ISSN 0006-2979, Biochemistry (Moscow), 2010, Vol. 75, No. 7, pp. 841-850. © Pleiades Publishing, Ltd., 2010.
Original Russian Text © A. V. Surdina, T. I. Rassokhin, A. V. Golovin, V. A. Spiridonova, A. M. Kopylov, 2010, published in Biokhimiya, 2010, Vol. 75, No. 7, pp. 954-965.
Originally published in Biochemistry (Moscow) On-Line Papers in Press, as Manuscript BM09-362, May 30, 2010.

Mapping the Ribosomal Protein S7 Regulatory Binding Site on mRNA of the *E. coli* Streptomycin Operon

A. V. Surdina[1], T. I. Rassokhin[1], A. V. Golovin[2], V. A. Spiridonova[1], and A. M. Kopylov[1,3]*

[1]*Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, 119991 Moscow, Russia; fax: (495) 939-3181; E-mail: kopylov.alex@gmail.com*
[2]*Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, 119991 Moscow, Russia*
[3]*Chemical Faculty, Lomonosov Moscow State University, 119991 Moscow, Russia*

Received December 17, 2009
Revision received March 23, 2010

Abstract—In this work it is shown by deletion analysis that an intercistronic region (ICR) approximately 80 nucleotides in length is necessary for interaction with recombinant *E. coli* S7 protein (r6hEcoS7). A model is proposed for the interaction of S7 with two ICR sites—region of hairpin bifurcations and Shine–Dalgarno sequence of cistron S7. A *de novo* RNA binding site for heterologous S7 protein of *Thermus thermophilus* (r6hTthS7) was constructed by selection of a combinatorial RNA library based on *E. coli* ICR: it has only a single supposed protein recognition site in the region of bifurcation. The SERW technique was used for selection of two intercistronic RNA libraries in which five nucleotides of a double-stranded region, adjacent to the bifurcation, had the randomized sequence. One library contained an authentic AG (−82/−20) pair, while in the other this pair was replaced by AU. A serwamer capable of specific binding to r6hTthS7 was selected; it appeared to be the RNA68 mutant with eight nucleotide mutations. The serwamer binds to r6hTthS7 with the same affinity as homologous authentic ICR of *str* mRNA binds to r6hEcoS7; apparent dissociation constants are 89 ± 43 and 50 ± 24 nM, respectively.

DOI: 10.1134/S0006297910070059

Key words: ribosomal biogenesis, S7 protein, streptomycin operon, bacteria, translation regulation, regulome, SELEX, SERW

Among numerous "omics" of systems biology, regulomics and interactomics occupy central places [1, 2]. Development of interactomics, which studies the interaction of biological molecules, defines the possibility of development of methods for purposeful interference into processes of specific interaction and thus creation of theoretical bases for directed synthesis of drugs of a new generation [3, 4]. This is especially important for design of antibacterial preparations.

Nucleic–protein interactions play an important role in many cell processes, among which biogenesis of ribosomes is of special interest. The ribosome is a multicomponent supramolecular RNA–protein complex capable of self-assembly *in vitro*. Rapidly growing bacterial cells spend about half of their resources for ribosome biogenesis, which includes synthesis of rRNA and ribosomal proteins (r-proteins) as well as their assembly [5, 6]. Ribosome biogenesis is rather strictly regulated by RNA–protein interactions [5, 7, 8].

Proteins S7 and S4 take part in biogenesis of the small ribosomal subparticle of *E. coli* [9]. S7 initiates the subparticle self-assembly *in vitro* via interaction with a local 16S rRNA region [10-12]. Protein S7 is translated from the *rpsG* cistron within the streptomycin (*str*) operon. The *str* operon includes cistrons of ribosomal proteins S12 (*rpsL*) and S7 (*rpsG*), translation elongation factor EF-G (*fus*), and one of two copies of the gene of transla-

Abbreviations: ICR (intercistronic region), region of *E. coli str* mRNA between cistrons S12 and S7; r6hEcoS7, recombinant protein S7 of *E. coli* with six *N*-terminal His residues; r6hTthS7, recombinant protein S7 of *Thermus thermophilus* with six *N*-terminal His residues; r-proteins, ribosomal proteins; SERF (Selection of Random RNA Fragments), selection of library of given RNA random fragments; SERW (Selection of Random RNA Windows), selection of library for basic RNA structure with randomized site several nucleotides in size; serwamer, RNA fragment obtained by SERW and capable of protein binding; *str*, streptomycin.

* To whom correspondence should be addressed.

Fig. 1. Scheme of *E. coli* K12 streptomycin operon. Coordinates are taken from GenBank (code U00096). Rectangles with protein names point to the operon genes. P_{tuf1} and P_{tuf2} designate promoters of one of two EF-Tu genes, that belongs to the *str* operon. Gene *yhe* L to the left from the P promoter and gene *chi* A to the right from EF-Tu are not incorporated in the *str* operon.

tion elongation factor EF-Tu (*tufA*) that has two of its own promoters (Fig. 1) [5, 7, 8].

Interest in investigation of the *str* operon is also defined by a practical aspect. Chromosomal mutations of S12 protein result in streptomycin dependence or resistance phenotype [13], which gave rise to the name of the operon. Since streptomycin has been used in clinical practice for a rather long time, mapping such mutations is necessary for medical population microbiology.

An interesting peculiarity of the *E. coli str* operon is the presence of an extended 96-nucleotide-long RNA region between cistrons S12 and S7 (intercistronic region, ICR). Although cistrons S12 and S7 are much separated physically, S7 translation is coupled with that of the foregoing S12, i.e. it proceeds only successively and subordinately. In the absence of 16S rRNA synthesis, when the rRNA regulome is switched off, the excess of expressed EcoS7 interacts with ICR of *str* mRNA and inhibits its own translation and translation of the following cistron of elongation factor EF-G [14, 15].

A present-day hypothesis suggests that the coupling of cistron S12 and S7 translation is assured by the specific spatial structure of ICR in which the terminating codon of cistron S12 is adjacent to the translation initiation site of cistron S7, which allows the terminating ribosome to "jump over" (to diffuse for a short distance) from cistron S12 to S7 [16]. Inhibition by S7 protein of coupled translation of cistrons S12 and S7 can be explained by two screen versions. According to the first, the protein stabilizes the ICR structure, which results in translation "retardation" at the S12 cistron terminus (and, possibly, to premature termination). According to the second version, the protein, on the contrary, destabilizes ICR structure, which positions the terminating ribosome far from the S7 cistron translation initiation site. One can choose between these two possibilities on the basis of properties of the binary complex of ICR with S7 protein.

It was shown earlier that S7 could form *in vitro* the binary complex with ICR [15-19]. Besides, the target serfamer, corresponding to the *str* mRNA ICR, was found using SERF in a search for possible regulatory *str* operon RNA regions by selection of random fragments of *E. coli str* mRNA for affinity to recombinant S7 protein [15]. Unlike authentic ICR, it contained nucleotide mutations that probably emerged due to multiple copying errors and were chosen in the course of affinity selection because they enhanced affinity to the protein.

The unique structure of the *E. coli* ICR makes it possible to use ICR as a characteristic marker for microorganism typing. For the same reason, the RNA–protein complex of ICR can be considered as a potential target for design of antibacterial agents with group specificity.

This work deals with identification of the minimal region in the *E. coli str* mRNA capable of interaction with recombinant r6hEcoS7 protein. In addition, the recognizable element for thermophilic protein homolog r6hTthS7 was created by selection of *de novo* RNA libraries. The RNA libraries were obtained on the basis of *E. coli* ICR basal structure by randomization of the five nucleotide-long fragment (SERW). A model of spatial structure of EcoS7 complex with *E. coli str* RNA ICR is proposed that incorporates all presently known experimental data.

MATERIALS AND METHODS

Isolation of recombinant *E. coli* protein (r6hEcoS7) from superproducing *E. coli* cells. The recombinant S7 protein r6hEcoS7 containing at its *N*-terminus 10 additional amino acids and six His residues [11], whose gene was cloned into vector pET28b+ (Novagen, Germany) at sites of cleavage by restriction endonucleases NdeI and EcoRI, was isolated as described previously [11, 16, 17, 20] with slight modifications. Cells were lysed for 3.5 h at 0°C in Rec-4 buffer: 20 mM Hepes-KOH, pH 7.5, 4 mM $MgAc_2$, 1 M NH_4Cl, 5 mM mercaptoethanol, and 6 M guanidine hydrochloride. The supernatant was applied onto Ni-NTA Agarose (Qiagen, Germany), and the resin was washed three times with Rec-4 and then with Rec-4 containing 20 mM imidazole. Protein absorbed on the resin was eluted with Rec-4 containing 1 M imidazole. The S7 was dialyzed against Rec-4 with 2 M urea, then against Rec-4 with 0.4 M NH_4Cl and stored at –70°C. Purity of isolated protein was characterized by standard gel electrophoresis according to Laemmli. For complex formation with RNA, the protein was dialyzed against buffer A containing 20 mM Tris-HCl, pH 7.4, 4 mM $MgAc_2$, 400 mM NH_4Cl, 0.2 mM EDTA, and 4 mM mercaptoethanol.

Isolation of *T. thermophilus* recombinant S7 protein (r6hTthS7) from superproducing *E. coli* cells. The recombinant S7 protein r6hTthS7 contains at its *N*-terminus 10 additional amino acids and six His residues; its gene was

cloned into pQE32 vector at the BamHI restriction site. The obtained construct was used to transform *E. coli* M15 strain containing plasmid pREP4 with *lac*-repressor gene.

For isolation of r6hTthS7 protein, cells were collected by centrifugation and washed with buffer A. The pellet was dissolved in buffer A with 6 M guanidine hydrochloride and left for 2 h at 0°C. After removal of debris, the supernatant was applied onto Ni-NTA agarose (Qiagen), the column was washed by buffer A with 20 mM imidazole, and r6hTthS7 was eluted with buffer A with 1 M imidazole. The protein was transferred into buffer A with 6 M urea, then dialyzed three times, 1 h each, against buffer A, and stored at −70°C. Purity of isolated protein was characterized with standard gel electrophoresis according to Laemmli.

Production of EcoStr mRNA deletion fragments. DNA templates for transcription of mRNA deletion fragments 143, 103, 84, 73, 61, AG, and AU were obtained by PCR with appropriate oligonucleotides from plasmid pDD143 containing a DNA copy of the *E. coli str* mRNA under control of T7 phage promoter. The 5′-terminal oligonucleotide primers contained the T7 phage promoter sequence. The ^{32}P-labeled fragments of EcoStr mRNA were obtained using *in vitro* transcription by T7 RNA polymerase (MBI, Fermentas, Lithuania) in the presence of α-[^{32}P]UTP according to standard technique and purified by PAGE in the presence of 7 M urea. Specific radioactivity of RNA was calculated after extraction and measurement of radioactivity and absorption at 260 nm; it was usually (30-50)·10^3 cpm/pmol.

Selection of combinatorial RNA library using r6hEcoS7 and r6hTthS7 proteins. The initial DNA library was obtained using PCR with the following primers: tm1, GACGAATTCTAATACGACTCACTATAGGGTAAGGCCAAACGTTTT; tm2, GGGTCAGGATTGTCCAAAACTCTA, on template TAAGGCCAAACGTTTTAACTTAAATGTCNNNNNAAACTCGTAGAGTTTTGGACAATCCTGA (the T7 phage promoter is underlined).

The RNA library was obtained by *in vitro* transcription by T7 phage RNA polymerase.

After complex formation with protein in appropriate conditions (Table 1), RNA was extracted from the complex on a filter by TE buffer with 6 M urea, treated by phenol and chloroform, and precipitated by ethanol. After precipitation, RNA was dissolved in deionized water, then 4 μl of 5-fold reverse transcription buffer, 2 μl of 10 mM dNTP mixture, 20 pmol of 3′-terminal primer, 0.5 μl (20 units) of RiboLock RNase inhibitor (Fermentas), and 1 μl (200 units) of M.MulV reverse transcriptase (Fermentas) were added. The mixture was incubated for 1 h at 42°C. The resulting cDNA of the enriched fraction of the combinatorial RNA library was amplified using PCR.

RNA complex formation with protein. Complex formation was carried out in 200 μl buffer containing 20 mM Tris-Ac, pH 7.6, 4 mM mercaptoethanol, 0.2% BSA solution supplemented with 15 mM $MgAc_2$, 200 mM NH_4Cl for r6hEcoS7 protein and 7 mM $MgAc_2$, 350 mM NH_4Cl for r6hTthS7. The RNA was heated at 95°C for 2 min and cooled on ice. Protein r6hTthS7 was heated at 65°C for 10 min and cooled on ice. Both components were heated separately for 30 min (at 37°C for r6hEcoS7 and at 42°C for r6hTthS7) and then together for 30 min at appropriate temperatures. Before sorption on membranes, mixtures were cooled to room temperature. Filter radioactivity was measured heterogeneously in water according to Cherenkov.

Table 1. Change in protein/RNA ratio during selection of serwamers to recombinant proteins S7 in experiments 1-4

Selection to protein r6hEcoS7		Selection to protein r6hTthS7	
1 (15 mM $MgCl_2$, 200 mM NH_4Cl)	2 (15 mM $MgCl_2$, 200 mM NH_4Cl)	3 (7 mM $MgCl_2$, 350 mM NH_4Cl)	4 (7 mM $MgCl_2$, 350 mM NH_4Cl; fragment AG, fragment AU)
	2 : 1		
	1 : 1	1 : 1	
	1 : 2	1 : 2	
1 : 2	1 : 2	1 : 2	
	1 : 3		
1 : 5	1 : 5	1 : 5	
1 : 5	1 : 5	1 : 5	
1 : 10	1 : 10	1 : 10	1 : 10
		1 : 10	
1 : 20	1 : 20	1 : 20	1 : 20
		1 : 20	
1 : 50	1 : 50	1 : 50	1 : 50
1 : 50			
1 : 100			1 : 100
1 : 100			
1 : 200			

In the case of SERW, complex formation was carried out in the volume of 200-1000 μl. The concentration of the RNA library was 100 nM.

Generation of complex formation isotherms. The extent of complex formation was measured by sorption on nitrocellulose filters (0.45 μm) at filtration rate 0.5 ml/min. Filter radioactivity was determined heterogeneously in water according to Cherenkov. Binding isotherms were calculated using linearization in Scatchard coordinates as described elsewhere [16].

Modeling spatial structure of S7 protein complex with *E. coli str* mRNA ICR. Coordinates of homolog protein

TthS7, whose structure was determined with resolution below 2 Å (PDB code 1RSS) [21], were used for modeling.

The structure of ICR was modeled on the basis of secondary structure proposed by us earlier [16]. The ICR tertiary structure was modeled on server http://12.192.230.173 [22]. The model of spatial arrangement of the ICR phosphorus atoms was obtained by service-modeling of the RNA tertiary structure using the program of one of authors of this work (A. V. Golovin). Atom coordinates for non-canonical structure in the supposed site of recognition by the protein were obtained from X-ray data on 16S rRNA structure within the small subparticle of *T. thermophilus* ribosomes [23].

Three-dimensional structures were combined using the PyMOL program [24].

RESULTS AND DISCUSSION

Deletion analysis of *E. coli* streptomycin operon fragment between cistrons S12 and S7 (ICR). Saito and Nomura showed previously [15] that the 400-nucleotide-long *str* mRNA fragment including the mRNA region between cistrons S12 and S7 is necessary for EcoS7 protein binding upon regulation of biogenesis of *E. coli* small ribosomal subunit. Further works showed that the physical dimensions of the *str* mRNA fragment for binding recombinant protein EcoS7 could be diminished to 230 nucleotides [12] and then to 143 nucleotides [17].

The *str* mRNA ICR is able to form a heterologous complex with recombinant thermophilic homolog protein S7 from *T. thermophilus* (r6hTthS7) [18]. In the case of heterologous complex, dimensions of ICR capable of protein binding were even smaller, only 103 nucleotides [19], but the level of the apparent binding constant for the heterologous complex was three times lower.

Next the characteristics of the potential binding site were obtained by three methods. Data on the chemical modification and directed mutagenesis of ICR [15] cannot be unambiguously interpreted due to the absence of the model of ICR spatial structure and its complex with S7 protein. The contact of the protein with residues U35 and U34, i.e. with the hairpin III, IV, and V bifurcation region, was shown by UV-induced chemical crosslinking (so-called "zero length" crosslinking) [17] (Fig. 2). A region with nucleotide sequence identical to that of 16S rRNA (hairpin 42) was found in the same region [15]. The second region of the *str* mRNA ICR with nucleotide sequence identical to that of 16S rRNA (hairpin 43) was found next to the hairpin bifurcation region (region III) [12].

Possible hypotheses concerning the mechanism of the coupled translation inhibition by S7 protein are associated with alteration of the ICR region II status [16]. Owing to this, system deletion analysis of ICR, including

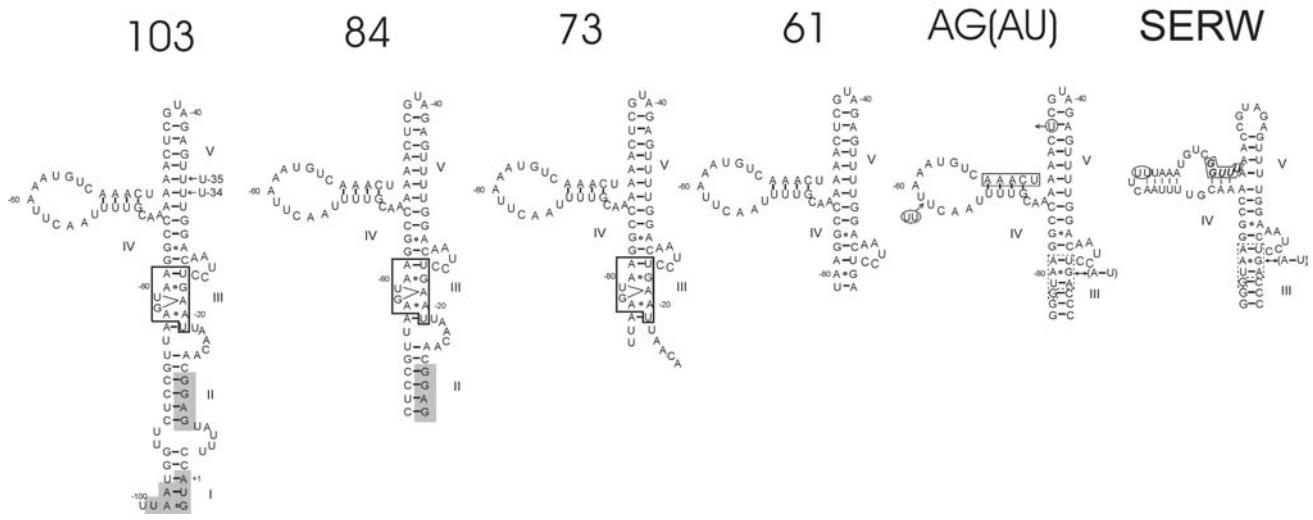

Fig. 2. Proposed secondary structure of a series of deletion fragments and their derivatives for the *E. coli str* operon region between cistrons S12 and S7 [16]. Residue A of initiator codon AUG of cistron S7 is taken as +1. Terminating codon S12, Shine–Dalgarno sequence, and initiating codon S7 are distinguished in gray tone. Figures above hairpins show the size of RNA fragments. The site of region III that is identical to 16S rRNA [16] is shown in the frame. AG(AU) fragments have a modified structure of region III and are terminated by the double-stranded region of three additional G-C pairs. A-G fragment contains an authentic mRNA pair (–82)A-G(–20). The substitute (–20)G/U was introduced into the A-U fragment to create the canonical A-U pair for region III stabilization in the ICR basement (shown by the two-way arrow). SERW designates the serwamer selected to the protein of r6hTthS7 libraries on the basis of *E. coli* AG(AU) fragments in which the site of five nucleotides (–53)AAACU(–49) shown in the frame, was randomized. Substitutions that emerged during selection are shown in oval frames for AG(AU) fragments: deletion U(–44) and insertion UU in front of C(–63) (shown in frames for serwamer) that resulted in the possibility of formation of modified secondary structure for ICR region IV.

region II, along with investigation of complex formation between the obtained series of fragments and S7 protein seem to be necessary.

In this work, systematic deletion analysis of the *str* mRNA ICR was carried out using PCR in the presence of a series of primers, where the forward primer in each pair included T7 promoter. The following series of *str* mRNA fragments were synthesized by T7 RNA polymerase from the obtained DNA templates: EcoStr mRNA143 of 143 nucleotides in length with coordinates –100, +43 (+1 corresponds to A of initiator codon AUG of cistron S7); EcoStr mRNA103 with coordinates –100, +3 (ICR itself); as well as shortened derivatives of ICR: EcoStr mRNA84 (–91, –8); EcoStr mRNA73 (–86, –14); EcoStr mRNA61 (–81, –21) (Fig. 2). As mentioned above, there are two hypotheses concerning the structure of the EcoS7 binding site. The ICR region III contains a site with short sequences identical to 16S rRNA [12]; besides, the bifurcation region of hairpins III and IV is exposed to proteins [17] (Fig. 2). Deletions were selected so that one of deletion fragments (mRNA73) contained both sites of possible recognition by the protein, while the other (mRNA61) contained only one supposed recognition site [15].

To obtain a series of the *str* mRNA ICR fragments, isotherms of complex formation with two recombinant proteins, homologous r6hEcoS7 and heterologous r6hTthS7, were determined.

Analysis of the *E. coli str* mRNA ICR deletion fragment binding to the r6hEcoS7 protein. Isotherms for r6hEcoS7 binding to EcoStr mRNA fragments were obtained under conditions identical to those for Eco16S rRNA [11]. Linearization of binding isotherms in Scatchard coordinates is shown in Fig. 3a. The Scatchard plot makes provides not only values for apparent dissociation constant ($K_{d\ app}$), but reveals the RNA fraction active in protein binding (Table 2), which depends on the procedure of sample preparation.

Deletion of 40 nucleotides (mRNA103) and 19 additional nucleotides (mRNA84) of the encoding part of cistron S7 from the 3′-end of the mRNA143 does not alter the binding parameters (Table 2). However, deletion of the next nine nucleotides from the 3′-end of intercistronic part, including the Shine–Dalgarno sequence of

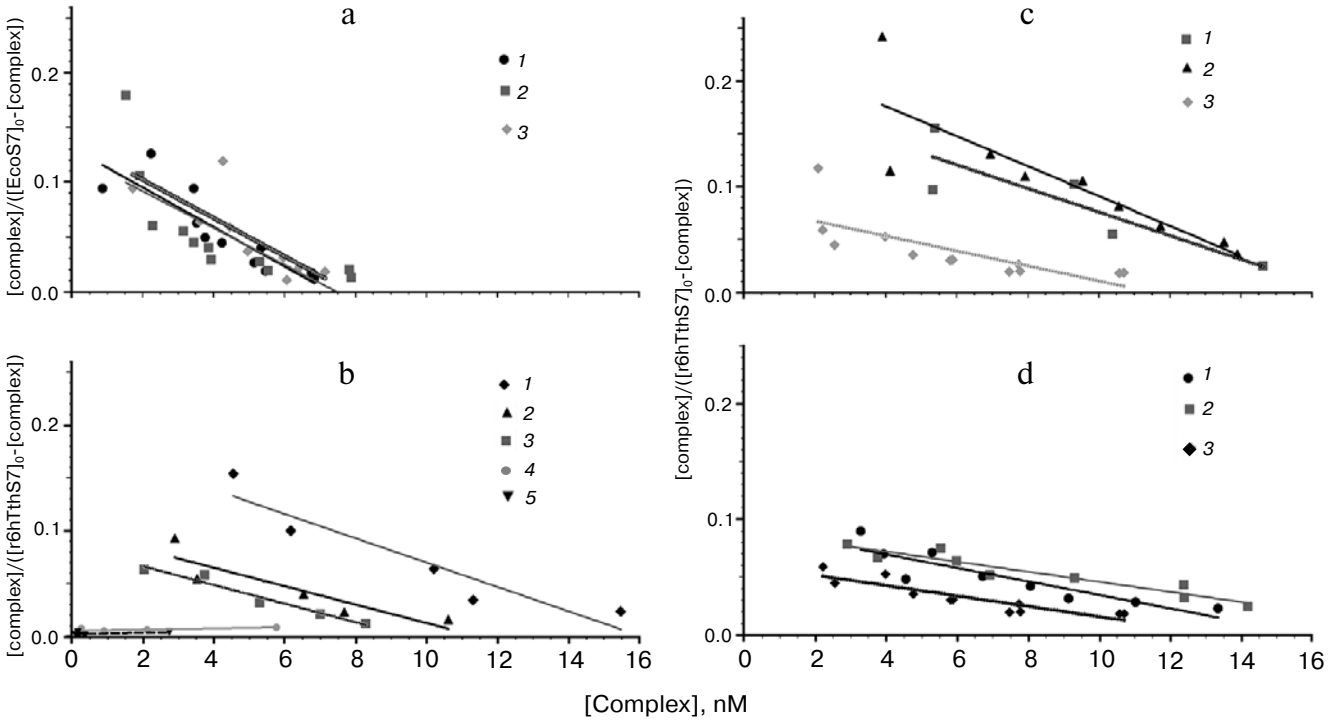

Fig. 3. Scatchard linearization of isotherm of r6hEcoS7 and r6hTthS7 protein binding to the EcoStr mRNA ICR fragments and their derivatives. Concentrations of EcoStr mRNA and their derivatives are 20 nM. a) r6hEcoS7: *1*, solid line) EcoStr mRNA143 ICR, $K_{d\ app}$ = 55 ± 12 nM; *2*, gray line) EcoStr mRNA103, $K_{d\ app}$ = 60 ± 25 nM; *3*, dotted line) EcoStr mRNA84, $K_{d\ app}$ = 50 ± 24 nM. b) r6hTthS7 binding to randomized fractions of AG(AU) fragments: *1*, gray line) enriched AG fragment fraction after 10 selection cycles (experiment 1), $K_{d\ app}$ = 87 ± 23 nM; *2*, solid line) enriched AU fragment fraction after four selection cycles (experiment 2), $K_{d\ app}$ = 115 ± 45 nM; *3*, dotted line) enriched AG fragment fraction after four selection cycles (experiment 2), $K_{d\ app}$ = 113 ± 16 nM; *4*, pale gray line) initial library of AG fragment, $K_{d\ app}$ is not determined; *5*, dashed line) initial AU fragment library, $K_{d\ app}$ is not determined. c) r6hTthS7 binding to serwamers from libraries: *1*, gray dotted line) serwamer from the AG fragment library, $K_{d\ app}$ = 89 ± 43 nM (experiment 4); *2*, solid line) serwamer from the AU fragment library, $K_{d\ app}$ = 70 ± 21 nM (experiment 4); *3*, dotted line) EcoStr mRNA84, $K_{d\ app}$ = 223 ± 42 nM. d) r6hTthS7: *1*, solid line) EcoStr mRNA143 ICR, $K_{d\ app}$ = 171 ± 41 nM; *2*, gray line) EcoStr mRNA103, $K_{d\ app}$ = 230 ± 31 nM; *3*, dotted line) EcoStr mRNA84, $K_{d\ app}$ = 182 ± 48 nM.

Table 2. Apparent dissociation constants ($K_{d\ app}$, nM) of complexes of r6hEcoS7 and r6hTthS7 proteins with intercistronic fragments EcoStr mRNA and serwamers and fraction of binding-competent RNA

Parameter	Protein	EcoStr-mRNA143	EcoStr-mRNA103	EcoStr-mRNA84	EcoStr-serwamer (fragments AG/AU)
$K_{d\ app}$	r6hEcoS7	55 ± 12	60 ± 25	50 ± 24	–
	r6hTthS7	171 ± 41	230 ± 31	202 ± 48*	89 ± 43/70 ± 21
Fraction of active RNA	r6hEcoS7	0.37	0.36	0.39	–
	r6hTthS7	0.80	1.00	0.68	0.79/0.85

* Average values of data shown in Figs. 3c and 3d are given.

cistron S7, results in loss of mRNA73 binding to r6hEcoS7 protein. The mRNA73 contains both variants of hypothetical protein binding sites [12, 14], probably because the loss of binding is the result of deletion of a previously unknown additional binding site that can be located in region II, i.e. in the region of duplex including the Shine–Dalgarno sequence (Fig. 2). The data correlate well with those obtained *in vivo* [14]: in the case of deletion of 14 nucleotides of region II, complementary to the Shine–Dalgarno sequence (deletion KS5, –96 –83), the level of cistron S7 coupled translation not only does not decrease, but it even increases somewhat when the S7 inhibitory effect is completely eliminated, evidently because the protein stops binding to such mRNA.

Up to the present time the possible existence of three *str* mRNA ICR structural determinants was considered that can define binding of S7 protein: two sites of identity with 16S rRNA [12, 14] and the site of UV-induced "zero length" crosslinking with the protein [17], which is overlapped with the site proposed by Saito and Nomura [15]. All these sites are grouped in the bifurcation region of the ICR top hairpins (Fig. 2).

The loss of the protein-binding ability upon deletion of double-stranded region II with the Shine–Dalgarno sequence can be explained by several factors. First, the Shine–Dalgarno sequence itself can be the determinant in recognition by the protein. This is supported by the fact of detection of translation initiation sites among low affinity serfamers during selection of the mRNA fragment library from the whole *str* operon [16]. Second, the duplex II structure as such can be important for binding because removal of the sequence complementary to the Shine–Dalgarno sequence results in loss of the effect of the coupled translation inhibition by S7 protein, possibly due to disappearance of protein binding [14]. Third, it can be supposed that the destruction of duplex II in the ICR basement results in destabilization of the whole ICR structure. In particular, duplex III in the absence of duplex II is not able to retain complex structure of the supposed protein binding site (the site of identity with 16S rRNA), which, in turn, results in structural alteration of the hairpin IV and V bifurcation region (crosslink region).

To test the last hypothesis concerning the possible role of region II in stabilization of the hairpin bifurcation region, the mRNA61 variant was synthesized with three additional G-C pairs to stabilize the basement of deleted ICR (AG fragment). This fragment contained one of the supposed protein binding sites [14]. Moreover, one more variant was synthesized, for which the authentic noncanonical A-G pair (–80/–22) was replaced by an A-U pair to eliminate the helix defect (AU fragment). Neither of the fragments (neither AG nor AU) bound r6hEcoS7 protein. This means that the role of duplex II is not limited to stabilization of intercistronic hairpin structure. More likely, region II is directly involved in the interaction with the protein.

Thus, the S7 protein binding to the *str* mRNA ICR requires two distal RNA regions: one in the hairpin bifurcation region and the other in duplex II in the basement of the intercistronic hairpin.

Comparison of three-dimensional models of *str* mRNA ICR and EcoS7 protein. The conclusion drawn in the previous section concerning the existence of two centers of protein S7 binding with the *str* mRNA ICR defined the necessity of design of the spatial model of the binary complex. For this aim, three-dimensional models of the protein and RNA were superposed: the X-ray model of homolog protein TthS7 and the spatial model of RNA ICR were designed on the basis of supposed secondary structure.

The detailed spatial structure of EcoS7 protein is now not known due to the low resolution of X-ray analysis of *E. coli* ribosomes [25]. Because of this, coordinates of homologous protein TthS7, whose structure has been determined at resolution below 2 Å (PDB code 1RSS), were used for modeling [21].

Modeling spatial structure of intercistronic region S12-S7 of E. coli str mRNA. Modeling spatial structure of the *str* mRNA ICR was based on the model of ICR secondary structure proposed by us earlier [16]. As already noted, two variants of region identity with 16S rRNA were found in the mRNA ICR primary structure [12, 14], and the second variant of the mRNA ICR identity with hairpin 43 of 16S rRNA [12] was considered as preferable (Fig. 2)

Table 3. Correlation of data on X-ray of *T. thermophilus* ribosome 30S subunits and 16S rRNA–protein S7 crosslinks obtained for 30S subunits of *E. coli* ribosomes in solution

No.	Crosslink with 16S rRNA/nucleotide	Crosslink with S7 protein/amino acid	Distance in Tth30S, Å	Reagent	Reagent dimensions, Å	References
1	16S rRNA	M115**	2.7	UV	0	[26]
2*	U1240	M115**	2.7	IT	5	[28, 29]
3	C1378	K75	3.2	IT	5	[29]

Note: Distances between macromolecules and nucleotides/amino acids in structure of 30S subunit of *T. thermophilus* ribosomes (code PDB 1RSS [21]) were estimated using Swiss PDB Viewer 3.7 program (http://au.expasy.org/spdbv/). Iminothiolane (IT) dimensions were determined by models designed using ACD/ChemSketch (www.acdlabs.com) program.

* A similar crosslink was identified in the small subunit of *B. stearothermophilus* ribosomes (Met115 Bst7) [26, 27].

** In papers by Urlaub et al. [26-29] Met115 is designated as Met114.

[16]. The RNA tertiary structure was modeled using the authors' program, for which coordinates of atoms of non-canonical structure in the supposed protein recognition site were obtained from data of X-ray analysis of 16S rRNA within the small ribosomal subunit of *T. thermophilus* [23].

When combining the two models of protein and RNA in space, their mutual orientation was chosen according to the following criteria. Only a single site of potential contact with the protein was identified in the mRNA ICR, it corresponding to the UV-induced protein crosslink [17] with U34 and U35 in the ICR, which are marked in Fig. 2 for mRNA103 and distinguished by black balls in Fig. 4.

Sites of contact of the protein with mRNA were not identified, but there are data mapping the RNA–protein crosslinks in the small *E. coli* ribosome subunit.

In the case of UV-irradiation of the small ribosomal subunit of *E. coli* ribosomes, S7 protein is among the first to be crosslinked (Table 3), the reaction involving methionine M115 [26]. It became possible to identify uracil U1240 in 16S rRNA crosslinked with S7 only in the case of irradiation in the presence of iminothiolane [27-29]. A similar crosslink (in the presence of iminothiolane) from the protein side was identified for the small subunit of *B. stearothermophilus* ribosomes (M115 Bst7) [26, 27]. The second protein site includes the β-loop residue K75 located near 16S rRNA and capable of crosslinking with C1378 in the presence of iminothiolane (Table 3) [29]. The two sites are distal relative to each other in the three-dimensional protein structure.

The model of interaction shown in Fig. 4 was chosen from possible orientations of two protein regions relative to supposed sites of binding with ICR of mRNA. Since in the case of direct UV irradiation of the small subunit, U1240 of 16S rRNA was activated and crosslinked with reactive M115 of S7 protein, it is highly probable that in the case of direct UV irradiation of the binary complex just the reactive M115 of S7 is crosslinked with U34 and U35 in ICR of RNA [17]. Therefore, M115 of S7 on the model of the binary complex is adjacent to U34 and U35 in the region of hairpin bifurcations (Fig. 4). The second RNA–protein contact can be formed by region II of ICR and β-loop of S7 protein. As seen in Fig. 4, there is very

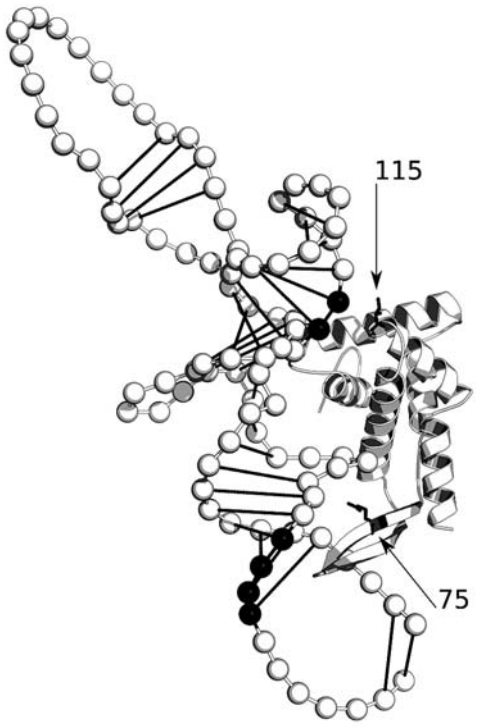

Fig. 4. Proposed model of spatial structure of S7 protein complex with EcoStr mRNA84. The structure of EcoS7 protein is taken from PDB (code 1RSS) [21]. Nucleotides are approximated by phosphorus atoms and shown by balls: black balls show two uridine residues crosslinked to the protein (in the center of the model) and four nucleotides of the Shine–Dalgarno sequence (at the bottom). Arrows point to amino acids crosslinked to 16S rRNA (Table 3). The three-dimensional structures are represented using the PyMOL program [24].

good steric correlation between two supposed contact regions of the protein and RNA in the model of the binary complex. The possible existence of a second contact between S7 protein and duplex II of mRNA explains the absence of binding of the protein to mRNA73 with deletion of this duplex.

Data on a possible role of the β-loop in S7 protein functioning do not contradict the suggested hypothesis. The spatial structure of TthS7 protein is homologous to the spatial structure of the HU/IHF family of DNA-binding proteins interacting with double-stranded DNA region, namely to β-loop [21, 30]. Besides, chromosomal deletion of a part of the R77-Y84 loop has no effect on regulation of *in vivo* expression of the *str* operon [31], whereas chromosomal deletion of the whole E73-E89 loop is lethal [32]; this points to functional significance of the E73-S76 region, for instance, of RNA–protein contact with K75.

The model agrees well with the mechanism proposed in the previous section for inhibition by S7 protein of coupled translation of two cistrons S12 and S7 by stabilization of ICR structure, which can cause abortive termination of cistron S12 translation.

Analysis of binding of *E. coli str* mRNA ICR deletion fragments to homologous protein r6hTthS7. The study of properties of heterologous RNA–protein complexes is of great interest because it makes it possible to compare a series of data on the RNA-binding activity of a set of homologous proteins from different sources. In such approach the amino acid sequence variations allowed by nature can be considered as an example of the protein's "natural mutagenesis" that, unlike directed mutagenesis, excludes essential alterations in protein structure. The drawback of such approach is also obvious—the variety of protein structures is restricted by their presence in nature. For heterologous complex the efficiency and specificity of RNA–protein interaction is defined by the protein structure variations in the contact zone as well as by alteration of the protein conformational mobility; the contribution of enthalpy and entropy components to the apparent dissociation constant is indicative of relative significance of these factors.

There are only few data in the literature concerning structure and properties of heterologous complexes. As to RNA–protein interactions in a ribosome, Nomura et al. showed already in their early works [33] the possibility of *in vitro* heterologous assembly of ribosomes using r-proteins and rRNA from different bacteria. Binary heterologous complexes are described for a number of r-proteins such as L1 [34] and S7 [18, 19].

In this work isotherms of r6hTthS7 heterologous binding to EcoStr mRNA fragments were obtained under conditions identical to those for complexes with Eco16S rRNA [11]. Binding isotherm linearization in Scatchard coordinates is shown in Fig. 3d. The apparent dissociation constant ($K_{d\ app}$) values and RNA portion active for protein binding are shown in Table 2. The following interesting regularities are revealed upon comparison of these parameters of homologous and heterologous complex formation of homologous proteins from mesophilic and thermophilic bacteria with the same series of *str* mRNA deletion fragments of mesophilic bacteria.

First, heterologous binding of thermophilic protein r6hTthS7 to *E. coli* ICR is only three times weaker than homologous r6hEcoS7 binding to *E. coli* ICR (Table 2). This means that the zone of RNA–protein contact is conservative.

Second, despite distinctions in absolute $K_{d\ app}$ values, RNA-binding ability of the two proteins is similar with regard to the series of *str* mRNA deletion fragments. Thus, EcoStr mRNA73 is not able to bind both homologous proteins. Perhaps the RNA–protein interactions for both homologous proteins can be described by the same kind of model.

Third, for the whole series of fragments the fraction of binding-competent RNA is much larger than for mesophilic proteins. Probably just the structural stability of thermophilic protein defines the efficiency of selection of the RNA target conformation from the total conformational landscape.

Design of single-center site in RNA for r6hTthS7 protein recognition. Detection of high potential for RNA–protein binding in thermophilic protein r6hTthS7 raises the following question: is it possible to design minimal RNA for thermophilic protein, for instance, on the basis of ICR having only one of the supposed recognition element in the region of hairpin bifurcation? To answer this question, a variant of a combinatorial method of selection (Systematic Evolution of Ligands by EXponential enrichment, SELEX) was used. The initial prototype of RNA structure was preserved including a part of the supposed conserved recognition site (III); five nucleotides of the site, directly adjacent to the hairpin bifurcation region (IV) and able to modulate protein binding (see below discussion of directed mutagenesis data), were randomized (Fig. 2). This resulted in generation of an RNA fragment mixture (RNA library), variations of base sequence of which were limited by a small "window". By analogy with the name of general SELEX method, this variant of selection was called SERW (Selection of Random RNA Window). A similar approach was successfully used previously for creation of RNA/DNA variants recognized by different proteins (cf. [35]).

In both variants of RNA fragments (AG and AU fragments), five nucleotides of the ICR double-stranded region IV, (–53)AAACU(–49), were randomized; these do not establish contacts with the proteins, but mutations in them influence binding to S7 protein (shown in the frame in Fig. 2, panel AG(AU)). To create the library, PCR from DNA template was used; the central part of this template was randomized, while the terminal sequences corresponded to DNA primers, one of which

contained a site of T7 promoter. An RNA library was obtained by transcription and used in cyclic selection for affinity to recombinant protein r6hTthS7.

A series of experiments was carried out for selection of two RNA libraries for isolation of so-called serwamers (from SERW), RNA fragments capable of efficient and specific protein binding. In this case the use for selected target products of the special term serwamer instead of the common term aptamer is defined by the following principal factor. In the case of the classical selection method (SELEX), aptamers are selected from the DNA/RNA library fragments in which base sequence is absolutely accidental (fully randomized). In this case the aptamer structure is defined and selected only by affinity to the target. In the case of the DNA/RNA fragment libraries having predetermined structure (or its framework, prototype) and in which only a small region is randomized, selection produces variants of predetermined structures. In such cases it is necessary to use terminology distinct from the classical one, like SERW and serwamer, respectively.

SERW was carried out according to methods earlier developed by us for a more complex selection variant [16]. The library of radioactive RNA fragments was obtained on DNA template using transcription by T7 RNA polymerase in the presence of α-[^{32}P]rNTP. After formation of a complex with recombinant protein, r6hTthS7 RNA–protein complexes were isolated by sorption on nitrocellulose membranes. RNA fragments were extracted from the membrane-bound complex, the cDNA copy was obtained, transcribed as described earlier, and used for the next selection cycle.

Selection to r6hEco7 protein. For homologous selection of the AG fragment library to mesophilic recombinant protein r6hEcoS7, two experiments were carried out under different conditions. In both experiments (1 and 2, Table 1) the buffer optimal for protein complex formation with 16S rRNA [11, 16] was used: 20 mM Tris-HCl, pH 7.6, 15 mM $MgCl_2$, 200 mM NH_4Cl, 4 mM mercaptoethanol. The two selection variants differed by the initial protein/RNA ratio because at high protein/RNA ratios serwamer losses are possible due to their extremely low representation in the initial library. Ten cycles of selection were carried out for each variant with gradual increase in the so-called selection strictness—the protein/RNA ratio changed in the range from 2 : 1 to 1 : 200 (Table 1). In both cases affinity of enriched RNA fraction to r6hEcoS7 protein did not change compared to the initial library.

Selection to r6hTthS7 protein. Due to significant identity to EcoS7 protein, protein TthS7 is capable of heterologous binding to fragments of EcoStr mRNA. And mentioned above, despite insignificant loss in affinity, thermophilic protein, unlike homologous mesophilic protein, is able to bind a large part of the RNA fraction, probably due to higher structural stability. These circumstances defined the choice of r6hTthS7 instead of r6hEcoS7 for selection of serwamers from the RNA library based on AG fragment. Conditions of complex formation corresponded to those chosen earlier for r6hTthS7 protein [11]. Ten selection cycles were carried out with gradual increase in selection strictness (experiment 3, Table 1). By the tenth cycle $K_{d\ app}$ of the enriched RNA fraction complex with r6hTthS7 protein was 87 ± 23 nM (Fig. 3).

The enriched library of the tenth cycle was cloned into pUC19 vector; the inserts into vector DNA of individual clones were sequenced. Seventy percent of the clones (13 out of 18) had identical base sequence designated as SERW (Fig. 2). This serwamer bound r6hTthS7 with high affinity: $K_{d\ app}$ = 89 ± 43 nM. It is interesting to note that in addition to randomized region, other regions of the serwamer RNA molecule changed as well: there appeared insertion of two U after C(–63) and deletion of U(–44) (the AG(AU) fragment, Fig. 2), i.e. mutations emerged in constant regions of the molecule. Mutations in the serwamer constant parts could be caused by different factors. Mutations can be the result of errors of chemical synthesis in the initial template DNA library; they can emerge due to accidental errors upon nucleic acid synthesis by polymerase with following selection due to high affinity to ribosomal protein.

To determine the factors responsible for emergence of mutations, two more selections were carried out to choose serwamers to r6hTthS7 protein under very strict conditions (Table 1) using two libraries—on the basis of AG and AU fragments. A significant enrichment of both libraries, comparable with results of the previous ten cycles, was achieved already at the fourth selection cycle: $K_{d\ app}$ of complexes of obtained enriched RNA fractions with r6hTthS7 protein were 89 ± 43 and 70 ± 21 nM for libraries on the basis of AG and AU fragments, respectively (Fig. 3 and Table 2). It appeared after cloning cDNA fragments into vector and sequencing individual clones of both libraries that the overwhelming majority of serwamers contain the same mutations in constant parts as the SERW serwamer from the first selection. Evidently alterations emerging in the constant part of the serwamer RNA upon selection do not depend on selection proper but are due to formation of specific structures with increased affinity to r6hTthS7 protein.

The proposed secondary structure of the obtained serwamer is shown in Fig. 2 and designated as SERW. Evidently mutations emerging upon selection can stabilize secondary structure of region IV, which should stimulate assembly of the hairpin bifurcation region and thus of the recognition site for S7 protein. The most surprising is the fact of coincidence of the character of found mutations with those observed earlier upon selection of RNA fragments of the complete (SERF) operon without any randomization; in the last case the same ICR region IV underwent mutation [16]. No doubt, such coincidence in

the character of mutations emerging upon different selection variants is not accidental. Most likely, mutations are necessary for creation of specific structural organization of the RNA-binding site of S7 protein.

Thus, the site of S7 protein binding to *str* mRNA is a small RNA hairpin in ICR that has two protein binding centers, the region of hairpin bifurcation and double-stranded region including the Shine–Dalgarno sequence of cistron S7. In the presence of additional stabilizing mutations interaction only with one binding site is possible, which is shown upon selection of RNA fragments for affinity to the protein. The possibility of creation and investigation of a set of RNA fragments for recognition by one protein opens new opportunities for the search for recognizable motifs in RNA.

The authors are grateful to A. A. Bogdanov for reading the manuscript and valuable comments.

This work was supported by the Russian Foundation for Basic Research (grants 07-04-01034a, 08-04-01540a, 08-04-90464).

REFERENCES

1. Kiemer, L., and Cesareni, G. (2007) *Trends Biotechnol.*, **25**, 448-454.
2. Lin, J., and Qian, J. (2007) *Exp. Rev. Proteom.*, **4**, 107-119.
3. Pommier, Y., and Marchand, C. (2005) *Curr. Med. Chem. Anticancer Agents*, **5**, 421-429.
4. Pommier, Y., and Cherfils, J. (2005) *Trends Pharmacol. Sci.*, **26**, 138-145.
5. Kaczanowska, M., and Ryden-Aulin, M. (2007) *Microbiol. Mol. Biol. Rev.*, **71**, 477-494.
6. Wilson, D. N., and Nierhaus, K. H. (2007) *Crit. Rev. Biochem. Mol. Biol.*, **42**, 187-219.
7. Nomura, M. (1999) *J. Bacteriol.*, **181**, 6857-6864.
8. Zengel, J. M., and Lindahl, L. (1994) *Prog. Nucleic Acid Res. Mol. Biol.*, **47**, 331-370.
9. Nowotny, V., and Nierhaus, K. H. (1988) *Biochemistry*, **27**, 7051-7055.
10. Kopylov, A. M. (2002) *Biochemistry (Moscow)*, **67**, 372-382.
11. Rassokhin, T. I., Golovin, A. V., Petrova, E. V., Spiridonova, V. A., Karginova, O. A., Rozhdestvensky, T. S., Brozius, Yu., and Kopylov, A. M. (2001) *Mol. Biol. (Moscow)*, **35**, 617-627.
12. Robert, F., and Brakier-Gingras, L. (2001) *Nucleic Acids Res.*, **29**, 677-682.
13. Lipin, M. Y., Stepanshina, V. N., Shemyakin, I. G., and Shinnick, T. M. (2007) *Clin. Microbiol. Infect.*, **13**, 620-626.
14. Saito, K., Mattheakis, L. C., and Nomura, M. (1994) *J. Mol. Biol.*, **235**, 111-124.
15. Saito, K., and Nomura, M. (1994) *J. Mol. Biol.*, **235**, 125-139.
16. Surdina, A. V., Rassokhin, T. I., Golovin, A. V., Spiridonova, V. A., Kraal, B., and Kopylov, A. M. (2008) *Biochemistry (Moscow)*, **73**, 652-659.
17. Golovin, A., Spiridonova, V., and Kopylov, A. (2006) *FEBS Lett.*, **580**, 5858-5862.
18. Spiridonova, V. A., Golovin, A. V., Drygin, D. Yu., and Kopylov, A. M. (1998) *Biochem. Mol. Biol. Int.*, **44**, 1141-1146.
19. Spiridonova, V. A., Rozhdestvensky, T. S., and Kopylov, A. M. (1999) *FEBS Lett.*, **460**, 353-356.
20. Karginov, A. V., Karginova, O. A., Spiridonova, V. A., and Kopylov, A. M. (1995) *FEBS Lett.*, **369**, 158-160.
21. Wimberly, B. T., White, S. W., and Ramakrishnan, V. (1997) *Structure*, **5**, 1187-1198.
22. Speshilov, G., and Golovin, A. V. (2008) RNA Tertiary Structure Modeling server on World Wide Web (http://12.192.230.173).
23. Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Morgan-Warren, R. J., Carter, A. P., Vonrhein, C., Hartsch, T., and Ramakrishnan, V. (2000) *Nature*, **407**, 327-339.
24. DeLano, W. L. The PyMOL Molecular Grapgics System (2002) on World Wide Web (http://www.pymol.org).
25. Berk, V., Zhang, W., Pai, R. D., and Cate, J. H. (2006) *Proc. Natl. Acad. Sci. USA*, **103**, 15830-15834.
26. Urlaub, H., Kruft, V., Bischof, O., Muller, E. C., and Wittmann-Liebold, B. (1995) *EMBO J.*, **14**, 4578-4588.
27. Urlaub, H., Kruft, V., and Wittmann-Liebold, B. (1995) *Methods in Protein Structure Analysis* (Atassi, M. Z., and Appella, E., eds.) Plenum Press, NY, pp. 275-282.
28. Urlaub, H., Thiede, B., Muller, E. C., Brimacombe, R., and Wittmann-Liebold, B. (1997) *J. Biol. Chem.*, **272**, 14547-1455.
29. Urlaub, H., Thiede, B., Muller, E. C., and Wittmann-Liebold, B. (1997) *J. Protein Chem.*, **16**, 375-383.
30. Hosaka, H., Nakagawa, A., Tanaka, I., Harada, N., Sano, K., Kimura, M., Yao, M., and Wakatsuki, S. (1997) *Structure*, **5**, 1199-1208.
31. Devaraj, A., Shoji, S., Holbrook, E. D., and Fredrick, K. (2009) *RNA*, **15**, 255-265.
32. Fredrick, K., Dunny, G. M., and Noller, H. F. (2000) *J. Mol. Biol.*, **298**, 379-394.
33. Nomura, M., Traub, P., and Bechmann, H. (1968) *Nature*, **219**, 793-799.
34. Kohrer, C., Mayer, C., Neumair, O., Grobner, P., and Piendl, W. (1998) *Eur. J. Biochem.*, **256**, 97-105.
35. Ikebukuro, K., Okumura, Y., Sumikura, K., and Karube, I. (2005) *Nucleic Acids Res.*, **33**, e108.