

# Structural basis of the non-coding RNA RsmZ acting as a protein sponge

Olivier Duss<sup>1</sup>, Erich Michel<sup>1</sup>, Maxim Yulikov<sup>2</sup>, Mario Schubert<sup>1</sup>, Gunnar Jeschke<sup>2</sup> & Frédéric H.-T. Allain<sup>1</sup>

**MicroRNA and protein sequestration by non-coding RNAs (ncRNAs) has recently generated much interest. In the bacterial Csr/Rsm system, which is considered to be the most general global post-transcriptional regulatory system responsible for bacterial virulence, ncRNAs such as CsrB or RsmZ activate translation initiation by sequestering homodimeric CsrA-type proteins from the ribosome-binding site of a subset of messenger RNAs. However, the mechanism of ncRNA-mediated protein sequestration is not understood at the molecular level. Here we show for *Pseudomonas fluorescens* that RsmE protein dimers assemble sequentially, specifically and cooperatively onto the ncRNA RsmZ within a narrow affinity range. This assembly yields two different native ribonucleoprotein structures. Using a powerful combination of nuclear magnetic resonance and electron paramagnetic resonance spectroscopy we elucidate these 70-kilodalton solution structures, thereby revealing the molecular mechanism of the sequestration process and how RsmE binding protects the ncRNA from RNase E degradation. Overall, our findings suggest that RsmZ is well-tuned to sequester, store and release RsmE and therefore can be viewed as an ideal protein ‘sponge’.**

In the last decade, ncRNAs have emerged in both eukaryotes and prokaryotes (in which they are termed small RNAs, or sRNAs) as important regulators in gene expression<sup>1–3</sup>. The bacterial Csr/Rsm system (for carbon storage regulator or repressor of secondary metabolism) is a well-characterized sRNA-based regulatory system that is present in about 75% of all bacterial species and affects over 20% of all mRNAs in *Escherichia coli*<sup>3–6</sup>. The homodimeric CsrA/RsmE proteins bind to the ribosome-binding site (RBS) of a subset of mRNAs and repress translation initiation<sup>5–7</sup>. Triggered by environmental signals, one or several sRNAs (for example CsrB or RsmZ) are transcribed, which can sequester the CsrA/RsmE proteins and therefore de-repress translation initiation. Both the sRNAs as well as the regulated mRNAs contain several GGA protein-binding motifs that bind the CsrA/RsmE proteins<sup>5–8</sup>. Although the RNA-binding surfaces of all the CsrA/RsmE homologues are highly conserved, the sRNAs are very diverse in length (100–500 nucleotides), predicted secondary structures, and numbers of GGA-binding motifs<sup>5,6</sup>.

We previously elucidated the molecular basis of translation repression by solving the solution structure of RsmE bound to a stem-loop RNA encompassing the Shine-Dalgarno sequence of the *hcnA* mRNA of *P. fluorescens*<sup>7</sup>. However, no information is available on the mechanism of protein sequestration by any sRNA. Here we report a combined NMR- and electron paramagnetic resonance (EPR) spectroscopy-based approach that enabled us to visualize and rationalize the cooperative assembly of several RsmE proteins on the sRNA RsmZ from *P. fluorescens* and to obtain the 70-kilodalton (kDa) solution structure of the first 72 nucleotides of RsmZ, RsmZ(1–72), bound to three homodimeric RsmE proteins. These structural data provide detailed insight into the mechanism of translation activation by the sRNA RsmZ and also suggest how coupling with sRNA degradation allows the release of the RsmE protein.

## Ordered RsmE–RsmZ RNP assembly

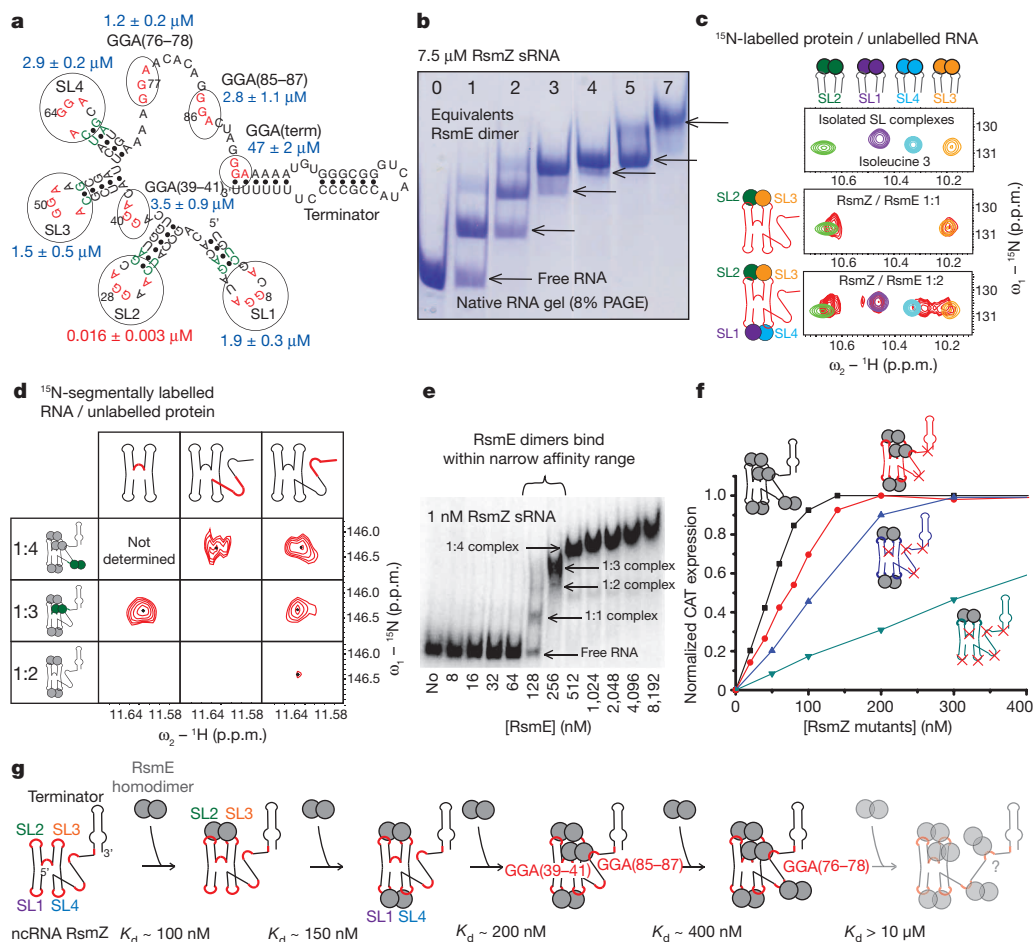
The 127-nucleotide sRNA RsmZ is composed of four stem-loops and a protective rho-independent terminator<sup>9</sup>. It contains eight GGA motifs, four being present in the loop of RNA hairpins and four located in single-stranded regions (Fig. 1a). In isolation, all eight GGA motifs of RsmZ

bind RsmE with dissociation constants ( $K_d$ ) ranging from 16 nM to 50  $\mu$ M (Fig. 1a)<sup>10</sup>. Interestingly, only the GGA motif within stem-loop 2 (SL2) binds RsmE with low-nM  $K_d$ , whereas all other GGA motifs have a  $K_d$  in the  $\mu$ M-range. Considering that RsmE is tightly bound to the RBS of the *hcnA* mRNA<sup>10</sup> ( $K_d = 180 \pm 40$  nM, see Methods), efficient translation activation by RsmZ can only be understood by investigating the binding of RsmE to the entire RsmZ sRNA.

We performed native polyacrylamide RNA gels by incubating the RsmZ sRNA with increasing stoichiometric amounts of RsmE protein. Notably, the sRNA can accommodate sequentially up to five RsmE dimers (see Fig. 1b). To determine in which order the different GGA motifs are bound by RsmE, we followed NMR chemical shift perturbations by titrating increasing stoichiometric amounts of <sup>2</sup>H-<sup>15</sup>N-labelled RsmE protein to the RNA RsmZ. By comparing the <sup>1</sup>H-<sup>15</sup>N heteronuclear single-quantum coherence spectra (HSQC) of RsmE bound to isolated GGA motifs with the spectra of the RsmE dimer bound to RsmZ(1–89) at a 1:1 or 2:1 ratio, we could conclude that the first RsmE protein dimer binds simultaneously the GGA motifs of SL2 and SL3 and the second RsmE dimer the motifs of SL1 and SL4 (Fig. 1c).

A similar approach to monitor the additional RsmE binding sites failed owing to strong spectral overlap of the protein resonances. We then prepared segmentally labelled RNAs<sup>11,12</sup> containing only one GGA motif isotopically labelled in each case (GGA<sub>39–41</sub>, GGA<sub>76–78</sub> or GGA<sub>85–87</sub>). When RsmE binds a GGA motif, the central guanosine imino proton forms an intermolecular hydrogen bond with the protein and becomes detectable by NMR spectroscopy. These experiments (Fig. 1d) revealed that the third RsmE dimer is simultaneously bound by the GGA motif in the single-stranded region between SL2 and SL3 (GGA<sub>39–41</sub>), and by GGA<sub>85–87</sub> in the single-stranded region between SL4 and the terminator (Fig. 1d, g). The fourth dimer is specifically bound by GGA<sub>76–78</sub>. We then estimated the dissociation constants for the binding of the several RsmE proteins using the entire RsmZ sRNA by electrophoretic mobility shift assays (EMSA). We found that the first four dimers all have submicromolar dissociation constants (100–400 nM) whereas the binding of the fifth RsmE protein is not detectable using EMSA (>10  $\mu$ M) (Fig. 1e, g). As all of the isolated binding sites except SL2 have micromolar

<sup>1</sup>Institute of Molecular Biology and Biophysics, ETH Zürich, CH-8093 Zürich, Switzerland. <sup>2</sup>Laboratory of Physical Chemistry, ETH Zürich, CH-8093 Zürich, Switzerland.



**Figure 1 | Sequential and specific binding of the RsmE protein to the RsmZ sRNA.** **a**, SL2 is the only high-affinity binding site in isolation. **b**, Up to five RsmE protein dimers can be accommodated by the RsmZ sRNA. **c**, **d**, Monitoring the binding of the first and second (**c**) and third and fourth (**d**) RsmE dimers by  $^1\text{H}$ - $^{15}\text{N}$  HSQC NMR spectroscopy observing the protein

or RNA side, respectively. **e**, Gel shift assay. **f**, *In vitro* translation assay demonstrating the functional importance for the binding of the four dimers with a submicromolar  $K_d$ . **g**, Schematic representation of RsmZ/RsmE RNP assembly.

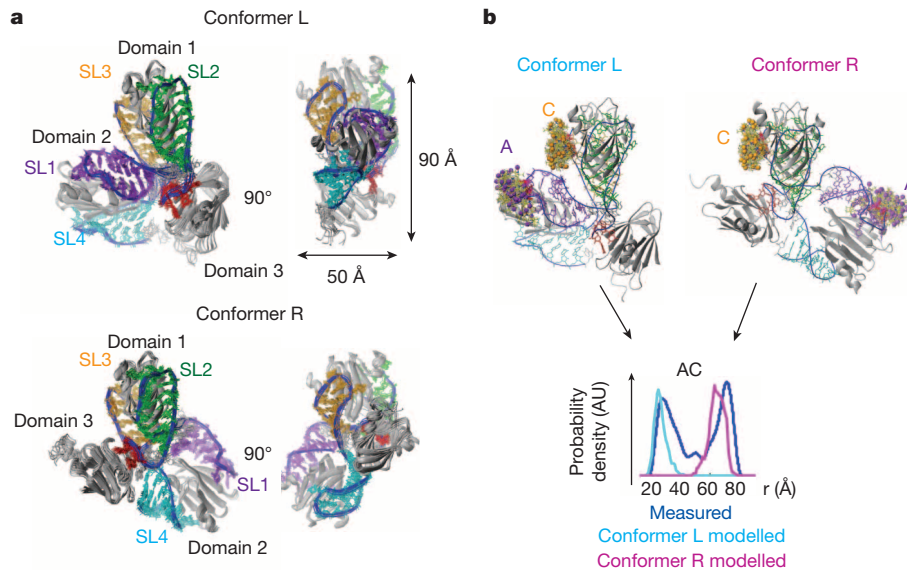
dissociation constants (Fig. 1a), this indicates that RsmE binding to RsmZ is cooperative. To determine which RsmE dimer binding sites on RsmZ are functionally important for activating translation initiation, we performed *in vitro* translation assays using a chloramphenicol acetyltransferase reporter gene fused to the 5'-untranslated region of the *hcnA* mRNA (Fig. 1f). Addition of RsmE at various concentrations of RsmZ sRNAs with mutations in different GGA motifs clearly demonstrated that the binding of all four dimers in the submicromolar affinity range are functionally important to reach optimal translation activation.

### 70-kDa RsmE–RsmZ solution structures

Although ordered, hierarchical and cooperative ribonucleoprotein (RNP) assembly is generally under evolutionary selection<sup>13</sup> as exemplified by the ribosome<sup>14,15</sup>, the spliceosome<sup>16</sup> or telomerase<sup>17</sup> RNP assemblies, it is rather unexpected to find that a sRNA such as RsmZ appears to be bound cooperatively by several RsmE dimers in such a well-defined and sequential manner considering that the primary role of this sRNA should be simply to sequester many CsrA/RsmE proteins. Does such an ordered assembly also lead to a well-defined tertiary structure that could reveal the molecular basis of the cooperative binding? To this end, we determined the 70-kDa solution structure of RsmZ(1–72) in complex with three RsmE protein dimers (Fig. 2a and Supplementary Discussion) using a novel approach for structure determination of large protein–RNA complexes in solution that combines NMR and EPR spectroscopy (Extended Data Fig. 1 and Methods)<sup>18</sup>.

We first determined the structures of the five RsmE–RNA subcomplexes (RsmE bound to the four stem–loops and to the linker between SL2 and SL3, each in isolation) at high resolution using NMR spectroscopy (Extended Data Fig. 1, step 1)<sup>10</sup>. Second, we checked experimentally that each individual subcomplex has the same structure in the full complex by comparing the NMR chemical shifts of the respective complexes (step 2). Third, we measured 21 distances by pulsed EPR spectroscopy between spin labels attached on the RNA of each subcomplex and used them as long-range distance constraints between the subcomplex structures (step 3). This latter step in the structure calculation was particularly difficult as we could see evidence for two conformations simultaneously present in solution, as evidenced by the appearance of two sets of NMR cross-peaks for the RNA resonances in the linker regions (Extended Data Fig. 2b, c and Supplementary Discussion) and by the presence of two well-separated peaks in some EPR distance distributions (Fig. 2b), or by broad EPR distributions that cannot arise from a single static structure (Extended Data Fig. 3b). Nevertheless, we could discriminate the distance constraints belonging to each of the two conformers and obtain well-defined structural ensembles for both conformers (Fig. 2a and Extended Data Table 1).

Both conformations adopt RNA folds that have not, to our knowledge, been reported before. Although four stem–loops are present, no coaxial stacking is observed between any pairs of them, unlike what is found in the structures of large RNAs or RNPs<sup>19</sup>. Rather, the RNPs seem to be arranged in a way that maximizes the distances between the three



**Figure 2 | Solution structures of both conformers of the 70-kDa complex between RsmZ(1–72) and three RsmE homodimers.** **a**, Structural ensembles of both conformations present in solution. **b**, Representative EPR distance

bound RsmE dimers (Fig. 2a). The structure ensembles reveal tertiary interactions that stabilize their global fold such as intermolecular contacts between SL4 and the third RsmE dimer in conformer R or intramolecular interactions between the SL1–SL2 and SL2–SL3 linker residues (Extended Data Fig. 4). Both conformers of the RNP structure have the shape of an oblate disc with a diameter of around 90 Å and a thickness of 50 Å (Fig. 2a). Knowing the detailed assembly pathway helped to explain how this complex adopts two distinct and almost equally populated conformations, which differ in the relative position of domain 2 (SL1 and SL4) and domain 3 in respect to the first RsmE dimer (domain 1) (Fig. 2a). These two conformations cannot interconvert unless RsmE dissociates from SL2 and SL3, strongly suggesting two parallel assembly pathways and a conformational selection mechanism for the binding of the first RsmE dimer (Extended Data Fig. 4g). This means that the determined  $K_d$  values (Fig. 1e) are averaged over the two corresponding conformations (Supplementary Discussion). We can speculate that the presence of two parallel and productive sequestration pathways might provide conformational flexibility for the sRNA, allowing efficient sequestration of RsmE from the RBS in various topological mRNA contexts.

### Positive and negative cooperativity

Almost the entire RsmZ sRNA becomes covered with proteins, illustrating the high binding capacity of this sRNA for the RsmE protein. The first three dimers all bind within a narrow affinity range ( $K_d$  of 100, 150 and 200 nM) despite very different individual binding sites. How is this tuned? The first RsmE protein dimer binds with a  $K_d$  of 100 nM, which might seem a lower affinity than expected considering that binding affinities of RsmE to the individual SL2 and SL3 were measured to be 16 nM and 1.5  $\mu$ M  $K_d$ , respectively (Fig. 1). We can explain this unexpected result by two observations. First, when studying the binding of the high-affinity SL2 in isolation, we saw by isothermal titration calorimetry (ITC) and NMR titration experiments that binding of the first dimer was accompanied by a negative allosteric effect reducing the affinity for binding of the second dimer (Fig. 3a, Extended Data Fig. 5 and Supplementary Discussion). Second, the linker between SL2 and SL3 does not have the optimal length to reach the highest affinity (Fig. 3b). A sRNA mutant in which the SL2–SL3 linker was shortened by three nucleotides resulted in more than threefold higher affinity for RsmE and enhanced translation activation compared to the wild-type linker (Fig. 3b). Overall, an inherent property of the RsmE dimer and the non-optimal linker length between SL2 and SL3 explain how the binding of

distribution indicating the presence of two conformations in solution. Shown are the modelled spin label conformations at the sites A and C in both conformers L and R.

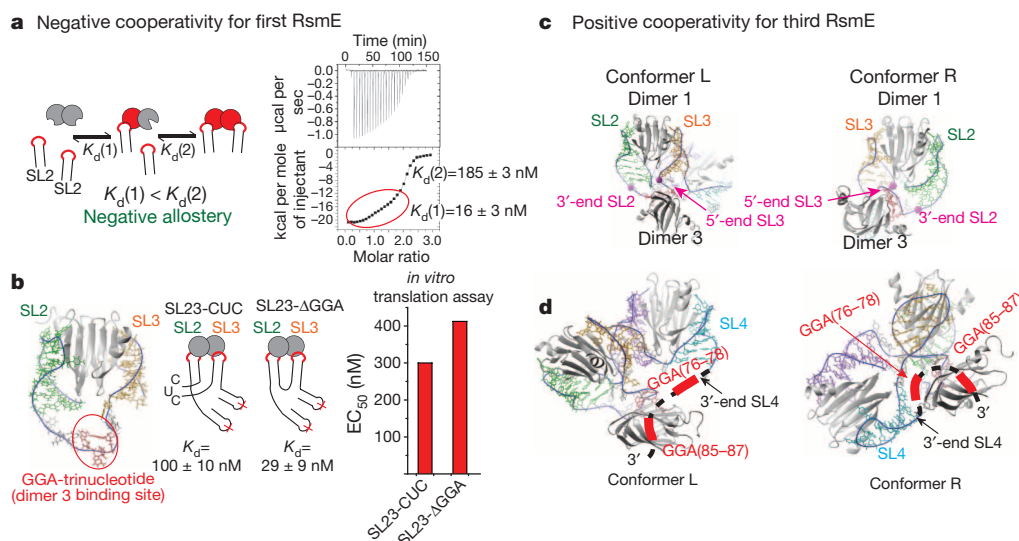
the first dimer is tuned to only 100 nM  $K_d$ . Our findings also suggest that the RNA linker between SL2 and SL3 in RsmZ has evolved to contain an additional binding site (domain 3) rather than to maximize the binding affinity and translation activation potential of the first dimer.

In contrary to the first RsmE dimer, the following two RsmE dimers bind with much higher affinity compared to the binding of isolated sites, indicating a positive cooperativity (Fig. 1). Notably, the binding affinity of the third dimer is strongly enhanced (25-fold, Extended Data Fig. 6a, b) by the binding of the first two dimers, which can be explained by two structural features. First, binding of the first RsmE dimer to SL2 and SL3 fixes the positions of the 5' and 3' ends of the third binding site and thus pre-form the latter (Fig. 3c). Second, in both conformers, binding of the second RsmE dimer to SL4 positions the 3' end of the sRNA such that GGA<sub>85–87</sub> is brought into close proximity to the second binding site on the third RsmE dimer (Fig. 3d). Overall, our findings demonstrate that the sRNA RsmZ has an architecture permitting a specific and sequential protein binding in which the binding of an RsmE dimer pre-organizes the binding of a consecutive RsmE dimer. This precise RNA scaffold is conserved in *Pseudomonas* (Extended Data Fig. 7), supporting the functional importance for the cooperative binding of multiple RsmE proteins within a narrow affinity range of 100–200 nM  $K_d$ .

### RNase protection and RsmE release

The presence of two compact and well-defined structures poses the question of how such a RNP could be turned over once translation activation is no longer required and the cell must return to its initial homeostasis phase. Bacterial RNA turnover is usually initiated by an endoribonucleolytic cleavage performed by RNase E<sup>20,21</sup>. In the free RNA, we could identify three specific cleavage sites that quite remarkably are located near GGA motifs (Fig. 4a and Extended Data Fig. 7e) and that are conserved in *Pseudomonas* (Extended Data Fig. 7d). We could further show that RsmE binding beyond the 1:1 stoichiometry progressively protects RsmZ from degradation by RNase E *in vitro* (Fig. 4b) and in *E. coli* whole-cell extracts containing the entire bacterial degradation machinery (Fig. 4c), consistent with previous *in vivo* findings<sup>22</sup>. In agreement with our RsmE binding data (Fig. 1), four RsmE dimers are required to strongly protect RsmZ from degradation (Fig. 4b–d). Our structural models explain the protection at this 4:1 stoichiometry as all three cleavage sites are bound by the protein (Fig. 4d). Thus, RsmZ binding by several RsmE dimers is not only an efficient way to sequester proteins but also to store them (Fig. 4c). Moreover, the RNA cleavage pattern proposes how RsmE could





**Figure 3 | Fine-tuning of the binding affinity of RsmZ for RsmE results from both negative and positive binding cooperativity.** **a**, A negative allosteric effect is induced upon binding of a first high-affinity SL2 to the RsmE protein homodimer. **b**, The SL23 linker is longer than required to bind dimer 1

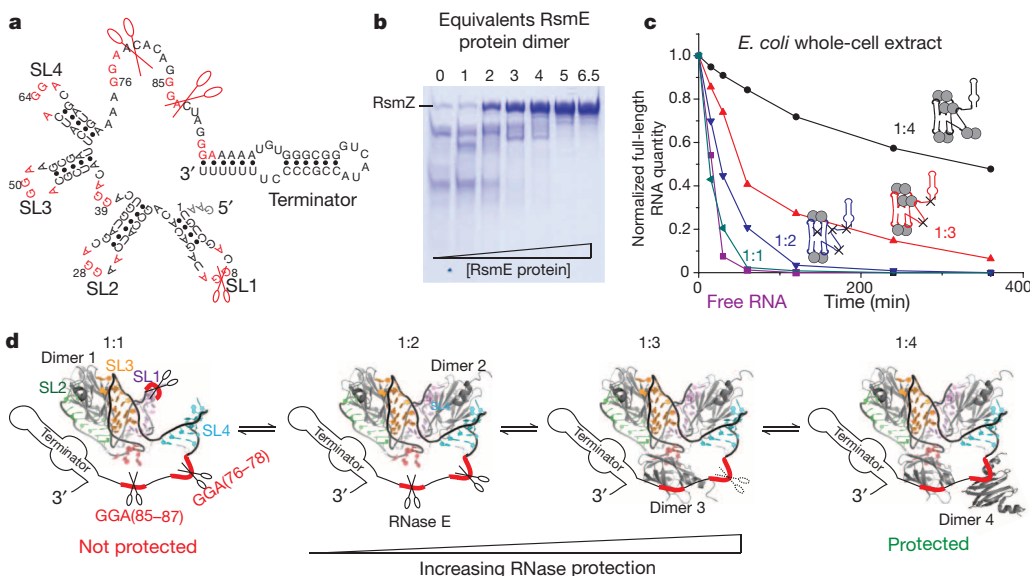
be released from RsmZ. The increased transcription of RsmZ or other RsmE titrating sRNAs or mRNAs would result in dissociation of the RsmZ-bound RsmE and its redistribution onto the newly transcribed sRNAs or mRNAs (at constant RsmE protein concentration<sup>22</sup>). The concomitance of some of the sites for RsmE binding and RNase E cleavage suggests that an RsmE dissociation followed by RNase E cleavage will weaken RsmE re-association and therefore lead to an unidirectional RsmE release (Figs 4d and 5). Therefore, RsmZ can be seen as a true protein ‘sponge’ because this ncRNA can sequester, store and, coupled with RNase E, release proteins (Fig. 5).

**Discussion**

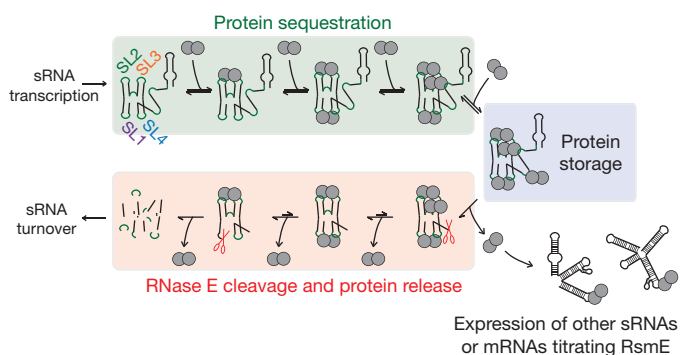
Here, we have structurally investigated how the ncRNA RsmZ sequesters the RsmE protein in order to activate translation. Unexpectedly, RsmZ binds sequentially up to five RsmE dimers, most of them cooperatively within a narrow affinity range (100–200 nM  $K_d$ ). Other sRNAs of the Csr/Rsm family have entirely different secondary structures, resulting in various assembly pathways and three-dimensional structures. As a

consequence, different sRNAs have distinct sequestration properties and therefore different physiological functions, illustrated by the finding that RsmZ, but not RsmY, overexpression coincides with impaired biofilm development in *P. aeruginosa*<sup>23</sup>. Thus, it is not surprising that many bacterial species contain different types of CsrA/RsmE sequestering sRNAs, which are expressed at different phases during cell growth<sup>24</sup> and can act either additively, redundantly or independently on different mRNA targets<sup>23,25,26</sup>, thereby causing different physiological reactions<sup>23,25</sup>.

This evolutionary conserved sequential and ordered assembly has several advantages compared to a random sequestration process. It prevents the formation of long-lived misassembled intermediates or aggregates<sup>13,17</sup>, which could either have poor sequestration capacity or be too stable for eventual degradation. Furthermore, this hierarchical assembly equips the sRNA with a well-defined and optimized binding process for RsmE that has been evolutionarily tuned to maximize its sequestration capacity and modulate the concentration range in which it efficiently sequesters RsmE. This sequential assembly also allows the storage of the protein, the RNP becoming insensitive to degradation



**Figure 4 | RsmE progressively protects the sRNA RsmZ from degradation.** **a**, Specific cleavage of RsmZ sRNA by RNase E. **b**, **c**, Protection of sRNA RsmZ by the RsmE protein against RNase E (**b**) or *E. coli* whole-cell extract (**c**). **d**, Structural basis for sRNA RsmZ protection (shown for conformer L).



**Figure 5 | The life of the ncRNA RsmZ as a protein sponge.** The sRNA RsmZ is transcribed and then sequesters the RsmE protein. When bound to RsmE, the sRNA is stable and can therefore store the protein. The transcription of more RsmZ sRNA or other RsmE-titrating sRNAs (RsmX or RsmY) or mRNAs (*hcnA* mRNA) could result in RsmE protein release from the RsmE-bound RsmZ sRNA, permitting its subsequent degradation.

after a 4:1 ratio of protein to RNA. Finally, coupling between the binding sites for RsmE and RNase E in the sRNA enables the release of the sequestered RsmE proteins from the RNP. The capacity for an RNA to sequester, store and release RNA is reminiscent of the recently discovered circular RNAs that act as microRNA sponges<sup>27,28</sup>. By analogy, RsmZ can be seen in light of our findings as a protein sponge (Fig. 5).

The mode of RNP assembly described here for RsmE is radically different from other cooperative assemblies found to date involving the binding of identical proteins such as hnRNP A1 (ref. 29) or the HIV-1 Rev<sup>30</sup> to RNA. Cooperative binding of RsmE to RsmZ does not involve protein-protein interactions but rather allosteric changes in the RNA allowing subsequent protein binding, a mode of binding that is more reminiscent to bacterial ribosome assembly<sup>14,15</sup>. As a comparison, it will be interesting to see whether the circular RNA sponges sequester microRNA in a random fashion or also bind in a cooperative, well-defined and regulated manner.

## METHODS SUMMARY

RsmE protein expression and purification<sup>7,10</sup>, RNA transcription, purification and preparation of segmental isotopically labelled<sup>11,12</sup> or spin-labelled RNAs<sup>18</sup> were performed as previously described or described in Methods. ITC binding experiments were conducted as described in Methods. EPR, NMR spectroscopy and the structure calculation are detailed in Methods, Extended Data Figs 1–3, 8 and 9, Extended Data Table 1 and Supplementary Methods and are described in more detail elsewhere<sup>10,18</sup>. *In vitro* translation<sup>31</sup> and RNase protection assays are outlined in Methods.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 4 October 2013; accepted 24 March 2014.

Published online 14 May 2014.

- Ponting, C. P., Oliver, P. L. & Reik, W. Evolution and functions of long noncoding RNAs. *Cell* **136**, 629–641 (2009).
- Storz, G., Vogel, J. & Wassarman, K. M. Regulation by small RNAs in bacteria: expanding frontiers. *Mol. Cell* **43**, 880–891 (2011).
- Waters, L. S. & Storz, G. Regulatory RNAs in bacteria. *Cell* **136**, 615–628 (2009).
- Papenfert, K. & Vogel, J. Regulatory RNA in bacterial pathogens. *Cell Host Microbe* **8**, 116–127 (2010).
- Romeo, T., Vakulskas, C. A. & Babitzke, P. Post-transcriptional regulation on a global scale: form and function of Csr/Rsm systems. *Environ. Microbiol.* **15**, 313–324 (2013).
- Lapouge, K., Schubert, M., Allain, F. H. & Haas, D. Gac/Rsm signal transduction pathway of  $\gamma$ -proteobacteria: from RNA recognition to regulation of social behaviour. *Mol. Microbiol.* **67**, 241–253 (2008).
- Schubert, M. *et al.* Molecular basis of messenger RNA recognition by the specific bacterial repressing clamp RsmA/CsrA. *Nature Struct. Mol. Biol.* **14**, 807–813 (2007).

- 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).
- 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).
- Duss, O., Michel, E., Diarra Dit Konte, N., Schubert, M. & Allain, F. H. Molecular basis for the wide range of affinity found in Csr/Rsm protein-RNA recognition. *Nucleic Acids Res.* <http://dx.doi.org/10.1093/nar/gku141> (21 February 2014).
- Duss, O., Lukavsky, P. J. & Allain, F. H. Isotope labeling and segmental labeling of larger RNAs for NMR structural studies. *Adv. Exp. Med. Biol.* **992**, 121–144 (2012).
- Duss, O., Maris, C., von Schroetter, C. & Allain, F. H. 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 (2010).
- Marsh, J. A. *et al.* Protein complexes are under evolutionary selection to assemble via ordered pathways. *Cell* **153**, 461–470 (2013).
- Woodson, S. A. RNA folding pathways and the self-assembly of ribosomes. *Acc. Chem. Res.* **44**, 1312–1319 (2011).
- Shajani, Z., Sykes, M. T. & Williamson, J. R. Assembly of bacterial ribosomes. *Annu. Rev. Biochem.* **80**, 501–526 (2011).
- Wahl, M. C., Will, C. L. & Luhrmann, R. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**, 701–718 (2009).
- Stone, M. D. *et al.* Stepwise protein-mediated RNA folding directs assembly of telomerase ribonucleoprotein. *Nature* **446**, 458–461 (2007).
- Duss, O., Yulikov, M., Jeschke, G. & Allain, F. H. EPR-aided approach for solution structure determination of large RNAs or protein-RNA complexes. *Nature Commun.* <http://dx.doi.org/10.1038/ncomms4669> (2014).
- Holbrook, S. R. Structural principles from large RNAs. *Annu. Rev. Biophys.* **37**, 445–464 (2008).
- Mackie, G. A. RNase E: at the interface of bacterial RNA processing and decay. *Nature Rev. Microbiol.* **11**, 45–57 (2013).
- Carpousis, A. J., Luisi, B. F. & McDowell, K. J. Endonucleolytic initiation of mRNA decay in *Escherichia coli*. *Prog. Mol. Biol. Transl. Sci.* **85**, 91–135 (2009).
- Reimann, 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).
- Petrova, O. E. & Sauer, K. The novel two-component regulatory system BfiSR regulates biofilm development by controlling the small RNA *rsmZ* through CafA. *J. Bacteriol.* **192**, 5275–5288 (2010).
- Kay, E. *et al.* Two GacA-dependent small RNAs modulate the quorum-sensing response in *Pseudomonas aeruginosa*. *J. Bacteriol.* **188**, 6026–6033 (2006).
- Bordi, C. *et al.* Regulatory RNAs and the HptB/RetS signalling pathways fine-tune *Pseudomonas aeruginosa* pathogenesis. *Mol. Microbiol.* **76**, 1427–1443 (2010).
- 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).
- Hansen, T. B. *et al.* Natural RNA circles function as efficient microRNA sponges. *Nature* **495**, 384–388 (2013).
- Memczak, S. *et al.* Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495**, 333–338 (2013).
- Okunola, H. L. & Krainer, A. R. Cooperative-binding and splicing-repressive properties of hnRNP A1. *Mol. Cell. Biol.* **29**, 5620–5631 (2009).
- Daugherty, M. D., Liu, B. & Frankel, A. D. Structural basis for cooperative RNA binding and export complex assembly by HIV Rev. *Nature Struct. Mol. Biol.* **17**, 1337–1342 (2010).
- Michel, E. & Wüthrich, K. High-yield *Escherichia coli*-based cell-free expression of human proteins. *J. Biomol. NMR* **53**, 43–51 (2012).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank G. A. Mackie for providing the RNase E (1–529) clone; N. Diarra Dit Konté and B. Alila for the help with RNA production; M. Blatter for help with structure calculations; G. Wider, T. Stahel, C. Maris and F. Damberger for help with NMR spectroscopy setup; G. Braach, Y. Nikolaev and M. Sattler for discussions. This work was supported by the Swiss National Science Foundation (SNF) grant numbers 3100A0-118118, 31003ab-133134 and 31003A-149921 to F.H.-T.A. and the SNF-NCCR structural biology Iso-lab.

**Author Contributions** O.D. prepared protein and RNA samples for NMR and EPR spectroscopy, collected and analysed the NMR experiments, designed the structure calculation protocol and performed structure calculations, conducted and analysed ITC experiments, gel shift and RNase E cleavage assays, designed the project and the experiments and wrote the manuscript. E.M. designed, performed and analysed *in vitro* translation and *E. coli* whole-cell extract RNase protection assays. M.Y. measured and analysed EPR data. M.S. performed initial NMR experiments. G.J. analysed EPR data and secured funding. 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.

**Author Information** Coordinates, NMR and EPR restraints have been deposited in the Protein Data Bank under accession numbers 2mf0 and 2mf1. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to F.H.-T.A. ([allain@mol.biol.ethz.ch](mailto:allain@mol.biol.ethz.ch)).

## METHODS

**RsmE protein expression and purification.** The RsmE protein was expressed and purified as reported earlier<sup>7</sup>. Deuterated RsmE was obtained by growing the cells in ~90–99% D<sub>2</sub>O.

**RNA transcription and purification.** RNA was produced by *in vitro* transcription from double-stranded DNA templates or linearized plasmids using T7 polymerase and was subsequently purified by denaturing HPLC followed by butanol extraction as previously described<sup>11,12</sup>.

**Segmental labelling of RNA.** Segmental isotope labelling of RNA has been performed by splinted T4 DNA ligation of isotopically labelled and non-labelled RNA fragments obtained by combined ribozyme and RNase H cleavage as previously described<sup>11,12</sup>.

**Spin-labelling of RNA.** Doubly spin-labelled RNAs for DEER measurements were prepared by first, attaching the 3-(2-iodoacetamido)-proxyl spin label to a small 4-thiouridine containing oligonucleotide (purchased from Dharmacon) according to a previously published protocol in ref. 32. Then, two small spin-labelled oligonucleotides were ligated with other larger unlabelled RNA fragments to obtain a doubly spin-labelled RNA using splinted T4 DNA ligation<sup>18</sup>. The larger unlabelled RNA fragments were either purchased from Dharmacon or obtained by site-specific RNase H cleavage as previously described<sup>12</sup>. After ligation, the doubly spin-labelled RNAs were purified with anion exchange HPLC under denaturing conditions<sup>12</sup>. An overview of all spin-labelled RNAs is given in Extended Data Fig. 3.

**Formation of protein–RNA complexes.** The complexes were prepared by mixing the RNA and the protein in the appropriate stoichiometric ratio. Samples of complexes for NMR measurements requiring mM concentrations were first mixed followed by buffer exchange and concentration in a Vivaspin centrifugation device. For NMR titration experiments or native polyacrylamide gels, the protein–RNA complexes were prepared by titrating a concentrated RNA solution (typically 1 mM) into concentrated protein solution both in a buffer containing 30 mM NaCl and 50 mM potassium phosphate, pH 7.2. NMR experiments were done using deuterated RsmE protein. The RsmZ(1–72)/RsmE 1:3 complex samples for DEER measurements were prepared by mixing a 25 μM doubly spin-labelled RsmZ(1–72) RNA dissolved in water with 3.5 equivalents of deuterated RsmE protein dimer and the amount of salts (NaCl and potassium phosphate, pH 7.2) such that the concentration in 30 μl volume results in 30 mM NaCl, 50 mM potassium phosphate, pH 7.2, and 100 μM in doubly spin-labelled RNA in complex. The samples were lyophilized and re-dissolved into 15 μl D<sub>2</sub>O and 15 μl entirely deuterated glycerol. Before DEER measurements, the samples were incubated at 37 °C for 1 h and the correct formation of the RsmZ(1–72)/RsmE 1:3 complexes was validated by native RNA polyacrylamide gels.

**NMR spectroscopy.** NMR spectra were acquired at 313 K on Bruker Avance III 500, 600, 700 or 900 MHz spectrometers equipped with cryo probes. 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 shift assignments and structure determinations of the five RsmE–RNA subcomplexes (RsmE bound to the four stem-loops and to the linker between SL2 and SL3, each in isolation) were performed by standard methods<sup>7,33</sup> and are described in details elsewhere<sup>10</sup>. In brief, non-exchangeable RNA proton resonances were assigned using <sup>1</sup>H–<sup>13</sup>H HSQC, <sup>1</sup>H–<sup>1</sup>H total correlation spectroscopy (TOCSY), 2D <sup>1</sup>H–<sup>1</sup>H nuclear overhauser effect spectroscopy (NOESY) ( $\tau_m = 150$  ms), 3D HC(C)H TOCSY and 3D <sup>13</sup>C-edited NOESY ( $\tau_m = 150$  ms) spectra at 313 K 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 (either A, U or G, C labelled)<sup>33</sup>. The RNA imino protons were assigned with 2D <sup>1</sup>H–<sup>1</sup>H NOESY experiments ( $\tau_m = 250$  ms) in 5% H<sub>2</sub>O/5% D<sub>2</sub>O (*v/v*) at 283 K. The <sup>1</sup>H, <sup>13</sup>C and <sup>15</sup>N chemical shifts of the RsmE protein in complex with the five RNA targets were assigned by standard methods<sup>7,33</sup>. The intra- and intermolecular NOEs were assigned on the basis of 2D <sup>1</sup>H–<sup>1</sup>H NOESY ( $\tau_m = 150$  ms), 3D (F1-edited, F3-filtered) NOESY spectra ( $\tau_m = 150$  ms) and 3D <sup>13</sup>C-edited NOESY ( $\tau_m = 150$  ms)<sup>33</sup> 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. The hydrogen-bonding distance restraints were based on the observation of an imino resonance of the corresponding base pair. Protein torsion angles were obtained by TALOS+<sup>34</sup>.

The <sup>1</sup>H–<sup>13</sup>C TROSY experiment recorded to observe the segmentally labelled RNAs in complex was an adapted version of the single scan, sensitivity-enhanced <sup>1</sup>H–<sup>15</sup>N TROSY from ref. 35.

**EPR spectroscopy.** Nitroxide–nitroxide distance measurements were performed with the four-pulse DEER experiment<sup>36</sup>. The distances between position pairs AB, CH, FH and GH were measured at X-band, as this provided sufficient data quality. Distances for all other position pairs were measured at Q-band.

The X-band pulse EPR measurements were performed with a Bruker EleXsys E580 X-band spectrometer with an ER4118X-MS3 resonator (split ring resonator,

mw frequency 9.2–9.6 GHz, maximum allowed diameter of a cylindrical sample 3 mm). The Q-band (34–36 GHz) measurements were performed at a home-built spectrometer<sup>37</sup> with a rectangular cavity which allows for oversized samples<sup>38</sup>. The DEER measurements were performed at 50 K. This temperature corresponds approximately to the optimum measurement conditions with respect to the longitudinal and transverse relaxation of nitroxide radicals. The sample temperature was stabilized with a He-flow cryostat (ER 4118 CF, Oxford Instruments).

In the DEER experiments at Q-band, all pulses were set to a duration of 12 ns, whereas at X-band the pump pulse of 12 ns and all detection pulses of 32 ns were used<sup>39</sup>. The first interpulse delay time in the DEER sequence was set to 400 ns in all cases. The second delay time (between the primary echo and refocusing pulse) was set according to the required length of the DEER trace, which was selected to provide sufficient range for background correction. Typical measurement time was 20–40 h for X-band measurements and 10–15 h for Q-band measurements, depending on the required length of the DEER trace. The offset between pump and detection frequencies was +65 MHz at X-band and –85 MHz at Q-band. The samples for DEER measurements were prepared as 100 μM solutions in 1/1 D<sub>2</sub>O/D-glycerol (*v/v*) mixture. For each sample about 30 μl of such mixture was placed into a quartz tube of 3 mm outer diameter and frozen by immersion into liquid nitrogen. After the preparation, the samples were stored in liquid nitrogen.

**DEER data processing.** Primary DEER time traces and background fit functions for all measured distance pairs are presented in Extended Data Fig. 9a. The corresponding form factor traces and their fits are shown in Extended Data Fig. 9b. Fitting of DEER data was performed with DeerAnalysis 2009 software<sup>40</sup>. To avoid possible artefacts, the range for the background fit was cut at least 100 ns before the end of the DEER trace. All traces were fitted with the unrestricted distance distribution model. In each case Tikhonov regularization<sup>41</sup> was performed and distributions corresponding to different regularization parameters were analysed. Although the smoothness of distance distributions varied with variation of the Tikhonov regularization parameter, the selection of the distance constraints (following section) was robust with respect to this variation.

A 3D background model was assumed in the background fits for all samples. In most cases, the obtained distance distribution was stable with respect to the change of the background model. In a few samples, indications of some degree of aggregation were present (CF, EG, BE: distance distributions with a ‘tail’ towards long distances). In most cases, instead of a long distance tail, the regularization procedure resulted in one or two additional peaks, typically in the distance range 8–10 nm (AB, AE, BD, BH, CD, CE, CH, DE, DH, EF, EG, FG, FH, GH). Such peaks are marked with asterisks in Extended Data Fig. 3b. As the distance range 8–10 nm is least trustworthy for the given lengths of DEER time<sup>42</sup>, and as the presence and positions of those peaks depended on the selection of the background model, these peaks were disregarded in the following analysis.

In contrast to those cases, the double peak pattern obtained for the AC and AG samples was stable with respect to the change of the background model and could not be attributed to any analysis imperfections. The presence of two very different dipolar modulation frequencies can also be seen directly in the primary time-domain data for these two samples.

**Selection of the distance constraints for the structure determination.** Separately positioned small peaks at the limit of the accessible distance range and the rise of the probability density  $P(r)$  at the very end of the distance interval were ignored for the determination of the distance constraints (Extended Data Fig. 3b). For all single-peak distance distributions, the maximum amplitude point was determined and the distance range covering 70% of the distribution area to the left and 70% of the area to the right of the distribution maximum was selected as the corresponding distance constraint. For the cases of clearly bimodal distributions (samples AC and AG) the two individual components were separated at the middle of the dip between them (5.1 nm for AC and 5.8 nm for AG). Accordingly, the part of the distance distribution for distances below this separating value was considered as the first peak and the part for distances above this separating value was considered as the second peak. The distance constraints for the two separate peaks were then determined in the same way (70% area condition) as for the single-peak distance distributions.

**Structure determination.** For each conformation L and R, we generated 2,500 preliminary structures of the full complexes by a simulated annealing protocol (consisting of 200,000 simulated annealing steps) using the CYANA package<sup>43</sup>. Starting from random coil conformations, we included simultaneously all the local constraints consisting of NOE distance, torsion angle and hydrogen-bond constraints from the isolated binding site complexes (Extended Data Fig. 2 and Supplementary Methods) and the global constraints consisting of 42 (2 × 21) DEER upper and lower distance restraints (Extended Data Fig. 3). The global EPR constraints were included into structure calculation using an efficient protocol that allows the transformation of DEER distance distributions into simple constraints for structure calculation<sup>18</sup>. First, the spin label distributions (‘radical clouds’) were calculated for each isolated binding site complex (RsmE homodimer bound to two singly spin-labelled RNAs)



using the torsion angle dynamics simulation program CYANA by restraining the local structure of the complex including the spin-labelled nucleotide but imposing no restraints on the flexible spin-label<sup>18</sup>. These 'radical clouds' (obtained by a superimposition of the lowest energy structures of this calculation) were then superimposed onto the corresponding sites in the full complex. Because the experimental and CYANA-predicted mean distances of distance distributions are in very good agreement<sup>18</sup>, we placed a dummy atom into the geometrical centres of the modelled spin-label distributions at each of the eight spin-label attachment sites on the full complex and then restrained each pair of dummy atoms by an upper and lower limit constraint, which was derived from a DEER distance distribution to confine 70% of the peak area from each side of the peak maximum (Extended Data Fig. 3b). These broad 70% distance ranges cover most of the distance distributions with reasonable probability but also account for uncertainties in the simulations of the spin-label distributions and in DEER data analysis (such as broadened wings in distance distributions due to Tikhonov regularization or possible low amplitude artefacts at the ends of the distance distributions)<sup>40,42</sup>.

To overcome the problem of determining the structures of two conformations simultaneously present in solution, we performed a separate structure calculation for each conformer. Only two of the 21 measured distance distributions (AC and AG) consist of two well separated peaks one of each belonging to one conformation, whereas the other distance distributions are not resolving the distance distributions of both conformations but rather are represented by broad distance distributions (see Extended Data Fig. 3b). Owing to geometrical reasons both low-distance peaks (conformer L) and both high-distance peaks (conformer R) in the AC and AG distance distributions must belong to the same conformation (see Supplementary Methods). Therefore, we performed for each conformation (L or R) a separate structure calculation in which only the two resolved distance distributions are providing distinct distance constraints whereas all broad unresolved distance distributions are used as same constraints for both conformations. For conformer R, we obtained a unique solution, whereas conformer L led initially to two possible solutions.

To select the correct solution for conformer L, we back-calculated all the 21 distance distributions for both possible solutions of conformer L together with the unique solution for conformer R (Extended Data Figs 8 and 9 and Supplementary Methods) assuming that the two conformations L and R are populated in a 1:2 ratio. This ratio was estimated from the relative integrals of the two separated peaks in the AC and AG distance distributions.

Then, for the correct solution of both conformations L and R, we selected from the 2,500 starting structures the 50 lowest energy structures and separately refined them in AMBER 9.0 (refs 44, 45). The complexes were refined in implicit solvent using the same distance (NOE and EPR), torsion angle and hydrogen-bond constraints as used in the CYANA-simulated annealing protocol<sup>33,44</sup>. The force field ff99 (ref. 46) was used along with the generalized Born model<sup>47</sup> to mimic the solvent. After AMBER refinement, we selected from the 30 lowest amber energy structures, the 20 lowest constraint energy structures. Finally, the dummy atoms were removed and the structures were energy minimized. The structural statistics for both structural ensembles (representing the two structures present in solution) are summarized in Extended Data Table 1. The Ramachandran statistics for conformer L (residues 1–59 for each protein chain): most favoured regions, 88%; additionally allowed regions, 11.4%; generously allowed regions, 0.5%; and disallowed regions, 0.2%; and for conformer R: most favoured regions, 87.8%; additionally allowed regions, 11.5%; generously allowed regions, 0.5%; and disallowed regions, 0.2%.

**ITC binding experiments.** The 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, pH 8.0). Concentrations were determined after dialysis using optical density absorbance at 260 and 280 nm for RNA or protein, respectively. Although the RNA or protein concentration in the syringe varied between 50–400  $\mu$ M, the concentration in the cell was 2–40  $\mu$ M depending on the expected binding affinity. ITC binding experiments were performed at 25 °C and typically consisted of 30–40 injections of 4–10  $\mu$ l with an injection speed of 2 s  $\mu$ l<sup>-1</sup> and a 5-min interval between additions. The stirring rate was 307 r.p.m. All measurements were repeated at least twice. Using Origin 7.0, the raw data was integrated, corrected for nonspecific heats and analysed according to a 1-site binding model. For SL2, the affinity could only be fitted with a 2-site binding model (see Fig. 3a). For the 5' UTR of the *hcnA* mRNA (see Extended Data Fig. 6d), ITC was performed with a construct missing the GGA motifs no. 1 and 2 (*hcnA\_GGA* nos. 3–5), because their affinity was determined to be in the mM range in isolation<sup>10</sup>. The ITC data for the constructs presented in Fig. 3b are shown in Extended Data Fig. 6c. The RNA sequences of all the constructs are summarized in Supplementary Methods.  $K_d$  determination of the isolated binding site constructs is described elsewhere<sup>10</sup>.

**Cell-free translation assay.** The cell-free expression vector pCFX100 was prepared by subcloning the 5' UTR of the *hcnA* mRNA into the pVEX1.3-CAT plasmid (Roche) and was amplified using a plasmid maxi prep kit (Macherey-Nagel). The presence of the 5' UTR of the *hcnA* mRNA allows RsmE/RsmZ-mediated modulation of translation of the encoded chloramphenicol acetyl transferase (CAT) reporter enzyme transcript. As a source for cell extract preparation, *E. coli* BL21 (DE3) Star cells were genetically modified by introducing a C-terminal (*His*)<sub>6</sub>-tag into the *csrA* gene following the procedure of Datsenko and Wanner<sup>48</sup>, resulting in the strain *E. coli* BL21 (DE3) Star *csrA::(His)*<sub>6</sub>. S30 cell extract was prepared on the basis of a previously described protocol<sup>31</sup> with an additional step that included removal of the endogenous *csrA* protein 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 following the previously described protocol<sup>31</sup>.

Analytical scale cell-free synthesis reactions (50  $\mu$ l) of the pCFX100 plasmid were set up in presence of various amounts of wild-type or mutant RsmZ sRNA and 200 nM RsmE protein from *P. fluorescens*. After 3.5 h of cell-free expression, the reaction mixture was centrifuged and placed on ice. 5  $\mu$ l of the reaction supernatant was thoroughly mixed with 495  $\mu$ l dilution buffer (100 mM Tris-HCl, pH 7.8, 1 mg ml<sup>-1</sup> BSA). 10  $\mu$ l of the diluted solution was then mixed with 990  $\mu$ l CAT reaction solution (100 mM Tris-HCl, pH 7.8, 0.5 mM DTNB, 50  $\mu$ M acetyl-CoA, 50  $\mu$ M chloramphenicol, 1 mg ml<sup>-1</sup> 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 which was obtained after complete saturation of RsmE in the reaction mixture with RsmZ RNA.

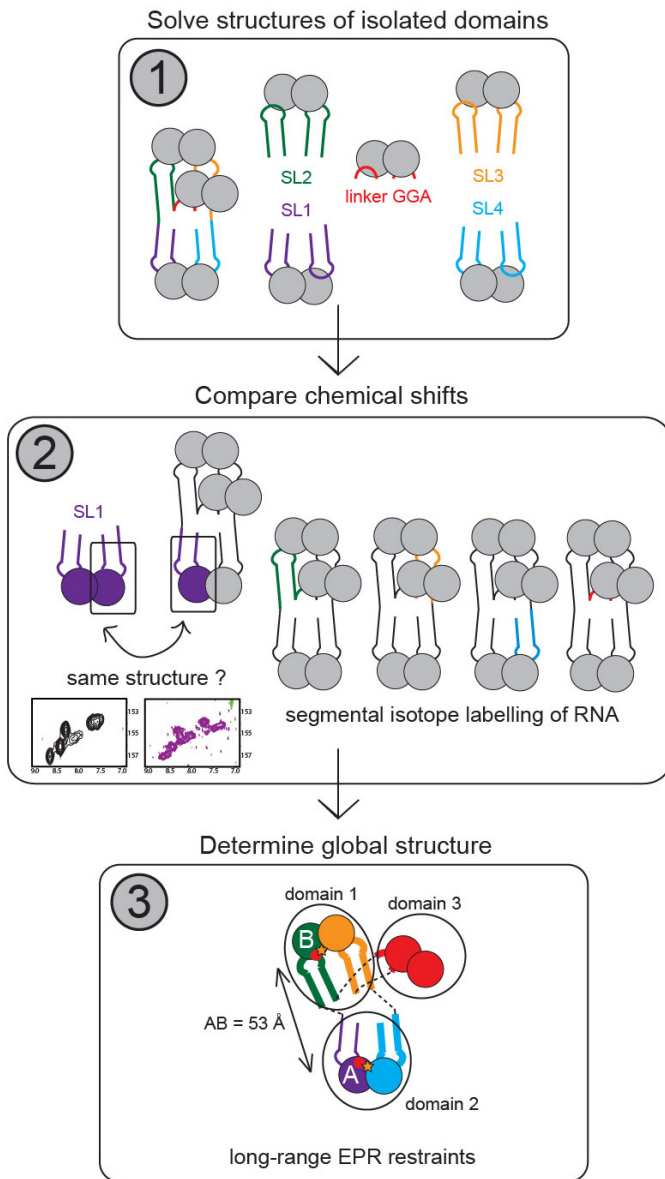
**RNase E cleavage assay.** *E. coli* RNase E (1–529) was purified as described previously<sup>49,50</sup>. 7.5–10  $\mu$ M *in vitro*-transcribed (triphosphorylated) RsmZ sRNA was incubated with increasing amounts of *E. coli* RNase E (1–529) (40–1,000 nM) in 25 mM Tris-HCl, pH 7.5, 10 mM MgCl<sub>2</sub>, 50 mM NaCl, 0.1 mM DTT and 5% glycerol for 6 h at 37 °C. The identity of the generated cleavage fragments was confirmed by mass spectrometry. RsmE protection was assayed using 330 nM RNase E (1–529) at the same conditions as above.

**Whole-cell extract RNase digestion assay.** S30 extract from *E. coli* BL21 (DE3) cells grown in LB medium was prepared as previously described<sup>31</sup> with the exception that pre-incubation of the cell extract was omitted. 1 nM of *in vitro*-transcribed wild-type or mutant RsmZ RNA was incubated with 6 nM RsmE protein dimer at 37 °C in a total volume of 50  $\mu$ l containing 30% (v/v) S30 extract, 37.4% (v/v) ERCN solution (energy regeneration mixture<sup>31</sup>), 6.4 mM magnesium acetate. 5  $\mu$ l aliquots were taken after various incubation times and the RNA was stepwise extracted using phenol and chloroform/isoamyl alcohol. The extracted RNA was then lyophilized and solubilized in 10  $\mu$ l loading buffer (12 M urea in TB buffer, toluidine blue, xylene cyanol, 25 mM EDTA) and stored at –20 °C until needed for urea-PAGE analysis.

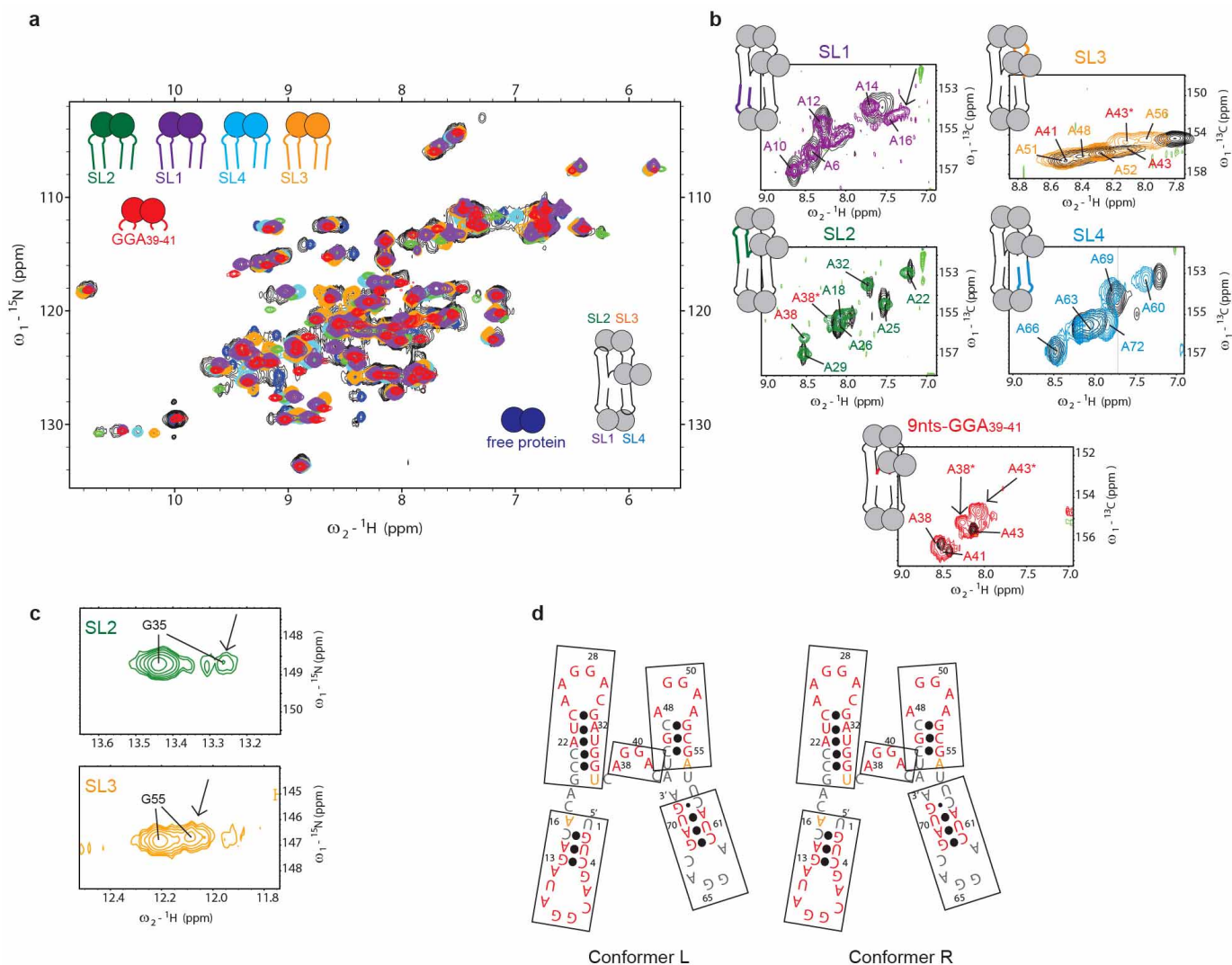
- Ramos, A. & Varani, G. A new method to detect long-range protein–RNA contacts: NMR detection of electron–proton relaxation induced by nitroxide spin-labeled RNA. *J. Am. Chem. Soc.* **120**, 10992–10993 (1998).
- Dominguez, C., Schubert, M., Duss, O., Ravindranathan, S. & Allain, F. H. Structure determination and dynamics of protein–RNA complexes by NMR spectroscopy. *Prog. Nucl. Magn. Reson. Spectrosc.* **58**, 1–61 (2011).
- Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **44**, 213–223 (2009).
- Weigelt, J. Single scan, sensitivity- and gradient-enhanced TROSY for multidimensional NMR experiments. *J. Am. Chem. Soc.* **120**, 10778–10779 (1998).
- Pannier, M., Veit, S., Godt, A., Jeschke, G. & Spiess, H. W. Dead-time free measurement of dipole–dipole interactions between electron spins. *J. Magn. Reson.* **142**, 331–340 (2000).
- Gromov, I. *et al.* A Q-band pulse EPR/ENDOR spectrometer and the implementation of advanced one- and two-dimensional pulse EPR methodology. *J. Magn. Reson.* **149**, 196–203 (2001).
- Tschaggelar, R. *et al.* Cryogenic 35GHz pulse ENDOR probehead accommodating large sample sizes: performance and applications. *J. Magn. Reson.* **200**, 81–87 (2009).
- Polyhach, Y. *et al.* High sensitivity and versatility of the DEER experiment on nitroxide radical pairs at Q-band frequencies. *Phys. Chem. Chem. Phys.* **14**, 10762–10773 (2012).
- Jeschke, G. *et al.* DeerAnalysis2006—a comprehensive software package for analyzing pulsed ELDOR data. *Appl. Magn. Reson.* **30**, 473–498 (2006).
- Jeschke, G., Panek, G., Godt, A., Bender, A. & Paulsen, H. Data analysis procedures for pulse ELDOR measurements of broad distance distributions. *Appl. Magn. Reson.* **26**, 223–244 (2004).
- Jeschke, G. in *Structure and Bonding* vol. 152 (eds Timmel, C. & Harmer, J.) (Springer, 2014, in the press).
- Güntert, P. Automated NMR structure calculation with CYANA. *Methods Mol. Biol.* **278**, 353–378 (2004).
- Case, D. A. *et al.* The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668–1688 (2005).
- v. AMBER 9. <http://ambermd.org/> (Univ. California, San Francisco, 2006).

46. Wang, J. M., Cieplak, P. & Kollman, P. A. 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 (2000).
47. Bashford, D. & Case, D. A. Generalized born models of macromolecular solvation effects. *Annu. Rev. Phys. Chem.* **51**, 129–152 (2000).
48. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA* **97**, 6640–6645 (2000).
49. Garrey, S. M. *et al.* Substrate binding and active site residues in RNases E and G: role of the 5'-sensor. *J. Biol. Chem.* **284**, 31843–31850 (2009).
50. Briant, D. J., Hankins, J. S., Cook, M. A. & Mackie, G. A. The quaternary structure of RNase G from *Escherichia coli*. *Mol. Microbiol.* **50**, 1381–1390 (2003).
51. Kaberdin, V. R. Probing the substrate specificity of *Escherichia coli* RNase E using a novel oligonucleotide-based assay. *Nucleic Acids Res.* **31**, 4710–4716 (2003).



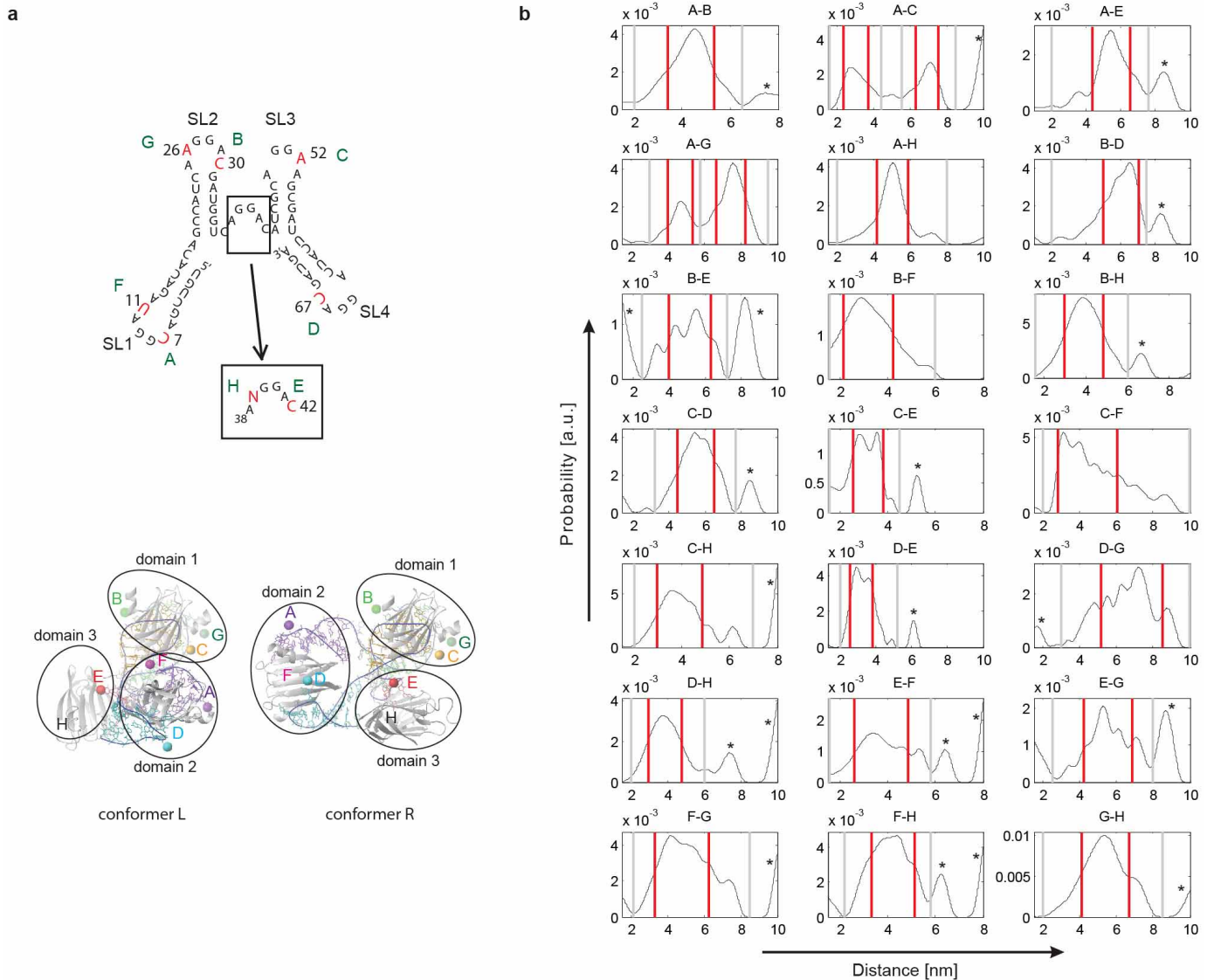


**Extended Data Figure 1 | General protocol for the structure determination of the full complex structure.** (1) The structures of the separate binding site complexes are solved. (2) Comparing the chemical shifts of the resonances in the isolated binding site complexes and the full complex reveals which parts of the isolated binding site complexes retain their structure in the full complex. Very similar A(N)GGAX binding motifs in all the five binding sites led to severe spectral overlap and thus required segmental labelling of a single binding site on the RsmZ sRNA at a time. (3) The global structure of the full complex is obtained by constraining the different domains with several EPR long-range distances measured between spin-labels attached at several positions on the RsmZ sRNA.



**Extended Data Figure 2 | Chemical shift comparison of RNA in isolated domain complexes or within the full 70-kDa complex.** **a**, Overlay of the protein amide  $^1\text{H}-^{15}\text{N}$  HSQC spectra of the isolated stem-loop complexes (SL1, violet; SL2, green; SL3, orange; SL4, cyan; and GGA<sub>39-41</sub>, red) and the 70-kDa RsmZ(1-72)/RsmE 1:3 complex (black). The free protein is shown in blue. **b**, The  $^1\text{H}-^{13}\text{C}$  transverse relaxation optimized spectroscopy (TROSY) spectra of the RNA aromatic adenine H2-C2 resonances of the segmentally labelled RNAs in complex (in colour) were shifted on both axes by 0.5 times the one-bond amide scalar coupling  $^1J_{\text{NH}}$  to be superimposed on the  $^1\text{H}-^{13}\text{C}$  HSQC spectra of the isolated binding sites in complex (black). The colours of the spectra correspond to the isotopically labelled RNA regions within the 70-kDa complex in the corresponding samples: SL1, 1-16; SL2, 17-40; SL3, 41-58; SL4, 59-72; 9nt-GGA<sub>39-41</sub>, 36-44. Note that nucleotides (nt) 36-40 are isotopically labelled in both SL2 and 9nt-GGA<sub>39-41</sub>, whereas nucleotides 41-44 are isotopically labelled in both SL3 and 9nt-GGA<sub>39-41</sub> constructs (red labels in

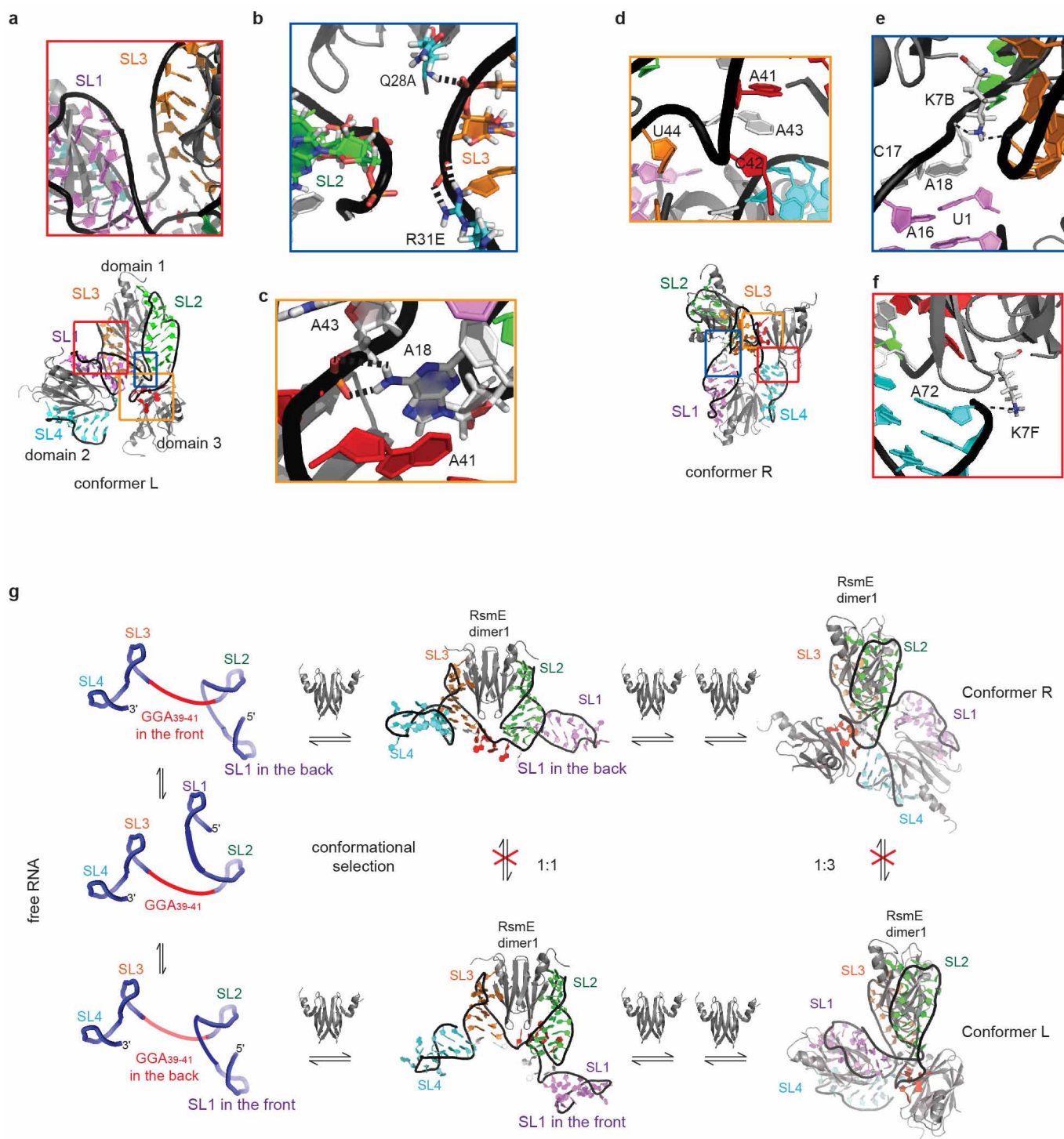
the SL2 and SL3 spectra). Double peaks due to the two conformations in the 70-kDa complex present in solution are indicated by arrows. See Supplementary Discussion for further explanation. **c**, Shown are the G35 and G55 imino regions in the  $^1\text{H}-^{15}\text{N}$  HSQC spectra of the full complex indicating two conformations. **d**, Depiction of chemical shift agreement between RNA in isolated domain and within the full 70-kDa complex. The RNA nucleotides having at least one assigned resonance with a combined chemical shift ( $\Delta\text{CCS}$ ) difference  $<0.12$  p.p.m. or  $<0.2$  p.p.m. between the isolated domain and full complex are coloured in red or orange, respectively. Black balls between two base-paired nucleotides indicate a protected imino observed in the  $^1\text{H}-^{15}\text{N}$  HSQC spectra of the full complex. The smaller black ball representing the G71 imino indicates that this imino was only weakly present. The regions for which the structures of the isolated domains were used as building blocks for the full complex are boxed. See Supplementary Methods for further explanation.



**Extended Data Figure 3 | Overview of measured EPR distances.** **a**, Location of all eight spin-label positions shown on the secondary structure of RsmZ (top) or on the three-dimensional structures of both conformations L and R (bottom). A ball represents the centre of the spin label distribution at the corresponding site. A common A(N)GGAX motif (where N and X stand for any nucleotide) is present at every binding site. The corresponding nucleotides N or X were replaced by a thioridine residue to which the spin-label was attached. Because the residues N and X are only unspecifically recognized by the protein<sup>7,10</sup>, the complex formation is not disturbed, evidenced by native PAGE gels confirming the formation of the 1:3 full complex and by double

electron-electron resonance (DEER) measurements of each isolated binding site in complex with one RsmE protein dimer<sup>18</sup>. **b**, Distance distributions obtained from DEER data (black). Grey and red vertical lines show the distance ranges considered in the EPR constraint determination (grey, 100% area; red, 70% area; see Methods). Red vertical lines indicate the regions that were taken as the EPR distance constraints. Asterisks indicate noise artefacts (B-E) and the features that appeared to be unstable upon variation of the background model. Note that for B-E, C-F and E-G, distance distributions are possibly affected by partial aggregation.

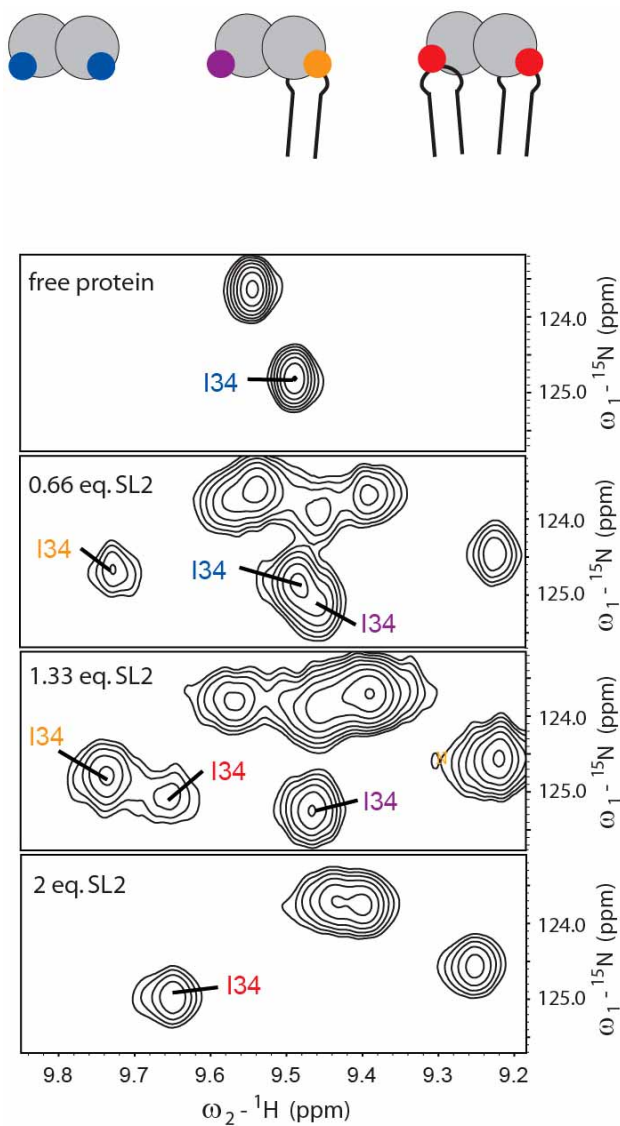




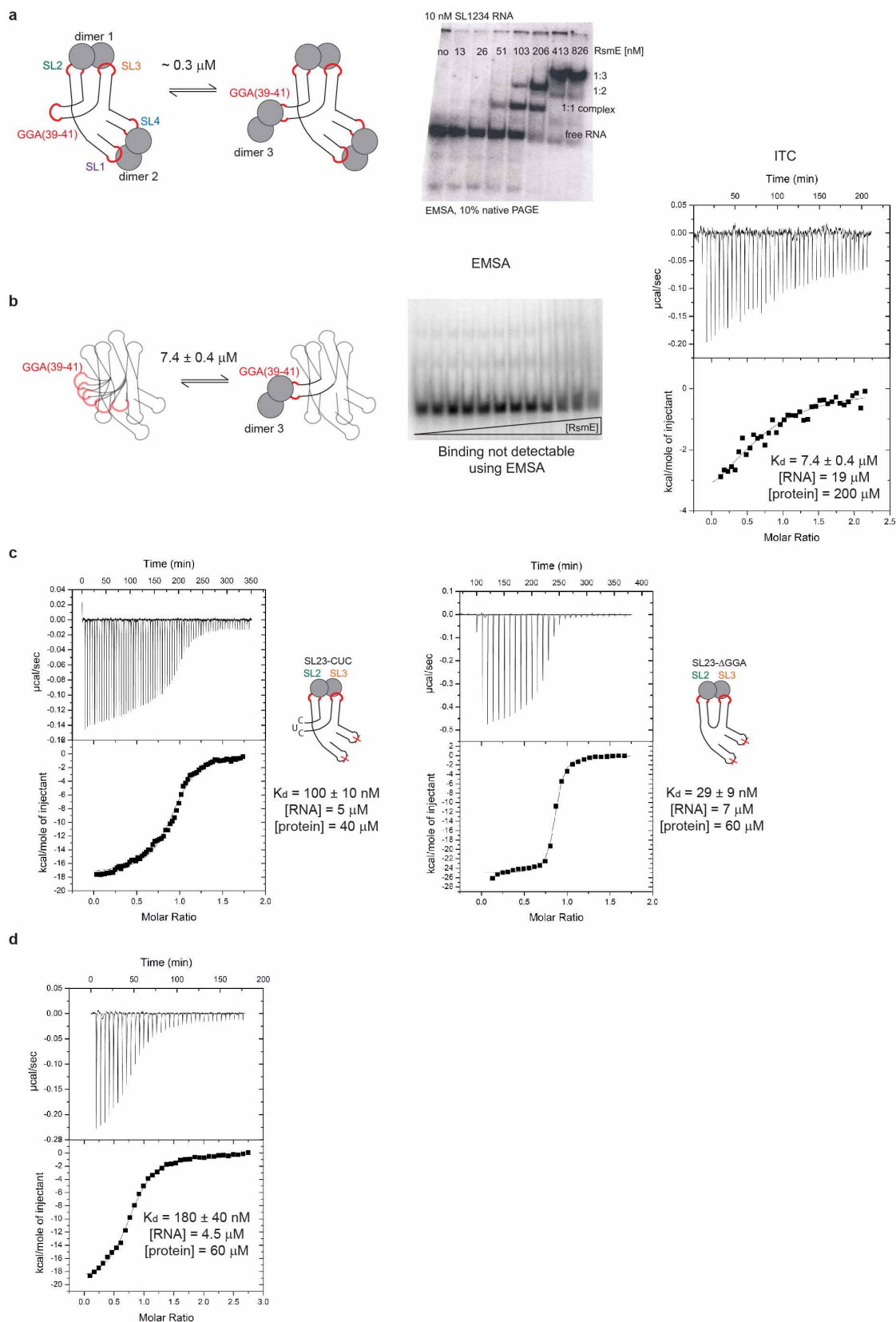
**Extended Data Figure 4 | Potential tertiary interactions stabilize the global fold in both conformations, which cannot directly interconvert.**

**a–f**, Conformer L (**a–c**) and conformer R (**d–f**). Protein chains A/B, C/D and E/F belong to RsmE protein dimers 1, 2 or 3, respectively. **a**, The RNA backbones of SL1 and SL3 approach each other up to 5.5 Å and thus, the high local charge density is probably stabilized by ions. **b**, The RNA backbones of SL2 and SL3 approach each other. Besides potential ions, the positively charged side chain of R31E (dimer 3) and the protein backbone amide of Q28A (dimer 1) might help to neutralize the high charge density. **c**, Potential hydrogen bonds and one stacking interaction between the conserved A41 and A18 could rationalize the packing of the third domain (GGA<sub>39–41</sub>, red residues) towards the RNA linker between SL1 and SL2 (A18). **d**, A base-stacking interaction between A41 and A43 might stabilize the positioning of the third

domain with respect to SL3. **e**, In several low-energy structures, the conserved C17 and A18 are stacking on the U1–A16 SL1 stem-closing base pair. While stacking, C17 and A18 form a non-canonical cSWC base pair. These base-pair and stacking interactions orient the stem of SL1 in respect to the RNA linker between SL1 and SL2 approaches the RNA backbone of SL3. This high charge density could be stabilized by the K7B protein side chain (dimer 1). **f**, The RNA backbone at the 3' end of SL4 could contact the K7F side-chain residue (dimer 3), thus also contributing to stabilization of the position of the third dimer. Overall, these potential interactions rationalize the presence of two well-defined structures. **g**, Two different possible RNA backbone conformations explain the formation of two conformations that cannot interconvert unless all proteins have dissociated.



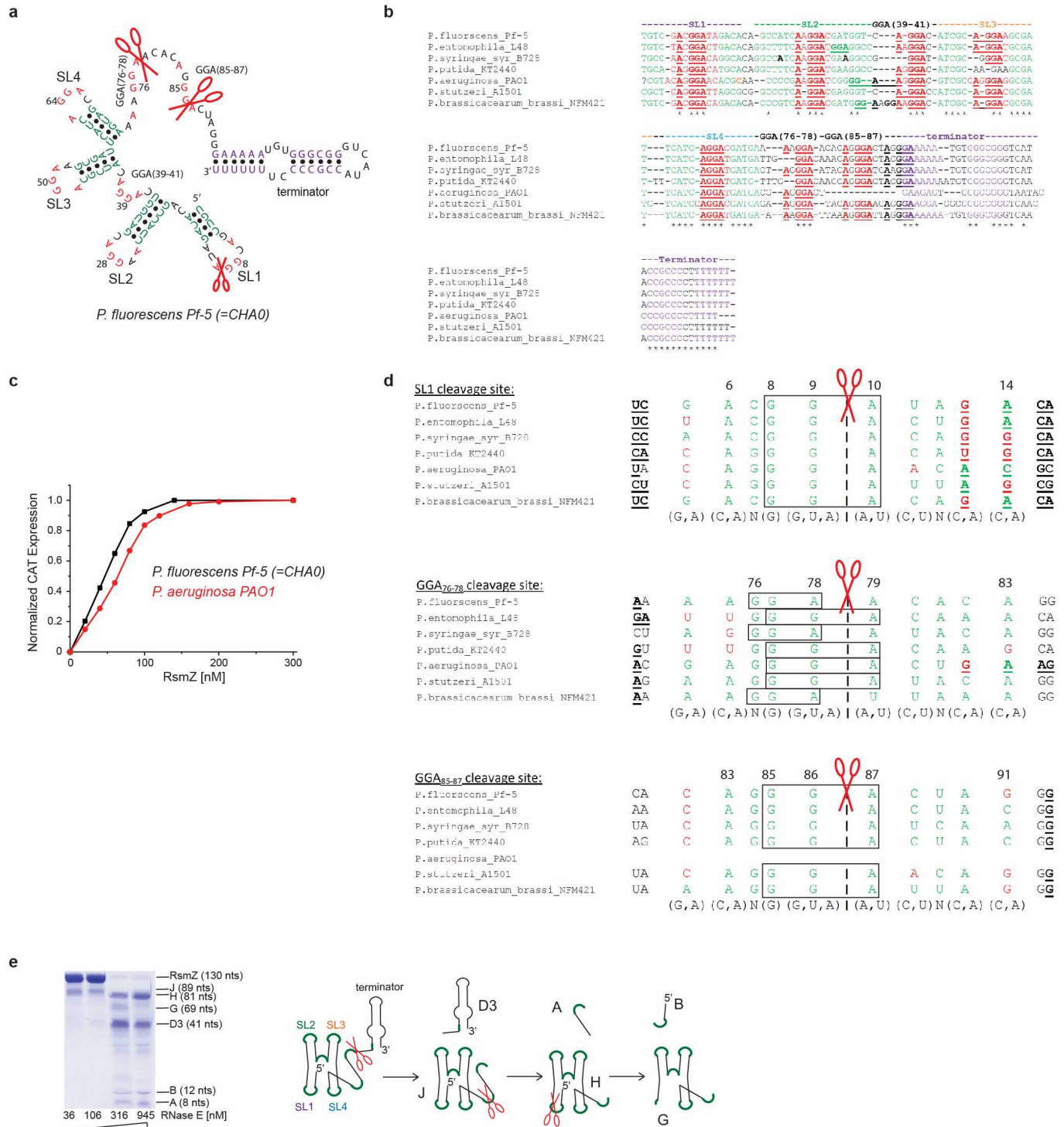
**Extended Data Figure 5 | NMR evidence for a structural and/or dynamic change upon binding of one SL2 RNA molecule to the RsmE protein homodimer.** Top, a schematic representation of the free protein, the singly RNA bound protein and the doubly bound protein form is depicted. The dots represent a certain protein resonance in the free protein (blue), on the free binding site when the other protein binding site is bound (magenta), in the bound form when the other protein binding site is unbound (orange) or in the bound form when the other protein binding site is also bound (red). Bottom, a zoom into a representative region of the protein  $^1\text{H}$ - $^{15}\text{N}$  HSQCs at all the titration points is shown (from top to bottom: free protein, 0.66, 1.33 and 2 equivalents of SL2 RNA). The resonance peaks of isoleucine I34 are labelled in colour according to the four possible unbound or bound states (same colour code as schematic representation (top)).



**Extended Data Figure 6 | ITC binding data and gel shift assays.** **a, b**, Binding of the first two RsmE dimers to RsmZ increases the binding affinity of the third dimer. **a**, Gel shift assay of an RNA construct only containing the first four stem-loops of RsmZ, RsmZ(1–72), shows that the binding of the third dimer has around 300 nM  $K_d$  when the first two dimers are already bound. **b**, When disrupting the four binding sites of the first two dimers (using SL1234\_ΔGGA\_GGA<sub>39–41</sub>, see Supplementary Methods), the binding affinity

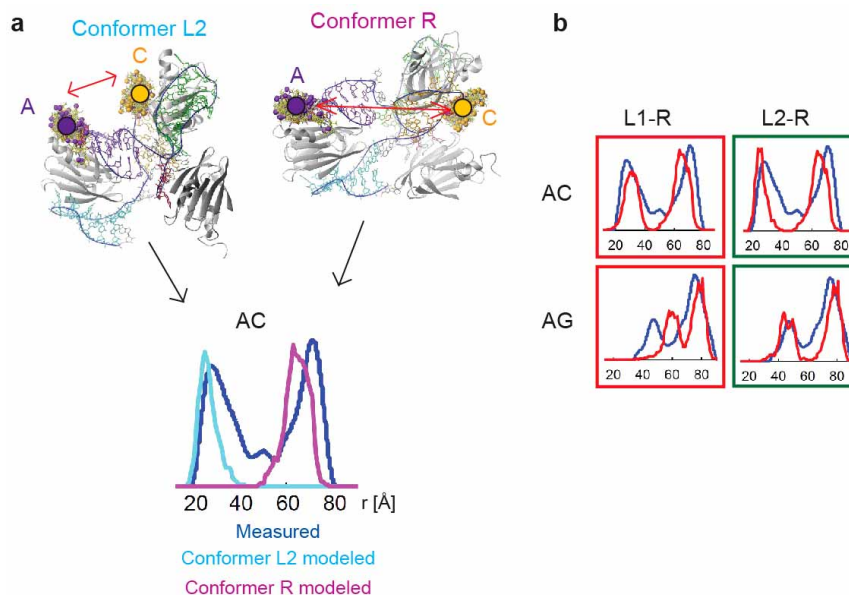
of the third dimer is reduced 25-fold. The affinity was determined by ITC, because no binding was detectable by gel shift assays. **c**, ITC titration curves showing that the SL23 linker is longer than required to bind the first dimer with the highest possible affinity (see Fig. 3b). **d**, ITC titration curve of RsmE binding to a truncated form of the 5' UTR of the *hcnA* mRNA missing the low-affinity GGA motifs nos. 1 and 2.





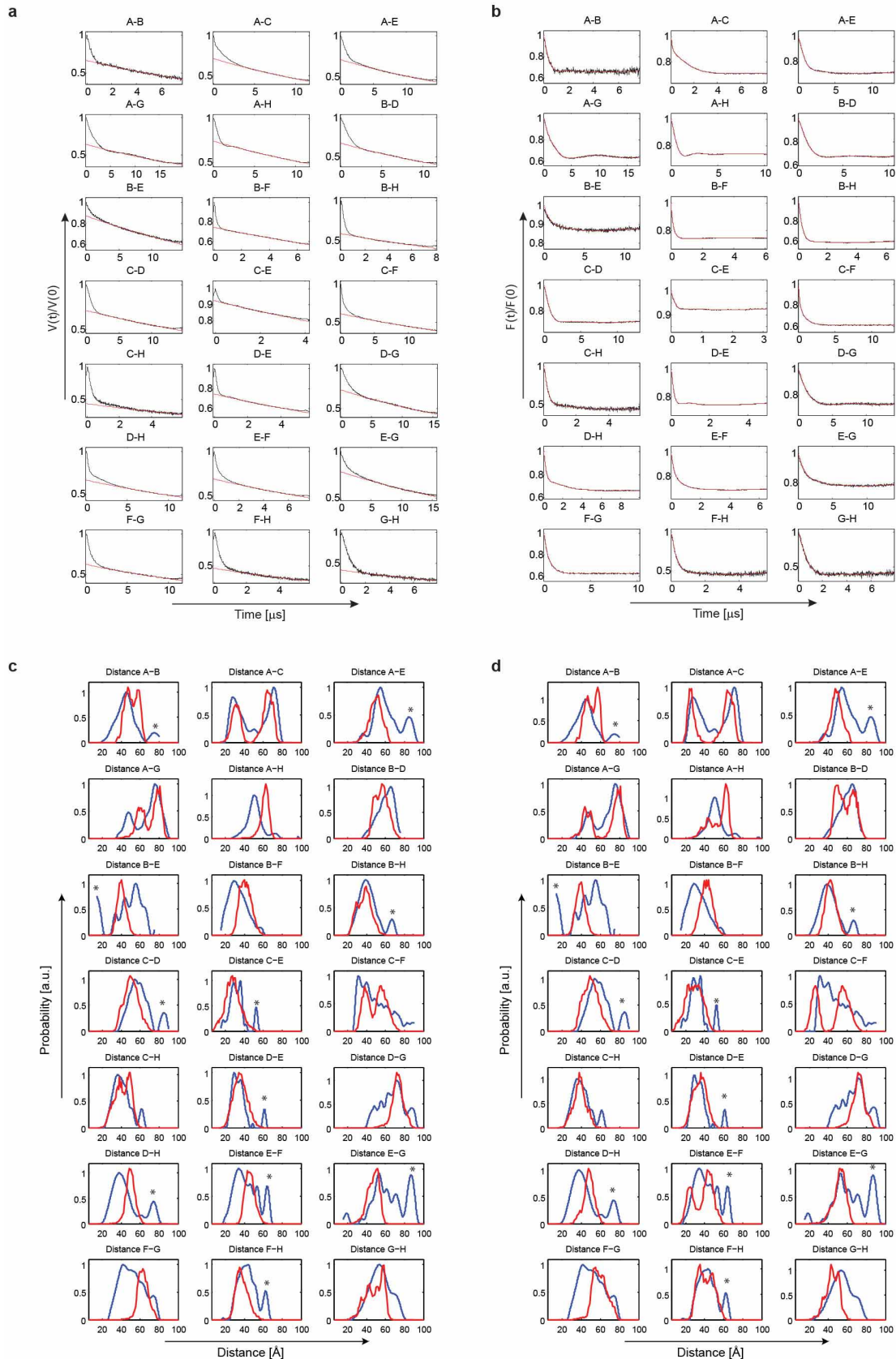
**Extended Data Figure 7 | Conservation of RsmZ sRNAs in *Pseudomonas*.**  
**a**, Secondary structure of the *P. fluorescens* Pf-5 RsmZ sRNA (which is identical to the CHA0 strain). **b**, Multiple sequence alignment of RsmZ sRNAs in different *Pseudomonas* species. The alignment was adapted manually according to the structural knowledge of the RsmZ sRNA. The GGA-binding motifs including all loop nucleotides are shown in red, nucleotides located in stem regions are coloured in green, and base-paired terminator nucleotides in purple. Nucleotides conserved within all the species are marked by stars. **c**, *In vitro* translation assay demonstrating a very similar translation activation potential for the RsmZ sRNAs in *P. fluorescens* and *P. aeruginosa*. The slightly smaller translation activation of RsmZ in *P. aeruginosa* is likely due to the missing GGA<sub>85-87</sub> motif. **d**, Conservation of RNase E cleavage sites in RsmZ sRNAs in *Pseudomonas*. The three RNase E cleavage sites are conserved in *Pseudomonas*, except that in *P. aeruginosa* the GGA<sub>85-87</sub> cleavage site is missing

because the terminator directly follows the GGA<sub>76-78</sub> motif. All RNase E cleavage sites overlap with GGA-binding motifs that bind dimer 2, dimer 3 and dimer 4 (boxed nucleotides). The nucleotides located in a stem are shown in bold and are underlined. The RNase E consensus cleavage site proposed by Kaberdin is shown below the multiple sequence alignments<sup>51</sup>. Nucleotides matching the RNase E cleavage consensus sequence are shown in green, the non-matching nucleotides in red. The two main cleavage sites in the single-stranded region between SL4 and the terminator are single-stranded and match the RNase E cleavage consensus sequence. The cleavage site in the loop of SL1 contains some nucleotides 3' to the cleavage site, which are not single-stranded and do not match the cleavage consensus sequence. This is in agreement with the RNase E cleavage site in SL1 being less active compared to the two main cleavage sites between SL4 and the terminator. **e**, Sequential RNase E degradation of the RsmZ sRNA *in vitro*.



**Extended Data Figure 8 | Selecting the correct cluster combinations by comparing measured and back-calculated distance distributions.** **a**, How to obtain the modelled distance distribution for a single spin label pair (AC as an example) for one specific structure? The modelled distance distribution is calculated by plotting the occurrence of distances between all the radical positions of spin label A to all the radical positions of spin label C. On the top, the ‘radical clouds’ (spin-label distributions) for spin labels A and C are superimposed onto one structure of cluster L2 (left) or cluster R (right). Bottom, the measured (blue) distance distribution is superimposed on the modelled one

of the L2 structure (cyan) and the R structure (magenta). **b**, Overview of measured (blue) and modelled (red) distance distributions for both possible cluster combinations for some representative spin label pairs (note that the AG and AC distance pairs are the best indicators to discriminate between the different cluster combinations (see Supplementary Methods)). The modelled distance distribution for the two clusters present are summed up and shown as a single distribution (red curve). All the distances are shown in Å. The best cluster combination (L2-R) is boxed in green. An overview of all 21 spin-label pairs for both possible solutions is shown in Extended Data Fig. 9c, d.



**Extended Data Figure 9 | Supplementary DEER data.** **a**, DEER primary time domain data (black) with 3D background fits (red). **b**, DEER form factor traces (black) and their best model free fits (red) with use of Tikhonov regularization (obtained with DEER analysis). **c**, **d**, Measured (blue) versus modelled spin label distributions (red) for cluster L1-R (**c**) or L2-R

(**d**) combinations (see Supplementary Methods and Extended Data Fig. 8 for explanations). Asterisks indicate noise artefacts (B-E) and the features that appeared to be unstable upon variation of the background model. Note that for BE, CF and EG, distance distributions are possibly affected by partial aggregation.



Extended Data Table 1 | NMR and refinement statistics for complexes

	Conformer L		Conformer R	
	Protein	RNA	Protein	RNA
<b>NMR distance and dihedral constraints</b>				
Distance restraints				
Total NOE (intramolecular)	4664	837	4664	873
Intra-residue	451	368	451	385
Inter-residue	4213	469	4213	488
Sequential ( $ i-j  = 1$ )	2091	353	2091	369
Non-sequential ( $ i-j  > 1$ )	2122	116	2122	119
Hydrogen bonds	96	46	96	48
Protein–protein or protein–nucleic acid intermolecular NOEs	1000	904	1000	904
Total dihedral angle restraints				
Protein	580	205	580	217
phi	288		288	
psi	292		292	
Nucleic acid				
Base pair				
Sugar pucker		29		29
Backbone				
Based on A-form geometry		176		188
EPR restraints (upper and lower limits)		42		42
<b>Structure statistics</b>				
Violations (mean and s.d.)				
Distance constraints ( $>0.2$ Å)	1.2 ± 1.0		2.5 ± 1.5	
Dihedral angle constraints ( $>5$ °)	0.0 ± 0.0		0.1 ± 0.3	
Max. dihedral angle violation (°)	0.61 ± 0.30		1.22 ± 2.14	
Max. distance constraint violation (Å)	0.25 ± 0.07		0.30 ± 0.06	
Deviations from idealized geometry				
Bond lengths (Å)	0.009		0.01	
Bond angles (°)	1.6		1.6	
Average pairwise r.m.s.d.* (Å)				
Protein (1-53, all six chains)				
Heavy	1.59 ± 0.57		1.24 ± 0.14	
Backbone	1.30 ± 0.66		0.83 ± 0.18	
RNA				
All RNA heavy (1-72)	1.83 ± 0.26		1.26 ± 0.25	
RNA binding site†	0.90 ± 0.17		0.97 ± 0.26	
Complex				
All complex heavy (C, N, O, P)‡	1.48 ± 0.51		1.19 ± 0.17	

\*Pairwise r.m.s.d. was calculated among 20 refined structures.

† Conformer L, 1–16, 19–36, 38–42, 44–56, 59–71; conformer R, 1–16, 19–36, 38–42, 44–56, 58–72.

‡ Protein: 1–53 and RNA same as in †.