Title
STRUCTURE-FUNCTION RELATIONS IN E. coli 16S RNA

Permalink
https://escholarship.org/uc/item/4x49p3pb

Authors
Thompson, J.F.
Hearst, J.E.

Publication Date
1983
Submitted to Cell

STRUCTURE-FUNCTION RELATIONS IN E. coli 16S RNA

John F. Thompson and John E. Hearst

January 1983

TWO-WEEK LOAN COPY
This is a Library Circulating Copy which may be borrowed for two weeks. For a personal retention copy, call Tech. Info. Division, Ext. 6782.
DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.
STRUCTURE-FUNCTION RELATIONS IN E. coli 16S RNA

John F. Thompson* and John E. Hearst

Laboratory of Chemical Biodynamics
Lawrence Berkeley Laboratory
and the
Department of Chemistry
University of California, Berkeley
Berkeley, California 94720

January 1983

*Present Address: Division of Biology & Medicine
Brown University
Providence, RI 02912

This manuscript was printed from originals provided by the authors.
INTRODUCTION

As a more thorough understanding of RNA secondary and tertiary structure has developed over the last several years, so has an appreciation of its importance in the function of the various RNP particles of which it is a part. For example, U1 RNA has been linked to mRNA splicing (Lerner et al., 1980; Rogers & Wall, 1980) while the RNA moiety of RNase P has been shown to be absolutely required for its activity (Kole et al., 1980). RNA has even been shown to be capable of making and breaking phosphodiester bonds in the complete absence of protein (Kruger et al., 1982). Along with these developments, the concept of ribosomal RNA being merely a framework on which ribosomal proteins can carry out their functions has been discarded. Indeed, speculations on the evolution of the protein synthesizing system have generally concluded that the RNA must have predated the protein components. The similarity in structure of protein-free 16S RNA in solution and 16S RNA in the 30S subunit (observed with psoralen crosslinking by Wollenzein et al., 1979; Thammana et al., 1979; Thompson & Hearst, 1983 and with electron microscopy) suggests that at least vestiges of the original catalytic structure remain. While E. coli rRNA may no longer be able to carry out protein-free translation, it is now generally accepted that it plays an active role in ribosomal functions. Unfortunately, the dearth of structural information has allowed formulation of only simple models for how RNA operates.

Even though the sequence of 16S RNA is known (Brosius et al., 1978; Carbon et al., 1979) and much of its secondary structure is agreed on (Noller and Woese, 1981; Stiegler et al., 1981; Zwieb et al., 1981), little progress has been made towards linking specific structures with function. Recent work with psoralen crosslinking of 16S RNA (Thompson & Hearst, 1983)
has confirmed parts of the secondary structure and also provided evidence for new interactions which appear to be functionally important. In the following, we will discuss how these structural features may be related to specific ribosomal mechanisms. We will concentrate on E. coli 16S RNA but eukaryotic 18S RNA will also be presented when its function appears to be substantially different. Reference to most ribosomal proteins will be tastefully omitted, primarily because their interactions with the RNA are poorly understood but also because we have approached the problem with the bias that they modulate the activity of the RNA rather than being the principal driving force behind it.

**mRNA Binding**: The role of 16S RNA in recognizing and binding mRNA in the initiation complex is well-established (Shine & Dalgarno, 1975; Steitz & Jakes, 1975); but, in eukaryotes, the sequence which has been implicated in mRNA binding has been deleted. Because both eukaryotic and prokaryotic small subunits perform essentially the same functions, there should be some compensating interaction between 18S RNA and mRNA. There are no apparent similarities in either the primary or secondary structure of eukaryotic mRNAs that would provide a basis for this.

The interaction 950-956/1507-1513, located by the psoralen crosslink GPs 956 x 1506 (see Thompson & Hearst (1983) for nomenclature), brings together two highly conserved regions in E. coli 16S RNA. In prokaryotes and eukaryotes, there is a number of modified bases located in both these parts of the RNA. In E. coli, there are a m^2G and a m^5C present in the region 550
bases from the 3' end. In eukaryotes, these have been replaced by the hyper-modified base amψ (Youvan and Hearst, 1981). 950-956/1507-1513 is conserved in eukaryotes thus placing amψ spatially near what corresponds to the Shine-Dalgarno sequence of prokaryotes and suggesting that it may have a role in mRNA recognition. The modifications present on amψ allow it to make specific interactions with the m7G cap structure found at the 5' end of all eukaryotic mRNAs. The negative charge delocalized on the carboxylic acid group of amψ can stabilize the positive charge delocalized on the imidazole nitrogens. Simultaneously, the amino group of amψ can interact with one of the negatively charged phosphate groups. Additional, indirect evidence for this interaction was presented by Thompson (1982). Efforts to obtain a more direct, experimental basis for the amψ – m7G interaction are now underway.

The equilibrium between 950-956/1507-1513 and 1506-1515/1520-1529 may also be involved in more complex intersubunit contacts. Azad (1979) has proposed an interaction between 5S RNA and the same region of 16S RNA (1509-1517) that pairs with the region near m2Gm5C. There is no firm evidence for this interaction and it was suggested by Schnare and Gray (1981) that it is not universal. However, stable base pairing of 5S and 18S RNA in solution has been observed (Oakden et al., 1977). The in vitro complex formed between D. melanogaster 18S and 5S RNAs can be crosslinked by HMT and large amounts of 5S co-purify with D. melanogaster 18S RNA even after two rounds of sucrose gradient centrifugation using standard purification protocols (Thompson, 1982). The fact that 30S subunits which contain EPs 956 x 1506 are less able to form 70S ribosomes than other cross-linked subunits (Thammana et al., 1979) further suggests that 5S pairs with
16S through this interaction.

How these interactions might alternate through the ribosomal cycle is not clear. For instance, one interaction might only occur during initiation while the others might switch during elongation. Only crosslinking results from ribosomes irradiated at specific points in translation will clarify this situation.

**Proofreading and tRNA Binding:** The total error rate in translation is simply a sum of the error rates of its component reactions. The theoretical and practical problems involved in the analysis of translational fidelity are reviewed by Kurland (1980) and Yarus (1979). Best estimates place the total error rate from all factors at one misincorporation per $10^4$ amino acids. The only step in translation which cannot be expected to easily yield this level of discrimination is tRNA binding via the codon-anticodon interaction. The difference in binding energies of two tRNAs which contain partially degenerate anticodons for a single codon is far too small to expect such accurate reading. To account for this, a number of models have been presented, all of which involve reading the anticodon twice to multiply small differences in binding. The lack of experimental data has, up until now, prevented formulation of a detailed physical model of this process which satisfactorily accounts for what little is known.

No part of 16S RNA has been associated with a proofreading function. Several proteins, however, are known to be involved in regulating translational fidelity. Elongation factor Tu, S4, S11, S12, and S17 have all been shown to profoundly affect the error rate (Gavriloa et al., 1981, and references therein). The characteristics of one of the long range crosslinks
observed, GPs 625 x 1420, suggest that it might have a role in proofreading and tRNA binding. The region near 1420 has been implicated in binding of tRNA to the P site (Taylor et al., 1981) and the region near 625 is part of the S8 binding site. S8, while not necessarily binding tRNA directly, has an effect on the amount bound (Nomura et al., 1969). Both regions are highly variable as would be expected for a proofreading function. The proofreading process requires energy and each species will have different requirements for optimizing the advantages of increased accuracy with the disadvantages of energy loss. Thus, the details of proofreading should be different even among closely related species with larger differences upon going from mitochondria to prokaryotes to eukaryotes. The region of E. coli 16S RNA around 580-660 varies considerably among prokaryotes and has been deleted entirely by mitochondria. Virtually all of the nucleotides which have been inserted into eukaryotic 18S RNA are found in this region. A similar behavior is seen around 1420. Mitochondria have shortened that helical stem while eukaryotes have expanded it.

Intuitively, one might expect eukaryotes to require the lowest error rate and hence devote more of the 18S RNA to that task. Eukaryotes synthesize many more proteins than prokaryotes and are thus more sensitive to error-induced damage. Mitochondria, on the other hand, are almost free of proofreading constraints. All proteins synthesized are multiple copy and only a few different ones are made. Indeed, all proteins which could propagate errors, of ribosomal proteins (with a couple exceptions) and polymerases, are synthesized outside the mitochondria. In some cases, mitochondria only read two of the anticodon nucleotides (Heckman et al., 1980), so that it would not be
surprising if they were to delete all or part of the proofreading apparatus. We are unable to rely on experimental results because measures of in vivo translational fidelity are extremely difficult to do and have been restricted to studies which detect only one or a few different misincorporated amino acids in a protein (Edelman & Gallant, 1977). Even these studies have to be viewed critically because *E. coli* ribosomes can reject nascent peptides which contain an error (Caplan & Menninger, 1979). These oligo-peptides are broken down rapidly in the cell and thus are difficult to measure quantitatively.

Lake (1979) has proposed a detailed model for what he terms the R (recognition) site of tRNA binding. The anticodon is read once in the R site. A conformational change in the tRNA occurs to bring it to the A site where the anticodon is read again. A primary reason for placing the R site on the exterior of the 30S sub-unit is the location of several tRNA binding and proofreading proteins there (including S8). For the reasons mentioned earlier, the cyclic interaction of 612-617/623-628 \[ \rightarrow \] 620-626/1420-1426 appears to be ideally suited for involvement in the process of moving a tRNA from the R site to the A site.

Lake (1981) proposes that the conformational change which brings the tRNA to the A site would occur solely in the tRNA with the only contact to the ribosomal complex being at the anticodon. This seems unlikely not only because of the weakness of some codon-anticodon interactions, but also because of the ease with which the process could be short-circuited. If the tRNA in the process of switching were to come off the mRNA, there would be nothing to prevent a new tRNA which had not undergone the initial screening at the R site from taking its place and moving into the A site. It is more likely that
there are multiple tRNA-protein and tRNA-rRNA contact points which ensure that the bound tRNA has all the important features of the cognate aminoacyl tRNA. In this way, other conformational changes in the ribosome could be tightly coupled to tRNA movement.

EF-Tu, which has been shown to recognize the 3' end of aminoacyl tRNA before binding the ribosome (reviewed by Weissbach, 1980), and the tRNA binding proteins on the exterior of the 30S subunit would make contact with tRNA bound to the R site. It would not be surprising if rRNA were also involved. The high variability of the 588-617/623-651 region suggests that it would not be directly involved in tRNA binding. There are, however, two nearby sequences of C-G-A-A that are highly conserved. Both of these stretches, located at 726-729 and 764-767 in E. coli, are present in all prokaryotes and eukaryotes and at least one is present in all mitochondria. C-G-A-A is complementary to the highly conserved TΨCG present in tRNA. While this sequence is not available for inter-RNA binding in solution, there is strong evidence that binding of a codon to tRNA makes this region more accessible (reviewed by Kim, 1978) and thus able to bind to 16S RNA or 5S RNA. Such an interaction would destroy contact between the D and TΨ loops of tRNA, also freeing other sequences for interaction. The conserved YGG sequence in the D loop could also be involved in interactions in the R site or it might remain free to allow specific binding to the A site upon switching.

There is a strongly conserved sequence in 16S RNA which would allow pairing of the Y-G-G sequence in the A site and subsequently in the P site. The sequence CCGm CmCCG (1399-1405 in E. coli) is present in all prokaryotes and eukaryotes and some mitochondria. In the A site, 1399-1401 would pair
with the exposed Y-G-G while the P site tRNA Y-G-G could pair with 1403-1405. This hypothesis is based partly on the data of Ofengand et al. (1982). They found that a modified base in the anticodon loop of a P site tRNA could crosslink to C-1400. This crosslinking was done with an empty A site so C-1400 would be available. When the proper codon was supplied for the tRNA (Ofengand & Liou, 1981), crosslinking was abolished. This indicates that the interaction between the anticodon and C-1400 is probably not functionally important. It does, however, establish that the P site is in very close proximity. The A site must also be very close because Johnson et al. (1982) found that the distance between the $^1$ position in tRNAs bound to the A and P sites is only 2-10 Å greater than the tRNA diameter. This implies that the A and P site tRNAs are in very close contact throughout their entire lengths because the anticodon loops and 3' ends also have to be quite close.

In order for the D loops to pair with 16S RNA as described, the tRNA would have to undergo a conformational change. The D and Tψ loops would have already separated in the R site. The D loop and stem would also have to twist slightly and move toward the anticodon loop. In such a conformation, the bases in the anticodon and D loops are in an antiparallel configuration. The A site is on the 3' side of of the mRNA; thus it is on the 5' side of the 16S RNA. A diagram of how this might happen is shown in Figure 1. The mRNA is necessarily kinked so the anticodon loops can base pair to adjacent codons on the mRNA. There is a larger distance allowable between the D loops because the base-paired regions are separated by an unpaired, modified nucleotide. The base and sugar methylations might somehow stabilize this kink.

Except for a few mitochondrial tRNAs and tRNAs involved in cell wall
synthesis, the sequence G-G is present in the same location in the D loop. The base 5' to this is usually a C, D, or U but occasionally an A. The pairing of G-1401 or G-1405 with this variable base holds additional potential for distinguishing between different tRNAs. The structures of G-C, G-U, G-D, and G-A pairs are all different and may change the orientation of a tRNA enough that, depending on the remainder of the structure, it could stabilize cognate and destabilize non-cognate tRNAs.

Translocation and Elongation: Very little is known about the mechanism of translocation and which parts of the ribosome are involved. The following model for translocation and movement of tRNAs through the ribosome is presented not as a definitive statement on how the ribosome works but as a way of accounting for our observations and those of other workers in the field. There is clearly much work to be done and this model should help point out weaknesses in our knowledge.

The role of tRNAs in translocation appears to be paramount. The distance which mRNA moves is determined by the tRNA (Thach & Thach, 1971; Gupta et al., 1971). Johnson et al. (1982) have also proposed that the energy for translocation comes from energy stored when the A site tRNA is tightly packed adjacent to the P site tRNA. In order for the tRNAs to be in such close contact, there must be other parts of the ribosome which prevent the tRNAs from escaping. Since this would necessarily have to be a cyclic interaction, RNA-RNA interactions seem likely to be involved. While intersub-unit interactions could play a role in this, we have no information which relates to this. A large part of the tRNA binding sites are localized on the 30S sub-
units so 16S RNA could certainly play a major role.

Brimacombe (1980), while not setting forth a specific model, proposed that 39-47/393-402 \(+\) 1055-1065/1186-1195 \(\rightleftharpoons\) 385-399/1052-1067 might somehow be involved in translocation. There is no direct evidence for Brimacombe's proposal but this seems to be exactly the type of cyclic, long range interaction necessary if translocation is to proceed as described above. The interactions shown in Figure 2 are certainly intricate enough to lock the P site tRNA in place. At the opposite end of the 30S sub-unit, similar inter-interactions would have to occur to lock the A site tRNA in place. These could include 950-956/1507-1513 and other interactions such as EPs 450 x 1540, EPs 510 x 1540, or EPs 0 x 1540 which have been mapped by electron microscopy (Wollenzein et al., 1979; Wollenzein & Cantor, 1982) but not known to sufficiently high resolution to describe in detail.

A chart showing our proposed model for the elongation process is shown in Figure 3. The role of elongation factors and conformational changes is shown, with the occupancy of the tRNA binding sites after each event listed.

When EF-G'GTP binds to the pretranslocation ribosome, it destabilizes the long range interactions and causes the short range base pairing of Figure 2B to occur. This provides an escape route for the P site tRNA. This tRNA is rapidly expelled from the P site because of electrostatic repulsion from the A site tRNA. The A site tRNA moves to the P site simultaneously because of the much greater affinity of peptidyl tRNAs for the P site. After tRNA movement, EF-G'GDP dissociates from the ribosome, catalyzed by GTP hydrolysis. The stabilization by EF-G of short range interactions is no longer a
factor as to which conformation is favored; thus the interdomain interactions shown in Figure 2 are re-established.

Evidence for an additional binding site for deacylated tRNA after the P site was found by Wettstein & Noll (1965). Additional support for this site, as well as a functional rationale for it, was provided by Nierhaus et al. (1980) and Rheinberger et al. (1981) on the basis of filter binding studies. Velocity sedimentation, however, has yielded ambiguous results on this point (Schmitt et al., 1982; Grajevskaja et al., 1982). Whether this is caused simply by differences in ribosomal preparations or by more serious inconsistencies is not clear. In any case, the D site (the E site has been renamed the D site for acronymic reasons) has been included in our model because, at present, the weight of evidence suggests it is real. The magnitude of the dissociation constant may be strongly dependent on the high method of ribosome preparation but even a dissociation rate might enhance accuracy in vivo. The presence of the deacylated tRNA in the D site accelerates the binding of the aminoacyl tRNA-\( \text{EF-Tu} \cdot \text{GTP} \) ternary complex to the ribosome. This occurs after an initial reading of the anticodon of the incoming tRNA. Once the anticodon has been interpreted as correct, \( \text{EF-Tu} \cdot \text{GTP} \) binds with high affinity to the short range interactions near the A site (946-955/1225-1235 + 1506-1515/1520-1529) and allows the R site tRNA to move into the A site. In the presence of \( \text{EF-Tu} \), this movement is irreversible and provides the non-equilibrium situation necessary for true proofreading to occur (Yarus, 1979; Kurland, 1980). Once in place, the tRNA anticodon is reread. If still deemed correct, \( \text{EF-Tu} \cdot \text{GDP} \) dissociates from the ribosome with hydrolysis of GTP. This allows the long range interactions to
reform (950-956/1507-1513) and locks the two tRNAs in place, correctly positioned for peptidyl transfer.

This model is necessarily incomplete but does account for all the data available on elongation at present. For instance, while four tRNA binding sites are proposed, only 2 or 3 are occupied at any one time. This agrees with the data of Rheinberger et al. (1981) who found 2 to 2.5 tRNAs bound during translation. It also includes the R and D sites which increase the fidelity of translation (Lake, 1981; Mierkow et al., 1980). The properties of non-cleavable GTP analogs in factor binding can be accounted for because the energy input is used solely to favor one direction in a conformational equilibrium. Non-enzymatic translation is possible because the same conformational equilibrium would be present in the absence of factors but would simply occur at a slower rate. In this case, the only energy input would be from peptide bond formation.

ACKNOWLEDGEMENTS

This work was supported by the Office of Energy Research, Office of Health and Environmental Research, Health Effects Research Division of the U. S. Department of Energy under Contract Number DE-AC03-76SF00098 and by the National Institutes of Health Grant # GM-11180.
FIGURE LEGENDS

Figure 1. Possible structure of two tRNA\textsubscript{s} interacting with both mRNA and 16S RNA in the A and P sites of the ribosome.

Figure 2. Possible conformational changes in 16S RNA. A) Long range interactions shown are supported by three different psoralen crosslinks (boxed) and one interaction found by Glotz et al (1981), 387-400/1053-1067. 1131-1144/1301-1317 is not present in any of the current models but similar structures can be drawn for other species. B) Short range interactions shown are from the newest version of the Noller & Woese (1981) model (H. Noller, personal communication).

Figure 3. Model for elongation and translocation. The model involves six steps, three of which involved structural transitions in the RNA. These transitions are described below.

1) Gate 1 is a steric barrier to tRNA movement at one end of the cleft in the 30S subunit. It separates the R and A sites and is associated with the following, and possibly other, transitions in secondary structure:

\[
\text{Gate 1 (closed) } \leftrightarrow \text{Gate 1 (open)} \]

950-956/1507-1513 \leftrightarrow 946-955/1225-1235 + 1506-1515/1520-1529

2) Gate 2 is a similar barrier at the other end of the cleft between the P and D sites. It is associated with the transition
described in Figure 2.

Gate 2 (closed) ↔ Gate 2 (open)

Figure 2A ↔ Figure 2B

3) The movement of the aminoacyl tRNA from the R site to the A site:

\[ \text{AA-tRNA (R site)} \leftrightarrow \text{AA-tRNA (A site)} \]
\[ 612-617/623-628 \leftrightarrow 620-626/1420-1426 \]

During the next step (tRNA expulsion), the reversal of this structural change in the 16S RNA occurs, leaving the aminoacyl tRNA in the A site and generating an empty R site.

In the above figure, four sites of tRNA binding are postulated. The R (recognition) site corresponds, in principle, to that described by Lake (1981). The physical attributes and position are not necessarily the same, however. The A and P sites are as usually proposed. The D (discharge) site corresponds to the E site of Rheinberger et al (1981). The name has been changed for acronymic reasons. AA refers to the aminoacyl tRNA, and D to the deacylated tRNA. The symbol \( \sim \) refers to the positions of the CCG sequences in the 16S RNA which base pair with the D loop of the tRNA (see Discussion).
REFERENCES


Grajewska, R. A., Ivanov, Y. V., & Saminsky, E. M. (1982). 70S Ribosomes of Escherichia coli have an additional site for deacylated tRNA


1795.


Figure 2
Figure 2
A Model for Protein Elongation

<table>
<thead>
<tr>
<th>R Site</th>
<th>Gate 1</th>
<th>A Site</th>
<th>P Site</th>
<th>Gate 2</th>
<th>D Site</th>
<th>Conformational Change</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>AA</td>
<td>P</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Peptidyl Transfer</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P</td>
<td>D</td>
<td></td>
<td></td>
<td>Translocation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>EFG·GTP binds</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Gate 2 opens</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A, P Site Regeneration</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>EFG·GDP dissociates</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Gate 2 closes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R Site Binding</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Gate 1 opens</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AA moves from R to A site</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>EFG·GDP dissociates</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R site regenerated</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Gate 1 closes</td>
</tr>
</tbody>
</table>

Figure 3
This report was done with support from the Department of Energy. Any conclusions or opinions expressed in this report represent solely those of the author(s) and not necessarily those of The Regents of the University of California, the Lawrence Berkeley Laboratory or the Department of Energy.

Reference to a company or product name does not imply approval or recommendation of the product by the University of California or the U.S. Department of Energy to the exclusion of others that may be suitable.