Bull. Inst. Chem. Res., Kyoto Univ., Vol. 65, No. 5-6, 1987

Possible Interaction Sites of Messenger RNA with Ribosome Deduced from Statistical Characterization of Base Sequences

Satoru TSUDA*, Tatsuo OOI*, and Kin-ichro Miura**

Received October 8, 1987

Base sequences near the initiation codon AUG were analyzed in terms of frequencies of singlet, doublet, and triplet at every position, and of specific base sequences. The characteristic features revealed here are :1) the leader sequence of prokaryotic and phage mRNA had a clear pattern of GG-rich and pyrimidine-poor character centered at the −10 region (i. e., SD sequence). 2) Another feature of the above mRNAs, —(A/U)—A—U—G—(G—C/A—A)— was found. 3) An optimal sequence for vertebrate mRNAs of —C—A—C—C—A—U—G—(G/A)— and that for eukaryotic virus of —C—A—A—(A/C/G)—A—U—G—G—A— were detected. 4) Other eukaryotes including plants and fungi had leader sequences of an A-rich and G-poor character. 5) Mitochondrial mRNAs appear to have an optimal sequence of —A—A—A—U—G—U—U—, implying that leader sequences are evidently different from those of prokaryote. 6) Chloroplast shows an A- and U-rich leader region and the typical SD-sequence was not commonly detected. 7) It is postulated that in the initiation step of the protein biosynthesis mRNA interacts with ribosome at two sites.

KEY WORDS: mRNA/ Initiation of translation/ Ribosomal RNA/ SD-sequence/

INTRODUCTION

On starting the biosynthesis of a polypeptide chain from a correct initiation position of an mRNA, some signal site in the mRNA must interact with ribosome to recognize the given position. This signal site must be located near the leader sequence region just before the initiation codon of a cistron in mRNA. The Shine-Dalgarno sequence (1) in a prokaryotic mRNA is an example; interaction of the sequence to the 3'-terminal part of prokaryotic 16S ribosomal RNA would be the initiation step as suggested also from several experiments (2-9). Since data on mRNAs have been accumulating during the last few years, we can collect such leader sequences from a DNA data bank, and analyze the statistical characteristics of the sequences in different types of organisms or organella as reported before (10). Because the increase in the nucleotide sequence data has been remarkable since then, we were able to perform more detailed analysis using accumulating data.

COLLECTION OF DATA

GenBank (release 32) was available to examine RNAs from prokaryotes, phages, mammals, other vertebrates, invertebrates, viruses, fungi, plants, and chloroplasts and mitochondria.

* 津田 諭, 大井龍夫, 三浦謹一郎 : Laboratory of Physical Chemistry of Enzyme, Institute for Chemical Research, Kyoto University, Uji 611.
** Department of Industrial Chemistry, Faculty of Engineering, University of Tokyo, Hongo, Japan.

The data from mammals and other vertebrates were combined into one as vertebrate, because the statistics of other vertebrates was very similar to that of mammals. The numbers of leader sequences collected were 374 for prokaryotes, 206 for phages, 400 for vertebrate, 59 for invertebrates, 476 for viruses, 27 for plants, 17 for fungi, 23 for chloroplasts, and 93 for mitochondria. From each mRNA, 60 bases upstream and downstream of the first letter of the initiation codon, which was determined experimentally, were collected, i.e., hypothetical mRNAs deduced from open reading frames were not included. When a leader sequence was less than 60 bases, the vacant positions were set blank. Therefore, the number of data at upstream positions from about the position −10 is sometimes less than the total number of the leader sequences collected, and we made necessary normalization in the statistics to correct blank positions in the sequences. For eukaryotes, some genes have introns in the leader sequence region, which were omitted according to the information of the positions of exons, i.e., mature mRNAs were used for the analysis.

SINGLET FEATURES OF LEADER SEQUENCES

The frequency of base singlets (4 kinds of base) against a position upstream and downstream of the initiation codon AUG is shown in Fig. 1 for types of prokaryote (*A*), phage (*B*), vertebrate (*C*), eukaryotic virus (*D*), mitochondria (*E*), fungi (*F*), chloroplast (*G*), and plant (*H*). Since the numbers of leader sequences collected for the types of prokaryote, phage, vertebrate, and eukaryotic virus are more than 200, the statistics are more meaningful than the rest of the data. Although those of other types, such as invertebrate, plant, fungi, and of organella would be less reliable because of the smaller numbers (less than 100), the data of species ((*E*) to (*H*)) must also contain some aspect reflecting the leader sequences of the species, being an A-rich and G-poor character in contrast to those of prokaryote (*A*) and vertebrate (*C*). In addition to characteristics deduced from the patterns shown in Fig. 1, more significant features could be detected from the analysis of distributions of doublet and triplet in the compiled sequences as described below.

FEATURE OF PROKARYOTE AND PHAGE LEADER SEQUENCES

The distribution of the bases along the prokaryotic sequences has a maximum of G, and a minimum of pyrimidine bases near the position −10 and an A-rich region appears before the initiation codon. This characteristic pattern corresponds to the Shine-Dalgarno sequence (1). The leader sequence may be characterized by the distribution of doublets (16 combinations using 4 kinds of base) as shown in Fig. 2 (*A*), where the frequency of doublet is plotted against a position near the initiation codon using the data for prokaryotes. The GG pair appears most frequently at the positions -9 and −10, the GA pair at the positions −7 and −8, and the AG pair at the positions −12 and −13, suggesting that an optimal sequence of AGGA is located in this region. The low frequency of pyrimidine bases near this region is consistent with the above statistics. In contrast to the GG-rich feature near the region centered at the position −10, the occurrence of G becomes low from the position −6 to the position −1, i.e., A and U are the major bases in the region just before the initiation codon. An optimal sequence in the region including the initiation codon, —(A/U)—A—U—G—G—C— or —(A/U)—A—U—G—A—

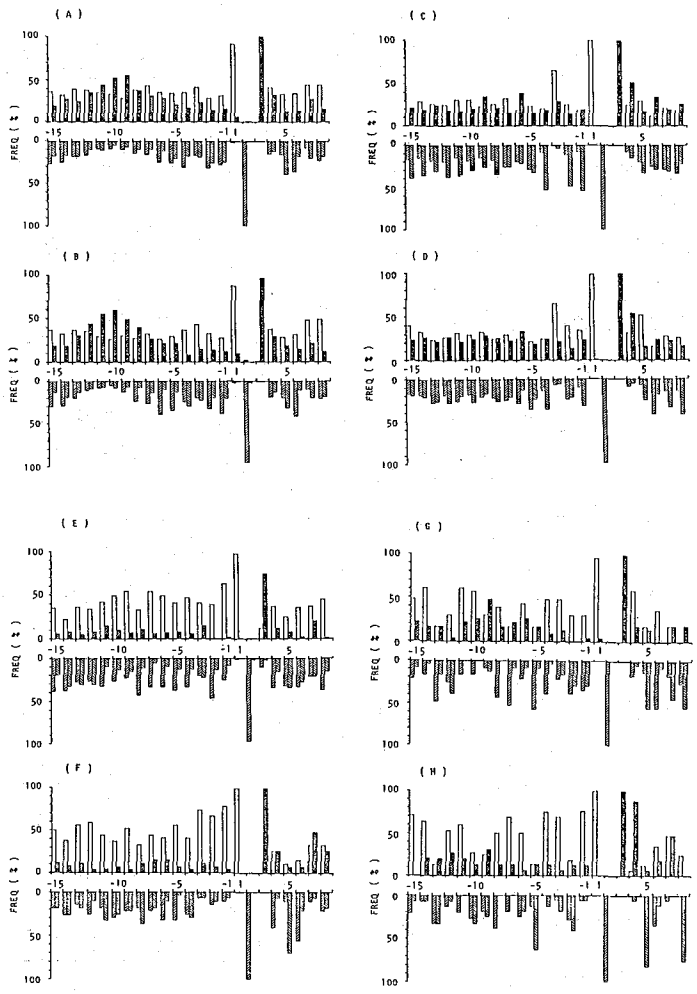

Fig. 1. Frequencies of occurrence of bases for prokaryote (*A*), phase (*B*), vertebrate (*C*), virus (*D*), mitochondria (*E*), fungi (*F*), chloroplast (*G*), and plant (*H*) are shown along the leader sequence (from the position −15 to −1), the initiation codon (from the position 1 to 3), and the following coding sequence (from the position 4 to 8). Those for adenine (A:□) and guanine (G:⊠) are shown in the upper half, and uracil (U:◰) and cytosine (C:◱) in the lower half.

A—, is deduced for prokaryotic mRNAs from the singlet and doublet data.

To confirm the results, frequencies of occurrence for triplets (64 combinations) were examined. Most of the patterns show a low level of the frequency as illustrated in Fig. 2 (*B*) for the case of GUC and UAC. Among patterns for all the combinations, those which contain any significant peak are demonstrated in Fig. 2(*B*). Since the peaks are at the position -10 for AGG, the position −9 for GGA, and the position −8 for GAG, the optimal sequence is deduced to be —A—G—G—A—C— from the position -12 to -8. Furthermore, a clear preference of GCU or AAA is shown immediately after the initiation codon.

The singlet pattern for leader sequences of phages is very similar to that of prokaryote (Fig. 1 (*B*)); the distribution of pyrimidine shows a sharper minimum near the position -10, and a

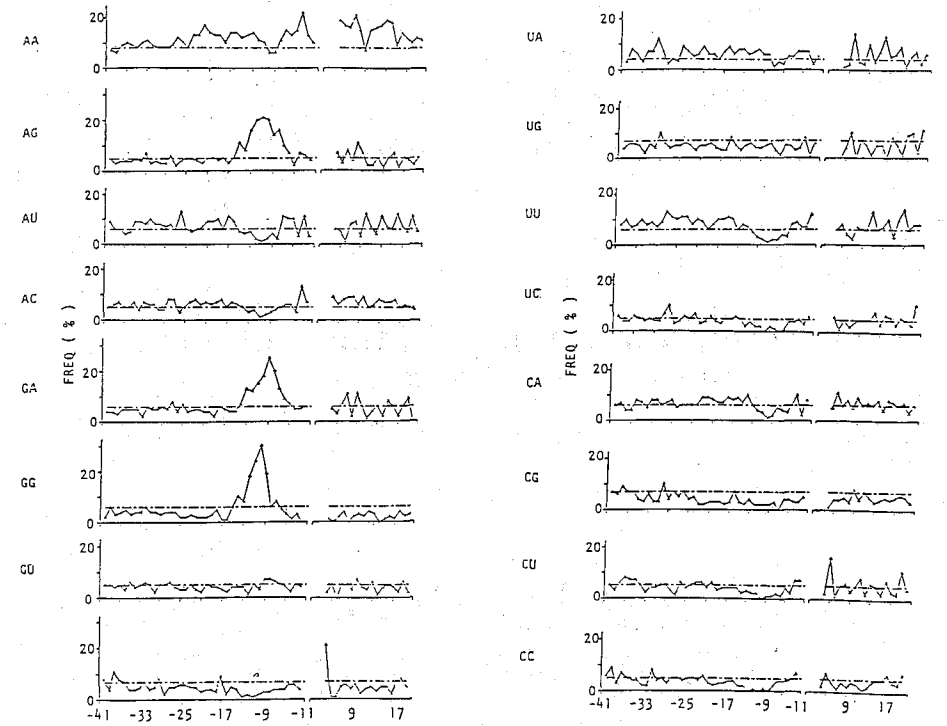
AA
AG
AU
AC
GA
GG
GU
UA
UG
UU
UC
CA
CG
CU
CC
FREQ (%)
20
0
-41 -33 -25 -17 -9 -11 9 17

Fig. 2 (A).

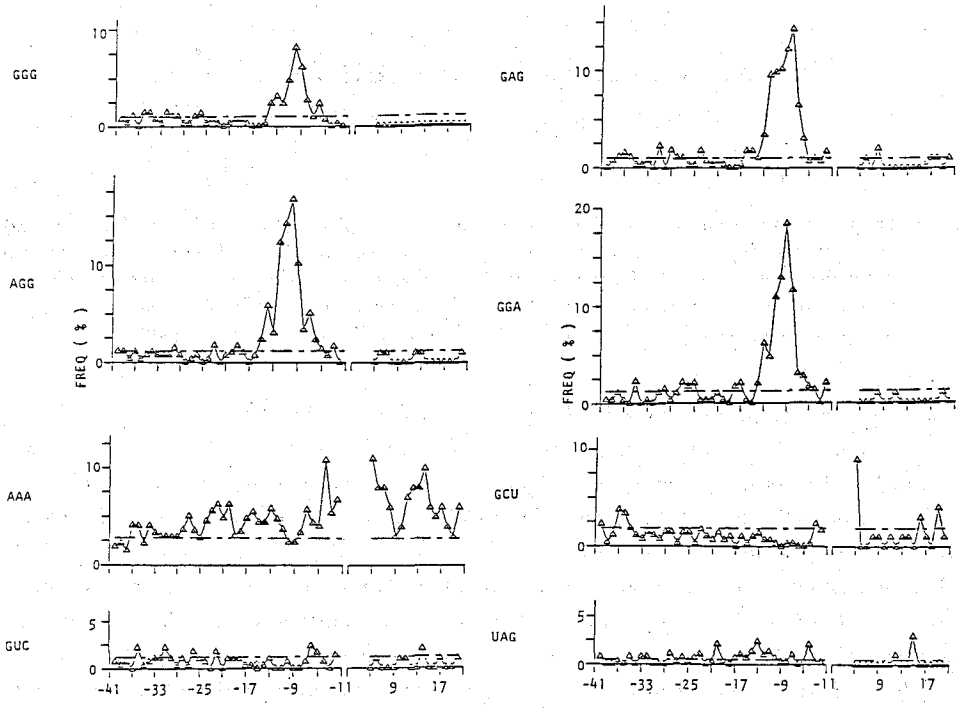
GGG
AGG
AAA
GUC
GAG
GGA
GCU
UAG
FREQ (%)
10
5
0
-41 -33 -25 -17 -9 -11 9 17

Fig. 2 (B).

maximum of G is located at the same position. From the doublet patterns (the figure is not shown), the frequency of the GG pair is found to be highest at the position −11 and still high at the position −10; frequencies of the AG and GA pairs are highest at the position -11 and -9, respectively. In the region from the position −13 to -9, there is almost no pyrimidine bases, i.e., the G-poor and A,U-rich feature in the downstream from the purine-rich region is very similar to that for prokaryotes. A small difference is the slight shift of the SD-region towards downstream compared with the pattern for prokaryote (Fig. 1 (*A*) and (*B*)). The triplet patterns, nevertheless, show the same characteristic feature for phage as that for prokaryote, i.e., the SD-sequence is deduced as —A—G—G—A—G— from the position −13 to −9 (one base shift compared with that for prokaryote), and a clear preference of GCU or AAA is found after the initiation codon.

FEATURES OF EUKARYOTE AND VIRUS LEADER SEQUENCES

The frequencies of singlets along the leader sequence for mature vertebrate mRNAs (Fig. 1 (*C*)) are close to random (or approximately 25%) except for a region from the position −4 to −1. This region has a preference for A at the position -3 and C at the positions −4, −2 and −1, indicating the characteristic of leader sequences of vertebrate mRNAs. The frequencies of doublets for the vertebrate leader sequences are demonstrated in Fig. 3 (*A*). We can recognize some characteristic features from the results; notably, the peaks which are observed for frequencies of AC, CA, and CC just before the initiation codon and those of GC, GA, and GU after the initiation codon. This implies that sequences near the initiation codon are specific; i.e., an optimal sequence —C—A—C—C—A—U—G—(G—C/A—A)— is deduced from patterns of the frequencies of singlets and doublets. Base frequencies at positions upstream from the position −6 did not show any significant character, and the occurrence seems random. In connection with this sequence, it is noteworthy to mention that the sequence —C—A—C—C—A—U—G— has been identified experimentally to be favourable to produce a good yield of a protein (14,15) in a eukaryotic system.

The singlet pattern for eukaryotic virus mRNAs is similar to that for vertebrates (Fig. 1 (*D*)); the difference is a preference of A at the positions -2 and -1 for virus, whereas C is more frequently located at these positions for vertebrate. The doublet frequencies for virus mRNAs show a clear difference in sequences near the initiation codon as domonstrated by patterns for the AA, AU, GA, and CC pairs in Fig. 3 (*B*). The optimal sequence deduced from the singlet and doublet frequencies is —C—A—A—X—A—U—G—(G/A)—A—U—, where X stands for A, C, and G.

Fig. 2. (*A*) Frequencies of occurrence of base doublets for prokaryote are shown along the leader sequence (from the position −40 to −2) and coding sequence (from the position 4 to 20). Doublets involving the initiation codon are not shown. Each horizontal chain line represents an average frequency of each doublet over the whole coding and noncoding region. (*B*) Frequencies of occurrence of base triplets for prokaryote are shown along the noncoding sequence (from the position −40 to −2) and coding sequence (from the position 4 to 20). The patterns for six triplets (AGG, GGG, AAA, GAG, GGA and GCU) show significant peaks, but those for the rest do not have any large peak as illustrated for two examples (GUC and UAG). Each horizontal chain line represents an everage frequency of each triplet over the whole coding and noncoding region.

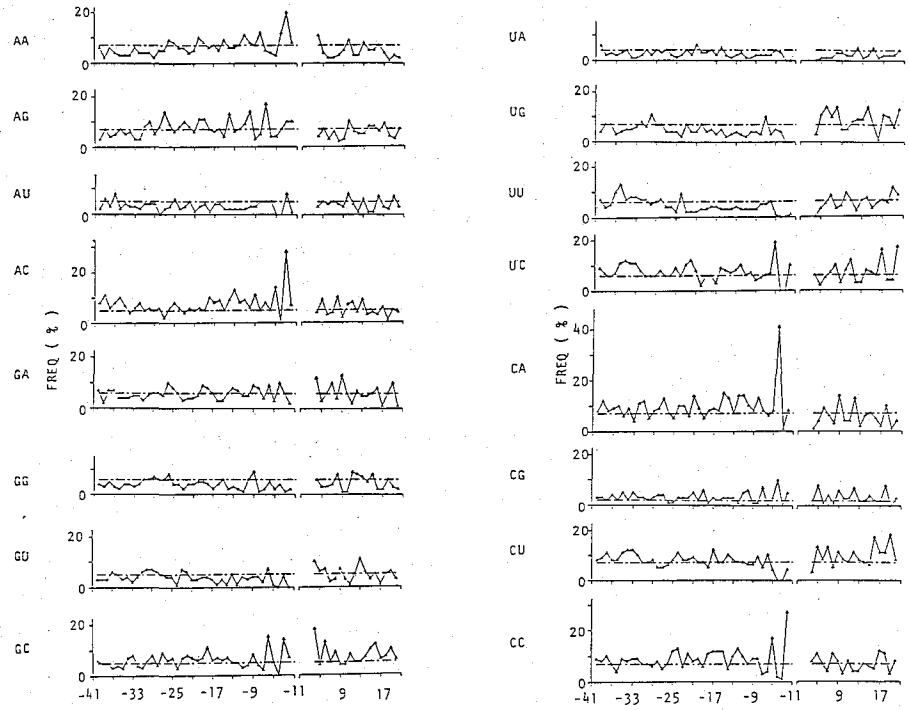
AA
AG
AU
AC
GA
GG
GU
GC
UA
UG
UU
UC
CA
CG
CU
CC
FREQ (%)
-41 -33 -25 -17 -9 -11 9 17

Fig. 3 (A).

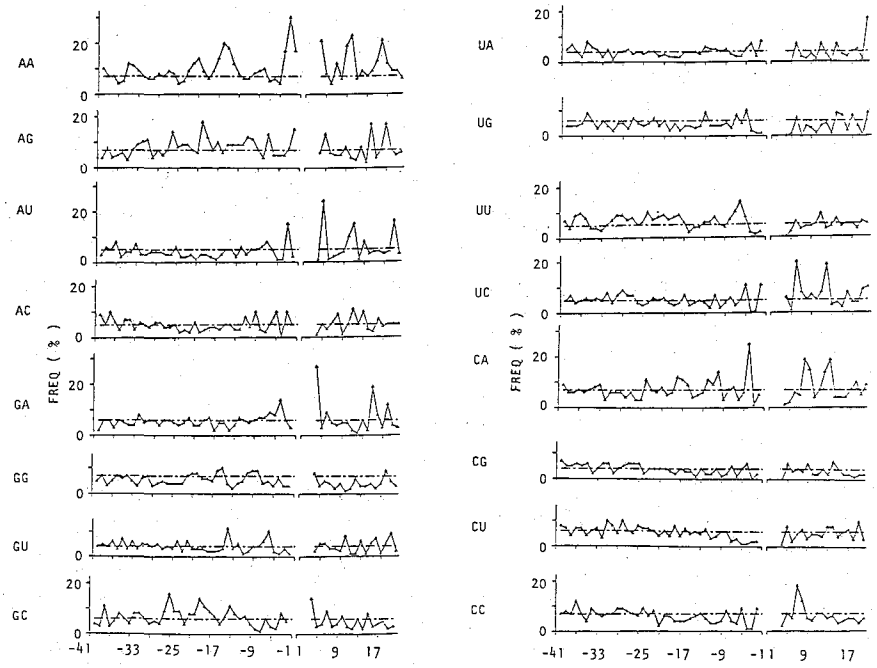
AA
AG
AU
AC
GA
GG
GU
GC
UA
UG
UU
UC
CA
CG
CU
CC
FREQ (%)
-41 -33 -25 -17 -9 -11 9 17

Fig. 3 (B).

FEATURES OF ORGANELLE LEADER SEQUENCES

Since generally only a few residues are located between a preceding termination codon and the next initiation codon in mitochondria, the length of leader sequences is very short. Therefore, the singlet frequency shown in Fig. 1 in the upstream region from the position −3 to −15 includes bases corresponding to the C-terminal region of the preceding protein. The frequencies of occurrence of doublets in the neighborhood of the initiaion codon for mitochondria were high for AA, AU, UA and UU, as expected from the singlet pattern in Fig. 1 (*E*). Since the length of leader sequences is short, the initiation step must occur with interaction at a site around the AUG; the characteristic sequence is deduced to be —(UA/AA)—A—U—G—(UU/AA)— from the frequencies of doublet. Apparently, this feature differs from those of prokaryotes and eukaryotes.

Although the number of data for chloroplast is too small to perform the statistical analysis, leader sequences of chloroplast mRNAs show a tendency to be rich for A and poor for G except at the position −9. The origin of only one peak for G at this position (Fig. 1 (*G*)) is due to the SD-like sequence found in some chloroplast mRNAs (16). This sequence, however, is not always found in the leader region of chloroplast mRNAs. Any notable feature was not found in a leader sequence in chloroplasts. This shows an evident difference from mRNAs of prokaryotes, which is condidered to be an origin of chloroplast from many aspects. The above features still hold when the number of data was increased about double using more recent results.

ENERGY ESTIMATION ON COMPLEX FORMATION OF LEADER SEQUENCES WITH 16S AND 18S rRNAS

The present characterization of leader sequences of mRNA for various organisms reveals their significant features for prokaryote, phage, vertebrate and eukaryotic virus types. Since some initiation complex of mRNA and ribosome is to be formed, interaction between a site in a leader sequence and the 3'-terminal part of prokaryotic 16S (or eukaryotic 18S) rRNA in ribosome may play a key role. We tried to calculate the smallest free energies originated from possible base pairings between the 3'-terminal part of rRNA of 20 bases long and a leader region of 40 bases long upstream of the initiation codon, for various leader sequences and both prokaryotic 16S and eukaryotic 18S rRNAs. If a leader sequence of a prokaryotic organism has a preference to bind 16S rRNA, the free energy computed by this method should be lower for 16S rRNA than for 18S rRNA and if there is some matching sequence in a eukaryotic mRNA, the binding to 18S rRNA would be more favourable than to 16S rRNA.

The results on prokaryote, eukaryote and its virus are listed in Table 1. For 81 sample

Fig. 3. (A) Frequencies of occurrence of base doublets for vertebrate are shown along the leader sequence (from the position −40 to −2) and coding sequence (from the position 4 to 20). Doublets involving the initiation codon are not shown. Each horizontal chain line represents an average frequency of each doublet over the whole coding and noncoding region. (B) Frequencies of occurrence of base doublets for virus are shown along the leader sequence (from the position −40 to −2) and coding sequence (from the position 4 to 20). Doublets involving the initiation codon are not shown. Each horizontal chain line represents an average frequency of each doublet over the whole coding and noncoding region.

Table 1. ENERGY ESTIMATION: 81, 67, and 46 leader sequences from prokaryote, eukaryote, and its virus, respectively, were aligned against 20 bases of the 3'-termianl part of 16S (11) and 18S (12) rRNA and the free energy due to base pairings for the alignment was calculated according to the procedure used for estimation of secondary structures in polynucleotide (13). The alignment was shifted by one base at a time to obtain the minimum free energy for the best matching of a leader sequence with the 3'-terminal part of 16S (and 18S rRNA). Numbers of leader sequences, which have lower free energies against the 3'-terminal regions of 16S rRNA (N_{16}) and of 18S rRNA (N_{18}), and numbers of those for which both free energies are the same ($N_{16=18}$), are listed for prokaryote, eukaryote, and eukaryotic virus.

Number	N_{tot}	N_{16}	$N_{16=18}$	N_{18}
prokaryote	81	44	18	19
eukaryote	67	22	18	27
virus	46	18	16	12

leader sequences of prokaryote, the 3'-terminal parts of 16S and 18S rRNA was found to interact at least with an energy of −4.4 Kcal/mol. The number of leader sequences which interact more strongly with the 3'-terminal part of 16S rRNA than that of 18S rRNA, was 44 out of 81; 18 had the same energy, and 19 favoured 18S rRNA. This is due to the specific purine-rich sequence shown in Figs. 1 and 2. Therefore, it is plausible that the leader part interacts favourably with the 3'-end of 16S rRNA. On the other hand, similar computations on 67 eukaryotic and 46 virus leader sequences did not show such a preference for a favourable interaction with 18S rRNA; 22 and 18 favoured 16S rRNA, 18 and 16 had the same energy, and 27 and 12 favoured 18S rRNA, respectively. Therefore, if we do not consider the formation of a secondary structure in rRNA to make a specific interaction site to a leader part or vice versa, a strong favourable interaction is not generally expected between a leader part upstream of the initiation codon and the 3'-terminal part of 18S rRNA. Probably this is the reason why any specific pattern of frequencies does not appear in the region from the position −20 to −5 as shown in Figs. 1 and 3. Since only a characteristic pattern is observed in the neighborhood of the initiation codon for leader sequences of vertebrate and virus mRNA, the interaction of this region to some site in 18S rRNA must occur in the initiation step for eukaryotes.

POSSIBLE BINDING REGIONS

The present results identify two characteristic sites in mRNAs; the SD-sequence for prokaryote and phage, and optimal sequences near the initiation codon for prokaryote, phage, vertebrate, and eukaryotic virus. These sequences must be a clue to find the location of interaction sites of mRNA and rRNA, when the data are compared with recent experimental results on the initiation step; e.g., formation of the complex and regulation of the efficiency (17, 18).

For prokaryotes and phages, the SD-sequence is usually accepted to be essential for the initiation step through interaction with the 3'-terminal part of 16S rRNA as supported by the present energy calculation. However, this site is not enough to account for the exceptional cases, e.g., as shown in Table 1, some leader sequences interact favorably with the 3'-terminal region of 18S rRNA. Here, we found another possible interaction site near the initiation

codon, i.e., —A—U—G—G—C—U— *or* A—U—G—A—A—A—. This suggests that the initiation codon itself is utilized as the initiation signal in the process of formation of an initiation complex of mRNA and ribosome; after the formation, the mRNA chain would move to the entry site of the ribosome. Since eukaryote and its virus have no sequence corresponding to the SD-sequence and the only possible site responsible for the interaction with rRNA is located near the initiation codon, the same or similar interaction site to that for eukaryotes is expected also for prokaryotes. If we assume that base pairings between nucleotides near the initiation codon and some part of 16S rRNA is the major interaction, the site should have a complementary sequence to the sequence, i.e., —A—G—C—C—A—U— or U—U—U—C—A—U—. The former site is found at the position 392–397 of 16S rRNA of E. coli (19) and Proteus vulgaris (20). Interestingly, this position is close to the 3'-terminus of 16S rRNA in the three dimensional structure of 16S rRNA according to the recent reports (21, 22); one in the 5'-terminal domain (D1 domain) and the other in the 3'-terminal domain (D4 domain). Therefore, the initiation complex for prokaryote could be achieved by two binding sites; the SD-sequence and the site including the initiation codon for mRNA, and the 3'-terminal part and the region about 400 bases downstream from the 5'-end of 16S rRNA, respectively (Fig. 4)

It is known in a eukaryotic system that the 5'-terminal cap structure of mRNA participates for the initial stage of binding of mRNA to ribosome (23, 24, 25, 26). Does the 5'-leader sequence of mRNA also participate in the binding step as in prokaryote? In eukaryote the 3'-terminal part of 18S rRNA lacks the sequence —C—C—U—C—C, which is the most important complementary sequence to the SD-sequence in the leader part of a prokaryotic mRNA (27). Instead, the 3'-terminal of 18S rRNA has the sequence —U—C—A—U—U—A

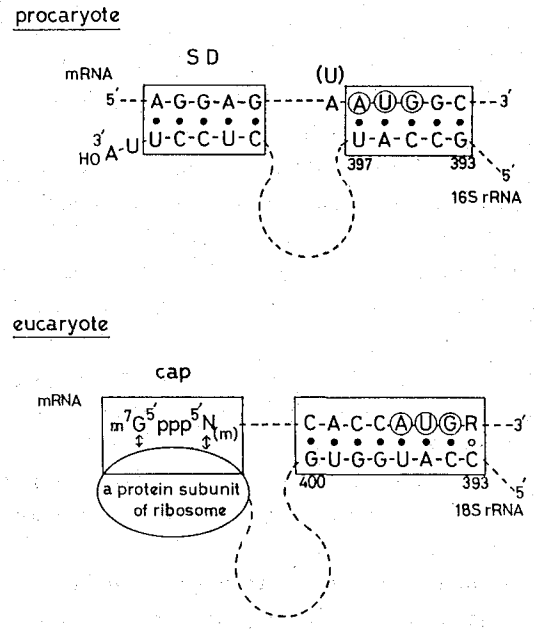

Fig. 4. Possible two binding sites are depicted schematically both for prokaryote and eukaryote. The cap site interacts probably with a protein subunit of ribosome shown by an ellipsoidal solid line. Dashed lines represent rRNAs.

OH, which contains the antiparallel complementary sequence to the initiation codon A—U—G(5' → 3') at the underlined C—A—U(5' → 3'). It is already suggested that this complementary relationship participates in the binding of mRNA to ribosome in eukaryote (28, 29, 30). It is notable that the next base of A—U—G, i.e., the position 4 is rich in purine, as it has a possibility to add more hydrogen bondings between this part (+1 to +4) in mRNA and the 3'-terminal of 18S rRNA to strengthen the pairings.

In the present study, the only characteristic sequence for vertebrate has been found near the initiation codon, and this sequence is —C—A—C—C—A—U—G—G—. The recent experimental results using single base substitutions near the AUG codon (15) indicate that this is the optimal sequence for the initiation. The initiation mechanism for vertebrate could be similar to that for prokaryote, because a complementary sequence, —C—C—A—U—G—G—U—G—, is found in D1 domain of 18S rRNA. The site is located at the position 393–400, 391-398, and 355-362 in 18S rRNA of rat (31), mouse (32) and *Xenopus* (33), respectively, and 16S rRNA does not contain this sequence. Since the possible site is located approximately at the same position as that for prokaryote, one may postulate that this site in 18S rRNA interacts with the initiation codon site to form an initiation complex. The specific sequence for eukaryotic virus —C—A—A—X—A—U—G—(G/A)—A, shows differences at the −1 and −2 position from that for vertebrate. The sequence resembles rather to that for invertebrate, —C—A—A—(A/G)—A—U—G—U—C, although the statistics are not so reliable because of the small number, 59, of leader sequences examined. By the experiments on the initiation complex formation in a eukaryotic protein synthesizing system, only the initiation codon AUG in an mRNA seems an essential region for binding with ribosome, and the 5'-terminal cap structure and the leader sequence, in some cases, enhance the formation of the initiation complex (34).

According to the present results, the sequence around the initiation codon A—U—G seems to be essential for the initial binding of mRNA with ribosome both in prokaryotes and in eukaryotes. In eukaryote, the 5'terminal cap of mRNA is initially important to approach ribosome (23,26,35). The next step is that the initiation codon A—U—G (or adding a few neighbouring bases) binds with ribosome by a specific part of rRNA. Kozak (36, 37) claims that mRNA attaches ribosome first at the cap, and then mRNA scans on the ribosome until A—U—G arrives at a proper position on ribosome. Alternatively, in the initial stage of the association of mRNA to ribosome in eukaryotes, two points binding by the cap and the sequence around the initiation codon in mRNA may occur as postulated here for prokaryotes (Fig. 4). Either scanning of mRNA or two points binding by the cap and the sequence around A—U—G occurs, the leader sequence also interacts with ribosome to modulate the efficiency of the formation of the initiation complex with mRNA and ribosome. A cross-linking experiments between two RNA molecules showed the presence of the binding between the leader sequence of mRNA and a part of rRNA (38). Furthermore, it is shown experimentally that the binding rate of formation of the initiation complex is dependent on the sequence of the 5'-leader part of mRNA (18).

The present two binding sites for the initiation complex both for prokaryotes and eukaryotes are deduced from the data analysis, but the postulate seems to be plausible because the analysis showed that the correct initiation of biosynthesis for a variety of organisms could not be exploited by a simple one sits binding.

ACKNOWLEDGMENT

This work was supported by research grants from the Ministry of Education, Science and Culture of Japan.

REFERENCES

(1) Shine, J. and Dalgarno, L. *Proc. Natl. Acad. Sci. US,* **71**, 1342–1346 (1974), *Nature,* **254**, 34–38 (1975).
(2) Spraque, K. J. and Steitz, J. A. *Nucleic Acids Res.,* **2**, 787–798 (1975).
(3) Steitz, J. A. and Jakes, K. *Proc. Natl. Acad. Sci. US,* **72**, 4734–4738 (1985).
(4) Steitz, J. A. and Bryan, R. A. *J. Mol. Biol.* **114**, 527–543 (1977).
(5) Steitz, J. A. and Steege, D. A. *J. Mol. Biol.* **114**, 545–558 (1977).
(6) Dunn, J. J., Buzash-Pollert, E. and Studier, F. W. *Proc. Natl. Acad. Sci. US,* **75**, 2741–2745 (1978).
(7) Taniguchi, T. and Weismann, C. *Nature,* **275**, 770–772 (1978).
(8) Taniguchi, T. and Weismann, C. *J. Mol. Biol.* **128**, 481–500 (1979).
(9) Eckhart, H. and Lührmann, R. *J. Biol. Chem.,* **254**, 11185–11188 (1979).
(10) Miura, K., Yamaguchi, K., Kohno, K., and Ooi, T. *Nucl. Acid Res. Symposium Series,* **11,** 121–124 (1982).
(11) Brosins, J., Palmer, M. L., Kennedy, P. L., and Moller, H. F. *Proc. Natl. Acad. Sci. UA,* **77**, 201–204 (1980).
(12) Salim, M. and Maden, B. E. H. *Nature,* **291**, 205–208 (1981).
(13) Ooi, T., and Takanami, M. *Biochim. Biophys. Acta,* **655**, 221–229 (1981).
(14) Morie, F., Lopez, B., Henni, T. and Godet, J. *EMBO J.* **4**, 1245–1250 (1985).
(15) Kozak, M. *Cell,* **44**, 283–292 (1986).
(16) Sugita, M., Kato, A., Shimada, H. and Sugiura, M. *Mol. Gen. Genet,* **194**, 200–205 (1984).
(17) Lomedico, P. T., and McAndrew, S. J. *Nature,* **229**, 221–226 (1982).
(18) Yamaguchi, K., Hidaka, S. and Miura, K. *Proc. Natl. Acad. Sci. US,* **79**, 1012–1016 (1982).
(19) Carbon, P., Ehresmann, C., Ehresmann, B. and Ebel, J.-P. *Eur. J. Biochem.,* **100**, 399–410 (1970).
(20) Carbon, P., Ebel, J.,-P., and Ehresmann, C., *Nucleic Acids Res.,* **8**, 2325–2333 (1984).
(21) Zwieb, C., Glotz, C. and Brimacombe, R. *Nucl. Acid Res.,* **9**, 3621–3640 (1981).
(22) Expert-Benzacon, A. and Wollenzien, P. C. *J. Mol. Biol.,* **185**, 53–64 (1985).
(23) Muthukrishnan, S., Morgan, M, Banerjee, A. K. and Shatkin, A. J. *Biochemistry* **15**, 5761–5768 (1976).
(24) Shimotohno, K., Kodama, Y. Hashimoto, J. and Miura, K. *Proc. Natl. Acad. Sci. US,* **74**, 2734–2738 (1977).
(25) Miura, K. *Adv. Biophys.,* **14**, 205–238 (1981).
(26) Shatkin, A. J. *Cell,* **9**, 645–653 (1976).
(27) Hagenbüchle, O., Santer, M., Steitz, J. A. and Mans, R. J. *Cell,* **13**, 55–563 (1978).
(28) Legon, S. *J. Mol. Biol.,* **106**, 37–53 (1976).
(29) Baralle, F. E. and Brownlee, G. G. *Nature,* **274**, 84–87 (1978).
(30) Baralle, F. E. *Cell,* **10**, 549–558 (1977).
(31) Chan, Y. L., Gutell, R., Noller, H. F., and Wool, I. G. *J. Biol. Chem.,* **259**, 224–230 (1984).
(32) Reynal, F., Michot, B., and Bachellerie, J.-P. *FEBS Lett.,* **169**, 263–268 (1984).
(33) Maden, B. E. H., Forbes, J. M., Stewart, M. A., and Eason, R. *EMBO J.,* **1**,597–601 (1982).
(34) Kohno, K., Sekine, M., Hata, T., Kumagai, I., and Miura, K. *Nature* (submitted).
(35) Zan-Kowalczekwska, M., Betner, M., Sierakowska, H., Szczesna, E., Fillipowicz, W., and Shatkin, A. J. *Nucleic Acids Res.,* **4**, 3065–3081 (1977).
(36) Kozak, M. *Nucl. Acid Res.,* **9**, 5233–5252 (1981).
(37) Kozak, M. *Cell,* **15**, 1109–1123 (1978).
(38) Nakashima, K. Darzynkiewiez, E. and Shatkin, A. J. *Nature* **286**, 226–230 (1980).