UNIVERSITY OF CALIFORNIA SAN DIEGO

In Vivo Evolution of a Trans-Splicing Group I Intron Ribozyme

A Thesis submitted in partial satisfaction of the requirements
for the degree Master of Science

in

Chemistry

by

Ishani Behera

Committee in charge:

Professor Ulrich Muller, Chair
Professor Thomas Hermann
Professor Brian Zid

2018

The Thesis of Ishani Behera is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

Chair

University of California San Diego

2018

To my parents

For being my biggest cheerleaders. For giving me everything and making me who I am. For being guiding lights with their wisdom, advice, support, and through their example.

To my family

For always placing an emphasis on education and showing what an impact it can have. For feeling like you are with me even though you are miles away. For always calling to make sure I am happy.

To my friends

For being there no matter what and supporting me in times of stress and need. For visiting me and helping me enjoy each and every moment.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge Professor Ulrich Muller for his support as the chair of my committee. His guidance throughout my time here at UCSD has been invaluable.

I would also like to acknowledge my fellow lab members- Arvin Akoopie, Kevin Sweeney, Logan Norrell, and Joshua Arriola- who have taught me new lab techniques, have given me advice not only for lab work but life, and for their company and friendship.

I would like to acknowledge Logan Norrell on his work in the results and materials and methods sections- the identification of promoter 5 and the first six rounds of the spliceozyme evolution. The results section was initially written by Ulrich Muller, including figures 2-5 and Table 1.

ABSTRACT OF THESIS

In Vivo Evolution of a Trans-Splicing Group I Intron Ribozyme

by

Ishani Behera

Master of Science in Chemistry

University of California San Diego, 2018

Professor Ulrich Muller, Chair

The *Tetrahymena* group I intron was one of the two first catalytic RNAs (ribozymes) to be discovered. Group I introns are sequences that can exist between exons in pre-mRNAs, that are able to self-splice and remove themselves when forming mature mRNAs in the absence of the spliceosome. The *Tetrahymena cis*-splicing ribozyme was engineered to accept substrate RNAs in *trans*. The designed *trans*-splicing ribozyme was termed the “spliceozyme”. The spliceozyme could in principle be used in therapeutic applications. However, there are two hurdles to overcome; the delivery of the ribozyme into cells and its efficiency once delivered. The focus of

this study is to improve the efficiency of the *trans*-splicing ribozyme in cells by evolving the ribozyme in *E. coli*. In a previous study, an evolution system in cells was developed for the spliceozyme which was used to generate a clone, W11, that increased product formation. However, this evolution used high ribozyme expression levels and focused on a single splice site. To improve the efficiency and sequence generality, the spliceozyme was evolved at low expression levels at two different splice sites with different flanking sequences. The results of the spliceozyme evolution in bacterial cells showed that specific mutations are able to improve spliceozyme efficiency in bacterial cells. This suggests that further analysis, and perhaps further evolution experiments, could generate spliceozymes with improved efficiency on a broad range of substrate sequences.

Introduction

The *cis*-splicing group I intron from *Tetrahymena thermophila* is one of the two first known catalytic RNAs (ribozymes) [1] [2]. Group I introns are sequences that can exist between exons in pre-mRNAs, that are able to self-splice and remove themselves when forming mRNAs in the absence of the spliceosome [1]. Group I introns use two consecutive trans-phosporylation reactions to self-splice [3]. With a Mg^{2+} cofactor, an exogenous guanosine initiates splicing through a nucleophilic attack by its free 3' hydroxyl group on the phosphate of the 5' splice site. The 5' splice site is based on recognition of a conserved wobble pair between the terminal U of the 5' exon and a conserved G within the intron's internal guide sequence (Figure 1) [4]. A conformational change, aided by the P10 helix, brings the 3' splice site to the active site which aligns the 5' and 3' exons. The free 3' hydroxyl group at the 5' splice site then attacks the 3' splice site, ligating the exons and releasing the intron. The 3' splice site is recognized by a conserved terminal G which is selected through hydrogen bonding and stacking interactions in the same site into which the endogenous G binds in the first step of the reaction cascade [5].

Group I introns from genera *Tetrahymena*, *Twort*, and *Azoarcus,* have been well characterized both biochemically and structurally with five high-resolution crystal structures [6] [7] [8] [9]. Although these ribozymes differ in their peripheral regions, the core structure including the active site is conserved. The active site contains the bound guanosine (either the exogenous G or the 3'-terminal G), which completes a base triple, sandwiched by three other base triples, and a metal. The *Tetrahymena* Group I intron provides the most detailed biochemical characterization [10]. The *Tetrahymena* ribozyme is robust with regard to different sequence contexts in the peripheral areas and is active *in vitro*, in bacteria, and in eukaryotes

including yeast, mouse, and human cells. The group I introns from *Pneumocystis carinii* [11], *Candida albicans* [12], *Didymium* and *Fuligo* [13], show similar potential but the well-studied *Azoarcus* ribozyme has very specific secondary structure requirements that make it less robust [14]. Therefore, most *trans*-splicing ribozyme studies focus on the *Tetrahymena* ribozyme which can be designed in different ways.

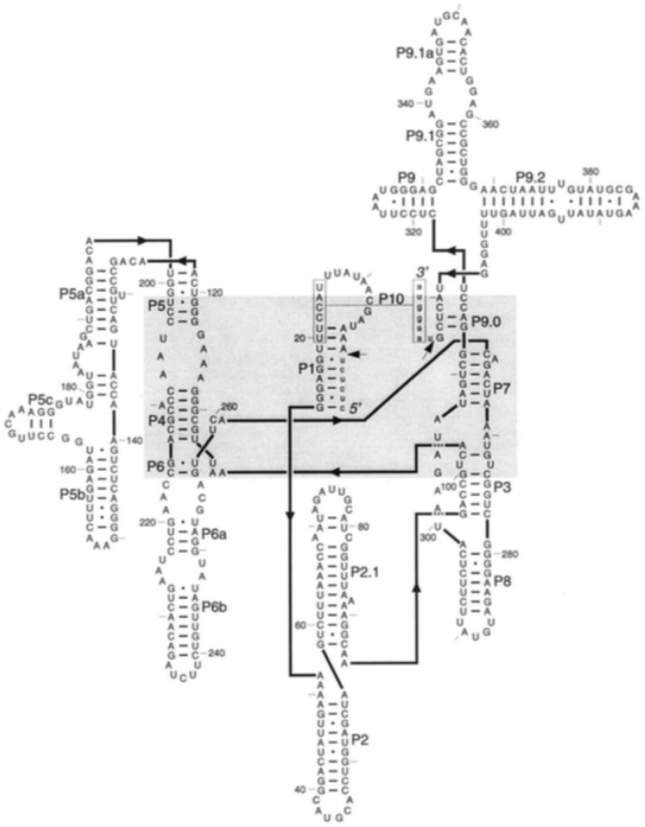


Figure 1. Secondary structure of the *Tetrahymena* group I intron ribozyme [4]. In all subsequent figures, the P4-P6 domain will be shoes to the right of the rest of the ribozyme for easier visualization of the secondary structures at the catalytic site.

Variants of the *cis*-splicing group I introns have been engineered to splice a substrate in *trans.* The first type of a *trans*-splicing ribozyme was designed based on the *Tetrahymena* Group I intron [15]. It could replace the 5' end of a substrate with the 3' end of the *trans*-splicing ribozyme. The *trans*-splicing ribozyme contacts the substrate by forming a P1 helix which docks into the active site allowing splicing at the 5' splice site. The free 3' hydroxyl group at the 5'

splice site is now free to attack the 3' end of the ribozyme resulting in the 5' portion of the substrate RNA and the 3' exon of the designed *trans*-splicing ribozyme [15] [16]. If a substrate RNA carries a mutation in its 3' portion, this *trans*-splicing ribozyme variant could be used to replace the 3' portion with a the wild-type sequence repairing the RNA.

In contrast to the designed *trans*-splicing at the 5' splice site, a *Pneumocystis carinii* ribozyme variant splices at the 3' splice site and replaces the substrate 5' end with the ribozymes' 5' exon [17]. The ribozyme recognizes the 3' splice site of the substrate by forming the P10 and P9.0 duplexes. This variant can then splice at the substrate 3' splice site, and replace the substrates 5' portion with the 5' exon of the ribozyme construct in the P1 duplex leading to the replacement of the substrates 5' portion. If there is a mutation in the 5' portion of an RNA, this variant could be used to replace it with the wild type sequence.

When both the 5' splice recognition and 3' splice recognition principles are combined in one ribozyme, it is possible to excise or insert a nucleotide in a substrate sequence. Variants based on the *Pneumocystis carinii* ribozyme have been engineered that *trans*-splice by using the P1 duplex as 5' splice recognition and the P9.0 and P10 duplexes as 3' splice recognition leading to the excision of internal sequences or an insertion in the internal sequence of an RNA [11]. The excision of longer fragments, up to 28 internal nucleotides of a substrate RNA has been shown *in vitro* [18] [19]. The efficiency of sequences longer than a single nucleotide is low in cells. [20]. The *trans*-splicing *Pneumocystis carinii* ribozyme could be used to remove insertion mutations or other frame-shift mutations in RNA. The insertion of short oligomeric sequences allows the insertion of modified RNAs into a target sequence *in vitro* for research purposes [21]. Thiol-modifications or isotopically labelled oligomers could be inserted for biochemical or NMR analysis.

When a different principle is used for the recognition of the 3' splice site, a *Tetrahymena* group I intron variant is able to excise a 100-nt-long intron from pre-mRNAs both *in vitro* and in bacterial cells [22]. This variant uses the P9.0 and P9.2 duplexes for recognition of the 3' splice site; the 5' splice site is still recognized with the P1 duplex (Figure 2). The P9.2 helix is formed in *trans* between the 3' terminus of the ribozyme and the target sequence of the substrate. The minimal length of both the *trans* P9.2 and P1 helix is about 6-7 base pairs. The two trans-phosphorylation reactions lead to the excision of the RNA segment between the 3' and 5' splice sites and the ligation of the upstream and downstream segments. These ribozymes are termed "spliceozymes" as they behave similarly to the spliceosome.

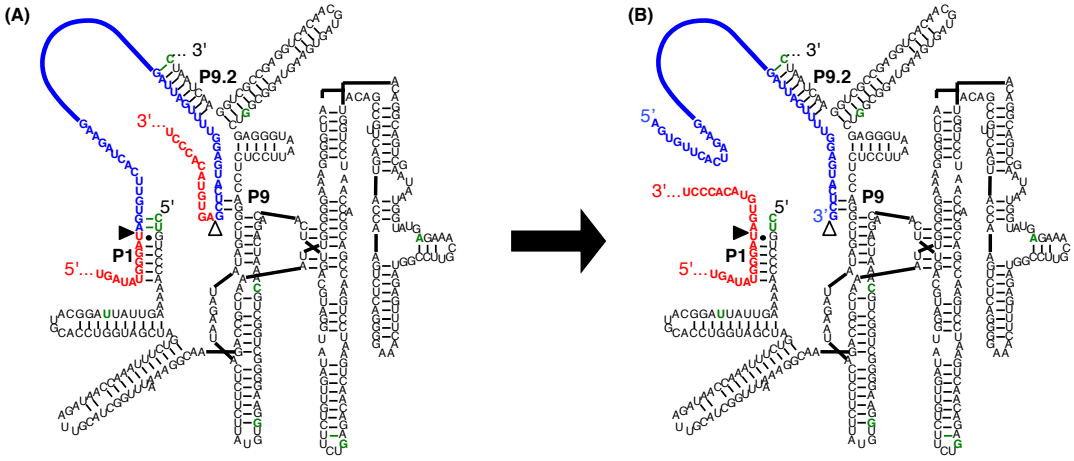


Figure 2. Secondary structure of 5' and 3' trans-splicing ribozyme or spliceozyme. The spliceozyme (black) base pairs with the pre-mRNA (red) and intron (blue) by forming the P1, P9.2, and P9 duplexes. **(A)** Before the splicing reaction, the pre-mRNA contains the 5' exon (red), an intron (blue), and the 3' exon (red). The 5' splice site is indicated by a filled triangle. The 3' splice site is indicated with an open triangle. **(B)** After the splicing reaction the 5' and 3' exons are joined together (red) while the intron (blue) is excised. [22]

Any uridine nucleotide in the substrate can be a possible target site for *trans*-splicing ribozymes due to the necessity of a G•U wobble pair. In order to predict efficient splice sites for the spliceozyme computationally, the binding free energies were calculated by summing computed energies of unfolding the target site, unfolding of the ribozyme 5' terminus, and the

binding of the target site and ribozyme 5' terminus [23]. The *trans*-splicing efficiency of 18 different splice sites were tested on a chloramphenicol acetyl transferase (CAT) mRNA. The results correlated very well with the free energy predictions, and additionally identified a new splice site in position 258 of the CAT gene, which turned out to be more efficient than any other splice site on CAT mRNA. The second-most active splice site was splice site 448, at nucleotide position 448 of the CAT gene.

To develop improved *trans*-splicing variants, an evolution system was designed in which the population generated and accumulated mutations that increase activity over multiple rounds of mutagenesis, selection, and amplification (Figure 3) [24] [25]. A *Tetrahymena* ribozyme variant was used to repair a CAT mRNA that was inactivated by a frameshift mutation in its' 3' portion. Ribozyme variants that were able to correctly replace the 3' portion would lead to *E. coli* growth on medium containing chloramphenicol. This *Tetrahymena* ribozyme variant was subjected to 21 rounds of evolution. The most efficient variant that was tested after the 21 rounds had four clustered mutations in the P6b loop. These mutations generated an accessible $(C)_5$ homopentamer sequence which recruited the Rho transcription factor. These mutations did not increase *trans*-splicing efficiency, but increased the translation of the spliced mRNA. This showed that an evolving population *in vivo* can interact with cellular components to benefit its evolutionary interest. Using a similar evolution system, the spliceozyme was improved through 10 rounds of evolution to generate a clone with improved splicing efficiency using a single splice site [26]. An intron was inserted into splice site 258 of CAT mRNA, the most efficient splice site, in order to inactivate it. The spliceozyme would have to remove this insertion thereby ligating the CAT mRNA together in order for *E. coli* to grow on medium containing chloramphenicol. The most efficient clone out of those that survived 10 rounds of evolution,

contained 10 mutations that increased product formation by reducing side product formation, reducing activity at the 5' splice site, and increasing activity at the 3' splice site.

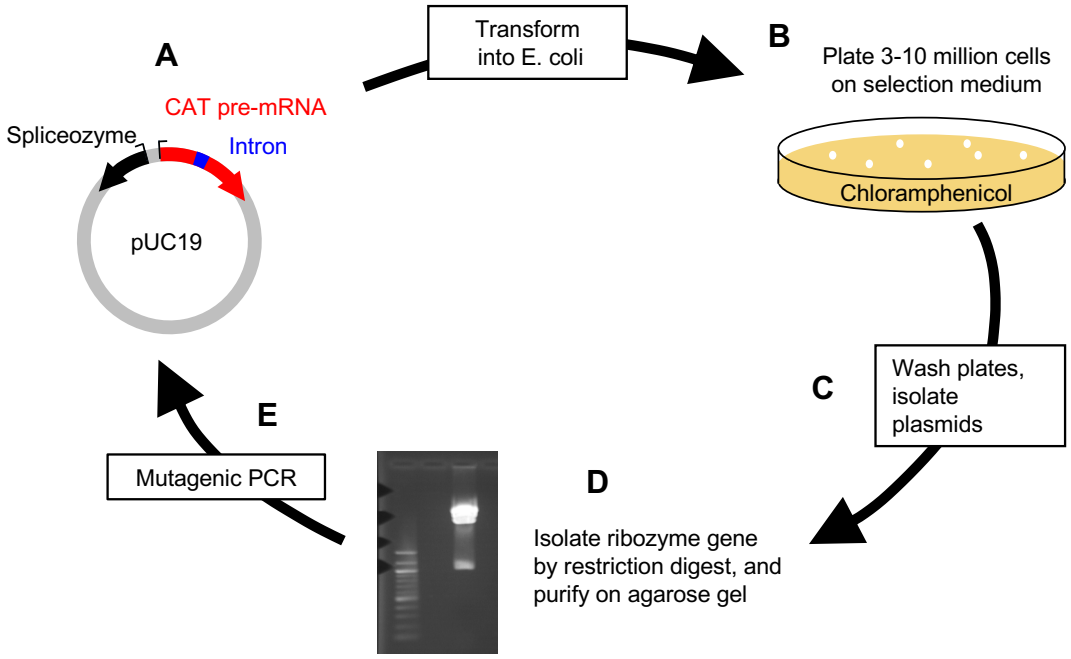


Figure 3. Evolution of spliceozymes in bacterial cells. **(A)** A library plasmid is generated by cloning two expression cassettes into a pUC19-based vector (grey). The expression of chloramphenicol acetyl transferase (CAT) pre-mRNA (red) is driven by a constitutive promoter. The mRNA is inactivated by an intron (blue). Spliceozyme (black) expression is driven by an IPTG-inducible promoter. The substrate binding sites on the spliceozyme are designed to recognize the two splice sites of CAT pre-mRNA and remove the intron. **(B)** After transformation of the library plasmid into E. coli and induction with IPTG, the bacterial library is plated on LB medium containing chloramphenicol. This allows only those bacterial cells to form colonies that contain spliceozymes able to remove the intron from CAT pre-mRNA. **(C)** The colonies are washed off the plates, and their plasmids isolated. **(D)** The spliceozyme genes are amplified from the isolated plasmids by PCR, and the PCR products separated from the plasmids by agarose gel electrophoresis. **(E)** The PCR products are used as templates for mutagenic PCR. This library of spliceozyme genes is cloned into a fresh library plasmid and subjected to additional evolution rounds.

Research aims to improve *trans*-splicing ribozymes not only from a biochemical perspective but for their ability to be used therapeutically. *Trans*-splicing ribozymes could be used therapeutically in two different ways. The first use for Group I intron ribozymes is to repair a disease-causing mutated mRNA by replacing the mutation portion of mRNA with the corrected sequence [15]. With current variants, the 3′ terminal end or the 5' terminal end can be replaced [17]. If there is an insertion, this can also be removed through the use of a *trans*-splicing

ribozyme [11]. Second, they can be used to selectively kill disease causing cells by replacing a portion of an RNA with a 'suicide gene' such as diphtheria toxin A or herpes simplex virus thymidine kinase (HSV-tk) [27][28][29]. This will cause apoptosis or lead to the cell death. Cancer cells can be killed by group I introns in this manner [28]. A *trans*-splicing ribozyme can be used to target sites on RNAs that are highly expressed in cancer cells and splice a 'suicide gene' into those sites. The suicide gene can encode for diphtheria toxin A, a cytotoxic peptide, or a termination sequence, triggering cell death of the cancer cell. A *trans*-splicing ribozyme, currently in clinical trials, targets hTERT RNA, a highly expressed RNA in cancer cells [29]. The ribozyme can splice in a suicide gene that encodes for herpes simplex virus thymidine kinase(HSV-tk) which converts the prodrug ganciclovir to a highly genotoxic phosphorylated ganciclovir which acts as a chain terminator of DNA synthesis. This causes selective induced cell death only in dividing cells such as those found in tumors.

Many other diseases can potentially be treated by *trans*-splicing ribozymes that occur due to mutations that cause alternative splicing. Some known alternative splicing disease include: familial dysautonoma, frontotemporal labor dementias, amyotrophic lateral sclerosis, Hutchinson-Gilford progeria syndrome, Medium-chain acyl-coA dehydrogenase (MCAD) deficiency, myotonic dystrophy, Prader-Willi Syndrome, Spinal muscular atrophy, β-thalassemia, and Tauopathies, which are caused by mutations at the 5' splice site, point mutations in regulatory sequences, exon skipping, or short repeat elements [30].

In the future, the spliceozyme could be used to target mRNAs that have a retained intron from the alternative splicing process. In an analysis of 96,822 alternative splicing events using 23 eukaryotic genomes, 12,337 (12.37%) were intron retention events [31]. For instance, an intron retention in the CACNA1H gene can lead to childhood absence epilepsy (CAE) and other

idiopathic generalize epilepsies. CAE occurs in children 4-10 years and can have ongoing effect into adulthood [32]. Children have absence seizures (4 to 20 seconds long) sometimes in the hundreds per days which can cause severe impairment of consciousness. CACNA1H regularly encodes for $Ca_v3.2$, a T-type member in the voltage-dependent calcium channel complex that helps mediate the influx of calcium ions into the cell upon membrane polarization. A common mutation in the CACNA1H gene is the retention of intron 23 which is 60 amino acids long (about 180 nucleotides) which can cause bursting behavior in neurons. The spliceozyme could be used to target this 180-nucleotide long sequence and excise the remainder of intron 23 from the CACNA1H gene. Another possible target for the spliceozyme could be SPINK5 (serine protease inhibitor Kazal-type 5), which when mutated, can lead to Netherton syndrome [33]. Netherton syndrome is a severe skin disease in children and young adults that causes skin inflammation, universal itch, severe dehydration, stunted growth, reddish skin, and increases susceptibility to infections and allergies. A c.283-12T>A mutation in intron 4 causes retention of the last 10 nucleotides of intron 4. A c.1820+53G>A in intron 19 and c.891C>T in exon 11 cause the retention of the first 54 nucleotides of intron 19. Any of these retentions cause a loss of function of SPINK5 leading to Netherton syndrome. The spliceozyme could be used therapeutically to excise either or both partial introns.

For the use of spliceozymes as tools in therapy, efficient delivery methods need to be developed. Currently, various delivery systems including adeno-virus vectors [34], enveloped protein nanocages [35], and modified Salmonella [36] are being developed but improvements to minimize toxicity, off-target effects, immune responses, and increase delivery to target tissues are necessary. Current adeno-virus vectors may be a possible delivery system for the spliceozymes. Additionally, the use of spliceozymes in therapy is limited by the efficiency,

accuracy and off-target effects, and general utility in cells. Efficiency is based on how much substrate becomes product. Spliceozyme accuracy depends on binding to the correct splice site and splicing at the correct base. A minimization of off-target effects is important to confirm that the spliceozyme targets the correct mRNA and not does not splice any others. This study aims to improve the use of spliceozymes for use in therapy by lowering the expression level to improve splicing efficiency and by switching splice sites to improve its specificity and general utility in *E. coli* cells.

Results

A previously developed spliceozyme [22] is able to remove a 100-nucleotide intron from the pre-mRNA of chloramphenicol acetyl transferase (CAT) in *E. coli* cells, with an efficiency high enough to support the growth of *E. coli* bacteria on medium containing chloramphenicol. This spliceozyme was evolved in *E. coli* cells to mediate growth at increased chloramphenicol concentrations (Figure 3) resulting in a spliceozyme variant termed W11, containing 10 mutations from the parent sequence [26]. This spliceozyme W11 is more efficient than the parent spliceozyme because it has an increased activity at the 3' splice site, and a decreased activity at the 5' splice site. This resulted in a 6-fold decrease of side products relative to products at shorter time points (20 min and 30 min), and a 2-fold increase in product formation at longer time points (60 minutes). These previous results showed that evolution in *E. coli* cells is able to fine-tune the sequence and reaction characteristics of the spliceozyme for improved performance.

To be useful in a clinical setting, the existing spliceozyme would have to be improved in two ways. First, the expression level of spliceozymes in the previous studies was about 10,000 molecules per cell, which would be too high for a clinical setting [25]. Second, the evolutionary optimization of the spliceozyme was done on a single splice site of CAT mRNA (splice site 258) with a single intron sequence [26]. This splice site is the most efficient in CAT mRNA for *trans*-splicing group I intron ribozymes, and most splice sites show much lower splicing efficiency [23]. In a clinical setting, the spliceozymes should be expressed at low intracellular concentrations, and work with similar efficiency at many different target sites. To identify spliceozymes that work efficiently at low expression levels, and less sequence dependence, we

evolved spliceozymes in *E. coli* cells at low expression levels, and with a swapping of splice sites and intron sequences over the course of the evolution.

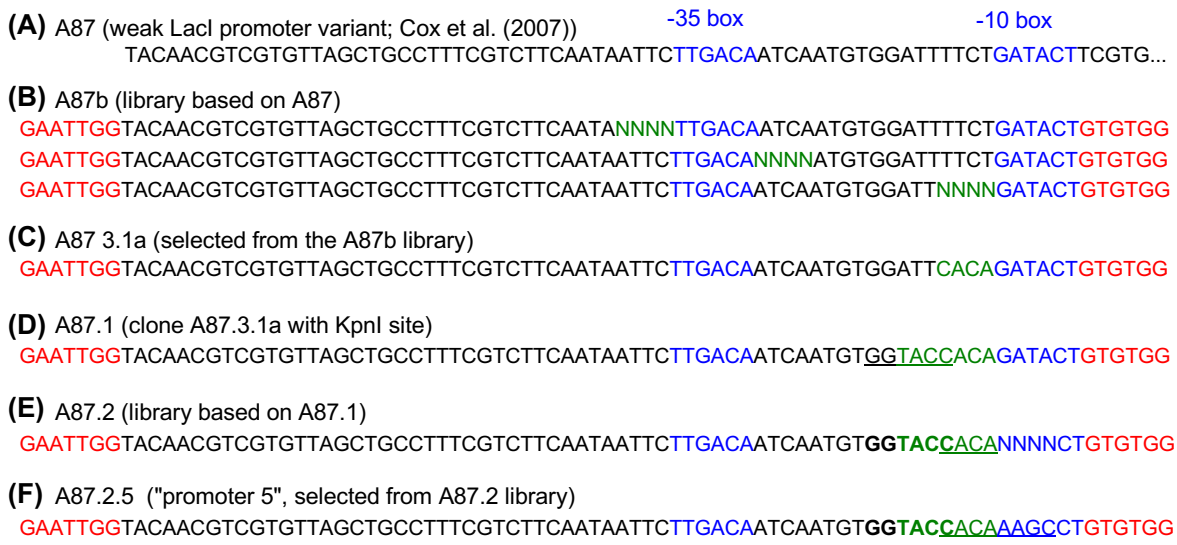


Figure 4. Development of the spliceozyme expression "promoter 5" that was used for the evolution. **(A)** the starting construct A87, which showed the weakest expression of five LacI promoter constructs in a previous publication (CoxElowitz07). **(B)** Library A87b, based on promoter A87, which was used to identify weaker promoters than A87. The red upstream was attached to facilitate cloning. The red downstream sequence was taken from our previous spliceozyme expression constructs. **(C)** Clone A87 3.1a resulted from one round of selection, and screening for weak growth, from the A87b library. **(D)** Clone 87.1 is identical to 87 3.1a, with exception that a KpnI cloning site was inserted into the promoter for easier cloning. This promoter mediated slightly stronger growth than A87 3.1a. **(E)** Library based on clone A87.2, with an NNNN sequence replacing the first four nucleotides of the -10 box. After one selection round, four clones were identified with weak growth. The spliceozyme expression in these clones was measured by RT-qPCR, showing an ~300-fold reduction compared to A87.1. **(F)** Clone A87.2.5 ("promoter 5") was chosen for the evolution due to its low expression level and the ability to mediate bacterial growth on LBcam.

Evolution conditions that would adjust the spliceozymes for low expression levels would be spliceozyme expression levels as low as possible - but high enough to mediate growth of *E. coli* cells on medium containing chloramphenicol. In addition, the spliceozyme promoter used in the evolution needed to be inducible by IPTG such that the bacteria could be grown in the absence of spliceozyme expression on medium containing ampicillin. This would only select for the presence of the plasmid and not for or against specific spliceozyme variants. To develop such promoter sequences, we started by testing five different IPTG-inducible lac promoters [38] to medicate weak, spliceozyme-mediated growth. The promoter closest to the desirable behavior

was promoter A87 (Figure 4A). In four steps, we converted promoter A87 to "promoter 5", which showed low expression level and the ability to mediate bacterial growth on LBcam medium.

In the first step, four positions in promoter A87 were randomized, in three different positions (Figure 2B). This library was screened for weak growth on LB medium containing chloramphenicol (4 mg/mL). This identified clone A87 3.1a (Figure 4C). In the second step, the restriction site KpnI was inserted into promoter A87 3.1a to allow easier cloning. This resulted in promoter A87.1 (Figure 4D), which mediated slightly stronger bacterial growth on medium containing chloramphenicol that A87 3.1a. From promoter A87.1, a library was generated by randomizing the first four nucleotides of the -10 promoter box (Figure 4E). From this library, four clones were identified as showing the desired, very weak growth on medium containing chloramphenicol. Promoter A87.2.5 ("promoter 5"; Figure 4F) was used for the spliceozyme evolution studies.

The spliceozyme expression level mediated by promoter 5 was characterized using RT-qPCR (Figure 5). The promoter trc2 (used in previous evolution experiments [25] [26] [22]) resulted in an average of 8.3 fewer qPCR cycles than promoter A87.2.5 ("promoter 5"), corresponding to >300-fold lower [39] spliceozyme concentration with promoter 5 (A87.2.5) than with promoter trc2. Because the trc promoter resulted in about 10,000 ribozyme copies per cell [25] the expression level from promoter 5 corresponds to ~30 spliceozyme molecules per cell. We were satisfied with this low expression level and the resulting selection pressure for the evolution of spliceozymes in *E. coli* cells described here. Evolutionary adaptation for better use of the spliceozymes at low expression levels might be achieved by reduced side product formation [26] or - even more exciting - in multi turnover splicing. Multi turnover splicing

should be possible for spliceozymes because in contrast to almost all previously developed *trans*-splicing group I intron ribozymes the spliceozymes do not get modified during intron removal.

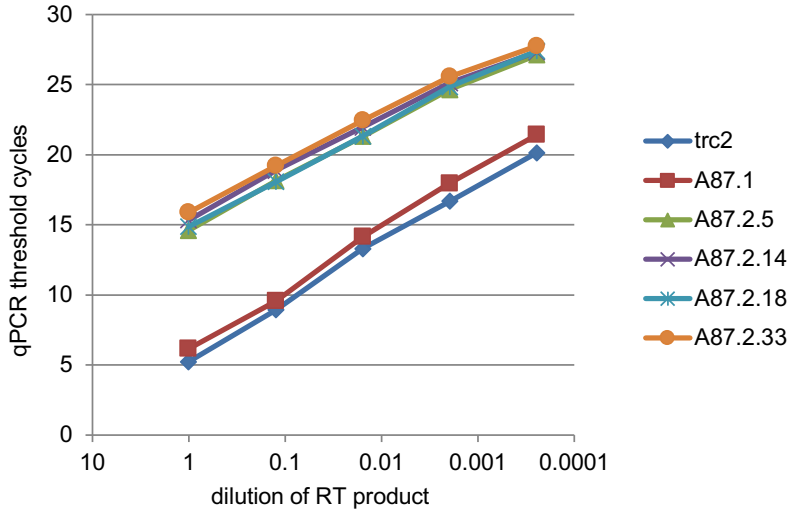


Figure 5. RT-qPCR analysis of the spliceozyme expression levels for different promoters. The number of qPCR cycles to cross a threshold of product concentration is plotted as a function of the dilution of the sample. Five different dilutions were used for each sample to reduce the impct of qPCR artifacts. The two promoters trc2 and the promoter A87.1 were shifted by an average of 8.3 qPCR cycles relative to the promoters in round A87.2. This correponds to >300-fold lower spliceozyme concentration with promoter 5 (A87.2.5) than with promoter trc2.

To evolve spliceozymes with broader sequence generality the spliceozyme was challenged with different substrate sequences in successive rounds of the evolution (Table 1). In rounds 1-6 of the evolution, the 100-nucleotide intron sequence, which is removed by the spliceozyme, was changed twice (from rounds 2-3 and from rounds 4-5). Additionally, in evolution rounds 7-9, both the splice site and the intron sequence were changed.

Table 1. Parameters during the 9 rounds of evolving spliceozymes in bacterial cells. The splice cite was switched from 258 to 338 after round 6. The intron was switched every two rounds until round 6, then every round. The colonies were estimated for every round, both on the selection plates with choloramphenicol (LBcam) as well as on plates that only selected for the presence of the plasmid (LBamp) Considering the insert ratio (the ration of colonies whose plasmids contain an insert, based on 18 arbitrary clones) the maximal effective complexity was calculated. The ration between the cell growth on LBcam and LBamp is a measure of how well the pools survived the selection pressure. The parameters that were not determined are labelled ND.

Round	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Splice Site	258	258	258	258	258	258	448	258	448
Intron	SiC10	SiC10	SiC9	SiC9	SiC4	SiC4	3C39	SiC8	4C6
Added [Mg2+]	10	10	10	10	10	0	0	0	0
[cam] ug/mL	4	4	4	4	4	4	4	8	12
Cfu(LBamp)	3.4E6	4.1E6	3.1E6	2.3E6	1.3E6	4.3E6	6.3E6	5.3E4	1.0E6
Insert ratio	0.94	0.89	0.89	0.94	0.94	1.00	0.89	0.94	0.94
Effective cfu	2.8E6	3.6E6	2.8E6	2.2E6	1.2E6	4.3E6	5.6E6	4.9E4	9.9E5
Plated LBcam	2.8E6	1.8E6	1.9E6	1.2E6	2.6E6	3.6E6	3.3E6	3.3E6	3.8E6
Effect. Compl.	**1.9E6**	**1.4E6**	**1.4E6**	**0.9E6**	**1.1E6**	**2.4E6**	**3.3E6**	**4.9E4**	**6.1E5**
Fract. Survive	0.55	0.49	0.40	0.38	0.35	0.17	ND	ND	ND
OD_{600} LBamp	46	28	43	40	51	55	ND	ND	ND
OD_{600} LBcam	11	6.5	6.0	4.0	9.5	3.6	ND	ND	ND
LBcam/LBamp	0.24	0.23	0.14	0.10	0.19	0.07	ND	ND	ND

The evolution experiment started from the single clone W11 which was the winner of the previous spliceozyme evolution experiment [26]. In evolution rounds 1-6 (but not 7-9), 10 cycles of mutagenic PCR [39] were used to introduce mutations into the spliceozyme (Table 1). In our experience, this generated about 2.4 mutations per ribozyme, which were reduced to about 1.1 mutations per ribozyme after the selection step [25]. Rounds 7-9 relied on the lower mutagenesis rate inherent to the used Taq DNA polymerase, which in our hands generated about 4-fold less mutations than mutagenic PCR. To adopt the evolving spliceozyme population to different intron sequences, these sequences were switched after round 2 from the previously used intron SiC10 [26] to the second-best intron SiC9, and after round 4 to the third-best intron SiC4 of a previous study [22].

We chose splice site 448 as target site in round 7 to adopt the evolving spliceozyme population to a new different splice site, and therefore new flanking sequences. Splice site 448 is the second-best splice site in CAT mRNA, less efficient than the best splice site 258 [37]. Splice site 448 had never been used for the spliceozymes, therefore it was necessary to identify sequences that could be used as introns in splice site 448: Most intron sequences are removed quite inefficiently from the splice site by the spliceozymes, and a functional evolution setup would require an intron that can be removed easily, at least for the initial rounds of evolution. To do this, a library was constructed by inserting a PCR product with 78 randomized positions into splice site 448 of the CAT expression cassette in the library plasmid, using spliceozyme W11. The selection setup was identical to the evolution setup described in Figure 3, only that the introns in the CAT gene provided the genetic diversity and were re-inserted into the plasmid library as opposed to the spliceozymes, and that no mutagenic PCR was used. In the first round of this selection experiment, the effective complexity at the selection step was 5.4×10^{11} different intron sequences of the plated bacterial library. Over four rounds of the selection procedure the most efficient intron sequences were enriched in the population.

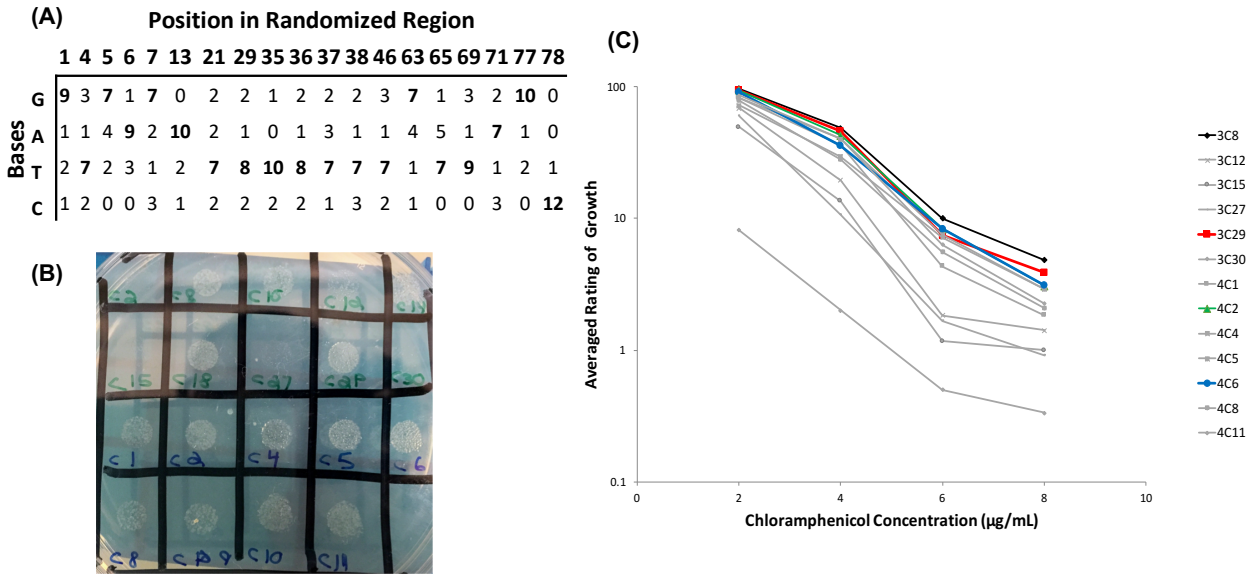


Figure 6. Determination of most efficient intron sequence at splice site 448. Sequencing analysis and semi-quantitative characterization of growth on medium containing chloramphenicol, mediated by plasmids containing the W11 spliceozyme and a CAT expression cassette with an intron library inserted after position 448. **(A)** A sequence analysis on 19 clones out of which 13 clones were unique sequences. Nucleotides in the randomized region (N_{78}) of the intron that occurred in more than 50% of the sequences are shown. The numbers show how many sequences had each nucleotide in those specific positions. The bases that appeared in more than 50% of sequences are bolded at their corresponding position. **(B)** 19 clones from the bacterial library were plated at the same density in 5 µL aliquots on four different concentrations of chloramphenicol. The plates were grown for 19 hours at 37°C. The 6µg/mL LBcam plate shown has19 clones plated. The growth was described by visually assigning numbers between 0 and 100, by two different persons, in triplicate. **(C)** The average of these values was plotted as a function of the chloramphenicol concentration.

Sequencing of 19 clones from rounds 3 and 4 resulted in 5 duplicates, and a total of 13 different intron sequences that were the correct length. The sequences of these 13 clones were compared showing there was a preference for certain nucleotides in 12 different positions of the randomized region (Figure 6A). At least 9 out of 13 sequences had a GC at positons N77 and N8 respectively which is complementary to the spliceozyme sequence they would bind to at these positions. These 13 clones were compared for their effect on bacterial growth on medium containing four different concentrations of chloramphenicol as well as their sequences (Figure 6B and C). Four clones showed the best growth, containing the introns 3C8, 3C29, 4C2, and 4C6. Intron 3C8 had possible contamination and could not be used for the evolution as a precaution. Intron 3C29 was used for evolution round 7 of the spliceozyme evolution, and intron

4C2 was used for evolution round 9, both with splice site 448. In evolution round 8, splice site 258 was used with intron SiC8 [22]. The reason behind this switching of splice sites and introns over the course of the evolution was to generate, and enrich for spliceozymes that showed less dependence on the substrate sequence.

Over the course of the evolution, the effective complexity of the evolving population had an upper limit of at least 49 thousand (Table 1). For rounds 1-6 the upper limit was 0.9 million. The fraction of surviving clones decreased from 24% in round 1 to 7% in round 6. This would signal a decrease in fitness of the evolving pool, and can be the result of a high mutational load [25] [40]. Nevertheless, the evolving pool also survived the additional evolution rounds 7-9, where it was challenged to switch splice sites and intron sequences. In Rounds 8 and 9 the effective complexity was lowered due to choosing LB-agar plate with higher chloramphenicol concentrations, 8 and 12 μg/mL respectively, as well as switching splice site every round.

After a total of nine rounds of evolution, the selected spliceozyme genes were re-cloned into fresh library plasmid, with intron SiC8 in splice site 258. The plasmids were transformed into fresh *E. coli* cells, and a total of 80 bacterial clones were screened for growth on medium containing 10 μg/mL chloramphenicol. Eight clones that showed stronger growth than spliceozyme W11 were chosen for further analysis.

The eight 'winners' of the evolution were challenged to splice in two different sequence contexts, splice site 258 (with intron SiC8) and splice site 448 (with intron 4C2). These growth assays were done in independent triplicate experiments, mostly in two triplicates, resulting in three to six data points for each clone (Figure 7). The eight clones showed very diverse behavior, with all clones better or on par with the parental spliceozyme W11 on either target sequence. For example, clone N3 mediated by far the most efficient growth on splice site 258, while clone C11

mediated the most efficient growth on splice site 448. The behavior of most clones was very robust between experiments, whereas clone 42 showed dramatic differences between experiments, perhaps mediated by metastable folding or annealing characteristics influenced by the mutations.

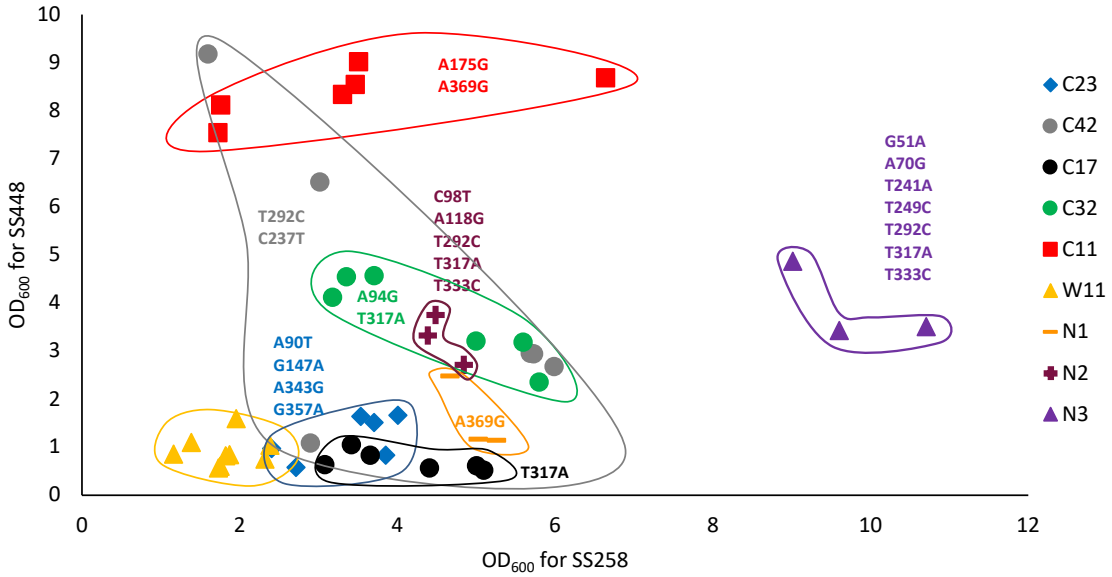
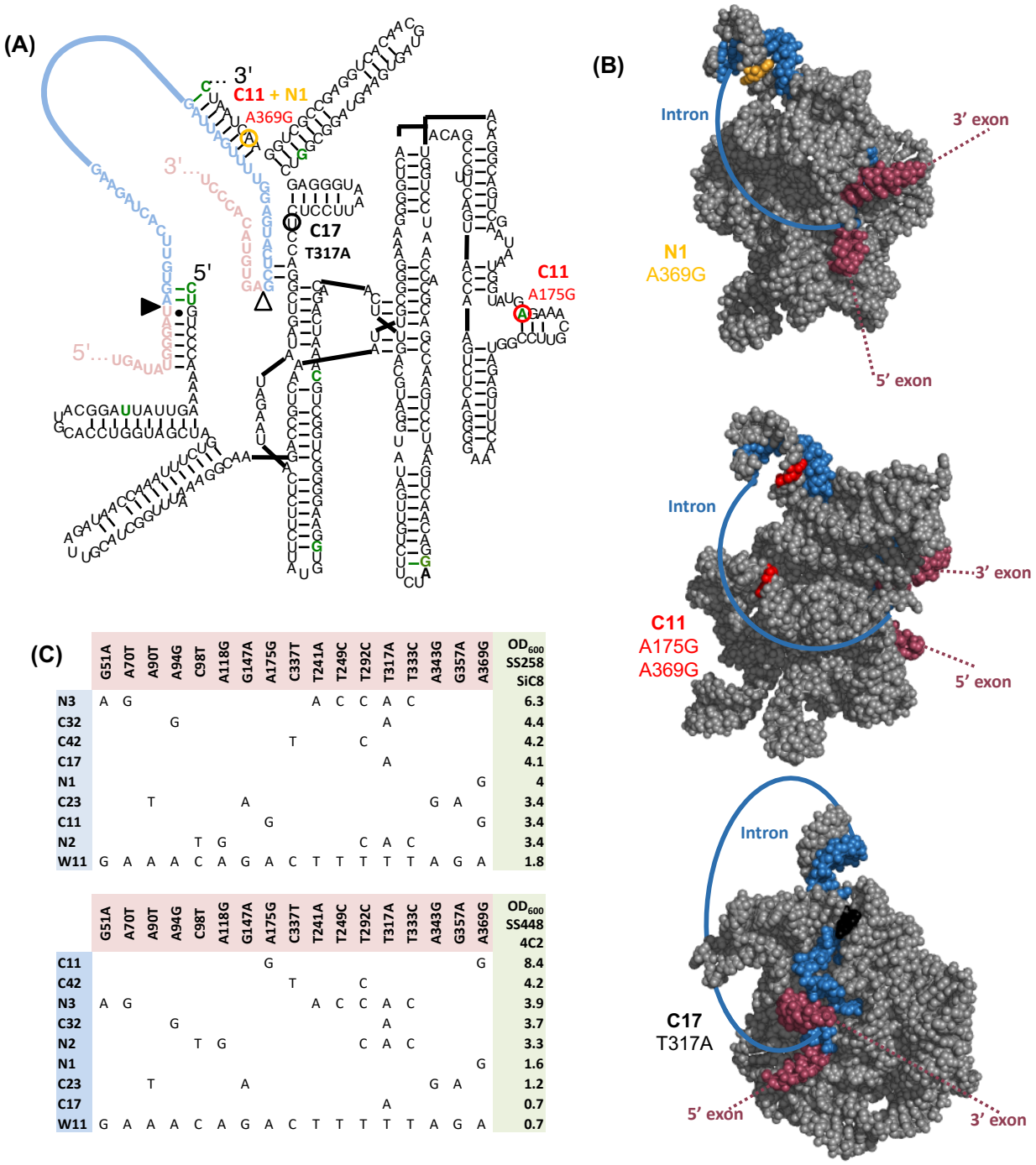


Figure 7. Growth of bacterial clones containing nine individual spliceozymes in two different sequence contexts. The nine clones analyzed were clone W11, the starting point of the evolution, and eight clones that were isolated from round 9 of the evolution experiment. Growth on LB_{cam} plates was measured by growing the bacterial cells for 19 hours on an LB plate with a chloramphenicol concentration of 10 µg/mL, suspending the cells in LB medium, and measuring the OD_{600} compared to LB medium. The growth (OD_{600}) in the sequence context of splice site 258 is shown on the x-axis, the growth in the sequence context of splice site 448 on the y-axis. The data from each clone are shown in colored symbols, with their legend on the right. The data points from each clone are encircled to help comparison of all data points.

In three cases, it was possible to ascribe the different bacterial growth on medium containing chloramphenicol to specific spliceozyme mutations (Figure 8). First, spliceozyme clone N1 carried only a single mutation compared to W11 (A369G), which resulted in much better growth with splice site 258 and SiC8. The growth of clone N1 with splice site 448 is within error of W11, suggesting that mutation A369G specifically improves splicing on splice site 258. Second, spliceozyme C11 carried mutation A175G, in addition to the same mutation as N1 (A369G). Growth of clone C11 is dramatically better with splice site 448, suggesting that mutation A175G

specifically improves splicing on splice site 448. On splice site 258, clone C11 shows a much larger variation than N1, suggesting that A175G leads not only to higher efficiency with splice site 448 but also to a destabilization of the beneficial effect of A369G with splice site 258. Third, clone C17 carried only the single mutation T317A, which caused a significant improvement on splice site 258 but no improvement on splice site 448. The T317A mutation also shows up in clones C32, N2, and N3 which all perform better on splice site 258 than W11. These three mutations are highlighted on the *Tetrahymena* Group I intron secondary structure (Figure 8A). The mutations, highlighted on 3D models of the *Tetrahymena* Group I intron, are all on the outside surface of the ribozyme suggesting possible interactions with outside factors (Figure 8B). The model suggests that mutations T317A and A369G may have an effect on the interaction with the intron portion close to the 3' splice site. For the other six clones the mutations were not related to other clones by point mutations and therefore did not allow ascribing the growth effects to individual mutations; this will require additional analysis (Figure 8C). Together, these results showed that spliceozyme mutations have specific improvements of bacterial growth on LBcam in the context of different substrate sequences.



(C)

	G51A	A70T	A90T	A94G	C98T	A118G	G147A	A175G	C337T	T241A	T249C	T292C	T317A	T333C	A343G	G357A	A369G	OD_{600} SS258 SiC8
N3	A	G								A	C	C	A	C				6.3
C32				G									A					4.4
C42									T			C						4.2
C17													A					4.1
N1																	G	4
C23			T				A								G	A		3.4
C11								G									G	3.4
N2					T	G						C	A	C				3.4
W11	G	A	A	A	C	A	G	A	C	T	T	T	T	T	A	G	A	1.8

	G51A	A70T	A90T	A94G	C98T	A118G	G147A	A175G	C337T	T241A	T249C	T292C	T317A	T333C	A343G	G357A	A369G	OD_{600} SS448 4C2
C11								G									G	8.4
C42									T			C						4.2
N3	A	G								A	C	C	A	C				3.9
C32				G									A					3.7
N2					T	G						C	A	C				3.3
N1																	G	1.6
C23			T				A								G	A		1.2
C17													A					0.7
W11	G	A	A	A	C	A	G	A	C	T	T	T	T	T	A	G	A	0.7

Figure 8. Mutations that appeared on clones in Round 9 of the evolution. The upstream and downstream exons (red) and the intron (blue) are depicted in the structures. **(A)** Mutations A369G (C11 and N1), A175G (C11), and T317A (C17) shown on secondary structure of spliceozyme (black). **(B)** Westhoff model of Tetrahymena Group I Intron (gray) with mutations highlighted for each clones N1 (A369G), C11 (A175G and A369G), and C17 (T317A). **(C)** All mutations that appeared in 8 clones tested in growth assays W11, as shown in Figure 7, arranged by growth on SS258 (top table) and SS448 (bottom table).

The difference in the effects of the spliceozyme mutations on different substrates could be due to (i) the difference between the two splice sites 258 and 448, (ii) due to the introns, or (iii) due to a combinatorial effect of splice sites and introns. To discriminate between these possibilities, we measured bacterial growth on LBcam with a combination of splice site and intron sequence that had been previously used in the evolution, SiC4 for splice site 258 and 3C29 for splice site 448, for all nine spliceozyme constructs. The results showed that there was no clear correlation between the splice sites when using different introns (Figure 9). For splice site 258, a trend line could be drawn between the growth on both introns, but clone C42 was a clear outlier to this trend. The clones generally grew better in both splice site 258 sequence contexts compared to the splice site 448 constructs. For splice site 448, two different trend lines could be drawn. All clones except W11 performed better on splice site 448 intron 4C2 than intron 3C29. The W11 results were very similar for both intron contexts, similar to the intron selection. The introns that were chosen from the splice site 448 intron selection may have been the best for W11, but not the other clones. More clones would need to be tested in order to see whether there is a significant correlation between splice site with different introns.

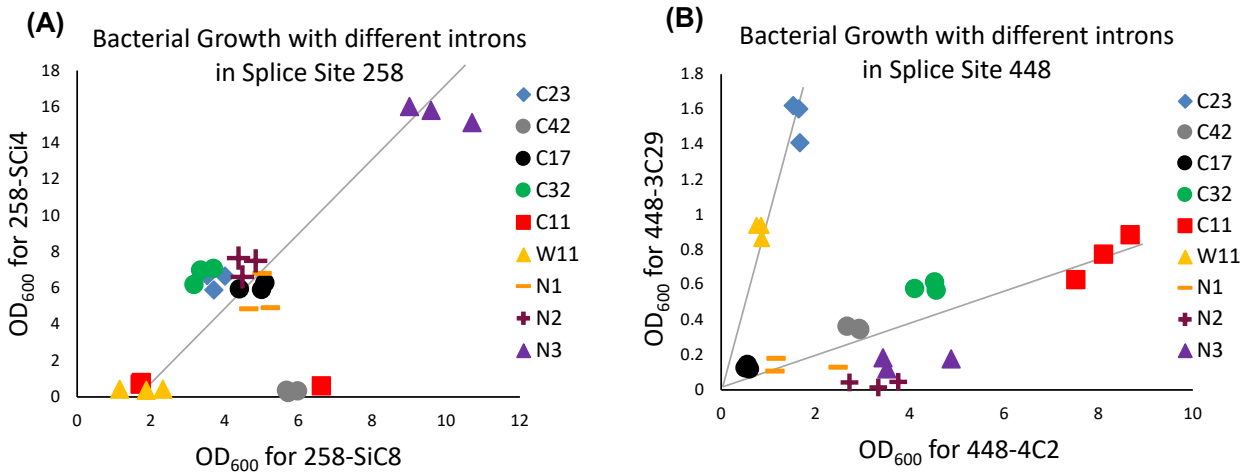
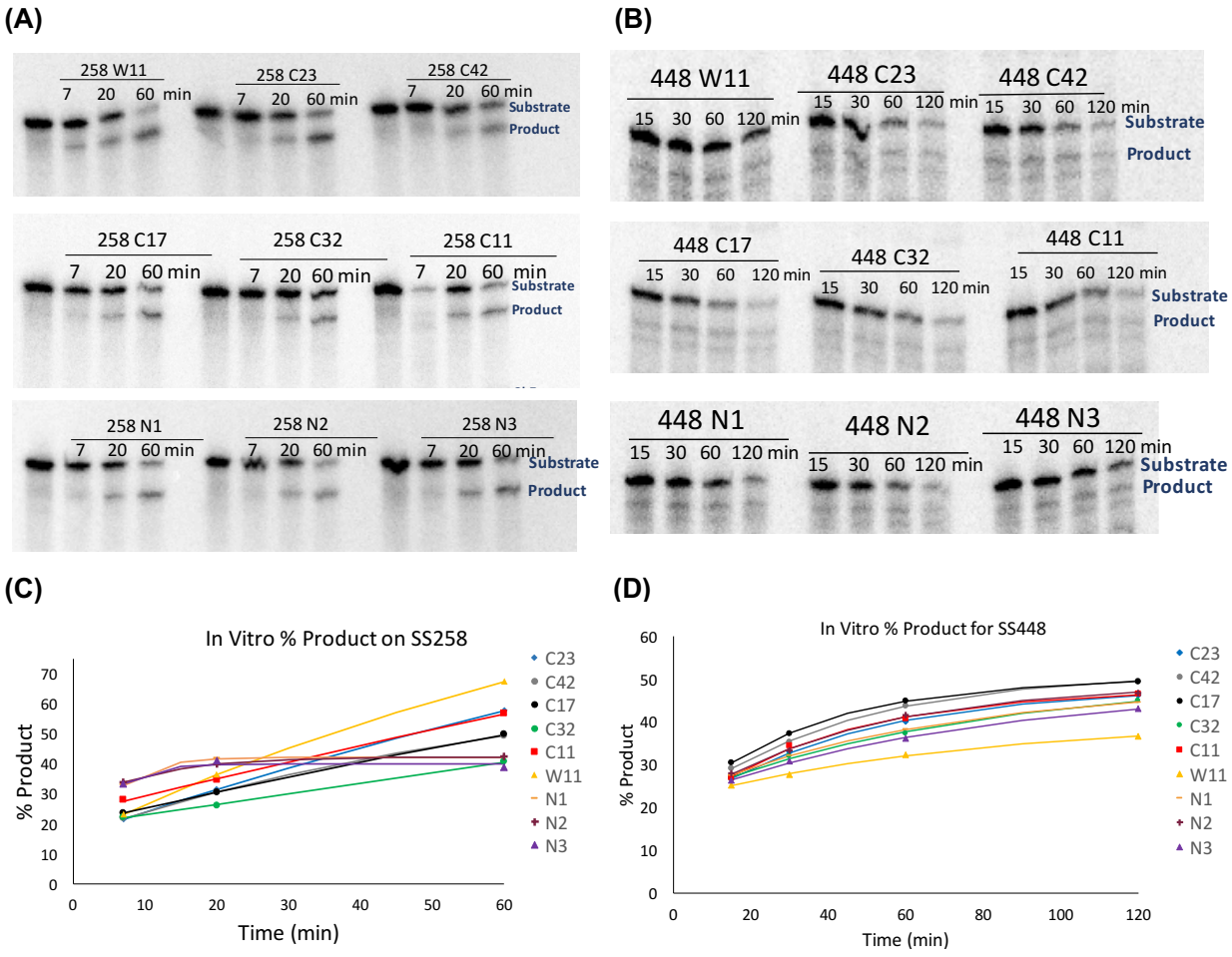


Figure 9. Growth of bacterial clones containing nine individual spliceozymes in two different sequence contexts with two different introns for each: Splice Site 258 with either introns Si4 and Si8 and Splice Site 448 with introns 3C29 and 4C2. The nine clones analyzed were clone W11, the starting point of the evolution, and eight clones that were isolated from round 9 of the evolution experiment. Growth on LB_{cam} plates was measured by growing the bacterial cells for 19 hours on an LB plate with a chloramphenicol concentration of 10 µg/mL, suspending the cells in LB medium, and measuring the OD_{600} compared to LB medium. **(A)** The growth (OD_{600}) in the sequence context of splice site 258 with intron Si8 is shown on the x-axis, the growth in the sequence context of splice site 258 with intron Si4 on the y-axis. The data from each clone are shown in colored symbols, with their legend on the right. **(B)** The growth (OD_{600}) in the sequence context of splice site 448 with intron 4C2 is shown on the x-axis, the growth in the sequence context of splice site 448 with intron 3C29 on the y-axis. The data from each clone are shown in colored symbols, with their legend on the right. The axis differ based on growth activity for each splice site and intron.

To test whether bacterial factors are necessary for the effects of the spliceozyme mutations in bacteria we performed *in vitro* assays of the spliceozyme variants on the two, respective substrate RNAs; the cat substrate with intron SiC8 at splice site 258 and the cat gene with intron 4C2 at splice site 448. In these assays, the same constructs were used as described previously for the *in vitro* characterization of the spliceozyme mutant W11 [26]. Substrate CAT pre-mRNA fragments are internally labelled with ^{32}P, allowing to follow the formation of products and side products after denaturing polyacrylamide gel electrophoresis (PAGE) and quantitative analysis of autoradiograms.



(E)

Rate constants (M^{-1} min^{-1})

	W11	C23	C42	C17	C32	C11	N1	N2	N3
SS 258	0.010	0.006	0.014	0.005	4.17×10^{-6}	0.003	0.216*	0.095	0.256*
SS 448	0.014	0.022	0.023	0.029	0.013	0.025	0.016	0.022	0.013

Figure 10. Substrate CAT pre-mRNA fragments are internally labelled with ^{32}P, allowing to follow the formation of products and side products after denaturing polyacrylamide gel electrophoresis (PAGE) and quantitative analysis of autoradiograms. **(A)** Time course of 8 spliceozymes clones and W11 reacting with CAT pre-mRNA at splice site 258. Samples were taken at 7,20, and 60 min. **(B)** Time course of 8 spliceozymes clones and W11 reacting with CAT pre-mRNA at splice site 448. Samples were taken at 15, 30, 60, and 120 min. **(C)** Plot of % product calculated from autoradiograms and against time in minutes for all 8 clones and W11 for splice site 258 gels. Markers are actual time points whereas lines are models based on calculated rates. **(D)** Plot of % product calculated from autoradiograms and against time (min) for all 8 clones and W11 for splice site 258 gels. Markers are actual time points values whereas lines are models based on calculated rates. **(E)** Rates calculated using a second order rate equation. Slower rates are shown in table. Values with "*"used a first order rate equation.

The results indicated that for splice site 258, none of the variants performed significantly better than W11 even after 60 min (Figure 10). The effects of N3 as seen in the *in vivo* growth

assays were not seen in the *in vitro* results Six variants and W11 used a second order reaction rate in order to describe the time course data. A first order reaction rate could only be applied to variants N1 and N3. Of those second order reaction rates, C42 was the fastest at 0.014, followed by W11 at 0.010, N2 at 0.095, C23 at 0.006, C17 at 0.005, C11 at 0.003, and C32 at 4.17 x 10^{-6} $M^{-1}min^{-1}$. For splice site 448, all the variants performed slightly better than W11 at 15, 30, 60 and 120 minutes. The effects of C11 as seen in the *in vivo* growth assays were not seen in the *in vitro* results. A second order rate reaction could be applied to the time points. Variant C17 had the fastest rate at 0.029, C11 at 0.025, C42 at 0.023, C23 and N2 at 0.022, N1 at 0.016, W11 at 0.014, and N3 and C32 at 0.013 $M^{-1}min^{-1}$.

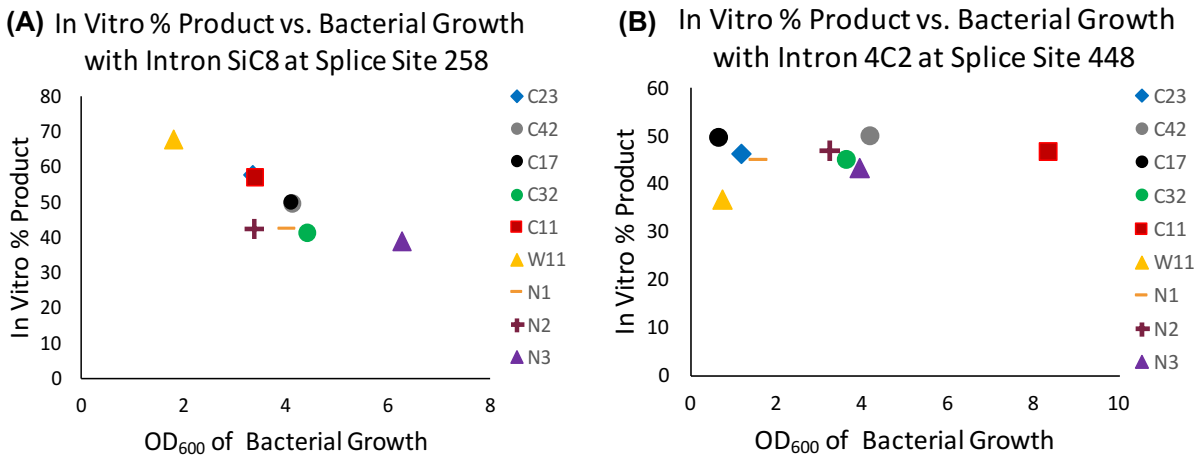


Figure 11. Correlation between *in vitro* % product and bacterial growth between same introns. **(A)** Data from bacterial growth assays plotted against data from % product quantified from autoradiograms of *in vitro* assays using intron SiC8 at splice site 258. **(B)** Data from bacterial growth assays plotted against data from % product quantified from autoradiograms of *in vitro* assays using intron 4C2 at splice site 448.

The *in vitro* assays with the spliceozyme RNA and CAT pre-mRNA and the *in vivo* growth assays showed no significant correlation. There are two possible interpretations of these results: Either bacterial factors are utilized by the spliceozymes, or the truncations of the substrate sequences to ~800 nucleotides prevented the replication of the mutation effects in vitro. This

question needs to be addressed in follow-up work, perhaps with the technically difficult use of the full-length CAT pre-mRNA of 1,070 nucleotides instead of the CAT pre-mRNA fragment.

Together, this study showed, based on an evolution experiment of spliceozymes in bacterial cells, that specific spliceozyme mutations are able to improve spliceozyme efficiency in bacterial cells and improve performance in difference substrate sequence contexts. This suggests that spliceozymes could be generated with improved efficiency on a wide range of substrate sequences with further analysis and experiments. It is currently unclear whether the evolved spliceozyme mutations require bacterial factors to exhibit their effects.

I would like to acknowledge Logan Norrell on his work in the results and materials and methods sections- the identification of promoter 5 and the first six rounds of the spliceozyme evolution. The results section was initially written by Ulrich Muller, including figures 2-5 and Table 1.

Discussion

This study showed, based on an evolution of spliceozymes in bacterial cells using lower expression levels and multiple splice sites, that specific spliceozyme mutations are able to improve spliceozyme efficiency in bacterial cells. Importantly, some mutations, and sets of mutations, improved the performance in several different substrate sequence contexts. This suggests that further analysis, and perhaps further evolutions, could generate spliceozymes with improved efficiency on a specific splice site or broad range of substrate sequences.

Evolutions could be done to generate spliceozymes which perform efficiently on an individual splice sites as was done for splice site 258 to generate W11 [26]. The spliceozyme went through 10 rounds of evolution to produce clone W11. In this study, an additional 6 rounds were completed with lower expression levels using clone W11 as a starting point. All 8 clones from Round 9 that were tested performed better than W11 in bacterial cells. If the selection pressure had been increased by lowering expression levels as well as increasing the chloramphenicol concentration during these 6 rounds, a more efficient clone may have been generated. The lower expression level alone was not enough selection pressure to produce a clone that was significantly more efficient. Similarly, in order to find a spliceozyme better suited for splice site 448, an evolution could be performed specifically for that splice site. Introns could be changed every two rounds to prevent intron sequence specificity and selection pressure could be increased the last three to four rounds of the evolution to generate a spliceozyme efficient for splice site 448.

In order to generate a clone that may be more general for different splice sites, W11 could be a starting point but splice sites could be changed more frequently earlier on in the evolution. Two rounds with promoter 5 for lower expression level could be done to make sure the spliceozyme mediates growth. Then, the splice site could be switched to 448 so that the clones do not become accustomed to splice site 258. The next best splice site, 405, could be added into the evolution as well [37]. Every two to three rounds the intron and splice site could be switched and the chloramphenicol concentration could be gradually increased to escalate the selection pressure. This way the spliceozyme does not get accustomed to a specific splice site or intron but has enough opportunity to gradually gain mutations that will increase efficiency across all three splice sites and introns. Overall the number of rounds needs to be increased and the splice sites need to be changed frequently.

In this study, mutations improved the performance of clones in bacterial cells but did not exhibit the same behavior *in vitro* (Figure 11). Clone C11, with mutations A369G and A175G, showed improvement on splice site 448 compared to W11 in bacterial cells but did not show a significant improvement in the *in vitro* assays. The growth of clone N1 (only has A369G mutation) with splice site 448 is within error of W11, suggesting that mutation A369G does not improve splicing with splice site 448. Mutation A175G can be examined more closely to determine its effect. An A175G mutation could be made using site directed mutagenesis on W11, and the growth on LBcam plates could be compared with clone C11 and clone W11. If this mutation causes more growth than W11 and similar growth to C11 on splice site 448, this would suggest that cellular factors may be an influence in the mutation effects. Both mutations on clone C11 are on the external portions of the *Tetrahymena* group I intron tertiary structure which could suggest interactions with cellular factors (Figure 8B). A protein pull down experiment and

further studies could be performed to determine which cellular factors are necessary and their effects.

It is currently unclear whether the evolved spliceozyme mutations require bacterial factors to exhibit their effects. The answer to this question will decide whether spliceozymes could be evolved in *E. coli* cells, or require the technically more difficult evolution in human cell lines for possible clinical applications. Nevertheless, the improved spliceozyme efficiency after evolution in bacterial cells suggests that it is only a technical question whether spliceozymes could be generated that are efficient enough for clinical applications.

Materials and Methods

Construction of spliceozyme/CAT pre-mRNA plasmids

The plasmid containing the spliceozyme gene and the CAT gene was based on previously described pUC19-based expression system, which includes a multiple cloning site with SacI, BamHI, SphI, and HindIII. The ribozyme gene was inserted between restriction sites SacI and BamHI and the CAT gene was inserted between SphI and HindIII as described in [22]. To prepare for the insertion of introns at position 448, a restriction enzyme site was introduced using a T423 mutation to create a PciI restriction site through PCR mutagenesis. Another PciI site in the origin gene was silently mutated through PCR mutagenesis so only one unique PciI site would remain in the plasmid.

Selection of Weak Promoter Sequences that still mediate growth

Five different promoter sequences derived from LacI operator, termed single-input gates (SIGs), were reported in [37]. These promoters were identified as B4, A81, A38, A87, and A52 with regulatory ranges given as 2.5×10^4, 1.5×10^4, 3.6×10^3, 2.6×10^2, and 6.2×10^2, respectively. W11 spliceozymes were PCR amplified with primers adding one of the five LacI promoters at the 5' end into pUC19b carrying a chloramphenicol gene with intron SiC10 at splice site 258.

After verification of the correct promoter sequences, growth curves for each LacI-derived promoter were performed in triplicate. An overnight culture of 5 mL LB with 100 μg/mL ampicillin was inoculated with a single colony and the next morning 5 μL of 1 M IPTG was

added to activate the LacI-derived promoters and downstream W11 spliceozymes for 1 h at 37°C with shaking. This induced starter culture was used to inoculate a 20 mL LB culture with 4 μg/mL chloramphenicol and 1 mM IPTG to a starting OD_{600} of 0.05 ± 0.01.

Further reduction of expression of the LacI A87 promoter was accomplished in four steps. First, four nucleotide randomized libraries flanking the 5'-end of the -35 box, the 3'-end of the -35 box, and the 5'-end of the -10 box were constructed using the following 5' primers and the 3' primer in purple for PCR. Four 5' primers were used including 5' GCTGCCTTTCGTCTT-CAATANNNNTTGACAATCAATGTGGATTTCTGATCTGTGTGTCTG, 5'GCTGCCTTTC-GTCTTCAATAATTCTTGACANNNNATGTGGATTTTCTGATACTGTGTGGCT, 5' GCT-GCCTTTCGTCTTCAATAATTCTTGACAATCAATGTGGATTNNNNGATACTGTGTGGCT G, 5' ACTGGGATCCGTACAACGTCGTCTTAGCTGCCTTTCGTCTTCAATA. The 3' primer used for PCR was 5' CGGAGCTCCAAAAAACCCTCAAGACCCGTTTAGAGGCCCCAAG-GGGTTATG. These N_4 libraries were analyzed by the same growth assay method and A87 3.1a was shown to have the slowest rate of growth.

Second, to assist cloning during the *in vivo* evolution, a KpnI site was added to A87 3.1 to produce A87.1. Third, another N_4 randomized library, A87.2 library, was constructed, this time within most nucleotides of the 5'-10 box. The weakest promoters were first selected by a growth assay in LB containing 4 μg/mL chloramphenicol and 1 mM IPTG. Fourth, Candidate clones were analyzed by RT-qPCR to determine levels of spliceozyme expression. To quantify the amount of spliceozymes being produced per cell, total RNA was isolated from logarithmically growing *E. coli* cells after induction with 1 mM IPTG, using the RNeasy kit (Qiagen). Two hundred nanograms of total RNA was reverse-transcribed using SuperScript III RT (Invitrogen) and a short oligonucleotide complementary to the 3'-end of the spliceozyme.

One-tenth of the reverse transcription reaction was used as template for the quantitative PCR. The PCR was performed using two sets of primers with the SYBR green master mix (Thermo-Fisher) on a Fast 7500 RT-PCR machine (Applied biosystems). The method started with a 10-min incubation at 95 °C. Each PCR cycle included incubations of 30 sec at 95°C, 30 sec at 57°C, and 30 sec at 72°C. Each PCR cycle included incubations of 30 sec numbers from three eightfold serially diluted samples confirmed the linearity of the assay.

Selection of efficient intron sequences for Splice Site 448

Plasmid libraries containing partially randomized introns were transformed into electrocompetent *E. coli* cells, which were plated on LB-agar containing 100 mg/mL ampicillin. The plates were washed with LB medium, after incubation at 37°C for 19 hours, and the bacterial libraries were frozen as glycerol cell stocks. The bacterial libraries were then diluted to approximately 3,000,000 cfu/mL with LB medium and induced with 1mM IPTG by shaking for 1 hour at 37°C. 100μL of this culture were plated onto each of 20 LB-agar plates containing 4μg/mL chloramphenicol, and allowed to grow for 17 hours at 37°C for the selective step. The plates were then washed with LB medium to create glycerol cell stocks. The intron sequences were isolated, amplified, and ligated back into fresh plasmids to create a library for the next round. These steps were completed again 4 times for a total of 4 rounds of selection.

Individual colonies from Rounds 3 and 4 were selected and analyzed for the sequences and efficiencies of individual introns. Each clone was transformed into electrocompetent *E. coli* cells, and plated on LB-agar containing 100 mg/mL ampicillin. After incubation at 37°C for 16 hours, the plates were washed with LB medium to generate a cell stock of each clone. Each clone was diluted to 2000 cfu/5μL. After inducing each dilution with 1mM IPTG by shaking at 37°C,

5μL of each clone was plated on a series dilution of LB-agar plates with either 2, 4, 6, or 8 μg/mL chloramphenicol and allowed to grow for 19 hours at 37°C. Each clone's growth was independently visually analyzed by two people and given a rating, 0 (no growth) to 100 (most growth across all plates). Both the plating and the analysis were done in triplicate. The ratings were plotted against chloramphenicol concentration in order to determine the efficiency of the introns. Quantitative measurements were taken by pipetting 10μl of fresh LB up and down on the 5μL growth spots and measuring the OD_{600}. This was not only laborious and time consuming but also inaccurate due to artifacts from the LB agar being added to rather than just the colonies themselves.

Spliceozyme Evolution

Rounds 1-6:

The evolution was completed essentially as described [24]. Briefly, spliceozyme constructs were randomized by mutagenic PCR and cloned into library plasmid. Library plasmids containing spliceozyme sequences were transformed into *E. coli* cells, plated on LB agar medium containing 100 μg/mL ampicillin (amp) and incubated at 37°C for 16 h to create pools. LB agar plates containing pools of *E. coli* cells were washed using liquid LB medium. The resulting medium was diluted to an OD_{600} of 0.0015 or 300,000 cfu/mL, induced with IPTG to a final concentration of 1 mM and shaken at 37°C for 1 h. *E. coli* cells were then plated on LB agar medium containing the indicated concentration of chloramphenicol (Table 1) and incubated at 37°C for 16 h to select for spliceozyme constructs able to mediate antibiotic resistance. LB agar plates were washed using liquid LB medium and plasmids were isolated by miniprep. Five or 10 clones were chosen for sequencing from each round of evolution.

Rounds 7-10:

The plasmid 448-3C29, 258-SiC4, and 448-4C6 were prepared for rounds 7,8, and 9 respectively. 5μg of each plasmid was digested at the KpnI and SacI sites and dephosphorylated.1nM of the R6 spliceozyme pool was amplified with 5' CAATGTGGTACCA-CAAAGCCTGTGTGGCTGGGGATAAAAGTTATTAGGC for Splice site 448 or 5' CAATGT-GGTACCACAAAGCCTGTGTGGCTGTCCCAAAAAGTTATTAGGC for splice site 258 and 3' primer GCCAGTGAATTCGAGCTC for 25 cycles. The pool was digested using KpnI and SacI at 30°C and the spliceozyme was ligated to the plasmid used for each round. The plasmid with spliceozyme insert was transformed into commercially available electrocompetant cells and plated LB-Agar plates containing 100 mg/mL ampicillin. After growing overnight, each plate was washed with 2mL of LB. An equal volume of glycerol was added to create a cell stock. The OD_{600} of the cell stocks were measured and then one cell stock was diluted to 3,000,000 cfu/mL. The dilutions were incubated at 37°C with a final concentration of 1mM IPTG for 1 hour with shaking. 100 μL was plated onto each LB-Agar plate containing 4 μg/mL chloramphenicol for Round 7. After growing for 20 hours. Each plate was washed with 3 mL of LB and centrifuged for 5 min at 5000g. The supernatant was removed and 5mL of fresh LB was added. This plasmid pool was miniprepped and named the R7 plasmid pool. For cycles 8 and 9, 1nM of the R7 and R8 pools respectively were amplified with the same primers described for 25 cycles. The products were run on a 1% agarose gel the correct sized band (536nt) was cut out to avoid any contaminating DNA. The gel piece was added to 200 μL of water and heated at 90°C for 5 min to melt the gel. 1μL of the melted gel was used for another 25 cycles of 100 μL PCR reaction with the same primers. The amplified spliceozyme was ligated to the respective cut and dephosphorylated plasmid for each round, transformed into electrocompetant cells, plated on

LBagar with 100 mg/mL ampicillin and made intro glycerol stocks. The stocks were then diluted to 3,000,000 cfu/mL and incubated with a final concentration of 1mM IPTG at 37°C for 1 hour with shaking to mediate antibiotic resistance. 100 μL were plated onto 5 plates of LB agar with either 4, 8, 12, 20, or 40 μg/mL chloramphenicol. After waiting 19 hours, the plates were compared and LB-agar plates with 8 μg/mL chloramphenicol were chosen to be washed four Round 8 and the 12 μg/mL for Round 9. These plates did not have a lawn of *E. coli* but mediated enough growth that there were around 1000's cfu/plate. Each plate was washed with 3 mL of LB and centrifuged for 5 min at 5000g. The supernatant was removed and 5mL of fresh LB was added. This plasmid pool was miniprepped.

Notes: Contamination was seen during this evolution and therefore steps were added to complete Rounds 8 and 9 of the evolution. The plasmids from the previous round were still appearing so that we had different sets of introns in our pool. This could be due to (1) contamination on the spreaders (2) Complete plasmid being replicated during PCR. To avoid contaminants from the spreaders: the spreaders were placed in bleach for a minimum of 15 minutes, washed with water, placed in a flame, and then left to cool before use. To avoid contamination from PCR, a step was added into the evolution cycle. Instead of 1 PCR step, the product was run through a 1% agarose gel, the correct sized band was excised, PCR was performed with this product and the evolution was continued. This step ensures that only the spliceozyme fragment was amplified.

Growth measurements in *E. coli*

Spliceozyme constructs were inserted into library plasmids: 258-SiC4, 258-SiC8, 448-3C29, and 448-4C2. Library plasmids containing spliceozyme sequences were transformed into

E. coli cells, plated on LB agar medium containing 100 μg/mL ampicillin (amp) and incubated at 37°C for 16 hours. A clone was selected off the plate and grown in a liquid culture containing100 μg/mL amp overnight. The OD_{600} of the overnight culture was measured and the culture was diluted to 50,000 cfu/mL and incubated with a final concentration of 1mM IPTG at 37°C for 1 hour with shaking to mediate antibiotic resistance. 100 μL was plated on a LB agar plate with 10 μg/mL chloramphenicol and grown for 19 hours at 37°C. The plates were then washed with 2mL of fresh LB and the OD_{600} was measured. This was done in independent triplicated experiments.

Spliceozymes and pre-mRNAs for in vitro work

The Spliceozymes and the CAT pre-mRNA substrate were made by in vitro transcription from PCR products with T7 RNA polymerase, as described in [37]. The PCR of the spliceozymes used the 5' primer **AATT**TAATACGACTCACTATTAAGTCCCAAAAAGTTA-TTAGG for splice site 258 and **AATT**TAATACGACTCACTATTAAGGGGATAAAAGTTA-TTAGG for splice site 448 (the T7 promoter is underlined and T7 enhancer is in bold). The PCR of the spliceozyme used the 3' primer TATGCGATTAGTTCCAGCGGC but clones that had a A369G mutation used the 3' primer TATGCGATTAG**C**TCCAGCGGC. PCR amplification of the template for T7 transcription of CAT pre- mRNA, the 5' primer **AATT**TAATACGACTCA-CTATAGGGAGGAGCTAAGGAAGCTAAAATG and the 3' primer CGCCCCGCCCTGCCA-CTCATC were used (the T7 promoter is underlined and T7 enhancer is in bold). Transcriptions were performed at 37°C for 3 hours in 40 mM Tris/HCl pH 7.9, 26 mM $MgCl_2$, 5 mM DTT, 0.01% Triton X-100, 2.5 mM spermidine and 2.5 mM of each NTP. The CAT pre-mRNA substrates were internally labelled using additional [^{32}P]-ATP and a decreased concentration of

1mM ATP. Transcribed spliceozymes and substrate CAT pre-mRNAs were purified by 7 M urea 5% polyacrylamide gel electrophoresis (PAGE), eluted in 300 mM NaCl, ethanol precipitated, re-dissolved in water. RNA concentrations were measured using the A_{260}.

In vitro reactions

In vitro reactions were done essentially as described in [22]. The final spliceozyme concentration was 300 nM and CAT pre-mRNA substrate concentrations were 100 nM. Reactions were incubated in 5 mM $MgCl_2$, 135 mM KCl, 50 mM MOPS/NaOH pH 7.0, 20 mM GTP, and 2 mM spermidine, at 37°C, in reaction volumes of 20 μL. Spliceozyme and substrate were pre-incubated separately in the reaction buffer for 10 minutes at 37°C before combining the solutions to start the reaction. Samples were taken in 2 μL aliquots after reaction times of 7, 20, and 60 minutes for spliceozymes being tested with Splice Site 258 and 15,30,60, and 120 minutes for clones being tested with Splice Site 448. In vitro reaction products were separated on 7 M urea 5% PAGE and visualized by phosphor-imaging. Bands were quantified on a phosphorimager (PMI, Bio-Rad) using the Bio Rad software Image Lab.

I would like to acknowledge Logan Norrell on his work in the materials and methods sections- the identification of promoter 5 and the first six rounds of the spliceozyme evolution.

References

1. Kruger, K.; Grabowski, P. J.; Zaug, A. J.; Sands, J.; Gottschling, D. E.; Cech, T. R., Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell* **1982,** *31* (1), 147-57.

2. Guerrier-Takada, C.; Gardiner, K.; Marsh, T.; Pace, N.; Altman, S., The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **1983,** *35*, 849-57.

3. Cech, T. R., Self-splicing of group I introns. *Annu Rev Biochem* **1990,** *59*, 543-68.

4. Cech, T. R.; Damberger, S. H.; Gutell, R. R., Representation of the secondary and tertiary structure of group I introns. *Nat Struct Biol* **1994,** *1* (5), 273-80.

5. Stahley, M. R.; Strobel, S. A., RNA splicing: group I intron crystal structures reveal the basis of splice site selection and metal ion catalysis. *Curr Opin Struct Biol* **2006,** *16* (3), 319-26.

6. Adams, P. L.; Stahley, M. R.; Kosek, A. B.; Wang, J.; Strobel, S. A., Crystal structure of a self-splicing group I intron with both exons. *Nature* **2004,** *430* (6995), 45-50.

7. Guo, F.; Gooding, A. R.; Cech, T. R., Structure of the Tetrahymena ribozyme: base triple sandwich and metal ion at the active site. *Mol Cell* **2004,** *16* (3), 351-62.

8. Golden, B. L.; Kim, H.; Chase, E., Crystal structure of a phage Twort group I ribozyme-product complex. *Nat Struct Mol Biol* **2005,** *12* (1), 82-9.

9. Lipchock, S. V.; Strobel, S. A., A relaxed active site after exon ligation by the group I intron. *Proc Natl Acad Sci U S A* **2008,** *105* (15), 5699-704.

10. Michel, F.; Westhof, E., Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J Mol Biol* **1990,** *216* (3), 585-610.

11. Dotson, P. P.; Sinha, J.; Testa, S. M., A Pneumocystis carinii group I intron-derived ribozyme utilizes an endogenous guanosine as the first reaction step nucleophile in the *trans* excision-splicing reaction. *Biochemistry* **2008,** *47* (16), 4780-7.

12. Dotson, P. P.; Johnson, A. K.; Testa, S. M., Tetrahymena thermophila and Candida albicans group I intron-derived ribozymes can catalyze the trans-excision-splicing reaction. *Nucleic Acids Res* **2008,** *36* (16), 5281-9.

13. Fiskaa, T.; Lundblad, E. W.; Henriksen, J. R.; Johansen, S. D.; Einvik, C., RNA reprogramming of alpha-mannosidase mRNA sequences in vitro by myxomycete group IC1 and IE ribozymes. *FEBS J* **2006,** *273* (12), 2789-800.

14. Dolan, G. F.; Müller, U. F., Trans-splicing with the group I intron ribozyme from Azoarcus. *RNA* **2014,** *20* (2), 202-13.

15. Sullenger, B. A.; Cech, T. R., Ribozyme-mediated repair of defective mRNA by targeted, trans-splicing. *Nature* **1994,** *371* (6498), 619-22.

16. Inoue, T.; Sullivan, F. X.; Cech, T. R., Intermolecular exon ligation of the rRNA precursor of Tetrahymena: oligonucleotides can function as 5' exons. *Cell* **1985,** *43*, 431-7.

17. Alexander, R. C.; Baum, D. A.; Testa, S. M., 5' transcript replacement in vitro catalyzed by a group I intron-derived ribozyme. *Biochemistry* **2005,** *44* (21), 7796-804.

18. Bell, M. A.; Johnson, A. K.; Testa, S. M., Ribozyme-catalyzed excision of targeted sequences from within RNAs. *Biochemistry* **2002,** *41* (51), 15327-33.

19. Bell, M. A.; Sinha, J.; Johnson, A. K.; Testa, S. M., Enhancing the second step of the *trans* excision-splicing reaction of a group I ribozyme by exploiting P9.0 and P10 for intermolecular recognition. *Biochemistry* **2004,** *43* (14), 4323-31.

20. Baum, D. A.; Testa, S. M., *In vivo* excision of a single targeted nucleotide from an mRNA by a *trans* excision-splicing ribozyme. *RNA* **2005,** *11* (6), 897-905

21. Johnson, A. K.; Sinha, J.; Testa, S. M., Trans insertion-splicing: ribozyme-catalyzed insertion of targeted sequences into RNAs. *Biochemistry* **2005,** *44* (31), 10702-10.

22. Amini, Z. N.; Olson, K. E.; Müller, U. F., Spliceozymes: ribozymes that remove introns from pre-mRNAs in *trans*. *PLoS One* **2014,** *9* (7), e101932.

23. Meluzzi, D.; Olson, K. E.; Dolan, G. F.; Arya, G.; Müller, U. F., Computational prediction of efficient splice sites for trans-splicing ribozymes. *RNA* **2012,** *18* (3), 590-602.

24. Olson, K. E.; Dolan, G. F.; Müller, U. F., *In vivo* evolution of a catalytic RNA couples trans-splicing to translation. *PLoS One* **2014,** *9* (1), e86473.

25. Olson, K. E.; Müller, U. F., An *in vivo* selection method to optimize trans-splicing ribozymes. *RNA* **2012,** *18* (3), 581-9.

26. Amini, Z. N.; Müller, U. F., Increased efficiency of evolved group I intron spliceozymes by decreased side product formation. *RNA* **2015,** *21* (8), 1480-9.

27. Ayre, B. G.; Köhler, U.; Goodman, H. M.; Haseloff, J., Design of highly specific cytotoxins by using trans-splicing ribozymes. *Proc Natl Acad Sci U S A* **1999,** *96* (7), 3507-12.

28. Kwon, B. S.; Jung, H. S.; Song, M. S.; Cho, K. S.; Kim, S. C.; Kimm, K.; Jeong, J. S.; Kim, I. H.; Lee, S. W., Specific regression of human cancer cells by ribozyme-

mediated targeted replacement of tumor-specific transcript. *Mol Ther* **2005,** *12* (5), 824-34.

29. Kim, Y. H.; Kim, K. T.; Lee, S. J.; Hong, S. H.; Moon, J. Y.; Yoon, E. K.; Kim, S.; Kim, E. O.; Kang, S. H.; Kim, S. K.; Choi, S. I.; Goh, S. H.; Kim, D.; Lee, S. W.; Ju, M. H.; Jeong, J. S.; Kim, I. H., Image-aided Suicide Gene Therapy Utilizing Multifunctional hTERT-targeting Adenovirus for Clinical Translation in Hepatocellular Carcinoma. *Theranostics* **2016,** *6* (3), 357-68.

30. Tazi, J.; Bakkour, N.; Stamm, S., Alternative splicing and disease. *Biochim Biophys Acta* **2009,** *1792* (1), 14-26.

31. Bhasi, A.; Pandey, R. V.; Utharasamy, S. P.; Senapathy, P., EuSplice: a unified resource for the analysis of splice signals and alternative splicing in eukaryotic genes. *Bioinformatics* **2007,** *23* (14), 1815-23.

32. Zhong, X.; Liu, J. R.; Kyle, J. W.; Hanck, D. A.; Agnew, W. S., A profile of alternative RNA splicing and transcript variation of CACNA1H, a human T-channel gene candidate for idiopathic generalized epilepsies. *Hum Mol Genet* **2006,** *15* (9), 1497-512.

33. Lacroix, M.; Lacaze-Buzy, L.; Furio, L.; Tron, E.; Valari, M.; Van der Wier, G.; Bodemer, C.; Bygum, A.; Bursztejn, A. C.; Gaitanis, G.; Paradisi, M.; Stratigos, A.; Weibel, L.; Deraison, C.; Hovnanian, A., Clinical expression and new SPINK5 splicing defects in Netherton syndrome: unmasking a frequent founder synonymous mutation and unconventional intronic mutations. *J Invest Dermatol* **2012,** *132*, 575-82.

34. Thomas, C. E.; Ehrhardt, A.; Kay, M. A., Progress and problems with the use of viral vectors for gene therapy. *Nat Rev Genet* **2003,** *4* (5), 346-58.

35. Votteler, J.; Ogohara, C.; Yi, S.; Hsia, Y.; Nattermann, U.; Belnap, D. M.; King, N. P.; Sundquist, W. I., Designed proteins induce the formation of nanocage-containing extracellular vesicles. *Nature* **2016,** *540* (7632), 292-295.

36. Bai, Y.; Gong, H.; Li, H.; Vu, G. P.; Lu, S.; Liu, F., Oral delivery of RNase P ribozymes by Salmonella inhibits viral infection in mice. *Proc Natl Acad Sci U S A* **2011,** *108* (8), 3222-7.

37. Cox, R. S.; Surette, M. G.; Elowitz, M. B., Programming gene expression with combinatorial promoters. *Mol Syst Biol* **2007,** *3*, 145.

38. Cadwell, R. C.; Joyce, G. F., Randomization of genes by PCR mutagenesis. *PCR Methods Appl* **1992,** *2* (1), 28-33.

39. Soll, S. J.; Díaz Arenas, C.; Lehman, N., Accumulation of deleterious mutations in small abiotic populations of RNA. *Genetics* **2007,** *175* (1), 267-75