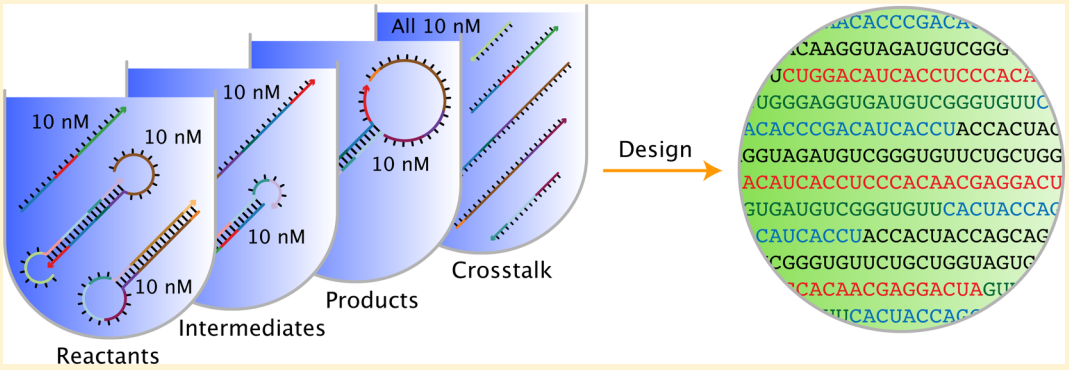
JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

Constrained Multistate Sequence Design for Nucleic Acid Reaction Pathway Engineering

Brian R. Wolfe,[†,#] Nicholas J. Porubsky,[‡,#] Joseph N. Zadeh,[†] Robert M. Dirks,[†,⊥] and Niles A. Pierce*[,†,§,∥]

[†]Division of Biology & Biological Engineering, California Institute of Technology, Pasadena, California 91125, United States
[‡]Division of Chemistry & Chemical Engineering, California Institute of Technology, Pasadena, California 91125, United States
[§]Division of Engineering & Applied Science, California Institute of Technology, Pasadena, California 91125, United States
[∥]Weatherall Institute of Molecular Medicine, University of Oxford, Oxford OX3 9DS, United Kingdom

Supporting Information



ABSTRACT: We describe 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. Each target test tube contains a set of desired "on-target" complexes, each with a target secondary structure and target concentration, and a set of undesired "off-target" complexes, each with vanishing target concentration. Optimization of the equilibrium ensemble properties of the target test tubes implements both a positive design paradigm, explicitly designing for on-pathway elementary steps, and a negative design paradigm, explicitly designing against off-pathway crosstalk. Sequence design is performed subject to diverse user-specified sequence constraints including composition constraints, complementarity constraints, pattern prevention constraints, and biological constraints. Constrained multistate sequence design facilitates nucleic acid reaction pathway engineering for diverse applications in molecular programming and synthetic biology. Design jobs can be run online via the NUPACK web application.

INTRODUCTION

Life is orchestrated by programmable biopolymers—DNA, RNA, and proteins—that execute complex self-assembly and disassembly processes to grow, regulate, and repair organisms. The emerging discipline of molecular programming is inspired by these biological proofs of principle and seeks to establish sequence design principles and algorithms that enable robust encoding of a desired molecular function into biopolymer sequences. To engineer dynamic self-assembly and disassembly processes, it is necessary to control not just equilibrium properties but also the kinetic pathways by which molecules interact. During the past decade, the programmable chemistry of nucleic acid base pairing has provided a fertile design space for engineering pathway-controlled self-assembly and disassembly processes.[1,2]

Molecular programmers engineer nucleic acid reaction pathways using an ever-increasing variety of small conditional DNA and RNA motifs (scDNAs and scRNAs) that exploit diverse design elements to interact and change conformation via prescribed hybridization cascades.[1,2] Modes of nucleating interactions include toehold/toehold,[3−10] loop/toehold,[11,12] loop/loop,[13,14] and template/toehold[12] hybridization. Modes of strand displacement include three-way branch migration,[3−5,7−11] four-way branch migration,[6,12,14,15] and spontaneous dissociation.[7,10,12] To exert control over the order of self-assembly and disassembly events, scDNAs are designed to coexist metastably (i.e., the molecules are kinetically trapped) or stably (i.e., the molecules are thermodynamically trapped),

Received: December 12, 2016
Published: February 13, 2017

DOI: 10.1021/jacs.6b12693
J. Am. Chem. Soc. 2017, 139, 3134−3144

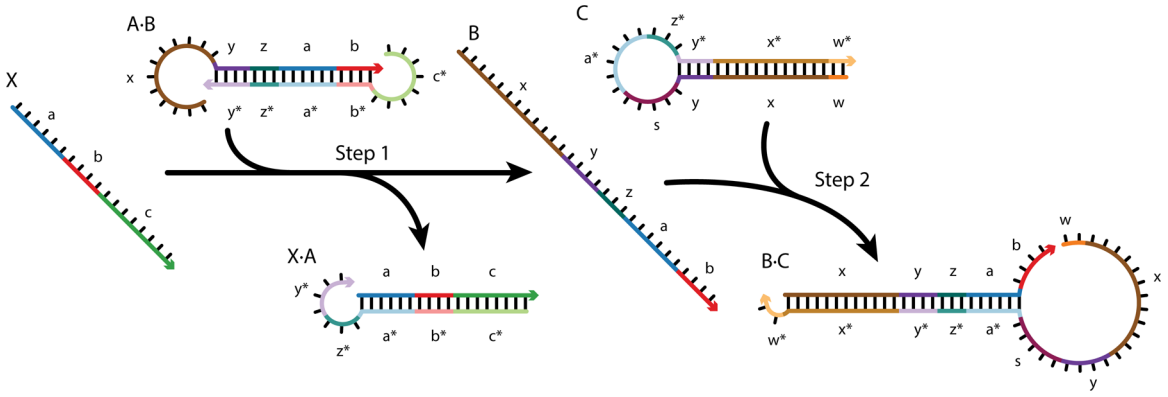


Figure 1. Reaction pathway for conditional Dicer substrate formation via shape and sequence transduction with small conditional RNAs (scRNAs).[12] scRNA A·B detects input X (comprising sequence 'a-b-c'), leading to production of Dicer substrate B·C (targeting independent sequence 'w-x-y-z'). Step 1: X displaces A from B via toehold-mediated three-way branch migration and spontaneous dissociation. Step 2: B assembles with C via loop/toehold nucleation and three-way branch migration to form Dicer substrate B·C. See reaction pathways for additional case studies in SI, section S2.1.

with the next step in the reaction pathway triggered either by a cognate molecular input detected from the environment or by a molecular output of a previous step in the reaction pathway. Principles for engineering conditional metastability include nucleation barriers,[4,8] topological constraints,[13,14] toehold sequestration,[5,7,8,10,12] and template unavailability,[12] while principles for engineering conditional stability include cooperativity[16] and sequence transduction.[12] These pathway-controlled self-assembly and disassembly reactions have been driven by the enthalpy of base pairing[3−6,8,10−13] and the entropy of mixing.[7,8,10,12] These design elements have enabled the rational design and construction of scDNAs executing diverse dynamic functions, including catalysis, signal amplification, sequence transduction, shape transduction, Boolean logic, and locomotion.[1,2]

Devising a new reaction pathway is akin to molecular choreography, requiring conception of both the scRNA participants (or equivalently scDNA participants) and the dance that they will execute via pathway-controlled self-assembly and disassembly operations. Owing to the modularity of scRNA function, new reaction pathways representing complex dynamic functions can often be fruitfully sketched by hand. Once a new reaction pathway has been choreographed, the task remains of encoding the intended pathway-controlled interactions and conformational changes into the sequences of the constituent scRNAs. To program this dynamic function, the nucleic acid sequences must be designed so that the molecules predominantly execute the desired on-pathway interactions while avoiding off-pathway alternatives. Here, we address the dual challenges of formulating and solving the sequence design problem for nucleic acid reaction pathway engineering.

DESIGN FORMULATION

Reaction Pathway Specification. Consider a set of nucleic acid molecules intended to execute a prescribed hybridization cascade. For example, the reaction pathway of Figure 1 describes scRNAs that, upon binding to input X, perform shape and sequence transduction to form a Dicer substrate targeting an independent output Y for silencing.[12] A reaction pathway specifies the *elementary steps* (each a self-assembly or disassembly operation in which complexes form or break) by which the molecules are intended to interact, the desired secondary structure for each on-pathway complex, and the complementarity relationships between sequence domains in the molecules. For example, in the reaction pathway of Figure 1, there are two elementary steps (Step 1: X + A·B → X·A + B, Step 2: B + C → B·C) involving six on-pathway complexes (X, A·B, X·A, B, C, B·C) and numerous sequence domains ('a*' complementary to 'a', 'b*' complementary to 'b', and so on).

In addition to specifying a set of desired on-pathway elementary steps, each reaction pathway also implicitly specifies a much larger set of off-pathway interactions, corresponding to undesired *crosstalk* between components within the pathway or with components from other unrelated reaction pathways. To perform sequence design for reaction pathway engineering, we formulate a multistate optimization problem to explicitly design for on-pathway elementary steps (a positive design paradigm) and against off-pathway crosstalk (a negative design paradigm).

Multistate Test Tube Design Ensemble. A multistate test tube design problem is specified as a set of target test tubes, Ω. Each tube, $h \in \Omega$, contains a set of desired *on-target complexes*, Ψ_h^{on}, and a set of undesired *off-target complexes*, Ψ_h^{off}. For each on-target complex, $j \in \Psi_h^{on}$, the user specifies a target secondary structure, s_j, and a target concentration, $y_{h,j}$. For each off-target complex, $j \in \Psi_h^{off}$, the target concentration is vanishing ($y_{h,j} = 0$) and there is no target structure ($s_j = \varnothing$). The set of complexes in tube h is then $\Psi_h \equiv \Psi_h^{on} \cup \Psi_h^{off}$, and the set of all complexes in multistate test tube ensemble Ω is $\Psi \equiv \cup_{h\in\Omega}\Psi_h$.

Consider specification of the multistate test tube ensemble, Ω, for the design of N orthogonal systems for a reaction pathway of M elementary steps. One *elementary step tube* is specified for each step $m = 0, ..., M$ for each system $n = 1, ..., N$ (treating formation of the initial reactants as a precursor "Step 0"). Additionally, a single *global crosstalk tube* is specified to minimize off-pathway interactions between the reactive species generated during all elementary steps of all systems. The total number of target test tubes is then $|\Omega| = (M + 1) \times N + 1$.

A detailed description of our approach for specifying target test tubes is provided in the Supporting Information (SI), section S2.2. To illustrate this approach, Figure 2a depicts target test tubes for the reaction pathway of Figure 1. There are three elementary step tubes, each containing on-target complexes corresponding to the products of the corresponding step: the Reactants tube (Step 0) contains on-targets X, A·B, and C; the Step 1 tube contains on-targets X·A and B; the Step

DOI: 10.1021/jacs.6b12693
J. Am. Chem. Soc. 2017, 139, 3134−3144

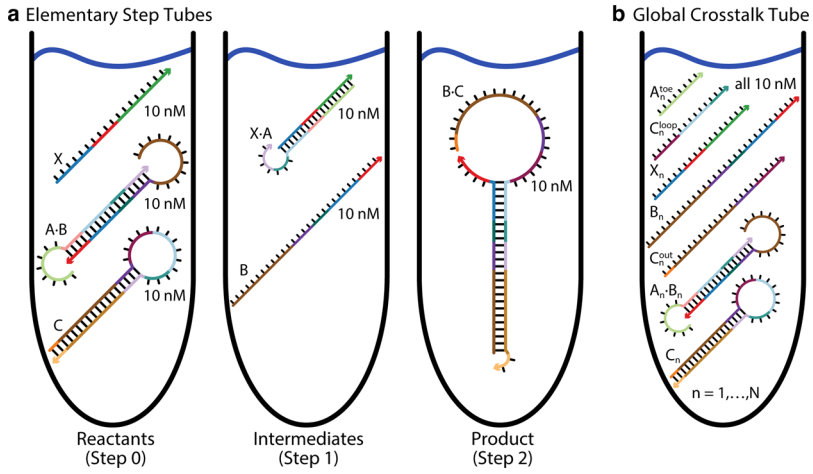


Figure 2. Target test tubes for reaction pathway of Figure 1 (conditional Dicer substrate formation via shape and sequence transduction with scRNAs). (a) Elementary step tubes. Reactants tube (Step 0): target X and scRNAs A·B and C. Step 1 tube: X·A and B. Step 2 tube: Dicer substrate B·C. Each target test tube contains the depicted on-target complexes corresponding to the on-pathway products for a given step (each with the depicted target secondary structure and a target concentration of 10 nM) as well as off-target complexes (not depicted) corresponding to on-pathway reactants and off-pathway crosstalk for a given step. To design N orthogonal systems, there are three elementary step tubes for each system $n = 1, ..., N$. (b) Global crosstalk tube. Contains the depicted on-target complexes corresponding to reactive species generated during Steps 0, 1, 2 as well as off-target complexes (not depicted) corresponding to off-pathway interactions between these reactive species. To design N orthogonal systems, the global crosstalk tube contains a set of on-targets and off-targets for each system $n = 1, ..., N$. Design conditions: RNA in 1 M Na^+ at 37 °C. See SI, section S2.2.6, for additional details on this case study including definition of the reactive species in the global crosstalk tube. See SI, section S2.2, for a general description of target test tube specification and additional case studies.

2 tube contains on-target B·C. Within each target test tube, each on-target complex is depicted by its target secondary structure labeled with its target concentration. Each elementary step tube also contains off-targets (with vanishing target concentration) of two varieties: reactants that are intended to fully convert into the on-pathway products, and off-pathway crosstalk products between these same reactants. Hence, these elementary step tubes are designed for full conversion of cognate reactants into cognate products and against local crosstalk between these same reactants.

To simultaneously design N orthogonal systems, three elementary step tubes of the type shown in Figure 2a are specified for each system. Furthermore, to design against off-pathway interactions within and between systems, a single global crosstalk tube is specified in Figure 2b. In the global crosstalk tube, the on-target complexes correspond to all reactive species generated during all elementary steps ($m = 0, 1, 2$) for all systems ($n = 1, ..., N$); the off-target complexes correspond to noncognate interactions between these reactive species (see SI, section S2.2 for details on defining reactive species for a given reaction pathway). Crucially, the global crosstalk tube ensemble omits the cognate products that the reactive species are intended to form (they appear as neither on-targets nor off-targets). Hence, all reactive species in the global crosstalk tube are forced to either perform no reaction (remaining as desired on-targets) or undergo a crosstalk reaction (forming undesired off-targets), providing the basis for minimization of global crosstalk during sequence optimization. To design 8 orthogonal systems for this reaction pathway, the total number of target test tubes is then $|\Omega| = 3 \times 8 + 1 = 25$.

These computational test tube ensembles have two conceptually interesting and practically significant properties. First, each target test tube isolates a different subset of the system components in local equilibrium, enabling optimization of kinetically significant states that would appear insignificant if all components were allowed to interact in a single ensemble. For example, the Step 1 tube of Figure 2a simultaneously optimizes for high-yield production of unstructured intermediate B and against appreciable formation of off-target dimer B·B, promoting rapid nucleation of the unstructured toehold in B with the loop of hairpin C during the next step of the reaction pathway. Second, for a tube containing a given set of system components, the cognate products of their interactions can be excluded from the ensemble (appearing as neither on-targets nor off-targets), enabling optimization for high-yield well-structured reactants and against crosstalk. For example, the Reactants tube of Figure 2a excludes the cognate product of Step 1 from the ensemble in order to optimize formation of initial reactants X, A·B, and C and discourage competing crosstalk interactions (e.g., X·X, A·A, X·C).

Design Objective Function. To provide a physically meaningful objective function for optimizing the equilibrium base-pairing properties of a single test tube of interacting nucleic acid strands, we previously derived the *test tube ensemble defect*,[17] C_h, quantifying the equilibrium concentration of incorrectly paired nucleotides over the ensemble of test tube h. Let

$$\mathcal{M}_h \equiv C_h / y_h^{\text{nt}} \in (0, 1) \tag{1}$$

denote the equilibrium fraction of incorrectly paired nucleotides in tube h. Here,

$$y_h^{\text{nt}} \equiv \sum_{j \in \Psi_h^{\text{on}}} |\phi_j| y_{h,j}$$

is the total concentration of nucleotides in tube h, where ϕ_j denotes the sequence of complex j. As $\mathcal{M}_h$ approaches zero, each on-target complex, $j \in \Psi_h^{\text{on}}$, approaches its target concentration, $y_{h,j}$, and is dominated by its target structure, s_j, and each off-target complex, $j \in \Psi_h^{\text{off}}$, forms with vanishing target concentration.

DOI: 10.1021/jacs.6b12693
J. Am. Chem. Soc. 2017, 139, 3134−3144

Table 1. Sequence Constraints

constraint type			constraint relation*	nucleotides
assignment	(ϕ^a)	$\in R_a^{\text{assignment}}$	$\equiv \{(q^1)\}$	1
match	(ϕ^a, ϕ^b)	$\in R_{a,b}^{\text{match}}$	$\equiv \{$(A,A), (C,C), (G,G), (U,U)$\}$	2
Watson−Crick	(ϕ^a, ϕ^b)	$\in R_{a,b}^{\text{WC}}$	$\equiv \{$(A,U), (C,G), (G,C), (U,A)$\}$	2
complementarity	(ϕ^a, ϕ^b)	$\in R_{a,b}^{\text{complement}}$	$\equiv \{$(A,U), (C,G), (G,C), (U,A), (G,U), (U,G)$\}$	2
composition	$(\phi^a, ..., \phi^b)$	$\in R_{a,...,b}^{\text{composition}}$	$\equiv \{(\phi^a, ..., \phi^b) \mid f^{\min} \le 1 - \sum_{i=a,...,b} d(\phi^i, q^1)/n \le f^{\max}\}$	$b - a + 1 = n$
similarity	$(\phi^a, ..., \phi^b)$	$\in R_{a,...,b}^{\text{similarity}}$	$\equiv \{(\phi^a, ..., \phi^b) \mid f^{\min} \le 1 - d((\phi^a, ..., \phi^b), (q^1, ..., q^n))/n \le f^{\max}\}$	$b - a + 1 = n$
pattern prevention	$(\phi^a, ..., \phi^b)$	$\in R_{a,...,b}^{\text{pattern}}$	$\equiv \{(\phi^a, ..., \phi^b) \mid (q^1, ..., q^n)$ is not a subsequence of $(\phi^a, ..., \phi^b)\}$	$b - a + 1 \ge n$
library	$(\phi^a, ..., \phi^b)$	$\in R_{a,...,b}^{\text{library}}$	$\equiv \{(q_1^1, ..., q_1^n), ..., (q_m^1, ..., q_m^n)\}$	$b - a + 1 = n$
window	$(\phi^a, ..., \phi^b)$	$\in R_{a,...,b}^{\text{window}}$	$\equiv \{(\phi^a, ..., \phi^b) \mid (\phi^a, ..., \phi^b)$ is a subsequence of $(q^1, ..., q^n)\}$	$b - a + 1 \le n$

*For user-specified $q^i \in$ {A,C,G,U,M,R,W,S,Y,K,V,H,D,B,N}.

Generalizing to the multistate test tube ensemble, the *multistate test tube ensemble defect*,

$$\mathcal{M} \equiv \frac{1}{|\Omega|} \sum_{h \in \Omega} \mathcal{M}_h \quad \in (0, 1) \tag{2}$$

quantifies the average equilibrium fraction of incorrectly paired nucleotides over the test tubes $h \in \Omega$. The goal is to design a set of sequences such that the multistate test tube ensemble defect, $\mathcal{M}$, satisfies the *stop condition*,

$$\mathcal{M} \le f_{\text{stop}} \tag{3}$$

for a user-specified value of $f_{\text{stop}} \in (0,1)$.

In some cases, the user may wish to alter the relative weighting of defect contributions within $\mathcal{M}$ to prioritize or deprioritize design quality for a portion of the design ensemble. To provide flexibility, we permit the user to define custom defect weights for the contribution of each nucleotide, complex, and test tube to $\mathcal{M}$ (see SI, section S1.6). With the default value of unity for all weights, this objective function is simply the multistate test tube ensemble defect (2). With custom weights, the physical meaning of the objective function is distorted in the service of adjusting design priorities.

Sequence Constraints. A nucleic acid reaction pathway imposes sequence constraints on its reactants (e.g., the complementary sequence domains 'a' and 'a*' in the reaction pathway of Figure 1). Furthermore, a molecular engineer may wish to impose a variety of additional sequence constraints (e.g., constraining GC content to optimize synthesis, or constraining input X, comprising sequence domains 'a-b-c' in Figure 1, to be a subsequence of a particular mRNA).

We provide a unified and extensible framework for imposing diverse types of sequence constraints on the design problem:

- *Assignment Constraint.* Nucleotide a is constrained to have a specified sequence (e.g., A, C, G, U or any of the IUPAC degenerate nucleotide codes; see Table S1).
- *Match Constraint.* Two nucleotides a and b are constrained to be identical (e.g., if a strand species appears in more than one on-target complex, corresponding nucleotides are constrained to have the same sequence in all complexes).
- *Watson−Crick Constraint.* Two nucleotides a and b are constrained to be Watson−Crick complements (by default, Watson−Crick constraints are implied for all base pairs present in on-target structures).
- *Complementarity Constraint.* Two nucleotides a and b are constrained to be Watson−Crick or wobble complements.
- *Composition Constraint.* Consecutive nucleotides a, ..., b are constrained to have a sequence composition in a specified range (e.g., a desired GC content can be achieved by constraining the fraction of S nucleotides to fall in the range $[f^{\min}, f^{\max}]$).
- *Similarity Constraint.* Consecutive nucleotides a, ..., b are constrained to be similar to a specified sequence of length $n = b - a + 1$ to a specified degree (e.g., the fraction of nucleotides matching an mRNA sequence can be constrained to fall in the range $[f^{\min}, f^{\max}]$).
- *Pattern Prevention Constraint.* Consecutive nucleotides a, ..., b are constrained not to contain a specified subsequence of length $n \le b - a + 1$ (e.g., prevention of GGGG, which is prone to forming G-quadruplexes[18] that are not accounted for in nearest-neighbor free energy models[19,20]).
- *Library Constraint.* Consecutive nucleotides a, ..., b are constrained to be selected from a specified library of m sequences of length $n = b - a + 1$ (e.g., a library of toehold sequences or a library of codons).
- *Window Constraint.* Consecutive nucleotides a, ..., b are constrained to be a subsequence of a specified source sequence of length $n \ge b - a + 1$ (e.g., the source sequence is an mRNA), or more generally, a subsequence of one of multiple specified source sequences.

Within this framework, each constraint is expressed as a constraint relation (Table 1). For some constraint relations, it is convenient to make use of the sequence distance function,

$$d(\phi, q) \equiv \sum_{a \in 1,...,|\phi|} \begin{cases} 0: \phi^a \in q^a \\ 1: \phi^a \notin q^a \end{cases}$$

between sequence ϕ and the constraint sequence q of equal length, which may contain degenerate IUPAC nucleotide codes (see Table S1). For example, d(ACGU, SSWW) = 2. Additional types of sequence constraints can be supported by specifying new constraint relations.

Constrained Multistate Test Tube Design Problem. Let $\phi_\Psi \equiv \phi_j \ \forall j \in \Psi$ denote the set of sequences for the complexes in Ψ and let $\mathcal{R}$ denote the set of user-specified sequence constraints. To design a set of sequences, ϕ_Ψ, for a given a nucleic acid reaction pathway, we specify on-target and off-target complexes within the set of target test tubes, Ω, to represent on-pathway elementary steps and off-pathway crosstalk. The constrained multistate test tube design problem is then:

$$\min_{\phi_\Psi} \mathcal{M} \text{ subject to } \mathcal{R} \tag{4}$$

DOI: 10.1021/jacs.6b12693
J. Am. Chem. Soc. 2017, 139, 3134−3144

where $\mathcal{M}$ is the multistate test tube ensemble defect over Ω. The sequence design algorithm seeks to iteratively reduce $\mathcal{M}$ while satisfying the constraints in $\mathcal{R}$, terminating sequence optimization upon satisfaction of the stop condition (3).

METHODS

Algorithm Overview. Our constrained multistate test tube design algorithm generalizes the test tube design algorithm of Wolfe and Pierce[17] to perform sequence design over an ensemble of an arbitrary number of target test tubes subject to diverse user-specified sequence constraints. The underlying physical model is based on nucleic acid secondary structure and nearest-neighbor free energy parameters (SI, section S1.1). The objective function, $\mathcal{M}$, is reduced via iterative mutation of a random initial sequence. Because of the high computational cost of calculating the objective function (SI, section S1.2), it is important to avoid direct recalculation of $\mathcal{M}$ in evaluating each candidate mutation. We exploit three concepts to enable efficient calculation of the objective function estimate, $\tilde{\mathcal{M}}$: using *test tube ensemble focusing*, sequence optimization initially focuses on only the on-target portion of each test tube ensemble (SI, section S1.3); using *hierarchical ensemble decomposition*, the structural ensemble of each on-target complex is hierarchically decomposed into a tree of conditional subensembles, yielding a forest of decomposition trees (SI, section S1.4); by calculating *conditional physical properties* over the conditional structural ensembles at any level within the decomposition forest, it is possible to efficiently estimate the equilibrium base-pairing properties of the multistate test tube ensemble (SI, section S1.5). Optional *defect weights* enable the user to adjust design priorities within this ensemble (SI, section S1.6). To minimize computational cost, candidate mutations are evaluated at the leaf level of the decomposition forest (SI, section S1.7). As optimized subsequences are merged toward the root level of the forest, any emergent defects are eliminated via ensemble redecomposition from the parent level on down and sequence reoptimization from the leaf level on up (SI, section S1.8). After subsequences are successfully merged to the root level, the exact objective function, $\mathcal{M}$, is calculated for the first time, explicitly checking for the effect of the previously neglected off-target complexes. Any off-target complexes observed to form at appreciable concentration are hierarchically decomposed, added to the decomposition forest, and actively destabilized during subsequent forest reoptimization (SI, section S1.9). When decomposition or focusing defects are encountered, hierarchical ensemble decomposition is performed using multiple exclusive split-points (SI, section S1.10). Throughout the sequence optimization process, whenever the sequence is initialized, mutated, or reseeded, we solve a *constraint satisfaction problem* to obtain valid sequences satisfying all constraints in $\mathcal{R}$ (SI, section S1.11). The algorithm flow is detailed in the pseudocode of Algorithm S1 (SI).

Implementation. The constrained multistate test tube design algorithm is coded in the C and C++ programming languages. The algorithm is available for noncommercial academic use as part of the NUPACK web application and source code (www.nupack.org).[21]

Sequence Design Trials. For each design problem, 30 independent design trials were performed. Design trials were run on a cluster of 2.53 GHz Intel E5540 Xeon dual-processor/quad-core nodes with 24 GB of memory per node. Each trial was run on one computational core using the default algorithm parameters of Table S2. Design quality is quantified by the multistate test tube ensemble defect $\mathcal{M}$. To design N orthogonal systems, all nucleotide, complex, and tube weights are left at the default value of 1 except for the global crosstalk tube, which is assigned a weight of N to prevent the effect of crosstalk from being diluted in the design objective function as the number of orthogonal systems increases. Data are typically plotted[22] as cumulative histograms over design trials. Relative design cost is quantified by dividing the cost of sequence design (cost_{des}) by the cost of a single evaluation of the multistate test tube ensemble defect ($\text{cost}_{\text{eval}}$). Designs are performed in 1 M Na^+ at 37 °C for RNA (conditional Dicer substrate formation case study) and in 1 M Na^+ at 25 °C for DNA (all other case studies). For each design trial, the stop condition is f_{stop} = 0.02 (i.e., no more than 2% of nucleotides incorrectly paired at equilibrium over the multistate test tube ensemble, Ω).

RESULTS

Reaction Pathway Engineering Case Studies. To examine algorithm performance, we consider a selection of reaction pathways from the molecular programming literature (SI, section S2.1):

- *Conditional self-assembly via hybridization chain reaction (HCR).* A single-stranded input X triggers self-assembly of metastable hairpins into a nicked double-stranded polymer.[4]
- *Boolean logic AND using toehold sequestration gates.* Detection of two independent single-stranded inputs X and Y triggers release of a single-stranded output.[5]
- *Self-assembly of a three-arm junction via catalytic hairpin assembly (CHA).* A single-stranded input X catalyzes self-assembly of a three-arm branched junction from metastable hairpins.[8]
- *Boolean logic AND using a cooperative hybridization gate.* Two independent single-stranded inputs X and Y cooperatively displace a single-stranded output.[16]
- *Conditional Dicer substrate formation via shape and sequence transduction.* Detection of a single-stranded input X leads to formation of a double-stranded Dicer substrate targeting an independent output sequence Y for silencing.[12]

For each reaction pathway, we define a set of target test tubes specifying 1, 2, 4, or 8 orthogonal systems intended to operate independently in the same sample (SI, section S2.2). These constrained multistate test tube design problems involve up to dozens of test tubes, dozens of on-target complexes, and thousands of off-target complexes (Table 2).

Table 2. Reaction Pathway Engineering Case Studies

reaction pathway	orthogonal systems	tubes $\lvert\Omega\rvert$	on-targets $\lvert\Psi^{\text{on}}\rvert$	off-targets $\lvert\Psi^{\text{off}}\rvert$
conditional self-assembly via hybridization chain reaction (HCR)	1	5	7	9
	8	33	56	520
Boolean logic AND using toehold sequestration gates	1	5	14	52
	8	33	112	3216
self-assembly of a three-arm junction via catalytic hairpin assembly (CHA)	1	6	11	27
	8	41	88	1588
Boolean logic AND using a cooperative hybridization gate	1	3	9	51
	8	17	72	2676
conditional Dicer substrate formation via shape and sequence transduction	1	4	9	26
	8	25	72	1580

Algorithm Performance for Constrained Multistate Test Tube Design. Figure 3 demonstrates the performance of the constrained multistate test tube design algorithm for the target test tubes of Figure 2, corresponding to conditional Dicer substrate formation via shape and sequence transduction with scRNAs. For each target test tube, the algorithm designs for the depicted on-target complexes (each with a target secondary

DOI: 10.1021/jacs.6b12693
J. Am. Chem. Soc. 2017, 139, 3134−3144

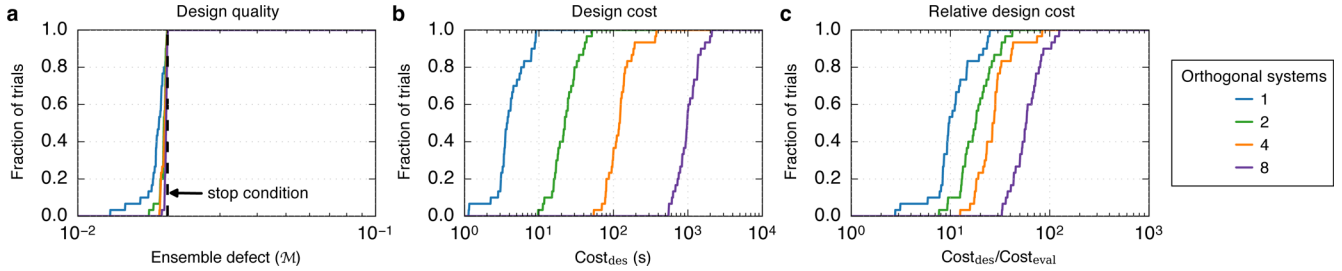


Figure 3. Algorithm performance for simultaneous design of N = 1, 2, 4, or 8 orthogonal systems using the target test tubes of Figure 2. (a) Design quality. The stop condition is depicted as a dashed black line. (b) Design cost. (c) Cost of sequence design relative to a single evaluation of the objective function. Case study: conditional Dicer substrate formation via shape and sequence transduction with scRNAs. See Figure S11 for comparison of all case studies.

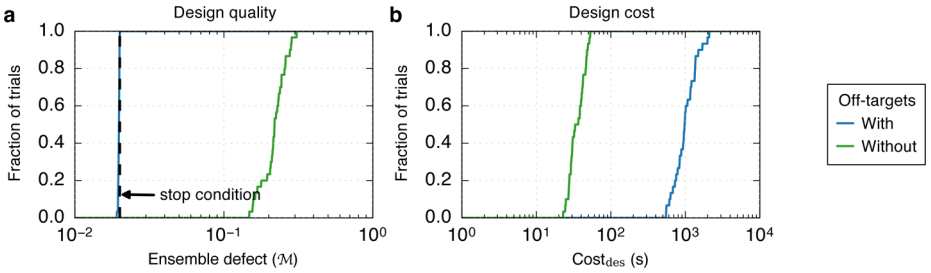


Figure 4. Importance of negative design in reducing crosstalk. Comparison of designs performed with or without off-targets in the design ensemble. (a) Design quality evaluated by calculating the multistate test tube ensemble defect ($\mathcal{M}$) over the ensemble containing off-targets. The stop condition is depicted as a dashed black line. (b) Design cost. Case study: conditional Dicer substrate formation via shape and sequence transduction with scRNAs (N = 8 orthogonal systems). See Figure S18 for comparison of all case studies.

structure and target concentration) and against all off-target complexes. The size of the design ensemble ranges from 4 tubes containing a total of 9 on-target and 26 off-target complexes for design of 1 system to 25 tubes containing a total of 72 on-target and 1580 off-target complexes for design of 8 orthogonal systems (Table 2). Sequences are designed subject only to implicit sequence constraints inherent to the sequence domain specification for each strand (e.g., match constraints for domain "a" appearing in multiple complexes, Watson−Crick constraints for complementary domains "a" and "a*"). Typical design trials achieve the desired design quality (stop condition $\mathcal{M} \leq 0.02$; panel a) and typical design costs range from seconds for the design of 1 system to minutes for simultaneous design of 8 orthogonal systems (panel b). The typical cost of design relative to the cost of analysis ranges from approximately 10 to 60 as the number of orthogonal systems increases from 1 to 8 (panel c). These rising relative design costs reflect the increasing challenge of designing against crosstalk as the number of orthogonal systems increases.

For the five engineering case studies of Table 2, typical design trials achieve the 2% stop condition (Figure S11). For simultaneous design of 8 orthogonal systems, the typical cost of design relative to the cost of analysis ranges from a factor of 60 to 1300. If desired, the design cost can be reduced by relaxing the design quality requirements; using f_{stop} = 0.05 instead of 0.02, the desired design quality is typically achieved with a relative design cost ranging from a factor of 30 to 300 (Figure S12).

At the conclusion of sequence design, the distribution of residual defects across target test tubes and across complexes within each target test tube depends on the idiosyncrasies of each case study (see SI, section S2.4), with the more challenging tubes and complexes retaining higher defects. If desired, nucleotide, complex, or tube defect weights can be adjusted from their default values of unity to prioritize or deprioritize reduction of the corresponding defect contributions to the design objective function, $\mathcal{M}$.

Importance of Negative Design in Reducing Crosstalk. The target test tubes summarized in Table 2 and detailed in SI, section S2.2 contain both on-target and off-target complexes, implementing both a positive design paradigm (designing for the target concentration of on-target complexes) and a negative design paradigm (designing against the formation of off-target complexes). Is it important to include off-targets in the design ensemble and explicitly destabilize these off-pathway interactions in order to arrive at sequence designs with low crosstalk? To examine this question, Figure 4 re-examines the simultaneous design of 8 orthogonal systems using either the full design ensemble (a total of 72 on-target and 1580 off-target complexes) or a reduced design ensemble that omits all off-target complexes, evaluating the quality of the resulting designs over the full ensemble. If the design ensemble contains no off-targets, the typical ensemble defect of the final sequence designs increases by an order of magnitude (from 2% to over 20%), emphasizing the importance of explicitly destabilizing off-targets. The other engineering case studies similarly emphasize the importance of negative design in reducing crosstalk (Figure S18).

Effect of Sequence Constraints. Figure 5 illustrates the effects of imposing explicit sequence constraints (composition: constraining GC content, pattern: preventing 4-nt stretches of any one nucleotide and 6-nt stretches of any two nucleotides, window: constraining input X and output Y to be subsequences of different mRNAs) on the design of a single system. Using any of these constraint types alone, typical design trials satisfy the 2% stop condition (panel a). For composition and pattern constraints, the cost of design is approximately 1 order of magnitude higher than the cost of analysis (panel c), while

DOI: 10.1021/jacs.6b12693
J. Am. Chem. Soc. 2017, 139, 3134−3144

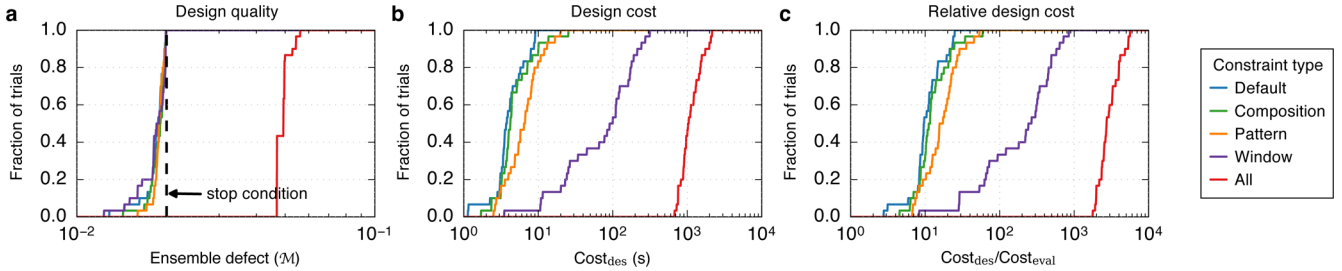


Figure 5. Algorithm performance including explicit sequence constraints. Default: implicit sequence constraints inherent to the reaction pathway (these constraints are also present in the other cases that follow). Composition constraint: fraction of S ∈ [0.45, 0.55]. Pattern constraint: prevent {AAAA, CCCC, GGGG, UUUU, KKKKKK, MMMMMM, RRRRRR, SSSSSS, WWWWWW, YYYYYY}. Window constraints: input X and output Y constrained to be subsequences of two different mRNAs (i.e., biological sequence constraints; see SI, section S2.6.1). All: all of the above constraints. (a) Design quality. The stop condition is depicted as a dashed black line. (b) Design cost. (c) Cost of sequence design relative to a single evaluation of the objective function. Case study: conditional Dicer substrate formation via shape and sequence transduction with scRNAs (N = 1 system). See Figure S19 for comparison of all case studies.

using window constraints to impose biological compatibility, the relative design cost increases to 2 orders of magnitude. Imposing all three constraint types simultaneously, the algorithm is typically unable to reduce the multistate ensemble defect below 5%, and the relative design cost increases to 3 orders of magnitude. It is noteworthy that imposing biological sequence constraints dramatically reduces the size of the design space: for the 39 nucleotides constrained by mRNAs X and Y, the number of feasible candidate sequences decreases from 3×10^{23} to 4×10^{6}; additionally imposing composition and pattern constraints decreases this number to 3×10^{4} (see SI, section S2.6.1). This dramatic reduction in sequence space increases both the challenge of jumping between feasible candidate sequences during the search process and the challenge of achieving a low multistate test tube ensemble defect for the final design. Sequence-constrained versions of the other engineering case studies reveal similar trends (Figure S19), illustrating that sequence constraints increase the degree of difficulty both in terms of design quality and design cost, and emphasizing the desirability of constraining the design space only to the extent necessary.

Robustness of Predictions to Model Perturbations. As the empirical parameter sets underlying nucleic acid secondary structure models[19,20,23−25] are further refined going forward, it is important that the predicted design quality is robust to perturbations in the parameters. Figure 6 demonstrates that the predicted quality is typically robust to 3% parameter perturbations (with typical ensemble defect comparable to the 2% stop condition), and even to 10% parameter perturbations (with the typical defect increasing to only 3%), but not to 30% parameter perturbations (with the typical defect increasing to 14%). Qualitatively similar trends are seen for the other case studies (Figure S20).

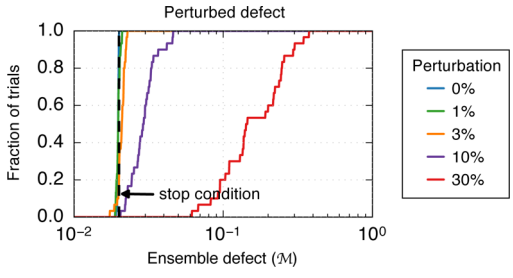


Figure 6. Robustness of design quality assessments to perturbations in model parameters. For each design trial, the median multistate test tube ensemble defect was calculated over 100 perturbed physical models (each parameter perturbed by Gaussian noise with a standard deviation of 0, 1, 3, 10, or 30% of the parameter modulus). The stop condition is depicted as a dashed black line. Case study: conditional Dicer substrate formation via shape and sequence transduction with scRNAs (N = 8 orthogonal systems). See Figure S20 for comparison of all case studies.

DISCUSSION

To enable sequence design for reaction pathway engineering, the *multistate test tube design* problem builds on three subsidiary design problems: *complex design*, *multistate complex design*, and *test tube design*. For each subsidiary design problem, we describe the design ensemble, define an ensemble defect that provides a physically meaningful objective function for optimizing sequence quality over this ensemble, and examine the extent to which the resulting formulation implements a positive design paradigm (stabilize on-targets) and a negative design paradigm (destabilize off-targets). These comparisons illuminate the conceptual properties of the current formulation and the manner in which it extends existing design capabilities.

Complex Design. For complex design, the goal is to design the equilibrium base pairing properties of a complex of (one or more) interacting nucleic acid strands. This subsidiary design problem is the foundation on which the other three design problems build, and has attracted the most algorithm development to date.[26−43]

Design Ensemble. For complex design, the user specifies an on-target complex, j, with on-target secondary structure, s_j. We may view complex design as a special case of multistate test tube design where the design ensemble comprises a single target test tube containing a single on-target complex and no off-target complexes (Table 3).

Design Objective Function. The *complex ensemble defect*,

Table 3. Nucleic Acid Sequence Design Ensembles

design problem	test tubes $\lvert\Omega\rvert$	per test tube $h \in \Omega$: on-target complexes $\lvert\Psi_h^{on}\rvert$	per test tube $h \in \Omega$: off-target complexes $\lvert\Psi_h^{off}\rvert$
complex design	1	1	0
multistate complex design	arbitrary	1	0
test tube design	1	arbitrary	arbitrary
multistate test tube design	arbitrary	arbitrary	arbitrary

DOI: 10.1021/jacs.6b12693
J. Am. Chem. Soc. 2017, 139, 3134−3144

$$n(\phi_j, s_j) = |\phi_j| - \sum_{\substack{1 \le a \le |\phi_j| \\ 1 \le b \le |\phi_j|+1}} P^{a,b}(\phi_j) S^{a,b}(s_j)$$

quantifies the equilibrium number of incorrectly paired nucleotides over the ensemble of complex j.[29,36] Here, $P(\phi_j)$ is the equilibrium base-pairing probability matrix and $S(s_j)$ is the target structure matrix for complex j (see SI, section S1.2). Let

$$\mathcal{N}_j \equiv n(\phi_j, s_j)/|\phi_j| \quad \in (0, 1)$$

denote the equilibrium fraction of incorrectly paired nucleotides in the ensemble of complex j. The goal is to design a sequence, ϕ_j, that satisfies the stop condition,

$$\mathcal{N}_j \le f_{\text{stop}}$$

for a user-specified value of $f_{\text{stop}} \in (0,1)$.

Design Paradigms. Optimization of the complex ensemble defect implements a positive design paradigm (designing for the on-target structure) and a negative design paradigm (designing against all off-target structures) within the ensemble of the on-target complex.[29,36] Both paradigms are crucial to achieve high-quality sequence designs with low complex ensemble defect.[29,36] Because the complex design ensemble is equivalent to a target test tube containing a single on-target complex and no off-target complexes, complex design implements only a positive design paradigm (designing for the on-target complex) within the ensemble of the target test tube (Table 4).

Table 4. Nucleic Acid Sequence Design Paradigms

	paradigms within complex		paradigms within test tube		
ensemble defect	positive	negative	positive	negative	states
complex	√	√	√		1
multistate complex	√	√	√		arbitrary
test tube	√	√	√	√	1
multistate test tube	√	√	√	√	arbitrary

Multistate Complex Design. The complex design ensemble comprises a single on-target complex, and hence does not address the multistate challenges inherent in reaction pathway engineering. To address this need, multistate complex design expands the design ensemble to include multiple on-target complexes.[44] For multistate complex design, the goal is to engineer the equilibrium base pairing properties of an arbitrary number of on-target complexes, each representing a reactant, intermediate, or product state along the reaction pathway.

Design Ensemble. For multistate complex design, the user specifies a set of on-target complexes, Ψ. For each on-target complex, $j \in \Psi$, the user specifies a target secondary structure, s_j.[44] We may view multistate complex design as a special case of multistate test tube design where the design ensemble contains an arbitrary number of target test tubes, each containing a single on-target complex and no off-target complexes (Table 3).

Design Objective Function. The *multistate complex ensemble defect*,

$$\mathcal{N} \equiv \frac{1}{|\Psi|} \sum_{j \in \Psi} \mathcal{N}_j \quad \in (0, 1)$$

quantifies the average equilibrium fraction of incorrectly paired nucleotides over the complexes $j \in \Psi$. The goal is to design a set of sequences, ϕ_Ψ, that satisfy the stop condition

$$\mathcal{N} \le f_{\text{stop}}$$

for a user-specified value of $f_{\text{stop}} \in (0,1)$.

Design Paradigms. Multistate complex design inherits the benefits and shortcomings of complex design for each reactant, intermediate, or product state along the reaction pathway, implementing positive and negative design paradigms within the ensemble of each on-target complex, but only a positive design paradigm within the ensemble of each target test tube (Table 4).

Test Tube Design. With complex design and multistate complex design, neither the concentration of the desired on-target complex(es), nor the concentrations of undesired off-target complexes are considered. As a result, sequences that are successfully optimized to stabilize a target secondary structure in the context of an on-target complex, may nonetheless fail to ensure that this complex forms at appreciable concentration when the strands are introduced into a test tube.[17] To address this critical shortcoming, test tube design expands the complex design ensemble to include off-target complexes.[17] For test tube design, the goal is to engineer the equilibrium base-pairing properties of a test tube of interacting nucleic acid strands.

Design Ensemble. For test tube design,[17] the user specifies a target test tube h containing a set of on-target complexes, Ψ_h^{on}, and a set of off-target complexes, Ψ_h^{off}. For each on-target complex, $j \in \Psi_h^{\text{on}}$, the user specifies a target secondary structure, s_j, and a target concentration, $y_{h,j}$. For each off-target complex, $j \in \Psi_h^{\text{off}}$, the target concentration is vanishing ($y_{h,j} = 0$) and there is no target structure (s_j=∅). We may view test tube design as a special case of multistate test tube design where the design ensemble contains a single target test tube containing arbitrary numbers of on- and off-target complexes (Table 3).

Design Objective Function. The *test tube ensemble defect*,

$$C(\phi_{\Psi_h}, s_{\Psi_h}, y_{h,\Psi_h}) = \sum_{j \in \Psi_h^{\text{on}}} [n(\phi_j, s_j)\min(x_{h,j}, y_{h,j}) + |\phi_j|\max(y_{h,j} - x_{h,j}, 0)] \tag{5}$$

quantifies the equilibrium concentration of incorrectly paired nucleotides over the ensemble of test tube h.[17] Here, $x_{h,j}$ is the equilibrium concentration of complex j in tube h (see SI, section S1.2). For each on-target complex, $j \in \Psi_h^{\text{on}}$, the first term in the sum represents the *structural defect*, quantifying the concentration of nucleotides that are in an incorrect base-pairing state within the ensemble of complex j, and the second term in the sum represents the *concentration defect*, quantifying the concentration of nucleotides that are in an incorrect base-pairing state because there is a deficiency in the concentration of complex j. For each off-target complex, $j \in \Psi_h^{\text{off}}$, the structural and concentration defects are identically zero, since $y_{h,j} = 0$. This does not mean that the defects associated with off-targets are ignored. By conservation of mass, nonzero off-target concentrations imply deficiencies in on-target concentrations, and these concentration defects are quantified by (5).[17] Using (1) to define the normalized test tube ensemble defect, $\mathcal{M}_h$, in terms of $C_h \equiv C(\phi_{\Psi h}, s_{\Psi h}, y_{h,\Psi_h})$, the goal is to design a set of sequences, $\phi_{\Psi h}$, that satisfy the stop condition,

$$\mathcal{M}_h \le f_{\text{stop}}$$

DOI: 10.1021/jacs.6b12693
J. Am. Chem. Soc. 2017, 139, 3134−3144

for a user-specified value of $f_{stop} \in (0,1)$.

Design Paradigms. Optimization of the test tube ensemble defect implements a positive design paradigm and a negative design paradigm at two levels (Table 4): 1) within the ensemble of each on-target complex (designing for the on-target structure and against all off-target structures),[29,36] and 2) within the ensemble of the target test tube (designing for the target concentration of each on-target complex and against the formation of all off-target complexes).[17] Both paradigms are crucial at both levels in order to achieve high-quality sequence designs with low test tube ensemble defect.[17,29,36]

Multistate Test Tube Design. The present work extends the conceptual benefits of test tube design to address the multistate demands of reaction pathway engineering.

Design Ensemble. The multistate test tube design ensemble generalizes each of the three subsidiary design ensembles, encompassing an arbitrary number of target test tubes, each containing arbitrary numbers of on- and off-target complexes (Table 3).

Design Objective Function. Likewise, the multistate test tube ensemble defect, $\mathcal{M}$, generalizes each of the three subsidiary ensemble defects. When the design ensemble comprises a single target test tube containing a single on-target complex and no off-target complexes, $\mathcal{M}$ reduces to $\mathcal{N}_j$, the normalized complex ensemble defect.[29,36] When the design ensemble comprises multiple target test tubes, each containing a single on-target complex and no off-target complexes, $\mathcal{M}$ reduces to $\mathcal{N}$, the multistate complex ensemble defect. When the design ensemble comprises a single target test tube containing arbitrary numbers of on- and off-target complexes, $\mathcal{M}$ reduces to $\mathcal{M}_h$, the normalized test tube ensemble defect.[17]

Design Paradigms. Multistate test tube design inherits the conceptual benefits of test tube design for each target test tube (reactant, intermediate, product, and global crosstalk), implementing positive and negative design paradigms both within the ensemble of each on-target complex and within the ensemble of each target test tube (Table 4).

CONCLUSIONS

Summary. Constrained multistate test tube design enables sequence design for nucleic acid reaction pathway engineering. The design ensemble uses a set of target test tubes to represent reactant, intermediate, and product states along the reaction pathway, as well as to model crosstalk between components. Each target test tube contains a set of desired on-target complexes (each with a target secondary structure and target concentration) and a set of undesired off-target complexes (each with vanishing target concentration). Sequence quality is quantified by the ensemble defect, representing the average equilibrium fraction of incorrectly paired nucleotides over the test tubes in the ensemble. Sequence optimization is performed by reducing the ensemble defect subject to user-specified sequence constraints. These sequence constraints can dramatically reduce the size of the design space (e.g., by constraining a sequence domain to be a subsequence of an mRNA). To effectively navigate the available sequence space, each valid candidate sequence is generated by solving a constraint satisfaction problem. The candidate sequence is then accepted or rejected based on an estimate of the ensemble defect calculated efficiently via hierarchical ensemble decomposition.[17]

Design Paradigms. Optimization of the ensemble defect implements a positive design paradigm and a negative design paradigm to optimize for on-pathway interactions and against off-pathway interactions at two levels: within the ensemble of each on-target complex (designing for the on-target structure and against all off-target structures),[29,36] and within the ensemble of each target test tube (designing for the target concentration of each on-target complex and against the formation of all off-target complexes).[17] Both paradigms are critical at both levels in order to robustly optimize ensemble properties over the design ensemble.[17,29,36]

NUPACK Software Suite. This work unifies and generalizes the complex design,[36] multistate complex design,[44] test tube design,[17] and (prepublication) multistate test tube design tools provided by the NUPACK software suite (nupack.org),[21] which have been used to engineer diverse nucleic acid structures, devices, and systems.[4,6,8,12,45−72]

Toward a Compiler for Molecular Programming. By enabling systematic reaction pathway engineering, constrained multistate test tube design takes an important step toward our long-term goal of developing a compiler for programming molecular function. Such a compiler would accept as input a reaction pathway specification and produce as output a set of nucleic acid sequences that implement the desired function.

Future Work. To achieve the goal of molecular compilation, additional work is needed to automate the specification of target test tubes given a desired reaction pathway. To engineer synthetic systems that operate in a biological context, where most of the nucleotides are native to a host organism rather than designed by the engineer, it may prove important to actively design against crosstalk with some or all of the native transcriptome, potentially requiring further advances in the sequence design formulation and algorithm. In situations where researchers are designing nucleic acid components intended to interact with proteins or small molecules, it may be desirable to incorporate user-specified energetic penalties or rewards into the design objective function in order to approximate molecular interactions that fall outside the scope of nucleic acid secondary structure models.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.6b12693.

Algorithm details, pseudocode, and default algorithm parameters; engineering case studies, including reaction pathways, specification of target test tubes, algorithm performance, residual defects, importance of negative design in reducing crosstalk, effect of sequence constraints, and robustness of predictions to model perturbations; and additional design studies, including performance for test tube design and performance for complex design; Figures S1−S24 and Tables S1−S3 (PDF)

AUTHOR INFORMATION

Corresponding Author

*niles@caltech.edu

ORCID

Niles A. Pierce: 0000-0003-2367-4406

Author Contributions

#B.R.W. and N.J.P. contributed equally.

DOI: 10.1021/jacs.6b12693
J. Am. Chem. Soc. 2017, 139, 3134−3144

Notes

The authors declare the following competing financial interest(s): US patents and pending patent applications.
[⊥]Deceased.

■ ACKNOWLEDGMENTS

This work was funded by the National Science Foundation via the Molecular Programming Project (NSF-CCF-0832824 and NSF-CCF-1317694), by the Gordon and Betty Moore Foundation (GBMF2809), by the Beckman Institute at Caltech (PMTC), by a Christensen Fellowship at St Catherine's College, University of Oxford, by the John Simon Guggenheim Memorial Foundation, by a Professorial Fellowship at Balliol College, University of Oxford, and by the Eastman Visiting Professorship at the University of Oxford.

■ REFERENCES

(1) Bath, J.; Turberfield, A. J. *Nat. Nanotechnol.* **2007**, *2*, 275−284.
(2) Zhang, D. Y.; Seelig, G. *Nat. Chem.* **2011**, *3*, 103−113.
(3) Yurke, B.; Turberfield, A. J.; Mills, A. P., Jr.; Simmel, F. C.; Neumann, J. L. *Nature* **2000**, *406*, 605−608.
(4) Dirks, R. M.; Pierce, N. A. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 15275−15278.
(5) Seelig, G.; Soloveichik, D.; Zhang, D. Y.; Winfree, E. *Science* **2006**, *314*, 1585−1588.
(6) Venkataraman, S.; Dirks, R. M.; Rothemund, P. W. K.; Winfree, E.; Pierce, N. A. *Nat. Nanotechnol.* **2007**, *2*, 490−494.
(7) Zhang, D. Y.; Turberfield, A. J.; Yurke, B.; Winfree, E. *Science* **2007**, *318*, 1121−1125.
(8) Yin, P.; Choi, H. M. T.; Calvert, C. R.; Pierce, N. A. *Nature* **2008**, *451*, 318−322.
(9) Zhang, D. Y.; Winfree, E. *J. Am. Chem. Soc.* **2009**, *131*, 17303−17314.
(10) Qian, L.; Winfree, E. *Science* **2011**, *332*, 1196−1201.
(11) Turberfield, A. J.; Mitchell, J. C.; Yurke, B.; Mills, A. P., Jr.; Blakey, M. I.; Simmel, F. C. *Phys. Rev. Lett.* **2003**, *90*, 118102.
(12) Hochrein, L. M.; Schwarzkopf, M.; Shahgholi, M.; Yin, P.; Pierce, N. A. *J. Am. Chem. Soc.* **2013**, *135*, 17322−17330.
(13) Bois, J. S.; Venkataraman, S.; Choi, H. M. T.; Spakowitz, A. J.; Wang, Z.-G.; Pierce, N. A. *Nucleic Acids Res.* **2005**, *33*, 4090−4095.
(14) Seelig, G.; Yurke, B.; Winfree, E. *J. Am. Chem. Soc.* **2006**, *128*, 12211−12220.
(15) Dabby, N. L. Synthetic molecular machines for active self-assembly: prototype algorithms, designs, and experimental study. Ph.D. Thesis, California Institute of Technology, 2013.
(16) Zhang, D. Y. *J. Am. Chem. Soc.* **2011**, *133*, 1077−1086.
(17) Wolfe, B. R.; Pierce, N. A. *ACS Synth. Biol.* **2015**, *4*, 1086−1100.
(18) Saini, N.; Zhang, Y.; Usdin, K.; Lobachev, K. S. *Biochimie* **2013**, *95*, 117−123.
(19) Mathews, D. H.; Sabina, J.; Zuker, M.; Turner, D. H. *J. Mol. Biol.* **1999**, *288*, 911−940.
(20) SantaLucia, J., Jr.; Hicks, D. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33*, 415−440.
(21) Zadeh, J. N.; Steenberg, C. D.; Bois, J. S.; Wolfe, B. R.; Pierce, M. B.; Khan, A. R.; Dirks, R. M.; Pierce, N. A. *J. Comput. Chem.* **2011**, *32*, 170−173.
(22) Hunter, J. D. *Comput. Sci. Eng.* **2007**, *9*, 90−95.
(23) Serra, M. J.; Turner, D. H. *Methods Enzymol.* **1995**, *259*, 242−261.
(24) SantaLucia, J., Jr. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 1460−1465.
(25) Koehler, R. T.; Peyret, N. *Bioinformatics* **2005**, *21*, 3333−3339.
(26) Hofacker, I. L.; Fontana, W.; Stadler, P. F.; Bonhoeffer, L. S.; Tacker, M.; Schuster, P. *Monatsh. Chem.* **1994**, *125*, 167−188.
(27) Flamm, C.; Hofacker, I. L.; Maurer-Stroh, S.; Stadler, P. F.; Zehl, M. *RNA* **2001**, *7*, 254−265.
(28) Dirks, R. M.; Pierce, N. A. *J. Comput. Chem.* **2003**, *24*, 1664−1677.
(29) Dirks, R. M.; Lin, M.; Winfree, E.; Pierce, N. A. *Nucleic Acids Res.* **2004**, *32*, 1392−1403.
(30) Andronescu, M.; Fejes, A. P.; Hutter, F.; Hoos, H. H.; Condon, A. *J. Mol. Biol.* **2004**, *336*, 607−624.
(31) Busch, A.; Backofen, R. *Bioinformatics* **2006**, *22*, 1823−1831.
(32) Aguirre-Hernández, R.; Hoos, H. H.; Condon, A. *BMC Bioinf.* **2007**, *8*, 34.
(33) Burghardt, B.; Hartmann, A. K. *Phys. Rev. E* **2007**, *75*, 021920.
(34) Gao, J. Z. M.; Li, L. Y. M.; Reidys, C. M. *Algorithms Mol. Biol.* **2010**, *5*, 27.
(35) Shu, W. J.; Liu, M.; Chen, H. B.; Bo, X. C.; Wang, S. Q. *J. Biotechnol.* **2010**, *150*, 466−473.
(36) Zadeh, J. N.; Wolfe, B. R.; Pierce, N. A. *J. Comput. Chem.* **2011**, *32*, 439−452.
(37) Avihoo, A.; Churkin, A.; Barash, D. *BMC Bioinf.* **2011**, *12*, 319.
(38) Ramlan, E. I.; Zauner, K. P. *BioSystems* **2011**, *105*, 14−24.
(39) Taneda, A. *Adv. Appl. Bioinform. Chem.* **2011**, *4*, 1−12.
(40) Levin, A.; Lis, M.; Ponty, Y.; O'Donnell, C. W.; Devadas, S.; Berger, B.; Waldispühl, J. *Nucleic Acids Res.* **2012**, *40*, 10041−10052.
(41) Matthies, M. C.; Bienert, S.; Torda, A. E. *J. Chem. Theory Comput.* **2012**, *8*, 3663−3670.
(42) Taneda, A. *Front. Genet.* **2012**, *3*, 36.
(43) Lyngsø, R. B.; Anderson, J. W. J.; Sizikova, E.; Badugu, A.; Hyland, T.; Hein, J. *BMC Bioinf.* **2012**, *13*, 260.
(44) Zadeh, J. N. Algorithms for nucleic acid sequence design. Ph.D. Thesis, California Institute of Technology, 2010.
(45) Patzel, V.; Rutz, S.; Dietrich, I.; Köberle, C.; Scheffold, A.; Kaufmann, S. H. E. *Nat. Biotechnol.* **2005**, *23*, 1440−1444.
(46) Penchovsky, R.; Breaker, R. R. *Nat. Biotechnol.* **2005**, *23*, 1424−1433.
(47) Salis, H. M.; Mirsky, E. A.; Voigt, C. A. *Nat. Biotechnol.* **2009**, *27*, 946−950.
(48) Choi, H. M. T.; Chang, J. Y.; Trinh, L. A.; Padilla, J. E.; Fraser, S. E.; Pierce, N. A. *Nat. Biotechnol.* **2010**, *28*, 1208−12.
(49) Li, B. L.; Ellington, A. D.; Chen, X. *Nucleic Acids Res.* **2011**, *39*, e110.
(50) Genot, A. J.; Zhang, D. Y.; Bath, J.; Turberfield, A. J. *J. Am. Chem. Soc.* **2011**, *133*, 2177−2182.
(51) Genot, A. J.; Bath, J.; Turberfield, A. J. *J. Am. Chem. Soc.* **2011**, *133*, 20080−20083.
(52) Choi, J.; Love, K. R.; Gong, Y.; Gierahn, T. M.; Love, J. C. *Anal. Chem.* **2011**, *83*, 6890−6895.
(53) Dong, J.; Cui, X.; Deng, Y.; Tang, Z. *Biosens. Bioelectron.* **2012**, *38*, 258−263.
(54) Nishimura, T.; Ogura, Y.; Tanida, J. *Appl. Phys. Lett.* **2012**, *101*, 233703.
(55) Schade, M.; Knoll, A.; Vogel, A.; Seitz, O.; Liebscher, J.; Huster, D.; Herrmann, A.; Arbuzova, A. *J. Am. Chem. Soc.* **2012**, *134*, 20490−20497.
(56) Vieregg, J. R.; Nelson, H. M.; Stoltz, B. M.; Pierce, N. A. *J. Am. Chem. Soc.* **2013**, *135*, 9691−9699.
(57) Genot, A. J.; Bath, J.; Turberfield, A. J. *Angew. Chem., Int. Ed.* **2013**, *52*, 1189−1192.
(58) Hamblin, G. D.; Hariri, A. A.; Carneiro, K. M. M.; Lau, K. L.; Cosa, G.; Sleiman, H. F. *ACS Nano* **2013**, *7*, 3022−3028.
(59) Santini, C. C.; Bath, J.; Tyrrell, A. M.; Turberfield, A. J. *Chem. Commun.* **2013**, *49*, 237−239.
(60) Choi, H. M. T.; Beck, V. A.; Pierce, N. A. *ACS Nano* **2014**, *8*, 4284−4294.
(61) Jiang, Y. S.; Bhadra, S.; Li, B. L.; Ellington, A. D. *Angew. Chem., Int. Ed.* **2014**, *53*, 1845−1848.
(62) Geary, C.; Rothemund, P. W. K.; Andersen, E. S. *Science* **2014**, *345*, 799−804.
(63) Green, A. A.; Silver, P. A.; Collins, J. J.; Yin, P. *Cell* **2014**, *159*, 925−939.

DOI: 10.1021/jacs.6b12693
J. Am. Chem. Soc. 2017, 139, 3134−3144

(64) Hu, J. M.; Yu, Y. J.; Brooks, J. C.; Godwin, L. A.; Somasundaram, S.; Torabinejad, F.; Kim, J.; Shannon, C.; Easley, C. J. *J. Am. Chem. Soc.* **2014**, *136*, 8467−8474.
(65) Machinek, R. R. F.; Ouldridge, T. E.; Haley, N. E. C.; Bath, J.; Turberfield, A. J. *Nat. Commun.* **2014**, *5*, 5324.
(66) Franco, E.; Giordano, G.; Forsberg, P. O.; Murray, R. M. *ACS Synth. Biol.* **2014**, *3*, 589−599.
(67) Koos, B.; Cane, G.; Grannas, K.; Löf, L.; Arngården, L.; Heldin, J.; Clausson, C. M.; Klaesson, A.; Hirvonen, M. K.; de Oliveira, F. M. S.; Talibov, V. O.; Pham, N. T.; Auer, M.; Danielson, U. H.; Haybaeck, J.; Kamali-Moghaddam, M.; Söderberg, O. *Nat. Commun.* **2015**, *6*, 7294.
(68) Zalatan, J. G.; Lee, M. E.; Almeida, R.; Gilbert, L. A.; Whitehead, E. H.; La Russa, M.; Tsai, J. C.; Weissman, J. S.; Dueber, J. E.; Qi, L. S.; Lim, W. A. *Cell* **2015**, *160*, 339−350.
(69) Galimidi, R. P.; Klein, J. S.; Politzer, M. S.; Bai, S. Y.; Seaman, M. S.; Nussenzweig, M. C.; West, A. P.; Bjorkman, P. J. *Cell* **2015**, *160*, 433−446.
(70) Raschle, T.; Lin, C. X.; Jungmann, R.; Shih, W. M.; Wagner, G. *ACS Chem. Biol.* **2015**, *10*, 2448−2454.
(71) Takahashi, M. K.; Watters, K. E.; Gasper, P. M.; Abbott, T. R.; Carlson, P. D.; Chen, A. A.; Lucks, J. B. *RNA* **2016**, *22*, 920−933.
(72) Lee, Y. J.; Hoynes-O'Connor, A.; Leong, M. C.; Moon, T. S. *Nucleic Acids Res.* **2016**, *44*, 2462−2473.

DOI: 10.1021/jacs.6b12693
J. Am. Chem. Soc. 2017, 139, 3134−3144