THE JOURNAL OF PHYSICAL CHEMISTRY Letters

Letter

pubs.acs.org/JPCL

Exploring the Mutational Robustness of Nucleic Acids by Searching Genotype Neighborhoods in Sequence Space

Qingtong Zhou,[†] Xianbao Sun,[‡] Xiaole Xia,[§] Zhou Fan,[∥,⊥,†,#] Zhaofeng Luo,[▽] Suwen Zhao,[†,#] Eugene Shakhnovich,[*,○] and Haojun Liang[*,‡]

[†]iHuman Institute, ShanghaiTech University, Shanghai 201210, China

[‡]CAS Key Laboratory of Soft Matter Chemistry, Collaborative Innovation Center of Chemistry for Energy Materials, Department of Polymer Science and Engineering, Hefei National Laboratory for Physical Sciences at Microscale, University of Science and Technology of China, Hefei, Anhui 230026, China

[§]Key Laboratory of Industrial Biotechnology, Ministry of Education, School of Biotechnology, Jiangnan University, Wuxi, Jiangsu 214122, China

[∥]Key Laboratory of Computational Biology, Max Planck Independent Research Group on Population Genomics, CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China

[⊥]University of Chinese Academy of Sciences, Beijing 100049, China

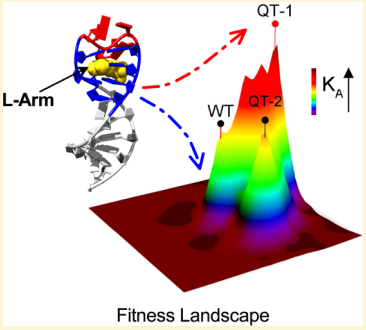
[#]School of Life Science and Technology, ShanghaiTech University, Shanghai 201210, China

[▽]School of Life Science, University of Science and Technology of China, Hefei, Anhui 230026, China

[○]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, United States

Supporting Information

ABSTRACT: To assess the mutational robustness of nucleic acids, many genome- and protein-level studies have been performed, where nucleic acids are treated as genetic information carriers and transferrers. However, the molecular mechanisms through which mutations alter the structural, dynamic, and functional properties of nucleic acids are poorly understood. Here we performed a SELEX in silico study to investigate the fitness distribution of the L-Arm-binding aptamer genotype neighborhoods. Two novel functional genotype neighborhoods were isolated and experimentally verified to have comparable fitness as the wild-type. The experimental aptamer fitness landscape suggests the mutational robustness is strongly influenced by the local base environment and ligand-binding mode, whereas bases distant from the binding pocket provide potential evolutionary pathways to approach the global fitness maximum. Our work provides an example of successful application of SELEX in silico to optimize an aptamer and demonstrates the strong sensitivity of mutational robustness to the site of genetic variation.



As one of the most important biological macromolecules, nucleic acids have diverse functions in encoding, transmission, and expression of genetic information. This diversity is due to the vast sequence space of nucleic acids. High dimensionality of sequence space provides multiple evolutionary pathways to evolve specific phenotypes under selection pressure. However, such pathways might pass through local fitness minima (valleys on fitness landscape) due to detrimental effects of mutation in immediate vicinity of evolved genotypes. To address the fitness effect of mutations, extensive studies have focused on understanding mutational robustness at the genome and protein levels. Previous analyses of DNA sequencing data and mutation accumulation and mutagenesis experiments have revealed that >90% of gene knockouts in *Escherichia coli* are nonlethal,[1] whereas in humans, most amino acid substitutions[2] have fitness effects, amounting to selection coefficients, in the range of 10^{-3} to 10^{-1}, and relatively few substitutions have effects greater than 0.1. The significant mutational robustness of cellular organisms could be explained by buffering mechanisms, including alternative metabolic pathways, genetic redundancy, and modularity. In a biological system without a buffering mechanism, such as some RNA viruses,[3,4] random nucleotide mutations can reduce fitness by an average of nearly 50%, with up to 40% mutations being lethal. These numbers are similar to those found for DNA viruses,[5] and both of these viruses exhibit greater mutational sensitivity than cellular organisms. The roles of nucleic acids are genetic information carriers and transferrers, but the in-depth mutational robustness of nucleic acids themselves, that is, how mutations alter the structural, dynamic, and functional

Received: November 26, 2016
Accepted: January 3, 2017
Published: January 3, 2017

DOI: 10.1021/acs.jpclett.6b02769
J. Phys. Chem. Lett. 2017, 8, 407−414

properties of nucleic acids, remains poorly understood. The exploration of nucleic acid sequence space is largely limited by the available experimental technologies. However, their reach is not sufficient to cover the vast sequence space, limiting the extent to which the mutational robustness of functional nucleic acids can be explored. Therefore, a comprehensive molecular-level analysis of the mutational effects on the structural, dynamic, and functional properties of nucleic acids will provide a solid basis for the understanding of molecular evolution of nucleic acids.

A nucleic acid can adopt distinct secondary or folded tertiary structures that bind targets potently and selectively. These structures, which are denoted aptamers, are generally identified from a random sequence library using Systematic Evolution of Ligands by Exponential Enrichment (SELEX).[6−8] Through the efforts of many researchers, SELEX technology has evolved rapidly, and the current technology enables the identification of a wide range of aptamer targets, ranging from small molecules and metal ions to proteins, biological cells, and tissues. One variant of SELEX is genomic SELEX,[9] which aims to identify genome-encoded nucleic acids with defined properties from a library consisting of short fragments from the human genome rather than random sequences. Through genomic SELEX, both ATP-binding aptamers and GTP-binding motifs were found to be encoded in genomic sequences,[10,11] which provides an interesting perspective on gene regulation. Taking the target-binding affinity as the fitness indicator, aptamers are fitness peaks in a sequence space and are surrounded by many genotypes, some of which have the ability to bind targets.[12,13] A 40-mer aptamer has 120 single mutant, 7020 double mutant, and 266 760 triple mutant genotype neighborhoods in the sequence space (consisting of $4^{40} \approx 10^{24}$ sequences). Although the initial SELEX library comprises up to 10^{18} sequences, it remains very difficult for SELEX to effectively identify functional sequences from an aptamer genotype neighborhood (Supporting Figure S1). Limited by a lack of high-throughput and parallel experimental technologies,[14−18] an exhaustive search is extremely difficult. Thus although the problem of finding a functional aptamer in a sequence space has been successfully addressed by SELEX, the greater challenge is to optimize the found aptamers toward better ligand-binding affinity or selectivity; that is, the inference of local fitness maximum to global fitness maximum in a sequence space, seems unsolvable. Consequently, the mutational robustness of nucleic acids is not fully understood.

Computational approaches are rapid, efficient, and parallelizable and have thus become important tools in nucleic acid research.[21−28] In our previous work,[29] we proposed a computational approach involving the application of SELEX in silico for aptamer selection and successfully identified six novel theophylline-binding RNA aptamers from 4^{13} sequences. In the present study, we selected the L-argininamide (L-Arm)-binding aptamer (the first solved 3D structure of a DNA aptamer[30]) as our research system and used SELEX in silico to predict the fitness (defined as the ligand-binding affinity K_d) of each aptamer genotype neighborhood in the sequence space. SELEX in silico is a two-step approach (Figure S2). The first step is secondary structure-based sequence screening, which aims to collect the sequences that can form the L-Arm-binding motif (Figure 1B) as an enhanced initial library. Then, molecular dynamics (MD) simulation based virtual screening is performed to enrich aptamer-like sequences from the enhanced initial library. All mutants that formed more than

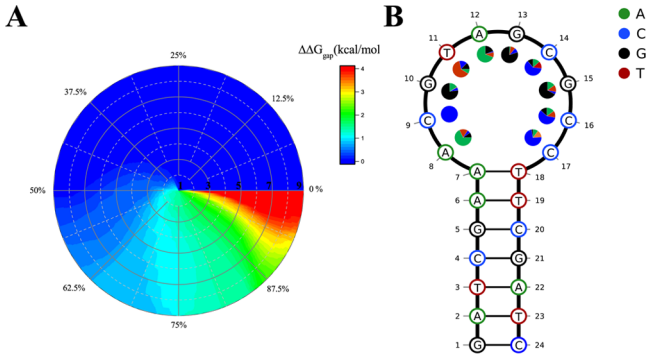


Figure 1. (A) Distribution of the free-energy gaps on target-motif formation in sequence space. The center of the polar plot is the WT L-Arm-binding DNA aptamer, the distance from the center indicates the corresponding Hamming distance of the mutants, the angle indicates the proportion of target motif foldable sequences in each sequence subspace, and the color represents the corresponding free-energy gap ($\Delta\Delta G_{gap}$). (B) Secondary structure of the WT aptamer. The base preferences at each position in the noncanonical region were calculated for the screened best 100 sequences, which were selected by SELEX in silico from the 2619 closest neighbors (whose Hamming distance to the WT aptamer is no greater than three).

five hydrogen bonds with L-Arm, or showed high stability of the binding complex (RMSD < 3 Å), or had comparable predicted binding free energies as the wildtype (WT) aptamer were retained for next round screening (Figure S4 and Table S1). The MD simulation of the WT aptamer was selected as reference during sequence selection of SELEX in silico, accompanied by the consideration of available computational resource and efficiency. The L-Arm-binding aptamer consists of a stem region (bases 1−7 and 18−24) and a noncanonical region (bases 8−17), which form the binding pocket[19,30−32] (Figure 1B). Base C9, which is stacked by a reversed Hoogsteen mismatch pair (A8−C17) and a Watson−Crick pair (G10•C16), forms two hydrogen bonds with L-Arm on its Watson−Crick edge. In the current study, we focused on the mutations in the binding pocket with the exception of C9, that is, on the mutations in bases 8 and 10−17. All of the mutants (4^9 = 262 144) were analyzed in the stage of secondary structure analysis, whereas in the MD-based virtual screening stage, the mutants with Hamming distances ranging from 1 to 3 to the WT aptamer (2619 mutants in total) were selected for in-depth analysis. Two novel functional genotype neighborhoods of L-Arm-binding aptamers were identified through SELEX in silico to exhibit comparable fitness (experimental K_d = 69.3 and 110.7 μM) to the WT aptamer (experimental K_d = 114.4 μM).Combined with previously reported data,[19] the constructed fitness landscape suggests that the mutational robustness of nucleic acids is generally low but infrequently high in certain evolutionary direction. The target-binding ability of nucleic acids is extremely sensitive to the sequence variation in or near the binding pocket, as expected, whereas bases distant from the binding pocket exhibit considerable tolerance to substitutions and represent a potential evolutionary pathway for approaching the global fitness maximum.

The minimum free energy (MFE) secondary structures of the 262 144 mutants can be grouped into 57 unique structural motifs, among which the L-Arm-binding motif (the target motif identified by SELEX in silico, Figure 1B) is the most populated (118,127 sequences). All of the mutants can fold into the target motif with varying energy penalties (Figure S3), and an average

DOI: 10.1021/acs.jpclett.6b02769
J. Phys. Chem. Lett. 2017, 8, 407−414

Table 1. Binding Affinity Estimates for the Wildtype L-Arm-Binding Aptamer and Its Genotype Neighborhoods

aptamer ID	sequence (5′ → 3′)	$\Delta G_{MM/PBSA}$ (kcal/mol)	$\Delta G_{MM/GBSA}$ (kcal/mol)	$-T\Delta S_{NMA}$ (kcal/mol)	$\Delta G_{theor-PB}$ (kcal/mol)[a]	$\Delta G_{theor-GB}$ (kcal/mol)[a]	K_d (μM)[b]	ΔH (kcal/mol)[c]
wild-type	GATCGAAACGTAGCGCCTTCGATC	−45.69(0.11)	−46.04(0.10)	21.42(0.05)	−24.27(0.12)	−24.62(0.11)	114.4(9.2)	−32.30(1.20)
QT-1	GATCGAAACGCGGTGCCTTCGATC	−56.15(0.13)	−55.72(0.17)	21.80(0.03)	−34.35(0.13)	−33.92(0.17)	69.3(6.1)	−33.79(1.23)
QT-2	GATCGAAACGTAGATCCTTCGATC	−47.52(0.12)	−46.43(0.10)	21.68(0.05)	−25.84(0.13)	−24.75(0.11)	110.7(12.6)	−32.11(1.70)
mutant-1 (A8G)	GATCGAAGCGTAGCGCCTTCGATC	−38.32(0.14)	−36.01(0.23)	20.52(0.15)	−17.80(0.21)	−15.49(0.27)	5622[d]	
mutant-2 (A8T)	GATCGAATCGTAGCGCCTTCGATC	−39.36(0.17)	−39.45(0.22)	21.08(0.10)	−18.28(0.20)	−18.37(0.24)	3797[d]	
mutant-3 (C9T)	GATCGAAATGTAGCGCCTTCGATC	−42.18(0.27)	−40.94(0.22)	21.63(0.11)	−20.55(0.29)	−19.31(0.25)	982[d]	
mutant-4 (G10A)	GATCGAAACATAGCGCCTTCGATC	−38.83(0.26)	−37.34(0.14)	20.07(0.16)	−18.76(0.30)	−17.27(0.21)	3797[d]	
mutant-5 (T11C)	GATCGAAACGCAGCGCCTTCGATC	−44.23(0.14)	−42.50(0.23)	21.90(0.17)	−22.33(0.22)	−20.60(0.29)	592[d]	
mutant-6 (A12G&G13T)	GATCGAAACGTGTCGCCTTCGATC	−47.56(0.24)	−43.29(0.21)	20.59(0.12)	−26.97(0.27)	−22.70(0.25)	83[d]	
mutant-7 (G13T)	GATCGAAACGTATCGCCTTCGATC	−49.52(0.15)	−46.99(0.17)	23.04(0.13)	−26.48(0.20)	−23.95(0.21)	83[d]	
mutant-8 (C14T)	GATCGAAACGTAGTGCCTTCGATC	−42.24(0.25)	−40.08(0.15)	23.66(0.26)	−18.58(0.36)	−16.42(0.30)	4512[d]	
mutant-9 (G15A)	GATCGAAACGTAGCACCTTCGATC	−45.96(0.18)	−44.31(0.19)	22.98(0.12)	−22.98(0.22)	−21.33(0.22)	592[d]	
mutant-10 (C16T)	GATCGAAACGTAGCGTCTTCGATC	−41.99(0.30)	−40.13(0.33)	21.83(0.22)	−20.16(0.37)	−18.30(0.39)	3797[d]	
mutant-11 (C17A)	GATCGAAACGTAGCGCATTCGATC	−35.32(0.20)	−36.06(0.30)	20.96(0.15)	−14.36(0.25)	−15.10(0.33)	8870[d]	
mutant-12 (A7T&T18A)	GATCGATACGTAGCGCCATCGATC	−36.62(0.18)	−34.43(0.11)	21.11(0.22)	−15.51(0.28)	−13.32(0.25)	3797[d]	
mutant-13 (A7G&T18C)	GATCGAGACGTAGCGCCCTCGATC	−37.65(0.32)	−37.86(0.21)	21.06(0.23)	−16.59(0.39)	−16.80(0.31)	3797[d]	
random-1 (A8C&T11C&C14G)	GATCGAACCGCAGGGCCTTCGATC	−36.67(0.19)	−36.32(0.15)	19.96(0.21)	−16.71(0.28)	−16.36(0.26)	NA[e]	
random-2 (A8G&A12G&C17T)	GATCGAAGCGTGGCGCTTTCGATC	−38.63(0.24)	−39.24(0.38)	21.44(0.29)	−17.19(0.38)	−17.80(0.48)	NA[e]	
random-3 (G10C&A12C&C16A)	GATCGAAACCTCGCGACTTCGATC	−41.24(0.23)	−37.49(0.15)	21.74(0.20)	−19.50(0.30)	−15.75(0.25)	NA[e]	

[a]Predicted binding free energy was calculated using the molecular mechanics energies combined with the Poisson−Boltzmann or generalized Born and surface area continuum solvation (MM/PBSA and MM/GBSA) with normal-mode analysis (NMA) methods. [b]Binding affinity was determined by circular dichroism study. [c]Thermodynamic parameter enthalpy ΔH was obtained from ITC experiment. [d]Binding parameters K_A and K_d for these mutants were taken from previous research[19] by the linear fitting between lgK_A and log salt concentration lg[NaCl].[19,20] [e]There are no changes in CD signal in the presence of L-Arm for these 1OLD random genotype neighbors, and thus K_A are referred to as NA (not active).

DOI: 10.1021/acs.jpclett.6b02769
J. Phys. Chem. Lett. 2017, 8, 407−414

free-energy gap $\Delta\Delta G_{gap}$ (the difference between the lowest secondary structure energy state ΔG_{MFE} and the target secondary structure state ΔG_{target}, defined as $\Delta G_{MFE} - \Delta G_{target}$) of 0.78 kcal/mol was obtained. Surprisingly, 84% of the mutants (220 721 sequences) have a lower $\Delta\Delta G_{gap}$ than that of the WT aptamer (1.44 kcal/mol). As observed in Figure 1A, the distribution of the $\Delta\Delta G_{gap}$ in the sequence subspace (each sequence subspace was composed of mutated sequences with the same Hamming distance to the WT aptamer and thus contained $3^n C_9^n$ sequences, where n is the Hamming distance) was found to be consistent, regardless of the value of n, indicating that the secondary structure of the DNA aptamer exhibits remarkable tolerance to base substitutions. This finding is different from that for the theophylline-binding RNA aptamer,[29] which has a complex secondary structure and is very sensitive to base substitutions, presented as a sharp peak on a rugged landscape. Similar to SELEX,[29] MD-based virtual screening approaches bias the initial library toward ligand binding by predicting the ligand-binding free energy. Figure S4 shows the distribution and cumulative count of the calculated binding free energy $\Delta G_{MM/PBSA}$ (black points) and $\Delta G_{theor-PB}$ (defined as $\Delta G_{MM/PBSA} - T\Delta S_{NMA}$) (red points), which was analyzed to explore the sequence enrichment of SELEX in silico (Table S1). After four rounds of MD-based virtual screening, the selected sequences were predicted to have noticeably lower ligand-binding free energy than random mutants. After two rounds of MD-based virtual screening, 100 of the 2370 sequences remained due to their high stability or low binding free energy, and the base preferences at each position were then calculated (Figure 1B). The percentage of the most populated bases ranged from the highest peak at the 13th base (cytosine, 80%) to the lowest peak at the 16th base (cytosine, 59%), whereas the reference values for the original and substituted bases among the 2619 mutants were approximately 68.3 and 10.6%, respectively. Although only the closest genotype neighborhoods in the sequence space (single, double, and triple mutants) were searched in the current study, the mutational effect appears highly position-dependent. At positions 10, 13, and 17, the original base is more dominant, whereas multiple mutations of the 14th or 16th base are allowable.

Ensembles of 20 simulations (Table 1 and Figure 2) were run to obtain sufficient sampling of the conformational space,[33] and the collected L-Arm-DNA complex snapshots were then subjected to MM/PB(GB)SA calculations and normal-mode analysis (NMA) to estimate the enthalpic and entropic contributions to the binding free energy, respectively. The snapshot-based normalized frequency distributions of $\Delta G_{MM/PBSA}$, $\Delta G_{MM/GBSA}$, and $-T\Delta S_{NM}$ presented well-defined Gaussian distributions (Figure S5–S7). The calculated binding free energies of the WT aptamer genotype neighbors agreed with experimental mutational effects reported in previous research.[19] Compared with the WT aptamer, most mutants have significantly higher calculated binding free energy, which are correctly predicted to bind the ligand with lower binding affinity, as found by experiment.[19] Interestingly, two mutants, Mutant-6 (A12G&G13T) and Mutant-7 (G13T), whose normalized frequency distribution $\Delta G_{MM/PBSA}$ is slightly shifted toward lower binding free energy relative to the WT (Figure 2A), bind the ligand more tightly than the WT aptamer. Surprisingly, in silico selected genotype neighborhood aptamer QT-1, the best one predicted by SELEX in silico, has lower predicted binding free energy (the mean $\Delta G_{MM/PBSA}$ was

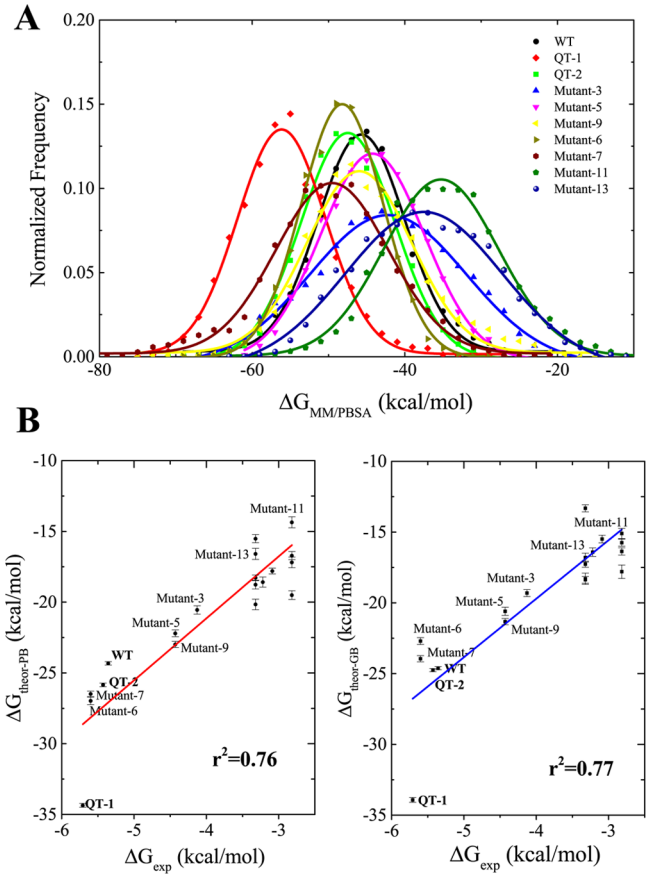


Figure 2. (A) Normalized frequency distribution for $\Delta G_{MM/PBSA}$ is shown in per snapshot for the WT aptamer and its genotype neighborhoods. The expected normal distribution given the same mean and standard deviation for each data set is shown by the lines. (B) Comparison between the experimental ΔG_{exp} (kcal/mol) and the theoretical predictions using MM/PBSA and normal-mode analysis ($\Delta G_{theor-PB}$, left) and MM/GBSA and normal-mode analysis ($\Delta G_{theor-GB}$, right). Error bars show the standard errors. The line represents a linear regression performed on each data set. See Supporting Text S1 for more computational details.

−56.15 kcal/mol) than the WT aptamer (−45.59 kcal/mol), while that of aptamer QT-2 was −47.52 kcal/mol. Similar effects were observed for $\Delta G_{MM/GBSA}$: Aptamer QT-1 was the strongest, followed by aptamer QT-2 and the WT aptamer. The calculated entropies, $-T\Delta S_{NM}$, of these three aptamers have coinciding mean (~21.6 kcal/mol) and standard deviation values. Thus two novel sequences (QT-1 and QT-2), which were identified through SELEX in silico from the aptamer closest neighborhood, were predicted to bind L-Arm as potently as the WT aptamer, and this finding was further experimentally verified (Supporting Text S2).

Circular dichroism (CD) has been extensively used in research on nucleic acids because of its sensitivity to the conformation of anisotropic molecules.[19,34−36] CD spectra were recorded by titrating the DNA aptamer at various concentrations of L-Arm (Figure 3). The WT aptamer displayed a positive peak at 280 nm in the CD spectra, whereas increasing concentrations of L-Arm decreased the molar ellipticity in this region (270−290 nm). This intensity change could indicate that the aptamer has changed its conformation to bind to the ligand, known as the induced-fit binding mechanism.[31,37] The sequences QT-1 and QT-2 exhibit similar changes in the CD

DOI: 10.1021/acs.jpclett.6b02769
J. Phys. Chem. Lett. 2017, 8, 407−414

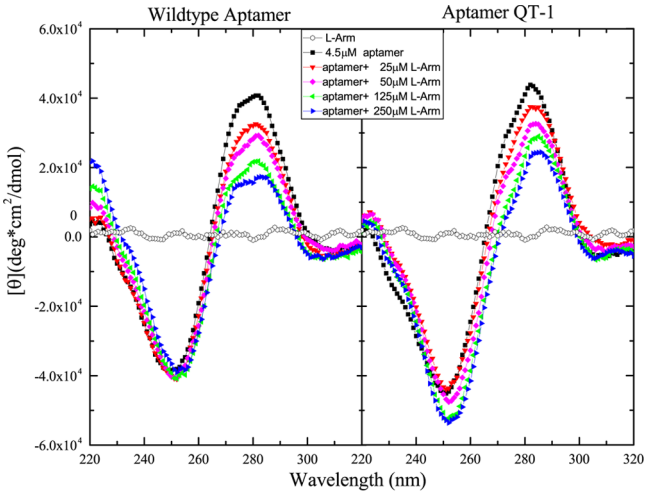


Figure 3. Circular dichroism (CD) spectra of the 4.5 μM L-Arm-binding DNA aptamers titrated with various concentrations of L-Arm in 10 mM sodium phosphate, 25 mM NaCl, pH 6.5. (Left) the WT aptamer; (Right) in silico screened aptamer QT-1.

spectra, which demonstrates that these sequences may bind L-Arm in a manner similar to that found for the WT aptamer. Conversely, for many randomly selected genotype neighborhoods and previously reported clone 12−28 mutants,[19] no changes in the CD signal were found in the presence of L-Arm, which is consistent with their extremely low ligand-binding affinity. To calculate the dissociation constant of the binding (K_d), we analyzed the CD spectra using the optical curve direct fitting method (Figure S8).[19,34,35] For the WT aptamer, the value of K_d was 114.4 ± 9.2 μM, which is similar to the previously reported value (~100[19] and 134.6 μM[34]for the longer 28-mer aptamer, 165.7 μM[35]for 24-mer 1OLD aptamer). The K_d of the aptamer QT-2 is similar to that of the WT aptamer (110.7 ± 12.6 μM), whereas the aptamer QT-1 exhibited strongest binding affinity with L-Arm (69.3 ± 6.1 μM), which is generally consistent with our computational prediction. To obtain the enthalpic contribution to the ligand-binding process, we performed "model-free" isothermal titration calorimetry (ITC) studies[35,38] to avoid any possible fitting bias (Figure S9). By integrating the corrected area under the peaks, the overall enthalpy of binding for the WT enthalpy was found to equal −32.30 ± 1.2 kcal/mol. In contrast, the aptamer QT-1 has a lower ΔH (−33.79 ± 1.23 kcal/mol) than that of QT-2 (−32.11 ± 1.7 kcal/mol). Comparing the experimental data and our prediction (Figure 2B), the coefficients of determination r^2 (0.76 for $\Delta G_{theor-PB}$ and 0.77 for $\Delta G_{theor-GB}$) were obtained, suggesting that MM/PB(GB)SA and NMA methods can accurately rank the ordering of ligand binding affinity of the mutants around the WT aptamer. As noted, the overall entropy change in a binding system[34,39−41] is a combination of the aptamer conformational changes, reorganization of the solvent environment, changes in the translational and conformational freedom of the ligand, and the release of counterion molecules. The development of binding free-energy calculation approaches, especially entropy estimation methods, will greatly facilitate the fast and accurate selection of functional nucleic acid sequences from the vast sequence space.

The aptamer QT-1 is a triple mutant (T11C&A12G&C14T) of the WT aptamer, whereas QT-2 is a double mutant (C14A&G15T). As observed from the binding conformations, the overall structure of the aptamer binding pockets has been retained in aptamers QT-1 and QT-2. As shown in Figure 4,

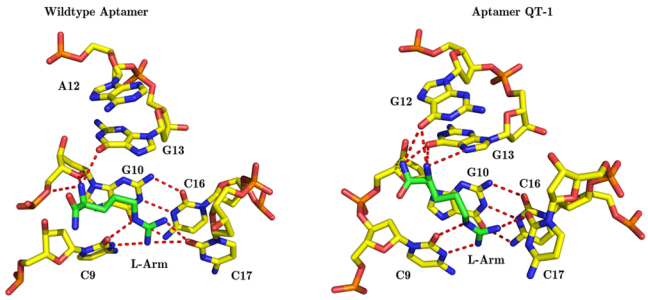


Figure 4. Comparison of the binding modes of L-Arm with the WT aptamer (Left) and in silico screened aptamer QT-1 (Right). The color of carbon in the aptamer was set to yellow, while for L-Arm it was blue. The red dashed lines indicate hydrogen-bonding interactions. Different from the WT aptamer by only three bases, the in silico screened aptamer QT-1 stabilizes L-Arm by constructing a closer binding pocket and forming extra hydrogen bonds between the base edges of G12 and L-Arm.

the guanidinium end of L-Arm was directed toward C16−C17 and forms two hydrogen bonds with the Watson−Crick edge of C9 of both the WT aptamer and its genotype neighborhood QT-1. The guanidinium-C9 pair was further stacked by a Watson−Crick pair G10•C16 and a reversed Hoogsteen mismatch pair (A8−C17). For the WT aptamer, the peptide linkage of L-Arm was directed toward A12 and forms a hydrogen bond with the sugar phosphate backbone of G10 and G13. However, the T11C mutation in the aptamer QT-1 weakens the occasional contact within T11-G15 and facilitates the folding of the T11-G15 loop segment toward L-Arm. The A12G mutation in particular successfully introduces an additional interaction between L-Arm and the carbonyl at C6 of guanine G12. These favorable interactions induced by mutations are conducive to the binding of L-Arm with the aptamer QT-1. Compared with aptamer QT-1, only one intermediate with lower ligand-binding affinity (G15A, K_A = 1689 M^{-1}) was found between the WT aptamer (K_A = 8000 M^{-1}) and aptamer QT-2 (C14A&G15T, K_A = 9033 M^{-1}) in our experimental fitness network (Figure S10). In the present study, we were able to screen only a small fraction of the full sequence space and likely have missed possible strong binders. In addition, some sequences that were discarded in the process of SELEX in silico might undergo significant conformational changes and bind the ligand with a novel binding mode not considered here (false-negatives).

On the basis of previous data[19] and the current study (Table S2), the fitness landscape was constructed to reflect how mutations alter the nucleic acid ligand-recognition ability.[42,43] As shown in Figure 5 and Figure S10, the WT aptamer is located at the origin of the $x-y$ plane, whereas the mutations that occurred in the noncanonical region were represented at different coordinate azimuths to the positions of the mutated bases. The height of each mutant was represented by its ligand binding constant K_A with a corresponding color. Surrounding the WT aptamer, there is one single mutant G13T, two double mutants (A12G&G13T and C14A&G15T), and a triple mutant (T11C&A12G&C14T) with equivalent fitness. Unsurprisingly, almost all of the mutations around the binding pocket (A8T, A8G, C9T, G10A, C16T, C17A, A7T&T18A, and A7G&T18C) abolished the ligand-binding, which suggested

DOI: 10.1021/acs.jpclett.6b02769
J. Phys. Chem. Lett. 2017, 8, 407−414

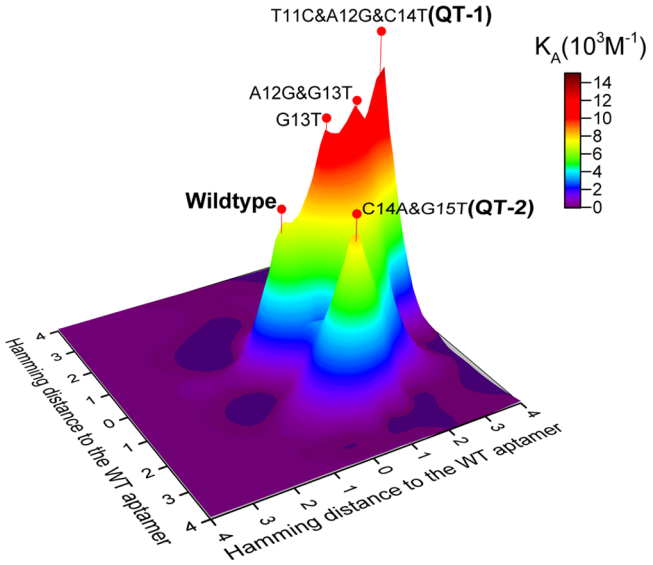


Figure 5. Experimental aptamer fitness landscape of ligand binding. The horizontal x and y axes represent the Hamming distance to the WT aptamer, and the vertical axis represents L-Arm binding affinity. Thirteen mutants from a previous work[19] and ten mutants in this work were used to construct the fitness landscape. The intensity of the fitness peak was represented by the binding constant K_A with corresponding height and color.

that these bases were conserved for L-Arm binding and display significant low mutational tolerance. This conclusion is reasonable because each base surrounding the pocket plays an indispensable role in maintaining the particular ligand binding mode as follows: C9 is the partner of the hydrogens bonds for the ligand, the Watson−Crick pair G10•C16 and reversed Hoogsteen mismatch pair A8−C17 can stack, and the Watson−Crick pair A7•T18 is the connector of the stem region and the noncanonical region of the aptamer. However, the bases located far from the binding pocket (T11, A12, G13, C14, and G15) show remarkable tolerance to mutations. The stepwise mutations G13T, A12G, and T11C&T13G&C14T can successfully evolve the WT aptamer to QT-1 without any fitness loss, which establishes an evolutionary beneficial pathway from one fitness peak to another higher fitness peak through local exploration. Thus from a macroscopic perspective functional nucleic acid aptamers are rare and evolutionarily isolated from one another in the sequence space, and the fitness landscape is a rugged "Badlands" landscape with multiple peaks.[12,13,19,29] Benefiting from the huge screened nucleic acid sequence library (up to 10^{18} sequences) and enrichment of ligand-binding nucleic acid, SELEX technology has greatly increased the probabilities of observing fitness peaks in sequence space. From a microscopic point of view, the majority of the mutants will lose their fitness, whereas only a few genotype neighborhoods in certain regions could be functional. The resulting fitness landscape is Fujiyama-like, while experimental parallel characterization approaches like microarrays[13−15] and computational approaches including SELEX in silico[29] could be adopted for its detailed exploration. The ligand-binding function of the nucleic acid is the central property determining the aptamer fitness in this study. More generally, other biological selective pressures could greatly affect the fitness landscape of a nucleic acid and its evolutionary dynamics. High-throughput screening in search for riboswitches with specific properties such as specific ligand-induced RBS (ribosome binding site) and dynamics in paired and unpaired dynamic states (ligand-free relatively slow translation and ligand-bound relatively fast translation)[44,45] identified synthetic riboswitches that show significant gene expression level change with/without the presence of the desired ligand. In principle, by building a suitable physical model and selecting appropriate physical chemistry parameters, computational approach[46,47] including SELEX in silico can further reveal the fitness landscapes of riboswitches and highlight their possible evolutionary dynamics.

In summary, we applied SELEX in silico to investigate the fitness distribution of nucleic acid genotype neighborhoods in a sequence space. Most mutants fail to bind the ligand with sufficient affinity, which is consistent with previous research for L-Arm-binding DNA aptamer[19] and other aptamers,[12,14,22,29] and this indicates that the aptamer is resistant to base substitutions and relies on the local sequence environment for target binding. Two novel aptamers were experimentally verified to exhibit comparable fitness to the WT aptamer. The experimental nucleic acid fitness landscape constructed based on the current work and previous research[19] suggests that the mutational robustness of nucleic acids is generally low but infrequently high in certain evolutionary direction. Our work provides an example of successful application of SELEX in silico for aptamer optimization and demonstrates the complexity of the mutational robustness of nucleic acids from a novel perspective.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpclett.6b02769.

Workflow of SELEX in silico, detailed experimental and simulation methods, and Supporting Figures S1−S7 and Tables S1 and S2. (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*E.S.: E-mail: shakhnovich@chemistry.harvard.edu.

*H.L.: E-mail: hjliang@ustc.edu.cn.

ORCID

Eugene Shakhnovich: 0000-0002-5087-5724

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The computations in this paper were run on the Odyssey cluster supported by the FAS Division of Science, Research Computing Group at Harvard University, the Shanghai Supercomputer Center, the Supercomputing Center of ShanghaiTech University, and USTC. This work was supported by the China Postdoctoral Science Foundation [2016M591721 to Q.Z.], the National Natural Science Foundation of China [51573175, 91427304, 21434007 to H.L.; 31570755 to Z.L.] and the National Institutes of Health [GM068670 to E.S.].

■ REFERENCES

(1) Baba, T.; Ara, T.; Hasegawa, M.; Takai, Y.; Okumura, Y.; Baba, M.; Datsenko, K. A.; Tomita, M.; Wanner, B. L.; Mori, H. Construction of Escherichia coli K-12 In-Frame, Single-Gene Knockout Mutants: the Keio Collection. *Mol. Syst. Biol.* **2006**, *2*, 2006.0008.

DOI: 10.1021/acs.jpclett.6b02769
J. Phys. Chem. Lett. 2017, 8, 407−414

(2) Eyre-Walker, A.; Keightley, P. D. The Distribution of Fitness Effects of New Mutations. *Nat. Rev. Genet.* **2007**, *8*, 610−8.
(3) Carrasco, P.; de la Iglesia, F.; Elena, S. F. Distribution of Fitness and Virulence Effects Caused by Single-Nucleotide Substitutions in Tobacco Etch Virus. *J. Virol.* **2007**, *81*, 12979−84.
(4) Sanjuan, R.; Moya, A.; Elena, S. F. The Distribution of Fitness Effects Caused by Single-Nucleotide Substitutions in an RNA Virus. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 8396−401.
(5) Domingo-Calap, P.; Cuevas, J. M.; Sanjuan, R. The Fitness Effects of Random Mutations in Single-stranded DNA and RNA Bacteriophages. *PLoS Genet.* **2009**, *5*, e1000742.
(6) Ellington, A. D.; Szostak, J. W. In Vitro Selection of RNA Molecules that Bind Specific Ligands. *Nature* **1990**, *346*, 818−22.
(7) Tuerk, C.; Gold, L. Systematic Evolution of Ligands by Exponential Enrichment: RNA Ligands to Bacteriophage T4 DNA Polymerase. *Science* **1990**, *249*, 505−10.
(8) Zhang, L.; Yang, Z.; Sefah, K.; Bradley, K. M.; Hoshika, S.; Kim, M. J.; Kim, H. J.; Zhu, G.; Jimenez, E.; Cansiz, S.; et al. Evolution of Functional Six-Nucleotide DNA. *J. Am. Chem. Soc.* **2015**, *137*, 6734−7.
(9) Lorenz, C.; von Pelchrzim, F.; Schroeder, R. Genomic Systematic Evolution of Ligands by Exponential Enrichment (Genomic SELEX) for the Identification of Protein-binding RNAs Independent of Their Expression Levels. *Nat. Protoc.* **2006**, *1*, 2204−12.
(10) Vu, M. M.; Jameson, N. E.; Masuda, S. J.; Lin, D.; Larralde-Ridaura, R.; Luptak, A. Convergent Evolution of Adenosine Aptamers Spanning Bacterial, Human, and Random Sequences Revealed by Structure-Based Bioinformatics and Genomic SELEX. *Chem. Biol.* **2012**, *19*, 1247−54.
(11) Curtis, E. A.; Liu, D. R. Discovery of Widespread GTP-Binding Motifs in Genomic DNA and RNA. *Chem. Biol.* **2013**, *20*, 521−32.
(12) Jimenez, J. I.; Xulvi-Brunet, R.; Campbell, G. W.; Turk-MacLeod, R.; Chen, I. A. Comprehensive Experimental Fitness Landscape and Evolutionary Network for Small RNA. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 14984−9.
(13) Knight, C. G.; Platt, M.; Rowe, W.; Wedge, D. C.; Khan, F.; Day, P. J.; McShea, A.; Knowles, J.; Kell, D. B. Array-Based Evolution of DNA Aptamers Allows Modelling of an Explicit Sequence-Fitness Landscape. *Nucleic Acids Res.* **2009**, *37*, e6.
(14) Katilius, E.; Flores, C.; Woodbury, N. W. Exploring the Sequence Space of a DNA Aptamer Using Microarrays. *Nucleic Acids Res.* **2007**, *35*, 7626−35.
(15) Cho, M.; Soo Oh, S.; Nie, J.; Stewart, R.; Eisenstein, M.; Chambers, J.; Marth, J. D.; Walker, F.; Thomson, J. A.; Soh, H. T. Quantitative Selection and Parallel Characterization of Aptamers. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 18460−5.
(16) Gotrik, M. R.; Feagin, T. A.; Csordas, A. T.; Nakamoto, M. A.; Soh, H. T. Advancements in Aptamer Discovery Technologies. *Acc. Chem. Res.* **2016**, *49*, 1903.
(17) Anthony, P. C.; Perez, C. F.; Garcia-Garcia, C.; Block, S. M. Folding Energy Landscape of the Thiamine Pyrophosphate Riboswitch Aptamer. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 1485−9.
(18) Li, C.; Qian, W.; Maclean, C. J.; Zhang, J. The Fitness Landscape of a tRNA Gene. *Science* **2016**, *352*, 837−40.
(19) Harada, K.; Frankel, A. D. Identification of Two Novel Arginine Binding DNAs. *EMBO J.* **1995**, *14*, 5798−811.
(20) Tao, J.; Frankel, A. D. Arginine-Binding RNAs Resembling TAR Identified by in Vitro Selection. *Biochemistry* **1996**, *35*, 2229−2238.
(21) Chushak, Y.; Stone, M. O. In Silico Selection of RNA Aptamers. *Nucleic Acids Res.* **2009**, *37*, e87.
(22) Hoinka, J.; Berezhnoy, A.; Dao, P.; Sauna, Z. E.; Gilboa, E.; Przytycka, T. M. Large Scale Analysis of the Mutational Landscape in HT-SELEX Improves Aptamer Discovery. *Nucleic Acids Res.* **2015**, *43*, 5699−707.
(23) Ahirwar, R.; Nahar, S.; Aggarwal, S.; Ramachandran, S.; Maiti, S.; Nahar, P. In Silico Selection of an Aptamer to Estrogen Receptor Alpha Using Computational Docking Employing Estrogen Response Elements as Aptamer-alike Molecules. *Sci. Rep.* **2016**, *6*, 21285.
(24) Lee, T. S.; York, D. M. Origin of Mutational Effects at the C3 and G8 Positions on Hammerhead Ribozyme Catalysis from Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **2008**, *130*, 7168−9.
(25) Yildirim, I.; Park, H.; Disney, M. D.; Schatz, G. C. A Dynamic Structural Model of Expanded RNA CAG Repeats: a Refined X-ray Structure and Computational Investigations Using Molecular Dynamics and Umbrella Sampling Simulations. *J. Am. Chem. Soc.* **2013**, *135*, 3528−38.
(26) Sponer, J.; Banas, P.; Jurecka, P.; Zgarbova, M.; Kührova, P.; Havrila, M.; Krepl, M.; Stadlbauer, P.; Otyepka, M. Molecular Dynamics Simulations of Nucleic Acids. From Tetranucleotides to the Ribosome. *J. Phys. Chem. Lett.* **2014**, *5*, 1771−1782.
(27) Manson, A.; Whitten, S. T.; Ferreon, J. C.; Fox, R. O.; Hilser, V. J. Characterizing the Role of Ensemble Modulation in Mutation-Induced Changes in Binding Affinity. *J. Am. Chem. Soc.* **2009**, *131*, 6785−93.
(28) Wan, H.; Chang, S.; Hu, J. P.; Tian, Y. X.; Tian, X. H. Molecular Dynamics Simulations of Ternary Complexes: Comparisons of LEAFY Protein Binding to Different DNA Motifs. *J. Chem. Inf. Model.* **2015**, *55*, 784−94.
(29) Zhou, Q.; Xia, X.; Luo, Z.; Liang, H.; Shakhnovich, E. Searching the Sequence Space for Potent Aptamers Using SELEX in Silico. *J. Chem. Theory Comput.* **2015**, *11*, 5939−46.
(30) Lin, C. H.; Patel, D. J. Encapsulating an Amino Acid in a DNA Fold. *Nat. Struct. Biol.* **1996**, *3*, 1046−50.
(31) Lin, P. H.; Tsai, C. W.; Wu, J. W.; Ruaan, R. C.; Chen, W. Y. Molecular Dynamics Simulation of the Induced-fit Binding Process of DNA Aptamer and L-Argininamide. *Biotechnol. J.* **2012**, *7*, 1367−75.
(32) Albada, H. B.; Golub, E.; Willner, I. Computational Docking Simulations of a DNA-Aptamer for Argininamide and Related Ligands. *J. Comput.-Aided Mol. Des.* **2015**, *29*, 643−54.
(33) Wright, D. W.; Hall, B. A.; Kenway, O. A.; Jha, S.; Coveney, P. V. Computing Clinically Relevant Binding Free Energies of HIV-1 Protease Inhibitors. *J. Chem. Theory Comput.* **2014**, *10*, 1228−1241.
(34) Lin, P. H.; Tong, S. J.; Louis, S. R.; Chang, Y.; Chen, W. Y. Thermodynamic Basis of Chiral Recognition in a DNA Aptamer. *Phys. Chem. Chem. Phys.* **2009**, *11*, 9744−50.
(35) Bishop, G. R.; Ren, J.; Polander, B. C.; Jeanfreau, B. D.; Trent, J. O.; Chaires, J. B. Energetic Basis of Molecular Recognition in a DNA Aptamer. *Biophys. Chem.* **2007**, *126*, 165−75.
(36) Liu, W.; Zheng, B.; Cheng, S.; Fu, Y.; Li, W.; Lau, T.-C.; Liang, H. G-quadruplex Formation and Sequence Effect on the Assembly of G-rich Oligonucleotides Induced by Pb^{2+} Ions. *Soft Matter* **2012**, *8*, 7017.
(37) Forster, U.; Weigand, J. E.; Trojanowski, P.; Suess, B.; Wachtveitl, J. Conformational Dynamics of the Tetracycline-Binding Aptamer. *Nucleic Acids Res.* **2012**, *40*, 1807−17.
(38) Ren, J.; Jenkins, T. C.; Chaires, J. B. Energetics of DNA Intercalation Reactions. *Biochemistry* **2000**, *39*, 8439−47.
(39) Mukherjee, A. Entropy Balance in the Intercalation Process of an Anti-Cancer Drug Daunomycin. *J. Phys. Chem. Lett.* **2011**, *2*, 3021−3026.
(40) Morton, C. J.; Ladbury, J. E. Water-Mediated Protein-DNA Interactions: the Relationship of Thermodynamics to Structural Detail. *Protein Sci.* **1996**, *5*, 2115−8.
(41) Spolar, R. S.; Record, J. M. T. Coupling of Local Folding to Site-Specific Binding of Proteins to DNA. *Science* **1994**, *263*, 777−84.
(42) Poelwijk, F. J.; Kiviet, D. J.; Weinreich, D. M.; Tans, S. J. Empirical Fitness Landscapes Reveal Accessible Evolutionary Paths. *Nature* **2007**, *445*, 383−6.
(43) Tannenbaum, E.; Deeds, E. J.; Shakhnovich, E. I. Equilibrium Distribution of Mutators in the Single Fitness Peak Model. *Phys. Rev. Lett.* **2003**, *91*, 138105.
(44) Lynch, S. A.; Desai, S. K.; Sajja, H. K.; Gallivan, J. P. A High-Throughput Screen for Synthetic Riboswitches Reveals Mechanistic Insights Into Their Function. *Chem. Biol.* **2007**, *14*, 173−84.
(45) Desai, S. K.; Gallivan, J. P. Genetic Screens and Selections for Small Molecules Based on A Synthetic Riboswitch that Activates Protein Translation. *J. Am. Chem. Soc.* **2004**, *126*, 13247−13254.

DOI: 10.1021/acs.jpclett.6b02769
J. Phys. Chem. Lett. 2017, 8, 407−414

(46) Espah Borujeni, A.; Mishler, D. M.; Wang, J.; Huso, W.; Salis, H. M. Automated Physics-Based Design of Synthetic Riboswitches from Diverse RNA Aptamers. *Nucleic Acids Res.* **2016**, *44*, 1−13.
(47) Domin, G.; Findeiss, S.; Wachsmuth, M.; Will, S.; Stadler, P. F.; Morl, M. Applicability of A Computational Design Approach for Synthetic Riboswitches. *Nucleic Acids Res.* **2016**, gkw1267.

DOI: 10.1021/acs.jpclett.6b02769
J. Phys. Chem. Lett. 2017, 8, 407−414