JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

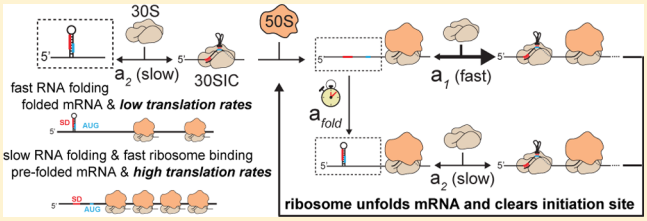
Translation Initiation is Controlled by RNA Folding Kinetics via a Ribosome Drafting Mechanism

Amin Espah Borujeni[†] and Howard M Salis*[,†,‡]

[†]Department of Chemical Engineering and [‡]Department of Biological Engineering, The Pennsylvania State University, University Park, Pennsylvania 16802, United States

Supporting Information

ABSTRACT: RNA folding plays an important role in controlling protein synthesis as well as other cellular processes. Existing models have focused on how RNA folding energetics control translation initiation rate under equilibrium conditions but have largely ignored the effects of nonequilibrium RNA folding. We introduce a new mechanism, called "ribosome drafting", that explains how a mRNA's folding kinetics and the ribosome's binding rate collectively control its translation initiation rate. During cycles of translation, ribosome drafting emerges whenever successive ribosomes bind to a mRNA faster than the mRNA can refold, maintaining it in a nonequilibrium state with an acceleration of protein synthesis. Using computational design, time-correlated single photon counting, and expression measurements, we demonstrate that slow-folding and fast-folding RNA structures with equivalent folding energetics can vary protein synthesis rates by 1000-fold. We determine the necessary conditions for ribosome drafting by characterizing mRNAs with rationally designed ribosome binding rates, folding kinetics, and folding energetics, confirming the predictions of a nonequilibrium Markov model of translation. Our results have widespread implications, illustrating how competitive folding and assembly kinetics can shape the gene expression machinery's sequence−structure−function relationship inside cells.



BACKGROUND AND RATIONALE

Inside cells, the self-assembly kinetics of DNA, RNA, and protein controls the rates of its cellular processes, such as transcription and translation. Since the discovery of the genetic code, it has been a long-standing question to elucidate how its sequence controls the rates of these cellular processes. While much of our understanding of biological self-assembly arises from cell-free, *in vitro* experiments where the system eventually reaches thermodynamic equilibrium, most cellular processes undergo continuous cycles and are maintained in non-equilibrium states.[1−4] In recent years, experimental and computational studies of RNA and protein folding kinetics have led to an increasing appreciation that the biologically relevant structure may not have reached equilibrium and that the kinetic rate of folding often has physiological outcomes.[1,4−7] However, it remains unclear how macromolecules could be maintained in nonequilibrium structures for long times, and whether those structures play a role in controlling cellular function. It also remains difficult to tune the folding kinetics of RNAs and proteins to perturb and study their effect on cellular processes. Here, we apply computational design, Markov modeling, time-resolved fluorescence measurements, and protein synthesis measurements to show how RNA folding kinetics affects the sequence-structure−function relationship controlling translation initiation inside living cells, which is the first step in protein synthesis.

We demonstrate a new mechanism, called "ribosome drafting", that explains how changes in RNA folding kinetics control a protein's synthesis rate by pushing the translation process into the maintenance of nonequilibrium RNA states. We illustrate how computational design can be used to tune the folding kinetics of RNA structures independent of their folding energetics. Our integrated experimental and computational results confirm a quantitative relationship between a mRNA's nucleotide sequence, its folding kinetics, and the rate of its assembly with the ribosome, directly controlling the frequency of translation initiation. Our results also demonstrate a rational design approach for studying assembly kinetics inside cells and provide the theoretical foundation for building more accurate nonequilibrium sequence-to-function models of cellular processes.

Translation initiation is a rate-limiting step in protein synthesis, whereby the 30S ribosomal subunit, $tRNA^{fMet}$, and initiation factors bind to a mRNA to form a 30S initiation complex (30SIC),[8,9] followed quickly by 50S recruitment, 70S assembly, and peptide elongation.[10,11] The 30SIC complex forms several attractive interactions with the mRNA to confer binding affinity and specificity; collectively, these interactions control the rate of ribosome assembly (Figure 1A). First, the ribosome's proteinaceous, positively charged platform domain forms electrostatic interactions with the mRNA's phosphate backbone at single-stranded RNA regions upstream of a start codon. These regions are called standby sites for their ability to

Received: February 8, 2016
Published: May 19, 2016

DOI: 10.1021/jacs.6b01453
J. Am. Chem. Soc. 2016, 138, 7016−7023

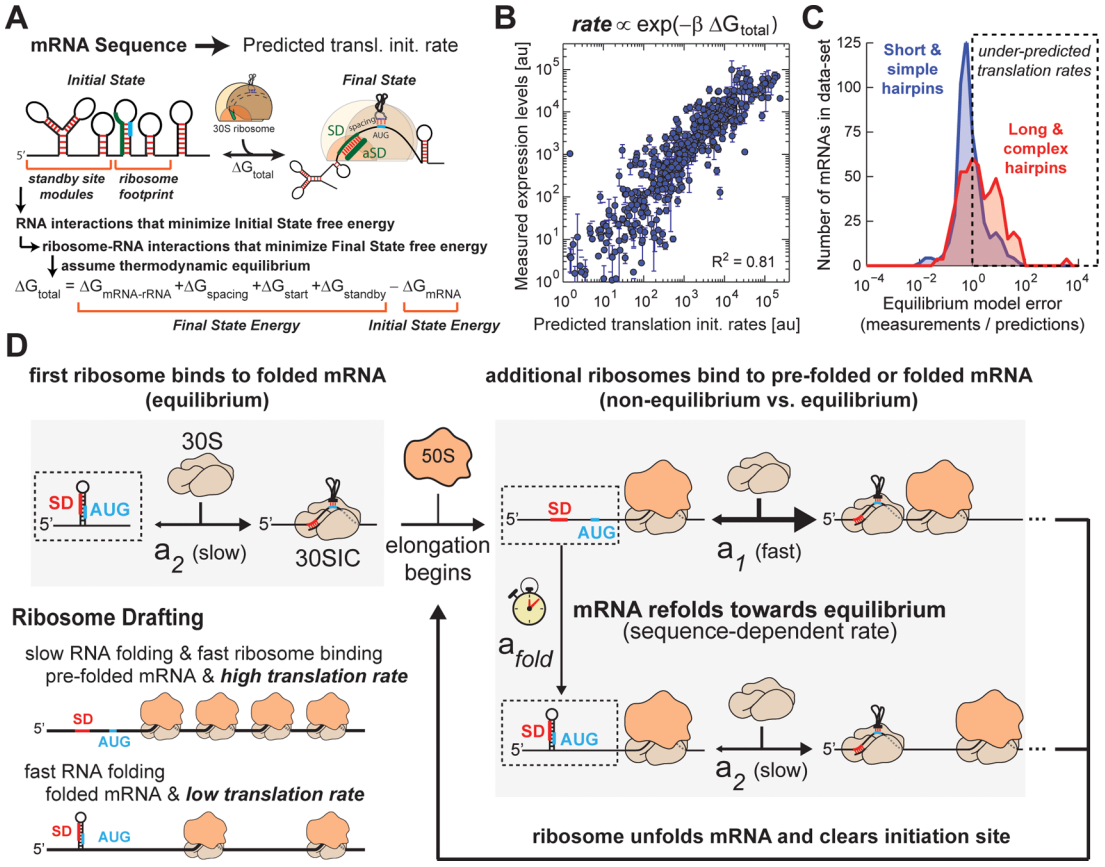


Figure 1. Equilibrium and nonequilibrium models of translation initiation. (A) The translation initiation rates of arbitrary mRNAs are calculated using a ribosome−mRNA free energy model and equilibrium binding conditions. (B) The equilibrium model was tested by comparing expression measurements on 485 diverse mRNAs to predicted translation rates (R^2 = 0.81, $p < 10^{-30}$). (C) The equilibrium model has higher error when predicting the translation rates of mRNAs that are likely to fold slowly because they contain long hairpins with mismatches, bulges, and other complex structures. (D) The ribosome drafting mechanism emerges when the ribosome's binding rate is faster than the mRNA's refolding rate. While the first ribosome binds to folded mRNA, successive ribosomes can bind much faster to prefolded mRNA, maintaining the mRNA in a nonequilibrium state and accelerating its translation initiation rate.

provide the ribosome with a preinitiation landing pad.[12,13] Second, the last nine nucleotides of the rRNA inside the 30SIC are solvent-accessible and can hydrogen bond to the mRNA region upstream-adjacent to the start codon, known as a Shine−Dalgarno (SD) sequence. SD sequences that are more complementary to the (anti-SD) rRNA sequence will form more hydrogen bonds and will bind more tightly.[14] Third, the anticodon loop of the $tRNA^{fMet}$ inside the 30SIC forms hydrogen bonds with the start codon. Fourth, the physical distance between the binding sites of the rRNA and $tRNA^{fMet}$, called spacing, modulates the 30SIC's binding entropy; nonoptimal spacing will either stretch or compress the 30SIC with a corresponding decrease in binding affinity.[15,16] Most importantly here, all RNA structures overlapping with the ribosome's footprint are unfolded prior to translation initiation, accelerated by the ribosome's helicase activity.[17] The unfolding of these RNA structures requires the ribosome to exert work, which will lower the probability that the ribosome binds to the mRNA. GTP hydrolysis, an external source of free energy, only occurs after the 30SIC has been assembled and translation has been initiated.

The probability that the 30SIC is bound to the mRNA, and its corresponding translation rate r, are determined by the change in free energy between the initially folded mRNA and the 30SIC's final state, according to $r \propto \exp(-\beta \Delta G_{total})$, where β is Boltzmann's constant.[13,16] ΔG_{total} is calculated using a sequence-dependent free energy model that quantifies the strengths of the several interactions between the ribosome and mRNA[9,11,13−16,18−21] (Figure 1A). Once translation initiates and protein synthesis begins, another 30S ribosome can bind to the mRNA to continue the cycle. Several cycles of translation typically take place per mRNA.

This model of translation initiation has been critically tested across 485 characterized mRNA sequences in diverse bacterial species with translation rates that varied by 100 000-fold[13,16,18,22] (Figure 1B). Model predictions have been used to tune protein expression levels, engineer genetic circuits, optimize metabolic pathways, and design translation-regulating riboswitches from diverse RNA aptamers.[22−27] For the purposes of this study, we can conceptually simplify the multiterm free energy model using two lumped terms: the 30SIC's binding free energy in the final state ($\Delta G_{complex}$) and the refolding energy of the mRNA's inhibitory structures ($\Delta G_{refold} < 0$). Together, the ribosome's total binding free energy is $\Delta G_{total} = \Delta G_{complex} - \Delta G_{refold}$ (Supporting Information).

However, the model assumes that, after translation has taken place, the mRNAs have sufficient time to refold toward their minimum free energy structures, reaching thermodynamic equilibrium. Within the 485 characterized mRNAs, there is a

DOI: 10.1021/jacs.6b01453
J. Am. Chem. Soc. 2016, 138, 7016−7023

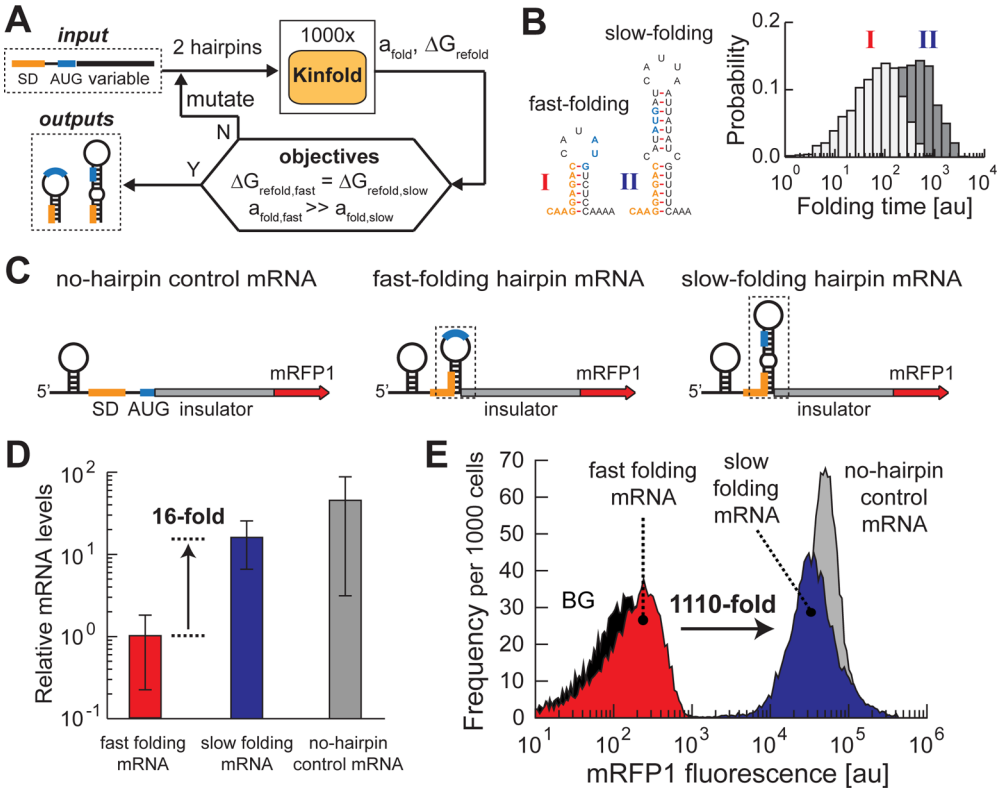


Figure 2. Demonstrating the ribosome drafting mechanism. (A) RNA hairpin sequences were computationally designed to have equivalent folding energetics but maximally different folding kinetics. τ is the mean folding time from 1000 RNA folding trajectories calculated using Kinfold.[33] (B) The structures and folding time distributions of two designed RNA hairpins with similar folding energies ($\Delta G_{refold,fast} = -7.6$, $\Delta G_{refold,slow} = -7.3$ kcal/mol), but different average folding times ($\tau_{fast} = 105$ au, $\tau_{slow} = 503$ au). (C) The expression systems used to measure the effect of RNA hairpins on mRNA levels and translation rates, including designed insulators to prevent undesired structures. The expression systems' (D) mRNA levels and (E) single-cell fluorescence levels were measured in *E. coli* DH10B cultures, comparing mRNAs with fast-folding and slow-folding hairpins to a control mRNA without an inhibitory hairpin. BG, background autofluorescence. Data points and error bars are the average and sd of four independent measurements on two separate days.

subclass of characterized mRNAs that fold into long hairpins with complex structures, including bulges and mismatches, where the equilibrium model's predictions are noticeably less accurate (Figure 1C). Given these observations, we suspected that these longer hairpins could feature slower folding kinetics, which could be responsible for changing the ribosome's binding free energy beyond the assumptions of an equilibrium thermodynamic model. While the potential influence of RNA folding kinetics on translation was considered 20 years ago,[28,29] recent studies have largely ignored its effects,[21,30−32] and a quantitative model or mechanism that relates RNA folding kinetics to translation initiation rate has yet to be proposed or tested.

RESULTS AND DISCUSSION

Initially, to determine whether RNA folding kinetics could play a role in controlling translation, we selected three mRNAs from our data set that each contain an inhibitory hairpin structure. We employed Kinfold,[33] a nucleotide-resolution RNA folding algorithm, to simulate 1000 stochastic folding trajectories and calculate folding time distributions. Two of the hairpins folded noticeably faster than the third with average folding times of 40 and 50 versus 670 in Kinfold time units. The equilibrium model accurately predicted the translation rates of the mRNAs with fast-folding hairpins; however, the translation rate of the mRNA with the slower folding hairpin was under-predicted by 10.3-fold (Figure S1). This result motivated the development and validation of a nonequilibrium mechanism that controls a mRNA's translation initiation rate.

A Ribosome Drafting Mechanism. We propose that ribosome drafting takes place anytime when the ribosome binds to a mRNA faster than the folding kinetics of its inhibitory RNA structures. While the first ribosome that initiates translation must unfold inhibitory RNA structures, successive ribosomes can bind to unfolded mRNAs if the structures fold slowly and if the ribosomes bind quickly (Figure 1D). The first ribosome that binds to a mRNA expends free energy to unfold the mRNA structures that overlap with the ribosome's footprint, lowering the mRNA's translation initiation rate. In contrast, with ribosome drafting, successive ribosomes do not need to unfold those same mRNA structures because they do not have sufficient time to refold, resulting in higher translation initiation rates. As a result, ribosome drafting accelerates protein synthesis.

In one sense, ribosome drafting is a kinetic race between the folding of the mRNA and its assembly with the ribosome. The race has the potential to be competitive because 30S ribosomes can prebind to upstream standby sites and need only undergo sliding and strand displacement to form a 30SIC,[13] which can minimize mass transfer as a rate-limiting step. In another sense, ribosome drafting requires the frequent and coordinated binding of ribosomes in succession; 30SICs must leave ribosome binding sites fast enough to allow new ribosome to bind. As its name suggests, ribosome drafting also decreases the

DOI: 10.1021/jacs.6b01453
J. Am. Chem. Soc. 2016, 138, 7016−7023

amount of work that successive ribosomes need to exert to initiate translation; the first ribosome must unfold the inhibitory mRNA structures, and the remaining ones draft behind it. Together, as ribosomes bind and translation initiates in continuous cycles, ribosome drafting causes the mRNA to rarely reach an equilibrium folded state. Instead, the continuous cycling of translation maintains the mRNA in a nonequilibrium state for its entire lifetime. Depending on the stochastic folding trajectory of the RNA, the RNA may have enough time to fold into an intermediate RNA structure or it may remain fully unfolded when the next ribosome binds. The maintenance of these nonequilibrium states is fully dependent on the ribosome's helicase activity during cycles of translation initiation and can be distinguished from kinetic traps that can occur during RNA folding in an isolated system.

Testing for Ribosome Drafting. In our first test, we computationally designed fast-folding and slow-folding mRNA hairpins that feature two-state folding with very different folding kinetics but equivalent folding energetics. Both RNA hairpins contain a high-affinity Shine–Dalgarno sequence ($\Delta G_{complex}$ = −8.3 kcal/mol), followed by a start codon. We also designed a no-hairpin mRNA control that contains the same Shine–Dalgarno sequence and start codon but lacks any mRNA structures that would inhibit translation rate. We introduced these sequences into our plasmid-based *mRFP1* expression system using identical constitutive promoters. We then utilized flow cytometry to record single-cell fluorescence levels inside transformed *Escherichia coli* DH10B cells maintained in the exponential growth phase (Figure 2). Under the same conditions, we also measured the mRNAs' levels using RT-qPCR.

To design fast-folding and slow-folding RNA hairpins, we combined Kinfold folding simulations with Monte Carlo computational optimization to efficiently search the variable nucleotide spaces (Figure 2A). The algorithm began with an initial pair of fully unfolded RNA sequences and simulated 1000 stochastic folding trajectories. From the ensemble of trajectories, the RNA hairpins' folding time distributions, average folding times, and equilibrium refolding free energies were calculated. The algorithm iterated by randomly inserting, deleting, or mutating nucleotides in the hairpins' variable regions, searching for sequences that maximized the differences in average folding times, while minimizing the differences in their refolding free energies. Sequence changes were accepted or rejected according to this mini-max objective function and the Metropolis criterion. To avoid introducing confounding variables, the algorithm only accepted RNA hairpin designs if their folding times are monoexponentially distributed, indicating two-state folding, if they contained no more than eight consecutive base pairings to minimize RNase-mediated degradation, and if the coding sequences after the start codons utilized only fast codons to maximize their translation elongation rates. We also prevented alternative RNA structures from forming by inserting an AC-rich insulator sequence, and only accepting RNA hairpin designs that folded into a compact ensemble around its minimum free energy structure (Supporting Information).

The computationally optimized RNA hairpin sequences had very similar calculated folding free energies ($\Delta G_{refold,fast}$ = −7.6, $\Delta G_{refold,slow}$ = −7.3 kcal/mol) but very different folding kinetics (average folding times τ_{fast} = 105 au, τ_{slow} = 503 au using Kinfold time units) (Figure 2B). If the RNA hairpins have sufficient time to refold, then the mRNAs' translation rates will be repressed. If the RNA hairpins do not have sufficient time to refold, then the mRNAs' translation rates will be much higher and similar to the no-hairpin control mRNA. According to our ribosome–mRNA free energy model, if the RNA hairpins do not have time to refold, the mRNAs' translation rate will be increased by 30.5-fold.

Single-cell fluorescence measurements using the mRFP1 reporter protein showed that the mRNA with a slow-folding RNA hairpin had a 1110-fold higher protein expression level than the mRNA with a fast-folding hairpin. This very large increase in protein expression level was accompanied by a 16-fold increase in mRNA level, indicating that the mRNA's translation rate increased by 69-fold (Figure 2C,D,E). We performed the same measurements on the no-hairpin control and found similar mRNA levels and only 1.6-fold higher protein expression levels compared with the slow-folding hairpin, suggesting that the ribosome did not need to exert much work to unfold the slow-folding RNA hairpin and initiate translation. These results provided initial evidence that the slow-folding RNA hairpin did not have sufficient time to refold during cycling of translation and that folding kinetics can greatly affect a mRNA's translation rate.

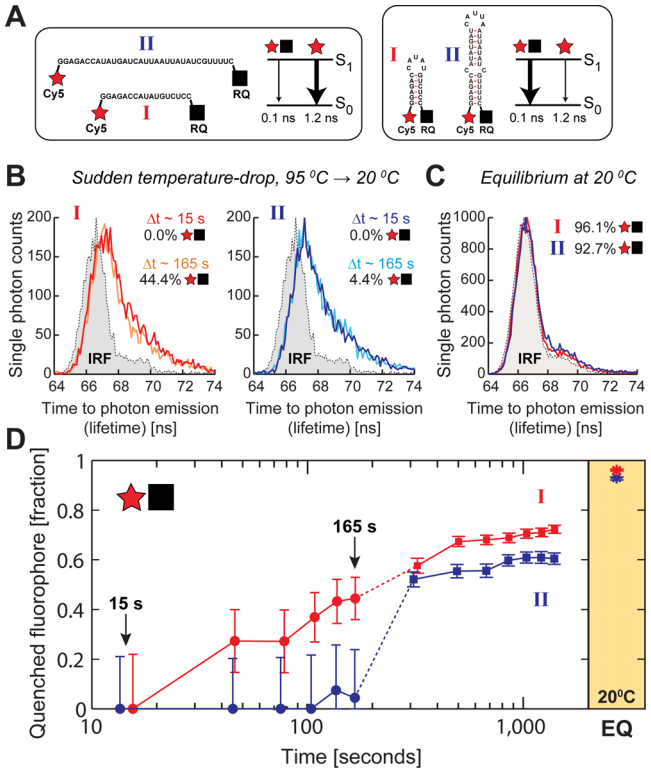


Figure 3. (A) The structural states and Jablonski diagrams for the (I) fast-folding and (II) slow-folding RNA hairpins, including their characteristic lifetimes and preferred emission pathways. (B) Kinetic TCSPC measurements showing lifetime distributions at two time points after RNA hairpins were subjected to a sudden temperature drop. (C) TCSPC measurements showing lifetime distributions after RNA hairpins were separately equilibrated at 20 °C. (D) The fraction of quenched fluorophore in the RNA solutions after a sudden temperature drop. Error bars show the 95% confidence interval. Circles and squares denote separate kinetic TCSPC runs with short and long time windows, respectively. Orange shaded section shows the equilibrium quenched fluorophore fraction at 20 °C. IRF: instrument response function.

DOI: 10.1021/jacs.6b01453
J. Am. Chem. Soc. 2016, 138, 7016–7023

We then employed time-correlated single photon counting (TCSPC) to measure the folding kinetics of the designed RNA hairpins in an *in vitro* system. RNA hairpin sequences were chemically synthesized with a 5′ terminal Cy5 fluorophore and a 3′ terminal Iowa Black RQ quencher (Figure 3A). A free Cy5 fluorophore emits photons with a characteristic lifetime of about 1.2 ns. When the RQ quencher contacts Cy5, photons are emitted much faster with a characteristic lifetime of about 0.1 ns. TCSPC determines the fraction of fluorophore that is either free or quenched by recording lifetime counts and measuring lifetime distributions, followed by convolution with the instrument's response function and fitting to multi-exponential functions (DecayFit v1.4)[34] (Supporting Information). Notably, a fluorophore's characteristic lifetimes are independent of RNA concentration and physiological temperature.

We performed kinetic TCSPC measurements on RNA solutions (100 mM NaCl, 5 mM $MgCl_2$, pH 7.0) at very low RNA concentrations (10 nM) to minimize self-dimerization. We acquired photon counts from individual temperature-dropped RNA solutions in runs where 95 °C RNA solution was injected into 4 °C buffer, and separately, where individual RNA solutions were equilibrated overnight at 20 °C. All lifetime distributions were globally fit together to biexponential functions, yielding characteristic lifetimes of 0.10 and 1.24 ns, with two relative amplitudes that sum to one. The global fit had a χ^2 test value of 1.5 (Supporting Information).

Just after a sudden temperature drop, the first kinetic TCSPC measurements show that both RNAs were unfolded; their lifetime distributions had characteristic times corresponding only to the free fluorophore. Over the next 165 s, the lifetime distributions for the first RNA, computationally designed to be fast-folding, skewed increasingly toward shorter lifetimes, with characteristic times corresponding to a mixture of free and quenched fluorophore. Over the same time period, the lifetime distributions for the second RNA, computationally designed to be slow-folding, did not appreciably change (Figure 3B). In contrast, after the RNAs were separately equilibrated at 20 °C, the lifetime distributions of both RNAs were highly skewed toward shorter lifetimes, indicating that the fluorophore was predominantly quenched (Figure 3C). When these lifetime distributions were globally fit to biexponential decays, the changes in amplitude provided a precise quantification of the RNAs' fraction of quenched fluorophore over the 25 min period (Figure 3D).

Based on these results, the computationally designed fast-folding RNA did indeed fold faster than the computationally designed slow-folding RNA. In these *in vitro* conditions, about 40% of the fast-folding RNA quenched its fluorophore within the first 3 min, while the slow-folding RNA remained largely unfolded. To compare, the slow-folding RNA required 5 min to begin quenching its fluorophore. Once folding began, both RNAs exhibited first-order folding kinetics with relatively constant rate parameters; the folding rate of the fast-folding RNA was about 30% higher than the folding rate of the slow-folding RNA. We provide further discussion of these results below.

Next, we investigated the competitive kinetic race that controls the extent of ribosome drafting within the *in vivo* environment inside cells. According to the mechanism, we should only expect an acceleration of translation rate when the mRNA's refolding kinetics are slower than the rate of ribosome binding. To quantify this relationship, we created the simplest possible nonequilibrium Markov model that describes the transition rates between the three most important ribosome-RNA states (Figure 4A). In state 1, the ribosome has just

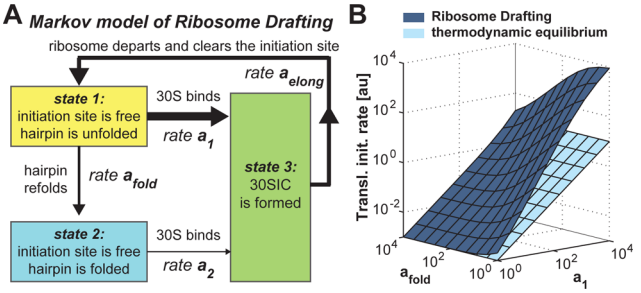


Figure 4. Markov model of ribosome drafting. (A) The model illustrates the kinetic race between RNA refolding and ribosome binding that partitions the system into folded and unfolded mRNA states with distinct translation rates. (B) We show how the ribosome's binding rate (a_1) and hairpin refolding rates (a_{fold}) control the translation rate according to either the (dark blue) nonequilibrium Markov model or the (light blue) equilibrium model of translation. Here, ΔG_{refold} is −15.35 kcal/mol, causing a_1 to be 1000-fold higher than a_2; a_{elong} is always faster than a_1.

departed the mRNA's initiation site, leaving it unfolded and unoccupied. Starting from state 1, the mRNA's initiation site can refold into an RNA hairpin structure with rate a_{fold}. Alternatively, starting from state 1, another ribosome can bind it with a fast rate a_1. If the mRNA refolds (state 2), then a ribosome will bind it with a slow rate a_2. According to our free energy model, the ribosome binding rates are controlled by the ribosome's interactions with the mRNA, according to $a_1 \propto \exp(-\beta\Delta G_{complex})$, and the refolding free energies of inhibitory mRNA structures ($\Delta G_{refold} < 0$), according to $a_2 = a_1 \exp(\beta\Delta G_{refold})$. From both states 1 and 2, when a ribosome binds the mRNA, the resulting ribosome−mRNA structures are identical (state 3). From state 3, the ribosome initiates translation, departs the initiation site, and begins to synthesize protein with a fast rate a_{elong}. The resulting Markov model is closed and recurrent, and its unique Master equation solution shows how ribosome drafting emerges when the ribosome's binding rate is similar or faster than the RNA structure's refolding rate (Figure 4B, Supporting Information).

We tested the extent of the ribosome's kinetic race with RNA folding by designing, constructing, and characterizing 18 mRNAs with systematically varied the ribosome binding rates a_1, fast or slow RNA hairpin refolding kinetics a_{fold}, and constant RNA refolding energetics (ΔG_{refold}). To vary the ribosome's binding rate, we designed six Shine−Dalgarno sequences that altered the ribosome's binding free energy ($\Delta G_{complex}$) from 4.7 to −10.3 kcal/mol and therefore varied a_1 by about 1000-fold (Figure 5A, Supporting Information). We first introduced these SD sequences into a no-hairpin control mRNA and measured reporter protein expression levels to confirm that the equilibrium model could accurately predict the translation rates of mRNAs in the absence of hairpin folding ($R^2 = 0.94$, $p = 0.0012$) (Figure 5B). To design fast-folding and slow-folding RNA hairpins, we repeated our computational hairpin design (Figure 2A) now embedding the six designed Shine−Dalgarno sequences into the hairpin sequences so that they will become sequestered when the RNA hairpin is folded. We then characterized their reporter expression levels as before. All RNA hairpins and their folding time distributions are shown

DOI: 10.1021/jacs.6b01453
J. Am. Chem. Soc. 2016, 138, 7016−7023

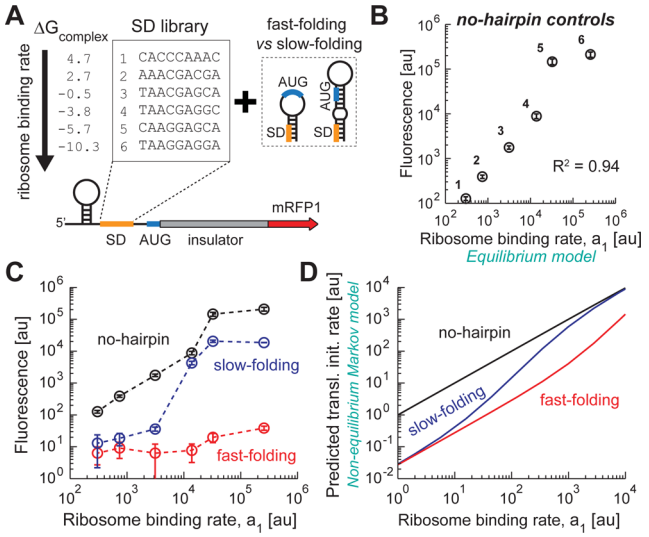


Figure 5. Extent of ribosome drafting is controlled by the ribosome's binding rate. (A) Pairs of mRNAs were designed with varying ribosome binding rates and thermodynamically equivalent fast-folding and slow-folding hairpins. (B) The measured expression levels of the no-hairpin control mRNAs were well-predicted by the thermodynamic model. (C) As the ribosome's binding rate was increased, the measured expression levels from slow-folding mRNAs increased significantly more than the fast-folding mRNAs and approached the no-hairpin control mRNAs. (D) The corresponding Markov model solutions with increasing ribosome binding rates. ΔG_{refold} = −8.0 kcal/mol; $a_{fold,fast}$ = 2000 au; $a_{fold,slow}$ = 20 au; a_{elong} is always faster than a_1. Data points and error bars are the average and sd of four independent measurements on two separate days.

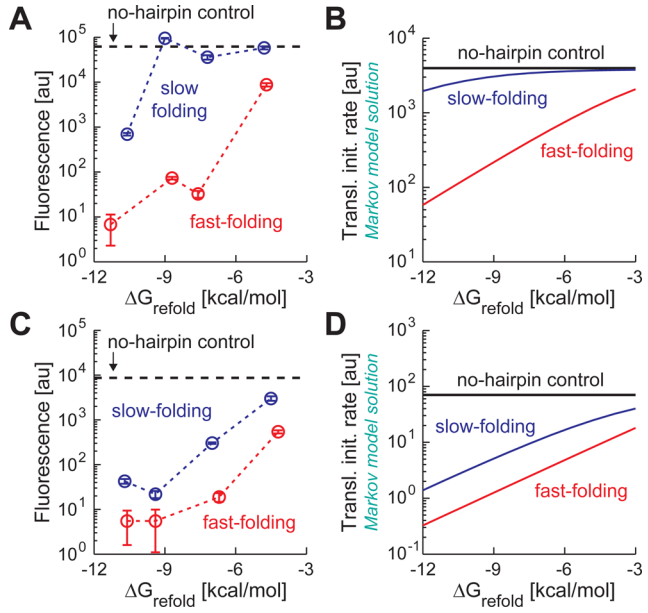


Figure 6. Ribosome drafting's impact on translation rate is controlled by the RNA hairpins' folding times and folding free energies. Thermodynamically equivalent fast-folding and slow-folding hairpins were designed with varying refolding energies (ΔG_{refold}). (A) The measured expression levels of these mRNAs with consensus SD sequences and high ribosome binding rates ($\Delta G_{complex}$ = −8.3 kcal/mol, a_1 = 4190 au). (B) The corresponding Markov model solution with a_1 = 4190 au and varying ΔG_{refold}. (C) The measured expression levels from these mRNAs with nonconsensus SD sequences and low ribosome binding rates ($\Delta G_{complex}$ = 0.8 kcal/mol, a_1 = 70 au). (D) The corresponding Markov model solution with a_1 = 70 au and varying ΔG_{refold}. RNA folding times in Markov models are the same as in Figure 5. Measured expression levels from no-hairpin control mRNAs are shown as horizontal dashed lines. Data points and error bars are the average and sd of four independent measurements on two separate days.

in Figure S3. All sequences, measurements, and calculations are included in the Supplementary Data.

We found that the nonequilibrium Markov model, utilizing the ribosome drafting mechanism, could explain how both the ribosome's binding rate and the RNA hairpin's folding kinetics controlled the mRNA's translation rate (Figure 5C,D). The model correctly predicted the observed sigmoidal relationship where slow-binding ribosomes were forced to unfold fully folded hairpins but where fast-binding ribosomes employed ribosome drafting to bind unfolded mRNAs with accelerated translation rates. As before, when a ribosome quickly bound to a slow-folding mRNA, its translation rate was similar to the no-hairpin control mRNA's translation rate. But when a ribosome slowly bound to a mRNA, its translation rate was unaffected by the folding kinetics of the mRNA's hairpin.

Next, we tested the impact of the RNA hairpins' refolding energetics on the ribosome drafting effect by designing 16 fast-folding and slow-folding inhibitory hairpins with varying refolding free energies (ΔG_{refold} from −11.3 to −4.7 kcal/mol) and inserting them into mRNAs that bound ribosomes with either fast or slow rates ($\Delta G_{complex}$ = −8.3 or 0.8 kcal/mol, respectively) (Figures S5 and S6). Our measurements illustrated that the ribosome drafting mechanism had a very large impact (a 103-fold expression level change) when ribosomes rapidly bound to a mRNA with a slow-folding and highly stable hairpin (ΔG_{refold} = −11.3 kcal/mol), compared with its equivalent fast-folding hairpin (Figure 6A). In contrast, when the same measurements were performed on a slow-folding but less stable hairpin (ΔG_{refold} = −4.7 kcal/mol), ribosome drafting had a much lower impact on the mRNA's translation rate (a 7-fold expression level change), which agrees well with the Markov model calculations (Figure 6B). In both cases, the large time-scale separation between ribosome binding and RNA folding prevented the system from reaching equilibrium, and the impact of the ribosome drafting mechanism was controlled by the hairpin's refolding free energy. However, when ribosomes very slowly bound to the mRNAs, their translation rates were controlled by the hairpins' refolding free energies and not their folding times (Figure 6C). A large change in folding time had a small effect on measured translation rates (4−6-fold expression level changes). Here, the system operated closer to equilibrium conditions, where both the nonequilibrium Markov model and the thermodynamic equilibrium model yielded similar predictions (Figure 6D and Figure S7).

Overall, the ribosome drafting mechanism explains how a mRNA's folding kinetics, refolding energetics, and ribosome binding rate will collectively control its translation rate, altering protein synthesis rates by over 1000-fold. By developing a Markov model and experimentally validating its predictions, we derived and confirmed the general conditions for when cycles of translation maintain mRNAs under nonequilibrium conditions. Ribosome drafting has the highest impact when the ribosome's binding rate is fast, the mRNA's folding kinetics are slow, and inhibitory hairpins are stable.

Intriguingly and perhaps controversially, our *in vitro* kinetic TCSPC measurements revealed a much slower rate of folding

DOI: 10.1021/jacs.6b01453
J. Am. Chem. Soc. 2016, 138, 7016−7023

for both the fast- and slow-folding RNA hairpins studied here, compared with previous RNA folding rate measurements.[35−41] We highlight several aspects of our experimental design that explain these differences. First, we ensured that the *initial state* of RNA folding was a homogeneous solution of extended RNA chains by heating and maintaining the RNAs at 95 °C until they were fully unstructured. We then applied a sudden temperature drop to initiate folding. Second, our RNA solutions were *intentionally dilute* to minimize the rate of self-dimerization and other bimolecular pathways that could accelerate RNA folding. Third, the RNA's *final folded state* is strictly defined as any structure that quenches the RNA's fluorophore. While the fully folded RNA hairpin will be the predominant final state, there may be short-lived structures that temporarily quench the fluorophore as well as long-lived structures that leave the fluorophore unquenched, such as misaligned RNA duplexes. It is also possible that our experimental design promotes compaction instead of zippering, whereby the RNA chain first forms non-native base pairings that result in a heterogeneous mixture of structures, followed by a slow and dynamically stiff folding pathway toward the final state.[39] Altogether, our experimental design has a wider reaction coordinate than previous measurements, which may explain the apparently slower rate in transitioning from initial to final state.

Importantly, because of the strictly defined initial and final states, our experimental design can be precisely mirrored in computational simulations for comparisons. Indeed, we found that corresponding Kinfold dynamic simulations quantitatively reproduced the fractions of quenched fluorophore in fast-folding and slow-folding RNAs (Figure S7). The initial state of these RNAs also closely mirrors the biologically relevant initial state of the mRNA; immediately after translation initiation, this portion of the mRNA exists in a fully unfolded state. However, we did not attempt to create an *in vitro* buffer that fully reproduces the *in vivo* conditions in which RNA folds; for example, macromolecular crowding inside intracellular environments has been reported to accelerate RNA folding kinetics.[42,43] Further investigation will be needed to determine how such experimental design factors influence folding rate.

As our folding simulations show, there is no strict relationship between an RNA hairpin's folding free energy and its folding kinetics; some weakly stable RNA structures fold quickly, while other energetically stable RNA hairpins fold slowly.[5,44] To truly understand the sequence−structure−function relationship controlling translation, it is necessary to simulate RNA folding and to incorporate RNA folding kinetics into a nonequilibrium model. Here, we computationally designed RNA hairpins that feature two-state folding and developed a conceptually simple three-state Markov model that couples together RNA folding with ribosome binding to successfully explain our measurements. By expanding this approach, it is now possible to develop more complex nonequilibrium Markov models that predict the translation rates of arbitrary mRNA sequences with multistate folding pathways.

More conceptually, we show that the assembly kinetics between a very large macromolecular complex (the ribosome) and unstructured polymeric RNA can actually be much faster than the folding kinetics of RNA into higher order structures. Here, we purposefully designed our slow-folding RNAs to possess a more torturous folding pathway with several reversible intermediate transitions that ultimately yielded a very stiff stochastic dynamical system. However, it is likely that slow-folding RNAs are ubiquitous since our optimization algorithm was able to readily find short RNA sequences that were predicted to fold slowly and the predicted RNA structures looked qualitatively similar to natural RNA structures. With additional algorithmic advances, it will be possible to determine how RNA folding kinetics influences genome-wide translation rates.

These results should have widespread implications because, while RNA plays such a central role in controlling several biological processes, the importance of RNA folding kinetics was largely thought to be limited to the processes' early, transient steps. Here, we show that differences in RNA folding rates can have significant and prolonged effects on a biological process well beyond the time scale of RNA folding, motivating investigations into other RNA-mediated processes, including viral particle assembly, transcriptional initiation and termination, translational coupling, metabolite and protein sensing, and dynamic regulation of gene expression levels.[18,21,45−48] With such definitive evidence, RNA folding kinetics should now be routinely included in the list of physical processes that control organism behavior, shaped by molecular evolution.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.6b01453.

Supplementary figures and detailed descriptions of the biophysical model of translation initiation, the Markov model of ribosome drafting, and experimental procedures (PDF)

Supplementary data for six pairs of hairpins sequestering a library of SD sequences with increasing strength, four pairs of hairpins sequestering a strong SD sequence, and four pairs of hairpins sequestering a weak SD sequence and the 485 mRNA dataset (XLS)

■ AUTHOR INFORMATION

Corresponding Author

*salis@psu.edu

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Jian Yang, Philip Bevilacqua, and members of their laboratories for technical advice. We also thank Penn State's Genomics core facility for their services. This research was supported by the Air Force Office of Scientific Research (Grant FA9550-14-1-0089), the Office of Naval Research (Grant N00014-13-1-0074), and an NSF Career Award (Grant CBET-1253641) to H.M.S.

■ REFERENCES

(1) Daniels, K. G.; Tonthat, N. K.; McClure, D. R.; Chang, Y. C.; Liu, X.; Schumacher, M. A.; Fierke, C. A.; Schmidler, S. C.; Oas, T. G. *J. Am. Chem. Soc.* **2014**, *136*, 822.

(2) Karbstein, K.; Doudna, J. a. *J. Mol. Biol.* **2006**, *356*, 432.

(3) Voelz, V. A.; Jäger, M.; Yao, S.; Chen, Y.; Zhu, L.; Waldauer, S. A.; Bowman, G. R.; Friedrichs, M.; Bakajin, O.; Lapidus, L. J.; Weiss, S.; Pande, V. S. *J. Am. Chem. Soc.* **2012**, *134*, 12565.

(4) Kim, H.; Abeysirigunawarden, S. C.; Chen, K.; Mayerle, M.; Ragunathan, K.; Luthey-Schulten, Z.; Ha, T.; Woodson, S. A. *Nature* **2014**, *506*, 334.

DOI: 10.1021/jacs.6b01453
J. Am. Chem. Soc. 2016, 138, 7016−7023

(5) Keller, B. G.; Kobitski, A.; Jäschke, A.; Nienhaus, G. U.; Noé, F. *J. Am. Chem. Soc.* **2014**, *136*, 4534.
(6) Behrouzi, R.; Roh, J. H.; Kilburn, D.; Briber, R. M.; Woodson, S. A. *Cell* **2012**, *149*, 348.
(7) Chakraborty, D.; Collepardo-Guevara, R.; Wales, D. J. *J. Am. Chem. Soc.* **2014**, *136*, 18052.
(8) Simonetti, A.; Marzi, S.; Jenner, L.; Myasnikov, A.; Romby, P.; Yusupova, G.; Klaholz, B. P.; Yusupov, M. *Cell. Mol. Life Sci.* **2009**, *66*, 423.
(9) Yusupova, G. Z.; Yusupov, M. M.; Cate, J. H. D.; Noller, H. F. *Cell* **2001**, *106*, 233.
(10) Ramakrishnan, V. *Cell* **2002**, *108*, 557.
(11) Schmeing, T. M.; Ramakrishnan, V. *Nature* **2009**, *461*, 1234.
(12) de Smit, M. H.; van Duin, J. *J. Mol. Biol.* **2003**, *331*, 737.
(13) Espah Borujeni, A.; Channarasappa, A. S.; Salis, H. M. *Nucleic Acids Res.* **2014**, *42*, 2646.
(14) de Smit, M. H.; van Duin, J. *J. Mol. Biol.* **1994**, *235*, 173.
(15) Chen, H.; Bjerknes, M.; Kumar, R.; Jay, E. *Nucleic Acids Res.* **1994**, *22*, 4953.
(16) Salis, H. M.; Mirsky, E. A.; Voigt, C. A. *Nat. Biotechnol.* **2009**, *27*, 946.
(17) Duval, M.; Korepanov, A.; Fuchsbauer, O.; Fechter, P.; Haller, A.; Fabbretti, A.; Choulier, L.; Micura, R.; Klaholz, B. P.; Romby, P.; et al. *PLoS Biol.* **2013**, *11*, e1001731.
(18) Tian, T.; Salis, H. M. *Nucleic Acids Res.* **2015**, *43*, 7137.
(19) Bentele, K.; Saffert, P.; Rauscher, R.; Ignatova, Z.; Blüthgen, N. *Mol. Syst. Biol.* **2013**, *9*, 675.
(20) Studer, S. M.; Joseph, S. *Mol. Cell* **2006**, *22*, 105.
(21) Duval, M.; Simonetti, A.; Caldelari, I.; Marzi, S. *Biochimie* **2015**, *114*, 18.
(22) Farasat, I.; Kushwaha, M.; Collens, J.; Easterbrook, M.; Guido, M.; Salis, H. M. *Mol. Syst. Biol.* **2014**, *10*, 731.
(23) Makino, T.; Skretas, G.; Kang, T.-H.; Georgiou, G. *Metab. Eng.* **2011**, *13*, 241.
(24) Lin, Z.; Xu, Z.; Li, Y.; Wang, Z.; Chen, T.; Zhao, X. *Microb. Cell Fact.* **2014**, *13*, 104.
(25) Ng, C. Y.; Farasat, I.; Maranas, C. D.; Salis, H. M. *Metab. Eng.* **2015**, *29*, 86.
(26) Kushwaha, M.; Salis, H. M. *Nat. Commun.* **2015**, *6*, 7832.
(27) Espah Borujeni, A.; Mishler, D. M.; Wang, J.; Huso, W.; Salis, H. M. *Nucleic Acids Res.* **2016**, *44*, gkv1289.
(28) Ma, C. K.; Koslesnikow, T.; Rayner, J. C.; Simons, E. L.; Yim, H.; Simons, R. W. *Mol. Microbiol.* **1994**, *14*, 1033.
(29) Poot, R. A.; Tsareva, N. V.; Boni, I. V.; van Duin, J. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 10110.
(30) Goodman, D. B.; Church, G. M.; Kosuri, S. *Science* **2013**, *342*, 475.
(31) Brar, G. A.; Weissman, J. S. *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 651.
(32) Seo, S. W.; Yang, J.-S.; Kim, I.; Yang, J.; Min, B. E.; Kim, S.; Jung, G. Y. *Metab. Eng.* **2013**, *15*, 67.
(33) Flamm, C.; Fontana, W.; Hofacker, I. L.; Schuster, P. *RNA* **2000**, *6*, 325.
(34) Preus, S.; Hildebrandt, L.; Noer, S. L.; Birkedal, V. *Biophys. J.* **2015**, *108*, 163a.
(35) Pörschke, D. *Biophys. Chem.* **1974**, *2*, 97.
(36) Ma, H.; Proctor, D. J.; Kierzek, E.; Kierzek, R.; Bevilacqua, P. C.; Gruebele, M. *J. Am. Chem. Soc.* **2006**, *128*, 1523.
(37) Proctor, D. J.; Ma, H.; Kierzek, E.; Kierzek, R.; Gruebele, M.; Bevilacqua, P. C. *Biochemistry* **2004**, *43*, 14004.
(38) Nagel, J.; Flamm, C.; Hofacker, I.; Franke, K.; De Smit, M.; Schuster, P.; Pleij, C. *Nucleic Acids Res.* **2006**, *34*, 3568.
(39) Chen, S.-J. *Annu. Rev. Biophys.* **2008**, *37*, 197.
(40) Zhuang, X.; Bartley, L. E.; Babcock, H. P.; Russell, R.; Ha, T.; Herschlag, D.; Chu, S. *Science* **2000**, *288*, 2048.
(41) Zarrinkar, P. P.; Williamson, J. R. *Science* **1994**, *265*, 918.
(42) Dupuis, N. F.; Holmstrom, E. D.; Nesbitt, D. J. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 8464.
(43) Kilburn, D.; Roh, J. H.; Guo, L.; Briber, R. M.; Woodson, S. A. *J. Am. Chem. Soc.* **2010**, *132*, 8690.
(44) Kuznetsov, S. V.; Ansari, A. *Biophys. J.* **2012**, *102*, 101.
(45) Chen, Y.-J.; Liu, P.; Nielsen, A. a. K.; Brophy, J. a. N.; Clancy, K.; Peterson, T.; Voigt, C. a. *Nat. Methods* **2013**, *10*, 659.
(46) Hao, C.; Li, X.; Tian, C.; Jiang, W.; Wang, G.; Mao, C. *Nat. Commun.* **2014**, 5.389010.1038/ncomms4890
(47) Mortimer, S. a.; Kidwell, M. A.; Doudna, J. a. *Nat. Rev. Genet.* **2014**, *15*, 469.
(48) Qi, L. S.; Larson, M. H.; Gilbert, L. a.; Doudna, J. a.; Weissman, J. S.; Arkin, A. P.; Lim, W. a. *Cell* **2013**, *152*, 1173.

DOI: 10.1021/jacs.6b01453
J. Am. Chem. Soc. 2016, 138, 7016−7023