RNA (1997), 3:255–268. Cambridge University Press. Printed in the USA.

The RNA binding site of S8 ribosomal protein of *Escherichia coli*: Selex and hydroxyl radical probing studies

HERVÉ MOINE,[1] CLAIRE CACHIA,[2] ERIC WESTHOF,[1] BERNARD EHRESMANN,[1] and CHANTAL EHRESMANN[1]

[1]UPR 9002 du CNRS, Institut de Biologie Moléculaire et Cellulaire, 67084, Strasbourg, France
[2]Laboratoire de Biophysique, Faculté de Pharmacie, 21033, Dijon, France

ABSTRACT

The RNA binding site of ribosomal protein S8 of *Escherichia coli* is confined to a small region within the stem of a hairpin in 16S rRNA (nt 588–605/633–651), and thus represents a model system for understanding RNA/protein interaction rules. The S8 binding site on 16S rRNA was suspected to contain noncanonical features difficult to prove with classical genetical or biochemical means. We performed in vitro iterative selection of RNA aptamers that bind S8. For the different aptamers, the interactions with the protein were probed with hydroxyl radicals. Aptamers that were recognized according to the same structural rules as wild-type RNA, but with variations not found in nature, were identified. These aptamers revealed features in the S8 binding site that had been concealed during previous characterizations by the high base conservation throughout evolution. Our data demonstrate that the core structure of the S8 binding site is composed of three interdependent bases (nt 597/641/643), with an essential intervening adenine nucleotide (position 642). The other elements important for the binding site are a base pair (598/640) above the three interdependent bases and a bulged base at position 595, the identity of which is not important. Possible implications on the geometry of the S8 binding site are discussed with the help of a three-dimensional model.

Keywords: base-triple; hydroxyl radical probing; in vitro selection; RNA structure

INTRODUCTION

The variety of intramolecular interactions that can be used by RNA to build specific structural features has been growing steadily during recent years. Noncanonical interactions seem to play a crucial role in building the three-dimensional motifs involved in the interactions between RNA and other ligands. The unraveling of such features remains one challenging goal in the study of RNA folding. For instance, site-directed mutagenesis is usually tedious and of limited use in such a process because the number of combinations to be tested is too large. Comparative sequence analysis by itself (Gutell et al., 1994) or substantiated with experimental structure probing (Michel et al., 1990) can be useful in demonstrating the existence of some noncanonical interactions. As an alternative approach, iterative in vitro genetic selection of RNA molecules that bind specific ligands (Ellington & Szostak, 1990) or Selex (Systematic Evolution of Ligands by EXponential enrichment [Tuerk & Gold, 1990]) represents an attractive way to investigate noncanonical features of RNA. This approach, which basically generates an artificial phylogeny, has been developed for a broad range of purposes (for review Conrad et al., 1996). The selection of variants able to adopt similar structural features with different primary structures can bring clues to the identification of noncanonical interactions, in particular because it allows testing of more variations than found in nature.

Studies from this laboratory (Mougel et al., 1987, 1993; Allmang et al., 1994) and from another laboratory (Gregory & Zimmermann, 1986; Gregory et al., 1988; Wu et al., 1994) have attempted to establish the structural requirements of ribosomal protein S8 binding on the 16S ribosomal RNA of *Escherichia coli*. Deletion experiments and structure probing have demonstrated that the recognition features are essentially contained in a composite hairpin structure extending from nt 588 to 651 in the *E. coli* 16S rRNA. The nucleotides essential for *E. coli* S8 (EcS8) binding were mapped in the center of the bottom stem (Mougel et al., 1987, 1993; Allmang et al., 1994) (Fig. 1).

Reprint requests to: Hervé Moine, UPR 9002, CNRS, 15 rue R. Descartes, 67084 Strasbourg cedex, France; e-mail: moine@ibmc.u-strasbg.fr.

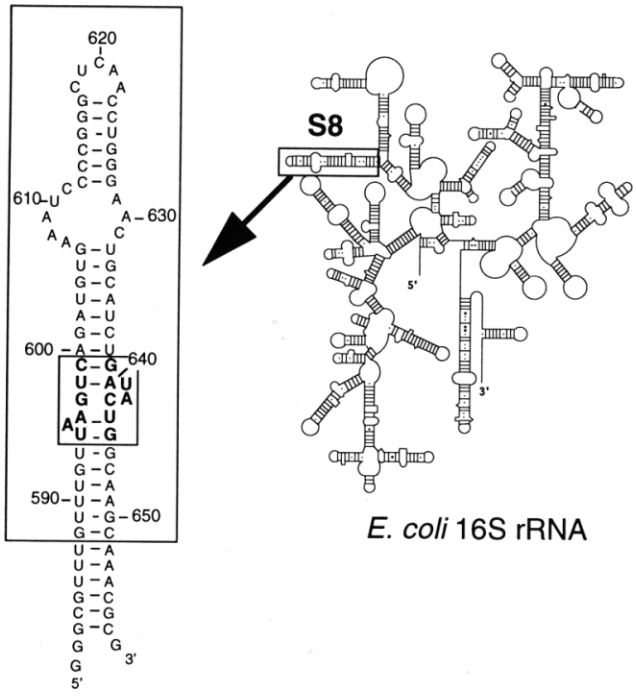

FIGURE 1. Secondary structure of the region containing the protein EcS8 binding site (at left) with its location on a schematic secondary structure of the 16S rRNA (at right). Nucleotides in bold were randomized in the present Selex experiments. Nucleotides outside the box correspond to sequence added for cloning purposes.

Other protection sites were identified elsewhere in the 16S rRNA (Powers & Noller, 1995). However, these additional protections are directly dependent upon binding of S8 to its anchoring site at 597–599/640–643 and are lost if this anchoring site is disrupted (such as by deletion of A 595) (H. Moine, unpubl.). The fact that the minimal hairpin structure (nt 588–651) is necessary and sufficient to provide an affinity identical to that of the whole 16S rRNA supports the proposal that these additional protections reflect a conformational change in the 16S rRNA upon binding of S8 rather than direct protection.

The anchoring site of protein S8 has been reported to consist either of a small asymmetrical internal bubble in the middle of an otherwise regular helix (Wu et al., 1994), or of a more constrained site involving non-classical Watson–Crick interactions with bulging nucleotides (Allmang et al., 1994). In both cases, the consequence is the generation of irregularities in the RNA backbone providing specific geometry and accessibility to essential atomic positions of specific residues. To investigate further the structure of the EcS8 binding site, we performed in vitro iterative selection on a 79-nt hairpin RNA encompassing nt 588–651 of *E. coli* 16S rRNA. Two windows of 6 and 7 nt, encompassing residues essential for the binding (594–599 and 639–645), were randomized on each strand of the hairpin structure. Two populations of RNA variants were selected with binding properties identical or similar to wild-type (wt) RNA: one population has features similar to those found in the natural phylogeny, whereas the other population possesses features not found in nature. The hydroxyl radical footprint of EcS8 on the different aptamers enabled us to show that variants of the second class, although lacking some of the universal features of prokaryotic EcS8 binding site, are actually recognized according to the same structural rules. Implications for the geometry of the EcS8 binding site on the RNA are presented.

RESULTS

Experimental strategy

Selex experiments were performed on a hairpin RNA encompassing the minimal binding site of EcS8 defined previously (Mougel et al., 1993). A 79-nt RNA was transcribed from a dsDNA product generated by PCR. This dsDNA contains the sequence corresponding to nt 588–651 from *E. coli* 16S rRNA under T7 promoter control with one restriction site present at each end for cloning purposes. It was verified that, despite the presence of additional sequence, the transcribed RNA binds EcS8 with the same dissociation constant (K_d) as the wt RNA (3 nM). Two windows of 6 and 7 nt on each strand of the stem, encompassing nt 594–599 and 639–645, were randomized (Fig. 1). These windows cover the nucleotides shown to be involved in EcS8 binding. The Selex experiments were started with a population of 67×10^6 RNA variants and at a concentration of 10^{-5} M (which represents an estimated 10^6 copies of each variant in the assay); four rounds of selection were performed in order to shift the affinity of the bulk RNA from a value estimated to be above 10^{-5} M to a value around 10^{-8} M at the concentration of EcS8 protein used (10^{-7} or 10^{-6} M) (Fig. 2). Lower concentrations of protein did not permit protein-specific selection because the RNA population was taken over rapidly by RNAs that bind to nitrocellulose. Although sequence of the bulk RNA was not random after four rounds of selection, no distinct sequence could be identified clearly from the RNA pool at this level.

Sorting out the RNA variants into two classes

After cloning into pUC18 vector (see Materials and Methods), 160 clones were analyzed over four independent Selex experiments. Among these clones, 83 individual clones bound EcS8 specifically. Thirty of them yielded RNAs with K_ds increased between 5- and 10-fold, or having a low plateau, or showing sequence heterogeneity. These species were not studied further. The rest (53 clones) had a K_d value undistinguishable from wt RNA. Sequence analysis revealed that 28 of them, representing 53% of the binding RNAs, contained wt sequences. This ratio remained constant

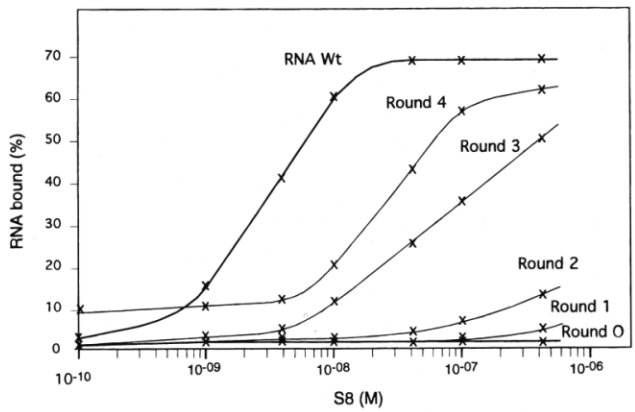

FIGURE 2. Typical binding curves of EcS8 on the various pools of RNAs used to perform the Selex experiments. RNA wt is fragment 588–651 of *E. coli* 16S rRNA with additional nucleotides (see Materials and Methods), but no random sequences. Binding curves of RNA pools from round 0 to round 4 are shown. One standard deviation was less than 5%.

throughout several distinct Selex experiments. The 25 remaining clones contained 19 distinct sequences (Table 1). In order to sort out the various species obtained, the sequences were compared with the consensus sequence (Fig. 3) derived from a pool of 2,496 procaryotic small subunit rRNAs (Maidak et al., 1996). The selected RNAs were also tested for their binding properties and their ability to compete with wt RNA. Typical competition curves are shown on Figure 4 and data are summarized in Table 1. However, an analysis of the footprints obtained with hydroxyl radicals in the presence of bound EcS8 (Fig. 5) allowed the classification of the various RNA species. Reaction of iron/EDTA with hydrogen peroxide generates hydroxyl radicals that induce the cleavage of the ribose/phosphate backbone independent of secondary structure (Imlay et al., 1988; Latham & Cech, 1989). Thus, despite the apparent lack of sequence homology of certain RNA species, the common protection pattern observed (Fig. 6) was a clear indication that these RNAs present similar structural features to the protein. Because the numerical position of the nucleotides building the binding site varies within the stem of the hairpin among different variants, we will hereafter refer to specific residues with respect to their position relative to the universally conserved A642 in the wt counterpart. This adenine residue was found to be conserved throughout most of the selected variants. Thus, numbering of the conserved nucleotides building the S8 binding site in the variants will hereafter appear as X. With this procedure, the RNA variants were classified into two distinct populations (Table 1).

Class I: Wt-like RNAs

Eight RNA variants (1.1–1.8) have sequences similar to the consensus obtained from natural phylogeny. Nucleotides apparently equivalent to the universally conserved nt G597, A642, and C643 (nucleotides shown in

TABLE 1. The two classes of RNA variants isolated with the selection procedure[a]

	#	Isolates	Sequences	Extra mutations	K_d Competition
	Wt	28×	UAAGUC GAUACUG		3 nM 1.0
Class I	1.1	2×	GUGUGA UCAAACU		= 1 ± 0.5
	1.2		AUGUUU AUAAACU		= 3 ± 1
	1.3		AUGUGU AUAAACU		= 3 ± 1
	1.4		CCGUCA UGAAACU		= 3 ± 1
	1.5		AAGUCG UGAAACU		= 3 ± 1
	1.6		AAGUCA UGAAACU		= 1 ± 0.5
	1.7		AUGUCG CGAAACU		= 1 ± 0.5
	1.8		AAGCGG CAAACCU	ΔC613–G627+U601+ΔC637	= 1 ± 0.5
Class IIa	2.1		AAUUCG CGACAGU		= 1 ± 0.5
	2.2		AAUUGG CCACAGU	C603+A615+ΔA621–U625+ΔA649	= 1 ± 0.5
	2.3		AAUUGG CUACAGU	ΔA649	= 1 ± 0.5
	2.4		AACUGG CCAAAGU	ΔA649	= 0.3 ± 0.1
	2.5	2×	GUCUGA CUUAUCG		= 0.3 ± 0.1
Class IIb	2.6		AUAUCA GUUUGAU	ΔG607–A608+ΔC611+ΔC623–G628	= 0.3 ± 0.1
	2.7		AAACGC GUUCCUG	ΔG604–A608+U612	= 3 ± 1
	2.8		UGGCAA UCGCAUG	ΔG604–A608+U612+ΔC647	= 3 ± 1
	2.9		UAGCAC GCUCCGC	ΔG604–A608+U612+C619+ΔC634	= 3 ± 1
	2.10	5×	ACCGGA UGGUCUG		= 0.3 ± 0.1
	2.11		ACCGGA UGGACUG		= 0.3 ± 0.1

[a]Sequence of the variants within the randomized window 594–599 and 639–645 is indicated together with the number of times each variant was isolated. Grey shadowed nucleotides are conserved from wt sequence. Black shadowed nucleotides are universally conserved among procaryotes. Mutations identified outside the Selex window are indicated. K_d is the apparent dissociation constant as determined by filter binding assays; "=" means a constant identical to the wt ($3 \pm 1 \times 10^{-9}$ M) was found. Competition strength of the different variants is given and is expressed as the ratio of mutant and wt RNA concentrations required to provide a 50% competition value (see Fig. 4 and text for details). Numbers represent the mean value of a minimum of three independent determinations.

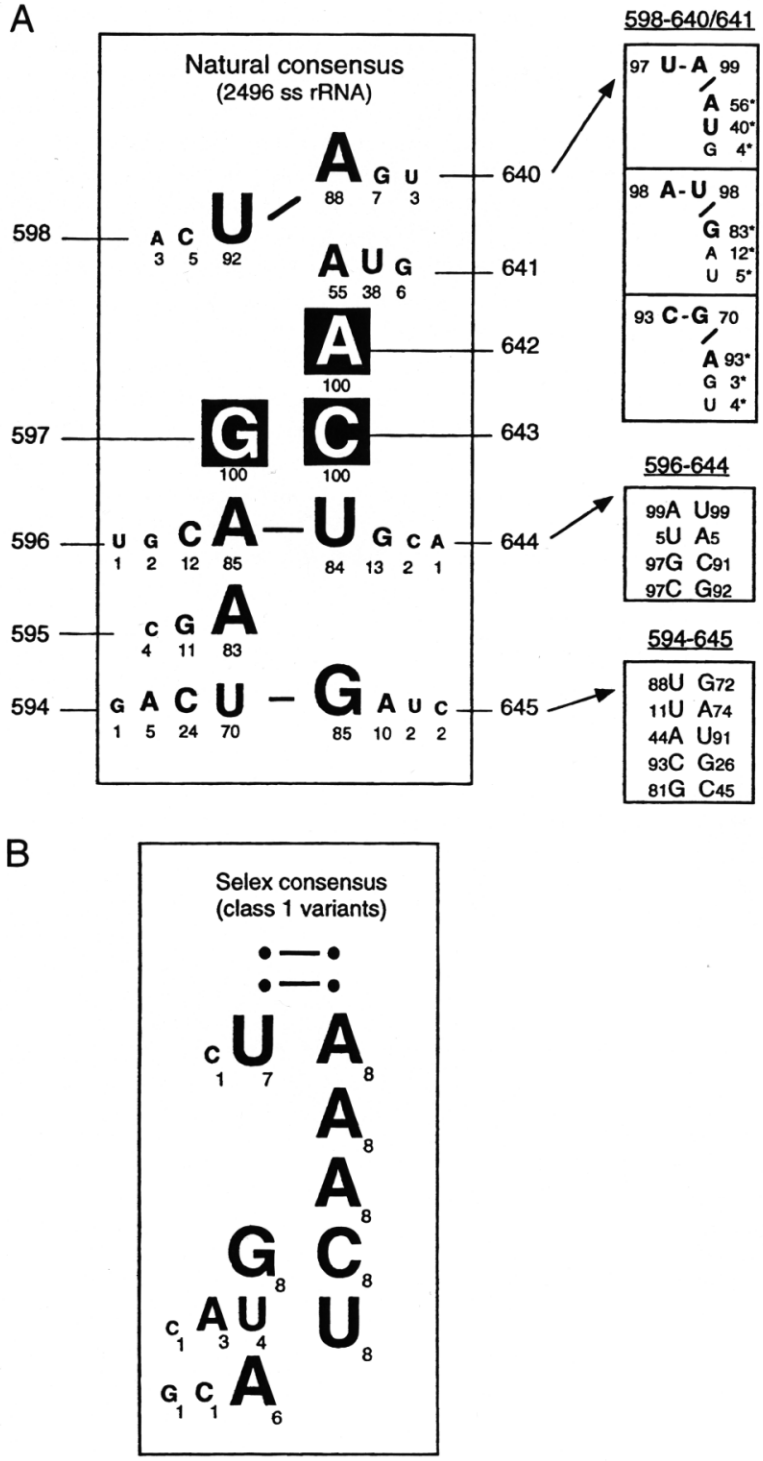

FIGURE 3. **A:** Consensus sequence of 16S rRNA regions 594–598 and 640–645 obtained from 2,496 prokaryotic small subunit rRNAs (sequences are from the Ribosome Database Project [Maidak et al., 1996]). The size of a particular base reflects its degree of conservation, the number next to the base is its percentage of occurrence over whole small subunit rRNAs determined as: (number of times this base appears/2,496) × 100. Covariation of bases is indicated in the boxes on the right as the percentage of occurrence of a given base when its Watson-Crick counterpart is present on the other strand; numbers marked with an asterisk indicate the percentage of occurrence of base 641 when the given Watson-Crick base pair at 598–640 is occurring. **B:** Consensus sequence obtained from the pool of class I variants (eight distinct sequences). The number of times a given base appeared is shown next to each base. Black dots linked by a bar indicate a nonconserved Watson-Crick base pair.

a black background in Table 1) were identified, but their positions are shifted downward in the helix for most of the variants. However, a few changes were found in these variants compared to wt (nucleotides not shadowed on Table 1). The homology between nucleotides from wt RNA and the conserved but shifted sequence pattern observed in class I RNA variants is demonstrated clearly by the fact that all these RNAs display the same hydroxyl radical footprints in the presence of EcS8 with respect to the positions of the universally conserved residues (upper part of Fig. 6). Notably, EcS8 induces the same footprint on the *Thermus thermophilus* RNA binding site (Fig. 6). Main protections in wt RNA were found at nt 598 and 599 on one strand and nt 640–644 on the other one. Minor sites of protection were also observed at the bottom of the stem (nt 588–589). This is consistent with the protection data of the corresponding phosphate groups

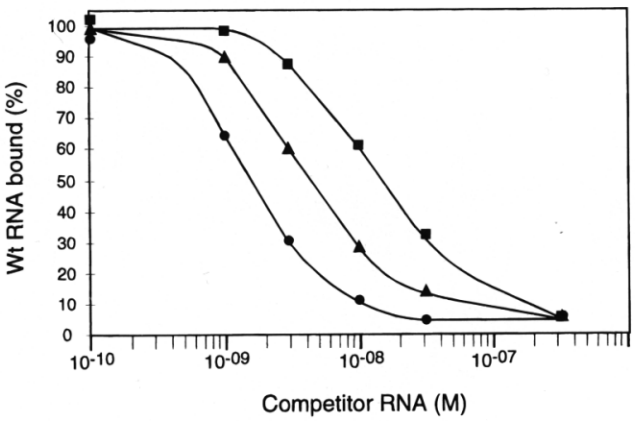

FIGURE 4. Competition between wt and the different RNA variants for complex formation with EcS8. Concentration of EcS8 was 3 nM. Wt RNA was uniformly labeled and added at a negligible concentration ($<10^{-11}$ M). Each curve is the mean value of four independent experiments, standard deviation was less than 5%. Increasing concentrations of unlabeled competitor RNA were added: ▲, wt RNA and RNAs 1.1, 1.6, 1.7, 1.8, 2.1, 2.2, 2.3; ●, RNAs 1.2, 1.3, 1.4, 1.5, 2.7, 2.8, 2.9; ■, RNAs 2.4, 2.5, 2.6, 2.10, 2.11.

toward ENU attack observed previously (Mougel et al., 1987). In all these variants, the major protection site is shifted by 1 nt along the stem within the randomized window (Fig. 6) with respect to the position of the shifted conserved residues. This is also accompanied by a simultaneous shift of the minor site, thus maintaining a constant distance between the two protection sites. Most likely, the selection exploited the last base pair of the bottom stem, which was not randomized, and used it as the conserved U594–G645 at the bottom of the EcS8 box. These observations indicate that EcS8 interacts with these RNA variants in the same way as with wt RNA. As expected, all these RNAs compete with wt RNA for EcS8 binding (Table 1). However, some of the RNAs (1.2–1.5) show a slightly decreased competition efficiency (by a factor of 3). Competition experiments provide a more sensitive assay than direct binding assay. This is rationalized by the observation that the excess of protein to RNA in the direct binding assay, in comparison with the equimolar amounts of RNA to protein in the competition assays, might minimize the effects due to the existence of an equilibrium between RNA conformers (Romaniuk, 1989). One can, therefore, postulate the existence of such conformers in the 1.2–1.5 variants.

The consensus site derived from these mutants was found consistent with the natural phylogenic variations (Fig. 3). Nucleotides 640 and 641 (conserved as adenines in 88 and 55% of the cases, respectively) are always adenines in class I variants. U598 (conserved at 92%) is replaced by a C in one of the RNAs (1.8). Nucleotides A595, A596, and U644 are also well conserved, both in nature (83–85%) and in class I variants. Therefore, U644 could potentially be paired with either A595 or A596. This is also the case for RNAs 1.5 and 1.6. However, nt 596 is replaced by U in RNAs 1.1, 1.2, 1.3, and 1.7, suggesting that A595 (or G595 in RNA 1.1) rather than A596 is paired with U644 in these RNAs. Conversely, the replacement of A595 by U in *E. coli* RNA, which does not affect EcS8 binding (Allmang et al., 1994), favors a A596–U644 base pair, consistent with probing experiments. A similar situation was observed in *T. thermophilus* RNA, in which G596 can pair with C644, leaving G595 as a bulged nucleotide. In the particular case of RNAs 1.4 and 1.8, no evident pairing was found. Therefore, the exact position and the nature of the bulged nucleotide (595 or 596) do not appear to be critical. Nevertheless, the presence of a bulged nucleotide was shown to be required (Allmang et al., 1994). The nature of base pairs 594–645 and 599–639, which flank both sides of the site, does not appear to be important, as already observed in natural RNAs.

Class II: RNAs differing from natural RNAs

The other population of variant RNAs (16 clones containing 11 different sequences) possesses features that were not found in natural procaryotic rRNAs. Indeed, all or some of the universally conserved nt G597, A642, and C643 are lacking and many or all nucleotides in the window of selection differ from wt. No obvious consensus could be deduced by thorough examination of the RNA sequences, although they all compete efficiently (with slight differences) with wt RNA for EcS8 binding (Table 1). This problem was solved by the analysis of the hydroxyl radical footprints induced by EcS8, which allowed us to map precisely the EcS8 binding site.

Remarkably, a similar topological distribution of the protected residues was found in RNAs 2.1, 2.2, 2.3, and 2.4. Although these RNAs were lacking obvious sequence homology with class I RNAs, their hydroxyl radical footprint induced by EcS8 was found identical or similar to the previous class (Fig. 6). Thus, as in class I RNAs, two distinct sites were observed: a major one is centered around an adenine residue equivalent to A642, and a minor one separated by five base pairs from the other one. Strikingly, the two other universally conserved nucleotides, G597 and C643 (shaded in Fig. 6), are substituted. G597 is replaced by U in RNAs 2.1, 2.2, and 2.3 and by C in RNA 2.4, whereas C643 is replaced by G in the four RNAs. This is an unexpected result because it was shown that the postulated G597–C643 pair could not be inverted (as observed in RNA 2.4) in the *E. coli* context (Mougel et al., 1993; Allmang et al., 1994). Therefore, another mutation should be responsible for the restored affinity. The most likely candidate is the nucleotide corresponding to U641 (also shaded in grey in Fig. 6), which is replaced by A in RNA 2.4. Moreover, U641 is replaced by C in RNAs 2.1, 2.2, and 2.3, in association with the U597 and G643 substitution. The U to C mutation at

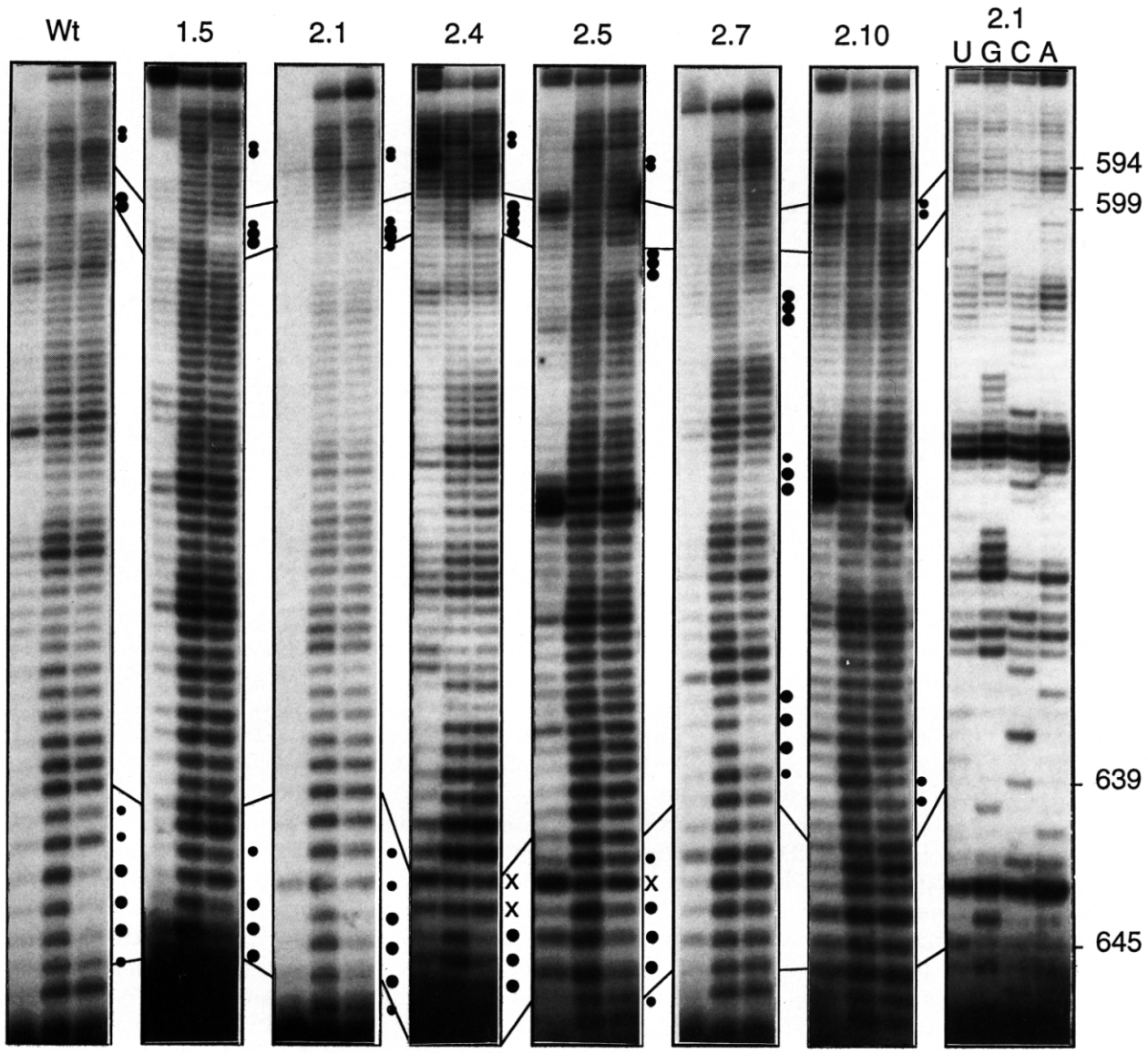

FIGURE 5. Autoradiographs presenting footprint of EcS8 RNAs as determined with hydroxyl radicals on wt RNA and on the main types of variants from class I and II. In each panel, from left to right: control lane without reagent, RNA alone, RNA + EcS8. Small and large dots point out strong and medium/weak protections, respectively. "X" represents undefined reactivity. The two windows of randomization, 645–639 and 599–594, are indicated.

position 641 was shown to induce a 10-fold decrease of affinity in the *E. coli* context (Mougel et al., 1993; Allmang et al., 1994). Our data reveal a possible covariation between nt 597, 641, and 643 (in grey in Fig. 6). This covariation does not exist in nature. Indeed, the natural consensus indicates only two possible combinations at those positions: G597/A641/C643 or G597/U641/C643. In the present class II variants, we found: U597/C641/G643 (RNA 2.1, 2.2, and 2.3) and C597/A641/G643 (RNA 2.4). It is noteworthy that both U598 and A640 were found conserved in these four variants.

In the case of RNA 2.5, the hydroxyl radical footprint is located on the 3′ strand of the randomized window, but unexpectedly downstream of the randomized window on the 5′ strand. Nevertheless, a class II-like binding site could be reconstructed by slipping the 5′ strand by 2 nt with respect to the 3′ strand (Fig. 6). In these conditions, the major site was found centered around the conserved A642, as in other RNAs. A second minor site was also found in the bottom of the stem, but irregularities in the stem were observed. If our assumption is correct, a third covariation is suggested: A597/U641/U643. It is noteworthy that, in this variant, nt U598 and A640 are substituted by A and U, respectively. From these observations, RNAs 2.1–2.5, which show covariations at positions 597/641/643, defined the subclass IIa.

The rest of the class II variants (class IIb) show more ambiguous hydroxyl radical footprints. However, a

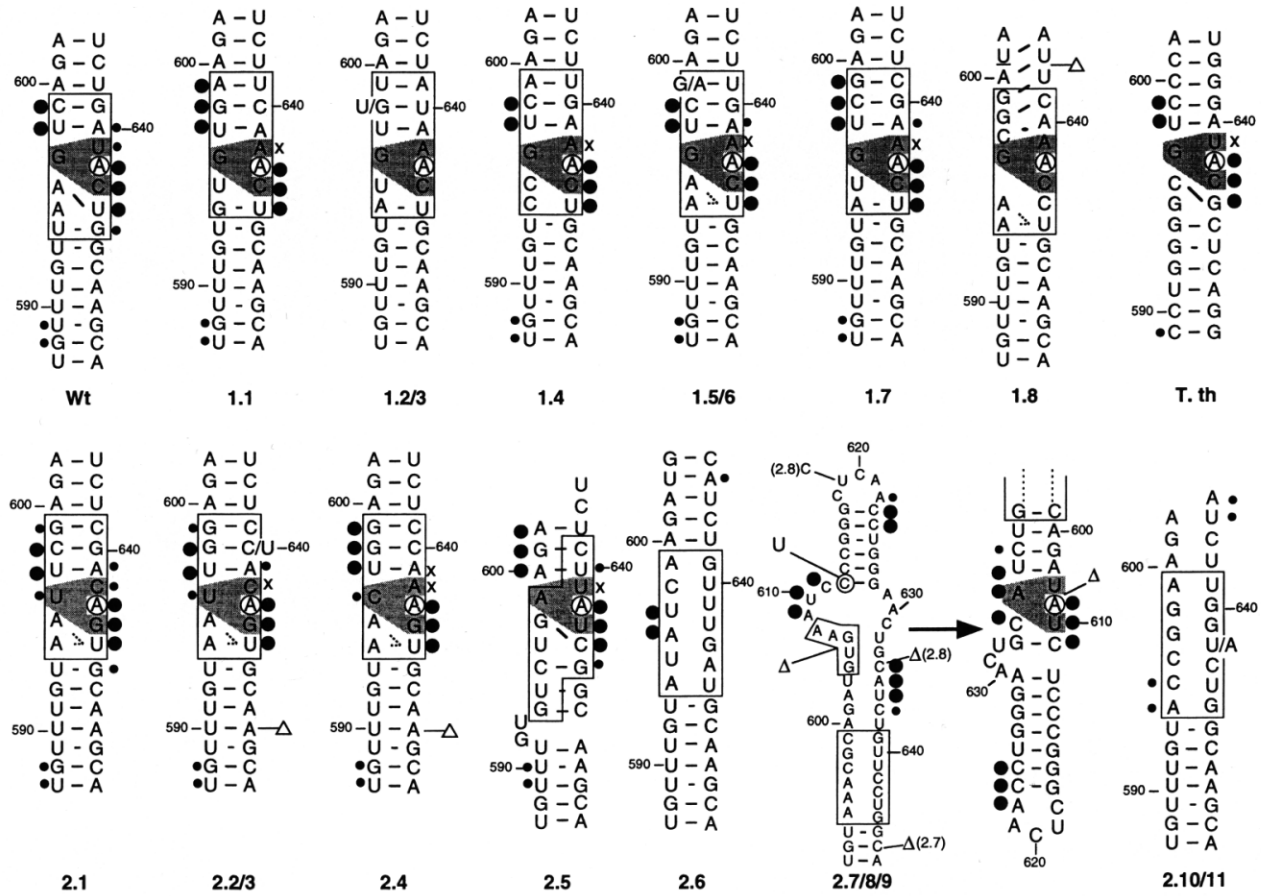

FIGURE 6. Hydroxyl radical footprints of EcS8 on the various RNA species are shown on the secondary structure. Large and small dots indicate strong and medium/weak protections. The window of randomization is boxed. The footprint obtained with an RNA fragment of *T. thermophillus* 16S rRNA is also shown (T. th). No other protection was found outside the indicated regions. Nucleotide A642 or its counterpart in variants is circled. Nucleotides 597/641/643 in wt or their variant counterparts are shadowed. Secondary structures are aligned as compared to the conserved nucleotides. "X" represents undefined reactivity. Footprint study was not performed on RNA variants 1.2, 1.3, and 1.8.

common feature presented by all these variants is an EcS8-induced protection at AUC triplets. Surprisingly, some variants (RNAs 2.7, 2.8, 2.9) showed no protection within the randomized window, but showed protection in three distinct areas located in the apical part of the hairpin structure. The appearance of new sites of protection in these variants is most likely related to concomitant mutations (Table 1), which might generate new sites of interaction with EcS8. The case of RNAs 2.7, 2.8, and 2.9, in which all protected nucleotides were localized outside of the randomized window, requires particular attention. In these RNAs, nt 604–608 are deleted and C612 is replaced by U, thus inducing a rearrangement of the secondary structure folding (Fig. 6). However, some analogies with class II binding site could be found, provided that the binding site is inverted with respect to the orientation of the stem (as indicated by the arrow in Fig. 6). As in the case of RNA 2.5, the protection pattern would suggest a combination analogous to A597/U641/U643. A third protection site is also observed, but it exhibits more extended and stronger protections than the corresponding site (588, 589) in the class IIa variants. Further investigations are needed to understand more completely the binding site of these variants and to determine if the homologies noticed with class IIa variants are relevant. RNAs 2.6, 2.10, and 2.11 also display a pattern of protection in and outside the randomized window. Variants 2.10 and 2.11 (which only differ by one base in the randomized window) show AUC and ACC protected triplets (Fig. 4). RNA 2.6 possesses two AUC protected triplets. These data are reminiscent of the features of 2.7, 2.8, and 2.9 variants. However, the presence of a third site of protection is absent in these variants.

Effect of the covariations at 597/641/643 in a wt context

Several concomitant mutations and deletions in class II RNA variants, probably due to errors of Taq DNA polymerase, were observed outside the randomized windows (Table 1). Although these mutations occurred

outside the defined EcS8 binding site, possible side effects of these mutations on binding could not be excluded. Indeed, mutations are probably responsible for the appearance of a new site in 2.7, 2.8, and 2.9 variants. Thus, in order to address this point and to definitively prove that the observed covariations are relevant, we introduced the three putative covariations, U/C/G, C/A/G, and A/U/U, at positions 597/641/643, respectively, in RNAs with wt context. We tested the effect of these mutations on the affinity for EcS8. Results are shown in Figure 7 and summarized in Table 2. The results unambiguously show that RNA mutants containing the combinations U/C/G and C/A/G bind EcS8 with the same K_d as wt RNA (Fig. 5). RNA mutants containing the combination A/U/U bind EcS8 with a slightly increased K_d (by a factor of 3), regardless of the nature of the base pair in position 596–644 (A-U or C-G). Two other combinations for 597/641/643 positions were also tested: U/U/G and U/A/G. These mutants are not able to promote efficient binding of EcS8 (Fig. 7). From our previous work, it was also shown that G/C/C does not allow efficient binding and that G/U/G induces a 10-fold increase of K_d (Mougel et al., 1993). These results demonstrate clearly the interdependency of the three positions 597/641/643.

TABLE 2. Effect of covariations at 597/641/643 on EcS8 binding (K_d)[a]

RNA variants	Sequences (594–599)–(639–645)	K_d (nM)
	597 641 643	
G/U/C (Wt)	UAA G UC GA U AC UG	3
U/C/G	UAA U UC GA C AG UG	3
C/A/G	UAA C UC GA A AG UG	3
U/U/G	UAA U UC GA U AG UG	300
U/A/G	UAA U UC GA A AG UG	300
A/U/U	UAA A UC GA U AU UG	9
A/U/U (C596/G644)	UA C A UC GA U AU GG	9
G/U/C (U640)*	UAA G UC GU U AC UG	300
G/U/C (A598)*	UAA G AC GA U AC UG	300
G/U/C (A598/U640)*	UAA G AC GU U AC UG	15
G/A/C*	UAA G UC GA A AC UG	3
C/U/G*	UAA C UC GA U AG UG	30
G/U/G*	UAA G UC GA U AG UG	>300
G/C/C*	UAA G UC GA C AC UG	>300
G/G/C (A598/U640)*	UAA G AC GU G AC UG	>300

[a]In the upper part of table are covariations reconstructed in wt context or associated with other mutations. In the lower part (indicated by *) are the K_ds of mutants reported from Mougel et al. (1993) and Allmang et al. (1994). Mutations are indicated by white letters on black background in sequences 594–599 and 639–645; positions 597, 641, 643 are indicated in grey. Numbers represent the mean value of a minimum of three independent determinations.

DISCUSSION

Natural and artificial phylogeny

Variant RNAs were selected for their capacity to bind EcS8, using Selex experiments. All the studied variants bound EcS8 with a wt affinity and were able to compete comparably well with wt RNA. A high proportion of wt RNA was isolated throughout the various Selex experiments. Because no difference in K_d was measured between wt RNA and many of the isolated variants (Table 1), these data are best explained by a contamination of the Selex assay by wt molecules rather than by the existence of a pressure for selecting wt species. Our results allowed us to characterize two distinct classes of RNA variants. RNAs from the first class share typical features with the procaryotic RNAs (Fig. 8). The consensus derived from eight different sequences was found consistent with the general consensus derived from the natural phylogeny, and the binding of EcS8 induces an identical hydroxyl radical footprint on these RNAs as on wt RNA. RNAs from the second class display features that are not found in nature. In particular, the two universally conserved nt G597 and C643 are not conserved. These RNAs could be divided into two subclasses. RNAs from class IIa show a covariation between nt 597, 641, and 643, whereas the universally conserved A642 is maintained (Fig. 8). Three different combinations, not observed in nature, U/C/G and C/A/G and A/U/U, were found and could be inserted in the *E. coli* context without significant reduction of affinity. It also appears that nt 598 and 640 obey particular constraints. Otherwise, a bulged nucleotide has to be present either at position 595 or 596. Again, the hydroxyl radical footprint was the one characteristic of the wt situation. The remaining RNAs form subclass IIb. How these RNAs bind EcS8 is more difficult to rationalize, because they show hydroxyl radical footprints differing from the wt, even though some of them might display some class IIa features. It is not yet clear whether these class IIb mutants represent true variants of the wt RNA that recognize the same regions of EcS8, or RNAs that recognize

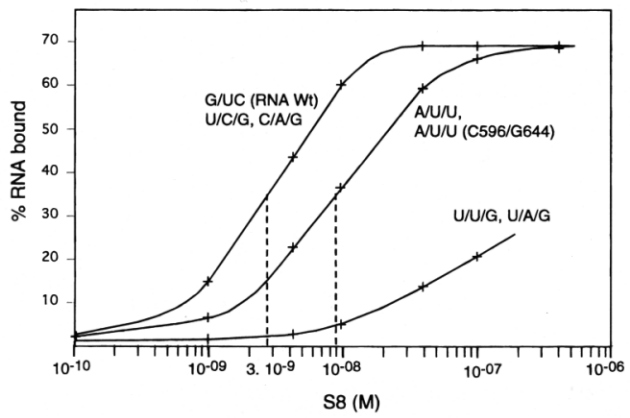

FIGURE 7. Saturation binding curves for interaction of EcS8 with wt RNA or with the RNA containing the different covariations at positions 597/641/643. Effect on binding of mutations at positions 596/644 is also shown. Apparent dissociation constants (K_d) are estimated as the concentration of protein required to give 50% saturation, according to Mougel et al. (1993).

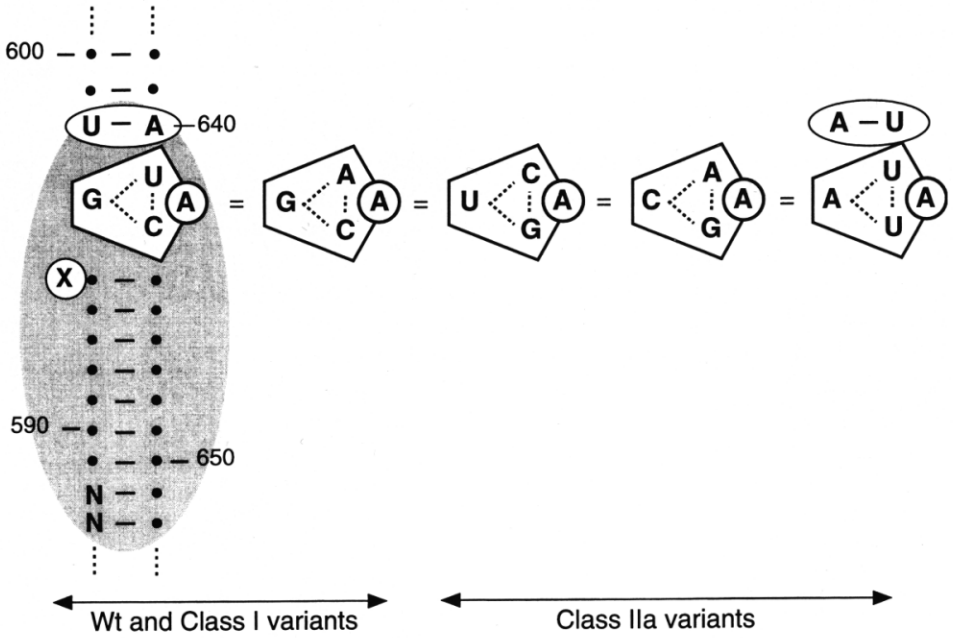

FIGURE 8. Consensus model of the EcS8 binding site. The interdependent nucleotides (597/641/643) are boxed. X is any of four bases making a bulge located at position 595 or 596. N indicates additional minor contacts on backbone (nature of base does not seem to be important). Dots linked by a bar indicates Watson-Crick base pairs not involved directly in the binding site.

the protein in a completely different manner. Because all these RNAs compete with wt RNA, they should recognize similar or overlapping regions on EcS8.

RNAs from class IIa obviously provide the most interesting information. Indeed, they bind EcS8 in a way similar to wt RNA, but the accommodation to the protein is achieved through different nucleotides than those generally used in nature. The most likely explanation is that the substituted nucleotides are able to build a common three-dimensional motif recognized by the protein. It should be remembered that these RNAs were solely selected on their ability to bind EcS8. In nature, circumstantial reasons as well as structural and/or functional constraints may have prevented the variations generated in our Selex experiments (e.g., other proteins or RNA sequences from 16S rRNA might interfere with the folding or binding).

Significance of the triple covariation

The observed covariation between nt 597, 641, and 643 could, in principle, be interpreted in different ways. It may reflect a structural interdependence between these residues, e.g., by forming a triple interaction providing a defined three-dimensional geometry. Such a conformation may be required for direct binding of the protein or for displaying other essential residues in the correct orientation for binding (e.g., A640). Alternatively, only two of them might see each other (e.g., nt 597 and 643), the third one being recognized by the protein (or other nucleotides). In the more extreme case, all three nucleotides may be structurally independent and required for binding by the protein. A similar situation is illustrated by the aptamers selected for arginine binding, in which conserved nucleotides form the binding pocket and three variant nucleotides provide hydrogen bonding for the ligand (Yang et al., 1996).

Nevertheless, a Watson–Crick base pair between nt 598 and 640 is clearly supported by natural covariations (Fig. 3) (Gutell et al., 1994). These nucleotides appear to be sensitive to the context, in particular to the nature of nt 641 (or the other way around) (Fig. 3). Indeed, the inversion of A598 and U640 in the *E. coli* context resulted in a fivefold reduction of affinity (Table 2), whereas this inversion has no effect in the RNA 2.5 context with the A/U/U covariations. A Watson–Crick pairing between U598 and A640 was not supported by the reactivity of A640 at N1 in the naked RNA (Mougel et al., 1987; Allmang et al., 1994). Possibly, this base pairing requires the bound EcS8 to be stabilized and the observed reduction of reactivity observed at A640(N1) in the presence of EcS8 may well account for such a stabilization. On the other hand, our previous and present results provide evidence in favor of an interaction between nt 597 and 643 (G-C, U-G or G-C, and A-U). Probing experiments support the existence of a G597–C643 base pair in wt RNA (Mougel et al., 1987; Allmang et al., 1994). Remarkably, the potential G597–C643 base pair could only be inverted without loss of affinity in the A641 context, and the U597–G643 pair is only allowed in the C641 context (whereas the combination G597/C641/C643 is not allowed). These observations, together with the observed covariations of nt 597, 641, and 643, favor the triple interaction hypothesis.

Three-dimensional folding and EcS8 binding

Because of the lack of stringent structural constraints, the identification of base triples is notoriously difficult

(for a discussion see Gautheret et al., 1995). In the present case, the Selex experiments clearly indicate an interdependence between the three residues at 597/641/643 and, accordingly, the interaction of these three residues in a base triple was tested by molecular modeling. A model for wt RNA accounting for these observations was constructed (Fig. 9). In this model, Watson-Crick base pairs were constructed between U598 and A640 and between A596 and U644. The base triple was constructed with a Watson-Crick base pair between G597 and C643, and U641 was located inside the deep groove of the helix with its O2 position forming hydrogen bonds with positions N4 and C5 of C643 as schematized on Figure 10. This triple can be denoted (G-C)•U. Residue A640 is bulging in the deep groove in an ill-defined conformation. For the other combinations of residues (597–641)•643, suggestions can be also given (Fig. 10). Thus, in the combination (G-C)•A, the most frequent in nature, the N3 of the A falls in a position very close to the N4 and C5 of C, leading to a similar base triple. For the combinations (U-G)•C and (C-G)•A, a possibility would be a contact between the Hoogsteen sites of G and N4(C) or C2(A). H-bonds implying C-H bonds have been observed recently in crystal structures (Wahl et al., 1996), as well as in molecular dynamics simulations of RNAs (Auffinger et al., 1996). In fact, for the last combination (A-U)•U, the third base U could form two H-bonds with the Watson-Crick paired U (N3-H•••O4 and O2•••H-C5), the *cis* equivalent to the *trans* U-U pair that was observed in a recent crystal structure (Wahl et al., 1996).

The slight differences in competition efficiencies observed between certain variants (Table 1; Fig. 4) might reflect the detailed geometries of the various triples. Bases flanking a triple were postulated to be able to play a compensatory role in stabilizing and maintaining orientation of the geometry of various triples (Gautheret et al., 1995). Such context effect of a neighboring base pair on a triple interaction was seen clearly for the RNA variant (2.5) containing the (A-U)•U triple, where A598/U640 above the triple seems to be required to gain affinity, whereas the other triples required the U598/A640 base pair. As a matter of fact, (A-U)•U triple cannot be displayed in exactly the same geometry than the other four combinations (Fig. 10).

Another point that emerges from our study is that EcS8 has two distinct anchoring sites. The major one, centered around the universally conserved A642, is highly specific. A weaker contact was observed on the backbone of the hairpin (ribose 589) located nine residues upstream on the 5′ strand. This latter site appears to be mainly base nonspecific. Due to the essentially helical structure of this RNA, both anchoring sites arise on the same side of the molecule. Thus, EcS8 would cover the width of a shallow and a deep groove spanning a

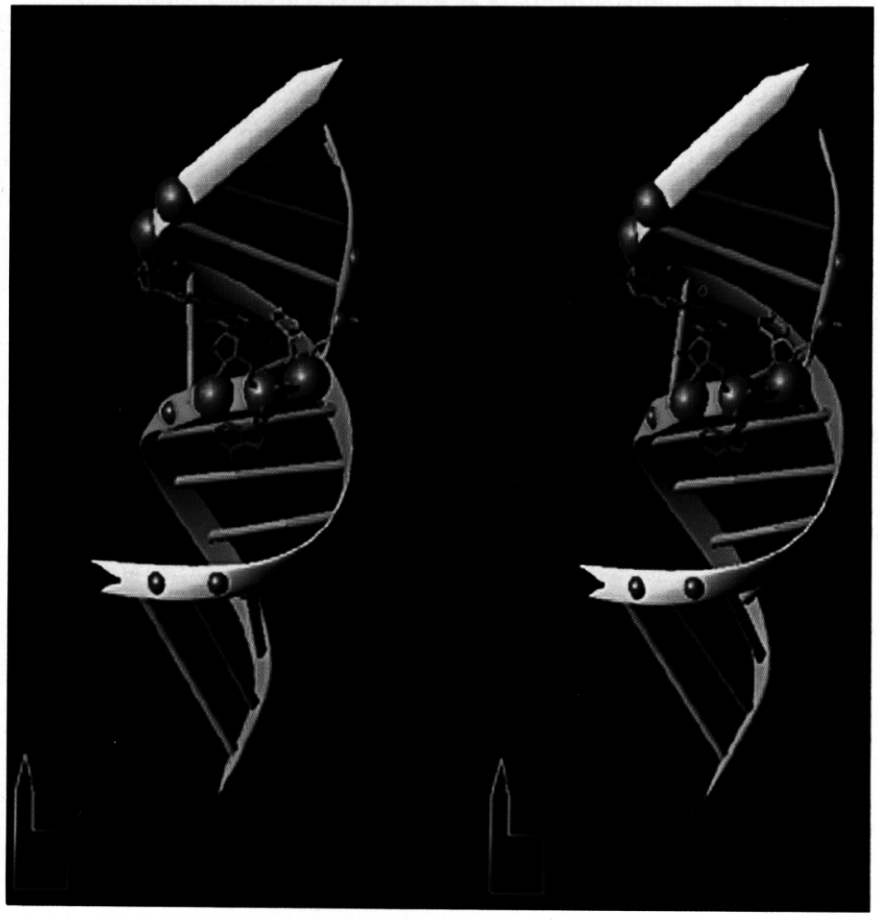

FIGURE 9. Stereo view of the binding site of protein EcS8 on 16S rRNA, encompassing nt 587–600 and 639–652, created by the program DRAWNA (Massire et al., 1994). 5′ and 3′ strands are in yellow and green, respectively. 5′ to 3′ orientation of the strands is indicated by arrow heads. Missing apex of the hairpin is at the top of the figure. The three interdependent nucleotides (597-641)•643 are represented in a triple interaction conformation. On the view, the shallow groove is on the left of the G597–C643 (in purple) and the universal A642 residue (in red) is lying within the deep groove with the third base of the triple U641 (in blue) on the right, also in the deep groove. Protections of the riboses by EcS8 toward hydroxyl radicals are indicated at the level of the corresponding phosphates, with strong and weak protections indicated on the backbone by large and small red spheres, respectively.

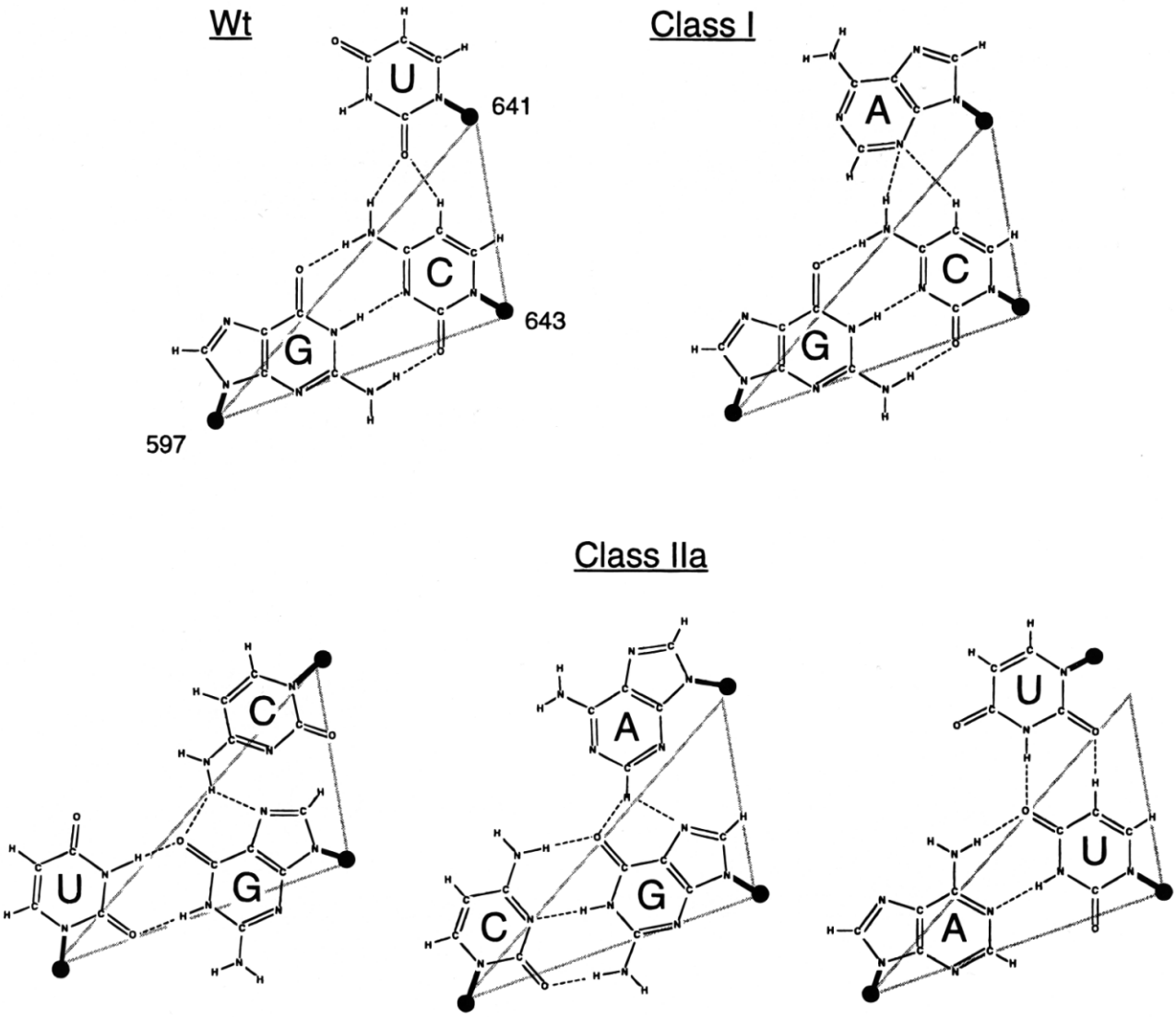

FIGURE 10. Proposed triple contacts between the three interdependent residues (597-643)•641 in the wt (see Fig. 9 for a stereo view), as well as in class I and class IIa variants. Triangles relating the C1′ carbon atoms of the ribose illustrate the geometrical conservation of the suggested interactions. C-H•••O or C-H•••N contacts are proposed on the basis of recent data emphasizing their importance (Auffinger et al., 1996). C-H•••O/N contacts implicate either a C5-H bond of a pyrimidine or a C2-H bond of an adenine, and the carbonyl at position 2 of an uridine. Those positions, in two cases C5 (Y), O2 (Y), and C2 (A), are the most frequently observed ones in C-H•••O contacts (Auffinger & Westhof, 1996). Notice the similar orientations of the glycosyl bonds in the wt and class I variant and, in the class IIa variants, the increased orientations and distances (for (A-U)•U variant).

length of about 30 Å. This model is compatible with the recent crystallographic structure of S8 from *Bacillus stearothermophilus*, where two RNA binding sites located on one side of the protein but at opposite ends were suggested (Davies et al., 1996).

MATERIALS AND METHODS

EcS8

Ribosomal protein EcS8 was prepared under nondenaturing conditions from 30S core particles of MRE600 *E. coli* cells according to Cachia et al. (1991). EcS8 protein was stored at 4 °C in 50 mM NH4-acetate, pH 7.5, 0.28 M NaCl.

Preparation of RNA pool with randomized S8 binding site

Double-stranded DNA template (121 base pairs long) containing two cassettes of 6 and 7 random nucleotides and extending from nt 588 to 651 of 16S rRNA was generated by PCR (Perkin Elmer Cetus) using three oligonucleotides synthesized on an Applied Biosystems apparatus (Fig. 11). Synthesis of dsDNA was performed using 100 pmol each of two amplification primers: primer 3′-(ATGCCGCGGATCCGCG TTTGCT), containing sequence of *Bam*H I restriction site, and primer 5′-(AGACCGGAATTCCTAATACGACTCACTA TAGGGCGTTTGTTTG), containing sequence for T7 RNA polymerase promoter and *Eco*R I restriction site. Oligonucle-

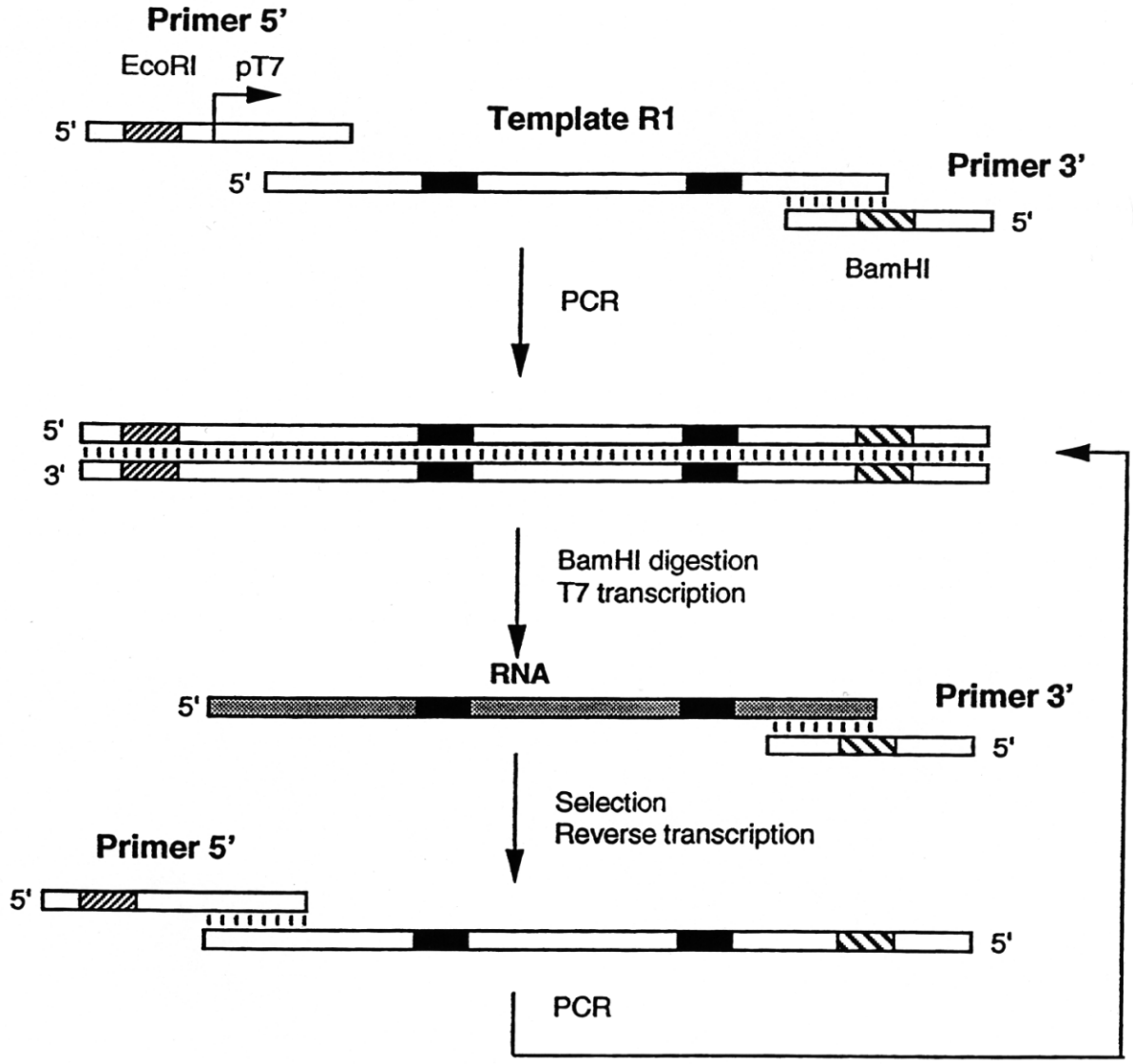

FIGURE 11. Strategy of selection used in this study (see details in text). Areas in black on the template symbolize the random regions.

otide (GGGCGTTTGTTTGTNNNNNNAGATGTGAAATCC CCGGGCTCAACCTGGGAACTGCATCTNNNNNNNG CAAGCAAACGCG, N being an equal ratio of all four bases) was used as template with 5 or 5×10^{-2} pmol amounts, representing, respectively, 44,500- or 445-fold each potential variant. Only 3′ primer could anneal to the template oligonucleotide R1 to produce minus-strand DNA. Completion of the minus strand of PCR product and synthesis of the plus strand was realized after annealing of 5′ primer to the extended 5′ primer (minus strand). The PCR reaction was conducted for 20 cycles with DNA polymerase Gold Star (Eurogentech). dsDNA was phenol/chloroform extracted and ethanol precipitated and washed with 80% ethanol. DNA template was then digested with *Bam*H I restriction enzyme because *Bam*H I sequence was shown to induce artefactual dimerization of RNA when present at the 3′ end of transcripts in conditions of binding of the EcS8 protein. DNA was then purified on a Sephadex G-50 (Pharmacia) minicolumn. The eluate was phenol/chloroform extracted and used directly in the following in vitro transcription. For the first round of selection, 100 pmol of template dsDNA were used in a 250-μL T7 transcription reaction for the preparation of the RNA pool. For the subsequent rounds, 10–100 pmol of template in 50 μL T7 transcription reaction were used. Transcripts were then digested with DNase I. In parallel, 2 pmol of DNA served as template in a T7 transcription with α^{32}P ATP (3,000 Ci/mmol), 20 μCi in 10-μL reaction. All transcripts were purified using an 8% PAGE under denaturing conditions.

Selection

RNAs were renatured in the EcS8 binding buffer, 50 mM Na-cacodylate, pH 7.5, 20 mM Mg-acetate, 250 mM K-acetate, 15 min at 40 °C, and then kept on ice. Selections were done at a protein concentration of 10^{-7} or 10^{-6} M in 50 μL. Complexes between EcS8 and RNA were formed during 10 min at 0 °C. These complexes were separated from free RNA by filtration through a nitrocellulose membrane (Millipore,

0.45 μ, HAWP) presoaked in binding buffer. Unbound RNAs were rinsed off with 5 mL of binding buffer at 0 °C. RNA was recovered from the filter after elution as described by Tuerk and Gold (1990). Controls without EcS8 were performed at each round of selection.

cDNA synthesis and subsequent rounds of selection

RNA recovered from the filter was subjected to reverse transcription using primer 3′ in 50 μL and the cDNA product was amplified with DNA polymerase as described previously (Tuerk & Gold, 1990). The dsDNA amplified product was then transcribed as described above to produce an RNA pool for the next round of selection.

Filter binding assay

K_ds at equilibrium were estimated for each population of RNA and for individual cloned species using α-^{32}P labeled RNA with the filter binding assays as described previously (Mougel et al., 1993). Before each round of selection, a sample of the selected population was tested for its EcS8 binding property. This sample was discarded if the background retention was greater than 10% of input (or 50% of the specific retention).

Competition experiments between RNAs

EcS8 binding on the different RNA species was further investigated by performing competition experiments between RNAs. A fixed negligible concentration of uniformly ^{32}P-labeled T7 transcribed wt RNA (about 10^{-11} M and 80,000 Cerenkov cpm) and various concentrations of cold RNA of the different species were renatured separately in the EcS8 binding buffer. These RNAs were then mixed with EcS8 protein at concentration 3 nM (= K_d) and incubated for 10 min in ice. Filter binding assays were performed as described above.

Structure probing with hydroxyl radicals

Five pmol of RNA were renatured in EcS8 binding buffer 15 min at 40 °C and incubated for 10 min in the presence of 10^{-6} M EcS8 at 0 °C. These complexes were then subjected to hydroxyl radical cleavage with 2 mM $(NH_4)_2Fe(SO_4)_2$, 4 mM EDTA, 1 mM DTE, 0.4% H_2O_2. The reaction mixture was incubated for 10 min at 0 °C and the modification was stopped by ethanol precipitation. The RNA was then phenol-extracted, ethanol precipitated, and washed with 80% ethanol before being vacuum dried. Control experiments without reagent and/or without EcS8 protein were run in parallel. Identification of cleavages was determined using primer extension with reverse transcriptase using primer P3′, as described previously (Mougel et al., 1987).

Cloning and sequencing

cDNAs of selected RNA species were cloned into pUC18 using the PCR-amplified ds-cDNA product. The dsDNA was hydrolyzed with *Eco*R I and *Bam*H I restriction enzymes and was ligated into pUC18 vector. The ligation mix was then used to transform competent DH1 cells (Hanahan, 1983). Sequencing of the minus strand of the individual clone dsDNA was performed on denatured dsDNA prepared from the transformed cells using reverse transcriptase and primer P3′.

Molecular modeling

The three-dimensional model integrating stereochemical and experimental data was constructed using several algorithms (Westhof, 1993) starting from the previously published model (Allmang et al., 1994).

ACKNOWLEDGMENTS

We thank Pascale Romby for comments and helpful discussions, Benoit Masquida for help with comparative sequence analysis, Flore Winter for technical assistance, Alexander A. Serganov for providing RNA from *T. thermophilus*, and J. Steve Lodmell for critical reading of the manuscript. This work was supported by grants from the CNRS.

Received November 18, 1996; returned for revision December 5, 1996; revised manuscript received January 2, 1997

REFERENCES

Allmang C, Mougel M, Westhof E, Ehresmann B, Ehresmann C. 1994. Role of conserved nucleotides in building the 16S rRNA binding site of *E. coli ribosomal protein S8*. *Nucleic Acids Res* 22:3708–3714.

Auffinger P, Louise-May S, Westhof E. 1996. Molecular dynamics simulations of the anticodon hairpin of tRNA^{Asp}: Structuring effects of C-H-O hydrogen bonds and of long-range hydration forces. *J Am Chem Soc 118*:1181–1189.

Auffinger P, Westhof E. 1996. H-bond stability in the tRNA^{Asp} anticodon hairpin: 3 ns of multiple molecular dynamics simulations. *Biophys J 71*:940–954.

Cachia C, Flamion PJ, Schreiber JP. 1991. Fast preparative separation of native core *E. coli* 30S ribosomal proteins. *Biochimie 73*:607–610.

Conrad RC, Giver L, Tian Y, Ellington AD. 1996. In vitro selection of nucleic acid aptamers that bind proteins. *Methods Enzymol 267*:336–367.

Davies C, Ramakrishnan V, White SW. 1996. Structural evidence for specific S8-RNA and S8-protein interactions within the 30S ribosomal subunit: Ribosomal protein S8 from *Bacillus stearothermophilus* at 1.9 Å resolution. *Structure 4*:1093–1104.

Ellington AD, Szostak JW. 1990. In vitro selection of RNA molecules that bind specific ligands. *Nature 346*:818–822.

Gautheret D, Damberger SH, Gutell RR. 1995. Identification of base-triples in RNA using comparative sequence analysis. *J Mol Biol 248*:27–43.

Gregory RJ, Cahill PBF, Thurlow DL, Zimmermann RA. 1988. Interaction of *Escherichia coli* ribosomal protein S8 with its binding sites in ribosomal RNA and messenger RNA. *J Mol Biol 204*:295–307.

Gregory RJ, Zimmermann RA. 1986. Site-directed mutagenesis of the binding site for ribosomal protein S8 within 16S ribosomal RNA from *Escherichia coli*. *Nucleic Acids Res 14*:5761–5776.

Gutell RR, Larsen N, Woese CR. 1994. Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. *Microbiol Rev 58*:10–26.

Hanahan D. 1983. Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol 166*:557–580.

Imlay JA, Chin SM, Linn S. 1988. Toxic DNA damage by hydrogen peroxide through the Fenton reaction in vivo and in vitro. *Science 240*:640–642.

Latham JA, Cech TR. 1989. Defining the inside and outside of a catalytic RNA molecule. *Science 245*:276–282.

Maidak BL, Olsen GJ, Larsen N, Overbeek R, Mccaughey MJ, Woese CR. 1996. The Ribosomal Database Project (RDP). *Nucleic Acids Res 24*:82–85.

Massire C, Gaspin C, Westhof E. 1994. A program for drawing schematic views of nucleic acids. *J Mol Graph 12*:201–206.

Michel F, Ellington AD, Couture S, Szostak JW. 1990. Phylogenetic and genetic evidence for base triple formation in the catalytic domain of group I introns. *Nature 347*:578–580.

Mougel M, Allmang C, Eyermann F, Cachia C, Ehresmann B, Ehresmann C. 1993. Minimal 16S rRNA binding site and role of conserved nucleotides in *Escherichia coli* ribosomal protein S8 recognition. *Eur J Biochem 215*:787–792.

Mougel M, Eyermann F, Westhof E, Romby P, Expert-Bezançon A, Ebel JP, Ehresmann B, Ehresmann C. 1987. Binding of *Escherichia coli* ribosomal protein S8 to 16S rRNA. A model for the interaction and the tertiary structure of the RNA binding site. *J Mol Biol 198*:91–107.

Powers T, Noller HF. 1995. Hydroxyl radical footprinting of ribosomal proteins on 16S rRNA. *RNA 1*:194–209.

Romaniuk PJ. 1989. The role of highly conserved single-stranded nucleotides of *Xenopus* 5S RNA in the binding of transcription factor IIIA. *Biochemistry 28*:1388–1395.

Tuerk G, Gold L. 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science 249*:505–510.

Wahl MC, Rao ST, Sundaralingam M. 1996. The structure of r(UUC GCG) has a 5′-UU-overhang exhibiting Hoogsteen-like trans U.U base pairs. *Nature Struct Biol 3*:24–31.

Westhof E. 1993. Modeling the three-dimensional structure of ribonucleic acids. *J Mol Struct 286*:203–211.

Wu H, Jiang L, Zimmermann RA. 1994. The binding site for ribosomal protein S8 in 16S rRNA and *spc* mRNA from *Escherichia coli*: Minimum structural requirements and the effects of single bulged bases on S8-RNA interaction. *Nucleic Acids Res* 22:1687–1695.

Yang Y, Kochoyan M, Burgstaller P, Westhof E, Famulok M. 1996. Structural basis of ligand discrimination by two related RNA aptamers resolved by NMR spectroscopy. *Science 272*:1343–1347.