

RESEARCH ARTICLE

Exploring of the feature space of *de novo* developed post-transcriptional riboregulators

Gert Peters¹, Jo Maertens¹, Jeroen Lammertyn², Marjan De Mey^{1*}

1 Centre for Synthetic Biology, Ghent University, Ghent, Belgium, **2** BIOSYST-MeBioS, KU Leuven, Leuven, Belgium

* Marjan.DeMey@UGent.be



Abstract

Metabolic engineering increasingly depends upon RNA technology to customly rewire the metabolism to maximize production. To this end, pure riboregulators allow dynamic gene repression without the need of a potentially burdensome coexpressed protein like typical Hfq binding small RNAs and clustered regularly interspaced short palindromic repeats technology. Despite this clear advantage, no clear general design principles are available to *de novo* develop repressing riboregulators, limiting the availability and the reliable development of these type of riboregulators. Here, to overcome this lack of knowledge on the functionality of repressing riboregulators, translation inhibiting RNAs are developed from scratch. These *de novo* developed riboregulators explore features related to thermodynamical and structural factors previously attributed to translation initiation modulation. In total, 12 structural and thermodynamic features were defined of which six features were retained after removing correlations from an *in silico* generated riboregulator library. From this translation inhibiting RNA library, 18 riboregulators were selected using a experimental design and subsequently constructed and co-expressed with two target untranslated regions to link the translation inhibiting RNA features to functionality. The pure riboregulators in the design of experiments showed repression down to 6% of the original protein expression levels, which could only be partially explained by a ordinary least squares regression model. To allow reliable forward engineering, a partial least squares regression model was constructed and validated to link the properties of translation inhibiting RNA riboregulators to gene repression. In this model both structural and thermodynamic features were important for efficient gene repression by pure riboregulators. This approach enables a more reliable *de novo* forward engineering of effective pure riboregulators, which further expands the RNA toolbox for gene expression modulation.

OPEN ACCESS

Citation: Peters G, Maertens J, Lammertyn J, De Mey M (2018) Exploring of the feature space of *de novo* developed post-transcriptional riboregulators. PLoS Comput Biol 14(8): e1006170. <https://doi.org/10.1371/journal.pcbi.1006170>

Editor: Shi-Jie Chen, University of Missouri, UNITED STATES

Received: November 8, 2017

Accepted: April 30, 2018

Published: August 17, 2018

Copyright: © 2018 Peters et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: The first author holds a Ph.D. Grant (121363) from the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen, www.vlaio.be). This research was also supported by the BOF-IOP project MLSB (BOF16/IOP/040) granted by the Bijzonder Onderzoeksfonds (www.ugent.be/en/research/funding/bof). The funders had no role in

Author summary

To allow reliable forward engineering of microbial cell factories, various metabolic engineering efforts rely on RNA-based technology. As such, programmable riboregulators allow dynamic control over gene expression. However, no clear design principles exist for *de novo* developed repressing riboregulators, which limits their applicability. Here, various

study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

engineering principles are identified and computationally explored. Subsequently, various design criteria are used in an experimental design, which were explored in an *in vivo* study. This resulted in a regression model that enables a more reliable computational design of repression small RNAs.

Introduction

Over the last decade, synthetic biology and systems biology spurred major advances in metabolic engineering, resulting in several economically competitive production processes for both bulk and fine chemicals from renewable resources, revolutionizing industrial biotechnology [1, 2, 3, 4, 5]. In this context, interfering with the native metabolism of the production host is a necessity to redirect the metabolic flux towards the product of interest with a view to maximizing productivity [6, 7]. Traditionally, tuning the cellular metabolism has been done through gene deletions, which is impossible for numerous essential genes often related to various biosynthetic pathways [8, 9]. As such, maximizing various production pathways requires tools able to specifically reduce gene expression. To this end, zinc fingers and transcription activator-like effectors were engineered to dynamically control transcription of a specific gene through DNA-binding proteins [10, 11]. These custom gene expression regulators are outperformed by recently emerged clustered regularly interspaced short palindromic repeats interference (CRISPRi), an adaptation of the type II clustered regularly interspaced short palindromic repeats (CRISPR) system controlling transcription through reversible binding of a RNA-guided deactivated Cas9 nuclease to DNA [12]. Various metabolic engineering efforts in multiple organisms used this CRISPRi technology to successfully repress a series of specific genes in a dynamic way, hereby ameliorating the desired product formation [13, 14, 15].

Alternative approaches to control gene expression on the translational level employ small RNAs (sRNAs) to repress protein production by blocking translational initiation, enabling metabolic flux redirection at will [9, 16, 17, 18]. Similar to the CRISPRi technology, which requires a small guide RNA (sgRNA) able to bind to the dCas9 protein, these types of sRNAs also require a protein binding RNA motif as they rely on the stabilizing Hfq protein. This Hfq based riboregulation has been used in various metabolic engineering efforts for gene repression and uses known Hfq binding motifs to combine with an antisense part, targeting specific mRNA. The dependence on coexpressed proteins might cause increased metabolic burden, which can lead to long term genetic instability and unexpected behaviour [14, 19, 20]. To reduce these undesired effects, gene expression modulation systems are preferred that solely rely on RNA [21]. These pure riboregulators require less cellular resources by avoiding the extra translation step, hereby lowering metabolic burden [21]. For example, riboregulator technology is successfully used to precisely downregulate specific genes, hereby redirecting metabolite fluxes towards the phenotype of interest [22, 23, 24]. Also, pure riboregulators, which do not require proteins, harness large potential for the construction of fast responding RNA circuitry [17, 25, 26, 27]. However, practical applications of pure riboregulators are the result of thorough laborious, inefficient and impractical screening of multiple designs comprising hundreds of nucleotides [22, 23].

Early pure riboregulators were designed to hybridize the translation initiation region. The RNA architecture in this region plays a pivotal role in the translation initiation process, enabling gene expression through RNA-RNA interactions [28]. This apparent link between RNA structure and biological function, in combination with the ease and reliability of RNA secondary structure prediction, drove several gene silencing efforts solely using RNA.

However, successful attempts to modulate gene expression using solely *trans* expressed RNA employed a variety of features [29, 30, 31]. As such, interfering with translation initiation using solely RNA-RNA interactions has been attributed to various features of the *trans* expressed RNA molecule [26, 29, 30, 32, 33, 34, 35].

These features are classified as either structural or thermodynamic features. Several structural features of riboregulators modulating translation initiation through RNA-RNA interactions include post-transcriptional ribosome binding site (RBS) occlusion [33, 34], formation of paired termini structures [32], and manipulation of the structural accessibility of the target site [29, 35]. Besides structural constraints, various thermodynamic features were previously used to design and optimize translation interfering riboregulators, mainly comprising formation and activation energies [26, 33]. Formation energies are typically obtained by estimating the minimum free energy (MFE) [33]. Despite the importance of activation energy, various estimation methods for the activation energy were previously used to create functional riboregulators [26, 33]. These methods rely on the initial monomeric structures and are based on the assumption that the unbound nucleotides in this state initiates the RNA-RNA complex formation [26, 33].

This broad range of employed features indicates the lack of consensus in literature, which limits the general applicability of the current design rules for pure riboregulators (without using coexpressed proteins). For instance, simply expressing the antisense strand does not fully repress gene expression on the post-transcriptional level [22, 23]. As such, various types of riboregulators suitable for metabolic engineering purposes were created using a number of different riboregulator design features, once again indicating the lack of consensus in literature on the development of riboregulators [26, 33, 36]. Overall, these riboregulators are either developed from a natural existing RNA regulator chassis or created *de novo*, the latter being the most interesting as this enables forward engineering in a broader context [33, 36]. Moreover, only activating riboregulators were created *de novo*, which limits the construction of genetic circuitry using solely RNA.

Here, we propose a framework for the *de novo* design of pure riboregulators, referred to as translation inhibiting RNAs (tiRNAs), which repress gene expression by blocking translation initiation. To develop this predictive framework, the influence of all features previously attributed to post-transcriptional gene modulation were analyzed in a design of experiments (DOE). This experimental design allows exploration of the feature landscape and evaluation of their influence on gene repression. First, using a library of tiRNA, all features were analyzed *in silico* to create a collection of features with maximal information content. Next, the performance of *de novo* designed tiRNAs was evaluated *in vivo*, and used to construct an ordinary least squares (OLS) and a partial least squares (PLS) model which links riboregulator features to gene repression.

Methods

Strains and growth conditions

Escherichia coli strain DH10B (Invitrogen) was used for both plasmid construction and fluorescence measurement purposes. Unless otherwise stated, all products were purchased from Sigma-Aldrich (Diegem, Belgium). For plasmid construction and fluorescence measurements strains were grown in lysogeny broth (LB) and MOPS EZ rich medium (Teknova, Bioquote, York, United Kingdom) at pH 7.4, respectively at 37°C with shaking. LB was composed of 1% tryptone-peptone (Difco, Erembodegem, Belgium), 0.5% yeast extract (Difco) and 1% sodium chloride (VWR, Leuven, Belgium). LB agar (LBA) plates contain the same

components as LB with the addition of 1% agar. If required, medium was supplemented with $100 \mu\text{g ml}^{-1}$ ampicillin and $50 \mu\text{g ml}^{-1}$ kanamycin.

Plasmids

pSilence plasmids were medium-copy vectors (pBR322 origin of replication (ori) and ampicillin resistance marker, originating from pSB6A1 [37]) with a truncated version (all nucleotides after the transcription start were removed) of the proD [38] as promoter and BBa_B1006 [39] as terminator for tRNA expression (see Figure A in [S1 Text](#) for more details), and pTarget plasmids were low-copy vectors (pSC101 ori and kanamycin resistance marker, originating from pCL1920 [40]) with proB [38] as promoter, *mKate2* [41] as reporter gene, rnpB T1 [39] as terminator, and the target 5' untranslated region (UTR) (see Figure B in [S1 Text](#) for more details). The reporter mKate2 was used due to its low background and good fluorescent protein properties (brightness and maturation time) [41]. A schematic overview of the two plasmid types used in this study (pSilence and pTarget) is shown in Figure C in [S1 Text](#).

The control plasmids used in this study were pBlank₁ and pBlank₂, which are the same vectors as the pSilence and pTarget plasmids, respectively. The pBlank₁ plasmid comprises only the vector backbone and pBlank₂ contains the *mKate2* open reading frame (ORF) and rnpB T1 [39] as terminator, thus without promoter and UTRs. All plasmids used in this study were constructed using Golden Gate [42] and CPEC [43] assembly. DNA oligonucleotides were commercially ordered from IDT (Leuven, Belgium) and DNA sequences of every constructed plasmid were verified using sequencing services (Macrogen Inc., Amsterdam, The Netherlands). All tRNA sequences used in this study are listed in Table A in [S1 Text](#). Details of the plasmids and DNA sequences used in this study are listed in Table B and C in [S1 Text](#), respectively.

In vivo fluorescence and optical density (OD) measurements

For *in vivo* assessment of translational inhibition, strains were plated on LBA plates containing $100 \mu\text{g ml}^{-1}$ ampicillin and $50 \mu\text{g ml}^{-1}$ kanamycin. After overnight incubation, three colonies were inoculated in $150 \mu\text{l}$ MOPS EZ rich medium, covered by a Breathe-Easy sealing membrane (Sigma-Aldrich), and grown overnight on a Compact Digital Microplate Shaker (Thermo Scientific) at 800 rpm and 37°C . Subsequently, these cultures were 1:100 diluted in $150 \mu\text{l}$ of fresh MOPS EZ rich medium and grown on a Compact Digital Microplate Shaker until late log phase (6 h) at 800 rpm and 37°C . Subsequently, fluorescence and OD were measured using a Tecan M200 pro microplate reader. Precultures were grown in Greiner bio-one (Vilvoorde, Belgium) polystyrene F-bottom 96 well plates. Fluorescence and OD measurements were performed after growth in Greiner bio-one (Vilvoorde, Belgium) black μclear 96 well plates. For measuring mKate2 expression an excitation wavelength and an emission wavelength of 588 nm and 633 nm were used, respectively. OD was measured at a wavelength of 700 nm to reduce bias in estimates of cell abundance [44].

Fluorescence data analysis

For fluorescence measurements, two types of controls were used on every 96-well microtiter plate, i.e., a MOPS EZ rich medium blank and *E. coli* DH10B cells without fluorescent protein expression (contains pBlank₁ and pBlank₂ plasmids). The medium blank was used to correct the background OD (OD_{bg}) of the medium. The fluorescence of the strain without fluorescent protein expression (FP_{bg}) was used to correct for the background fluorescence of *E. coli*. For all

strains fluorescence per OD was calculated as follows:

$$\left(\frac{FP}{OD}\right)_{corrected} = \frac{FP - FP_{bg}}{OD - OD_{bg}} \quad (1)$$

The relative protein expression was defined as follows:

$$\text{Relative protein expression (\%)} = \frac{\left(\frac{FP}{OD}\right)_{corrected} \text{ with riboregulator}}{\left(\frac{FP}{OD}\right)_{corrected} \text{ without riboregulator}} \times 100 \quad (2)$$

Feature quantification using RNA bioinformatics

For each tiRNA candidate 12 features were calculated (see Table 1 for detailed definitions), which were determined based on various previous riboregulator efforts described in literature [26, 29, 33, 34, 35]. The tiRNA features are classified in two main groups: thermodynamic properties and structural constraints. All intra- and intermolecular interactions between RNA molecules were predicted using RNAfold [45] and RNAcofold [46], respectively. Both RNA secondary structure prediction algorithms were available through the Vienna RNA package [47] and were used with only the options -noLP -d2 and an accuracy of 10^{-100} , all other settings are set to the default setting. Suboptimal structures of RNA molecules are drawn with probabilities equal to their Boltzmann weights using RNAsubopt [48]. The intermolecular binding between the unbound part of the tiRNA and the UTR is estimated by the RNAup

Table 1. Detailed definitions of all features (free energy of the tiRNA monomer (EA), free energy of the tiRNA-tiRNA dimer (EAA), free energy of the tiRNA-UTR dimer (EAB), formation energy of the tiRNA-tiRNA dimer (FAA), formation energy of the tiRNA-UTR dimer (FAB), total seed energy (ETS), intermolecular binding seed energy (EIS), probability availability of UTR (PAU), RBS coverage of length 5 (RBS5), RBS coverage of length 11 (RBS11), paired termini (PT), and the length of the translation inhibiting RNA (tiRNA) (L) used in the initial *in silico* screening of the tiRNA library for repression of a target untranslated region (UTR). This UTR contains a ribosome binding site (RBS) and controls the coding DNA sequence (CDS) of the reporter protein. All binding probabilities of monomers and dimers are respectively derived from base pairing probability matrices estimated by RNAfold [45] and RNAcofold [46].

Name	Definition of the tiRNA feature
EA	ΔG_{tiRNA} ; free energy of the tiRNA monomer, calculated using RNAfold [45].
EAA	$\Delta G_{tiRNA-tiRNA}$; free energy of the tiRNA-tiRNA dimer, calculated using RNAcofold [46].
EAB	$\Delta G_{tiRNA-UTR}$; free energy of the tiRNA-UTR dimer, calculated using RNAcofold [46].
FAA	Formation energy of the tiRNA-tiRNA dimer; $\Delta G_{tiRNA-tiRNA} - 2 \Delta G_{tiRNA}$.
FAB	Formation energy of the tiRNA-UTR dimer; $\Delta G_{tiRNA-UTR} - \Delta G_{tiRNA} - \Delta G_{UTR}$. With ΔG_{UTR} defined as free energy of the UTR (including first 50 nucleotides of the CDS RNA), calculated using RNAfold [45].
ETS	Average minimal total energy (gains from intermolecular binding and needs to 'open' the binding site) for the binding of unbound parts of the tiRNA monomer to the target UTR, calculated using RNAup [49] for 100 suboptimal structures randomly drawn from Boltzmann ensemble [48].
EIS	Average minimal energy gained from intermolecular binding of unbound parts of the tiRNA monomer to the target UTR, calculated using RNAup [49] for 100 suboptimal structures randomly drawn from Boltzmann ensemble [48].
PAU	Weighted average of the relative number of unbound nucleotides in the UTR monomer with the relative number of nucleotides bound by the tiRNA molecule in the tiRNA-UTR dimer complex as weights.
RBS5	The RBS coverage (relative number of bound nucleotides) in the region $C_{RBS}-2$ to $C_{RBS}+2$, where C_{RBS} is defined as the weighted average of the nucleotides in the UTR bound by the 16S rRNA.
RBS11	The RBS coverage (relative number of bound nucleotides) in the region $C_{RBS}-5$ to $C_{RBS}+5$, where C_{RBS} is defined as the weighted average of the nucleotides in the UTR bound by the 16S rRNA.
PT	Average number of bound nucleotides between the first and the second half of the a tiRNA sequence calculated for 100 suboptimal structures randomly drawn from Boltzmann ensemble [48].
L	The length of the tiRNA (nt)

<https://doi.org/10.1371/journal.pcbi.1006170.t001>

algorithm [49]. All calculations were done using Python scripting on an Intel Xeon E5-2670 (2.60GHz) Linux (Debian) server. Details on the quantification of thermodynamic and structural properties of tiRNA molecules are available in Supplementary Methods 1.1.

Statistical calculations and experimental design

All statistical calculations and analyses were performed in R. Unless otherwise stated, error bars represent the standard deviation ($n = 3$). All coefficient of determinations (R^2 s) were calculated using the hydroGOF package in R.

Experimental design. The 2^{6-2} fractional factorial design, which comprises solely UTR₁, was generated using the R package FrF2 [50]. In the DOE, the -1 and 1 state of the factors were defined as the 0.1 and 0.9 p-quantiles of the original feature distribution, respectively. The center points are designed to be [0, 0, 0, 0, 0, 0], where 0 represents the average of the absolute values of the tiRNA features in the initial library of 1,500,000 possible tiRNA candidates. For each feature in the experimental design all data points (X_i) were centered and scaled based on the 0.1 and 0.9 p-quantiles ($q_{X,0.1}$ and $q_{X,0.9}$) of the original distribution of feature X (Eq 3).

$$\tilde{X}_i = \frac{X_i - (q_{X,0.9} + q_{X,0.1})/2}{(q_{X,0.9} - q_{X,0.1})/2} \quad (3)$$

The centered features \tilde{X} were only used in the analysis of the DOE. All data points of the 2^{6-2} fractional factorial design are shown in Table D in [S1 Text](#). The features FAB and EIS were multiplied by -1 to obtain positive regression coefficients as these were expected to be negatively correlated with tiRNA performance.

Regression models

Ordinary least squares (OLS) regression. The OLS regression was done in R. The OLS regression model was calibrated using the absolute (unprocessed) values of FAB for all data points, including data from target UTR₂. Eq 4 depicts the linear relationship obtained from the OLS regression, where Y_j is the relative protein expression when tiRNA j is present, $X_{j,1}$ is feature FAB of tiRNA j , β_0 and β_1 are regression coefficients and ϵ_j is an error term.

$$\log_{10}(Y_j) = \beta_0 + \beta_1 X_{j,1} + \epsilon_j \quad (4)$$

Partial least squares (PLS) regression. The PLS regression was done in R with the package pls [51]. The PLS model was validated by splitting the data set from UTR₁ and UTR₂ in a training set and validation set (5:1 ratio). Subsequently, the training set was used to create the model by leave-one-out cross validation where predictors were scaled prior to regression (by dividing each variable the sample standard deviation). In PLS regression the matrix of predictors \mathbf{X} is decomposed into orthogonal score matrix \mathbf{T} (projection of \mathbf{X}) and loadings matrix \mathbf{P} , circumventing potential collinearities in the data set:

$$\mathbf{X} = \mathbf{T} \mathbf{P} \quad (5)$$

Subsequently, \mathbf{Y} is regressed on the scores \mathbf{T} (and not \mathbf{X}). The specific PLS algorithm used is kernel PLS, which was described by Dayal et al [52].

Results and discussion

The *trans* expressed tiRNAs are *de novo* designed to inhibit translation initiation of a gene of interest, the rate-limiting step in translation [53], as depicted in Fig 1A. Contrary to previous

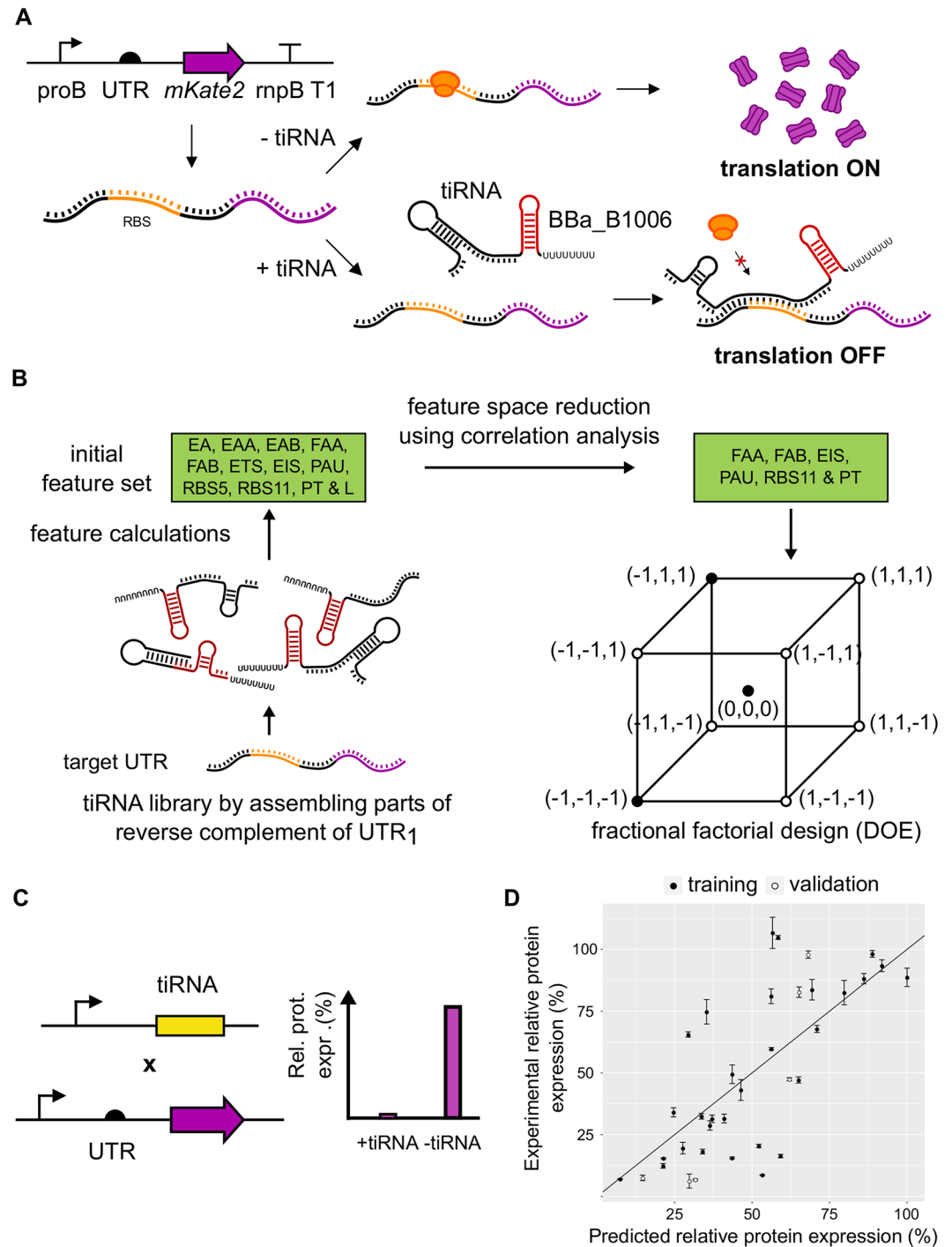


Fig 1. A) Schematic overview of the translation inhibiting RNA (tiRNA) working mechanism B) Workflow for the *in silico* selection of the tiRNAs comprising the design of experiments (DOE) to unravel design principles. The defined tiRNA features (free energy of the tiRNA (EA), free energy of the tiRNA-tiRNA dimer (EAA), free energy of the tiRNA-UTR dimer (EAB), formation energy of the tiRNA-tiRNA dimer (FAA), formation energy of the tiRNA-UTR dimer (FAB), total seed energy (ETS), intermolecular binding seed energy (EIS), probability availability of UTR (PAU), RBS coverage of length 5 (RBS5), RBS coverage of length 11 (RBS11), paired termini (PT), and the tiRNA length (L)) are calculated for a tiRNA library created based on a specific target 5' untranslated region (UTR). C) *In vivo* assessment of the tiRNA in the designed experiment D) Linking features to tiRNA performance through modeling. The fluorescence (FP) per optical density (OD) for a strain was calculated as follows: $(FP/OD)_{corrected} = (FP - FP_{bg}) / (OD - OD_{bg})$ with FP_{bg} = fluorescence of the strain without fluorescent protein expression and OD_{bg} = OD of the medium. The relative protein expression (%) was

defined as the (FP/OD)corrected in presence of the riboregulator divided by the (FP/OD)corrected in the absence of the riboregulator.

<https://doi.org/10.1371/journal.pcbi.1006170.g001>

efforts to construct repressing riboregulators, these RNA devices are constructed from scratch without a functional chassis, which is often based on a natural occurring RNA regulation device [26, 54, 55]. To enable reliable forward engineering of tiRNAs, a workflow to improve the *de novo* development of repressing riboregulators through DOE guided exploration of the sequence space was developed and optimized (see Fig 1B–1D). First, possibly important features for translational inhibiting riboregulators are derived from literature. Secondly, the number of features are reduced by removing correlations. Subsequently, this reduced set of tiRNA properties is used in an experiment designed to unravel design principles to build effective tiRNA molecules. In the DOE, tiRNAs are constructed that explore the feature space in an intelligent way. Ultimately, thoroughly analyzing the performance of the constructed tiRNAs with varying features can improve the knowledge on the structure–function relationship, which correlates to better predictability of *de novo* created riboregulators [26, 33, 36].

Identification of determinative features of repressing riboregulators

In total, 12 potentially determinative features of efficient tiRNA were identified and derived from literature (see Fig 2 and Table 1 for more details). These 12 features represent all design principles previously used in riboregulator construction. Five out of the 12 identified features are based on structural properties. Namely, two features are defined to quantify RBS coverage, i.e. RBS5 and RBS11, which is the average base pairing probability in the region of the RBS with length 5 and 11, respectively. The third feature quantifies the amount of paired termini (PT) and the availability of the UTR is evaluated by the PAU property. The last structural feature is determined by the length of the tiRNA (L). The remaining seven defined features are based on properties relating to thermodynamics. The energy required for the formation of the tiRNA-tiRNA dimer and the tiRNA-UTR dimer are defined as FAA and FAB, respectively. These formation energies are calculated based on the estimated Gibbs free energy of the final

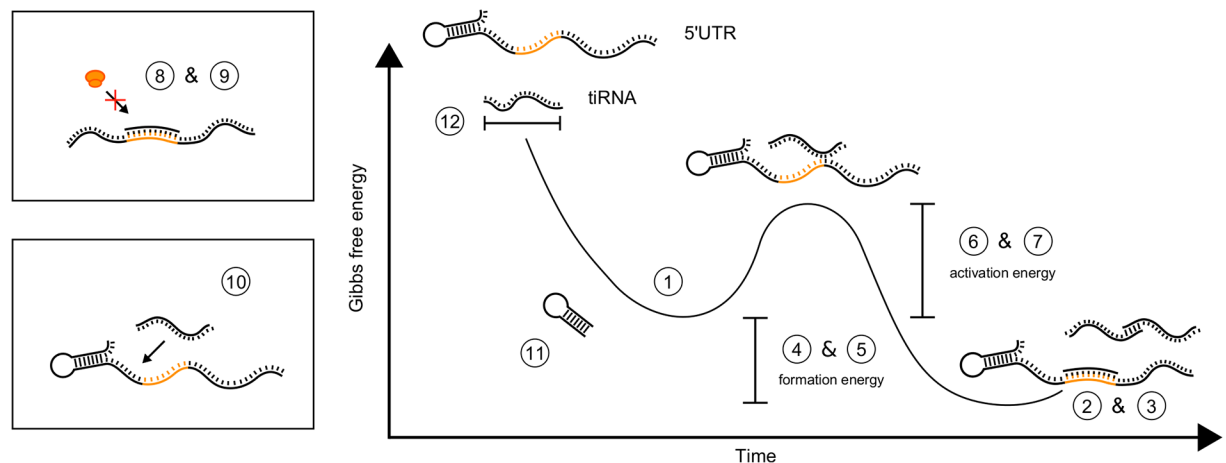


Fig 2. Schematic overview of all potentially determinative features of translation inhibiting RNAs (tiRNAs): 1) free energy of the tiRNA monomer (EA), 2) free energy of the tiRNA-UTR dimer (EAB), 3) free energy of the tiRNA-tiRNA dimer (EAA), 4) formation energy of the tiRNA-UTR dimer (FAB), 5) formation energy of the tiRNA-tiRNA dimer (FAA), 6) intermolecular binding seed energy (EIS), 7) total seed energy (ETS), 8) RBS coverage of length 5 (RBS5), 9) RBS coverage of length 11 (RBS11), 10) probability availability of UTR (PAU), 11) paired termini (PT), and 12) tiRNA length (L).

<https://doi.org/10.1371/journal.pcbi.1006170.g002>

dimer and both initial monomer states, which are described by the EA, the EAA, and the EAB features. In addition, two features describe the activation energy: intermolecular binding seed energy (EIS) and the total seed energy (ETS). Notably, most previous approaches to specify design rules for riboregulators only take the MFE structures into account, simplifying the Boltzmann ensemble of RNA secondary structures and the corresponding complex dynamic energy landscape of regulatory RNAs [56]. The workflow followed here to improve the *de novo* development of repressing riboregulator through DOE guided exploration of the sequence space is depicted in Fig 1B. Here, simplifications were minimized by taking the Boltzmann ensemble into account as much as possible.

Feature space reduction using correlation analysis

A library of 1,500,000 unique possible tiRNA sequences with length 20, 30 or 40 nucleotides (nt) was created *in silico* based on UTR₁ (see Table E in S1 Text) using custom perl code. To generate this library, sequences were generated by combining a random number of different (randomly chosen) parts (length ≥ 2 nt) of the reverse complement of UTR₁ and keeping the order of occurrence of these different parts, as effective riboregulators typically contain parts of the reverse complement of the target UTR.

The amount of correlations between the various features was reduced by analyzing existing correlations between all features, and subsequently removing correlations above a set threshold of 0.75. This was done by calculating the Pearson correlation coefficients (PCCs) (see Fig 3). The correlations between FAA, EA, and EAA, between FAB and EAB, and between EIS and ETS are caused by one or more features being used in the calculation of another feature. Also, RBS5 and RBS11 are correlated, which can be explained by the fact that the RBS region covered by RBS5 is also covered by RBS11. Finally, the length of tiRNA (L) is correlated with EAB as the stability of the tiRNA-UTR complex increases (lower Gibbs free energy) with the tiRNA length. The feature space was reduced, while minimizing information loss, by removing correlations between various features. To this end, one feature of a set of correlated features was selected ($|\text{PCC}| > 0.75$). The reduced set of the tiRNA features X_i with limited correlations, i.e. FAA, FAB, EIS, PAU, RBS11, and PT, was used in a DOE to unravel the features with the most influence on the repression efficiency of these pure riboregulators (see Fig 1B). More specifically, a fractional factorial 2-level design was used with a resolution of IV (2^{6-2} design), comprising two center points and 16 factorial points. After rescaling all features (see Section and Table D in S1 Text for details), the 18 best suiting data points were selected from the library of 1,500,000 tiRNA candidates. The density of all tiRNA features of the complete constructed library with the 0.1 and 0.9 p-quantiles is depicted in Figure D in S1 Text. Because the features are calculated based on the sequence of a generated tiRNA candidate, the factors cannot be set to a specific value. Instead, suitable sequences were selected from the tiRNA candidate library based on the residual sum of squares (RSS) between the real data point of the experimental design and the actual features at this specific candidate (overall, average RSS is 0.59). The selected tiRNA sequences (one feature was selected from features with a PCC above 0.75) with their corresponding theoretical data point are depicted in Table A in S1 Text and Fig 4A, respectively.

In vivo analysis of tiRNA performance

Subsequently, the performance of these 18 tiRNAs in the DOE was evaluated *in vivo* as depicted in Fig 4A. In this experimental setup, the tiRNA molecules are expressed from pSilence plasmids carrying the pBR322 ori, which have an approximately fourfold higher copy number compared to the pSC101 ori of the pTarget plasmids utilized for UTR expression [37],

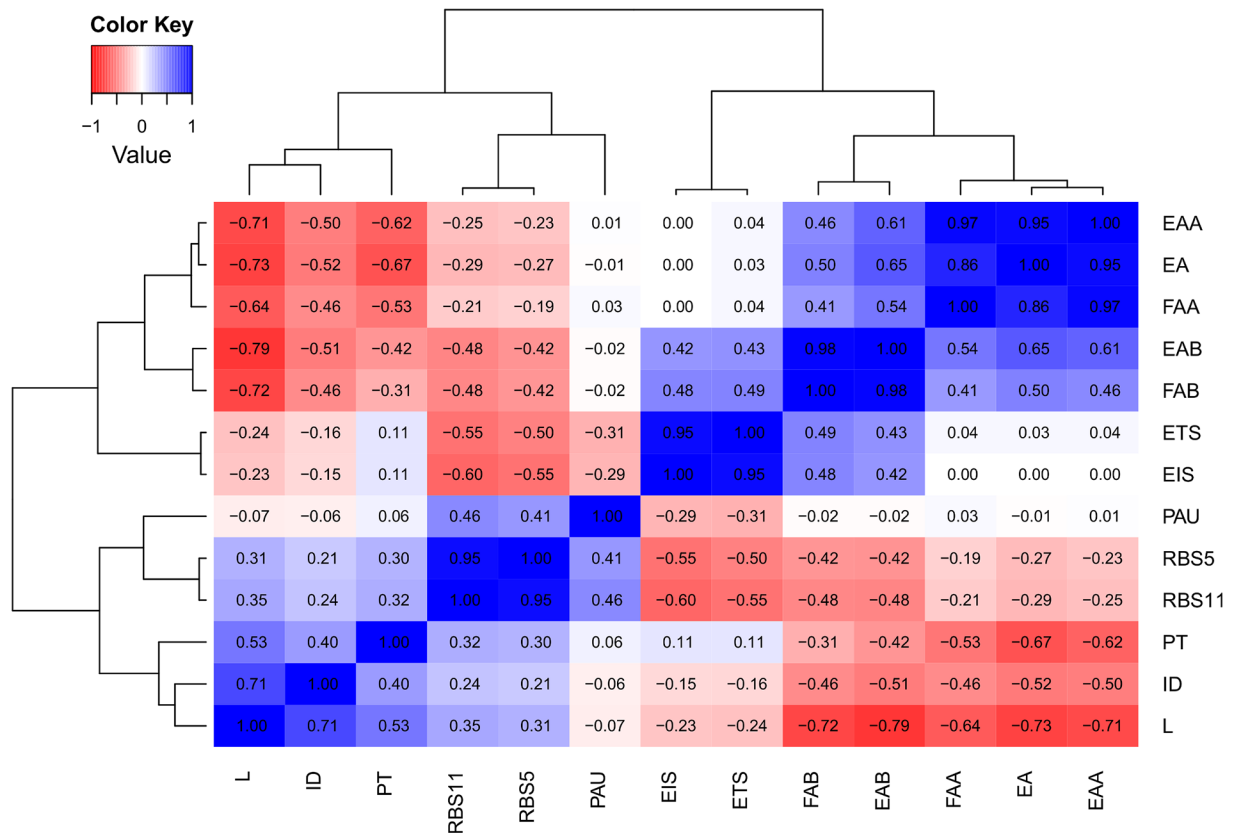


Fig 3. Heatmap of the Pearson correlation coefficients (PCCs) between all features of the translation inhibiting RNA (tiRNA) library, with L = length, EA = free energy of the tiRNA monomer, EAA = free energy of the tiRNA-tiRNA dimer, EAB = free energy of the tiRNA-UTR dimer, FAA = formation energy of the tiRNA-tiRNA dimer, FAB = formation energy of the tiRNA-UTR dimer, ETS = total seed energy, EIS = intermolecular binding seed energy, PAU = probability availability of UTR, RBS5 = RBS coverage of length 5, RBS11 = RBS coverage of length 11, and PT = paired termini (see Table 1 and Fig 2). The fluorescence (FP) per optical density (OD) for a strain was calculated as follows: $(FP/OD)_{corrected} = (FP - FP_{bg}) / (OD - OD_{bg})$ with FP_{bg} = fluorescence of the strain without fluorescent protein expression and OD_{bg} = OD of the medium. The relative protein expression (%) was defined as the $(FP/OD)_{corrected}$ in presence of the riboregulator divided by the $(FP/OD)_{corrected}$ in the absence of the riboregulator.

<https://doi.org/10.1371/journal.pcbi.1006170.g003>

and are under the control of the proD promoter, which showed 8.4 fold higher transcriptional activity compared to the proB promoter used for UTR expression [38]. The overall higher relative tiRNA expression (compared to its target UTR) was chosen based on the fact that *trans* acting sRNA typically require relatively higher expression of the sRNA compared to its target, in both natural and synthetic sRNA regulation systems [57, 58].

To enlarge the data set, the pSilence plasmids were co-transformed with the pTarget plasmid containing UTR₁ or UTR₂ (a truncated version of UTR₁, see Table E in S1 Text), respectively, evaluating the repression efficiency of the tiRNAs in the DOE. Compared to UTR₁, UTR₂ results in 3.3 times less production of fluorescent protein in absence of any riboregulator (see Figure E in S1 Text) although the thermodynamic stability of UTR₁ is much higher than UTR₂ (-27.6 and -17.3 kcal/mol, respectively), which is in contrast to previous studies inversely relating translation to the Gibbs free energy of the UTR [59, 60]. Moreover, the UTR₂ forms much less base pairs in the region around the Shine-Dalgarno (SD) sequence (see Figure E in S1 Text), making the RBS possibly more accessible. However, the removal of a terminal stem loop in the 5' UTR could decrease mRNA stability by exposing the RNA to RNases, resulting in a decrease in fluorescence [61, 62].

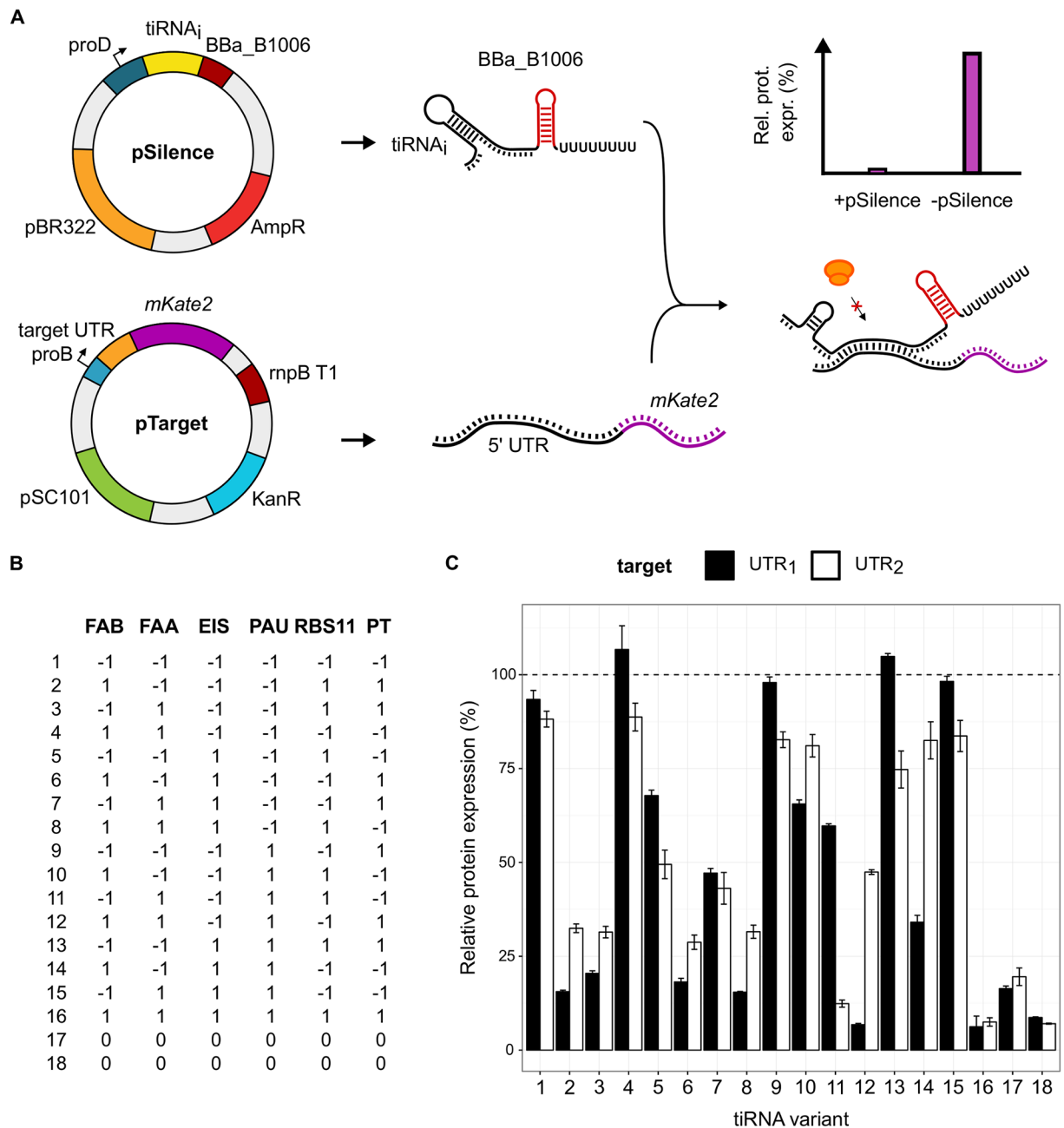


Fig 4. The results from the designed experiment to unravel the principles for translation inhibiting RNA (tiRNA) design. (A) The practical execution of the design of experiments (DOE). All tiRNAs, representing a data point in the DOE, are coexpressed with the target untranslated region (UTR) and the riboregulator efficiency is determined. (B) All tiRNAs with corresponding feature data point in the DOE, where the features are calculated using UTR₁. (C) The tiRNA performance, expressed in relative protein expression, where lower expression represents more effective tiRNAs. The performance was evaluated using both UTR₁ and UTR₂. The 100% relative protein expression represents the protein expression in absence of the tiRNA. Error bars represent standard deviation (n = 3).

<https://doi.org/10.1371/journal.pcbi.1006170.g004>

The outcome of the designed experiment is depicted in Fig 4C. In this DOE tiRNA₁ ([-1,-1,-1,-1,-1,-1]) does not possess any of the features and serves as a control when combined with UTR₁ to determine the burden of the promoter used, which is not significantly different from the strain containing pBlank₁ and pTarget₁. This is in accordance with

literature determining the burden of RNA riboregulators, which is low compared to other types of gene expression regulation [21]. The activity of the *de novo* designed riboregulators shows that almost all tiRNAs were active. Specifically, eight of the 18 tiRNAs targeting UTR₁ inhibit the translation initiation of UTR₁ with more than 75%. The most repressing tiRNAs reduce protein expression of UTR₁ to about 6% of the original expression level. The highest dynamic range of translation repression among all data points is 16, which is higher than previously described repressing riboregulators [26]. Moreover, these tiRNA are created *de novo*, without using a naturally occurring functional chassis.

Overall, the repression levels on UTR₂ are comparable to those of UTR₁, indicating that the truncated part distal to the RBS BBa_B0032 is less important for riboregulator activity. There is a clear difference in repression efficiency between tiRNA₁ ([-1,-1,-1,-1,-1,-1]) and tiRNA₁₆ ([1, 1, 1, 1, 1, 1]), showing the importance of at least one of the selected features for translational inhibition. Another interesting fact is the high repression rate of tiRNA₁₇ and tiRNA₁₈, which are the center points of the DOE. The good performance of these center points indicate that choosing less extreme values can also result in effective translational repression.

Linking features to tiRNA activity

To unravel underlying design principles of repressing tiRNAs, an OLS linear regression analysis was performed in a first approach. To this end, a linear regression model was applied using all data points in the experimental design. All relative expression percentages from all data points of the experimental design are plotted against all normalized features (with only UTR₁ as target) in Figure F in [S1 Text](#). The predicted MFE secondary structure of the 18 different tiRNA:UTR complexes for UTR₁ and UTR₂ is depicted in Figure G and H in [S1 Text](#), respectively. Relative expression percentages plotted against all absolute features for all data (including the repression percentages of UTR₂) are depicted in Figure I in [S1 Text](#). Only two factors in the linear model had a significant influence, namely FAB ($p < 0.05$) and PT ($p < 0.1$). Factors FAB and PT also had significant influence on several other reported riboregulator systems with or without the aid of Hfq [9, 18, 26, 32, 33, 36]. When using only these two features in a linear regression model, only the factor FAB was significant ($p < 0.05$), while factor PT turned out to be not significant ($p > 0.1$). As the factor FAB is based on thermodynamic properties, it was hypothesized that the relation between FAB and the relative protein expression is exponential. Therefore, a linear model was used to relate the logarithmic of relative protein expression percentage to the tiRNA feature FAB (see [Eq 4](#)). The outcome of this OLS regression is depicted in [Fig 5](#). Despite the significant influence of FAB in the DOE, this basic model is still unable to explain all tiRNA functionality which is reflected by the fact that the majority of the data points are not within the 95% confidence interval of the OLS model.

Data driven approaches using regression methods have previously been successful in biological engineering [63, 64] and, more specifically, forward design of various RNA devices [26, 36, 65]. Therefore, in a second approach, PLS regression was performed. To maximize the information possibly linked to tiRNA activity, the 12 defined features were included.

To perform the PLS regression, all data points from UTR₁ and UTR₂ were split into two subsets: one set used for model calibration, i.e. training set, and one independent set used for model validation, i.e. test set. The latter set was selected by randomly picking tiRNAs from three groups which are ordered based on the averaged gene expression of both UTR₁ and UTR₂. The test data comprised tiRNA₉, tiRNA₁₂, and tiRNA₁₈, all other tiRNAs were used in the training set. Before regression, the absolute values of the tiRNA features were scaled through division by the sample standard deviation. Model calibration was done using all 30 data points from the training set and uses 12 features ($k = 12$) describing tiRNA performance. The final

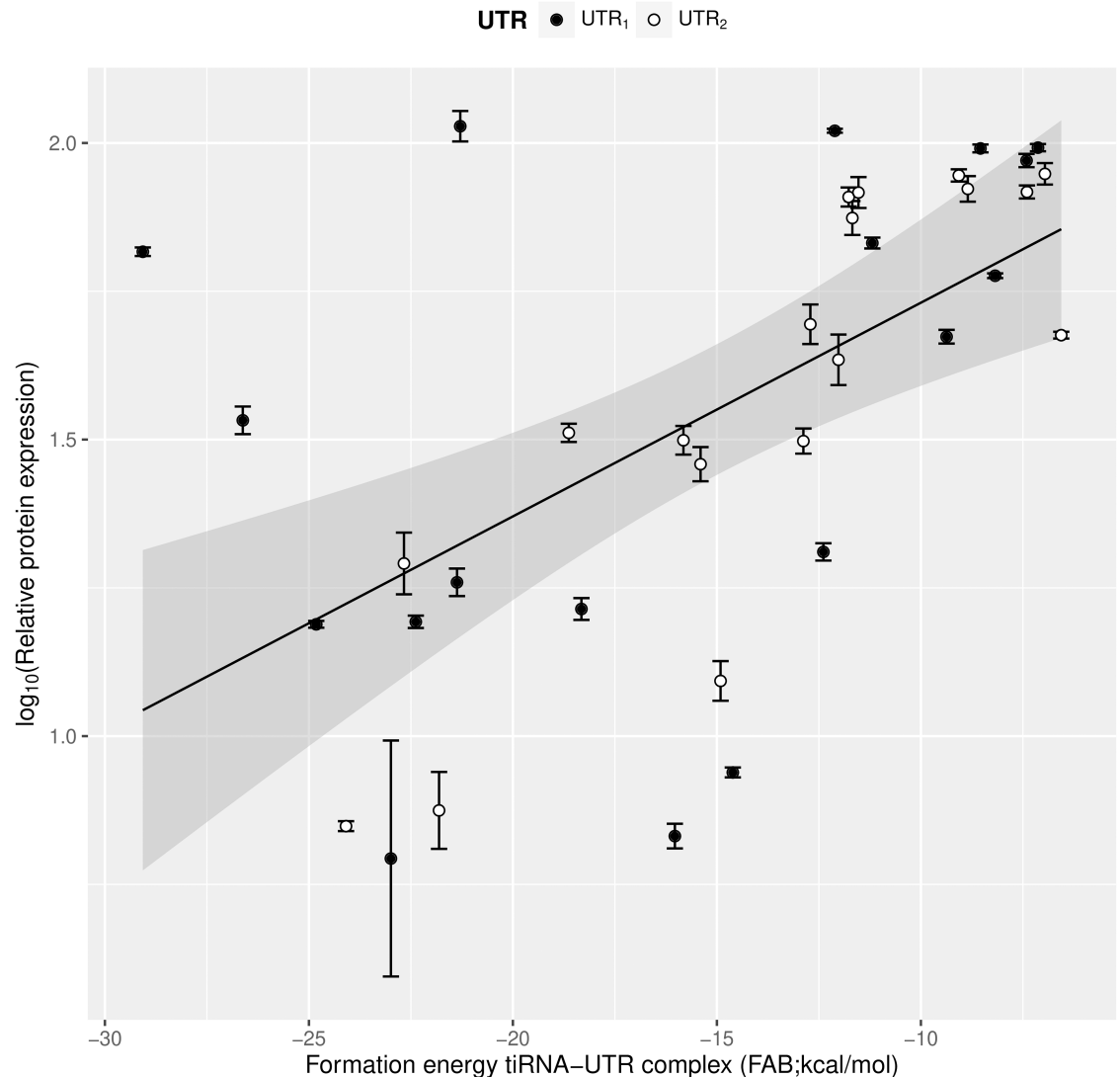


Fig 5. Plot of the ordinary least squares (OLS) regression of the linear model, linking \log_{10} of the relative protein expression to the translation inhibiting RNA (tiRNA) feature formation energy of the tiRNA-UTR dimer (FAB). All data points were used, including the effect of the tiRNAs on both untranslated region 1 (UTR₁) and UTR₂. The gray area depicts the 95% confidence interval of the OLS linear regression. Error bars represent the standard deviation ($n = 3$).

<https://doi.org/10.1371/journal.pcbi.1006170.g005>

model contained 4 latent variables and was selected based on the root mean squared error of prediction and the explained Y variance. By using the training set, a final PLS model contains 63.9% of the X variance, which explained 50.4% variance of the response variable and a R^2 (describing the model efficiency) of 0.50 (see Fig 6). To validate this PLS model, the independent validation set was used to assess the quality of the PLS model. The R^2 of this validation set was 0.69, indicating that the model successfully explains tiRNA activity. To identify the most important factors in the PLS regression model, all estimated regression coefficients are calculated (see Table F in S1 Text for all coefficients and scaling factors). The regression coefficients of the 12 tiRNA features are shown in Figure J in S1 Text. The cumulative loadings of the 4 components and the biplot of the first two components are depicted in Figure K and L in S1 Text, respectively. From these estimates the formation energy of the UTR-tiRNA complex is again inversely correlated to the final protein expression as both regression coefficients of EAB

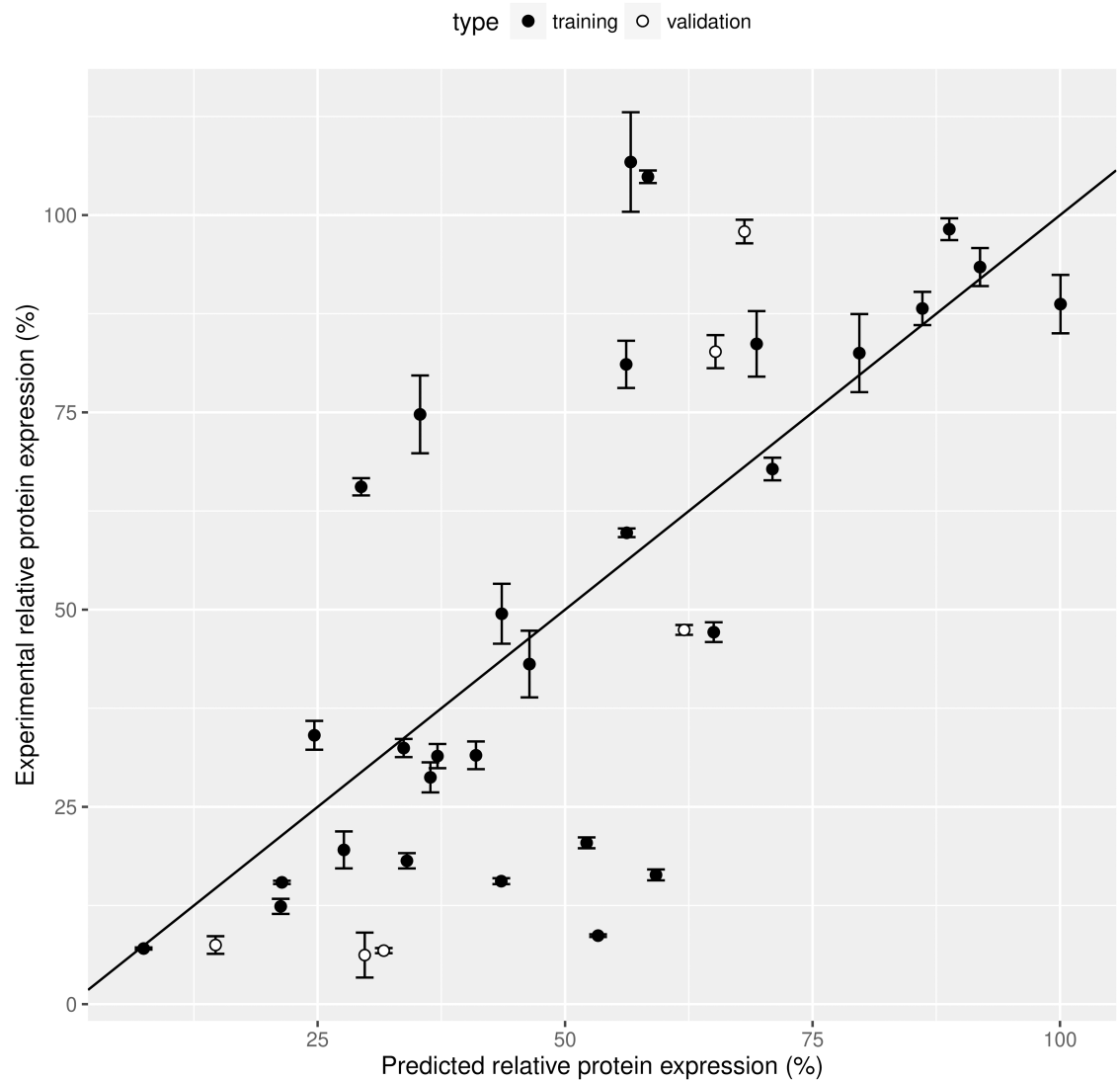


Fig 6. Validation of partial least squares (PLS) regression model predicting relative protein expression from 12 predictors (features of translation inhibiting RNA (tiRNA)). Plot of experimental versus predicted relative protein expression via PLS model for the training set (used for model calibration) and the validation set (used to test the model efficiency; coefficient of determination (R^2) equal to 0.69). Error bars represent the standard deviation ($n = 3$).

<https://doi.org/10.1371/journal.pcbi.1006170.g006>

and FAB are positive. This link between dimer stability and riboregulator performance was also previously observed in other RNA devices [26, 33]. Other observations are the negative relation between FAA and protein expression, indicating that a stable tiRNA-tiRNA dimer complex decreases tiRNA efficiency. Besides these thermodynamic factors, structural features PAU and PT are inversely correlated to protein expression. Thus, as described in literature [29, 32, 35], target nucleotide availability and the number of paired termini (linked to RNA stability) in the riboregulator monomer is important for repression efficiency. Contrary to previous studies, activation energy and total RBS occlusion has a rather limited influence on gene repression.

Overall, the PLS modelling approach employed here successfully predicts tiRNA activity based on the described 12 features, which were defined based on literature. However, various

features used in previously described efforts were quantified using different methods [26, 33]. This lack of standardized methods to determine thermodynamic and structural features of riboregulators complicates forward engineering of riboregulators. Also, the diverse range of features required to explain tRNA functionality is an indication of the complex nature of the regulatory mechanism of riboregulation. As such, RNA regulation might require properties unknown today, which can be discovered using recently developed technologies allowing detailed structural analysis of riboregulators with a high-throughput. For instance, SHAPE-Seq allows *in vivo* characterization of RNA structure by coupling chemical probing techniques to next-generation sequencing technology [66, 67].

Conclusions

The developed approach allows *de novo* design of translation inhibiting riboregulators, which further broadens the RNA regulator toolbox. From the 18 constructed tRNAs molecules designed in the DOE eight tRNAs repressed protein production with more than 75%. The riboregulators described in this paper do not require any coexpressed proteins, which increases their applicability to build complex genetic circuitry. For instance, it allows to reconstitute a RNase III site (resulting in RNA degradation [68]) or interference with guide RNAs of a CRISPR system to obtain complex biological functions. To further improve riboregulator design several basic modelling approaches were employed. However, these basic efforts were unable to fully explain tRNA performance, indicating the complexity of riboregulator repression. Previous efforts often rely on several criteria to engineer riboregulators of various types with varying success [26, 33, 36, 69]. Based on these efforts, 12 features were defined and used in a DOE to explore the tRNA feature space. Subsequently, to improve the reliability of *de novo* forward engineering of repressing riboregulators, a sequence-function model was constructed to link tRNA functionality to the defined tRNA features. To this end, both structural and thermodynamic tRNA features were used in a PLS regression model, which was evaluated using an independent test set (R^2 equal to 0.69). The success of this data driven approach indicates the importance of machine learning techniques in modern synthetic biology to grasp the ever increasing complexity of biological design. Furthermore, the complex nature of riboregulation and the limited knowledge of the underlying working mechanisms makes engineering RNA devices challenging. To this end, novel technologies (for instance SHAPE-Seq) enable high-throughput study of the structure-function relationship of various types of riboregulators in detail by combining RNA structural probing techniques and next-generation sequencing technology, allowing better prediction of riboregulator performance [66, 67].

Supporting information

S1 Text. A single pdf file containing all supplementary methods, figures, and tables. (PDF)

Author Contributions

Conceptualization: Gert Peters, Jo Maertens, Jeroen Lammertyn, Marjan De Mey.

Data curation: Gert Peters.

Formal analysis: Gert Peters.

Funding acquisition: Gert Peters, Jo Maertens, Jeroen Lammertyn, Marjan De Mey.

Investigation: Gert Peters.

Methodology: Gert Peters, Jo Maertens, Marjan De Mey.

Project administration: Marjan De Mey.

Resources: Marjan De Mey.

Software: Gert Peters.

Supervision: Jo Maertens, Jeroen Lammertyn, Marjan De Mey.

Validation: Gert Peters, Jo Maertens, Jeroen Lammertyn, Marjan De Mey.

Visualization: Gert Peters, Jo Maertens, Marjan De Mey.

Writing – original draft: Gert Peters.

Writing – review & editing: Gert Peters, Jo Maertens, Jeroen Lammertyn, Marjan De Mey.

References

1. Paddon CJ, Westfall PJ, Pitera DJ, Benjamin K, Fisher K, McPhee D, et al. High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature*. 2013; 496(7446):528–32. <https://doi.org/10.1038/nature12051> PMID: 23575629
2. Yim H, Haselbeck R, Niu W, Pujol-Baxley C, Burgard A, Boldt J, et al. Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. *Nature Chemical Biology*. 2011; 7(7):445–452. <https://doi.org/10.1038/nchembio.580> PMID: 21602812
3. Atsumi S, Hanai T, Liao JC. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature*. 2008; 451(7174):86–89. <https://doi.org/10.1038/nature06450> PMID: 18172501
4. Pirie CM, De Mey M, Prather KLJ, Ajikumar PK. Integrating the protein and metabolic engineering toolkits for next-generation chemical biosynthesis. *ACS Chemical Biology*. 2013; 8(4):662–672. <https://doi.org/10.1021/cb300634b> PMID: 23373985
5. Ajikumar PK, Xiao WH, Tyo KE, Wang Y, Simeon F, Leonard E, et al. Isoprenoid pathway optimization for Taxol precursor overproduction in *Escherichia coli*. *Science*. 2010; 330(6000):70–74. <https://doi.org/10.1126/science.1191652> PMID: 20929806
6. Biggs BW, De Paepe B, Santos CNS, De Mey M, Ajikumar PK. Multivariate modular metabolic engineering for pathway and strain optimization. *Current Opinion in Biotechnology*. 2014; 29:156–162. <https://doi.org/10.1016/j.copbio.2014.05.005> PMID: 24927371
7. Jones JA, Toparlak ÖD, Koffas MA. Metabolic pathway balancing and its role in the production of biofuels and chemicals. *Current Opinion in Biotechnology*. 2015; 33:52–59. <https://doi.org/10.1016/j.copbio.2014.11.013> PMID: 25445548
8. Cho SH, Haning K, Contreras LM. Strain engineering via regulatory noncoding RNAs: not a one-blueprint-fits-all. *Current Opinion in Chemical Engineering*. 2015; 10:25–34. <https://doi.org/10.1016/j.coche.2015.07.008>
9. Na D, Yoo SM, Chung H, Park H, Park JH, Lee SY. Metabolic engineering of *Escherichia coli* using synthetic small regulatory RNAs. *Nature Biotechnology*. 2013; 31(2):170–174. <https://doi.org/10.1038/nbt.2461> PMID: 23334451
10. Lee JY, Sung BH, Yu BJ, Lee JH, Lee SH, Kim MS, et al. Phenotypic engineering by reprogramming gene transcription using novel artificial transcription factors in *Escherichia coli*. *Nucleic Acids Research*. 2008; 36(16):e102–e102. <https://doi.org/10.1093/nar/gkn449> PMID: 18641039
11. Politz MC, Copeland MF, Pfeleger BF. Artificial repressors for controlling gene expression in bacteria. *Chemical Communications*. 2013; 49(39):4325–4327. <https://doi.org/10.1039/c2cc37107c> PMID: 23230569
12. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*. 2013; 152(5):1173–1183. <https://doi.org/10.1016/j.cell.2013.02.022> PMID: 23452860
13. Cress BF, Toparlak OD, Guleria S, Lebovich M, Stieglitz JT, Englaender JA, et al. CRISPRPathBrick: modular combinatorial assembly of type II-A CRISPR arrays for dCas9-mediated multiplex transcriptional repression in *E. coli*. *ACS synthetic biology*. 2015; 4(9):987–1000. <https://doi.org/10.1021/acssynbio.5b00012> PMID: 25822415
14. Lv L, Ren YL, Chen JC, Wu Q, Chen GQ. Application of CRISPRi for prokaryotic metabolic engineering involving multiple genes, a case study: Controllable P (3HB-co-4HB) biosynthesis. *Metabolic Engineering*. 2015; 29:160–168. <https://doi.org/10.1016/j.ymben.2015.03.013> PMID: 25838211

15. Cleto S, Jensen JV, Wendisch VF, Lu TK. Corynebacterium glutamicum metabolic engineering with CRISPR interference (CRISPRi). *ACS synthetic biology*. 2016; 5(5):375–385. <https://doi.org/10.1021/acssynbio.5b00216> PMID: 26829286
16. Man S, Cheng R, Miao C, Gong Q, Gu Y, Lu X, et al. Artificial trans-encoded small non-coding RNAs specifically silence the selected gene expression in bacteria. *Nucleic Acids Research*. 2011; 39(8):e50–e50. <https://doi.org/10.1093/nar/gkr034> PMID: 21296758
17. Peters G, Coussement P, Maertens J, Lammertyn J, De Mey M. Putting RNA to work: Translating RNA fundamentals into biotechnological engineering practice. *Biotechnology advances*. 2015; 33(8):1829–1844. <https://doi.org/10.1016/j.biotechadv.2015.10.011> PMID: 26514597
18. Hoynes-O'Connor A, Moon TS. Development of design rules for reliable antisense RNA behavior in *E. coli*. *ACS synthetic biology*. 2016; 5(12):1441–1454. <https://doi.org/10.1021/acssynbio.6b00036> PMID: 27434774
19. Shachrai I, Zaslaver A, Alon U, Dekel E. Cost of unneeded proteins in *E. coli* is reduced after several generations in exponential growth. *Molecular cell*. 2010; 38(5):758–767. <https://doi.org/10.1016/j.molcel.2010.04.015> PMID: 20434381
20. Gorochowski TE, van den Berg E, Kerkman R, Roubos JA, Bovenberg RA. Using synthetic biological parts and microbioreactors to explore the protein expression characteristics of *Escherichia coli*. *ACS synthetic biology*. 2013; 3(3):129–139. <https://doi.org/10.1021/sb4001245> PMID: 24299494
21. Ceroni F, Algar R, Stan GB, Ellis T. Quantifying cellular capacity identifies gene expression designs with reduced burden. *Nature methods*. 2015; 12(5):415–418. <https://doi.org/10.1038/nmeth.3339> PMID: 25849635
22. Yang Y, Lin Y, Li L, Linhardt RJ, Yan Y. Regulating malonyl-CoA metabolism via synthetic antisense RNAs for enhanced biosynthesis of natural products. *Metabolic Engineering*. 2015; 29:217–226. <https://doi.org/10.1016/j.ymben.2015.03.018> PMID: 25863265
23. Solomon KV, Sanders TM, Prather KL. A dynamic metabolite valve for the control of central carbon metabolism. *Metabolic Engineering*. 2012; 14(6):661–671. <https://doi.org/10.1016/j.ymben.2012.08.006> PMID: 23026120
24. Kim JY, Cha HJ. Down-regulation of acetate pathway through antisense strategy in *Escherichia coli*: Improved foreign protein production. *Biotechnology and Bioengineering*. 2003; 83(7):841–853. <https://doi.org/10.1002/bit.10735> PMID: 12889024
25. Takahashi MK, Chappell J, Hayes CA, Sun ZZ, Kim J, Singhal V, et al. Rapidly characterizing the fast dynamics of RNA genetic circuitry with cell-free transcription–translation (TX-TL) systems. *ACS synthetic biology*. 2014; 4(5):503–515. <https://doi.org/10.1021/sb400206c> PMID: 24621257
26. Mutalik VK, Qi L, Guimaraes JC, Lucks JB, Arkin AP. Rationally designed families of orthogonal RNA regulators of translation. *Nature Chemical Biology*. 2012; 8(5):447–454. <https://doi.org/10.1038/nchembio.919> PMID: 22446835
27. Callura JM, Dwyer DJ, Isaacs FJ, Cantor CR, Collins JJ. Tracking, tuning, and terminating microbial physiology using synthetic riboregulators. *Proceedings of the National Academy of Sciences*. 2010; 107(36):15898–15903. <https://doi.org/10.1073/pnas.1009747107>
28. de Smit MH, Van Duin J. Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. *Proceedings of the National Academy of Sciences*. 1990; 87(19):7668–7672. <https://doi.org/10.1073/pnas.87.19.7668>
29. Shao Y, Wu Y, Chan CY, McDonough K, Ding Y. Rational design and rapid screening of antisense oligonucleotides for prokaryotic gene modulation. *Nucleic Acids Research*. 2006; 34(19):5660–5669. <https://doi.org/10.1093/nar/gkl715> PMID: 17038332
30. Stefan A, Schwarz F, Bressanin D, Hochkoeppler A. Shine-Dalgarno sequence enhances the efficiency of *lacZ* repression by artificial anti-*lac* antisense RNAs in *Escherichia coli*. *Journal of Bioscience and Bioengineering*. 2010; 110(5):523–528. <https://doi.org/10.1016/j.jbiosc.2010.05.012> PMID: 20646957
31. Nakashima N, Tamura T. Conditional gene silencing of multiple genes with antisense RNAs and generation of a mutator strain of *Escherichia coli*. *Nucleic Acids Research*. 2009; 37(15):e103–e103. <https://doi.org/10.1093/nar/gkp498> PMID: 19515932
32. Nakashima N, Tamura T, Good L. Paired termini stabilize antisense RNAs and enhance conditional gene silencing in *Escherichia coli*. *Nucleic acids research*. 2006; 34(20):e138–e138. <https://doi.org/10.1093/nar/gkl697> PMID: 17062631
33. Rodrigo G, Landrain TE, Jaramillo A. De novo automated design of small RNA circuits for engineering synthetic riboregulation in living cells. *Proceedings of the National Academy of Sciences*. 2012; 109(38):15271–15276. <https://doi.org/10.1073/pnas.1203831109>

34. Isaacs FJ, Dwyer DJ, Ding C, Pervouchine DD, Cantor CR, Collins JJ. Engineered riboregulators enable post-transcriptional control of gene expression. *Nature Biotechnology*. 2004; 22(7):841–847. <https://doi.org/10.1038/nbt986> PMID: 15208640
35. Johnson E, Srivastava R. Volatility in mRNA secondary structure as a design principle for antisense. *Nucleic Acids Research*. 2013; 41(3):e43–e43. <https://doi.org/10.1093/nar/gks902> PMID: 23161691
36. Green AA, Silver PA, Collins JJ, Yin P. Toehold switches: de-novo-designed regulators of gene expression. *Cell*. 2014; 159(4):925–939. <https://doi.org/10.1016/j.cell.2014.10.002> PMID: 25417166
37. Lee TS, Krupa RA, Zhang F, Hajimorad M, Holtz WJ, Prasad N, et al. BglBrick vectors and datasheets: a synthetic biology platform for gene expression. *Journal of biological engineering*. 2011; 5(1):1.
38. Davis J, Rubin A, Sauer R. Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic acids research*. 2011; 39(3):1131–1141. <https://doi.org/10.1093/nar/gkq810> PMID: 20843779
39. Cambray G, Guimaraes JC, Mutalik VK, Lam C, Mai QA, Thimmaiah T, et al. Measurement and modeling of intrinsic transcription terminators. *Nucleic Acids Research*. 2013; 41(9):5139–5148. <https://doi.org/10.1093/nar/gkt163> PMID: 23511967
40. Lerner CG, Inouye M. Low copy number plasmids for regulated low-level expression of cloned genes in *Escherichia coli* with blue/white insert screening capability. *Nucleic acids research*. 1990; 18(15):4631. <https://doi.org/10.1093/nar/18.15.4631> PMID: 2201955
41. Shcherbo D, Murphy C, Ermakova G, Solovieva E, Chepurnykh T, Shcheglov A, et al. Far-red fluorescent tags for protein imaging in living tissues. *Biochem J*. 2009; 418:567–574. <https://doi.org/10.1042/BJ20081949> PMID: 19143658
42. Engler C, Kandzia R, Marillonnet S. A one pot, one step, precision cloning method with high throughput capability. *PLoS one*. 2008; 3(11):e3647. <https://doi.org/10.1371/journal.pone.0003647> PMID: 18985154
43. Quan J, Tian J. Circular polymerase extension cloning of complex gene libraries and pathways. *PLoS one*. 2009; 4(7):e6441. <https://doi.org/10.1371/journal.pone.0006441> PMID: 19649325
44. Hecht A, Endy D, Salit ML, Munson MS. When Wavelengths Collide: Bias in Cell Abundance Measurements due to Expressed Fluorescent Proteins. *ACS synthetic biology*. 2016; 5(9):1024–1027. <https://doi.org/10.1021/acssynbio.6b00072> PMID: 27187075
45. Hofacker IL, Fontana W, Stadler PF, Bonhoeffer LS, Tacker M, Schuster P. Fast folding and comparison of RNA secondary structures. *Monatshefte für Chemie/Chemical Monthly*. 1994; 125(2):167–188. <https://doi.org/10.1007/BF00818163>
46. Bernhart SH, Tafer H, Mückstein U, Flamm C, Stadler PF, Hofacker IL. Partition function and base pairing probabilities of RNA heterodimers. *Algorithms for Molecular Biology*. 2006; 1(1):3. <https://doi.org/10.1186/1748-7188-1-3> PMID: 16722605
47. Lorenz R, Bernhart SH, Zu Siederdisen CH, Tafer H, Flamm C, Stadler PF, et al. ViennaRNA Package 2.0. *Algorithms for Molecular Biology*. 2011; 6(1):1. <https://doi.org/10.1186/1748-7188-6-26>
48. Wuchty S, Fontana W, Hofacker IL, Schuster P, et al. Complete suboptimal folding of RNA and the stability of secondary structures. *Biopolymers*. 1999; 49(2):145–165. [https://doi.org/10.1002/\(SICI\)1097-0282\(199902\)49:2<145::AID-BIP4>3.0.CO;2-G](https://doi.org/10.1002/(SICI)1097-0282(199902)49:2<145::AID-BIP4>3.0.CO;2-G) PMID: 10070264
49. Mückstein U, Tafer H, Hackermüller J, Bernhart SH, Stadler PF, Hofacker IL. Thermodynamics of RNA–RNA binding. *Bioinformatics*. 2006; 22(10):1177–1182. <https://doi.org/10.1093/bioinformatics/btl024> PMID: 16446276
50. Grönmping U. R package FrF2 for creating and analyzing fractional factorial 2-level designs. *Journal of Statistical Software*. 2014; 56(1):1–56.
51. Mevik BH, Wehrens R, et al. The pls package: principal component and partial least squares regression in R. *Journal of Statistical software*. 2007; 18(2):1–24. <https://doi.org/10.18637/jss.v018.i02>
52. Dayal B, MacGregor JF, et al. Improved PLS algorithms. *Journal of chemometrics*. 1997; 11(1):73–85. [https://doi.org/10.1002/\(SICI\)1099-128X\(199701\)11:1%3C73::AID-CEM435%3E3.0.CO;2-%23](https://doi.org/10.1002/(SICI)1099-128X(199701)11:1%3C73::AID-CEM435%3E3.0.CO;2-%23)
53. Laursen BS, Sørensen HP, Mortensen KK, Sperling-Petersen HU. Initiation of protein synthesis in bacteria. *Microbiology and Molecular Biology Reviews*. 2005; 69(1):101–123. <https://doi.org/10.1128/MMBR.69.1.101-123.2005> PMID: 15755955
54. Lucks JB, Qi L, Mutalik VK, Wang D, Arkin AP. Versatile RNA-sensing transcriptional regulators for engineering genetic networks. *Proceedings of the National Academy of Sciences*. 2011; 108(21):8617–8622. <https://doi.org/10.1073/pnas.1015741108>
55. Liu CC, Qi L, Lucks JB, Segall-Shapiro TH, Wang D, Mutalik VK, et al. An adaptor from translational to transcriptional control enables predictable assembly of complex regulation. *Nature Methods*. 2012; 9(11):1088–1094. <https://doi.org/10.1038/nmeth.2184> PMID: 23023598

56. Strobel EJ, Watters KE, Loughrey D, Lucks JB. RNA systems biology: uniting functional discoveries and structural tools to understand global roles of RNAs. *Current opinion in biotechnology*. 2016; 39:182–191. <https://doi.org/10.1016/j.copbio.2016.03.019> PMID: 27132125
57. Kleckner N. Regulating Tn10 and IS10 transposition. *Genetics*. 1990; 124(3):449. PMID: 17246512
58. Meyer S, Chappell J, Sankar S, Chew R, Lucks JB. Improving fold activation of small transcription activating RNAs (STARs) with rational RNA engineering strategies. *Biotechnology and bioengineering*. 2016; 113(1):216–225. <https://doi.org/10.1002/bit.25693> PMID: 26134708
59. de Smit MH, van Duin J. Control of Translation by mRNA Secondary Structure in *Escherichia coli*: A Quantitative Analysis of Literature Data. *Journal of Molecular Biology*. 1994; 244(2):144–150. <https://doi.org/10.1006/jmbi.1994.1714> PMID: 7966326
60. Salis HM, Mirsky EA, Voigt CA. Automated design of synthetic ribosome binding sites to control protein expression. *Nature Biotechnology*. 2009; 27(10):946–950. <https://doi.org/10.1038/nbt.1568> PMID: 19801975
61. Baker KE, Mackie GA. Ectopic RNase E sites promote bypass of 5'-end-dependent mRNA decay in *Escherichia coli*. *Molecular Microbiology*. 2003; 47(1):75–88. <https://doi.org/10.1046/j.1365-2958.2003.03292.x> PMID: 12492855
62. Dasgupta S, Fernandez L, Kameyama L, Inada T, Nakamura Y, Pappas A, et al. Genetic uncoupling of the dsRNA-binding and RNA cleavage activities of the *Escherichia coli* endoribonuclease RNase III—the effect of dsRNA binding on gene expression. *Molecular microbiology*. 1998; 28(3):629–640. <https://doi.org/10.1046/j.1365-2958.1998.00828.x> PMID: 9632264
63. Jonsson J, Norberg T, Carlsson L, Gustafsson C, Wold S. Quantitative sequence-activity models (QSAM)—tools for sequence design. *Nucleic acids research*. 1993; 21(3):733–739. <https://doi.org/10.1093/nar/21.3.733> PMID: 8441682
64. De Mey M, Maertens J, Lequeux GJ, Soetaert WK, Vandamme EJ. Construction and model-based analysis of a promoter library for *E. coli*: an indispensable tool for metabolic engineering. *BMC Biotechnology*. 2007; 7(1):34. <https://doi.org/10.1186/1472-6750-7-34> PMID: 17572914
65. Bonde MT, Pedersen M, Klausen MS, Jensen SI, Wulff T, Harrison S, et al. Predictable tuning of protein expression in bacteria. *Nature methods*. 2016; 13:233–236. <https://doi.org/10.1038/nmeth.3727> PMID: 26752768
66. Watters KE, Abbott TR, Lucks JB. Simultaneous characterization of cellular RNA structure and function with in-cell SHAPE-Seq. *Nucleic acids research*. 2016; 44(2):e12–e12. <https://doi.org/10.1093/nar/gkv879> PMID: 26350218
67. Takahashi MK, Watters KE, Gasper PM, Abbott TR, Carlson PD, Chen AA, et al. Using in-cell SHAPE-Seq and simulations to probe structure–function design principles of RNA transcriptional regulators. *RNA*. 2016; 22(6):920–933. <https://doi.org/10.1261/ma.054916.115> PMID: 27103533
68. Zamore PD. Thirty-three years later, a glimpse at the ribonuclease III active site. *Molecular cell*. 2001; 8(6):1158–1160. [https://doi.org/10.1016/S1097-2765\(01\)00418-X](https://doi.org/10.1016/S1097-2765(01)00418-X) PMID: 11885596
69. Chappell J, Takahashi MK, Lucks JB. Creating small transcription activating RNAs. *Nature Chemical Biology*. 2015; 11:214–220. <https://doi.org/10.1038/nchembio.1737> PMID: 25643173