Supplementary Materials

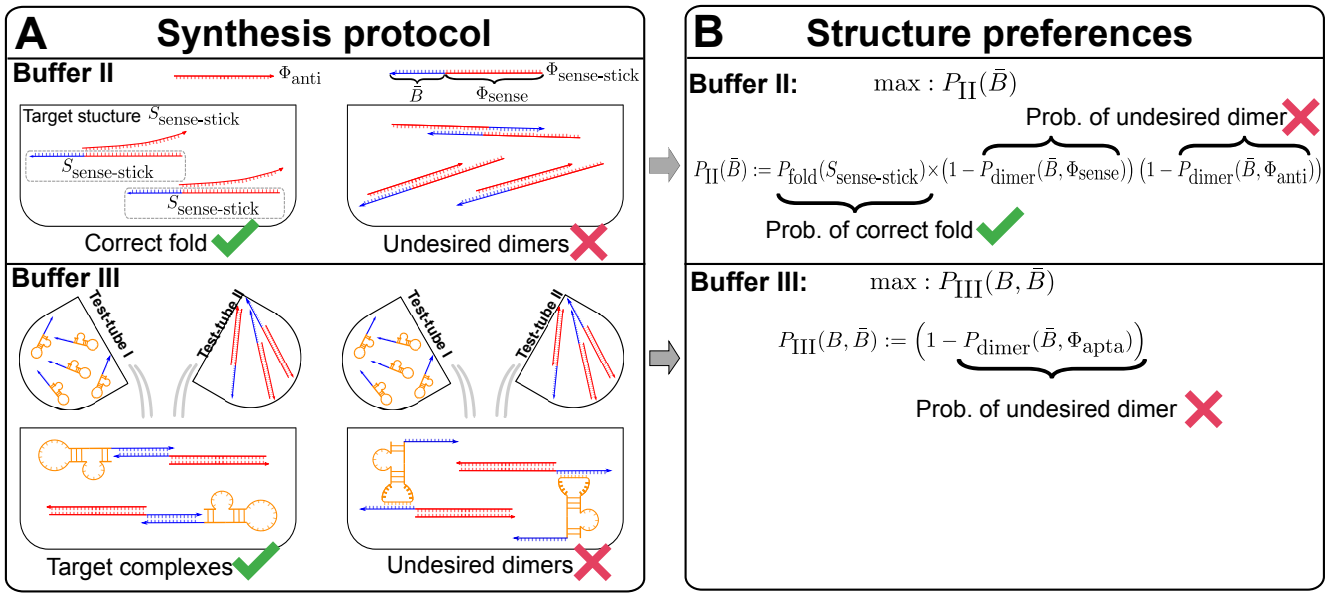

Fig. S1: Structure preferences in Buffer II and III for dsRNA cargoes. (A) The target structure and undesired dimer in Buffer II and III. (B) Converting structure preferences in (A) into to objective functions.

A Structure preferences for double stranded RNA cargoes

AptaBlocks can also design sticky bridges for dsRNA cargoes, such as siRNA. The structure preferences in Buffer II and Buffer III are different from the ones introduced in the main manuscript. Let Φ_{sense} and Φ_{anti} represent the sense and anti-sense parts of an siRNA. Assume the sticky bridge $\bar{B}$ is conjugated to the sense part Φ_{sense} of the siRNA and let $\Phi_{\text{sense-stick}}(\bar{B})$ be the conjugation between Φ_{sense} and $\bar{B}$. In Buffer II, we expect structure $S_{\text{sense-stick}}$ (shown in Fig. S1 Buffer II) be the dominant structure for $\Phi_{\text{sense-stick}}(\bar{B})$ and therefore the probability $P_{\text{fold}}(S_{\text{sense-stick}}|\Phi_{\text{sense-stick}}(\bar{B}))$ should be maximized. At the same time, undesired dimerizations between the sticky bridge $\bar{B}$ and either the sense part Φ_{sense} of the siRNA or anti-sense part Φ_{anti} of the siRNA should be avoid. Considering all those preferences, we maximize

$$P_{\text{II}}(\bar{B}) := P_{\text{fold}}(S_{\text{sense-stick}}|\Phi_{\text{sense-stick}}(\bar{B})) \times \big(1 - P_{\text{dimer}}(\bar{B}, \Phi_{\text{sense}})\big) \times \big(1 - P_{\text{dimer}}(\bar{B}, \Phi_{\text{anti}})\big) \quad (1)$$

For Buffer III, we only need to prevent dimerzation between the sticky bridge $\bar{B}$ and the aptamer Φ_{apta}. Hence, we define $P_{\text{III}}(B, \bar{B})$ as

$$P_{\text{III}}(B, \bar{B}) := 1 - P_{\text{dimer}}(\bar{B}, \Phi_{\text{apta}}). \tag{2}$$

B Stacking hybridization model for RNA-RNA interaction

In this paper, to reduce time complexity but still achieve reasonable results, we propose to minimize $P^{\theta}_{\text{stacking}}(\mathcal{U}, \mathcal{V}|\mathbb{S})$ instead of $P_{\text{dimer}}(\mathcal{U}, \mathcal{V})$ (time complexity $O(|V|^6)$) [1, 2]) to prevent dimerization between $\mathcal{U}$ and $\mathcal{V}$. In fact, we are not interested in computing the exact thermodynamic probability $P_{\text{dimer}}(\mathcal{U}, \mathcal{V})$ that $\mathcal{U}$ and $\mathcal{V}$ dimerize. Instead, our goal is to avoid strong dimerization between $\mathcal{U}$ and $\mathcal{V}$. To reduce time complexity, we only consider stacking hybridization between two RNA strands without considering other types of interacting structures between them. Hence, we compute $P_{\text{stacking}}(\mathcal{U}, \mathcal{V})$ rather than $P_{\text{dimer}}(\mathcal{U}, \mathcal{V})$. Furthermore, to avoid strong binding between two RNA strands, we propose to minimize $P^{\theta}_{\text{stacking}}(\mathcal{U}, \mathcal{V})$ that $\mathcal{U}$ and $\mathcal{V}$ interact by stacking hybridization with free energy less than θ. Based on the following lemma, we know that minimizing $P^{\theta}_{\text{stacking}}(\mathcal{U}, \mathcal{V}|\mathbb{S})$, the upper bound of $P^{\theta}_{\text{stacking}}(\mathcal{U}, \mathcal{V})$, can eventually force $P^{\theta}_{\text{stacking}}(\mathcal{U}, \mathcal{V})$ to decrease. Here, $P^{\theta}_{\text{stacking}}(\mathcal{U}, \mathcal{V}|\mathbb{S})$ is the probability that $\mathcal{U}$ and $\mathcal{V}$ interact by stacking hybridization with free energy less than θ conditioned on that $\mathcal{U}$ and $\mathcal{V}$ actually interact only by stacking hybridization.

Lemma 1. $P^{\theta}_{\text{stacking}}(\mathcal{U}, \mathcal{V}) \leq P^{\theta}_{\text{stacking}}(\mathcal{U}, \mathcal{V}|\mathbb{S})$

Proof. We define $\mathbb{C}$ to represent all conformations that $\mathcal{U}$ and $\mathcal{V}$ interact by stacking hybridization with free energy less than θ and $\mathbb{S}$ to represent all conformations that $\mathcal{U}$ and $\mathcal{V}$ interact by stacking hybridization. Based on the definitions, we know $\mathbb{C} \subseteq \mathbb{S}$. Then we have

$$\begin{aligned} P^{\theta}_{\text{stacking}}(\mathcal{U}, \mathcal{V}) &= P(\mathbb{C}) \\ &= P(\mathbb{C} \cap \mathbb{S}) \\ &= P(\mathbb{C}|\mathbb{S})P(\mathbb{S}) \\ &\leq P(\mathbb{C}|\mathbb{S}) \\ &= P^{\theta}_{\text{stacking}}(\mathcal{U}, \mathcal{V}|\mathbb{S}). \end{aligned} \tag{3}$$

The third line in Eq. (3) is derived based on Kolmogorovs definition. The fourth line in Eq. (3) is derived based on $P(\mathbb{S}) \leq 1$. □

The probability $P^{\theta}_{\text{stacking}}(\mathcal{U}, \mathcal{V}|\mathbb{S})$ can be efficiently computed as following. Let U and V be the sequences of two RNA molecules and assume $|U| < |V|$. The free energy of the stacking hybridization of length w between $\mathcal{U}$ staring at U_k and $\mathcal{V}$ staring from $V_{k'}$ is $\Delta G^s(U_k, V_{k'}, w)$. The energetics of RNA-RNA interactions can be viewed as a stepwise process, therefore, $\Delta G^s(U_k, V_{k'}, w)$ consists of the energy for exposing the binding site in the appropriate conformation for individual molecules and the energy gain from hybridization at the binding site [3]. Then the partition function of stacking hybridization between $\mathcal{U}$ and $\mathcal{V}$ can be calculated by

$$Z(U, V) = \sum_{k<|U|} \sum_{k'<|V|} \sum_{w<|U|} \exp\left[-\Delta G^s(U_k, V_{k'}, w)/k_B T\right]. \tag{4}$$

Based on the same framework, the partition function of stacking hybridization with free energy smaller than θ can be computed by

$$Z^{\theta}(U, V) = \sum_{k<|U|} \sum_{k'<|V|} \sum_{w<|U|} \exp\left[-\Delta G^s(U_k, V_{k'}, w)/k_B T\right] \times I(\Delta G^s(U_k, V_{k'}, w) < \theta), \tag{5}$$

where $I(a < b)$, an indicator function, equals 1 when $a < b$ and 0 otherwise. Both $Z^{\theta}(U, V)$ and $Z(U, V)$ can be computed in $O(|V|^3)$. Then we can compute the probability $P^{\theta}_{\text{stacking}}(\mathcal{U}, \mathcal{V}|\mathbb{S})$ using (6) in $O(|U|^2|V|)$.

$$P^{\theta}_{\text{stacking}}(\mathcal{U}, \mathcal{V}|\mathbb{S}) = \frac{Z^{\theta}(U, V)}{Z(U, V)}. \tag{6}$$

C Computation details

C.1 Computation of P_{fold}

Given a sequence ϕ and its structural ensemble Ω, the partition function

$$Q(\phi) = \sum_{s \in \Omega} \exp\left(-\Delta(\phi, s)/RT\right), \tag{7}$$

can be used to compute the equilibrium probability [4, 5] of any subset $\Gamma \in \Omega$ of the structural ensemble Ω:

$$P_{\text{fold}}(\Gamma|\phi) = \frac{Q(\phi, \Gamma)}{Q(\phi, \Omega)}, \tag{8}$$

R is the universal gas constant and T is the temperature. Γ could be a specific structure or a set of structures of interest. For example, $S_{\text{apta-stick}}$ in $P_{\text{fold}}\left(S_{\text{apta-stick}}|\Phi_{\text{apta-stick}}\left(B\right)\right)$ in Eq. (1) main

text is a set of structures where the sticky bridge part B is constrained to be unpaired and the structures of the rest of the strand are unconstrained.

C.2 Computation of the probability of the preserved original structure

The probability of the preserved original structure for aptamters is defined by $P_{\text{fold}}\left(S_{\text{apta-stick}}|\Phi_{\text{apta-stick}}\left(B\right)\right)$ in Eq. (1) main text. For given B, $P_{\text{fold}}\left(S_{\text{apta-stick}}|\Phi_{\text{apta-stick}}\left(B\right)\right)$ can be easily computed by (8). Similarly, the probability of the preserved original structure for cargoes is defined by $P_{\text{fold}}\left(S_{\text{cargo-stick}}|\Phi_{\text{cargo-stick}}\left(\bar{B}\right)\right)$ in Eq. (2) main text. For given $\bar{B}$, $P_{\text{fold}}\left(S_{\text{cargo-stick}}|\Phi_{\text{cargo-stick}}\left(\bar{B}\right)\right)$ can also be easily computed by (8).

D Implementation details

D.1 Designing sticky bridges for single aptamer-cargo pair

We first generated 100 aptamer sequences of size 10 drawn from a uniform nucleotide distribution. In the same fashion, we generated 100 ssRNA cargo sequences of size 10. Next, we randomly combined those 100 aptamers and 100 cargoes without replacement to obtain 100 aptamer-cargo pairs. For each aptamer-cargo pair, we set the length of the sticky bridge to 10 base pairs. Two sequence strands in the sticky bridges are respectively concatenated to aptamers and cargoes. The design scenario is similar to Fig. 1.

Other than the length of the sticky bridge, AptaBlocks and Nupack Design need one more parameter, respectively. AptaBlocks needs λ to balance between structure preferences and energy enforcement. We assigned $\lambda = 1.0, 1.5, 2.0$. Nupack Design needs a random seed for initialization.

For each simulated aptamer-cargo pair, we ran AptaBlocks 10 times for each λ and ran Nupack Design 10 times using a different random seed. Then we computed the averages of the probabilities of preserving the structures of the aptamers and the cargoes, the probabilities of hybridization between the designed sticky bridges, and the probabilities of incurring undesired dimers. The average of these results is summarized and detailed in Fig. 2.

D.2 Length of sticky bridges

We utilized the same simulation dataset generated in D.1 but we varied the length of the sticky bridges from 6 base pairs to 11 base pairs. For each aptamer-cargo pair, we ran AptaBlocks 10

times using $\lambda = 1.0$ and Nupack Design 10 times using 10 different random seeds. We computed the averages of the probabilities of preserving the structures of the aptamers and the cargoes, the probabilities of hybridization between the designed sticky bridges, and the probabilities of incurring undesired dimers over the 10 designs. We associated the average probabilities to each aptamer-cargo pair and display the averages over 100 aptamer-cargo pairs in Fig. 3.

D.3 Designing universal sticky bridges

We designed universal sticky bridges for one aptamer and multiple ssRNA cargoes using AptaBlocks. We utilized the same simulation dataset generated in D.1, but for each aptamer-cargo pair, we used the cargo sequence as the seed to generate additional cargoes of identical size. For this, we first, set all cargo sequences to the seed cargo sequence. Then, we randomly selected several locations on the sequences according to the desired sequence similarity and mutated that position. Here, sequence similarity is defined as the ratio of the number of matching bases between two sequences to the length of the sequences. Furthermore, for cargoes that need to be delivered by the same aptamer, we ensure that the pairwise sequence similarity between any two cargoes is the same.

For each design, we ran AptaBlock 10 times using $\lambda = 1.0$ and Nupack Design 10 times with 10 random seeds, and computed the average performance. For a total of 100 designs, we computed the average over the average performance for each design and report the result in Fig. 5. Notably, we computed the probabilities of preserving the original structures rather than computing the probabilities for aptamers and cargoes separately as shown in Fig. 4.

E Comparison based on the secondary structure model with pseudoknots

We compared AptaBlocks with Nupack Design on preserving structures of aptamers and cargoes using a secondary structure model with pseudoknots. The Fig. S2 shows their performance. Both AptaBlocks and Nupack Design do not consider pseudoknots in their designs. But Nupack Design used the same energy parameters used by Nupack Analysis which we used to compute the probability of preserving structures considering pseudoknots. Although Nupack Design has an advantage by using the same energy parameters, AptaBlocks still outperforms this method.

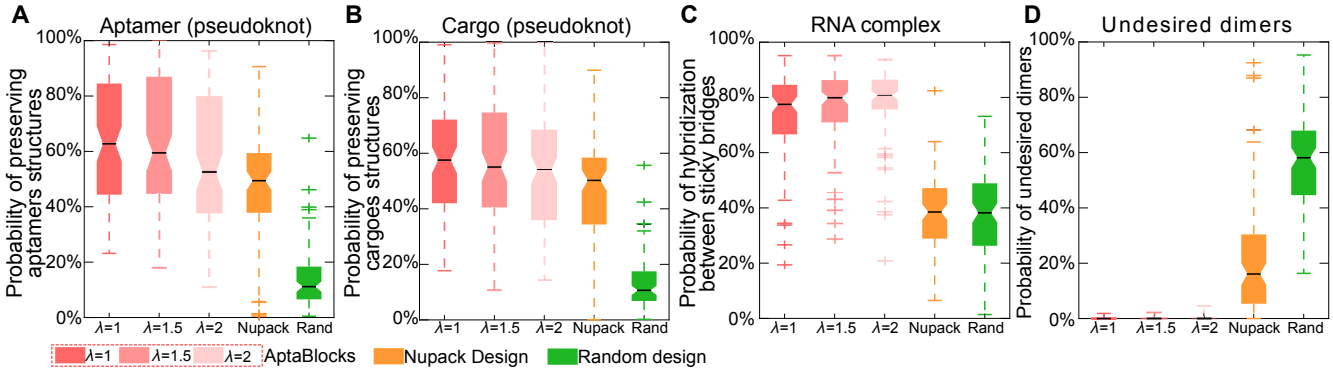

Fig. S2: Comparison of the competing algorithms. (A) Comparison on conserving the secondary structure of aptamers. The probability of sticky bridges being unpaired is computed using a secondary structure model with pseudoknots. (B) Comparison on conserving the secondary structure of cargoes. The probability of sticky bridges being unpaired is computed using a secondary structure model with pseudoknots. (C) Comparison on binding affinity. The binding affinity is approximated by computing the probability of hybridization between sticky bridges using RNA-RNA interaction model without pseudoknots. (D) Comparison on incurring undesired dimers. The probability of undesired dimers is computed by our proposed model in supplementary materials B. Larger λ indicates that AptaBlocks focuses more on minimizing binding free energy between sticky bridges.

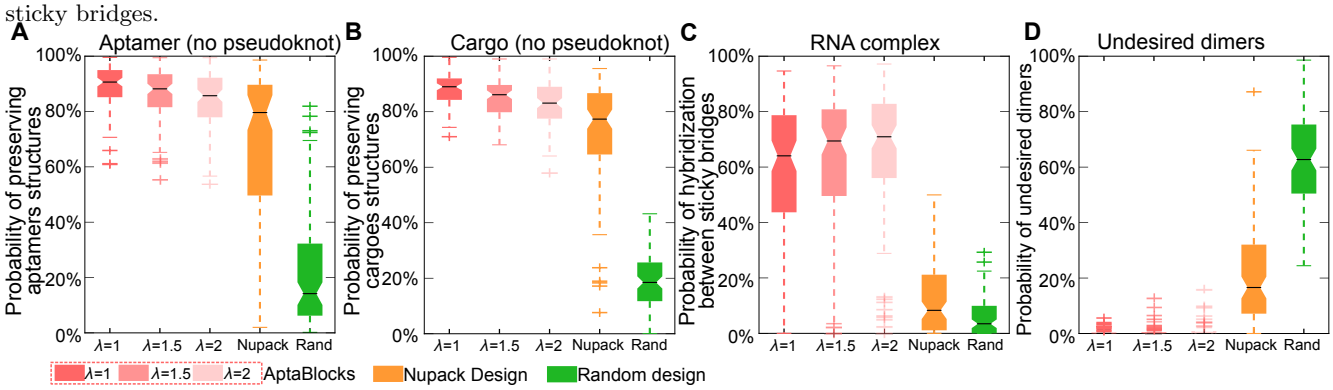

Fig. S3: Comparison of the competing algorithms. (A) Comparison of conserving the secondary structure of aptamers. The probability of sticky bridges being unpaired is computed using a secondary structure model without pseudoknots. (B) Comparison of conserving the secondary structure of cargoes. The probability of sticky bridges being unpaired is computed using a secondary structure model without pseudoknots. (C) Comparison of binding affinity. The binding affinity is approximated by computing the probability of hybridization between sticky bridges using an RNA-RNA interaction model. (D) Comparison on incurring undesired dimers. Larger λ indicates that AptaBlocks focuses more on increasing the melting temperature of sticky bridges.

F Test on simulation data for aptamers with stable structures

We used the following procedure to generate aptamer sequences with stable structures. Many real world truncated aptamers are typically from 20 to 30 nt [6, 7], therefore, we first set the length of the simulated aptamers to be 30 nt. Then, we randomly generated 1,000,000 sequences with 30 nt and ranked them based on their predicted minimum free energy in ascending order. In the end, we selected the top 25% of the sequences and randomly picked 100 sequences as the apatamer sequences for the experiments. In addition, we randomly generate 100 cargo sequences with 10 nt. The comparisons between AptaBlocks and other approaches are detailed in Fig. S3. Similar to the trend shown in main text Fig. 2, Figs. S3A and B show that AptaBlocks outperforms other approaches on preserving structures of aptamers and cargoes. Despite that AptaBlocks' design does not take pseudoknots into account, it performed quite well on the energy model with pseudoknots

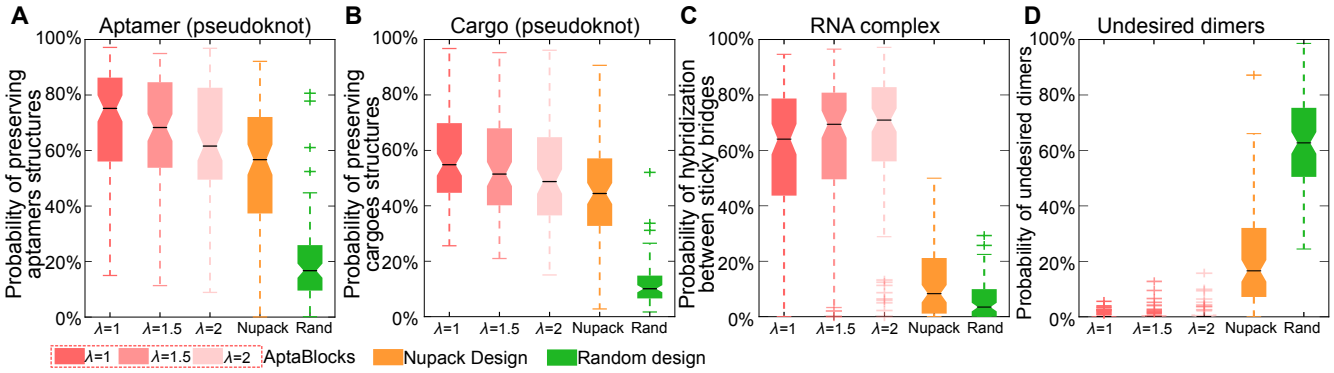

Fig. S4: Comparison of the competing algorithms. (A) Comparison of conserving the secondary structure of aptamers. The probability of sticky bridges being unpaired is computed using a secondary structure model considering pseudoknots. (B) Comparison of conserving the secondary structure of cargoes. The probability of sticky bridges being unpaired is computed using a secondary structure model considering pseudoknots. (C) Comparison of binding affinity. The binding affinity is approximated by computing the probability of hybridization between sticky bridges using an RNA-RNA interaction model. (D) Comparison on incurring undesired dimers. Larger λ indicates that AptaBlocks focuses more on increasing the melting temperature of sticky bridges.

(see Fig. S4), again significantly outperforming its competitors. Furthermore, we observe that the sticky bridges designed by AptaBlocks have higher probabilities to hybridize than other designs by a large margin as shown in Fig. S3C, but lower probabilities to incur undesired dimerizations as shown in Fig. S3D.

G Nupack Design

Nupack Design [8, 9] is a framework for designing the sequences of multiple nucleic acid strands intended to hybridize in solution via a prescribed reaction pathway. Sequence design is formulated as a multistate optimization problem using a set of target test tubes to represent reactant, intermediate, and product states of the system, as well as to model crosstalk between components. Tailored to our sticky bridge design problem, the aptamer conjugated with one part of the sticky bridge and the cargo conjugated with the other part of the sticky bridge are considered as reactants and the aptamer-sticky bridge-cargo complex is consider as the target, on which the target structure preference can be imposed.

References

[1] Fenix W D Huang et al. “Partition function and base pairing probabilities for RNA-RNA interaction prediction”. In: *Bioinformatics* 25.20 (2009), pp. 2646–2654. arXiv: 0903.3283.

[2] Hamidreza Chitsaz et al. “A partition function algorithm for interacting nucleic acid strands”. In: *Bioinformatics* 25.12 (2009), pp. 365–373.

[3] Ulrike Mückstein et al. “Thermodynamics of RNA-RNA binding”. In: *Bioinformatics* 22.10 (2006), pp. 1177–1182.

[4] Robert M Dirks and Niles a Pierce. “Partition Function and Base-Pairing Probability Algorithms for Nucleic Acid Secondary Structure including Pseudoknots”. In: *J Comput Chem* 24 (2003), pp. 1664–1677.

[5] J S McCaskill. “The equilibrium partition function and base pair binding probabilities for RNA secondary structure.” In: *Biopolymers* 29.6-7 (1990), pp. 1105–1119.

[6] Meriem Baouendi et al. “Solution structure of a truncated anti-MUC1 DNA aptamer determined by mesoscale modeling and NMR”. In: *FEBS Journal* 279.3 (2012), pp. 479–490.

[7] Eun Hee Lee et al. “Highly Sensitive Detection of Bisphenol A by NanoAptamer Assay with Truncated Aptamer”. In: *ACS Applied Materials and Interfaces* 9.17 (2017), pp. 14889–14898.

[8] Brian R Wolfe et al. “Constrained Multistate Sequence Design for Nucleic Acid Reaction Pathway Engineering”. In: *Journal of the American Chemical Society* 139.8 (2017), pp. 3134–3144.

[9] Joseph N Zadeh et al. “Software News and Updates NUPACK : Analysis and Design of Nucleic Acid Systems”. In: *Journal of Computational Chemistry* 28.1 (2009), pp. 73–86.