Volume 13 Number 15 1985 Nucleic Acids Research

Nucleotide sequences of two serine tRNAs with a GGA anticodon: the structure-function relationships in the serine family of *E. coli* tRNAs

H.Grosjean[1], K.Nicoghosian, E.Haumont[1], D.Söll[2] and R.Cedergren

Département de Biochimie, Université de Montréal, Montréal, Québec H3C 3J7, Canada, [1]Département de Biologie Moléculaire, Université Libre de Bruxelles, B-1640 Rhode-St-Genèse, Belgique, and [2]Department of Molecular Biology and Biophysics, Yale University, New Haven, CT, USA

Received 29 April 1985; Revised and Accepted 22 July 1985

ABSTRACT

We have determined the nucleotide sequence of the major species of E.coli $tRNA^{Ser}$ and of a minor species having the same GGA anticodon. These two tRNAs should recognize the UCC and UCU codons, the most widely used codons for serine in the highly expressed genes of E.coli. The two sequences differ in only one position of the D-loop. Neither tRNA has a modified adenosine in the position 3'-adjacent to the anticodon. This can be rationalized on the basis of a structural constraint in the anticodon stem and may be related to optimization of the codon-anticodon interaction. Comparison of all E.coli serine tRNAs (and that encoded by bacteriophage T_4) reveals characteristic (possibly functional) features. Evolutionary analysis suggests an eubacterial origin of the $T_4 tRNA^{Ser}$ gene and the existences of a recent common ancestor for the $tRNA^{Ser}_{GGA}$ and $tRNA^{Ser}_{GUC}$ genes.

INTRODUCTION

The degeneracy of the genetic code is such that most of the twenty amino acids are specified by more than one code word. In particular, serine is encoded by six codons arranged in two blocks that differ in the first and the second codon, i.e., the quartet UCN (where N = C,U,G or A) and the doublet (AGY where Y = C or U). In order to read these six codons during mRNA decoding, tRNAs with different anticodons are required. In mitochondria, only two tRNA isoacceptors are needed, since one having a UGA anticodon reads the entire family of UCN codons and another tRNA with a GCU anticodon reads the two AGY codons. In the eukaryotic or prokaryotic cytoplasm a minimum of four isoacceptors are needed. The analysis of more than 33 $tRNAs^{Ser}$ from different origins shows that there are mainly five anticodon types: IGA, U*GA, (U* is modified), GGA, CGA, and GCU. In addition to these isoacceptors (having different anticodons), other isoacceptor species have been found which differ in only the modified nucleoside levels (isocoding species) (for reviews see refs 1-2). In E.coli three types of $tRNA^{Ser}$ were sequenced; they had the anticodon U^*GA, GCU and CGA (3). The object of this article is to present the sequence of the fourth type of E.coli serine tRNA (GGA anticodon) which

reads UCC and UCU codons.

Some time ago, one of us (D.S.) reported the identification and the purification of five serine tRNAs from E.coli K12. A major species (called Ser I) and a minor (Ser V) exhibited a coding response to poly UC and were taken to have the GGA anticodon (4). Subsequently, these two tRNAs were shown to form anticodon-anticodon complexes with E.coli $tRNA_U^{*Gly}_{CC}$ and yeast $tRNA_U^{*Arg}_{CU}$ (5-6). We report here the complete primary structure of these two tRNAs. Since these sequences complete the structural work on the serine tRNA family of E.coli, it is now possible to consider the hypothesis concerning the role of these tRNAs in regulation and mRNA decoding.

MATERIALS AND METHODS

The two $tRNA^{Ser}_{GGA}$ (corresponding to $tRNA^{Ser}_{I}$ and $tRNA^{Ser}_{V}$) were isolated from E.coli K12, strain CA 244 as previously published (4). Due to the inconsistencies of nomenclature in the serine tRNAs (see Table 1), we will distinguish them by their anticodon sequence. The tRNAs have been maintained at -80°C, since that time.

3'-labelling of 1 µg samples was performed by the incubation of tRNA with [^{32}P]pCp (3000 Ci/mmole) and T4-RNA ligase (from P.L. Biochemicals) as described in (7). Purification followed on 0.08 cm thick, 50 cm long gels made of 15% polyacrylamide with 50 mM Tris-boric acid buffer at pH 8.3 containing 7M urea and 10 mM EDTA. Both tRNA samples gave a single radioactive band, and their sequences were determined by the now standard rapid gel sequencing techniques (8).

The modified nucleosides were determined by a variation of the Gupta and Randerath method (9). A sample of 5 µg of each tRNA was partially hydrolyzed in 5 µl of deionized formamide at 100°C for 4 min. Chromatographic analysis was done by the separation of nucleoside diphosphates on PEI-cellulose thin layer plates in 0.55 M ammonium sulfate at 4°C or by the separation of nucleoside monophosphates (produced by an additional digestion by P1 nuclease) on two-dimensional cellullose thin layer plates.

RESULTS AND DISCUSSION

The serine $tRNA_{GGA}$ sequences and their analysis

The two serine tRNAs are composed of 88 nucleotides, 16 of which are located in the long extra arm (Fig.1a). These tRNAs along with the E.coli $tRNA^{Ser}_{UGA}$ have the shortest extra arms among all sequenced serine tRNAs (3). The two tRNAs differ by one nucleotide; there is a C in position 20 of

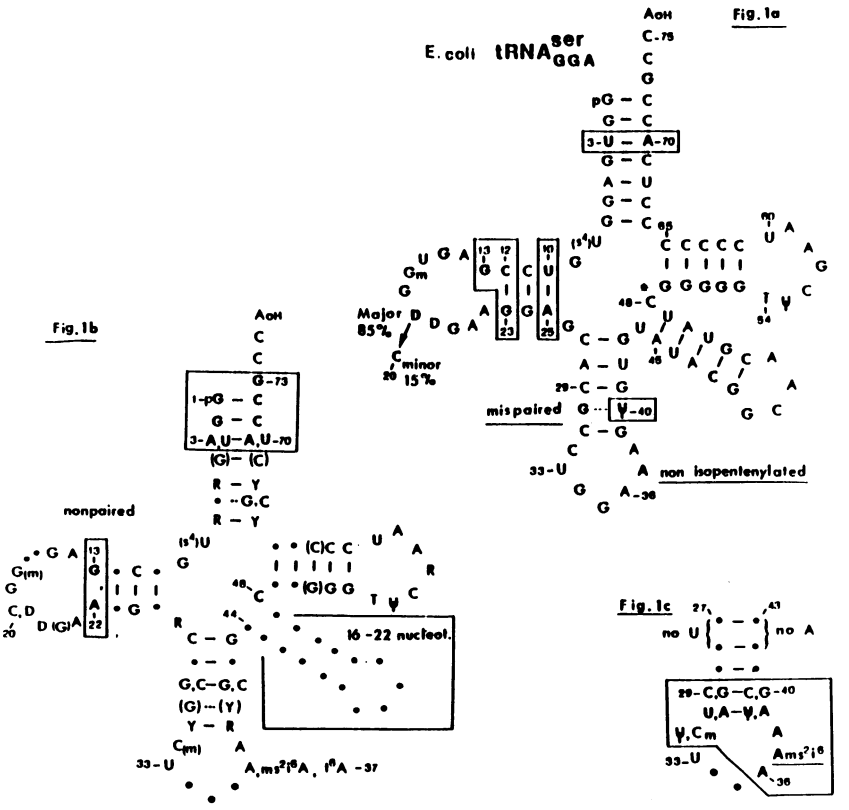

Figure 1. a) The two sequenced $tRNA^{Ser}_{GGA}$ isoacceptors: position 20 is D in major and C in minor species. A37 is unmodified. C* is a mixture of C and an unknown derivative. Boxed-in positions correspond to nucleotides which are unique in that they occur in less than 10% of all other eubacterial tRNAs (10).

b) Composite of E.coli $tRNAs^{Ser}$ (with T4 sequence): identical nucleotides are those found in all sequences. Y is pyrimidine, R is purine. Nucleotides in parenthesis are present in six out of seven sequences analyzed. Boxed-in regions indicate characteristic features of the serine tRNA family.

c) Consensus sequence in anticodon loops and stems of fifty eubacterial tRNAs having fully modified $ms^2i^6A_{37}$ (3,21).

$tRNA^{Ser}_{V}$ and a D in this position of $tRNA^{Ser}_{I}$. It is unusual although not altogether without precedence that one nucleotide change can so affect the mobility of tRNA on BD-cellulose chromatography (3). A more surprising finding is that neither of the two tRNAs contains a modified adenosine in the position 3'-adjacent to the anticodon. Often this A_{37} is modified to either a 6-threonylcarbamoyl (t^6A) or a 2-methylthio-6-isopentenyl derivative (ms^2i^6A) in E.coli (39) as in other serine tRNA isoacceptors (see Table I).

Table 1 : Codon and anticodon usage in serine tRNA

Species number	Relative amount (%)	Anticodon + 3'planning base	Expected recognized codon(s)	Frequency* H	Frequency* L	Gene symbol,map position
I (4)	40	G G A . A	U C C	16	9	unknown
V (4)	5	G G A . A	U C U	18	7	unknown
IV (4) I (36)	19	V G A . ms^2i^6A	U C A	1	7	serT, 16 min(28)
2 (27)	5	C G A . ms^2i^6A	U C G	2	13	serT, 43 min(37)
II (4) 3a(29,30)	13	G C U . t^6A	A G C	9	12	serW. 61 min(28)
III (4) 3b(29,30)	19	G C U. A	A G U	2	11	

Heavy letters correspond to "wobble" base of anticodon or the corresponding third letter of the codon. *Frequency of each codon is expressed per thousand codons for highly expressed genes (column H) and lowly expressed genes (column L) (24). Amount of serine tRNAs (VGA and GCU) relative to other tRNA isoacceptors of E.coli can be found in ref.(38). Numbers in brackets correspond to the reference list. V is uridine-5-oxyacetic acid; ms^2i^6A is N-6-(Δ2-isopentenyl)2-methylthioadenosine; t^6A is N-[9-(β-D-ribofuranosyl)purin-6-ylcarbamoyl] threonine.

Position 48 is likely a mixture including C and another derivative that is not m^5C.

In order to highlight particularities of the two tRNAs reported here, we compared them to the statistical tRNA maps previously published (10). These maps show the distribution of the four parent nucleotides in each position of the cloverleaf and thereby can easily be used to note unusual structural features. Boxed-in positions in Fig.1a denote the nucleotides of these sequences which occur in less than 10% of all other prokaryotic elongator tRNAs. Attention is thereby drawn to the U_3-A_{70} base pair, a sizeable portion of the D-stem, the unmodified A_{37} and finally the unusual pseudouridine at position 40. The U_{10}-A_{25} pair is seen in only two other tRNAs: $tRNA^{Ser}_{UGA}$ from S. pombe and $tRNA^{His}_{GUG}$ from bacteriophage T5, aside from some occurrences in mitochondrial tRNAs. This feature may have structural implications, since nucleotides in these positions are involved in tertiary structure stabilization (2,11). Equally rare is the G_{30}-Ψ_{40} pair. Finally these tRNAs have only three base pairs in the D-stem, a general feature of tRNAs having long extra arms (3).

Similarities in E.coli serine tRNAs.

Since all E.coli serine tRNAs are aminoacylated equally well by the single seryl-tRNA synthetase present in E.coli (4,12), we have made a composite from these tRNAs (Fig. 1B). Although the $tRNA^{Ser}$ encoded by the bacteriophage T4 has been included in the compilation, the two T5 serine tRNAs

have been excluded (see below). If we ignore invariant or semi-invariant positions common to most prokaryotic elongator tRNAs (10), those nucleotides that are characteristic of the serine family are located at the end of the acceptor stem (boxed nucleotides). This region includes the U_3-A_{70} base pair and the fourth nucleotide, the subject of a structure-function correlation (13,14). The nucleotides at the end of the CCA arm are also implicated in the mischarging mutation of tyrosine tRNA by glutamine, glutamine tRNA by tryptophan (15) and the suppressor lysine tRNA (16). Equally noteworthy in this position is the rare G_3-U_{70} mispair which is restricted to alanine tRNAs (10). Perhaps this region, at least partially, defines the recognition site for a class of tRNA synthetases such as in the aforementioned examples. Using the same reasonning Thorbjarnardottir et al have proposed different sites in leucine tRNAs to be responsible for synthetase recognition (17). If these hypotheses are correct it follows that the elusive synthetase recognition site can differ among tRNA families.

On the other hand, the structural factors recognized by a given synthetase are more likely defined in the three-dimensional structure rather than by a primary sequence (18a,18b). In that view, the large extra arm and the lack of the fourth base pair in the D-stem may also play a role in the specific interaction with the seryl- and leucyl-tRNA synthetases. The structural variability would exclude the length and the sequence of the extra arm from playing a major role. In the same way, the anticodon region of the $tRNA^{Ser}$ would not be involved, since both its sequence and its hydrophobic nature generated by the 3'-adjacent adenosine (or a modification thereof) vary.

The modification of A_{37}

Earlier studies suggested that the major species of $tRNA^{Ser}_{GGA}$ had no methylthio-isopentenyl modification at the position adjacent to the anticodon: this tRNA, contrary to many others, including E.coli $tRNA^{Ser}_{UGA}$ had no cytokinin activity in plant extracts (19-20). Due to the limited amount of E.coli $tRNA^{Ser}_{V}$, no firm conclusion could be made as to the presence of the modification (19). Originally $ms^2i^6A_{37}$ was thought to be found 3'-adjacent to all anticodons ending with A_{36} in E.coli (1), however, we have found no modified A_{37} in the serine tRNAs reported here. More recent evidence however suggests that A_{36}, A_{37}, A_{38} and the five base pairs of the anticodon stem are involved in the recognition of A_{37} by the isopentenylation enzyme (31). Although our two tRNAs do have the same three A's 36-38, the very unusual pair G_{30}-Ψ_{40} is at variance with the five base pair requirement and could perturb the conformation of the anticodon loop

sufficiently so that enzymatic modification is no longer possible. This hypothesis could be tested in a direct way by changing Ψ_{40} for C_{40} by site-directed mutagenesis. Further examination of the 15 sequenced tRNAs from E.coli which have the fully modified $ms^2i^6A_{37}$ (see Fig.1C) would indicate that the presence of a A-Ψ or a U-A at positions 31 and 39 rather than the C_{31}-G_{39} in our $tRNA^{Ser}$ is strongly correlated with modification. Again this correlation could be verified effectively by site-directed mutagenesis.

The evolution of the serine tRNA family

Given the complete set of six E.coli $tRNAs^{Ser}$ and three bacteriophage tRNAs from T_4 and T_5 which presumably operate with the same protein synthesizing machinery, we have compared them to deduce their evolutionary origin. Based on a simple test which we have previously described (2,22a), the T4 tRNA is homologous to the E.coli tRNAs, whereas the T5 tRNAs are not statistically related. This result demonstrates the origin of the T4 to be within the Gram negative bacteria. Parallel analysis with the leucine tRNA family reveals that T5 tRNA genes may be related to other bacterial phyla (Cedergren, unpublished results). The high degree of similarity of the $tRNA^{Ser}$ and $tRNA^{Ser}$ from E.coli, more similar than isocoding species (tRNAs having the same anticodon) from E.coli and B.subtilis, is unexpected. Normally, tRNAs coding in a different box of the genetic code should have had a very early common ancestor, in order to have read their codons continuously since the establishment of the present genetic code. However this data supports a strong case for a recent gene duplication event followed by divergence. The regular distribution of differences along the entire length of the two molecules would argue against a gene conversion event (22b).

The close relationship of these tRNAs suggested the examination of their extra arm area.

UAUACGG-CAA----CGUAUC in $tRNA^{Ser}_{GGA}$

UAUGCGGUCAAAAGCUGCAUC in $tRNA^{Ser}_{GCU}$

Above is the optimal alignment of these regions for positions 44 to 48 in the conventional tRNA numbering system. Indeed, in spite of the length difference of the two loops, strong similarity can be observed; the optimal alignment shown infers three base replacements and two insertion/deletions of one and four bases (underlined). Since overall length in this region does not seem to be a major consideration, DNA polymerase slippage during replication of one of the duplicated genes (perhaps in the A stretch) is conceivable.

Multiplicity and decoding in the serine tRNA family

For several years we have entertained a model concerning the role of modified nucleosides in the anticodon region in modulating codon-anticodon interactions so that the lifetime of all tRNAs in the decoding site of the ribosome are similar (optimal codon-anticodon theory (2,23,24). Since base composition and nucleotide sequence of anticodons vary, the need for a modulating effect will vary from one tRNA to another: as a general rule, tRNAs with anticodons containing mostly A and/or U require strengthened stabilization by hypermodification of the 3'-adjacent base, while tRNAs with G-C rich anticodons are naturally less dependent on hypermodified adjacent nucleotides (or other features of the anticodon loop and stem, ref.25). Each organism has a distinct tRNA population. E.coli has at least six $tRNAs^{Ser}$ which are encoded by at least four genes (see Table 1). The minor species $tRNA^{Ser}_{CGA}$, a product of the serU gene is not essential, since it can be mutated to a suppressor (26,27). Two of the other $tRNA^{Ser}$ are redundant in that they duplicate decoding responses. The two $tRNA^{Ser}_{GCU}$ isoacceptors originate from the same gene (serW) (28) as the redundancy emanates from an incomplete modification of A_{37} to t^6A (29,30). Because this hypermodification stabilizes energetically weak U-A and U-G base pairs (5), we expect that the serine codon, AGU, will be more efficiently read by the tRNA containing the t^6 modification; the AGC codon, on the other hand, may be decoded equally well by the modified or unmodified tRNA. In addition the AGU codon may be avoided in highly expressed genes of E.coli (23,24). Differential codon reading by isocoding $tRNAs^{Lys}$ having different levels of hyper-modification of A_{37} has been demonstrated in a reticulocyte cell free system (31).

As demonstrated in this paper, $tRNA^{Ser}_{GGA}$ is also redundant. The major species is seven times more abundant than the minor and makes up 40% of the total $tRNA^{Ser}$ in E.coli. Differing in only one position, it is possible that the two tRNAs are transcribed from the same gene, if one assumes the presence of a new type of insertase. Alternatively, this heterogeneity could be evidence for a region of the gene subject to transcriptional (hot spot) mutation. We do not know if this variance in $tRNA^{Ser}_{GGA}$ has any incidence on the coding property of these tRNAs.

Recent studies on anticodon-anticodon complexes reveal that $tRNA^{Ser}_{GGA}$ forms a particularly stable complex with $tRNA^{Gly}_{UCC}$: the rate constants for association and dissociation are similar to that of other tRNA pairs stabilized by $ms^2i^6A_{37}$ (6). Thus the methyl-2-thio group is indeed a major factor in the stabilization of anticodon interactions of

$ms^2i^6A_{37}$ containing tRNAs. $tRNA^{Ser}_{GGA}$ is intrinsically "sticky" and may not need a hypermodified A_{37} to efficiently read the UCC and UCU codon. The frequency of these codons is very high particularly in highly expressed and other genes (Table I). A direct consequence of this logic that the reading of the predominant serine codons UCC, UCU and AGC in highly expressed mRNA should not be affected by the E.coli miaA mutation (32), which are unable to make either i^6A or ms^2i^6A. Only the translation of mRNA containing the rarely used codon UCA and UCG, which are decoded by $tRNAs^{Ser}$ containing fully hypermodified $ms^2i^6A_{37}$, should be affected. E.coli miaA mutants have only slightly impaired growth, but are drastically derepressed for the enzymes of tryptophan, phenylalanine and tyrosine biosynthesis. In the same mutants, metabolism of serine, leucine and cysteine are affected little, if at all (33). To rationalize these facts consider 1) that the aromatic aminoacids have only one type of tRNA, and 2) that these have A-U rich anticodons which are the most sensitive to the stabilizing effect of the flanking $ms^2i^6A_{37}$. The less efficient tRNA, lacking the ms^2i^6 group in miaA mutants mimics amino acid starvation thus turning on the biosynthesis of these genes which are regulated by the attenuation mechanism (34).

On the contrary, the most frequently used codons for serine and leucine are read by tRNAs that are not dependent on the isopentenylating enzyme; while cysteine is infrequently used in most proteins and thus, should not influence much overall protein synthesis. Similar arguments may apply to the growth of E.coli in iron deficient media which results in a drastic reduction of methylthiolation of i^6A (39). The methyl-2-thio group is indeed a major factor in the stabilization of anticodon-anticodon interactions of $ms^2i^6A_{37}$ containing tRNAs (6). An intriguing possibility exists that because of its unique Ψ_{40}, (which is also present in $tRNA^{Leu}_{CAG}$) E.coli $tRNA^{Ser}_{GGA}$ is a target for regulation by a specific pseudouridylate synthetase similar in effect to hisT mutants in Salmonella typhimurium (35). Consequently, a judicious use of certain codons together with a careful control of built-in features of the anticodon loop, including modification of adjacent positions, may be part of a coordinated structure modification strategy that allows cells to adapt to physiological stresses.

ACKNOWLEDGEMENTS

The authors would like to thank Barry MacLennan for providing information on the identification of adenosine derivates and Damien Labuda for discussion on the interpretations of our data. This work was supported by the Natural

Science and Engineering Research Council and the Fonds de coopération Belgique-Québec (project no. 3.1.10).

REFERENCES

1. Nishimura, S. (1979), in "Transfer RNA: structure, properties and recognition" - Schimmel, P.R., Söll, D. and Abelson, J.N. eds) pp. 59-79, Cold Spring Harbor Lab.
2. Cedergren, R.J., Sankoff, D., Larue, B. and Grosjean, H. (1981) in C.R.C. Critical Reviews in Biochem. vol. 11, 35-104.
3. Gauss, D.H. and Sprinzl, M. (1984) Nucl. Acids. Res. 12, r1-r131.
4. Roy, K.L. and Söll, D. (1970) J. Biol. Chem. 245, 1394-1400.
5. Weissenbach, J. and Grosjean, H. (1981) Eur. J. Biochem. 116, 207-213.
6. Houssier, Cl. and Grosjean, H. (1985) J. Biomolec. Struct. and Dynamics, submitted.
7. Peattie, D.E. (1979) Proc. Natl. Acad. Sci. USA 76, 1760-1764.
8. Gu, X.R., Nicoghosian, K., Cedergren, R.J. and Wong, J.T.-F. (1983) Nucl. Acids Res. 11, 5433-5442.
9. Gupta, R.C. and Randerath, K. (1979) Nucl. Acids Res. 6, 3443-3458.
10. Grosjean, H., Cedergren, R.J. and McKay, W. (1982) Biochimie 64, 387-397.
11. Goddard, J.P. (1977) Progress. Biophys. Mol. Biol. 32, 233- .
12. Sundharadas, G., Katze, J.R., Söll, D., Konigsberg, W., Lengyel, P. (1968) Proc. Natl. Acad. Sci. USA 61, 693-700.
13. Crothers, D.M., Seno, T. and Söll, D. (1972) Proc. Natl. Acad. Sci. USA 69, 3063-3067.
14. Uemura, H., Imai, M., Ohtsuka, E., Ikehara, M. and Söll, D. (1982) Nucl. Acids Res. 10, 6531-6539.
15. Celis, J.E. and Piper, P.W. (1982) Nucl. Acids Res. 10, r83-r91.
16. Prather, N.E., Murgola, E.J. and Mims, B.H. (1984) J. Mol. Biol. 172, 177-184.
17. Thorbjarnardottir, S., Dingermann, T., Rafnar, T., Andresson, O.S., Söll, D. and Eggertsson, G. (1985) J. Bact. 161, 219-222.
18a. Schimmel, P.R. and Söll, D. (1979) Ann. Rev. Biochem. 48, 601-648.
18b. Thuriaux, P. (1983) Biochimie 65, 585-588.
19. Armstrong, D.J., Burrows, W.J., Skoog, F., Roy, K.L. and Söll, D. (1969) Proc. Natl. Acad. Sci. USA 63, 834-841.
20. Bartz, J., Söll, D., Burrows, W.J. and Skoog, F. (1970) Proc. Natl. Acad. Sci USA 67, 1448-1453.
21. Tsang, T.H., Buck, M. and Ames, B.N. (1983) Biochim. Biophys Acta 741, 180-196.
22a Cedergren, R.J., Larue, B., Sankoff, D., Lapalme, G. and Grosjean, H. (1980) Proc. Natl. Acad. Sci. USA 77, 2791-2796.
22b Munz, P., Amstutz, H., Kohli, J. and Leupold, V. (1982) Nature 300, 225-228.
23. Grosjean, H., Sankof, D., Jou, M.J., Fiers, W. and Cedergren, R.J. (1978) J. Mol. Evol. 12, 113-119.
24. Grosjean, H. and Fiers, W. (1982) Gene 18, 199-209.
25. Yarus, M. (1982) Science 218, 646-652.
26. Thorbjarnardottir, S., Uemura, H,. Dingerman, T., Rafnak, T., Thorsteinsdottir, S., Söll, D. and Eggertsson, G. (1985) J. Bact. 161, 207-211.
27. Steege, D.A. (1983) Nucl. Acids Res. 11, 3823-3832.
28. Ikemura, T. and Ozeki, H. (1977) J. Mol. Biol. 117, 419-446.
29. Yamada, Y. and Ishikura, H. (1975) Biochim. Biophys. Acta 402, 285-287.
30. Ish-Horowicz, D. and Clark, B.F.C. (1973) J. Biol. Chem. 248, 6663-6673.
31. Smith, W.E., McNamara, A.L., Rice, M. and Hatfield, D.L. (1981) J. Biol. Chem. 256, 10033-10036.

32. Eisenberg, S.P., Soll, L. and Yarus, M. (1979) J. Mol. Biol. 135, 111-126.
33. Buck, M. and Ames, B.N. (1984) Cell 36, 523-531.
34. Yanofsky, Y.C. (1981) Nature 289, 751-758.
35. Turnbough, C.L., Neill, R.J., Landsberg, R. and Ames, B.N. (1979) J. Biol. Chem. 254, 5111-5119.
36. Ishikura, H., Yamada, Y. and Nishimura, S. (1971) Biochem. Biophys. Acta 228, 471-481, also (1971) FEBS Letters 16, 68-70.
37. Hoffman, E.P. and Wilhelm, R.C. (1970) J. Bacteriol. 103, 32-36.
38. Ikemura, T. (1982) Cold Spring Harbor Symposia vol. 48, 1087-1097.
39. McLennan, B.D., Buck, M., Humphreys, J. and Griffiths, E. (1981) Nucl. Acids Res. 11, 2629-2640.

NOTE ADDED:

During the processing of this manuscript we have learned of the determination of the E.coli $tRNA^{Ser}_{GGA}$ sequence reported here, by Fisher, W. and Sprinzl, M. (1985 Banz tRNA Workshop).