Journal of Biomolecular Structure & Dynamics, ISSN 0739-1102
Volume 6, Issue Number 3 (1988), ©Adenine Press (1988).

Prediction of the Recognition Sites on 16S and 23S rRNAs from *E.coli* for the Formation of 16S-23S rRNA Complex

T.A. Thanaraj, A.S. Kolaskar* and M.W. Pandit**
Centre for Cellular and Molecular Biology
Hyderabad - 500 007, India

Abstract

Interactions between RNA molecules have been postulated to play an important role in the assembly of ribosomes. Using the sequence analysis and the search of continuous complementary regions on 16S rRNA and 23S rRNA, the recognition sites involved in the formation of ribosome of *E.coli* are postulated. The number of postulated sites was narrowed down by taking available experimental data. The suggestive evidence for correct postulation is obtained from sequence comparison studies of 16S and 23S rRNAs from various species. The sites 891-899 and 1195-1203 on 16S rRNA along with the corresponding complementary sites 1904-1912 and 760-768 on 23S rRNA are predicted to be the most probable candidates for the sites of recognition between 16S and 23S rRNAs. The possibility of the involvement of the additional site 630-638 on 16S rRNA with its complementary site 2031-2039 on 23S rRNA cannot be ruled out.

Introduction

In order to understand the structure and the function of ribosomes, several attempts have been made in recent years using interesting experimental approaches and theoretical modeling (1-12). From these studies, an interesting possibility arises namely that primary interactions in the formation of a ribosome are the interactions between 16S and 23S rRNA molecules; the interactions between protein and rRNA molecules although important may not be sufficient alone without the intervention of nucleic acids. Further, the complex formed between 16S and 23S rRNAs being identical under given conditions suggests the involvement of specific sites on 16S and 23S rRNAs in the complex formation. It is, therefore, important to find out the sites of interaction on the 16S and 23S rRNA molecules which are responsible for the recognition of each other during the formation of the complex. Several chemical studies on 16S and 23S rRNAs of *E.coli* have indicated the involvement of nucleotides in the complex formation which are not in a base-paired state (3,4,13-16). The sites of recognition, we feel, should satisfy the following criteria: (i) oligonucleotides consisting

*Current Address: Dept of Zoology, University of Poona, Ganeshkhind, Pune 411 077, India

**Author to whom correspondence should be addressed.

of the recognition site(s) should be free for intermolecular interactions and, therefore, should not be involved in intramolecular base pairing, (ii) these sites should lie mostly on the surface to facilitate ready access to the other interacting partner, and (iii) the length of such oligonucleotides should be sufficiently long to generate a stable complex.

Thus, sufficiently long contiguous stretches of oligonucleotides which have a complementary sequence available on the other molecule will be the most likely candidates as recognition sites. Such long oligonucleotides, if there are any, will get base-paired with the complementary stretches on another molecule and form a stable complex through sufficiently large number of hydrogen bonds. Therefore, a method was developed to find out such long stretches of oligonucleotides and using available experimental data the recognition sites were predicted. Further, as these sites are expected to remain conserved during evolution, sequence comparison studies on 16S and 23S rRNAs from six species were carried out. The results discussed in the subsequent sections indicate that ours is probably the first prediction of the recognition sites on 16S and 23S rRNAs; and there are at least two such sites with a possibility of the third one, all having a length of nine nucleotides.

Method

A computer program - COMPLEMENT - was developed to find out continuous complementary stretches in a sequence as well as between two sequences of given nucleic acids. The program searches and arranges the complementary stretches in a decreasing order of length of oligonucleotides and avoids the stretches which have already appeared in the larger complementary oligonucleotides. The program was used to find the intra- and intermolecular complementary oligonucleotides of 16S and 23S rRNAs of *E.coli*. It was found that the maximum length of intermolecular complementary regions was of the order of nanonucleotides. A search was carried out to find which of these nanonucleotides are free from intramolecular complementation. The direct search as well as the search which made use of other simple criteria gave negative results. Therefore, these nanonucleotides were subjected to the following criterion in order to pick up the recognition sites. The oligonucleotide stretch in question should satisfy the dictates of experimental data obtained from chemical/biochemical studies such as Psoralen cross-linking, kethoxal modification, bisulphite reactivity, nuclease digestion, etc. (3,4,13-16); and thus, the oligonucleotide stretch should be free from intramolecular base-pairing so that it can have a potential to complement intermolecularly. Further, sequences of 16S and 23S rRNAs, from distantly related species (Table III) were aligned using the program ALIGN (17,18). Then the studies were carried out to check the extent of conservation of selected recognition sites in 16S and 23S rRNAs from the selected species.

Results and Discussion

It was found that the maximum length of intramolecular stretch in 23S rRNA was of the order of decanucleotide while in case of 16S rRNA, it was of the order of nanonucleotide. However, the search for intermolecular complementary region

Table I
Intra- and intermolecular complementation in rRNA molecules of *E.coli* at deca- and nanonucleotide level

Length	Complementary stretches between		
	16S-16S	23S-23S	16S-23S
10		(16-25) / (515-524) (71-80) / (1167-1176) (260-269) / (2643-2652)	
9	(392-400) / (1053-1061) (413-421) / (414-422) (430-438) / (431-439) (863-871) / (864-872)	(88-96) / (2472-2480) (95-103) / (1915-1923) (246-254) / (1291-1299) (519-527) / (520-528) (561-569) / (2432-2440) (728-736) / (2603-2611) (792-800) / (1460-1468) (1052-1060) / (2516-2524) (1074-1082) / (1977-1985) (1505-1513) / (2101-2109) (1590-1598) / (2187-2195)	(307-315) / (767-775) (630-638) / (2031-2039) (653-661) / (2001-2009) (783-791) / (2295-2303) (891-899) / (1904-1912) (1156-1164) / (33-42) (1195-1203) / (760-768) (1204-1212) / (2411-2419) (1404-1412) / (1823-1831) (1444-1452) / (1846-1854) (1464-1472) / (2564-2572)

showed that the maximum length of such oligonucleotides was of the order of nanonucleotide. Intra- and intermolecular complementary stretches found in 16S and 23S rRNA molecules at the level of deca- and nanonucleotides are given in Table I. The last column of the Table I indicates that there are 11 nanonucleotides of 16S rRNA which are having complementary stretches available on 23S rRNA. It can be seen from the table that none of the nanonucleotides which are potentially available for intermolecular complementation, has intramolecular complementary region of the same length. Therefore, these stretches can act as recognition sites, provided these oligonucleotides are in single stranded conformation and hence available for intermolecular complementation. When checked, all the 11 stretches of the 16S rRNA/23S rRNA were found to have intramolecular complementary oligonucleotides of length 4 and thus do not appear to be available completely for intermolecular complementation.

Therefore, we further scrutinized these sites by making use of experimental data already available in the literature which gives the information about the regions on 16S and 23S rRNAs which necessarily exist in a single stranded form. The results of this scrutiny are given in Table II. These results show that there are 3 sites on 16S rRNA which do exist entirely in a single stranded form (shown by single asterisk in Table II in which case all the 9 nucleotides of these stretches were found to be in a single stranded form experimentally). It was interesting to see that there also exist 3 sites (shown by double asterisk in the Table II) on 23S rRNA which are experimentally proved to exist in a single stranded form and, in addition are complementary to such sites in 16S rRNA. Thus out of 11 nanonucleotides on 16S rRNA for which complementary stretches are available on 23S rRNA, at least three pairs of nanonucleotides have a potential to act as sites of recognition between 16S and 23S rRNA for the formation of their complex. This cannot be said with certainty for other nanonucleotides.

Table II
Selection of potential recognition sites based on the experimental evidence

Stretch No.	Nanonucleotide stretch on 16S rRNA	No. of bases existing in a single stranded form (exptl. evidence)	Nanonucleotide stretch on 23S rRNA complementary to the stretch on 16S rRNA	No. of bases existing in single stranded form (exptl. evidence)
1	(307-315)	2	(767-775)	2
2	(630-638)*	9	(2031-2039)**	9
3	(653-661)	0	(2001-2009)	0
4	(783-791)	1	(2295-2303)	0
5	(891-899)*	9	(1904-1912)**	9
6	(1156-1164)	0	(33-41)	0
7	(1195-1203)*	9	(760-768)**	9
8	(1204-1212)	0	(2411-2419)	0
9	(1404-1412)	2	(1823-1831)	0
10	(1444-1452)	1	(1846-1854)	0
11	(1464-1472)	1	(2564-2572)	0

*, **, are the postulated sites of recognition on 16S and 23S rRNAs, respectively.

If the three pairs of nanonucleotides mentioned above are really the sites of recognition between 16S rRNA and 23S rRNA then these sites would play a crucial role in the assembly of ribosomes - an essential component of the protein-translation machinery; and, therefore, one would expect such sites to remain conserved in the rRNAs from various species even if they are evolutionarily distant from each other (3,28). In order to test the validity of this argument we have scrutinized these sites in case of 16S rRNA, by comparision through alignment of rRNA-sequences derived from 2 species each from 3 selected classes of organisms - chloroplast, eubacteria and archaebacteria. The choice of rRNAs was governed by the length of rRNA of these species which was kept approximately within +/− 5% of the length of rRNAs from *E.coli*. The analysis of 23S rRNA was restricted to only few species of rRNA because of such limitations. As can be seen from Table III, the analysis of conserved bases with respect to these sites points out that in all the species examined, the site 891-899 of 16S rRNA is almost invariant (7 out of 9 nucleotides) and the site 1195-1203 has at least 5 nucleotides conserved. In case of the site 630-638 although a few nucleotides were conserved, none of the nucleotides was common to all the species studied. The conservation of all nucleotides in the species of *Proteus vulgaris* (eubacteria) is understandable as it is well known that *E.coli* is evolutionarily closely related to eubacteria. Studies on the sites on 23S rRNA from *E.coli* with sequences from chloroplast and eubacteria also show that all the sites postulated to be recognition sites are almost conserved (at least 7 out of 9 nucleotides). Both sequence alignment studies and available experimental data clearly point out that there are at least two sites (891-899 and 1195-1203) on 16S rRNA which have a potential to specifically interact with the corresponding complementary sites on 23S rRNA to form a complex of 16S-23S rRNA in *E.coli*. However, there is a likelihood of the third site (630-638 on 16S rRNA) being involved in such interactions. All these postulated recognition sites are shown in Table IV. It is known (4,7,29) that there are several nucleotides which are involved in the interaction between 16S and 23S rRNAs. However, these nucleotides are quite distant from each other with respect to their positions in the

Table III
Data on the conservation of postulated recognition sites on 16S and 23S rRNAs

Recognition site on 16S rRNA from *E.coli*	No. of bases from the sites of *E.coli* 16S rRNA conserved in							Recognition site on 23S rRNA from *E.coli*	No. of bases from the sites of *E.coli* 23S rRNA conserved in			
	Chlor-oplast	Eubacteria		Archaebacteria			All species		Chlor-oplast	Eubacteria		All species
	Ma	Tb	Pv	An	Hv	Hm			Ma	Tb	Bs	
630-638	2	2	9	3	2	2	0	2031-2039	7	7	8	7
891-899	7	7	9	7	8	8	7	1904-1912	9	9	9	9
1195-1203	9	9	8	9	5	5	5	760-768	8	8	7	7
Reference	19	20	21	22	23	24			25	26	27	

Ma: Maize Tb: Tobacco; Pv: Proteus vulgaris; An: Anacystis nidulans; Hv: Halobacterium volcanii; Hm: Halococcus morrhua; Bs: Bacillus stearothermophillus.

Table IV
Postulated recognition sites (nanonucleotides) between 16S and 23S rRNAs

Site No.	Location		Sequence	
	16S	23S	16S	23S
1	630-638	2031-2039	5′-ACUGCAUCU-3′	5′-AGAUGCAGU-3′
2	891-899	1904-1912	5′-UACGGCCGC-3′	5′-GCGGCCGUA-3′
3	1195-1203	760-768	5′-CAAGUCAUC-3′	5′-GAUGACUUG-3′

Site No. 1 is tentative (see text).

sequence. Therefore, it should be kept in mind that although such nucleotides may play an important role in stabilizing the complex of 16S and 23S rRNAs, they may not be the primary sites of recognition.

In conclusion, it is possible by using theoretical approach supported by experimental evidence to predict the sites of recognition on 16S and 23S rRNA molecules of *E.coli* which are involved in the formation of their complex. Our studies indicate that there are atleast two sites (with a possibility of the third one) each on 16S rRNA as well as on 23S rRNA which can act as the sites of recognition between 16S and 23S rRNAs for the formation of their complex. This, therefore, suggests further the following theoretical framework for the various steps involved in the assembly of ribosomes: (i) freshly synthesized 16S and 23S rRNAs start folding into meta-stable structures which consist of long oligonucleotide stretches on their surfaces without getting involved into intramolecular hydrogen bonding; and, therefore, are available for intermolecular hydrogen bonding, (ii) when 16S and 23S rRNA molecules come closer, these specific sites on the surface recognise each other and form a stable complex through a large number of hydrogen bonds and stacking of base pairs in the form of short helical conformations, (iii) simultaneously, during the complex formation,

internal meta-stable structures of 16S and 23S rRNA molecules get altered and the complex of two nucleic acids assumes modified thermodynamically stable conformations, and (iv) out of such conformations, the complex then finally may get locked into one conformation through its interaction with various proteins and small nucleic acids involved in the ribosomal assembly. Involvement of specific recognition sites on 16S and 23S rRNAs in the formation of their complex can only be confirmed through experimental approach. In our forthcoming paper, which will be published elsewhere, we have proposed the secondary structure for the complex of 16S and 23S rRNAs, based on the graph theoretical approach.

The strategy used in building up the structures for such a complex takes into account the recognition sites postulated in the present paper.

Acknowledgement

We acknowledge the EMBL, Heidelberg, W.Germany for providing the nucleic acid sequence data.

References and Footnotes

1. K. Nagano and M. Harel, *Prog. Biophys. Molec. Biol. 48*, 67 (1986).
2. A. Liljas, *Prog. Biophys. Molec. Biol. 40*, 161 (1982).
3. R. Brimacombe and W. Stiege, *Biochem. J. 229*, 1 (1985).
4. H.F. Noller, *Ann. Rev. Biochem. 53*, 119 (1984).
5. A. Expert-Bezancon and P.L. Wollenzien, *J. Mol. Biol. 184*, 53 (1985).
6. A.S. Kolaskar, T.A. Thanaraj and M.W. Pandit, *J. Biosci. 8*, 791 (1985).
7. Y.G. Chu, P.L. Wollenzien and C.R. Cantor, *J. Biomol. Struct. Dyn. 1*, 647 (1983).
8. A.S. Spirin,in *Progress in Nucleic Acid Research and Molecular Biology*, Ed., W.E. Cohn and K. Moldave, Acad. Press, New York, Vol. 32, p. 75 (1985).
9. D.P. Burma, B. Nag and D.S. Tewari, *Proc. Natl. Acad. Sci. USA 80*, 4875 (1983).
10. D.P. Burma, A.K. Srivastava, S. Srivastava, D. Dash, D.S. Tewari and B. Nag, *J. Biosci. 8*, 757 (1985).
11. D.P. Burma, A.K. Srivastava, S. Srivastava, D. Dash, *J. Biol. Chem. 260*, 10517 (1985).
12. D. Dash, S. Mohanti and D.P. Burma, *J. Biosci. 11*, 561 (1987).
13. P. Maly and R. Brimacombe, *Nucleic Acids Res. 11*, 7263 (1983).
14. H.F. Noller and C.R. Woese, *Science 212*, 403 (1981).
15. W. Herr and H.F. Noller, *Biochem. 17*, 307 (1978).
16. P.L. Wollenzien, R.F. Murphy and C.R. Cantor, *J. Mol. Biol. 184*, 67 (1985).
17. S.B. Needleman and C.D. Wunsch, *J. Mol. Biol. 48*, 443 (1970).
18. M.O. Dayhoff, R.M. Schwartz and B.C. Orcutt, in *Atlas of protein sequence and structure*, Ed. M.O. Dayhoff, NBRF, Washington, USA, Vol. 5 suppl. 3, p. 345 (1979).
19. Z.S. Schwarz and H. Koessel, *Nature 283*, 739 (1980).
20. H. Tohdoh and M. Sugiura, *Gene 17*, 213 (1982).
21. P. Carbon, J.P. Ebel and C. Ehresmann, *Nucleic Acids Res. 9*, 2325 (1981).
22. N. Tomioka and M. Sugiura, *Mol. Gen. Genet. 191*, 46 (1983).
23. R. Gupta, J.M. Lanter and C.R. Woese, *Science 221*, 656 (1983).
24. H. Latters and R.A. Garrett, *EMBO J. 3*, 1613 (1984).
25. K. Edwards and H. Koessel, *Nucleic Acids Res. 9*, 2853 (1981).
26. F. Takaiwa and M. Sugiura, *Eur. J. Biochem. 124*, 13 (1982).
27. J. Kop, V. Wheaton, R. Gupta, C.R. Woese and H.F. Noller, *DNA 3*, 347 (1984).
28. H.F. Noller, J. Kop, V. Wheaton, J. Brosius, R.R. Gutell, A.M. Kopylov, F. Dohme, W. Herr, D.A. Stahl, R. Gupta and C.R. Woese, *Nucleic Acids Res. 9*, 6167 (1981).
29. W. Steiege, J. Atmadja, M. Zobawa and R. Brimacombe, *J. Mol. Biol. 191*, 135 (1986).

Date Received: May 4, 1988

Communicated by the Editor R.H. Sarma