acids Res*, **50**, 35–45.

30. Craig,M.E., Crothers,D.M. and Doty,P. (1971) Relaxation kinetics of dimer formation by self complementary oligonucleotides. *Journal of Molecular Biology*, **62**, 383–401.

31. Jones,M.S., Ashwood,B., Tokmakoff,A. and Ferguson,A.L. (2021) Determining Sequence-Dependent DNA Oligonucleotide Hybridization and Dehybridization Mechanisms Using Coarse-Grained Molecular Simulation, Markov State Models, and Infrared Spectroscopy. *J. Am. Chem. Soc.*, **143**, 17395–17411.

32. Hertel,S., Spinney,R.E., Xu,S.Y., Ouldridge,T.E., Morris,R.G. and Lee,L.K. (2022) The stability and number of nucleating interactions determine DNA hybridization rates in the absence of secondary structure. *Nucleic Acids Research*, **50**, 7829–7841.

33. Schaeffer,J.M. (2013) Stochastic Simulation of the Kinetics of Multiple Interacting Nucleic Acid Strands. 10.7907/JEBY-6X69.

34. Flamm,C., Fontana,W., Hofacker,I.L. and Schuster,P. (2000) RNA folding at elementary step resolution. *RNA*, **6**, 325–338.

35. Menssen,R.J., Kimmel,G.J. and Tokmakoff,A. (2021) Investigation into the mechanism and dynamics of DNA association and dissociation utilizing kinetic Monte Carlo simulations. *The Journal of Chemical Physics*, **154**, 045101.

36. Metropolis,N., Rosenbluth,A.W., Rosenbluth,M.N., Teller,A.H. and Teller,E. (1953) Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, **21**, 1087–1092.

37. Wetmur,J.G. and Davidson,N. (1968) Kinetics of renaturation of DNA. *Journal of Molecular Biology*, **31**, 349–370.

38. Law,S.M., Eritja,R., Goodman,M.F. and Breslauer,K.J. (1996) Spectroscopic and Calorimetric Characterizations of DNA Duplexes Containing 2-Aminopurine. *Biochemistry*, **35**, 12329–12337.

39. Woodside,M.T., Anthony,P.C., Behnke-Parks,W.M., Larizadeh,K., Herschlag,D. and Block,S.M. (2006) Direct Measurement of the Full, Sequence-Dependent Folding Landscape of a Nucleic Acid. *Science*, **314**, 1001–1004.

40. Manuel,A.P., Lambert,J. and Woodside,M.T. (2015) Reconstructing folding energy landscapes from splitting probability analysis of single-molecule trajectories. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 7183–7188.

41. Ashwood,B., Jones,M.S., Lee,Y., Sachleben,J.R., Ferguson,A.L. and Tokmakoff,A. (2023) Molecular insight into how the position of an abasic site and its sequence environment influence DNA duplex stability and dynamics Biophysics.

42. Rauzan,B., McMichael,E., Cave,R., Sevcik,L.R., Ostrosky,K., Whitman,E., Stegemann,R., Sinclair,A.L., Serra,M.J. and Deckert,A.A. (2013) Kinetics and Thermodynamics of DNA, RNA, and Hybrid Duplex Formation. *Biochemistry*, **52**, 765–772.

43. Menssen,R.J. and Tokmakoff,A. (2019) Length-Dependent Melting Kinetics of Short DNA Oligonucleotides Using Temperature-Jump IR Spectroscopy. *J. Phys. Chem. B*, **123**, 756–767.

44. Rejali,N.A., Ye,F.D., Zuiter,A.M., Keller,C.C. and Wittwer,C.T. (2021) Nearest-neighbour transition-state analysis for nucleic acid kinetics. *Nucleic Acids Research*, **49**, 4574–4585.

45. Wang,Z.-X. (1995) An exact mathematical expression for describing competitive binding of two different ligands to a protein molecule. *FEBS Letters*, **360**, 111–114.

46. Zhang,D.Y. and Winfree,E. (2009) Control of DNA Strand Displacement Kinetics Using Toehold Exchange. *J. Am. Chem. Soc.*, **131**, 17303–17314.

47. Walton,T., Zhang,W., Li,L., Tam,C.P. and Szostak,J.W. (2019) The Mechanism of Nonenzymatic Template Copying with Imidazole-Activated Nucleotides. *Angew. Chem. Int. Ed.*, **58**, 10812–10819.

48. Zhou,L., O'Flaherty,D.K. and Szostak,J.W. (2020) Template-Directed Copying of RNA by Non-enzymatic Ligation. *Angew. Chem. Int. Ed.*, **59**, 15682–15687.

49. Zhou,L., Kim,S.C., Ho,K.H., O'Flaherty,D.K., Giurgiu,C., Wright,T.H. and Szostak,J.W. (2019) Non-enzymatic primer extension with strand displacement. *Elife*, **8**.

50. Marzi,M.J., Ghini,F., Cerruti,B., de Pretis,S., Bonetti,P., Giacomelli,C., Gorski,M.M., Kress,T., Pelizzola,M., Muller,H., *et al.* (2016) Degradation dynamics of microRNAs revealed by a novel pulse-chase approach. *Genome Res*, **26**, 554–65.

51. Mariani,A., Bonfio,C., Johnson,C.M. and Sutherland,J.D. (2018) pH-Driven RNA Strand Separation under Prebiotically Plausible Conditions. *Biochemistry*, **57**, 6382–6386.

52. Tjhung,K.F., Shokhirev,M.N., Horning,D.P. and Joyce,G.F. (2020) An RNA polymerase ribozyme that synthesizes its own ancestor. *Proc. Natl. Acad. Sci. U.S.A.*, **117**, 2906–2913.

53. Duzdevich,D., Carr,C.E., Ding,D., Zhang,S.J., Walton,T.S. and Szostak,J.W. (2021) Competition between bridged dinucleotides and activated mononucleotides determines the error frequency of nonenzymatic RNA primer extension. *Nucleic Acids Research*, **49**, 3681–3691.

54. Carninci,P., Yasuda,J. and Hayashizaki,Y. (2008) Multifaceted mammalian transcriptome. *Current Opinion in Cell Biology*, **20**, 274–280.

55. Yoshida,T., Naito,Y., Yasuhara,H., Sasaki,K., Kawaji,H., Kawai,J., Naito,M., Okuda,H., Obika,S. and Inoue,T. (2019) Evaluation of off-target effects of gapmer antisense oligonucleotides using human cells. *Genes Cells*, **24**, 827–835.

56. Lindow,M., Vornlocher,H.-P., Riley,D., Kornbrust,D.J., Burchard,J., Whiteley,L.O., Kamens,J., Thompson,J.D., Nochur,S., Younis,H., *et al.* (2012) Assessing unintended hybridization-induced biological effects of oligonucleotides. *Nat Biotechnol*, **30**, 920–923.

57. Hentze,M.W., Castello,A., Schwarzl,T. and Preiss,T. (2018) A brave new world of RNA-binding proteins. *Nat Rev Mol Cell Biol*, **19**, 327–341.

58. Smith,L., Andersen,K.B., Hovgaard,L. and Jaroszewski,J.W. (2000) Rational selection of antisense oligonucleotide sequences. *European Journal of Pharmaceutical Sciences*, **11**, 191–198.