Computer simulation of chaperone effects of Archaeal C/D box sRNA binding on rRNA folding

Ruud J. W. Schoemaker and Alexander P. Gultyaev*

Section Theoretical Biology, Leiden Institute of Biology, Leiden University, Kaiserstraat 63, 2311 GP Leiden, The Netherlands

ABSTRACT

Archaeal C/D box small RNAs (sRNAs) are homologues of eukaryotic C/D box small nucleolar RNAs (snoRNAs). Their main function is guiding 2′-*O*-ribose methylation of nucleotides in rRNAs. The methylation requires the pairing of an sRNA antisense element to an rRNA target site with formation of an RNA–RNA duplex. The temporary formation of such a duplex during rRNA maturation is expected to influence rRNA folding in a chaperone-like way, in particular in thermophilic Archaea, where multiple sRNAs with two binding sites are found. Here we investigate possible mechanisms of chaperone function of *Archaeoglobus fulgidus* and *Pyrococcus abyssi* C/D box sRNAs using computer simulations of rRNA secondary structure formation by genetic algorithm. The effects of sRNA binding on rRNA structure are introduced as temporary structural constraints during co-transcriptional folding. Comparisons of the final predictions with simulations without sRNA binding and with phylogenetic structures show that sRNAs with two antisense elements may significantly facilitate the correct formation of long-range interactions in rRNAs, in particular at elevated temperatures. The simulations suggest that the main mechanism of this effect is a transient restriction of folding in rRNA domains where the termini are brought together by binding to double-guide sRNAs.

INTRODUCTION

The formation of secondary and tertiary structures of large RNA molecules is a dynamic process, characterized by multiple alternative structures and various folding pathways [for a review see (1)]. The kinetic capturing of RNA into a long-lived, metastable state may frequently occur because the energy barriers separating the alternative structures are rather high, in particular at the level of secondary structures.

A requirement to avoid kinetic trapping into misfolded structures should favour certain folding pathways that reliably lead to the functional structure. Owing to the presence of alternative pathways in large RNAs, a kinetic partitioning mechanism of RNA folding (2,3) divides a population of molecules into a fraction that rapidly folds directly to the native structure and fraction(s) of slowly folded molecules trapped in the intermediates. At the level of tertiary structure, this kind of trapping is common and well studied in a number of ribozymes; however, in some of them selection has been able to evolve sequences that can preferentially follow the direct pathway (4,5). The requirement for efficient folding pathways in the formation of functional secondary structures also determines selective pressures that suppress potential non-productive paths and alternative structures or favour quickly formed hairpins (6–10). An important factor in paving the optimal folding pathway at the steps of both secondary and tertiary structure formation is co-transcriptional folding that diminishes conformational complexity and may differ significantly from the refolding of the full-length RNA chain [e.g. (11–14)].

The natural RNA sequences can also be adapted so as to have folding pathways leading to functional secondary structures that do not correspond to the global free energy minimum. In some molecules relatively slow kinetics of refolding between alternative structures can even have a regulatory function (15–17). In principle, the refolding time of a RNA structure grows with RNA size and it is estimated that for relatively large RNA domains (>100 nt) biologically significant secondary structures frequently deviate from the lowest free energy state which is never reached (18).

The sensitivity of a RNA structure to the folding process leads to a possibility to influence the formation of functional structures by RNA chaperones, i.e. molecules that modulate folding pathways (19). Many RNAs perform their functions in large RNA–protein complexes (e.g. ribosome or spliceosome) and a number of RNA-binding proteins have been shown to have RNA-chaperone activity [reviewed in (20)]. Interactions

*To whom correspondence should be addressed. Tel: +31 71 5274814; Fax: +31 71 5274900; Email: gultyaev@rulsfb.leidenuniv.nl

of small RNA (sRNAs) molecules with large RNAs may also modulate their folding pathways. In particular, the post-transcriptional modification of rRNAs is known to require binding of multiple small nucleolar RNAs (snoRNAs) in eukaryotes or snoRNA-like sRNAs in archaebacteria. This kind of binding has been suggested to have a chaperone-like function as well (21–23).

Indeed, some snoRNAs have been shown to participate in rRNA maturation by pairing to rRNA precursors (pre-rRNA) and facilitating proper pre-rRNA folding [e.g. (24–26)]. However, chaperone-like properties may be suspected in many other snoRNAs and sRNAs, in particular so-called C/D box RNAs, primarily responsible for guiding the 2′-O-ribose methylation of rRNA nucleotides by C/D box ribonucleoprotein complexes (27,28). This modification requires the formation of a duplex formed by a snoRNA antisense element and a rRNA target site, which is usually longer than 10 bp. The formation of such a duplex is expected to compete with intramolecular rRNA folding.

Furthermore, the existence of C/D box snoRNAs and sRNAs with two antisense elements interacting with the regions located closely in the rRNA secondary structure indicates possible chaperone effects due to constraints in rRNA folding imposed by the binding of a single molecule to the two sites (25,29,30). Such simultaneous base pairing of the two guide regions to a single target molecule was shown for interactions between a number of archaeal double-guide sRNAs and model oligonucleotides (31). Moreover, it was shown that for maximal methylation activity of archaeal dual sRNAs the simultaneous binding to both target sequences and symmetrical juxtaposition of two ribonucleoprotein complexes associated with the conserved boxes are required (31,32). Double-guide sRNAs are abundant in thermophiles and there is a correlation between living temperature of thermophilic archaea and the number of sRNAs they have, indicating to a possibly important role of archaeal sRNAs in assisting rRNAs to cope with increased folding problems at high temperatures (28,29,33). It should be noted that high temperatures are expected not only to decrease the thermodynamic stability of functional structure, but also to diminish the differences between free energies of alternative structures, therefore increasing folding uncertainty (8), so some structural constraints may have a stronger influence on the folding process.

This kind of constraints may be especially important at the early stages of rRNA folding during transcription. Eukaryotic snoRNAs, involved in pre-rRNA processing, were shown to interact with their targets co-transcriptionally (26,34). The binding of the U3 snoRNA to the 5′ end of growing nascent pre-rRNA transcripts can be even visualized in electron micrographs in so-called terminal knobs, corresponding to the SSU processome, which are not formed in the absence of the U3 snoRNA (34). Apparently, the C/D box nucleotide-modifying snoRNAs also bind pre-rRNA at the early stages of its synthesis, and ribose methylations in rRNA occur very quickly after the co-transcriptional cleavage of the 3′ external transcribed spacer and before the complete processing of the primary transcript [for a review see (35,36)]. Archaeal guide RNAs function similarly to their eukaryotic analogues and were shown to modify rRNA in the eukaryotic nucleus (37).

To examine the possible chaperone-like role of C/D box sRNAs in archaea, we performed computer simulations of rRNA-folding pathways in the presence of these molecules. The calculations were done using the genetic algorithm for RNA folding (6), able to predict biologically important RNA-folding pathways (16,38,39). The effect of transient binding of a given sRNA on rRNA folding was approximated by creating temporary topological restrictions on the base pairing in the rRNA region involved in the interaction. The restrictions included the prohibition of intramolecular pairing of the rRNA sites paired to the sRNA antisense elements and forcing topologies with closely located ends of rRNA regions between two sites bound to a single sRNA. These constraints were imposed in the folding simulation for the growing transcribed rRNA chain and were removed in the subsequent full-length refolding simulation. Such an implementation attempts to mimic the co-transcriptional functioning of snoRNAs. The comparison of the final predicted rRNA secondary structure with the one computed in the absence of sRNAs and with the phylogenetically proven rRNA structure allows one to identify a chaperone effect, which should be reflected in a better prediction in the sRNA presence.

MATERIALS AND METHODS

Genetic algorithm

The details of RNA-folding simulations using a genetic algorithm were described previously (6). In summary, at every iteration the algorithm generates a population of alternative RNA structures for an intermediate length of the transcript. In the course of one iteration, new structures are generated by randomized disrupting and adding of some stems in the previous folding of each alternative. The new population is produced by selecting the most stable structures. Furthermore, the length of the RNA chain is gradually increased to simulate the folding of a synthesized transcript. At every step the program displays the most stable folding in the population found so far, which represents the simulated pathway.

The algorithm is implemented in the package STAR for RNA structure predictions (6,40). The thermodynamic parameters used for the RNA secondary structure elements were taken from the version 2.3 set of Turner and co-workers (41) (http://www.bioinfo.rpi.edu/~zukerm/rna/energy/). The calculations were performed in the temperature range of 37–90°C. Reliable predictions at higher temperatures were not possible, owing to secondary structure melting. With the available thermodynamic parameters, extrapolations to higher temperatures are not correct and even the lowest free energy states of hyper-thermophilic rRNA are essentially single-stranded when computed at the optimum growth temperature (8).

Unless otherwise stated, simulations were performed with populations of 10 structures in the genetic algorithm. For the predictions of the final structures of full-length RNAs, the simulations were continued until population convergence (i.e. all structures become equivalent).

Implementation of s(no)RNA binding in the folding simulations

The association and dissociation of s(no)RNA molecules to and from the rRNA were simulated by a series of program
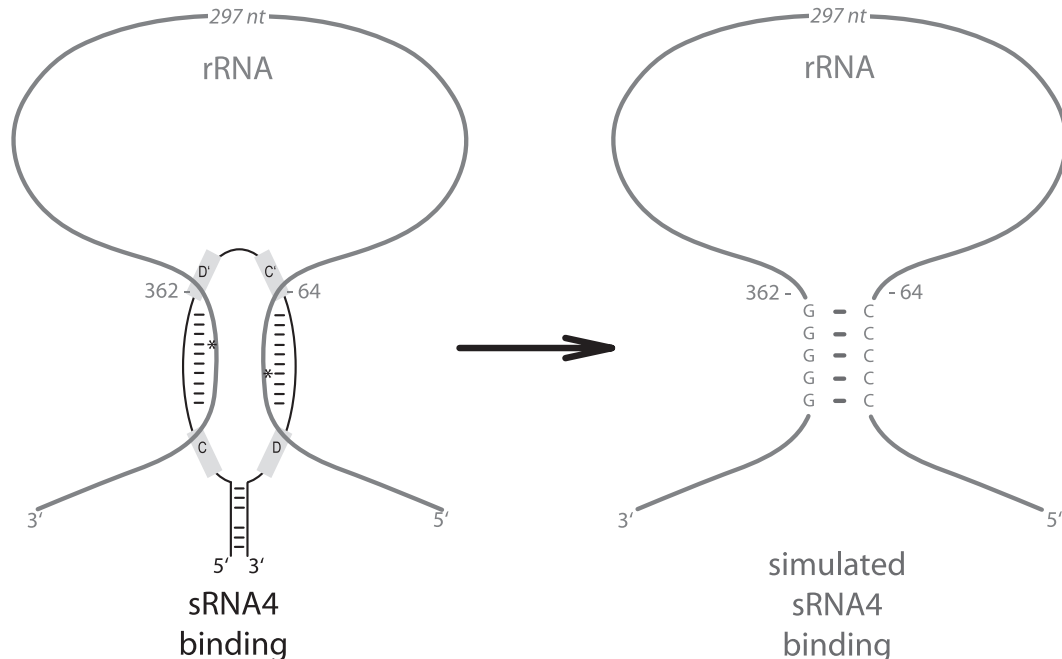
Figure 1. Simulation of the binding of double-guide sRNA to its two target rRNA regions, exemplified for the *P.abyssi* sRNA4 binding. The artificial GC-rich stem is forced in the constrained domain folding simulation, mimicking close proximity of the domain ends. Methylated nucleotides are indicated by asterisks.

runs, where the structures yielded at a particular simulation (e.g. with a snoRNA bound) were transferred to the next program run (e.g. without the binding). The s(no)RNA binding to rRNA was simulated in an rRNA-folding pathway by prohibiting the pairing of the nucleotides in the binding site region to any other rRNA region. The transfer of structures from one program run to another was done by 'forcing' the structure yielded by the first run into the following one, so that the latter simulation started from the structure folded by the previous step. Depending on the specific step of folding simulation, two options of such a forcing were used: 'strong' forcing (the forced structure was not allowed to be disrupted) and 'weak' forcing (the forced structure was allowed to be disrupted, once a conformation with a lower free energy had been found).

The effect of a single snoRNA-binding site was implemented by simulating the folding by two program runs, as follows:

(i) In the first run transcription is simulated by growing the rRNA in the 5′–3′ direction, while keeping the complementarity region of the snoRNA occupied by the snoRNA (the complementarity region is therefore prohibited to pair).

(ii) The complete structure from the first part is weakly forced into a second run, which simulates the non-growing chain being refolded without the snoRNA. This part is continued until the 10 populations of the algorithm have converged.

Simulations of double-site binding

An effect of co-transcriptional binding of a snoRNA molecule with two complementary regions to rRNA, with subsequent snoRNA release, was simulated as follows.

(i) Transcription is simulated by growing the folded rRNA up to the 3′ end of the second complementary region, while keeping both complementarity regions bound to the snoRNA.

(ii) The refolding of the domain between the two complementary regions is simulated separately with forcing a configuration that brings the ends of domain together. For instance, the binding of a single *Pyrococcus abyssi* sRNA4 molecule to its two target regions 54–63 and 363–372 of the SSU rRNA (Figure 1) is assumed to constrain the folding of domain 64–362, favouring a topology with closely located domain ends. First of all, the domain structure yielded by the previous simulation is weakly forced. The connecting of the ends is mimicked by adding artificial sequences to the domain ends: five guanines upstream of the domain and five cytosines downstream of it (Figure 1). The pairing of these sequences is strongly forced. Although this imposes some constraint on the domain folding, such an implementation of sRNA double-site binding does not prohibit any base pairing inside the domain. The simulation is continued until the 10 populations of the algorithm have converged.

(iii) After the refolding of the constrained domain, the simulation of rRNA folding continues with the chain growing until the very 3′ end, while the snoRNA presence is still assumed. Therefore, the structure of the domain yielded by the step (ii) is strongly forced and the complementarity regions are kept single-stranded. Of course, artificial terminal sequences [step (ii)] are removed.

(iv) After the rRNA chain has been completely folded, refolding after the release of the snoRNA is simulated by weakly forcing the whole structure of step (iii), while no part is left occupied by the snoRNA. The simulation

is continued until all 10 structures in the population have converged. Thus all strong constraints implemented at the intermediate steps are removed at this last refolding step.

Such an implementation of the binding of a double-guide sRNA to its two target sites in rRNA assumes that the sRNA–rRNA complex exists only during rRNA transcription, thus simulating a transient, chaperone-like character of the accompanying structural constraints. In principle, a more realistic model should take into account the dynamic competition between intramolecular and intermolecular pairings that would lead to shorter or longer lifetimes of this complex. However, such a model would require the incorporation of currently unknown thermodynamic parameters of sRNA association and dissociation and RNA concentrations.

No additional assumptions were made on the pairing of methylated nucleotides, because they are known to occur in both double- and single-stranded rRNA regions (35). Thus during existence of sRNA–rRNA complex [steps (i–iii)] they were assumed to be engaged in RNA–RNA duplexes (Figure 1), while no restrictions were imposed on these nucleotides at the last sRNA-free step (iv). This is consistent with the situation *in vivo*.

Consensus structures

Every simulation of genetic algorithm may be viewed as one trajectory along some folding pathway and repeated calculations may follow other pathways. The relative frequencies of prediction of particular structures in repeated simulations can be used for rough estimates of the probabilities of the formation of the structures (39). We have used this feature of the algorithm to derive some consensus structure predictions. For each particular case, the whole procedure of rRNA folding was repeated three times and the most frequently folded stems (present in at least two of the three structures) were strongly forced in one final folding simulation for the full-length rRNA. The structure gained from that simulation was considered to be the final structure. The three independent rRNA-folding simulations were also used to calculate the standard deviations of the numbers of correctly predicted stems and base pairs.

Other methods of producing consensus structures were explored such as selecting stems present in three out of three structures, in three out of five, in five out of five, in five out of nine or in nine out of nine. These simulations did not improve results significantly, while some of them consumed much more time. Another approach tested was to determine the most diverged three structures out of five and to use these to make a consensus structure by collecting their matching stems in the final simulation. This also had a minimal effect and did not produce better results.

Energy minimization

The predictions of rRNA structure using free energy minimization were performed using the Mfold server (42) with the version 2.3 of the thermodynamic parameters, these parameters are the same as used with the genetic algorithm (see above). In order to compare statistics of mfold predictions with that of folding simulations, the first three (sub)optimal structures yielded by Mfold were used to calculate the SDs.

Sequence data

The comparative analysis structures of the 16S RNAs were taken from the European Ribosomal rRNA Database (43) (http://www.psb.ugent.be/rRNA/index.html).

The following sequences were used: *Saccharomyces cerevisiae* (accession no. V01335), *Archaeoglobus fulgidus* (accession no. X05567), *P.abyssi* (accession no. AJ248283).

The sequences of the archaeal sRNAs with corresponding complementarity regions on the rRNA were taken from The Methylation Guide snoRNA Database (33) (http://lowelab.ucsc.edu/snoRNAdb/), data on yeast snoRNAs from The Yeast SnoRNA Database (44) (http://www.bio.umass.edu/biochem/rna-sequence/Yeast_snoRNA_Database/main.html).

RESULTS

The simulations of the transient prohibition of secondary structure formation in sRNA-paired regions of Archaeal 16S RNAs do not show significant effects of sRNA binding on the final rRNA structure

A competition between the mutually exclusive pairings of sRNA target regions in rRNA either to sRNA antisense elements or other regions in the rRNA sequence may be viewed as the most straightforward mechanism of sRNA influence on the rRNA-folding pathway. It may be expected that sRNA binding during rRNA transcription would favour base pairing excluding nucleotides in the sRNA-bound regions. If the energy barrier to disrupt these intermediate structures would be relatively high, they would not refold, even after sRNA release. Such a mechanism was implemented in our folding simulations by prohibiting the intramolecular pairing of sRNA targets during the growing rRNA chain folding, while the subsequent refolding simulation was done without these restrictions (Materials and Methods).

The simulations for *A.fulgidus* and *P.abyssi* 16S rRNA folding in the presence of various sRNAs did not show significant improvement as compared to the predictions done without sRNA binding. No effect was observed in either simulations for binding of only one type of sRNA or using the entire pool of all known sRNAs for a given species. For instance, implementing the binding of the only known 16S-rRNA-binding sRNA of *A.fulgidus*, sRNA2, into simulations (by restricting 24 nt in the two target regions to fold during transcription) did not change the number of correctly predicted stems after sRNA release [47 correctly predicted stems at 37°C and 41 at 65°C, the latest being the optimal growth temperature (28)], while the number of correctly predicted base pairs deviated only slightly (254 versus 256 at 37°C and 235 versus 242 at 65°C with and without sRNA, respectively).

Usually, incorporation of this sRNA-binding model resulted in a small decline in the quality of final predictions at lower temperatures, while at higher temperatures sometimes nonsignificant improvements were observed. Thus, simultaneous implementation of all 25 known *P.abyssi* sRNAs (total 430 nt in the targets) into the simulation at 37°C resulted in 45 correct stems containing 249 bp, while the simulation for the bare 16S rRNA yielded 49 stems with 288 bp. At 80°C, in both cases 38 correct stems were predicted, with 242 and 237 correct base pairs in simulations with sRNA effect and without, respectively.

In general, the temperature increase lowered the number of predicted stems due to the melting of secondary structure with available thermodynamic parameters. Similar to the rRNA secondary structure predictions by energy minimization (8), the structures produced at very high temperatures (>90°C) contained very few double-stranded regions. At these temperatures, the predictions by the genetic algorithm did not essentially deviate from the lowest energy states predicted by the Mfold program.

The simulations of transient constrained folding of Archaeal 16S RNA domains between two binding sites of a single sRNA molecule suggest that some sRNAs with two antisense elements may guide rRNA folding at high temperatures

The implementation of constrained folding of domains located between the binding sites of sRNA molecules with two antisense elements (Materials and Methods) turned out to have some effect on rRNA-folding pathways. The effects varied for different sRNAs and temperatures. For instance, the *A.fulgidus* sRNA2 binding (positions 11–22 and 845–856 in 16S rRNA) improves the predictions for 16S rRNA folding in the temperature range of about 45–75°C, while at lower or higher temperatures the predictions do not deviate significantly from those yielded by bare *A.fulgidus* 16S rRNA simulations (Figure 2A). The *P.abyssi* sRNA 4 seems to guide 16S rRNA

Figure 2. Temperature dependence of prediction quality in 16S rRNA folding simulations with domains constrained by double-site sRNA binding. (**A**) *A.fulgidus* sRNA2; (**B**) *P.abyssi* sRNA4. Closed squares—simulations of corresponding rRNAs without sRNA binding, open diamonds—simulations with implemented sRNA binding. Error bars correspond to the standard deviations calculated from the three repeated simulations (Materials and Methods).

folding at temperatures of about 75–80°C, but at other temperatures the results fluctuate for simulations both with and without sRNA4, with positive or negative differences (Figure 2B). Similar results were obtained for some other *P.abyssi* sRNAs (data not shown): at high temperatures the simulations of rRNA folding were better in the presence of sRNAs than in their absence. This is consistent with the expectation that chaperone effects of sRNAs are more important at high temperatures (28,29,33).

The *Pyrococcus* species, a hyper-thermophile, has a very high number of sRNAs (30). The snoRNA database (http://lowelab.ucsc.edu/snoRNAdb/) contains 25 *P.abyssi* sRNAs targeting the 16S rRNA, with 15 of them being double-guide molecules with antisense elements associated with both the D and D′ box motifs. Two of them (sRNA8 and sRNA 19) each have three binding sites at the 16S rRNA with two additional possible configurations of double-site binding. We simulated the potential sRNA-binding effects for all 17 variants. The locations of corresponding double-guide sRNA-binding sites in 16S rRNA structure are shown in Figure 3. As the most pronounced chaperone-like effect of several *P.abyssi* sRNAs on rRNA folding was observed at 80°C, a temperature which is close to the optimal growth conditions of *P.abyssi*, we performed simulations at this temperature.

As seen in Table 1, the majority of the double-guide sRNA molecules seems to direct 16S RNA folding towards the phylogenetic structure. The simulations suggest that this chaperone-like effect is determined by possible constraints in the domain between two binding sites rather than by restrictions on the pairing of nucleotides bound to sRNA antisense elements, because simulations with the same sRNAs assuming binding of two separate molecules, as described in previous section, did not produce any significant effect (data not shown).

The attempts to combine various *P.abyssi* sRNAs did not produce any essential improvements (data not shown). This is probably due to the additional approximations that have to be made to implement the effects of the sRNA combinations to the procedure. For instance, the folding protocol used here, allows one to implement only the sequential effects with overlapping constrained domains, while in nature these molecules could exert their influence simultaneously. Also, many of the tested sRNAs seem to improve predictions in the same regions of the molecule (Discussion).

Using the simulations with the constrained domain procedure, we have also tested possible effects of double-site C/D snoRNA binding in yeast. In yeast, there are only two such molecules with both antisense elements binding to the 18S rRNA: U14 and snR41. The U14-binding sites (positions 83–95 and 410–423 of *S.cerevisiae* 18S rRNA) are topologically close to those of *P.abyssi* sRNA4 (54–63 and 363–372), the one showing the strongest effect on *P.abyssi* rRNA-folding simulations (Table 1). It has been previously suggested that U14 has a chaperone role due to the presence of two complementarity regions, only one of which being important for rRNA modification (25,45). The simulations of the yeast 18S rRNA folding in the presence of U14 turned out to be variable (data not shown). While some of them were better than those produced for bare rRNA, others did not show any improvement. Even smaller effects were noticed for snR41 binding.
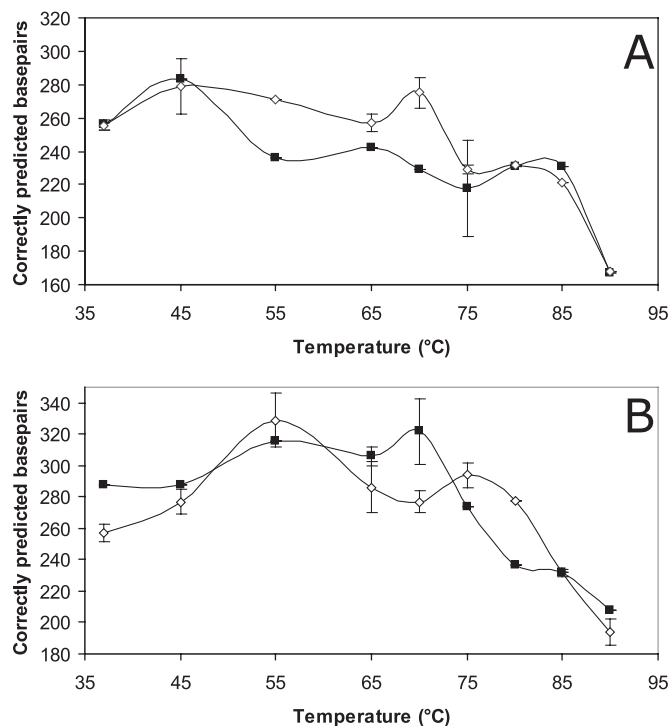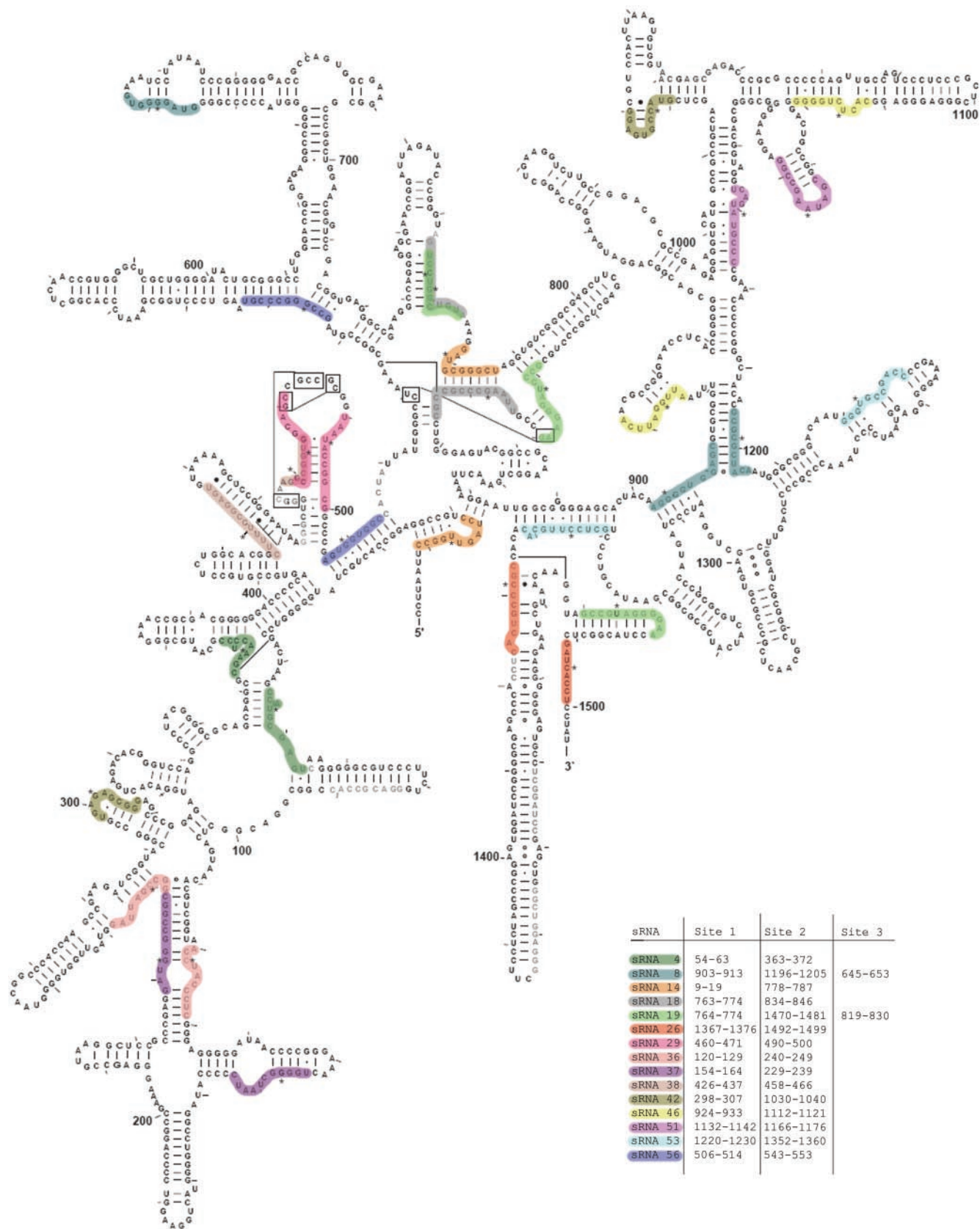
Figure 3. The location of binding sites of *P.abyssi* double-guide sRNAs on 16S rRNA. The data on binding sites are taken from The Methylation Guide snoRNA Database (http://lowelab.ucsc.edu/snoRNAdb/), the 16S rRNA structure is from (29). Binding sites of a given sRNA are indicated in the same colour, methylated nucleotides are shown by asterisks.

sRNA	Site 1	Site 2	Site 3
sRNA 4	54–63	363–372	
sRNA 8	903–913	1196–1205	645–653
sRNA 14	9–19	778–787	
sRNA 18	763–774	834–846	
sRNA 19	764–774	1470–1481	819–830
sRNA 26	1367–1376	1492–1499	
sRNA 29	460–471	490–500	
sRNA 36	120–129	240–249	
sRNA 37	154–164	229–239	
sRNA 38	426–437	458–466	
sRNA 42	298–307	1030–1040	
sRNA 46	924–933	1112–1121	
sRNA 51	1132–1142	1166–1176	
sRNA 53	1220–1230	1352–1360	
sRNA 56	506–514	543–553	

Table 1. The number of correctly predicted stems and base pairs in *P.abyssi* 16S RNA upon binding of double-guide sRNAs at 80°C

sRNA	Binding sites	Stems	Base pairs
None (16S)		38	237
sRNA4	54–63/363–372	44	278
sRNA8	903–913/1196–1205	42	263
sRNA8	645–653/1196–1205	36 ± 4	226 ± 25
sRNA14	9–19/778–787	42 ± 1	263 ± 4
sRNA18	763–774/834–846	43 ± 1	280 ± 3
sRNA19	764–774/819–830	45 ± 3	280 ± 19
sRNA19	764–774/1470–1481	42	263
sRNA26	1367–1376/1492–1499	38	239
sRNA29	460–471/490–500	44 ± 1	280 ± 2
sRNA36	120–129/240–249	41 ± 2	256 ± 13
sRNA37	154–164/229–239	42 ± 3	264 ± 22
sRNA38	426–437/458–466	41 ± 2	255 ± 9
sRNA42	298–307/1030–1040	38	236
sRNA46	924–933/1112–1121	38	238
sRNA51	1132–1142/1166–1176	42 ± 1	265 ± 5
sRNA53	1220–1230/1352–1360	36	230
sRNA56	506–514/543–553	41	256 ± 1

A comparison of rRNA structure predictions made by folding simulations and energy minimization at different temperatures

It is remarkable that the implementation of transient sRNA binding in RNA-folding simulations by a genetic algorithm improves the predictions of rRNA native structures, mostly at elevated temperatures, at which free energy barriers for refolding are lowered. This suggests that the kinetic trapping in metastable structures may even be important at rather high temperatures. On the other hand, the temperature is expected to increase the uncertainty of secondary structure formation, characterized by the base pairing probability distribution and to lower the equilibrium probability of the lower free energy conformation (8). In this respect, to distinguish between kinetic and equilibrium factors in rRNA folding, it is interesting to compare the quality of rRNA comparative structure predictions at different temperatures produced by the kinetic simulations using standard genetic algorithm simulations without sRNA binding (6) with that of equilibrium structures yielded by free energy minimization (42).

The predictions yielded by the two approaches for *A.fulgidus* and *P.abyssi* 16S rRNAs were compared (Figure 4). In general, the behaviour of the two programs is comparable: at some conditions the energy minimization yields better predictions, while folding simulations are better in other cases. Both approaches exhibit an obvious decline in the quality at high temperatures due to secondary structure melting. However, this decline seems to be less drastic in genetic algorithm simulations: at 80–85°C folding simulations result in slightly better predictions for both 16S rRNAs, as compared with the lowest energy structures (Figure 4), indicating that the native structure may indeed be kinetically favoured. Apparently, at some temperatures such a kinetic preference for the native structure formation can be further enhanced by sRNA binding during the folding process. Of course, at extremely high temperatures this preference disappears because then both folding simulations and energy minimization predict very similar structures, that are mostly single-stranded.
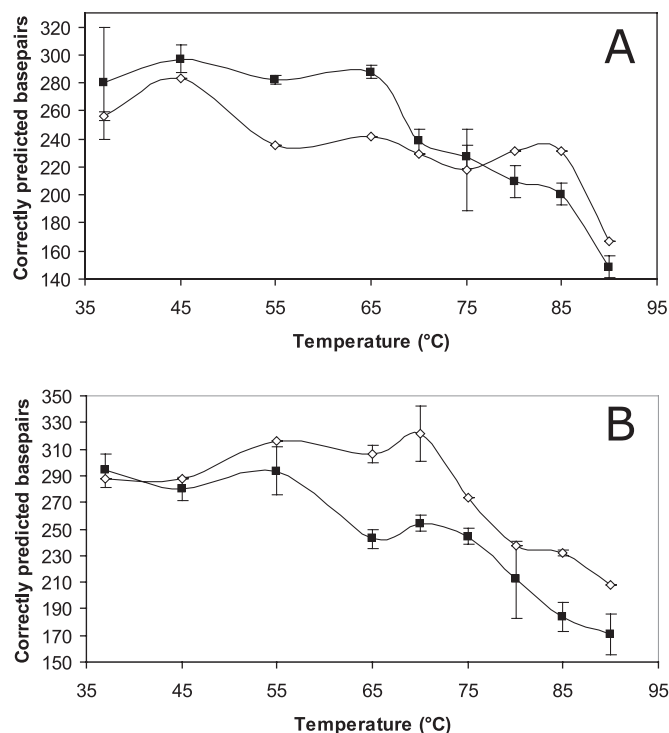

Figure 4. Comparison of the genetic algorithm (open diamonds) and MFOLD (closed squares) predictions for *A.fulgidus* (**A**) and *P.abyssi* (**B**) 16S rRNAs. Error bars in genetic algorithm data correspond to the SDs calculated from three repeated simulations, those for Mfold are computed from the three best (sub)optimal structures (Materials and Methods).

Archaeal sRNAs with two antisense elements mostly assist in long-range secondary structure formation

The comparison of the 16S rRNA structure predictions, yielded by the simulations in the presence of sRNAs, shows that the effects of all sRNAs are mostly located in the central regions of 16S rRNA. Figure 5 shows that the implementation of *P.abyssi* sRNAs into folding simulations at 80°C leads to the correct prediction of many long-range pairings that are not predicted without the sRNA effects. For instance, it is remarkable that the prediction of the bare *P.abyssi* 16S rRNA has no correct stems with a distance between two halves of a stem >100 nt (the largest is 71 nt), while the introduction of sRNA4 into the simulation leads to correct prediction of five long-range stems with >200 nt between the complementary stem parts. Implementation of the only *A.fulgidus* sRNA2 also improves predictions in the 16S rRNA core at 70°C (data not shown). This suggests that the efficient formation of the central parts of archaeal 16S rRNA secondary structure at high temperatures may be improved by assistance of sRNAs.

It should be noted that the relatively poor predictions for long-range interactions in 16S rRNA, obtained in the absence of sRNA binding, are not a specific feature of the genetic algorithm. The results of energy minimization at high temperatures are not better (Figure 4) and a systematic study of the minimum free energy structures of 16S rRNAs (46) shows that the majority of long-range stems (distance >100 nt) are not predicted.

A comparison of the simulated folding pathways shows that the binding of some sRNAs guide co-transcriptional formation
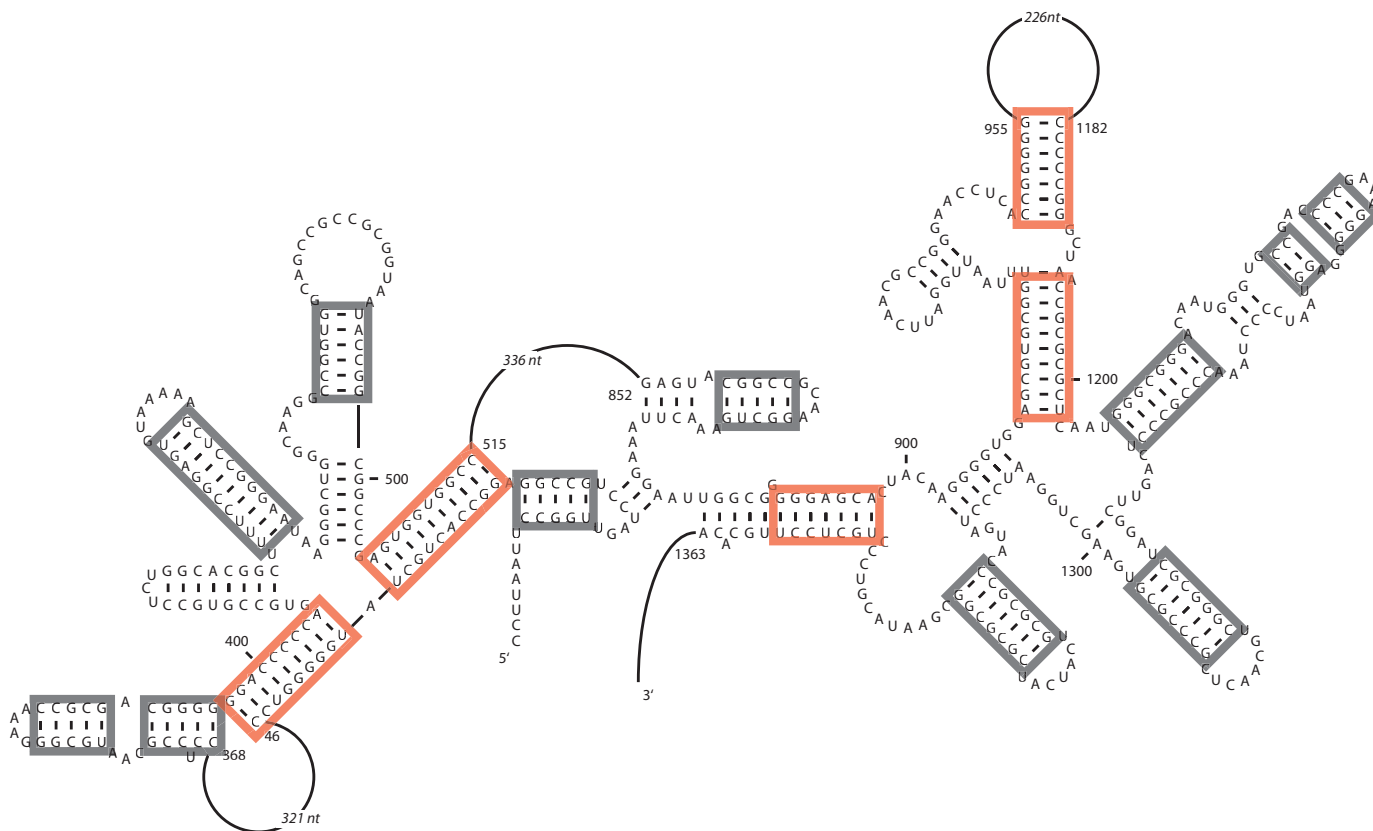
Figure 5. Effect of sRNA binding implementation on predictions of long-range interactions in the *P.abyssi* 16S rRNA structure. Black rectangles indicate stems that are predicted both with and without sRNA binding, red rectangles—stems that are only predicted with sRNA binding.

of important long-range interactions, which are not disrupted after sRNA release. As an example, the simulated effect of *P.abyssi* sRNA4 binding on the folding of 16S rRNA 5′-domain is shown in Figure 6. In the absence of sRNA binding, the domain structure is not predicted correctly and in the final prediction the interior of the domain is paired to the sequences in the 3′ major domain (interactions 38...294/ 1027...1187, Figure 6A). sRNA4 binding to the complementarity regions 54–63 and 363–372 during transcription guides the formation of stems 38–46/397–405 and 27–36/506–515 that close the domain in the nascent rRNA transcript (Figure 6B). In the subsequent steps of simulated pathway upon rRNA elongation, these stems and the bound sRNA4 prevent incorrect long-range pairing, thereby also favouring the correct 3′-domain formation (Figure 5). Finally, the simulated release of sRNA4 only leads to a minor rearrangement of 16S secondary structure with the formation of stems incompatible with transient sRNA interaction, such as the stem 368–373/392–397 (Figure 6B, inset). At this step, the barrier for the disruption of the stems closing the domain (38–46/ 397–405 and 27–36/506–515) is apparently too high and these long-range interactions are present in the final prediction.

DISCUSSION

The presented simulations of archaeal 16S rRNA folding with transient sRNA binding suggest that archaeal C/D box sRNAs

with two antisense elements assist in rRNA folding at high temperatures. This is seen in a comparison of the quality of 16S rRNA predictions, yielded by simulations in the presence of various sRNAs, to those produced without sRNAs or by free energy minimization. The quality of all predictions, estimated by comparison to well-proven comparative structures, drops with temperature due to melting of the secondary structure. However, such a decline is less when the binding of certain sRNAs is introduced to simulations at the stage of folding the growing transcribed rRNA polynucleotide chain. Thus, the simulations support the idea that archaeal double-site sRNAs in thermophilic organisms are RNA chaperones helping to solve RNA-folding problems at high temperatures (28,29).

Furthermore, the simulations show the most significant influence on RNA folding upon binding of sRNAs with two antisense elements, which are especially abundant in hyper-thermophiles. The modelling of a simple delay of intra-molecular pairing in the sRNA-bound rRNA regions did not produce significant effects. In contrast, the simulations of transiently constrained folding in the domains, flanked by rRNA regions bound to a single sRNA molecule, did yield folding pathways leading to predictions that are more similar to rRNA comparative structures.

Our implementation of sRNA binding in the folding algorithm envisages that an sRNA molecule brings together the ends of an rRNA domain between binding sites, guides the folding of this domain to a locally (meta)stable structure and
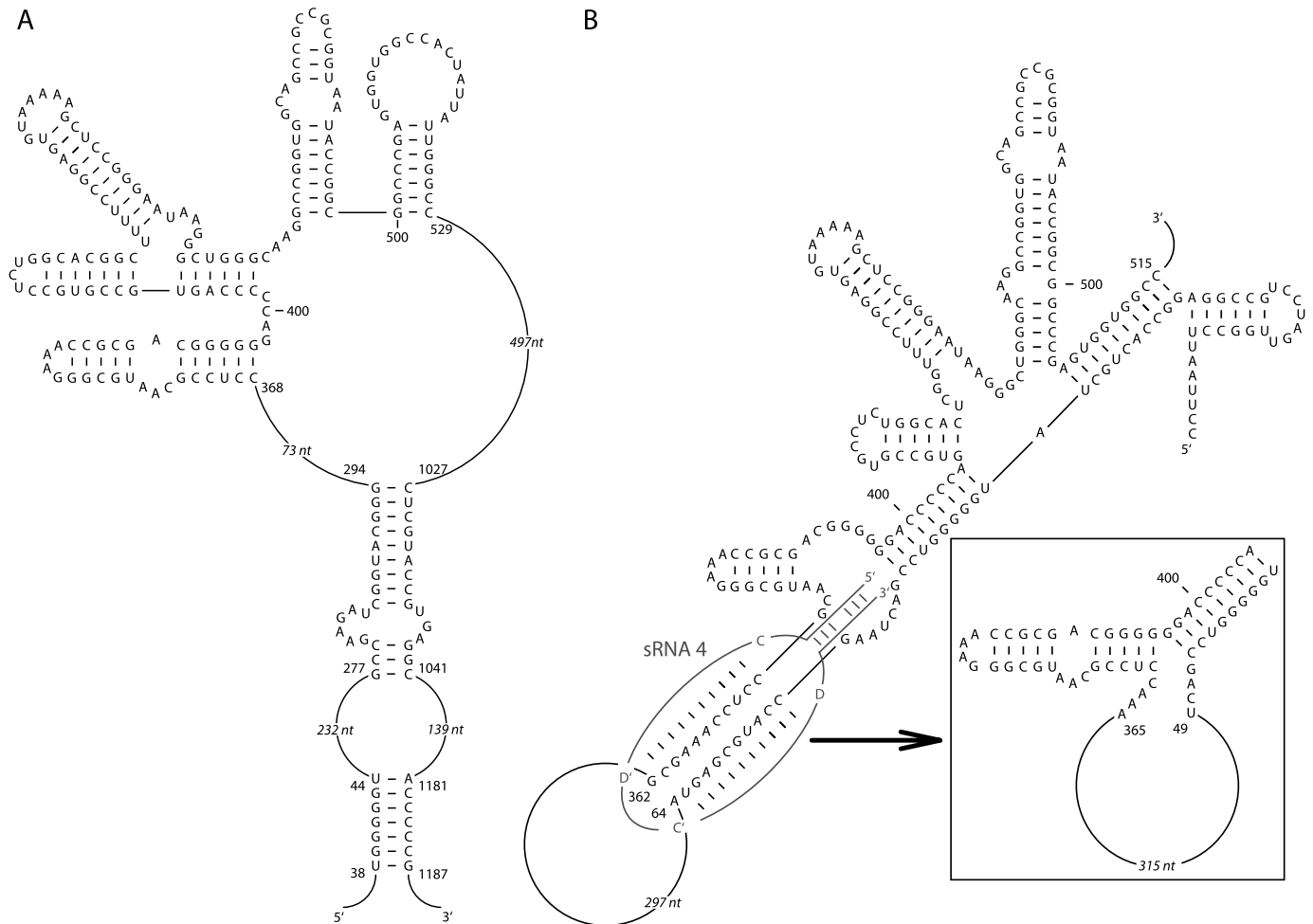
Figure 6. Comparison of folding predictions for *P.abyssi* 16S rRNA 5′-domain in the absence (**A**) and presence (**B**) of sRNA4. The inset shows the refolding of the structure after sRNA4 release.

restricts its interactions with other regions of the rRNA. Although the model assumes that these restrictions only exist during transcription and disappear upon the sRNA release, they seem to be sufficient to lead to metastable structures that are not completely refolded in the full-length rRNA. The effect of such co-transcriptional sRNA binding is mostly seen in an efficient formation of long-range inter-actions in the rRNA, which are otherwise not formed, owing to either kinetic or thermodynamic reasons, as they are not predicted by folding simulations in the absence of sRNAs or by energy minimization.

Such a picture of sRNA-assisted rRNA folding is consistent with a number of facts. It is known that eukaryotic snoRNAs co-transcriptionally bind to the rRNA precursors (22,26,34) and it is reasonable to suspect that this is true for archaeal sRNAs. Transient secondary structures, formed during pre-rRNA transcription and influenced by interactions with spacer sequences and snoRNAs, are important for functional rRNA structure formation [e.g. (24,47–50)]. The conformational rearrangements, occurring co-transcriptionally, are very sensitive to the conditions: accelerated transcription of bacterial 16S rRNA was shown to result in misfolded molecules (12). Slow transcription and transcriptional pausing can increase

folding efficiency of other ribozymes as well (51–53). From a mechanistic point of view, the co-transcriptional constraints implemented in our sRNA-binding model are somewhat similar to the effects of slower transcription, because in both cases interactions involving the 3′-proximal parts of folded RNA are delayed while upstream domains are allowed to fold. For instance, both computer simulations and experiments on co-transcriptional folding of the delta ribozyme (52,53) suggest that the presence of a downstream attenuator sequence determines a limited time-window, during which the ribozyme core can fold correctly. In our implementation the constraints imposed on the domains between the sRNA-binding sites also prevent certain domain interactions to the upstream regions. This 'locks' such domains in a restricted number of conformations, somewhat similar to the model suggested for protein-assisted RNA folding in some group I introns (54). Recent computational analysis of low energy structures predicted for *Escherichia coli* 16S rRNA with topological constraints imposed by binding of ribosomal proteins suggests that such constraints may facilitate rRNA functional structure formation (55).

Interestingly, the chaperone-like effect of *P.abyssi* sRNAs on rRNA folding is mostly seen at higher temperatures

($\sim$80°C), than the maximal effect of the lone *A.fulgidus* sRNA 2 (70°C), approximately following the difference in the optimum growth temperatures for these species, 95–100°C and 65–70°C, respectively (28). It should be noted here that our simulations at extremely high temperatures (>85°C) do not represent the real environmental conditions for hyper-thermophiles, where high pressure may be expected to stabilize RNA structures with high melting points (56). In the absence of accurate thermodynamic parameters that take pressure into account, it can be speculated that calculations at moderately high temperatures may serve as an approximation for the combined effect of an elevated temperature and a high pressure.

A specific adaptation of archaeal rRNA sequences, increasing reliability of folding at high temperatures, has also been observed in equilibrium properties of alternative secondary structures (8). In addition to relatively high G+C contents, which prevents melting at high temperatures, thermophilic rRNAs have base pairing probability distributions indicating to more well-defined structures than those of other species. These distributions, calculated at the same temperature, differ from those computed for organisms living at lower temperatures. This explains relatively good predictions obtained for archaeal 16S and 23S rRNAs by energy minimization at the standard temperature (37°C) conditions (8,46). However, the calculations of base pairing probabilities for higher temperatures indicate less defined structures (8). It is remarkable that the sRNA-chaperone effect is seen just at high temperatures, while at lower temperatures implementation of sRNA binding into folding simulation can even decrease the prediction quality. Thus, the sRNA functioning in facilitating correct rRNA folding seems to be adapted to the optimum growth temperature.

The binding of proteins assisting in RNA tertiary structure formation is known to exert sometimes opposite effects depending on the conditions and/or folding steps (57,58). Interestingly, the RNA-chaperone activity of the *E.coli* protein StpA was found to have different effects on functional RNA structure formation in some mutants at different temperatures (58). In this case, however, the RNA-chaperone activity was pronounced at lowered temperatures, while the opposite effect was observed at 37°C. On the other hand, the mechanism of the StpA chaperone effect, namely the non-specific loosening of misfolded tertiary structures (59), is different from sRNA specific binding that can influence secondary structure formation. Proteins that have RNA-chaperone activity due to specific RNA-stabilizing interactions exist as well. Moreover, it was shown that one of such proteins, required for the stabilisation of the catalytic core of some group I introns, can be replaced by a peripheral RNA structure (60).

Our simulations demonstrate that sRNA effects on RNA folding depend on the location of the binding sites. The strongest effects (42 or more correctly predicted stems in case of *P.abyssi* 16S rRNA) were observed for sRNAs with both binding sites located within one of the four major domains (61) of the 16S rRNA secondary and tertiary structure or flanking two or more of these domains. Approximately one-half of the 17 tested *P.abyssi* sRNA-binding topologies yielded such an effect (Table 1). The binding sites of the three most efficient *P.abyssi* sRNAs (sRNA19, sRNA4, sRNA29) are located within either the 5'domain (nt 1–522)

or the central one (nt 523–876). Interestingly, *P.abyssi* sRNA 14 and *A.fulgidus* sRNA 2 bind 16S rRNA (complementarity regions 9–19/778–787 and 11–22/845–856 correspondingly) at topologically comparable positions flanking the 5'domain and a large part of the central one. Such a location explains why these sRNAs may guide the folding of long-range interactions in these domains and indeed, they both exhibit considerable effects on the final rRNA structures (Table 1). On the other hand, mostly no effect was observed for sRNAs with two binding sites located within the interiors of different domains. Also, the binding of the *P.abyssi* sRNA26 with two sites located in the 3'-proximal minor domain did not produce any effect.

These results are consistent with the hypothesis that co-transcriptional binding of double-guide sRNAs would stabilize the domains located between two binding sites. The analysis of simulated folding pathways (see e.g. Figure 6) shows that such effects may also be due to non-specific resolving of misfolded conformations by the barriers created for the interactions of a domain locked in between binding sites with other 16S rRNA regions. For many sRNAs the simulations also reveal sRNA influence rather far from the domain locked in between the binding sites, like e.g. stabilization of the pairing 891–897/1351–1357, 911–919/1194–1202 and 949–955/1182–1188 (Figure 5) in the *P.abyssi* 16S rRNA by sRNA4 (complementarity regions 54–63 and 363–372). Surprisingly, we did not reveal any dependence of observed effects on the length of sequence between binding sites on rRNA. A strong influence on simulations was observed for both sRNAs with relatively closely located sites and those with distant sites.

With the exception of plant snoRNAs (30), snoRNAs with two antisense elements complementary to the same rRNA are relatively rare in eukaryotes as compared to thermophilic archaeal sRNAs (29,33). For some of them, in particular those with complementarities without a modification function, a chaperoning function has been suggested [e.g. (25,45,62)]. In plants, such a function could be linked to the requirement to maintain the native rRNA folding upon exposure to high temperatures (30). Our preliminary simulations for yeast U14 snoRNA binding to 16S rRNA reveal some effects at both low and high temperatures, but the simulated folding pathways turned out to be less reproducible as compared with those simulated for thermophilic archaea. This may be caused either by sequence features of eukaryotic rRNAs that determine less defined structures (8) or by the inability of the used folding protocol to properly handle the partition of predicted folding pathways due to higher barriers at lower temperatures. Apparently, more accurate calculations are needed for reliable predictions of eukaryotic snoRNA–rRNA folding intermediates at physiological temperatures.

It should be noted that the model used is a rather rough approximation of the folding process. The genetic algorithm simulations may reveal some important long-lived folding intermediates, but they do not implement the real kinetics of folding. More realistic simulations of RNA-folding kinetics require computation of the transition rates between all local minima in the free energy landscape, which may be approximated by clustering similar configurations (63–65). Computationally, this is a challenging task, which is currently restricted to molecules of about 400 nt at the level of stems

as elementary steps (64) and of the size of tRNAs (76 nt) at the level of base pairs (65).

Furthermore, our model of sRNA-assisted rRNA folding implements forced temporary sRNA binding instead of more realistic competition between intra- and intermolecular base pairings. Such a computation should incorporate the concentrations of both rRNAs and sRNAs and rate constants for sRNA binding and dissociation. The algorithm for the prediction of RNA–RNA hybridization with account of concentration effects has been proposed recently (66), but it is only considering equilibrium structures.

Despite the very approximate nature of our model, it does show significant improvements of structure predictions upon sRNA binding implementation and therefore may serve as a 'proof-of-the-principle' for the mechanism of sRNA-induced rearrangements of the secondary structure in archaeal rRNAs.

ACKNOWLEDGEMENTS

REFERENCES

1. Woodson,S.A. (2000) Recent insights on RNA folding mechanisms from catalytic RNA. *Cell Mol. Life Sci.*, **57**, 796–808.
2. Thirumalai,D. and Woodson,S.A. (1996) Kinetics of folding of proteins and RNA. *Acc. Chem. Res.*, **29**, 433–439.
3. Pan,T., Thirumalai,D. and Woodson,S.A. (1997) Folding of RNA involves parallel pathways. *J. Mol. Biol.*, **273**, 7–13.
4. Fang,X.W., Pan,T. and Sosnick,T.R. (1999) Mg2+-dependent folding of a large ribozyme without kinetic traps. *Nature Struct. Biol.*, **6**, 1091–1095.
5. Su,L.J., Brenowitz,M. and Pyle,A.M. (2003) An alternative route for the folding of large RNAs: apparent two-state folding by a group II intron ribozyme. *J. Mol. Biol.*, **334**, 639–652.
6. Gultyaev,A.P., van Batenburg,F.H. and Pleij,C.W.A. (1995) The computer simulation of RNA folding pathways using a genetic algorithm. *J. Mol. Biol.*, **250**, 37–51.
7. Higgs,P.G. (1995) Thermodynamic properties of transfer-RNA—a computational study. *J. Chem. Soc. Faraday Trans.*, **91**, 2531–2540.
8. Huynen,M., Gutell,R. and Konings,D. (1997) Assessing the reliability of RNA folding using statistical mechanics. *J. Mol. Biol.*, **267**, 1104–1112.
9. Schultes,E.A., Hraber,P.T. and LaBean,T.H. (1999) Estimating the contributions of selection and self-organization in RNA secondary structure. *J. Mol. Evol.*, **49**, 76–83.
10. Gultyaev,A.P., van Batenburg,F.H. and Pleij,C.W.A. (2002) Selective pressures on RNA hairpins *in vivo* and *in vitro*. *J. Mol. Evol.*, **54**, 1–8.
11. Kramer,F.R. and Mills,D.R. (1981) Secondary structure formation during RNA synthesis. *Nucleic Acids Res.*, **9**, 5109–5124.
12. Lewicki,B.T., Margus,T., Remme,J. and Nierhaus,K.H. (1993) Coupling of rRNA transcription and ribosomal assembly *in vivo*. Formation of active ribosomal subunits in *Escherichia coli* requires transcription of rRNA genes by host RNA polymerase which cannot be replaced by bacteriophage T7 RNA polymerase. *J. Mol. Biol.*, **231**, 581–593.
13. Heilman-Miller,S.L. and Woodson,S.A. (2003) Effect of transcription on folding of the *Tetrahymena* ribozyme. *RNA*, **9**, 722–733.
14. Meyer,I.M. and Miklos,I. (2004) Co-transcriptional folding is encoded within RNA genes. *BMC Mol. Biol.*, **5**, 10.
15. Ma,C.K., Kolesnikow,T., Rayner,J.C., Simons,E.L., Yim,H. and Simons,R.W. (1994) Control of translation by mRNA secondary structure: the importance of the kinetics of structure formation. *Mol. Microbiol.*, **14**, 1033–1047.
16. Gultyaev,A.P., Franch,T. and Gerdes,K. (1997) Programmed cell death by hok/sok of plasmid R1: coupled nucleotide covariations reveal a phylogenetically conserved folding pathway in the hok family of mRNAs. *J. Mol. Biol.*, **273**, 26–37.
17. Poot,R.A., Tsareva,N.V., Boni,I.B. and van Duin,J. (1997) RNA folding kinetics regulate translation of phage MS2 maturation gene. *Proc. Natl Acad. Sci. USA*, **94**, 10110–10115.
18. Morgan,S.R. and Higgs,P.G. (1996) Evidence for kinetic effects in the folding of large RNA molecules. *J. Chem. Phys.*, **105**, 7152–7157.
19. Herschlag,D. (1995) RNA chaperones and the RNA folding problem. *J. Biol. Chem.*, **270**, 20871–20874.
20. Schroeder,R., Barta,A. and Semrad,K. (2004) Strategies for RNA folding and assembly. *Nature Rev. Mol. Cell Biol.*, **5**, 908–919.
21. Bachellerie,J.P., Michot,B., Nicoloso,M., Balakin,A., Ni,J. and Fournier,M.J. (1995) Antisense snoRNAs: a family of nucleolar RNAs with long complementarities to rRNA. *Trends Biochem. Sci.*, **20**, 261–264.
22. Steitz,J.A. and Tycowski,K.T. (1995) Small RNA chaperones for ribosome biogenesis. *Science*, **270**, 1626–1627.
23. Gerbi,S.A. (1995) Small nucleolar RNA. *Biochem. Cell Biol.*, **73**, 845–858.
24. Hughes,J.M.X. (1996) Functional-base-pairing interaction between highly conserved elements of U3 small nucleolar RNA and the small ribosomal subunit RNA. *J. Mol. Biol.*, **259**, 645–654.
25. Liang,W.Q. and Fournier,M.J. (1995) U14 base-pairs with 18S rRNA: a novel snoRNA interaction required for rRNA processing. *Genes Dev.*, **9**, 2433–2443.
26. Peculis,B.A. (2001) snoRNA nuclear import and potential for cotranscriptional function in pre-rRNA processing. *RNA*, **7**, 207–219.
27. Bachellerie,J.P. and Cavaille,J. (1997) Guiding ribose methylation of rRNA. *Trends Biochem. Sci.*, **22**, 257–261.
28. Dennis,P.P., Omer,A. and Lowe,T. (2001) A guided tour: small RNA function in Archaea. *Mol. Microbiol.*, **40**, 509–519.
29. Gaspin,C., Cavaille,J., Erauso,G. and Bachellerie,J.P. (2000) Archaeal homologs of eukaryotic methylation guide small nucleolar RNAs: lessons from the *Pyrococcus* genomes. *J. Mol. Biol.*, **297**, 895–906.
30. Barneche,F., Gaspin,C., Guyot,R. and Echeverria,M. (2001) Identification of 66 box C/D snoRNAs in *Arabidopsis thaliana*: extensive gene duplications generated multiple isoforms predicting new ribosomal RNA 2′-O-methylation sites. *J. Mol. Biol.*, **311**, 57–73.
31. Ziesche,S.M., Omer,A.M. and Dennis,P.P. (2004) RNA-guided nucleotide modification of ribosomal and non-ribosomal RNAs in Archaea. *Mol. Microbiol.*, **54**, 980–993.
32. Tran,E.J., Zhang,X. and Maxwell,E.S. (2003) Efficient RNA 2′-O-methylation requires juxtaposed and symmetrically assembled archaeal box C/D and C′/D′ RNPs. *EMBO J.*, **22**, 3930–3940.
33. Omer,A.D., Lowe,T.M., Russell,A.G., Ebhardt,H., Eddy,S.R. and Dennis,P.P. (2000) Homologs of small nucleolar RNAs in Archaea. *Science*, **288**, 517–522.
34. Dragon,F., Gallagher,J.E.G., Compagnone-Post,P.A., Mitchell,B.M., Porwancher,K.A., Wehner,K.A., Wormsley,S., Settlage,R.E., Shabanowitz,J., Oshelm,Y. *et al.* (2002) A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis. *Nature*, **417**, 967–970.
35. Maden,B.E. (1990) The numerous modified nucleotides in eukaryotic ribosomal RNA. *Prog. Nucleic Acid Res. Mol. Biol.*, **39**, 241–303.
36. Venema,J. and Tollervey,D. (1999) Ribosome synthesis in *Saccharomyces cerevisiae*. *Annu. Rev. Genet.*, **33**, 261–311.
37. Speckman,W.A., Li,Z.-H., Lowe,T.M., Eddy,S.R., Terns,R.M. and Terns,M.P. (2002) Archaeal guide RNAs function in rRNA modification in the eukaryotic nucleus. *Curr. Biol.*, **12**, 199–203.
38. Gultyaev,A.P., van Batenburg,F.H. and Pleij,C.W.A. (1995) The influence of a metastable structure in plasmid primer RNA on antisense RNA binding kinetics. *Nucleic Acids Res.*, **23**, 3718–3725.
39. Gultyaev,A.P., van Batenburg,F.H. and Pleij,C.W.A. (1998) Dynamic competition between alternative structures in viroid RNAs simulated by an RNA folding algorithm. *J. Mol. Biol.*, **276**, 43–55.
40. Abrahams,J.P., van den Berg,M., van Batenburg,E. and Pleij,C.W.A. (1990) Prediction of RNA secondary structure, including pseudoknotting, by computer simulation. *Nucleic Acids Res.*, **18**, 3035–3044.
41. Walter,A.E., Turner,D.H., Kim,J., Lyttle,M.H., Muller,P., Mathews,D.H. and Zuker,M. (1994) Coaxial stacking of helixes

enhances binding of oligoribonucleotides and improves predictions of RNA folding. *Proc. Natl Acad. Sci. USA*, **91**, 9218–9222.

42. Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.

43. Wuyts,J., Perriere,G. and Van de Peer,Y. (2004) The European ribosomal RNA database. *Nucleic Acids Res.*, **32**, D101–D103.

44. Samarsky,D.A. and Fournier,M.J. (1999) A comprehensive database for the small nucleolar RNAs from *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **27**, 161–164.

45. Dunbar,D.A. and Baserga,S.J. (1998) The U14 snoRNA is required for 2′-*O*-methylation of the pre-18S rRNA in *Xenopus* oocytes. *RNA*, **4**, 195–204.

46. Konings,D.A. and Gutell,R.R. (1995) A comparison of thermodynamic foldings with comparatively derived structures of 16S and 16S-like rRNAs. *RNA*, **1**, 559–574.

47. Pardon,B. and Wagner,R. (1995) The *Escherichia coli* ribosomal RNA leader nut region interacts specifically with mature 16S RNA. *Nucleic Acids Res.*, **23**, 932–941.

48. Besancon,W. and Wagner,R. (1999) Characterization of transient RNA-RNA interactions important for the facilitated structure formation of bacterial ribosomal 16S RNA. *Nucleic Acids Res.*, **27**, 4353–4362.

49. Cote,C.A., Greer,C.L. and Peculis,B.A. (2002) Dynamic conformational model for the role of ITS2 in pre-rRNA processing in yeast. *RNA*, **8**, 786–797.

50. Liiv,A. and Remme,J. (2004) Importance of transient structures during post-transcriptional refolding of the pre-23S rRNA and ribosomal large subunit assembly. *J. Mol. Biol.*, **342**, 725–741.

51. Pan,T., Artsimovitch,I., Fang,X.W., Landick,R. and Sosnick,T.R. (1999) Folding of a large ribozyme during transcription and the effect of the elongation factor NusA. *Proc. Natl Acad. Sci. USA*, **96**, 9545–9550.

52. Isambert,H. and Siggia,E.D. (2000) Modeling RNA folding paths with pseudoknots: application to hepatitis delta virus ribozyme. *Proc. Natl Acad. Sci. USA*, **97**, 6515–6520.

53. Diegelman-Parente,A. and Bevilacqua,P.C. (2002) A mechanistic framework for co-transcriptional folding of the HDV genomic ribozyme in the presence of downstream sequence. *J. Mol. Biol.*, **324**, 1–16.

54. Solem,A., Chatterjee,P. and Caprara,M.G. (2002) A novel mechanism for protein-assisted group I intron splicing. *RNA*, **8**, 412–425.

55. Favaretto,P., Bhutkar,A. and Smith,T.F. (2005) Constraining ribosomal RNA conformational space. *Nucleic Acids Res.*, **33**, 5106–5111.

56. Dubins,D.N., Lee,A., Macgregor,R.B.Jr and Chalikian,T.V. (2001) On the stability of double stranded nucleic acids. *J. Am. Chem. Soc.*, **123**, 9254–9259.

57. Webb,A.E. and Weeks,K.M. (2001) A collapsed state functions to self-chaperone RNA folding into a native ribonucleoprotein complex. *Nature Struct. Biol.*, **8**, 135–140.

58. Grossberger,R., Mayer,O., Waldsich,C., Semrad,K., Urschitz,S. and Schroeder,R. (2005) Influence of RNA structural stability on the RNA chaperone activity of the *Escherichia coli* protein StpA. *Nucleic Acids Res.*, **33**, 2280–2289.

59. Waldsich,C., Grossberger,R. and Schroeder,R. (2002) RNA chaperone StpA loosens interactions of the tertiary structure in the td group I intron *in vivo*. *Genes Dev.*, **16**, 2300–2312.

60. Mohr,G., Caprara,M.G., Guo,Q. and Lambowitz,A.M. (1994) A tyrosyl-tRNA synthetase can function similarly to an RNA structure in the *Tetrahymena* ribozyme. *Nature*, **370**, 147–150.

61. Wimberly,B.T., Brodersen,D.E., Clemons,W.M.C.Jr, Morgan-Warren,R.J., Carter,A.P., Vonrhein,C., Hartsch,T. and Ramakrishnan,V. (2000) Structure of the 30S ribosomal subunit. *Nature*, **407**, 327–339.

62. Vitali,P., Royo,H., Seitz,H., Bachellerie,J.-P., Huttenhofer,A. and Cavaille,J. (2003) Identification of 13 novel human modification guide RNAs. *Nucleic Acids Res.*, **31**, 6543–6551.

63. Xayaphoummine,A., Bucher,T., Thalmann,F. and Isambert,H. (2003) Prediction and statistics of pseudoknots in RNA structures using exactly clustered stochastic simulations. *Proc. Natl Acad. Sci. USA*, **100**, 15310–15315.

64. Xayaphoummine,A., Bucher,T. and Isambert,H. (2005) Kinefold web server for RNA/DNA folding path and structure prediction including pseudoknots and knots. *Nucleic Acids Res.*, **33**, W605–W610.

65. Wolfinger,M.T., Svrcek-Seiler,W.A., Flamm,C., Hofacker,I.L. and Stadler,P.F. (2004) Efficient computation of RNA folding dynamics. *J. Phys. A Math. Gen.*, **37**, 4731–4741.

66. Hackermuller,J., Meisner,N.C., Auer,M., Jaritz,M. and Stadler,P.F. (2005) The effect of RNA secondary structures on RNA-ligand binding and the modifier RNA mechanism: a quantitative model. *Gene*, **345**, 3–12.