Supplementary Information

Highly motif- and organism-dependent effects of naturally occurring hammerhead ribozyme sequences on gene expression

Lena A. Wurmthaler[1], Benedikt Klauser[1], and Jörg S. Hartig[1*]

[1] Department of Chemistry, and Konstanz Research School Chemical Biology (KoRS-CB), University of Konstanz, Germany

* To whom correspondence should be addressed. Tel: +49-7531-884575; Email: joerg.hartig@uni-konstanz.de

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

CONTENT

Supplementary Note 1: Characterisation of selected HHR sequences

We selected eight HHR motifs representing all three topologies by type 1 the mouse gut metagenome HHR [1] and the fungus *Yarrowia lipolytica* HHR motifs [1], the type 2 HHR motifs originate from *Roseburia intestinalis* (*Clostridia*, gram-positive bacteria) [2] and sewage microbiome sequencing data [3] (SewR3-00810s-1), and the type 3 HHR motifs of the satellite Lucerne Transient Streak Virus (-) [4] (sLTSV(-)), the Chrysanthemum Chlorotic Mottle Viroid (-) [5] (CChMVd(-)), *Arabidopsis thaliana* [6], and from marine metagenome sequences [1]. The selection of HHR motifs is depicted in the main text **Fig. 1** and **Supplementary Tab. S1**. We analysed all HHR sequences for alternative start codons in bold. Sequences were inserted as shown in the *in vivo* gene expression constructs in **Tab. S4**.

Supplementary Note 2: Artifical gene expression systems to study intracellular HHR activity in human cells, *S. cerevisiae* and *E. coli*

Plasmid construct design for gene expression in eukaryotic cell systems

We aimed for the application of artificial genetic systems that allow us to investigate the selected HHR motifs in a clear genetic context decoupling the folding structure of the HHR motifs from the influence of neighbouring sequences. To evaluate intracellular activity of the selected HHR motifs are evaluated by comparison of expression levels of reporter gene transcripts with an inserted actively cleaving wild-type HHR motif to its inactive version (**Fig. 1**). Therefore, a single A to G mutation was introduced to inactivate ribozyme-dependent self-cleavage of the RNA backbone [7]. In eukaryotic cells mRNA lifetime is prolonged by the integrity of the two structural stabilisers 5´-cap and poly(A) tail [8]. Insertion of a ribozyme into either the 5´-UTR or 3´-UTR of a eukaryotic reporter gene leads to efficient downregulation of mRNA levels [7, 9, 10]. Autocatalytic cleavage of the mRNA destabilises the transcript due to missing poly(A) tail or 5´-cap. De-adenylated mRNAs are rapidly degraded by the cytoplasmic exosome [8].

Eukaryotic ribosomes possess a unique scanning mechanism, which causes translation initiation at the first AUG start codon reached [11]. We analysed the chosen HHR representatives for the occurrence of any AUG triplet and identified at least one per sequence, except for the marine metagenome HHR (**Tab. S1**). To prevent any inhibitory effect on reporter gene expression by alternative start codons, we decided to insert the HHR sequences into the 3´-UTR of a reporter gene in eukaryotic cell systems. Importantly, to exclude the formation of alternative secondary structures we chose adenosine-rich flanking sequences that reduce the likelihood of interference with HHR structure formation in both yeast and mammalian contexts.

Plasmid construct design for gene expression in mammalian cell line

For a transiently transfected human HeLa S3 cell line system a dual luciferase reporter gene construct was utilised. The psi-CHECK2 vector was used (**Fig. 3A**). The *Renilla* luciferase hRluc expression depends on ribozyme inactivity. The expression of firefly luciferase encoded on the same plasmid is ribozyme-independent. The orthogonal read-out of both luciferase activities enables normalisation of reporter gene expression and hence reduces noise coming from varying transfection efficiencies, fluctuating cell numbers, and environmental fluxes. The insertion areas of the ribozyme motifs are depicted in **Tab. S4**.

Plasmid construct design for gene expression in yeast

In yeast the ribozyme constructs were developed for controlling the expression of the Gal4 transcription factor by a ribozyme (**Fig. 3B**). An engineered form of the low-copy number pBT3 plasmid, a yeast two hybrid assay (Y2H) plasmid with a split-ubiquitin system that is normally employed in membrane protein-protein interactions studies was used. The open reading frame (ORF) of the fusion protein composed by the C-terminal half of ubiquitin (Cub) and the artificial transcription factor LexA-VP16, was removed by restriction digest and the ORF of the GAL4 transcription factor was inserted on the plasmid. The gene construct is under the control of the constitutive CYC1 promoter. GAL4 activates the transcription of a genomically encoded lacZ gene whose expression levels can be easily quantified in a chemoluminescence assay (see Methods). The genetic setup was successfully used by our group before [12]. The long spacer sequence in the 3´-UTR between the ribozyme and the terminator is what remains on the plasmid after insertion of the GAL4 transcription factor coding sequence. It does not contain any functional RNA sequence or ORF. *In vivo* experiments are conducted in the *S. cerevisiae* MaV203 strain, which features a chromosomally encoded lacZ gene under control of a Gal4-inducible promoter. MaV203 cells are mutated in their endogenous GAL4 and GAL80 genes coding for the Gal4 transcription

factor and its repressor. Cleavage of the Gal4 mRNA by a catalytically active HHR results in rapid mRNA degradation and, thus, low Gal4 protein levels and reduced lacZ expression levels. In contrast, maximal Gal4 mRNA levels are reached, when the inserted HHR is inhibited by the insertion of an A to G mutation, which results in strong lacZ expression. β-Galactosidase expression levels are quantitatively determined by enzymatic analysis of the β-galactosidase activity. The insertion areas of the ribozyme motifs are depicted in **Tab. S3**.

Plasmid construct designs for gene expression in *E. coli*

In case of the bacterial genetic context in *E. coli* we inserted the chosen HHR motifs without additional spacer sequences into the 3´-UTR (**Fig. 4A**) or the 5´-UTR (**Fig. 5A**) of an enhanced green fluorescence protein (eGFP) reporter gene which is under control of a constitutive promoter. The constructed plasmids are based on the pQE-TriSystem (QIAGEN) that features an ampicillin resistance and the eGFP gene. The constitutive J06 promoter inserted was modified from Anderson library (http://parts.igem.org/Promoters/Catalog/Anderson) and drives eGFP transcription. Sequences of the HHR motifs were subcloned into the 3´- or 5´-UTR of the eGFP reporter gene. Here we chose the insertion position for HHR motifs 31 nt downstream of the stop codon for the 3´-UTR constructs. For 5´-UTR constructs we tried to prevent any discussed problems [13-17] by insertion of the HHR motifs 12 nt upstream of the SD site (AAGGAG) and 25 nt upstream of the to start codon (AUG). The distance to the upstream promoter is 11 nt. In bacteria the Shine Dalgarno site is necessary for initiation of translation, therefore alternative start codons within the 5´-UTR should not influence reporter gene expression. Alternative open reading frames (ORFs) and other factors could be ruled out as described in **Supplementary Note 5 and 6**. Expression of eGFP in the *E. coli* system can be directly quantified. The insertion area of the ribozyme motifs is depicted in **Tab. S4**.

Supplementary Note 3: Flow cytometric singe-cell fluorescence (FC) analysis of into 3´- and 5´-UTR inserted HHR motifs in *E. coli* reporter system with overexpression of eGFP reporter gene.

Flow cytometric single-cell analysis was performed to confirm the ribozyme mediated effects on gene expression in *E. coli* which we determined by gene expression assays before when either inserted into the 3´-UTR (**Fig. 4**) or into the 5´-UTR (**Fig. 5**) of the eGFP reporter gene. For flow cytometric single-cell analysis *E. coli* bacterial cultures were grown as described for gene expression analysis to stationary phase. 1 ml of culture was harvested, pelleted and washed in 1x PBS. Finally, the pellet was dissolved in 1.5 ml 1x PBS and analysed with a FACScalibur cell analyser (BD Biosciences Singapore, FlowKon facility) by using a 488-nm argon laser for excitation. Fluorescence of 100,000 cell counts was detected through FL-1 band pass filter with photon multiplier tube voltage of 880. Additionally, forward scatter (FSC) and side scatter (SSC) were measured. Signals were amplified with the logarithmic mode. All cytometric data were gated and analysed with CellQuestPro software. Populations were compared in histograms with counts plotted against FL-1 signal. We could detect shifts of homogeneous populations due to their active or inactive HHR counterpart inserted into 5´-UTR or 3´-UTR of eGFP reporter gene (**Fig. S2**). First an eGFP control, the parental expression construct lacking ribozyme sequences was used for comparison to a non eGFP expressing plasmid system in TOP 10 *E. coli* cells (**Fig. S2A**). The samples with active sLTSV(-) and CChMVd(-) HHR motifs inserted in the 3´-UTR of eGFP reporter (**Fig. S3B**) gene showed a decrease in eGFP expression compared to its inactive counterpart. However, no shift was detected for the other 3´-UTR inserted HHR constructs tested. In the 5´-UTR of reporter gene eGFP inserted active *Y. lipolytica*, *R. intestinalis* and sLTSV(-), but also *A. thaliana* HHR motifs showed less eGFP intensity compared to their inactive HHR populations (**Fig. S3C**). The others, especially the mouse gut metagenome and SewR3-00810s-1 HHR motifs, show an increase in eGFP signal compared to the culture with the inactive HHR motif. *Y. lipolytica* and CChMVd(-) motif constructs do not show a difference between the two

populations according to their eGFP signal intensity. With an exception of the *A. thaliana* HHR motif construct, bulk measurements of eGFP signal show increased eGFP signals of the active compared to the inactive counterpart. The flow cytometry analysis confirms the tendencies analysed by the gene expression assays in bulk.

Supplementary Note 4: Quantitative real-time PCR analysis of mRNA levels in *E. coli:*

RNA source, total RNA extraction and reverse transcription

Total RNA was isolated using the QIAGEN RNeasy Mini Kit with additional RNAprotect Bacteria Reagent and RNase-Free DNase Set. Total RNA was purified from late exponential growth cultures, cultivated in 1 ml LB-Carb media, 24-well system, inoculated from overnight culture to OD_{600} of 0.1. Samples were grown in biological triplicates at 37 °C, 200 rpm until an OD_{600} of 1.2. For checking the protein levels of eGFP, a gene expression assay as described before was carried out (data not shown). According to the QIAGEN protocol $5x10^8$ cells were harvested and incubated with doubled volume of RNAprotect Bacteria Reagent for 5 min at room temperature. The cells were pelleted by centrifugation for 15 min 5000 x g and supernatant were carefully removed, before freezing at – 80 °C until use. Twenty samples were further treated in one go, first thawed on ice. Cell pellet was lysed by addition of 13.5 U proteinase K (Fermentas) and 60,000 U lysozyme (Roth) in 1x TE buffer, carefully resuspended and incubated 10 min at room temperature. RNA was purified using the provided protocols from QIAGEN for bacterial cells. An on-column DNA digestion was performed during RNA purification following the provided protocol of QIAGEN. Finally, RNA was eluted with 30 µl DEPC treated water and nucleic acids were quantified first by Tecan infinite M200 with Nano Quant Plate (software Tecan i-control). Purity of samples was given by A_{260}/A280 ratio of > 2. Second RNA was separated by 0.8% agarose gel electrophoresis and

bands were visualised by a peqGreen dye in the Amersham Imager 600 with band volume detection (GE Healthcare, data not shown). The ratios of 23S rRNA to 16S rRNA were around ≥ 1.3 for 5´-UTR and ≥ 2.0 for 3´-UTR RNA samples, pointing towards good RNA integrity [18]. Synthesis of cDNA and non-reverse transcribed controls (NTCs) were directly performed using 1 µg total RNA and the QuantiTect® Reverse Transcription kit (QIAGEN). The samples were stored at – 20 °C.

Primer specificity and efficiency for semi-quantitative RT-PCR

The data set of quantitative RT-PCR (qPCR) presented derived from second time thawed cDNA samples. For detection of eGFP target gene forward primer 5´-GAAGGAGATATACCATGGGCCATCA and reverse primer 5´-GCTGAACTTGTGGCCGTTTAC (target length 164 nt) were used. The reference gene ssrA (miscRNA) was amplified by forward primer 5´-ACGGGGATCAAGAGAGGTCAAAC and reverse primer 5´-CGGACGGACACGCCACTAAC (target length 65 nt). qPCR runs were performed on a TOptical thermocycler (Biometra, Analytic Jena) in 96 well plates with a final volume of 10 µl with 0.5 µl cDNA from synthesis in GoTaq® qPCR Master Mix (Promega). The following thermocycler parameters were used: initial denaturation at 95 °C, 2 min; denaturation 95 °C, 15 s, annealing and extension at 60 °C, 60 s in 35 cycles. A first dilution row showed inhibition for 100 % cDNA used, however a 10-fold diltuion of cDNA could be used for further analysis. Primer efficiency and specificity were tested within a dilution row of 10 % to 0.01 % of cDNA synthesis concentration (**Fig. S4A**). The qPCR efficiency (*E*) was calculated with the slope of dilution curves depicted, according to the equation: $E = 10^{[-1/slope]}$. E values of 1.97 and 2.06 were determined for ssrA and eGFP, respectively (see Fig S3). Specificity of qPCR was determined by melting curve analysis performed after the final extension, starting from 65 °C up to 98 °C. Melting temperatures were 81.8 °C for ssrA and 85.6 °C for eGFP. All following samples were measured with the same primer and enzyme

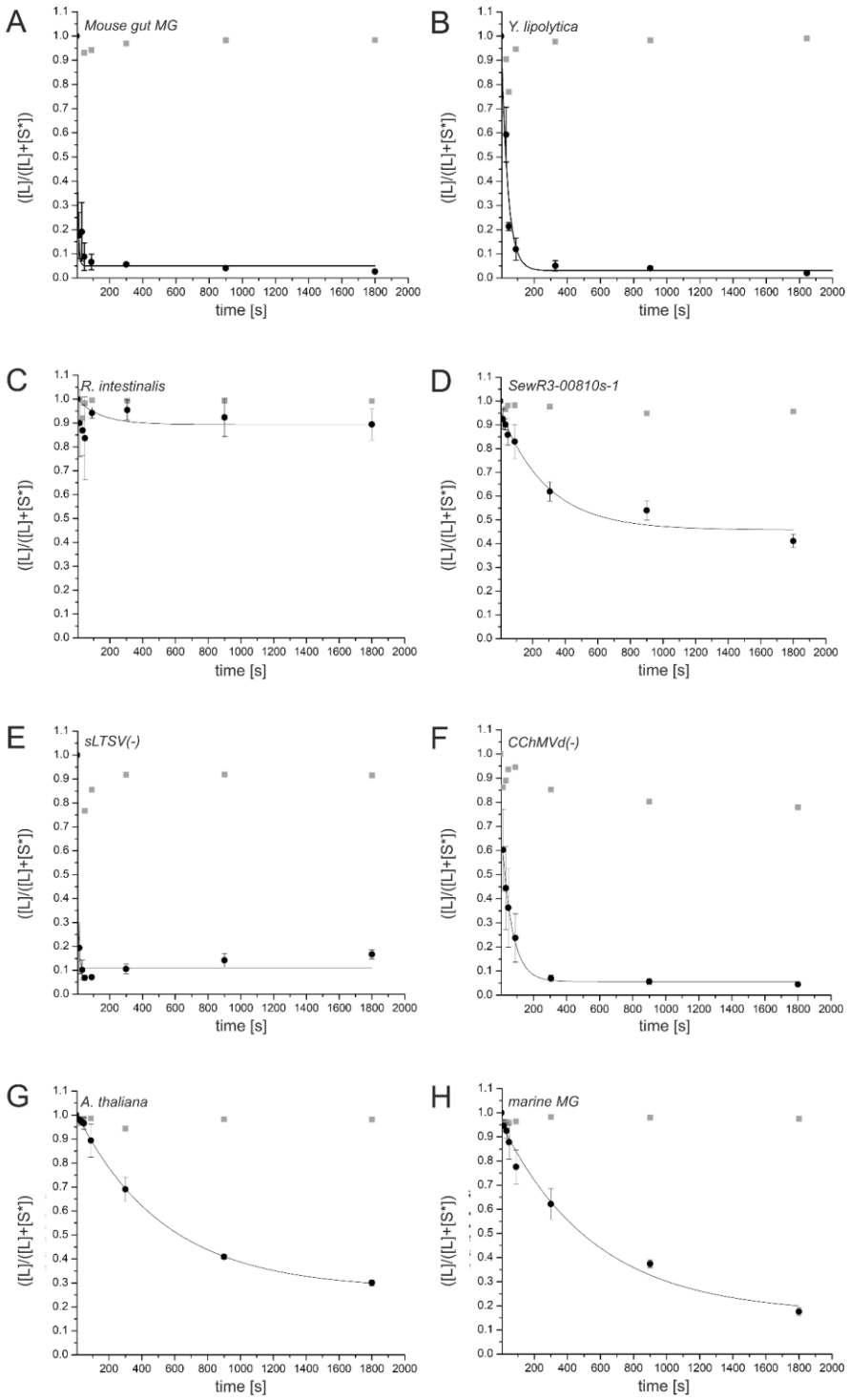
mixes for PCR reaction. C_qs were determined using the qPCR 3.2 software (Biometra, Analytik Jena). A water control was added to each plate. The parental expression construct lacking ribozyme sequences was used for comparison to a non eGFP expressing plasmid system in TOP 10 *E. coli* cells and was added to every qPCR run on the plate as additional internal control. Outlier samples were determined by their melting curves or by comparing with No-template control (NTC) values (distance at least ~10 C_q). Samples were further analysed by delta-C_q method, leading to a $\Delta C_q = C_q$ *(sample)* $- C_q$ *(reference gene).* Established and comparable primer efficiencies allowed for further description of relative expression ratios to reference gene given by the formula ratio = $2^{-\Delta Cq}$, with $\Delta Cq = Cq$(eGFP) - Cq(ssrA), and $2^{-\Delta\Delta Cq}$ indicating the relative changes between constructs containing of active versus inactive HHR motifs (**Fig. 4C and 5C**). Significance was tested with unpaired t-test two-tailed with not significant (ns), $p < 0.05$ (*) and $p < 0.01$ (***) (**Fig. S4B und S4C).**

Supplementary Note 5: Considerations regarding 5´-UTR-inserted HHR motifs in *E. coli*

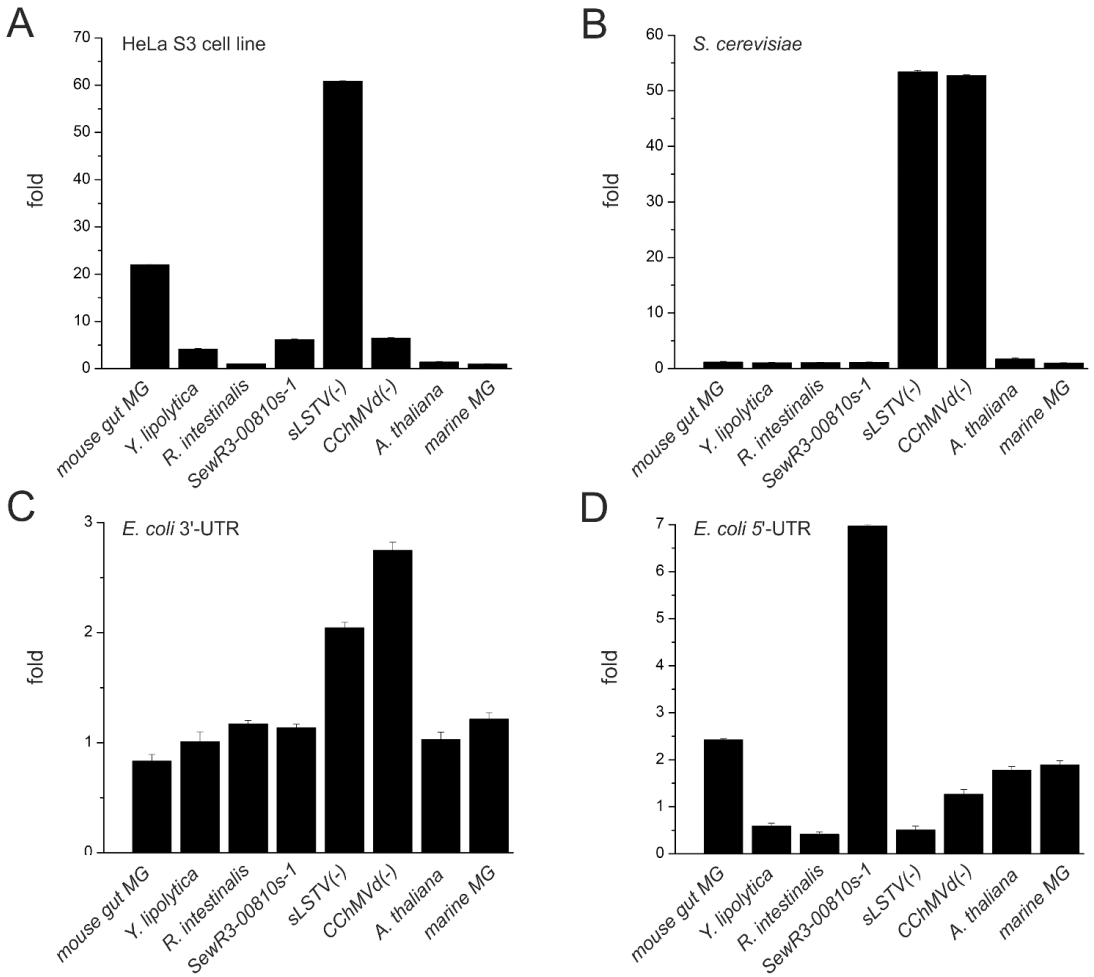
Insertion of ribozymes or secondary structures into 5´-UTRs for controlling of downstream gene expression is discussed. In bacteria the start codon as well as the Shine Dalgarno (SD) site is necessary for initiation of translation. Therefore, alternative start codons within the 5´-UTR should not influence reporter gene expression. Alternative open reading frames (ORFs) could be ruled out as described in **Supplementary Note 6**. We also folded the designed 5´-UTRs of the mRNAs including the complete 5´-end and the HHR motifs until the start codon for detecting alternative secondary structures or for folds that would sequester the SD site [16] via mfold (http://unafold.rna.albany.edu/). In the constructs used for our study, HHR motifs were folded properly and the SD site was not involved in secondary structures formations (data not shown).

Supplementary Note 6: Alternative open reading frames (ORFs) by insertion of HHR motifs with containing start codons in 5´-UTR of *E. coli*

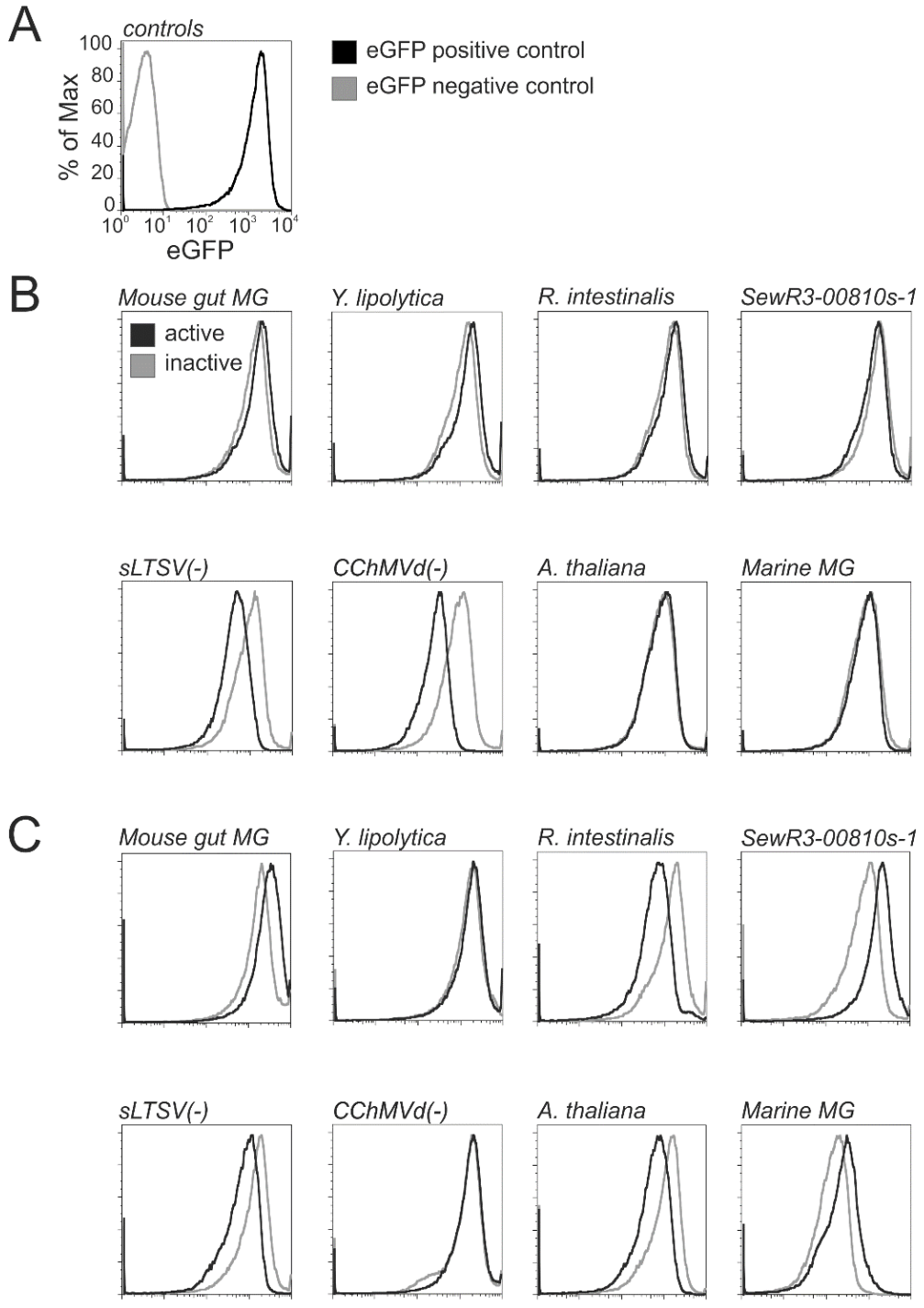
For inserted HHR motifs in the 5´-UTR of the eGFP reporter gene in *E. coli* plasmid constructs, we checked all designed 5´-UTRs for alternative translation starts by the translation webtool *http://web.expasy.org/translate/* to rule out the generation of upstream ORFs that could interfere with expression of the eGFP ORF. Additionally, we manually searched for possible alternative SD sites upstream to introduced start codons. Importantly, we were unable to identify alternative / upstream ORFs to the eGFP reporter gene in the utilised constructs.



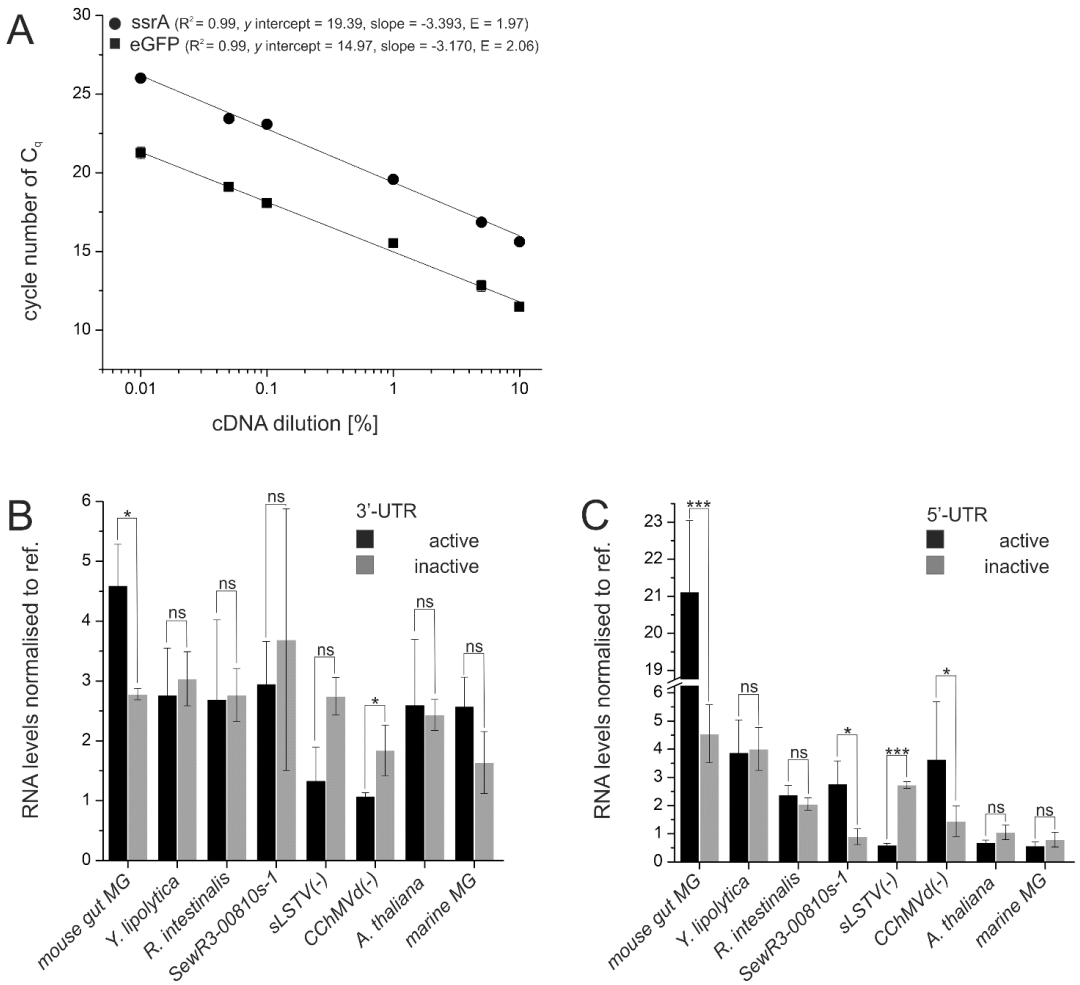
Supplementary Figure S1. Ribozyme cleavage kinetics of intramolecular HHR motifs. *In vitro* cleavage kinetics for the investigated HHR motifs (A-H) determined during *in vitro* transcription. Full-length transcript (L) and cleaved transcript (S) with normalisation (*) to A-residues within the transcript which are visualised via ^{32}P-imaging. The data were fit using mono-exponential fit, k_{obs} [min^{-1}] was calculated (Table 1). Error bars represent experimental triplicates.



Supplementary Figure S2. Fold changes of expression calculated from bulk measurements for expression systems in different host organisms. (A) mammalian HeLaS3 cell line, (B) *S.cerevisiae* MAV203 strain, in (C) *E. coli* TOP10 3´-UTR to eGFP and (D) *E. coli* TOP10 5´-UTR to eGFP. Fold changes in gene expression levels were calculated in (A - C) as the ratio of inactive divided by active expression state and as the ratio of active divided by inactive expression state (D). Error bars represent the propagation of uncertainty of three biological replicates.



Supplementary Figure S3. Comparative flow cytometric single-cell analysis to study uniformity of bacterial populations under influence of inserted HHR motif constructs. Cultures of the *E. coli* TOP10 strain were grown at 37 °C and Flow cytometric single-cell analysis was performed for (A) controls with and without eGFP, (B) 3´-UTR and (C) 5´-UTR insertions. One replicate of an active HHR construct (black) and one of the inactive HHR construct (grey) are depicted as representatives for the biological replicates run in at least triplicates.



Supplementary Figure S4. RNA level analysis by quantitative real-time PCR. (A) The amplification rate (E) for GFP and ssrA primer pairs were tested and calculated with the slope of dilution curves depicted, according to the equation: $E = 10^{[-1/slope]}$. Transcripts showed values of E(ssrA), 1.97 and E(eGFP), 2.06, displaying doubling of RNA. According to that RNA levels can be calculated with $2^{-\Delta Cq}$. RNA levels for (B) representatives of HHR motifs inserted into the 3´-UTR of *E. coli* reporter system and (C) representatives of HHR motifs inserted into the 5´-UTR of the *E. coli* reporter system. Cultures of the *E. coli* strain were grown at 37 °C until OD_{600} of 1.2. Transcript levels of eGFP were determined from determined by qRT-PCR using RNA purified from these cultures. RNA level values normalised to reference gene, equal to $2^{-\Delta Cq}$ are depicted. Black bars: reporter gene containing catalytically active HHRs, grey bars: reporter gene with catalytically inactive HHR. Error bars represent the standard deviation of experiments performed in technical triplicates of biological triplicates. Significance was tested with unpaired t-test two-tailed with not significant (ns), $p < 0.05$ (*) and $p < 0.01$ (***).

Supplementary Table S1. DNA sequences of the selected HHR motifs. Start codons are highlighted in red and 3´-end after cleavage is underlined.

HHR motif	**Sequence with potential start codons coloured in red, cleaved 3'end underlined**
Mouse gut metagenome	GGTACCGAATAAATCCCCTG**ATG**AGCAACG**GTG**AGAGCCGGCGAAACTACCCAAACAAG GGTAGTCGGGATAGTACCATAA
Yarrowia lipolytica	GGGGGACTGGCTGCCCTG**ATG**AGAACAAACCC**ATG**ACTAGCGTCGAAACATCAACA**GTG** GGGGCTGTTG**GTG**TCGGCAGCCACTAGTCATAA
Roseburia intestinalis	ATATTAAGCCGAAACGCTCCGATCTGGAGCGTCAGCTGCGGACCGGTCGCCGCGGCTCTG ACG**ATG**GCAGACCAA
SewR3-00810s-1	AGGAAGAAGAGGATAACAGTCGAAACTCCCTTGACCGGGAGTCTACCGGGCTTGCTCTCC CGGTACTG**ATG**AGACAGAGCAACA
sLTSV(-)	GACGT**ATG**AGACTGACTGAAACGCCGTCTCACTG**ATG**AGGCC**ATG**GCAGGCCGAAACGT C
CChMVd(-)	TTCCAGTCGAGACCTGAA**GTG**GGTTTCCTG**ATG**AGGCT**GTG**GAGAGAGCGAAAGCTTTAC TCCCGCACAAGCCGAAACTGGAA
Arabidopsis thaliana	TGGTC**GTG**ATCTGAAACTCGATCACCTG**ATG**AGCTCAAGGCAGAGCGAAACCA
Marine metagenome	GC**GTG**TCGGCCACGGCCCCTTCTGGACCTCGTCC**GTG**GCCCTGACGAGTAGGGTCCAGAG GGGACGAAACACGC

Supplementary Table S2. Genomic location and blasts of the HHR motif selection in their original organisms. Genomic locations of selected HHR motifs in their original organism context were blasted using nucleotide BLAST tool (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Alignments and gene annotations were extracted from blast results. Localisations and gene surrounding are demonstrated in the following table. All of the selected HHR motifs have a natural intergenic location. For the three HHR motifs of the Mouse gut metagenome, SewR3-00810s-1, and Marine metagenome no origin-localisations could be done so far.

HHR motif	Organism type	HHR motif type (intragenic, intergenic, repeat element, etc.	position	gene surrounding			
				upstream gene	distance to upstream gene [nt]	distance to downstream gene [nt]	downstream gene
Mouse gut metagenome	DNA from fecal samples of mice	no characteristics known so far					
Y. lipolytica	fungus	interspersed repeat elements, intergenic HHRs within *Y. lipolytica* WSH-Z06 chromosomes YALIOA, YALIOF, YALIOC, YALIOD and *Y. lipolytica* CLIB122 chromosomes A, D, C and F	*Y. lipolytica* CLIB122, chromosome A, 1073559 to 1073650	complement YALI0_A10461g, pseudo, part. start, part. stop	800	4633	complement CAG83873.1
			Y. lipolytica CLIB122, chromosome A, 926754 to 926845, 2 mutation sites	complement YALI0_A09361g, pseudo, part. start, part. stop	859	1741	complement CAG83827.1
			Y. lipolytica CLIB122, chromosome D, 1990467 to 1990558, 1 mutation site	YALI0_D16054t	263	4515	YALI0_D16038g, pseudo, part. start, part. stop
			Y. lipolytica CLIB122, chromosome D, 1831981 to 1832072, 2 mutation sites	complement CAG81032.1	4133	2719	CAG81031.1
			Y. lipolytica CLIB122, chromosome C, 2071686 to 2071777, 1 mutation site	divergent CAG82161.1.1	7306	2915	complement tRNA-Ser
R. intestinalis	saccharolytic, butyrate-producing, gram-positive bacterium	single intergenic HHR between two hypotheticla genes, associated with bacteriophage related sequences	R. Intestinalis M50/1 draft genome, 3976828 to 3976902	CBL10707.1 hypothetical protein	-	30	CBL10708.1 hypothetical protein
SewR3-00810s-1	uncultured organism clone	no characteristics known so far					
sLTSV(-)	lucerne transient streak virus	form HHRs in both polarity strands, likely involved in rolling circle mechanism	-				
CChMVd(-)	chrysanthemum chlorotic mottle viroid	form HHRs in both polarity strands, likely involved in rolling circle mechanism	-				
A. thaliana	plant	HHR-containing sequences are expressed in various tissues, found in intergenic in tandem like position	A. thaliana chromosome 4, 15027830 to 15027882	complement SDG4	214	545	MUS81 (endonuclease)
			A. thaliana chromosome 4, 15032895 to 15032843, 2 mutations	AT4G30872 (RNA)	83	245	complement MUS81 (endonuclease)
Marine metagenome	DNA from marine environment	no characteristics known so far					

Supplementary Table S3. Transcription templates for *in vitro* ribozyme cleavage kinetics.

HHR motif	**PCR product for transcription (bold = T7 promoter, cleavage site marked by (-), HHR sequence underlined, bold grey = inactivation site (A to G)**
R. intestinalis	GAAAT**TAATACGACTCACTATAGGGAG**caaaATATTAAGCCGA**A**ACGCTCCGATCTG GAGCGTC-AGCTGCGGACCGGTCGCCGCGGCTCTGACGATGGCAGACCAA
SewR3-00810s-1	GAAAT**TAATACGACTCACTATAGGGAG**caaaAGGAAGAAGAGGATAACAGTCGA**A**A CTCCCTTGACCGGGAGTC-TACCGGGCTTGCTCTCCCGGTACTGATGAGACAGAGCAACA
A.thaliana	GAAAT**TAATACGACTCACTATAGGGAG**caaaTGGTC-GTGATCTGAAACTCGATCACCTGATGAGCTCAAGGCAGAGCGA**A**ACCA
Marine metagenome	GAAAT**TAATACGACTCACTATAGGGAG**caaaGCGTGTC-GGCCACGGCCCCTTCTGGACCTCGTCCGTGGCCCTGACGAGTAGGGTCCAGAGGGGA CGA**A**ACACGC
Mouse metagenome	GAAAT**TAATACGACTCACTATAGGGAG**caaaGGTACCGAATAAATCCCCTGATGAGC AACGGTGAGAGCCGGCGA**A**ACTACCCAAACAAGGGTAGTC-GGGATAGTACCATAAATGGTAAGTGGAGAACCTATTGGGTATTAAAGAGGAGAAGGT ACC
Y. lipolytica	GAAAT**TAATACGACTCACTATAGGGAG**caaaGGGGGACTGGCTGCCCTGATGAGAAC AAACCCATGACTAGCGTCGA**A**ACATCAACAGTGGGGGCTGTTGGTGTC-GGCAGCCACTAGTCATAAATGGTAAGTGGAGAACCTATTGGGTATTAAAGAGGAGAA GGTACC
sLTSV(-)	GAAAT**TAATACGACTCACTATAGGGAG**caaaGACGTA-TGAGACTGACTGAAACGCCGTCTCACTGATGAGGCCATGGCAGGCCGA**A**ACGTCATG GTAAGTGGAGAACCTATTGGGTATTAAAGAGGAGAAGGTACC
CChMVd(-)	GAAAT**TAATACGACTCACTATAGGGAG**caaaTTCCAGTC-GAGACCTGAAGTGGGTTTCCTGATGAGGCTGTGGAGAGAGCGAAAGCTTTACTCCCG CACAAGCCGA**A**ACTGGAAATGGTAAGTGGAGAACCTATTGGGTATTAAAGAGGAGA AGGTACC

Supplementary Table S4. Insertion area of the selected HHR motifs in expression systems for mammalian, *S. cerevisiae* and *E. coli* constructs. HHR motifs depicted in Supplementary Table 1 are inserted as shown here. Start or stop triplets are depicted in bold. Additional flanking areas are *italic*. SD site (-13 to -7) is underlined.

Organism expression system	**Insertion area of HHR motifs in 5'3' direction**
Mammalian cell line (3'-UTR)	**TAA**TTCTAGGCGACTAGT*AAACAAACAAA* - **HHR motif** – *AAAAAGAAAAATAAAAA*GCGGCCGC
***S. cerevisiae* (3'-UTR)**	**TAA**TAATACTAGT*AAACAAACAAA* - **HHR motif** – *AAAAAGAAAAATAAAAA*GCGGCCGC
***E. coli* (3'-UTR)**	**TAA**GGATCCGGCTGCTAACAAAGCCCGAAAGGAA- **HHR motif** – GCTGAGTTGGCTG
***E. coli* (5'-UTR)**	**Promoter**-ACTAGTGGACG- **HHR motif** –TTTAGCTTTAAG-<u>AAGGAG</u>ATATACC**ATG**

Supplementary References

1. Perreault J, Weinberg Z, Roth A, Popescu O, Chartrand P, Ferbeyre G, Breaker RR. Identification of hammerhead ribozymes in all domains of life reveals novel structural variations. PLoS computational biology 2011; 7:e1002031.
2. Hammann C, Lupták A, Perreault J, de la Peña M. The ubiquitous hammerhead ribozyme. RNA (New York, NY) 2012; 18:871-85.
3. Jimenez RM, Delwart E, Lupták A. Structure-based search reveals hammerhead ribozymes in the human microbiome. The Journal of biological chemistry 2011; 286:7737-43.
4. Forster AC, Symons RH. Self-cleavage of plus and minus RNAs of a virusoid and a structural model for the active sites. Cell 1987; 49:211-20.
5. Navarro B, Flores R. Chrysanthemum chlorotic mottle viroid: unusual structural properties of a subgroup of self-cleaving viroids with hammerhead ribozymes. Proceedings of the National Academy of Sciences of the United States of America 1997; 94:11262-7.
6. Przybilski R, Graf S, Lescoute A, Nellen W, Westhof E, Steger G, Hammann C. Functional hammerhead ribozymes naturally encoded in the genome of *Arabidopsis thaliana*. The Plant cell 2005; 17:1877-85.
7. Yen L, Svendsen J, Lee JS, Gray JT, Magnier M, Baba T, D'Amato RJ, Mulligan RC. Exogenous control of mammalian gene expression through modulation of RNA self-cleavage. Nature 2004; 431:471-6.
8. Meaux S, Van Hoof A. Yeast transcripts cleaved by an internal ribozyme provide new insight into the role of the cap and poly(A) tail in translation and mRNA decay. RNA (New York, NY) 2006; 12:1323-37.
9. Ketzer P, Haas SF, Engelhardt S, Hartig JS, Nettelbeck DM. Synthetic riboswitches for external regulation of genes transferred by replication-deficient and oncolytic adenoviruses. Nucleic acids research 2012; 40:e167.
10. Ausländer S, Ketzer P, Hartig JS. A ligand-dependent hammerhead ribozyme switch for controlling mammalian gene expression. Molecular bioSystems 2010; 6:807-14.
11. Kozak M. Adherence to the first-AUG rule when a second AUG codon follows closely upon the first. Proceedings of the National Academy of Sciences of the United States of America 1995; 92:7134.
12. Klauser B, Atanasov J, Siewert LK, Hartig JS. Ribozyme-Based Aminoglycoside Switches of Gene Expression Engineered by Genetic Selection in *S. cerevisiae*. ACS synthetic biology 2014; 4:516-25.
13. Hui MP, Foley PL, Belasco JG. Messenger RNA Degradation in Bacterial Cells. Annual Review of Genetics 2014; 48:537-59.
14. Smolke CD, Keasling JD. Effect of copy number and mRNA processing and stabilization on transcript and protein levels from an engineered dual-gene operon. Biotechnology and bioengineering 2002; 78:412-24.
15. Celesnik H, Deana A, Belasco JG. Initiation of RNA decay in *Escherichia coli* by 5' pyrophosphate removal. Molecular cell 2007; 27:79-90.
16. Isaacs FJ, Dwyer DJ, Ding C, Pervouchine DD, Cantor CR, Collins JJ. Engineered riboregulators enable post-transcriptional control of gene expression. Nature biotechnology 2004; 22:841-7.
17. Espah Borujeni A, Channarasappa AS, Salis HM. Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. Nucleic acids research 2014; 42:2646-59.
18. Sambrook J, Fritsch EF, Maniatis T. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press, 1989.