

# Molecular basis for the wide range of affinity found in Csr/Rsm protein–RNA recognition

Olivier Duss, Erich Michel, Nana Diarra dit Konté, Mario Schubert and Frédéric H.-T. Allain\*

Institute of Molecular Biology and Biophysics, ETH Zürich, 8093 Zürich, Switzerland

Received December 27, 2013; Revised January 22, 2014; Accepted January 24, 2014

## ABSTRACT

The carbon storage regulator/regulator of secondary metabolism (Csr/Rsm) type of small non-coding RNAs (sRNAs) is widespread throughout bacteria and acts by sequestering the global translation repressor protein CsrA/RsmE from the ribosome binding site of a subset of mRNAs. Although we have previously described the molecular basis of a high affinity RNA target bound to RsmE, it remains unknown how other lower affinity targets are recognized by the same protein. Here, we have determined the nuclear magnetic resonance solution structures of five separate GGA binding motifs of the sRNA RsmZ of *Pseudomonas fluorescens* in complex with RsmE. The structures explain how the variation of sequence and structural context of the GGA binding motifs modulate the binding affinity for RsmE by five orders of magnitude (~10 nM to ~3 mM,  $K_d$ ). Furthermore, we see that conformational adaptation of protein side-chains and RNA enable recognition of different RNA sequences by the same protein contributing to binding affinity without conferring specificity. Overall, our findings illustrate how the variability in the Csr/Rsm protein–RNA recognition allows a fine-tuning of the competition between mRNAs and sRNAs for the CsrA/RsmE protein.

## INTRODUCTION

Regulation by small non-coding RNAs (sRNAs) is crucial for orchestrating global changes in bacterial gene expression (1–3). The best studied small regulatory RNAs in bacteria function by Hfq chaperone-assisted base pairing with target messenger RNAs (mRNAs) (4,5). They often work by binding to the ribosome binding site (RBS), thereby repressing mRNA translation and through recruiting RNase E via Hfq, they can target the mRNA for degradation (6). Another important group of sRNAs

do not act on mRNAs directly but function by sequestering the CsrA-type protein from the RBS of a subset of mRNAs and thus, activate translation initiation (1,7,8).

The Csr system has been characterized as regulating global pathways involved in central carbon metabolism, cell motility, biofilm formation, quorum sensing, the production of extracellular products and is viewed as the most important post-transcriptional regulator of bacterial pathogenesis (7–10). Homologues of the CsrA protein have been found to be widely distributed among bacteria (encoded by ~75% of all species) (9). The orthologues of CsrA (carbon storage regulator), such as RsmA or RsmE (regulator of secondary metabolism), have high sequence identity and similarity for all protein amino acid side-chains contributing to RNA recognition (11). Remarkably, even the very recently identified orthologues RsmN/RsmF protein from *Pseudomonas aeruginosa* has a conserved RNA recognition surface, despite a distinctly different polypeptide fold when compared with the domain-swapped dimeric structure of the CsrA/RsmE homologues (12,13). The 15 kDa CsrA/RsmE homodimer is able to bind, using two identical binding sites, to RNA sequences containing a central GGA motif flanked by additional nucleotides which are also bound (11,14). In general, the 5'-UTR of the CsrA/RsmE-regulated mRNAs contain between one and several GGA motifs that bind CsrA/RsmE with different affinities (10,15–17). Often, a high affinity GGA motif overlaps directly with the RBS and can bind CsrA/RsmE efficiently, preventing the docking of the RBS to the 30S ribosomal subunit and thus inhibiting translation initiation. It is also possible that the GGA motif overlapping the RBS has only low affinity for CsrA/RsmE and is therefore not able to efficiently repress ribosome loading (16). However, other GGA motifs located upstream of the RBS with high affinity for CsrA can recruit the homodimeric protein, thereby increasing its local concentration to enable cooperative binding to the RBS and efficient translation repression (16). The de-repressor sRNAs also contain several GGA binding motifs, typically located in hairpin loops (14,18–21), also in single-stranded regions or even buried within secondary structures.

\*To whom correspondence should be addressed. Tel: +41 44 63 33940; Fax: +41 44 63 31294; Email: allain@mol.biol.ethz.ch

We have previously solved the solution nuclear magnetic resonance (NMR) structure of a stem-loop encompassing the RBS of the *Pseudomonas fluorescens* CHA0 *henA* mRNA in complex with RsmE (11). The structure shows how the guanine and adenine bases of the **ACGGAU** hexanucleotide loop of the 20 nt stem-loop RNA (20 nt-RBS RNA) are specifically recognized by the protein backbone and thus, the shape of the protein, while the RNA stem is semi-specifically recognized on its major groove side by protein side-chains. Although this structure demonstrates the molecular basis for the CsrA/RsmE protein recognition by a specific GGA motif, which is close to the high affinity SELEX-derived consensus sequence (14), it is unclear how other GGA motifs in different sequence and structural contexts would be recognized by the CsrA/RsmE protein. To deepen the understanding of CsrA/RsmE recognition by the diverse class of GGA binding motifs, we solved the solution structures of five different GGA motifs, all located within the sRNA RsmZ of *P. fluorescens*, in complex with the RsmE protein. We show how the same protein can recognize different RNA targets by plasticity of both protein and RNA structure. The sequence and structural context of a GGA motif allows modulation in the affinity by more than five orders of magnitude. Overall, we provide the framework for predicting the binding affinity of a certain GGA motif for the CsrA/RsmE protein according to its sequence and structural context. Using this framework, we can successfully predict the effect of sRNA mutations on translation activation by the *P. fluorescens* sRNA RsmX *in vitro* and rationalize the sequential binding of RsmE to the 5'-UTR of the *P. fluorescens henA* mRNA.

## MATERIALS AND METHODS

### Protein and RNA sample preparation

The RsmE protein homo-dimer was expressed recombinantly and purified with a C-terminal histidine tag as reported previously (11). The RNA was obtained by *in vitro* transcription from double-stranded DNA templates using T7 polymerase and was subsequently purified by denaturing high-performance liquid chromatography (HPLC) followed by butanol extraction as reported earlier (22,23). Due to its small size, the 9 nt-GGA<sub>39-41</sub> RNA was first transcribed as a longer precursor RNA, which was subsequently cut into smaller fragments by sequence-specific RNase H cleavage. Protein-RNA complexes used for NMR experiments were prepared by mixing the protein and RNA at a 1:1 ratio with a typical concentration of 1 mM in a buffer containing 30 mM NaCl and 50 mM potassium phosphate at pH 7.2.

### NMR spectroscopy

NMR spectra were acquired at 313 K except for 2D <sup>1</sup>H-<sup>1</sup>H NOESY experiments in H<sub>2</sub>O, which were performed at 283 K to observe RNA imino protons in stem regions of RNA hairpins. The spectra were recorded on Bruker Avance III 500, 600, 700 or 900 MHz spectrometers equipped with a cryo probe. All spectra were processed

with Topspin 2.1 or 3.0 and analysed in Sparky 3.0. The <sup>1</sup>H, <sup>13</sup>C and <sup>15</sup>N chemical shifts of the RsmE protein in complex with the 5 RNA targets were assigned by standard methods (24,25). The RNA imino protons were assigned with 2D <sup>1</sup>H-<sup>1</sup>H NOESY experiments ( $\tau_m = 250$  ms) in 95% H<sub>2</sub>O/5% D<sub>2</sub>O (v/v) at 283 K. Non-exchangeable RNA proton resonances were assigned using 2D <sup>1</sup>H-<sup>1</sup>H NOESY ( $\tau_m = 150$  ms), <sup>1</sup>H-<sup>1</sup>H TOCSY, <sup>1</sup>H-<sup>13</sup>C HSQC, 3D <sup>13</sup>C-edited NOESY ( $\tau_m = 150$  ms) and 3D HC(C)H TOCSY spectra in D<sub>2</sub>O with samples, in which the RsmE protein was only <sup>15</sup>N-labelled and the RNA nucleotide-specifically <sup>13</sup>C, <sup>15</sup>N-labelled (with either A, U or G, C labelled) (25). Due to severe spectral overlap of the RNA resonances for SL3 in complex with RsmE, a third cytosine-only <sup>13</sup>C, <sup>15</sup>N-labelled RNA in complex with only <sup>15</sup>N-labelled RsmE protein was prepared. For the 9 nt-GGA<sub>39-41</sub> RNA in complex with RsmE, it was sufficient to isotopically label the RNA uniformly for full assignment. The intra- and intermolecular nuclear Overhauser effect (NOE) signals were assigned based on 2D <sup>1</sup>H-<sup>1</sup>H nuclear Overhauser enhancement spectroscopy (NOESY) ( $\tau_m = 150$  ms), 3D <sup>13</sup>C-edited NOESY ( $\tau_m = 150$  ms) and 3D (F1-edited, F3-filtered) NOESY spectra ( $\tau_m = 150$  ms) (25) of samples in which either the protein was <sup>13</sup>C, <sup>15</sup>N-labelled and the RNA unlabelled, or the protein only <sup>15</sup>N-labelled and the RNA nucleotide-specific <sup>13</sup>C, <sup>15</sup>N-labelled as described above. The NOEs were semi-quantitatively classified according to their intensities in the 2D and 3D NOESY spectra. Hydrogen bonding distance restraints were based on the observation of an imino resonance of the corresponding base pair. Angle restraints of the sugar pucker conformations were extracted from <sup>1</sup>H-<sup>1</sup>H TOCSY spectra. Protein torsion angle restraints were obtained from TALOS+ (26).

The heteronuclear <sup>1</sup>H-<sup>15</sup>N NOE experiments were recorded in an interleaved fashion, recording alternatively one increment for the NOE or the reference spectrum (27). The relaxation delay was 2 s and the <sup>1</sup>H pre-saturation delay 3 s for the NOE experiment, while a 5 s relaxation delay was used in the reference experiment.

### Structure calculation and refinement

Preliminary structures were generated by a simulated annealing protocol using the CYANA package (28) including manually assigned NOE distance, torsion angle and hydrogen bond constraints as summarized in Supplementary Table S1. A total of 999 structures were generated starting from random coil RNA and protein chains using 20 000 simulated annealing steps. An ensemble of the 50 lowest target energy structures was selected and further refined in AMBER 9.0 (29,30). The complex was refined in implicit solvent using the same distance, torsion angle and hydrogen bond constraints as used in the CYANA simulated annealing protocol. The force field ff99 (31) was used along with the generalized Born model (32) to mimic the solvent. The 20 lowest energy structures were selected. The structural statistics are presented in Supplementary Table S1.

### Isothermal titration calorimetry

The isothermal titration calorimetry (ITC) binding experiments were conducted on a VP-ITC instrument from MicroCal. The calorimeter was calibrated according to the manufacturer's instructions. The RNA and the protein samples were dialysed against the same buffer batch (300 mM NaCl, 50 mM potassium phosphate at pH 8.0). Concentrations were determined after dialysis using optical density absorbance at 260 and 280 nm for RNA or protein, respectively. The RNA (syringe, 100–600  $\mu$ M concentration) was titrated into the RsmE protein (cell, 5–30  $\mu$ M dimer concentration). ITC binding experiments were performed at 298 K and typically consisted of 30–40 injections of 4–10  $\mu$ l with an injection speed of 2 s/ $\mu$ l and a 5 min interval between additions. The stirring rate was 307 rpm. All measurements were repeated at least twice. Using Origin 7.0, the raw data was integrated, corrected for non-specific heats and analysed according to a one-site or two-site model. The RNA sequences of all the constructs are summarized in Supplementary Table S2.

### Binding affinity determination by NMR spectroscopy

In the slow exchange regime on the NMR timescale, the resonances corresponding to the free and bound macromolecules are simultaneously present as two separated peaks if their chemical environments change upon complex formation. Their relative integrals directly relate to the fraction of free and bound macromolecule (protein or RNA) present in the sample. Knowing the total concentration of macromolecules (protein and RNA), the dissociation constant can directly be determined using the equation:

$$K_d = (\text{Protein})(\text{RNA})/(\text{Complex}) \quad (1)$$

As only a single well-separated resonance (imino of the second G in the GGA motif forming an intermolecular hydrogen bond) is observable for the unlabelled RNA in complex, the dissociation constants with their associated errors for the hcnA-GGA#1 and hcnA-GGA#2 RNAs binding to RsmE (Figure 6) were determined using several amide resonance pairs of free and bound protein at a total concentration of protein and RNA of 0.54 mM.

### Cell-free translation assay

The 5'-UTR of the *hcnA* gene was subcloned into the vector pIVEX1.3-CAT (Roche) directly upstream of the chloramphenicol acetyl-transferase (CAT) open reading frame, yielding the vector pCFX100, which was amplified using a plasmid maxi prep kit (Macherey-Nagel). S30 cell extract was obtained from *Escherichia coli* BL21 (DE3) Star cells, which were genetically modified by introducing a C-terminal (His)<sub>6</sub>-tag in the *csrA* gene following the procedure by Datsenko and Wanner (33), resulting in the strain *E. coli* BL21 (DE3) Star *csrA::His*<sub>6</sub>. In addition to the previously described extract preparation protocol (34), the endogenous *csrA* protein was removed by passing the cell lysate directly over Ni-NTA beads. The obtained

S30 extract devoid of *csrA* was then used for *in-vitro* translation of the CAT reporter gene from pCFX100 according to the previously described protocol (34). The RsmX sRNA was prepared by *in vitro* transcription from a linearized plasmid and purified by denaturing HPLC as described previously (22,23).

Analytical scale cell-free expressions (50  $\mu$ l) of the pCFX100 plasmid were set up in presence of various amounts of wild-type or mutant RsmX sRNA and 100 nM dimeric RsmE protein from *P. fluorescens*. After 3 h of cell-free expression, the reaction mixture was centrifuged and placed on ice. A quantity of 5  $\mu$ l of the reaction supernatant was thoroughly mixed with 495  $\mu$ l dilution buffer (100 mM Tris-HCl at pH 7.8, 1 mg/ml BSA). A total of 10  $\mu$ l of the diluted solution was then mixed with 990  $\mu$ l CAT reaction solution (100 mM Tris-HCl at pH 7.8, 0.5 mM DTNB, 50  $\mu$ M acetyl-CoA, 50  $\mu$ M chloramphenicol, 1 mg/ml BSA) and the increase in absorbance at 412 nm was followed for 20 min on a Cary 300 Bio spectrophotometer. The expression levels of the reporter enzyme were derived from the slope of absorbance at 412 nm against time and were then normalized to the largest value that was obtained after complete saturation of RsmE in the reaction mixture with RsmX RNA.

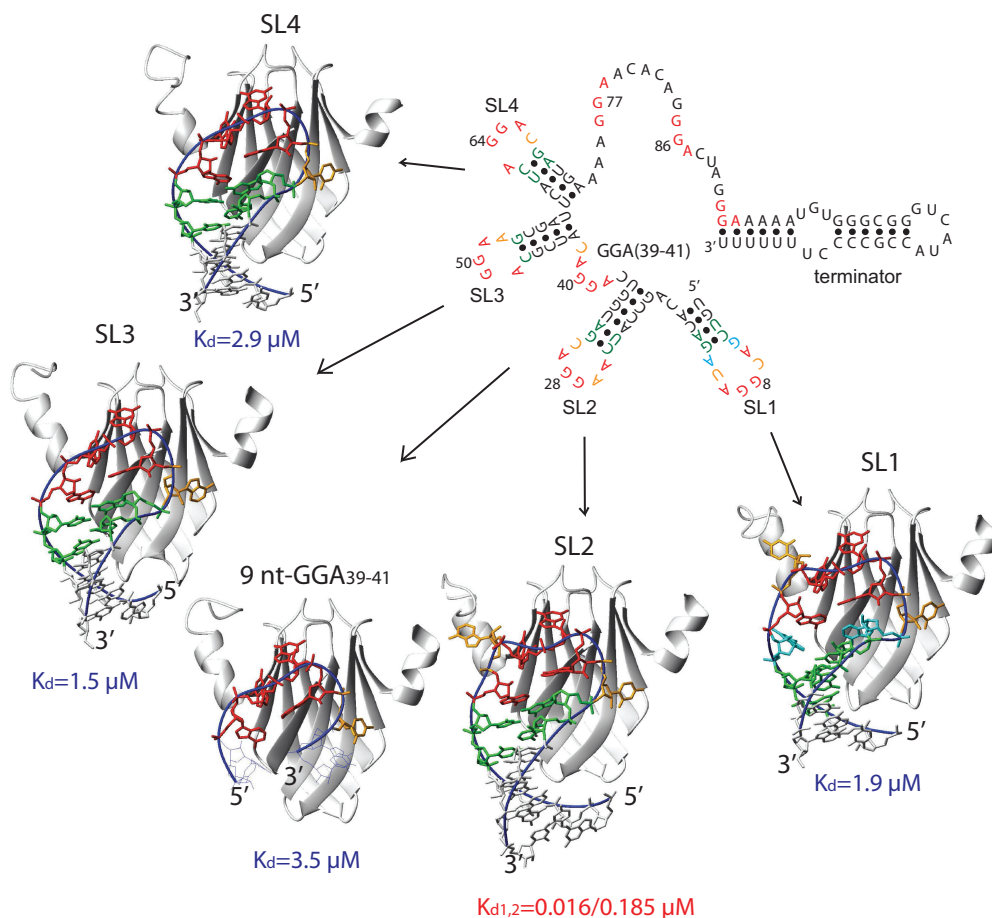
## RESULTS

### Solution structures of five different GGA motifs of RsmZ bound to RsmE

The sRNA RsmZ from *P. fluorescens* CHA0 is a sRNA of 127 nt composed of four stem-loops (SLs) and a  $\rho$ -independent terminator (19). It contains eight GGA motifs that are predicted to bind RsmE (Figure 1). Here, we have solved the NMR solution structures of the four isolated stem-loops and of the single-stranded region between SL2 and SL3 of the RsmZ sRNA in complex with the RsmE protein from *P. fluorescens* CHA0. All four stem-loops of RsmZ contain a conserved A(N)GGAX motif in the loop on top of a stem closed by two base pairs, a C-G followed by a U-A, except SL3 that has a G-C instead of a U-A as penultimate stem-closing base pair (Figure 1). However, the loop length varies between 5 and 8 nt. SL3 and SL4 have the shortest loop sequence of 5 nt, in which the nucleotide N in A(N)GGAX is missing. SL2 contains 6 nt in the loop like the 20 nt-RBS RNA. The difference between the two stem-loops occurs at the nucleotides N and X and the lower part of the stem. SL1 has the longest loop (8 nt), in which two additional nucleotides are inserted between the ANGGAX sequence and the stem. The RNA sequence between SL2 and SL3 (9 nt-GGA<sub>39-41</sub>) is single-stranded and is also missing the nucleotide N of the consensus motif.

First, we characterized the binding of these RNAs to RsmE with NMR chemical shift titrations and ITC-binding experiments. All the protein and RNA resonances in the 4 SL complexes are in slow exchange relative to the NMR chemical shift timescale (Supplementary Figure S1). Unexpectedly, the complex with the single-stranded





**Figure 1.** NMR solution structures of five different GGA motifs of RsmZ in complex with RsmE. The predicted secondary structure of the RsmZ sRNA and the lowest energy structure of each protein–RNA complex with its corresponding  $K_d$  (determined by ITC) are presented. The second RNA molecule binding to the homo-dimeric RsmE protein is not shown for simplicity. The conserved A(N)GGAX motif is shown in red, the CG/UA stem closing base pair in green, the looped-out nucleotides N and X in orange and the inserted nucleotides between the A(N)GGAX loop and the CG-UA stem closing base pair in SL1 are shown in cyan. The RNA nucleotides in the complex structures are for SL1: 1–16, SL2: 19–36, SL3: 43–57, SL4: 58–72 and 9 nt-GGA<sub>39-41</sub>: 36–44 (U36, C37, A43 and U44 are unstructured and shown in grey line representation). For SL2 binding to the homo-dimeric RsmE protein, the ITC binding data could only be fitted using a two-site binding model (both  $K_d$  values shown), whereas the binding for the other complexes could be fitted using a one-site binding model only. An overview of all ITC titration curves is presented in Supplementary Figure S2.

9 nt-GGA<sub>39-41</sub> RNA is also in slow exchange indicating a slow rate of complex dissociation ( $<10/s$ ). The binding affinities differ by one to two orders of magnitude. Whereas SL2 has an affinity of  $16 \pm 3$  and  $185 \pm 3$  nM for the binding of the first and second RNA molecule to the dimeric RsmE protein, respectively, the other RNA targets have affinities ranging from 1.5 to  $3.5 \mu\text{M}$  (Figure 1 and Supplementary Figure S2).

To understand the molecular basis for these differences in affinity, we determined the structures of the corresponding protein–RNA complexes. The high quality of the spectra allowed us to collect enough intra- and intermolecular NOEs to solve the structures of all five complexes with an RMSD of 0.75–1.2 Å. The structural statistics are shown in Supplementary Table S1. All the RNA nucleotides are well defined in the stem–loop RNA complexes (Supplementary Figure S3). In the single-stranded 9 nt-GGA<sub>39-41</sub> complex structure, the nucleotides flanking the common A(N)GGAX binding motif, are not well defined

and are not recognized by the RsmE protein (Supplementary Figure S3). They are flexible as can be judged from the narrower line width of their aromatic and sugar resonances. An overview of all the structures is shown in Figure 1.

#### Common binding mode

The conserved A(N)GGAX motif is recognized in an identical manner in all five structures, also found in the structure of the 20 nt-RBS RNA bound to RsmE (11) (Supplementary Figure S4). As the recognition of the A(N)GGAX motif is largely achieved by many hydrogen bonds between the RNA bases and the protein backbone, the recognition is sequence specific and these bases cannot be substituted by others and still be accommodated by the given protein fold. In contrast to the core A(N)GGAX motif, the nucleotides N and X within and the nucleotides adjacent to it are variable and allow for a modulation of the binding affinity.

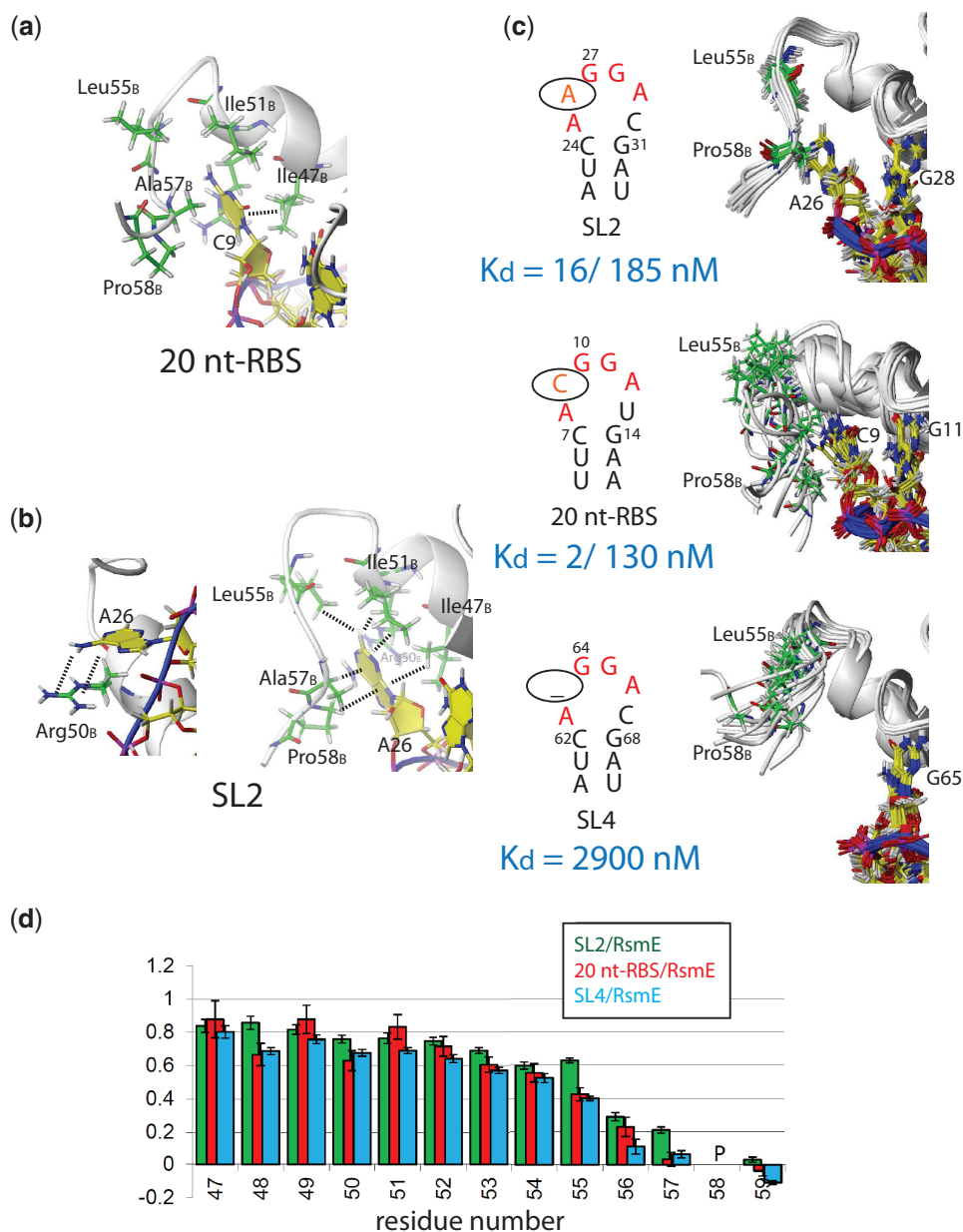


### Nucleotide N in A(N)GGAX contributes to binding affinity

The presence of the looped out nucleotide N contributes to a one to three orders of magnitude increase in binding affinity (Figure 2). If the looped out residue is a cytosine as in 20 nt-RBS or SL1, hydrophobic interactions of the base with the side-chain of Ile47<sub>B</sub> are present (Figure 2a). In contrast, when a larger adenine is looped-out as in SL2, more hydrophobic contacts are observed between the H2 proton and the larger surface of the purine base of adenosine A26 with the aliphatic side-chains of Ile47<sub>B</sub>, Arg50<sub>B</sub>,

Ile51<sub>B</sub>, Leu55<sub>B</sub>, Ala57<sub>B</sub> and Pro58<sub>B</sub> (Figure 2b). Interestingly, this adenine also stacks on Arg50<sub>B</sub> (Figure 2b, left); this interaction is not possible for the cytosine which is too small to stack with Arg50<sub>B</sub>. Obviously, all these interactions are lost if there is no base looped out as in SL3, SL4 or the ssRNA 9 nt-GGA<sub>39-41</sub> (Figure 2c, bottom image), rationalizing the significant loss in affinity when the nucleotide N in A(N)GGAX is not present.

Despite the larger number of hydrophobic contacts and the stacking of adenine A26 on Arg50<sub>B</sub>, the affinity of SL2 is very similar to that of 20 nt-RBS for RsmE (both RNA



**Figure 2.** Recognition of the looped-out nucleotide N in A(N)GGAX. (a and b) Structural details of interactions important for recognition of the nucleotide N in 20 nt-RBS (a) and SL2 (two views) (b). (c) NMR structural ensemble of C-terminus of RsmE contacting A26 in SL2 (top), C9 in 20 nt-RBS (center) and no base in SL4 (bottom). The  $K_d$  values are indicated for the different RNAs. For the 20 nt-RBS and SL2, the two  $K_d$  values correspond to the first and second RNA molecule binding to the homo-dimeric RsmE protein, respectively. (d)  $^{15}\text{N}$ -heteronuclear NOE values of the C-terminus of RsmE in complex with SL2 (green), 20 nt-RBS (red) and SL4 (cyan). The error bars were determined according to Farrow *et al.* (35).

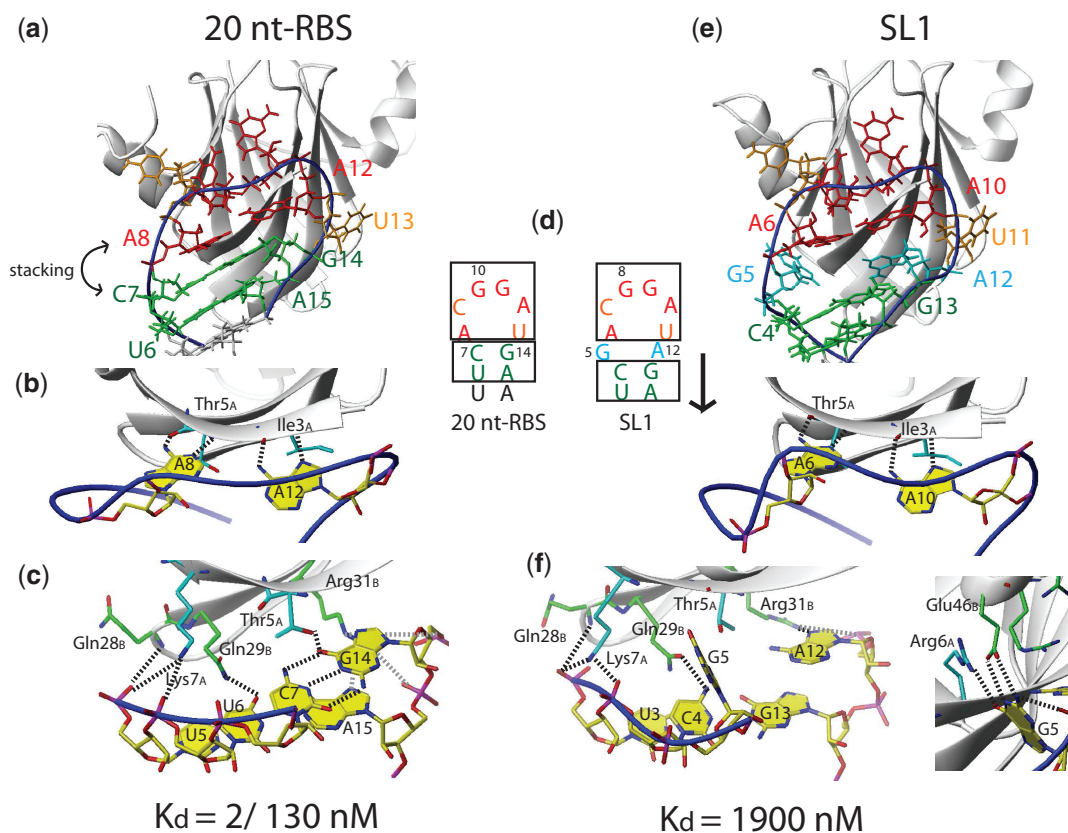
sequences in the ANGGAX loop are identical in the region that contacts the protein except for the two looped out nucleotides N and X). A static structure allows a rationalization of the enthalpic contribution to the free energy of binding but does not provide any insight into the entropic contribution to binding, which might explain why the affinity of SL2 is not significantly higher than that of the 20 nt-RBS for RsmE. Remarkably, ITC-binding experiments show a significantly higher binding enthalpy for SL2 compared with the 20 nt-RBS RNA and SL4 (Supplementary Figure S5). However, the favourable binding enthalpy is compensated by more unfavourable binding entropy for SL2 compared with the other two RNAs.

We then addressed the question of why SL2 binding is entropically less favoured when compared to binding of SL4 and the 20 nt-RBS RNA to RsmE. For this, we recorded  $^{15}\text{N}$ -heteronuclear NOE experiments for all three complexes, which report on the protein backbone dynamics in the ps–ns timescale (36). The  $^{15}\text{N}$ -heteronuclear values show that if the looped out nucleotide base N is an adenine (SL2), the C-terminal nucleotides Leu55 to Ala57 become more rigid than when N is a cytosine (20 nt-RBS) or N is absent (SL4)

(Figure 2). These results suggest that the additional hydrophobic contacts made by a looped out adenine (a gain in enthalpy) are counteracted by unfavourable entropic contributions to the free energy of binding, at least partially due to the residues in the C-terminal  $\alpha$ -helix becoming more ordered. This supports earlier findings that conformational entropy can significantly influence the free energy of binding (37). In summary, the presence but not the identity of the looped out nucleotide N of the A(N)GGAX motif results in a significant gain in binding affinity for *P. fluorescens* CHA0 RsmE. However, the energetic origin of this gain depends on the identity of the base of the looped out nucleotide.

#### Nucleotides located between the ANGGAX hexa-loop and the stem decrease binding affinity

In SL1, the insertion of two additional nucleotides between the ACGGAU loop and the stem leads to a 10-fold decrease in affinity (compare SL1 with 20 nt-RBS, which have identical loop and stem sequences except for the two inserted nucleotides; Figure 3d). On the one hand, this insertion leads to a lower number of hydrogen bond contacts from the protein side-chains with the functional groups of the bases in the major groove of the stem when



**Figure 3.** Effect of the two additional G5 and A12 nucleotides inserted between the stem and the loop in SL1. G5 and A12 in SL1 shift down the CG-UA stem closing base pairs by one register. (d) This disrupts hydrogen bonds that would be present between the protein and the major groove of the 20 nt-RBS (e.g. to Thr5<sub>A</sub>), compare (c) and (f). However, other hydrogen bonds between the SL1 RNA and the protein are formed due to the looping out of the G5 base (f, right). The binding of the two adenines in ANGGAX [A8 and A12 in 20 nt-RBS, (b)] is cooperatively supported by their stacking onto the stem (closing with the C7-G14 base pair), (a). In SL1, this stabilization by stacking of the two adenines (A6 and A10 in SL1) is not possible, because of the looped-out G5 (e). For the 20 nt-RBS, the two  $K_d$  values correspond to the binding of the first and second RNA molecule to the homo-dimeric RsmE protein, respectively.

the GC-UA closing base pair is shifted down by one register in SL1 compared with the 20 nt-RBS (compare Figure 3c and f). For example, the hydrogen bond of the threonine Thr5<sub>A</sub> side-chain with the carbonyl O6 of the guanine of the CG stem closing base pair is absent in SL1. However, this loss in hydrogen bond contacts between protein and the RNA major groove of SL1 is compensated by new hydrogen bonds of the looped-out G5 base with the side-chains of Arg6<sub>A</sub> and Glu46<sub>B</sub> (Figure 3f, right). Although the number of hydrogen bonds between the protein and the RNA cannot explain the affinity difference between SL1 and the 20 nt-RBS RNA for RsmE, the lower affinity of SL1 for RsmE likely arises from the fact that the looping out of the G5 base in SL1 results in a loss of the favourable stacking interactions with the first adenine (A6) of the A(N)GGAX motif and with C4 in the stem (Figure 3e), contacts which are present in all other SLs (see stacking of C7 onto A8 and U6 in the 20 nt-RBS, Figure 3a). This stacking interaction might be important for positioning and cooperatively stabilizing the A(N)GGAX adenine (A8 in the 20 nt-RBS RNA) at an optimal place for forming strong hydrogen bond interactions with the backbone of the protein (Figure 3b).

#### Adaptation of protein side-chains or RNA conformation for recognition of different RNA sequences by the same protein

While the presence of a looped-out nucleotide N in A(N)GGAX or the insertion of two nucleotides between the ANGGAX hexa-loop and the stem influence the binding affinity for RsmE by one to three orders of magnitude, several RNA constructs have comparable affinities despite significant structural differences (low micromolar affinity, Figure 1). SL3 and SL4 have different nucleotides in the upper part of their stems, whereas the 9 nt-GGA<sub>39-41</sub> RNA is even missing the stem entirely. Comparing the structures of these RNA constructs shows that not only the protein side-chains, but also the RNA can adapt to each other in such a way that various RNA sequences can be bound by RsmE with similar thermodynamic stability.

The 5'-nucleotide of the penultimate stem closing base pair is specifically recognized by the glutamine Gln29<sub>B</sub> side-chain independent of its identity (Figure 4a-c). In SL2, SL4 and 20 nt-RBS, the uracil base of the penultimate stem closing base pair is hydrogen bonded by its O4 carbonyl to the Gln29<sub>B</sub> HE2 protons (Figure 4d), whereas the same protons are contacting the N7 of the guanine G46 in SL3 (Figure 4e). Although we have not solved a structure containing an adenine at this position, it is likely that an adenine would also be recognized by its N7. The SELEX-derived consensus sequence has selected an adenine at this position (14). In contrast, in SL1 the OE1 carbonyl group of the Gln29<sub>B</sub> side-chain rather than its HE2 protons contact the amino group of the cytosine C4 (Figure 4f). This nicely demonstrates that the same side-chain can recognize any base through a slight movement or rotation, thereby interacting with a different functional group. Although the glutamine Gln29<sub>B</sub> side chain recognizes the functional groups of bases, the resulting hydrogen bonds do not contribute to any base

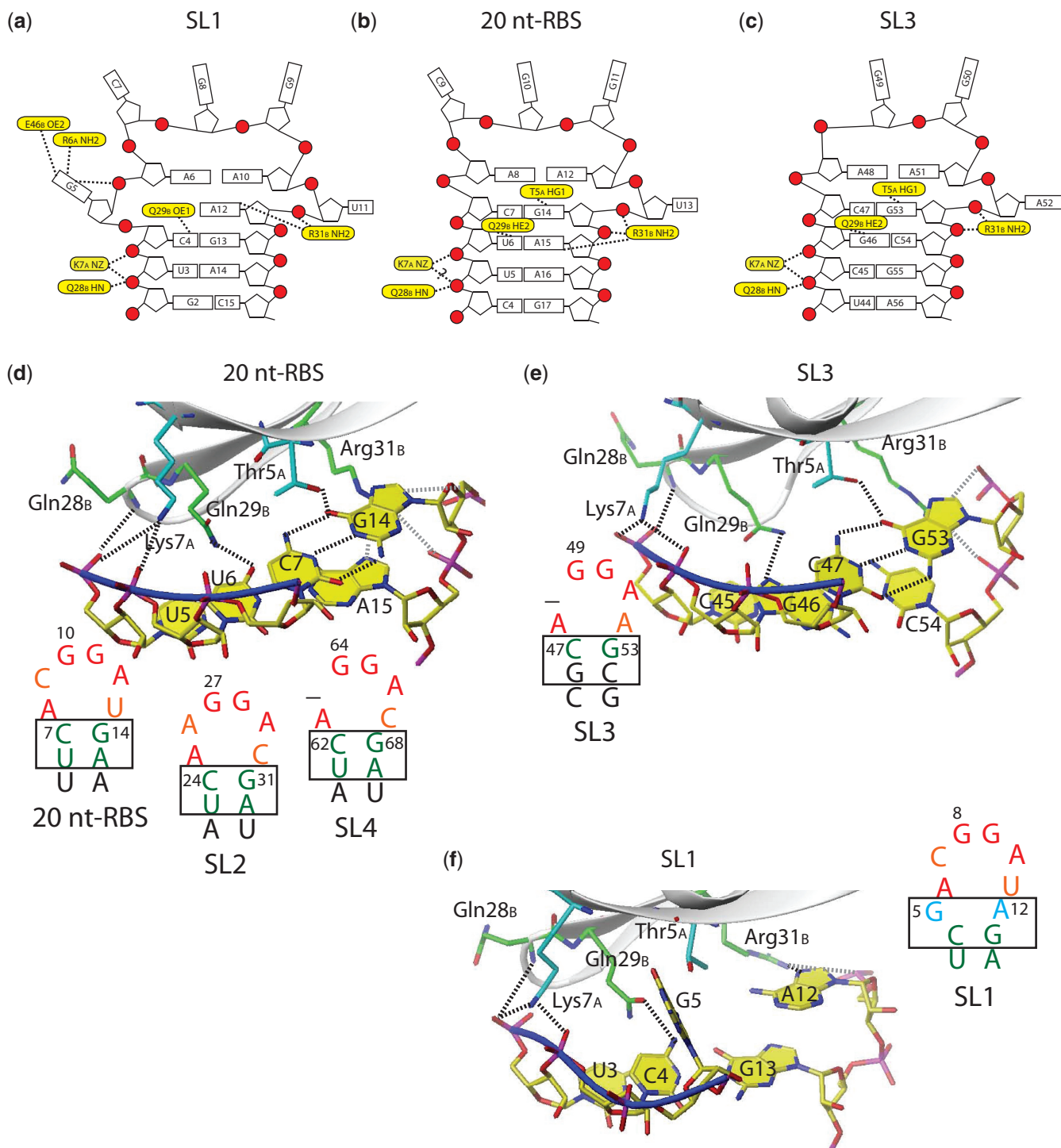
discrimination at this position. Similarly, the arginine Arg31<sub>B</sub> protein side-chain can adapt its conformation to form hydrogen bonds to different RNA bases located in the stem (Figure 4 and Supplementary Text).

Not only can the protein side-chains adapt to recognize different RNA sequences but so can the RNA. In SL2 and 20 nt-RBS, arginine Arg44<sub>B</sub> contacts the phosphate backbone of the RNA loop but is not involved in stacking interactions with the RNA (Figure 5b). In contrast, in SL1, SL3, SL4 and the 9 nt-GGA<sub>39-41</sub> RNA, Arg44<sub>B</sub> stacks on the adenine of the A(N)GGAX motif (Figure 5a, c and d). This is possibly due to a slight rearrangement of this adenine base, when either no looped out nucleotide N is present (SL3 and SL4), no stem is present (9 nt-GGA<sub>39-41</sub> RNA) or two additional nucleotides are inserted between the stem and the A(N)GGAX loop (SL1). Intriguingly, in the single-stranded 9 nt-GGA<sub>39-41</sub> RNA, the Arg44<sub>B</sub> side-chain is not only stacking on the adenine A38 of the A(N)GGAX motif, but is also forming a hydrogen bond to the A38 2'-hydroxyl group (Figure 5d). This is only possible because the sugar pucker conformation is in C2'-*endo* and not C3'-*endo* like in all stem-loop RNA targets (Figure 5d). Experimental NMR evidence for the C2'-*endo* conformation comes from the 2D <sup>1</sup>H-<sup>1</sup>H TOCSY spectrum showing strong correlation peaks between H1' and both H2' and H3' (25). In addition, the chemical shifts of the C1' and C4' of adenine A38 are clearly indicating a C2'-*endo* conformation (38,39). Overall, the structures of both the RNA and the protein can adapt their structure such that RsmE can bind to different RNA targets with similar affinity.

#### GGA motifs within secondary structures bind with mM affinity but show slow exchange with respect to the NMR time scale

We determined structures of GGA motifs located either in loops of hairpins or single-stranded regions with affinities for RsmE covering more than two orders of magnitude (~10–3500 nM). Yet, other GGA motifs are partially or entirely buried within secondary structures and would therefore not be expected to bind the CsrA/RsmE protein. Two such GGA motifs are present in SL1 of the 5'-UTR of the *henA* mRNA in *P. fluorescens* (17). As NMR spectroscopy allows detecting weak interactions, we individually titrated two mutant RNA constructs containing either of the two GGA motifs into RsmE. We detected resonances with chemical shifts characteristic of complex formation (Figure 6 and Supplementary Figure S1). Strikingly, however, by titrating two equivalents of the mutant RNAs into one equivalent of RsmE protein homo-dimer, we could simultaneously observe the free and the bound protein amide resonances and notably also the free and bound RNA imino resonances (Figure 6). This shows that first, the dissociation constants of these complexes are in the range of the protein and RNA concentrations used during the NMR chemical shift titration (~0.5 mM) and second, that the complexes are in the slow exchange regime compared with the NMR chemical shift time scale. This finding is unexpected because low-affinity

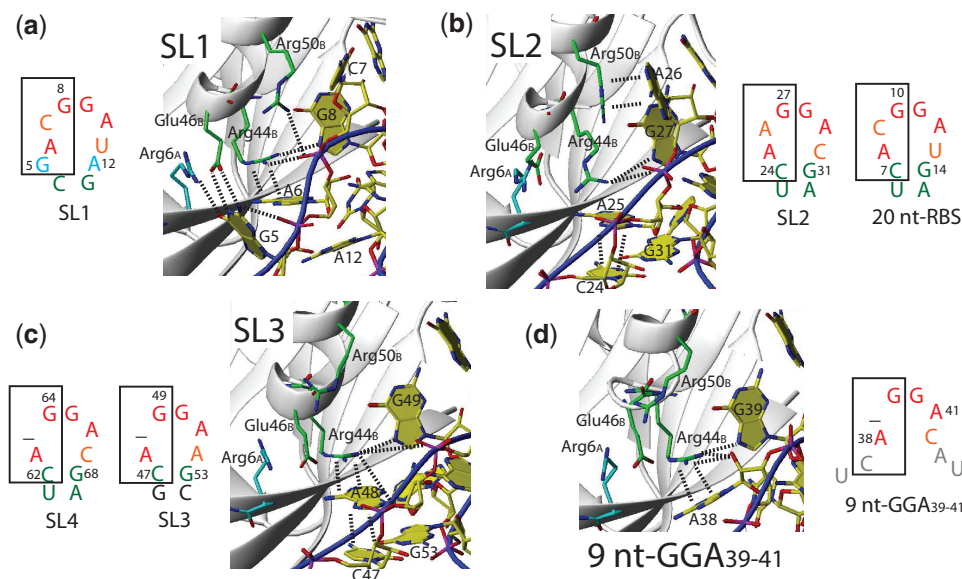




**Figure 4.** Differential RNA stem recognition. (a–c) Schematic representation of intermolecular RNA–protein interactions, yellow residues are involved in potential hydrogen bond contacts. (d–f): Structural details of interactions important for recognition. (d) 20nt-RBS (which shows same contacts as SL2 and SL4), (e) SL3. (f) SL1.

protein–RNA complexes are usually found to be in the fast exchange regime due to a high rate of complex association and dissociation [ $k_{on}$  in the order of  $10^6$ – $10^8$ /Ms (25,40)]. Thus, a low-affinity complex being in slow exchange on the NMR chemical shift time scale implies a small apparent association rate (41). A small apparent association rate can be rationalized by the following two scenarios: the

RsmE protein can only bind to a single-stranded GGA motif (presented as single-stranded region or hairpin loop). Thus, binding of RsmE to a GGA motif partially or entirely buried within a secondary structure can only occur during the short time window in which the secondary structure is transiently opened and thus, is present in an accessible, binding competent state (conformational



**Figure 5.** Differential hairpin loop recognition. (a–d): Structural details of interactions important for recognition. (a) SL1, (b) SL2 (which shows same contacts as 20 nt-RBS). (c) SL3 (which shows same contacts as SL4). (d) 9 nt-GGA<sub>39-41</sub>.

selection). This decreases the fraction of productive complex formation events and thereby the apparent association rate. The second possibility is that RsmE binds non-specifically to the buried GGA-binding motif, which is followed by a slow structural rearrangement (induced fit) forming the cognate protein–RNA complex, again having a small apparent association rate as a consequence.

In summary, we demonstrate that not only the sequence surrounding a certain GGA motif contributes to the modulation of the binding affinity for RsmE but that the secondary structural context influences the range of binding affinities of the Csr/RsmE protein–RNA recognition system to extend over five orders of magnitude.

#### Framework for predicting binding affinity of GGA motifs for CsrA/RsmE

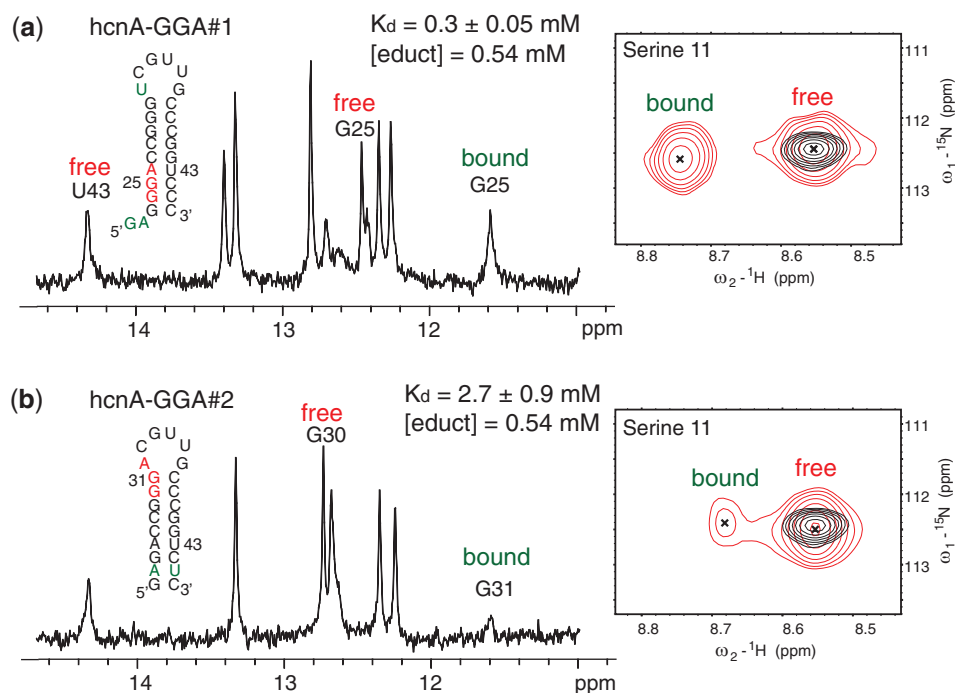
Our study allows the proposal of a high affinity binding consensus sequence, which is defined by a hexa-loop of the form ANGGAX placed directly on top of a stem (Figure 7a). Modifying this high affinity binding motif leads to a decrease in binding affinity. The occlusion of the GGA motif within a base paired region has the largest impact in reducing the binding affinity. Smaller variations in binding affinity (10 to 1000-fold) can be achieved by omitting the nucleotide N in A(N)GGAX (SL3 and SL4), inserting additional nucleotides between the ANGGAX hexa-loop and the stem (SL1) or entirely lacking a stem (23 nt-GGA<sub>76-78</sub> and 9 nt-GGA<sub>39-41</sub>, Figure 7a).

Having a framework for predicting the affinity of a particular GGA motif for the CsrA/RsmE protein, we asked ourselves if we could predict the binding of RsmE to other mRNAs and sRNAs eventually making predictions of the effect of certain sRNA or mRNA mutations on translation activation or repression, respectively.

First, we aimed to predict the binding of RsmE to the 5'-UTR of the *hcnA* mRNA from *P. fluorescens*

(Figure 7b). Aside from two GGA motifs binding with low mM affinity (Figure 6), the 5'-UTR contains three additional GGA motifs. On the basis of the GGA motif sequences, we expect that the GGA#3, #4 and #5 motifs in isolation bind RsmE with intermediate, intermediate–low and with high affinity, respectively. These predictions were verified by ITC (Figure 7b and Supplementary Figure S2). We performed NMR chemical shift titrations to observe the binding of <sup>15</sup>N-labelled RsmE to an unlabelled minimal fragment of the *hcnA* mRNA 5'-UTR containing the three GGA#3–#5 motifs. Unsurprisingly, the first RsmE protein dimer binds simultaneously to GGA#3 and GGA#5, whereas a second RsmE dimer binds to GGA#4, which is the weakest binding motif. These findings suggest that one RsmE dimer binding to both GGA#3 and GGA#5, which overlaps with the RBS, is responsible for translation repression of the *hcnA* mRNA. These observations are in agreement with previous *in vivo* results demonstrating that mutations in GGA#3 and #5 significantly reduced translation repression by RsmE, while mutations in GGA#4 only slightly affected repression (11,17).

Next, we investigated the effect of single point mutations in different GGA motifs of the sRNA RsmX on translation activation of the *hcnA* mRNA of *P. fluorescens* (Figure 7c). The RsmX sRNA in *P. fluorescens* contains six GGA motifs which are predicted to include one high, two intermediate and three low affinity sites for RsmE. By positioning the 5'-UTR of *hcnA* mRNA upstream of the reporter gene CAT, we tested by cell-free expression the activation potential of wild-type RsmX compared with three RsmX mutants in which either the predicted high, intermediate or low affinity GGA motif was mutated in order to abolish its binding. Notably, mutating the site predicted to bind with high affinity strongly reduced translation activation, while the predicted intermediate and low affinity site mutants had only a slight or no effect



**Figure 6.** Low affinity GGA motifs are in slow exchange on the NMR chemical shift time scale. **(a and b)** Determination of the binding affinity of the two GGA motifs located in SL1 of the 5'-UTR of the *hcnA* mRNA ( $K_d$ 's in the mM range) could be performed by NMR because the resonances are in the slow exchange regime permitting to observe the free RNA, the free protein and the complex at the same time (for details see experimental part). Shown is the imino RNA region of the one-dimensional proton NMR spectrum (left) and the serine 11 amide region of the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum (right) of the *hcnA*-GGA#1 (a) or the *hcnA*-GGA#2 RNA (b) in complex with RsmE. In the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum, the free RsmE protein (black) is superimposed onto the protein-RNA complex (red). The full spectra are shown in Supplementary Figure S1. The initial educt concentrations were 0.54 mM for the RNA and the protein (monomer concentration). All the spectra were measured at 313 K. The GGA binding motifs are coloured in red, mutations or non-native nucleotides are coloured in green. For the *hcnA*-GGA#1, a GA dinucleotide was placed at the 5'-end to prevent quadruplex formation observed when omitting the GA. To conserve the secondary structure in the *hcnA*-GGA#2 RNA, the mutation of the GGA to AGA was compensated with a mutation of C45 to U45 in the opposite strand of the stem.

on translation activation, respectively. These observations establish that strong CsrA/RsmE binding sites on sRNAs more efficiently sequester the RsmE protein from the RBS of mRNAs and are therefore more competent in activating translation initiation.

In conclusion, these two examples show that our framework for correlating a certain GGA motif to its binding affinity for CsrA/RsmE can be used to successfully predict the function/effect of certain GGA motifs/mutations on translation activation or repression.

## DISCUSSION

We have elucidated and compared the solution complex structures of RsmE bound to six different target RNAs, which are the four SLs and the single-stranded region between SL2 and SL3 of the sRNA RsmZ (this study) and a 20 nt RNA SL encompassing the RBS of the *hcnA* mRNA in *P. fluorescens* (11). All RNA target sequences contain a common  $\text{A(N)GGAX}$  binding motif but the context of the sequence modulates the binding affinity for RsmE by more than two orders of magnitude. When the GGA motif lies partially or entirely within a base paired region, the affinity can further decrease by two to three orders of magnitude, resulting in a total modulation of the binding affinity of the Csr/Rsm system to more than

five orders of magnitude. The degree of binding affinity attenuation is given by the extent by which the RNA secondary structure has to be disrupted in order to bind RsmE (compare in Figure 7 the three low affinity binding constructs). Remarkably, all the tested GGA binding motifs are in slow exchange in respect to the NMR chemical shift time scale independent of their binding affinity. This can be explained in that the rate of complex dissociation is slow ( $<10/\text{s}$ ) for all complexes of this family of protein-RNA interactions and that the binding affinity covering five orders of magnitude is significantly affected by the variable rate of complex association. The slow rate of complex dissociation is provided by the conserved  $\text{A(N)GGAX}$  binding motif, which is recognized by many hydrogen bonds to the backbone of the RsmE protein. As this recognition involves binding of the nucleobases by the backbone and not the side-chains of the protein, this recognition is specific and these bases cannot be replaced by other ones and still be accommodated by the given protein fold. In contrast to the core  $\text{A(N)GGAX}$  motif, the nucleotides N and X and the nucleotides adjacent to the core motif are variable and allow for a modulation of the binding affinity.

We present a toolkit for predicting the binding affinity of different GGA motifs for the Csr/Rsm proteins (Figure 7a). While a  $\text{ANGGAX}$  hexa-loop directly





placed on top of a stem contributes to high affinity binding, omitting the nucleotide N in A(N)GGAX (SL3 and SL4), inserting additional nucleotides between the ANGGAX hexa-loop and the stem (SL1), removing the entire stem (GGA<sub>76-78</sub> and GGA<sub>39-41</sub>) or more severely, occluding GGA motifs within secondary structures lead to lower binding affinity. Interestingly, AGGGA and AAGGA pentaloop sequences have very recently been shown to have high affinity for RsmE, thus suggesting that omitting the nucleotide X in A(N)GGAX when the nucleotide N is a purine also constitutes a high affinity binding target for CsrA/RsmE (42).

Unexpectedly, omitting a stem when N in A(N)GGAX is missing does not change the binding affinity considerably. The single-stranded 9 nt-GGA<sub>39-41</sub> RNA has a similar affinity like SL4, although the 9 nt-GGA<sub>39-41</sub> RNA contains an identical loop sequence but does not have a stem (Figure 7a) and therefore lacks all the hydrogen bond contacts with the RsmE protein provided by the stem. Furthermore, it is expected that the binding of a single-stranded RNA is entropically disfavoured compared with binding of a pre-formed stem-loop RNA. Yet, this discrepancy can be explained at least partially by the additional H-bond from Arg44 to the A38 2'-hydroxyl group which is cooperatively enforced by the Arg44/A38 stacking interaction. The importance of this interaction is supported by a recent finding that a single-stranded DNA (ssDNA) (missing the 2'-hydroxyl group) containing a GGA motif did not bind the homologous CsrA protein, whereas a GGA motif located in the loop of a hairpin DNA did (43). Interestingly, this additional H-bond can only form because of a structural adaptation of the RNA. The adenine A38 of the A(N)GGAX in the single-stranded 9 nt-GGA<sub>39-41</sub> RNA has a C2'-endo sugar pucker conformation, while an equivalent adenine stacking on a stem has a sugar pucker C3'-endo conformation. RNA adaptation leading to recognition of different RNA sequences with equal affinity has been demonstrated recently. The solution structures of the oligonucleotides 5'-UCCAGU-3' and 5'-UGGAGU-3' in complex with the RRM domain of SRSF2 revealed that the cytosines having an 'anti'-base conformation are recognized very similarly as the guanines having a *syn* conformation (44). One arginine side-chain recognizes either the Watson-Crick edge of the cytosine or the Hoogsteen edge of the guanine.

Interestingly, our structures reveal that also adaptation of the protein side-chains allows the recognition of different RNA sequences. For example, Gln29 can recognize all four bases by a simple adjustment of its side-chain position and orientation. Alternative side-chain conformations have also been suggested to contribute to the degenerate recognition of polypyrimidine tracts by U2AF<sup>65</sup> (45), the recognition of multiple RNA targets by the specific fem-3 binding factor (FBF) protein of the PUMILIO/FBF (PUF) family (46) or three different RNA targets by Lin28 (47). Very recently, the X-ray crystal structures of Pot1pC in complex with its cognate and several non-cognate ssDNA ligands have nicely demonstrated that changes in nucleic acid and protein structure (such as rotation of bases or protein side-chains)

allows for alternative H-bond networks or stacking interactions and thus for the accommodation of several different DNA sequences by the same protein with thermodynamic equivalence (48). Thus, adaptation of protein and nucleic acid seems to be a general mechanism contributing to the binding affinity in protein-nucleic acid recognition, without conferring specificity, supporting recent findings that non-specific and specific RNA-binding modes may not differ fundamentally (49).

Notably, despite their distinct protein folds, the orthologues RsmE protein from *P. fluorescens* (this study) and the RsmN protein from *P. aeruginosa* (12) bind SL2 from the corresponding RsmZ sRNA in an almost identical fashion (Supplementary Figure S6). Simply, the looped-out adenine (corresponding to nucleotide N of the A(N)GGAX motif) has a different structure. This is due to the  $\alpha$ -helix, which is only present in RsmE but not in RsmN.

In contrast to the CsrA/RsmE system, in which the protein binding surface is almost identical in all the homologous proteins and the variety of RNA sequences are recognized by the same protein by adaptation of protein side-chain and RNA conformations, the coat proteins of single-stranded bacteriophages recognize their cognate RNA targets by very distinct binding modes. The co-evolution of coat protein and corresponding RNA structure ensures that each coat protein specifically binds only its cognate RNA target and discriminates against other RNA targets (50,51).

Fine-tuning of the binding affinity for CsrA/RsmE over more than five orders of magnitude allows RNA (mRNA or the competing sRNA) to adjust its affinity according to its specific function in the cell. Different affinities of mRNAs for CsrA/RsmE allow for a fine-tuning of translation repression if CsrA/RsmE binds to the RBS or close to it. Strong binding of CsrA/RsmE to the RBS increases competition with the 30S ribosomal subunit for binding to the RBS, hence stronger translation repression (Figure 7b) (11,17). Likewise, strong CsrA/RsmE binding sites on a sRNA result in a more efficient translation de-repression by an increased sequestration capability of the sRNA (Figure 7c). Furthermore, different binding affinities for CsrA/RsmE could also modulate the lifetime of mRNAs or sRNAs, which are stabilized from degradation upon CsrA/RsmE binding. The CsrA protein has been shown to increase the lifetime of the *flhDC* mRNA by binding to its 5'-UTR (52,53) and mutations of three or five out of the seven GGA motifs in the sRNA RsmY in *P. fluorescens* have lowered its stability significantly (20). We speculate that strong binding of CsrA/RsmE to a potential RNase E cleavage site would protect the RNA more efficiently against endonucleolytic attack because CsrA/RsmE would better compete with RNase E for the binding/cleavage site.

Considering the estimated total concentrations of the CsrA protein in the range of 6–17  $\mu$ M in *E. coli* (54), it is legitimate that the GGA motif overlapping the RBS (nM affinity) is binding RsmE under physiological conditions (11,17). In contrast, the two GGA motifs located partially within the stem of SL1 of the *hcnA* mRNA are not assumed to be relevant for binding RsmE *in vivo*

(dissociation constants of 300  $\mu\text{M}$  and 2.7 mM). It has been shown that chromosomally transcribed mRNAs are not homogeneously expressed in the cellular space of *E. coli* (55). Thus, locally, the concentration of CsrA might largely exceed concentrations in the low to intermediate  $\mu\text{M}$  range and binding might also be relevant with  $K_d$  values in the low mM range. Interestingly, mutations disrupting these GGA motifs had an impact on *hcnA* gene expression (17). It is possible that these mutations simply destabilize SL1 leading to a reduced stability of the *hcnA* mRNA (56). Another possibility is that modulation of the secondary structure of the *hcnA* mRNA could lead to an exposure of the buried GGA motifs and thus significantly increase their affinity for the CsrA/RsmE protein. *In vivo*, the secondary structure (or even the tertiary structure) of an mRNA could be modulated by several factors that could vary in a cell state-specific manner. It is credible that sRNAs base pair with the 5'-UTR of the mRNA as has been proposed for RsmY and the *hcnA* mRNA 5'-UTR in *P. fluorescens* (20). In addition, it is conceivable that protein factors such as Hfq (57), small molecules (58), ions (e.g.  $\text{Mg}^{2+}$ ) (59), the temperature (60) or co-transcriptional folding (61) could modulate the accessibility of buried GGA motifs by stabilization of alternative secondary structures. Specifically, the possibility that an mRNA is transcribed in the presence of CsrA/RsmE could allow the binding of the protein to a GGA motif for which the complementary secondary structural element has not been transcribed yet. A rearrangement of the entire secondary structure and the binding of otherwise inaccessible GGA binding motifs could be the consequence.

In conclusion, this work exemplifies the enormous diversity in the Csr/Rsm protein-RNA recognition and provides the basis for the next level of binding complexity, which is the assembly of several CsrA/RsmE homodimeric proteins on a sRNA or mRNA containing more than one GGA motif on a single molecule.

## ACCESSION NUMBERS

Atomic coordinates, NMR chemical shifts and NMR restraints for the structures of the SL1/RsmE, SL2/RsmE, SL3/RsmE, SL4/RsmE and 9 nt-GGA<sub>39-41</sub>/RsmE complexes have been deposited in the Protein Data Bank under accession codes 2mfc, 2mfe, 2mff, 2mfg and 2mfh, respectively.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We thank David Witmer for the help with RNA production; Dr Markus Blatter for help with structure calculations; Prof. Gerhard Wider, Thea Stahel, Dr Christophe Maris and Dr Fred Damberger for help with NMR spectroscopy setup; Dr Fionna Loughlin, Irene Beusch and Georg Braach-Dorn for reading the manuscript;

Prof. Karine Lapouge and Prof. Dieter Haas for helpful discussions.

O.D. prepared protein and RNA samples for NMR spectroscopy, collected and analysed the NMR experiments, performed structure calculations, conducted and analysed ITC experiments, designed the project and the experiments and wrote the manuscript. E.M. designed, performed and analysed *in vitro* translation assays. N.D. prepared initial protein and RNA samples. M.S. performed initial NMR experiments. F.H.-T.A. designed and supervised the study, secured funding and wrote the manuscript. All authors discussed the results, commented on and approved the manuscript.

## FUNDING

The Swiss National Science Foundation [3100A0-118118, 31003ab-133134 and 31003A-149921 to F.H.-T.A]; the SNF-NCCR structural biology Iso-lab. Funding for open access charge: NAR Editorial Board members are entitled to one free paper per year in recognition of their work on behalf of the journal.

*Conflict of interest statement.* None declared.

## REFERENCES

- Waters, L.S. and Storz, G. (2009) Regulatory RNAs in bacteria. *Cell*, **136**, 615–628.
- Storz, G., Vogel, J. and Wassarman, K.M. (2011) Regulation by small RNAs in bacteria: expanding frontiers. *Mol. Cell*, **43**, 880–891.
- Gripenland, J., Netterling, S., Loh, E., Tiensuu, T., Toledo-Arana, A. and Johansson, J. (2010) RNAs: regulators of bacterial virulence. *Nat. Rev. Microbiol.*, **8**, 857–866.
- Vogel, J. and Luisi, B.F. (2011) Hfq and its constellation of RNA. *Nat. Rev. Microbiol.*, **9**, 578–589.
- Aiba, H. (2007) Mechanism of RNA silencing by Hfq-binding small RNAs. *Curr. Opin. Microbiol.*, **10**, 134–139.
- Wagner, E.G. (2009) Kill the messenger: bacterial antisense RNA promotes mRNA decay. *Nat. Struct. Mol. Biol.*, **16**, 804–806.
- Romeo, T., Vakulskas, C.A. and Babitzke, P. (2013) Post-transcriptional regulation on a global scale: form and function of Csr/Rsm systems. *Environ. Microbiol.*, **15**, 313–324.
- Lapouge, K., Schubert, M., Allain, F.H. and Haas, D. (2008) Gac/Rsm signal transduction pathway of gamma-proteobacteria: from RNA recognition to regulation of social behaviour. *Mol. Microbiol.*, **67**, 241–253.
- Papenfort, K. and Vogel, J. (2010) Regulatory RNA in bacterial pathogens. *Cell Host Microbe*, **8**, 116–127.
- Timmermans, J. and Van Melderen, L. (2010) Post-transcriptional global regulation by CsrA in bacteria. *Cell. Mol. Life Sci.*, **67**, 2897–2908.
- Schubert, M., Lapouge, K., Duss, O., Oberstrass, F.C., Jelesarov, I., Haas, D. and Allain, F.H. (2007) Molecular basis of messenger RNA recognition by the specific bacterial repressing clamp RsmA/CsrA. *Nat. Struct. Mol. Biol.*, **14**, 807–813.
- Morris, E.R., Hall, G., Li, C., Heeb, S., Kulkarni, R.V., Lovelock, L., Silistre, H., Messina, M., Camara, M., Emsley, J. *et al.* (2013) Structural rearrangement in an RsmA/CsrA ortholog of *Pseudomonas aeruginosa* creates a dimeric RNA-binding protein, RsmN. *Structure*, **21**, 1659–1671.
- Marden, J.N., Diaz, M.R., Walton, W.G., Gode, C.J., Betts, L., Urbanowski, M.L., Redinbo, M.R., Yahr, T.L. and Wolfgang, M.C. (2013) An unusual CsrA family member operates in series with RsmA to amplify posttranscriptional responses in *Pseudomonas aeruginosa*. *Proc. Natl Acad. Sci. USA*, **110**, 15055–15060.



14. Dubey, A.K., Baker, C.S., Romeo, T. and Babitzke, P. (2005) RNA sequence and secondary structure participate in high-affinity CsrA-RNA interaction. *RNA*, **11**, 1579–1587.
15. Brencic, A. and Lory, S. (2009) Determination of the regulon and identification of novel mRNA targets of *Pseudomonas aeruginosa* RsmA. *Mol. Microbiol.*, **72**, 612–632.
16. Mercante, J., Edwards, A.N., Dubey, A.K., Babitzke, P. and Romeo, T. (2009) Molecular geometry of CsrA (RsmA) binding to RNA and its implications for regulated expression. *J. Mol. Biol.*, **392**, 511–528.
17. Lapouge, K., Sineva, E., Lindell, M., Starke, K., Baker, C.S., Babitzke, P. and Haas, D. (2007) Mechanism of hcnA mRNA recognition in the Gac/Rsm signal transduction pathway of *Pseudomonas fluorescens*. *Mol. Microbiol.*, **66**, 341–356.
18. Babitzke, P. and Romeo, T. (2007) CsrB sRNA family: sequestration of RNA-binding regulatory proteins. *Curr. Opin. Microbiol.*, **10**, 156–163.
19. Heeb, S., Blumer, C. and Haas, D. (2002) Regulatory RNA as mediator in GacA/RsmA-dependent global control of exoproduct formation in *Pseudomonas fluorescens* CHA0. *J. Bacteriol.*, **184**, 1046–1056.
20. Valverde, C., Lindell, M., Wagner, E.G. and Haas, D. (2004) A repeated GGA motif is critical for the activity and stability of the riboregulator RsmY of *Pseudomonas fluorescens*. *J. Biol. Chem.*, **279**, 25066–25074.
21. Kay, E., Dubuis, C. and Haas, D. (2005) Three small RNAs jointly ensure secondary metabolism and biocontrol in *Pseudomonas fluorescens* CHA0. *Proc. Natl Acad. Sci. USA*, **102**, 17136–17141.
22. Duss, O., Lukavsky, P.J. and Allain, F.H. (2012) Isotope Labeling and Segmental Labeling of Larger RNAs for NMR Structural Studies. *Adv. Exp. Med. Biol.*, **992**, 121–144.
23. Duss, O., Maris, C., von Schroetter, C. and Allain, F.H. (2010) A fast, efficient and sequence-independent method for flexible multiple segmental isotope labeling of RNA using ribozyme and RNase H cleavage. *Nucleic Acids Res.*, **38**, e188.
24. Zwahlen, C., Legault, P., Vincent, S.J.F., Greenblatt, J., Konrat, R. and Kay, L.E. (1997) Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: Application to a bacteriophage lambda N-peptide/boxB RNA complex. *J. Am. Chem. Soc.*, **119**, 6711–6721.
25. Dominguez, C., Schubert, M., Duss, O., Ravindranathan, S. and Allain, F.H. (2011) Structure determination and dynamics of protein-RNA complexes by NMR spectroscopy. *Prog. Nucl. Magn. Reson. Spectrosc.*, **58**, 1–61.
26. Shen, Y., Delaglio, F., Cornilescu, G. and Bax, A. (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR*, **44**, 213–223.
27. Zhu, G., Xia, Y., Nicholson, L.K. and Sze, K.H. (2000) Protein dynamics measurements by TROSY-based NMR experiments. *J. Magn. Reson.*, **143**, 423–426.
28. Güntert, P. (2004) Automated NMR structure calculation with CYANA. *Methods Mol. Biol.*, **278**, 353–378.
29. Case, D.A., Cheatham, T.E. 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K.M. Jr, Onufriev, A., Simmerling, C., Wang, B. and Woods, R.J. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem.*, **26**, 1668–1688.
30. Case, D.A., Darden, T.A., Cheatham, T.E. III, Simmerling, C.L., Wang, J., Duke, R.E., Luo, R., Merz, K.M., Pearlman, D.A., Crowley, M. et al. (2006) *AMBER 9*. University of California, San Francisco.
31. Wang, J.M., Cieplak, P. and Kollman, P.A. (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.*, **21**, 1049–1074.
32. Bashford, D. and Case, D.A. (2000) Generalized born models of macromolecular solvation effects. *Annu. Rev. Phys. Chem.*, **51**, 129–152.
33. Datsenko, K.A. and Wanner, B.L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA*, **97**, 6640–6645.
34. Michel, E. and Wüthrich, K. (2012) High-yield *Escherichia coli*-based cell-free expression of human proteins. *J. Biomol. NMR*, **53**, 43–51.
35. Farrow, N.A., Muhandiram, R., Singer, A.U., Pascal, S.M., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Forman-Kay, J.D. and Kay, L.E. (1994) Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by <sup>15</sup>N NMR relaxation. *Biochemistry*, **33**, 5984–6003.
36. Ishima, R. (2012) Recent developments in (<sup>15</sup>N) NMR relaxation studies that probe protein backbone dynamics. *Top. Curr. Chem.*, **326**, 99–122.
37. Frederick, K.K., Marlow, M.S., Valentine, K.G. and Wand, A.J. (2007) Conformational entropy in molecular recognition by proteins. *Nature*, **448**, 325–329.
38. Aeschbacher, T., Schmidt, E., Blatter, M., Maris, C., Duss, O., Allain, F.H., Güntert, P. and Schubert, M. (2013) Automated and assisted RNA resonance assignment using NMR chemical shift statistics. *Nucleic Acids Res.*, **41**, e172.
39. Aeschbacher, T., Schubert, M. and Allain, F.H. (2012) A procedure to validate and correct the <sup>13</sup>C chemical shift calibration of RNA datasets. *J. Biomol. NMR*, **52**, 179–190.
40. Auweter, S.D., Fasan, R., Reymond, L., Underwood, J.G., Black, D.L., Pitsch, S. and Allain, F.H. (2006) Molecular basis of RNA recognition by the human alternative splicing factor Fox-1. *EMBO J.*, **25**, 163–173.
41. Latham, M.P., Zimmermann, G.R. and Pardi, A. (2009) NMR chemical exchange as a probe for ligand-binding kinetics in a theophylline-binding RNA aptamer. *J. Am. Chem. Soc.*, **131**, 5052–5053.
42. Lapouge, K., Perozzo, R., Iwaskiewicz, J., Bertelli, C., Zoete, V., Michielin, O., Scapozza, L. and Haas, D. (2013) RNA pentaloop structures as effective targets of regulators belonging to the RsmA/CsrA protein family. *RNA Biol.*, **10**, 1031–1041.
43. Koharudin, L.M., Boelens, R., Kaptein, R. and Gronenborn, A.M. (2013) A NMR guided approach for CsrA-RNA crystallization. *J. Biomol. NMR*, **56**, 31–39.
44. Daubner, G.M., Clery, A., Jayne, S., Stevenin, J. and Allain, F.H. (2012) A syn-anti conformational difference allows SRSF2 to recognize guanines and cytosines equally well. *EMBO J.*, **31**, 162–174.
45. Thickman, K.R., Sickmier, E.A. and Kielkopf, C.L. (2007) Alternative conformations at the RNA-binding surface of the N-terminal U2AF(65) RNA recognition motif. *J. Mol. Biol.*, **366**, 703–710.
46. Wang, Y., Opperman, L., Wickens, M. and Hall, T.M. (2009) Structural basis for specific recognition of multiple mRNA targets by a PUF regulatory protein. *Proc. Natl Acad. Sci. USA*, **106**, 20186–20191.
47. Nam, Y., Chen, C., Gregory, R.I., Chou, J.J. and Sliz, P. (2011) Molecular basis for interaction of let-7 microRNAs with Lin28. *Cell*, **147**, 1080–1091.
48. Dickey, T.H., McKercher, M.A. and Wuttke, D.S. (2013) Nonspecific recognition is achieved in Pot1pC through the use of multiple binding modes. *Structure*, **21**, 121–132.
49. Guenther, U.P., Yandek, L.E., Niland, C.N., Campbell, F.E., Anderson, D., Anderson, V.E., Harris, M.E. and Jankowsky, E. (2013) Hidden specificity in an apparently nonspecific RNA-binding protein. *Nature*, **502**, 385–388.
50. Rumnieks, J. and Tars, K. (2013) Crystal structure of the bacteriophage Qbeta coat protein in complex with the RNA operator of the replicase gene. *J. Mol. Biol.*, **10.1016/j.jmb.2013.08.025**.
51. Chao, J.A., Patskovsky, Y., Almo, S.C. and Singer, R.H. (2008) Structural basis for the coevolution of a viral RNA-protein complex. *Nat. Struct. Mol. Biol.*, **15**, 103–105.
52. Yakhnin, A.V., Baker, C.S., Vakulskas, C.A., Yakhnin, H., Berezin, I., Romeo, T. and Babitzke, P. (2013) CsrA activates flhDC expression by protecting flhDC mRNA from RNase E-mediated cleavage. *Mol. Microbiol.*, **87**, 851–866.
53. Wei, B.L., Brun-Zinkernagel, A.M., Simecka, J.W., Pruss, B.M., Babitzke, P. and Romeo, T. (2001) Positive regulation of motility and flhDC expression by the RNA-binding protein CsrA of *Escherichia coli*. *Mol. Microbiol.*, **40**, 245–256.
54. Gudapaty, S., Suzuki, K., Wang, X., Babitzke, P. and Romeo, T. (2001) Regulatory interactions of Csr components: the RNA binding protein CsrA activates csrB transcription in *Escherichia coli*. *J. Bacteriol.*, **183**, 6017–6027.

55. Montero Llopis,P., Jackson,A.F., Sliusarenko,O., Surovtsev,I., Heinritz,J., Emonet,T. and Jacobs-Wagner,C. (2010) Spatial organization of the flow of genetic information in bacteria. *Nature*, **466**, 77–81.
56. Blumer,C. and Haas,D. (2000) Iron regulation of the hcnABC genes encoding hydrogen cyanide synthase depends on the anaerobic regulator ANR rather than on the global activator GacA in *Pseudomonas fluorescens* CHA0. *Microbiology*, **146**(Pt 10), 2417–2424.
57. Sorger-Domenigg,T., Sonnleitner,E., Kaberdin,V.R. and Blasi,U. (2007) Distinct and overlapping binding sites of *Pseudomonas aeruginosa* Hfq and RsmA proteins on the non-coding RNA RsmY. *Biochem. Biophys. Res. Commun.*, **352**, 769–773.
58. Winkler,W.C. and Breaker,R.R. (2005) Regulation of bacterial gene expression by riboswitches. *Annu. Rev. Microbiol.*, **59**, 487–517.
59. Woodson,S.A. (2005) Metal ions and RNA folding: a highly charged topic with a dynamic future. *Curr. Opin. Chem. Biol.*, **9**, 104–109.
60. Narberhaus,F., Waldminghaus,T. and Chowdhury,S. (2006) RNA thermometers. *FEMS Microbiol. Rev.*, **30**, 3–16.
61. Pan,T. and Sosnick,T. (2006) RNA folding during transcription. *Annu. Rev. Biophys. Biomol. Struct.*, **35**, 161–175.