Genomic SELEX for Hfq-binding RNAs identifies genomic aptamers predominantly in antisense transcripts

C. Lorenz[1], T. Gesell[2], B. Zimmermann[1], U. Schoeberl[1], I. Bilusic[1], L. Rajkowitsch[1], C. Waldsich[1], A. von Haeseler[2] and R. Schroeder[1,*]

[1]Department of Biochemistry and [2]Center for Integrative Bioinformatics in Vienna, Max F. Perutz Laboratories, Dr Bohr-Gasse 9/5, University of Vienna, Medical University of Vienna and University of Veterinary Medicine, Vienna, Austria

ABSTRACT

An unexpectedly high number of regulatory RNAs have been recently discovered that fine-tune the function of genes at all levels of expression. We employed Genomic SELEX, a method to identify protein-binding RNAs encoded in the genome, to search for further regulatory RNAs in *Escherichia coli*. We used the global regulator protein Hfq as bait, because it can interact with a large number of RNAs, promoting their interaction. The enriched SELEX pool was subjected to deep sequencing, and 8865 sequences were mapped to the *E. coli* genome. These short sequences represent genomic Hfq-aptamers and are part of potential regulatory elements within RNA molecules. The motif 5′-AAYAAYAA-3′ was enriched in the selected RNAs and confers low-nanomolar affinity to Hfq. The motif was confirmed to bind Hfq by DMS footprinting. The Hfq aptamers are 4-fold more frequent on the antisense strand of protein coding genes than on the sense strand. They were enriched opposite to translation start sites or opposite to intervening sequences between ORFs in operons. These results expand the repertoire of Hfq targets and also suggest that Hfq might regulate the expression of a large number of genes via interaction with *cis*-antisense RNAs.

INTRODUCTION

Systematic searches for functional RNAs in both prokaryotes and eukaryotes have revealed an astonishingly high number of regulatory RNAs in a variety of organisms. Computational predictions (1–4) and cDNA-based experimental strategies (5–8) have been successfully employed to identify over 100 small RNAs (sRNAs) in *Escherichia coli* and many other bacteria. Most of these sRNAs are transcribed from intergenic regions in response to environmental stress, act *in trans* and show only short and partial complementarity to their target mRNAs (9,10). Most sRNAs are regulatory molecules targeting mainly mRNAs, but also protein activities can be regulated (11). Regulation of gene expression by sRNAs occurs mainly by modulation of translation and/or stability of the target mRNAs (11). Many of the *E. coli* sRNAs require the regulatory protein Hfq for function (12,13).

The *E. coli* host factor Hfq was originally described as essential factor for the replication of the RNA phage Qß (14). *Escherichia coli hfq* mutants display a very broad phenotype, because Hfq is involved in regulating the expression of many regulatory genes (15). For example, Hfq regulates the expression of the *rpoS* gene, which codes for the stationary phase sigma factor σ^S (16); it controls the stability of several mRNAs (17,18) and small regulatory RNAs (sRNAs) (9,19); and it is involved in the riboregulation of mRNAs (20–22). In addition, Hfq was shown to act as a virulence factor in several bacterial

pathogens (23–25). Hfq is an Sm-like protein, which assembles into a ring-shaped homo-hexamer and stimulates RNA–RNA interactions (21,26–29). Hfq was shown, for example, to accelerate the association of the small regulatory RNA DsrA with its target mRNA *rpoS* (30–33). Because of Hfq's characteristics and its role in the modulation of small RNA function, it is a perfect candidate for the isolation of novel regulatory RNAs and it has already been used to pull down RNAs via co-immunoprecipitation (6,32,34). Hfq exerts many functions in the cell and is viewed as a general and pleiotropic regulator. However, the functions of Hfq are still elusive in many organisms. Hfq is not essential for growth, but the adaptation to changing environmental conditions is hampered in the absence of Hfq.

RNA-mediated regulation most often concerns downregulation or complete silencing of genetic elements. If Hfq is controlling, among many other things, the repression of genes, then repressed genes might be targets of Hfq. In order to assess further functions of Hfq, we developed an alternative and complementary approach to search for additional Hfq-binding RNAs, independent of their expression levels. Instead of relying on enriching transcripts from cell extracts, our SELEX-based approach detects regions of the genome that encode RNA domains binding Hfq with high affinity irrespective of their expression. SELEX has been used extensively in the 90's before whole genomes were sequenced, but today, with so many available genomes and with the development of massive sequencing techniques, Genomic SELEX has the potential to become a major alternative approach for the detection of functional and/or regulatory elements within RNAs, even when these are not transcribed. We have recently evaluated the SELEX procedure aiming at analysing the impact of SELEX-specific non-selective steps on the selected RNAs (35). In that study, we concluded that the SELEX-imposed requirements on the sequences are much weaker in comparison to the enrichment of genomic aptamers, when specific characteristics confer high-affinity interactions.

Using an *E. coli* genomic library and the global regulator protein Hfq as bait, we performed nine rounds of SELEX and the enriched pool was subjected to 454 sequencing. The vast majority of the identified Hfq aptamers map to the antisense strand opposite of translation initiation sites and in intervening sequences between ORFs in operons. *In silico* and chemical footprinting analyses revealed sequence motifs that are significantly enriched in antisense strands of the *E. coli* genome. We demonstrate that Genomic SELEX combined with pyrosequencing and rigorous analysis is a powerful approach to identify regulatory RNAs motifs termed 'genomic aptamers'. We propose that Hfq might regulate gene expression by targeting *cis*-antisense RNAs more frequently than hitherto assumed.

MATERIAL AND METHODS

Genomic SELEX

Via random priming, we constructed a representative PCR library of the *E. coli* genome containing overlapping sequences from 50 to 500 nt in length. Library fragments are flanked by primer annealing sites for amplification and are preceded by a T7 promoter for transcription of the library into RNA (36).

For selection of Hfq-binding RNAs, we incubated Hfq with the respective RNA pool for 30 min at room temperature using near-physiological buffer conditions (150 mM NaCl, 0.8 mM $MgCl_2$, 0.5 mM DTT and 50 mM Tris–HCl pH 7.5). Subsequently, we employed filter binding to separate bound from unbound species, recovered Hfq-binding RNAs via urea-mediated denaturation followed by phenol/chloroform extraction and amplified selected sequences via RT–PCR (35). For the next SELEX cycle, obtained PCR products were again *in vitro* transcribed into RNA.

Yeast three-hybrid system

We used the yeast three-hybrid system, as described by Bernstein *et al.* (37), to evaluate Hfq-RNA binding in a cellular environment. Therefore plasmids pIIIa-MS2-2-RNAX, expressing the hybrid MS2-RNA sequences, and pACT2-HFQ, expressing an Hfq-Gal4AD fusion protein, were transformed into the yeast strain YBZ-1. *HIS3* and *lacZ*, encoded by YBZ-1, served as reporter genes to monitor Hfq-RNA binding as well as to estimate the strength of the interaction. Hfq-independent reporter gene activation was assessed by eliminating pACT2-HFQ from YBZ-1, followed by a second β-gal assay.

E. coli strains and growth conditions

In this study, *E. coli* wild-type strains B and K12 MG1655 were used. Cells were grown at 37°C under the following conditions: exponential phase—growth in LB to OD_{600} 0.4; stationary phase—growth in LB 16–18 h; acidic stress—growth in LB to OD_{600} 0.4, addition of [2-(N-morpholino)ethanesulfonic acid] (MES) to a final concentration of 170 mM, growth for another 40 min; oxidative stress—growth in LB to OD_{600} 0.4, addition of H_2O_2 to a final concentration of 80 nM, growth for another 15 min; M9: growth in M9 minimal medium supplemented with 0.2% glucose to OD_{600} 0.4; hyperosmotic stress: growth in M9 minimal medium supplemented with 0.4% glycerol to OD_{600} 0.4, addition of NaCl to a final concentration of 300 mM, growth for another 20 min.

RNA isolation and RT–PCR

For total RNA isolation hot-phenol procedure was used as described by Jahn *et al.* (38) Alternatively, thaw-freeze protocol was used. Briefly, cells were harvested by centrifugation and resuspended in VD buffer (10 mM Tris/HCl, pH 7.3, 60 mM NH_4Cl, 6 mM β-mercaptoethanol, 2 mM MgAc) with the addition of 0.4 mg/ml of lysozyme and 50 U DNaseI (NEB). After incubation on ice for 30 min, three to four cycles of freezing and thawing followed. Lysates were cleared by centrifugation and RNA extracted by conventional phenol–chloroform extraction followed by ethanol precipitation. Total RNA was treated with DNaseI (NEB) as recommended by the manufacturer.

Expression analysis and primer walking were performed using One Step RT-PCR Kit (Qiagen). Manufacturer's instructions were followed except that 2–4 μg of total RNA was used. Strand specificity of reactions was provided by adding only reverse primer into the RT reaction and forward primer was added immediately prior to PCR. Absence of DNA contamination was confirmed for each reaction by omitting the addition of any primer into the RT step of the reaction. Primers for primer walking were designed to map down- and up-stream from the Hfq binding region in a stepwise manner giving products increasing in size. For reverse transcription primers as-cydD-R1 (5′-GGGTTATGCAGGATGGCC GG-3′), as-cydD-R2 (5′-CTGTTCGCTATTACTGTTG GATG-3′), as-cydD-R3 (5′-GCCAGCGAACAAGAAT TACAAG-3′) and as-cydD-R4 (5′-GCGTTTTTATCCT CCGGCATTC-3′). Primers used for subsequent amplification were as-cydD-F1 (5′-CACACTTAATTCCGCGT AACG-3′), as-cydD-F2 (5′-GAGCAGCGTCACAATTG CCAG-3′), as-cydD-F3 (5′-GGTAAAGATGATCGAGC GTATC-3′), as-cydD-F4 (5′-GCACCAAAAATGGTCA GCTCAG-3′) and as-cydD-F5 (5′-CCGTTAAGTCAGA GATACGTAC-3′).

RNA preparation for structural probing

Standard PCR was applied to produce templates from three different loci of *E. coli* K12 genomic DNA using the following primer sequences: ptsLfow1: 5′-CGTAAT ACGACTCACTATAGGTGAGCAGTGTCTTACGA AC-3′, ptsLrev1: 5′-CACGGTCAGGTCGCCACC-3′, ccafow1: 5′-CGTAATACGACTCACTATAGGCCACC CACCAGATAAATCTTC-3′, ccaLrev1: 5′-CGCAAAC GCAAAGATCGCTG-3′, iclRfow1: 5′-CGTAATACGA CTCACTATAGGATTGCATTAGCTAACAATAAAA ATG-3′, iclRrev1: 5′-TCTGCCGCGTTTCGCGGG-3′ (Thermo). Underlined nucleotides indicate the T7-promoter sequence. The PCR products were used for *in vitro* RNA transcription according to the user manual's guidelines of the RNA MaxX High Yield Kit (Stratagene). The RNAs were purified on an 8% denaturing (7 M Urea) PAGE in 1× TBE (89 mM Tris base, 89 mM Boric acid, 2 mM EDTA pH 8.0), eluted into 1× elution buffer (40 mM MOPS pH 6.0, 1 mM EDTA pH 8.0, 250 mM NaOAc pH 5.0), precipitated and resuspended in ME buffer (10 mM MOPS pH 6.0, 1 mM EDTA pH 8.0) to be stored at −20°C.

In vitro DMS modification

RNA structure determination was performed in the presence of increasing amounts of $MgCl_2$ in a total volume of 50 μl. The RNA (2 pmol) in 10× binding buffer (final concentration: 80 mM Cacodylic acid buffer pH 8.0, 150 mM NaCl) was denatured at 95°C for 1 min and after 2 min at room temperature $MgCl_2$ was added to a final concentration of 0, 1, 10 mM. The RNA was incubated at room temperature for 20 min to ensure correct folding before the treatment with dimethysulphate, (DMS, 21 mM final concentration). The modification reaction was left to proceed for 20 min at room temperature and was stopped by addition of β-mercaptoethanol

(57 mM β-ME final concentration). After precipitation, the sample was resuspended in 6 μl ddH₂O to be used for reverse transcription. For Hfq footprinting assays the standard protocol was slightly adapted, starting with an input of 4 pmol RNA in 10× binding buffer with DTT (0.5 mM DTT final concentration) but identical denaturing and folding procedures. Then Hfq (final concentration 400 nM or 1.2 μM Hfq₆) was added to the sample and incubated for 30 min at room temperature prior to DMS and -mercaptoethanol treatment as described above. The samples were subsequently purified by phenol extraction, ethanol precipitated and resuspended in 6 μl ddH₂O for primer extension. DMS (1 pmol) modified or untreated RNA was subjected to reverse transcription. 5′-end labelled oligonucleotides (^{32}P-γ-ATP by Hartmann Analytics; PCR reverse primers were used) were used for primer extension experiments described in ref. (39) and the samples were subsequently separated on 8% denaturing polyacrylamide gel (7 M urea) and detected using the Storm Phosphoimager (Molecular Dynamics) and Image Quant.

Electrophoretic mobility shift assays

In vitro transcribed and gel purified RNAs were 5′-end labelled with γ-^{32}P-ATP. Labelled RNA (0.01 nM) was incubated at room temperature for 30 min in 50 mM Tris–HCl pH 7.5, 150 mM NaCl, 8 mM $MgCl_2$ and 0.5 mM DTT with various amounts of Hfq₆ as indicated. RNA–protein complexes were separated on 5% native polyacrylamide gels over-night at 3 V/cm and 4°C. Gels were dried and visualized on a Phosphoimager and quantified using the software ImageQuant1.4. Primers used to clone and transcribe RNA were as for structural probing.

Quantitative RT–PCR

Reverse transcription was performed with 2 μg of total RNA isolated from exponential phase or acidic stress using Omniscript RT Kit, Qiagen. Briefly, RNA was mixed with 1 μM RT-primer, 0.5 mM dNTPs, 10 U RNasin, Promega, 3 U reverse transcriptase and 1× buffer in total volume of 20 μl. Reaction was incubated at 50°C for 30 min followed by enzyme inactivation at 95°C, 15 min. Absence of genomic DNA was confirmed by omitting addition of the enzyme in the RT-reaction. To test absence of unspecific cDNA synthesis, no primer control (NPC) was introduced in which addition of any primer in RT reaction was omitted.

Quantitative PCR was done with Mesa Green qPCR MasterMix Plus for SYBR, Eurogentec on an Abi Prism 7000 machine. Briefly, 1 μl of RT reaction was mixed with 0.1 μM primers and 1× master mix in total volume of 25 μl. All experiments were repeated twice from two independent RNA preparations. All reactions were performed in three technical replicates. RNA copy number was estimated from the standard curve created from series of corresponding PCR products of known starting concentration.

Reverse transcription of sense transcript was obtained using cca-R (5′-TGGCGCAACAGACGATTT-3′)

and htrG-R (5′-TAAGTTGTTTCAACGGGATCCA-3′) primers, while reverse transcription of antisense transcript was preformed using primers cca-F (5′-CGTCGCGATCT GACCATTAA-3′) and htrG-F (5′-CGGCGAGGAAGT GACCTTATT-3′). For subsequent amplification of cca and htrG amplicon primers cca-R and cca-F, and htrG-R and htrG-F were used, respectively. All primers were designed using Primer Express software.

Sequencing and computational analysis

The sequences from round 9 of Hfq selection and round 9 of a control neutral selection were tagged and sent to 454 sequencing among other pools as described in Zimmermann *et al.* (submitted). In total, 9991 and 5894 of these sequences were identified as deriving from round 9 of Hfq and control selection, respectively. Primer and T7 promoter sequences were removed. These sequences were then aligned to the *E. coli* K12 genome (GenBank accession NC_000913) using vmatch (40), taking only alignments with an *E*-value of at most $1e-10$, leaving 8865 Hfq sequences and 5853 control sequences.

All sequences were analysed for overlap and orientation to known features, based on all available annotations from GenBank (41) and EcoCyc (42). Clusters of reads were computed by grouping all overlapping sequences, including sequences linked by a third sequence, and not overlapping. Base-level enrichment for each type of feature (CDS sense, CDS antisense, etc.) was calculated by summing the number of aligned reads that overlapped the feature at each position of the feature. Start and stop codon positional enrichment was calculated by summing the number of reads overlapping each position up- and down-stream of every translation signal, and dividing the number of reads at each position from the Hfq sequences by the control sequences. The enrichment was visualized using R (43).

For minimum free energy predictions (MFE), we used Zuker's algorithm (44,45) and the partition function algorithm by McCaskill (46), which calculates base pair probabilities in the thermodynamic ensemble as implemented in version 1.6.5 of Vienna RNA Package RNAfold (47). For visualizations of the secondary structures several options of the Vienna RNA packages were helpful (http://www.tbi.univie.ac.at/~ivo/RNA).

RESULTS

Genomic SELEX for the isolation of RNAs with high affinity to Hfq

We constructed an *E. coli* genomic library (36,48), which we used to isolate and identify high-affinity Hfq-binding RNAs. This library was transcribed into RNA and used for several consecutive cycles of SELEX for enrichment of Hfq-binding RNAs via filter binding as described in the 'Materials and Methods' section (Figure 1A). We selected high affinity Hfq-binding RNAs over 9–10 iterative SELEX cycles, applying increasingly stringent conditions (Figure 1B). Then we cloned and sequenced a total number of 112

RNAs resulting from cycles 8 (minus competitor) and 9 (plus tRNA competitor).

To investigate whether *in vitro*-selected sequences bind Hfq in a cellular environment, we introduced RNAs resulting from SELEX cycle 8 into the yeast three hybrid system (37) and monitored Hfq-binding via expression of the reporter genes HIS3 and *lacZ* (Figure 1C). Both tolerance to 3-AT (3-aminotriazole is a competitive inhibitor of the HIS3 gene product) and speed and intensity of colour development in the *lacZ* assay are indicators of the strength of Hfq–RNA interactions. We picked and sequenced 58 clones, which showed dark blue staining. As all clones investigated were able to activate *lacZ* in an Hfq-dependent manner, we conclude that *in vitro* selected RNA sequences do bind to Hfq in a cellular environment.

Initial sequencing of the 170 clones derived from the different three pools enriched from Genomic SELEX and yeast three-hybrid revealed 108 unique RNAs overlapping in all three pools. Most surprising was the observation that the vast majority (85) mapped antisense to protein coding genes, only nine mapped to intergenic regions and 14 within mRNAs. These results clearly show that the sequences were not exhausted and most importantly, known Hfq-binding RNAs were not detected. We therefore performed deep sequencing using the 454 technique. Since some of the sequences were shared among the three pools, we chose only to sequence the SELEX pool of cycle 9.

After eliminating the primers and those reads that did not match to the *E. coli* genome unambiguously, 8865 reads were mapped to the *E. coli* genome. Each read was grouped into a cluster based on overlap with other reads. In Figure 1D, the bar demonstrates the number of reads contained in cluster regions of the genome. One thousand and five hundred and twenty-two individual clusters were obtained. Among these clusters, approximately half (775) contain only one read, which we take as an indication that the sequencing is still not exhausted. Table 1 lists those clusters that contained more than 50 reads. The chromosomal location and the function of the gene product are indicated. It is worth mentioning that most of the enriched aptamers are antisense to genes coding for proteins involved in the cell's interaction with the environment. The most highly enriched aptamers map to the universal stress protein gene (upsG), enterobacterial common antigen (ECA), transporters and membrane proteins.

Analysis of deep sequenced SELEX pool 9

We evaluated the selection outcome by comparing the selected sequences with known Hfq-binding RNAs, both sRNAs and mRNAs and their targets. We obtained three known sRNAs, Tpk70, OmrA/B and GcvB and the targets of two of these sRNAs, *ompT* mRNA targeted by OmrA/B and *argT* and *dppA* mRNAs targeted by GcvB (49,50). We also obtained reads mapping to two tRNAs, *hisR* and *leuV*. Hfq had been reported to bind tRNAs (51). The following known Hfq-binding mRNAs were selected, *thrL*, *yjbJ*, *glpFK* and
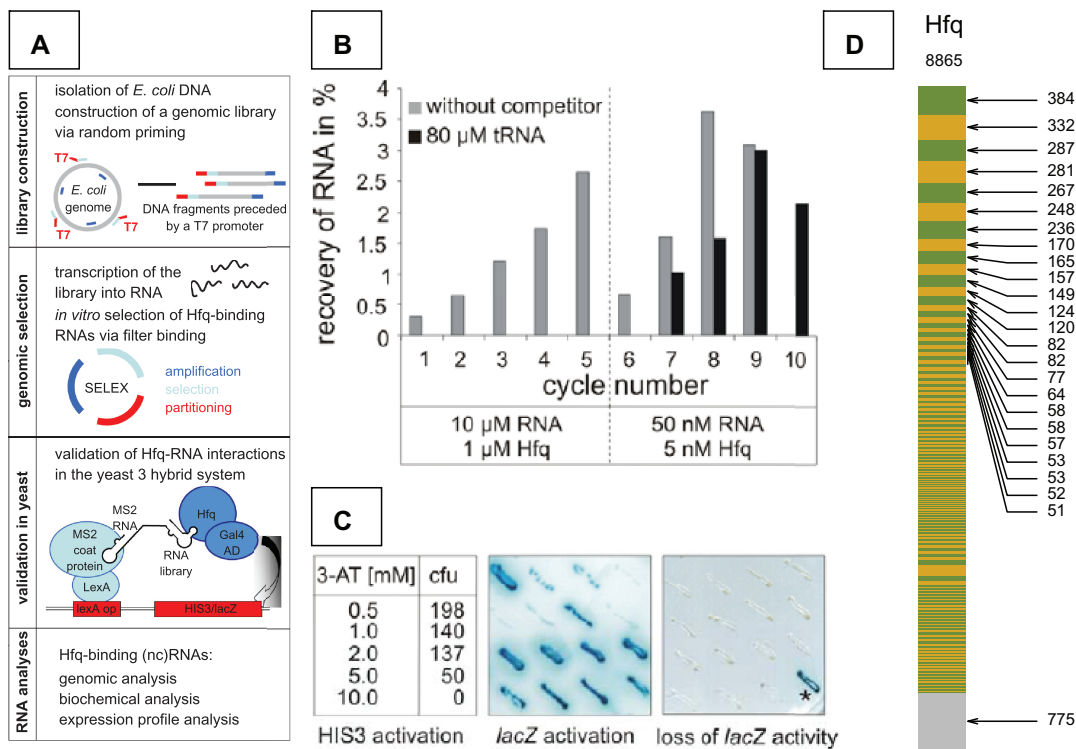
Figure 1. Genomic SELEX for Hfq-binding aptamers. (**A**) Schematic outline for our approach to isolate novel Hfq-binding ncRNAs. After construction of a genomic library via random priming, DNA fragments are *in vitro* transcribed and subjected to multiple rounds of SELEX to isolate Hfq-binding RNAs (36,48). Interactions of isolated transcripts with Hfq are evaluated in a cellular environment using the yeast three hybrid system (37). (**B**) Enrichment profile for the selection of Hfq-binding RNAs, showing the percentage of recovered RNA per cycle. RNAs were selected via filter binding. Low stringency conditions (10 μM RNA and 1 μM Hfq) were applied for the first five consecutive cycles to firmly establish a population of binding species. Then stringency was increased by reducing RNA and Hfq concentrations (50 nM RNA and 5 nM Hfq) and two parallel selections, one without (grey bars) and one with additional tRNA competitor (80 μM, black bars), were performed to select for high-affinity Hfq-binding sequences. We obtained recovery rates of up to 3.0 and 3.6% for SELEX with and without competitor, respectively. Since a 10-fold molar excess of RNA over protein was used to support competition, these values correspond to 30 and 36% of the total amount of RNA able to bind Hfq, assuming a single binding site on the protein. (**C**) Identification of HIS3 and *lacZ* activating clones in the yeast three hybrid system (37). The number of colony forming units (cfu) obtained for growth on medium supplemented with increasing 3-AT concentrations (left panel) as well as a representation of clones that display different strengths of *lacZ* activation, as indicated by colour intensity (middle panel), are shown. Prior to sequencing dark blue clones were tested for false-positive Hfq-independent reporter gene activation. In a second β-gal-assay, performed after withdrawal of Hfq from the system, all colonies remained white, indicating that they have lost their former ability to activate *lacZ* (right panel, asterisk denotes the positive control). (**D**) Sequences from the enriched pool of cycle 9 were deep sequenced using the 454 technique. Reads (8865) were mapped to the *E. coli* genome. Coloured bars represent the number of reads from individual clusters. In grey are the clusters with a single read. The depth of the clusters with 50 or more reads is indicated.

glgT mRNAs. Also targets of sRNAs known to bind Hfq are present in our selected pool, for example the *glmSU* mRNAs, which are up-regulated by the sRNAs GlmZ and GlmY (52). In addition, five mRNAs were obtained, which are regulated by sRNAs and hence might also interact with Hfq: *tna* and *gadX* mRNAs targeted by the GadY sRNA, *rbsD* mRNA (DsrA), *sdh* mRNA (RhyB) and *fhl*A mRNA (OxyS).

We further compared our sequences with RNAs found to be differentially expressed in Hfq-depleted cells and hence thought to be regulated by Hfq directly or indirectly. Fifty two of our clusters are among the down-regulated and 42 among the up-regulated genes in the microarray analysis of Hfq-depleted cells (53). For example, the mRNA of one of our strongest hits, the *ptsL/manX* asRNA, is 8-fold down-regulated in the absence of Hfq, whereas the *ygiM/cca* mRNA is 4-fold up-regulated.
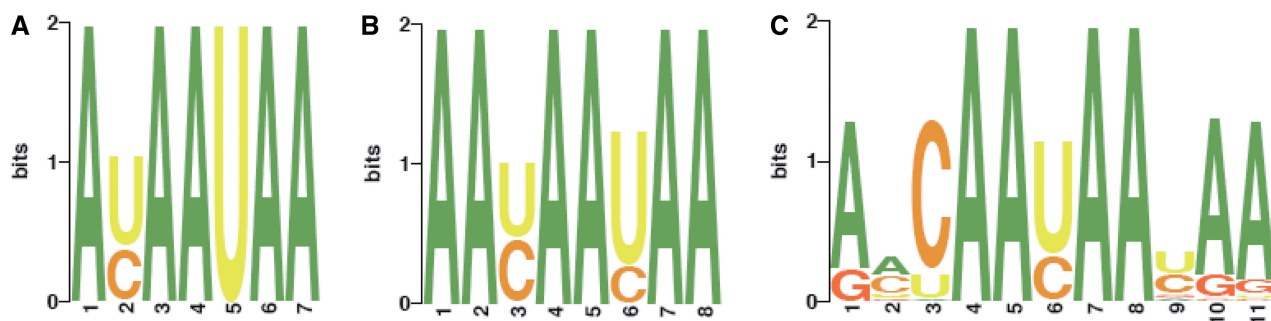
From these results, we conclude that the enrichment of Hfq-binding RNAs was successful, and that we obtained a large but non-exhaustive number of Hfq-binding aptamers, which map precisely to the *E. coli* genome. Many of the most abundant and best known Hfq-binding small RNAs have not been enriched. This might be due to a lower affinity of these RNAs to Hfq than the low nanomolar affinity of the selected aptamers, or due to secondary-structure elements, which often lead to reverse transcriptase stops.

Identification of Hfq-binding motifs

To discover a possible common motif in the Hfq-binding sequence data, the sequences were subjected to motif recognition. We have excluded all clusters with less than two reads, to ensure that we do not include any artefacts in the motif search. The remaining clusters had an average sequence length of 90bp. For this collection of clusters,

Table 1. Clusters of Hfq aptamers with more than 50 reads are listed. Cluster ID: identification number of cluster; Reads: number of obtained sequences matching with indicated cluster; Chromosome starting and ending base number is indicated as well the Gene name and its b-number

Cluster ID	Reads	Chromosome start	Chromosome end	Gene	b-number	Orientation	Product
86	384	640 717	640 856	uspG	b0607	Antisense	Universal stress protein UP12
944	332	1 002 579	1 002 746	ycbU; ycbV	b0942; b0943	Antisense; antisense	Predicted fimbrial-like adhesin Protein;predicted fimbrial-like adhesin protein
1399	287	3 967 056	3 967 166	wzzE	b3785	Antisense	Entobacterial Common Antigen (ECA) polysaccharide chain length modulation protein
1117	281	1 900 602	1 900 713	manX	b1817	Antisense	Fused mannose-specific PTS enzymes: IIA component/IIB component
635	267	3 729 119	3 729 264	xylF	b3566	Sense	D-xylose transporter subunit
1351	248	3 673 658	3 673 756	yhjE	b3523	Antisense	Predicted transporter
354	236	2 334 580	2 334 737	yfaA	b2230	Antisense	Predicted protein
1237	170	2 863 240	2 863 409	ygbN	b2740	Antisense	Predicted transporter
998	165	1 241 194	1 241 315	cvrA	b1191	Sense	Predicted cation/proton antiporter
1044	157	1 439 032	1 439 135	ydbJ	b4529	Antisense	Predicted protein
1277	149	3 199 820	3 199 959	ygiM; cca	b3055; b3056	Antisense; antisense	Predicted signal transduction protein (SH3 domain); fused tRNA nucleotidyl transferase/2′3′-cyclic phosphodiesterase/2′-nt and phosphatase
1238	124	2 890 603	2 890 749	ygcN	b2766	Antisense	Predicted oxidoreductase with FAD/NAD(P)-binding domain
946	120	1 012 183	1 012 338	pqiA	b0950	Antisense	Paraquat-inducible membrane protein A
1006	82	1 266 785	1266871	ychA	b1214	Antisense	Predicted transcriptional regulator
311	82	2 056 314	2 056 462	yeeO	b1985	Antisense	Predicted multidrug efflux system
1518	77	4 604 053	4 604 123	yjjZ; leuV	b4567; b4368	Antisense; sense	Predicted protein; tRNA–Leu
405	64	2 581 119	2 581 278	nudK	b2467	Antisense	Predicted NUDIX hydrolase
1208	58	2 662 835	2 662 948	yfhR	b2534	Antisense	Predicted peptidase
285	58	1 921 761	1 921 916	yebZ	b1840	Antisense	Predicted inner membrane protein
812	57	301 934	302 029	Intergenic	–	–	–
1118	53	1 908 964	1 909 053	yebQ	b1828	Antisense	Predicted transporter
430	53	2 736 022	2 736 135	pheA	b2599	Sense	Fused chorismate mutase P/prephenate dehydratase
674	52	3 955 917	3 956 023	ilvC	b3774	Sense	Ketol-acid reductoisomerase, NAD(P)-binding
871	51	632 630	632 729	ybdH	b0599	Sense	Predicted oxidoreductase

Figure 2. Hfq-binding motif(s). Predicted Hfq-binding motifs determined by MEME (52,53) as most significant on a remaining cluster set with at least three reads in one cluster. The height of a letter at the given position (*x*-axis) indicates the frequencies scaled relative to the information content (measure of conservation) at each position. MEME was run with fixed motif length (A) 7 nts, (B) 8 nts and (C) 11 nts.

MEME (version 4.1.1) (54,55) was applied with the default option, searching the given strand only and a model which assumes zero or more motif occurrences per sequence (TCM). Based on the initial results, we reran MEME again with fixed length 7, 8 and 11 nt (Figure 2, A–C, respectively). The following motifs were determined as most significant: 5′-AYAATA-3′ with an *E*-value of $7.00e-04$ and 5′-AAYAAYAA-3′ with an *E*-value of $1.8e-05$, where Y represents pyrimidines (C or U). Both motifs are included in the 11-mer motif with an *E*-value of $3.4e-17$ (Figure 2). The representation of these related motifs within *E. coli* was analysed. The 7-mer was found 365 times within protein coding genes and 712 times on strands opposite to protein coding genes. Hence, the factor of the 7-mer motif for asRNAs over mRNAs is 1.95. The antisense/sense ratio for the 8-mer motif is 1.3.

Determination of the Hfq-binding motif by DMS footprinting

To confirm that the predicted Hfq-binding motif is indeed the binding site for Hfq, we determined the interaction site of Hfq in the aptamers by dimethylsulphate (DMS) footprinting. This chemical reagent methylates the N1 position of adenosine and the N3 position of cytosine, in case they are accessible. DMS structural probing was performed for the three selected sequences from the *cca* asRNA, *ptsL/manX* asRNA and *icIR* mRNA. First, the secondary structures of the Hfq-binding aptamers were analysed by computational prediction via calculation of the MFE structure and the

base pairing probability matrix to take suboptimal structures into account (RNAfold program of the Vienna RNA package, 5).

Figure 3A shows as an example of the DMS footprinting data for Hfq binding for the *cca* asRNA. Acrylamide gels were quantified (Figure 3B) and protected bases are marked by green dots, whereas enhancements of DMS modification upon Hfq binding are indicated in orange. The secondary-structures probed by DMS were plotted to the secondary-structure predictions of RNAfold (Figures 3C–E). The DMS pattern of all RNAs shows distinct differences upon incubation with Hfq, both protections and enhancements could be observed. The *cca* asRNA, although highly structured harbouring two
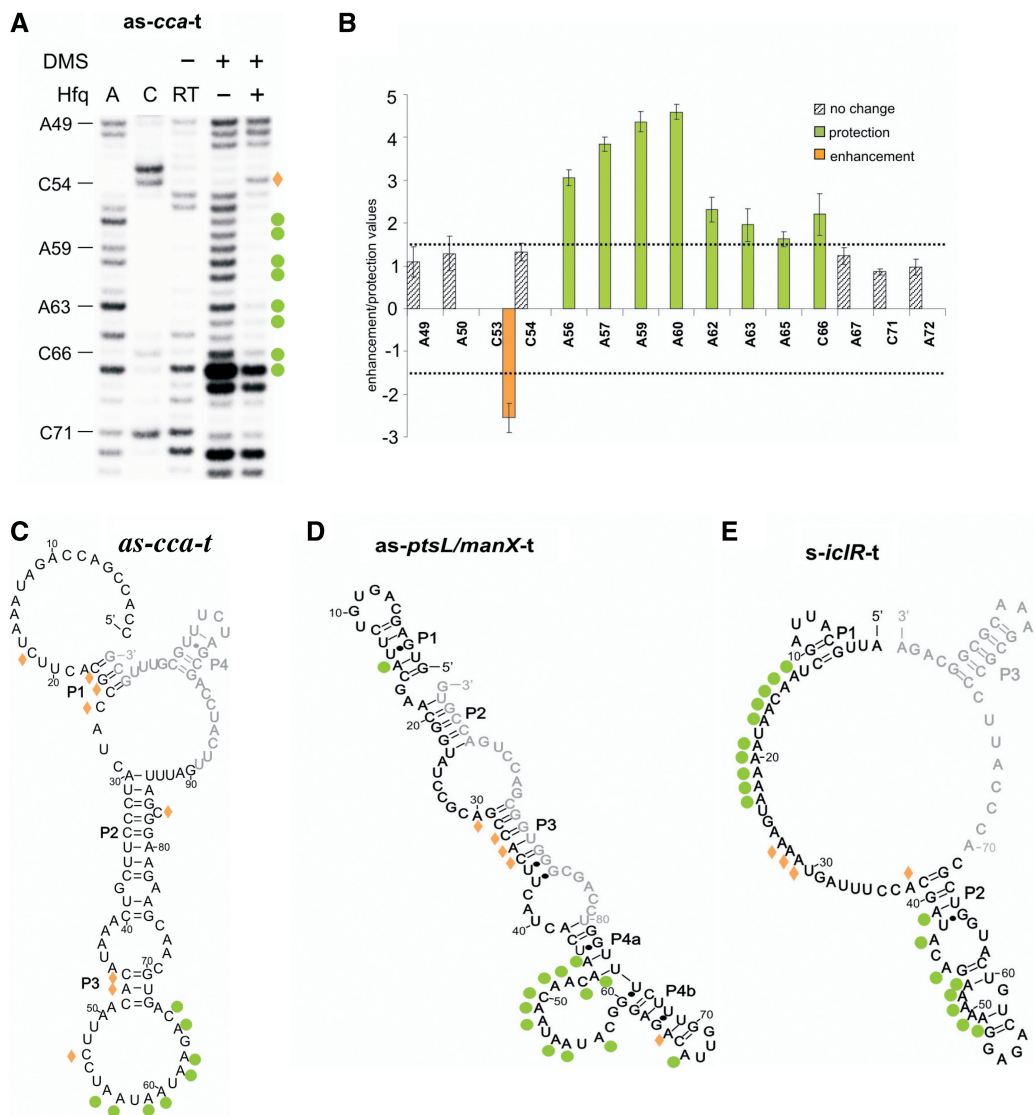

Figure 3. DMS footprinting of Hfq-binding motifs. (**A**) Footprinting data of the Hfq-binding site in the *cca* asRNA. A representative region of an 8% denaturing gel shows the cluster of protections (green dots) and one enhancement (orange diamond) upon Hfq addition [+Hfq lane; (0.4 μM Hfq₆)]. A, C are sequencing lanes, RT is the control not treated with DMS, − or +Hfq lanes are treated with DMS. (**B**) Calculated protection and enhancement (ratio of DMS −Hfq/DMS +Hfq), defined by a value of ≥1.5 for protections (green) and ≤−1.5 for enhancements (orange) are presented for the selected region. Dashed bars correspond to nucleotides without changes in the DMS pattern. (**C**) Secondary structure of the selected *cca* asRNA aptamer was predicted by the RNAfold algorithm (78). Footprinting results are shown on the predicted secondary-structure. Green circles are protection from DMS upon Hfq binding, orange squares indicate enhancements. (**D**) Footprinting the Hfq-binding site in the *ptsL/manX* asRNA, as in (C). (**E**) Footprinng the Hfq-binding site in the *icIR* mRNA, as in (C).

extended loop regions, has one discrete cluster of ten nucleotides A56–C66 (labelled green in Figure 3C) residing in the otherwise very accessible loop, that are protected from methylation in the presence of Hfq. Furthermore, induced fit upon binding of Hfq was observed, detectable by weak enhancements (labelled orange throughout Figure 3) in P1 and P3 regions. Also further weak enhancements at single-stranded cytosines showed higher accessibility upon protein interaction.

Figure 3D shows the footprinting results for the *ptsL/manX* asRNA. Strongly modified and hence accessible nucleotides are located between A24 and A30 and from A45 to C58, the latter representing the stretch of highest accessibility in this RNA, which maps perfectly to the secondary-structure prediction. A similar change of modification pattern upon Hfq binding as seen for the *cca* asRNA could also be observed for this RNA. We detected a region of eleven successive protections in the loop J4a/4b (A45–A57) and enhancements in P3, again indicating an induced fit upon Hfq binding.

For the third tested aptamer, located in the *iclR* mRNA, again a large accessible loop was detected, which is protected upon Hfq binding between position A13 and A23 (Figure 3E). In addition, a second accessible domain between positions A41 and A52 is protected. Either the sequences harbour a second Hfq binding site or a stabilization of the suboptimal secondary-structure could occur. However, the latter possibility is rather unlikely since most of the affected nucleotides are predicted to be accessible even if the stem-loop conformation forms.

From the DMS footprinting data, we conclude that the predicted enriched motif 5′-AAYAAYAA-3′, is indeed an interaction site for Hfq, because in all three tested RNAs, the motif is present and protected from DMS modification after incubation with Hfq. However, we are aware that the predicted motifs may change if regions with a different coverage (e.g. clusters of size <2) are included in our motif analysis.

The binding affinities for four selected RNAs (as-cca, as-ptsL, ig-proX/ygaX and as-metH/icIR) were determined via electrophoresis mobility shift assays. *In vitro* transcribed 5′ γ-^{32}P labelled RNAs (0.01 nM) were incubated with increasing amounts of Hfq$_6$ as indicated in Figure 4. $K_{1/2}$ values ranged from 0.2 nM for as-metH/icIR to 11 nM for ig-proX/ygaX. The Hill coefficients suggest that more than one Hfq$_6$ complex binds to the RNAs. These affinities are indeed stronger than those of known Hfq-binding small RNAs. For example, the well-studied DsrA RNA binds with a K_d of 18 nM Hfq$_{12}$ (or 36 nM Hfq$_6$) and the *rpoS* mRNA 146 Hfq$_6$ (30). The RNA with the highest affinity to Hfq known is the *sodB* mRNA, which has a K_d of 2 nM (56). The 5′ UTR of this mRNA has the exact Hfq-binding motif we identified here.

Expression profile of *cis*-antisense RNAs

As the identified *cis*-antisense RNAs originate from a genomic DNA library, we next asked under which conditions they are expressed, if at all. We isolated total RNA from cells exposed to different growth conditions and
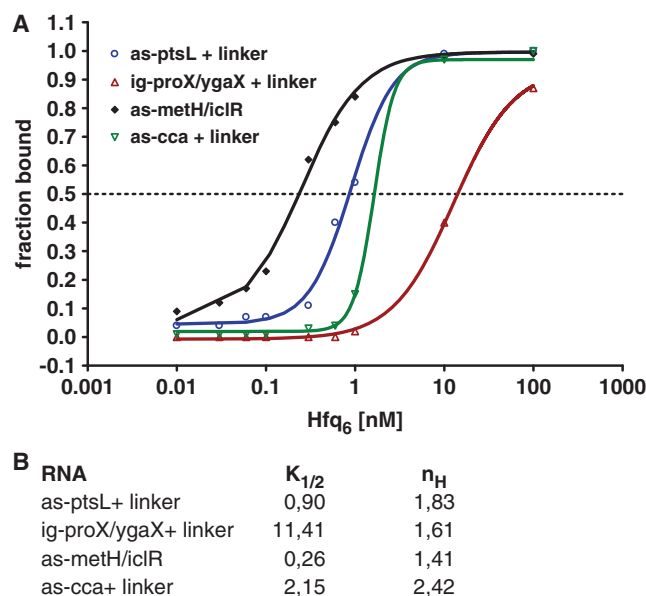

Figure 4. Determination of binding affinities of Hfq aptamers. Binding properties of Hfq aptamers were determined via electrophoretic mobility shift assays. γ-^{32}P labeled RNAs (0.01 nM) (as indicated in the figure) were incubated at room temperature for 30 min with increasing amounts of Hfq$_6$. RNA–protein complexes were separated on 5% native polyacrylamide gels. Complexes were quantified with a PhosphoImager and analysed using Imagequant 1.4 software. Analysed RNAs are indicated; 'linker' means that the aptamer was extended on the 3′-end with its genomic sequence as primer binding sites for reverse transcription. (B) Determined $K_{1/2}$ values and Hill coefficients.

studied expression patterns via RT–PCR. We analysed 11 transcripts (*cca, ptsL, ydbJ, ybdQ, mtlA, ptsI, pqi5, sprT, phoR, ybjJ* and *ftsQA* asRNAs) and found all of them to be expressed (Figure 5A). While they all show individual expression profiles, we repeatedly found that expression in stationary phase is lower than during exponential growth (9 out of 11 RNAs). For the remaining two RNAs (*ptsL* and *ptsI* asRNAs), expression is equally low both in stationary and exponential phase. Most are much higher expressed under stress than under optimal conditions. Since we observed expression for all transcripts investigated, we assume that most, if not all, of the *cis*-antisense RNAs identified in our screen are expressed likewise, suggesting that transcription from the antisense strand of protein coding genes might be much more prevalent than hitherto presumed. In addition, genome-wide low level transcription from opposite protein coding regions has been observed in a tiling array study of *E. coli* antisense genes, although expression of detected transcripts was not examined further (57).

A strong argument that the selected Hfq aptamers are indeed part of expressed RNAs that bind Hfq also *in vivo* comes from a recent report on a global study of Hfq-binding RNAs in *Salmonella typhimurium*. A high throughput approach was performed to identify Hfq target RNAs by co-immunoprecipitation of Hfq followed by deep sequencing (32). With this approach, a small set of antisense transcripts was detected. We compared the *E. coli* genes containing genomic Hfq
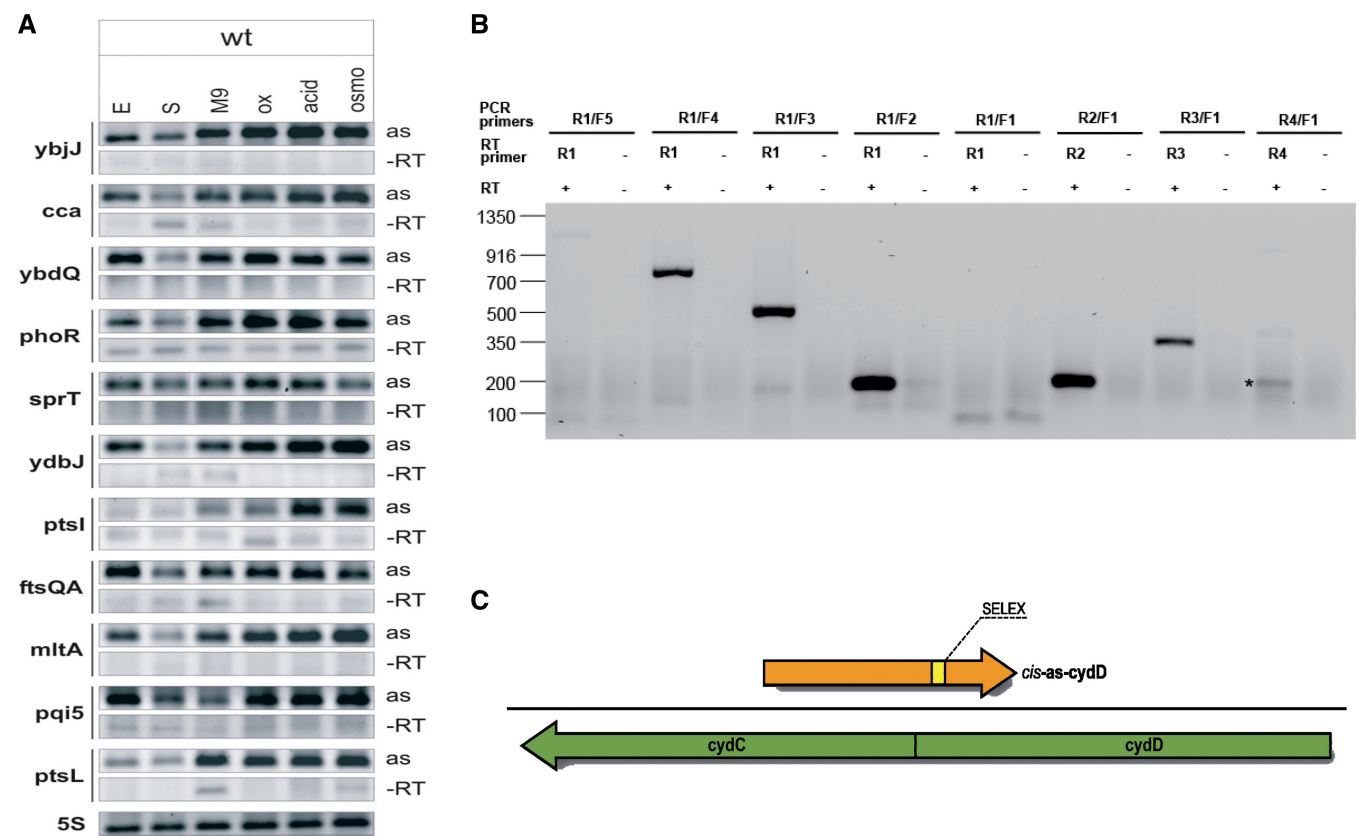
Figure 5. Characterization of *cis*-antisense transcripts. (**A**) Strand-specific RT–PCR analysis of *cis*-encoded antisense RNAs (*ybjJ*, *cca*, *ybdQ*, *phoR*, *sprT*, *ydbJ*, *ptsI*, *ftsQA*, *mltA*, *pqi5*, *ptsL* asRNAs) and 5S rRNA, using 4 µg total RNA isolated from *E. coli wt* growing under the following expression conditions: exponential phase (E), stationary phase (S), growth in minimal medium (M9), growth under oxidative (ox), acidic (acid) or hyperosmotic (osmo) stress. −RT specifies the negative control. Samples were taken from the exponential phase of PCR amplification. (**B**) Primer walking for size estimation of *cis*-antisense RNAs. Strand-specific RT–PCR of *cis*-encoded *cydD* asRNA, using 2 µg total RNA isolated from *E. coli wt* growing under hyperosmotic stress. Estimation of 5′- and 3′-ends was obtained with different combination of reverse and forward primers that map up- and down-stream of the Hfq binding fragment in a stepwise manner. Expected sizes of PCR products: R1F5,1062 bp; R1F4, 760 bp; R1F3, 510 bp; R1F2, 197 bp; R1F1, 56 bp; R2F1, 208 bp; R3F1, 396 bp; R4F1, 982 bp. Minimal length of the as-*cydD* transcript is 1073-nt. −RT specifies the negative control. Asterisk denotes unspecific PCR product. (**C**) Schematic representation of *cydD* asRNA relative to corresponding protein coding genes.

aptamers on the antisense strand with the *Salmonella* antisense RNAs and found 93 antisense RNAs to overlap in both screens. Note, that many *Salmonella* genes cannot be annotated to *E. coli* genes.

We further estimated the approximate size of the *cis*-antisense RNAs, which contained Hfq aptamers by strand specific RT–PCR and primer walking. As a typical example we show in Figure 5B the result for the *cydD* asRNA. The *cydD* asRNA is at least 1073 base pairs long (Figure 5C). We further tested three additional asRNAs (*cca*, *tldD* and *yjeH*) and all are at least 1 kb long (data not shown).

Next, we aimed to quantify expression of *cca* asRNA more precisely using qRT–PCR. In addition, this approach allowed us to quantify previously reported unspecific cDNA synthesis (58) and assess the portion to which it influences target signal. Unspecific cDNA synthesis can be a consequence of back-looping or other internal structure of the target RNA. Otherwise, it can be an effect in which short degradation products of either RNA or DNA serve as random oligonucleotides which lead to general reverse transcription which then compromises strand specificity. To monitor this artefact, we introduced

a novel control—NPC in which addition of any primer during the RT is omitted.

We quantified levels of sense and antisense *cca* transcript in two different growth conditions; exponential phase of growth and under acidic stress. Based on size of *cca* asRNA derived from primer walking (data not shown), two primer pairs were designed giving amplicons on each end of both opposing transcripts. Resulting *cca* and *htrG/ygiM* amplicons lie towards the 5′- and 3′-ends of the antisense transcript, respectively, and inversely in case of sense transcript. Resulting expression level of sense transcript is considerably high and ranges from 10×10^6 to 27×10^6 copies while antisense transcript level is $\sim 2 \times 10^5$ copies per microgram of total RNA (Figure 6). Most importantly, unspecific cDNA synthesis gives rise to only 4000 copies per microgram and hence does not influence detection of endogenous transcripts.

Considerably similar levels of *cca* asRNA obtained from both amplicons (Figure 6) imply that indeed they belong to the same endogenous transcript. This serves as a verification of long antisense transcripts, existence of which we are reporting here.

Location of Hfq aptamers within *E. coli* genes

From the total reads shown in Figure 1D, 84% are located within annotated *E. coli* genes, whereby 15% map to the sense strand of mRNAs and 69% map to the antisense
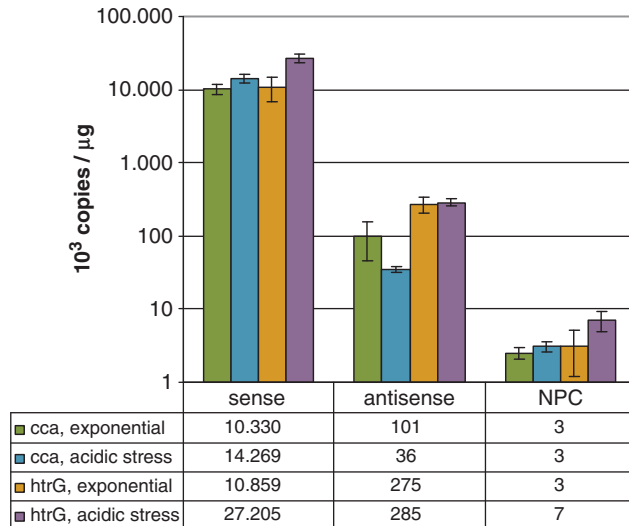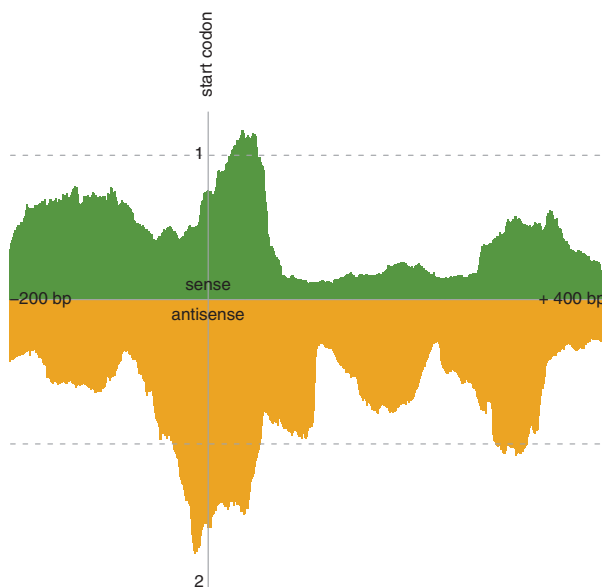

Figure 6. Quantitative strand specific RT–PCR of sense-antisense RNA pair. Copy number of sense and antisense *htrG/ygiM-cca* transcript was determined using qRT–PCR. Each transcript is represented by two independent amplicons (cca and htrG). Expression was analysed in exponential growth phase and under acidic stress. Values are average of two biological replicates, for each of which three technical replicates were preformed. NPC denotes no primer control.

strand. Thus, Hfq aptamers are 4-fold more frequent on the antisense strand than the sense strand. Reads (9.9%) were in regions that are not annotated and only 0.5% within non-coding RNAs.

We further looked more closely asking whether Hfq aptamers were evenly distributed within genes or whether there are positions where they are enriched. Figure 7A shows the relative enrichment of Hfq genomic aptamers as compared a control 'neutral' selection (35) relative to their distance to the translation start and stop sites. Most interesting is that coding regions are unenriched, whereas there is a peak surrounding the start site and reaching into a few codons into the coding region. There is a strong enrichment of Hfq aptamers antisense to the translation start site. Figure 7B shows the distance of Hfq aptamers relative to the translation termination site. Here again the enrichment in the antisense strand is remarkable when compared to the sense strand, suggesting that codon regions should be devoid of Hfq aptamers. There is a slight increase of aptamers mapping downstream of the stop codon. However, when looking closer into this phenomenon, we observed that most of these are located opposite to intervening sequences between two ORFs in polycistronic mRNAs.

A typical example of such an Hfq aptamer is the *ycbU/ ycbV* asRNA (Figure 8A). This aptamer lies opposite to a 4 kb polycistronic RNA coding for the ycb operon containing genes *ycbS, T, U, V* and *F* and the aptamer lies opposite to the intervening sequence between *ycbU* and *V*. A homologue of this large antisense RNA has recently
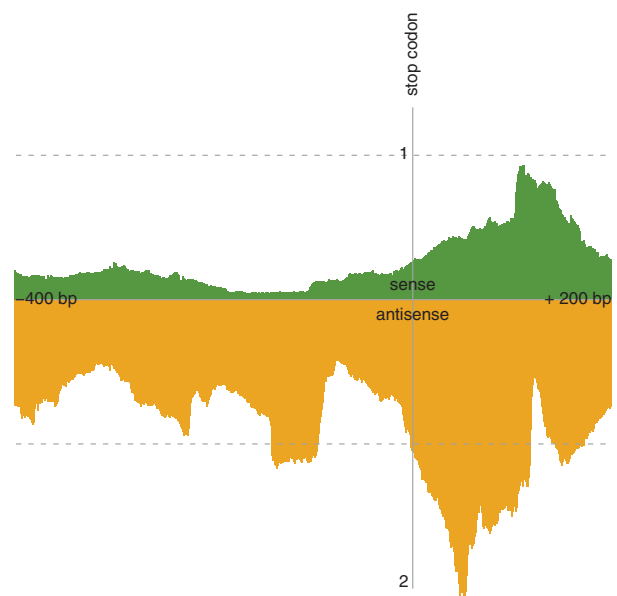

Figure 7. Enrichment of genomic Hfq aptamers surrounding start and stop codons. Positional fold enrichment of genomic aptamers as compared to a control 'neutral' selection surrounding translation signal codons. The upper part, in green, of the diagram shows aptamers in the sense strand and the lower part, in yellow, aptamers on the antisense strand. (**A**) Shows enrichment surrounding start codons and (**B**) shows enrichment surrounding stop codons.
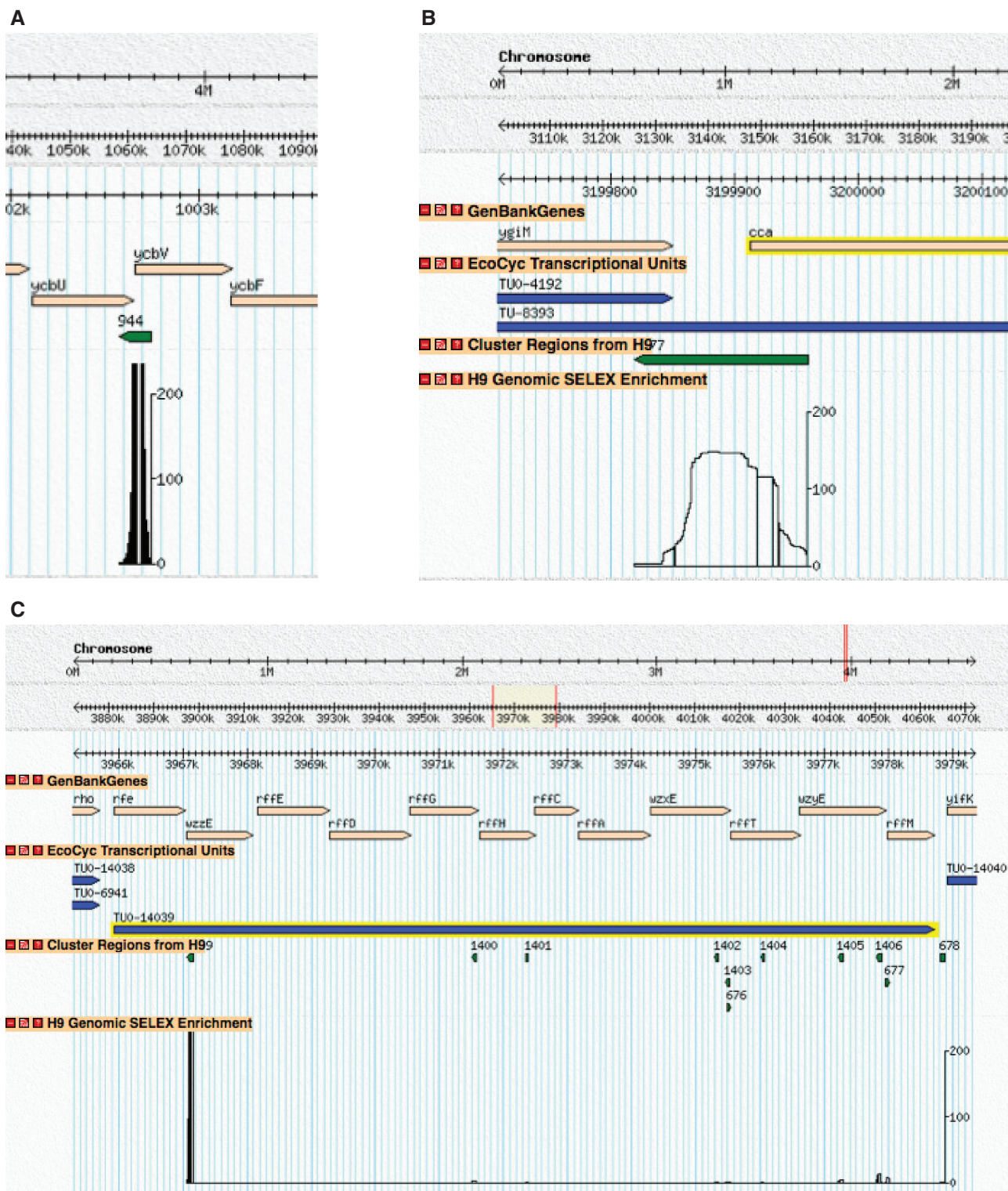
Figure 8. Examples of mapping sites of genomic Hfq aptamers in the *E. coli* genome. In this figure, selected regions of the *E. coli* K-12 genome where genomic aptamers were detected are shown in a genome browser window. Features are drawn as thick, coloured rectangles, each with a point at one end indicating the genomic strand of the feature. The scale across the top indicates the position in the genome, peach-coloured features represent open reading frames, blue features (where present) indicate transcriptional units (complete operons), green features indicate genomic aptamers and histogram data at the bottom indicates enrichment level at each base of a genomic aptamer. (**A**) *ucbU/V* asRNA exactly opposite to the intervening sequence between genes *ucbU* and *ucbV* in the ucb operoen. (**B**) *cca* asRNA opposite to the intervening sequence between the *yhiM/htrG* mRNA and the *cca* mRNA and (**C**) the ECA operon with its 12 genes and 11 Hfq aptamers in green.

been detected in a tiling array screen in *Listeria monocytogenes* (59). It's mRNA codes for a predicted fimbrial-like adhesin protein.

Another typical example of the location of Hfq aptamers is the *cca* aptamer located opposite to the intervening sequence between the *htrG/ygiM* and the *cca* genes (Figure 8B). This antisense RNA was also pulled down in *Salmonella* (32) and the mRNA is up-regulated in the absence of Hfq in *E. coli* (53).

A third example is the *wzzE* asRNA located antisense to the 5′-end of the coding sequence. *wzzE* codes for the modulator of the polysaccharide chain length of the Enterobacteria common antigen (ECA) and lies in a 13-kb-long polycistronic operon coding for 12 genes. These genes code for proteins involved in the biosynthesis of the ECA and most of the antigens interact with hosts. Interestingly, we obtained 11 genomic Hfq aptamers within this operon, most antisense but for example there is one position where we obtained genomic aptamers on both strands just close to the 3′-end of the *wzxE* gene (Figure 8C).

DISCUSSION

Genomic SELEX is a powerful tool to identify genomic aptamers

The first small regulatory RNAs were initially discovered by chance or due to their high abundance (60). Their relevance was only appreciated later when systematic computational searches looking for secondary structure, sequence conservation and orphan promoters and terminators in intergenic regions revealed a large number of small RNAs in many different bacteria (61). More recently, small RNAs have been discovered by cDNA cloning (RNomics) (7), by co-immunoprecipitation coupled to microarray analyses (6) and more lately by deep sequencing of total RNA (32). Here we describe an alternative and complementary approach to detect regulatory domains within RNAs, Genomic SELEX. While computational prediction is a knowledge-based approach, cloning and deep sequencing is dependent on the expression levels of the RNAs, limiting the discovery to those RNAs that have already known properties or are expressed at the analysed conditions. Genomic SELEX overcomes both limitations, as the initial library pool used to select RNAs is derived from genomic DNA, and no predictions have to be made. The selection procedure does, however, favour the selection of sequences with lower structural stability and specific nucleotide content, but the positive selection forces exceed the biases imposed by the SELEX procedure (35). The key step in Genomic SELEX is the selection of the protein that serves as bait. We chose the regulator protein Hfq, due to its properties. Hfq displays chaperone activities (27,29,30,62) and binds a very large number of quite different RNAs with variable affinities (13,18,63). We enriched the SELEX pool aiming at isolating high affinity binding RNAs in the low nanomolar range. Deep sequencing of the enriched SELEX pool revealed 1522 individual aptamers with varying representation in the pool. The most enriched

sequences were present over 100 times, and 775 were obtained only once. We assume that by further sequencing more sequences will be obtained.

Therefore, we conclude that genomic SELEX is a valuable approach to identify regulatory domains within RNAs provided the choice of the bait is appropriate. We performed in parallel to the Hfq selection, a selection for high affinity binders to another *E. coli* protein with RNA chaperone activity, StpA, which does not bind RNAs specifically and no RNAs were enriched (64). This indicates that only specific binders survive the procedure and are enriched.

Genomic aptamers

RNAs enriched via Genomic SELEX do not represent bona fide RNA molecules, but represent domains within putative RNAs that provide high affinity to the ligand used as bait. We therefore define the term 'genomic aptamer' as a domain within RNAs that recognizes specific ligands and as a consequence might act as regulatory element. The aptamer domains within riboswitches are typical examples of genomic aptamers that bind the metabolite of choice and induce a regulatory switch (65). The outcome of our Genomic SELEX against the regulator protein Hfq is a collection of short sequences that bind with high affinity to Hfq and have the potential to act as regulatory elements.

Genomic Hfq aptamers are highly enriched in the antisense strand of protein coding genes

The most unexpected outcome of our Hfq aptamer analysis is their mapping to the antisense strand of protein coding genes. Until recently, most antisense RNAs in *E. coli* were thought to act *in trans* and to have only limited complementarity to their target RNAs. Mainly plasmid-encoded small RNAs were transcribed from *cis*-antisense strands resulting in transcripts with potentially extensive complementarity. Recently, a few examples have been reported for chromosomally encoded homologues of the plasmid hok/sok and hok/sok-like systems. These antisense RNAs are called 'antitoxin' RNAs, as they inhibit the expression of short toxic peptides (66). They act by base pairing across the ribosome binding site blocking initiation of translation and often activating cleavage and leading to degradation of the whole mRNA. Two examples of *cis*-acting antisense RNAs are the SymR RNA, which is complementary to the Shine-Dalgarno (SD) sequence and extends into the coding region of the *symE* mRNA across the AUG start codon (67) and the sok RNA, which is complementary to the SD sequence of the *mok* mRNA, which needs to be translated to mediate translation of the *hok* mRNA (68,69). Further examples of antitoxin RNAs are the Sib and OhsC RNAs (70) and the IstR RNA(71), which act by competing with the ribosome by base pairing to the ribosome standby site ~100 nucleotides upstream of the SD site (72). These antitoxin RNAs, are from their site of expression and mode of action, clearly *cis*-acting antisense RNAs reminiscent of the Hfq-aptamers identified in this study. Like the majority of the Hfq aptamers, they map

antisense to protein coding genes just opposite of the translation initiation site spanning the SD sequence and the AUG start codon. It is, however, not known whether Hfq plays a role in promoting the interaction of the anti-toxin RNAs with their target RNAs. However, it is conceivable that these RNAs need some feature to make base pairing more efficient. For example, some of these RNAs are highly structured and their base pairing is facilitated by loop–loop interactions, which might contain a U-turn loop structure (73). Alternatively, for less structured RNAs, which cannot present a loop to facilitate base pairing, a Hfq binding site might be able to accelerate the interaction of the mRNA with its antisense RNA.

In line with this idea is the recent finding that a short motif 5′-AAYAA-3′ within the *rpoS* mRNA leader greatly affects the annealing rates of the sRNA DsrA to its target *rpoS* mRNA. Hfq accelerated the annealing of DsrA to the *rpoS* mRNA more than 50-fold. Deletion of the 5′-AAYAA-3′ motif from the rpoS leader mRNA had only a slight effect on the binding affinity to Hfq, but it practically abolished Hfq's capacity to accelerate DsrA/rpoS annealing (31). The 5′-AAYAA-3′ motif is contained within the identified motif of the genomic Hfq aptamers. This would indicate that these motifs could be involved in promoting annealing of the antisense RNAs with their target sequences.

Enrichment of Hfq aptamers opposite to translation start sites and to intervening sequences between ORFs of polycistronic transcripts

From the analysis of the position of Hfq-aptamers relative to start and stop codons three interesting observations become apparent. (i) Significantly few Hfq-aptamers map to the coding regions on the sense strand compared to the antisense strand. This suggests that there might be a reason for coding regions to be devoid of Hfq-binding sites. This is in line with a very recent report that the MicC sRNA targets the *ompD* mRNA downstream of its start codon with a 12 base pair complementary region in codons 23–26. This sRNA–mRNA interaction does not inhibit translation initiation but leads to an RNase E-dependent degradation of the mRNA (74). Thus, the presence of Hfq-binding sites within codon regions might be a destabilizing factor that mRNAs should avoid. (ii) The genomic aptamers are enriched antisense of the SD and start codons suggesting that they might be able to regulate translation. This is reminiscent of the *cis*-encoded antitoxin RNAs (66,67,69), and finally (iii) there is an enrichment opposite to intervening sequences between ORFs in operons, suggesting that Hfq aptamers might influence the processing and individual expression of genes within operons. There is one example of a small *cis*-antisense RNA, GadY, where its pairing to the region between the *gadXW* mRNAs leads to cleavage between the two ORFs and to stabilization of *gadX* mRNA (75,76).

The physiological roles of Hfq are still not completely understood. It is clear that Hfq serves as an accelerator of RNA annealing between regulatory RNAs and their targets, but for example in *Staphylococcus aureus*, Hfq lacks the C-terminal domain required for interaction with A-rich domains of mRNAs and it is not required for the function of regulatory RNAs. Our results suggest that Hfq might be involved in establishing a balance between pairs of sense/antisense transcripts. Because antisense transcripts are much less abundant than sense RNAs, we must assume that their role can only manifest when their expression is induced or when their cognate mRNA is repressed. A possible role for these antisense transcripts containing Hfq aptamers might be to inhibit translation of their cognate mRNAs, when their concentration is not higher than that of their cognate antisense pair. Our study expands the repertoire of Hfq targets to a new class of molecules, large *cis*-antisense transcripts. Their functions remain to be revealed, but our results point to a potential regulatory function in translation and in the differential expression of individual genes within operons.

REFERENCES

1. Argaman,L., Hershberg,R., Vogel,J., Bejerano,G., Wagner,E.G., Margalit,H. and Altuvia,S. (2001) Novel small RNA-encoding genes in the intergenic regions of Escherichia coli. *Curr. Biol.*, **11**, 941–950.
2. Chen,S., Lesnik,E.A., Hall,T.A., Sampath,R., Griffey,R.H., Ecker,D.J. and Blyn,L.B. (2002) A bioinformatics based approach to discover small RNA genes in the Escherichia coli genome. *Biosystems*, **65**, 157–177.
3. Rivas,E., Klein,R.J., Jones,T.A. and Eddy,S.R. (2001) Computational identification of noncoding RNAs in E. coli by comparative genomics. *Curr. Biol.*, **11**, 1369–1373.
4. Wassarman,K.M., Repoila,F., Rosenow,C., Storz,G. and Gottesman,S. (2001) Identification of novel small RNAs using comparative genomics and microarrays. *Genes Dev.*, **15**, 1637–1651.
5. Tjaden,B., Saxena,R.M., Stolyar,S., Haynor,D.R., Kolker,E. and Rosenow,C. (2002) Transcriptome analysis of Escherichia coli using high-density oligonucleotide probe arrays. *Nucleic Acids Res.*, **30**, 3732–3738.
6. Zhang,A., Wassarman,K.M., Rosenow,C., Tjaden,B.C., Storz,G. and Gottesman,S. (2003) Global analysis of small RNA and mRNA targets of Hfq. *Mol. Microbiol.*, **50**, 1111–1124.
7. Vogel,J., Bartels,V., Tang,T.H., Churakov,G., Slagter-Jager,J.G., Huttenhofer,A. and Wagner,E.G. (2003) RNomics in Escherichia

coli detects new sRNA species and indicates parallel transcriptional output in bacteria. *Nucleic Acids Res.*, **31**, 6435–6443.

8. Kawano,M., Storz,G., Rao,B.S., Rosner,J.L. and Martin,R.G. (2005) Detection of low-level promoter activity within open reading frame sequences of Escherichia coli. *Nucleic Acids Res.*, **33**, 6268–6276.

9. Gottesman,S. (2004) The small RNA regulators of Escherichia coli: roles and mechanisms. *Annu. Rev. Microbiol.*

10. Storz,G., Altuvia,S. and Wassarman,K.M. (2005) An abundance of RNA regulators. *Annu. Rev. Biochem.*, **74**, 199–217.

11. Waters,L.S. and Storz,G. (2009) Regulatory RNAs in bacteria. *Cell*, **136**, 615–628.

12. Valentin-Hansen,P., Eriksen,M. and Udesen,C. (2004) The bacterial Sm-like protein Hfq: a key player in RNA transactions. *Mol. Microbiol.*, **51**, 1525–1533.

13. Brennan,R.G. and Link,T.M. (2007) Hfq structure, function and ligand binding. *Curr. Opin. Microbiol.*, **10**, 125–133.

14. Franze de Fernandez,M.T., Eoyang,L. and August,J.T. (1968) Factor fraction required for the synthesis of bacteriophage Qbeta-RNA. *Nature*, **219**, 588–590.

15. Tsui,H.C., Leung,H.C. and Winkler,M.E. (1994) Characterization of broadly pleiotropic phenotypes caused by an hfq insertion mutation in Escherichia coli K-12. *Mol. Microbiol.*, **13**, 35–49.

16. Muffler,A., Fischer,D. and Hengge-Aronis,R. (1996) The RNA-binding protein HF-I, known as a host factor for phage Qbeta RNA replication, is essential for rpoS translation in Escherichia coli. *Genes Dev.*, **10**, 1143–1151.

17. Vytvytska,O., Moll,I., Kaberdin,V.R., von Gabain,A. and Blasi,U. (2000) Hfq (HF1) stimulates ompA mRNA decay by interfering with ribosome binding. *Genes Dev.*, **14**, 1109–1118.

18. Folichon,M., Arluison,V., Pellegrini,O., Huntzinger,E., Regnier,P. and Hajnsdorf,E. (2003) The poly(A) binding protein Hfq protects RNA from RNase E and exoribonucleolytic degradation. *Nucleic Acids Res.*, **31**, 7302–7310.

19. Sledjeski,D.D., Whitman,C. and Zhang,A. (2001) Hfq is necessary for regulation by the untranslated RNA DsrA. *J. Bacteriol.*, **183**, 1997–2005.

20. Masse,E. and Gottesman,S. (2002) A small RNA regulates the expression of genes involved in iron metabolism in Escherichia coli. *Proc. Natl Acad. Sci. USA*, **99**, 4620–4625.

21. Zhang,A., Wassarman,K.M., Ortega,J., Steven,A.C. and Storz,G. (2002) The Sm-like Hfq Protein Increases OxyS RNA Interaction with Target mRNAs. *Mol. Cell*, **9**, 11–22.

22. Masse,E., Escorcia,F.E. and Gottesman,S. (2003) Coupled degradation of a small regulatory RNA and its mRNA targets in Escherichia coli. *Genes Dev.*, **17**, 2374–2383.

23. Robertson,G.T. and Roop,R.M. Jr (1999) The Brucella abortus host factor I (HF-I) protein contributes to stress resistance during stationary phase and is a major determinant of virulence in mice. *Mol. Microbiol.*, **34**, 690–700.

24. Sonnleitner,E., Hagens,S., Rosenau,F., Wilhelm,S., Habel,A., Jager,K.E. and Blasi,U. (2003) Reduced virulence of a hfq mutant of Pseudomonas aeruginosa O1. *Microb. Pathog.*, **35**, 217–228.

25. Christiansen,J.K., Larsen,M.H., Ingmer,H., Sogaard-Andersen,L. and Kallipolitis,B.H. (2004) The RNA-binding protein Hfq of Listeria monocytogenes: role in stress tolerance and virulence. *J. Bacteriol.*, **186**, 3355–3362.

26. Schumacher,M.A., Pearson,R.F., Moller,T., Valentin-Hansen,P. and Brennan,R.G. (2002) Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein. *EMBO J.*, **21**, 3546–3556.

27. Moller,T., Franch,T., Hojrup,P., Keene,D.R., Bachinger,H.P., Brennan,R.G. and Valentin-Hansen,P. (2002) Hfq. A Bacterial Sm-like Protein that Mediates RNA-RNA Interaction. *Mol. Cell*, **9**, 23–30.

28. Kawamoto,H., Koide,Y., Morita,T. and Aiba,H. (2006) Base-pairing requirement for RNA silencing by a bacterial small RNA and acceleration of duplex formation by Hfq. *Mol. Microbiol.*, **61**, 1013–1022.

29. Rajkowitsch,L. and Schroeder,R. (2007) Dissecting RNA chaperone activity. *RNA*, **13**, 2053–2060.

30. Lease,R.A. and Woodson,S.A. (2004) Cycling of the Sm-like protein Hfq on the DsrA small regulatory RNA. *J. Mol. Biol.*, **344**, 1211–1223.

31. Soper,T.J. and Woodson,S.A. (2008) The rpoS mRNA leader recruits Hfq to facilitate annealing with DsrA sRNA. *Rna*, **14**, 1907–1917.

32. Sittka,A., Lucchini,S., Papenfort,K., Sharma,C.M., Rolle,K., Binnewies,T.T., Hinton,J.C. and Vogel,J. (2008) Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. *PLoS Genet.*, **4**, e1000163.

33. Sittka,A. and Vogel,J. (2008) A glimpse at the evolution of virulence control. *Cell Host Microbe*, **4**, 310–312.

34. Sittka,A., Sharma,C.M., Rolle,K. and Vogel,J. (2009) Deep sequencing of Salmonella RNA associated with heterologous Hfq proteins in vivo reveals small RNAs as a major target class and identifies RNA processing phenotypes. *RNA Biol.*, **6**, 266–275.

35. Zimmermann,B., Gesell,T., Chen,D., Lorenz,C. and Schroder,R. (2010) Monitoring Genomic Sequences During SELEX Using High-Throughput Sequencing: Neutral SELEX. *PLoSONE*, **5**, e9169.

36. Lorenz,C., F.,V.P. and Schroeder,R. (2006) Genomic systematic evolution of ligands by exponential enrichment (genomic SELEX) for the identification of protein-binding RNAs independent of their expression level. *Nature Protocols*, **1**, 2204–2212.

37. Bernstein,D.S., Buter,N., Stumpf,C. and Wickens,M. (2002) Analyzing mRNA-protein complexes using a yeast three-hybrid system. *Methods*, **26**, 123–141.

38. Jahn,C.E., Charkowski,A.O. and Willis,D.K. (2008) Evaluation of isolation methods and RNA integrity for bacterial RNA quantitation. *J Microbiol Methods*, **75**, 318–324.

39. Waldsich,C. and Schroeder,R. (2005) In Hartman,R.K., Bindereif,A., Schoen,A. and Westhof,E. (eds), *Handbook of RNA Biochemistry*, Vol. 1. Wiley-VCH Verlag GmbH, Weinheim, pp. 229–237.

40. Abouelhoda,M., Kurtz,S. and Ohlebusch,E. (2004) Replacing suffix trees with enhanced suffix arrays. *J. Discrete Algorithms*.

41. Benson,D., Lipman,D.J. and Ostell,J. (1993) GenBank. *Nucleic Acids Res.*, **21**, 2963–2965.

42. Karp,P.D., Riley,M., Paley,S.M., Pellegrini-Toole,A. and Krummenacker,M. (1999) Eco Cyc: encyclopedia of Escherichia coli genes and metabolism. *Nucleic Acids Res.*, **27**, 55–58.

43. R Development Core Team (2008) Vienna, Austria.

44. Zuker,M. and Stiegler,P. (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.*, **9**, 133–148.

45. Zuker,M. (2000) Calculating nucleic acid secondary structure. *Curr. Opin. Struct. Biol.*, **10**, 303–310.

46. McCaskill,J.S. (1990) The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, **29**, 1105–1119.

47. Hofacker,I.L., Fontana,W., Stadler,P.F., Bonhoeffer,S., Tacker,M. and Schuster,P. (1994) Fast folding and comparison of RNA secondary structures. *Monatsh. Chem.*, **125**, 167–188.

48. Singer,B., Shtatland,T., Brown,D. and Gold,L. (1997) Libraries for genomic SELEX. *Nucleic Acids Res.*, **25**, 781–786.

49. Urbanowski,M.L., Stauffer,L.T. and Stauffer,G.V. (2000) The gcvB gene encodes a small untranslated RNA involved in expression of the dipeptide and oligopeptide transport systems in Escherichia coli. *Mol. Microbiol.*, **37**, 856–868.

50. Guillier,M. and Gottesman,S. (2006) Remodelling of the Escherichia coli outer membrane by two small regulatory RNAs. *Mol. Microbiol.*, **59**, 231–247.

51. Lee,T. and Feig,A.L. (2008) The RNA binding protein Hfq interacts specifically with tRNAs. *Rna*, **14**, 514–523.

52. Urban,J.H., Papenfort,K., Thomsen,J., Schmitz,R.A. and Vogel,J. (2007) A conserved small RNA promotes discoordinate expression of the glmUS operon mRNA to activate GlmS synthesis. *J. Mol. Biol.*, **373**, 521–528.

53. Guisbert,E., Rhodius,V.A., Ahuja,N., Witkin,E. and Gross,C.A. (2007) Hfq modulates the sigmaE-mediated envelope stress response and the sigma32-mediated cytoplasmic stress response in Escherichia coli. *J. Bacteriol.*, **189**, 1963–1973.

54. Bailey,T.L. and Elkan,C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, **2**, 28–36.
55. Bailey,T.L., Williams,N., Misleh,C. and Li,W.W. (2006) MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.*, **34**, W369–W373.
56. Geissmann,T.A. and Touati,D. (2004) Hfq, a new chaperoning role: binding to messenger RNA determines access for small RNA regulator. *EMBO J.*, **23**, 396–405.
57. Selinger,D.W., Cheung,K.J., Mei,R., Johansson,E.M., Richmond,C.S., Blattner,F.R., Lockhart,D.J. and Church,G.M. (2000) RNA expression analysis using a 30 base pair resolution Escherichia coli genome array. *Nat. Biotechnol.*, **18**, 1262–1268.
58. Haddad,F., Qin,A.X., Giger,J.M., Guo,H. and Baldwin,K.M. (2007) Potential pitfalls in the accuracy of analysis of natural sense-antisense RNA pairs by reverse transcription-PCR. *BMC Biotechnol.*, **7**, 21.
59. Toledo-Arana,A., Dussurget,O., Nikitas,G., Sesto,N., Guet-Revillet,H., Balestrino,D., Loh,E., Gripenland,J., Tiensuu,T., Vaitkevicius,K. *et al.* (2009) The Listeria transcriptional landscape from saprophytism to virulence. *Nature*, **459**, 950–956.
60. Wassarman,K.M., Zhang,A. and Storz,G. (1999) Small RNAs in Escherichia coli. *Trends Microbiol.*, **7**, 37–45.
61. Livny,J. and Waldor,M.K. (2007) Identification of small RNAs in diverse bacterial species. *Curr. Opin. Microbiol.*, **10**, 96–101.
62. Moll,I., Leitsch,D., Steinhauser,T. and Blasi,U. (2003) RNA chaperone activity of the Sm-like Hfq protein. *EMBO Rep.*, **4**, 284–289.
63. Mikulecky,P.J., Kaw,M.K., Brescia,C.C., Takach,J.C., Sledjeski,D.D. and Feig,A.L. (2004) Escherichia coli Hfq has distinct interaction surfaces for DsrA, rpoS and poly(A) RNAs. *Nat. Struct. Mol. Biol.*, **11**, 1206–1214.
64. Mayer,O., Lorenz,C. and Schroeder,R. (2006) Transient interactions rather than tight binding determines the RNA chaperone activity of the E. coli protein StpA. *Nucleic Acids Res.*, **35**, 1257–1269.
65. Wakeman,C.A., Winkler,W.C. and Dann,C.E. 3rd (2007) Structural features of metabolite-sensing riboswitches. *Trends Biochem. Sci.*, **32**, 415–424.
66. Fozo,E.M., Hemm,M.R. and Storz,G. (2008) Small toxic proteins and the antisense RNAs that repress them. *Microbiol. Mol. Biol. Rev.*, **72**, 579–589, table of contents.
67. Kawano,M., Aravind,L. and Storz,G. (2007) An antisense RNA controls synthesis of an SOS-induced toxin evolved from an antitoxin. *Mol. Microbiol.*, **64**, 738–754.
68. Gerdes,K., Nielsen,A., Thorsted,P. and Wagner,E.G. (1992) Mechanism of killer gene activation. Antisense RNA-dependent RNase III cleavage ensures rapid turn-over of the stable hok, srnB and pndA effector messenger RNAs. *J. Mol. Biol.*, **226**, 637–649.
69. Gerdes,K. and Wagner,E.G. (2007) RNA antitoxins. *Curr. Opin. Microbiol.*, **10**, 117–124.
70. Fozo,E.M., Kawano,M., Fontaine,F., Kaya,Y., Mendieta,K.S., Jones,K.L., Ocampo,A., Rudd,K.E. and Storz,G. (2008) Repression of small toxic protein synthesis by the Sib and OhsC small RNAs. *Mol. Microbiol.*, **70**, 1076–1093.
71. Vogel,J., Argaman,L., Wagner,E.G. and Altuvia,S. (2004) The small RNA IstR inhibits synthesis of an SOS-induced toxic peptide. *Curr. Biol.*, **14**, 2271–2276.
72. Darfeuille,F., Unoson,C., Vogel,J. and Wagner,E.G. (2007) An antisense RNA inhibits translation by competing with standby ribosomes. *Mol. Cell*, **26**, 381–392.
73. Franch,T., Petersen,M., Wagner,E.G., Jacobsen,J.P. and Gerdes,K. (1999) Antisense RNA regulation in prokaryotes: rapid RNA/RNA interaction facilitated by a general U-turn loop structure. *J. Mol. Biol.*, **294**, 1115–1125.
74. Pfeiffer,V., Papenfort,K., Lucchini,S., Hinton,J.C. and Vogel,J. (2009) Coding sequence targeting by MicC RNA reveals bacterial mRNA silencing downstream of translational initiation. *Nat. Struct. Mol. Biol.*, **16**, 840–846.
75. Opdyke,J.A., Kang,J.G. and Storz,G. (2004) GadY, a small-RNA regulator of acid response genes in Escherichia coli. *J. Bacteriol.*, **186**, 6698–6705.
76. Tramonti,A., De Canio,M. and De Biase,D. (2008) GadX/GadW-dependent regulation of the Escherichia coli acid fitness island: transcriptional control at the gadY-gadW divergent promoters and identification of four novel 42 bp GadX/GadW-specific binding sites. *Mol. Microbiol.*, **70**, 965–982.
77. Vecerek,B., Rajkowitsch,L., Sonnleitner,E., Schroeder,R. and Blasi,U. (2008) The C-terminal domain of Escherichia coli Hfq is required for regulation. *Nucleic Acids Res*, **36**, 133–143.
78. Hofacker,I.L. (2003) Vienna RNA secondary structure server. *Nucleic Acids Res.*, **31**, 3429–3431.