Chapter 7

Ribozyme-mediated engineering of circular mRNA and its functional *in vivo* and *in vitro* translation

Van Lieshout, J.F.T., Vroom, W., De Vos, W.M., Van der Oost, J.

submitted

Abstract

The instability of prokaryotic mRNA in some instances may result in a decreased overall yield of gene expression. Hence, better insight in the features that determine mRNA stability, and the ability to engineer stable messenger variants, might contribute to optimized production of certain proteins. The aim of the present study was to explore novel strategies for stabilization of mRNA by cyclization. A series of pET9d-derived plasmids has been generated, including a reporter (*Pyrococcus furiosus* β-glycosidase, CelB) and a ribozyme (*Tetrahymena thermophila* self-splicing intron), and used to transform the LacZ-deficient *Escherichia coli* JM109(DE3). The design of the constructs allowed for a simple *in vivo* selection for ribozyme-catalyzed mRNA processing, as the cyclization of the RNA molecule resulted in re-ligation of truncated *celB*-fragments. Subsequent translation of this cyclic mRNA resulted in a functional β-glycosidase enzyme, as was judged by blue staining after growth on LB agar supplemented with X-Gal. In addition, β-glycosidase activity was demonstrated in lysates of liquid cultures of the selected recombinants. Moreover, analysis of total RNA of these recombinants by RT-PCR confirmed the presence of covalently closed circular mRNA. Also transcription *in vitro* resulted in cyclic mRNA that showed resistance to degradation by exonuclease. Several variant constructs have been designed aiming at optimal splicing, ligation, and subsequent transformation of the cyclic transcript. Although the present cyclic variants did not give rise to an enhanced protein production compared to the linear construct, this engineering exercise showed potential applicability for its improvement of mRNA stability.

Introduction

The *Tetrahymena thermophila* self-splicing intron is inserted in the 26S ribosomal RNA (rRNA) gene of this unicellular eukaryote, where it autocatalyses its excision from the primary rRNA transcript. This ribozyme has been classified as a Group I intron, based on its conserved sequence elements that form distinctive secondary and tertiary structures by intramolecular Watson-Crick base-pairing (Fig. 7.1). In general, Group I introns catalyse self-splicing from their primary transcript in two sequential trans-esterification reactions, resulting in excision and circularization of the intron fragment, and ligation of the exon fragments (Fig. 7.2A) [1].

Thorough analysis has identified the essential sequence elements required for splicing and accurate recognition of the splice site by the intron. These elements include the 5' region of the intron, the internal guide sequence (IGS) which pairs with the 5' exon to form the Paired region 1 (P1) [2, 3]. This structure is needed for recognition of the conserved 5' exon sequence. A GU wobble base pair at the 5' splice site (Fig. 7.1B) determines the exact location of the cleavage site and is required for the initiation of the splicing reaction [4]. The sole requirements for splicing appear to be the presence of both monovalent and divalent cations (usually Mg^{2+}) and of guanosine or one of its 5′-phosphorylated forms (GMP, GDP, or GTP) [1]. These characteristics allow efficient splicing *in vivo* in *Escherichia coli* as well as *in vitro*.

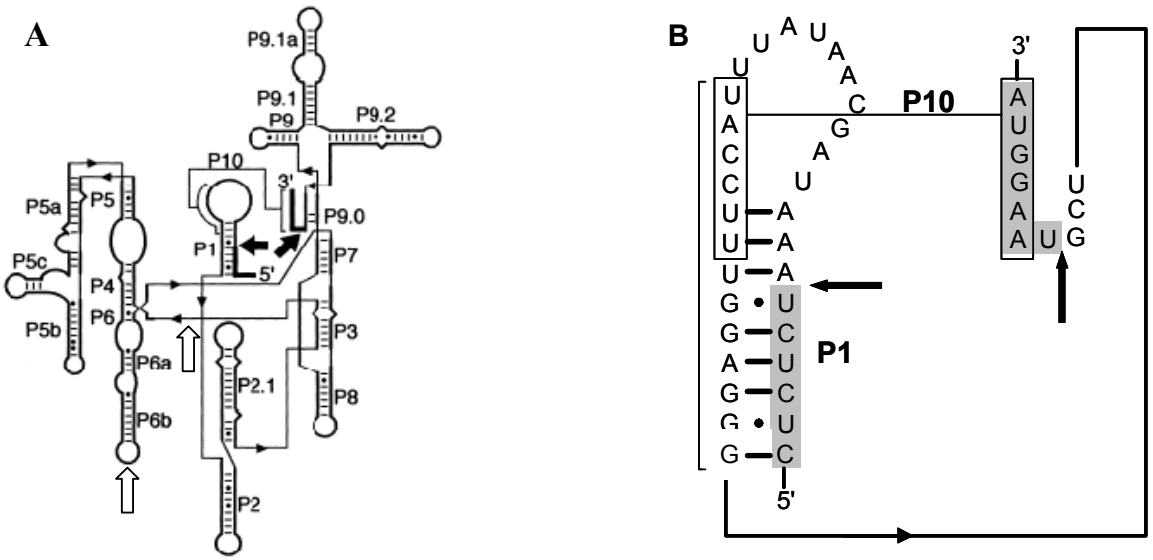

Figure 7.1 Secondary structure of the ribozyme from *T. thermophila*. **(A)** Open arrows indicate truncation sites for constructs pWUR96 and pWUR97. This figure is reprinted, with modifications, with permission from [5]. **(B)** A close-up of the P1 and P9 stem loops. Closed arrows indicate the 5' and 3' splice sites. Exon nucleotides are indicated in grey. The boxed nucleotides align to form the P10 helix, required in the second step of the splicing reaction. Adapted from [6], with permission.

Previous experiments with the Group I intron of the bacteriophage T4 (td intron) [7, 8] and the Group I introns of *Anabaena* [9, 10] have shown the possibility to divide the ribozyme into two parts and to rearrange the exon and intron sequences in such a way that the 3'-portions of both ribozyme and exon is followed by the 5' halves of exon and ribozyme (Fig. 7.2B). Such a DNA construct has been shown to yield a circular exon RNA after splicing both *in vitro* and *in vivo*. Furthermore, it has been demonstrated that a circular GFP-encoding transcript, including the sequential elements needed for translation initiation (RBS, ribosome binding site; DB, downstream box), can be effectively translated in *E. coli*, yielding the properly folded GFP [8]. An interesting observation was the apparent accumulation of circular mRNA, suggesting enhanced messenger stability, most likely due to increased exonuclease resistance [8]. In the same study, deletion of the stop-codon resulted in the "non-stop translation" of extremely long poly-proteins from circular messengers, proving the ability of the *E. coli* ribosomes to bind the circular RNA, and to continue translation for at least ten rounds. Unfortunately, translation efficiency of these circular mRNAs proved to be rather low, 20% compared to translation of their linear equivalent. Considering that circular messengers were found to accumulate *in vivo* [8, 11], the translational efficiency per RNA molecule may be even less.

In this study, the described techniques have been applied using the group I intron from *Tetrahymena thermophila* for generating circular transcripts of a β-glycosidase reporter protein. Several unprecedented designs have been made with the goal to establish an efficient system for *in vivo* and *in vitro* translation of circular mRNA. Furthermore, the increased stability of circular mRNA towards degradation by exonucleases has been analyzed.

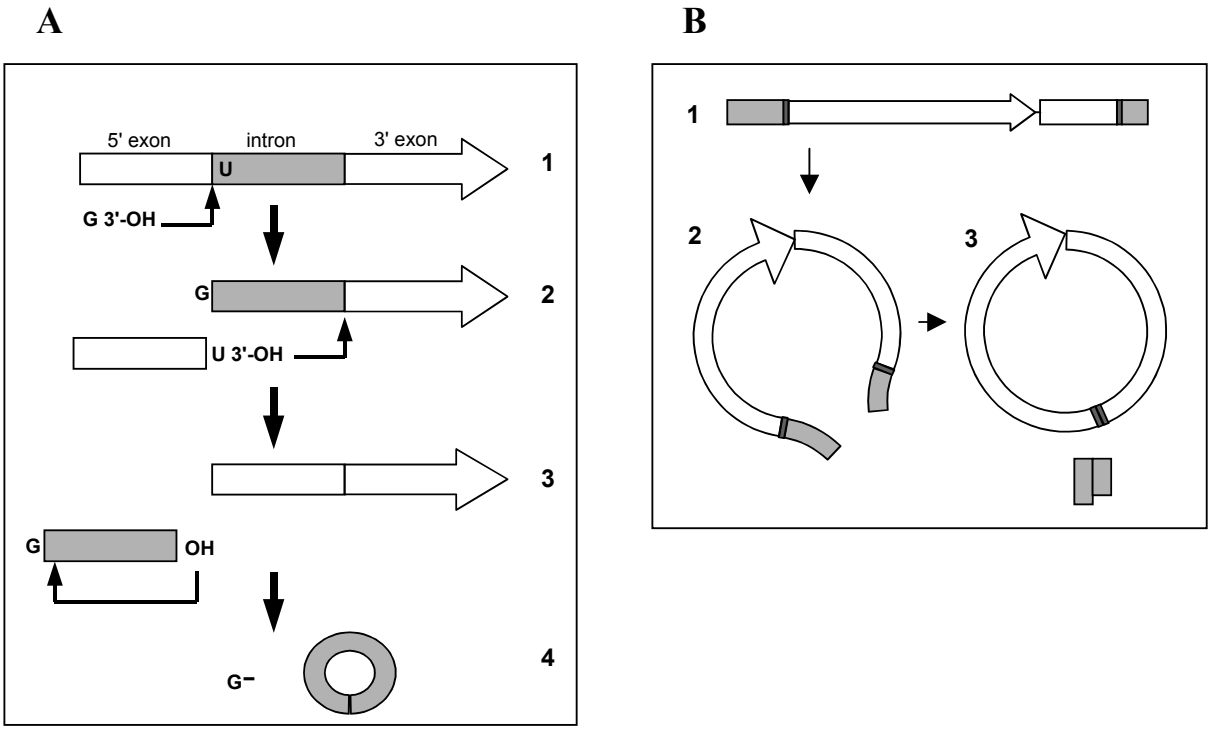

Figure 7.2 Schematic representation of the splicing mechanism of a Group I intron **(A)** and after engeneric rearrangement, yielding a circular exon **(B)**. Exon sequence is drawn white, intron sequence light grey. The dark grey fragments are the mutated exon boundaries in the *celB* gene. **(A)** The reaction is initiated by the nucleophilic attack of a 3' hydroxyl of a guanosine cofactor at the 5' splice site (**1**). The exon-intron phosphodiester bond is cleaved and the guanosine forms a 3',5'-phosphodiester bond at the 5' end of the intron. The liberated 3' hydroxyl of the 5' exon then makes a nucleophilic attack at the 3' splice site to form the ligated exons and release the intron with the non-encoded guanosine (**2**). Nucleotides close tot the 5' end of the intron are realigned on the "internal guide sequence" (IGS) and the highly conserved 3' terminal guanosine of the intron makes a nucleophilic attack at a phospho-diester bond between nucleotides 15 and 16 or between nucleotides 19 and 20 within the intron in a reaction that is analogous to the first step of splicing (**3**). The intron is circularised (**4**) and a small fragment containing the non encoded guanosine is released (adapted from [5]). **(B) 1** represents the transcribed, unspliced messenger. **2** shows both parts of the intron realigned *in vivo*. Finally, the splicing reaction yields a ligated, circular exon and a complex of two intron fragments (**3**).

Materials and Methods

Bacterial hosts and vectors - The T7 expression vector pET9d was obtained from Novagen. *Escherichia coli* JM109(DE3) (Stratagene) was used as an expression host for the pET-derivatives. *E. coli* was grown in TY medium in a rotary shaker at 37°C. Kanamycin was added to a final concentration of 30 μg/ml.

Cloning and expression - The *celB* gene coding for *Pyrococcus furiosus* β-glycosidase previously cloned in the pET9d-derived pLUW511 [12] was used as a template in PCR reactions. The intron has been PCR-amplified from *T. thermophila* genomic DNA (a kind gift from Dr. Hackstein, Radboud University Nijmegen). The primers that were used for the different constructs are listed in Table 7.1. All mutations and permutations were generated using overlap extension PCR [13]. The generated set of permuted genes was then ligated into pET9d after digestion with *Nco*I and *Bam*HI,

Table 1. Primers used in this study

[a]s = sense, a = antisense

[b]sequences are given from 5'-3'

primer[a]	primer sequence[b]	Description
BG64 a	TACGGAAATTATGTTGCC	anneals at position 235 in *celB*
BG238 s	GCGCGCCATGGCAAAGTTCCCAAAAAACTTCATGTTTG	forward primer *celB*
BG239 a	CGCGCGGATCCCTACTTTCTTGTAACAAATTTGAGG	reverse primer *celB*
BG1372 s	CTCTCTAAATAGCAATATTTACCTTTGG	forward primer *Tth* intron
BG1373 a	TACCTTACGAGTACTCCAAAACTAATCAATATAC	reverse primer *Tth* intron
BG1374 s	CAGAGTCCTCTCTTAAGGTACTAGAGAAAATTGCCAACATGGAG	introducing mutation in *celB*, to create pWUR94
BG1375 a	CTAGTACCTTAAGAGAGGACTCTGGAACGTCTACGGAAATTATG	introducing mutation in *celB*, to create pWUR94
BG1397 s	GGAGTACTCGTAAGGTACTAGAGAAAATTGCCAAC	connecting 3' half of *celB* with 5' end of *Tth* intron, to create pWUR95
BG1398 a	ATATTGCTATTTAGAGAGGACTCTGGAACGTCTAC	connecting 5' half of *celB* with 3' end of *Tth* intron, to create pWUR95
BG1399 s	CCAGAGTCCTCTCTAAATAGCAATATTTACCTTTG	connecting 3' end of *Tth* intron with 5' half of *celB*, to create pWUR95
BG1400 a	TTCTCTAGTACCTTACGAGTACTCCAAAACTAATC	connecting 5' end of *Tth* intron with 3' half of *celB*, to create pWUR95
BG1401 s	AAGAAAGTAGGAAGGAGATATACCATGGCAAAGTT	circular permutation pWUR96 and pWUR97, with rbs of pET9d
BG1402 a	ATGGTATATCTCCTTCCTACTTTCTTGTAACAAATTTGAGGTC	circular permutation pWUR96 and pWUR97, with rbs of pET9d
BG1403 s	GCGCGCTCTAGAAATTGCGGGAAAGGGGTCAACAG	circular permutation pWUR96
BG1404 a	CGCGCGCTCAGCTGACGGTCTTGCCTTTTAAACCG	circular permutation pWUR96
BG1405 s	GCGCGTCTAGATTCTGTTGATATGGATGCAGTTCAC	circular permutation pWUR97, pWUR115 and pWUR124
BG1406 a	CGCGCGCTCAGCGATCTGTTGACTTAGGACTTGGC	circular permutation pWUR97, pWUR115 and pWUR124
BG1436 s	ACCTCAAATTAGGAGGTAGAAAATGAGGATCCGGCTGCTAACAAAGCC	introducing ATGA motif and rbs in pWUR94, to create pWUR114
BG1437 s	CCTCAAATTAGGAGGTAGAAAATGAAGTTCCCAAAAAACTTCATGTTTGG	introducing ATGA motif and rbs in pWUR97, to create pWUR115
BG1438 a	TCATTTTCTACCTCCTAATTTGAGGTCTGCGAGGTGAGC	introducing ATGA motif and rbs in pWUR94 and pWUR97, to create pWUR114 and pWUR115
BG1486 a	CCTAACTACTTTCTTGTAACAAATTTGAGGT	introducing rbs from pWUR115 into pWUR97, to create pWUR124
BG1495 s	CAAATTTGTTACAAGAAAGTAGTTAGGAGGTAGACCATGGCAAAGTTCCCAAAAAACT	introducing rbs from pWUR115 into pWUR97, to create pWUR124
pET rev a	CCCGTCCTGTGGATATCCGG	pET reverse primer, used with BG1436, to create pWUR114

or *Xba*I and *Blp*I (pWUR97), and constructs were used to transform to *E. coli* JM109(DE3). Sequence analysis of all constructs was done by the dideoxynucleotide chain termination method with a Li-Cor automatic sequencing system (model 4000L). The expression of the *celB* gene was monitored by blue-white-screening of colonies grown on plates containing 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-Gal) and in cell lysates by a discontinuous activity assay with pNp-β-D-glucopyranoside (pNp-Glu) as a substrate.

Enzymatic assays - Standard enzymatic assays were performed at 90°C in 60 mM citrate buffer (pH 7.0) with pNp-Glu (final concentration 3mM) as a substrate. An amount of 495 µl was preheated in a 1.5 ml Eppendorf vial. The reaction was started by the addition of 5 µl of heat-stable cell-free extract (soluble fraction of the cell-free extract after 30 minutes at 80°C). After exactly 7 minutes, the reaction was stopped by the addition of 1.0 ml of ice cold 0.5 M Na_2CO_3. This causes the pH to rise to about 9-10, terminating the reaction and enhancing the specific absorption coefficient of the liberated nitrophenol. (ε_{pNp}= 18.3 mM^{-1} at pH = 9.8). The absorption of the reaction mixture was measured at 405 nm. Protein concentrations were determined according to Bradford [14]. All activities were corrected for spontaneous hydrolysis and for hydrolysis by cell free extracts of *E. coli* cells harboring pET9d without a gene insert.

***In vitro* expression -** All *in vitro* reactions were performed using the Novagen EcoPro System, according to the corresponding protocol. Briefly, DNA-template samples were extracted twice with an equal amount of TE-buffered phenol:CIAA (1:1; CIAA is 24 parts chloroform, 1 part isoamyl alcohol) to remove possible RNAse. After extracting once with CIAA NaOAC was added to a final concentration of 0.3 M. The DNA was precipitated with 2.5 volumes of 96% ice-cold ethanol and kept at –20 °C for 60 minutes. After centrifugation of the sample for 30 minutes at maximum speed the pellet was washed with 70% ethanol and resuspended in 20 µl DEPC-treated sterile milliQ water. For *in vitro* expression, 5 µl of this resuspended DNA was used per 50 µl of total reaction volume.

RNA - For all RNA work, RNAse free solutions and glassware were used. All plasticware was handled wearing gloves and sterilized in an autoclave. Solutions were treated with 0.1% DEPC for at least 12 hours at 37 °C, after which they were sterilized in an autoclave. DNA-samples are treated with a phenol-extraction to remove possible RNase. The isolation of total RNA was performed with a Qiagen RNeasy mini-kit. The RNA was eluted from the Qiagen RNA column with 30 µl of sterile DEPC-treated milliQ water. Contaminating DNA was removed using the Qiagen RNase-Free DNase Set.

RT-PCR was performed with the Ambion RLM-Race kit, using MMLV reverse transcriptase. The Qiagen-isolated RNA was directly used in the reverse transcription experiment as described in the RLM-Race protocol. However, 4 µl RNA was used in the reverse transcription reaction, instead of the recommended 2 µl. Also, 2 µl RT-product was used in 50 µl amplification reaction, instead of the recommended 1 µl. Two primer sets were used to prove the presence of

circular mRNA. Primer set 1 consisted of primers BG238 and BG239, primer set 2 consisted of primers BG64 and BG1374 (Table 7.1). Amplification of cDNA in the second part of the protocol was performed with an annealing temperature of 52 ºC.

Exonuclease - The stability of circular mRNA towards degradation by exonuclease was examined by adding a specific exonuclease (phosphodiesterase II, Sigma) to the *in vitro* expression reaction. Samples were taken at different time points, and the reaction was stopped by placing the sample on ice. The amount of enzyme added to the *in vitro* reaction was 7 x 10^{-3} units per 30 µl reaction mix. This corresponds to 7 µl of a 10 mU/µl stock of phosphodiesterase in milliQ. The expression of *celB* was determined by a discontinuous activity assay as described earlier, but with an incubation period of 15 minutes at 90 °C.

Results

The self-splicing activity of the *Tetrahymena thermophila* ribozyme was applied for the production of circular RNA, i.e. by rearranging the sequences of the ribozyme intron and a reporter exon (Fig. 7.2B and 7.4). The *celB* gene of the hyperthermophilic archaeon *Pyrococcus furiosus* was used as a reporter gene to monitor efficient and accurate splicing of the ribozyme and the consecutive translation of the circularized *celB* messenger *in vivo*. The *celB* gene encodes a heat-stable β-glycosidase [15]. Functional *celB* expression in a β-glucosidase (LacZ) mutant of *E. coli* can easily be verified by blue-staining of recombinant colonies grown on X-Gal containing agar plates, and, in solution, by an activity assay with pNp-Glu as a substrate. The translated and properly folded CelB monomers assemble into a tetramer with four active subunits [15].

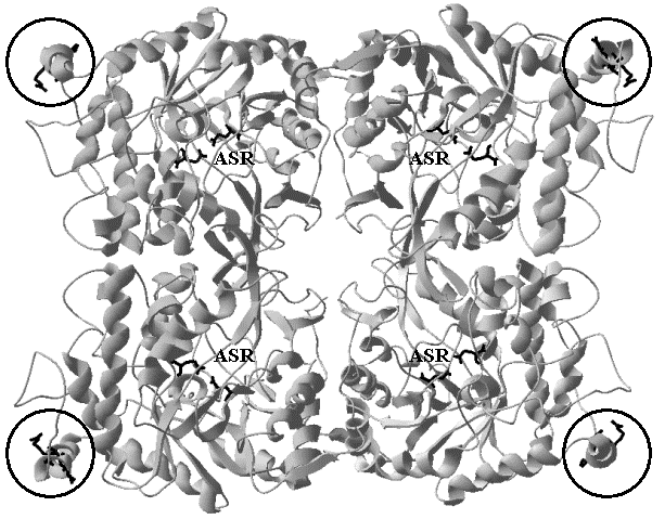

Figure 7.3 Model of the 3D-structure of CelB. Ribbon model of *Pyrococcus furiosus* β-glucosidase CelB, viewed along one of the 2-fold axes of the tetramer [12]. Active site residues (marked with ASR) and mutated residues (in encircled areas) are depicted in black.

Previous research [6, 16] has shown the importance of the flanking exon boundaries for an efficient and accurate recognition of the splice site. Therefore the *celB* gene was adapted for the insertion of the group I intron from *T. thermophila*. To ensure that such a mutation of the coding sequence of the gene would not disrupt the activity of the translated enzyme, a site at the surface of the tetramer and far from the active sites of the enzyme was chosen (Fig. 7.3). Using PCR, a 15 base pair substitution was introduced in the *celB* gene to create exon boundaries at the 5' and 3' flanking regions of the selected intron insertion site, yielding pWUR94 after cloning into *Nco*I/*Bam*H1 digested pET9d The introduced flanking regions were similar to those in the 26S rRNA gene (Fig. 7.4A), and as such should perfectly pair with the internal guide sequence (IGS) of the intron to ensure proper folding at the splice sites. The intron was inserted in the 5' and 3' flanking regions of pWUR94, by the overlap extension method, resulting in pWUR95. Both constructs were used to transform *E. coli* JM109(DE3) cells, and colonies were screened for blue-staining on X-Gal supplemented agar plates. Cell-free extracts of cultures of the different recombinants were prepared and incubated at 80 ºC for 30 minutes; these conditions result in enrichment of the heat-stable CelB because of denaturation of most *E. coli* proteins. The level of production of CelB was compared by discontinuous activity assays and analysis on SDS-PAGE (Fig. 7.5). Functional expression of active CelB was observed for both constructs, confirming i) that the introduced mutation did not impair enzyme activity, and ii) the capacity of the *Tth* ribozyme to perform self-splicing in *E. coli* JM109(DE3), with the introduced mutations in the exon boundaries in the *celB* gene. The increase in specific activity of the samples after heat incubation, indicated that it was in fact the heat stable CelB that was responsible for the β-glycosidase activity measured.

Next, a circular permutation of the gene construct was made subsequently to create a DNA sequence that would lead to the production of circular RNA, provided that the splicing reaction still occurs. Analysis of the known 2-dimensional structure of the *Tth* group I intron (Fig. 7.1A) showed that dividing the intron in the P6b stem at position 238 (indicated with an open arrow in Fig. 7.1A) would most probably result in an active ribozyme even when divided by a 1.5 kb gene. The bacteriophage T4 td intron has previously been truncated in the P6 region as well, yielding an active ribozyme [11]. Moreover, division of the intron in the P6b stem would result in large complementary strands with the potential to strongly interact. Using partially overlapping primers, both parts of the gene construct with the mutated *celB* and the ribozyme were PCR-amplified and ligated in a consecutive PCR reaction, finally resulting in pWUR97 (Fig. 7.4). This construct contains, from 5’ to 3’: the 3’ portion of the ribozyme, the 3’ exon with a stop codon (TAG), a ribosome binding site (AGGAGAT), a start-codon (ATG), the 5’ exon and finally the 5’ portion of the ribozyme.

A second truncation site in the intron was made in the joining region of stem 3 and 4 (at position 104 from the 5' end of the intron; indicated with an open arrow in Fig. 7.1A), yielding pWUR96. Motivation for this truncation site was the reported high-affinity self-assembly of separately expressed domains of the *Tth* ribozyme (P4-P6 and P3-P9), resulting in an active ribozyme [17].

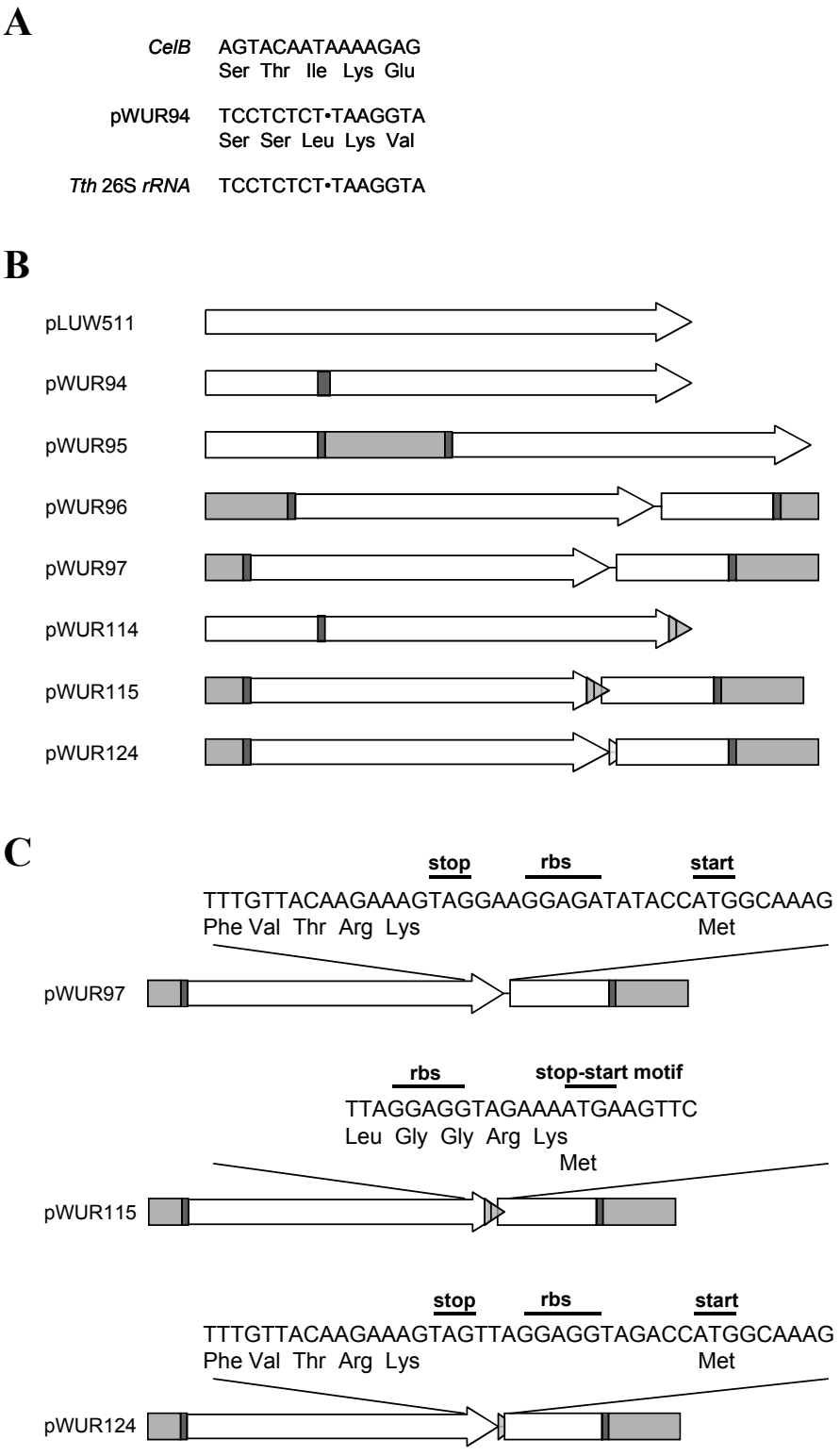

Figure 7.4 Overview of all DNA constructs. **(A)** sequence alignment of the flanking exon boundaries in the 26S rRNA of *Tetrahymena thermophila* (*Tth* 26S *rRNA*), the one that has been introduced in *celB* (pWUR94) and the original sequence of *celB* at that position. **(B)** and **(C)** Schematic presentation of the different constructs. Only the DNA insert of the vectors is shown. Colors as in Fig. 2. The introduced ribosome binding site (rbs) in pWUR114, 115 and 124 is depicted in outlined diamond pattern, the introduced ATGA motif is depicted in dark horizontal pattern. **(C)** detailed sequence view of the stop-start region of the 3 constructs with different designs of ribosome binding site (rbs) or stop-start motif. Stop- and start-codons are indicated.

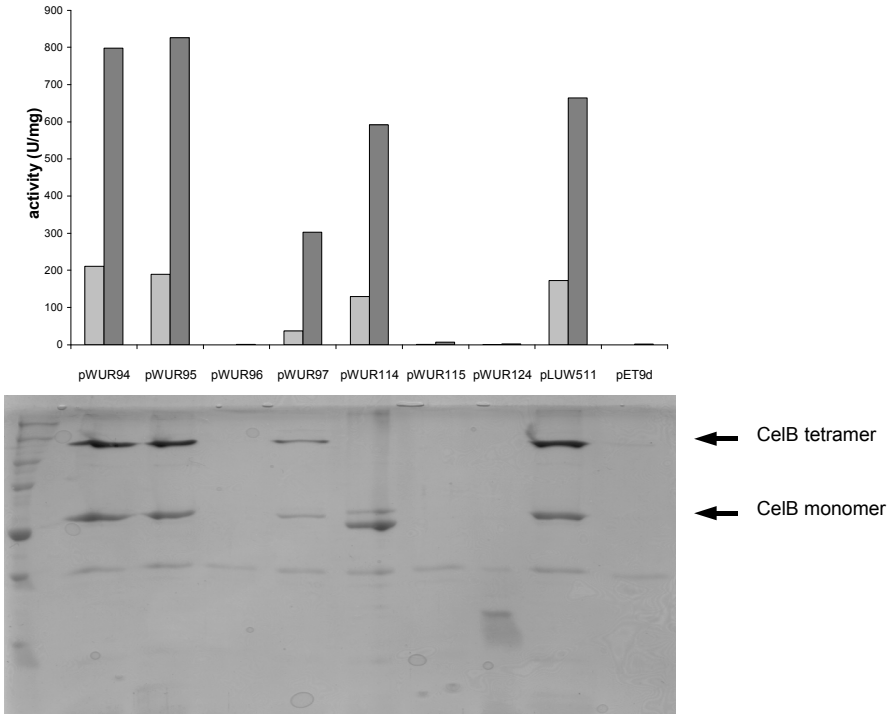

Figure 7.5 Expression and activity of CelB from different constructs. The upper panel shows the specific activity of CelB in Units per mg protein in cell-free extract, CFE (depicted in light grey) and heat-stable cell-free extract, HSCFE (depicted in dark grey). The lower panel shows the corresponding SDS-PAGE analysis of the heat stable cell-free extracts of the different constructs. Arrows indicate the tetramer (216 kDa) and monomer (54 kDa) of CelB. Lane 1, Biorad precision marker; Lane 2, pWUR94; Lane 3, pWUR95; Lane 4, pWUR96; Lane 5, pWUR97; Lane 6, pWUR114; Lane 7, pWUR115; Lane 8, pWUR124; Lane 9, pLUW511; Lane 10, pET9d;

A recent study by Ishida *et al.* (2002) [18] has shown how an upstream small open reading frame (encoding a leader peptide) may enhance the expression of an adjacent gene in *E. coli*, especially when the stop and start codons overlap (ATGA). Introducing an ATGA-motif with a 4-basepair overlap between the stop codon of a leader open reading frame and the start codon of the downstream coding region of a gene, increased protein expression more than 3-fold, compared to the expression from a construct with a 10 nt gap between the leader ORF and the start codon of the gene. The authors proposed that the enhanced expression of the downstream gene was a result of "efficient ribosome recruitment" [18]. In analogy, pWUR115 was designed with an ATGA motif and a RBS introduced at the 3' end of the *celB* coding region (Fig. 7.4). The design of construct pWUR115 with respect to the RBS and the spacing towards the start-codon was based on previous research on the role of the Shine-Dalgarno (SD) sequence in prokaryotes [19]. To introduce the RBS at the 3' end of *celB*, three amino acids had too be substituted. To evaluate the influence of this mutation on the activity of the enzyme, the same mutation was introduced in pWUR94, resulting in pWUR114. PCR using partially overlapping primers was used to create pWUR115 and pWUR114 with the described mutations in the 3' end of the sequence. The importance of the exact sequence of the SD region for ribosome recruitment, and therefore translation initiation, has been shown before [18]. To examine the influence of the RBS and its directly flanking region on the expression from pWUR115, this sequence was replaced in pWUR97, resulting in pWUR124 (Fig. 7.4).

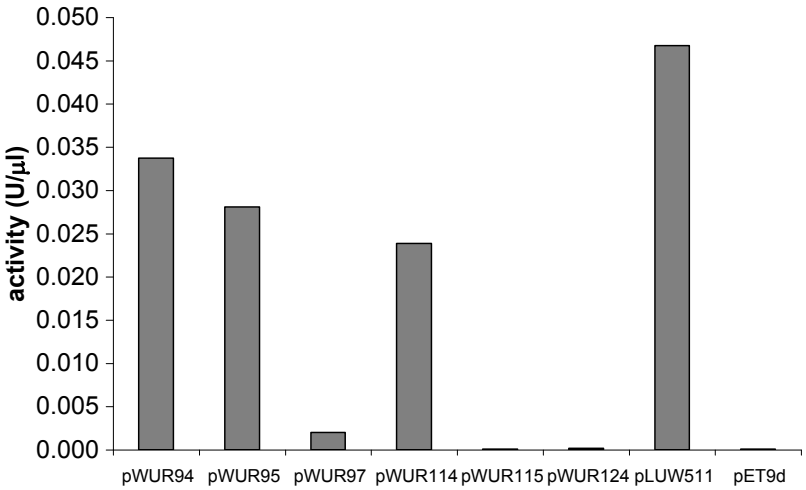

Figure 7.6 *In vitro* expression. The expression levels *in vitro* are indicated by the activity in Units per µl of the *in vitro* transcription/translation reaction.

No functional expression of CelB was observed for constructs pWUR96, pWUR115 and pWUR124, while pWUR97 yielded lower levels of active protein compared to wild-type *celB* and pWUR94 and pWUR95. Analysis on SDS-PAGE revealed the presence of CelB tetramer and monomer in the different heat-stable cell-free extracts (Fig. 7.5), as has been observed before [20]. All samples were incubated at 99°C for 10 minutes, prior to loading on gel to dissociate the CelB tetramer. However, to some extent the tetramer was still present on gel, except for sample pWUR114 (Fig. 7.5C, lane 6).

In addition to *in vivo* heterologous expression, all constructs were tested for transcription, splicing and translation *in vitro*, using the Novagen Ecopro system, which is a prokaryotic *in vitro* expression system based on a bacterial extract. Expression levels of CelB were examined using a discontinuous activity assay, showing similar trends as observed *in vivo* (Fig. 7.6).

An alternative assay to prove the presence of circular RNA was performed by RT-PCR [21]. RNA was purified from cultures in the mid log phase, and used as a template in a reverse transcription reaction by MMLV reverse transcriptase to produce the corresponding cDNA. Two sets of primers for the RT-reaction were designed such that only covalently closed circular RNA (cccRNA) could give rise to cDNA with both primer sets. Linear, unspliced RNA would only give rise to amplification with one primer set (set 1, grey dashed line in Fig. 7.7A), and RNA from the wild-type gene would give rise to amplification from the other primer set (set 2, black dotted line in Fig. 7.7A). The cDNA produced in the RT-reaction was consecutively amplified in a normal PCR reaction with *Taq* polymerase to produce sufficient amounts of DNA to visualize by Ethidium Bromide staining after electrophoresis on an agarose gel. For pWUR97, this analysis indeed showed the presence of amplified DNA products of the expected size for both primer sets, thereby confirming the presence of circular mRNA (Fig. 7.7B). The same reaction with pWUR96-RNA showed only amplification from primer set 1, indicating the presence of unspliced mRNA.

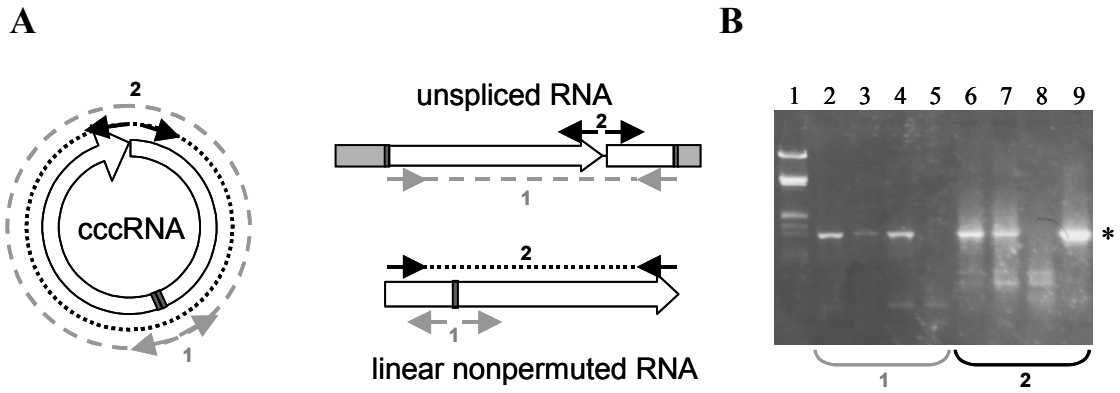

Figure 7.7 RT-PCR. **(A)** Two sets of primers can amplify different products, depending on the template present. Only cccRNA will give rise to a product with both primersets. Colors as in Fig. 2. **(B)** Gel electrophoresis after RT-PCR, the asterix indicates the expected product (1.4 kb). Lane 1, DNA marker; Lanes 2-5 represent the products with primer set 1, with in Lane 2, pWUR97; Lane 3, pWUR97; Lane 4, pWUR96; Lane 5, pLUW511; Lanes 6-9 represent the products obtained with primer set 2, with in Lane 6, pWUR97; Lane 7, pWUR97; Lane 8, pWUR96; Lane 9, pLUW511;

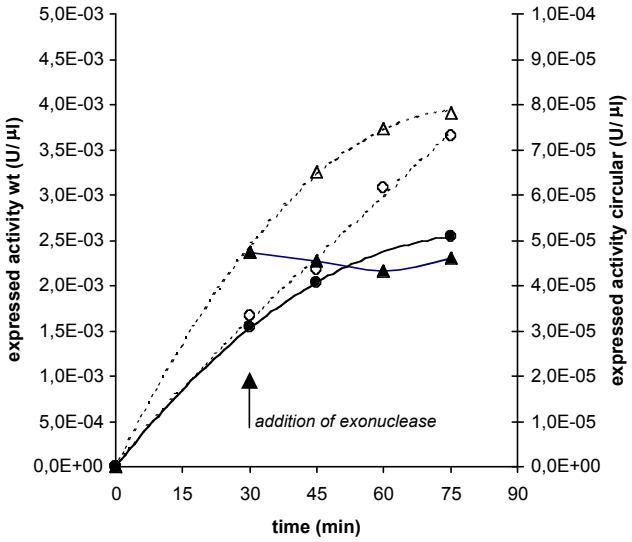

Figure 7.8 Stability of circular mRNA. *In vitro e*xpression of *celB* from linear and circular messenger RNA, after addition of exonuclease (phosphodiesterase II, Sigma). The lines with triangles show the results from the wild-type *celB* gene (left axis), the lines with closed circles show the results from circular RNA from construct pWUR97 (right axis). After 30 minutes of in vitro expression, the exonuclease was added to both samples. The dashed lines show the increase in expressed CelB activity in the absence of exonuclease.

pLUW511-RNA gave rise to a product with primer set 2, indicating the presence of normal, non-permutated messengers, but no circular messenger.

The stability of circular and linear mRNA against exonuclease degradation was compared. During *in vitro* expression, a specific exonuclease (phosphodiesterase II, Sigma) was added to the reaction and samples were taken at different time points. The increase in CelB-activity after addition of exonuclease was measured and compared to the CelB-activity in samples without added exonuclease (Fig. 7.8). The results indicated an enhanced yield in the case of the circular messenger, strongly suggesting a stabilization of the circular mRNA.

Discussion

In the present study we have shown that a rearranged group I intron, derived from the ciliate *T. thermophila,* can be used to synthesize large circular mRNAs in *E. coli*, and that this circular transcript is functionally translated to an active β-glycosidase. Additional analyses, using SDS-PAGE and *in vivo* and *in vitro* activity assays, were instrumental for assessing the influence of the constructs' design on the expression level of *celB*.

We could show that introduction of an intron in the *celB* gene, as in pWUR95, did not significantly decrease the expression level of the enzyme. This confirmed results of Guo and Cech [6], showing that the eukaryotic splicing mechanism of the intron can easily occur *in vivo* in *E. coli*,. Unlike the latter study, however, the original "intron insertion site" introduced in the *celB* gene proved to be effective for an accurate and efficient splicing reaction, without decreasing the enzyme activity or stability.

Furthermore, we could demonstrate that the success of a circular permutation of construct pWUR95 strongly depends on the actual site of the permutation. Truncating the intron between stem P3 and P4, as in construct pWUR96, did not give rise to the production of CelB. This is most likely due to a disturbed splicing mechanism of the transcribed mRNA. The failure of this construct is unexpected given the results of Doudna and Cech [17], and probably indicates the effect of the additional sequence (*celB* fragments) that may affect the correct assembly of the ribozyme domains. On the other hand, a division in stem-loop P6, as in pWUR97, yielded functional CelB expression. This successful re-assembly of the ribozyme domains agrees well with analogous truncation experiments with T4 td intron [11]. The production of active CelB by cells harbouring pWUR97 confirmed that the transcribed messenger can indeed perform autocatalytic splicing and ligation *in vivo*, and that the created circular messenger can be translated by the prokaryotic translation machinery. Although translation of circular mRNAs has been shown once before with the T4 Group I intron [8], the here described circular messenger is slightly larger in size, it concerns an enzyme-encoding transcript, and the unprecedented design of shuffled gene fragments is a convenient tool for screening functional construct designs.

Comparison of the expression levels of the permutated construct and the wild type gene revealed a decrease in activity of approximately 90%; this is comparable with the GFP production from a circular messenger using the T4 td intron [8]. Since the introduction of an intron in the native *celB* gene does not affect expression levels, the truncation of the intron most likely results in a decreased splicing efficiency of the messenger because of sub-optimal assembly of the ribozyme domains. This may result in lower levels of accurately processed mRNAs, or may even slow down the overall splicing and ligation reactions, allowing nucleases to degrade the linear messenger. Alternatively, efficiency of translation initiation could be decreased in case of a circular messenger. Ribosome recruitment on a circular messenger might be less efficient than recruitment on a linear messenger. This could also negatively affect protein production levels.

In vitro expression experiments revealed expression profiles for the different constructs that agree very well with the trends observed *in vivo*. Comparison of the wild-type gene and the permutated construct showed an approximate 95% decrease in functional CelB production, which is

slightly worse than observed *in vivo*. The anticipated increased stability of the circular RNA does not appear to be an extra advantage *in vitro*.

Apart from the technical challenge, a motivation of generating circular messengers was to enhance *in vivo* expression, because of anticipated longer messenger lifetime, and maybe even further by increased translation efficiency via a "rolling circle mechanism". However, these effects could neither be demonstrated in the case of CelB (this study) nor GFP [8]. A likely explanation for this observation is that the transcript stability in the case of these two reporters is not a limiting factor for protein production.

To accomplish rolling-circle translation of the *celB* messenger, the construct was engineered to introduce a 4-basepair overlap in the stop and start codon of the *celB* gene, as is observed in many natural poly-cistronic messengers in prokaryotes. This should allow the ribosome to continuously read and translate without dissociating from the circular mRNA. However, pWUR115, did not yield active CelB. To create the 4-basepair overlap in stop and start codon (ATGA-motif), radical changes had to be introduced at the C-terminus of CelB, and to some extent in the SD-sequence. Two constructs were designed to evaluate these changes introduced in the DNA-sequence of pWUR115: pWUR114 and pWUR124 (Fig. 7.7AB). In construct pWUR114, the mutation in the 3' end of the *celB* gene was introduced in the mutated *celB* gene, without intron sequence. Although some activity was measured in the cell-extracts of cells containing this construct, it was approximately 50% less than the expression from the wild-type gene. On SDS-PAGE, the overexpressed protein appeared to migrate slightly faster than the wild-type CelB. Moreover, the very stable tetramer of the enzyme was still present in the lanes with the wild-type CelB, but did not appeare in the pWUR114 lane. This suggests that either the tetramer of the enzyme was not present at all, or it was less stable than the tetramer of the wild-type CelB and had therefore completely dissociated during the heat-incubation of the sample, prior to loading on gel. The C-terminal end of the CelB monomer has previously been reported to play a crucial role in tetramer formation and stability [20]. A mutation in this part of the protein might therefore indeed decrease the stability of the tetramer. Since only the tetramer of the protein is known to be active [20], this might also explain the decreased enzyme activity in the cell-extract.

The influence of the changed SD-sequence was evaluated with construct pWUR124, in which the SD-sequence and spacer between stop- and start-codon of construct pWUR97 was replaced by the SD-sequence as present in pWUR115 (Fig. 7.7). The lack of activity of this construct (which has a more perfect match with the anti-SD of *E. coli*) is puzzling, as also sequence analysis did not indicate mistakes. This might reflect the importance of the sequence directly upstream of the gene for protein expression [22]. However, the introduced ribosome binding site is known to be functional in *E. coli*. An alternative explanation might be that the introduced substitutions enable the formation of some a-specific interaction with the ribozyme domains that disrupt its splicing and/or ligation activity. Future research is required to draw final conclusions on the effect of the 4-basepair overlap in an ATGA-motif. Random mutagenesis of the SD-sequence might lead to variants with enhanced efficiency, as well as applying the proposed design on a gene with a less crucial 3' end.

Addition of exonuclease apparently resulted in an immediate stop of CelB production from a linear construct (Fig. 7.8). This is in agreement with the expected degradation of the linear messenger RNA by the enzyme, thus destroying any template for the translation reaction. Addition of exonuclease to a circular RNA-template decreased CelB production as well, but certainly not to the same extent. The decrease in production compared to the samples without exonuclease can be explained by assuming the degradation of primary, unspliced, and therefore linear, transcript. However, the already present pool of circular mRNA remains available as a template for translation, and as such for functional CelB production.

A method was developed to prove the presence of a specific circular RNA by means of RT-PCR. Although the presence of an active enzyme itself is a strong indication for a successful intron splicing and an effective translation of the messenger, the RT-PCR method provides elegant means to confirm the presence of circular RNA. Moreover, in case of disturbed messenger translation, RT-PCR might still reveal the presence of circular RNA. Finally, the increased stability of circular mRNA towards exonuclease degradation is demonstrated.

Instead of optimizing the translational efficiency of the circular messenger, another way of improving protein expression in future research could be to improve the splicing efficiency of the permutated messenger *in vivo*. Studies by Guo and Cech (2002) [6] indicated that the flanking regions are extremely important: base pairing should not be too strong, but proper base pairing of the 5' exon to the IGS to form the P1 stem loop, as well as base pairing of the 3' exon to the IGS to form the P10 stem loop, are essential. Analysis of the secondary structure of the *Tth* 26S rRNA, shows a highly organised structure which most likely is crucial for optimal intron splicing. Previous research has in fact revealed that the minimum length for optimal splicing of the *Tth* 26S rRNA, is 145 nucleotides upstream and 85 nucleotides downstream of the intron [23]. More research might reveal essential sequential elements in this region for an improved intron splicing, and might be useful to explain some of the efficiency-related matters that were encountered in the present study.

A rather interesting question that remains is whether the ribosome is capable of translating the cirularized mRNA more than once without dissociating from its RNA template. Evaluating the expression of different constructs described in this study does not give a definite answer to this question. In the study of Perriman and Ares [8] a design was made in which the stop codon was deleted, resulting in a continuous circular open reading frame; this indeed resulted in long repetitive strings of proteins, indicating the occurrence of rolling circle translation. Alternative designs presented here did not lead to the anticipated result, for reasons that are not understood at present.

The system described in this study may be useful for detailed analysis of ribozyme-catalyzed splicing, as well as of ribosome-mediated translation. Apart from these fundamental processes, the system may have potential for improved expression of genes that have very instable messengers, or for certain *in vitro* applications that require mRNA templates with enhanced stability.

Acknowledgments

This research was supported by the Technology Foundation (STW), applied science division of NWO and the technology programme of the ministry of Economic Affairs.

References

1. **Cech, T. R.** (1990) Self-splicing of group I introns, *Annu. Rev. Biochem.* **59**, 543-568.
2. **Been, M. D. & Cech, T. R.** (1986) One binding site determines sequence specificity of *Tetrahymena* pre-rRNA self-splicing, trans-splicing, and RNA enzyme activity, *Cell.* **47**, 207-216.
3. **Waring, R. B., Towner, P., Minter, S. J. & Davies, R. W.** (1986) Splice-site selection by a self-splicing RNA of *Tetrahymena*., *Nature*, 133-139.
4. **Strobel, S. A. & Cech, T. R.** (1995) Minor groove recognition of the conserved G.U pair at the *Tetrahymena* ribozyme reaction site, *Science.* **267**, 675-679.
5. **Tanner, N. K.** (1999) Ribozymes: the characteristics and properties of catalytic RNAs, *FEMS Microbiol. Rev.* **23**, 257-275.
6. **Guo, F. & Cech, T. R.** (2002) *In vivo* selection of better self-splicing introns in *Escherichia coli*: the role of the P1 extension helix of the *Tetrahymena* intron, *RNA.* **8**, 647-658.
7. **Perriman, R.** (2002) Circular mRNA encoding for monomeric and polymeric green fluorescent protein, *Methods Mol. Biol.* **183**, 69-85.
8. **Perriman, R. & Ares, M., Jr.** (1998) Circular mRNA can direct translation of extremely long repeating-sequence proteins *in vivo*, *RNA.* **4**, 1047-1054.
9. **Puttaraju, M. & Been, M. D.** (1992) Group I permuted intron-exon (PIE) sequences self-splice to produce circular exons, *Nucleic Acids Res.* **20**, 5357-5364.
10. **Puttaraju, M. & Been, M. D.** (1996) Circular ribozymes generated in *Escherichia coli* using group I self-splicing permuted intron-exon sequences, *J. Biol. Chem.* **271**, 26081-26087.
11. **Ford, E. & Ares, M., Jr.** (1994) Synthesis of circular RNA in bacteria and yeast using RNA cyclase ribozymes derived from a group I intron of phage T4, *Proc. Natl. Acad. Sci. U S A.* **91**, 3117-3121.
12. **Kaper, T., Lebbink, J. H., Pouwels, J., Kopp, J., Schulz, G. E., van der Oost, J. & de Vos, W. M.** (2000) Comparative structural analysis and substrate specificity engineering of the hyperthermostable beta-glucosidase CelB from *Pyrococcus furiosus*, *Biochemistry.* **39**, 4963-4970.
13. **Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K. & Pease, L. R.** (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction, *Gene (Amsterdam).* **77**, 51-59.
14. **Bradford, M. M.** (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, *Anal. Biochem.* **72**, 248-254.
15. **Voorhorst, W. G., Eggen, R. I., Luesink, E. J. & de Vos, W. M.** (1995) Characterization of the celB gene coding for beta-glucosidase from the hyperthermophilic archaeon *Pyrococcus furiosus* and its expression and site-directed mutation in *Escherichia coli*, *J. Bacteriol.* **177**, 7105-7111.
16. **Hagen, M. & Cech, T. R.** (1999) Self-splicing of the *Tetrahymena* intron from mRNA in mammalian cells, *EMBO J.* **18**, 6491-6500.
17. **Doudna, J. A. & Cech, T. R.** (1995) Self-assembly of a group I intron active site from its component tertiary structural domains, *RNA.* **1**, 36-45.
18. **Ishida, M. & Oshima, T.** (2002) Effective structure of a leader open reading frame for enhancing the expression of GC-rich genes, *J. Biochem. (Tokyo).* **132**, 63-70.
19. **Ma, J., Campbell, A. & Karlin, S.** (2002) Correlations between Shine-Dalgarno sequences and gene features such as predicted expression levels and operon structures, *J. Bacteriol.* **184**, 5733-5745.

20. **Pouwels, J., Moracci, M., Cobucci-Ponzano, B., Perugino, G., van der Oost, J., Kaper, T., Lebbink, J. H., de Vos, W. M., Ciaramella, M. & Rossi, M.** (2000) Activity and stability of hyperthermophilic enzymes: a comparative study on two archaeal beta-glycosidases, *Extremophiles.* **4**, 157-164.
21. **Mackie, G. A.** (2000) Stabilization of circular rpsT mRNA demonstrates the 5'-end dependence of RNase E action in vivo, *J. Biol. Chem.* **275**, 25069-25072.
22. **Sprengart, M. L. & Porter, A. G.** (1997) Functional importance of RNA interactions in selection of translation initiation codons, *Mol. Microbiol.* **24**, 19-28.
23. **Woodson, S. A.** (1992) Exon sequences distant from the splice junction are required for efficient self-splicing of the *Tetrahymena* IVS, *Nucleic Acids Res.* **20**, 4027-4032.