Molecular Microbiology (2003) **50**(4), 1111–1124 doi:10.1046/j.1365-2958.2003.03734.x

Global analysis of small RNA and mRNA targets of Hfq

Aixia Zhang,[1] Karen M. Wassarman,[2] Carsten Rosenow,[3] Brian C. Tjaden,[4†] Gisela Storz[1*] and Susan Gottesman[5*]

[1]*Cell Biology and Metabolism Branch, National Institute of Child Health and Development, Bethesda MD 20892, USA.*

[2]*Department of Bacteriology, University of Wisconsin, Madison, WI 53706, USA.*

[3]*Affymetrix, Santa Clara, CA 95051, USA.*

[4]*Department of Computer Science, University of Washington, Seattle, WA 98195, USA.*

[5]*Laboratory of Molecular Biology, National Cancer Institute, Bethesda, MD 20892, USA.*

Summary

Hfq, a bacterial member of the Sm family of RNA-binding proteins, is required for the action of many small regulatory RNAs that act by basepairing with target mRNAs. Hfq binds this family of small RNAs efficiently. We have used co-immunoprecipitation with Hfq and direct detection of the bound RNAs on genomic microarrays to identify members of this small RNA family. This approach was extremely sensitive; even Hfq-binding small RNAs expressed at low levels were readily detected. At least 15 of 46 known small RNAs in *E. coli* interact with Hfq. In addition, high signals in other intergenic regions suggested up to 20 previously unidentified small RNAs bind Hfq; five were confirmed by Northern analysis. Strong signals within genes and operons also were detected, some of which correspond to known Hfq targets. Within the *argX-hisR-leuT-proM* operon, Hfq appears to compete with RNase E and modulate RNA processing and degradation. Thus Hfq immunoprecipitation followed by microarray analysis is a highly effective method for detecting a major class of small RNAs as well as identifying new Hfq functions.

Introduction

The recognition of the roles of small, non-coding RNA regulators in bacteria, archaea and eukaryotes has greatly expanded over the past few years (reviewed in Gottesman, 2002; Grosshans and Slack, 2002; Storz, 2002; Wassarman, 2002; Massé *et al*., 2003). A subset of these small RNAs act via short, interrupted basepairing interactions with target mRNAs. How do these small RNAs find and anneal to their targets? In *Escherichia coli*, at least part of the answer lies in their association with and dependence upon the RNA chaperone, Hfq. The abundant Hfq protein was identified originally as a host factor for RNA phage Qβ replication (Franze de Fernandez *et al*., 1968), but later *hfq* mutants were found to exhibit multiple phenotypes (Brown and Elliott, 1996; Muffler *et al*., 1996). These defects are, at least in part, a reflection of the fact that Hfq is required for the function of several small RNAs including DsrA, RprA, Spot42, OxyS and RyhB (Zhang *et al*., 1998; Sledjeski *et al*., 2001; Massé and Gottesman, 2002; Møller *et al*., 2002). All of these small RNAs are believed to act by complementary pairing with target messages and Hfq has been shown to promote annealing of OxyS and Spot42 RNAs to their target mRNAs *in vitro* (Møller *et al*., 2002; Zhang *et al*., 2002). In addition to its role in facilitating small RNA function, Hfq also has been found to contribute to regulation of *ompA* mRNA stability and the polyA tailing of some mRNAs (Hajndsorf and Regnier, 2000; Vytvytska *et al*., 2000).

Hfq is a highly conserved protein encoded within many bacterial genomes (Sun *et al*., 2002). It oligomerizes into a hexameric ring structure, and both sequence and structural analyses show a significant similarity to archaeal and eukaryotic Sm and Sm-like proteins integral to premRNA splicing and RNA degradation complexes (Schumacher *et al*., 2002). Hfq does not have a precise target sequence but appears to bind unstructured AU rich sequences, frequently close to more structured RNA regions (Vytvytska *et al*., 2000; Zhang *et al*., 2002; Brescia *et al*., 2003), similar to, although not as specific as, the binding sites defined for eukaryotic Sm and Sm-like proteins.

We recently carried out a genome-wide search for new small RNAs in *E. coli* and identified 17 novel RNAs (Wassarman *et al*., 2001). As a component of the characterization of these new small RNAs, we tested for potential interactions with Hfq using a co-immunoprecipitation analysis. We found significant enrichment for eight of the novel small RNAs in the samples selected with Hfq-specific antiserum, indicating they interact with Hfq (Wassarman *et al*., 2001). Other small RNAs were either poorly enriched, or not detected among the selected

Accepted 21 July, 2003. *For correspondence. E-mail storz@helix.nih.gov; Tel. (+1) 301 402 0968; Fax (+1) 301 496 3875; or E-mail susang@helix.nih.gov; Tel. (+1) 301 496 3524; Fax (+1) 301 496 3875. †Present address. Computer Science Department, Wellesley College, Wellesley, MA 02481, USA.

RNAs, indicating they bind Hfq inefficiently or not at all. To directly detect and identify more members of the Hfq-dependent family of small RNAs, we carried out microarray analysis of *E. coli* RNAs which co-immunoprecipitate with Hfq. To evaluate the reliability and sensitivity of this approach, we first analysed 46 known small RNAs and found that at least 15 specifically interact with Hfq, in agreement with and extending previous studies. Microarray results predicted the presence of 20 novel Hfq-binding small RNAs; five were confirmed to be bound by Hfq by Northern analysis. We also found that Hfq associates with a number of mRNAs and operon mRNAs, suggesting this approach also may be useful for identification of RNAs that are targets of Hfq, either directly or indirectly through the action of an associated small RNA. Finally, we have identified an association of Hfq with precursors of the *proM* tRNA, suggesting yet other roles for Hfq within the cell.

Results

Microarray detection of RNAs which co-immunoprecipitate with Hfq

Wild-type cells were grown under three different conditions, LB (exponential phase and stationary phase) and minimal glucose medium. Extracts from these cells were prepared and subjected to immunoprecipitation with either Hfq-specific serum or control preimmune serum. Immunoprecipitated RNAs were identified by direct hybridization on DNA microarrays; these microarrays carry 15 oligonucleotide probes (25-mers) within each gene and most intergenic (Ig) regions. Hybridization of RNAs to oligonucleotide probes was detected directly with antibodies specific for RNA:DNA hybrids (see *Experimental procedures*). This novel method significantly improved the sensitivity of detection and avoided problems inherent in the labelling of small, structured RNAs. Two separate experiments were carried out in LB (coded E1, E2 for the exponential phase samples and S1, S2 for the stationary phase samples); the experiment in minimal glucose media (M) was a single trial.

Corrected signals for individual probes were calculated, and the regions with two or more adjacent probes above a given expression level were rated as described in *Experimental procedures*. If the average signal for a group of probes was equal to or greater than 10 000 in two duplicate experiments (E1 and E2 or S1 and S2) or in the single minimal media experiment (M), the region was rated 5. If the first condition was not met, and the signal from two of the duplicate experiments or the single minimal growth experiment was equal to or greater than 5000, the region was rated 4 (see *Experimental procedures* for further explanation of ratings).

Microarray detection of known small RNAs which bind Hfq

A total of 46 small RNAs have been described in *E. coli*, as a result of earlier work as well as recent genome-wide searches (Argaman *et al*., 2001; Rivas *et al*., 2001; Wassarman *et al*., 2001; Chen *et al*., 2002) (Table 1). We began our analysis by evaluating the ability of these small RNAs to be detected in the Hfq co-immunoprecipitated samples using the microarray method and rating as described above. Seventeen RNAs were rated 5, four RNAs were rated 4, and the rest had lower scores. Several of these 46 small RNAs had previously been tested for their ability to bind Hfq; 12 were shown to efficiently interact with Hfq, five did not bind Hfq and one gave partial binding (Sledjeski *et al*., 2001; Wassarman *et al*., 2001; Møller *et al*., 2002; Zhang *et al*., 2002). These data provided a reference set for evaluating the sensitivity and specificity of the microarray approach. Of the 12 previously identified Hfq-binding RNAs, 11 were rated 5 in our microarray tests. Of those previously found not to bind Hfq or only show partial binding, all except tmRNA (see below) had a score of 4 or lower.

To directly assess Hfq binding to seven previously untested small RNAs, Northern analysis was carried out on the RNA extracted from Hfq immunoprecipitation samples (Fig. 1). Three of the four RNAs which were rated 5 or 4 were confirmed to bind Hfq (DicF, SraD and MicF) and none of the three RNAs tested that were rated 3 or lower (RydB, IS092, and RNase P) were positive for Hfq binding. Thus, for previously detected small RNAs, a rating of 5 was a good predictor of Hfq-binding.

There are two exceptions where we see microarray scores of 5, yet do not find specific interactions with Hfq as assessed by Northern analysis, 4.5S (Fig. 1) and tmRNA (Wassarman *et al*., 2001). Notably, these two RNAs are present at very high levels within the cell, and therefore it is possible these RNAs give high scores on microarrays as a result of their abundance. High signal was observed for both RNAs in the immune and preimmune samples (*Experimental procedures* and data not shown), and 4.5S is detected in both the anti-Hfq and preimmune samples upon overexposure of the Northern blot. Alternatively, it is possible that the 4.5S and tmRNA or their precursors do exhibit transient or low-efficiency binding to Hfq. We note that the RNase P and 6S RNAs, which also are abundant, did not give a high score in the microarray, even before correction for the preimmune samples (data not shown), although they too can be detected in overexposed Northern blots in both immune and preimmune samples. Thus, it is not clear if the high scores for tmRNA and 4.5S reflect specificity.

There also is one exception where a known Hfq-binding RNA gives a low microarray signal, RprA RNA. This anomaly can be explained by re-examining the raw expression

Table 1. Hfq immunoprecipitation of known small RNAs.

Ig or b#	Start	Length	Cons	Name	Flanking genes	Strand	RNA Size	Score	Microarray experiments ≥10 000	Hfq binding confirmed
10	14077	91		tpke11	*dnaK/dnaJ*	> > >	370	5	S1; S2; M	
b0455	475596	300	5	4.5S (*ffs*)	*ybaZ/ybaA*	< > >	114	5	E1; E2; S1; S2; M	**no**
528	887180	180	4	RybB	*ybjK/ybjL*	> < <	80	5	E1; E2; S1; S2; M	yes (d)
b1574	1647063	570	5	DicF	*rzpQ/dicB*	> > >	53	5	E1; E2; S1; S2	**yes**
1157	1920997	395	4	RyeB	*pphA/yebY*	< < <	100	5	S2; M	yes (d)
b1954	2023233	300	5	DsrA	*dsrB/yedP*	< < >	85	5	E1; E2; S1; S2; M	yes (c)
1292	2165049	278	4	RyeE	*yegQ/orgK*	> > <	86	5	E1; S1; M	yes (d)
b2621	2753397	783	5	tmRNA (*ssrA*)	*smpB/intA*	> > >	363	5	E1; E2; S2; M	no (d)
1666	2812755	153	4	SraD	*luxS/gshA*	< > <	~70	5	E1; E2; S1; S2; M	**yes**
1736	2940590	353	5	GcvB	*gcvA/ygdI*	< > <	205	5	E1, E2, S1, S2	
1755	2974037	584	4	RygA (PAIR2)	*aas/galR*	< < >	89	5	E1, E2, S1, S2, M	yes (d)
1755	2974037	584	4	RygB/SraE (PAIR2)	*aas/galR*	< < >	83	5	S1, S2	yes (d)
1992	3348110	223	4	RyhA/SraH	*elbB/arcB*	< > <	45	5	E1; E2; S1; S2	yes (d)
2116	3578437	332	4	RyhB (SraI)	*yhhX/yhhY*	< < >	90	5	M	yes (d)
2340	3983621	681	4	RyiA/SraJ	*aslA/hemY*	< > <	210	5	E1; E2	yes (d)
b3864	4047330	386	5	Spot42 (*spf*)	*polA/yihA*	> > <	109	5	E1; E2; S1; S2; M	yes (b,d)
2440	4155800	269	5	OxyS	*argH/oxyR*	> < >	109	5	E1; E2; S1; S2; M	yes (a)
1372	2310770	738	5	MicF	*ompC/yojN*	< > >	93	4		**yes**
1725	2922136	624	5	CsrB	*ygcC/syd*	< < <	360	4	E1; S1	
2163	3662249	370	1	IS183	*yhiW/yhiX*	< > <	85, 100	4		
2381	4048313	614	4	CsrC (RyiB/SraK/tpk2)	*yihA/yihI*	< > >	245	4	S2	no (d)
103	189507	367	5	t44	*map/rpsB*	< > >	135	3		
274	457922	190	4	SraA	*clpX/lon*	> < >	120	3		
508	852161	245	4	RybA	*ybiP/ybiQ*	> < >	205	3	S1	no (d)
674	1145858	159	4	SraB	*yceF/yceD*	< > >	149–168	3		
b2911	3053959	302	5	6S (*ssrS*)	*ygfE/ygfA*	> > >	184	3		no (d)
b3123	3267466	803	5	RNase P (*rnpB*)	*yhaC/yhaD*	> < <	377	3		**no**
2513	4275510	548	4	RyjA/SraL	*soxR/yjcD*	> < >	140	3	S1	no (d)
855	1434918	369	1	MicC (ISO63)	*ompN/ydbK*	< > <	100	2		
1579	2689182	497	5	tke1	*yfhK/purL*	< < <	150, 180	2		
1808	3054807	394	1	QUAD1c (RygC)	*ygfA/serA*	> > <	107, 139	2		
1892	3192539	425	4	QUAD1d (tp8)	*yqiK/rfaE*	> < <	110, 145	2		
1921	3235936	286	4	SraF/IS160/tpk1	*ygjR/ygjT*	> > >	189	2		
64	122854	163		tp2	*pdhR/aceE*	> < >	60, 120	1		
834	1403671	332	1	IS061	*abgR/ydaL*	> < <	85	1		
1055	1762411	550	4	RydB/tpe7	*sufA/ydiH*	< < <	60	1		**no**
1057	1768208	404	5	RprA	*ydiK/ydiL*	> > >	105	1		yes (d)
1157	1920997	395	4	RyeA/SraC	*pphA/yebY*	< > <	275	1		partial (d)
1193	1985803	442	1	IS092	*yecJ/yecR*	< < >	150	1		**no**
1254	2069231	174	1	IS102	*yeeP/flu*	> > >	180	1		
1289	2151151	740	4	QUAD1a (RyeC)	*yegL/yegM*	< > >	143	1		
1289	2151151	740	4	QUAD1b (RyeD)	*yegL/yegM*	< > >	139	1		
1473	2494583	497	4	tpke70	*ddg/yfdZ*	> < <	40	1		
1560	2651357	823	4	PAIR3 (RyfA/tp1)	*sseA/sseB*	> > <	300	1		
1968	3308879	180	4	SraG	*pnp/rpsO*	< > <	146–174	1		
2070	3483456	301	5	crpT	*yhfA/crp*	< < >	300	1		

The definitions of the ig (Intergenic regions) and b numbers (genes) are found at http://arep.med.harvard.edu/ExpressDB/EDS37/GAPS_webpages/GAPS_main.htm. Start gives the nucleotide address of the beginning of the intergenic region. Cons. gives the conservation score previously assigned to intergenic regions of >180 nt (Wassarman *et al*., 2001); 5 was used for a previously identified small RNA, 4 is high conservation for >80 nt between the Ig region in *E. coli* and that in *Salmonella* and/or *Klebsiella*; 3 is high conservation over 60–80 nt; 2 is moderate conservation over >65 nt; 1 indicates little or no conservation. All published names of the small RNAs are given, as are both flanking genes and the orientations of the flanking genes and the small RNA (> = clockwise and < = counterclockwise). The size of the RNA observed on Northern blots also is given. The signal on the microarray experiment was scored as described in *Experimental procedures*, and experiments giving the highest signal (≥10 000) are listed. The results from other studies [(a): (Zhang *et al*., 2002) (b): (Møller *et al*., 2002) (c): (Sledjeski *et al*., 2001) (d): (Wassarman *et al*., 2001)], as well as from analyses done in this work, are listed under Hfq binding confirmed. Information on function and characteristics of RNAs comes from multiple sources and is summarized in (Wassarman *et al*., 1999; Wassarman, 2002; Hershberg *et al*., 2003).

data. Only three microarray probes assay the RprA transcript. Both the 5′ and 3′ probes showed a high signal, but no data was available for the central probe. Because of the lack of high signal for two adjacent probes, the region was given a low rating. Thus one potential problem for the detection of small RNAs that are complementary to only a small number of probes on the microarrays is the loss of signal for one probe. Although this is not usually a serious problem for longer transcripts, this may cause a particularly short RNA to be missed. Modifica-

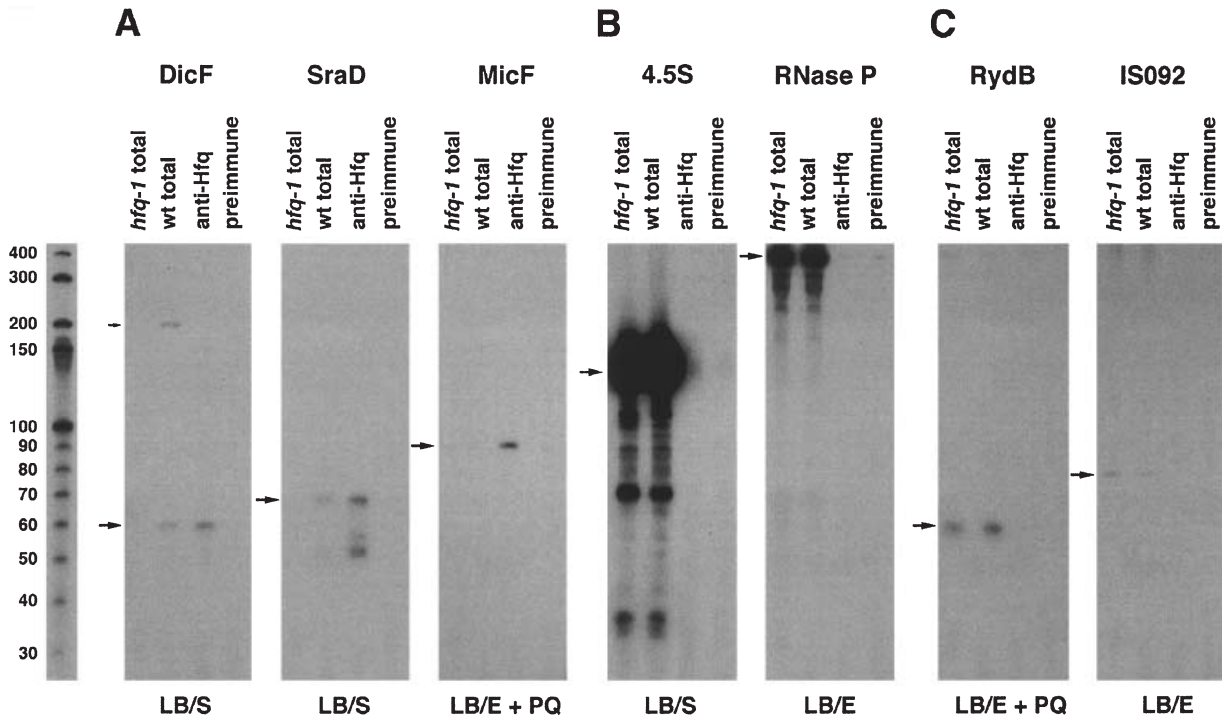

Fig. 1. Hfq binding to known small RNAs. Cell extracts were prepared from MC4100 or MC4100 *hfq-1* cells collected under three different growth conditions: exponential growth in LB medium (LB/E), exponential growth in LB medium treated with 0.5 mM paraquat for 20 min (LB/E + PQ), known to induce MicF RNA (Chou *et al*., 1993), and stationary phase in LB medium (LB/S). Immunoprecipitations were carried out on the MC4100 extracts with Hfq antiserum or preimmune serum and compared to total RNA from 1/10 extract equivalents of the wild-type and *hfq-1* mutant strains. RNAs were fractionated on 8% polyacrylamide urea gels and analysed by Northern hybridization using probes specific to the indicated RNAs. The bands corresponding to the small RNAs are denoted by the large arrows, and a DicF RNA precursor is indicated by the small arrow. RNA molecular weight markers prepared as described in *Experimental procedures* were run with each set of samples. One lane of markers is shown for comparison.
A. Small RNAs exhibiting specific binding to Hfq (DicF, SraD, MicF).
B. Abundant small RNAs showing no specific binding to Hfq (4.5S, RNase P).
C. Low abundance small RNAs showing no binding to Hfq (RydB, IS092). Note that the RNA in IS092 is about 85 nt, smaller than previously described (Chen *et al*., 2002).

tions in the data analysis to take into account such a situation (two positive probes bracketing a single negative or missing probe) should increase the sensitivity of our RNA detection, but also would be expected to increase the number of false positives.

The sensitivity of the microarray detection of co-immunoprecipitated RNAs is surprisingly strong; it enriches for Hfq-binding small RNAs that are expressed at low levels, allowing their detection. For example, OxyS, an RNA induced to high levels only after oxidative stress, and DsrA, an RNA made preferentially at low temperature (Wassarman, 2002; Massé *et al*., 2003), were not detected in previous microarray analysis (Tjaden *et al*., 2002) and are known to be poorly expressed under the growth conditions used here. However, they both gave strong signals in the co-immunoprecipitation microarray experiments (Table 1).

The advantage of examining the signal from individual probes is illustrated for Spot42 (Fig. 2A) and for the intergenic (Ig) region that contains two small RNAs, RygA and RygB (Fig. 2B). Whereas Spot42 is known to bind Hfq (Møller *et al*., 2002), not every probe covering the *spf* gene gave a signal in the microarray analysis (Fig. 2A), possibly because of poor hybridization such as observed for RprA above. The peak of Hfq binding is for a set of overlapping oligonucleotides. For the RygA and RygB Ig region, there are two separate peaks of Hfq binding, supporting the previous finding that each of these two homologous RNAs individually bind Hfq and are expressed under somewhat different conditions (Fig. 2B). For RygB RNA, whereas three probes are contained within the RNA, only two give a high signal. Thus the requirement for only two adjacent probes showing very high expression increased the sensitivity of detection for this small RNA, as well as others.

Identification of new small RNAs which co-immunoprecipitate with Hfq

Given the high correspondence between Ig regions show-

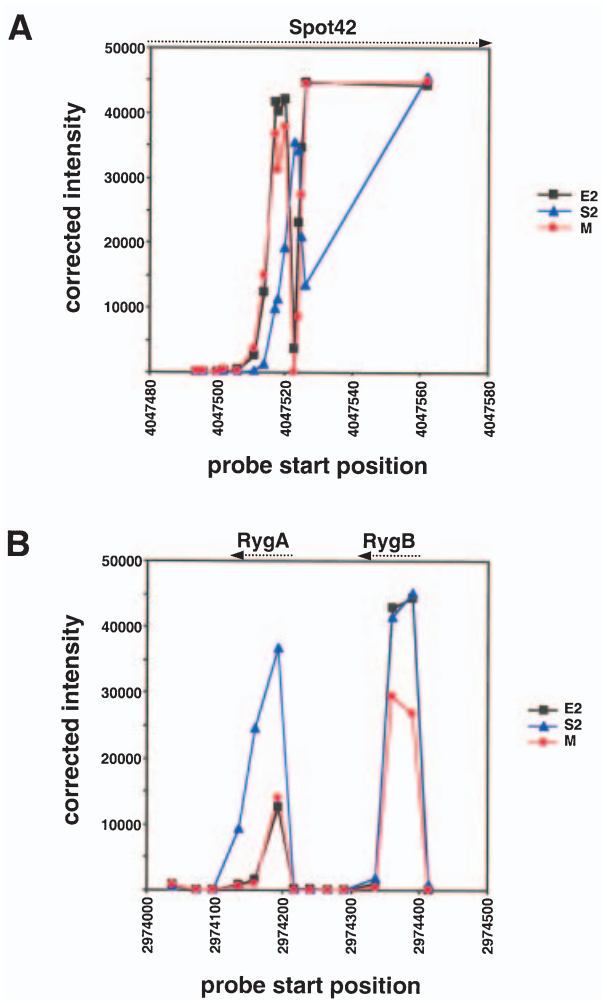

Fig. 2. Microarray signals for regions encoding Spot42 and RygA and RygB RNAs. The signal intensity for each oligonucleotide probe in a given experiment was plotted against the genomic start position of that probe. Control signals from the preimmune samples are not shown, but were low for both Ig regions. E: exponential phase, LB; S: stationary phase, LB; M: exponential phase, minimal media. Dotted lines indicate orientations and positions of genes encoding RNAs.
A. Spot 42 RNA. The closely spaced points indicate that these probes are overlapping, starting just one or two nucleotides apart.
B. Ig 1775 encoding RygA and RygB. The *rygA* (also called *sraE*) and *rygB* genes are in the same, 584 nt intergenic region, in the following order and orientation relative to the neighbouring genes: < *aas* < *rygA* < *rygB* > *galR* (Argaman *et al*., 2001; Wassarman *et al*., 2001).

ing a strong signal and small RNAs known to be bound by Hfq encoded within these regions, we next examined the possibility that our microarray analysis of Hfq co-immunoprecipitated RNAs could be used to predict the existence of previously undiscovered Hfq-binding small RNAs. The microarray data set was divided to specifically examine ratings for the non-protein coding regions of the genome (Ig regions) that were greater than 100 nucleotides in length. Seventy-six such Ig regions were rated as 5. We then discarded candidates in which the signal was likely to correspond to an Ig region encompassed within an operon, a 5′ untranslated region of an mRNA, or, in a few cases, a 3′ untranslated region. In all cases, these assignments were done by comparing the location and orientation of the positive probes in the Hfq immunoprecipitation, relative to flanking genes, as well as previous transcriptome analysis of *E. coli* (Tjaden *et al*., 2002). In the transcriptome analysis, total RNA isolated from *E. coli* grown under 13 different conditions was analysed for co-expression of the Ig regions with flanking genes. From this analysis, likely 5′ UTRs were identified for many genes with the position of the 5′ end predicted from those probes that showed co-expression with the flanking gene. We found that a number of these previously detected 5′ and 3′ UTRs gave strong signals in our Hfq co-immunoprecipitation, suggesting that they may be target mRNAs (see below). These candidates were not further considered here, although it is certainly possible that in some cases, a small RNA, either synthesized under the same conditions as the adjacent gene or processed from an mRNA, may be present. One example of such a case was found (see below, Ig1179).

After this analysis, 20 Ig regions which were rated 5 remained (Table 2). One of these (Ig957) was among the regions previously tested for small RNAs and was found to encode a 31 amino acid peptide, *yneM* (Wassarman *et al*., 2001), but on the opposite strand to that on which we detected Hfq binding. Only two of the 19 remaining Ig regions met the two major criteria for our previous search for small RNAs [Ig size of ≥180 nt and high conservation (rating of 4 in Table 2, conservation column)] (Wassarman *et al*., 2001). However, four of the regions (Ig45, Ig309, Ig877 and Ig1205) had previously been proposed to encode small RNAs; one on the basis of putative promoter and terminator sequences (S. Chen, pers. comm.) and three on the basis of detection of transcription in the intergenic region in the transcriptome analysis (Tjaden *et al*., 2002). An arbitrary sampling of nine of the candidate regions was tested by Northern analysis. Four regions (Ig45, Ig309, Ig453 and Ig877) were found to express small transcripts that were co-immunoprecipitated with Hfq (Fig. 3 and Table 2). For one region (Ig2665), no signal was detected in the Hfq immunoprecipitation sample though the levels of this small RNA were decreased in the strain lacking Hfq, consistent with Hfq affecting its stability (Fig. 3), as it does for many previously described small RNAs (see, for instance, DicF and SraD in Fig. 1). Possibly only a small proportion of this RNA binds to Hfq at any one time, making it difficult to detect. For the remaining four regions (Ig388, Ig496, Ig1085 and Ig1205), no transcripts were detected even in the whole cell extracts (data not shown). Possibly small RNAs encoded by these latter four regions are only expressed under very specific conditions or accumulate

Table 2. Ig regions showing highest Hfq co-immunoprecipitation signal and characteristics of small RNAs.

Ig[a]	Start	Length	Cons	Name	Flanking genes	Strand	Score	Microarray experiments ≥10 000	Northern
45	77300	321	1	RyaA	*yabN/yabM*	< > >	5	E1, E2, S1, S2, M	sRNA
309	506307	203	1	RybC	*ybaK/ybaP*	< > <	5	E1, E2, S1, S2, M	sRNA
388	643191	229	1		*rnk/rna*	< < <	5	S1, S2, M	No transcript
453	764272	103	low	RybD	*sucD/farR*	> > <	5	E1, E2, S1, S2, M	sRNA
470	785908	157	low		*aroG/gpmA*	> > <	5	M	
496	836657	228	2		*ybiC/ybiJ*	> < <	5	E1, E2, S1, S2, M	No transcript
877	1489454	244	1	RydC	*cybB/ydcA*	> < >	5	E1, E2, S2	sRNA
957	1620541	440	4[b]		*ydeE/ydeH*	> < <	5	E1M	
959	1622522	275	1		*ydeI/ydeJ*	< > >	5	S1, S2	
973	1636692	362	2		*cspI/ydfP*	< > <	5	S1, S2	
1085	1808072	150	some		*yniC/ydjM*	> < >	5	E1, E2, S1, S2	No transcript
1140	1894770	183	4		*yeaB/sdaA*	> > >	5	E1, S1, S2	
1179	1956157	390	3	RyeF	*yecK/cutC*	< < <	5	S1, S2, M	sRNA
1205	1994856	229	1		*sdiA/yecC*	< > <	5	M	No transcript
1456	2468480	298	1		*yfdI/tfaS*	> < >	5	E1, E2, S1, S2, M	
1598	2723766	322	low		*kgtP/rrfG*	< > <	5	S1, S2	
1658	2798494	247	4		*ygaM/nrdH*	> < >	5	S1, S2, M	
2335	3979885	102	high		*yifK/argX*	> > >	5	S1, S2	
2408	4103900	148	high		*cpxP/yiiP*	> > >	5	E1, E2, S2, M	
2451	4173409	114	high		*thrT/tufB*	> > >	5	E1, E2	
2665	4525548	130	some	RyjB	*sgcA/sgcQ*	< > <	5	E2, S1, S2, M	sRNA[a]

a. This small RNA was not detected in the Hfq-immunoprecipitate by Northern analysis, but is less abundant in an *hfq-1* mutant (see Fig. 3 and text).
b. This Ig region encodes an mRNA, on the opposite strand to that detected to give Hfq binding (Wassarman *et al*., 2001).
All Ig regions of greater than 100 nt rated 5 and not consistent with an operon, a 5' untranslated region (UTR) or a 3' UTR are shown. Column designations as for Table 1, except results of Northern blots indicated in Northern column, in place of Hfq binding confirmed column. Conservation is as for (Wassarman *et al*., 2001) for Igs of ≥180 nt and is described in the legend to Table 1; shorter Igs were rated roughly as high (well conserved in *Salmonella*), some (short region conserved) or low. Names for detected new small RNAs of unknown function use the same convention as in (Wassarman *et al*., 2001) (for *ryxZ*, x indicates 10' region of the chromosomal map in which small RNA is found (a:0–9'; b:10–19'; c:20–29, etc.), and z is a letter assigned for each small RNA in that 10' interval). This is similar to the nomenclature for unknown open reading frames.

to very low levels, sufficient to be detected by the very sensitive microarray approach, but not by our Northern analysis under limited growth conditions, as was observed for the OxyS and DsrA RNAs. One other region (Ig1179) with a rating of 5 that showed features of a 3′ untranslated region also was found to express a distinct small RNA bound by Hfq, possibly processed from a larger RNA (Fig. 3). This Ig region was added to Table 2 to allow comparison with the others.

Our data suggest the existence of a number of additional Hfq-binding small RNAs in *E. coli*, many present in non-conserved regions of the chromosome. The relatively high yield of new small RNAs from this approach (five small RNAs detected out of 10 regions tested under a limited number of expression conditions) also supports co-immunoprecipitation as a viable approach for detecting small RNA partners for Hfq in other organisms and, more generally, other families of RNAs that bind a common protein.

mRNAs also co-immunoprecipitate with Hfq

In addition to signals from small RNAs, a number of selected RNAs corresponding to single genes and operons were significantly enriched by Hfq co-immunoprecipitation (Table 3). In general, we found almost no case of a signal mapping to within a protein-coding gene that was as strong as the signal seen for Ig regions encoding small RNAs. In addition, there was more variability from experiment to experiment for many of these mRNAs. This may reflect more transient binding of Hfq to mRNA targets, lower mRNA expression levels, and/or greater instability of longer messages during the time-frame required for the immunoprecipitation procedure. In general, whereas the genes represented 46% of the total signals at the lowest Hfq binding level (≥200), they represented only 28% at the highest level (≥10 000).

Some of the positive gene signals correspond to known Hfq targets such as *ompA* (rated as 4, and therefore not in Table 3) or known small RNA targets such as *nlpD* (encoded on the same transcript as *rpoS*) (Sledjeski *et al*., 1996; Brown and Elliott, 1997; Majdalani *et al*., 1998; Vytvytska *et al*., 1998). Other known small RNA targets, however, were either not detected, or detected only at lower expression levels. For 60 genes with a score of 4 or 5, approximately half showed reasonably high signals from the associated Ig region. Table 3 gives all genes with a rating of 5 and indicates those with flanking Ig regions that also showed strong Hfq binding signals.

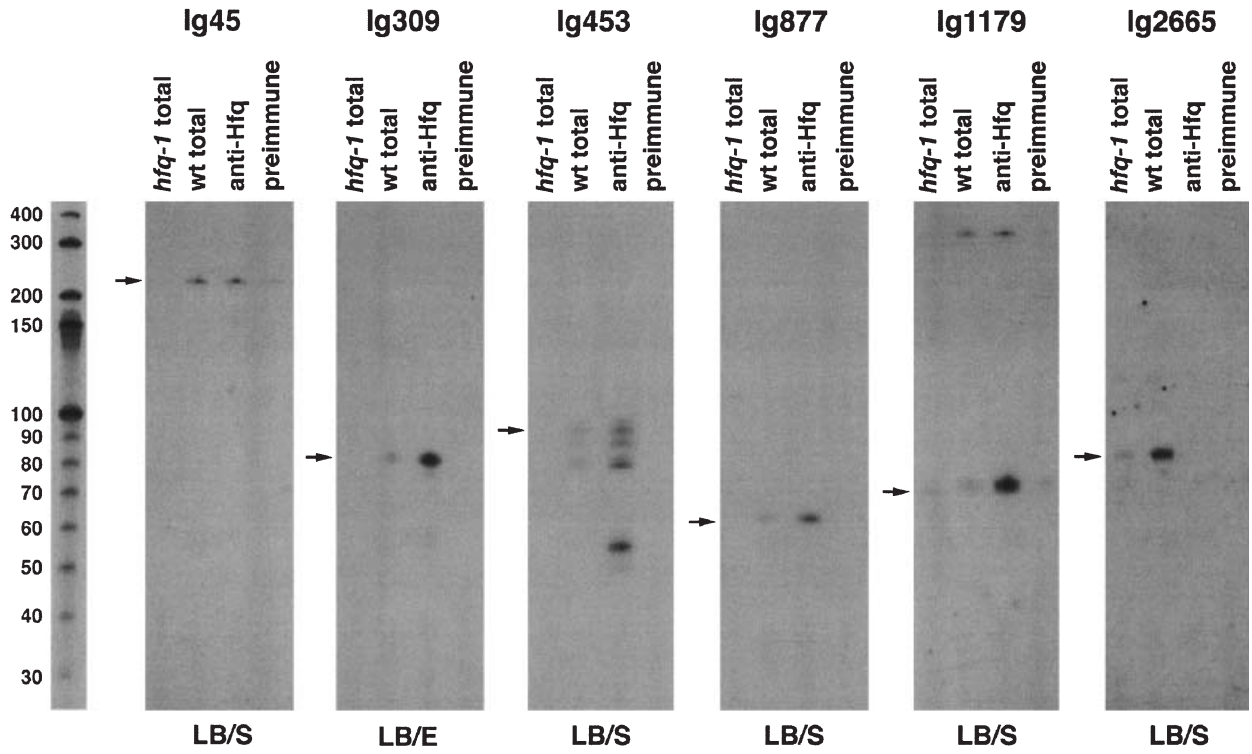

Fig. 3. Hfq binding to small RNAs encoded in Ig regions. Extracts were prepared from MC4100 or MC4100 *hfq-1* cells grown in LB to exponential phase (LB/E) or stationary phase (LB/S), and immunoprecipitations and Northern blots were carried out as described in Fig. 1. Probes specific for the Ig regions shown, for the strand showing binding in the microarray experiments, were generated as described in *Experimental procedures*.

A few genes and operons giving high signals in the microarray analysis were examined further by primer extension and Northern blot analysis (Fig. 4). Because of the unstable nature of many *E. coli* mRNAs, especially in the immunoprecipitated extracts, we chose to test Hfq binding to *ompA* and *dps* by a more sensitive primer extension assay. Both full-length RNAs were found to immunoprecipitate with the Hfq-specific serum but not the preimmune serum. However, the efficiency of selection was much lower compared to what we routinely observe for the small RNAs. This is probably due in part to the less stable nature of the mRNAs during the time required for the immunoprecipitation. It also is possible that only a small proportion of the full-length message targets are bound by Hfq at any one time, while it appears that essentially all of a small RNA population is stably bound. The *hfq-1* mutant strain had higher levels of the *ompA* transcript than the wild-type strain, in agreement with previous studies that reported Hfq promotes *ompA* instability (Vytvytska *et al*., 1998; 2000). For *dps*, we found the most abundant Hfq-binding species was approximately 30 nt shorter than the full-length *dps* mRNA (Fig. 4A). Although this shorter product could be derived from a second promoter or from a degradation intermediate, we favour the possibility that the enriched product is a processing or degradation intermediate because of the role of Hfq in tRNA processing described below. We also suggest that processing products or intermediates of some other mRNAs or even small RNAs may be the primary species bound by Hfq.

Several mRNAs found to bind to Hfq come from complex operons with known post-transcriptional regulation (see *rplC* and *rplL* in Table 3). The operon containing *rpsJ* and *rplC*, which showed multiple Hfq-bound fragments including those of some Ig regions, is the ribosomal operon that encodes S10, L4, and a number of other ribosomal proteins. This operon is autoregulated by binding of L4 to the 5′ UTR that results in transcription termination and translational repression (Zengel and Lindahl, 1996). However, the Hfq-bound fragments with the strongest signal were far from this known regulatory region and might suggest other control mechanisms. The long transcript containing *rplL* also encodes L7/L12, which regulates this operon. Northern analysis of these two ribosomal protein operons revealed two types of Hfq effects (Fig. 4B). First, the amount of mRNA recovered was significantly decreased in the *hfq-1* mutant, consistent with Hfq binding inhibiting the degradation of the message. Second, some degradation intermediates did not accumulate in the Hfq immunoprecipitate, possibly because Hfq inhibits cutting at specific sites or because Hfq only binds a subset of the intermediates.

Table 3. Genes/operons showing highest co-immunoprecipitation signal.

b#	Start	Gene name	Protein size	Score	Microarray experiments gene ≥10 000	Microarray experiments 5′ Ig ≥5000	Microarray experiments 3′ Ig ≥5000	Hfq binding confirmed
b0001	190	*thrL*	21 aa	5	E1; E2; S2; M	E1; E2; S1; S2; M		
b0623	656515	*cspE*	69 aa	5	E1; S1; S2		E1; E2; S1; S2; M	
b0812	847631	*dps*	167 aa	5	S1; S2	S1; S2; M		yes
b0814	849673	*ompX*	171 aa	5	E1; S1; S2	E2; S1; S2		
b0880	921589	*cspD*	74 aa	5	S1; S2	E1; E2; S1; S2; M		
b0881	922136	*clpS*	106 aa	5	E1; S1; S2; M			
b0953	1014938	*rmf*	55 aa	5	S1; S2			
b1051	1113030	*msyB*	124 aa	5	S1; S2			
b1061	1120465	*dinI*	81 aa	5	E1; E2; S1	E1; E2; M		
b1142	1200292	*ymfH*	103 aa	5	S1; S2			
b1178	1226294	ycgK	133 aa	5	M			
b1205	1258014	*ychH*	92 aa	5	S1; S2	S2		
b1259	1313880	*yciG*	59 aa	5	S1; S2			
b1283	1341134	*osmB*	72 aa	5	S1; S2			
b1480	1553850	*sra*	45 aa	5	S1; S2	S2	E1, E2, M	
b1677	1755445	*lpp*	78 aa	5	E1; E2; S1; S2; M	S1		
b1957	2026210	*yodC*	60 aa	5	S1; S2			
b2266	2378742	*elaB*	101 aa	5	S1; S2			
b2597	2735174	*yfiA*	113 aa	5	E1; S1; S2	E2; S1; S2; M	E1; S1; S2	
b2672	2798155	*ygaM*	113 aa	5	S1; S2			
b2695	2816575	*serV*	tRNA	5	E2; S1; S2; M	E2		
b2742	2865637	*nlpD*	379 aa	5	E1; E2; S1; S2; M		E1; E2; S1; S2; M	yes
b2910	3053632	*ygfE*	109 aa	5	S1; S2		E1; S1; S2	
b3049	3189755	*glgS*	66 aa	5	S1; S2			
b3239	3383492	*yhcO*	90 aa	5	S1; S2	S1; S2		
b3320	3449934	*rplC*	209 aa	5	E1; E2	E1; E2	E1	yes
b3509	3653596	*hdeB*	108 aa	5	S1; S2	S1; S2	S2	
b3510	3654038	*hdeA*	110 aa	5	S1; S2		S1; S2	
b3555	3717107	*yiaG*	96 aa	5	S1; S2	S1; S2; M	E1; E2; S1; S2	
b3766	3947945	*ilvL*	32 aa	5	E1; M	E1; S2	S2	
b3781	3963376	*trxA*	109 aa	5	S1; S2; M	S1		
b3799	3980438	*proM*	tRNA	5	E1; S1; S2; M	E1; S1		yes
b3914	4103401	*cpxP*	166 aa	5	S1; S2		E1; E2; S2; M	
b3976	4172967	*thrU*	tRNA	5	E2; S1; S2			
b3986	4178138	*rplL*	121 aa	5	E1; E2; S1; S2; M			yes
b4045	4256816	*yjbJ*	69 aa	5	S1; S2	S1	S1; S2	
b4217	4437164	*ytfK*	68 aa	5	E1; S1; S2	E1, S1	E1; S1; S2	

Columns are titled as for Table 1; only genes showing ratings of 5 are shown. If the microarray signal from the flanking 5′- and 3′ intergenic regions, on the same strand as the gene, were ≥5000, the experiment(s) in which this was detected are indicated.

A role for Hfq in tRNA processing?

We noted that one tRNA operon, encoding the *argX*, *hisR*, *leuT* and *proM* tRNAs, gave multiple high signals for Hfq co-immunoprecipitation, especially from the Ig region (Ig2337) near the end of the operon. Northern analysis with a *proM*-specific probe revealed that certain precursors (marked by the asterisks), but not the mature tRNA product (heavy band), are enriched in the Hfq co-immunoprecipitated samples (Fig. 4B). One product (indicated by **) also is under-represented in the *hfq-1* mutant RNA sample. These results suggest a role for Hfq in processing of certain tRNA operons.

The sizes and pattern of the precursors detected in Fig. 4B were reminiscent of the known pattern of processing intermediates within this operon [summarized in Fig. 5A (Ow and Kushner, 2002; Li and Deutscher, 2002)] indicating that the enriched products might reflect incomplete RNase E processing intermediates. To test this possibility, cells encoding a temperature-sensitive RNase E (*rne-50*) were grown at the permissive temperature to mid-exponential phase; the temperature then was raised to the non-permissive 43°C for 15 min, and the RNA co-immunoprecipitated with Hfq under these different conditions was probed for *proM* (Fig. 5B). We note a great increase in partial processing intermediates in the RNase E mutant at the non-permissive temperature. These intermediates also are enriched in the Hfq immunoprecipitate from this strain, consistent with Hfq binding to intermediates not yet cut by RNase E (Fig. 5B). Finally, the pattern of *proM*-hybridizing bands was compared among wild-type cells, an *rne-50* mutant, an *hfq-1* mutant and an *rne-50 hfq-1* double mutant (Fig. 5C). Higher levels of intermediates are detected in both the *rne-50* single and *rne-50 hfq-1* double mutant strains at 43°C, indicating that, in the absence of RNase E, intermediates accumulate. How-

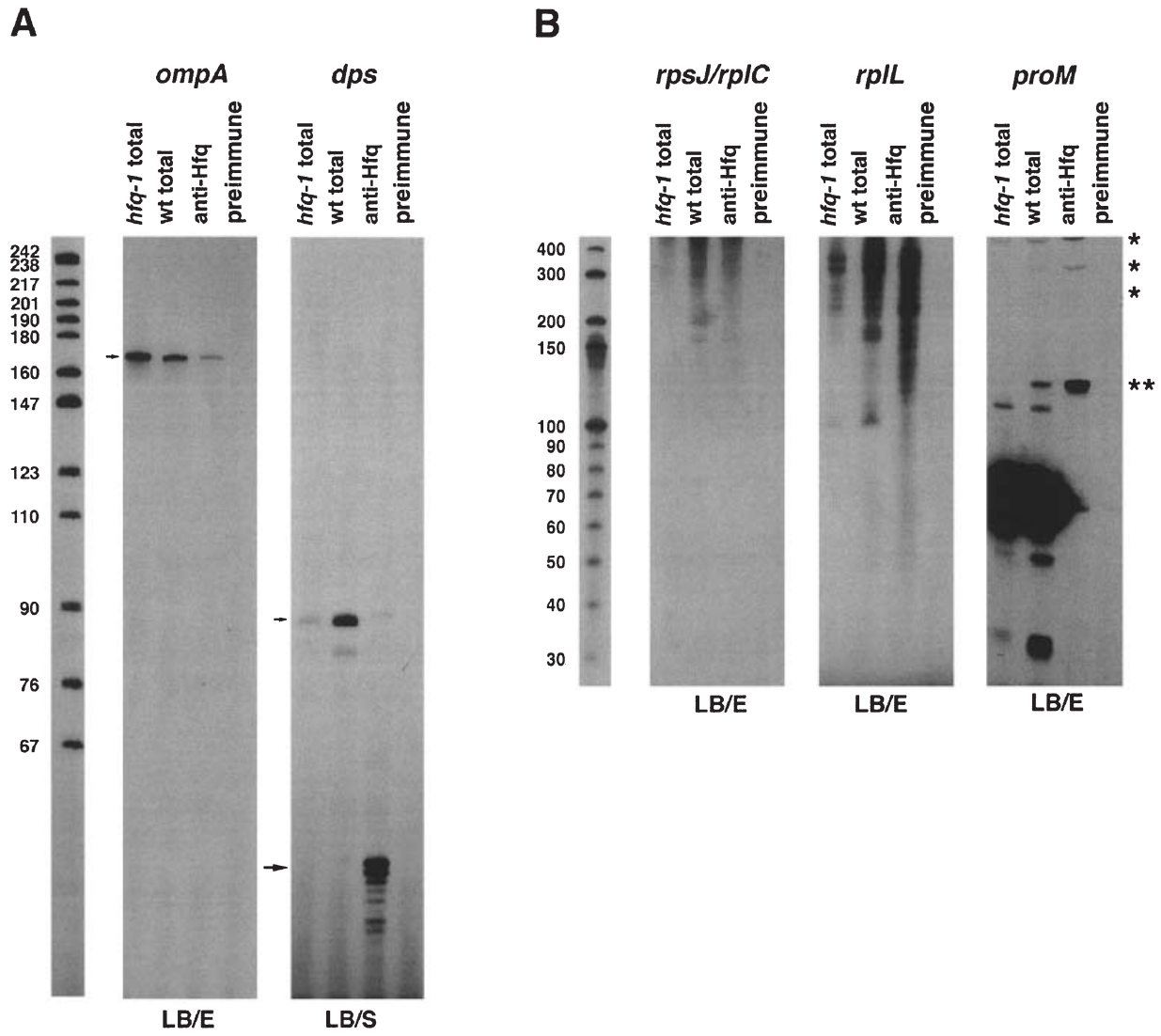

Fig. 4. Hfq binding to mRNAs and tRNAs. Extracts were prepared from MC4100 or MC4100 *hfq-1* cells grown in LB to exponential phase (LB/E) or stationary phase (LB/S) and immunoprecipitations were carried out as described in Fig. 1.
A. Single gene operons showing Hfq binding. Total and immunoprecipitated RNA samples were analysed for the *ompA* and *dps* mRNAs by primer extension analyses. Sizes of primer extension products estimated from the DNA size markers (left lane). Small arrows indicate bands consistent with sizes expected from known promoters. Large arrow indicates shorter primer extension product observed in immunoprecipitations from stationary phase samples.
B. Multigene operons showing Hfq binding. Total and immunoprecipitated RNA samples were analysed for *rpsJ/rplC*, *rplL* and *proM* RNAs by Northern blots using probes covering the Ig region between *rpsJ* and *rplC* as well as part of each gene, the entire *rplL* gene, and *proM* together with the preceding Ig and 3′ end of transcript. Processing intermediates for the *proM* tRNA that are immunoprecipitated with Hfq are indicated by asterisks. Full-length transcripts for the larger (*rpsJ/rplC*, *rplL*) operons were not resolved on our gels.

ever, in the *hfq-1* mutant, we no longer observe the most prominent intermediate detected in the wild-type strain (indicated by **), and the levels of all the intermediates in the *rne-50* single mutant are decreased in the *rne-50 hfq-1* double mutant strain. These patterns are consistent with Hfq binding to the transcript for the operon slowing cleavage by RNase E and perhaps other nucleases; in the absence of Hfq, this cleavage occurs more rapidly.

We note that two other tRNA genes, *serV* and *thrU*, were rated as 5 in our microarray assay (Table 3); we have not looked in more detail at these genes. The processing of the two tRNAs may be impacted by Hfq, although, for these tRNAs, high signals also were detected in the pre-immune samples, raising the possibility they may be false positives. For other tRNAs not detected in our microarray assay, RNase E cleavage may be rapid enough to efficiently outcompete Hfq binding or these precursors may not be bound by Hfq.

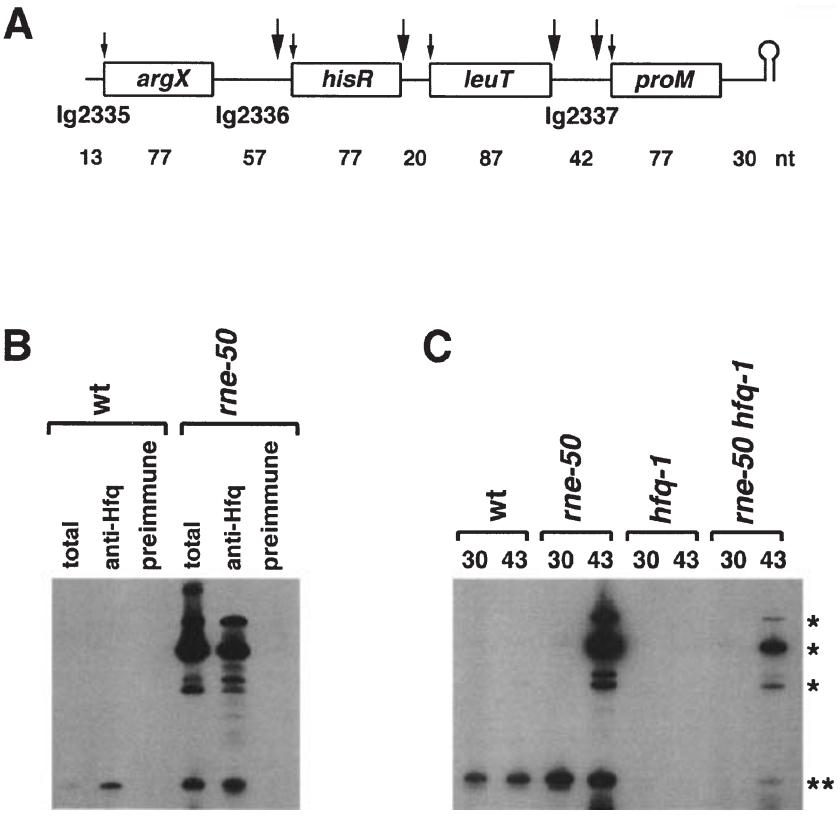

Fig. 5. Role of Hfq in processing of *argX-hisR-leuT-proM* transcript.
A. Schematic of *argX-hisR-leuT-proM* and processing sites. Genes are indicated by boxes. The sizes of the genes and Ig regions are given below. Cleavage sites for RNase E (Ow and Kushner, 2002; Li and Deutscher, 2002) are denoted by heavy arrows; cleavage sites for RNAse P are denoted by small arrows. Ig2337 also was found to bind Hfq.
B. *proM* transcripts immunoprecipitated from *rne*+ and *rne-50* strains. Extracts were prepared from *rne*+ and *rne-50* derivatives of MC4100 grown to mid-exponential phase at low temperature (30°C) and then shifted to the non-permissive temperature (43°C) for 15 min. Immunoprecipitations were carried out as described in Fig. 1.
C. *proM* transcripts in *rne*+, *rne-50*, *hfq-1* and *rne-50 hfq-1* strains. Cultures of *rne*+, *rne-50*, *hfq-1* and *rne-50 hfq-1* derivatives of MC4100 were grown to mid-exponential phase at 30°C. The cultures were then split; half were maintained at 30°C while the other half was shifted to 43°C for 15 min. Total RNA was extracted from all four strains grown at each temperature. For both B and C, Northern blots were carried out as described in Fig. 1, and all samples were probed with the *proM* probe described for Fig. 4B. The bottom of the gel, containing the very abundant mature tRNA, has been cut off to allow easier visualization of the precursor bands. The bands marked with asterisks are the same size as those marked in Fig. 4B.

Discussion

In this paper, we report that immunoprecipitation of RNAs with Hfq together with the identification of the selected RNAs by direct binding to microarrays is a sensitive and effective method for detecting small RNAs bound by Hfq. The experiments here show that at least 15 of the previously known *E. coli* small RNAs fall into this class of RNAs (Table 1). Five new small RNAs were identified using this approach, and we predict that at least 10 additional Ig regions showing strong signals on the DNA microarrays will encode small RNAs bound by Hfq. The sensitivity of the method is apparent in the ability to detect small RNAs that are poorly expressed under the growth conditions used for the immunoprecipitation, such as the OxyS and DsrA RNAs. Part of the sensitivity undoubtedly is a result of the approach of direct hybridization followed by detection with antibody. Because of the small size and structured nature of the regulatory RNAs, cDNA synthesis with random primers is likely to have led to underrepresentation of small RNAs in previous microarray experiments. In addition, the direct hybridization approach allowed the use of very small amounts of selected RNA, facilitating the analysis of immunoprecipitation experiments of this sort.

Can common features be discerned that identify the Hfq-binding small RNAs? The primary characteristic suggested previously by studies of the OxyS, DsrA and Spot42 RNAs, that the Hfq-binding small RNAs function by basepairing with target mRNAs, is reinforced further by our findings. Two additional small RNAs, DicF and MicF, that were found to bind Hfq (Table 1 and Fig. 1) also are known to function by basepairing with their mRNA targets (Andersen and Delihas, 1990; Tetart and Bouché, 1992). For small RNAs in Table 1 whose mode of action is known, those that do not basepair with target mRNAs (4.5S, tmRNA, CsrC, RNase P, and 6S) do not bind Hfq. Some small RNAs that act by basepairing may not bind or use Hfq, but it seems increasingly clear that we can expect the newly identified Hfq-binding RNAs to function by basepairing with target mRNAs. Hfq is a relatively abundant protein, estimated to be up to 30 000–60 000 molecules (or 10 000–20 000 multimers) per cell (Kajitani *et al.*, 1994). It is not clear whether Hfq availability plays a significant role in limiting the functions of bound small RNAs.

Most of the small RNAs identified previously are conserved between *E. coli* K12 and *Salmonella* species, in part because conservation was used to identify many of them (Table 1 and see Hershberg *et al.*, 2003). However, several of the predicted Hfq-binding small RNAs, including some of those confirmed by Northern analysis, are in non-conserved regions of the chromosome, as defined by BLAST analysis of the intergenic region (Table 2). Thus, not

all small RNAs bound by Hfq are conserved. With a few exceptions, the Hfq-binding small RNAs are in the range of 80–110 nucleotides; many, if not all, have well-defined rho-independent terminator sequences. The identification of other common characteristics will require both better definition of the 5′ and 3′ ends and secondary structures of the newly discovered Hfq-binding small RNAs, as well as further analysis of the functions of this set of RNAs.

The microarray approach also allowed us to detect mRNAs that co-immunoprecipitate with Hfq. Previously identified mRNA targets of Hfq (*ompA*) as well as targets for small RNA action (*nlpD*) were detected. These results suggest that other mRNAs showing strong microarray signals in Hfq co-immunoprecipitation (Table 3) also have their stability and/or translation potential affected by Hfq, either directly or indirectly via interaction with a small RNA. Although it is possible some of the signals from apparent mRNAs in Table 3 represent small RNAs that are processed out of larger messages, we suggest most are a result of the enrichment of mRNAs.

Through our initial characterization of this class of Hfq targets, we found some additional, unsuspected roles for Hfq. Most striking was the finding that *proM* tRNA precursors co-immunoprecipitated with Hfq (Figs 4B and 5). These intermediates also were less abundant in the total RNA from an *hfq-1* mutant strain. The pattern suggests that Hfq may act to slow cleavage by RNase E at a number of sites within the operon. This is intriguing because the recognition sites for Hfq and RNase E are very similar [single-stranded poly AU stretches adjacent to a stem-loop, reviewed in (Kennell, 2002; Kushner, 2002)], suggesting the possibility that Hfq and RNase E might compete at some sites. An examination of other tRNA operons under conditions of RNase E limitation and/or mutations affecting other RNases should clarify the universality of Hfq participation in tRNA processing. Because Hfq is dispensible, it is not required for tRNA processing. It remains to be determined whether the Hfq binding to tRNA precursors we observe has a biological consequence. Eukaryotic and archaeal Sm-like proteins have been implicated in tRNA and rRNA processing as well, although it is unclear that the roles in those cases are parallel to what we have described (Toro *et al*., 2001; Kufel *et al*., 2002; 2003). The appearance of ribosomal protein operon mRNAs among those co-immunoprecipitated with Hfq, and the general loss of these messages in an *hfq* mutant, also may reflect competition between Hfq and other processing or degradation systems. However, it is possible that *hfq* mutations have an indirect effect on these operons.

We note that many of the genes detected as strong Hfq co-immunoprecipitation signals encode very short proteins (including two that encode the very short leader peptides *thrL* and *ilvL*). The significance of this observation is not yet clear. If it is a true reflection of Hfq mRNA targets, it might suggest unique regulation of mRNA folding, stability or translation for short genes. For example, Hfq binding to RNAs encoding leader peptides may suggest a role for Hfq in modulating attenuation. However, it is possible that these regions were detected preferentially because the density of probes is higher for a short gene. If Hfq binding protects a portion of an mRNA from degradation, an increase in probe density could increase the chance of detecting the protected fragment with two adjacent probes.

In summary, we have described a sensitive method for detecting Hfq-binding RNAs, in particular small RNAs. This microarray detection of Hfq-immunoprecipitated RNAs should be applicable to any organism in which Hfq exists. The approach also should allow the identification of transcripts bound to other RNA binding proteins, and represents a variation on the ‘ribonomics’ approach to identify the targets of mRNA-binding proteins in eukaryotes (Tenenbaum *et al*., 2002). Using this method we have determined that a large number of the small RNAs in *E. coli* efficiently bind to Hfq. We propose that all these small RNAs act by RNA-RNA basepairing; defining their ‘antisense’ targets is the next challenge.

Experimental procedures

Microarray analysis

Preparation of RNA samples. Cell extracts were prepared as described (Wassarman and Storz, 2000) from MG1655 cells grown in LB medium at 37°C to exponential (OD_{600} = 0.4; experiments E1 and E2) or stationary phase (overnight cultures of 16–18 h; experiments S1 and S2), or in M63 medium supplemented with 0.2% glucose and 0.002% vitamin B1 at 37°C to exponential phase (OD_{600} = 0.4; experiment M). E1 and E2 represent independently grown cultures, independent immunoprecipitations, and independent RNA extractions, as do S1 and S2. Immunoprecipitations were carried out according to Wassarman and Storz (2000) using 20 μl Hfq antiserum or matched preimmune serum (Zhang *et al*., 2002), 15 mg of protein-A-sepharose (Amersham Biosciences, Piscataway, NJ) and 400 μl cell extract per immunoprecipitation reaction. Five parallel immunoprecipitated samples were pooled from each experiment for the microarray analysis. Immunoprecipitated RNA was isolated from immunoprecipitated pellets by extraction with phenol:chloroform:isoamyl alcohol (50:50:1), followed by ethanol precipitation. The RNA pellet was dissolved in DEPC H_2O.

RNA hybridization and detection. High-density oligonucleotide probe arrays on which the complete *Escherichia coli* MG1655 genome sequence is represented were used [*E. coli* Genome Array (sense); Affymetrix, Santa Clara, CA]. Each previously annotated open-reading frame (ORF) (Blattner *et al*., 1997) has 15 oligonucleotide probes designed to be complementary to the sense strand; each lg region greater than 40 bp is interrogated with 15 probes on each of the

strands. The sequences of the oligonucleotide probes and their genomic locations, as well as the assignment of ig and b numbers (see Table 1) are given at http://arep.med.harvard.edu/ExpressDB/EDS37/GAPS_webpages/GAPS_main.htm. The intergenic regions were defined based on the genes originally found by Blattner *et al*. (1997) and therefore usually lie between two ORFs.

The RNA samples were hybridized directly to the high density oligonucleotide arrays without labelling or cDNA synthesis. The RNA samples were added to the hybridization solution [1× MES (100 mM MES, 1 M NaCl, 20 mM EDTA, 0.01% Tween 20, pH 6.6, 0.1 mg ml^{-1} herring sperm DNA, 0.5 mg ml^{-1} BSA and 50 pM of the control Biotin-oligonucleotide B2 (GeneChip® Expression Analysis Technical Manual or at http://www.Affymetrix.com)], heated to 99°C for 5 min and then incubated at 45°C for an additional 5 min before being placed in the microarray cartridge. Hybridization was carried out at 45°C for 16 h on a rotary mixer at 60 r.p.m. Following hybridization, the sample solution was removed and the array was washed as recommended in the technical manual (Affymetrix).

Hybridization was assayed using antibodies specific to the RNA:DNA hybrid and the Digene HC *Express*Array™ Kit (Digene, Gaithersburg, MD). This antibody was developed for experimental research applications and is not a commercial product. Briefly, 25 µl of an RNA:DNA polyclonal antibody (1.3 mg ml^{-1}) was resuspended in 475 µl of matrix solution, loaded on the array and incubated at 25°C for 20 min. After removal of the first antibody stain and 10 wash cycles, the array was incubated with the second antibody stain, containing 0.036 mg ml^{-1} biotin rabbit anti-goat IgG and 0.4 mg ml^{-1} rabbit IgG in 1× MES. RNA:DNA hybrids then were fluorescently labelled by incubating with 10 mg ml^{-1} streptavidin-phycoerythrin (Molecular Probes, Eugene, OR) and 2 mg ml^{-1} BSA in 1× MES. The arrays were read at 570 nm with a resolution of 3 mm using a laser scanner (Affymetrix).

Data analysis

To identify transcripts, we modified an algorithm previously developed for transcriptome analysis (Tjaden *et al*., 2002). Microarray suite 5.0 (MAS 5.0) was used for intensity analysis. All files were normalized to a target intensity of 500 to generate a.*cel* file, which contains the intensity information for each perfect match (PM) and the corresponding mismatch (MM) probe on the array. The PM-MM values for the corresponding control experiment (immunoprecipitated with preimmune serum) were also calculated, and subtracted from the experimental values for each probe. Adjacent probes (two or more probes) in which the PM-MM value, after subtraction of the preimmune value (corrected intensity), was greater than a given threshold were identified and a transcript intensity calculated based on the average of the corrected intensity. As described in (Tjaden *et al*., 2002), bad probes (about 2% of the total) and single low probes surrounded by high signals were excluded from the analysis. In order to optimize sensitivity for detection of short transcripts within the Ig regions and avoid lowering the apparent signal by averaging over large regions, the initial cut-off for average intensity was chosen to be very high (≥10 000); additional analysis was done with lower average intensities, and those for ≥5000 are used in some cases for the analysis below. Experiments in which a particular region (Ig or gene) gave ≥10 000 for duplicates (E1 and E2, S1 and S2) or for the single M experiment were rated 5 in tables; those in which at least duplicate experiments (E1 and E2; S1 and S2) or the single M experiment were ≥5000 were rated 4. Those with expression levels ≥1000 in duplicate experiments were rated 3, those with expression levels ≥200 were rated 2, and those <200 were rated 1. There were more than 8000 entries with expression levels ≥200 in at least one experiment, more than 2000 with expression levels ≥1000, around 600 with expression levels ≥5000, and 335 with expression levels ≥10 000. The complete set of data calculated at each of these expression levels can be found at http://cs.wellesley.edu/~btjaden/sRNAs/.

For many small RNAs rated as 5, there was little or no signal from the preimmune serum immunoprecipitates. Some, such as tmRNA, encoded by *ssrA*, and 4.5S, encoded by *ffs*, had high preimmune signals in the E1 and S1 experiments (for one peak probe, *ffs* gave a signal of 9300 for the preimmune sample and 33 300 for the immune sample; *ssrA* gave a signal of 11 341 for the preimmune sample and 43 000 for the immune sample). It is not yet clear whether the signals from tmRNA and 4.5S represent false positives or transient binding of precursors to Hfq.

Northern analysis

To construct plasmids for generating *in vitro* transcribed RNA probes used for Northern analysis, indicated genes or Ig regions were amplified by PCR from MC4100 genomic DNA using specific primers (see http://dir2.nichd.nih.gov/nichd/cbmb/segr/segrPublications.html for the sequences of all oligonucleotide primers). The PCR fragments then were digested with *Eco*RI and *Hin*dIII, and cloned into the corresponding sites of pGEM2 (Promega, Madison, WI). All DNA manipulations were carried out using standard procedures, and all clones were confirmed by sequencing.

The *E. coli* strain MC4100 *hfq-1*::Ω (GSO81) was described previously (Zhang *et al*., 2002). To examine the effects of the temperature sensitive *rne-50* mutation, *rne*$^+$ and *rne-50* alleles linked to *zce-726*::Tn*10* [EM1272 and EM1274, derived from AC23 (Vanzo *et al*., 1998) by Eric Massé], were moved into MC4100 by P1 transduction, generating GSO88 and GSO89, respectively. Subsequently, the *hfq-1*::Ω mutant allele (Tsui *et al*., 1994) was P1 transduced into both strains generating GSO90 and GSO91. Cell extracts were prepared as described (Wassarman and Storz, 2000) from ~2×10^{10} cells (20 OD_{600}) grown as described in the figure legends. Immunoprecipitations were carried out according to Wassarman and Storz (2000), using 20 µl Hfq antiserum or preimmune serum, 24 mg of protein-A-sepharose and 200 µl extract per immunoprecipitation reaction. RNA was isolated from immunoprecipitated pellets or, for total RNA samples, from 20 µl cell extract, by extraction with phenol:chloroform:isoamyl alcohol (50:50:1), followed by ethanol precipitation. The RNA pellet was dissolved in 10 µl DEPC H_2O. The RNA molecular weight markers ≤150 nucleotides were generated using the Decade Marker System (Ambion, Austin, TX) and end-labelled using T4 polynucleotide kinase and γ^{32}P-ATP. The 100–1000 nucleotide markers were generated and uniformly labelled using T7

RNA polymerase, α-^{32}P-ATP, and the Perfect RNA Marker Template Mix (Novagen, Madison, WI).

The MicF, 4.5S, RNase P, IS092, Ig45, Ig309, Ig388, Ig496, Ig1085, Ig1179, Ig1205, Ig2665, *rpsJ*/*rplC*, *rplL* and *proM* RNAs were analysed by Northern hybridization using RNA probes as described (Wassarman *et al*., 2001). RNA samples (2 μl) were fractionated on 8% polyacrylamide urea gels and transferred to Hybond N membranes (Amersham Pharmacia Biotech, Piscataway, NJ) or to GeneScreen Plus membranes (NEN Life Science Products, Boston, MA) for the *proM* blots in Fig. 4B. Uniformly ^{32}P labelled RNA probes were generated by *in vitro* transcription from plasmids linearized with *Eco*RI using T7 RNA polymerase. Hybond N membranes were prehybridized and hybridized in buffer containing 50% formamide, 1.5× SSPE, 1% SDS, 0.5% dry milk at 50°C or, in the case of Ig2665, in ULTRAhyb buffer (Ambion) at 68°C. GeneScreen Plus membranes were prehybridized and hybridized in Hybrisol I solution (Intergen, Purchase, NY) at 45°C. The DicF, SraD, RyhB, Ig453 and Ig877 RNAs were analysed by Northern hybridization using labelled oligonucleotide probes. RNA samples (2 μl) were fractionated on 8% polyacrylamide urea gels and transferred to Zeta-Probe membranes (Bio-Rad Laboratories, Hercules, CA). Northern membranes were prehybridized and hybridized in ULTRAhyb-Oligo buffer (Ambion) at 45°C. All membranes were washed twice with 2× SSC/0.1% SDS at room temperature followed by two 25 min washes with 0.1× SSC/0.1% SDS at 50°C (or 68°C for Ig2665) for Hybond N membranes and 45°C for GeneScreen Plus and Zeta-Probe membranes.

Primer extension analysis

The *ompA* and *dps* mRNAs were analysed by primer extension analysis using primers specific to the indicated genes as described (Zhang *et al*., 1998). RNA samples (2 μl) were incubated with 0.5 pmol of 5′-^{32}P-end labelled primer for 5 min at 65°C, and then quick-chilled on ice. After the addition of dNTPs (1 mM each) and AMV reverse transcriptase (10 U, Life Sciences, St Petersburg, FL), the reactions were incubated for 1 h at 42°C. The cDNA products then were fractionated on 8% polyacrylamide urea gels. The DNA molecular weight markers were generated by end-labelling a pBR322-Msp I digest (New England Biolabs, Beverly, MA) with T4 kinase.

Acknowledgements

We would like to thank R. Mukherjee Saxena for technical assistance and A. Tracy, G. Miyada and J. Lazar for providing the detection assay protocol, M. Kawano for providing the RNA markers, W. Outten for providing the DNA markers and A. Carpousis, E. Massé and M. Winkler for providing strains. We also thank S. Chen for sharing unpublished data and D. Court, R. Weisberg, D. Fitzgerald, and members of our laboratories for comments on the manuscript.

References

Andersen, J., and Delihas, N. (1990) *micF* RNA binds to the 5′ end of *ompF* mRNA and to a protein from *Escherichia coli*. *Biochemistry* **29:** 9249–9256.

Argaman, L., Hershberg, R., Vogel, J., Bejerano, G., Wagner, E.G.H., Margalit, H., and Altuvia, S. (2001) Novel small RNA-encoding genes in the intergenic region of *Escherichia coli*. *Curr Biol* **11:** 941–950.

Blattner, F.R., Plunkett, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., *et al*. (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* **277:** 1453–1474.

Brescia, C.C., Mikulecky, P.J., Feig, A.L., and Sledjeski, D.D. (2003) Identification of the Hfq-binding site on DsrA RNA: Hfq binds without altering DsrA secondary structure. *RNA* **9:** 33–43.

Brown, L., and Elliott, T. (1996) Efficient translation of the RpoS sigma factor in *Salmonella typhimurium* requires Host Factor I, an RNA-binding protein encoded by the *hfq* gene. *J Bacteriol* **178:** 3763–3770.

Brown, L., and Elliott, T. (1997) Mutations that increase expression of the *rpoS* gene and decrease its dependence on *hfq* function in *Salmonella typhimurium*. *J Bacteriol* **179:** 656–662.

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.

Chou, J.H., Greenberg, J.T., and Demple, B. (1993) Posttranscriptional repression of *Escherichia coli* OmpF protein in response to redox stress: positive control of the *micF* antisense RNA by the *soxRS* locus. *J Bacteriol* **175:** 1026–1031.

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

Gottesman, S. (2002) Stealth regulation: biological circuits with small RNA switches. *Genes Dev* **16:** 2829–2842.

Grosshans, H., and Slack, F.J. (2002) Micro-RNAs: small is plentiful. *J Cell Biol* **156:** 17–21.

Hajndsorf, E., and Regnier, P. (2000) Host factor Hfq of *Escherichia coli* stimulates elongation of poly (A) tails by poly (A) polymerase I. *Proc Natl Acad Sci USA* **97:** 1501–1505.

Hershberg, R., Altuvia, S., and Margalit, H. (2003) A survey of small RNA-encoding genes in *Escherichia coli*. *Nucleic Acids Res* **31:** 1813–1820.

Kajitani, M., Kato, A., Wada, A., Inokuchi, Y., and Ishihama, A. (1994) Regulation of the *Escherichia coli hfq* gene encoding the host factor for phage Q beta. *J Bacteriol* **176:** 531–534.

Kennell, D. (2002) Processing endoribonucleases and mRNA degradation in bacteria. *J Bacteriol* **184:** 4645–4657.

Kufel, J., Allmang, C., Verdone, L., Beggs, J.D., and Tollervey, D. (2002) Lsm proteins are required for normal processing of pre-tRNAs and their efficient association with La-homologous protein Lhp1p. *Mol Cell Biol* **22:** 5248–5256.

Kufel, J., Allemang, C., Petfalski, E., Beggs, J., and Tollervey, D. (2003) Lsm proteins are required for normal processing and stability of ribosomal RNAs. *J Biol Chem* **278:** 2147–2156.

Kushner, S.R. (2002) mRNA decay in *Escherichia coli* comes of age. *J Bacteriol* **184:** 4658–4665.

Li, Z., and Deutscher, M.P. (2002) RNase E plays an essen-

tial role in the maturation of *Escherichia coli* tRNA precursors. *RNA* **8:** 97–109.

Majdalani, N., Cunning, C., Sledjeski, D., Elliott, T., and Gottesman, S. (1998) 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* **95:** 12462–12467.

Massé, 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.

Massé, E., Majdalani, N., and Gottesman, S. (2003) Regulatory roles for small RNAs in bacteria. *Curr Opin Microbiol* **6:** 120–124.

Møller, T., Franch, T., Hojrup, P., Keene, D.R., Bachinger, H.P., Brennan, R., and Valentin-Hansen, P. (2002) Hfq: a bacterial Sm-like protein that mediates RNA–RNA interaction. *Mol Cell* **9:** 23–30.

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

Ow, M.C., and Kushner, S.R. (2002) Initiation of tRNA maturation by RNase E is essential for cell viability in *E. coli*. *Genes Dev* **16:** 1102–1115.

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.

Schumacher, M.A., Pearson, R.F., Møller, 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.

Sledjeski, D.D., Gupta, A., and Gottesman, S. (1996) The small RNA, DsrA, is essential for the low temperature expression of RpoS during exponential growth in *Escherichia coli*. *EMBO J* **15:** 3993–4000.

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

Storz, G. (2002) An expanding universe of noncoding RNAs. *Science* **296:** 1260–1263.

Sun, X., Zhulin, I., and Wartell, R.M. (2002) Predicted structure and phyletic distribution of the RNA-binding protein Hfq. *Nucleic Acids Res* **30:** 3662–3671.

Tenenbaum, S.A., Lager, P.J., Carson, C.C., and Keene, J.D. (2002) Ribonomics: identifying mRNA subsets in mRNP complexes using antibodies to RNA-binding proteins and genomic arrays. *Methods* **26:** 191–198.

Tetart, F., and Bouché, J.P. (1992) Regulation of the expression of the cell-cycle gene *ftsZ* by DicF antisense RNA. Division does not require a fixed number of FtsZ molecules. *Mol Microbiol* **6:** 615–620.

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.

Toro, I., Thore, S., Mayer, C., Basquin, J., Seraphin, B., and Suck, D. (2001) RNA binding in an Sm core domain: X-ray structure and functional analysis of an archaeal Sm protein complex. *EMBO J* **20:** 2293–2303.

Tsui, H.-C.T., Leung, H.-C.E., 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.

Vanzo, N.E., Li, Y.S., Py, B., Blum, E., Higgins, C.F., Raynal, L.C., Krisch, H.M., and Carpousis, A.J. (1998) Ribonuclease E organizes the protein interactions in the *Escherichia coli* RNA degradosome. *Genes Dev* **12:** 2770–2781.

Vytvytska, O., Jakobsen, J., Balcunaite, G., Andersen, J., Baccarini, M., and von Gabain, A. (1998) Host Factor I, Hfq, binds to *Escherichia coli ompA* mRNA in a growth-rate dependent fashion and regulates its stability. *Proc Natl Acad Sci USA* **95:** 14118–14123.

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.

Wassarman, K.M. (2002) Small RNAs in bacteria: diverse regulators of gene expression in response to environmental changes. *Cell* **109:** 141–144.

Wassarman, K.M., and Storz, G. (2000) 6S RNA regulates *E. coli* RNA polymerase activity. *Cell* **101:** 613–623.

Wassarman, K.M., Zhang, A., and Storz, G. (1999) Small RNAs in *Escherichia coli*. *Trends Microbiol* **7:** 37–45.

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.

Zengel, J.M., and Lindahl, L. (1996) A hairpin structure upstream of the terminator hairpin required for ribosomal protein L4-mediated attenuation control of the S10 operon of *Escherichia coli*. *J Bacteriol* **178:** 2383–2387.

Zhang, A., Altuvia, S., Tiwari, A., Argaman, L., Hengge-Aronis, R., and Storz, G. (1998) The *oxyS* regulatory RNA represses *rpoS* translation by binding Hfq (HF-1) protein. *EMBO J* **17:** 6061–6068.

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.