universität wien

DIPLOMARBEIT

Titel der Diplomarbeit

Characterization of the Hfq binding cis antisense RNA as-htrG-cca

angestrebter akademischer Grad

Magister der Naturwissenschaften (Mag. rer.nat.)

Verfasser:	Peter Hajdusits
Matrikel-Nummer:	a0200955
Studienrichtung:	A 490 Diplomstudium Molekulare Biologie
Betreuerin / Betreuer:	Univ. Prof. Dr. Renée Schroeder

Wien, am 26.11.2009

Table of Contents

1) Abstract....4
2) Zusammenfassung....6
3) Introduction....8
3.1) Central Dogma of Molecular Biology....8
3.2) The catalytic function of RNA....9
3.3) The RNA world hypothesis....11
3.4) Noncoding RNAs....12
3.5) Bacterial noncoding RNAs....13
3.5.1) ncRNAs acting via basepairing (asRNAs)....13
3.5.2) ncRNAs directly modifying the function of proteins....18
3.6) Long noncoding RNAs....18
3.7) tRNA processing and the htrG-cca operon....20
3.8) Hfq, a multifunctional protein....22
3.9) Genomic SELEX....24
4) Aims of the project....26
5) Preliminary Data....27
5.1) gemonic SELEX with Hfq....27
5.2) genomic SELEX outcome....28
5.3) Hfq binding motif....28
5.4) Antisense transcript as-cca....29
5.5) as-cca Hfq binding sequence....31
6) Results....32
6.1) Strand specific primer walking....32
6.1.1) Primer design....32
6.1.2) 3' end mapping....36
6.1.3) 5' end mapping....37
6.2) RT-PCR with two flanking primers....39
6.3) Simultaneous mapping of 5' and 3' ends (Circularization) 41
6.4) Northern Blots....46
6.4.1) Northern Blots with radiolabeled RNA probes....46
6.4.2) Northern Blots with radiolabeled LNA probes....47
6.4.3) simultaneous hybridization with two LNA probes....51
7) Discussion....54
7.1) htrG-cca locus and the CCA adding enzyme....54
7.2) Quality of the RNA....55
7.3) Primer design....56
7.4) Comparison of the three methods....57
7.4.1) Strand specific primer walking....57
7.4.2) Simultaneous mapping of 5' and 3' ends (Circularization)....58
7.4.3) Northern Blots....60
7.5) Multiple Bands....61
7.6) Computational Analysis....64
7.6.1) Promoter sequence analysis....64
7.6.2) Transcription terminator sequence analysis....64
7.7) Multiple transcripts....66
7.8) Expression pattern....68
7.9) Possible Functions....69
8) Conclusions and Future Prospects....71

9) Materials and Methods....73
9.1) Used Bacterium strain and Media....73
9.2) Growth and Stress conditions....73
9.3) Isolation of total RNA from E. coli Cells....74
9.4) Primer Walking....76
9.5) Simultaneous mapping of 5' and 3' ends (Circularization)....78
9.6) Northern hybridization....82
10) Appendix....86
10.1) Table PRIMER....86
10.2) circularized sequences....89
11) References....91
12) CV....97
13) Lebenslauf....100
14) Acknowledgements....103

1. ABSTRACT

Cis antisense RNAs are encoded opposite of protein coding genes and they act on their target RNAs via base-pairing. In bacteria, the already known cis-as-RNAs are between 50nt and 300nt long and their major role is post transcriptional inhibition of the target RNA's functions. This could have a positive or a negative effect on fundamental cellular processes like transcription, translation or mRNA stability.

A genomic SELEX for RNAs that bind Hfq with high specificity successfully identified 85 cis antisense RNA fragments in the bacterium Escherichia Coli. For further investigation, I chose the Hfq binding cis antisense RNA htrG-cca, lying opposite to the genes htrG and cca. The bacterial cca gene encodes the enzyme tRNA-nucleotidytransferase which repairs the 3' terminal ends of transfer RNAs (tRNAs) whereas htrG's function is only speculative yet.

The aim of my thesis was to analyze and characterize the cis antisense RNA as-htrG-cca. In a first experiment, based on Reverse Transcriptase PCR (RT-PCR), I narrowed down the 5' and 3' ends of the investigated transcript. The calculated distance between the two ends was at least 2205nt. Surprisingly, the outcome of a second experiment in which I circularized the RNA by head to tail ligation, suggested the existence of many shorter antisense transcripts at the same locus all with varying length. To collect more reliable data regarding the length and abundance of as-htrG-cca, I performed a northern blot with sequence specific probes. Two probes were specific for the antisense and two for the sense transcripts. Thereby I identified a single 1200nt long antisense transcript, expressed in cells grown until logarithmic phase. Additionally I could confirm the actual length of the htrG mRNA, which was ~900nt long and hence also the hitherto unknown length of its ~170nt long 3' UTR. Both, the antisense as well as the sense transcripts were low abundantly expressed. As a possible explanation for the discrepancies of the achieved results, concerning the transcript's length I would assume that there is one 1200nt long antisense RNA which is sufficiently abundant to be recognized with northern blot. At the same time, there are many low abundant RNAs of different length at the same locus, only detectable with the highly sensitive RT-PCR based methods.

Concluding my abstract, I could successfully characterize the cis antisense RNA as-htrG-cca. Its ~1200 nt length and the proof that it is expressed from the antisense strand were the key

findings of my thesis. The novelty of these results was that until now there was no such huge cis antisense RNA identified. This is an evidence for a new group of long cis antisense RNAs.

2. ZUSAMMENFASSUNG

Cis antisense RNAs liegen gegenüber von Protein kodierenden Genen am entgegengesetzten DNA Strang und wirken auf ihre Ziel-RNAs durch Basenpaarung. Die bisher bekannten bakteriellen cis antisense RNAs sind zwischen 50 und 300 Nukleotide lang. Ihre maßgebliche Rolle widerspiegelt sich in der post-transkriptioneller Inhibition der Funktionen ihrer Ziel-RNAs. Diese Inhibition kann einen positiven oder einen negativen Effekt auf fundamentale Prozesse wie Transkription, Translation oder mRNA Stabilität haben.

Durch ein genomic SELEX (engl. für: Systematische Evolution von Liganden durch exponentielle Anreicherung) Verfahren für RNAs, welche das Protein Hfq mit einer hohen Spezifität binden, konnten im Bakterium Escherichia Coli unter anderem 85 cis antisense RNA Fragmente identifiziert werden.

Für weitere Untersuchungen wählte ich die Hfq bindende cis antisense RNA htrG-cca, welches gegenüber der Gene htrG und cca liegt. Das bakterielle Gen cca kodiert für das Enzym tRNA-Nukleotidyltransferase, welches die 5' Enden der Transfer-RNA (tRNA) repariert. Über htrGs Funktion ist bisher nichts Konkretes bekannt.

Das Ziel meiner Diplomarbeit war es, die cis antisense RNA as-htrG-cca zu untersuchen und zu charakterisieren. In einem ersten Experiment, welches auf die Technik 'Reverse Transkriptase Polymerase Kettenreaktion' (RT-PCR) basiert, konnte ich die 5' und die 3' Enden des zu untersuchenden Transkripts eingrenzen. Die errechnete Distanz zwischen den zwei Enden war mindestens 2205 Nukleotide. Erstaunlicherweise hat das Resultat eines zweiten Experiments - bei dem ich die RNAs durch 3' zu 5' Ligation zirkularisiert habe - ergeben, dass auf demselben Genlokus mehrere, jedoch kürzere antisense Transkripte liegen, die alle unterschiedliche Längen haben. Um aber verlässlichere Ergebnisse über die Länge und Quantität von as-htrG-cca zu sammeln, habe ich einen Northern Blot mit sequenz-spezifischen Sonden durchgeführt. Zwei Sonden waren spezifisch für das sens Transkript und zwei für das antisense Transkript. Dadurch wurde ein einzelnes, 1200nt langes antisense Transkript gefunden, welches nur in logarithmischen Zellen exprimiert wird. Weiters konnte ich die tatsächliche Länge der htrG mRNA bestätigen. Diese war ~900 Nukleotide lang und wies eine 170nt lange 3' untranslatierte Region (UTR) auf, welche bisher nicht bekannt war. Sowohl das sens, als auch das antisense Transkript sind niedrig exprimiert.

Die Länge der Transkripte betreffend würde ich die Unstimmigkeiten in den unterschiedlichen Untersuchungsmethoden folgendermaßen deuten: auf dem htrG-cca Lokus ist ein 1200nt langes Transkript exprimiert, welches in ausreichender Konzentration vorliegt um durch ein Northern Blot Verfahren detektiert zu werden. Zugleich gibt es mehrere, niedrig exprimierte RNAs mit unterschiedlichen Längen, welche man nur durch die hochempfindliche RT-PCR Methode nachweisen kann.

Zusammenfassend konnte ich die antisense RNA as-htrG-cca erfolgreich beschreiben und charakterisieren. Ihre ~1200nt Länge und der Beweis, dass sie tatsächlich auf dem antisense Strang exprimiert wird, waren die wichtigsten Erkenntnisse meiner Arbeit. Die Neuheit dieser Resultate war, dass bis dato keine so große cis antisense RNA bekannt war. Dies könnte ein Indiz für eine neue Gruppe von 'großen cis antisense RNAs' sein.

3. INTRODUCTION

3.1 Central Dogma of Molecular Biology

The long lasted dogma of molecular biology, DNA makes RNA makes protein has been true for more than 50 years and is still accepted. The DNA's function as a holder of genetic information has been indisputable since 1952. However, the knowledge about RNA's diversified functions has been marginal, not to say constricted for a longer period of time. It was upgraded several times and is still under construction. New RNAs were discovered and 'old' RNAs obtained new functions.

After the proof that DNA is the holder of genetic information in 1952 with the Hershey and Chase experiments (Hershey and Chase, 1952) and the publication of the most important scientific paper in the 20th century, the Nobel Prize winning paper „Molecular Structure of Nucleic Acids“, by James Watson and Francis Crick (Crick and Watson, 1953) the major pressure in science lay on the discovery of a model explaining the flow of genetic information, carried by DNA, to protein synthesis. Only a few years later in 1956 Francis Crick pronounced the „Central Dogma of molecular biology“ and published it in 1970 in Nature (Crick, 1970). It claims that genetic information is solely transferred by the three biopolymers DNA, RNA and protein, which are constructed by using biopolymers as a template. DNA can be copied to DNA (replication) or to mRNA (transcription) and proteins can be synthesized from the mRNA as template (translation). The basic idea is that the information flow is only one-way, from DNA to protein.

At this time it was believed, that the origin of life strongly correlated exclusively with DNA, because it carried all the information we need to live and to develop. In this theory, the RNA molecules had only the function of transmitting the signal from the DNA to protein. However, in the last decade the knowledge about the function of RNA expanded and RNA developed to a more important molecule than ever has been thought.

After the discovery and characterization of functional RNAs like ribosomal RNAs, transfer RNAs, messenger RNAs etc., the importance of functional RNAs as a DNA inde-

pendent molecule had been established. Thus during the last decades more and more details were unveiled. It became clear that besides RNA's messenger function there were several others which took part in the regulation of the most important molecular processes. They are the so called regulatory and catalytic RNAs. Here is a short retrospection about their discovery:

3.2 The catalytic function of RNA

The first to be discovered were the self splicing RNAs in 1982. The authors around Tom Cech observed autocatalytic rearrangements of RNA molecules in Terahymena ribosomal RNA (Kruger et al., 1982). The 413bp long ribosomal RNA (rRNA) intervening sequence (IVS) underwent self-splicing when incubated under certain conditions in vitro. The excision reaction, though performed without the involvement of enzymes and small nucleolar RNPs (snRNPs), contained several enzyme-like properties that broke and reformed phosphodiester bonds.

The second newcomer was RNaseP, first described in 1971 by Sidney Altman and coworkers (Robertson et al., 1972), although its catalytic RNA moiety was not discovered until 1983. It was clear that RNaseP was a key player in tRNA maturation but it took almost 12 years to recognize that RNaseP is an endonuclease composed of RNA and protein, cleaving tRNA precursors in buffers, where its protein counterpart alone showed no activities (Guerrier-Takada et al., 1983). For their 'discovery of catalytic properties of RNA' both Tom Cech and Sidney Altman were awarded with the Nobel Prize for chemistry in 1989. They found that RNA could fold into complex enzymes and execute biochemical reactions, which was believed to be constricted to protein enzymes. Thus RNA was not just a messenger but an active participant in the chemistry of life (Thomas R Cech at nobleprize.org, 1989). Both, RNaseP as well as the self splicing RNAs have been described in more detail over the years and are manifested in nowadays understanding of molecular biology.

The next crucial step in unveiling new functions of RNA was the discovery of RNA-catalyzed RNA polymerization. The idea that RNA could independently replicate without the help of proteins was amazing, because it strengthened the evidence that an early self-

replicating system could have consisted of RNA alone. The phenomenon of RNA catalyzed RNA polymerization has been first reported 1989, although protein based RNA replication by RNA-dependent RNA-Polymerases (RdRPs) were long before known in RNA viruses (Landers, 1974). However, a new class of catalytic RNAs was found to splice together multiple oligonucleotides in Tetrahymena (Doudna et al., 1989).

And again it was Tom Cech who also played an important role in this part of the story. His group published that the above mentioned ribozymes could also catalyze RNA chain elongation with a specificity for 3‘-5‘ linked dinucleotides (Young et al., 1989). A few years later a group around P.D. Bartel isolated new ribozymes from a large pool of random sequences (Bartel et al., 1993) and thereby they found one ribozyme that could successfully produce an RNA 6-mer by using the same reaction as protein enzymes which catalyze RNA polymerization. The nucleoside triphosphates joined the growing RNA chain by 3',5' phosphodiester linkages (Ekland et al., 1996).

The latest breakthrough in the field of catalytic RNAs was the characterization of the ribosome catalytic center. Harry Noller (Noller et al., 1992) predicted and Tom Steitz (Nissen et al., 2000) proved that the catalytic center of the ribosome is composed exclusively out of RNA and the peptidyltransferase center, responsible for the peptide bond formation, is exclusively the 23S ribosomal RNA. Steitz also recognized a significant contribution of the alignment of tRNA CCA ends in the active center to its catalytic power (Moore and Steitz, 2003). It is more than worth mentioning that Steitz, together with the crystallographers Ada Yonath and Venkatraman Ramakrishnan contributed most to our today‘s knowledge that the ribosome is a ribozyme, regulated by and functioning through RNAs. This Trio was rewarded with the Nobel Prize in Chemistry 2009 "for studies of the structure and function of the ribosome". All these and other not mentioned discoveries lead to the hypothesis that RNA could have played THE major role in the origin of life and evolution.

3.3 RNA world hypothesis

The notion „RNA world“ was coined 1986 by Walter Gilbert, who postulated that in early life neither RNA nor protein were required in a primitive system if RNA was a catalyst. RNA can overtake the function of DNA as store of genetic information as well as the catalytic functions of proteins. (Gilbert, 1986).
Evidence for this theory was amongst others given by the fact that DNA is produced by the reduction of RNA molecules (2' -OH residue -> 2' -H), by the necessity of RNA primers for DNA replication and by the previously mentioned catalytic functions of RNA.

According to this theory, one could presume that in the beginning of evolution, when fast turnover was required for responding to fast environmental changes, RNA molecules had three major advantages: First of all a **lower energy cost**, as smaller RNA molecules are 'cheaper' than big multidomain proteins and the long way through translation can be skipped. Secondly the **speed of response**, as catalytic RNAs work faster and could respond immediately, whereas the gene expression factory (DNA -> RNA -> protein) is rather inert. This is comparable to neuronal signaling (fast) versus hormonal signaling (slow). The third advantage is the **higher mutation rate**. Faster mutation of a delicate molecule like RNA favors faster evolution which was necessary to leave the primordial soup. Later, when multicellular life emerged, constancy became more important and proteins overtook RNA‘s catalytic function and the more stable DNA was introduced as the cell‘s hard-disk for information storage.

Nevertheless one question still remains unsolved: "Was RNA the first genetic material or was it preceded by one or more simpler genetic materials?" (Orgel, 1998). However, we can be quite sure that RNA‘s function was as indispensable during evolution as it is in an ancient modern world.

3.4 Noncoding RNAs

Besides the protein coding mRNAs there are several non protein coding RNAs that regulate the complex network of gene expression in very diverse manners. In the last few years a tremendous number of scientific papers have been published on regulatory RNAs: micro RNAs, (miRNAs) noncoding RNAs (ncRNAs), small interfering RNAs (siRNAs), cis antisense RNAs are only a few of them. Their most striking commonalty is that they are all noncoding, which means they do not code for proteins but have regulatory functions.

The pioneers in this field were Britten and Davidson when they postulated that noncoding RNAs could function via sequence specificity. They discovered that an "activator RNA" (ncRNA) could form a sequence specific complex with a "receptor gene" (promoter) which is in turn linked to a "producer gene" (coding region) and so changes the expression of that gene (Britten and Davidson, 1969). After this, noncoding RNAs became a serious object of study. A second huge landmark on that field was for sure the Nobel Prize winning paper of Craig Mello and Andrew Fire introducing RNA interference in Caenorhabditis elegans (Fire et al., 1998). They were the first logically explaining the molecular mechanism of RNA interference, although the phenomenon of post transcriptional gene silencing was long before observed in plants.

Fire and Mello stated that experimentally introduced dsRNA in C.Elegans could interfere with its endogenous gene function. This was the consequence of a simple antisense mechanism where the injected RNA hybridized with the endogenous mRNA and so downregulated its function. The general mechanism is that artificially introduced small interfering RNAs (siRNAs) are processed by the enzyme Dicer to 20nt - 22nt ssRNA molecules which form a duplex with the target mRNA. The mRNA is then degraded by the RNA induced silencing complex (RISC). These findings were the beginning of a new field in molecular biology that concentrated on the post transcriptional silencing of genes. RNAi has become a powerful tool for gene inhibition in plants, worms and flies.

The terms ncRNA (non-coding RNA), snmRNA (small non-messenger RNA), fRNA (functional RNA), sRNA (small RNA), eRNA (effector RNA) or RNAreg (RNA regulator) are of-

ten used in an equivalent meaning. I will use the most general term 'noncoding RNA (ncRNA)', since the other terms implicate a categorical size or function.

3.5 Bacterial noncoding RNAs

Bacterial noncoding RNAs have been identified in the 1970s, but their characterization is the work from recent years. Studies from the last decade have described many hitherto unknown bacterial ncRNAs (Vogel and Wagner, 2007; C. Lorenz, unpublished data). In general they are small (50 - 300nt), even though longer transcripts have also been found recently (see 3.6 long noncoding RNAs), diffusible and highly structured (Brantl, 2007).

In bacteria, there are two different ways how noncoding RNAs can be expressed: constitutively and environmentally induced. Constitutively expressed ncRNAs are principally needed for development and maintenance of the cell. They regulate the functioning of fundamental cellular processes and the expression of endogenous genes and proteins. On the other hand, environmentally induced ncRNAs are basically required in stress responses, but also play a role in their defense mechanisms (e.g. toxin-antitoxin).

Furthermore, there are two different ways how bacterial ncRNAs can act:
- via basepairing with its target (I will term them antisense RNAs - asRNAs) or
- via directly modifying the function of a protein (protein binding RNAs - PBRs)

3.5.1 ncRNAs acting via basepairing (asRNAs)

asRNAs (reviewed in Brantl, 2007) were first discovered on *Escherichia Coli (E.Coli)* plasmid ColE1, involved in colicin E1 synthesis and immunity to colicin killing (Sherratt, Warren G,1979; Tomizawa et al., 1981). Later, asRNAs were also discovered in the genome of pro- and eucaryotes. They are expressed either constitutively (plasmid encoded) or under certain environmental conditions (chromosomal encoded).

In procaryotes, two types of antisense RNAs are distinguished, regarding their location relative to the target genes: trans encoded asRNAs and cis encoded asRNAs.

Trans encoded RNAs act apart from their transcription site and do not need full complementarity with their targets for functioning. Usually they lie in intergenic regions and work by partially unfolding stem loops. One asRNA can act on multiple mRNAs and one mRNA can be regulated by more than one asRNA.

One nice example is DsrA (Majdalani et al., 1998; reviewed in Storz et al., 2004), a 87nt long asRNA, expressed at low temperatures. It has complementarities with 5 different genes: hns, rpoS, argR, ilvlH and rbsD. Two of them, hns and rpoS are known to be regulated by DsrA. hns, a global transcriptional regulator, is inhibited and rpoS, an alternative sigma factor, is upregulated (Lease et al., 2000). The mode of action of DsrA relies on differential basepairing with multiple target mRNAs, as DsrA secondary structure exhibits three adjacent stem loops (Fig. 1). Stem one has rpoS complementarity whereas stem 2 has complementarities to two regions of the hns mRNA. One is near the ribosome binding site and one just before the stop codon. The third stem is a transcription terminator. Stems 1 and 2 can attack their target sequences by melting out independently to basepair with the respective mRNA and so inhibiting or upregulating their targets.

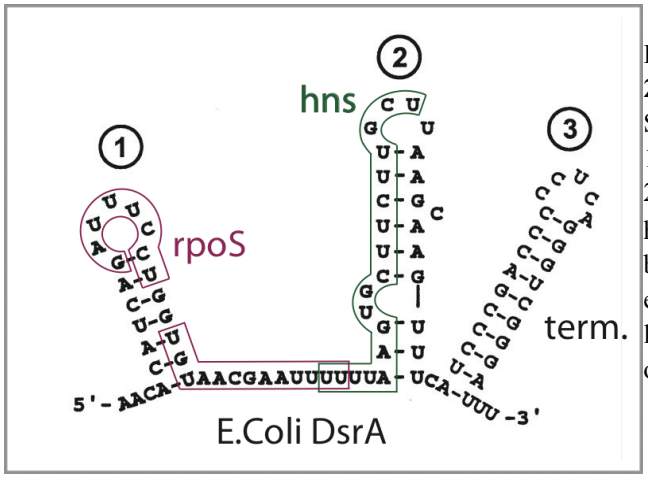


Fig. 1: E.Coli DsrA (modified from Lease et al., 2000)
Secondary structure of DsrA is indicated. Stem loop 1 has rpoS complementarity (magenta) and stem loop 2 has complementarities (green) to two regions of the hns mRNA. The 3rd stem loop is a terminator. By binding of the stem loop 1 to the 5' UTR of rpoS its expression is upregulated. Upon basepairing of stem loop 2 with hns mRNA, its central region is looped out and the gene expression is inhibited.

Cis encoded asRNAs lie in the same DNA region, only opposite to the RNA they act upon and therefore exhibit full complementarity with their target. The majority of cis acting RNAs was

found in plasmids, phages and transposons (Brantl, 2002), but the reason for this bias lies most probably in the easier accessibility of plasmids in general. With the help of novel screening methods more and more cis asRNAs have also been found recently in the bacterial chromosome (Vogel and Wagner, 2007). Their major role is post transcriptional inhibition of the target RNA's functions. This could have a positive or negative effect on fundamental cellular processes like transcription, translation, or mRNA stability (Brantl, 2007). In contrast to trans encoded RNAs, where the action of Hfq (see 3.8 Hfq, a multifunctional protein) is needed either for duplex formation or to be protected against degradation, cis encoded RNAs usually do not require additional proteins for binding (Brantl, 2007).

Six different modes of actions from the two categories of cis asRNAs are shown in Fig 2 (reviewed in Brantl, 2007). The plasmid encoded cis-asRNAs target members of the gene families involved in the regulation of plasmid replication or bacterial conjugation. One representative is the RNAIII/II system, encoded on plasmids of gram positive bacteria. Nascent rep mRNA can adopt 2 mutually exclusive conformations depending on presence or absence of asRNA RNAIII. In its presence, it anneals with rep and a premature terminator stem loop is formed upstream of its Shine-Dalgarno (SD) sequence, **inhibiting transcription and translation** of rep (-> no replication). In the absence of RNAIII, rep can fold into its natural secondary structure and no terminator loop is formed, therefore translation and replication can occur (Fig 2a).

A second system of cis as RNAs is affecting **mRNA translation** by blocking the ribosome binding site (RBS). The FinP / traJ system is encoded on the F-plasmid. traJ mRNA is the activator of the conjugal transfer operons and is controlled by FinP antisense RNA. The protein FinO affects FinP half-live by protecting it against RNase III cleavage and consequently promotes its annealing to traJ RBS. The FinP/traJ duplex is then cleaved by RNaseIII and traJ translation is blocked. In the absence of FinO, the translation of traJ leads to conjugation (Fig 2b)

A third system, Inc/repZ, acts on an **activating pseudoknot formation** of the repZ mRNA, required to prevent a terminator loop at the rep RBS. Binding of the asRNA Inc leads to inhibition of repZ translation and hence, replication (Fig 2d).

A fourth system, RNA II and its antisense RNA I, encoded on plasmid ColE1 is inhibiting plasmid replication by **preventing primer maturation** (Fig 2c).

The majority of today known cis-asRNAs encoded on bacterial chromosomes seems to affect translation or mRNA stability.

Antisense RNA mediated mRNA degradation is used by the txpA/RatA system of *B.Subtilis*. The txpA mRNA, encoding a killer toxin, is translated in absence of the asRNA RatA. Otherwise RatA anneals with txpA and the duplex is degraded by RNaseIII which results in the inhibition of toxin synthesis (Fig 2e)

The RNAs GadY/gadX in *E.Coli* exclude a **stabilizing function** by inducing an RNase III cleavage into the long and unstable gadXW mRNA, resulting in two shorter but stable transcripts. gadXW mRNA is transcribed in one piece and it is rapidly degraded in its long form. If the asRNA GadY is expressed, basepairing with gadXW leads to cleavage of the double strand region of gadXW/gadY hybrid, resulting in gadX and gadW mRNAs both involved in acid response (Fig 2f).

The initial contact of two RNAs can occur between two complementary loops (kissing loop complex) or between a loop and a single stranded region. In the case of the kissing loop complex, simple helix progression is impossible due to accumulating torsional stress. Therefore they need a second contact site distal of the initial one to avoid this limitation. The final result is often a duplex, degraded by RNaseIII. However, full duplex formation is too slow to have a decent biological effect, so many RNAs form complexes that involve only a small number of basepairs with their targets. In this case RNaseIII action is not always required for control (Brantl, 2007).

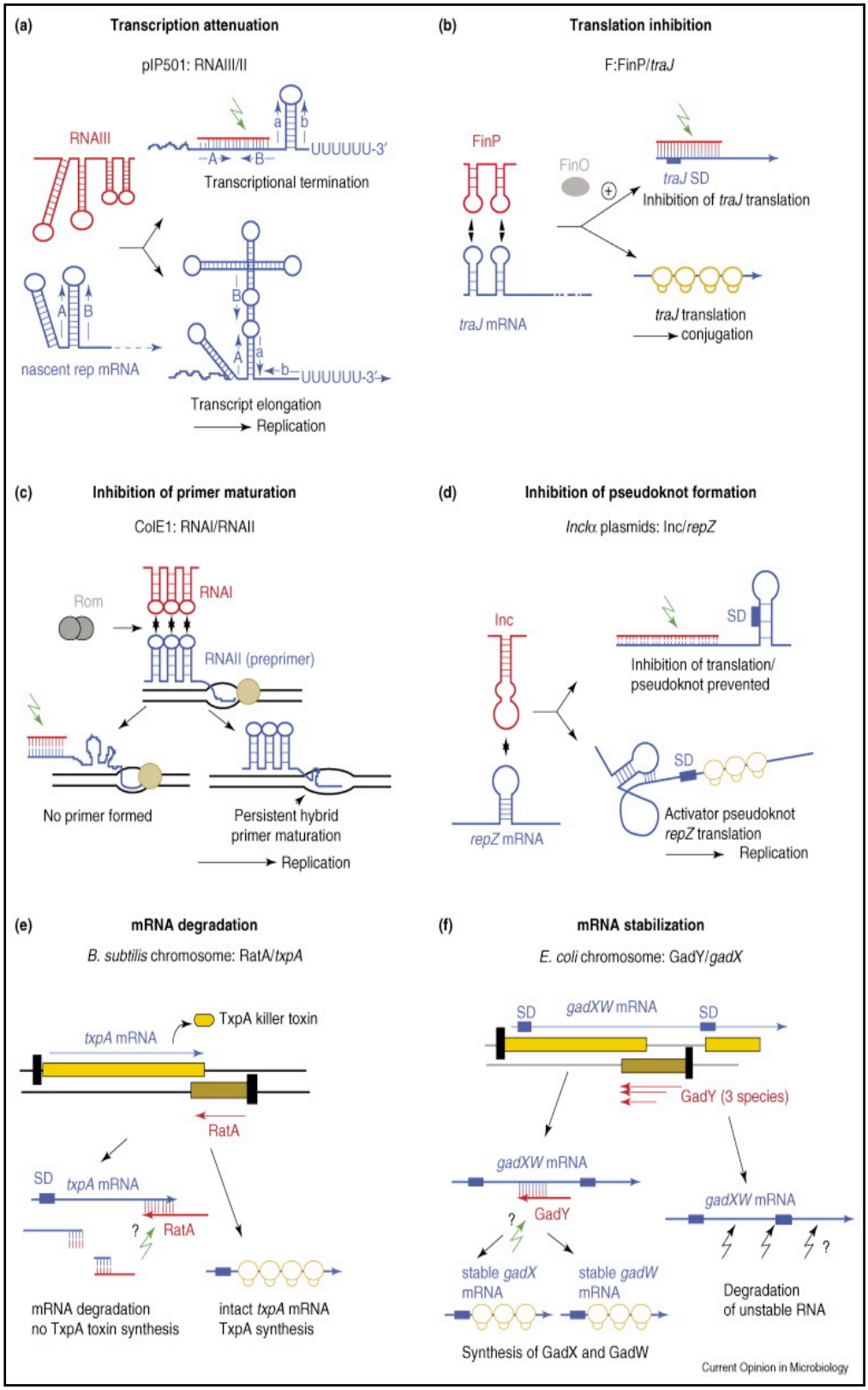


Fig. 2: regulatory mechanism of cis-encoded antisense RNAs (Brantl, 2007)
Details are described in the text. Antisense RNAs are red, sense RNAs blue. Black rectangles depict promoters, yellow and brown boxes denote sense and antisense genes. Black arrows indicate RNase action, green arrows, the action of RNaseIII (adapted from Brantl, 2007).

3.5.2 ncRNAs directly modifying the function of proteins

The second rather small group of ncRNAs directly binds target proteins and thereby modify their functions (protein binding RNAs). Although the precise mechanism has not been explored, ncRNAs directly binding to proteins have been found in pro- and eucaryotes. In both cases it is believed that the ncRNAs antagonize the nucleic acid interaction, normally made by the protein, affecting its function.

One representative is the *E.Coli* 6S RNA. It is a 184nt long, highly abundant ncRNA and binds the holoenzyme of housekeeping RNA polymerase, preventing transcription of σ^{70} dependent promoters. This highly conserved RNA forms a strong complex with the σ^{70} dependent form of the RNA Polymerase ($E\sigma^{70}$), thereby inhibiting transcription at many promoters. The key factor for inhibition is its conserved secondary structure, a mostly double stranded region with a large, single stranded bulge. The similarity to promoters in their melted conformation in the open complex formed during transcription initiation, suggests a competition with 6S RNA for promoter binding (Wassarman and Saecker, 2006).

An analog to E.Coli 6S RNA is mouse B2 RNA, although it inhibits transcription at a later step. It is a 178nt Pol III transcript, derived from short interspersed elements (SINEs). B2 is upregulated under stress conditions and associates with RNA Pol II, thereby inhibiting its function by preventing the formation of the active preinitiation complex (PIC) (Allen et al., 2004; Espinoza CA, et al., 2004).

3.6 Long noncoding RNAs

Long noncoding RNAs range between 300 and several thousand nucleotides. They lack ORFs and usually exhibit a low level of conservation among species. Although recent methods have identified a lot of ncRNA transcripts in bacteria, the number of well understood long ncRNAs is marginal. Whereas in eucaryotes, several long noncoding RNAs have been identified and the best known (Xist, Kcnq1, Air) control chromatin silencing and genomic imprinting.

The best studied RNA is probably Xist (Wutz and Jaenisch, 2000), a long ncRNA affecting the inactivation of all but one X chromosome (XCI). Xist is 15-17 kb long in mice and ~19 kb long in humans, spliced, polyadenylated and is solely associated with the inactive X chromosome (Xi). Upregulation of Xist RNA results in coating and irreversible inactivation of one X chromosome, suggesting changes in chromatin modifications. Two ncRNAs, Tsix and Xite, lie antisense to Xist and act negatively on its expression. Both transcripts are upregulated in the active X chromosome (Xa) and downregulated in Xi. A different long ncRNA is Kcnq1, affecting genomic imprinting (Thakur et al., 2004). Genomic imprinting which is an inheritance process where an allele is expressed in a parent-of-origin-specific manner, independent from the Mendelian rules. Imprinted genes are either expressed from the maternal allele, inherited from the mother, or from the paternal allele, inherited from the father (Pandey et al., 2008). Kcnq1 is encoded on the methylated KCNQ1 locus and is closely associated with the human Beckwith-Wiedemann syndrome (BWS), an overgrowth related disorder. Almost all genes of the KCNQ1 locus are expressed maternally except the paternally expressed Kcnq1ot1 ncRNA which is the antisense counterpart to Kcnq1. Kcnq1ot1 is a 91kb long nuclear transcript of RNA Polymerase II and is important for the bidirectional silencing of the Kcnq1 domain, again by chromatin remodeling (Pandey et al., 2008; Zwart et al., 2001). A similar situation is found at the igfr2 locus located on human chromosome 6 and in mouse on chromosome 17 which provides another example of ncRNA mediated regulation within imprinted loci (Umlauf et al., 2008). It has three protein coding genes which are maternally inherited. The paternally imprinted long noncoding RNA Air (for antisense Igf2r RNA) is a 108kb long Pol II transcript, antisense to the igf2r locus. Similar to Kcnq1ot1, it drives the epigenetic silencing of neighboring genes (Prasanth et al., 2007).

One of the aspects of my diploma thesis deals with the long noncoding cis antisense RNA htrG-cca which is assumed to interfere with the function of the cca gene, encoding the tRNA nucleotidyltransferase in E.Coli. tRNA nucleotidyltransferase plays a remarkable role in transferRNA (tRNA) processing in eucaryotes as well as in procaryotes, whereas the enzyme's function differs in those two domains.

3.7 tRNA processing and the htrG-cca operon

tRNA, also a functional noncoding RNA, is the metaphorical link between RNA and protein. During translation, tRNA transfers the amino acid to the growing polypeptide chain at the site of protein synthesis in the ribosome. The major function of tRNAs lies in peptide bond formation and builds, together with the 23S rRNA, the active center of the ribosomes. They have a cloverleaf shape in their primary structure and the shape of a capital L in their tertiary 3D structure. tRNAs bind the amino acids respective to their anticodons at their 3'end and transfer them to a growing polypeptide chain during translation. The 3‘ terminal sequence of mature tRNAs in all organisms is always 'CCA' and plays an important role in its functions (Schmidt et al., 1975). It is either genome encoded and already present on the pre tRNA transcripts as in *E.Coli* and most other bacteria, or it is additionally transfered by an enzyme like in prokaryotes (tRNA nucleotidyltransferase).

In both *E.Coli* and in eucaryotes, the tRNAs are processed by the interaction of several endo- and exonucleolytic cleavages (Schürer et al., 2001). In *E.Coli* an exonuclease removes nucleotides from the 3‘ end of the precursor tRNA then the endonuclease RNaseP generates its mature 5‘ end. The remaining nucleotides at the 3‘ end are endonucleolytically removed till the A of the CCA end (Fig 3A). In eucaryotes, the process differs in the maturation of the 3' end. Again two endonucleolytic cleavages are involved. The first, like in procaryotes, is performed by RNaseP and generates the mature 5‘ end. The second, carried out by a 3‘ endonuclease, generates the 3' end, ready for CCA addition. The latter enzyme varies from organism to organism, but in most cases it is tRNase Z (Späth, 2007). The 3‘ CCA terminus is then added by a third enzyme, the tRNA nucleotidyltransferase (Fig 3B). Finally, splicing and RNA editing maturate the tRNA to its eventual shape and the tRNA aminoacyltransferase is then ready to add the particular amino acid.

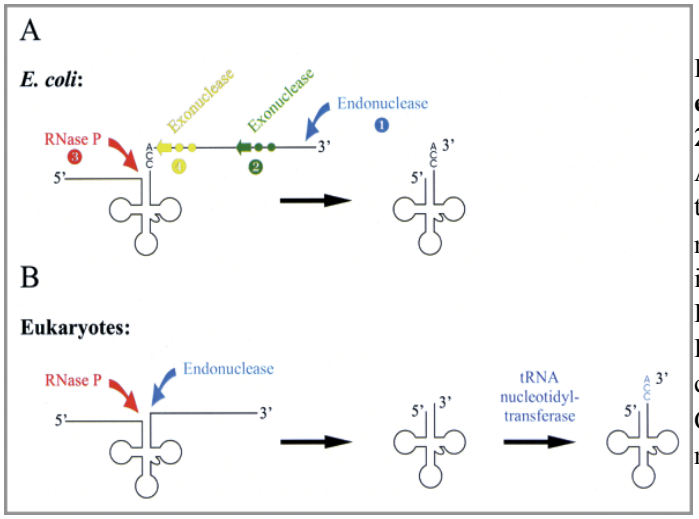


Fig. 3: comparison of tRNA maturation in eucaryotes and E.Coli (from Schürer et al., 2001)
A: In E.Coli, RNaseP maturates the 5' end and the 3' end is formed by the action of one Endonuclease and two Exonucleases. The 3' CCA end is already encoded in the E.Coli genome.
B: In Eucaryotes, tRNAs are maturated by RNaseP cleavage at the 5' end and an endonucleolytic cleavage on the 3'end. The terminal CCA end is additionally added by the tRNA nucleotidyltransferase.

Since the 3' terminal CCA triplet is already encoded in the *E.Coli* tRNA sequence, the role of the tRNA-nucleotidyltransferase does not lie in its biosynthesis but in its repair (Reichert and Mörl, 2000). An *E.Coli* strain with a defect nucleotidyltransferase is still viable, but has a slower growth rate simply because of the accumulation of defect tRNAs (Zhou and Deutscher, 1987).

As described above, the two major processing enzymes in tRNA maturation are RNaseP at the 5' end and the tRNA nucleotidyltransferase at the 3' end. The function of the latter makes a huge difference between eucaryotes, where the loss of the enzyme is fatal, and procaryotes, where the cells are still viable despite its loss. Actually, it is an evolutional mystery how the tRNA nucleotidyltransferase evolved from an auxiliary enzyme to a leading factor in tRNA processing and why the originally encoded CCA terminus got lost in the tRNA gene.

In E.Coli the tRNA nucleotidyltransferase is encoded on the htrg-cca operon. It is a 52kD protein (Shofield, 1977) encoded on the cca gene, which lies on the htrg-cca operon. The htrG-cca operon lies on the + stand of the *E.Coli* genome, encoding two genes htrG and cca. It has two different promoters from which two mRNAs are transcribed. HtrgGp is the promoter of the shorter, ~900nt, htrG transcript and HtrgGp2 promotes the long, over 1.9kb, transcript htrG-cca. The two promoters are 62bp apart of each other (Fig. 4).

The function of htrG (synonyms: EcfG, YgiM) is not exactly known yet, but it is predicted to be a small signal transduction protein (SH3 domain). It might be required for cell growth above 43°C (Dartigalongue et al., 2001).

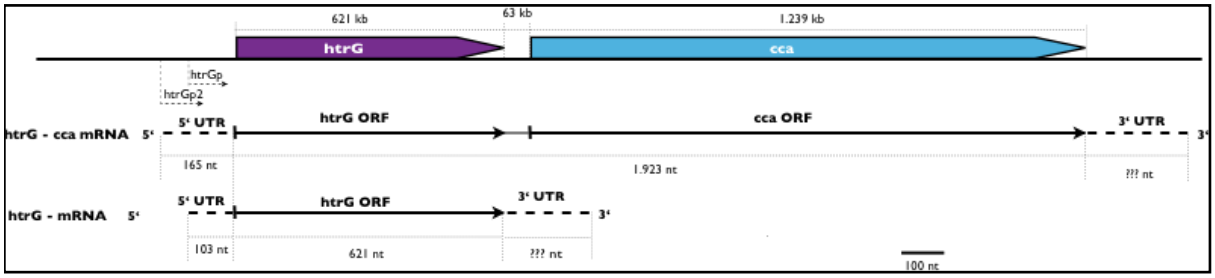


Fig. 4: htrG-cca operon
htrG cca operon lies on the + strand of E.Coli in between the genes ygiF and BacA (not shown). Two RNAs are transcribed from two different promoters. HtrG mRNA, is transcribed as a monocistronic RNA from the promoter htrGp. HtrG is also transcribed together with cca as a polycistronic mRNA from promoter htrGp2. The 3' UTR regions were not mapped. (http://regulondb.ccg.unam.mx)

3.8 Hfq, a multifunctional protein

Hfq is a highly abundant, 102 amino acid residue (11.2 kD) protein found in approximately half of the gram positive and gram negative bacteria (Brennan and Link, 2007). It was first discovered as host factor for Phage Qβ, and was required for the replication of its RNA plus-strand. The crystal structure of hitherto 4 different organisms, 3 apoproteins and one with a bound AU5G heptamer, shows that Hfq is a doughnut shaped homohexamer with an outer diameter of 70Å and a thickness of 25Å. The central pore is 8Å to 12Å wide in the apoprotein. It is a structural homolog of the eukaryotic Sm and Sm like protein family. One hast to mention that the peptides of Hfq in the X-ray structures lack the C-terminal domain. The full size protein structure is still unknown.

Today‘s knowledge is that Hfq binds RNA with high affinity to poly A and AU rich sequences, preceded or followed by a stem loop (Storz et al., 2004; Brennan and Link, 2007). A crystal structure of S.aureus Hfq, bound to an artificial AUUUUUG heptamer RNA, indicates one binding mechanism of adenosine and uridine stretches. Hfq is an RNA binding protein and there is evidence that it could work as an RNA chaperone, specific for AU rich sequences. It promotes the basepairing between DsrA and rpoS and accelerates strand displacement between the ncRNA and its target mRNA.

However, its importance is exhibited in its null mutant, which has pleiotropic effects for the cell: decreased growth rates; sensitivity to UV light, mutagens, and oxidants as well as

increased cell death. It is known that Hfq plays an important role in the function of several cis and trans acting antisense ncRNAs like DsrA, OxyS, RprA and Spot42 by facilitating base-pairing between ncRNA and mRNA (Brennan and Link, 2007)

There are two assumptions how Hfq could function: One states that it increases the accessibility of certain bases by unwinding the RNA and the other claims that Hfq increases local target RNA concentration by simultaneously binding two RNAs. (Rajkowitsch et al., 2007).

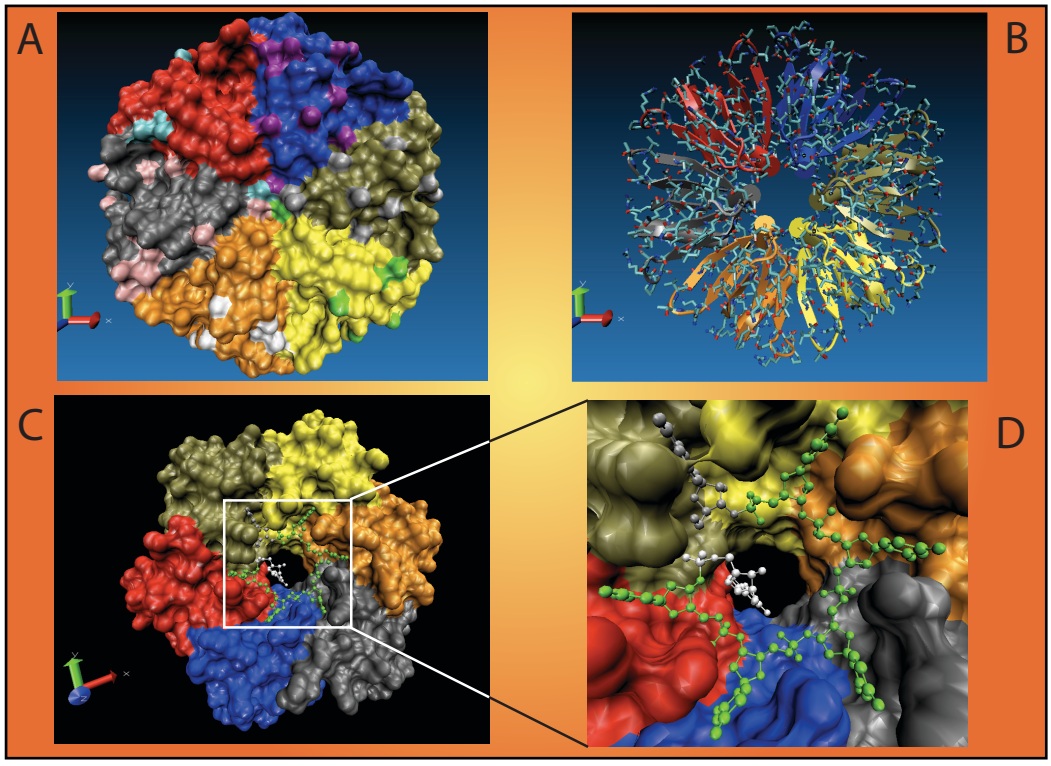


Fig. 5: 3D Model of S.Aureus and E.Coli Hfq protein.
A and B: Structure of S.Aureus Hfq (1KQ2) with a bound AUUUUUG heptamer in surface representation. The different subunits are represented in different colors, Fig. B is a zoom of the central region of Fig. A. The AU5G RNA is bound to the central pore region of Hfq in a circular manner, expanding through the pore. Adenosine is grey, Uracil green and Guanosine is white.
C and D: Structure of E.Coli Hfq (1HK9) displayed in surface representation (C) and in a mix of stick and ribbon representation (D).
Figures modeled with Visual Molecular Dynamics (University of Illinois)

Identification of novel ncRNAs is a difficult task, not only because of the limited knowledge about them but also because of their fast turnover. Usually they exist for a short time and in small concentrations. Another difficulty is the restriction to one growth condition. Conventional methods for finding novel ncRNAs like RNomics or microarray applications use cDNA, which is restricted to a specific growth and/or stress condition, so only one expression/stress condition can be observed at a time. A useful method to avoid the aforementioned problems is genomic SELEX (Originally developed by Craig Tuerk and Larry Gold; Turek, 1993).

3.9 Genomic SELEX

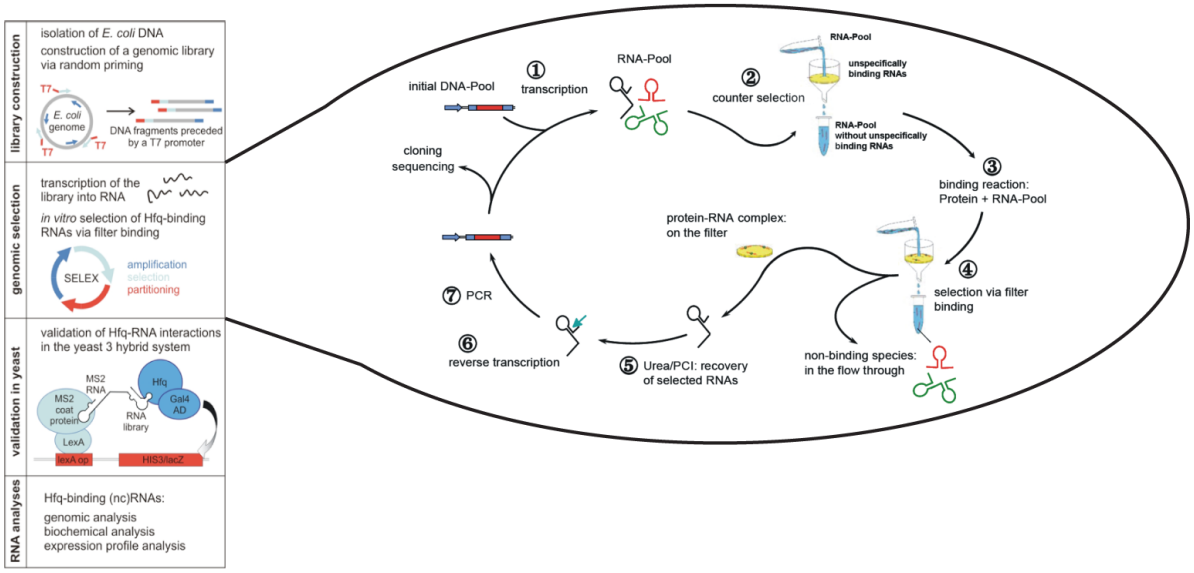


Fig. 6: genomic SELEX (adapted and modified from C. Lorenz)
Schematic overview of genomic SELEX approach, using Hfq to isolate novel RNAs. The four major parts are: library construction via random priming - genomic selection - validation of the RNAs in yeast three hybrid system and RNA analyses using BLASTn, gelshift assays and RT-PCR.
Genomic selection is additionally depicted in a bubble. Its major steps are: transcription of the library into RNA (1); counter selection to remove unspecifically bound RNAs (2); binding of the RNAs to Hfq (3) and selection via filter binding (4); the protein-RNA complex, bound to the filter is recovered (5), reverse transcribed (6) and amplified by PCR (7).

Genomic SELEX (systematic evolution of ligands by exponential enrichment) is an expression state independent approach for selecting novel RNAs. The expression state is independ-

ent, because the starting point for the selection is the genomic DNA and not the total RNA isolated from different growth and stress conditions.

Genomic SELEX is a modified version of in vitro SELEX, which is based on the three major mechanisms of evolution: variation, selection and amplification. A random pool is used for finding RNAs binding with high affinity to a bait protein. In our case the random pool was replaced by a genomic library of *E.Coli*. After transcription of the library, RNAs with high affinities to the protein are selected via filter binding and non binding species are eliminated from the pool after several rounds of SELEX (Fig. 6).

4. AIMS OF THE PROJECT

As the key finding of the Hfq genomic SELEX project was the identification of 85 cis antisense RNAs the aim of my project was the further investigation of one of them: the Hfq binding cis antisense RNA 'as-htrG-cca'. Since the cca gene codes for the tRNA nucleotidyltransferase (see 3.7 tRNA processing and the htrG-cca operon), which repairs the 3' CCA ends of the tRNAs, the antisense lying as-cca sequence might be interesting for its regulation.

The project was sub classified in three parts. First, a rough idea of the length was established by primer walking, a method based on RT-PCR. Here the 3' and 5' ends of the transcripts were narrowed down to a few basepairs (6.1 Strand specific primer walking). In a second experiment the RNA was circularized to map the 3' and 5' ends of the transcript simultaneously (6.2 Simultaneous mapping of 5' and 3' ends). Finally a northern blot with radiolabeled strand-specific probes was performed to investigate the expression of the RNA and to prove the strand specificity as well (6.3 Northern Blot).

5. PRELIMINARY DATA

(C. Lorenz, Dissertation; U. Schöberl, Diploma Thesis)

5.1 Genomic SELEX with Hfq

In a recent work done by Christina Lorenz, the potential ncRNA binding protein Hfq was used to isolate and identify novel RNAs by genomic SELEX. A representative genomic PCR library was constructed via random priming and additionally a T7 promoter was inserted for adjacent transcription. The advantage in this step is that the whole, in our case *E.coli*, genome is represented, including transcriptionally silent and noncoding regions. As bacterial ncRNAs are between 50 and 500 nt, the respective lengths were selected and transcribed into RNA by T7 polymerase. Ideally the RNA library should now contain overlapping sets of fragments covering the whole genome. For the selection of high affine ncRNAs the pool was then exposed to Hfq, bound to a filter. So the non binding species were excluded from the pool. Then the selected RNA pool was eluted, revers transcribed into DNA and amplified via PCR before subjecting to the second round of SELEX. After the 6th round of selection, a competitor tRNA was added to the remaining pool to enhance the stringency and to remove low affinity binding RNAs. Then 4 more rounds were performed, resulting in a pool of high affinity Hfq binding RNAs.

After SELEX, 112 RNAs resulting from cycle 8 (minus competitor) and cycle 9 (plus tRNA competitor) were cloned and sequenced. To validate that in vitro selected sequences bind Hfq in a cellular environment, RNAs resulting from SELEX cycle 8 were introduced into the yeast three hybrid system and Hfq-binding was monitored via expression of the reporter genes HIS3 and lacZ. Then clones, positive in this reporter assay, were chosen for further analysis and additional 58, to the previous 112, were sequenced (170 in total).

5.2 Genomic SELEX outcome

Of the 170 analysed sequences 107 mapped to individual transcripts. Of the 107, 85 were located antisense to protein coding sequences (cis antisense), 9 clones mapped to intergenic regions and 13 to mRNAs of *e.coli* (Fig. 7).

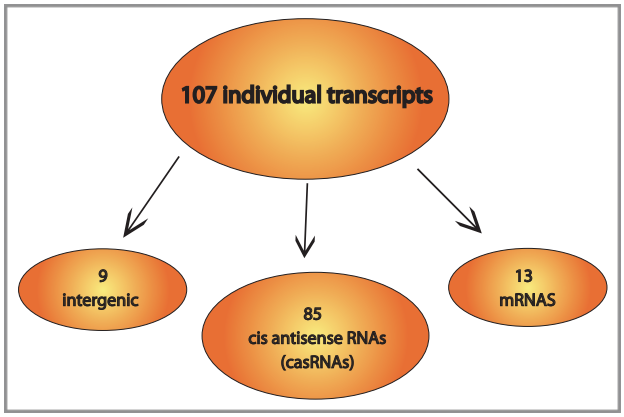


Fig 7: SELEX outcome
107 sequences mapped to individual transcripts. Out of those 107, 85 sequences (79,5%) mapped antisense to mRNAs (cis antisense RNAs); 13 (12%) to mRNAs and 9 (8,5%) to intergenic sequences

To date most of the identified ncRNAs map to intergenic sequences (trans encoded RNAs) and build a category of highly structured short RNAs involved in multi target regulatory processes (Wagner, 2007).
In this work (C. Lorenz, Dissertation), as well as in previous works (Storz, 2005), a remarkable number of cis encoded RNAs was found in the E.Coli genome. Until now, the role of these cis-as RNAs was considered to be of low importance. Through the discovery of more and more cis-as RNAs the general opinion is changing insofar that the role of cis encoded regulation is thought to be of more importance then hitherto assumed.

5.3 Hfq binding motif

A second work connected to the above mentioned was done by Ursula Schauer and Tanja Gesell and concerned the structural approach of the selected RNAs. She investigated the structural properties of four RNAs (as-ptsL, as-cca, as-metH/iclR, ig-proX/ygaX) via

chemical probing and probable structural changes between secondary structure prediction (Hofacker, Vienna RNA package) and DMS pattern. The DMS pattern of as-cca-tail (tail represents an artificial binding site) in the presence of Hfq is shown in Fig. 8.

As a result of DMS footprinting an Hfq binding site was hypothesized to be a double-A pattern, 5'-xnAAxAAxAAxn- 3'. This pattern was found in three out of four RNAs and may have a key function in high affinity Hfq binding.

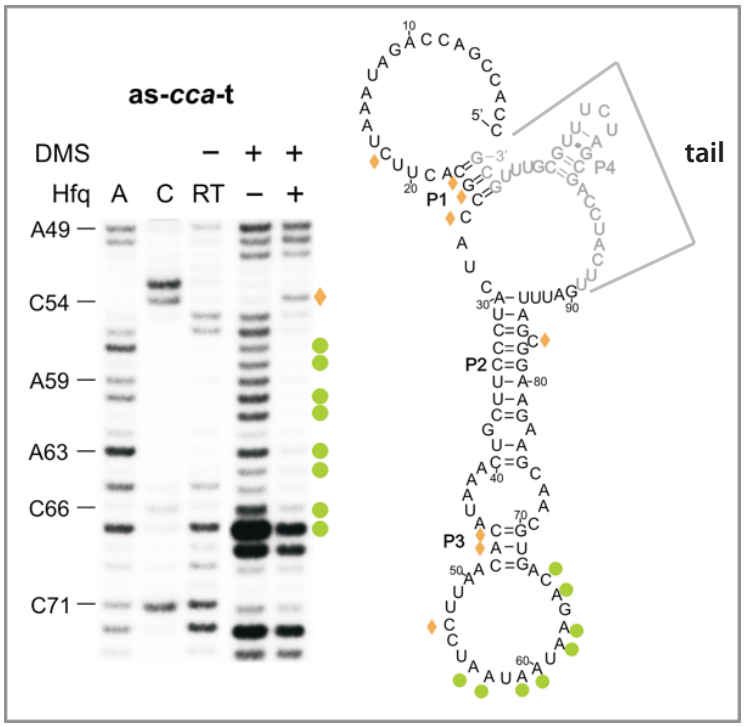


Fig. 8: DMS pattern of as-cca-tail in the presence of Hfq.
left: A representative region of an 8% denaturing polyacrylamid gel to detect DMS modifications. Sequencing lanes A and C and -RT were not treated with DMS. DMS was added to lanes - and + Hfq. Green dots and orange diamonds label protections or enhancements induced by Hfq.
right: The minimum free energy (MFE) plot of as-cca-tail. The Hfq binding site, represented by green dots, reaches from A56 to C66. 'Tail' labels the nucleotides added to the 3' end which served as a primer binding site. (Fig. modified from U. Schauer, dipoma thesis)

5.4 antisense transcript as-cca

In my Diploma thesis I analyzed the cis antisense RNA as-cca, discovered by the SELEX experiment of Christina Lorenz. As the cca gene codes for the tRNA nucleotidytransferase (see 3.7 tRNA processing and the htrG-cca operon), which repairs the 3' CCA ends of the tRNAs, the antisense lying as-cca sequence might be interesting for its regulation. The SELEX

sequence as-cca is 91nt long and lies on the opposite strand (see 3.7 tRNA processing and the htrG-cca operon) between the two genes htrG and cca, both encoded in one operon. Its sequence is shown in 5.5. It is known that as-cca is expressed in the logarithmic phase, stationary phase, in M9 minimal medium as well as under oxidative and acidic stress conditions (Fig. 9a). Therefore the author believes its function lies in stress response.

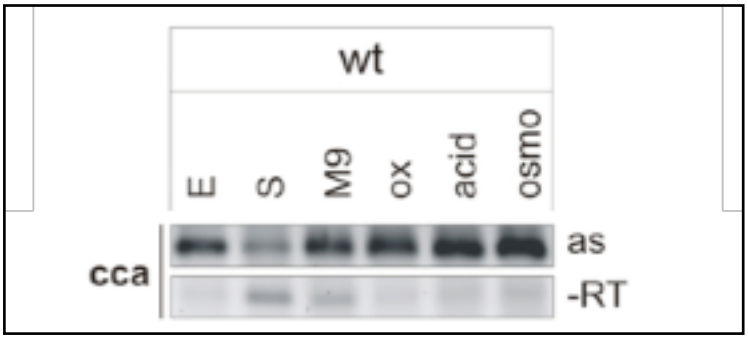


Fig. 9a: as-cca expression pattern
RT-PCR with primers specific for as-cca. as-cca is expressed under logarithmic phase (E), stationary phase (S), in M9 minimal medium (M9) as well as under oxidative (ox) and acidic (acid) stress conditions. The lower lane represente the -RT controls. (modified from C. Lorenz Disseration)

Furthermore as-cca exhibited a very high affinity to Hfq protein, verified by gel mobility shift assays. Here a complete shift at 1 nM Hfq was observed (U. Schauer, Diploma Thesis).

As Hfq tightly binds to the isolated cis antisense RNAs (Lorenz, unpublished data) and RNA duplexes are conceivable targets for RNase III mediated degradation (RNaseIII paper), the effect of Hfq and RNaseIII was tested on as-cca. While expression levels remained unchanged during exponential growth, transcript levels were reduced in hfq+ E.coli cells, compared to cells lacking Hfq (hfq-) in stationary phase, indicating that the presence of Hfq down regulates as-cca abundance (Fig 9, left). In stationary phase both as-cca and mRNA levels were reduced in an rnc+ background when compared to cells lacking RNaseIII (rnc-), indicating RNase III mediated duplex degradation (Fig. 9b, right)

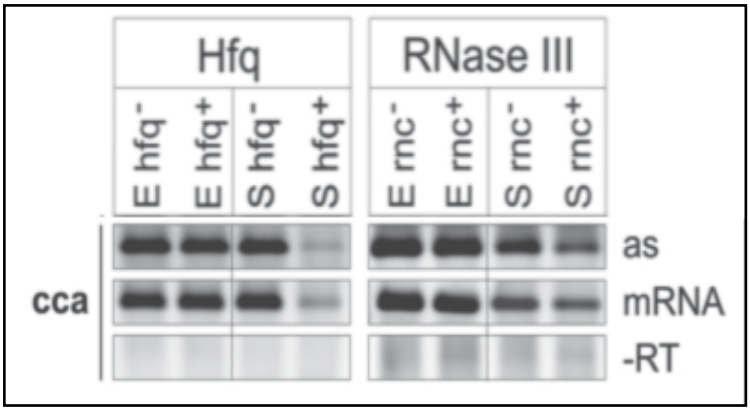


Fig. 9b: effect of Hfq and RNaseIII

RT PCR with primers specific for as-cca and cca mRNA. The expression levels during exponential growth and stationary phase were compared in matters of hfq^+ / hfq^- background as well as rnc^+ / rnc^- background.
left: lanes 1 + 2: expression levels of as-cca and mRNA remain unchanged during exponential (E) growth but are reduced in hfq^+ cells compared to hfq^- cells during stationary (S) phase (lanes 3 + 4).
right: lanes 1 and 2: expression levels of as-cca and mRNA remain unchanged during exponential growth (E). lanes 3 and 4: in stationary phase (S) as-cca and mRNA levels were reduced in rnc^+ cells compared to rnc^- cells. The lowest lane indicate the -RT controls of each reaction. (modified from C. Lorenz unpublished data)

5.5 as-cca Hfq binding sequence

5' CCACCGACCAGAUAAAUCUUCACGCCAUCAUCCCUUCGUCAAAUACACAAU UCCU**AAUAAUAA**GACAGUGCAACGAAGAAGGCGAUUUAGU **3'**

AAUAAUAA ... Hfq binding site (see also Fig. 8)

6. RESULTS

6.1 Strand specific primer walking

The main objective of my project was to determine the length and abundance of the recently isolated antisense RNA as-cca (original name given by C. Lorenz - I renamed the transcript to as-htrG-cca). Some of the cis antisense RNAs, identified by the SELEX approach were bigger than expected (see preliminary data), thus it was assumable that the antisense cca-htrG RNA also fell in the category of long cis antisense RNAs.
A straight forward method to estimate the approximate length of an RNA is primer walking, which I therefore used. Given only a small known region of an unknown RNA one can narrow down its 5' and 3' ends by simply performing RT-PCR reactions. A schematic diagram of the method is shown in Fig. 10. The known region of my antisense transcript was the Hfq binding sequence (Hfq b.s.), identified by genomic SELEX. It is 91nt long and lies on the minus strand between the htrG and cca genes (see preliminary data). One difficulty was that the RNA of interest, as-cca-htrG, lies antisense to a known mRNA so that reverse transcription turned out to be more challenging. To reverse-transcribe only the antisense RNA, I had to use strand specific primers instead of random hexamers.

6.1.1 Primer design

I designed sets of PCR primers inside and outside of the Hfq binding sequence based on the genomic DNA sequence of this region. Reverse primers (R) were complementary to the antisense RNA and the forward primers (F) complementary to the resulting cDNA after reverse transcription. The Hfq b.s. specific primers were fixed so that they overlap the known sequence, because Hfq b.s. was unfortunately not suitable for designing optimal primers. These primers were named R^{Hfq} for the reverse primer and F^{Hfq} for the forward primer. The flanking primers for the 5` and 3` flanking regions of the known Hfq b.s. were named R^{fl}, for reverse flanking and F^{fl}, for forward flanking. They were designed in two steps. First sets of primers in a range of about 200-300 bp steps (set $R^{fl}2$, $R^{fl}3$ and $R^{fl}4$ at the 3` end and set $F^{fl}2$, $F^{fl}3$, $F^{fl}4$

and $F^{fl}5$ at the 5` end) than the sets were further restricted to 20-50bp ($R^{fl}2$-1, $R^{fl}2$-2 and so on) intervals. The primer sets are summarized on the table PRIMER.

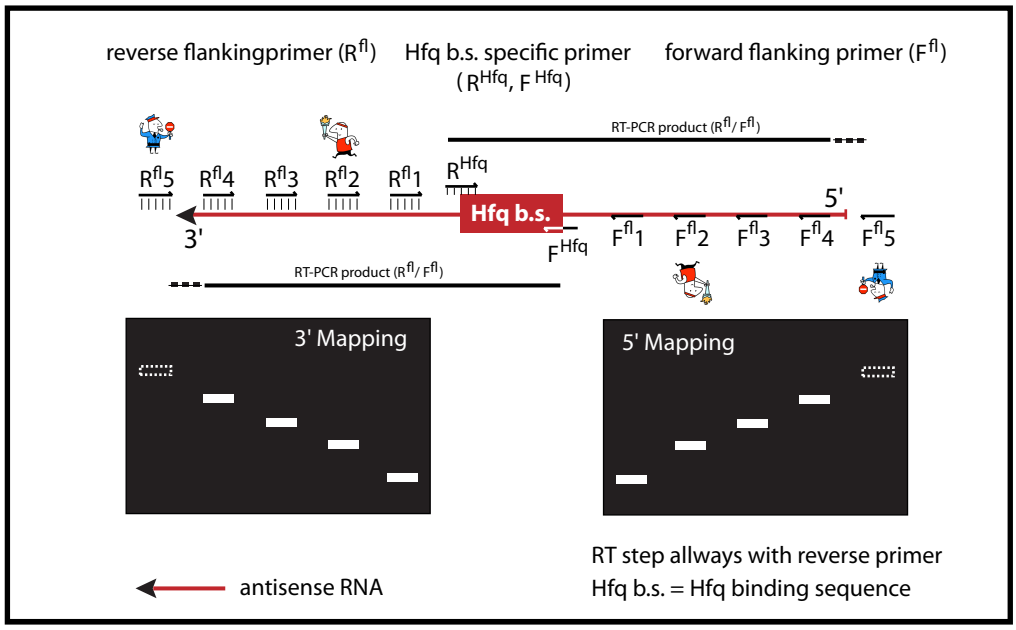
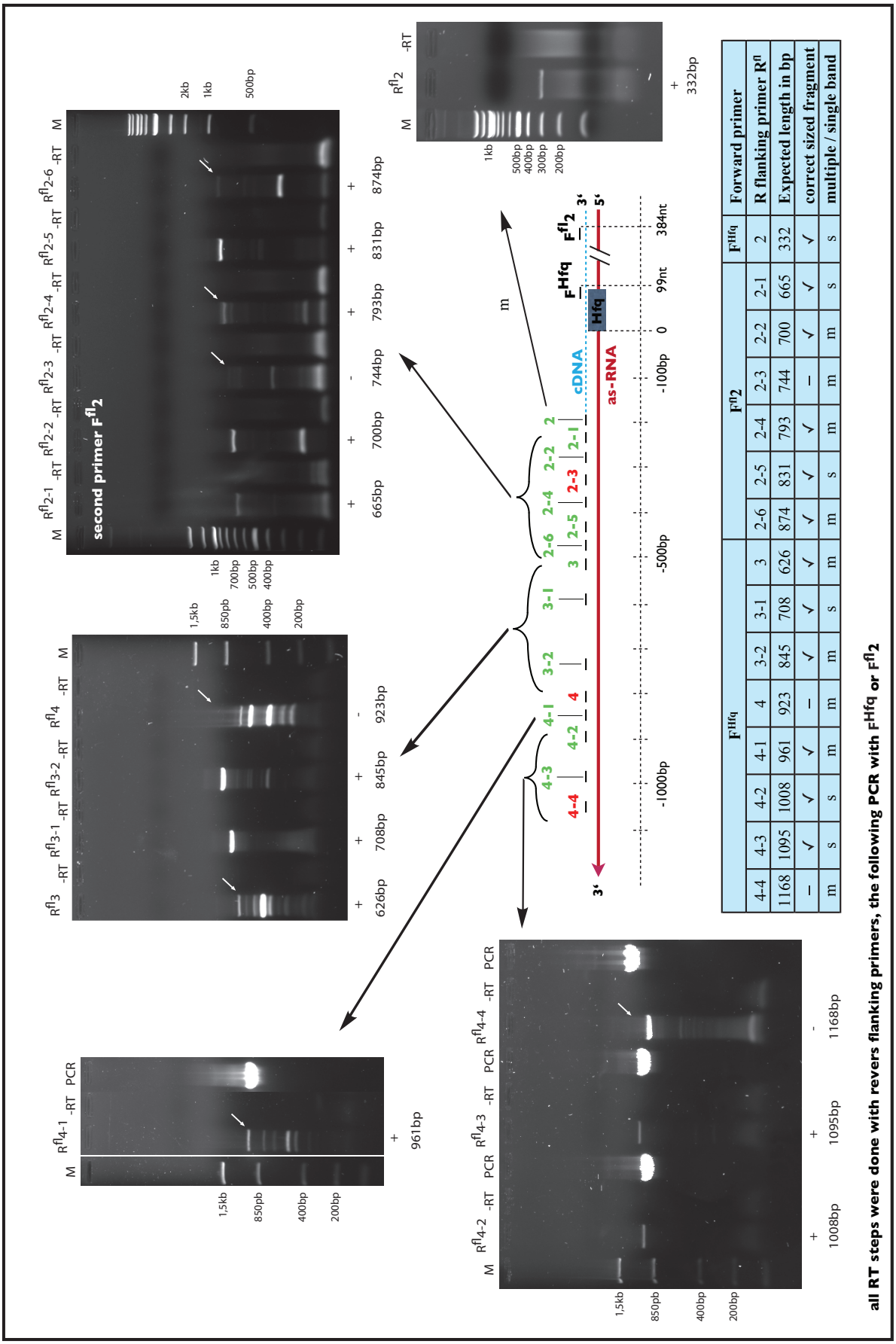


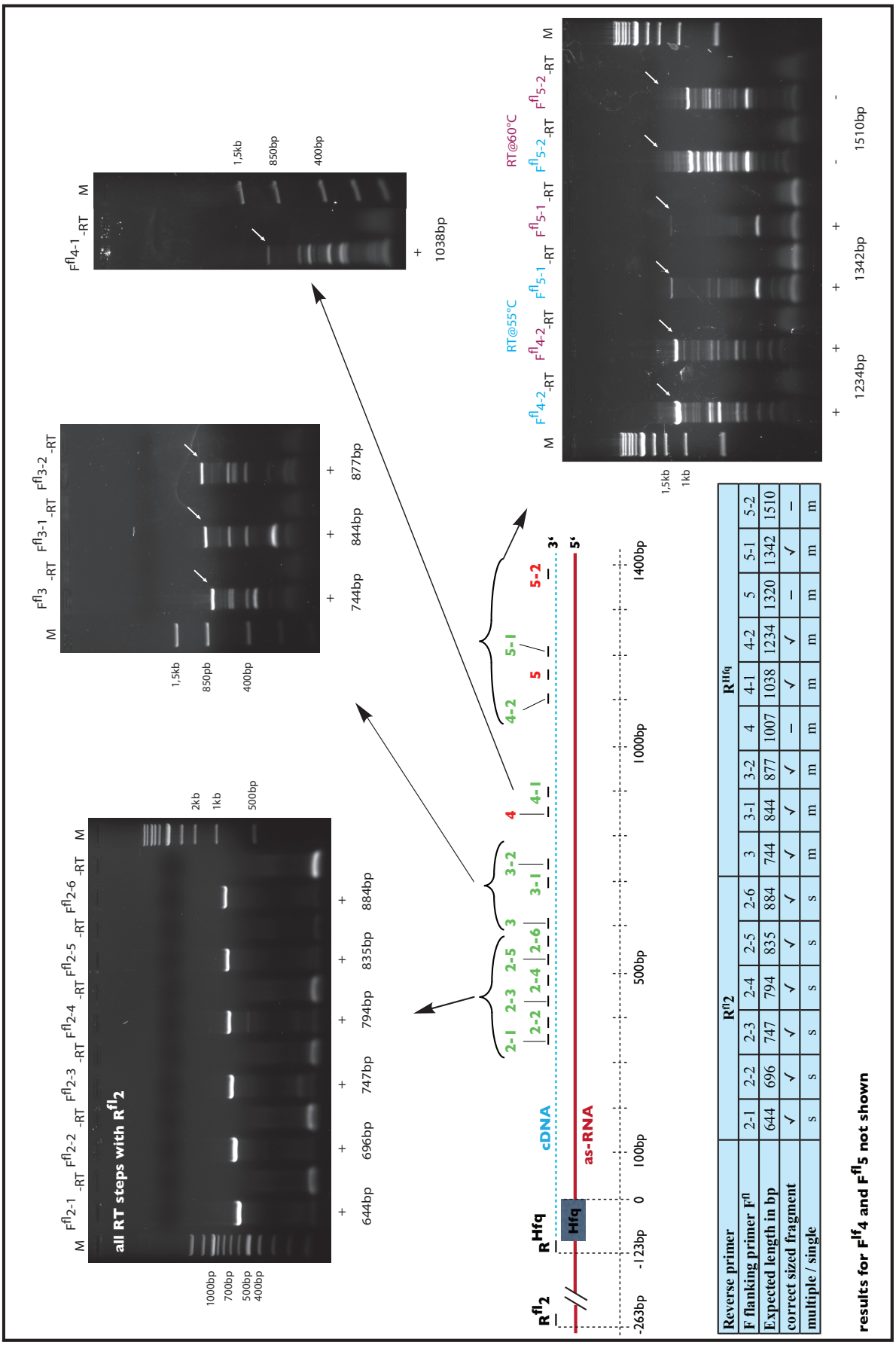
Fig 10: Strand specific primer walking:
The known region of the antisense RNA is symbolized by the red box (Hfq b.s.) whereas the unknown 3' and 5' regions are marked by the red arrow. Two primers were designed, overlapping the known region, I will refer to them as 'Hfq b.s. specific primers' (R^{Hfq}, F^{Hfq}), and several primers outside the known region, named 'flanking primers' (R^{fl}, reverse flanking and F^{fl}, forward flanking). Both reverse primers (R^{fl} + R^{Hfq}) are complementary to the antisense RNA and the forward primers (F^{fl} + F^{Hfq}) complementary to the resulting cDNA after reverse transcription.
The RT step for 5' mapping was done with Hfq b.s. specific primer R^{Hfq} and the following PCR reactions beginning with the 3' flanking primer $F^{fl}1$, continuing stepwise with $F^{fl}2$, $F^{fl}3$, and so on. For 3' mapping, the RT steps were done with series of flanking primers R^{fl} and the PCR amplification with Hfq b.s. specific primer F^{Hfq}. If loading the resulting RT-PCR products on an agarose gel (black rectangle), a similar band pattern should be observable. When the flanking primer has its complementary sequence on the antisense RNA (3' end) or on its respective first strand cDNA (5' end), a DNA band will be amplified. If the flanking primer does not lie on the antisense strand (or first strand cDNA) anymore, then the amplification cannot work and no band is visible. The end of the transcript was then narrowed down to the region between the last 'positive primer' (resulted in a band - white rectangle) and the first 'negative primer' (resulted in no band - white dotted rectangle).

In Fig. 11A and 11B I displayed the respective agarose gels to all used primers. For every RT-PCR reaction I also performed a negative control with the same primer pairs, which is the direct amplification of the total RNA without reverse transcription (labeled with -RT). I loaded the negative control next to its respective RT-PCR lane. Partially (only for bigger fragments), I also performed a positive control with the same primer pairs, which is a standard PCR reaction with genomic DNA.



F Hfq								F fl 2							F Hfq	Forward primer
4-4	4-3	4-2	4-1	4	3-2	3-1	3	2-6	2-5	2-4	2-3	2-2	2-1	2	R flanking primer R fl
1168	1095	1008	961	923	845	708	626	874	831	793	744	700	665	332	Expected length in bp
–	✓	✓	✓	–	✓	✓	✓	✓	✓	✓	–	✓	✓	✓	correct sized fragment
m	s	s	m	m	m	s	m	m	s	m	m	m	s	s	multiple / single band

all RT steps were done with revers flanking primers, the following PCR with F^{Hfq} or $F^{fl}2$



Reverse primer	R^fl2								R^Hfq					
F flanking primer F^fl	2-1	2-2	2-3	2-4	2-5	2-6	3	3-1	3-2	4	4-1	4-2	5	5-1	5-2
Expected length in bp	644	696	747	794	835	884	744	844	877	1007	1038	1234	1320	1342	1510
correct sized fragment	√	√	√	√	√	√	√	√	√	–	√	√	–	√	–
multiple / single	s	s	s	s	s	s	m	m	m	m	m	m	m	m	m

results for F^lf4 and F^lf5 not shown

Fig. 11: 3' and 5' end mapping
The used primers for 3' and 5' mapping are shown in schematic diagrams on both figures A and B. The known region of as-htrG-cca RNA is indicated by red dotted arrows. Their respective agarose gels are presented all around.
In the diagrams all those primers, which resulted in a correct sized band were marked in green and primers which failed to succeed in red. For better understanding, a scale of length was added, showing the relative distance to Hfq b.s (grey box) as well as a table, showing the most important parameters of the experiment.
RT-PCR reactions were done with 3-5µg of total RNA (for more detail see materials and methods) and the RT-PCR lane is marked with the relevant primer's name. The expected band size is indicated for every primer pair, underneath its RT-PCR-lane. White arrows indicate the expected bands. The negative control is labeled with '-RT' and shown to its respective RT-band's right side. A positive control, labeled with 'PCR' was only done for primers R^{fl}4-1 to R^{fl}4-4.
A: 3' end mapping
The agarose gels are shown in a counterclockwise order, starting with primer R^{fl}2 on the right bottom and ending with the gel showing R^{fl}4-2, ..., R^{fl}4-4 on the left bottom. Reverse transcription was done with the flanking reverse primers R^{fl}2, R^{fl}2-1, ..., R^{fl}4-4, complementary to the antisense RNA (asRNA), and for the following PCR steps Hfq b.s. specific primer F^{Hfq} was added as second primer. Sole exception is primer set R^{fl}2, where forward primer F^{fl}2 was added instead of F^{Hfq}.
B: 5' end mapping
Fig 11B displays the resulting agarose gels in a clockwise order starting with primer F^{fl}2-1 on the left and ending with F^{fl}5-1 on the right bottom. Reverse transcription was done with R^{Hfq}, except for primer set F^{fl}2 (F^{fl}2-1, ..., F^{fl}2-6) where the RT was done with reverse primer R^{fl}2. For the following PCR, the forward primers F^{fl}2-1, ..., F^{fl}5-1 were added, all complementary to the resulting cDNA (blue dotted line). In the last picture (right bottom) the incubation temperatures of the RT step are indicated in different colors. For primer pairs F^{fl}4-2, F^{fl}5-1 and F^{fl}5-2, the RT-PCR was performed at 55°C (blue) and at 60°C (violet).
M - DNA length marker, -RT - negative control, PCR - positive control

6.1.2 3' end mapping

For the mapping of the 3' end I used the designed flanking primers (R^{fl}2, ..., R^{fl}4-4) for the reverse-transcription. In the subsequent PCR reaction the respective Hfq b.s. specific primer (F^{Hfq}) was added as second primer. For the R^{fl}2 primer sets (R^{fl}2-1, ..., R^{fl}2-6) forward primer F^{fl}2 was added as second primer, for every other reaction I used R^{Hfq} as Hfq b.s. specific primer. The resulting gel pictures are depicted in Fig 11A.

On the first picture, I show the amplification with the R^{fl}2/F^{Hfq} primer pair. The length of the amplified fragment agreed with the expected 332 basepairs. On the second gel, the amplification with the R^{fl}2 primer set (R^{fl}2-1/F^{fl}2, R^{fl}2-2/F^{fl}2, R^{fl}2-3/F^{fl}2, R^{fl}2-4/F^{fl}2, R^{fl}2-5/F^{fl}2, R^{fl}2-6/F^{fl}2) is shown. Three RT-PCR reactions with the primer pairs R^{fl}2-3/F^{fl}2, R^{fl}2-4/F^{fl}2 and R^{fl}2-6/F^{fl}2 resulted in multiple bands, however R^{fl}2-4/F^{fl}2 and R^{fl}2-6/F^{fl}2 exhibited the correct sized band (white arrows) among the others. R^{fl}2-3/F^{fl}2 did not contain the band with the ex-

pected size (the white arrow displays the expected position of the band). Of the primer pairs shown on the next picture (top middle) $R^{fl}3\text{-}1/F^{Hfq}$ gave rise to a single band, $R^{fl}3/F^{Hfq}$, $R^{fl}3\text{-}2/F^{Hfq}$ exhibited multiple bands, but including the expected band with the correct size and $R^{fl}4/F^{Hfq}$ failed to succeed. The last two gel pictures (left side in Fig 11A) show the primer set $R^{fl}4$ ($R^{fl}4\text{-}1$, ..., $R^{fl}4\text{-}4$). For those reactions, a positive control was performed as well. While the RT-PCR with $R^{fl}4\text{-}2/F^{Hfq}$ and $R^{fl}4\text{-}3/F^{Hfq}$ gave a single band of the correct size, primer pair $R^{fl}4\text{-}1/F^{Hfq}$ resulted in multiple bands, including the band with the expected size and $R^{fl}4\text{-}4/F^{Hfq}$ gave a band of incorrect size (inconsistent with positive control).

Summing up, positive single bands were achieved with reverse primers $R^{fl}2$, $R^{fl}2\text{-}1$, $R^{fl}2\text{-}2$, $R^{fl}2\text{-}5$, $R^{fl}3\text{-}1$, $R^{fl}3\text{-}2$, $R^{fl}4\text{-}2$, and $R^{fl}4\text{-}3$ with their respective forward primers. Besides that I got multiple bands with primers $R^{fl}2\text{-}4$, $R^{fl}2\text{-}6$, $R^{fl}3$ and $R^{fl}4\text{-}1$, all containing the band with the expected size. Primer pairs $R^{fl}2\text{-}3/F^{fl}2$, $R^{fl}4/F^{Hfq}$ and $R^{fl}4\text{-}4$ failed to give a correct band. The last positive primer which worked was $R^{fl}4\text{-}3$, 996nt proximal to the Hfq b.s., whereas its neighboring primer $R^{fl}4\text{-}4$, 1069nt proximal to Hfq b.s., did not result in a band anymore (Fig. 11A).

6.1.3 5' end mapping

For the 5' end mapping I used the Hfq b.s. specific primer R^{Hfq} for the reverse transcription and the flanking primers $F^{fl}2$, ..., $F^{fl}5\text{-}2$ were added to the subsequent PCR reactions as second primers. For the $F^{fl}2$ primer sets ($F^{fl}2\text{-}1$, ..., $F^{fl}2\text{-}6$) reverse primer $R^{fl}2$ was used for reverse transcription. All gel pictures are shown in Fig. 11B.
On the first gel picture (left top) I show the RT-PCR with primer set $F^{fl}2$ ($R^{fl}2/F^{fl}2\text{-}1$, $R^{fl}2/F^{fl}2\text{-}2$, $R^{fl}2/F^{fl}2\text{-}3$, $R^{fl}2/F^{fl}2\text{-}4$, $R^{fl}2/F^{fl}2\text{-}5$, $R^{fl}2/F^{fl}2\text{-}6$). All pairs resulted in single bands of the correct size. On the next two pictures the results of primer pairs $R^{Hfq}/F^{fl}3$, $R^{Hfq}/F^{fl}3\text{-}1$, $R^{Hfq}/F^{fl}3\text{-}2$, $R^{Hfq}/F^{fl}4\text{-}1$ are shown. Unlike the previous ones, these primer pairs gave rise to multiple bands. Again all of them included the correct sized bands, indicated with a white arrow. The last picture (right bottom) displays primer pairs $R^{Hfq}/F^{fl}4\text{-}2$, $R^{Hfq}/F^{fl}5\text{-}1$ and $R^{Hfq}/F^{fl}5\text{-}2$.

Due to the problems with multiple bands, I tried to optimize the reaction by performing the reverse transcription at higher temperatures. Therefore, the RT step was performed at the previously used 55°C (blue marked) and additionally at 60°C (violet marked). Although primer pairs $R^{Hfq}/F^{fl}4$-2, $R^{Hfq}/F^{fl}5$-1 still showed multiple bands at both temperatures, I could reduce the intensity of the multiple bands, so that the correct sized one was strongest. The primer pair $R^{Hfq}/F^{fl}5$-2 also gave rise to multiple bands, but this time, the correct sized band was missing (white arrow indicates the expected size). This means, the last primer resulting in a correct band was $F^{fl}5$-1, 1219nt distal to Hfq b.s.

Recapitulated, positive single bands were scored with forward primers $F^{fl}2$-1, $F^{fl}2$-2, $F^{fl}2$-3, $F^{fl}2$-4, $F^{fl}2$-5, $F^{fl}2$-6 with their respective reverse primers. Again I got unexpected multiple bands for the primers $F^{fl}2$-7, $F^{fl}3$, $F^{fl}3$-1, $F^{fl}3$-2, $F^{fl}4$-1, $F^{fl}4$-2 and $F^{fl}5$-1. The primer pairs $R^{Hfq}/F^{fl}4$, $R^{Hfq}/F^{fl}5$ (results not shown) and $R^{Hfq}/F^{fl}5$-1 completely failed to succeed. Therefore, the last positive primer on the 5' end was $F^{fl}5$-1, 1219bp distal to Hfq b.s., whereas primer $F^{fl}5$-2, 1387bp distal to Hfq b.s., failed to succeed.
For the reliability of the experiment, it has to be mentioned that all negative controls (-RT band) as well as all performed positive controls (PCR with genomic DNA) exhibited the correct result.

According to my results, the 5' end lies between primers $F^{fl}5$-1 and $F^{fl}5$-2 and the 3' end between primers $R^{fl}4$-3 and $R^{fl}4$-4. By walking the transcript to both ends, I suggest that the transcript is at least 2205nt (last positive primers $R^{fl}4$-3 and $F^{fl}5$-1) long but shorter than 2547nt (first negative primers $R^{fl}4$-4 and $F^{fl}5$-2) in compliance with the condition that there is only one antisense transcript in this region.

Fig. 12A displays all used primers during primer walking as well as the minimum and maximum size of the as-cca-htrG RNA, according to primer walking experiment. In the diagram, green painted primers indicate primers, which contained the correct sized band and the red colored primers indicate primers, failed to succeed. For 3' mapping, 12 out of 15 reverse

primers worked on a region of 996bp, distal from Hfq b.s. (primer R^{fl}4-3). For 5' mapping, 13 out of 16 forward primers worked inside a region, 1219bp proximal to Hfq b.s. (primer F^{fl}5-1)

6.2 RT-PCR with two flanking primers

To make sure that the as-cca-htrG RNA is one big continuous transcript, I performed RT-PCR reactions with two flanking primers, one lying distal and one proximal to Hfq b.s. I used following primer pairs: R^{fl}2/F^{fl}3, R^{fl}3/F^{fl}2, R^{fl}2-5/F^{fl}3, R^{fl}2/F^{fl}4, R^{fl}4/F^{fl}2, R^{fl}3/F^{fl}3, whereupon the reverse transcription was always performed with the reverse primers only (see Fig. 12B).

Primer pair R^{fl}2/F^{fl}3 gave rise to a 911bp single band, R^{fl}3/F^{fl}2 resulted in a 945bp sized band and the largest product I could achieve with two flanking primers was a 1159bp single band with R^{fl}2-5/F^{fl}3 (Fig. 12B). RT PCR with the primer pairs R^{fl}3/F^{fl}3, R^{fl}2/F^{fl}4, R^{fl}4/F^{fl}2 did not work (results not shown).
All performed negative controls (described above) were negative.
In Fig. 12B, I summarized the outcome of these experiments.

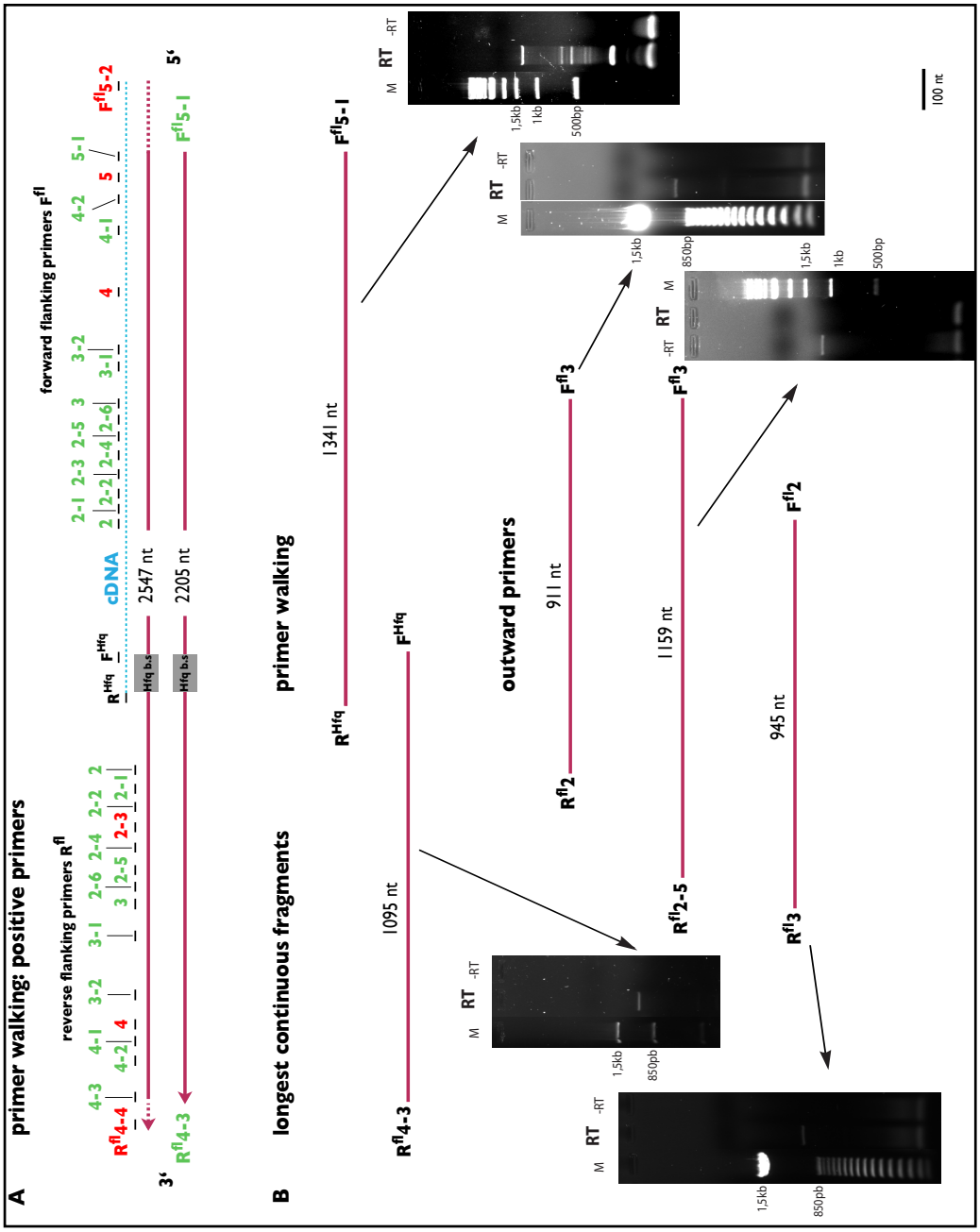


Fig 12: positive primers and longest continuous fragments

A positive primers: shows the the success of all used primers during the primer walking experiment in one graph. Primers are indicated by short black lines true to scale, Hfq b.s is symbolized by the grey box. Primers, which gave a correct sized band are represented in green, primers which failed, in red. The shorter arrow indicates the minimum size of the as-htrG-RNA (2205bp - $R^{fl}4\text{-}3/F^{fl}5\text{-}1$) and the longer arrow indicates the maximum size of the transcript (2547bp - $R^{fl}4\text{-}4/F^{fl}5\text{-}2$).

B The longest continuous fragments: achieved with RT-PCR are indicated by red linear slopes. Labels of respective reverse primers are shown on their left side, those of the forward primers on the right. Resulting agarose gels are shown as well, black arrows are pointing towards them. A DNA length marker (M), the RT-PCR product (middle lane) and the -RT negative control (-RT) are shown. The first two RT-PCR products were the result of the primer walking experiment, the 1341bp long fragment came from 5' mapping (primer pair $R^{Hfq}/F^{fl}5\text{-}1$) and the 1095bp long fragment from the 3' mapping (primer pair $R^{fl}4\text{-}3/F^{Hfq}$). The last three lines represent the resulting bands using two flanking primers for RT-PCR.

6.3 Simultaneous mapping of 5' and 3' ends (Circularization)

After I had a rough idea about the length of as-htrG-cca RNA and additionally a suspicion that there could be more than one antisense transcript on that locus, I wanted to map the exact 5' and 3' ends of the/these transcript/s. This would be important for further analysis of the region, like promoter and terminator sequence searches as well as a crucial step toward mechanism discovery.

An elegant method for the simultaneous mapping of the 5’ and 3’ end of a transcript is RNA circularization, followed by RT-PCR. This method was partially developed in the Schroeder lab (Gonzales et al., unpublished). The intention of this experiment is to circularize every single RNA molecule by ligating the entire total-RNA extract (transcriptome) with T4 Ligase and perform an RT-PCR with inverted primers. Inverted means, turning the primer inside out, that is, the reverse complement of the primer (e.g. 5’-ACGTA-3’ -> 5’-TACGT-3’). This means that the primers are designed in reverse complement direction, so that the reverse transcription goes into the opposite direction (outside form the known region). Both ligated ends should then lie in between the used primers and should be revealed after sequencing the PCR product (see also Fig. S1 in Appendix).

For a better understanding of the procedure, I would like to mention that the total RNA was treated with Tobacco Acidic Phosphatase (TAP) prior the ligation reaction. This was necessary for the removal of the 5' Tri-Phosphate cap, which is present on all unmodified RNA Polymerase transcripts. This step is mandatory for the T4 ligase reaction, which is performed in 4 different dilutions (for detailed description see materials and methods). The control reactions for the RT-PCR as well as for the correct ligation were following: RT-PCR of a ligated, non TAP treated RNA as a control for TAP-treatment (c), RT-PCR of non ligated, TAP treated RNA as a ligation control (neg) and as a positive control an RT-PCR of total RNA with normal, non-inverted primers (pos). The obligatory -RT control was the same as for primer walking.

First I tested the method on the highly abundant E.coli 6S RNA, which is a 183nt long non-coding RNA directly binding to RNA polymerase holoenzyme to inhibit transcription (Wassarman, 2007). I designed inverted primers to the 6S RNA sequence so that the resulting 154nt long fragment would be clearly visible on the agarose gel (primer pairs see table PRIMER). As shown in Fig. 13, the agarose gel exhibited the expected 154nt long band (black arrow) for the RT-PCR (lane 2-5) and the -RT control (lane 8) was negative, underlining the RNA origin of the RT-PCR bands. Also the ligation controls (lanes 6+7) were negative, implicating a correct result.

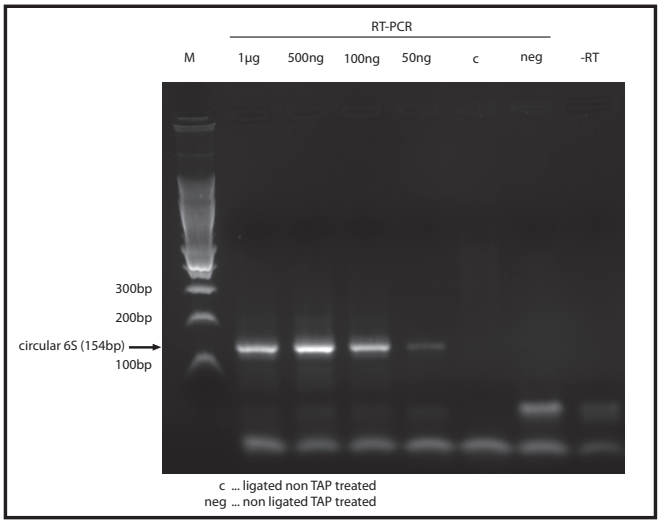


Fig 13: circularization of 6S RNA

The ligation of the total RNA was done in 4 different concentrations and the following RT PCR was performed with primer pairs 6S-R-inv and 6S-F-inv, at which 6S-R-inv was used for reverse transcription. 10µl of the reaction was loaded on a 2% agarose gel including a DNA length marker (M - lane1).

The lanes 2-5 show the 154bp RT-PCR fragment (black arrow) of the ligates; 1µg, 500ng, 100ng and 50ng indicate the ligation concentration of the total RNA.

Lane 6: RT-PCR of a ligated, non TAP treated RNA (c), Lane 7: non ligated TAP treated RNA (neg), Lane 8: -RT control.

After successful circularization of the 6S RNA I adapted this method for the mapping of ascca-htrG RNA. I inverted the primers $R^{fl}2$ and $F^{fl}2$ (see table PRIMER) so that the relevant Hfq b.s was positioned between those primers and all RNAs in this region would be attained for amplifying.

Several attempts and variations failed to get a result. Neither performing the RT step on higher temperatures for more primer specificity, nor using different primer pairs brought any reasonable outcome. The agarose gels showed either no bands or several controls were false (results not shown). After hopeless attempts however, I accidentally used both primers, $R^{fl}3$-

inv and $F^{fl}3$-inv simultaneously in the reverse transcription step after ligation and achieved nevertheless an interesting result:

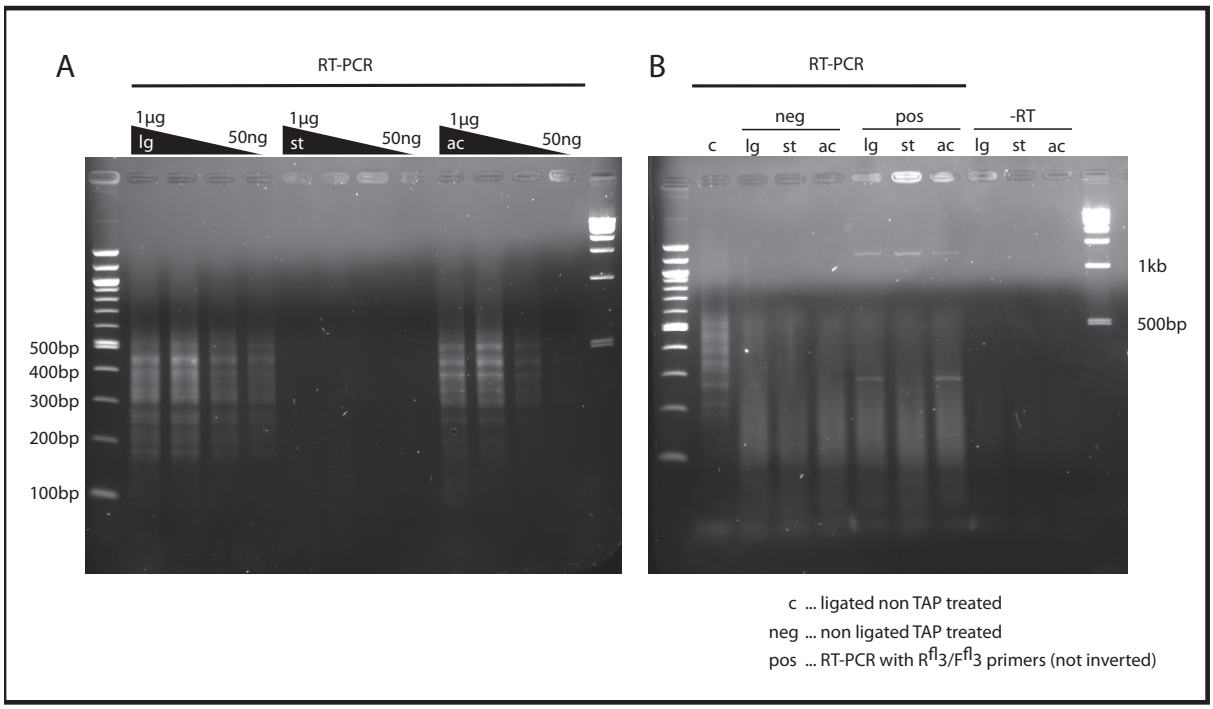


Fig 14: dual mapping of as-cca-htrg RNA with inverted primers

The ligation of the total RNA was done in 4 different concentrations (1µg, 500ng, 100ng and 50ng indicated by black triangles in Fig 14A) and the following RT PCR was performed with primers $R^{fl}3$-inv and $F^{fl}3$-inv, whereas the reverse transcription was done in the presence of both, forward and revere primers! 10µl of the reaction was loaded on a 1% agarose gel including a DNA length marker (M).

A RT-PCR reactions of the ligates: Lanes 2-5 show the RT-PCR of the ligates of total RNA extracted from logarithmic phase (lg), lanes 6-9 the ligates of total RNA extracted from stationary phase (st) and the lanes 10-13 the ligates of total RNA extracted after acidic stress treatment (ac). The resulting bands of logarithmic phase and acidic stress ligates showed multiple bands, all in the range between 100 and 600 bp, the RT-PCR from stationary phase completely failed to succeed.

B control reactions: As ligation and RT-PCR controls, following reactions were loaded on a second gel: lane 2: control for TAP-treatment (c); lanes 3-5: ligation control (neg) for 3 different conditions (lg, st, ac); lanes 6-8: positive control (pos) for 3 different conditions (lg, st, ac), performed with primers $R^{fl}3/F^{fl}3$; lanes 9-11: -RT control for 3 different conditions (lg, st, ac). The neg and the -RT controls were negative, the positive control gave the expected band at 1239bp and the c (ligated non TAP treated) control gave rise to an unidentifiable smear.

The agarose gel showed the RT-PCR outcome of the circularized RNAs (Fig. 14A), isolated from 3 different growth or stress conditions (lg: logarithmic phase, st: stationary phase, ac: acidic stress) and ligated in various concentrations (1µg, 500ng, 100ng, 50ng). Log phase and acidic stress ligates exhibited multiple bands after RT-PCR (lanes 2-5 for lg and 10-13 for ac in Fig. 14A) whereas stationary phase (lanes 6-9 in Fig. 14A) completely failed to succeed. The aforesaid multiple bands covered a range of 50 - 600nt and their intensity attenuated with decreasing ligation concentrations. As a sign of correctness, both RT controls (Fig. 14B: lanes 6-11) as well as the ligation controls (Fig. 14B: lanes 3-5) were correct.

For further investigation, I pooled the RT-PCR samples of the four ligation reactions of each stress condition and cut out the fragments between 50nt and 500nt. After cloning into a Topo-TA vector I picked several colonies and isolated the plasmid DNA (Miniprep). A second PCR control with $R^{fl}3$-inv/$F^{fl}3$-inv primers was done and 13 clones were sequenced (results not shown). The sequence files are depicted in Appendix 11.2.

After aligning the resultant sequence files with the E.Coli K12 genome via BLASTn search, all 13 sequences showed homology to the cca-htrG locus. After further investigation, I identified two different proximal ends varying in the 13 sequences. 4 out of 13 exactly mapped to the Htrgp1 promoter (capital A in Fig. 15), which could be the 3' end of htrG-cca mRNA, but the remaining 9, mapped to no hitherto identified transcription start or end noted in this region. The newly identified end lied 138bp upstream of htrg-cca promoter Htrgp2, 77bp upstream of htrG promoter Htrgp1 and is 643bp downstream of Htrg b.s. (capital B in Fig. 15). The distal ends mapped to several different loci in that region (capital C and D in Fig. 15). Those 4 transcripts, which had their proximal end similar to Htrgp2, ranged between 1577nt and 1716nt in total length (capital C in Fig. 15). The other 9 transcripts ranged between 1359nt and 1564nt in total length (capital D in Fig. 15). Nevertheless, none of the 13 different distal ends mapped to a known transcription start or end but were all in the range of 219 bp (Δ219 in Fig. 15).

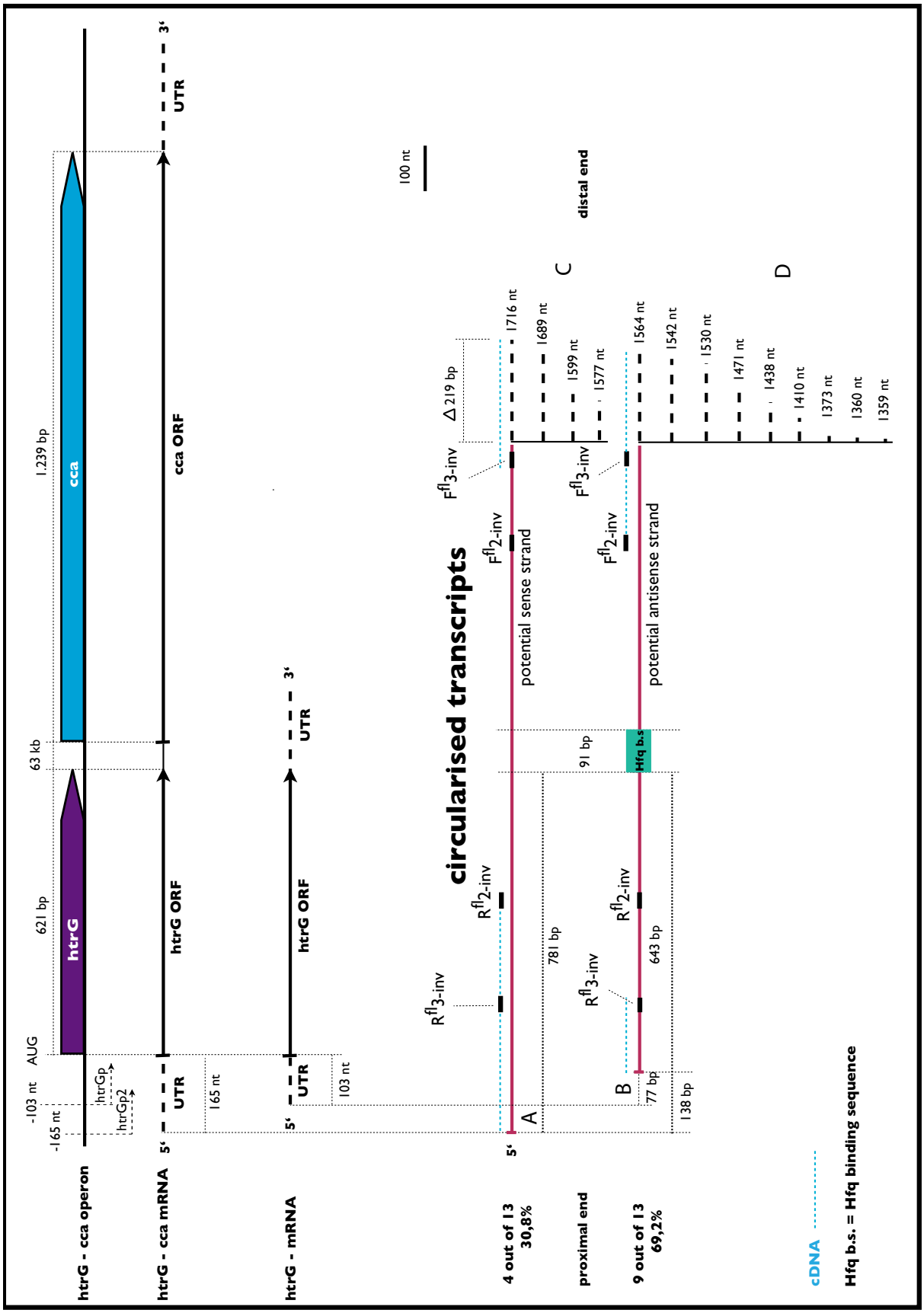

htrG - cca operon
htrG - cca mRNA
htrG - mRNA
-103 nt
AUG
-165 nt
621 bp
63 kb
1.239 bp
htrG
cca
htrGp
htrGp2
5'
UTR
165 nt
htrG ORF
cca ORF
3'
UTR
103 nt
100 nt
circularised transcripts
$R^{fl}3$-inv
$R^{fl}2$-inv
781 bp
643 bp
91 bp
Hfq b.s
potential sense strand
potential antisense strand
$F^{fl}2$-inv
$F^{fl}3$-inv
Δ 219 bp
A
B
C
D
4 out of 13
30,8%
proximal end
9 out of 13
69,2%
distal end
77 bp
138 bp
1716 nt
1689 nt
1599 nt
1577 nt
1564 nt
1542 nt
1530 nt
1471 nt
1438 nt
1410 nt
1373 nt
1360 nt
1359 nt
cDNA
Hfq b.s. = Hfq binding sequence

Fig 15: outcome of the circularization experiment

The two graphs explain the mapping of the circularized transcripts in relation to the genomic background of the htrG-cca locus.

The top graph shows the htrG-cca locus with its two promoters and is identical with Fig 5.

The bottom graph displays the circularized fragments achieved by dual mapping with primers $F^{fl}3$-inv and $R^{fl}3$-inv. Their total lengths are indicated on their right. The reverse transcription was accidentally performed with both primers in one reaction, 4 out of 13 sequences (30,8%) mapped with their proximal ends to the 5' end of the htrG-cca mRNA (A) which means those four most probably originate from the htrG-cca mRNA. 9 out of 13 (69,2%) mapped with their proximal end (B) to a region that did not mapped to a thitherto known gene start or end. Its position relative to the known promoters are indicated by length markers. Those 9 sequences could in fact originate from the antisense strand.

The distal ends of all 13 sequences mapped to different regions, but were all in a range of Δ219 bp (C+D). They are indicated by dotted lines on the proximal end.

6.4 Northern Blots

6.4.1 Northern blots with radiolabeled RNA probes

Although the two previously done experiments for mapping the RNA showed different results, both indicated the existence of more than one transcript on the antisense strand of the htrG-cca locus. If this was true, then the following experiments should confirm this theory. To determine the existence, length and stand specificity of an RNA molecule, Northern Blot with strand specificity is the most effective technique to do so. In my case, I wanted to investigate these three parameters for the as-cca-htrG RNA/RNAs on one side and its sense htrG-cca mRNAs on the other.

By loading the total RNA from different growth or stress conditions to a denaturing formaldehyde agarose gel, the RNA is separated by size. After blotting the gel to a nitrocellulose membrane, the latter is then hybridized with a labeled probe, complementary to the antisense transcript of interest. If the hybridization was successful, a signal in form of a band should then be visible on the resulting radioblot, revealing the transcript's existence and size. The signal's intensity, compared to another band, which can be the sense strand for example, hints the transcript's relative abundance.

I followed the northern blot protocol using between 10µg and 30µg total RNA per lane for the denaturing formaldehyde gel and blotted the separated RNA to a positively charged nylon membrane (further details see material and methods). I hot-transcribed the probe in the presence of two α-^{32}P labeled NTPs (α-^{32}P-UTP and α-^{32}P-GTP), 7,5pmole (45µCi) each, so that theoretically one radiolabeled nucleotide was built in per 10 cold nucleotides (body labeling). Unfortunately, the specific activity of the probe was not measurable due to a defect of the scintillation counter at that time. The antisense htrG-cca probe was complementary to an 81nt long region, inside the Hfq b.s. (for the corresponding primers see table PRIMER).

After hybridization and excessive development of the radioblot, I could not detect specific signals, but the two big ribosomal RNAs (18S and 23S) after exaggerated adjustment of the greyscale (results not shown). Further experiments with enhanced amounts of RNA per lane and using end labeled DNA oligo probes with higher specific activities (equimolar concentrations of DNA probe and ɣ-32P-GTP) did not bring the expected result either. In some cases only a big smear was visible along the lanes, but in almost all cases the ribosomal RNA bands gave the solemn signal (results not shown).

6.4.2 Northern blots with radiolabeled LNA probes

LNA™s (locked nucleic acids) contain an artificially added methylene bridge between the 2' and 4' carbon atoms of their backbone. The molecule is locked in its 3'-endo structural conformation which enhances its binding specificity and thermal stability. Usually, the LNA™ probes are designed so that every 3rd nucleotide is LNA™, the rest is DNA.

As the previously done northern blots with radiolabeled DNA or RNA probes failed to succeed, I tested a different species of nucleic acids for the probe's configuration. It is known that LNA™ probes have a 10 times higher detection rate and a significantly higher specificity to RNA than DNA oligo probes (www.exiqon.com). Another advantage of using LNA™ probes is that it allows hybridizing the membrane at higher temperatures to reduce unspecific binding and the background noise on the membrane (Petersen, 2003).

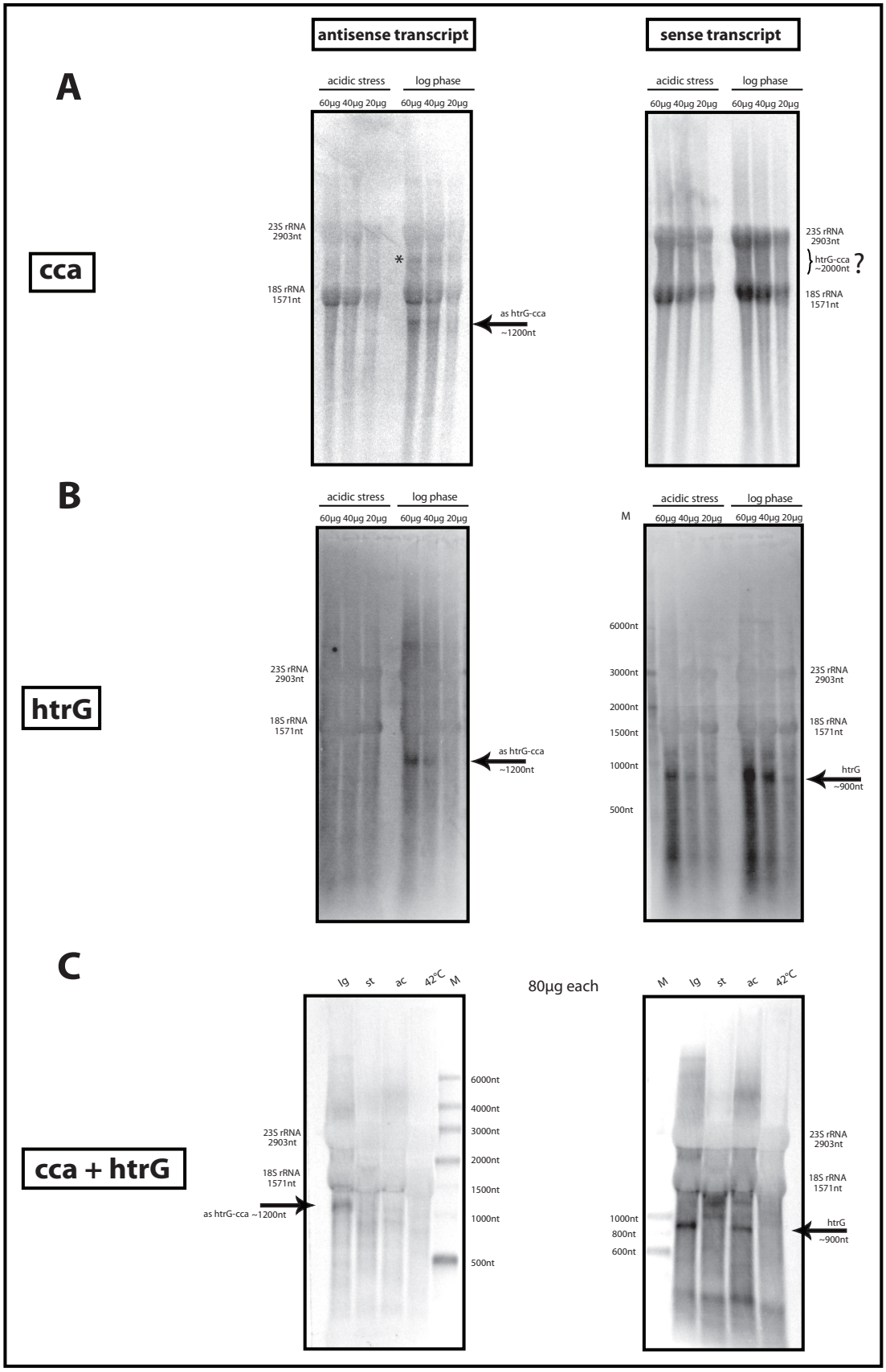
Referring to the genetic background of the htrG-cca locus and the previously done strand specific primer walking experiments, I designed two LNA™ probe complements (sense and antisense) on two different locations (see table PRIMER).
The sense and antisense cca probes (cca and as-cca) would hybridize to their respective strand on the cca locus and the htrG probes (htrG and as-htrG) to the htrG locus. The northern blots with cca and htrG probes should then show the bands, commensurate with the cca mRNA (~2000nt) and the htrG mRNA (~900nt) (see also Fig. 17). The bands for as-htrG and as-cca probes were expected to be multiple according to the primer walking and circularization experiments.

For the denaturing agarose gel I loaded 20µg, 40µg and 60µg total RNA per lane, isolated from E.Coli B strain grown at logarithmic phase and from cells exposed to acidic stress. The quality of the RNA seemed to be good, as the ribosomal RNA bands were intact on the ethidium-bromide stained gel. The efficiency of the overnight capillary transfer was over 90% for all blots, only the 23S rRNA band was slightly visible on the remaining gel. The LNA™ probes were end-labeled with a molar excess of γ-^{32}P-ATP to raise its detection threshold (8,33 pmole γ-^{32}P-ATP, that is 50µCi, for 7,5pmole LNA probe) and hybridized to the membrane at increased temperature (70°C - 75°C - see materials and methods).

The resulting northern blots are pictured in Fig. 16. The blots for the antisense transcripts are displayed on the left and the blots hybridized with the sense probes on the right. The denaturing agarose gel was run and blotted in one piece. After blotting, the membrane was cut in two halves and one was hybridized with the sense probe and the other one with the antisense probe. This meant that sense and antisense radioblots in Fig. 16 originated from the same agarose gel.
After intense washing of the membrane, I could detect one antisense transcript with the as-cca probe at ~1.200nt (black arrow in Fig. 16A left side) in the logarithmic phase (lanes 5 -7) but not in cells grown under acidic stress (lanes 1-3). The strongest signal came from the 5th lane (60µg RNA) and diminished in lanes 6 and 7. Besides that a very faint band was also visible at ~2000nt in lanes 5 and 6 (black asterisk).

The cca mRNA, which was supposed to be ~2000nt long, was not detectable at any condition or growth state (Fig. 16A right). Only the ribosomal RNA bands gave a signal at 1571nt (18S rRNA) and 2903nt (23S rRNA) but not in between (bold question mark). Just like the as-cca probe, as-htrG also detected a ~1200nt long transcript in the logarithmic phase (black arrow) but not under acidic stress (Fig. 16B left). Again the strongest signal was in lane 5 (60µg) and faded in lanes 6 (40µg) and 7 (20µg). As pictured in Fig. 16B (right side) the sense htrG probe detected one ~900nt long sense transcript (black arrow), expressed in the logarithmic phase and under acidic stress condition. The signal was stronger for the log phase (lanes 5-7) compared to acidic stress (lanes 1-3). As the sense htrG transcript was expected to be ~900nt the positive signal most probably belongs to the htrG mRNA. Therefore I also determined the length of its hitherto unknown 3' UTR, as the 5' UTR and the ORF are 724nt in total. It is ~170nt long and its 3'-end overlaps with the 5' start of the cca RNA in circa 50nt (see Fig. 17).

Fig 16: northern blot with radiolabeled LNA probes:
Northern blots specific for the antisense transcripts of the htrg-cca locus are shown on the left side and northern blots specific for the sense transcript of the htrg-cca locus are depicted on the right side. The agarose gels of the radioblots from A, B and C, were run and blotted in one piece before the membrane was cut into two halves.
The figures A and B show the result of the northern blot experiments with single antisense LNA probes (as-cca in A and as-htrG in B) on the left and the sense LNA probes (cca in A and htrG in B) on the right. For the agarose gel, total RNA isolated form E.coli under two different conditions (log phase and acidic stress) was loaded in three different concentrations (20µg, 40µg, 60µg).
A cca: The as-cca LNA probe gave rise to a ~1200nt long transcript, detectable only in the logarithmic phase. There is also a faint band around 2000nt visible (asterisk).The sense cca LNA probe did neither detect the expected 2000nt long htrG-cca transcript (question mark) nor any other transcripts. Only the ribosomal RNA bands were visible.
B htrG: The as-htrG LNA detected a ~1200nt long transcript, expressed only in the logarithmic phase. The sense cca LNA probe detected the ~900nt long htrG mRNA in both conditions. The signal was strongest in the lanes for 60µg total RNA but was still visible in the lanes for 20µg total RNA.
C cca + htrG: Northern blots, hybridized with both LNA probes simultaneously (as-cca + as-htrG on the left; cca + hrtG on the right). For the agarose gel 80µg total RNA, isolated form E.coli under four different conditions (lg - logarithmic phase, st - stationary phase, ac - acidic stress, 42°C - temperature stress) was loaded. Both antisense LNA probes detected the same antisense transcript as they detected in their separate experiments (A and B): a ~1200nt long RNA, expressed only in the logarithmic phase (left). The same was true for the sense probes: The ~900nt long transcript, expressed in logarithmic phase and under acidic stress, is the htrG mRNA. The ~2000nt sense htrG-cca transcript could not be detected again (right).


antisense transcript
sense transcript
A
cca
acidic stress
log phase
60µg 40µg 20µg
60µg 40µg 20µg
23S rRNA
2903nt
18S rRNA
1571nt
*
as htrG-cca
~1200nt
htrG-cca
~2000nt
?
B
htrG
M
6000nt
3000nt
2000nt
1500nt
1000nt
500nt
htrG
~900nt
C
cca + htrG
lg
st
ac
42°C
M
80µg each
4000nt
800nt
600nt

6.4.3 Simultaneous hybridization with two LNA probes

As the previous northern blots gave the interesting result that both antisense probes detected a transcript with 1200nt in length similar expression patterns, I wanted to check whether both probes hybridized to the same transcript. If this was the case, as-cca and as-htrG probes would give the same signal when hybridized together in one blot. I therefore simultaneously hybridized a new membrane with both as-probes. This time I loaded 80µg total RNA per lane, isolated from E.coli B at logarithmic phase (lg), stationary phase (st) as well as from cells grown under acidic stress (ac) and temperature stress (42°C).

The antisense transcript was detected with both probes in one hybridization experiment only in the log phase (Fig. 16C left, lane 1). It gave an explicit signal at ~1200nt. The other growth and stress conditions (lanes 2-4) resulted in no defined signal, only background noise was visible.
To complete the experiment, I also dual hybridized a new membrane with both sense probes. I used the same amount of total RNA as mentioned above and also the same expression/stress conditions. Again the former experimental data was confirmed as shown in Fig. 16C (right). The resulting blot gave the htrG signal at ~900nt in the logarithmic phase (lane 2) and acidic stress (lane 4). Stationary phase and temperature stress at 42°C had no band in that size. The cca signal was completely missing in all four conditions.

In summary I could detect one antisense cca-htrG transcript. It is around 1200nt long and expressed in cells grown until logarithmic phase. Its expression could not be proven for stationary phase, acidic stress and for temperature stress. Due to the fact that both as-htrG and as-cca hybridized to the same transcript and according to the knowledge of the distribution of my antisense probes, the location of the antisense transcript could be bordered to 837nt (continuous red line in Fig. 17). Only the distribution of the remaining 363nt from total 1200nt is still unknown. However these results would definitely go along with the results of the primer walking experiments. The longest continuous fragment achieved with two flanking primers

($R^{fl}2$-5/$F^{fl}3$) was 1159bp long and both antisense LNA probes lie inside the fragment. A summary of the results is shown in Fig. 18

I also detected the htrG mRNA as a 900nt long transcript with a ~170nt long 3' UTR (Fig. 17) It was expressed in the logarithmic phase and under acidic stress conditions. Expression in stationary phase and under temperature stress could not be proven.

The cca mRNA, which was supposed to be transcribed together with htrG and should have resulted in a ~2000nt long transcript but could not be detected under any condition I used.

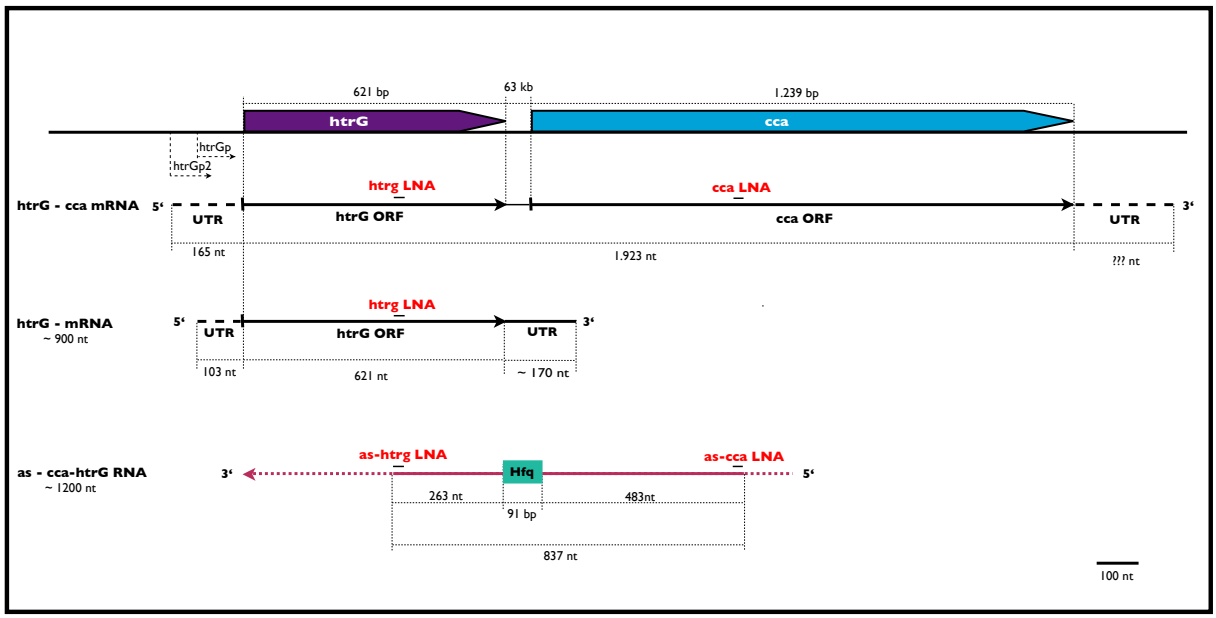


Fig. 17: outcome of the northern blot experiment

The graph schematically shows the results of the northern blots with radiolabeled LNA probes.

The probes (short black lines) were designed to cover a region of 837bp. The sense htrG LNA probes should detect the short htrG mRNA as well as the long hrtG-cca transcript, whereas the sense cca LNA probe should only detect the long hrtG-cca transcript. On the other hand the complementary antisense LNA probes, as-htrG and as-cca, were designed to hybridize to the same region as the sense probes, but on the antisense RNA, represented in red (as-cca-htrG RNA).

The htrG mRNA was ~900nt long, with a ~170nt long thitherto unknown 3'UTR (see Fig 16 B left). The cca LNA did not gave any results (see Fig 16 A left)

The antisense LNA probes hybridized to the same mRNA and gave rise to a ~1200nt long antisense transcript. This result could be achieved in an experiment where both antisense LNA probes were used separately (Fig 16 A+B right) and also in an experiment where the two antisense LNA probes were hybridized together in one step (Fig 16 C right). This fact proves that the existence of the 837nt (continuous red line) between the two probes can be taken for granted. The 91 nt long Hfq binding sequence lies in the middle of it and the distribution of the remaining ~370nt is still unknown (red dotted line).

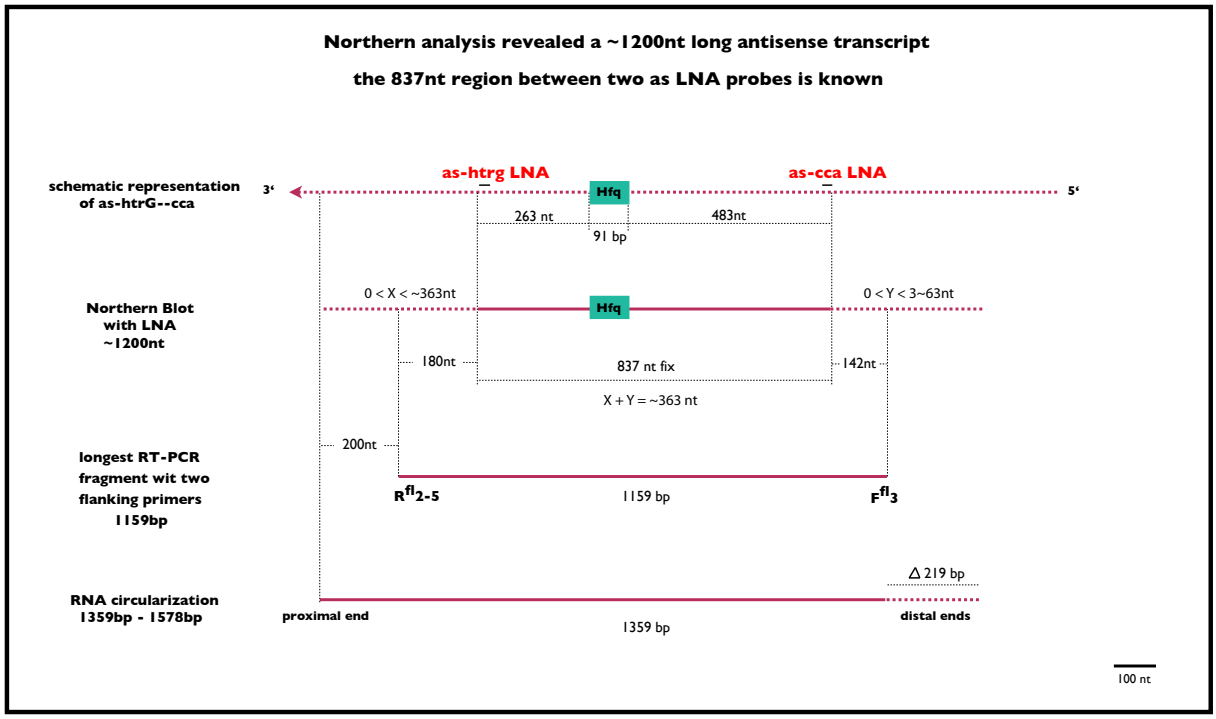


Fig. 18: summary of the results

The graph summarizes the results the primer walking (longest continuous fragment), RNA circularization (proximal and distal ends) and the Northern Blot for the antisense transcript.
The northern analysis revealed a ~1200nt long antisense transcript. The 837nt long region between the LNA probes is known. The distribution of the remaining ~363nt is unknown (X + Y = 363nt).
The longest continuous fragment achieved by RT-PCR with two flanking primers was 1159bp long ($R^{fl}2\text{-}5/F^{fl}3$). This would go along with the northern blot results, since the two antisense LNA probes lie inside the fragment.
The results of the circularization experiment are shown in the bottom. The potential 3' end lies 643nt proximal to Hfq binding sequence. The multiple 5' ends are indicated by "Δ219".

7. DISCUSSION

7.1 htrG-cca locus and the CCA adding enzyme

There is a lot more transcription going on than originally thought. In yeast, 85% of the genome is transcribed in rich media, with many non-annotated transcripts overlapping known genes in antisense direction (David et al., 2006). With the detection of more and more new classes of ncRNAs (Storz et al., 2005; Brantl, 2007; Vogel and Wagner, 2007), it became clear that the expressed fraction of the genome, independent of the organism is higher than previously thought. Also a hidden fraction, represented by CUTs, was found to be expressed and rapidly degraded (Wyers, et al., 2005).
Performed in our lab, a genomic SELEX approach in E.coli identified 107 potential new ncRNAs (C.Lorenz, unpublised data) and also revealed new possible functions of ncRNAs. After the discovery of the above mentioned ncRNAs in E.Coli, their characterization became the biggest issue. 107 new Hfq binding RNAs were selected and 85 mapped antisense to mRNA coding regions (C.Lorenz, unpublished data). For my Diploma thesis I chose the characterization of the 91nt long, Hfq binding sequence as-cca (original name by C. Lorenz), lying between the already known genes htrG (b3055) and cca (b3056) on the opposite (antisense) stand.

Cca's function in adding C-C-A to tRNAs 3' termini (Schmidt, 1975) as well as its sequence (Cudny, 1986) had been discovered many years ago, however its correct function in repairing tRNA termini was not characterized until the work of Mario Mörl (Reichert and Mörl, 2000). The second gene expressed on that locus, htrG (synonyms: ecfG, ygiM), is still cryptic. There is hardly any literature about the gene and it is only known that it is a predicted signal transduction protein (SH3 domain) (www.ncbi.nlm.nih.gov/sites/entrez). One publication (Dartigalongue et al., 2001) indicated that a htrG knockout mutant displays temperature sensitive growth above 43°C.

Cca is transcribed together with htrG as a polycistronic RNA of ~2,2kb length with an unknown 3' end (cca-htrG). Additionally htrG is also transcribed as a monocistronic mRNA from a different promoter, 62nt upstream of htrG-cca promoter. HtrG's 3' end is not mapped as well. Furthermore, the antisense strand of the htrG-cca locus was completely unexplored until now.

7.2 Quality of the RNA

I extracted total RNA from e.coli using two different methods: Hot phenol extraction (Gross Lab protocol modified) and a second method using Chloramphenicol (Cam). The hot phenol extraction is more invasive/aggressive but the RNase contamination level is kept low during the isolation procedure. At first, the bacterial culture is centrifuged in the presence of an RNase stop solution (5% Phenol in EtOH). After weakening the cell wall with Lysozyme, hot phenol (65°C) is immediately added to the lysate which breaks up the rigid cell wall and inactivates all degrading enzymes like RNases, DNases or proteases. Additionally the acidic RNA phenol excludes the DNA from the aqueous phase, limiting the contamination of the extract with DNA. The second method is gentler and has the advantage that Cam stabilizes the ribosomes by binding and less RNA gets lost. Cam is added to the bacterial culture before centrifuging and the phenol extraction is preceded by a one hour DNase I treatment.

The hot phenol extraction method turned out to be the better choice based on the quality of the RNA (see Fig. 19). While the quality-indicating ribosomal RNAs seemed to be intact in the first method, they were more degraded in the second, assuming greater RNase contamination due to longer duration before phenol extraction. Also, a performed RT-PCR, as a control for RNA's quality was not satisfactory at all (results not shown).

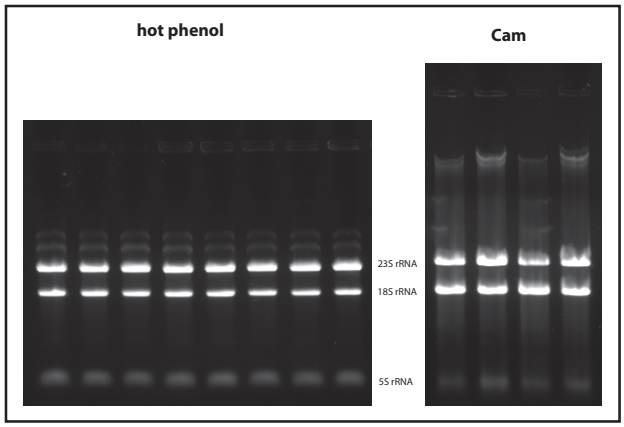


Fig. 19: Quality control of the RNA
4µg total RNA was loaded on a 1% agarose gel.
In the hot phenol extraction method (left) the 23S rRNA is 3 times more intense than the 18S rRNA.
In the extraction method using Chloramphenicol (right) the two ribosomal bands are about the same intensity. 5S RNA, as a quality control of small RNAs is visible in both methods.

7.3 Primer design

All primers were designed with the "oligo" primer analysis software (Oligo, Version 5.0 by Molecular Biology Insights Inc.). This application is based on the nearest neighbor thermodynamic data and finds single primers to a specific genomic region as well as primer pairs which are compatible to use. It analyzes the GC content, secondary structure (looping) and internal stability of an oligo, the dimerization frequency of a DNA or RNA primer and the false priming sites compared to a reference sequence. For the primer walking experiment I used the genomic sequence of E.Coli K12 htrG-cca operon with additional 1000bp upstream and downstream ("3198229 - 3202151" in genome databank) as reference. For finding optimal LNA probes, I used the whole E.Coli K12 (AC_000091) genome as reference.

The main criteria for designing primers were minimal false priming efficiencies on one hand and low dimerization ability on the other. I also tried to keep the internal stability of the 3' end low, avoiding GC pairs in the last 6 nucleotides (GC clamp), as the oligo hybridizes from 3' to 5' direction. Due to limitation to a certain sequence, the false priming efficiency of the primer was sometimes above the threshold and avoiding the dimerization of several primer pairs was also unsuccessful in some cases, because I had many forward primers (flanking primers) for one reverse primer (internal primer) and vice versa. Anyway with the help of this software I could improve my RT-PCR reactions.

7.4 Comparison of the three methods

To identify the length of the given transcript I used three different methods: primer walking, circularization and northern blot with radiolabeled LNA probes. All three methods rely basically on the ability of DNA oligos to anneal to complementary regions on the RNA (sequence specificity).

7.4.1 Strand specific primer walking

According to the primer walking experiment, the 5' end of the as-htrG-cca transcript lies between 1219nt and 1387nt proximal of Hfq binding sequence and the 3' end between 996nt and 1069nt distal from Hfq b.s. (see Fig. 11). This proposes a total length of the transcript which is at least 2205nt but less than 2547nt. However, it was not possible to amplify the overall length in one piece, a 1159nt fragment, achieved with the primer pair $R^{fl}2\text{-}5/F^{fl}3$, was the largest product using two flanking primers. However this fragment overlapped with the Northern Blot results, since the two antisense LNA probes lay inside the fragment.

Primer walking is a technique for nearly estimation of an RNA fragment's length via narrowing down its 3' and 5' ends by RT-PCR, starting from a given sequence. The essential step here is the annealing of the primer-oligo to the RNA at relatively low temperatures, hence the annealing temperature of the primer is restricted by the low incubation temperature of the reverse transcriptase (42°C - 55°C). In my case, one difficulty was that the as-cca-htrG RNA, which I analyzed, lay antisense to a known mRNA. Therefore, I had to use strand specific primers instead of random hexamers, where the annealing temperatures are significantly higher than for the latter mentioned (see table PRIMER). The usage of strand specific primers was essential, because if using random hexamers, every RNA molecule is transcribed into cDNA during reverse transcription. Under these circumstances, the analysis of one particular strand is not possible anymore. The strand specificity is only given, if one primer, complementary to the strand of interest, is used during reverse transcription.

Another problem was the contamination of the sample with genomic DNA, as the isolation of great amounts of total RNA also entails high amounts of genomic DNA. That is why an efficient DNase digestion step always had to precede the RT-PCR reaction (see materials and methods). This was indeed achieved, since all -RT controls (direct amplification of the RNA without RT step) in several performed RT-PCRs were negative.

7.4.2 Simultaneous mapping of 5' and 3' ends (Circularization)

The result of this experiment was the detection of 13 different sequences with two different distal, but 13 different proximal ends. One distal end (contained in 4 out of 13 clones) was identical with the promoter HtrGp2, the 5' start of the htrG-cca transcript. The other identified end (contained in the remaining 9 clones) did not map to any thitherto known transcription start or end (see Fig. 15).

Simultaneous mapping via RNA circularization detects both ends of the RNA in one experiment and supplies the exact length and position of the transcript. First the total E.Coli RNA (transcriptome) is treated with Tobacco Acidic Phosphatase (TAP) to maintain the next step, the self ligation of the 5' and 3' ends (Gonzales et al., unpublished). The resulting fragments are circular with a 5' to 3' ligation. A strand specific RT-PCR is performed with inverted primers and the outcome is cloned and sequenced (see also Fig. S1 in Appendix). The resulting sequence files can be observed on the computer and the 5' and the 3' ends of the transcripts can be easily identified with a conventional BLAST search.

As I mentioned before, this method requires a ligation step prior to the RT-PCR reaction. The bottle neck of this step might be the correct maintenance of the components of the ligation reaction (concentration of loose ends, concentration of DMSO). Furthermore, the enzyme T4 ligase requires a high dilution rate of the total RNA and moreover, the enzyme itself is very heat and agitation sensitive as well.

A second difficulty might be the transcript's abundance and the fact, that more than 80-85% of the transcriptome is composed of ribosomal RNA. The less abundant a transcript is, the more

improbable is to ligate together its ends and the more, in this case, 'junk' RNA is included in the reaction the less probable is the ligation of the correct 3' and 5' ends.
The dual mapping of my antisense transcript turned out to be more difficult because I had minimum three different transcripts on one locus, two on the sense strand (htrG-cca mRNA and htrG mRNA) and at least one on the antisense strand (as-hrtG-cca). So the RT-PCR had to be, again, done with inverted strand specific primers that would cover all the presumptive transcripts on the antisense strand. Therefore I first used the inverted primers $F^{fl}2$-inv and $R^{fl}2$-inv, because primer pair $R^{fl}2/F^{fl}2$ (not inverted!) gave a fantastic RT-PCR band (data not shown) and all the relevant RNAs in that region would be for sure covered, according to the primer walking experiment. Neither with the aforementioned primers nor with any other primers could I produce satisfying results. Only if I performed the reverse transcription after the ligation step (see materials and methods) with the $F^{fl}3$-inv and $R^{fl}3$-inv primer pair simultaneously in one reaction, I surprisingly got a result. The problem in the theory is that if both primers are given to the same RT reaction at once, then the strand, from which the RT-PCR product originates, is only retraceable under certain circumstances.
In my case, the fact that the resulted transcripts contained only two different distal ends (see Fig. 15) one could easily trace back the strand since 4 transcripts mapped perfectly to the transcription start of the sense htrG-cca transcript (htrGp). This implies that they were expressed from the sense strand. The rest, which distal ends did not map to any transcription start or end most probably originated from the antisense strand.

An explanation of the fluctuating proximal ends would be the position of the primers in the circularization experiment. Since the primers I used for circularization were inverted, they would only amplify regions heading outward, which is toward the ligated 5' and 3' ends of the transcripts. This means that the position of the inverted primers was limiting. The previously mentioned results were achieved with primer pairs $F^{fl}3$-inv/$R^{fl}3$-inv, lying more outward than the $F^{fl}2$-inv/$R^{fl}2$-inv primers (see Fig. 15). This denotes that all possible transcripts lying in-between the $F^{fl}3$-inv/$R^{fl}3$-inv primers were not accounted for the results. This would not exclude the existence of one or more possible transcript, lying in between the $F^{fl}3$-inv/$R^{fl}3$-inv primers and could explain the ~1200nt long antisense transcript showed with northern blot.

However, the fact that the distal ends of all transcripts mapped to different regions could either lead to the assumption that there are multiple transcripts at this locus (see 7.7 Multiple transcripts) or that they were products of truncated transcripts.

7.4.3 Northern Blots

I could detect one ~1200nt long transcript, which lay antisense to the htrG-cca gene. Both antisense LNA probes as-cca and as-htrG gave indistinguishable results, for which I supposed detected the same transcript. After the hybridization of a fresh blot with both antisense LNA probes at the same time gave identical results, I was sure that as-htrG-cca was one continuous transcript. The sense htrG RNA was ~900nt long with a ~170nt long thitherto unknown 3' UTR. The cca mRNA which was supposed to be ~2000nt long was not detectable at any condition or growth state. This could be a sign for low abundance or for inductive transcription at conditions I haven't used.

The northern blot approach is, apart from the labeling of the probe, an enzyme free application, relying on the hybridization of the probe to the RNA. A successful northern blot should prove the existence of a transcript and give us an insight into its abundance, its size, strand specificity and also into its expression pattern. If using northern blot for qualitative and quantitative analysis, one has to consider that its detection level is much lower than those of RT-PCR based methods, because no amplification step is involved. A strong signal is therefore composed of the amount of the total RNA per lane on the agarose gel, the specific activity of the probe and the equilibrium of the annealing reaction which is: $probe_{free}$ <---> $probe_{annealed}$. For the detection of low abundant transcripts all three conditions have to be optimal. Since the steady state for LNA probes is on the 'annealed' side (LNA_{free} <--->>> $LNA_{annealed}$), I preferred LNA probes to DNA/RNA probes. The LNA molecule is locked in its 3'-endo structural conformation and therefore has a reduced structural flexibility.
Another benefit using LNA is the hybridization temperature. I used LNA probes at temperatures, where DNA or RNA probes would not work anymore. The temperature I used to hy-

bridize the membrane with the LNA probes (70°C - 75°C) was 10°C - 15°C higher than for riboprobes (RNA-probes), which had also a positive effect on the background noise.

The low abundance of the antisense, but also of the sense transcripts was the major difficulty I had to handle. My first experiments with riboprobes, synthesized by hot transcription, completely failed to bring any result even though I used excessive amounts of total RNA per lane and harsh methods for labeling the probes (2 different radiolabeled dNTPs in high concentration - see materials and methods). The strong background signal of the ribosomal RNA dominated the radioblot. Only hybridization of the membrane with LNA probes gave the expected results.

7.5 Multiple bands

A commonality of the RT-PCR based results was nevertheless the appearance of multiple bands on the agarose gels, ergo some primer pairs produced more than one band. Noticeable was also that if there was more than one band per lane, the largest band was always the correct sized one (see Fig. 11A+B) and that the other bands were smaller. A common striking pattern was not detectable among the multiple bands. As all my -RT controls were negative, I can exclude that the multiple bands originated from genomic DNA. To explain this phenomenon, a couple of theories could be proposed, as also depicted in Fig. 20.

Mispriming during reverse transcription (Fig. 20A - top)

If the primer's 3' end, where the annealing to the RNA template initiates, has a strong affinity to a different region of the transcript (like a GC clamp), then it is likely that the RT enzyme takes that misprimed region as the starting point for reverse transcription. If this happens, the resulting bands exhibit a different size than expected.

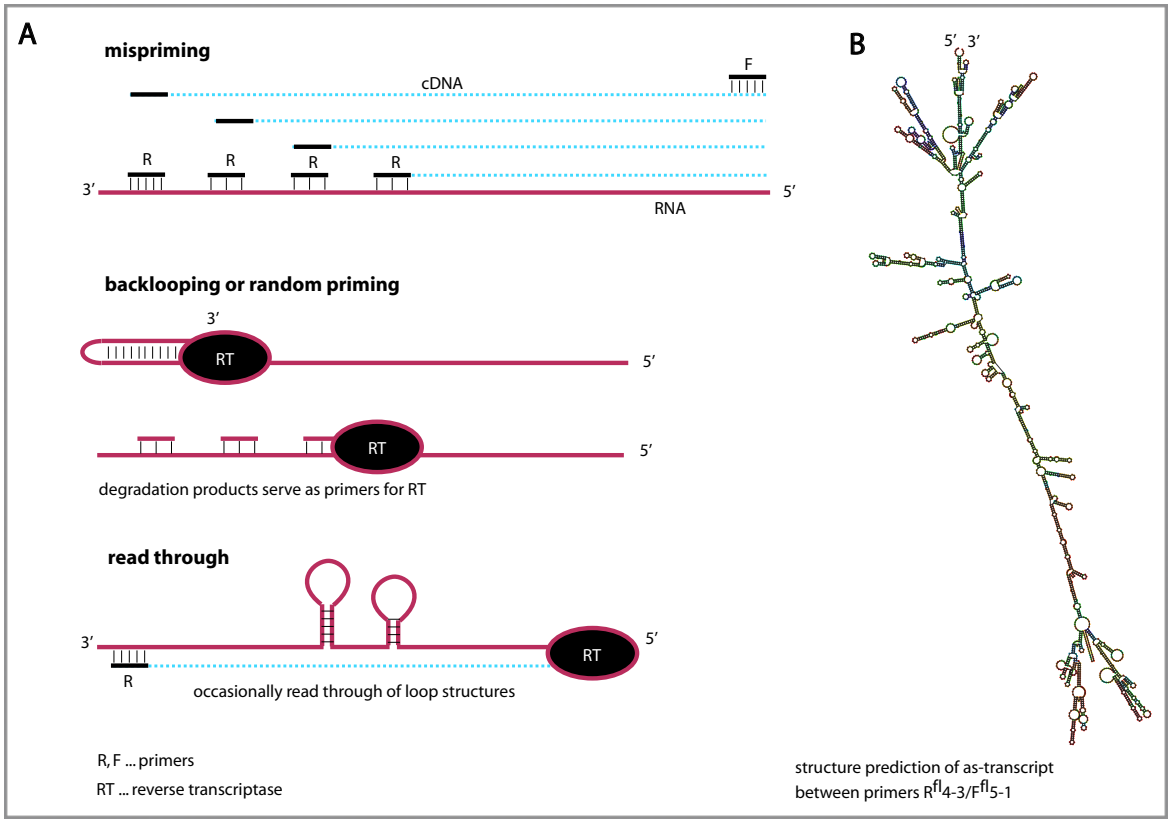


Fig. 20: Multiple band theory

A: Three different models, how multiple bands can occur after RT-PCR are indicated.
mispriming: The primer, used for RT-PCR, could anneal to unspecific regions of the RNA thereby producing smaller cDNA fragments, but containing the correct primer sequence for the second strand synthesis.
backlooping or random priming: The 3' end of the RNA could backloop and serve as a primer for the RT enzyme. A second theory is that degradation products could serve as primers for the RT.
read through: highly structures of the RNA could be occasionally read through and skipped in the resulting cDNA, thereby producing cDNAs of different length.
B: The predicted secondary structure of the antisense transcript between the two flanking primers R^{fl}4-3 and F^{fl}5-1 (MFE Plot). Both were the last flanking primers, resulting in a band in the primer walking experiment. predicted by Vienna RNA Fold web server (Hofacker et al., 1994)

Since the annealing temperature of the RT primer, which was limited by the incubation temperature of the RT enzyme, was 15°C - 20°C lower than the primer's calculated Tm (see table PRIMER), mispriming could have lead to more than one primer binding site. This theory would propose that between the original primer binding sites, there are several other, less efficient binding sites, but if the annealing temperature is low enough, the primer would still anneal and produce a cDNA with the correct primer sequence.

One experiment that especially strengthens this theory was done in the course of the primer walking experiments. For the primer pairs R^{Hfq}/F^{fl}4-2 and R^{Hfq}/F^{fl}5-1 (Fig. 11B) I used a special RT enzyme (Thermoscript by Invitrogen), which was designed to tolerate higher incubation temperatures, and performed the RT step on 55°C and 60°C. Although the multiple bands did not vanish completely they became fainter after raising the temperature by only 5°C.

Random priming (Fig. 20A - middle)

A work from 2007 stated that there is a significant production of primer-independent cDNA synthesis using an RNase H- RT enzyme. They show that a PCR product was amplified from RT reactions that were carried out without addition of RT-primer (Haddad et al., 2007).

Similar results were observed in our lab by Ivana Bilusic. She did a strand-specific reverse transcription with a primer, specific for a certain gene. Using this cDNA as a template, she could amplify PCR fragments with primer pairs, specific for completely different genes. The correctness of the experiment was given by negative RT- controls as well as complete inactivation of the RT-enzyme (Ivana Bilusic personal communication).

A possible explanation for this phenomenon would be random priming by short degradational products of RNA/DNA or self priming by back looping.

Read through (Fig. 20A - bottom)

Highly structured regions in the RNA, as stem-loops, could be read through by the RT enzyme and produce a smaller cDNA with deleted sites in the middle, but with the correct flanking primer sequences. As the region, where the as-htrG-cca RNA was predicted, is highly structured (Fig. 20B), the possibility that the RT enzyme skips several loops is given. As a counter-argument, the RT reaction always preceded a 5 minute denaturation step of the RNA template and the RT-primer (see materials and methods).

An adequate experiment to see whether the smaller fragments in the multiple bands originated from the same locus would have been a southern blot with the RT-PCR products or the isolation of several bands and sequencing them.

7.6 Computational analysis

7.6.1 Promoter sequence analysis

A promoter sequence analysis was performed with 3 different web applications (Promscan[ref], BPROM[ref], BDGP[ref]). As reference I fed the sequence of the HtrG-cca locus, started 415bp upstream of htrG transcription start and ended 632bp downstream of cca ORF end (3198711-3200481 of the e.coli genome).

Summing up the results of all three applications 15 promoters were found on the sense strand, expressing htrG-cca, and 18 on the antisense strand, expressing as-htrG-cca. In Fig. 21, all the potential promoter regions are depicted. A peculiarity of the results is that there are several hotspots for transcription starts, found by more than one application, on both strands. Unfortunately, no promoter sequence was found ~1200nt upstream of the predicted 3' end of the antisense transcript, which would have strengthened the results of the circularization experiment (see also Fig. 15).

7.6.2 Transcription terminator sequence analysis

I also performed a web search for potential rho independent transcription terminators with the web software FindTerm (Softberry, Inc.). As reference I fed the sequence of the htrG-cca locus, started 415bp upstream of htrG transcription start and ended 632bp downstream of cca ORF end (3198711-3200481 of e.coli genome), to the program. Again several potential terminator sequences were found all over the locus, but I concentrated on the region concordantly with the circularization, primer walking and northern blot results.

On the antisense strand a terminator loop structure was found 20nt upstream of the 3' end found with RNA circularization. This would be a relevant evidence that the 3' end of the antisense transcript was 643nt downstream of Hfq binding site. A second potential terminator loop structure was found by observing the region proximal to the primer R^{fl}2-5, which was the reverse primer, I could amplify a 1159bp long RT-PCR fragment with. This would fit the longest continuous RT-PCR fragment, into the Northern Blot results. On the sense strand, the found terminator loop was 930nt downstream of the htrG transcription start, which would per-

fectly fit into the northern blot results (Fig. 17). Their typical terminator loop structure, calculated by the Vienna RNAfold WebServer (Hofacker et al., 1994) are shown in Fig. 21 as well.

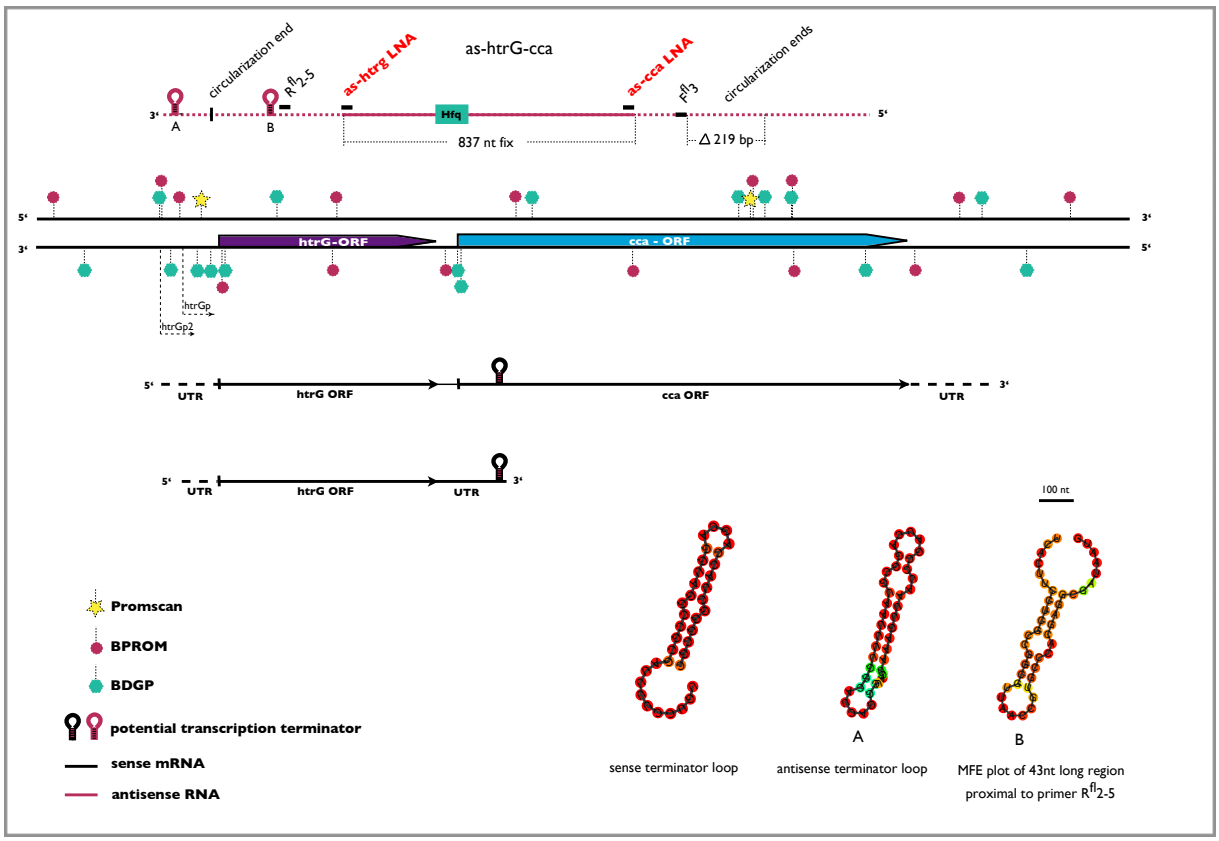


Fig. 21: Computational analysis

The antisense transcript (as-htrG-cca) is displayed at the top of the graph and the sense transcripts (htrG, htrG-cca) by two black arrows. The results of the primer walking, RNA circularization and Northern Blot experiments are indicated in the top. The known region between the two antisense LNA probes is drawn in a continuous red line (837nt fix), the unknown flanking regions in a dotted red line. The location of the primers of the longest continuous RT-PCR fragment are also indicated ($R^{fl}2$-5, $F^{fl}3$). The proximal end of the circularization is indicated by a vertical black line and 'Δ219' marks the position of the multiple distal ends.

Promoter prediction: The presumptive promoters, predicted with three different web servers are indicated by a red dot (BPROM), green hexagon (BDGP) and yellow asterisk (Promscan). The positive DNA strand, expressing htrG and cca mRNA is indicated by the lower black line. The negative (antisense) strand, expressing the antisense RNA, is indicated by the upper black line.

Terminator prediction: The predicted terminator loops are indicated by comic stem loops. The predicted secondary structures of the terminator loop sequences are shown in the right bottom.

The terminator sequences were predicted with the web server FindTerm (Softberry, Inc.), structure prediction by Vienna RNA Fold web server (Hofacker et al., 1994).

As the predicted terminator sequence of the htrG mRNA lay in the open reading frame (ORF) of the cca gene, the question came up, how the cca transcript avoids its transcription termina-

tion. There must be a mechanism which prevents transcription termination of cca. Possibilities would be that RNA chaperones bind the pre-transcribed RNA and prevent loop structure folding. A second possibility would be that 5' leader sequences align with the forming loop structure and thereby prevent transcription termination, in a riboswitch like manner (Tucker et al., 2005). Also the involvement of an antisense transcript is conceivable.

7.7 Multiple transcripts

Taking together the experimental and computational results, the existence of multiple sense and antisense transcripts on the htrG-cca locus is indicated in many points.
First, the primer walking experiment would suggest a possibility of at least 3 different transcripts on that locus, as the total length of the antisense RNA was mapped to minimum 2205nt, but it was not possible to amplify the overall length in one piece. Only a 1159nt long region could be amplified with two flanking primers at once (Fig. 12). Second, the RNA circularization experiment revealed only one possible 3' end, but the individual circularized transcripts had several 5' ends, which was an evidence for multiple transcripts. Even parts of the northern blot result would fit into this multiple transcript theory, as faint bands appear in addition to the expected ones (Fig. 16). Multiple transcripts, however, would furthermore implicate different promoters distributed all over the locus. This was calculated by computational analysis. The reason for these phenomena could be as multiple as the transcripts.
For sure all results rely amongst others on the quality of the experiments as well as on the quality of the components (RNA, primers, enzymes, etc.). The chance that there was a problem in these parameters is always given, but as several control reactions were all right I assume that the experimental error is negligible.
A possible explanation to these phenomena comes from the bakers yeast (Saccharomyces Cerevisiae), which would explain the inconsistency among the experiments identifying the length of the transcripts. In a work, done by Alain Jacquier's group, the authors first described a new form of low abundant Pol II transcripts, rapidly degraded by a polyadenylation assisted degradation mechanism (Wyers et al., 2005). In a yeast strain, deficient of the nuclear exosome subunit Rrp6, polyadenylated transcripts from presumably silent intergenic regions un-

expectedly accumulated. Interestingly most of the newly discovered transcripts were not mediated by PAP I, the classical polyadenylation machinery, but by the Trf4/Air1 or Air2 complex, also taking part in the quality control pathway of initiator tRNAmet (Vanácová et al., 2005). Expectedly the polyadenylation of most cryptic transcripts was abolished in a Trf4 mutant strain leading to their further stabilization (Wyers et al., 2005).
Another distinctive feature was their termination. Arigo et al., described that proteins Nrd1 and Nab3, required for the termination of sn/snoRNAs, is required for the termination of cryptic unstable transcripts as well. Nrd1/Nab3 colocalize with genomic regions expressing CUTs and are a prerequisite for rapid degradation by the yeast nucleolar exosome (TRAMP complex) (Arigo et al., 2006). The comparison of the termination mechanism of Pol II transcripts, sno RNAs and CUTs is reviewed by Lykke-Andersen (Lykke-Andersen et al., 2007)

In more recent works it also became clear that CUTs represented the largest group of hidden transcripts in yeast (Gudipati, 2008), and most of them originated from bidirectional promoters in nucleolar free regions (Xu et al., 2009). Anyhow, their function is as cryptic as the transcripts themselves. All these works show that next to the already known transcription of 'detectable' RNAs, there is also excessive hidden transcription going on, resulting in unstable transcripts which are degraded rapidly by the exosome. Therefore the hidden transcripts stay low abundant.

As all the previously mentioned work was performed in a eucaryotic organism, we can just speculate about the existence of 'CUT like transcripts' in E.Coli. If there is something of the kind as cryptic transcription in E.Coli, its composition and mechanism would be distinct, due to differences in eukaryotic and prokaryotic transcription and degradation machinery. All three methods I used for determination of the transcript's length had a very different detection limit which ranged from theoretical one RNA molecule in RT-PCR to 10.000 copies in northern blots (http://www.ambion.com/techlib/append/mRNA_detection_chart.html). It is possible that there is a 1200nt long antisense transcript on the htrG-cca locus, detected by northern blot, and additionally an undefined number of low level transcripts, detected only by RT-PCR.

The computational analyses would also strengthen the multiple transcript theory, since several promoter sites were found on the htrG-cca locus in both directions. It is possible that there is low abundant transcription going on all over the htrG-cca locus but due to low abundance, it is hardly detectable.

7.8 Expression pattern

As already mentioned in the introduction, the advantage of genomic SELEX is that the selection considers all RNAs independent of their growth states (Lorenz, et al., 2006). The disadvantage is that one has to determine under which growth conditions the selected RNAs are expressed.

For the northern blot experiments I used total RNAs, expressed under four different growth conditions: RNA extracted from bacteria growing at 37°C to logarithmic phase, from bacteria growing at 37°C to stationary phase, from bacteria growing at 37°C to logarithmic phase then for another 40 minutes under acidified environment and RNA from bacteria growing at 37°C to logarithmic phase then another 3 hours at 42°C. For the circularization experiment I used RNA from logarithmic phase, stationary phase and acidic stress. For the primer walking I only used total RNA from logarithmic phase.

As for the circularization experiment I used both primers in one RT-reaction, one cannot distinguish a hundred per cent, which transcripts are expressed under which conditions. It is only clear that the experiment worked for logarithmic phase and for acidic stress. As the amplification of the antisense strand was achieved during primer walking experiment, I can suggest that the antisense hrtG-cca transcript is expressed in the log phase. However, in the northern blot experiments the total RNA from 4 expression conditions was loaded next to each other. As shown in Fig. 16 C, the antisense htrG-cca was only detectable under logarithmic phase, but not under any other condition.

For the sense transcripts the situation was different. It is thought that two transcripts, htrG and cca, are transcribed from the htrG-cca locus as a long polycistronic mRNA. Additionally the htrG is transcribed as a monocistronic mRNA as well. The htrG-cca transcript was not detectable under any condition I used. This would be logical, as the cca gene is thought to work on damaged tRNAs (Schürer et al., 2001). Maybe the cca gene is only expressed after a large amount of damaged tRNAs accumulated. Here, additional growth conditions, like oxidative stress (Evans et al., 2004; Gong et al, 2006) or X-irradiation, would have probably been necessary to detect the htrG-cca RNA. The monocistronic htrG mRNA was indeed expressed, namely in log phase and acidic stress. The mRNA is more abundant than its antisense counterpart, as the transcript from log phase RNA is already detectable at 20µg total RNA (Fig. 16B). The transcript detected under acidic stress conditions is about the same quantity as the antisense transcript.

Although there is no plausible explanation for the differences in the expression patterns yet, I for sure identified a 1200nt long, thitherto unknown transcript, lying antisense to the htrG-cca mRNA. As the new transcript is an antisense RNA lying opposite of a coding region (cis) I also identified a member of a new class of long-cis-antisense RNAs.
Furthermore I first showed the existence of the htrG mRNA by northern blot and determined its length which is 900nt with an estimated fluctuation of 20nt-40nt (vertical diffusion of the RNA marker band in the gel). Moreover, if the terminator sequence analysis was correct, I could also map its 3' end.

7.9 Possible Functions

Summing up the functions of the already known cis acting antisense RNAs, their major role is post transcriptional inhibition of target RNA functions affecting fundamental cellular processes (reviewed in Brantl, 2007). This may be also correct for long-cis-antisense RNA, although their length would exclude some of the known mechanisms as inhibition of pseudoknot formation (Inc/RepZ), inhibition of primer maturation (RNAI/RNAII) or RNA stabilization (GadY/GadX) - see introduction.

The most probable function of these RNAs would be translation inhibition by mRNA degradation after duplex formation. If two transcripts exhibit 100 per cent sequence complementarity they might need a chaperone like Hfq to anneal to each other and inhibit their functions. RNAseIII might also play a role in this mechanism. As in my case the operating RNAs are over 1000nt in length, a full duplex formation of the sense and antisense RNA seems to be unrealistic. Probably a local double strand RNA formation is sufficient for their mode of action.

A recent work of Mario Mörl's lab indicated that Hfq influences not only the Poly-A-Polymerase protein in E.Coli (Hajnsdorf et al., 2000), but also the CCA adding enzyme (Scheibe et al., 2007). They performed a CCA-adding assay with the tRNA nucleotidyltransferase in the presence and absence of Hfq. The result was that Hfq accelerated the addition of CCA to the model tRNA (yeast $tRNA^{Phe}$, lacking CCA triplet) end. It was also shown that Hfq interacts with tRNA substrate independent of the presence or absence of the CCA end. This indicates a function of Hfq, next to the CCA adding enzyme, in the restoration of tRNA ends.

How antisense cca-htrG could function on the CCA adding enzyme is only speculative. As a probable mechanism, I would suggest that as-htrG-cca acts on the sense mRNAs (htrG, htrG-cca) in a sequence specific inhibitory manner. Since as-htrG-cca contains an Hfq binding sequence (Schöberl U., unpublished data) and this sequence binds Hfq in vitro (Schöberl U., unpublished data), it is most probable that Hfq also binds the full length as-HtrG-cca transcript in vivo. This would indicate that the antisense cca-htrG together with the CCA adding enzyme and Hfq regulate the terminal CCA repair of tRNAs in E.Coli. The low abundance of as-htrG-cca would suggest a fine tuning function.

For the identification of the mode of action on the htrG-mRNA, it would be inevitable to improve our knowledge about the htrG gene itself and its function. Only then could we propose a speculative mode of functioning which I believe is similar to cca.

8. CONCLUSIONS AND FUTURE PROSPECTS

CONCLUSIONS

Concluding my thesis, I was the first to analyze the cis antisense RNA as-htrG-cca.
I showed its existence and its actual length via northern blot, which was ~1200nt, and proved that it is encoded on the antisense strand. According to the knowledge of the distribution of my northern blot probes, the location of the antisense transcript could be bordered to 837nt. The distribution of the remaining 363nt from total 1200nt is still unknown.
Furthermore I also showed that there is a significant difference in analyzing the transcript via RT-PCR or via northern blot. The more sensitive RT-PCR techniques indicated the existence of many cis antisense RNAs on that locus. On the other hand, northern analysis revealed only one transcript. This could imply that along with many low abundant as-htrG-cca transcripts, there is one 1200nt long asRNA, which is sufficiently abundant to be detected by northern blot. However the novelty of my findings was that - independent from their identification method - all the mentioned transcripts were bigger than 1000nt. Until now there was no such big cis antisense RNA identified, which may be an evidence for a new group of long cis antisense RNAs.

However, at this point of the project we are not able to tell, whether as-cca-htrG exhibit a specific function or it is just an evolutionary rudiment. Further investigations will be necessary to validate one of the two possibilities.

Together with the project's actual ambitions, a second finding revealed the length of the htrG mRNA with its 170nt 3' UTR. I was the first to show the ~900nt long htrG mRNA via northern blot.

FUTURE PROSPECTS

All these findings bring up the question of what the function of long cis antisense RNAs may be in general. The functions of already known cis antisense RNAs allow the assumption that their major role could lie in the post transcriptional inhibition of the target RNA's functions. This could affect fundamental cellular processes like transcription, translation or mRNA stability. However, one has to consider that the actual length of the hitherto known cis as RNA differ from my findings by almost 10 fold. Their huge lengths could be obstructive for fast action as they are bulky and slow.

The subsequent step in the ongoing htrG-cca project will be the exact mapping of the as-htrG-cca RNA. The calculation of presumptive promoter and terminator sequences was the first step to do so.
Considering the whole "Hfq binding cis asRNA" project, the future work will be dealing with the investigation of the of the remaining 84 cis asRNAs. There is a legitimate probability that most of the latter mentioned RNAs exhibit similar lengths and build, together with as-htrG-cca, a novel group of long cis antisense RNAs. Their commonalty of Hfq binding may also entail a commonalty in function. A part of the remaining SELEX RNAs is currently analyzed in Renée Schroeder's lab by Ivana Bilusic. Her findings will hopefully provide more information about the little explored cis antisense RNAs.
The second big issue will be the bioinformatical approach to analyze sequence and structure of these RNAs. With today's databases of functional RNA motifs and secondary structures there is a good chance to succeed. An already performed secondary structure computation of the SELEX RNAs (Hfq binding sequences) by Tanja Gesell was the first step.
Only if we manage to gain a broad basis knowledge about the long cis antisense RNAs, we can fancy with the prospect that one day we will discover their functions.

9. MATERIALS AND METHODS

9.1 Used Bacteria strain and Media

9.1.1 Bacteria

E.coli B

BHLα bacteria cell (NEB)

9.1.2 Growth medium (Luria Broth LB)

10g Bacto-Tryptone, 5g Bacto-Yeast extract and 10g NaCl in 900 ml H_2O, adjust pH to 7.0 then adjust volume to 1L.

Sterilize by autoclaving.

9.2 Growth and Stress conditions

The liquid bacteria cultures were always grown in a shaking incubator at the indicated temperatures while shaking at 250rpm!

Overnight Culture:

A single E.coli K12 colony was inoculated in 5ml LB medium as starter culture. Cells were grown at 37°C over night (O/N).

Growth conditions:

The O/N culture was diluted 1:100 in fresh LB media and divided into two tubes. One of them was let to grow until logarithmic phase (OD_{600} ~ 0,4 - 0,5) before the cells were harvested. The other tube was incubated until **stationary phase (**OD_{600} ~ 4,0) before cells were harvested.

Temperature stress:

The O/N culture was diluted 1:100 in fresh LB media.

Bacteria was grown at 37°C to OD_{600} ~ 0,4, then at 42°C for another 3h.

Acidic Stress:

The O/N culture was diluted 1:100 in fresh LB media.

Bacteria was grown to OD_{600} ~ 0,3 - 0,4, then MES (2-(N-morpholino)ethanesulfonic acid) was added to a final concentration of 170 mM. This acidified the medium to pH 5. The cells were incubated for another 40 min at 37°C before harvesting.

9.3 Isolation of total RNA from E. coli Cells

9.3.1 Hot phenol extraction (modified Gross Lab Protocol)

Bacteria cultures were treated with 1/10 volume ice cold stop solution (5% water saturated phenol, pH < 4, in ethanol) to repress RNA degradation and centrifuged (8.000 rpm for 2 min at 4°C). The pellet was immediately frozen in liquid nitrogen. The cells were lysed by re-suspending in 4 ml TE buffer (10 mM Tris-Cl, 1 mM EDTA) containing 2 mg/ml lysozyme. 400µl 10% SDS was added and incubated at 64°C for 2 min. After incubation, 440µl 1 M NaOAc (pH 5,2) was added and the sample was transferred into 15 ml falcon tubes containing the same volume of phenol.

The tubes were incubated for 7 min in a 64°C water bath while inverting them 10 times every 40 sec or so. Then the samples were chilled on ice and centrifuged for 15 min at 10,000 rpm at 4°C. The aqueous phase was phenol extracted again and precipitated in the presence of 0.3 M NaOAc pH 5.2, 1 mM EDTA and 2.5 Vol. $EtOH_{abs}$ for 30 min at -80°C. The precipitate was centrifuged at 10.000rpm for 45 min at 4°C and the pellet was washed in 80% ice cold EtOH, air dried and dissolved in 800µl nuclease free H_2O.

The samples were further purified in 2 ml Phase Lock Gel Tubes (5-PRIME) once with Phenol:Chloroform:Isoamylalcohol (24:24:1) - PCI, then with Chloroform:Isoamylalcohol (24:1) CI and finally precipitated again as written above. The pellet was washed 2 times with ice cold 80% EtOH and dissolved in 200µl RNase free water.

9.3.2 RNA extraction using Chloramphenicol (Cam)

45ml bacteria culture was poured into a chilled, 50 ml Falcon Tube including 200µl Chloramphenicol (25 mg/ml) and centrifuged for 5 min at 5000rpm. The cells were re-suspended in 1 ml ice cold TM Buffer (10 mM Tris HCl pH7 and 10 mM $MgCl_2$) and transferred into a 1,5

ml Eppendorf tube, containing 5µl Cam (25 mg/ml). Again the sample was spun down for 5 min at 5000 rpm and the supernatant was completely removed before the pellet was shock frozen in liquid nitrogen. The frozen pellet was carefully re-suspended in 150 µl buffer A (150µl TE buffer, 1,5µl 1 M DTT, 0,75µl RNase Inhibitor, 4µl 10 mg/ml lysozyme) by vortexing. The pellet was dissolved by 3 cycles of freezing the cells in liquid nitrogen and immediately thawing them in a 37°C water bath, vortexing vigorously between the cycles. After the pellet was dissolved, 20µl solution B (200 mM $MgAc_2$, 7U DNase I, 10u RNase inhibitor) was added and the suspension was incubated on ice for 1h. 2µl stop solution (1µl 0,2 M CH_3COOH, 1µl 10% SDS) was added and the samples were incubated for 5 min at room temperature.

The sample was transferred into 2 ml Phase Lock Gel Tubes and extracted with Phenol:Chlorophorm:Isoamylalcohol (24:24:1) once, then with Chlorophorm:Isoamylalcohol (24:1). The extracted sample was finally precipitated in 2.5 volumes EtOH and centrifuged by full speed for 45 min at 4°C. The pellet was washed 2 times with 80% EtOH and re-suspended in 30µl RNase free water.

(Protocol was kindly provided by Oliver Mayer)

9.3.3 DNase I treatment

Working with RNA, the removal of all contaminating DNA from the samples is indispensable for achieving unbiased results. This is attained by intense DNase I treatment. Usually, high amounts of RNA implicate a high degree of DNA contamination so the sample has to be digested with high quantity of DNase I for a long time. Upscaling the DNase reaction can be problematic too. The trouble with DNases is that they start digesting the RNA after a certain time as well.

As a conclusion, I always kept the incubation time below one hour, and repeated the DNase treatment twice with a phenol extraction step in between.

The DNase I treatment was performed at 37°C in the presence of RNase inhibitor (NEB).

I used following setup:

500 - 1000µg total RNA, 40U DNase I (NEB), 1x DNase I buffer (10 mM Tris-HCl, 2.5 mM $MgCl_2$, 0.5 mM $CaCl_2$), 160U RNase inhibitor (NEB). The reaction volume was chosen so that the volume of Glycerol (contained in Enzyme and RNase Inhibitor) was <10%.

9.4 Primer Walking

9.4.1 Primer design

All primers were designed with the "oligo" primer analysis software (Oligo, Version 5.0 by Molecular Biology Insights Inc.). This application is based on the nearest neighbor thermodynamic data and finds single primers to a specific genomic region as well as primer pairs which are compatible to use. The program analyzes the GC content, secondary structure (looping) and internal stability of an oligo, the dimerization frequency of a DNA or RNA primer, and the false priming sites compared to a reference sequence.

For the primer walking experiment I used the genomic sequence of E.coli K12 htrG-cca operon with additional 1000bp upstream and downstream ("3198229 - 3202151" bp in E.coli genome) as reference (www.ecogene.com).

The main criteria for designing primers were the minimizing of false priming efficiencies on one hand and reducing the dimerization ability of the primers pairs on the other. I also tried to keep the internal stability of the 3' end of the primers low, avoiding GC pairs in the last 6 nucleotides (GC clamp), since the oligo hybridizes to the template at the 3' end first and then proceeding to 5' direction.

Even though the false priming efficiencies of certain primers could not always be kept low I could improve the quality of my RT-PCR reactions with the help of this software.

9.4.2 RT-PCR

9.4.2.1 One step RT-PCR (QIAGEN, 210212)

Kit was used according to manufacturer's instructions.

9.4.2.2 RT-PCR with M-MLV-RT (Promega, M1701)

4 µg total RNA + 100 pmole reverse primer were denatured at 94°C for 5 min, then immediately chilled on ice. The reaction was performed in presence of 1x M-MLV buffer, 30u RNase inhibitor, 12,5 nmole dNTPs and 200U M-MLV reverse transcriptase. The incubation time for the cDNA synthesis was 1 hour at 50°C.

As M-MLV Reverse Transcriptase has a low RNase H activity, 5U of RNase H were added and incubated for another 20 min to degrade the RNA template.

After incubation, the enzymes were inactivated 15 min at 75°C.

10µl of the reaction was added to the subsequent PCR master mix and the second-strand synthesis was performed as described in section 9.4.3).

9.4.2.3 RT-PCR with Thermoscript™ RT-PCR kit (Invitrogen, 12236-014)

'Thermoscript RT' is an avian reverse transcriptase and has a higher thermal stability then M-MLV RT. It can generate cDNA at temperatures between 50°C and 65°C.

The kit was used according to manufacturer's instructions.

9.4.3 PCR amplification

The PCR reaction was performed in 50µl reaction volume as follows:

1x PCR Buffer (NEB), 0,2 mM dNTPs, 3 mM $MgCl_2$, 0,4 µM forward primer, 0,4 µM reverse primer, 100 pg DNA template (or 1/5 vol/vol RT-template), 2U TAQ-Polymerase (NEB)

The reaction was mixed gently and the PCR reaction was performed in a Thermocycler (VWR, Qattro Chass) with heated lid.

The PCR steps were 5min denaturation at 95°C, 30 cycles (X" denaturation at 95°C, Y" annealing at Tm-5°C, Z" polymerization at 72°C,), 10' extension step at 72°C, store at 4°C. Denaturation step (X) was 1 minute per 1000 bp, annealing step (Y) was 30 sec, elongation step

(Y) was 1 minute per 1000 bp. The annealing temperature was 5°C lower than the average Tm of the two primers.
After amplification, 10µl sample was mixed with 1.5µl 6X loading dye (Promega) and checked on agarose gel.

9.5 Simultaneous mapping of 5' and 3' ends (Circularization)

The intention of this experiment is to circularize every single RNA molecule in the entire total-RNA extract by ligating with T4 ssRNA Ligase and perform an RT-PCR reaction with inverted primers (see also Fig. S1 in Appendix). The T4 RNA Ligase catalyzes the ATP-dependent intra- and intermolecular formation of phosphodiester bonds between 5'-phosphate and 3'-hydroxyl termini of oligonucleotides, single-stranded RNA and DNA.

9.5.1 CAP sequence removal with TAP

Since all fresh transcribed RNAs contain a 5' phosphate cap which makes the ligation slow, the first step was the removal of the 5' phosphate group by Tobacco Acidic Phosphatase (TAP). TAP hydrolyzes the phosphoric acid anhydride bonds in the triphosphate bridge of the cap structure, releasing the cap nucleoside and generating a 5´-phosphorylated terminus on the RNA molecule. I used 12µg total RNA with 20U TAP enzyme (NEB) in 1x manufacturer's buffer (NEB). Additionally I added 20U RNase inhibitor (NEB) to the sample and incubated the reaction at 37°C for 45 min. After incubation, the enzyme was removed by PCI and CI extraction and subsequent EtOH precipitation.

9.5.2 T4 RNA ligase reaction

The TAP treated RNA was denatured at 95°C for 5 min and the ligation reaction was prepared as following:
Four different quantities of total RNA (1µg, 500µg, 100µg and 50µg) were ligated with 40U T4 RNA Ligase I in the presence of 1x manufacturer's buffer (NEB), 1U DNase I (NEB), 20U RNase inhibitor (NEB) and 10% (vol/vol) DMSO.

The incubation time was 1h at 37°C. Again, the sample was purified with PCI and CI extraction and after EtOH precipitation finally dissolved in 10µl RNase-free ddH_2O.

The following RT-PCR reactions (see also 9.4.2 RT-PCR) were performed with inverted primers. Inverted means the reverse complement of the primer (e.g. 5'-ACGTA-3' -> 5'-TACGT-3'). The primers are designed in reverse complement direction, so that the reverse transcription goes into the opposite direction (outside form the known region). The reverse transcription was done with the inverted forward primer and after inactivation of the RT-enzyme, the second primer (inverted reverse primer) was added and a straight forward PCR reaction was performed as described in section 9.4.3.

The following controls were performed: RT PCR with TAP treated non ligated total RNA as negative ligation control (c). RT PCR with ligated non TAP treated total RNA as negative TAP-treatment control (neg). RT PCR with ligated, TAP treated total RNA with normal orientation primers as positive RT-PCR control (pos). A PCR reaction (without RT step) with total RNA and inverted primers as a control for DNA contamination (-RT). After RT PCR the samples were loaded on an agarose gel and analyzed. If all the control reactions were all correct then the samples were cloned into TA vectors and sequenced.

9.5.3 Agarose gel electrophoresis

I used 1% agarose gels for expected fragment lengths over 100bp and 2% agarose gels for smaller fragments.

1g low MP Agarose was dissolved in 100 ml 1xTBE buffer by heating up in the microwave oven. After the agarose was completely dissolved, the solution was allowed to cool down to about 60°C. Then 0,5 µg/ml Ethidium Bromide was added, mixed and immediately poured into a gel-tray or to a plexiglass plate (surface tension gel). The gel was allowed to cool down at room temperature until it was solid.

The samples were loaded in Bromophenol Blue containing loading dye and the gel was run with 100 V in 1x TBE running buffer until the tracking dye reached the bottom of the gel.

9.5.3.1 Recovery of the fragments from the agarose gel

For recovery of fragments from Agarose gels I used the Wizard®SV Gel and PCR Clean-Up System by Promega (Cat No:A9340) according manufacturer's protocol.

The Wizard® SV Gel and PCR Clean-Up System are based on the ability of DNA to bind to silica membranes in the presence of chaotropic salts. After electrophoresis, the band of interest was excised and dissolved in the presence of guanidine isothiocyanate (Membrane Binding Solution applied by kit). DNA fragments were isolated using centrifugation to force the dissolved gel slice through the membrane while simultaneously binding the DNA on the surface of the silica. After washing the isolated DNA fragment the DNA is eluted in water. (www.promega.com)

9.5.4 TA cloning

9.5.4.1 Ligation

I used pGEM-T cloning Kit by Promega (A3600) according to manufacturer's instructions.

The pGEM®-T Vector is linearized vector with a single 3´-terminal thymidine at both ends. The T-overhangs at the insertion site greatly improve the efficiency of ligation of PCR products by preventing re-circularization of the vector and providing a compatible overhang for PCR products generated by TAQ thermostable polymerases. (www.promega.com)

For vector : insert ratio I chose 1:1, 1:3, 1:5. The whole ligation mix was used up for transformation.

9.5.4.2 Transformation

For the transformation I used self-made BHLα competent cells.

I mixed 100µl cells and 10µl ligation very gently and incubated 20 min on ice then heat shocked the sample for 90 sec at 42°C and chilled the tube immediately for another 5 minutes. 900µl sterile LB media was added and incubated 45 min at 37°C with 200rpm shaking.

The cells were centrifuged and the overflow was discarded. The pellet was resuspended in the remaining drop of the supernatant and spread on LB amp (100 µg/ml) + IPTG (0,5 mM) +XGal (80 µg/ml) plates and incubated O/N at 37°C.

Single colonies were picked up and used for plasmid DNA purification.

9.5.4.3 Competent cells

One colony of BHLα bacteria cell was inoculated in a 5µl LB medium and incubated at 37°C overnight while vigorous shaking.

The O/N stock culture was diluted 1:100 in fresh LB medium to a total volume of 100 ml and incubated at 37°C while shaking until the medium reached OD_{600} ~ 0,4 - 0,5. Then the cells were cooled down and kept on ice for 10 min before being pelleted by centrifugation (4000 rpm for 10 min at 4°C). The pellet was resuspended very gently in 50 ml ice-cold 0,1 M $CaCl_2$. Then the cells were kept on ice for 30 min before being pelleted again, the way described above. The cells were resuspended in 4 ml ice-cold 0,1 M $CaCl_2$ and stored on ice O/N (the longer, the better). Glycerol was added to a final concentration of 20%. The competent cells were aliquoted in 1.5 ml reaction tubes, 200µl each, and immediately frozen in liquid nitrogen. The competent cells can be stored at -80°C for a few months.

9.5.5 Plasmid DNA purification

Single white colonies from the O/N incubated culture plates were inoculated in 5 ml LB medium containing 100 µg/ml Ampicillin and incubated O/N at 37°C while shaking.

Next day 2 ml cells were centrifuged for 5 min at full speed. Plasmid DNA was isolated using Wizard® *Plus* SV Minipreps DNA Purification System (Promega Cat No:A1340).

9.5.6 Sequencing of the circularized fragments

The purified plasmid DNA was send to sequencing to the company 'VBC-Biotech'.

9.6 Northern hybridization

9.6.1 Probe labeling

9.6.1.1 Hot transcription

For labeling I used the MEGAshortscript kit from Ambion.

The T7 promoter (see Table PRIMER) sequence was synthesized together with two Hfq specific primers (F^{T7} and R^{T7}, see Table PRIMER) and used to amplify fragments: T7-R^{T7}/F^{T7} specific for sense strand and T7-F^{T7}/R^{T7} specific for antisense strand. 200nM resulting PCR products were used as DNA templates for the T7 transcription.

(The α-^{32}P labeled UTP and GTP was ordered from the company Hartmann Analytic)

For the hot transcription I used 7,5 pmole (45µCi) of α-^{32}P-UTP and α-^{32}P-GTP each (both 6000Ci/mmol) with 75 pmole cold UTP/GTP. The nonlabeled nucleotides (dATP and dCTP) were used in non limiting concentrations (150nmole each) which means that every 10th nucleotide should be labeled.

The reaction was performed following the manufacturer's instruction.

For recovery of the hot transcribed RNA probes 12% denaturing PAGE was used.

40% acrylamide/bis-acrylamide (19:1), was dissolved in 1x TBE (final conc.) buffer in the presence of 8 M Urea (final conc.).

10µl 10% APS (Ammonium persulfate) and 1µl TEMED (N,N,N',N'-Tetramethylethylenediamine) per 1 ml gel solution were added and the gel was immediately poured between two glass plates. After solidifying, the samples were mixed with a loading dye, loaded and the gel was run at 25 W until the Bromophenol Blue tracking dye reached the bottom of the gel.

After the gel run, the apparatus was disassembled and a phospho-imager screen was adjusted to the gel. After 10min exposure, the screen was scanned and printed to a DIN-A4 paper. The paper was adjusted behind the gel, so one could see the location of the RNA probe

For recovery of the RNA fragments, I cut out the RNA probe band, incubated it in 600µl TE Buffer in a 1.5 ml tube over night while shaking. Next day I centrifuged the tube for 5 min at

13.000g, and precipitated the supernatant with 0.3 M NaOAc, and 2.5 Vol. $EtOH_{abs.}$ Subsequently the labeled and purified LNA probe in X µl water was dissolved and with Geiger counter checked.

9.6.1.2 Design and γ labeling of the DNA/LNA probes

LNA™s (locked nucleic acids) contain a methylene bridge between the 2nd and 4th carbon atoms of their backbone. The molecule is locked in its 3'-endo structural conformation which enhances its binding specificity and thermal stability. Usually, the LNA™ probes are designed so that every 3rd nucleotide is LNA™, the rest is DNA.

For the design of the LNA™ probes I again used 'Oligo' software but this time I searched the whole genomic sequence of the E.coli K12. The probes were generated that their false priming sites among the genome were insignificant and their secondary structures excluded internal looping or homodimerization. The LNA probe sequences are displayed in Table PRIMER.

7.5 pmole LNA oligo was labeled with 8.33 pmole (50µCi) γ-^{32}P-ATP (6000Ci/mmol) in the presence of 1x PNK Buffer (NEB), 2u PNK enzyme (NEB) and 20 mM DTT.
The reaction was mixed and incubated at 37°C for 45 min then the PNK enzyme was heat inactivated (98°C for 5 min) in a thermal cycler with heated lid.

To remove the non-incorporated nucleotides I used G-25 Sephadex® columns (MicroSpin by GE Healthcare). The reaction was filled up to an end volume of 50µl and purified according to the manufacturer‘s manual.
After purification, the activity of the probe was compared to the activity of the unincorporated nucleotides, using a Geiger counter to verify the success of the labeling reaction.

9.6.2 Northern blot (Modified form Sambrook et al., 1989)

9.6.2.1 Formaldehyde Agarose Gel

For 200 ml denaturing agarose gel:

Weigh 2.6 g MP agarose and melt in 147 ml sterile distilled water.

Cool the melted agarose solution to about 60 °C and add successively 20 ml 10x MOPS gel buffer and 33 ml 37 % formaldehyde. Mix well. Pour the agarose onto a gel bed under a fume hood.

10x MOPS gel-running buffer:

Dissolve 200 mM MOPS and 50 mM NaOAc in DEPC treated H_2O, add 10mM EDTA and adjust pH with NaOH to 7.0 and fill up with H_2O to the required volume.

Sample loading dye:

55% deionized formamide, 0.015% SDS, 0.015% Bromophenol Blue, 0.015% Xylene Cyanol, 0.3 mM EDTA, 26.8% formaldehyde, 1.5x MOPS buffer, 50 µg/ml Ethidium Bromide; add at least equivolume of sample buffer to the sample, heat 15 min at 75°C than chill immediately on ice.

RNA Formaldehyde Agarose Gel Electrophoresis:

The gel and the samples were prepared as described above, and 30 - 80 µg total RNA from different expression stages, dissolved in loading buffer was loaded per lane. Additionally an RNA ladder was applied. The gel was run in 1x MOPS Buffer at 20W until the Bromophenol Blue dye reached the bottom of the gel. Then a gel picture was taken under UV transilluminator.

9.6.2.2 Capillary transfer:

For the transfer of the separated RNA to the membrane, I used a Capillary transfer blotter.

The gel was equilibrated in 2x SSC for 5-10 min. The apparatus was assembled as follows (from bottom to top): 2 layers of Whatman paper - gel - NYTRAN SPC (Whatman, Cat No: 340-260216294) Membrane - 2 layers of Whatman paper - paper towels - lid.

The reservoir of the apparatus was filled with 10x SSC and the RNA was blotted O/N.

Next day the membrane was crosslinked with a Stratagene UV-crosslinker (auto-crosslink program) and dried.

9.6.3 Hybridization of the membrane with LNA/DNA probes

The membrane was pre-hybridized for at least 1h in 15 ml of the following mix at 70°C:
5x SSC, 7% SDS, 1x Dennhardt‘s solution (0.02% Ficoll 400, 0.02% polyvinylpyrrolidone, 0.02% bovine serum albumin), 20 mM sodium phosphate buffer (Na_2HPO_4 + NaH_2PO_4), 300 µg denatured salmon sperm in 30 ml hybridization mix (10 mg/ml).
Then the entire amount of the labeled probe was added to 10 ml of fresh hybridization mix and the membrane was hybridized over night at 70°C -75°C (LNA probes can be hybridized up to 80°C) in a hybridization oven while rotating.
Next day the hybridization buffer was discarded, the membrane was washed once in 5x SSC plus 5% SDS buffer for 15 min at 70°C and then washed again for 5 min in 1x SSC plus 1% SDS at 95°C. The membrane was regularly checked with the Geiger Counter to get a rough idea of the signal to background ratio. If the signal at the presumptive spot was higher than the background signal, the membrane was immediately wrapped in Saran wrap and exposed to a Phosphor screen for: 4h (pre-check). If the background signal was still too high after 4h exposure (ribosomal RNA bands stronger than the signal), than the last washing step (1x SSC, 1% SDS at 95°C) was repeated. If the background signal was acceptable the exposition was continued O/N maximum for 1-2 weeks. The phosphor screens were scanned with the Phosphoimager (STORM).

10. APPENDIX

10.1 Table PRIMER

Primer name	Sequence 5‘-NNNNNN-3‘	Tm
	Primers for primer walking	
	Hfq b.s. specific primers	
F^{Hfq}	CCACCGACCAGATAAATCTTCACGCC	64.3
R^{Hfq}	CGCCTTCTTCGTTGCACTGTC	57.2
	flanking primers	
set $F^{fl}2$		
$F^{fl}2$	TCTCGCCGTTATCGTCCTGGGCC	66.6
$F^{fl}2$-1	CGCCCAGACCGTTGTACGGG	61.9
$F^{fl}2$-2	GGATCTTCGCCAAAAGCGG	57.0
$F^{fl}2$-3	CCGAGGTGGGCATAACGCGCAGC	69.6
$F^{fl}2$-4	GGGTCATCTCGCGCATCAACGCC	68.0
$F^{fl}2$-5	CCATACCCGTTCAGGCGTCAGG	62.2
$F^{fl}2$-6	GGAAGAACACCTGTGGATTGCGGG	64.2
set $F^{fl}3$		
$F^{fl}3$	CCAAACAGTGCGTCAATTTCCGGG	64.9
$F^{fl}3$-1	GACATCGACCTGCGGACTCAG	56.9
$F^{fl}3$-2	TTTACCGAGATCGTGGCATAAAGTTG	59.4
set $F^{fl}4$		
$F^{fl}4$	CAGCCACCAGTCTGGCTAAATCGC	62.7
$F^{fl}4$-1	ATTGGGAAGGTGTGGATGAGAT	53.8
$F^{fl}4$-2	GACGGCTTTTGTCGGCACTGACTG	63.6
set $F^{fl}5$		
$F^{fl}5$	CGTTGTTCCTTCCAGCTGGCTACC	61.8
$F^{fl}5$-1	TACTCATTCAGGCTTTGGGCAACG	61.1
$F^{fl}5$-2	AGTGTCGGATGCTTGGCGTTGTCA	64.4
set $R^{fl}2$		
$R^{fl}2$	GCAGAGCGACAGCGTGATCAACGGG	69.1
$R^{fl}2$-1	CGATAACACCTGGAATCAGCGC	58.3
$R^{fl}2$-2	CAGGTCAAAACCCTGACCG	52.6
$R^{fl}2$-3	GCACTGAGCCAAGCCTGCGCTCCC	70.2
$R^{fl}2$-4	GTGAAAGACAGCTCTGGCCGTACCG	63.8
$R^{fl}2$-5	CCTTATTACAAACTGACGCCAACACC	58.9
$R^{fl}2$-6	CATTATCGCCTCGTGGGCACGG	64.9

Primer name	Sequence 5‘-NNNNNN-3‘	Tm
set $R^{fl}3$		
$R^{fl}3$	CGAACTGAATACCTGGGTCCGCAGCG	68.6
$R^{fl}3$-1	AATTACGCCTGATCGGATTAACTTTA	55.9
$R^{fl}3$-2	TGACGGGCGTAAAAGTTTGAAGC	59.2
set $R^{fl}4$		
$R^{fl}4$	CGATGAACGGCGAACTTAATGCG	62.3
$R^{fl}4$-1	GCCATGGCCTTTACTTATCGGTTATG	59.9
$R^{fl}4$-2	CCTCAACGGCACTGTGATTAACAATA	58.2
$R^{fl}4$-3	AGTTATCCGGCGTTTCGTAGTAAATA	55.9
$R^{fl}4$-4	GTCGAGGGCGATTTCAATTT	53.3
	Primers for circularisation:	
$F^{fl}2$-inv	GGCCCAGGACGATAACGGCGAGA	66.6
$R^{fl}2$-inv	CCCGTTGATCACGCTGTCGCTCTGC	69.1
$F^{fl}3$-inv	CCCGGAAATTGACGCACTGTTTGG	64.9
$R^{fl}3$-inv	CGCTGCGGACCCAGGTATTCAGTTCG	68.6
6S-F-inv	GGTATGAAATATCGGCTCAGGGG	57.6
6S-R-inv	GCATGCTCGGTCCGTCCGAGAAGCC	70.7
	Primers for northern blot hot transcription	
T7 promoter	CAACGTAATACGACTCACTATAg...	
T7-F^{T7}	CAACGTAATACGACTCACTATAGCCACCGACC	76.4
	AGATAAATCTTCACGCC	
T7-R^{T7}	CAACGTAATACGACTCACTATAGCGCCTTCTT	74.4
	CGTTGCACTGTC	
F^{T7}	GCCACCGACCAGATAAATCTTCACGCC	
R^{T7}	GCGCCTTCTTCGTTGCACTGTC	
	LNA probes	
as-cca	AT[+G]AC[+C]CA[+T]GC[+G]GG[+T]GA[+A]CT	82
cca	AG[+T]TC[+A]CC[+C]GC[+A]TG[+G]GT[+C]AT	79
as-htrg	GC[+G]CA[+C]CG[+C]AG[+A]AA[+T]GC[+A]GC	84
htrg	GC[+T]GC[+A]TT[+T]CT[+G]CG[+G]TG[+C]GC	81
[+N] is LNA		

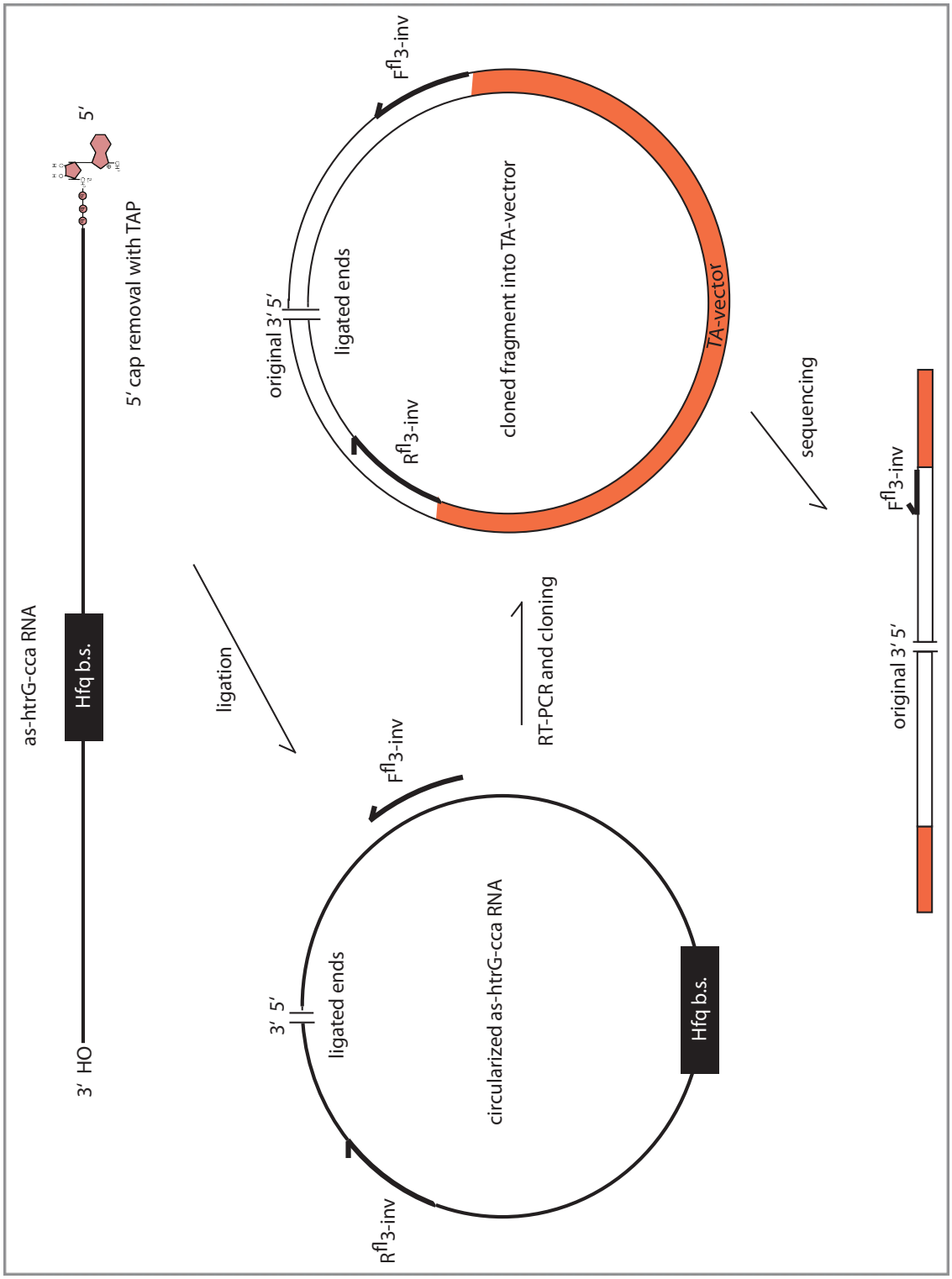


Fig S1: Circularization

The antisense transcript is drawn as a black line, containing the Hfq binding sequence (Hfq b.s.), a 5' cap and a 3' OH group. First, the 5' cap is removed with the enzyme Tobacco Acidic Phosphatase (TAP). Next step is the head to tail ligation of the total RNA. The ligated ends are marked with 5' and 3'. An RT-PCR is then performed with the two inverted flanking primers $F^{fl}3$-inv and $R^{fl}3$-inv. The RT-PCR outcome is then cloned into a Topo TA vector and transformed into BHLa competent cells. The isolated plasmid (Miniprep) is then sent to sequencing. The result is the head to tail ligated sequence of as-cca-htrG.

10.2 Circularized fragments (raw data):

```
3'-5'                          ... ligated ends
CCAAACAGTGCGTCAATTTCCGGG       ... Ffl3-inv primer
...CCGGCCGCCATGGCCGCGGGA       ... cloning vector fragment
```

PH1
GTTATTGTCATGGAAAGTGGAACGATAGTAGTGGCATCAGTGTCGCTACGCAAAGCATTCGACATCAA
TCGGAATCCTCTGTGTCAGGTCTTGCCACTGCTTCAAACTTTTACGCCCGTCAAATTGCCCGTTGCAT
GGCAATTTGGCGCAAAATAC3'-5'GAAACGGACATCGACCTGCGGACTCAGCATCGCCGCCATTGAG
AGCGTCATTAAGGTATGAATACCCGTATCGATTTCCGGATGCCA

PH7
GTTATTGTCATGGAAAGTGGAACGATAGTAGTGGCATCAGTGTCGCTACGCAAAGCATTCGACATCAA
TCGGAATCCTCTGTGTCAGGTCTTGCCACTGCTTCAAACTTTTACGCCCGTCAAATTGCCCGTTGCAT
GGCAATTTGGCGCAAAATAC3'-5'TCAGCATCGCCGCCATTGAGAGCGTCATTAAGGTATGAATACC
CGTATCGATTTCCGGATGCCACTTGGCAGGGGCCGGAACGC

PH5
GTTATTGTCATGGAAAGTGGAACGATAGTAGTGGCATCAGTGTCGCTACGCAAAGCATTCGACATCAA
TCGGAATCCTCTGTGTCAGGTCTTGCCACTGCTTCAAACTTTTACGCCCGTCAAATTGCCCGTTGCAT
GGCAATTTGGCGCAAAATAC3'-5'GCGCAGACGCTGGCATAGTTGTTCCACTAACTTAACACCCGCC
GGGCCATGACCATGATGACGCGGCCAGAGTTCTGGCGGCGTCA

PH4
GTTATTGTCATGGAAAGTGGAACGATAGTAGTGGCATCAGTGTCGCTACGCAAAGCATTCGACATCAA
TCGGAATCCTCTGTGTCAGGTCTTGCCACTGCTTCAAACTTTTACGCCCGTCAAATTGCCCGTTGCAT
GGCAATTTGGCGCAAAATAC3'-5'TAACTTAACACCCGCCGGGCCATGACCATGATGACGCGGCCAG
AGTTCTGGCGGCGTCAGCCCTTTACCGAGATCGTGGCATAA

PH1
GCGACGTCGCATGCTCCCGGCCGCCATGGCCGCGGGATTCGCTGCGGACCCAGGTATTCAGTTCGTCG
GAAACATAGCGCGTTTCTTCAGCGTGTGAGACGGCAGTCGCGCTAAGTGCGAGTAAAGTTAATCCGAT
CAGGCGTAATTTTGGCATCAGGCTGTCGTTATTGTCATGGAAAGT3'-5'CAGCATCGCCGCCATTGA
GAGCGTCATTAAGGTATGAATACCCGTATCGATTTCCGGATGCCACTTGGCAGGGGCCGGAACGCCAA
ACAGTGCGTCAATTTCCGGGAA

PH3
CGACGTCGCATGCTCCCGGCCGCCATGGCCGCGGGATTCGCTGCGGACCCAGGTATTCAGTTCGTCGG
AAACATAGCGCGTTTCTTCAGCGTGTGAGACGGCAGTCGCGCTAAGTGCGAGTAAAGTTAATCCGATC
AGGCGTAATTTTGGCATCAGGCTGTCGTTATTGTCATGGAAAA3'-5'GAACGCCAAACAGTGCGTCA
ATTTCCGGGAA

PH6
CCGACGTCGCATGCTCCCGGCCGCCATGGCCGCGGGATTCGCTGCGGACCCAGGTATTCAGTTCGTCG
GAAACATAGCGCGTTTCTTCAGCGTGTGAGACGGCAGTCGCGCTAAGTGCGAGTAAAGTTAATCCGAT

CAGGCGTAATTTTGGCATCAGGCTGTCGTTATTGTCATGGAAAGT3'-5'ATAAAGTTGCGAAACGGA
CATCGACCTGCGGACTCAGCATCGCCGCCATTGAGAGCGTCATTAAGGTATGAATACCCGTATCGATT
TCCGGATGCCACTTGGCAGGGGCCGGAACGCCAAACAGTGCGTCAATTTCCGGGAA

PH8
GGGCCGACGTCGCATGCTCCCGGCCGCCATGGCCGCGGGATTCGCTGCGGACCCAGGTATTCAGTTCG
TCGGAAACATAGCGCGTTTCTTCAGCGTGTGAGACGGCAGTCGCGCTAAGTGCGAGTAAAGTTAATCC
GATCAGGCGTAATTTTGGCATCAGGCTGTCGTTATTGTCATGGAAAGT3'-5'ACCCGCCGGGCCATG
ACCATGATGACGCGGCCAGAGTTCTGGCGGCGTCAGCCCTTTACCGAGATCGTGGCATAAAGTTGCGA
AACGGACATCGACCTGCGGACTCAGCATCGCCGCCATTGAGAGCGTCATTAAGGTATGAATACCCGTA
TCGATTTCCGGATGCCACTTGGCAGGGGCCGGAACGCCAAACAGTGCGTCAATTTCCGGGAA

PH9
ATGCTCCGGCCGCCATGGCCGCGGGATTCGCTGCGGACCCAGGTATTCAGTTCGTCGGACACATAGCG
CGTTTCTTCAGCGTGTGAGACGGCAGTCGCGCTAAGTGCGAGTAAAGTTAATCCGATCAGGCGTAATT
TTGGCATCAGGCTGTCGTTATTGTCATGGAAAGT3'-5'ATAGTTGTTCCACTAACTTAACACCCGCC
GGGCCATGACCATGATGACGCGGCCAGAGTTCTGGCGGCGTCAGCCCTTTACCGAGATCGTGGCATAA
AGTTGCGAAACGGACATCGACCTGCGGACTCAGCATCGCCGCCATTGAGAGCGTCATTAAGGTATGAA
TACCCGTATCGATTTCCAGATGCCACTTGGCAGGGGCCGGAACGCCAAACAGTGCGT

PH10
CCGACGTCGCATGCTCCCGGCCGCCATGGCCGCGGGATTCGCTGCGGACCCAGGTATTCAGTTCGTCG
GAAACATAGCGCGTTTCTTCAGCGTGTGAGACGGCAGTCGCGCTAAGTGCGAGTAAAGTTAATCCGAT
CAGGCGTAATTTTGGCATCAGGCTGTCGTTATTGTCATGGAAAGT3'-5'ATGACCATGATGACGCGG
CCAGAGTTCTGGCGGCGTCAGCCCTTTACCGAGATCGTGGCATAAAGTTGCGAAACGGACATCGACCT
GCGGACTCAGCATCGCCGCCATTGAGAGCGTCATTAAGGTATGAATACCCGTATCGATTTCCGGATGC
CACTTGGCAGGGGCCGGAACGCCAAACAGTGCGTCAATTTCCGGGAA

PH11
ACGTCGCATGCTCCCGGCCGCCATGGCCGCGGGATTCGCTGCGGACCCAGGTATTCAGTTCGTCGGAA
ACATAGCGCGTTTCTTCAGCGTGTGAGACGGCAGTCGCGCTAAGTGCGAGTAAAGTTAATCCGATCAG
GCGTAATTTTGGCATCAGGCTGTCGTTATTGTCATGGAAAGT3'-5'ACGCCAAACAGTGCGTCAATT
TCCGGGAA

PH12
GACGTCGCATGCTCCCGGCCGCCATGGCCGCGGGATTCGCTGCGGACCCAGGTATTCAGTTCGTCGGA
AACATAGCGCGTTTCTTCAGCGTGTGAGACGGCAGTCGCGCTAAGTGCGAGTAAAGTTAATCCGATCA
GGCGTAATTTTGGCATCAGGCTGTCGTTATTGTCATGGAAAGT3'-5'TAAGGTATGAATACCCGTAT
CGATTTCCGGATGCCACTTGGCAGGGGCCGGAACGCCAAACAGTGCGTCAATTTCCGGGAA

PH13
GGGCGACGTCGCATGCTCCCGGCCGCCATGGCCGCGGGATTCGCTGCGGACCCAGGTATTCAGTTCGT
CGGAAACATAGCGCGTTTCTTCAGCGTGTGAGACGGCAGTCGCGCTAAGTGCGAGTAAAGTTAATCCG
ATCAGGCGTAATTTTGGCATCAGGCTGTCGTTATTGTCATGGAAAGT3'-5'TTGGCAGGGGCCGGAA
CGCCAAACAGTGCGTCAATTTCCGGGAA

11. REFERENCE

ALLEN TA, Von Kaenel S, Goodrich JA, Kugel JF; The SINE-encoded mouse B2 RNA represses mRNA transcription in response to heat shock. Nat Struct Mol Biol (2004) vol. 11 (9) pp. 816-21

ARIGO JT, Eyler DE, Carroll KL, Corden JL; Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3. Molecular Cell (2006) vol. 23 (6) pp. 841-51

BARTEL PD and Szostak J; Isolation of new ribozymes from a large pool of random sequences. Science (1993) vol. 261 (5127) pp. 1411-8

BRANTL S; Antisense RNAs in plasmids: control of replication and maintenance. Plasmid (2002) vol. 48 (3) pp. 165-73

BRANTL S; Regulatory mechanisms employed by cis-encoded antisense RNAs. Current Opinion in Microbiology (2007) vol. 10 (2) pp. 102-109

BRENNAN RG and Link TM; Hfq structure, function and ligand binding. Current Opinion in Microbiology (2007) vol. 10 (2) pp. 125-33

BRITTEN RJ and Davidson EH; Gene regulation for higher cells: a theory. Science (1969) vol. 165 (891) pp. 349-57

CECH TR; at nobleprize.org, 1989

CRICK F; Central Dogma of Molecular Biology. Nature (1970) pp. 3

CUDNY H, Lupski JR, Godson GN, Deutscher MP; Cloning, sequencing, and species relatedness of the Escherichia coli cca gene encoding the enzyme tRNA nucleotidyltransferase. J Biol Chem (1986) vol. 261 (14) pp. 6444-9

DARTIGALONGUE C, Missiakas D, Raina S; Characterization of the Escherichia coli sigma E regulon. J Biol Chem (2001) vol. 276 (24) pp. 20866-75

DAVID L, Huber W, Granovskaia M, Toedling J, Palm CJ, Bofkin L, Jones T, Davis RW, Steinmetz LM; A high-resolution map of transcription in the yeast genome. Proc Natl Acad Sci USA (2006) vol. 103 (14) pp. 5320-5

DOUDNA JA and Szostak JW; RNA-catalysed synthesis of complementary-strand RNA. Nature (1989) vol. 339 (6225) pp. 519-22

EKLAND EH and Bartel PD; RNA-catalysed RNA polymerization using nucleoside triphosphates. Nature (1996) vol. 382 (6589) pp. 373-6

ESPINOZA CA, Allen TA, Hieb AR, Kugel JF, Goodrich JA; B2 RNA binds directly to RNA polymerase II to repress transcript synthesis. Nat Struct Mol Biol (2004) vol. 11 (9) pp. 822-9

EVANS MD and Cooke MS; Factors contributing to the outcome of oxidative damage to nucleic acids. Bioessays (2004) vol. 26 (5) pp. 533-42

FIRE A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC; Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature (1998) vol. 391 (6669) pp. 806-11

GILBERT W; The RNA World. Nature (1986) vol. 319 (6055) pp. 618

GONG X, Tao R, Li Z; Quantification of RNA damage by reverse transcription polymerase chain reactions. Anal Biochem (2006) vol. 357 (1) pp. 58-67

GUDIPATI RK, Villa T, Boulay J, Libri D; Phosphorylation of the RNA polymerase II C-terminal domain dictates transcription termination choice. Nat Struct Mol Biol (2008) vol. 15 (8) pp. 786-94

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) vol. 35 (3 Pt 2) pp. 849-57

HADDAD F, Qin AX, Giger JM, Guo H, Baldwin KM; Potential pitfalls in the accuracy of analysis of natural sense-antisense RNA pairs by reverse transcription-PCR. BMC Biotechnol (2007) vol. 7 pp. 21

HAJNSDORF E and Régnier P; Host factor Hfq of Escherichia coli stimulates elongation of poly(A) tails by poly(A) polymerase I. Proc Natl Acad Sci USA (2000) vol. 97 (4) pp. 1501-5

HERSHEY AD and Chase M; Independent functions of viral protein and nucleic acid in growth of bacteriophage. J Gen Physiol (1952) vol. 36 (1) pp. 39-56

HOFACKER IL, SCHUSTER P, Fontana W, Stadler PF; From sequences to shapes and back: a case study in RNA secondary structures. Proc Biol Sci (1994) vol. 255 (1344) pp. 279-84

KRUGER K, Grabowski PJ, Zaug AJ, Sands J, Gottschling DE, Cech TR; Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. Cell (1982) vol. 31 (1) pp. 147-57

LANDERS TA, Blumenthal T, Weber K; Function and structure in ribonucleic acid phage Q beta ribonucleic acid replicase. The roles of the different subunits in transcription of synthetic templates. J Biol Chem (1974) vol. 249 (18) pp. 5801-8

LEASE RA and Belfort M; A trans-acting RNA as a control switch in Escherichia coli: DsrA modulates function by forming alternative structures. Proc Natl Acad Sci USA (2000) vol. 97 (18) pp. 9919-24

LORENZ C, von Pelchrzim F, Schroeder R; Genomic systematic evolution of ligands by exponential enrichment (Genomic SELEX) for the identification of protein-binding RNAs independent of their expression levels. Nat Protoc (2006) vol. 1 (5) pp. 2204-12

LYKKE-ANDERSEN and Jensen TH; Overlapping pathways dictate termination of RNA polymerase II transcription. Biochimie (2007) vol. 89 (10) pp. 1177-82

MAJDALANI N, Cunning C, Sledjeski D, Elliott T, Gottesman S; DsrA RNA regulates translation of RpoS message by an anti-antisense mechanism, independent of its action as an antisilencer of transcription. Proc Natl Acad Sci USA (1998) vol. 95 (21) pp. 12462-7

MOORE PB und Steitz TA; After the ribosome structures: how does peptidyl transferase work?. RNA (2003) vol. 9 (2) pp. 155-9

NISSEN P, Hansen J, Ban N, Moore PB, Steitz TA; The structural basis of ribosome activity in peptide bond synthesis. Science (2000) vol. 289 (5481) pp. 920-30

NOLLER HF, Hoffarth V, Zimniak L; Unusual resistance of peptidyl transferase to protein extraction procedures. Science (1992) vol. 256 (5062) pp. 1416-9

ORGEL LE; The origin of life--a review of facts and speculations. Trends Biochem Sci (1998) vol. 23 (12) pp. 491-5

PANDEY RR, Mondal T, Mohammad F, Enroth S, Redrup L, Komorowski J, Nagano T, Mancini-Dinardo D, Kanduri C; Kcnq1ot1 Antisense Noncoding RNA Mediates Lineage Specific Transcriptional Silencing through Chromatin-Level Regulation. Molecular Cell (2008) vol. 32 (2) pp. 232-246

PETERSEN M; LNA: a versatile tool for therapeutics and genomics. Trends in Biotechnology (2003) vol. 21 (2) pp. 74-81

PRASANTH KV and Spector DL; Eukaryotic regulatory RNAs: an answer to the 'genome complexity' conundrum. Genes & Development (2007) vol. 21 (1) pp. 11-42

RAJKOWITSCH L and Schroeder R; Dissecting RNA chaperone activity. RNA (2007) vol. 13 (12) pp. 2053-60

REICHERT AS and Mörl M; Repair of tRNAs in metazoan mitochondria. Nucleic Acids Research (2000) vol. 28 (10) pp. 2043-8

ROBERTSON HD, Altman S, Smith JD; Purification and properties of a specific Escherichia coli ribonuclease which cleaves a tyrosine transfer ribonucleic acid presursor. J Biol Chem (1972) vol. 247 (16) pp. 5243-51

SCHEIBE M, Bonin S, Hajnsdorf E, Betat H, Mörl M; Hfq stimulates the activity of the CCA-adding enzyme. BMC Mol Biol (2007) vol. 8 pp. 92

SCHMIDT FJ; A novel function of Escherichia coli transfer RNA nucleotidyltransferase. Biosynthesis of the C-C-A sequence in a phage T4 transfer RNA precursor. J Biol Chem (1975) vol. 250 (21) pp. 8399-403

SCHÜRER H, Schiffer S, Marchfelder A, Mörl M; et al., This is the end: processing, editing and repair at the tRNA 3'-terminus. Biol Chem (2001) vol. 382 (8) pp. 1147-56

SHERRATT DJ, Dougan G, Saul M, Sunar B, Twigg A, Warren G; Structure and function of plasmid ColE1 and related plasmids. Contrib Microbiol Immunol (1979) vol. 6 pp. 100-10

SHOFIELD P and Williams KR; Purification and some properties of Escherichia coli tRNA nucleotidyltransferase. J Biol Chem (1977) vol. 252 (15) pp. 5584-8

SPÄTH B, Canino G, Marchfelder A; tRNase Z: the end is not in sight. Cell Mol Life Sci (2007) vol. 64 (18) pp. 2404-12

STORZ G, Opdyke JA, Zhang A; Controlling mRNA stability and translation with small, noncoding RNAs. Current Opinion in Microbiology (2004) vol. 7 (2) pp. 140-4

THAKUR N, Tiwari VK, Thomassin H, Pandey RR, Kanduri M, Göndör A, Grange T, Ohlsson R, Kanduri C; An antisense RNA regulates the bidirectional silencing property of the Kcnq1 imprinting control region. Mol Cell Biol (2004) vol. 24 (18) pp. 7855-62

TOMIZAWA J, Itoh T, Selzer G, Som T; Inhibition of ColE1 RNA primer formation by a plasmid-specified small RNA. Proc Natl Acad Sci USA (1981) vol. 78 (3) pp. 1421-5

TUCKER BJ, and Breaker RR; Riboswitches as versatile gene control elements. Curr Opin Struct Biol (2005) vol. 15 (3) pp. 342-8

TUREK C and MacDougal-Waugh; In vitro evolution of functional nucleic acids: high affinity RNA ligands of HIV-1 proteins. Gene (1993) vol. 137 (1) pp. 33-9

UMLAUF D, Fraser P, Nagano T; The role of long non-coding RNAs in chromatin structure and gene regulation: variations on a theme. Biol Chem (2008) vol. 389 (4) pp. 323-31

VANÁCOVÁ S, Wolf J, Martin G, Blank D, Dettwiler S, Friedlein A, Langen H, Keith G, Keller W; A new yeast poly(A) polymerase complex involved in RNA quality control. PLoS Biol (2005) vol. 3 (6) pp. e189

VOGEL J and Wagner EG; Target identification of small noncoding RNAs in bacteria. Current Opinion in Microbiology (2007) vol. 10 (3) pp. 262-70

WASSARMAN KM and Saecker RM; Synthesis-mediated release of a small RNA inhibitor of RNA polymerase. Science (2006) vol. 314 (5805) pp. 1601-3

WATSON JD and Crick F; Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature (1953) vol. 171 (4356) pp. 737-8

WUTZ A and Jaenisch R; A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. Molecular Cell (2000) vol. 5 (4) pp. 695-705

WYERS F, Rougemaille M, Badis G, Rousselle JC, Dufour ME, Boulay J, Régnault B, Devaux F, Namane A, Séraphin B, Libri D, Jacquier A; Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. Cell (2005) vol. 121 (5) pp. 725-37

XU Z, Gagneur J, Perocchi F, Clauder-Münster S, Camblong J, Guffanti E, Stutz F, Huber W, Steinmetz LM; Bidirectional promoters generate pervasive transcription in yeast. Nature (2009) vol. 457 (7232) pp. 1033-7

YOUNG B and Cech TR; Specificity for 3',5'-linked substrates in RNA-catalyzed RNA polymerization. J Mol Evol (1989) vol. 29 (6) pp. 480-5

ZHU L and Deutscher MP; tRNA nucleotidyltransferase is not essential for Escherichia coli viability. EMBO J (1987) vol. 6 (8) pp. 2473-7

ZWART R, Sleutels F, Wutz A, Schinkel AH, Barlow DP; Bidirectional action of the Igf2r imprint control element on upstream and downstream imprinted genes. Genes & Development (2001) vol. 15 (18) pp. 2361-6

Computational analysis:

PROMSCAN: STUDHOLME DJ and Dixon R. 2003. Domain architectures of sigma54-dependent transcriptional activators. J. Bacteriol. 185:1757-67

BDGP: REESE M, Lawrence Berkeley Laboratory, CA, U.S.A.

BPROM: SOFTBERRY, INC.; NY 10549, USA www.softberry.com

RNAfold Web Server: http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi

Books:

SAMBROOK J, **Fritsch, E.F and Maniatis, T** (1989) In: Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 8.64.

Homepages:

-REGULON DATABASE: http://regulondb.ccg.unam.mx

-EXIQON: www.exiqon.com

-htrG at NCBI:
http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene&cmd=Retrieve&dopt=full_report&list_uids=947555

-Northern Blot detection rate (Ambion):
http://www.ambion.com/techlib/append/mRNA_detection_chart.html

12. CV

Name:	Peter HAJDUSITS
private address:	Penzingerstraße 39/7, A-1140 Vienna, Austria
Professional address:	Max F. Perutz Laboratories, Department of Biochemistry, University of Vienna, Dr. Bohrgasse 9, A-1030 Vienna, Austria
Date of Birth:	19th April 1982
Place of Birth:	Szeged, Hungary
Nationality:	Austria (since 2000)
Email:	peter.hajdusits@univie.ac.at

Education

since 2002	Studies of Molecular Biology at the University of Vienna
2001-2002	Austrian Army, military service
2001	Matura, Abitur school leaving examination (equal to A-levels)
1993-2001	Gymnasium (equal to High School) BG BRG BORG Eisenstadt Kurzwiese 1
1991-1993	Elementary School Vienna, Austria
1989-1991	Elementary School Szeged, Hungary

Additional Work Experience

September 2007 - November 2009	Diploma Thesis Max F. Perutz Laboratories, University of Vienna Prof. Dr. Renée Schroeder *"Characterization of the Hfq binding cis antisense RNA as-htrG-cca"*
August - September 2006	internship at the Institute of Molecular Pathology Vienna (IMP) Dr. Barry Dickson *"Neural circuit development in Drosophila Melanogaster"*
March - July 2006	research assistant at the Institute of Molecular Pathology Vienna (IMP) Dr. Tim Clausen *"Overexpression and purification of the intramembrane protease YaeL (RseP)"*
February - March 2006	internship at the Max Perutz Laboratories (University of Vienna) Prof. Dr. Renée Schroeder *"Analysis of selected RNA polymerase II binding snRNAs"*
August-September 2005	internship at the Biological Research Centre Szeged (BRC), Hungary Dr. Ferenc Nagy *"Molecular mechanisms of light and clock-signaling pathways in Arabidopsis Thaliana"*
July-August 2004	research assistant at ARC Seibersdorf Research GmbH Mag. Silvia Fluch *"Isolation of total RNA from Populus Nigra, fluorescent labeling and hybridization with Microarray-Chips"*
July-August 2003	research assistant at ARC Seibersdorf Research GmbH Mag. Silvia Fluch *"Isolation of random EST clones, purification, isolation and sequencing"*
Summer 1999, 2000, 2002	internships at the Austrian Research Center ARC Seibersdorf Research GmbH Dr. Maria Berényi *"Training of laboratory skills and standard techniques in molecular biology"*

Scientific communications

***Talks*:**

January 2008 — *"Characterization of novel Hfq binding antisense RNAs"* progress report at the BACRNA meeting, Januray 15th - Januray 17th 2008, Ein Gedi, Israel

Posters:

July, August 2008 — *"Cis-antisense RNAs (casRNAs) With High Affinity To Hfq"* Peter HAJDUSITS, Christina Lorenz, Ursula Schoeberl, Oliver Mayer, Tanja Gesell, Christina Waldsich, Arndt von Haeseler, Renée Schroeder
13th annual meeting of the RNA Society, July 28th - August 3rd 2008, Berlin, Germany

13. LEBENSLAUF

Name:	Peter HAJDUSITS
Private Adresse:	Penzingerstraße 39/7, A-1140 Wien, Österreich
Institut Adresse:	Max F. Perutz Laboratories, Department of Biochemistry, Universität Wien, Dr. Bohrgasse 9, A-1030 Wien, Österreich
Geburtstag:	19. April 1982
Geburtsort:	Szeged, Ungarn
Nationalität:	Österreich (seit 2000)
Email:	peter.hajdusits@univie.ac.at

Ausbildung

2002 - gegenwärtig	Diplomstudium Molekulare Biologie an der Universität Wien
2001 - 2002	Österreichisches Bundesheer, Präsenzdienst
2001	Matura, Abitur
1993 - 2001	Gymnasium BG, BRG, BORG Eisenstadt Kurzwiese 1
1991 - 1993	Volksschule 1180 Wien
1989 - 1991	Volksschule Szeged, Ungarn

Zusätzliche Ausbildungsbegleitende Tätigkeiten

September 2007 - November 2009	Diplomarbeit am Max F. Perutz Laboratories der Universität Wien Prof. Dr. Renée Schroeder *"Characterization of the Hfq binding cis antisense RNA as-htrG-cca"*
August - September 2006	Praktikum am Institute of Molecular Pathology, Wien (IMP) Dr. Barry Dickson "*Neural circuit development in Drosophila Melanogaster"*
März - Juli 2006	Wissenschaftlicher Mitarbeiter am Institute of Molecular Pathology, Wien (IMP) Dr. Tim Clausen "*Overexpression and purification of the intramembrane protease YaeL (RseP) "*
Februar - März 2006	Praktikum am Max F. Perutz Laboratories der Universität Wien Prof. Dr. Renée Schroeder "*Analysis of selected RNA polymerase II binding snRNAs"*
August - September 2005	Praktikum am Biological Research Center Szeged (BRC) Dr. Ferenc Nagy "*Molecular mechanisms of light and clock-signaling pathways in Arabidopsis Thaliana"*
Juli - August 2004	Wissenschaftlicher Mitarbeiter am ARC Seibersdorf Research GmbH Mag. Silvia Fluch "*Isolation of total RNA from Populus Nigra, fluorescent labeling and hybridization with Microarray-Chips"*
Juli - August 2003	Wissenschaftlicher Mitarbeiter am ARC Seibersdorf Research GmbH Mag. Silvia Fluch "*Isolation of random EST clones, purification, isolation and sequencing"*
Sommer 1999, 2000, 2002	Praktikum am Austrian Research Centers Seibersdorf Research GmbH (ARCS) Dr. Maria Berényi *"Aneignung von gängigen Techniken in der Molekularbiologie"*

Wissenschaftliche Kommunikation

Reden:

Jänner 2008	*"Characterization of novel Hfq binding antisense RNAs"* Tätigkeitsbericht am BACRNA meeting, 15. - 17. Jänner 2008, Ein Gedi, Israel

Posters:

Juli, August 2008	*"Cis-antisense RNAs (casRNAs) With High Affinity To Hfq"* Peter HAJDUSITS,Christina Lorenz, Ursula Schoeberl, Oliver Mayer, Tanja Gesell, Christina Waldsich, Arndt von Haeseler, Renée Schroeder 13th annual meeting of the RNA Society, 28. Juli - 3. August 2008, Berlin, Germany

14. ACKNOWLEDGEMENTS

I would like to thank Renée for her advise and the opportunity to learn a lot and to write my Diploma thesis in her lab.

Moreover, I would like to thank all the Schroeder group members (present and past) for their support and and all the beer we had together.

A very special "Thank You" goes to my love Susanne and my beloved family, my brother Bence, my mother Maria and my father Josef for their indefatigable support, for their strong backing and for being always there when I needed help.

Last but not least I'd like to thank Martin, Rita, Maike, Poldi, Joe and all my friends who helped me a lot in this stressful time.