doi:10.1038/nature13271

Supplementary Methods

Which parts of the isolated domains have the same structure in isolation as in the 70 kDa quaternary complex?

After having determined the structures of the separated domains[10], it has to be verified which parts also retain their structures in the context of the full complexes. Chemical shifts of nuclei are very sensitive to their chemical environment[52,53]. It has been shown that chemical shifts are sensitive to report on specific RNA helical conformations[54]. Therefore, comparing the chemical shifts of the resonances in the isolated domains and the domains in the full complex is a valid approach to identify if their structures are identical or very similar in both environments.

On the protein side, only a few resonances are isolated and many of them cannot unambiguously be assigned to a specific monomer bound to its stem-loop. Indeed most protein resonances have identical chemical shifts in the six monomers of the three homo-dimeric proteins and thus, prevent an unambiguous sequential resonance assignment using triple resonance experiments. Nevertheless, the ^{1}H-^{15}N HSQC spectra of the protein in complex with the isolated SLs or with the full RsmZ(1-72) RNA are matching well (Extended Data Fig. 2a) implying that the restraints for the protein structures in complex with the isolated SLs can be used for the full complex.

The situation is more difficult on the RNA side. Because the full complex contains 5 very similar ANGGAX RsmE binding motifs the resonance overlap is dramatic on the RNA side. Therefore, we segmentally labelled the RsmZ(1-72) RNA, by isotopically labelling only one protein binding site on the RNA at a time leaving the rest of the RNA unlabelled[11,12]. The resonances of the iminos, the aromatic H2 protons of the adenines (Extended Data Fig. 2b), the aromatic H8 protons of the purines as well as many H1′ sugar protons had a good line shape in the ^{1}H-^{15}N or ^{1}H-^{13}C HSQC or TROSY spectra and were mostly well separated in the segmentally labelled RNAs in the full complex. However, all the other sugar resonances (H2′ to H5″) and many H5 and H6 protons of pyrimidine bases within the stems were broadened beyond detection due to strong proton-proton dipolar relaxation induced by close protons. Only the H5 and H6 protons of the looped out pyrimidine bases

N and X of the ANGGAX motifs were visible in the ^{1}H-^{13}C HSQC or TROSY spectra indicating that they are more flexible compared to pyrimidines located in the stems. For many nucleotides (mostly in the loop and the upper part of the stem) we could unambiguously assign 2-3 ^{1}H-^{13}C or ^{1}H-^{15}N correlations per nucleotide that correspond to resonances in the isolated SL/RsmE complexes. Their identical or very similar chemical shifts compared to the isolated SL complexes report the same or very similar structure in the isolated domain and the full complex (Extended Data Fig. 2d).

In the lower parts of the stems, fewer resonances could unambiguously be assigned because of the following reasons: First, most pyrimidines H5 and H6 were broadened beyond detection in the stems (see above). Second, the chemical shift degeneracy is much more pronounced in the RNA stems compared to the nucleotides making contacts to the protein. For example, the H8-C8 resonance pairs of G46, G53 and G55 in SL3 resonate at the same frequency. Finally, the chemical shifts of the full complex resonances are not identical anymore to the isolated domains in the lower parts of some stems. This is expected because the stem orientation might (slightly) be changed in the full complexes in comparison to the isolated SL complexes. Even slight helical conformational variations can lead to significant chemical shift changes[54]. Furthermore, the base-pairs in the lower parts of the stems that are present in the separate SL complexes and also in the free RsmZ RNA at low temperature might transiently open at high temperature (40 °C) resulting in imino resonances that are not observable in the spectra of the full complex.

In SL1, SL2 and SL3, strong ^{1}H-^{15}N imino correlations with identical or very similar chemical shifts compared to the isolated complexes were present throughout the stem down to G2, G35 and G55, respectively. Although, the iminos of G35 and G55 were splitted into two peaks due to the two global conformations, the presence of strong imino peaks even at 40 °C and the very similar chemical shifts strongly indicate that the subsequent base-pairs in these three stems are also formed, namely: U1-A16, G19-U36 and U44-A56. The formation of these base-pairs is further supported for SL1 and SL3 in that the chemical shifts for the C2-H2 correlations in adenine A16 and A56, respectively are characteristic of a stacked base with a similar chemical environment as in the isolated stem-loop complexes (Extended Data Fig. 2b)[52]. Finally, for both conformations L and R, preliminary structure calculations in which the restrains for the stem closing base-pairs U1-A16, G19-U36 and U44-A56

were omitted, suggested the formation of these base-pairs as the RNA backbone was very similar when restraining or not restraining the base-pairs. Also, the target functions of the structures generated with or without these three additional base-pair were very similar indicating that formation of these base-pairs is in agreement with the global fold of the complex.

G71 is the last imino present in the stem of SL4. Although the chemical shift is very similar in the full complex and in the isolated SL4 in complex, the imino is very weak. Therefore, during preliminary structures calculations, we used the restraints from the isolated SL4 complex including the C59-G71 base-pair. For conformer R, however, these preliminary structure calculations suggested that the following U58-A72 base-pair is likely also formed. Furthermore, structure calculations including this base-pair resulted in very similar target function energies supporting the formation of this base-pair in conformer R.

Finally, for the single-stranded sequence between SL2 and SL3, several resonances were identical or very similar for the full and the isolated complexes, indicating that A38 to C42 form the same structure in isolation as in the full complex. At this point, it has to be mentioned that only one of the two global conformations has the same chemical shifts for A38 H2 in the isolated domain and the domain in context of the full complex (Extended Data Fig. 2b, red spectrum). However, it is likely that the structure of the A38 nucleotide is similar in both global conformations. A38 is the first adenine in the conserved **A**(N)GGAX motif and is almost identically recognized in all 5 isolated complex structures[7,10]. The reason for a different chemical shift of the adenine A38 in both global conformations is likely due to a different structure of the previous cytosine C37 nucleotide in both global conformations providing a different chemical environment sensed by the adenine A38 H2.

Overall, for the structure calculation of the full complexes, we implemented all the RNA restraints from the isolated complexes for the following residues (Extended Data Fig. 2d): U1-A16, G19-U36, A38-C42, U44-A56 and C59-G71. For conformer R, in addition, we also used the constraints for the stem-closing base-pair of SL4, U58-A72.

Due to geometrical reasons both low distance peaks in the AC and AG distance distributions belong to conformer L, while the high distance peaks belong to conformer R.

The spin labels at positions C and G are both located on the same domain 1, and have therefore a fixed distance of around 25 Å between the centers of their spin label distributions. Therefore, it is geometrically not possible that the left AC peak (30 Å) belongs to the same conformation as the right AG peak (75 Å). It is only possible that the two left peaks belong to one conformer (conformer L) and the two right peaks correspond to the other conformer (conformer R). Note that doing different structure calculations for testing all four possible AC/AG distance combinations, two could directly be discriminated due to high target function energies.

Selecting the correct cluster combinations by comparing measured and back-calculated distance distributions.

To determine which of the two possible solutions for conformer L represents the correct one, we modelled the distance distributions for both solutions of conformer L in combination with the unique solution of conformer R and compared them to the experimental data (Extended Data Figs. 8 and 9). To model a certain distance distribution, we superimposed the two "radical clouds" (spin-label distributions) corresponding to the two spin labels onto one structure of a certain cluster (structural ensembles belonging to the same solution) and measured all the possible distances from each radical position of spin label 1 two each radical position of spin label 2 (Extended Data Fig. 8a) with a home-written Matlab script, which uses several routines from the MMM package[55] (www.epr.ethz.ch/software/index). We did the same for all the structures in a specific cluster and summed up all the distances. After normalization we could obtain the modelled distance distribution for one single cluster. The same was performed for the cluster of the other conformation and the two distance distributions summed up to yield the back-calculated distance distribution for a certain cluster combination (Extended Data Fig. 8b). By putting a weight of 33 % and 67 % for conformer L and R, respectively (based on the relative integrals of the two peaks in the AC and AG distance distributions), we could identify the best cluster combination (Extended Data Figs. 8b and 9c, d). Some distance distributions were not in perfect agreement in both possible combinations (mean

distance around 10 Å off). This is still in agreement with the experimental data as for some spin label positions we found a deviation from the modelled distance distributions also for the model systems[18]. Note that for rotamer library approaches a comparison to 92 experimental DEER distance distributions on eight proteins with known crystal structures revealed a standard deviation for the predicted mean distances of 3.2 Å[56]. Still, in some cases, significantly larger deviations of ~10 Å and even slightly beyond that value were found. Looking at the model systems[18], the most trustable spin label positions are A, G and C, indicating that the AC and AG distance distributions are the best indicators to discriminate between the different cluster combinations (Extended Data Fig. 8b). Furthermore, the AC and AG distance distributions are the only ones with two separated peaks.

Overview RNA constructs used.

All the RNA constructs used in this study are shown (from 5′ to 3′). The GGA binding motifs are underlined.

Isolated binding sites of RsmZ sRNA:

SL1: GG GUGUCGACGG AUAGACACCC

SL2: GG GCCA UCAA GGAC GAUG GU CC

SL3: GGG AUCGCAGGAA GCGAUCCC

SL4: GGGUCAUCAGG ACGAUGACCC

9nts-GGA$_{39\text{-}41}$: UCAGGAC AU

23nts-GGA$_{76\text{-}78}$: GAGAAAGGAACACAGAGACUAGG

23nts-GGA$_{85\text{-}87}$: GAGAAAAGAACACAGGGACUAGG

GGA$_{term}$: GAGCUAGGGAAAAAUGUGGGCGGGUCAUACCGCCCCUUUUUUU

Other constructs of RsmZ sRNA:

SL23_CUC: UGCUUCGGCA CAGCCAUCAA GGACGAUGGU CACUCCAUCG

CAGGAAGCGA UUCACUUCGGUGA

SL23_ΔGGA: UGCUUCGGCA CAGCCAUCAA GGACGAUGGU CACAUCGCAG GAAGCGAUUC ACUUCGGUGA

SL1234_ΔGGA_GGA$_{39\text{-}41}$: GGC UGUGCUUC GGCACACAGCCA CUUCGGUG GUCAGGAC AUCG CUUCGG CGAUUCACU UCGGUGA

RsmZ(1-72): UGUCGACGGA UAGACACAGC CAUCAAGGAC GAUGGUCAGG ACAUCGCAGG AAGCGAUUCAUCAGGACGAU GA

RsmZ(1-89): GAA UGUCGACGGA UAGACACAGC CAUCAAGGAC GAUGGUCAGG ACAUCGCAGG AAGCGAUUCA UCAGGACGAU GAAAAGGAAC ACAGGGACU

RsmZ-wt: GAA UGUCGACGGA UAGACACAGC CAUCAAGGAC GAUGGUCAGG ACAUCGCAGG AAGCGAUUCAUCAGGACGAU GAAAAGGAAC ACAGGGACUA GGGAAAAAUG UGGGCGGGUC AUACCGCCCCUUUUUUU

Constructs from *hcnA* mRNA 5′ UTR:

hcnA-GGA#3-5: GAGCAUGGACGGCGGGACGCCGGGUACCCCAUU CAUUUUUCACGGAUGAA

Supplementary Discussion

Justification for solving the solution structure of RsmZ(1−72) in complex with three RsmE protein dimers

Compared to the 127 nt *wild type* RsmZ sRNA, the RNA construct used to determine the solution structure of RsmZ bound to three RsmE protein dimers is missing the second GGA motif ($GGA_{85–87}$) binding to the third RsmE protein dimer.

We have several experimental indications that the binding of the $GGA_{85–87}$ motif to the third RsmE dimer does not have a large influence in the sequestration potential of the third dimer. First, gel shift assays of the RsmZ(1−72) construct used to determine the structure shows a K_d of around 300 nM for the binding of the third dimer (see Extended Data Fig. 6a), while the *wild type* RsmZ sRNA has a K_d of around 200 nM for the binding of the third dimer (see Fig. 1e). Second, the *in vitro* translation activation potential of the RsmZ(1−72) sRNA and a construct binding the three dimers with all six GGA motifs is comparable. Furthermore, the RsmZ sRNA from *P. aeruginosa* is missing $GGA_{85–87}$ and has a comparable translation activation potential as the RsmZ sRNA from *P. fluorescens* used in this study (see Extended Data Fig. 7b,c). Finally, although our structures have been solved with a missing $GGA_{85–87}$ motif, the global structures in both conformations allow a perfect modeling for the binding of the $GGA_{85–87}$ motif to the free binding site of the third dimer (see Fig. 3d as an illustration). This strongly indicates that the global structure is very similar with and without the $GGA_{85–87}$ bound to the third dimer.

NMR evidence for two conformations present in solution

In the spectra of the segmentally labelled RNAs in complex, we observed two peaks for resonances at the lower parts of most stem-loops, or differently stated, at most connections between the three domains in the full complex indicating the presence of two different global conformations of the full complex that are in slow exchange on the NMR time scale. In the segmentally labelled RNA in complex, in which segment U36-U44 is isotopically labelled, the resonances of A38 and A43

appeared each as two well separated peaks (see Extended Data Fig. 2b, red spectrum). In the segmentally labelled SL1 complex, A16 H2 is splitted into two peaks. In SL2, the imino of G35 is broad indicating two conformations with slightly different chemical shifts (see Extended Data Fig. 2c). In SL3, the imino of G55 shows two peaks with slightly different chemical shifts. Only in SL4, we do not have evidence for two conformations because the last imino that can be observed in the stem (G71), is too weak to reveal a potential second conformation. Overall, the occurrence of two peaks for many nucleotides located at the connections between the rigid domains indicates that these two conformations are not local but involve rearrangements of whole domains, which is in agreement with the DEER data (Extended Data Fig. 3b).

The determined K_d's are an average of the two conformers

While we can distinguish two conformations with NMR and EPR spectroscopy (Fig. 2b and Extended Data Fig. 2b, c), the native RNA gels (Fig. 1b) and gel shift assays (Fig. 1e) show only one band for both conformations for each intermediate on the assembly pathway. This suggests that the global shape is similar for both conformations of each intermediate on the assembly pathway. Our structures show that this is indeed the case for the 1:3 complex (Fig. 2a). We cannot isolate one of the two conformations and the determined K_d values are therefore an average of both conformations.

To illustrate the meaning of an average K_d value, we can consider the simple case of an RNA interconverting between two conformations (L and R), which both can bind one protein and in which protein binding prevents a direct interconversion between the two forms (see Extended Data Fig. 4g as an illustration). In such a case, the K_d's for the binding of the protein in the two parallel assembly pathways and the average K_d are defined by:

$$Kd\ (conformer\ L) = \frac{[RNA,L][protein]}{[complex,L]} \quad (1)$$

$$Kd\ (conformer\ R) = \frac{[RNA,R][protein]}{[complex,R]} \quad (2)$$

$$Kd\ (average) = \frac{([RNA,L]+[RNA,R])[protein]}{[complex,L]+[complex,R]} \quad (3)$$

Assuming that the K_d of conformer L is 10-fold higher (for example because the on-rate for binding of the protein to the RNA conformer L is 10-fold lower than to RNA conformer R), the two protein-bound conformations can still be equally populated, when the free RNA conformer L is 10-fold higher populated than conformer R. This shows that the K_d's for the binding of the protein in the two individual assembly pathways can be significantly different, although the two protein-bound conformations are equally populated.

While, the experimentally determined average K_d does not characterize the binding of the individual pathways at the molecular level, it provides the biological relevant affinity describing the total sequestration potential of the sRNA independent of the conformation it is present.

Negative allostery for binding of the second high affinity RNA molecule to the RsmE protein homo-dimer

Interestingly, for the high affinity SL2, the ITC data can only be fitted using a two-site binding model, in which binding of the first SL2 molecule to the RsmE homo-dimer binds with about a 10-fold higher affinity as binding of the second SL2 molecule ($K_D^{(1)}$ = 16 ± 3 nM and $K_D^{(2)}$ = 183 ± 3 nM) (Fig. 3a). The lower affinity for the second SL2 binding to RsmE might come from sterical hindrance by the long stem of the first SL2 molecule bound to RsmE. Alternatively, the binding of one SL2 molecule could lead to structural changes on the unbound binding site decreasing its binding affinity (Fig. 3a). To distinguish between these two possibilities, we performed an NMR chemical shift titration by titrating increasing amounts of unlabelled SL2 RNA to ^{15}N-labelled RsmE protein. Strikingly, for most protein resonances, we could observe two sets of "free protein" resonances and two sets of bound protein resonances during the course of the chemical shift titration (Extended Data Fig. 5). The two sets of "free protein" resonances correspond to the unbound protein binding site when the other RNA binding site is either free or occupied. In analogy, the two sets of bound protein resonances correspond to the bound protein side when the other binding site is either free or bound. This observation strongly supports a negative allosteric effect involving slight global structural rearrangements and/or dynamical changes leading to a reduced binding affinity for the second SL2

molecule binding to RsmE. Interestingly, we did not detect such behaviour for any of the low μM range affinity RNA targets. Overall, we conclude that a high affinity SL is able to induce an experimentally observable structural and/or dynamical change on the protein and thus, is exerting a negative allosteric effect.

Supplementary References

52 Aeschbacher, T. *et al.* Automated and assisted RNA resonance assignment using NMR chemical shift statistics. *Nucleic Acids Res.* **41**, e172 (2013).

53 Aeschbacher, T., Schubert, M. & Allain, F. H. A procedure to validate and correct the 13C chemical shift calibration of RNA datasets. *J. Biomol. NMR* **52**, 179-190 (2012).

54 Bullock, S. L., Ringel, I., Ish-Horowicz, D. & Lukavsky, P. J. A'-form RNA helices are required for cytoplasmic mRNA transport in Drosophila. *NSMB* **17**, 703-709 (2010).

55 Polyhach, Y., Bordignon, E. & Jeschke, G. Rotamer libraries of spin labelled cysteines for protein studies. *Phys Chem Chem Phys* **13**, 2356-2366 (2011).

56 Jeschke, G. Conformational dynamics and distribution of nitroxide spin labels. *Prog Nucl Mag Res Sp* **72**, 42-60 (2013).