Molecular basis of messenger RNA recognition by the specific bacterial repressing clamp RsmA/CsrA

Mario Schubert[1], Karine Lapouge[2], Olivier Duss[1], Florian C Oberstrass[1], Ilian Jelesarov[3], Dieter Haas[2] & Frédéric H-T Allain[1]

Proteins of the RsmA/CsrA family are global translational regulators in many bacterial species. We have determined the solution structure of a complex formed between the RsmE protein, a member of this family from *Pseudomonas fluorescens*, and a target RNA encompassing the ribosome-binding site of the *hcnA* gene. The RsmE homodimer with its two RNA-binding sites makes optimal contact with an 5′-$^A/_U$CANGGANG$^U/_A$-3′ sequence in the mRNA. When tightly gripped by RsmE, the ANGGAN core folds into a loop, favoring the formation of a 3-base-pair stem by flanking nucleotides. We validated these findings by *in vivo* and *in vitro* mutational analyses. The structure of the complex explains well how, by sequestering the Shine-Dalgarno sequence, the RsmA/CsrA proteins repress translation.

In translational control of bacterial gene expression, the ability of ribosomes to access mRNAs at the ribosome-binding site (RBS) is crucial. Repression of translation initiation occurs when an RBS is occluded either by proteins or by base-pairing RNAs. In one important mechanism, RNA-binding proteins of the regulator of secondary metabolism (RsmA)/carbon storage regulator (CsrA) family bind target mRNAs at the RBS; the resulting translational repression can be relieved by small RNAs having high affinity for RsmA/CsrA proteins. These noncoding RNAs have repeated, unpaired GGA motifs, which are essential for binding and are expressed under the positive control of the GacS/GacA two-component system in many Gram-negative bacteria[1–8].

The GacS/GacA system of *Pseudomonas* spp. globally induces the expression of extracellular secondary metabolites and lytic enzymes, via sequestration of RsmA and related proteins that act as translational repressors. In pathogenic species, exoproducts controlled by GacS and GacA can be virulence factors. In plant-beneficial species with biocontrol properties, such exoproducts contribute to the protection of plants from disease caused by fungi or nematodes. In *Escherichia coli*, the homologous BarA/UvrY system controls central carbon metabolism, the production of a storage compound (glycogen) and the formation of a biofilm polysaccharide, poly(*N*-acetyl-glucosamine). The translational regulator that responds to the BarA/UvrY system in *E. coli* is CsrA[1–4]. Homologs of the *rsmA* and *csrA* genes have been found in more than 150 Gram-negative and Gram-positive species, and some species contain more than one homolog. Recently, crystal and solution structures of the highly similar RsmA/CsrA proteins from *E. coli*, *Pseudomonas aeruginosa* and *Yersinia enterocolitica* have been determined. These proteins are homodimers in which the five β-strands of each subunit are intertwined, with α-helices forming wing-like structures near the C termini[9–11]. Alanine-scanning mutagenesis of *E. coli* CsrA has identified two regions (residues 2–7 and 40–47) that are important for RNA binding *in vivo*[12]. Furthermore, *in vitro* selection experiments indicate that CsrA binds preferentially to RNA hairpin structures with an ANGGAN loop, indicating that both the RNA sequence and the secondary structure are important for CsrA recognition[13]. However, the molecular basis of RNA recognition by RsmA/CsrA proteins remains to be determined.

In the plant-beneficial soil bacterium *Pseudomonas fluorescens* CHA0, the RsmA protein and its homolog RsmE translationally repress the expression of genes encoding biocontrol factors—for example, the *hcnA* gene (encoding hydrogen cyanide synthase subunit A). At high cell population densities, the GacS/GacA system induces the expression of three small RNAs (RsmX, RsmY and RsmZ), which sequester RsmA and RsmE, resulting in *hcnA* expression[14–18]. To aid understanding of how this family of proteins recognizes mRNA, we present the structure of a complex formed between the *P. fluorescens* RsmE protein and its *hcnA* mRNA target sequence overlapping the RBS.

RESULTS

RsmE–*hcnA* RNA complex formation

We chose to study the RsmE protein, which is a homodimer of 64 residues (**Fig. 1a**), because of its favorable solubility properties *in vitro*. As the mRNA target, we chose a 12-nucleotide sequence containing the RBS of the *hcnA* gene (**Fig. 1b**). The free 12-nucleotide *hcnA* sequence did not form a stable stem-loop structure, as only weak imino protons could be observed at 276 K and none at room temperature (**Supplementary Fig. 1** online). In contrast, two imino

[1]Institute of Molecular Biology and Biophysics, ETH Zürich, CH-8093 Zürich, Switzerland. [2]Département de Microbiologie Fondamentale, Université de Lausanne, CH-1015 Lausanne, Switzerland. [3]Biochemisches Institut der Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland. Correspondence should be addressed to D.H. (dieter.haas@unil.ch) or F.H.-T.A. (allain@mol.biol.ethz.ch).
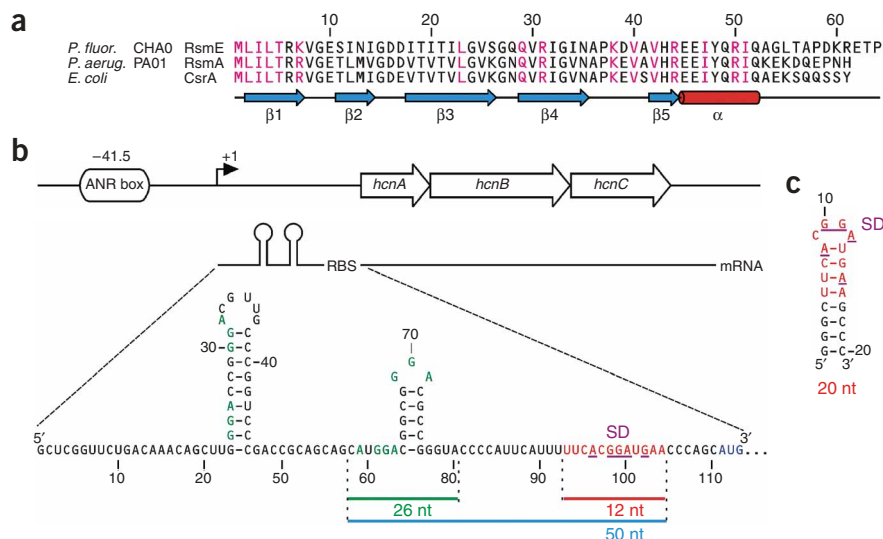
Figure 1 Secondary structure of RsmE and genetic organization of the *hcnA* 5′ untranslated mRNA. (**a**) Alignment of RsmE of *P. fluorescens* CHAO, RsmA of *P. aeruginosa* PAO1 and *E. coli* CsrA sequences. Magenta, residues involved in RNA recognition (ref. 12 and this study). Secondary structures are shown schematically below alignment. (**b**) Transcription of the *P. fluorescens hcnABC* operon is under control of the anaerobic regulator of nitrate respiration and arginine fermentation (ANR) transcription factor, which binds the ANR box[14]. Red, the 12-nucleotide *hcnA* sequence involved in RsmE binding; green, other potential RsmE-binding sites; blue, AUG *hcnA* start codon; underline, Shine-Dalgarno sequence (SD) of the RBS. (**c**) Predicted secondary structure of the 20-nucleotide *hcnA* sequence used for structure determination of the RsmE–RNA complex.

signals (from G11 and G14) appeared upon addition of RsmE protein, suggesting that base pairs are formed in the 12-nucleotide *hcnA* sequence upon RsmE binding (**Supplementary Fig. 1**). The ^{15}N HSQC spectrum of RsmE was altered substantially upon RNA binding (**Supplementary Fig. 2** online). The protein resonances are in slow exchange relative to the NMR timescale, indicating strong binding of RsmE to the 12-nucleotide *hcnA* sequence. To determine the structure of the RsmE–*hcnA* RNA complex by NMR spectroscopy, we extended the 12-nucleotide *hcnA* sequence by four G-C pairs to isotopically label the RNA and to form a stable stem-loop (**Fig. 1c**). The resulting 20-nucleotide *hcnA* sequence has a secondary structure resembling that of RNAs selected as high-affinity ligands for CsrA[13]. The NMR spectra of RsmE bound to the 12-nucleotide *hcnA* sequence and to the 20-nucleotide sequence were almost identical, indicating that the protein recognizes the two RNAs in the same way (**Supplementary Figs. 1** and **2**).
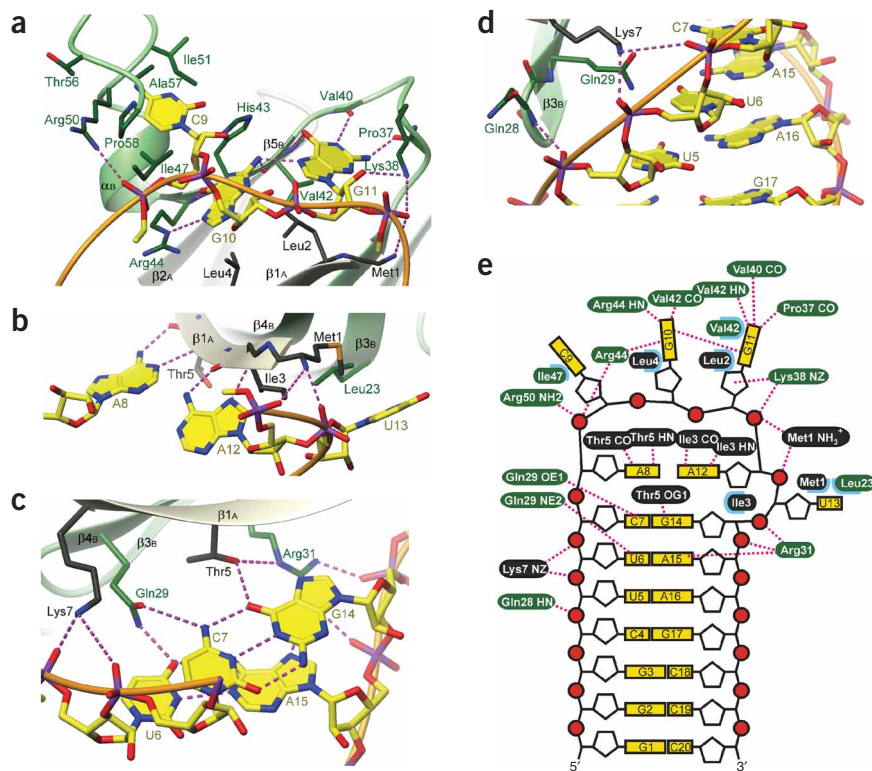
Solution structure of the RsmE–*hcnA* complex

The complex of RsmE with the 20-nucleotide *hcnA* sequence was studied at low salt concentrations and pH 7.2. Under these conditions, RsmE bound the RNA in a 1:1 molar ratio, resulting in two RNA molecules

bound per protein homodimer. Assignments of 132 intermolecular protein-protein NOEs and a measured correlation time τ_C of 11.2 ns at 40 °C support the presence of a 29-kDa homodimeric complex. Using 1,207 NOE-derived distance restraints (including 220 intermolecular RNA-protein restraints per monomer), we determined the solution structure of the RsmE–*hcnA* RNA complex (**Fig. 2**). The protein-RNA interface is well-defined (**Supplementary Figs. 3** and **4** online), allowing us to elucidate the molecular basis of RsmE–*hcnA* mRNA recognition.
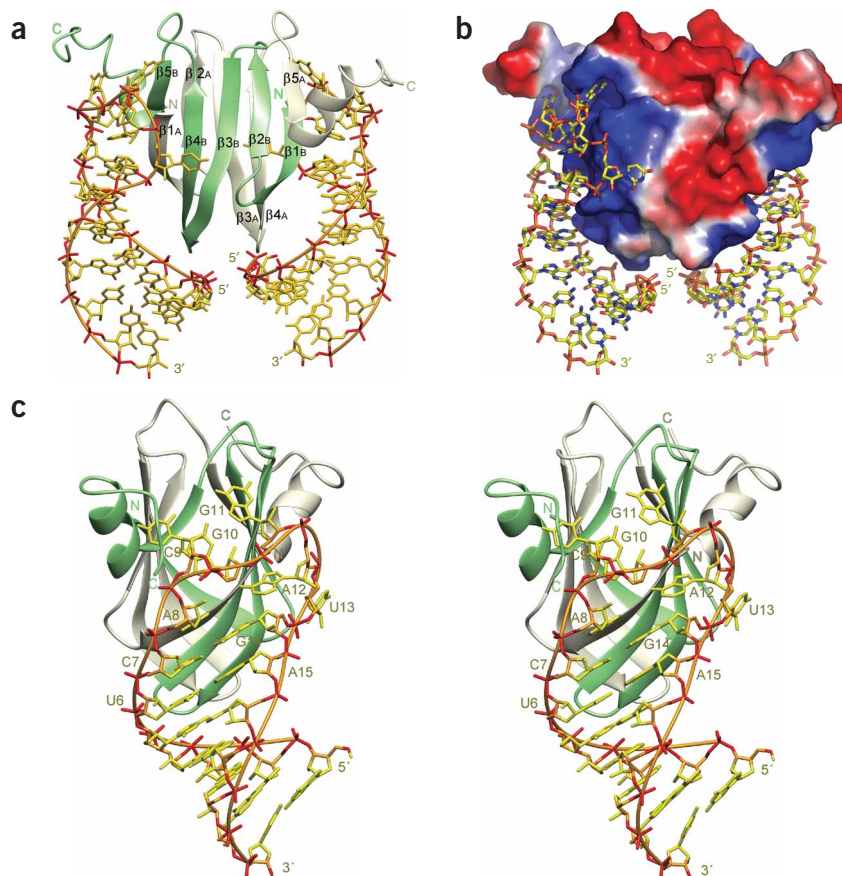
Figure 2 NMR solution structure of the RsmE–*hcnA* RNA complex. (**a**) Solution structure of the 2:2 complex of RsmE with 20-nucleotide *hcnA* sequence. Protein ribbons for each monomer are shown in green and gray. Heavy atoms of the two RNAs are shown in yellow and red. An orange ribbon linking the phosphates is also shown. (**b**) Surface representation of the RsmE dimer in complex, colored by electrostatic potential (blue, positive; red, negative). (**c**) Stereo view of one 20-nucleotide *hcnA* sequence bound to the edge of the RsmE dimer sandwich, omitting the second RNA molecule in the background; a representative structure is shown.

Figure 3 Details of the RsmE–*hcnA* RNA complex NMR structure. (**a**–**d**) Structural details of interactions important for recognition. Black and green, side chains and backbone of RsmE monomers A and B, respectively; magenta dashed lines, possible hydrogen bonds. (**e**) Schematic representation of intermolecular RNA-protein interactions, colored as in **a**–**d**. Cyan, hydrophobic interactions.

The RsmE–*hcnA* RNA complex has C_2 symmetry and consists of a protein dimer with two RNA molecules bound at spatially separated sites (**Fig. 2a**). In the RNA complex, RsmE adopts virtually the same fold as it does in the free form: two subunits (A and B) form two intertwined antiparallel β-sheets ($β1_A$-$β4_B$-$β3_B$-$β2_B$-$β5_A$ and $β1_B$-$β4_A$-$β3_A$-$β2_A$-$β5_B$; throughout, subscript A or B denotes the subunit) facing one another, followed by one α-helix in each subunit[9–11]. The RNAs are bound on a highly positively charged surface formed by the edges of the β-sandwich, the $β1_A$/$β5_B$ and the $β1_B$/$β5_A$ edge, and the region around the β3-β4 and β4-β5 loops (**Fig. 2a,b**). The bound RNAs form the expected stem-loops but not the predicted tetraloops (**Fig. 1c**); instead, the loops contain six unpaired nucleotides (A8 to U13; **Fig. 2c**). The protein dimer interacts with all nucleotides of each hexanucleotide loop as well as with the two C7-G14 and U6-A15 base pairs, on the major-groove side of the base pairs. For simplicity, here we show only the RNA bound to the $β1_A$/$β5_B$ edge of the sandwich to describe the interface (**Fig. 2c**).

RsmE recognizes an $^A/_U$CANGGANG$^U/_A$ consensus sequence

The G10 and G11 bases are packed against the hydrophobic core at the $β1_A$/$β5_B$ edge consisting of the $Leu2_A$, $Leu4_A$ and $Val42_B$ side chains, and the Watson-Crick edges of G10 and G11 are specifically recognized by hydrogen bonds to the backbone of $β5_B$ and the preceding $β4_B$-$β5_B$ loop (**Fig. 3a**). G10 and G11 are coplanar and interact via an intramolecular hydrogen bond between the G10 amino group and the G11 N7 (**Fig. 3a**). A8 and A12 are also coplanar and stack on the C7-G14 base pair closing the stem, but they do not interact with one another (**Fig. 3b**). However, they are specifically recognized through

hydrogen bonds to the backbone of $β1_A$. The $Thr5_A$ main chain recognizes the Watson-Crick edge of A8, whereas the $Ile3_A$ main chain recognizes the Hoogsteen edge of A12 (**Fig. 3b**). C9 and U13 are looped out and are accommodated by two hydrophobic patches, one located at the C terminus of chain B ($Ile47_B$, $Ile51_B$, $Ala57_B$ and $Pro58_B$) and the other around the N terminus of chain A ($Met1_A$, $Leu23_B$), respectively (**Fig. 3a,b**). Unlike the other four bases of the hexanucleotide loop, C9 and U13 are not recognized sequence specifically by RsmE. The phosphate backbone of the hexanucleotide loop is stabilized by four positively charged lysine and arginine side chains ($Arg31_B$, $Lys38_B$, $Arg44_B$ and $Arg50_B$) and also by the N-terminal $Met1_A$ NH_3^+, which interacts with the A12 and U13 phosphate oxygens (**Fig. 3**). The two closing base pairs of the stem, C7-G14 and U6-A15, are contacted on the major-groove side by hydrogen bonds from the side chains of $Thr5_A$, $Gln29_B$ and $Arg31_B$ of $β1_A$ and $β4_B$ (**Fig. 3c**). The contact between $Gln29_B$ and the C7 amino group specifically recognizes the presence of a C7-G14 base pair, whereas the protein contacts with U6-A15 would also be allowed by an A-U base pair but would be tolerated less well by other base pairs. Finally, the $Lys7_A$ and $Arg31_B$ side chains and the amide of $Gln28_B$ interact with the phosphate backbone of the stem, extending the contacts down to the U5-A16 base pair (**Fig. 3c–e**). Note that the four added G-C pairs in the 20-nucleotide *hcnA* sequence do not contact RsmE.

In vivo and in vitro studies of RsmE-RNA interactions

We next tested repression of *hcnA* expression by RsmE *in vivo*, using an *rsmE*-overexpressing construct (pME6851). This system provides a robust assay of RsmE function and is more reliable than would be a comparison between the wild-type strain and an *rsmE*$^-$ mutant[17]. In our assay, overexpressed RsmE repressed *hcnA* expression three-fold (**Table 1**). In support of the structure, mutations in the sequence-specifically recognized nucleotides A8, G10, G11 and A12 (**Fig. 1c**) abolished both induction by GacA and repression by RsmE *in vivo*. These nucleotides are part of the Shine-Dalgarno sequence in the *hcnA* RBS. Moreover, mutation in C7 strongly diminished regulation by GacA and RsmE (**Table 1**). Within the RsmA/CsrA family, all side chains contributing to sequence-specific RNA recognition are either identical or similar[10] (**Fig. 1a**), suggesting that these proteins have a common RNA-recognition mode. The structure explains well the previously proposed RsmA/CsrA recognition sequences, in particular the ANGGA consensus sequence, which occurs repeatedly in target mRNAs[5,14,19,20], as well as the 5′-ACANGGANGU-3′ motif found in RNAs selected *in vitro* for CsrA binding[13]. Furthermore, the structure is entirely consistent with a previous *in vitro* study[13] showing that the ANGGAN hexanucleotide loop needs to be placed on a short stem to form a preferred CsrA target.

The structure also agrees with the RNA-recognition surface inferred from alanine-scanning mutagenesis of CsrA and *in vivo* expression data[12], although the interaction surface in the RsmE–*hcnA* RNA

Table 1 Effects of mutations in the *hcnA* RBS on translation regulation

Plasmid	Target sequence	β-galactosidase activity (Miller units × 10³)				GacA induction factor[a]	RsmE repression factor[b]
		CHA0 (wild-type)	CHA89 (*gacA*)	+pME6001 (empty vector)	+pME6851 (overexpressing *rsmE*)		
pME6533	Wild-type	7.0 ± 1.0	0.3 ± 0.1	3.5 ± 0.7	1.1 ± 0.3	23	3
pME6624	C7A	30.0 ± 1.0	16.0 ± 0.2	20.0 ± 1.4	10.4 ± 1.0	1.9	1.9
pME6629	A8U	16.0 ± 1.0	14.5 ± 2.0	10.5 ± 1.5	12.0 ± 2.5	1.1	0.9
pME7633	G10A G11C	1.6 ± 0.5	1.4 ± 0.3	1.2 ± 0.2	0.87 ± 0.15	1.1	1.3
pME6638	A12U	0.46 ± 0.04	0.36 ± 0.04	0.32 ± 0.02	0.20 ± 0.04	1.3	1.6

Shown are effects of mutations in the *hcnA* RBS on regulation of the expression of an *hcnA'-'lacZ* translational fusion by GacA and RsmE. β-galactosidase expression of various *hcnA'-'lacZ* fusion constructs was measured in *P. fluorescens* strains CHA0 (wild-type), CHA89 (*gacA*::Km[r]), CHA0/pME6001 (vector control) and CHA0/pME6851 (equivalent to pME6001 overexpressing RsmE) when the cells (grown in nutrient yeast broth at 30 °C) reached an A_{600} of 2.5. Values are averages of triplicate measurements ± s.d. [a]The GacA induction factor is the ratio of the expression of each fusion construct in CHA0 to that in CHA89. [b]The RsmE repression factor is the ratio of the expression of each fusion in CHA0/pME6001 to that in CHA0/pME6851.

complex is much larger than predicted. This is due to the numerous intermolecular contacts involving protein main chain atoms that could not be deciphered using an alanine-scanning approach, which detects the effects of side chain changes. We verified the functional importance of two RsmE amino acid side chains that recognize G10, namely Leu4 and Arg44. Both an L4A and an R44A RsmE mutant partially lost the ability to repress *hcnA* mRNA *in vivo* (**Fig. 4a**). Moreover, an R6A RsmE mutant, which lacks the Arg6-Glu46 salt bridge[10,11] and therefore is predicted to have a dislocated α-helix, was totally devoid of biological activity (**Fig. 4a**). After checking by NMR that the protein mutants L4A and R44A were properly folded (**Supplementary Fig. 5** online), we tested the effects of the mutations *in vitro* by isothermal titration calorimetry (ITC) experiments with the 20-nucleotide *hcnA* sequence. Whereas wild-type RsmE bound tightly to this sequence with 1:1 stoichiometry and a K_d of 85 nM, the L4A

mutant protein had a 35-fold lower affinity for the 20-nucleotide *hcnA* sequence, and the affinity of the R44A mutant was below the detection limit of ITC (**Fig. 4b,c**). This indicates that the *in vivo* effects of the mutations were due to weak binding to the RNA targets.

DISCUSSION

Earlier reports used sequence similarity to predict a KH domain fold for RsmA/CsrA family members[20]. The protein structures of RsmA and CsrA revealed that the RsmA/CsrA fold is entirely different from that of a KH domain[9–11]. Our RsmE–RNA complex structure further confirms that the mode of RNA recognition by the RsmA/CsrA fold is indeed different. It is notable that sequence-specific recognition of the unpaired nucleotides A8, G10, G11 and A12 is mediated mostly by carbonyl oxygens and amides of the protein main chain. This implies that the fold of the homodimeric RsmA/CsrA proteins is itself

Figure 4 *In vivo* and *in vitro* functional studies of RsmE-RNA interaction. (**a**) Cell density–dependent β-galactosidase expression of a chromosomal *hcnA'-'lacZ* translational fusion in *P. fluorescens* strains CHA1027/pME6032 (*rsmA*::ΩKm *rsmE*::ΩHg *hcnA'-'lacZ,* containing empty vector; ◆), CHA1027/pME7618 (equivalent to pME6032 overexpressing wild-type RsmE; □), CHA1027/pME9502 (equivalent to pME6032 overexpressing RsmE(L4A); ■), CHA1027/pME9503 (equivalent to pME6032 overexpressing RsmE(R6A); ●) and CHA1027/pME9504 (equivalent to pME6032 overexpressing RsmE(R44A); ○). Each data point represents the average from three different cultures ± s.d. (**b,c**) ITC binding experiments with RsmE mutants and the 20-nucleotide *hcnA* RNA. For measurements of wild-type (WT) protein or L4A mutant, protein concentration in the cell was 10 μM; RNA concentration in the syringe was 180 μM or 150 μM, respectively. During measurement of binding of the R44A mutant, protein concentrations in the cell was 35 μM; RNA concentration in the syringe was 590 μM. Raw calorimetric output from experiments with the 20-nucleotide *hcnA* RNA are shown in **b**, with thermograms of L4A and R44A shifted on the *y*-axis for clarity.

Binding isotherms describing formation of complex between RsmE and the 20-nucleotide *hcnA* sequence are shown in **c**. Data points represent integrated heats after normalization for molar concentration. Continuous lines are nonlinear fits according to a binding model assuming noninteracting sites. More complicated binding models do not improve the fitting statistics. (**d**) Competition of [α-^{33}P]UTP-labeled 20-nucleotide *hcnA* sequence (60 nM) for binding to His[6]-RsmE with various unlabeled RNA competitors (RsmX, RsmY, RsmZ, *carA* 5′ leader mRNA or *hcnA* 5′ leader mRNA[16,18]).
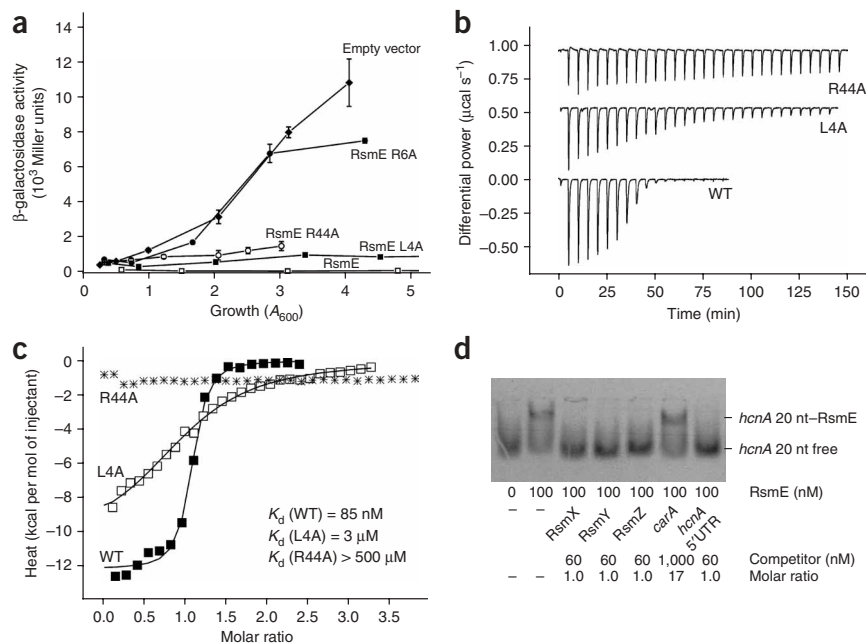
responsible for their RNA binding specificity. This distinguishes the RsmA/CsrA family from other small RNA-binding domains such as the RRM domains, where sequence specificity is achieved primarily through protein side chains[21].

By binding specifically to the 5′-$^A/_U$CANGGANG$^U/_A$-3′ consensus sequence, which closely matches the ideal 5′-AAGGAGGU-3′ Shine-Dalgarno sequence[22–24], the proteins of the RsmA/CsrA family can globally regulate the expression of numerous genes at the level of translation. As mentioned above, CsrA represses genes involved in gluconeogenesis, glycogen metabolism and biofilm formation in *E. coli*[4,25], whereas RsmA and RsmE repress genes responsible for the production of secondary metabolites and exoenzymes in *P. fluorescens*[14,17]. Five nucleotides of the *hcnA* Shine-Dalgarno sequence (underlined: <u>ACGGAUG</u>) are buried in the complex, either by contacts with the RsmE protein (<u>ACGGA</u>UG) or by base-pairing in the stem induced by protein binding (ACGGAU<u>G</u>). The structure explains well how the RsmA/CsrA proteins, by clamping the Shine-Dalgarno sequence, prevent 16S ribosomal RNA binding to the Shine-Dalgarno sequence and repress translation. This is reminiscent of the mechanism by which riboswitches[7] and RNA thermometers[26] regulate gene expression post-transcriptionally at the Shine-Dalgarno sequence, except that in the case of RsmA/CsrA, protein-RNA interactions replace RNA-RNA interactions.

The RsmA/CsrA–RNA complex is unusual in that the protein is a homodimer containing two RNA-binding sites. Is there a biological function associated with the presence of these two binding sites? Examination of several mRNAs regulated by RsmA/CsrA proteins reveals multiple putative binding sites upstream of the Shine-Dalgarno sequence in *E. coli* and *Pseudomonas* spp. For instance, in the 5′ untranslated region (5′ UTR) of *hcnA* in *P. fluorescens*, there are four GGA motifs upstream of the Shine-Dalgarno sequence (**Fig. 1b**). When all four motifs are mutated, translational regulation of *hcnA* by the Gac/Rsm system is entirely lost (K.L. and D.H., unpublished data). It seems that the upstream motifs as well as the motif overlapping the Shine-Dalgarno sequence are required for effective regulation by the Gac/Rsm system. In the 5′ UTR of the *E. coli glgC* mRNA, two binding sites for CsrA have been mapped, one at the Shine-Dalgarno sequence and one in an upstream stem-loop containing a GGA motif in the loop[27]. We modeled the two binding sites of the *glgC* 5′ UTR into our structure, and the model shows that binding of a single dimer to both binding sites is sterically possible (**Supplementary Fig. 6** online). This could explain the observation that mutations in the upstream site reduce the affinity of CsrA for the mRNA[27]. Moreover, in the small noncoding RNAs RsmX, RsmY, RsmZ, CsrB and CsrC, there are numerous GGA motifs that could be binding sites for RsmA/CsrA proteins. With regard to the *hcnA* 5′ UTR, we found that RsmX, RsmY and RsmZ could all competitively remove RsmE from its complex with the 20-nucleotide *hcnA* sequence, whereas an unrelated RNA (*carA* 5′ UTR) could not (**Fig. 4d**). This shows that an RNA containing multiple binding sites for the protein has higher affinity than one containing a single binding site. We also investigated this by NMR measurement of a complex between RsmE and two additional constructs derived from the *hcnA* 5′ UTR: a 50-nucleotide RNA containing the Shine-Dalgarno sequence plus two upstream GGA motifs, and a 26-nucleotide RNA containing solely the two upstream GGA motifs (**Fig. 1b**). Preliminary NMR data obtained with these complexes showed that RsmE can bind both GGA motifs in the 26-nucleotide *hcnA* sequence, whereas in the complex with the 50-nucleotide sequence, RsmE binds the Shine-Dalgarno sequence and only one of the two upstream GGA motifs (**Supplementary Fig. 7** online). This demonstrates that the RsmE dimer binds at two different locations within the *hcnA* 5′ UTR. In conclusion, RsmA/CsrA-RNA

recognition is complex, as it depends on at least two RNA-recognition sequences as well as their spatial relationship and potentially their binding cooperativity. Although our structure explains well the first level of complexity, only a structure of multiple RsmA/CsrA proteins bound to noncoding RNA or to a longer 5′ UTR RNA will explain the second level of complexity.

METHODS

Bacterial strains, DNA manipulations and growth conditions. The bacterial strains, plasmids and oligonucleotides used in this study are listed in **Supplementary Table 1** online. Details of DNA manipulation and growth conditions are given in **Supplementary Methods** online.

RsmE–*hcnA* RNA complex formation. The His-tagged RsmE protein (His$_6$-RsmE) was overexpressed in *E. coli* BL21(DE3)+RIL/pME7013 at 37 °C in minimal medium M9 containing 1 g l^{-1} ^{15}NH$_4$Cl and 4 g l^{-1} glucose (for ^{15}N-labeled proteins) or 1 g l^{-1} ^{15}NH$_4$Cl and 2 g l^{-1} ^{13}C-glucose (for ^{13}C,^{15}N labeled proteins). For structure determination, the 20-nucleotide *hcnA* sequence (5′-GGGCUUCACGGAUGAAGCCC-3′) was used with three different labeling schemes. Unlabeled RNA samples and two ^{13}C,^{15}N-labeled RNA samples (with only G and U or with only C and A labeled) were produced by *in vitro* run-off transcription with T7 polymerase and purified by anion-exchange high-pressure liquid chromatography under denaturing conditions. The complexes were prepared by titrating the concentrated RNA solution, of typically 10 mM, into an ∼0.5 mM solution of RsmE in a buffer of 300 mM NaCl and 50 mM K$_2$HPO$_4$ (pH 8.0) until a 1:1 stoichiometry was reached. Subsequently, the buffer was exchanged to 30 mM NaCl and 50 mM K$_2$HPO$_4$ (pH 7.2) with a Centricon device (5 kDa molecular mass cutoff membrane, Vivascience; see **Supplementary Methods** for more details).

Preparation of RsmE mutants. The His-tagged RsmE containing either an L4A or R44A mutation was overexpressed in *E. coli* BL21(DE3)+RIL (containing plasmids pME9502 and pME9504, respectively) in minimal medium M9 containing 1 g l^{-1} ^{15}NH$_4$Cl and 4 g l^{-1} glucose, and purified by the same protocol as for the wild-type RsmE.

NMR spectroscopy. NMR spectra were acquired on DRX-500, DRX-600 and Avance 900 Bruker spectrometers equipped with inverse triple-resonance probes and pulse-field gradient accessory (see **Supplementary Methods** for details).

Structure calculation and refinement. Preliminary structures of the RsmE–RNA complex were obtained by a simulated-annealing protocol using the DYANA package[28] and manually assigned NOE distance constraints. DYANA was used to generate 100 structures starting from random RNA and protein starting structures with 30,000 simulated-annealing steps. Initial calculations with an RsmE monomer and one RNA did not converge. Over 100 NOE distance restraints can be satisfied only by an RsmE dimer with a CsrA fold and were thus classified as unambiguous intermolecular restraints. This is supported by amide signals protected from hydrogen exchange at the dimerization interface (between strands β1$_A$ and β4$_B$, β2$_A$ and β5$_B$, β1$_B$ and β4$_A$, and β2$_B$ and β5$_A$). At later stages of the refinement, hydrogen bond restraints were added, including six intermolecular ones (three from slowly exchanging amides and three from substantial chemical shift perturbations in the carbonyl, see **Supplementary Table 2** online). An ensemble of 20 structures, selected on the basis of the lowest target function, served for the refinement in AMBER 7.0 (ref. 29). The complex was refined in implicit solvent using NOE-derived distances, torsion angles and hydrogen bond restraints as summarized in **Table 2**. In all AMBER calculations, force field 98, based on the force field of ref. 30, was used along with the generalized Born model[31] to mimic solvent. The Ramachandran plot shows 81.1% of the residues in the most favored regions, 18.4% in the additionally allowed regions, 0.0% in the generously allowed regions and 0.4% in the disallowed regions. More details are found in **Supplementary Methods**.

Isothermal titration calorimetry. ITC experiments were performed on a VP-ITC instrument (MicroCal). The calorimeter was calibrated according to the manufacturer's instructions. Samples of protein and RNA were prepared in and thoroughly dialyzed against the same batch of buffer (300 mM NaCl,

Table 2 NMR and refinement statistics

	Protein	RNA
NMR distance and dihedral constraints (per protein–RNA subunit)		
Distance restraints		
Total NOEs (intramolecular)	470	385
Intra-residue	22	196
Inter-residue	448	189
Sequential ($|i − j| = 1$)	232	134
Nonsequential ($|i − j| > 1$)	216	55
Hydrogen bonds (intramolecular)	18	19
Protein-protein intermolecular NOEs	132	
Protein-RNA intermolecular NOEs	220	
Protein-protein intermolecular hydrogen bonds	10	
Protein-RNA intermolecular hydrogen bonds	6	
Total dihedral angle restraints		80
Sugar pucker		6
Backbone[a]		74
Structure statistics		
Violations (mean ± s.d.)		
Number of distance restraint violations > 0.2 Å	1.60 ± 0.52	
Number of dihedral angle violations > 5°	0.60 ± 1.26[b]	
Max. dihedral angle violation (°)	11.8 ± 31.6[c]	
Max. distance constraint violation (Å)	0.25 ± 0.03	
Deviations from idealized geometry		
Bond lengths (Å)	0.009	
Bond angles (°)	1.5	
Average pairwise r.m.s. deviation (Å)[d]		
Protein (residues Met1–Ala53 of both subunits)		
Heavy	1.31 ± 0.15	
Backbone	0.79 ± 0.15	
RNA (residues G3–C18 bound to protein subunit A)		
All RNA heavy atoms	0.80 ± 0.26	
Complex		
All heavy atoms in complex (C, N, O, P)	1.20 ± 0.15	

The AMBER energies of the ten refined structures were −10,298 ± 63 kcal mol⁻¹.
[a]Based on A-form geometry derived from high-resolution crystal structures, where $\alpha = 270°–330°$, $\beta = 150°–210°$, $\gamma = 30°–60°$, $\delta = 50°–110°$, $\varepsilon = 180°–240°$, $\zeta = 260°–320°$. These restraints were used only for the double-helical region. [b]Violations in three models. [c]Single violation of an α angle in C20 in one structure. [d]Pairwise r.m.s. deviation was calculated among ten refined structures.

50 mM K$_2$HPO$_4$ (pH 8.0)) to minimize artifacts owing to minor differences in buffer composition. Concentration was determined after dialysis. The sample cell (1.4 ml) was loaded with 5–18 μM protein (10–35 μM binding sites); RNA concentration in the syringe was 150–600 μM. Titration experiments were done at 25 °C and typically consisted of 20–30 injections, each of 10 ml volume and 10 s duration, with a 5-min interval between additions. Stirring rate was 300 r.p.m. Raw data were integrated, corrected for nonspecific heats, normalized for the molar concentration and analyzed according to a 1:1 binding model assuming a single set of identical binding sites.

Purification of His-tagged RsmE for gel mobility shift assay. The His-tagged RsmE protein was overexpressed in *E. coli* BL21(DE3)/pME7609 and purified by affinity chromatography on nickel–nitrilotriacetic acid agarose beads (Qiagen) according to the manufacturer's instructions. After dialysis, the protein was further purified on cellulose phosphate (Sigma) by elution with 50 mM Tris-HCl (pH 7.5) and 10% (w/v) glycerol, concentrated on a Centricon membrane (10-kDa cutoff; Millipore) and stored at −20 °C. The purity of the protein was ≥90%, as judged from SDS-Tris-glycine PAGE.

Preparation of *in vitro* transcripts and electrophoretic mobility shift assays. The following transcripts were synthesized with a T7 transcription kit (Fermentas) on linear DNA templates: 20-nucleotide *hcnA* sequence, *hcnA*

leader (nucleotides +1 to + 151 of *hcnA* containing the 110-nucleotide untranslated leader and the first 41 nucleotides of the *hcnA* coding sequence), *rsmX*, *rsmY*, *rsmZ* and the *carA* leader. The templates were produced by PCR on pME6533 (*hcnA*), pME7318 (*rsmX*), pME6919 (*rsmY*), pME6920 (*rsmZ*) and pME6926 (*carA*) DNA with the oligonucleotides hcn20 and hcn20rev for the 20-nucleotide *hcnA* sequence, hcn and hcnrev for *hcnA*, PTZ and PTZeco for RsmX, and PTZ and PTZrev for RsmY, RsmZ and the *carA* leader (**Supplementary Table 2**). The 20-nucleotide *hcnA* transcript was radioactively labeled in the presence of [α-³³P]UTP and used for binding reactions with the His$_6$-RsmE protein; unlabeled RNA competitors were included when appropriate. The reaction mixtures (10 μl) were incubated at 30 °C for 30 min. Samples were loaded on nondenaturing 10% polyacrylamide gels in 0.1 M Tris, 0.1 M H$_3$BO$_3$ and 0.0025 M EDTA (pH 8.3) and run at 10 mA for 4 h. Radioactive bands were visualized by autoradiography after the gels were dried.

β-galactosidase assays. *P. fluorescens* strains were grown at 30 °C with shaking (180 r.p.m.) in 50-ml flasks containing 20 ml of nutrient yeast broth (NYB) supplemented with 0.05% (w/v) Triton X-100. Specific activities were determined by the Miller method[32]. To induce expression, 1 mM IPTG was added to cultures of strains containing pME6032 derivatives.

Accession code. Protein Data Bank: Coordinates have been deposited with accession code 2JPP. Biological Magnetic Resonance Data Bank (http://www.bmrb.wisc.edu/): the chemical shifts of the RsmE protein in complex with *hcnA* RNA have been deposited with accession number 15257.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

AUTHOR CONTRIBUTIONS
Protein and RNA samples for structural studies were prepared by M.S., O.D. and F.C.O. Functional experiments and cloning were carried out by K.L. and analyzed by K.L. and D.H. NMR data were analyzed by M.S., O.D. and F.H.-T.A. ITC measurements were performed by I.J. The structure calculation and refinement was carried out by M.S. The manuscript was written by M.S., D.H. and F.H.-T.A.

COMPETING INTEREST STATEMENT
The authors declare no competing financial interests.

1. Babitzke, P. & Romeo, T. CsrB sRNA family: sequestration of RNA binding regulatory proteins. *Curr. Opin. Microbiol.* **10**, 156–163 (2007).
2. Haas, D. & Defago, G. Biological control of soil-borne pathogens by fluorescent pseudomonads. *Nat. Rev. Microbiol.* **3**, 307–319 (2005).
3. Heeb, S. & Haas, D. Regulatory roles of the GacS/GacA two-component system in plant-associated and other gram-negative bacteria. *Mol. Plant Microbe Interact.* **14**, 1351–1363 (2001).
4. Romeo, T. Global regulation by the small RNA-binding protein CsrA and the non-coding RNA molecule CsrB. *Mol. Microbiol.* **29**, 1321–1330 (1998).
5. Valverde, C., Lindell, M., Wagner, E.G. & Haas, D. A repeated GGA motif is critical for the activity and stability of the riboregulator RsmY of *Pseudomonas fluorescens*. *J. Biol. Chem.* **279**, 25066–25074 (2004).
6. Majdalani, N., Vanderpool, C.K. & Gottesman, S. Bacterial small RNA regulators. *Crit. Rev. Biochem. Mol. Biol.* **40**, 93–113 (2005).
7. Winkler, W.C. & Breaker, R.R. Regulation of bacterial gene expression by riboswitches. *Annu. Rev. Microbiol.* **59**, 487–517 (2005).
8. Winkler, W.C. Riboswitches and the role of noncoding RNAs in bacterial metabolic control. *Curr. Opin. Chem. Biol.* **9**, 594–602 (2005).
9. Gutierrez, P. *et al.* Solution structure of the carbon storage regulator protein CsrA from *Escherichia coli*. *J. Bacteriol.* **187**, 3496–3501 (2005).

10. Heeb, S. *et al.* Functional analysis of the post-transcriptional regulator RsmA reveals a novel RNA-binding site. *J. Mol. Biol.* **355**, 1026–1036 (2006).

11. Rife, C. *et al.* Crystal structure of the global regulatory protein CsrA from Pseudomonas putida at 2.05 Å resolution reveals a new fold. *Proteins* **61**, 449–453 (2005).

12. Mercante, J., Suzuki, K., Cheng, X., Babitzke, P. & Romeo, T. Comprehensive alanine-scanning mutagenesis of *Escherichia coli* CsrA defines two subdomains of critical functional importance. *J. Biol. Chem.* **281**, 31832–31842 (2006).

13. Dubey, A.K., Baker, C.S., Romeo, T. & Babitzke, P. RNA sequence and secondary structure participate in high-affinity CsrA-RNA interaction. *RNA* **11**, 1579–1587 (2005).

14. Blumer, C., Heeb, S., Pessi, G. & Haas, D. Global GacA-steered control of cyanide and exoprotease production in Pseudomonas fluorescens involves specific ribosome binding sites. *Proc. Natl. Acad. Sci. USA* **96**, 14073–14078 (1999).

15. Heeb, S., Blumer, C. & Haas, D. Regulatory RNA as mediator in GacA/RsmA-dependent global control of exoproduct formation in *Pseudomonas fluorescens* CHA0. *J. Bacteriol.* **184**, 1046–1056 (2002).

16. Kay, E., Dubuis, C. & Haas, D. Three small RNAs jointly ensure secondary metabolism and biocontrol in *Pseudomonas fluorescens* CHA0. *Proc. Natl. Acad. Sci. USA* **102**, 17136–17141 (2005).

17. Reimmann, C., Valverde, C., Kay, E. & Haas, D. Posttranscriptional repression of GacS/GacA-controlled genes by the RNA-binding protein RsmE acting together with RsmA in the biocontrol strain *Pseudomonas fluorescens* CHA0. *J. Bacteriol.* **187**, 276–285 (2005).

18. Valverde, C., Heeb, S., Keel, C. & Haas, D. RsmY, a small regulatory RNA, is required in concert with RsmZ for GacA-dependent expression of biocontrol traits in *Pseudomonas fluorescens* CHA0. *Mol. Microbiol.* **50**, 1361–1379 (2003).

19. Wang, X. *et al.* CsrA post-transcriptionally represses pgaABCD, responsible for synthesis of a biofilm polysaccharide adhesin of *Escherichia coli. Mol. Microbiol.* **56**, 1648–1663 (2005).

20. Liu, M.Y., Yang, H. & Romeo, T. The product of the pleiotropic *Escherichia coli* gene csrA modulates glycogen biosynthesis via effects on mRNA stability. *J. Bacteriol.* **177**, 2663–2672 (1995).

21. Auweter, S.D., Oberstrass, F.C. & Allain, F.H. Sequence-specific binding of single-stranded RNA: is there a code for recognition? *Nucleic Acids Res.* **34**, 4943–4959 (2006).

22. Shine, J. & Dalgarno, L. The 3′-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proc. Natl. Acad. Sci. USA* **71**, 1342–1346 (1974).

23. Ma, J., Campbell, A. & Karlin, S. Correlations between Shine-Dalgarno sequences and gene features such as predicted expression levels and operon structures. *J. Bacteriol.* **184**, 5733–5745 (2002).

24. Yusupova, G.Z., Yusupov, M.M., Cate, J.H. & Noller, H.F. The path of messenger RNA through the ribosome. *Cell* **106**, 233–241 (2001).

25. Jackson, D.W. *et al.* Biofilm formation and dispersal under the influence of the global regulator CsrA of *Escherichia coli. J. Bacteriol.* **184**, 290–301 (2002).

26. Narberhaus, F., Waldminghaus, T. & Chowdhury, S. RNA thermometers. *FEMS Microbiol. Rev.* **30**, 3–16 (2006).

27. Baker, C.S., Morozov, I., Suzuki, K., Romeo, T. & Babitzke, P. CsrA regulates glycogen biosynthesis by preventing translation of glgC in *Escherichia coli. Mol. Microbiol.* **44**, 1599–1610 (2002).

28. Guntert, P., Mumenthaler, C. & Wuthrich, K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273**, 283–298 (1997).

29. Case, D.A. *et al. AMBER Version 7* (University of California, San Francisco, 2002).

30. Cornell, W.D. *et al.* A 2nd generation force-field for the simulation of proteins, nucleic-acids, and organic-molecules. *J. Am. Chem. Soc.* **117**, 5179–5197 (1995).

31. Bashford, D. & Case, D.A. Generalized born models of macromolecular solvation effects. *Annu. Rev. Phys. Chem.* **51**, 129–152 (2000).

32. Miller, J.H. *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA, 1972).