AN *IN VIVO* APPROACH TO tRNA IDENTITY

Thesis by
Jennifer Normanly

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

1989
(Submitted December 20, 1988)

ACKNOWLEDGEMENTS

I may never have contemplated pursuing a Ph.D. in Biology had it not been for the support and guidance of several people to whom I am grateful; Steve Wilson for suggesting that I go into science in the first place, and who provided me with constant encouragement and moral support, Frank Talamantes for giving me a shove in the right direction, Harry Noller for letting me loose in his lab, and Jeff Prince who redefines the word meticulous, and who set the standards for precision which I can only hope to live up to. I also wish to mention my parents, Ardis and Jerry, and siblings Brian and Colleen, who, while they where never quite sure *exactly* what I was doing, cheerfully supported me anyway, my aunt Nora who kept me well fed through college, and my grandmother Laura Naur who bankrolled me during most of my undergraduate days.

My time spent as a graduate student has been enjoyable, made so by my various labmates and associates; my dear friends Eric Phizicky and Beth Grayhack, who took me under their wing, and Joey Zwass, a truely zany person, who provided many amusing diversions. Richard Ogden was a great deal of help to me early in my graduate career, as were Richard Schwartz and Peter Johnson. David Horowitz, who distracted me from my research countless times with stimmulating debate, is to be commended for his unflagging efforts to keep me grammatically correct. George Komatsoulis has been a very amiable labmate, always quick to be of help when asked. I have been enriched for having had the opportunity to collaborate with Jeffrey Miller, who not only added to my development as a scientist, but made sure that the dreary diet afforded by a graduate student stipend, was supplemented with more than one memorable meal. I have also enjoyed knowing fellow graduate student, Lynn Kleina, who slogged it out with me in the trenches of DHFR purification. Several members of the staff at Caltech, Nancy Gill, Patricia Bateman, Mary Marsh, Connie Katz and Cathy Bragg, have very graciously dealt with this hapless graduate student, ensuring smooth sailing for me in a beaurocratic sense.

Finally, I am grateful to my advisor, John Abelson, who has enthusiastically supported and guided me in my research, while allowing me a great deal of indepedence. His intrepid approach to science is one I hope to emulate.

ABSTRACT

A leucine-inserting tRNA has been transformed into a serine-inserting tRNA by changing 12 nucleotides. Only 8 of the 12 changes are required to effect the conversion of the leucine tRNA to serine tRNA identity. The 8 essential changes reside in basepair 11-24 in the D stem, basepairs 3-70, 2-71 and nucleotides 72 and 73, all of the acceptor stem.

Functional amber suppressor tRNA genes were generated for 14 species of tRNA in *E. coli*, and their amino acid specificities determined. The suppressors can be classified into three groups, based upon their specificities. Class I suppressors, $tRNA^{Ala2}_{CUA}$, $tRNA^{GlyU}_{CUA}$, $tRNA^{HisA}_{CUA}$, $tRNA^{Lys}_{CUA}$, and $tRNA^{ProH}_{CUA}$, inserted the predicted amino acid. The Class II suppressors, $tRNA^{GluA}_{CUA}$, $tRNA^{GlyT}_{CUA}$, and $tRNA^{Ile1}_{CUA}$ were either partially or predominantly mischarged by the glutamine aminoacyl tRNA synthetase (AAS). The Class III suppressors, $tRNA^{Arg}_{CUA}$, $tRNA^{AspM}_{CUA}$, $tRNA^{Ile2}_{CUA}$, $tRNA^{Thr2}_{CUA}$, $tRNA^{Met(m)}_{CUA}$ and $tRNA^{Val}_{CUA}$ inserted predominantly lysine.

TABLE OF CONTENTS

INTRODUCTION

tRNA Identity

Jennifer Normanly and John Abelson

Department of Biology, California Institute of Technology, Pasadena CA 91125

[Annual Review of Biochemistry (1989) vol. 58, in press]

BACKGROUND

In the translation of genetic information from nucleic acid to protein, transfer RNA plays a crucial intermediate role. Each codon is read by a unique tRNA that has been aminoacylated with the appropriate amino acid. Because of degeneracy in the genetic code, in general there is more than one tRNA for each amino acid (1). Although there are some exceptions to the rule, there is a single aminoacyl tRNA synthetase (AAS) for each amino acid. An AAS must correctly recognize and activate an amino acid and then join it to each of the cognate set of tRNAs. By virtue of the fact that they must function interchangeably on the ribosome during protein synthesis, all tRNAs have similar primary (2), secondary (3), and tertiary (4, 5) structures (Figure 1). Within these constraints each cognate set of tRNAs contains distinct elements that mediate correct recognition by the AAS. These elements comprise the *identity* of a tRNA. Identity elements can be both positive and negative. Positive elements are those features of the tRNA that the cognate AAS recognizes directly, and negative elements are those features that block the recognition by other AASs.

Although tRNAs comprise a homogeneous set of molecules, AASs are remarkably diverse. The quaternary structures α, α_2, α_4, and $\alpha_2\beta_2$ have all been observed (6, 7). AAS subunit sizes range from 300 to 900 amino acids. The details of the interaction between these diverse enzymes and their tRNA substrates will not be known until high resolution X-ray structures of tRNA-AAS complexes have been solved. Fortunately, there is the prospect that several such complexes will be solved in the near future (8, 9). Already, we have a remarkably detailed description of the amino acid and ATP binding sites for the *B. stearothermophilus* tyrosine AAS (10) and the *E. coli* methionine AAS (11). It is not our purpose to review the large body of information on the structure and properties of the AAS, or on the indirect efforts that have been employed to map their interactions with tRNA. These have been well reviewed before in these volumes (6, 7). Rather, we wish to describe the results of fresh and energetic approaches to the problem of tRNA identity, which have appeared in the last few years.

Past and Present Approaches to Defining tRNA Identity.

We will describe two new approaches to the determination of tRNA identity. Both have their roots in previous work but are now more powerful because of advances in gene synthesis technology. In the first, attempts are made to alter with the fewest changes, the identity of a tRNA while retaining complete biological function. Those changes responsible for the redirection of AAS recognition constitute identity elements. We term this type of experiment an identity swap. In the second, tRNA variants are generated by altering the sequence of a tRNA, and their properties as a substrate for the AAS are determined in vitro. This approach has been greatly facilitated by the ability to synthesize tRNA in vitro with bacteriophage T7 RNA polymerase.

The identity swap has its origins in the study of amber suppressors. An amber suppressor is a mutant tRNA that can recognize and suppress the chain terminating effects of a UAG codon, by virtue of a change in the tRNA's anticodon (12, 13). Suppressors have greatly facilitated the application of genetics to the study of tRNA, and have been utilized extensively in bacteria (14), T4 bacteriophage (15, 16), and yeast (17). Such work has led to a formal proof of the secondary structure of tRNA (14) and to information about the mechanism of tRNA processing (18-20). Suppressors have also been used to address the tRNA identity problem. Attempts were made to alter the specificity of the *E. coli* $tRNA^{Tyr}$ suppressor, $su^{+}3$, by selecting mutants of this tRNA that could suppress amber mutations at positions in which tyrosine was not an acceptable amino acid. This approach was taken independently by J. D. Smith and S. Brenner in Cambridge and by H. Ozeki and Y. Shimura in Kyoto nearly 20 years ago (21, 22). Both groups obtained the same results. The $su^{+}3$ mutants that were isolated inserted glutamine instead of tyrosine into amber codons (23). Sequence analysis revealed changes in the fourth base from the 3' end, $A_{72} \rightarrow G$ (22) and in the first and second base pairs of the acceptor stem 1-72 and 2-71 (21). This was an exciting result, but it did not prove to be universally applicable to the study of tRNA identity. Despite many attempts, $su^{+}3$ mutants that inserted an amino acid other than glutamine or tyrosine were never found (23). It is likely that such transformations would have required more than one change in the tRNA sequence and therefore could not be obtained by genetic selection. This approach was abandoned for fifteen years.

We have returned to this approach armed with an automated DNA synthesizer, which allows the rapid synthesis of altered suppressors *de novo*, as in the pioneering work of Khorana (24). The synthetic genes are assembled from oligonucleotides and inserted in

cassette form into a plasmid, which supplies a promoter, either the inducible *lac* promoter (25) or the constitutive *lpp* promoter (26, 27), and a transcription terminator. Usually a screen is employed, in which the suppression of an amber mutation in a gene requires the insertion of a specific amino acid to allow growth. The amino acid specificity of the suppressor is determined by sequencing the protein product of a suppressed amber mutant gene. For this purpose the *E. coli* dihydrofolate reductase (DHFR) gene with an amber mutation at codon ten is used. DHFR is easily purified by affinity chromatography on a methotrexate resin. X-ray crystallographic data (28-29) revealed that residue ten of DHFR lies on the surface of the protein, well removed from the active site, and it appears that any amino acid is acceptable at this position. If more than one amino acid is inserted by the suppressor, it can be readily detected by this method. Protein sequence data acquired in this manner reveal the outcome of the competition among all 20 AAS in the cell for the suppressor. The drawbacks to this method are twofold. First, the anticodon of the tRNA must be CUA to allow recognition of amber codons; so the contribution of the anticodon to tRNA identity cannot be assessed. As described below, the anticodon is a strong identity element for some tRNAs. Second, the effectiveness of the suppressor tRNA as a substrate for the AAS can only be measured by the efficiency of suppression. Multiple factors, such as the ability of the tRNA to be transcribed, processed, modified, and accepted by the ribosome, can affect the observed suppression efficiency, and in this method cannot be distinguished from the efficiency of aminoacylation.

In the second method developed by Uhlenbeck (30), tRNA genes and their variants are synthesized and placed under the control of the bacteriophage T7 promoter so that the first nucleotide of the in vitro transcript is the mature 5' end of the tRNA. Alternatively, synthetic tRNAs with 5' precursor sequences can be processed correctly by incubation with the RNA component of RNAse P (31; L. Schulman, unpublished). A Bst N1 restriction endonuclease recognition site (5'CC·AGG) is encoded at the 3' end of the gene. Run-off transcription of Bst N1 cleaved DNA generates the correct CCA terminus. Prior to the availability of this methodology, the study of tRNA was hindered in that most variants were not processed or were unstable in vivo. This technique not only overcomes such obstacles but also allows the easy purification of the transcript. Reactions can be scaled up and milligrams of some transcripts have been produced for NMR studies (Hall *et al*, unpublished). These synthetic tRNAs do not contain the normal base modifications found in native tRNA (and are therefore excellent substrates for the study of any modifying enzyme). They are, nonetheless, good substrates for the AAS. Typically, the K_m values

for the substrates are one- to four-fold higher than for fully modified tRNA. V_{max} values for the unmodified substrates are at most, two-fold lower, resulting in values of the specificity constant ($kcat/K_m$) that are five- to seventeen-fold lower. In contrast, the $kcat/K_m$ for a typical misacylation is orders of magnitude lower (6). These results indicate that the modified nucleotides are not likely to play a large role in recognition by the AAS, as their absence does not significantly diminish the efficiency of aminoacylation.

In order to determine that the synthetic tRNAs are properly folded, Sampson and Uhlenbeck (32) have exploited the phenomenon of lead-mediated cleavage of tRNA. This reaction is mediated by a tightly bound Pb^{++} ion whose position is determined by the tertiary interactions between the D-loop and T-loop (33, 34). In native yeast $tRNA^{Phe}$, cleavage takes place at residue 17. Unmodified $tRNA^{Phe}$ is cleaved at the same position, with similar kinetics, indicating that the synthetic molecule has folded properly. The in vitro approach is attractive in that one can quantitate the effects of individual nucleotide changes on the recognition of the tRNA by cognate and non-cognate AAS. The method does not allow assessment of the effect of competition between one AAS and the other 19 AASs. In the end, it will be desirable to apply both the in vivo and in vitro approaches to determine the identity elements of any tRNA.

The Role of the Anticodon in tRNA Identity.

Since the anticodon is directly correlated with the identity of any tRNA species, it is the most logical identity element. Indeed, one of the earliest recognition hypotheses stated that the "coding triplet (anticodon) plays an important, perhaps decisive role in the interaction with the synthetase" (35). Early evidence to support this hypothesis came from a combination of chemical modification and oligonucleotide interference experiments as well as from analysis of mutant tRNAs, and suggested that for a number of tRNAs, the anticodon can be implicated in recognition. Among them are *E. coli* $tRNA^{Gly}$, $tRNA^{Lys}$, $tRNA^{Arg}$, $tRNA^{Glu}$, $tRNA^{Met(f)}$, $tRNA^{Trp}$, yeast $tRNA^{Gly}$, $tRNA^{Val}$, and beef pancreatic $tRNA^{Trp}$ (reviewed in 36).

The opposing notion, that the anticodon could not be a recognition element, was derived from the observation that the anticodon could be altered to create amber suppressors with no effect on amino acid specificity in at least four tRNAs. Additionally, translation of the codons (six each) for serine, leucine, and arginine requires a set of tRNAs with different nucleotides in two out of three positions for leucine and arginine, and in all three for serine. The idea that the anticodon is not an identity element seems to have held

sway, in some quarters at least, because a few textbooks still state this unequivocally. Nonetheless, it is now clear that the anticodon is a recognition element for tRNA identity in perhaps a majority of tRNAs.

The strongest evidence in support of this view comes from L. Schulman's work on *E. coli* $tRNA^{Met}$. Experiments conducted over a period of 15 years have demonstrated conclusively that the CAU anticodon of *E. coli* $tRNA^{Met}$ is an essential recognition element for the methionine AAS. Chemical modification experiments have shown that the anticodon nucleotides C_{34} and A_{35} cannot be modified without loss of methionine acceptor activity in vitro (37). The anticodon of $tRNA^{Met}$ was altered by the judicial insertion of oligoribonucleotides into the tRNA, which had been cleaved specifically by RNAse. These experiments revealed that substitution of C_{34} with a U, A, or G lead to a decrease in the aminoacylation rate by four to five orders of magnitude (38). Altering the other two positions of the anticodon (A_{35} and U_{36}) affect AAS recognition as well, decreasing the rate of aminoacylation one to four orders of magnitude to that of normal (39). Substitutions of any position in the anticodon with nucleotides containing functional groups in common with those of the wildtype nucleotides were tolerated more than if nucleotides, lacking similar functional groups were substituted. The conclusion was that the synthetase must be interacting with specific functional groups in the anticodon (40). In addition, chemical crosslinking experiments demonstrated that C_{34} of the anticodon is uniquely crosslinked to Lys 465 in the methionine AAS (41).

More recently, Schulman has conducted identity swap experiments in the in vitro system (42). In these experiments genes for both $tRNA^{Trp}$ and $tRNA^{Val}$ of *E. coli* were synthesized with the methionine (CAU) anticodon, transcribed in vitro, and the kinetics of acylation by the methionine AAS and valine AAS were determined for the synthetic tRNAs (Table 1). $tRNA^{Val}$ (CAU) was aminoacylated by the methionine AAS with a V_{max}/K_m similar to native $tRNA^{Met}$. $tRNA^{Trp}$ (CAU) was also acylated by the methione AAS, but the V_{max}/K_m was down ten-fold. Conversely, $tRNA^{Met}$ with the valine anticodon (UAC) was acylated by the valine AAS with a V_{max}/K_m value 1/10 of that for normal $tRNA^{Val}$. In comparison, the value for acylation of $tRNA^{Met}$ by the valine AAS was four orders of magnitude lower than that for $tRNA^{Val}$. These results clearly show that the anticodon is a major, if not the sole, recognition element for $tRNA^{Met}$ and is important for $tRNA^{Val}$ identity as well.

Theories about the role of the anticodon in AAS recognition have generally been influenced by the existence and the properties of amber suppressors. The demonstration that suppressors for $tRNA^{Tyr}$, $tRNA^{Ser}$, $tRNA^{Gln}$, $tRNA^{Leu}$, were correctly charged, indicated that not all AASs recognize the anticodon nucleotides of their cognate tRNAs. A more complete picture of the effects of the amber anticodon on AAS recognition has been obtained from our efforts in collaboration with J. Miller and coworkers to construct a complete bank of 20 amber suppressors (27, 43 Normanly *et al*, unpublished). These suppressors have been generated by synthesizing the genes of isoacceptors from the 20 species of tRNA in *E. coli*, each containing a CUA anticodon. As in the identity swap experiments described above, the tRNA genes are inserted in cassette form into a plasmid containing the constitutive *lpp* promoter and a terminator. The amino acid specificity of each suppressor was determined as described above by sequencing the suppressed product of the amber mutation at position ten in DHFR. Tables 2 and 3 give a summary progress report of this project. Table 2 lists the species of tRNA, which exist as amber alleles, derived either genetically or by synthetic means. Table 3 classifies the suppressors by their amino acid specificity. Seven of the synthetic suppressors insert the correct amino acid. Together with the four original suppressors, 11 of 20 tRNAs in *E. coli* can tolerate the amber anticodon without misacylating. In some cases suppressor tRNAs are mischarged by the glutamine AAS. The synthetically derived $tRNA^{GluA}_{CUA}$ inserts 20% glutamine. From previous work (44) we know that $tRNA^{Met(f)}_{CUA}$ is mischarged by the glutamine AAS in vitro, and the $su^{+}7$ allele of $tRNA^{Trp}$ inserts a mixture of glutamine and tryptophan in vivo (45). The amber allele of one isoacceptor of $tRNA^{Gly}$, $tRNA^{GlyU}_{CUA}$, was correctly charged in vivo, another $tRNA^{GlyT}_{CUA}$, was misacylated by the glutamine AAS (46; Normanly *et al*, unpublished). We were surprised to find that six of the suppressors generated in this work, $tRNA^{Arg}_{CUA}$, $tRNA^{AspM}_{CUA}$, $tRNA^{Ile2}_{CUA}$, $tRNA^{Met(m)}_{CUA}$, $tRNA^{Thr2}_{CUA}$ and $tRNA^{Val}_{CUA}$, were misacylated by the lysine AAS. There are also examples of suppressors that are misacylated by both the glutamine and lysine AASs. Such is the case for $tRNA^{Ile1}_{CUA}$, as well as for some mutant suppressor tRNAs (46-48).

Apparently, only the glutamine AAS and the lysine AAS misacylate amber suppressors. Each tends to misacylate a separate set. Because the CUA anticodon is recognized by both the glutamine AAS and the lysine AAS, there must be negative elements that prevent acylation by these enzymes. Some suppressors have both elements, some only one, and others have neither. It is interesting to note that there are two lysine AAS genes in *E. coli*, one of which is preferentially expressed during heat shock (49). Preliminary

experiments designed to detect which of the enzymes is responsible for the observed misacylation indicated that both are.

A Single Base Pair is a Major Identity Determinant for tRNA^{Ala}.

Hou and Schimmel set out to isolate mutants of a tRNA^{Ala} amber suppressor, which could not be recognized by the alanine AAS using the in vivo approach (50). An impressive collection of synthetic mutants was generated with multiple changes throughout the tRNA. Only those mutants having in common a change in the G-U pair at position 3-70 failed to insert alanine. This finding is consistent with two previous observations. First, earlier work by Murgola had indicated that a G-U pair at this position in a tRNA^{Lys} missense suppressor caused the insertion of either alanine or glycine (51). Second, the G-U pair at this position is conserved in virtually all of the tRNA^{Ala} that have been sequenced. To test the significance of the G_3-U_{70} pair, Hou and Schimmel performed two identity swap experiments. tRNA^{Cys} and tRNA^{Phe} amber suppressors were synthesized, each containing the G-U pair at position 3-70. The altered cysteine suppressor inserted entirely alanine (Table 4). The altered phenylalanine suppressor inserted predominately alanine (63%) and some phenylalanine (37%). This result is certainly one of the most dramatic identity shift experiments to date in that it implicated a single G-U pair as the major identity element for the alanine AAS.

The ability of the G_3-U_{70} pair to direct alanine AAS recognition has been tested in other tRNAs with mixed results. Superimposing the G-U pair at position 3-70 of a lysine suppressor results in complete conversion to alanine identity (48). However, the same alteration to a glycine suppressor tRNA (GlyT) results in the insertion of predominantly glutamine, a small amount of glycine, and no detectable levels of alanine (46). It shoud be noted that the glycine amber suppressor with the normal G-C base pair at position 3-70 inserts predominantly glutamine (46; Normanly *et al*, unpublished). In the case of tRNA^{GlyT}, simply superimposing the G-U pair at position 3-70 is not sufficient to counteract the identity determinants for the glutamine AAS. The glycine to alanine identity swap experiment might be more successful if the amber allele of another isoacceptor of tRNA^{Gly} which is not misacylated, were used instead.

McClain and Foss approached the problem of tRNA^{Ala} from a different angle (46). McClain has developed an algorithm (52, 53), which compares the sequences of a tRNA with its isoacceptors as well as isoacceptors of all 19 other tRNAs. This algorithm selects those positions in a tRNA, which are correlated with its isoacceptors and anticorrelated

with the other tRNA sequences. Subjecting $tRNA^{Ala}$ to this type of analysis resulted in the prediction that four nucleotides were significant to $tRNA^{Ala}$ in addition to the anticodon. These were the G_3-U_{70} pair, as well as G_{20} and C_{60} (Figure 2). The phenylalanine suppressor was altered to include these nucleotides, and a change was incorporated at position 51 to increase transcriptional efficiency (54). They also made changes at positions 16 and 17, which were designed to remove phenylalanine identity. In the DHFR assay, the resulting suppressor inserted 96% alanine and 4% lysine, indicating that the phenylalanine identity determinants had been removed.

They examined the contribution of C_{60} to alanine identity, and found that altering this putative identity element in $tRNA^{Ala}$ did not abolish alanine acceptance, nor did it give rise to acylation by the alanine AAS when superimposed upon $tRNA^{Phe}$. In a similar fashion, the contribution of G_{20} to alanine identity was determined to be minimal. Alanine acceptance was not abolished when G_{20} was altered in $tRNA^{Ala}$, nor did G_{20} transplanted into $tRNA^{Phe}$ result in alanine insertion. It would appear that these elements of alanine identity predicted by the algorithm do not have a strong effect upon alanine AAS recognition.

$tRNA^{Phe}$ containing the G_3-U_{70} base-pair alone, inserted 24% alanine and 76% phenylalanine. This result was the inverse of that obtained by Hou and Schimmel (63% alanine and 37% phenylalanine). The reason for this discrepancy is not clear, because both groups used the same expression system and the same specificity assay; namely, suppression of an amber mutation in DHFR. In any case, the G_3-U_{70} pair is clearly the dominant determinant of Ala identity, and is the strongest contributor to alanine identity predicted by McClain's algorithm.

The G-U pair as an identity element raises an interesting question. Does the alanine AAS recognize the functional groups of these bases via interactions in the deep groove of the A-helix or is it the perturbations to the helix stability and structure imposed by the G-U "wobble" pair that are recognized? To answer this question McClain has now constructed alanine suppressors that contain each of 16 possible base combinations at position 3-70 (48). All the variants had substantially lowered suppression efficiencies, and some of the base combinations yielded inactive suppressors. The four Watson-Crick base pairs at this position yielded suppressors that inserted lysine and glutamine. $tRNA^{Ala}$ containing the mispairs G-A, C-A, and U-U at position 3-70 all inserted 90% alanine and, though less efficient, were similar to G-U at this position. Other variants, containing A-C and C-C, C-U or U-G at position 3-70 inserted ~80% alanine, along with lysine and glutamine. The

mispairs, G-G, A-A, and A-G led to inactive suppressors, and the variant with a U-C mispair inserted predominantly glutamine. These results suggest that the alanine AAS is recognizing a structural perturbation imparted by the G-U pair and not specific bases. However, not all mispairs are acceptable; so the perturbation that is recognized must be sequence dependent. (It should be noted that NMR and X-ray crystallographic data reveal that a G-U base pair does not have a dramatic effect upon the overall structure of the acceptor stem helix.) Translocation of the G-U pair to position 4-69 led to partial alanine insertion, but only if the 3-70 pair was A-U and not G-C. G-U at position 2-71 had no activity. More questions are raised by these observations than are answered. The results emphasize the fact that there is insufficient data on the sequence dependence of RNA structure. Among the structural variations that one can imagine is a bending of the helix or even a separation of the strands.

The tRNA$^{Leu \rightarrow Ser}$ Identity Swap.

We have used the suppressor approach to define the elements of tRNA^{Ser} identity (25). Serine was chosen for several reasons. There are six serine codons, AGU/C and UCN (where N denotes any nucleotide) recognized by at least four different tRNAs. One of the original amber suppressors (su$^+$1) inserted serine; so in all, the serine AAS can recognize tRNAs with at least five anticodons. It is unlikely that there are any serine recognition elements in the anticodon although this assumption has not yet been tested in vitro. We set out to change the identity of a tRNA^{Leu} amber suppressor to serine. An examination of the six *E. coli* tRNA^{Ser} sequences revealed fourteen nucleotides conserved in all tRNA^{Ser} sequences but not found in the tRNA$^{Leu}_5$ starting sequence. Additionally, the base pair at position 3-70, though not conserved (A-U or U-A) was found by McClain's statistical analysis to be correlated with serine and anticorrelated with other sequences (52, 53). This base pair was included as a potential element of serine AAS recognition (Figure 3a). The conserved nucleotides in the acceptor stem and in the D-stem and loop were included in the original trial construction. In addition, the size of the D-loop is highly conserved in the six tRNA^{Ser} sequences. Changes were made in that region to conform to the serine D-loop size. The first trial construction contained 12 changes from the starting tRNA$^{Leu}_5$ amber suppressor (Figure 3b). The resultant suppressor, tRNA$^{Leu \rightarrow Ser}$, was able to suppress an amber mutation in the β–lactamase gene, which requires the insertion of serine to produce an active enzyme. However, the suppression efficiency of tRNA$^{Leu \rightarrow Ser}$, 1%, was quite low compared to the 60 to 100% efficiency of the starting tRNA^{Leu} suppressor. Using the

DHFR assay, the amino acid specificity of the $tRNA^{Leu \rightarrow Ser}$ suppressor was determined to be serine.

Following our initial successful attempt at the identity swap, we then set out to determine the minimum number of changes required to effect the transformation. This has now been accomplished and the results are shown in Figure 3c. The minimum number of base changes required is eight: residing in the first three base pairs of the acceptor stem, nucleotide 73 and the 11-24 base pair in the D-stem. The resulting suppressor containing these minimal changes is approximately 40% efficient. The efficiency was regained when the unnecessary changes in the D-loop were discarded. Interestingly, we find that only an A-U is acceptable at position 3-70 even though some serine tRNAs contain U-A at this position. We speculate that the A-U is acting as a deterrant to the leucine AAS. We are now using this information to conduct further identity swaps. Perhaps other serine identity elements exist that were already present in $tRNA^{Leu}_5$. By this iterative procedure we hope to arrive at a clear understanding of $tRNA^{Ser}$ identity. The *E. coli* serine AAS has been purified and crystallized by R. Leberman in Grenoble (55). The ultimate understanding of $tRNA^{Ser}$ recogniton will come when the manner in which these identity elements are recognized by the enzyme has been defined.

$tRNA^{Phe}$ Recognition.

As in the case of *E. coli* $tRNA^{Met}$, the anticodon has been implicated in the recognition of yeast $tRNA^{Phe}$ by the yeast phenylalnine AAS. Early work by Zachau, in which the anticodon of yeast $tRNA^{Phe}$ was removed, resulted in a tRNA that was aminoacylated by the phenylalnine AAS at a level 20% that of wildtype $tRNA^{Phe}$ (55a). Using recombinant RNA technology, Uhlenbeck and co-workers have demonstrated that substitution of any of the yeast $tRNA^{Phe}$ anticodon nucleotides diminishes aminoacylation ($kcat/K_m$ was decreased 3- to 10-fold) (56). In addition, transplanting the phenylalanine anticodon onto $tRNA^{Tyr}$ leads to aminoacylation of the hybrid molecule by the phenylalnine AAS (57). As the aminoacylation of the hybrid tRNA was fairly weak, they speculated that there must be phenylalanine recognition elements outside of the anticodon as well.

Using the in vitro system, transcription of synthetic tRNA genes with T7 RNA polymerase, a large number of variants have been constructed and their properties as substrates have been tested (30, 58). This is by far the largest and most comprehensive set of tRNA variants ever constructed. At this point nearly all of the non-universal bases in $tRNA^{Phe}$ have been changed and the variants have been tested as substrates. Experiments

in which the nucleotides involved in tertiary interactions were altered, revealed that changes in these bases can be tolerated as long as the overall tertiary interactions are not disrupted. For example the G_{19}- C_{56} tertiary Watson-Crick base pairs can be altered to C_{19}- G_{56} without affecting recognition by the phenylalnine AAS. Either of the single mutations, $G_{19}\rightarrow$ C or $C_{56}\rightarrow$ G leads to a five-fold drop in kcat/K_m, indicating that the tertiary structure of the tRNA must be maintained but that neither G_{19} or C_{56} is recognized directly. When G_1-C_{72} is altered to either G-U or A-U the kcat/K_m drops four- and seven-fold, respectively (58a). It is not clear yet whether G_1- C_{72} is important as a recognition element for the phenylalanine AAS, or as a structural feature. The results of this study led to the proposal that five nucleotides, all in single strand regions, are recognition elements for the phenylalanine AAS. They are the three anticodon bases, G_{34}, A_{35}, and A_{36}, G_{20} in the D-loop, and A_{73} the fourth base from the 3' end (Figure 4). Several tests of the model were undertaken. First, the *E. coli* $tRNA^{Phe}$, a rather poor substrate for the yeast phenylalnine AAS, contained all of the putative identity elements except the one at position 20. Changing U_{20} to G brought kcat/K_m to within a factor of two for that of yeast $tRNA^{Phe}$. Identity swaps based on the conserved bases were also carried out[1], changing yeast $tRNA^{Met}$, $tRNA^{Arg}$, and $tRNA^{Tyr}$ to good stubstrates for the phenylalnine AAS. It appears that each of the five bases contributes independently to the specificity. Thus, a tRNA containing four of the five recognition elements has a kcat/K_m one order of magnitude lower than wild type, and a substrate with only three recognition elements has a kcat/K_m which is two orders of magnitude lower. Some years ago Dudock and collaborators reported the mischarging of a number of tRNAs by the phenylalanine AAS (59). With knowledge of the recognition elements and thus relative contributions to specificity, Sampson and Uhlenbeck were able to reconcile that data nicely. Of the yeast tRNAs that have been sequenced none have more than three of the recognition elements so that the in vivo specificity can be explained if one also assumes a contribution to overall specificity due to competition with the cognate AAS. The five recognition elements are located far apart from one another in the tRNA tertiary structure so the phenylalanine AAS must contact the entire surface of the tRNA molecule.

[1]Changes were incorporated elsewhere as well, either to facilitate the runoff transcription reaction, or to introduce $tRNA^{Phe}$ structural elements where they differed from the starting tRNA.

McClain and Foss have used the suppressor approach to investigate *E. coli* $tRNA^{Phe}$ identity. From their results it appears that U_{20} and A_{73} are implicated in $tRNA^{Phe}$ identity. In addition, small effects are observed in changing the anticodon stem base pairs 27-43, and 28-42 (47).

$tRNA^{Gln}$ identity.

It is pertinent to discuss what is known about the recognition of *E. coli* $tRNA^{Gln}$. In a short time we will know more about the interaction of this tRNA with its AAS than for any other system. At this point (October 1988) Steitz and Söll have generated $tRNA^{Gln}$-AAS co-crystals that refract to high resolution. Data has been collected and a map is being constructed. Where do we expect the contacts will be made?

First, as we have already mentioned, it is likely that the glutamine AAS recognizes the anticodon. The su^+7 allele of *E. coli* $tRNA^{Trp}$, which inserts both glutamine and tryptophan in vivo, has a single change in the anticodon from CCA to CUA implicating the U in the middle position (45). In vitro, the su^+7 tRNA is a better substrate for the glutamine AAS by about six orders of magnitude (60). Changing the anticodon of *E. coli* $tRNA^{Met(f)}$ from CAU to CUA leads to misacylation by the glutamine AAS in vitro, at levels similar to that seen for su^+7 tRNA (44). The su^+3 allele of *E. coli* $tRNA^{Tyr}$ does not mischarge in vivo but elevated levels of the glutamine AAS can lead to mischarging in vitro (61).

We also anticipate that the glutamine AAS will interact with the 3' end of the tRNA. Mutants of su^+3 $tRNA^{Tyr}$ with alterations at the 3' end result in recognition by the glutamine AAS. One of these mutants, $A_{73} \rightarrow G$, is particularly effective (21-23). $tRNA^{Gln}$ has a G at position 73. Rogers and Söll have also described mutational changes of the $tRNA^{Ser}$ in the acceptor base pairs 1-72 and 2-73, which convert this tRNA to glutamine specificity in the DHFR assay (62). Note that these are also positions recognized by the serine AAS so competition between the two enzymes was decreased by making these changes.

The Variable Pocket - A Potential Site in tRNA for AAS Recognition.

In discussing the structure of yeast $tRNA^{Phe}$, Klug pointed out that the cluster of nucleotides formed by the interaction of the D-loop and T-loop forms a patch that arches out from the molecular surface of the molecule (63). The nucleotides involved in this patch are 16, 17, 20, 59, and 60 (Figure 1). The patch is variable in that these nucleotides are

not conserved among different tRNAs, and there can be insertions or deletions in the D-loop which could potentially change the configuration of this pocket. Klug explicitly suggested that this variable pocket may "form part of a recognition system for different tRNAs, perhaps for sorting into classes for synthetase discrimination."

We have already seen that G_{20} in yeast $tRNA^{Phe}$ is a recognition element for the phenylalanine AAS (58). In an analysis of the six *E. coli* $tRNA^{Arg}$ sequences, McClain observed a strong correlation of A_{20} with arginine identity. Consequently, a phenylalnine to arginine identity swap was attempted (64). The A_{20} was inserted into the phenylalanine D-loop causing U_{20} to become $U_{20.1}$. This suppressor was not active. However, with the variable pocket model in mind, residues 16, 17, and 59 were altered in an attempt to create the variable pocket characteristic of $tRNA^{Arg}_{ACG-1}$. This suppressor inserts 75% arginine, 5% lysine and 6% tyrosine. It is a significantly better arginine suppressor than one constructed simply by changing the anticodon of $tRNA^{Arg}_{ACG-1}$ to CUA, which inserts 37% Arg and 55% lysine. Insertion of A_{20} and A_{59} alone had virtually the same effect. Taken together with the $tRNA^{Phe}$ results, this finding certainly highlights residue 20 as an important synthetase recognition element. Given the exposed configuration of the variable pocket, other synthetases may well use different residues in this group as recognition elements.

tRNA Specificity Defined by Nucleotide Modification.

In general, tRNA modifications do not play a large role in tRNA identity. That this is so is most clearly demonstrated by the near normal substrate activity of unmodified transcripts. Indeed, even a $tDNA^{Phe}$ can be specifically aminoacylated (65). There is, however, a very interesting case in which a tRNA modification plays a crucial role not only in AAS recognition but in codon recognition as well. *E. coli* $tRNA^{Ile}_2$ recognizes the codon AUA. It was therefore a surprise when it was discovered that the anticodon in the DNA sequence of the $tRNA^{Ile}_2$ gene is CAT (66), the anticodon for methionine. The same sequence was found in the bacteriophage T4 $tRNA^{Ile}$ gene (67, 68), and in a spinach chloroplast $tRNA^{Ile}$ gene (69). Recall that C_{34} in $tRNA^{Met}$ is a crucial recognition element. Thus, we have the dilemma that this $tRNA^{Ile}$ should not only miscode, but it should also mischarge. The dilemma was resolved by the discovery that the C in $tRNA^{Ile}_2$ is modified to lysidine (Figure 5). This alteration affects recognition by the synthetase; the unmodified tRNA is recognized by the methionine AAS and the modified tRNA by the isoleucine AAS (66).

Lysidine also changes the coding capacity of the tRNA so that it recognizes AUA and not AUG. Why is this bizarre route taken, rather than simply making that anticodon UAU? Muramatsu *et al*, suggest that the evolution of the lysidine mechanism may have allowed a change in the genetic code converting AUA from methionine to isoleucine specificity. Once the mechanism was fixed it was retained except in some mitochondria where AUA has reverted to methionine specificity. Muramatsu *et al*, suggest that this was due to the loss of the lysidine biosynthetic mechanism (66) .

CONCLUSION

Much progress has been made recently with the application of two powerful methods to the study of tRNA identity. The in vitro approach assesses the contribution of specific nucleotides to tRNA identity by examining the aminoacylation kinetics of specifically altered tRNAs, providing a very clear picture of recognition elements for a given tRNA. Adding to this approach, determination of the in vivo specificity of altered tRNAs provides clues into the physiological relevance of specific nucleotides, i.e., what effect they have upon the competition between all 20 AAS. The experiments reviewed here have revealed the following: a) tRNA identity is defined by a relatively small number of elements, and b) the anticodon appears to be an important identity element for a majority of tRNAs.

Identity elements may vary in the extent of their evolutionary conservation. For example the G_3-U_{70} base pair so important for alanine identity is present in virtually every $tRNA^{Ala}$ sequenced to date, over a wide phylogenetic spectrum. The same appears to hold true for the A_{20} found to be involved in arginine identity. The determinants of serine tRNA identity, however, are conserved only in bacterial genes.

There are likely to be classes of tRNAs in regards to AAS recognition. On the simplest level, tRNAs can be classified according to the effect that an alteration in the anticodon has upon AAS recognition. Cognate recognition is either a) normal, or b) diminished/abolished. Within these subsets further divisions can be made. For example, both *E. coli* $tRNA^{Met(m)}$ and yeast $tRNA^{Phe}$ fall into the latter class. Yet for $tRNA^{Met(m)}$ the anticodon appears to be the sole identity element, while for yeast $tRNA^{Phe}$ two other nucleotides are important for complete phenylalanine acceptance. Similarly, *E. coli* $tRNA^{Ser}$ and $tRNA^{Ala}$ fall into class one, i.e., the anticodon does not appear to play a role in cognate AAS recognition. While eight identity elements have been identified for $tRNA^{Ser}$, a single basepair is primarily responsible for $tRNA^{Ala}$ identity.

The phenomena of lysine and glutamine mischarging remain a mystery. The composition of the anticodon must in large part direct mischarging by these two AASs (e.g. a U at position 35 is a recurring theme in glutamine misacylation), but, as they do not misacylate universally, there must be identity determinants or deterrents outside of the anticodon that dictate which tRNAs can be readily misacylated. Both of these AASs may have rather loose specificity, and, subsequently, behave as "default" AAS for tRNAs that are not well charged by their cognate AAS. Sorting out identity determinants from deterrents will no doubt prove to be a difficult task; however, a systematic application of both the in vitro and in vivo approaches will go a long way towards solving the problem.

Recent progress in the crystallography of tRNA-AAS complexes (8, 9) promises to give insight into the interaction between the two molecules, and will undoubtedly suggest new pathways to take in the study of the recognition problem.

ACKNOWLEDGEMENTS

We wish to thank Jeff Sampson, Olke Uhlenbeck, LaDonne Schulman, and Bill McClain for sharing their results prior to publication, and to David Horowitz, Olke Uhlenbeck and Jeff Sampson for critical reading of the manuscript. The authors were supported by grants from NIH (GM32637) and ONR (N00014-86-K-0755).

REFERENCES

1. Crick, F. H. C. *J. Mol. Biol.* 1966. 19:548-55
2. Sprinzl, M., Hartmann, T., Meissner, F., Moll, J., Vorderwülbecke, T. 1987. *Nucleic Acids Res.* 15:r53-r188
3. Holley, R. W., Apgar, J., Everett, G. A., Madison, J. T., Marquisee, M., *et al*, 1965. *Science.* 147:1462-65
4. Kim, S. H., Sussman, J. L., Suddath, F. L., Quigley, G. J., McPherson, A. *et al.* 1974. *Proc. Natl. Acad. Sci. USA* 71:4970-74
5. Klug, A., Robertus, J. D., Ladner, J. E., Brown, R. S., Finch, J. T. 1974. *Proc. Natl. Acad. Sci. USA* 71:3711-15
6. Schimmel, P. R., and Söll, D. 1979 *Ann. Rev. Biochem.* 48:601-48.
7. Schimmel, P. R. 1987. *Ann. Rev. Biochem.* 56:125-58.
8. Podjarny, A., Rees, B., Thierry, J .C., Cavarelli, J., Jesior J. C. *et al*, 1987. *J. Biomol. Struct. Dynam.* 5:187-98.

9. Perona, J. J., Swanson, R., Steitz, T. A., Söll, D. 1988. *J. Mol. Biol.* in press.
10. Fersht, A. R., Shi, J. P., Knill-Jones, J., Lowe, D. M., Wilkinson, A. J., *et al*, 1985. *Nature*. 314:235-38
11. Zelwer, C., Risler, J. L., Brunie, S. 1982. *J. Mol. Biol.* 155:63-81
12. Goodman, H. M., Abelson, J., Landy, A., Brenner, S., Smith, J. D. 1968. *Nature*. 217:1019-24
13. Steege, D. A., Söll, D. 1979. in *Biological Regulation and Development* ed. R. F. Goldberger, I:433-85. New York, Plenum Press
14. Smith J. D. in *Nonsense Mutations and tRNA Suppressors*. 1979. eds J. E. Celis, J. D. Smith, pp.109-25. London, Academic Press 349 pp.
15. Wilson, J. H., Abelson, J. N. 1972. *J. Mol. Biol.* 69:57-73
16. McClain, W. H., Wilson, J. H., Seidman, J. G. 1988. *J. Mol. Biol.* 203: in press.
17. Kurjan, J., Hall, B. D., Gillam, S., Smith, M. 1980. *Cell*. 20:701-9
18. Altman, S. in *Nonsense Mutations and tRNA Suppressors* 1979. eds J. E. Celis, J.D. Smith, pp. 173-89. London, Academic Press 349 pp.
19. Willis, I., Hottinger, H., Pearson, D., Chisholm, V., Leupold, U. *et al* 1984 *EMBO* 3:1573-80
20. Nishikura, K., Kurjan, J., Hall, B. D., DeRobertis, E. M. 1982. *EMBO*. 1:263-68
21. Hooper, M. L., Russell, R. L, Smith, J. D. 1972. *Febs Lett*. 22:149-56
22. Shimura, Y., Aono, H., Ozeki, H., Sarabhai, A., Lamfrom, H., Abelson, J. 1972. *Febs Lett*. 22:144-48
23. Ghysen, A., Celis, J. E. 1974. *J Mol. Biol.* 83:333-51
24. Khorana, H. G. 1979. *Science* 203:614-25
25. Normanly, J., Ogden, R.C., Horvath, S.J., Abelson, J. 1986. *Nature*. 321:213-19.
26. Masson, J.-M., Miller, J.H. 1986. *Gene*. 47:179-83.
27. Normanly, J., Masson, J. -M., Kleina, L. G., Abelson, J., Miller, J. H. 1986. *Proc. Natl. Acad. Sci. USA*. 83:6548-52
28. Bolin, J. T., Filman, D. J., Matthews, D. A., Hamlin, R. C., Kraut, J. 1982. *J. Biol. Chem*. 257:13650-662
29. Filman, D. J., Bolin, J. T., Matthews, D. A., Kraut, J. 1982. *J. Biol. Chem*. 257:13663-672

30. Sampson, J. R., Uhlenbeck, O. C. 1988. *Proc. Natl. Acad. Sci. USA.* 85:1033-37
31. Samuelsson, T., Boren, T., Johansen, T. -I., Lustig, F. 1988. *J. Biol. Chem.* 263:13692-9
32. Sampson, J. R., Sullivan, F. X., Behlen, L. S., DiRenzo, A. B, Uhlenbeck, O. C. 1987. *Cold Spring Harbor Symp. Quant. Biol.* 52:267-79
33. Brown, R. S., Dewan, J. C., Klug, A. 1985. *Biochemistry.* 24:4785-801
34. Dirheimer, G., Ebel, J. -P., Bonnet, J., Gangloff, J., Keith, G., *et al.* 1972. *Biochimie.* 54:127-44
35. Engelhardt, W. A, Kisselev, L. L. 1966. in *Current Aspects of Biochemical Energetics*. eds. N. O.Kaplan, E. O.Kennedy, pp.213-15 New York, Academic Press
36. Kisselev, L. L. 1985. in *Progress in Nucleic Acids Research and Molecular Biology*. eds. W. E. Cohn, K. Moldave, v32:237-66. Florida, Academic Press.
37. Schulman, L. H. in *Transfer RNA: Structure, Properties and Recognition.* 1979. eds. P. R. Schimmel, D. Söll, and J. N. Abelson.pp.311-24. New York, Cold Spring Harbor Laboratory Press
38. Schulman, L. H., Pelka, H., Susani, M. 1983. *Nucleic Acids Res.* 11:1439-55
39. Schulman, L. H., Pelka, H. 1983. *Proc. Natl. Acad. Sci. USA.* 80:6755-59
40. Schulman, L. H., Pelka, H. 1984. *Fed. Proc.* 43:2977-80
41. Valenzuela, D., Leon O., Schulman, L. H. 1984. *BBRC.* 119:677-84
42. Schulman, L. H., Pelka, H. 1988. *Science* in press
43. Miller, J. H., Normanly, J., Masson, J. -M., Kleina, L. G., Abelson, J. 1987. in *Integration and Control of Metabolic Processes: Pure and Applied Aspects*.pp. 571-81. Cambridge University Press, England
44. Schulman, L. H. Pelka, H. 1985. *Biochemistry.* 24:7309-14
45. Celis, J. E., Coulondre, C. Miller, J. H. 1976 *J.Mol. Biol.* 104:729-34
46. McClain, W. H., Foss, K. 1988. *Science*, 240:793-96.
47. McClain, W. H., Foss, K. 1988. *J. Mol. Biol.* 202:697-709
48. McClain, W. H., Foss, K., Chen, Y. -M., Schneider, J. 1988. *Science.* in press
49. Hirshfield, I. N. Bloch, P. L., Van Bogelen, R. A., Neihardt, F. C. 1981. *J. Bacteriol.* 146:345-51
50. Hou, Y.-M., Schimmel, P. R. 1988. *Nature.* 333:140-45.
51. Prather, N. E., Murgola, E. J. Mims, B. H. 1984 *J. Mol. Biol.* 172:177-84
52. Atilgan, T., Nicholas, H. B., McClain, W. H. 1986. *Nucleic Acids Res.* 14:375-80

53. McClain, W. H., Nicholas, H. B. 1987. *J. Mol. Biol.* 194:635-42

54. McClain, W. H., Guerrier-Takada, C., Altman, S. 1987. *Science.* 238:527-30

55. Leberman, R., Berthet-Colominas, C., Cusack, S., Härtlein, M. 1987. *J. Mol. Biol.* 193:423-25

55a. Thiebe, R., Harbers, K., Zachau, H. G. 1972. *Eur. J. Biochem.* 26:144-52

56. Bruce, A. G., and Uhlenbeck, O.C. 1982. *Biochemistry.* 21:3921-27

57. Bare, L., and Uhlenbeck, O.C. 1985. *Biochemistry.* 24:2354-60

58. Sampson, J. B., DiRenzo, A., Behlen, L., Uhlenbeck, O. C. 1988. *Science.* in press

58a. Uhlenbeck, O. C., Wu, H. -N., Sampson, J. R. 1987. in *Molecular Biology of RNA, New Perspectives.* eds. M. Inouye, B. S. Dudock, pp. 285-94. Academic Press

59. Roe, B., Sirover, M., Dudock, B. 1973. *Biochemistry.* 12:4146-54

60. Yarus, M., Knowlton, R., Soll, L. 1977. in *Nucleic acid-protein recognition,* pp. 391-408 New York, Academic Press

61. Hoben, P., Uemura, H., Yamao, F., Cheung, A., Swanson, R., *et al,* 1984. *Fed. Proc.* 43:2972-76

62. Rogers, M. J., and Söll, D. 1988. *Proc. Natl. Acad. Sci. USA.* in press

63. Ladner, J. E., Jack, A., Robertus, J. D., Brown, R. S., Rhodes, D., *et al* 1975. *Proc. Natl. Acad. Sci. USA.* 72:4414-18

64. McClain, W. H., Foss, K. 1988. *Science.* 241:1804-7

65. Khan, A. S., Roe, B. 1988. *Science.* 241:74-9

66. Muramatsu, T., Nishikawa, K., Nemoto, F., Kuchino Y., Nishimura, S. *et al.* 1988. *Nature.* in press.

67. Fukada, K., Abelson, J. 1980. *J. Mol. Biol.* 139:377-91

68. Mazzara, G. P., Plunkett, G., III, McClain, W. H. 1981. *Proc. Natl. Acad. Sci. USA.* 78:889-92

69. Kashdan, M. A., Dudock, B. S. 1982. *J. Biol. Chem.* 257:11191-94

70. Thorbjarnardottir, S., Dingermann, T., Rafnar, T., Andresson, O. S., Söll, D. *et al,* 1985. *J. Bacteriol.* 161:219-22

71. Murgola, E. J. 1985. *Ann. Rev. Genetics.* 19:57-80

72. Murgola, E. J., Hijazi, K. A. 1983. *Mol. Gen. Genet.* 191:132-37

73. Murgola, E. J., Prather, N. E., Pagel, F. T., Mims, B. H., Hijazi, K. A. 1984. *Mol. Gen. Genet.* 193:76-81

Table 1. Kinetic parameters for aminoacylation of tRNAs with *E.coli* methionine and valine AAS.

tRNA	Anticodon	V/K_m	Relative V/K_m
Methionine AAS			
Met(f)*	CAU	1.3	$7x10^5$
Met(m)*	CAU	3.7	$2x10^6$
Met(m)	CAU	1.7	$9x10^5$
Val	CAU	1.3	$7x10^5$
Val	UAC	$2x10^{-6}$	1.0
Trp	CAU	0.1	$5x10^4$
Trp	CCA	$2x10^{-6}$	1.0
Valine AAS			
Val*	UAC	4.0	$5x10^5$
Val	UAC	1.5	$2x10^5$
Met(m)	UAC	0.16	$2x10^4$
Met(m)	CAU	$8x10^{-6}$	1.0

*Native tRNA

Data from (42).

Table 2. Amber suppressor tRNA genes in *E. coli*[*].

Derived by genetic means	Derived by synthetic means	
Ser (su^+1)[a]	Ala2	Ile1
Gln (su^+2)[a]	Arg	Ile2
Tyr (su^+3)[a]	AspM	Leu5[g]
Leu (su^+6)[b]	Asn[#]	Lys
Trp (Su^+7)[c]	Cys[f]	Met(m)
GlyU[d]	GlyU	Phe[f]
GlyT[e]	GlyT	Pro
	GluA	Thr2
	HisA	Val

*The specificity of these tRNAs is not indicated.

[#]Not a functional suppressor, specificity unknown.

From (Normanly *et al*, unpublished) unless otherwise indicated.

[a](13)

[b](70)

[c](45)

[d](71)

[e](72, 73)

[f](27)

[g](25)

Table 3. tRNAs classified by the effect of a CUA upon AAS recognition.

Class I No affect upon charging	Class II Glutamine AAS mischarging	Class III Lysine AAS mischarging
Ala2	Ile1	Ile2
GlyU	GlyT	Arg
Cys	Met(f)*	Met(m)
Phe	GluA	AspM
ProH	Trp (su^+7)	Thr2
HisA		Val
Lys		
Ser (su^+1)		
Gln (su^+2)		
Tyr (su^+3)		
Leu (su^+6;also synthetic Leu5)		

* Demonstrated in vitro only (44)

Table 4. Conversion to $tRNA^{Ala}$ identity.

Starting tRNA identity (*E.coli*)	Nucleotide changes	tRNA identity (% amino acid inserted)
Cys	$(C_3\text{-}G_{70})\rightarrow$ (G-U)	Ala,100[a]
Gly	$C_{70}\rightarrow U$	Gln,95;Gly,5[b]
Phe	$(C_3\text{-}G_{70})\rightarrow$ (G-U)	Ala,63;Phe,37[a] Phe,76;Ala,24[b]
Phe	$(C_3\text{-}G_{70})\rightarrow$ (G-U); $U_{16}\rightarrow C$; $C_{17}\rightarrow U$; $U_{20}\rightarrow G$; $U_{60}\rightarrow C$; $U_{51}\rightarrow C$	Ala,94;Lys,6[b]
Lys	$C_{70}\rightarrow U$	Ala,95[c]

a (50)
b (46)
c (48)

Table 5. Aminoacylation kinetics of wildtype and mutant $tRNA^{Phe}$ transcripts.

tRNA	Nucleotide alterations	Relative kcat/Km
yeast Phe	wildtype	1.0
E. coli Phe	wildtype	.04
E. coli Phe	$U_{20}\rightarrow G$	.52
yeast Met→ Phe	$A_{20}\rightarrow G$; $C_{34}\rightarrow G$; $U_{36}\rightarrow A$; $(G_{49}\text{-}C_{65})\rightarrow$(C-G); $A_{59}\rightarrow U$;	.68
yeast Arg→ Phe	$C_{20}\rightarrow G$; $U_{34}\rightarrow G$; $C_{35}\rightarrow A$; $U_{36}\rightarrow A$; $C_{59}\rightarrow U$; $C_{73}\rightarrow A$;$U\rightarrow_{17}$*	.64
yeast Tyr→ Phe	$(U_{20},U_{20\text{-}1},U_{20\text{-}2})\rightarrow G$; $A_{13}\rightarrow C$; $A_{22}\rightarrow G$; $(C_{12}\text{-}G_{23})\rightarrow$ (U-A); $(C_1\text{-}G_{72})\rightarrow$(G-C); $U_{35}\rightarrow A$; $A_{46}\rightarrow G$; $(U_2\text{-}A_{71})\rightarrow$(C-G); $(C_3\text{-}G_{70})\rightarrow$(G-C)	1.5

*Represents an insertion

Data from (58)

FIGURE LEGENDS

Figure 1. Secondary and tertiary structure of tRNA. In the cloverleaf structure (left) invariant nucleotides that are involved in stabilizing tertiary interactions are indicated; the lines indicate tertiary interactions between bases. Dots indicate base pairing. (Right) The L-model for tRNA tertiary structure. (4, 5)

Figure 2. Cloverleaf secondary structure of $tRNA^{Ala}$. The G-U basepair at position 3-70, along with bases 20 and 60 (indicated by open circles) were predicted to be important alanine identity elements by statistical analysis (46).

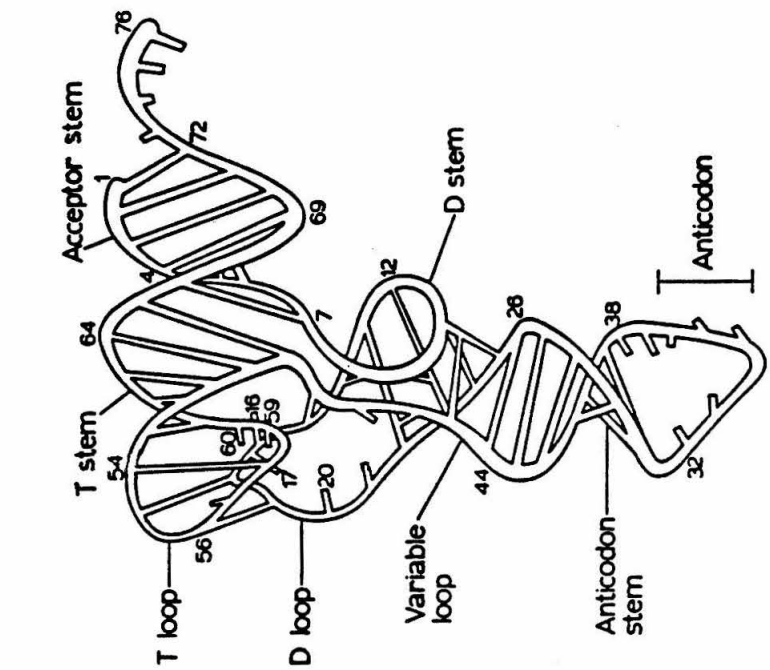
Figure 3a. Composite *E. coli* $tRNA^{Ser}$. Bases highly conserved in serine tRNAs and not found in $tRNA^{Leu}_{5CUA}$ are indicated.

Figure 3b. Mutations introduced into $tRNA^{Leu}_{5CUA}$ to effect the conversion from leucine to serine identity result in the mutant $tRNA^{Leu \rightarrow Ser}$.

Figure 3c. The minimum number of nucleotides required to effect the leucine to serine conversion (25; Normanly & Abelson, unpublished).

Figure 4. Nucleotides postulated to be important for yeast $tRNA^{Phe}$ identity (58).

Figure 5. Chemical structure of Lysidine.


Acceptor stem
T stem
T loop
D loop
D stem
Variable loop
Anticodon stem
Anticodon

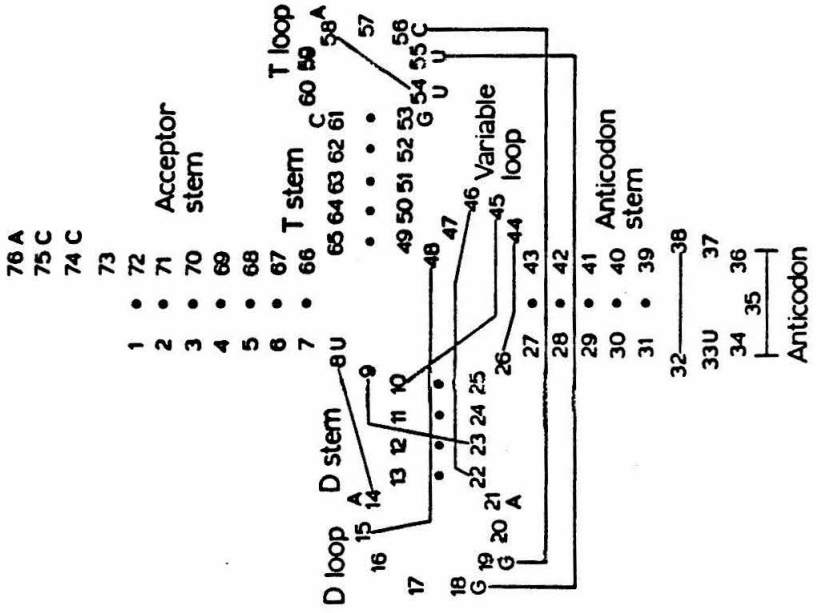

Acceptor stem
T stem
T loop
D loop
D stem
Variable loop
Anticodon stem
Anticodon

Figure 1

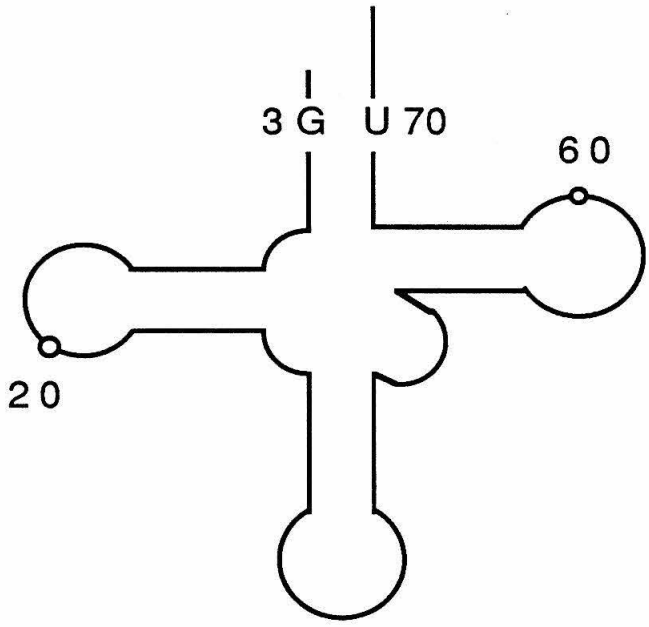


Figure 2

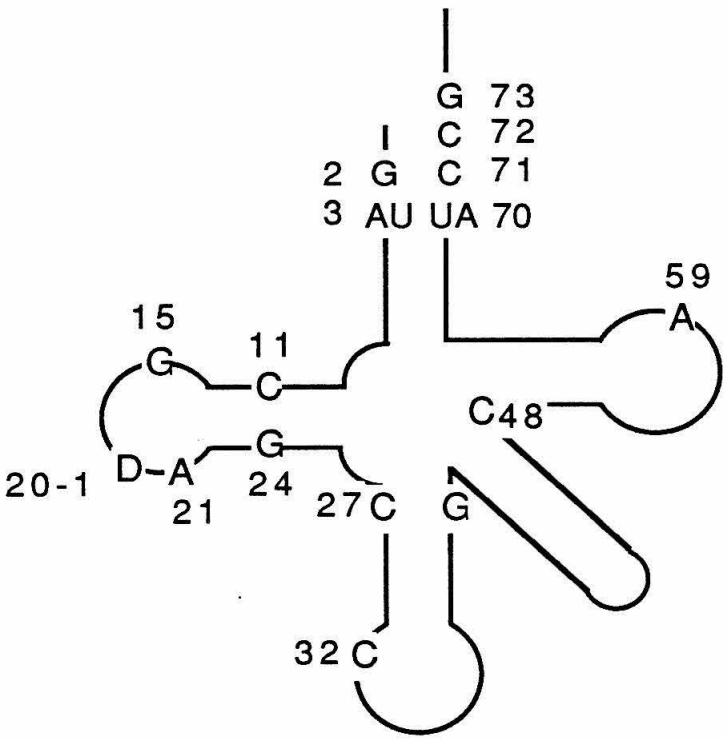


Figure 3a

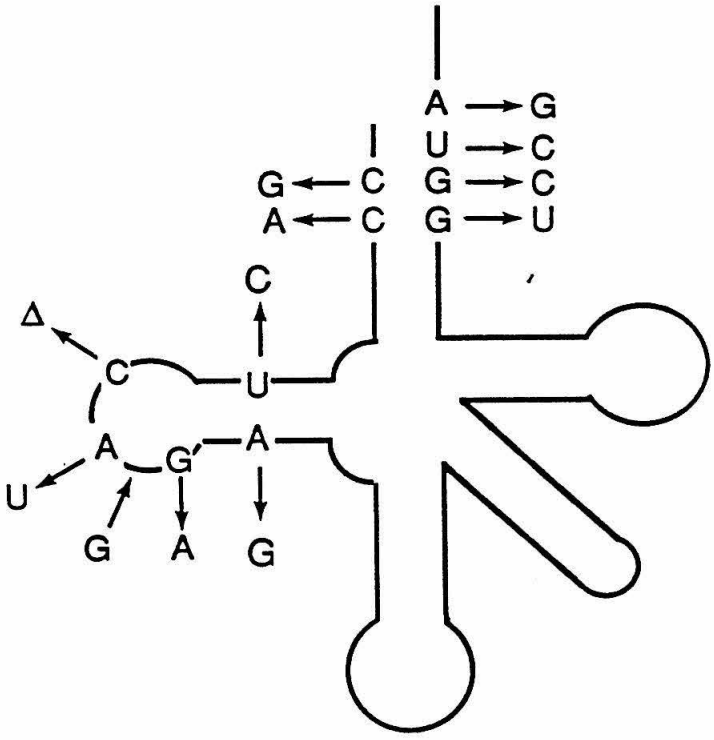


Figure 3b

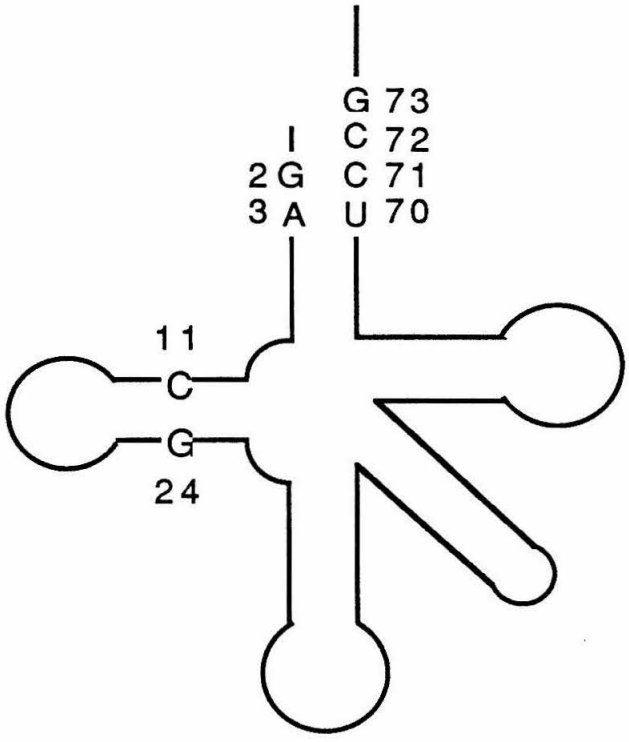


Figure 3c

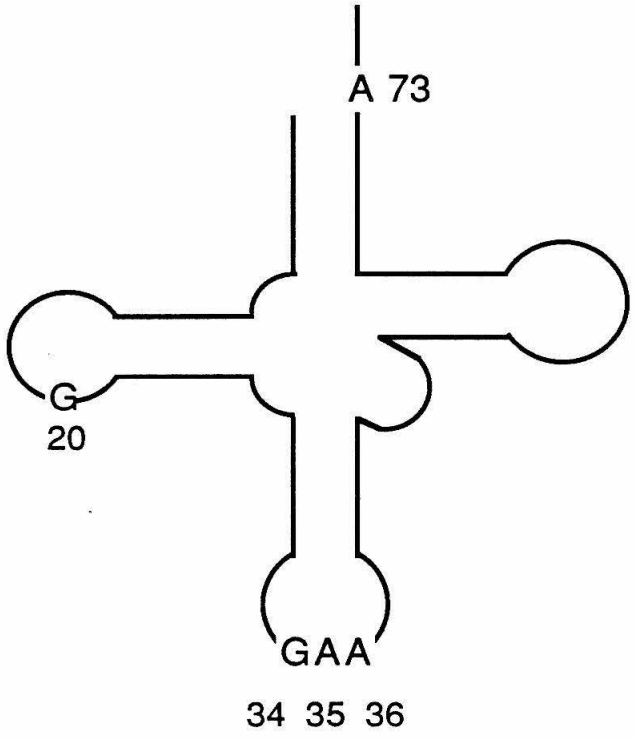


Figure 4

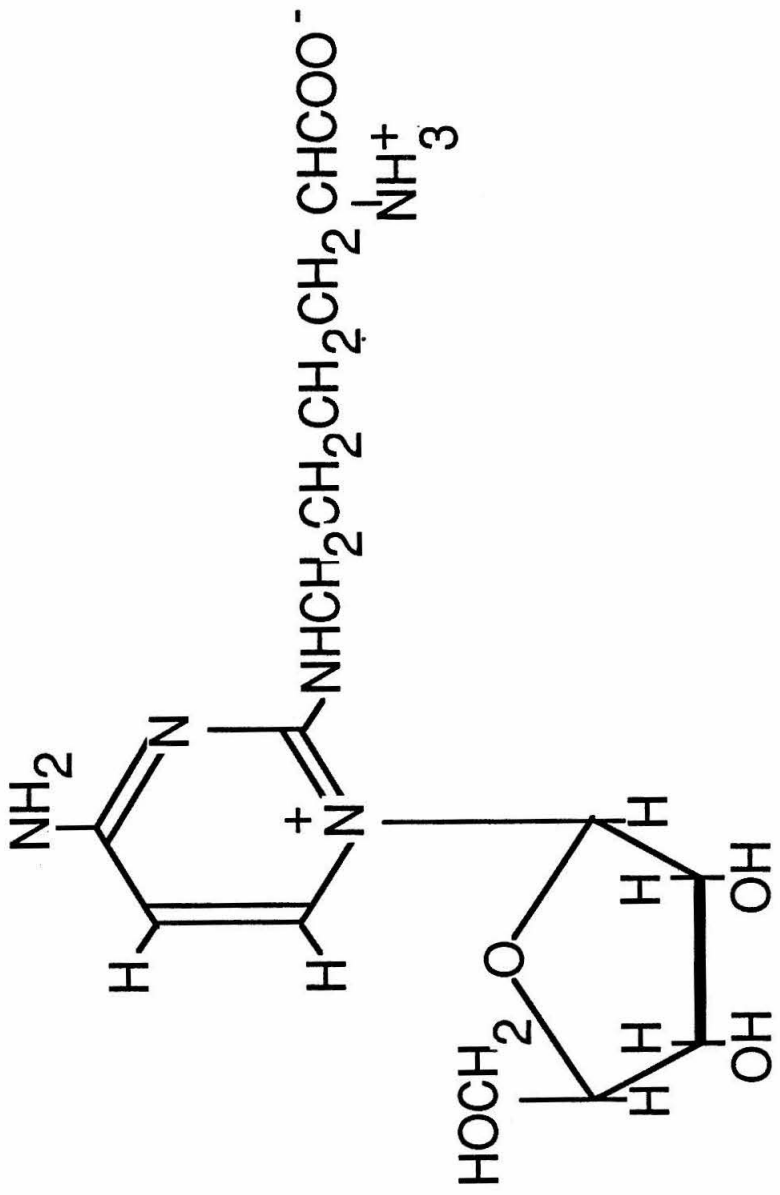


Figure 5

CHAPTER I

Changing the identity of a tRNA

Jennifer Normanly, Richard C. Ogden[1], Suzanna J. Horvath and John Abelson

Division of Biology, California Institute of Technology, Pasadena, California 91125, USA
[1]Agouron Institute, 505 Coast Boulevard South, La Jolla, California 92037 USA

[Nature (1986) 321:213-219]

ABSTRACT

A leucine transfer RNA has been transformed into a serine transfer RNA by changing 12 nucleotides. This result indicates that a limited set of residues determine tRNA identity.

INTRODUCTION

The rules for specificity in protein synthesis are to be found in the specific interactions between transfer RNA molecules and aminoacyl tRNA synthetases (AAS) (1). There exists a single AAS for each of the 20 amino acids, but because of the degeneracy of the genetic code and the limited range of allowed codon-anticodon interactions (2), there are generally multiple tRNAs specifying each amino acid. The accurate transfer of information from messenger RNA to protein therefore relies on each AAS recognizing and acylating its cognate set of tRNAs while failing to acylate all others. All tRNAs so far sequenced (and the number is greater than 300) (3), can be folded into the familiar cloverleaf structure (4) and share a set of absolutely conserved nucletotides. The X-ray crystal structures of several tRNAs (5-8) show that these constant nucleotides are involved in stabilizing the tertiary structure of the tRNA molecule, therefore, by extrapolation, the tertiary structures in the cell are very similar. In some way, within the variations allowed by this framework, there exists a set of distinguishing features that constitute amino-acid specificity, that is, the identity of each tRNA.

By contrast, aminoacyl tRNA synthetases are quite variable in structure; they range in size from 300 to 1,000 amino-acid residues and are found in various quaternary forms (9). Despite the variability of these enzymes, a unifying model has been proposed for the interaction between the tRNA molecules and AAS. Summarizing diverse chemical and physical studies involving several cognate complexes (10) (see also refs 9, 11, 12), Rich and Schimmel proposed that all AAS interact with tRNA along the inside diagonal of the L structure (5,6); the regions of contact inferred from these studies extend from the acceptor stem through the D stem to the anticodon loop. While the range of AAS contacts may be similar for all tRNA substrates, the rules for recognition lie in the detailed interactions and these must, of course, be idiosyncratic. For example, while the anticodon is clearly an important recognition element for the *Escherichia coli* $tRNA^{fMet}$ synthetase (13-15), this seems less likely to be the case for the serine and leucine synthetases because there are six

codons specifying each of their respective amino acids, requiring a set of isoacceptor tRNAs with different anticodons.

A genetic approach to understanding AAS-tRNA recognition was taken over 15 years ago (16-19). In those experiments, a genetic selection was devised to isolate mutant tRNAs having altered specificity. In the initial experiments, the *E. coli* amber suppressor gene *supF* was used. This suppressor is a mutant allele of one of three $tRNA^{Tyr}$ genes in *E. coli* and can recognize the amber codon UAG because it has an altered anticodon, CUA (20) (for a review of nonsense suppression, see ref. 21). Among the amber mutants in bacterial and phage genes, there are some that can be suppressed by *supD* (serine) and *supE* (glutamine) but not by *supF*. Selection of *supF* mutants that could suppress this class of amber resulted in $tRNA^{Tyr}$ alleles with altered specificity. The mutations occurred in the first and second base pairs of the acceptor stem and in the fourth nucleotide from the 3' end of the tRNA (16,17,22-25). In all cases, the amino acid inserted by the mutant tRNAs was glutamine. This interesting result was somewhat obscured later by the finding that amber suppressor alleles of the $tRNA^{Trp}$ gene also insert glutamine (26). One interpretation of these results is that the glutamine synthetase is indiscriminate and charges substrates that are not recognized by other synthetases. Other alterations in $tRNA^{Tyr}$ specificity were never observed, perhaps because too many changes are required. The approach remains an attractive one, however, because the alteration of identity with retention of tRNA function, by definition, reveals elements that determine specificity.

The advances in oligonucleotide synthesis allowing rapid gene synthesis (27) have prompted a return to the genetic approach to studying AAS-tRNA interaction. If multiple changes are required to alter the specificity of a tRNA, they cannot be selected but they can be constructed. This is the approach we have taken here.

METHODS

Strains, Media, and Enzyme Assays. JM101is (F' *traD36, proAB, lacI^{q}Zm15/Δ(lac proB) supE thi*). XAC-A16 is (F'*lacI$_{am}$-Z/argE$_{am}$, nalA, rif, Δ(lac proB) metB, ara*). The strains carrying *supD, supE, supF,* and *supP* are isogenic; (*lac, proB, metB, argE$_{am}$, rif, nalA, ara)* . MC1061/9100.1 is F'(*lacI^q'ZYA tn5/ lac, galU, galK, rpl Δ(ara ABIOC-leu$_{7679}$)hsr$^-$hsm$^+$*). L broth was made from 10g/l Bactotryptone, 5g/l yeast extract, 8g/l NaCl. L plates contained 15g/l Bacto agar as well. β-lactamase activity was

assayed as described by Dalbadie-McFarland *et al.* (51) using ampicillin as a substrate. β-galactosidase activity was assayed as described by Miller (44).

Gene Synthesis. Deoxyoligonucleotides were synthesized on an Applied Biosystems 380A DNA synthesizer. Three to five A_{260} units of oligonucleotide were electrophoresed through a 20% acrylamide/7M urea gel, visualized by ultraviolet shadowing, excised and eluted in 0.3M NaCl, 10mM Tris-HCl pH 7.4, 1mM EDTA, 1% phenol at 37°C overnight. The DNA was precipitated with 100% ethanol, washed with 70% ethanol, dried and resuspended in dH_2O. Phosphorylations were carried out on 2ug of each oligonucleotide in 70mM Tris-HCl pH 7.6, 10mM $MgCl_2$, 5mM dithiothreitol (DTT), 100uM ATP, 100uCi ^{32}P-ATP, with four units of polynucleotide kinase for 1 h at 37°C, followed by phenol/chloroform extraction and ethanol precipitation. 80 pmol of each phosphorylated oligonucleotide were incubated together in 100mM NaCl at 80°C for 5 min, then allowed to cool to room temperature over a period of 3 hr. The mixture was made 50mM in NaCl and combined with digested vector in a 10:1 insert to vector (mass) ratio. The vector in this case was M13mp8 containing a synthetic *rrnC* terminator (40) inserted into the Pst I-Hind III sites. Use of this transcription terminator was suggested by the work of Yarus *et al.*(60). Ligation was carried out in 50mM Tris-HCl, pH 7.6, 10mM $MgCl_2$, 20mM DTT, lmM ATP, 50ug/ml bovine serum albumin with 1 unit of T4 DNA ligase at 15°C for 12 hr. The ligation mixture was used to transform competent JM101. Single-stranded DNA from the resulting plaques was immobilized on nitrocellulose filters (61) and screened for the presence of inserts by first prehybridizing in 10x Denhardt's reagent (62), 0.2% SDS, 6x SSC (63), 1mM EDTA for 2 h at 65°C, then hybridizing with $4x10^5$ Cerenkov c.p.m./ml of one of the internal oligonucleotides in 5x DEnhardt's reagent, 6x SSC at 50°C for 12 hr. The filters were washed in 6x SSC at room temperature for 5 min, then at 50°C for 5 min and autoradiographed. Single-stranded DNA was prepared from positive candidates and sequenced according to the method of Sanger *et al.* (64).

Plasmid Construction. To construct plasmid pBR322-βam, the Eco RI-Pst I fragment of pBR322 was subcloned into M13mp8. Oligonucleotide-directed mutagenesis (65) of the β-lactamase gene was achieved using the 19-mer $^{5'}$TAAAAGTCTACATCATTGG$^{3'}$. After

verification of the mutant sequence, the Eco RI-Pst I fragment was transferred back into pBR322, yielding plasmid pBR322-βam. Pvu II fragments containing the synthetically constructed $tRNA^{Leu}_{UAG}$ or $tRNA^{Leu\rightarrow Ser}_{UAG}$ gene together with the M13mp8-derived *lac* promoter and rrnC terminator were subcloned into the Pvu II site of pBR322-bam to yield pBR322-βam:$tRNA^{Leu}_{UAG}$ and pBR322-βam:$tRNA^{Leu\rightarrow Ser}_{UAG}$, respectively.

To construct plasmid pDA3-12, the 19-mer $^{5'}$ACGTAAAGCTTAAACCGTC$^{3'}$ was used to create a Hind III site in between the ribosome binding site and Pribnow box of the *E. coli fol* gene (C. R. Morris, unpublished results) encoded by a 1 kilobase fragment contained in M13mp8 (66). This was done to facilitate subcloning of the gene into a *tac* expression vector. Subsequently, the 23-mer $^{5'}$GATAACGCGATTCTACGCTAACG$^{3'}$ was used for oligonucleotide-directed mutagenesis of the *fol* gene. The 247-base pair (bp) Eco RI-Hind III fragment of ptac12H (53) containing the *tac* promoter was ligated into the Eco RI-Hind III site of pBR322 to create pTAC10. Subsequently, the DHFR amber, encoded by a 1,057-bp Hind III fragment, was subcloned into the Hind III site of pTAC10, yielding pDA3-12. The 392-bp Pvu II fragment from pBR322-βam:$tRNA^{Leu}_{UAG}$ or pBR322-βam:$tRNA^{Leu\rightarrow Ser}_{UAG}$ containing the *lac* promoter, synthetic tRNA gene and *rrnC*-derived terminator, was subcloned into the Pvu II site of pDA3-12 to yield pDA3-12:Leu or pDA3-12:Leu→Ser.

DHFR Purification. Two liters of MC1061/9100.1 carrying pTAC101 (the same construct as pDA3-12, except that the *fol* gene is wild type) in L broth were incubated at 37°C with aeration to an A_{600} of 0.4-0.6, isopropyl thiogalactoside was added to 1 mM and the culture was incubated as before for an additional 4 hr. Cells were harvested and lysed (67) and the crude lysate was extracted with ammonium sulphate (68). The 90% ammonium sulphate-saturated insoluble fraction was brought to a protein concentration of 20 mg/ml and dialysed with 3 ml of methotrexate-Sepharose (69) overnight. The resin was batch-washed and DHFR eluted with folic acid (69). To remove nucleic acid and folate, peak fractions were dialized against 50mM potassium phosphate pH 8.0, 1mM DTT and applied to a DEAE-Sephacel column (1ml per mg protein) equilibrated in the same buffer. DHFR was eluted in a linear gradient of 0-0.4M KCl, dialized against water and lyophilized. Mutant DHFR was purified in a similar manner but on a larger scale (12 l for the $tRNA^{Leu\rightarrow Ser}_{UAG}$ substituted mutant) and the peak fractions from the DEAE column were passed through methotrexate affinity and DEAE columns a second time. Protein

sequence analysis was carried out on an Applied Biosystems Model 470A gas-phase protein sequenator (58) by the USC Microchemical Core Laboratory.

RATIONALE

To define the specific interaction between *E. coli* serine-specific tRNAs and their cognate AAS, we set out to transform a leucine-specific amber suppressor tRNA into an amber suppressor tRNA recognized exclusively by the serine AAS. We assume that there exists a limited set of changes one could make to a leucine tRNA that would entirely block its recognition by the leucine AAS and allow recognition solely by the serine AAS. The choice of this particular transformation was in part dictated by the fact that serine and leucine amber suppressor alleles exist in *E. coli*, and because all serine and leucine tRNAs fall into the type II structure class. Type II tRNAs have large extra loops in a stem-loop configuration and therefore lack the triple-base tertiary interactions, which are found in type I tRNAs (28). In addition, it is possible that the extra loop is a distinct structural feature recognized by the AAS for type II tRNAs, although this notion has not been addressed experimentally.

Six different $tRNA^{Ser}$ isoacceptors recognized by the *E. coli* serine AAS have been sequenced (five *E. coli* $tRNA^{Ser}$ species and a T4 $tRNA^{Ser}$; refs 29-34). Comparison of these sequences provides the information necessary to design the transformation. Such an approach was used previously to delineate sequences important in the recognition of $tRNA^{Phe}$ by the yeast phenylalanine AAS (9,35). Inspection of the serine tRNA sequences reveals 35 positions in which a nucleotide is absolutely conserved (Figure 1). It is likely that the specific residues that are necessary and sufficient for recognition of the set of $tRNA^{Ser}$ molecules by serine AAS are limited to these 35 nucleotides. Of these, 14 occur in all tRNAs (the constant nucleotides) and an additional seven are present in $tRNA^{Leu}_5$. The remaining 14, detailed in Figure 2a, are concentrated in the acceptor stem and D arm, together with one each in the TΨC loop, extra arm, and anticodon loop, and two comprising a base pair in the anticodon stem.

Figure 2b shows the proposed changes for the leucine to serine transformation, the justifications for which are as follows. We chose to focus only on the acceptor stem and D stem and loop as these domains have been shown to be in contact with the synthetase in all cases studied. In this initial exercise, we have ignored the TΨC stem-loop, the extra arm and the anticodon stem and loop. Although the anticodon has been strongly implicated in

some tRNA-AAS interactions (13-15,26,36,37), the existence of at least six isoaccepting species of serine tRNA, recognizing six different codons, and the existence of a serine tRNA suppressor allele, argues against a crucial role for the anticodon region in the case of serine tRNA recognition. Accordingly, the conserved $tRNA^{Ser}$-specific bases at positions 32 in the anticodon loop and the base pair 27-43 in the anticodon stem were not incorporated into the $tRNA^{Leu}_5$ framework. We have thus superimposed the conserved serine nucleotides at positions 2, 71, 72 and 73 in the acceptor stem as well as positions 20-1, and 21 in the D loop, and base pair 11-24 in the D stem.

Looking at probable tertiary interactions in the leucine and serine tRNAs, there are five base-pair interactions common to both type I and II tRNAs that appear to be responsible for stabilization of the tertiary structure (5,6,28); four of these, G_{18}-U_{55}, G_{19}-C_{56}, U_{54}-A_{58} and U_8-A_{14}, involve constant nucleotides, and are therefore not directly affected by the alterations we propose. The fifth, a transtertiary base pair between residues 15 and 48, is G_{15}-C_{48} in all serine tRNAs. As this tertiary interaction is supplied in $tRNA^{Leu}_5$ (albeit by an A-U pair), we elected not to alter bases 15 and 48. A difference in primary structure between $tRNA^{Ser}$ and $tRNA^{Leu}$ which could account for a subtle difference in tertiary structure between the two sets of tRNAs, lies in the position of the constant nucleotides G_{18} and G_{19} in the D loop. Note in Figure 1 how the positions of G_{18} and G_{19} in $tRNA^{Ser}$ are offset by one nucleotide in comparison with G_{18} and G_{19} in $tRNA^{Leu}_5$. The question at this point becomes whether or not to maintain rigorously the leucine tRNA structure, and merely superimpose the predicted serine determinants on the leucine scaffold, or to alter the nature of the tertiary interactions to resemble those of a serine tRNA. As the position of G_{18} and G_{19} is a highly conserved feature of the serine tRNAs, we felt it may well be important for serine AAS recognition, especially in view of its location within the structure. Through an insertion and a deletion in the D loop of $tRNA^{Leu}_5$, these bases are kept in the same position relative to the D stem that is seen in all six $tRNA^{Ser}$ sequences.

Finally, the G-C to A-U change in the third base pair of the acceptor stem was not dictated by absolute conservation of that sequence among serine tRNAs but was suggested by a statistical study of *E. coli* tRNA sequences (W. H. McClain and H. B. Nicholas, personal communication) in which it was found that an A-U or U-A base pair at that position is significantly correlated with all $tRNA^{Ser}$ sequences and is anti-correlated with

other tRNA sequences. The rest of the proposed changes are compatible with the significant serine-specific residues detected in McClain's study.

To effect this transformation of a leucine amber suppressor tRNA, we first synthesized a leucine amber suppressor tRNA gene and then altered it at the residues indicated in Figure 2b.

SYNTHESIS OF tRNA^{LEU} SUPPRESSOR GENE

The sequence of the starting leucine amber suppressor gene was based on the sequence of *E. coli* tRNA$^{Leu}_5$ with substitution of the anticodon to CTA. tRNA$^{Leu}_5$ recognizes the codon UUA and has a different sequence from the leucine amber suppressor *supP*, as became apparent from studies published after the present work was started (38,39). The synthetic gene, tRNA$^{Leu}_{UAG}$, is comprised of six overlapping oligonucleotides, which were annealed and ligated into the M13 vector mp8 containing a synthetic transcription terminator (see Figure 3). In this construction, transcription of the tRNA gene is directed from the *lac* promoter and is expected to terminate within the sequence derived from the ribosomal RNA cistron *rrnC* (40). No provision was made for the inclusion of natural 5' and 3' flanking sequences, as the nucleases that process tRNA precursor transcripts are not sequence-specific; instead, they recognize conserved features of secondary and tertiary structure common to all tRNAs (41). It was therefore anticipated that a tRNA could be correctly processed from any flanking sequence context.

To ascertain that the synthetic gene produced a functional tRNA, we assayed for suppression of an amber mutation in an *E. coli* strain carrying the tRNA$^{Leu}_{UAG}$ gene. The amber mutation resides in the N-terminal *lacI* portion of a *lacI-Z* fusion gene (42). Activity of the *lacZ* gene product, β-galactosidase, is unaffected by the *lacI* portion of the fusion, hence quantitation of β-galactosidase activity gives a measure of the extent of suppression of the amber mutation. The tRNA$^{Leu}_{UAG}$ gene, together with its terminator, was subcloned into plasmid pEMBL8$^+$ (ref. 43) and transformed into the *lacI-Z* amber strain XAC-A16. Suppression efficiency was determined directly by assaying for β-galactosidase activity (44). When tested in this way, the tRNA$^{Leu}_{UAG}$ suppressor is 60% efficient, producing substantial levels of β-galactosidase.

SYNTHESIS OF $tRNA^{LEU \rightarrow SER}_{UAG}$ GENE

For construction of the $tRNA^{Leu \rightarrow Ser}_{UAG}$ gene, four additional oligonucleotides were synthesized and combined with two of the $tRNA^{Leu}_{UAG}$ gene chains as shown in Figure 3. This gene was then transferred to the pEMBL8$^+$ vector and its activity assayed using the *lacI-Z* amber strain. The $tRNA^{Leu \rightarrow Ser}_{UAG}$ gene is active but produces a much weaker suppressor than its parent, exhibiting a 0.5-1% efficiency of suppression. By contrast, the *supD* suppressor (serine) is 35% efficient.

To determine the specificity of this suppressor, two approaches were taken. The first approach uses an amber mutant in the β-lactamase gene that is functionally suppressed only by a serine suppressor. This approach allows the selection of serine amber suppressors. If our design of a suppressor with serine identity had been wrong by one or two nucleotides, the serine-specific amber could have been used to select an active serine suppressor by spontaneous or induced mutation. Alternatively, this mutant could be used to select stronger variants of a weak serine suppressor. In the second approach the $tRNA^{Leu \rightarrow Ser}_{UAG}$ suppressor was used to suppress an amber mutation near the beginning of the *E. coli* dihydrofolate reductase gene *fol*, and subsequently the suppressed protein was sequenced.

SERINE-REQUIRING AMBER MUTATION

The active site of β-lactamase contains an essential serine residue (45,46), which participates in the cleavage of the β-lactam ring of ampicillin and thereby confers antibiotic resistance to cells harbouring the β-lactamase gene *amp*. By oligonucleotide mutagenesis we converted the AGC serine codon at residue 68 (residue 70 according to Ambler's nomenclature, ref. 47) of pBR322 *amp* gene to TAG (Figure 4a). Figure 4b shows that this mutation can be suppressed in an *E. coli supD* strain (*supD* inserts serine) but not by *supE* (glutamine), *supF* (tyrosine) or *supP* (leucine). The only other amino acids, which might be expected to substitute for serine at this position based on some knowledge of the reaction mechanism (48), are cysteine and threonine. These substitutions have been made (46,49-51) and in the case of cysteine resulted in an enzyme with 1/30 of wild-type activity. In the case of threonine, the enzyme was inactive (51). The active-site amber mutation therefore meets our requirements as apparently it can be functionally suppressed

only by a serine-inserting suppressor.

The synthetic suppressor genes $tRNA^{Leu}_{UAG}$ or $tRNA^{Leu\rightarrow Ser}_{UAG}$, together with the *lac* promoter and *rrnC* terminator, were ligated into the *PvuII* site of pBR322-βam. Each plasmid was introduced into *E. coli* and the transformants were tested for growth on tetracycline and ampicillin. Only the $tRNA^{Leu\rightarrow Ser}_{UAG}$ suppressor allowed growth on ampicillin (Figure 4c). Futhermore, the cells containing $tRNA^{Leu\rightarrow Ser}_{UAG}$-suppressed β-lactamase are fully viable in the presence of the drug at 80 μg ml^{-1}. We conclude that $tRNA^{Leu\rightarrow Ser}_{UAG}$ is a functional tRNA that inserts serine. This experiment cannot, however, exclude the possibility that the tRNA inserts other amino acids also. The only way to determine the specificity of suppression is to sequence through the site of the amber mutation in a suppressed protein.

$tRNA^{LEU\rightarrow SER}_{UAG}$ SPECIFICITY

To determine the specificity of the $tRNA^{Leu\rightarrow Ser}_{UAG}$ suppressor, we needed to use it for suppression of an amber mutation close to the beginning of a gene whose product is relatively small, easy to purify, and present at high concentration. The *E. coli* dihydrofolate reductase (DHFR) gene, *fol*, proved an attractive candidate. We linked the *fol* gene, the sequence of which is known (52), to the *tac* promoter (53) in the expression vector pTAC10 (Figure 5). The 159-amino acid DHFR protein represents approximately 10% of the total protein in fully induced strains of *E. coli* harbouring this plasmid, and is easily purified by methotrexate affinity chromatography. An amber mutation at a site near the N-terminus would allow sequence determination of the suppressed DHFR by Edman degradation (54) of the intact protein. Residues 2-8 constitute a strand of a β-pleated sheet and are part of the NADP binding site. We chose position 10, a valine residue found in a loop on the surface of the protein, because this residue is not phylogenetically conserved and X-ray crystallographic data reveal that it is well removed from the active site (55-57). Substitution of any residue at this site should not affect the purification of DHFR on a methotrexate affinity column. Another point to consider is that the efficiency of amber suppression and therefore the yield of suppressed protein is dependent on the context of the amber codon. UAG codons followed by A residue are usually better suppressed than those followed by a G (42). Consequently, by oligonucleotide-directed mutagenesis, we altered the DHFR sequence GTA-GAT of codons 10 and 11 to TAG-AAT, changing the protein

sequence from Val-Asp to amber-Asn. This altered *fol* gene, together with its *tac* promoter, was transferred to plasmids carrying either the $tRNA^{Leu}_{UAG}$ or $tRNA^{Leu \rightarrow Ser}_{UAG}$ genes (Figure 5a). Figure 5b documents the purification of DHFR from a strain harbouring the wild-type gene under control of the *tac* promoter (pTAC101).

Dihydrofolate reductase was purified from four strains: one carrying pTAC101, another a *supD* strain harbouring the DHFR amber mutant in the *tac* expression vector (pDA3-12), and the third and fourth carrying pDA3-12:Leu and pDA3-12:Leu→Ser, respectively. Each purified protein was sequenced by Edman degradation using an automatic gas-phase sequencer (58). In interpreting these results (Figure 6), it should be considered that the chromosomal copy of the *fol* gene is present in these strains, leading to fractional yield (usually small) of Val-Asp at positions 10 and 11. This sequence is very clearly seen when wild-type DHFR is sequenced but when the DHFR amber is suppressed by the $tRNA^{Leu}_{UAG}$ suppressor, the sequence is unambiguously Leu-Asn at positions 10 and 11. Overproduction of the plasmid-encoded DHFR is probably greater than 100-fold and the $tRNA^{Leu}_{UAG}$ suppressor is efficient. When the DHFR amber is suppressed by *supD*, serine is clearly present in position 10, demonstrating that the well known problems in detecting serine in Edman degradation have been overcome. Here, valine is clearly not detected at position 10 but there is some aspartate at position 11; we believe this results from deamidation of asparagine.

When the DHFR amber is suppressed by the $tRNA^{Leu \rightarrow Ser}_{UAG}$ suppressor, there is a good yield of serine at position 10, but also a detectable (12-15%) amount of valine. The presence of valine is due to the fact that the $tRNA^{Leu \rightarrow Ser}_{UAG}$ suppressor is weak, giving rise to a lower ratio of suppressed protein to wild-type protein. There is also a small amount (10%) of leucine at position 10; this is probably due to carry-over of the leucine from postion 8, but the experimental method used cannot exclude the possibility that the $tRNA^{Leu \rightarrow Ser}_{UAG}$ suppressor inserts small amounts of leucine or valine. We could detect no other amino acid at position 10, and conclude that we have changed a strong leucine suppressor to a weak serine suppressor.

DISCUSSION

Comparison of the sequences of six serine tRNAs, which are substrates for the *E. coli* serine AAS reveals several conserved nucleotides. Here we altered an amber suppressor allele of $tRNA^{Leu}_5$ at 12 positions to yield a tRNA sequence containing a subset of these conserved serine residues-those located in the acceptor stem and in the D stem and loop. The resultant molecule is a weak suppressor, which inserts serine in response to amber nonsense codons. The specific acquisition of $tRNA^{Ser}$ identity was established by a highly discriminatory functional assay, while the loss of $tRNA^{Leu}$ identity in exchange for unique acquisition of $tRNA^{Ser}$ identity was established by protein sequencing. Our study demonstrates that a limited number of nucleotides by virtue of their nature and position on a tRNA scaffold, are sufficient to dictate cognate synthetase recognition. It is noteworthy that this result was achieved with only primary and secondary structure information about the two tRNAs and no knowledge of the AAS structures.

A number of questions remain of course. We have converted a strong suppressor into a weak suppressor. This reduction in efficiency may be the result of a weak interaction between the suppressor tRNA and serine AAS. Alternatively, it may reflect a perturbation in the general tRNA structure resulting in poorer competition with the release factors. THere are two possible routes to producing a better suppressor with serine specificity. First, we can further alter the sequence by including those changes in the anticodon stem and TΨC loop that were conserved in all serine tRNAs, but which we chose to ignore. Our predictions about the importance of specific tertiary interactions in AAS recognition also warrant a more systematic investigation. Perhaps a better method, however, is to mutagenize the gene and select stronger serine suppressors by requiring suppression of the β-lactamase active-site amber in the presence of increasingly higher levels of ampicillin. When a strong serine suppressor is generated, the sequence of the tRNA will be determined. At that point, we can begin a systematic analysis of the identity switch to determine which changes were necessary. Eventually, the minimum set of changes required to transform $tRNA^{Leu}$ to $tRNA^{Ser}$ can be determined. This information will then be used to effect a different transformation, for example, $tRNA^{Tyr}$ to $tRNA^{Ser}$. It will clearly be necessary to gather supporting *in vitro* kinetic and thermodynamic data for the various tRNAs and AAS involved. This iterative method can lead to a more detailed

understanding of the essence of serine tRNA identity.

ACKNOWLEDGEMENTS

We thank Joe Kraut for assistance in the design of the DHFR amber mutant, Lynn Williams of the USC Microchemical Core Laboratory for protein sequencing and Jeffrey Miller for discussion and bacterial strains. The manuscript was improved by comments from Olke Uhlenbeck, Paul Schimmel, LaDonne Schulman and Bill McClain. J.N. thanks Marge Strobel and Eric Phizicky for technical advice, and Rich Schwartz and Peter Johnson for advice and interest. This work was supported in part by an NIH grant (GM32637) to J. A., an NSF grant (DMBA3-18244) to R. C. O., an NIH grant (GM06965) to S. J. H. and by a National Research Service Award (T32GM07616) from the National Institute of General Medical Sciences to J. N.

REFERENCES

1. Chapeville, F. *et al.* 1962. *Proc. Natl. Acad. Sci. U.S.A* 48:1086-92
2. Crick, F. H. C. 1966. *J. Mol. Biol.* 19:548-55
3. Sprinzl, M., Moll, J., Meissner, F., Hartmann, T. 1985. *Nucleic Acids Res.* 13:r1-49
4. Holley, R. W. *et al.* 1965. *Science* 147:1462-65
5. Robertus, J. D. *et al.* 1974. *Nature* 250:546-51
6. Kim, S. H. *et al.* 1974. *Science* 185:435-40
7. Schevitz, R. W. *et al.* 1974. *Nature* 278:188-90
8. Westhof, E., Dumas, P., Moras, D. 1985. *J. Mol. Biol.* 184:119-45
9. Schimmel, P. R., Söll, D. 1979. *Ann. Rev. Biochem.* 48:601-48
10. Rich, A., Schimmel, P. R. 1977. *Nucleic Acids Res.* 4:1649-65
11. Schimmel, P. R. 1977. *Acts.Chem.Res.* 10:411-18
12. Schimmel, P. R. 1979. in *Transfer RNA: Structure, Properties and Recognition* (eds. Schimmel, P. R., Söll, D., Abelson, J.) Cold Spring Harbor Laboratory, New York pp. 297-310
13. Schulman, L. H., Pelka, H., Susani, M. 1983. *Nucleic Acids Res.* 11:1439-55
14. Schulman, L. H., Pelka, H. 1983. *Proc. Natl. Acad. Sci. U.S.A.* 80:6755-9
15. Schulman, L. H., Pelka, H. 1985. *Biochemistry* 24:7309-14

16. Hooper, M. L., Russell, R. L., Smith, J. D. 1972. *Febs. Lett.* 22:149-55
17. Shimura, Y. *et al.* 1972. *Febs. Lett.* 22:144-48
18. Ghysen, A., Reyes, O., Allende, C. C., Allende, J. E. 1973. in *Gene Expression and Its Regulation* (eds Kenney *et al.*) Plenum Press, New York pp. 209-16
19. Ozeki, H. *et al.* 1980. in *Transfer RNA: Biological Aspects* (eds. Schimmel, P. R., Söll, D., Abelson, J.) Cold Spring Harbor Laboratory, New York pp. 341-62
20. Goodman, H. M., Abelson, J., Landy, A., Brenner, S., Smith, J. D. 1968. *Nature* 217:1019-24
21. Steege, D. A., Söll, D. 1979. in *Biological Regulation and Development* Vol. I (ed. Goldberger, R. F.) Plenum Press, New York pp. 433-85
22. Smith, J. D., Celis, J. E. 1973. *Nature new Biol.* 243:66-71
23. Celis, J. E., Hooper, M. L., Smith, J. D. 1973. *Nature new Biol.* 244:261-64
24. Inokuchi, H., Celis, J. E., Smith, J. D. 1974. *J. Mol. Biol.* 85:187-92
25. Ghysen, A., Celis, J. E. 1974. *J. Mol. Biol.* 83:333-51
26. Yaniv, M., Folk, W. R., Berg, P., Soll, L. 1974. *J. Mol. Biol.* 86:245-60
27. Khorana, H. G. 1979. *Science* 203:614-25
28. Brennan, T., Sundaralingam, M. 1976. *Nucleic Acids Res.* 3:3235-3251
29. Steege, D. A. 1983. *Nucleic Acids Res.* 11:3823-32
30. Yamada, Y., Ishikura, H. 1973. *Febs. Lett.* 29:231-34
31. Ish-Horowicz, D., Clark, B. F. C. 1973. *J. Biol. Chem.* 248:6663-73
32. Ishikura, H., Yamada, Y., Nishimura, S. 1971. *Febs. Lett.* 16:68-70
33. McClain, W. H., Barrell, B. G., Seidman, J. G. 1975. *J. Mol. Biol.* 99:717-32
34. Grosjean, H., Nicoghosian, K., Haumont, E., Söll, D., Cedergren, R. 1985 *Nucleic Acids Res.* 13:5697-706
35. Roe, B., Sirover, M., Dudock, B. 1973. *Biochemistry* 12:4146-54
36. Bruce, G. A., Uhlenbeck, O. C. 1982. *Biochemistry* 21:3921-26
37. Kisselev, L. L. 1985. *Prog. Nucleic Acids Res. Mol. Biol.* 32:237-66
38. Yoshimura, M., Inokuchi, H., Ozeki, H. 1984. *J. Mol. Biol.* 177:627-44
39. Thorbjarnardottir, S. *et al.* 1985. *J. Bact.* 161:219-22
40. Young, R. A. 1979. *J. Biol.Chem.* 254:12725-31
41. Abelson, J. 1979. *Ann. Rev. Biochem.* 48:1035-69
42. Miller, J. H., Albertini, A. M. 1983. *J. Mol. Biol.* 164:59-71
43. Dente, L., Cesareni, G., Cortese, R. 1983. *Nucleic Acids Res.* 11:1645-55
44. Miller, J. H. 1972. in *Experiments in Molecular Genetics* Cold Spring Harbor

Laboratory, New York pp. 352-55

45. Knott-Hunziker, V., Waley, S. G., Orlek, B. S., Sammes, P. G. 1979. *Febs. Lett.* 99:59-61
46. Dalbadie-McFarland, G. *et al.* 1982. *Proc. Natl. Acad. Sci. U.S.A.* 79:6409-13
47. Ambler, R. P. 1980. *Phil. Trans. R. Soc.* B289:321-31
48. Fisher, J., Belasco, J. G., Khosla, S., Knowles, J. R. 1980. *Biochemistry* 19:2895-901
49. Sigal, I. S., Harwood, B. G., Arentzen, R. 1982. *Proc. Natl. Acad. Sci. U.S.A.* 79:7157-60
50. Sigal, I. S., DeGrado, W. F., Thomas, B. J., Petteway, S. R. 1984. *J. Biol. Chem.* 259:5327-32
51. Dalbadie-McFarland, G., Neitzel, J. J., Richards, J. H. 1986. *Biochemistry* 25:332-8
52. Smith, D. R., Calvo, J. M. 1980. *Nucleic Acids Res.* 8:2255-74
53. Amann, J. E., Brosius, J., Ptashne, M. 1983. *Gene* 25:167-178
54. Edman, P., Begg, G. 1967. *Eur. J. Biochem.* 1:80-91
55. Bolin, J. T., Filman, D. J., Matthews, D. A., Hamlin, R. C., Kraut, J. 1982. *J. Biol. Chem.* 257:13650-62
56. Matthews, D. A. *et al.* 1977. *Science* 197:452-55
57. Filmann, D. J., Bolin, J. T., Matthews, D. A., Kraut, J. 1982. *J. Biol. Chem.* 257:13663-72
58. Hewick, R. M., Hunkapiller, M. W., Hood, L. E., Dreyer, W. J. 1981. *J. Biol. Chem.* 256:7990-7
59. Yamaizumi, Z., Kuchino, Y., Harada, F., Nishimura, S., McCloskey, J. A. 1980. *J. Biol. Chem.* 255:2220-5
60. Yarus, M., McMillan, C., Cline, S., Bradley, D., Snyder, M. 1980. *Proc. Natl. Acad. Sci. U.S.A.* 77:5082-96.
61. Maniatis, T., Fritsch, E. F., Sambrook, J. 1982. in *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory, New York pp. 320-21
62. Denhardt, D. T. 1966. *Biochem. Biophys Res. Com.* 23:641-6
63. Maniatis, T., Fritsch, E. F., Sambrook, J. 1982. in *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, New York p. 447
64. Sanger, F., Nicklen, S., Coulson, A. R. 1977. *Proc. Natl. Acad. Sci. U.S.A.* 74:5463-7

65. Newman, A. J., Ogden, R. C., Abelson, J. 1983. *Cell* 35:117-25
66. Villafranca, J. E. *et al.*, 1983. *Science* 222:782-8
67. Baccanari, D., Phillips, A., Smith, S., Sinski, D., Burchall, J. 1975. *Biochemistry* 14:5267-73
68. Baccanari, D. P., Averett, D., Briggs, C., Burchall, J. 1977. *Biochemistry* 16:3566-72
69. Baccanari, D. P., Stone, D., Kuyper, L., 1981. *J. Biol. Chem.* 256:1738-47

FIGURE LEGENDS

Figure 1. Serine tRNAs recognized by *E. coli* serine AAS (29-34) compared with *E. coli* $tRNA^{Leu5}$ (ref. 59 and K. Jones B. Reid, personal communication). Boxed nucleotides indicated absolute conservation among the five $tRNA^{Ser}$ sequences. The conserved serine nucleotides that differ from $tRNA^{Leu5}$ have been underlined in the $tRNA^{Leu5}$ sequence. A fifth *E. coli* serine tRNA has been sequenced (not shown) and is identical to that of *E. coli* 4, except for residue 20, which is a C (39). Acc. stem, acceptor stem. A. C., anticodon.

Figure 2. a)Cloverleaf representation of conserved *E. coli* $tRNA^{Ser}$ residues. The indicated nucleotides are those that are absolutely conserved in all six serine tRNAs and not found in $tRNA^{Leu5}$. b) $tRNA^{Leu5}$ with the changes proposed to confer $tRNA^{Ser}$ specificity. The modified nucleotides are those present in $tRNA^{Leu5}$. Note that the changes in the D loop at positions 20-2 and 21 are equally well represented by insertion of an A between residues 21 and 22. Thus there are actually 11 independent changes.

Figure 3. The $tRNA^{Leu}_{UAG}$ gene was constructed from six oligonucleotides containing the $tRNA^{Leu}_{UAG}$ sequence flanked by Eco RI and Pst I restriction endonuclease cohesive ends at the immediate 5' and 3' ends of the gene respectively. Junctions between oligonucleotides are indicated. The four oligonucleotides containing the appropriate sequence alterations for construction of $tRNA^{Leu\rightarrow Ser}_{UAG}$ exactly replace four of the $tRNA^{Leu}_{UAG}$ oligonucleotides, and are shown above and below the $tRNA^{Leu}_{UAG}$ sequence. The sequence of the terminator, synthesized as two chains, is indicated.

Figure 4a. Sequence of the serine-requiring amber mutation in the *amp* gene of pBR322, generated by oligonucleotide-directed mutagenesis. Pvu II fragments containing the synthetically constructed $tRNA^{Leu}_{UAG}$ or $tRNA^{Leu\rightarrow Ser}_{UAG}$ gene together with the M13mp8-derived *lac* promoter and rrnC terminator were subcloned into the Pvu II site of pBR322-βam to yield pBR322-βam:$tRNA^{Leu}_{UAG}$ and pBR322-βam:$tRNA^{Leu\rightarrow Ser}_{UAG}$, respectively.

Figure 4b. Suppression of the serine-requiring amber mutation. Isogenic *E. coli* strains carrying the *supD, supE, supF* and *supP* alleles were transformed with pBR322-βam and

plated on L plates containing either 20ug/ml tetracycline or 80ug/ml ampicillin.

Figure 4c. Suppression of serine-requiring amber mutation by $tRNA^{Leu \rightarrow Ser}_{UAG}$. XAC-A16 was transformed with either pBR322-βam:$tRNA^{Leu}_{UAG}$ or pBR322-βam:$tRNA^{Leu \rightarrow Ser}_{UAG}$ and plated as in (b), except that IPTG was added to 1mM. β-lactamase activity in the latter strain was 2% the level seen in a strain harbouring wild-type pBR322 and 5% the level seen when this mutation was suppressed by *supD*.

Figure 5 a) Sequence of the *fol* amber. Asp 11(GAT) was changed to Asn (AAT) to improve the context of the amber mutation. Pvu II fragments encoding either $tRNA^{Leu}_{UAG}$ or $tRNA^{Leu \rightarrow Ser}_{UAG}$ along with the *lac* promoter and *rrnC* terminator were subcloned into the Pvu II site of plasmid pDA3-12 to yield plasmids pDA3-12:Leu and pDA3-12:Leu→Ser, respectively. b) DHFR purification visualized by Coomassie blue staining of an SDS-polyacrylamide gel. Lane 1, crude extract; Lane 2, 90% ammonium sulphate-saturated insoluble fraction; Lane 3, flow through from methotrexate sepharose column; Lane 4, folic-acid eluate from methotrexate sepharose; Lane 5, peak protein fraction from DEAE sephacel column.

Figure 6. Protein sequence analysis of wild-type DHFR (first column of graphs), DHFRam/Leu suppressor (second column), DHFRam/Ser suppressor (third column), and DHFRam/Leu→Ser suppressor (fouth column). The picomolar yields of phenylthiohydantoin (PTH) derivatives of valine, aspartic acid, asparagine, leucine and serine are plotted against the residue number for wild-type DHFR, isolated from *E. coli* carrying pTAC101, and mutant DHFR isolated from strains carrying pDA3-12:Leu, or pDA3-12:Leu→Ser, as well as from a *supD* strain carrying pDA3-12.

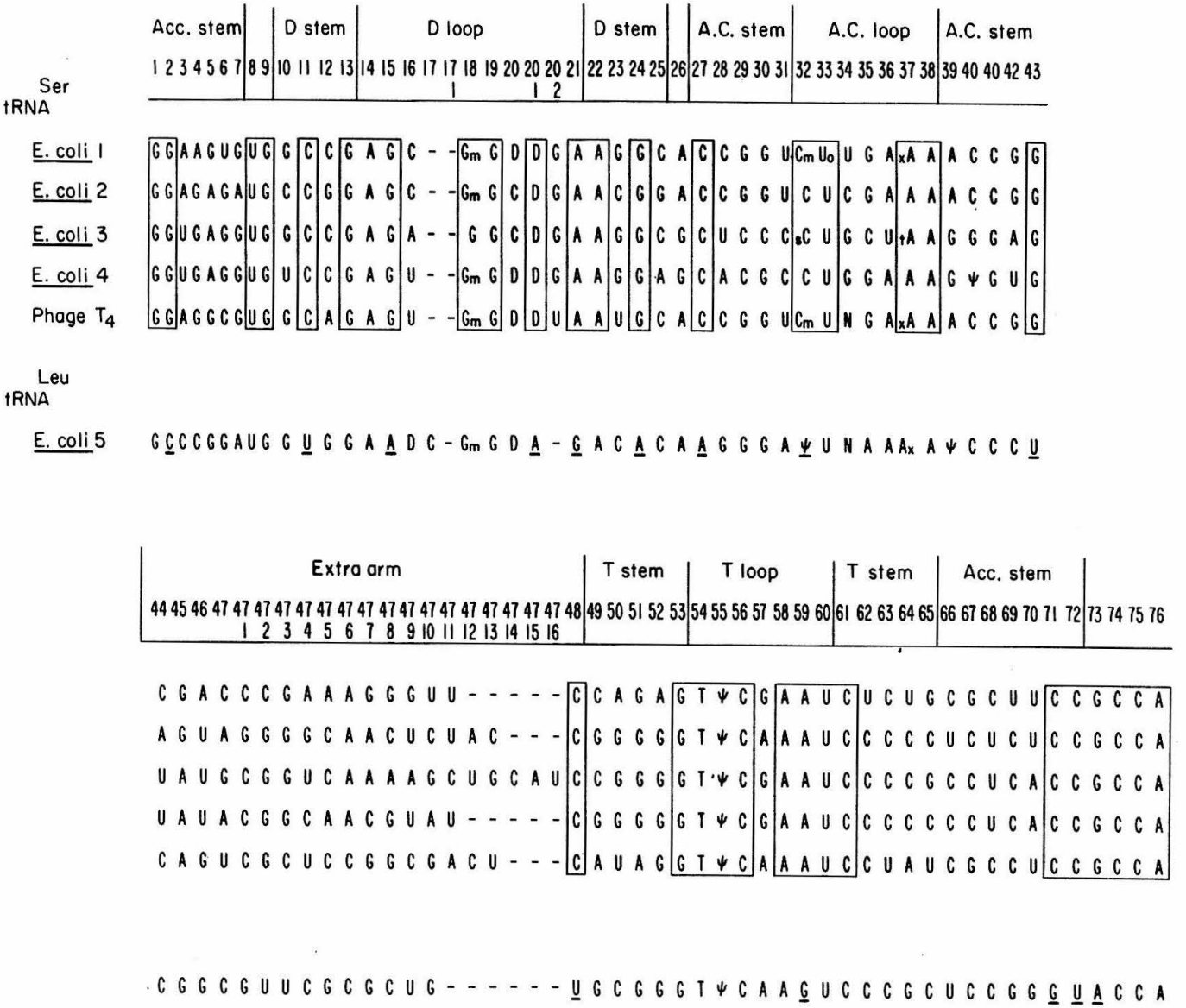

Acc. stem
D stem
D loop
D stem
A.C. stem
A.C. loop
A.C. stem
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 17 18 19 20 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 40 42 43
Ser tRNA
E. coli 1
E. coli 2
E. coli 3
E. coli 4
Phage T4
Leu tRNA
E. coli 5
Extra arm
T stem
T loop
T stem
Acc. stem
44 45 46 47 47 47 47 47 47 47 47 47 47 47 47 47 47 47 47 47 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76

Figure 1

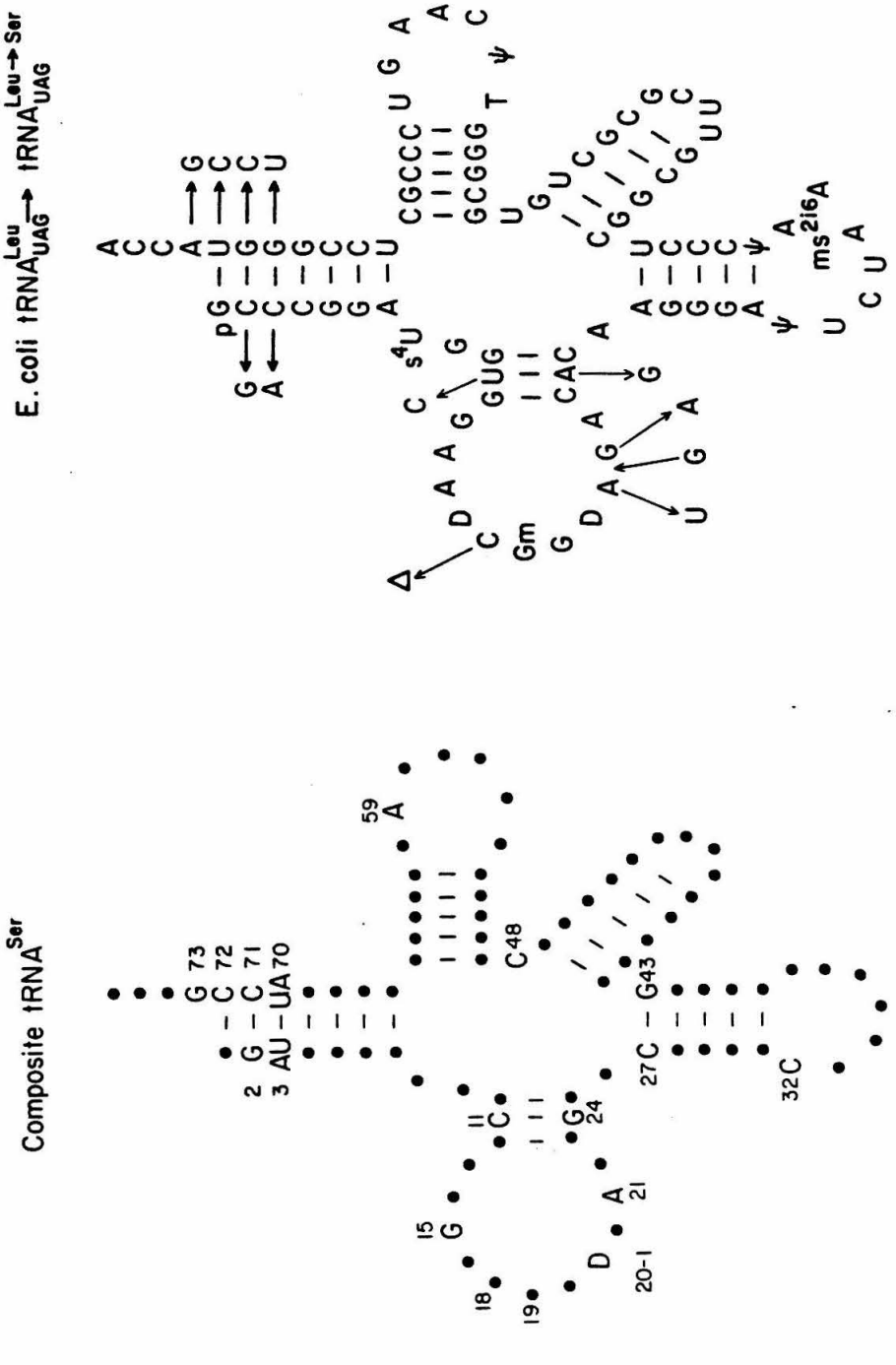


Figure 2

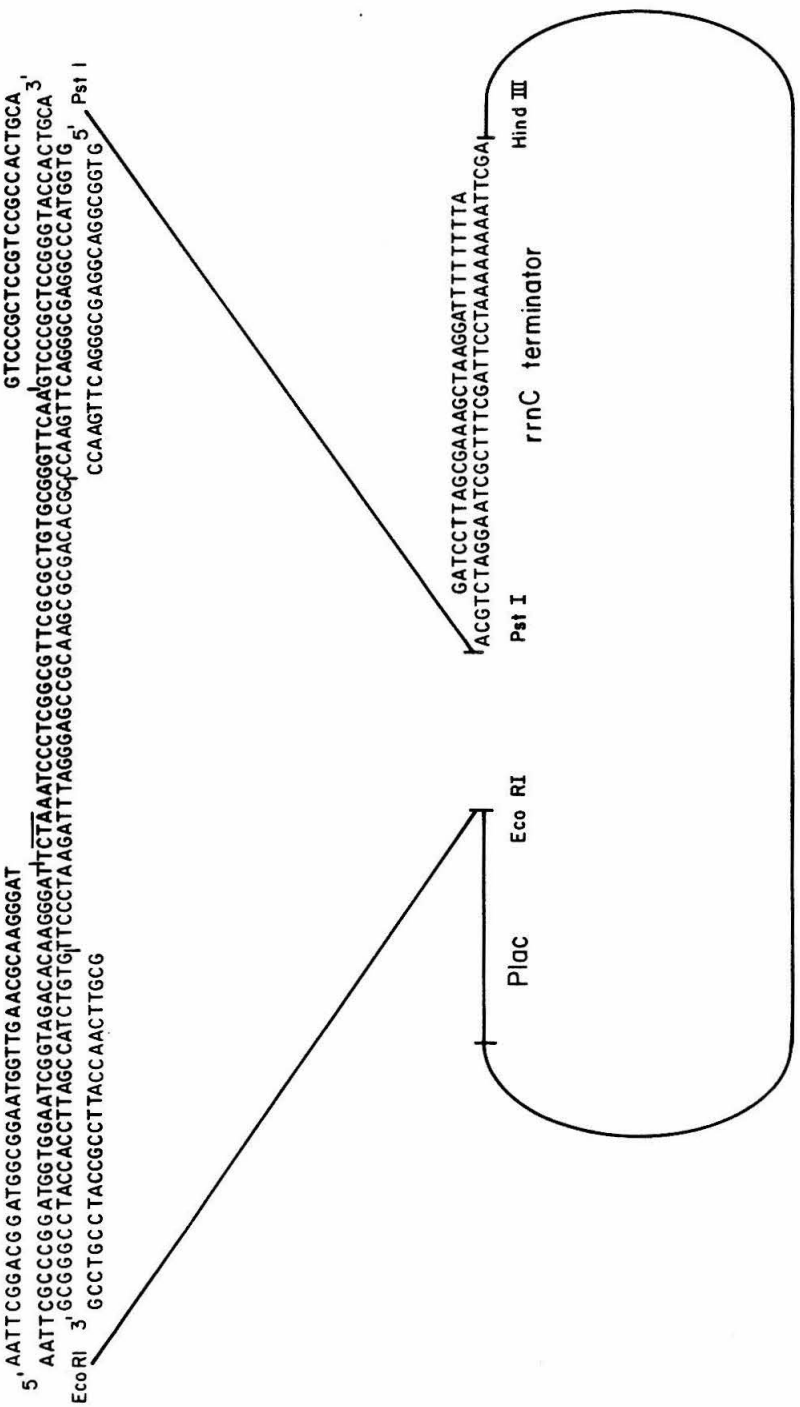

5' AATTCGGACGGATGGCGGAATGGTTGAACGCAAGGGAT
AATTCGCCCGGATGGTGGAATCGGTAGACACAAGGGATTCTAAATCCCTCGGCGTTCGCGCTGTGCGGGTTCAATGTCCCGCTCCGGTACCACTGCA 3'
GCGGGCCTACCACCATCTGTGTTCCCTAAGATTTAGGGAGCCGCAAGCGCGACACGCCCAAGTTCAGGGCGAGGCCCATGGTG 5' Pst I
EcoRI 3' GCCTGCCTACCGCCTTACCAACTTGCG
CCAAGTTCAGGGCGAGGCAGGCGGTG
GTCCCGCTCCGTCCGCCACTGCA
GATCCTTAGCGAAAGCTAAGGATTTTTTTTA
ACGTCTAGGAATCGCTTTCGATTCCTAAAAAAAATTCGA
Pst I
Hind III
rrnC terminator
Plac
Eco RI

Figure 3

pBR322 β-lactamase amber

	60								Active site		70
β wt:	Pro	Glu	Glu	Arg	Phe	Pro	Met	Met	Ser	Thr	Phe
	CCC	GAA	GAA	CGT	TTT	CCA	ATG	ATG	AGC	ACT	TTT
β am:	...	...	...	...	...	...	...	...	TAG	...	...

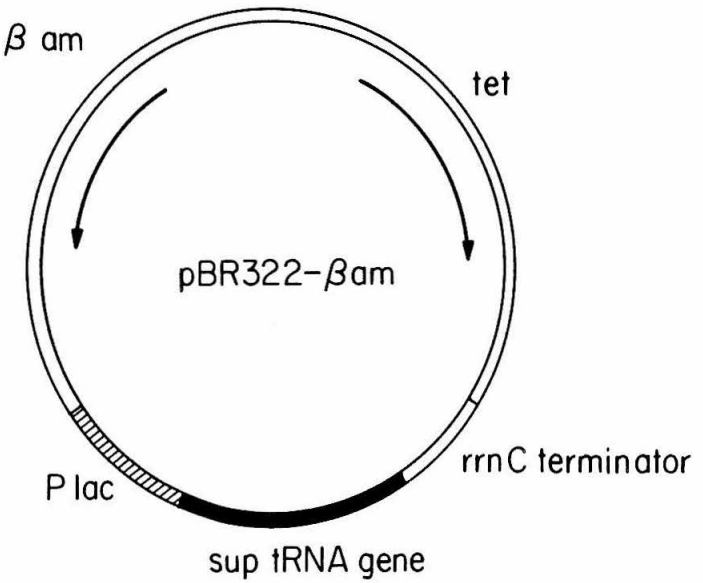


Figure 4a

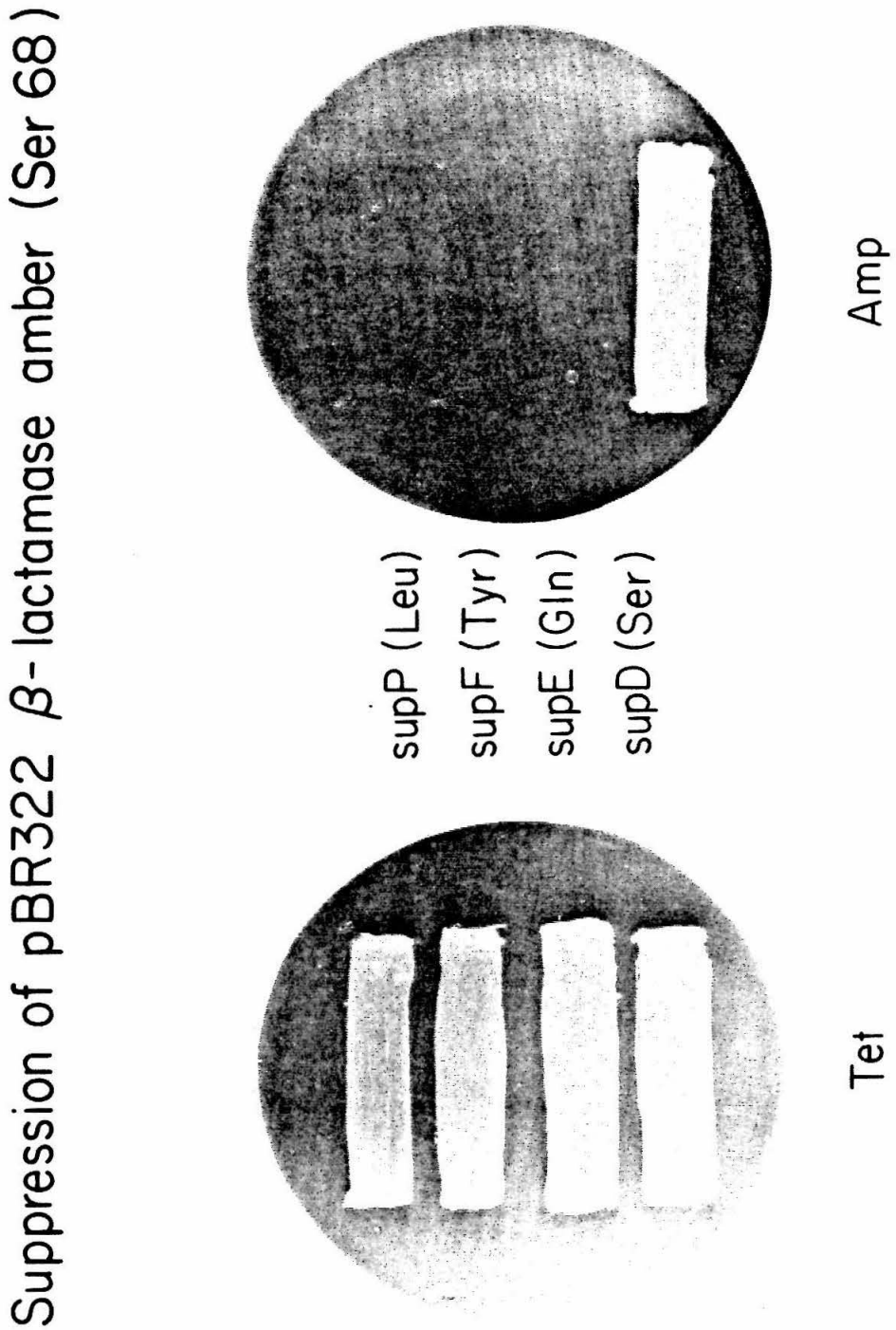

Suppression of pBR322 β- lactamase amber (Ser 68)
supP (Leu)
supF (Tyr)
supE (Gln)
supD (Ser)
Amp
Tet

Figure 4b

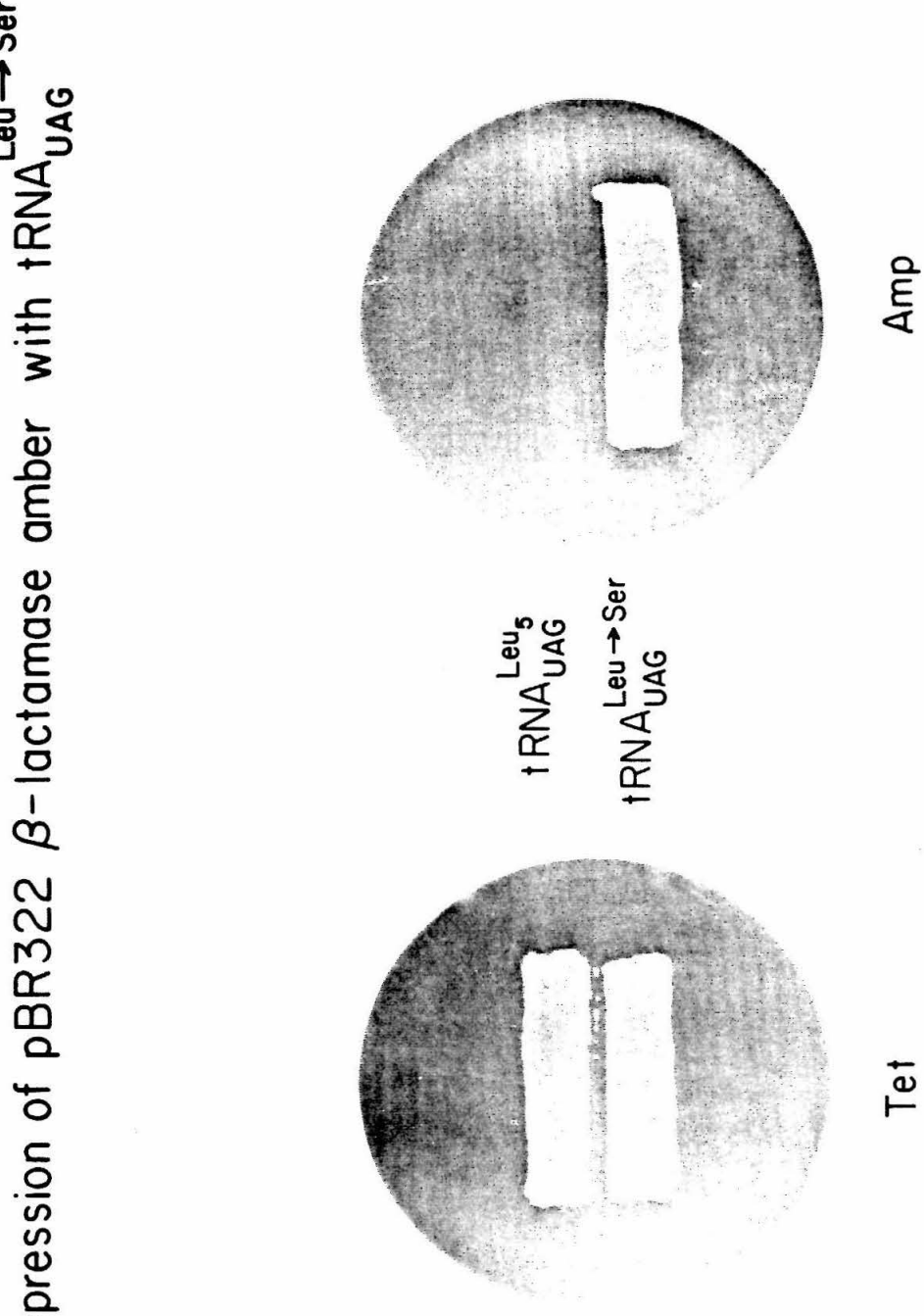
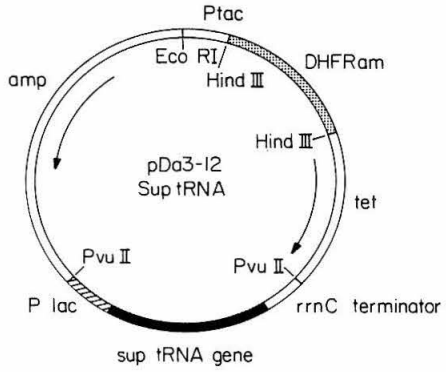

Suppression of pBR322 β-lactamase amber with tRNA$^{Leu \rightarrow Ser}_{UAG}$
tRNA$^{Leu_5}_{UAG}$
tRNA$^{Leu \rightarrow Ser}_{UAG}$
Amp
Tet

Figure 4c

DHFR amber

	2									12
DHFR wt:	Ile	Ser	Leu	Ile	Ala	Ala	Leu	Ala	Val	Asp Arg
	ATC	AGT	CTG	ATT	GCG	GCG	TTA	GCG	GTA	GAT CGC
DHFR am:	...	...	...	...	...	...	...	...	am	Asn ...
	...	...	...	...	...	...	...	...	TAG	AAT ...



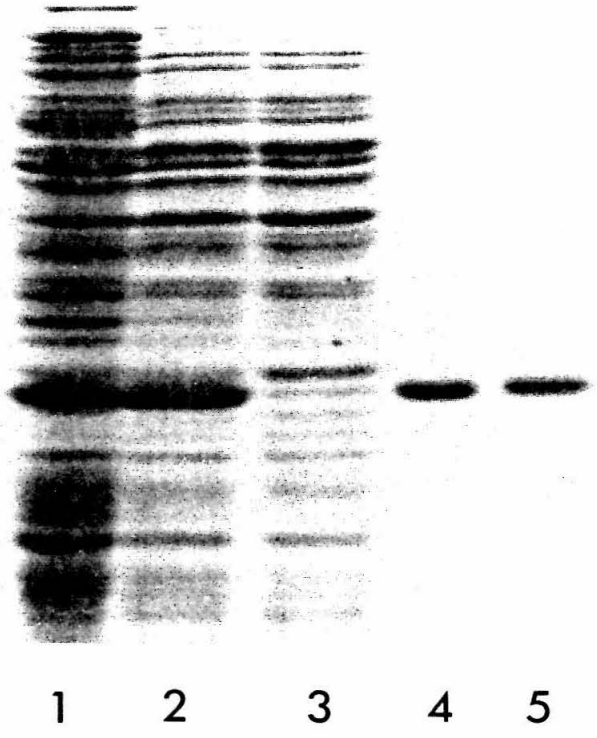


Figure 5

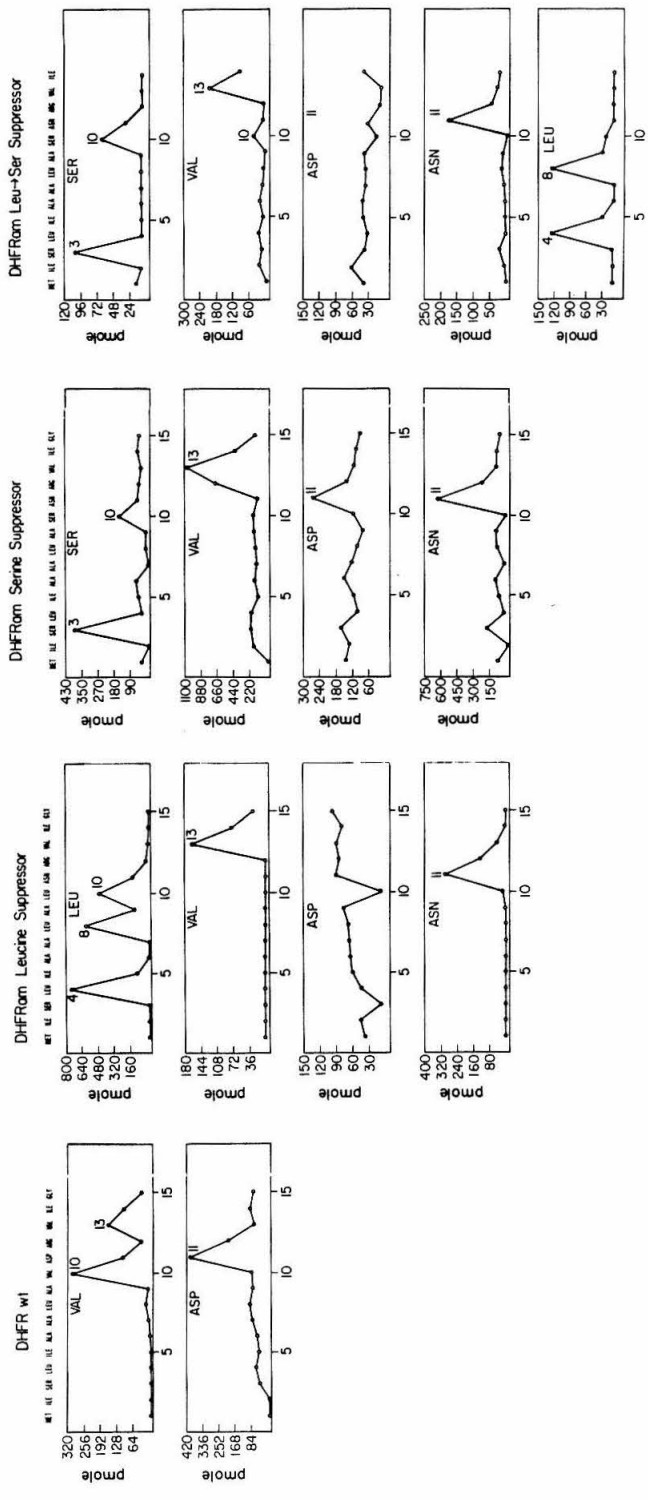

DHFR wt
DHFRam Leucine Suppressor
DHFRam Serine Suppressor
DHFRam Leu→Ser Suppressor
pmole
VAL
ASP
LEU
SER
ASN

Figure 6

CHAPTER II

Eight base changes are sufficient to convert a leucine-inserting tRNA into a serine-inserting tRNA

INTRODUCTION

The identity of a tRNA is defined by the elements within the tRNA that are responsible for the correct aminoacylation of the tRNA by its cognate aminoacyl tRNA synthetase (AAS). The correct aminoacylation of a tRNA is the outcome of two types of interactions, 1) the positive interaction between a tRNA and its cognate AAS, and 2) the negative interaction between a tRNA and all 19 noncognate AASs. The determinants of these positive and negative interactions are in the nucleotides of any given tRNA. One method that has been used to determine which nucleotides comprise a tRNA's identity is referred to as an identity swap. In this approach, the identity of a tRNA is altered with the fewest changes, while retaining its complete biological function. Those changes responsible for redirection of AAS recognition constitute identity elements.

In an effort to define the identity elements of *E. coli* serine tRNA, we have developed a system to facilitate the identity swap approach. Amber suppressor tRNA genes containing the alterations predicted to redirect AAS recognition are assembled from oligonucleotides and inserted in cassette form into a plasmid, which supplies the inducible *lac* promoter and a transcription terminator (1). First, the biological function of the mutant tRNA is established by assaying the efficiency at which the tRNA can suppress a *LacZ* amber mutation. This mutation, in a non-essential portion of *LacZ*, can tolerate any amino acid. Second, we screen specifically for the insertion of serine using a mutant of the *amp* gene, which contains an amber mutation in the essential, active-site serine residue (2). Finally, the amino acid specificity of the supperssor is determined by sequencing the protein product of a suppressed amber mutant gene. For this purpose the *E. coli* dihydrofolate reductase (DHFR) gene, *fol*, with an amber mutation at codon ten is used. DHFR is easily purified by affinity chromatography on a methotrexate resin. X-ray crystallographic data (4, 5) reveals that residue ten of DHFR lies on the surface of the protein, well removed from the active site, and it appears that any amino acid is acceptable at this position. If more than one amino acid is inserted by the suppressor, they can be readily detected by this method. Protein sequence data acquired in this manner reveal the outcome of the competition among all 20 AAS in the cell for the suppressor.

Previously, we reported the conversion of a leucine-inserting amber suppressor tRNA into a serine-inserting amber suppressor tRNA (1). Comparison of six *E. coli* serine-accepting tRNAs revealed 16 nucleotides, which were highly conserved among, or highly correlated with, serine tRNAs, yet not found in the leucine tRNA that we used as a

starting tRNA. Twelve of the 16 nucleotides were superimposed upon the amber suppressor allele of $tRNA^{Leu5}$, by constructing a tRNA gene from synthetic oligonucleotides that encoded the leucine tRNA with the twelve base changes that we wished to incorporate. The resultant tRNA, $tRNA^{Leu \rightarrow Ser}_{UAG}$ was a weak suppressor (1% efficient as determined by β-galactosidase assays) yet it inserted predominantly serine (Table 1).

Here, we describe our efforts to determine the minimum number of base changes necessary to effect the leucine to serine conversion, and to increase the suppression efficiency of the tRNA. We speculated that not all of the 12 nucleotide changes were necessary for the conversion and may have been responsible for the low suppression efficiency of $tRNA^{Leu \rightarrow Ser}_{UAG}$. By reverting the mutations singly or in groups, we determined that a minimum of eight base changes were required to effect the leucine to serine conversion. They were basepair C_{11}-G_{24} in the dihydrouridine (D) stem, base pairs G_2-C_{71}, A_3-U_{70}, and nucleotides C_{72} and G_{73}, all in the acceptor stem. These eight nucleotides, when incorporated into a leucine tRNA, yield a tRNA that inserts exclusively serine with 40% suppression efficiency.

METHODS

Gene Construction. The tRNA genes were constructed from synthetic oligonucleotides (made by the Caltech Microchemical Facility), which were assembled as described previously (1), and cloned into the vector pJN10. pJN10 is derived from plasmids pBR322 (6) and $pEMBL8^+$ (7). The PvuI-PvuII fragment from $pEMBL8^+$ containing a portion of the *amp* gene, the replication origin, and the polylinker, was ligated to the PvuI-PvuII fragment of pBR322 that carries the *tet* gene. In pJN10, the *amp* gene lacks a Pst I site, and, additionally, contains an amber mutation at its active site serine. Prior to constructing pJN10, the Eco RI site in pBR322 was removed by digesting pBR322 with Eco RI, filling in the "sticky ends" with Klenow and ligating. tRNA genes are ligated into the Eco RI-Pst I sites located in the polylinker region of pJN10, which are flanked by the *lac* promoter and *rrnC* terminator, 5' and 3', respectively. Plasmid pJN10, into which a suppressor tRNA gene has been cloned, will be referred to as pJN10:sup.

Determination of suppression efficiency: Suppression efficiencies were determined by assaying β-galactosidase activity (7) in strain XAC-1 (F'$lacI_{373}$ $lacZ_{u118am}$ $proB^{+}/\Delta(lacproB)_{x111}$ *nalA, rif, ara,* $argE_{am}$), which had been transformed with pJN10:sup.

Specificity assay: purification of DHFR: DHFR was isolated from XAC-1 carrying pJN10:sup and plasmid pDAYQ. pDAYQ encodes the gene conferring resistance to chloramphenicol as well as the gene encoding DHFR, *fol*, with an amber mutation at residue ten (Normanly and Abelson, unpublished). The *fol* gene is under the control of the *tac* promoter (8). Cells were cultured in L broth (10g/l bactotryptone, 10g/l NaCl, 5g/l yeast extract) supplemented with 5ug/ml tetracycline and 30ug/ml chloramphenicol to an A_{600} of 0.4-0.8, at which time they were diluted fifty-fold into minimal M9 Glycerol medium (7), which had been supplemented with 5ug/ml tetracycline, 30ug/ml chloramphenicol, and lmM IPTG. The culture was incubated, shaking, at 37°C for 12 to 18 hr., and the cells harvested. DHFR was purified from the cells as described in (1). N-terminal sequence analysis of DHFR was carried out at either the USC Microchemical Core Laboratory or the Caltech Microchemical Facility.

RESULTS

Figure 1a shows the nucleotides that are highly conserved in serine-accepting tRNAs (or as in the case of basepair 3-70, highly correlated) and which are not found in $tRNA^{Leu5}_{UAG}$, our starting tRNA. Figure 1b shows the original leucine to serine conversion, $tRNA^{Leu\rightarrow Ser}_{UAG}$ (1), with the 12 changes indicated. We first examined the contribution of the mutations in the D loop. In $tRNA^{Leu\rightarrow Ser}_{UAG}$, we had shifted the position of the conserved nucleotides G_{18} and G_{19}, based upon the observation that the position of these nucleotides (relative to the D stem),which is highly conserved in serine tRNAs, differs from the position of G_{18} and G_{19} in leucine tRNAs. We predicted that this structural feature may play a role in serine AAS recognition, and therefore, altered the leucine D loop to introduce a serine-like configuration (Figure 1b). In mutant LSM6 (Figure 2) the D loop has the original $tRNA^{Leu5}$ sequence and configuration, while the mutations in

the D stem and acceptor stem are retained. This suppressor inserted exclusively serine at high efficiency (Table 1).

We next examined the contribution of the mutations in the D stem (basepair 11-24) by generating a mutant that contained mutations in the acceptor stem only, and none in the D loop or stem. This variant, LSM4 (Figure 3), while an efficient suppressor, inserted a mixture of amino acids: serine (16%), glutamine (39%) and leucine (38%). Clearly, basepair C_{11}-G_{24} plays a role in serine acceptance. In all serine tRNAs base pair 11-24 is a C-G and in $tRNA^{Leu5}_{UAG}$ is a U-A. We wondered if a U-G basepair would be acceptable at this position, thereby pinpointing G_{24} as the identity element. Mutant LSM11 (Figure 4) contained a U-G at position 11-24, along with the original mutations in the acceptor stem. This variant suppressed the β-lactamase amber sufficiently to enable growth in the presence of ampicillin; however, the specificity assay revealed that LSM11 inserts predominantly leucine (62%), glutamine (33%) and only 1% serine. From these data we established that alterations to the D loop were unnecessary, and from the standpoint of suppression efficiency, most likely deleterious. Base pair C_{11}-G_{24}, however, plays a key role in serine AAS acceptance.

We next set out to determine which of the original changes in the acceptor stem were actually necessary by reverting one base pair or nucleotide at a time, retaining the remainder of the acceptor stem changes as well as the essential C-G pair at position 11-24.

In mutant LSM10, position 73 has been changed back to an A, while the other acceptor stem mutations, and the mutations at basepair 11-24 have been retained (Figure 5). This tRNA variant inserts exclusively leucine. In mutant LSM12 (Figure 6) basepair 2-71 has the original leucine sequence, C-G, and in mutant LSM13 (Figure 7) position 72 has the leucine sequence U. Both of these mutants insert predominantly leucine and glutamine (Table 1). In mutant LSM9 (Figure 8) the A-U at 3-70 has been changed to the C-G found in $tRNA^{Leu5}_{UAG}$. This tRNA inserts predominantly leucine (72%), some serine (20%) and glutamine (6%), indicating that A_3-U_{70} is necessary for complete serine identity. The A-U pair at position 3-70 is not highly conserved in serine tRNAs, some have a U-A at this position. A mutant, LSM14, with a U-A at position 3-70, along with the other acceptor and D stem mutations (Figure 9) inserts predominantly leucine (65%) along with glutamine (19%) and serine (11%). It would appear that a U-A pair at position 3-70 results in a loss of serine identity. This is odd considering that some serine tRNAs have a U_3-A_{70} pair. It

is possible that an A_3-U_{70} pair plays less a role in serine identity but rather acts as a deterrent to the leucine or glutamine AASs.

From these reversion studies it is apparent that all of the original changes made in the acceptor stem were necessary to convert a leucine-accepting tRNA into a serine-accepting tRNA. Including basepair 11-24, the minimum number of changes to effect the leucine to serine conversion is eight, illustrated in Figure 2. This tRNA (LSM6) is also a very efficient suppressor, indicating biological function.

It is possible that in this initial experiment, some of the serine identity elements would not have been discovered if, for example, they happened to be present in the $tRNA^{Leu5}_{UAG}$ that we used as a starting tRNA. To further define the elements of serine identity, we wished to convert another tRNA to serine acceptance. Therefore, we attempted to convert a tyrosine tRNA to a serine tRNA. We chose a tyrosine tRNA because in *E. coli* this species of tRNA, like the leucine and serine tRNAs, belongs to the type II class. Type II tRNAs are distinguished from type I tRNAs (in *E. coli* type I tRNAs constitute the remainder of tRNA species) primarily in the size of their extra loop, which is larger, 9 to 21 nucleotides, as opposed to 4 to 5 nucleotides for type I tRNAs (9). As the function of the extra arm is not understood, we elected to remain within the type II class for our identity swap experiments.

Figure 10 shows the nucleotides highly conserved in serine tRNAs but not found in the su^+3 allele of $tRNA^{Tyr}$ (10). The leucine to serine identity swap experiments revealed that alterations in the D loop were unnecessary for generating a serine accepting tRNA; furthermore, it was not necessary to make alterations in the anticodon stem of $tRNA^{Leu5}_{UAG}$ to obtain serine identity. We therefore chose to make only two changes to the tyrosine tRNA, those at postion 9 and 73. This mutant, Tyr→Ser1 (Figure 11), enabled growth in the presence of ampicillin, yet inserted predominantly glutamine (71%), tyrosine (26%) and only 1% serine, as determined by the specificity assay (Table 1). While we may have introduced a serine determinant, we also either introduced a glutamine identity element or removed a deterrent. Also, Tyr→Ser1 retained tyrosine identity elements. This result did not come as a complete suprise to us, as a tyrosine amber suppressor (su^+3) with the same alteration at position 73 as in the Tyr→Ser1 mutant, inserts glutamine (11-13). We made two other attempts at a tyrosine to serine identity swap; Tyr→Ser2 (Figure 12) has alterations at all seven of the positions highly conserved in serine that are not found in tyrosine. Tyr→Ser3 (Figure 13) has changes at positions 9, 73, and basepair 3-70. Neither

mutant inserts more than 1% serine, both insert predominantly glutamine and some tyrosine (Table 1).

DISCUSSION

Previously, we reported the conversion of a leucine-inserting tRNA into a serine-inserting tRNA, brought about by making 12 base changes in the leucine tRNA (1). We have now determined which of those 12 changes were necessary, i. e., the minium number of changes required to effect the leucine to serine identity conversion. Only eight base changes are necessary, those at positions 11-24 in the D stem, 2-71, 3-70, 72, and 73 in the acceptor stem.

In making the original $tRNA^{Leu \rightarrow Ser}_{UAG}$, we shifted the position of G_{18} and G_{19} in the leucine tRNA to conform to the configuration of a serine D loop. We found that not only was this alteration unnecessary for serine AAS recognition, it had a deleterious effect upon the suppression efficiency of the tRNA. We have also found that, while serine tRNAs may have either an A-U or U-A at position 3-70, in the leucine to serine identity conversion, an A_3-U_{70} pair is essential, while a U_3-A_{70} pair results in charging by the glutamine and leucine AASs. Possibly, the A_3-U_{70} acts less as an identity element for serine AAS, but in the context of the $tRNA^{Leu \rightarrow Ser}_{UAG}$ sequence, acts as a deterrent to the leucine and glutamine AASs. It would be interesting to see what effect an A_3-U_{70} pair would have upon the identity of a serine tRNA which normally has a U-A at position 3-70.

Identity swap experiments have been used to define the identity elements for *E. coli* $tRNA^{Ala}$ (14-16). A G-U pair at position 3-70 is unique to alanine tRNAs, and all alanine tRNAs that have been sequenced to date have this G_3-U_{70} pair. Superimposing this G_3-U_{70} pair onto an amber alleles of *E. coli* $tRNA^{Cys}$ or $tRNA^{Lys}$, results in complete alanine identity (14, 16). When the G_3-U_{70} is superimposed upon a phenylalanine amber suppressor tRNA, the tRNA inserts predominantly alanine into amber codons, however it inserts some phenylalanine as well (14, 15). Yet when the G_3-U_{70} pair is superimposed upon a glycine amber suppressor tRNA, there is no conversion to alanine identity (15). These experiments highlight one of the problems encountered when attempting to define tRNA identity, and help to explain our failure to effect a complete conversion from tyrosine to serine identity. Merely superimposing predicted identity elements onto a tRNA is not

sufficient to effect a complete conversion of identity. One has to contend with the positive or negative recognition elements for all other AAS, which also reside in the tRNA being studied.

While we were able to introduce serine identity elements into the tyrosine tRNA sequence (all the mutants; Tyr→Ser1 Tyr→Ser2 and Tyr→Ser3 were able to adequately suppress the serine requiring amber mutation in the *amp* gene), tyrosine identity elements remain in the tRNA. Additionally, we either introduced glutamine identity elements or removed glutamine AAS deterrents. As stated above, the introduction of a G at position 73 of a tyrosine amber suppressor (su^+3, the same allele used in our experiments) results in glutamine mischarging. Yet, in a leucine amber suppressor, an A to G change at position 73 does not result in glutamine mischarging. In fact, the phenomenon of glutamine mischargine is not new. A number of amber suppressor tRNAs are mischarged by the glutamine AAS. There must be a context effect; that is, there must be other glutamine determinants or a lack of glutamine deterrents present in those tRNAs that are mischarged by the glutamine AAS, while the opposite must be true for tRNAs that are not mischarged. Sequence comparison between tRNAs that are, or are not, mischarged by the glutamine AAS does not reveal any uniquely conserved nucleotides.

In summary, identity swap experiments have proven instructive about the nature of serine tRNA identity, however, we will not be able to fully refine our model until we have a better understanding of the positive and negative identity elements of other tRNAs, glutamine tRNAs in particular.

REFERENCES

1. Normanly, J., Ogden, R. C., Horvath, S. J., Abelson, J. 1986. *Nature* 321:213-19
2. Dalbadie-McFarland, G., *et al.* 1982. *Proc. Natl. Acad. Sci. U.S.A.* 79:6409-13
3. Bolin, J. T., Filman, D. J., Matthews, D. A., Hamlin, R. C, Kraut, J. 1982. *J. Biol. Chem.* 257:13650-62
4. Filman, F. J., Bolin, J. T., Matthews, D. A., Kraut, J. 1982. *J. Biol. Chem.* 257:13663-72
5. Bolivar, F., Rodriguez, R. L., Greene, P. J., Betlach M. C., Heynecker, H. L., Boyer, H. N. 1977. *Gene* 2:95-113
6. Dente, L., Cesareni, G., Cortese, R. 1983. *Nucleic Acids Res.* 11:1645-55

7. Miller, J. H. 1972. in *Experiments in Molecular Genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
8. Amann, E., Brosius, J., Patashne, M. 1983. *Gene* 25:167-78
9. Brennan, T., Sundaralingam, M. 1976. *Nucleic Acids Res.* 3:3235-51
10. Goodman, H. M., Abelson, J., Landy, A., Brenner, S., Smith, J. D. 1968. *Nature* 217:1019-24
11. Shimura, Y., Aono, H., Ozeki, H., Sarabhai, A., Lamfrom, H., Abelson, J. 1972. *Febs. Lett.* 22:144-8
12. Hooper, M. L., Russel, R. L., Smith, J. D. 1972. *Febs. Lett.* 22:149-55
13. Ghysen, A., Celis, J. E. 1974. *J. Mol. Biol.* 83:333-51
14. Hou, Y. -M., Schimmel, P. R. 1988. *Nature* 333:140-5
15. McClain, W. H., Foss, K. 1988. *Science* 240:793-6
16. McClain, W. H., Foss, K. 1988. *Science* in press

Table 1.

$tRNA_{UAG}$	Suppression Efficiency (given in % wild type β-gal)	Growth on amp	Specificity (given in % amino acid inserted at $DHFR_{am10}$)
Leu	52-59	-	Leu (99)
Ser	34-52	+	Ser (92)
Leu→Ser	0.5-1	+	Ser (90),Leu(10)
LSM6	33-49	+	Ser (92)
LSM4	35-48	+	Leu (38),Gln (39),Ser (16)
LSM11	11-19	+	Leu (62),Gln (33),Ser (<1)
LSM10	20-35	-	Leu (99)
LSM12	5-9	-	Leu (91),Gln (9)
LSM13	12	+	Leu (15), Gln (78),Ser (<1)
LSM9	11-12	+	Leu (72),Gln (6),Ser (20)
LSM14	14-16	+	Leu (65),Gln (19),Ser (11)
Tyr	18-25	-	Tyr (95)
Tyr→Ser 1	6-8	+	Gln (71),Tyr (26),Ser (1)
Tyr→Ser 2	35-44	+	Gln (71),Tyr (29),Ser(<1)
Tyr→Ser 3	21-23	+	Gln (82),Tyr (13), Ser (1)

Values for percent amino acid inserted may not total 100%, because background levels of less than 5% for any amino acid are not included with the exception of serine. While the values given for % serine inserted is below 5% in some cases, the ability to confer ampicillin resistance indicates that the mutant must be inserting some low level of serine. Estimates for % Gln inserted include Glu. amp; ampicillin

FIGURE LEGENDS

Figure 1a. Composite of *E. coli* serine tRNAs. Nucleotides highly conserved in or correlated with serine tRNAs and not found in $tRNA^{Leu5}_{UAG}$ are indicated. Adapted from (1).

Figure 1b. $tRNA^{Leu5}_{UAG}$ with twelve base changes predicted to convert the tRNA to serine identity. Adapted from (1).

Figure 2. Mutant LSM6

Figure 3. Mutant LSM4

Figure 4. Mutant LSM11

Figure 5. Mutant LSM10

Figure 6. Mutant LSM12

Figure 7. Mutant LSM13

Figure 8. Mutant LSM9

Figure 9. Mutant LSM14

Figure 10. Composite of *E. coli* serine tRNAs. Nucleotides highly conserved in serine tRNAs and not found in $tRNA^{Tyr}$ ($su^{+}3$) are indicated.

Figure 11. Mutant Tyr→Ser 1

Figure 12. Mutant Tyr→Ser 2

Figure 13. Mutant Tyr→Ser 3

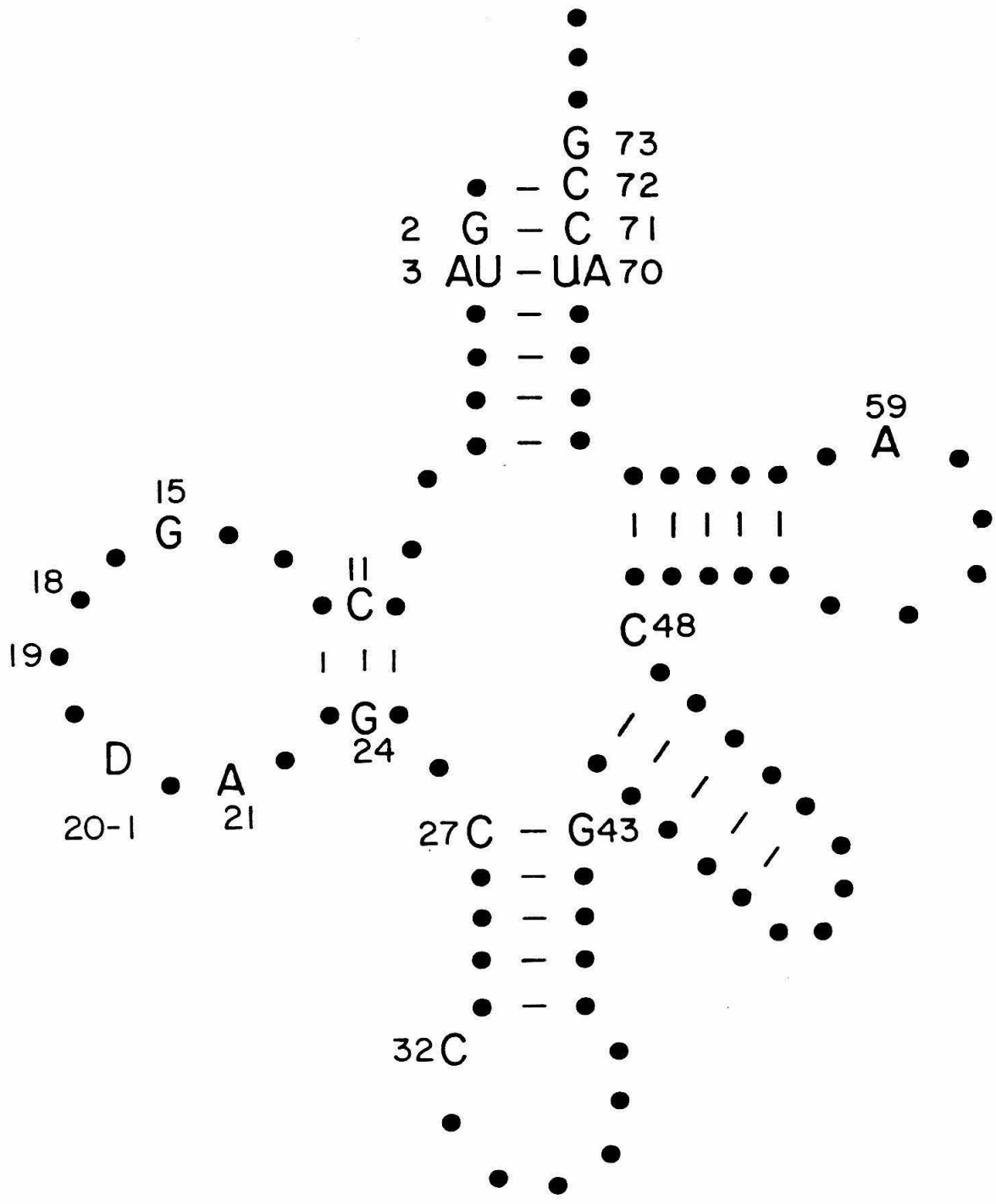

G 73
C 72
2 G – C 71
3 AU – UA 70
59
A
15
G
18
11
C
19
C 48
G
24
D
A
20-1
21
27 C – G 43
32 C

Figure 1a

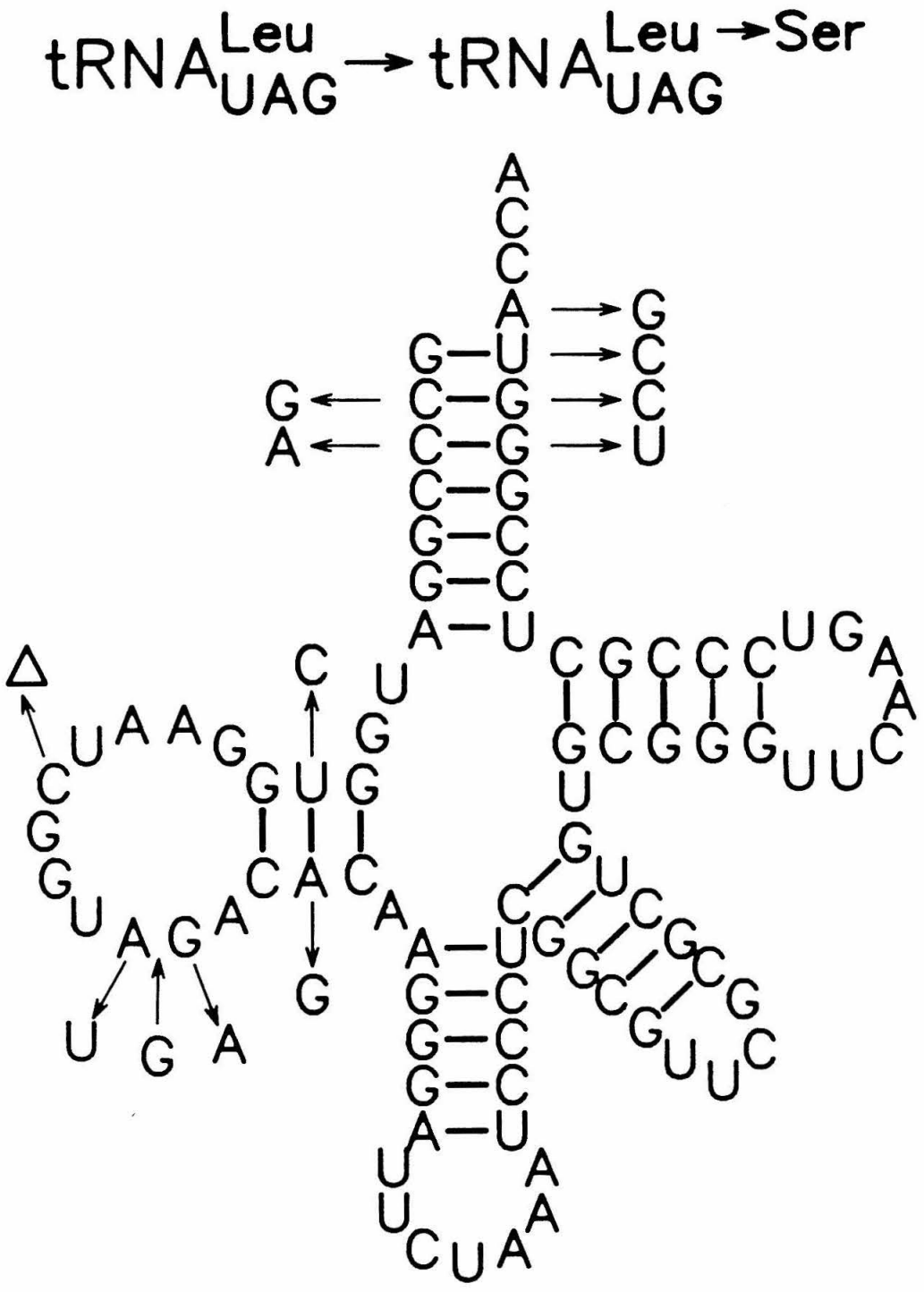


Figure 1b

LSM6

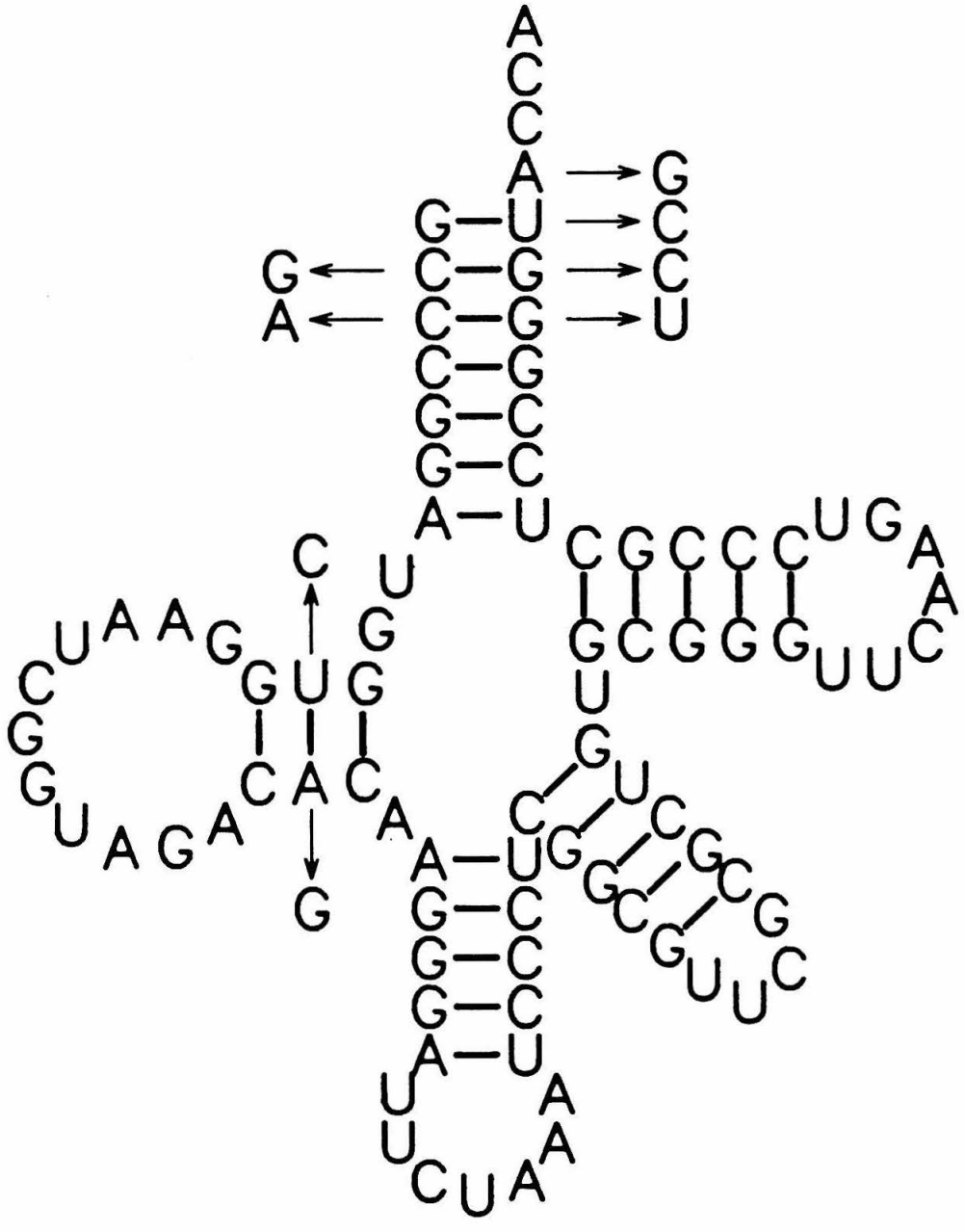


Figure 2

LSM4

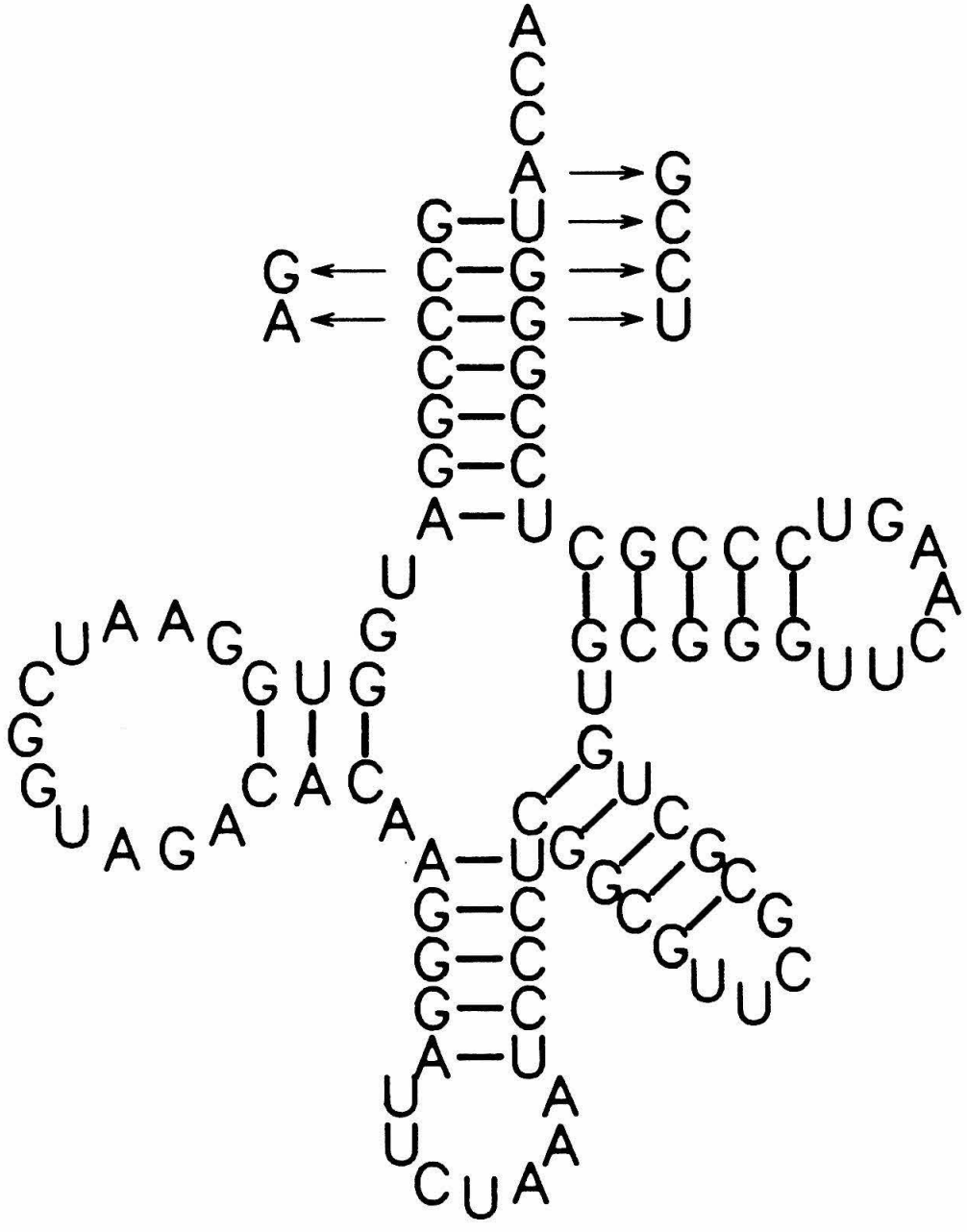


Figure 3

LSM11

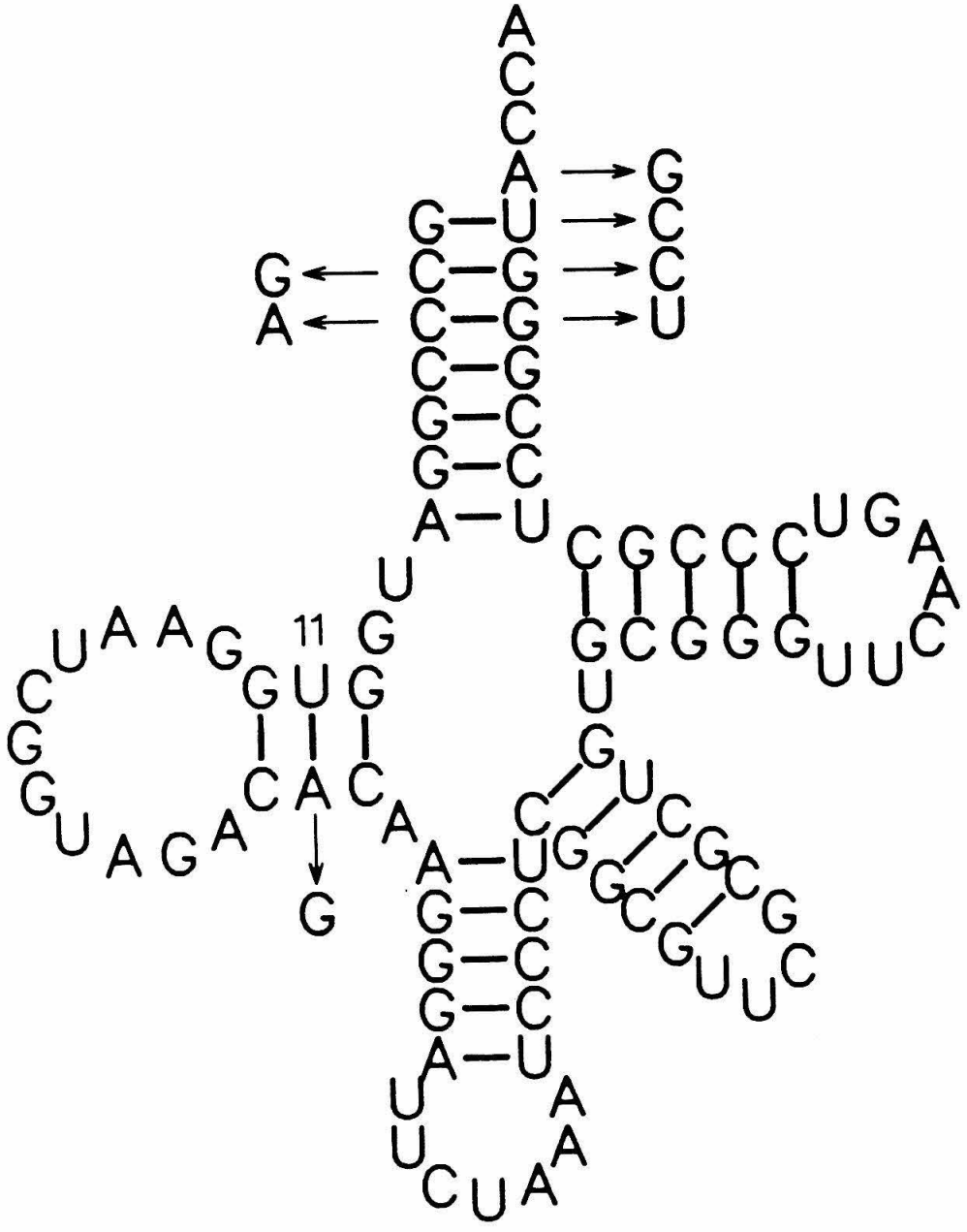


Figure 4

LSM10

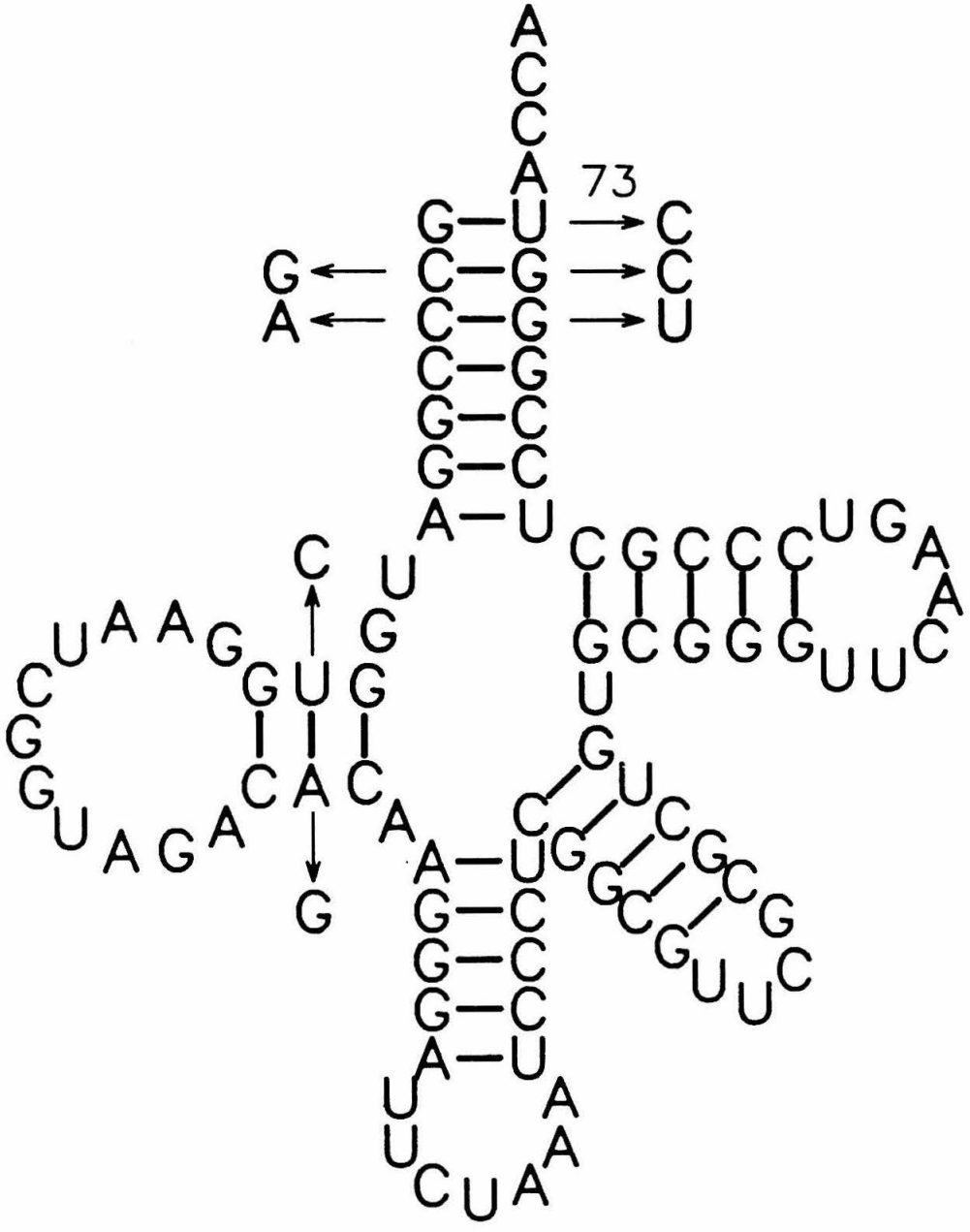


Figure 5

LSM12

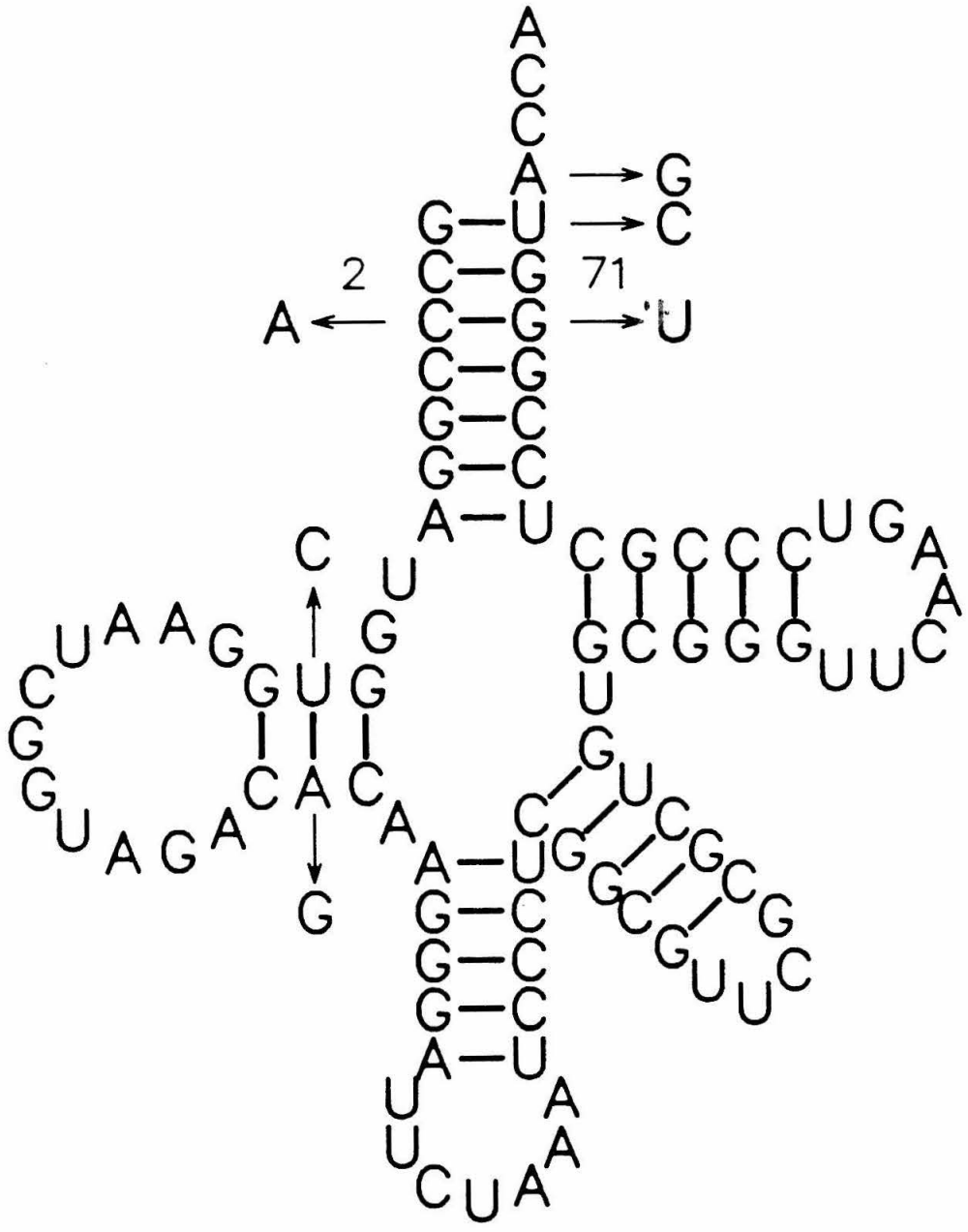


Figure 6

LSM13

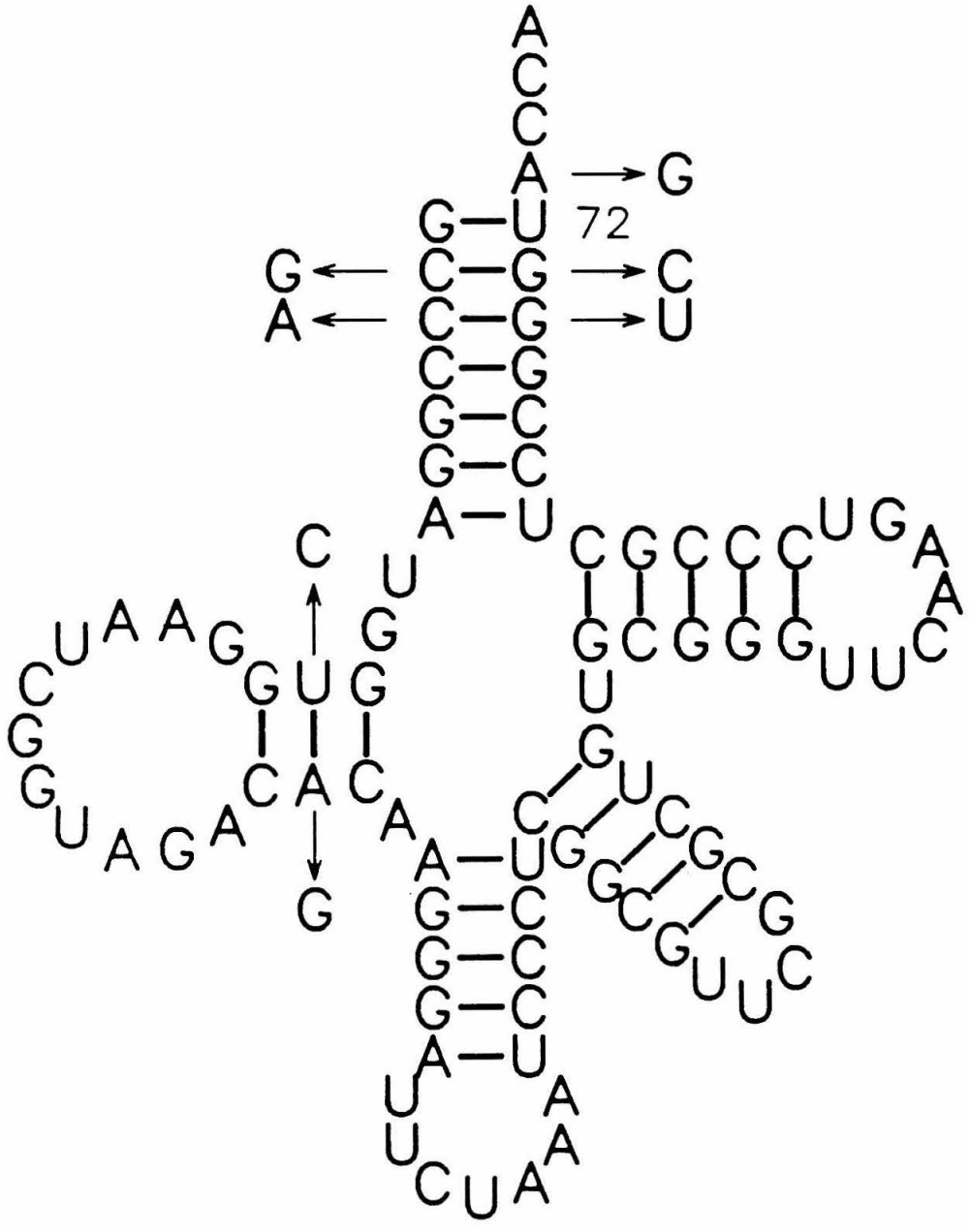


Figure 7

LSM9

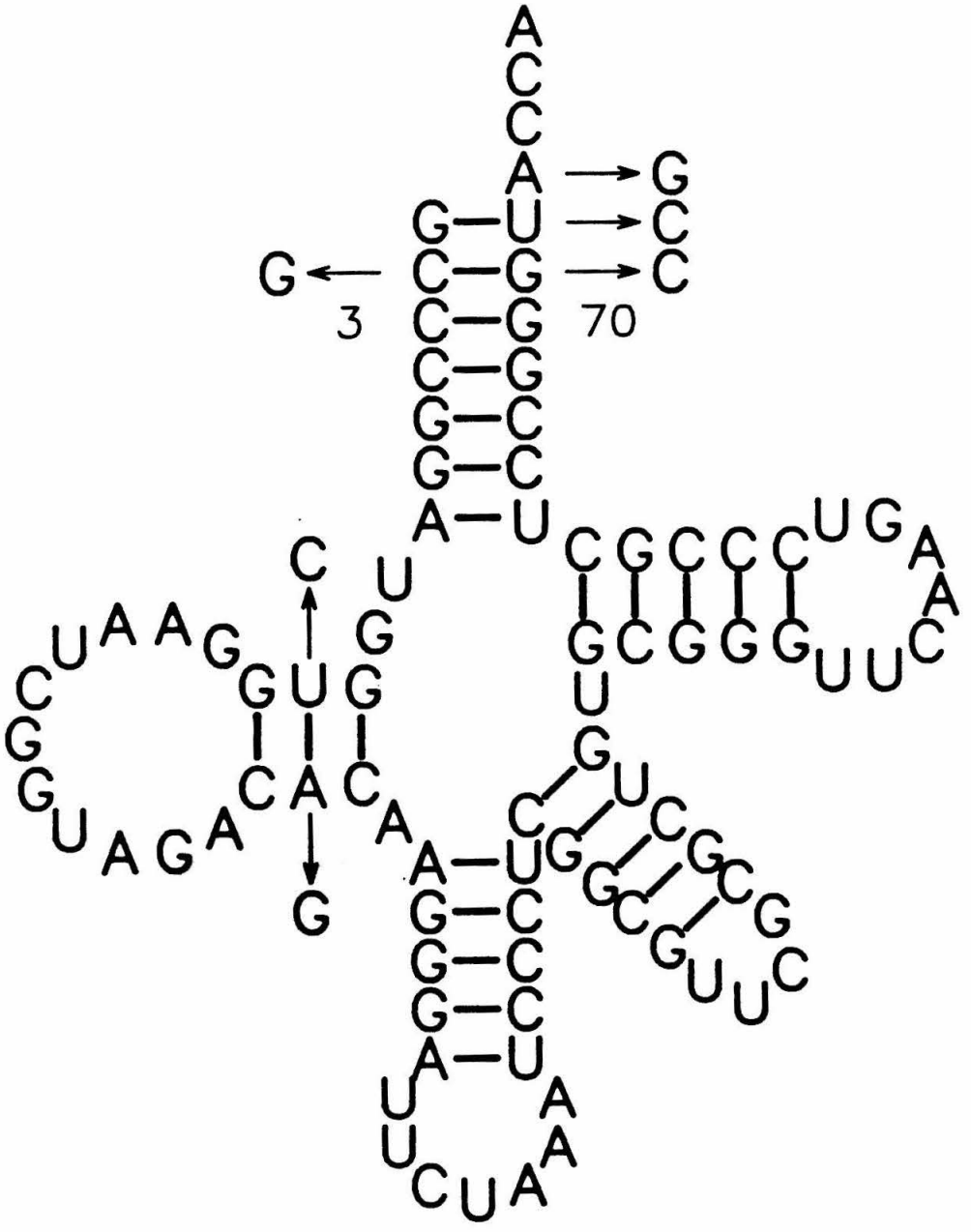


Figure 8

LSM14

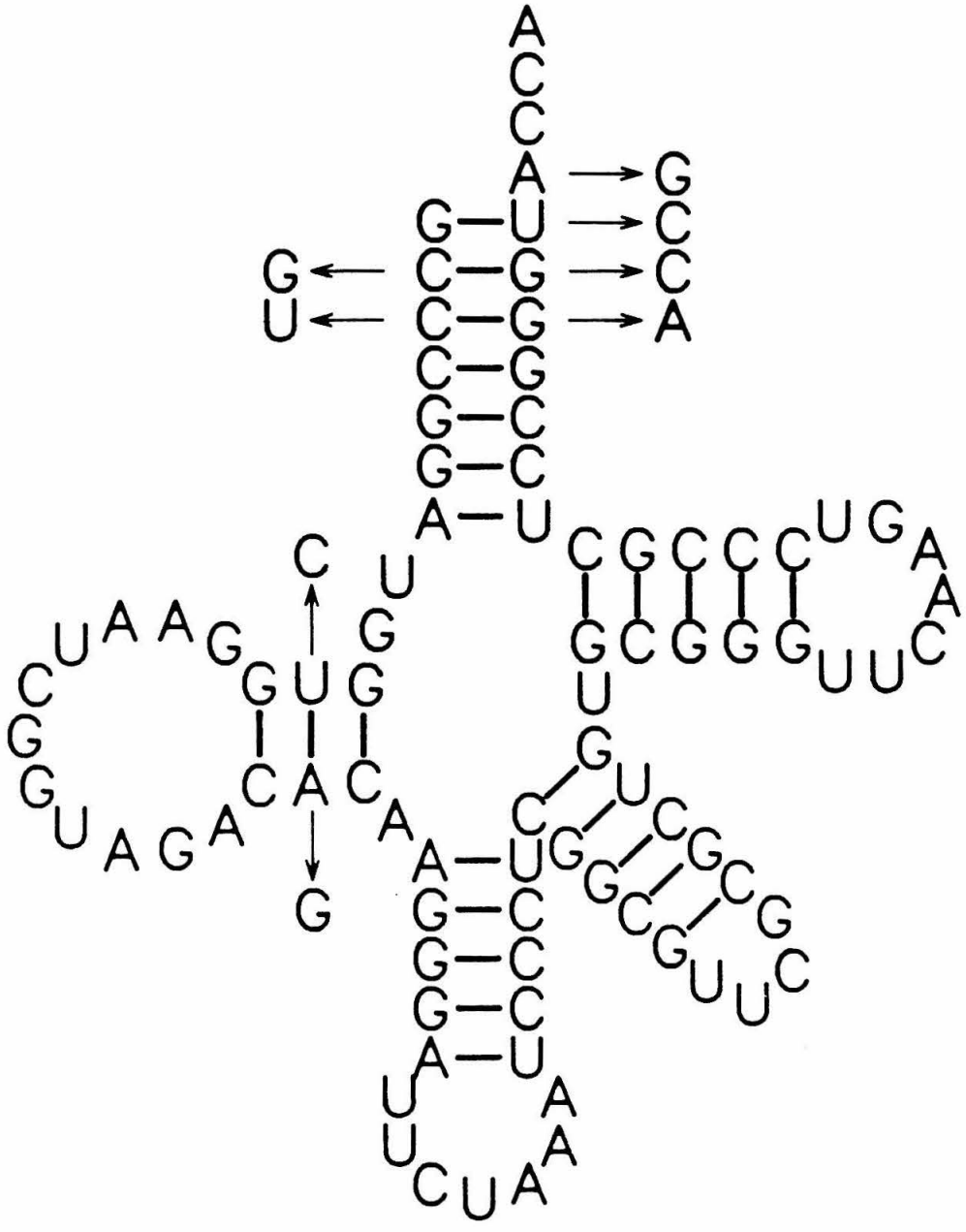


Figure 9

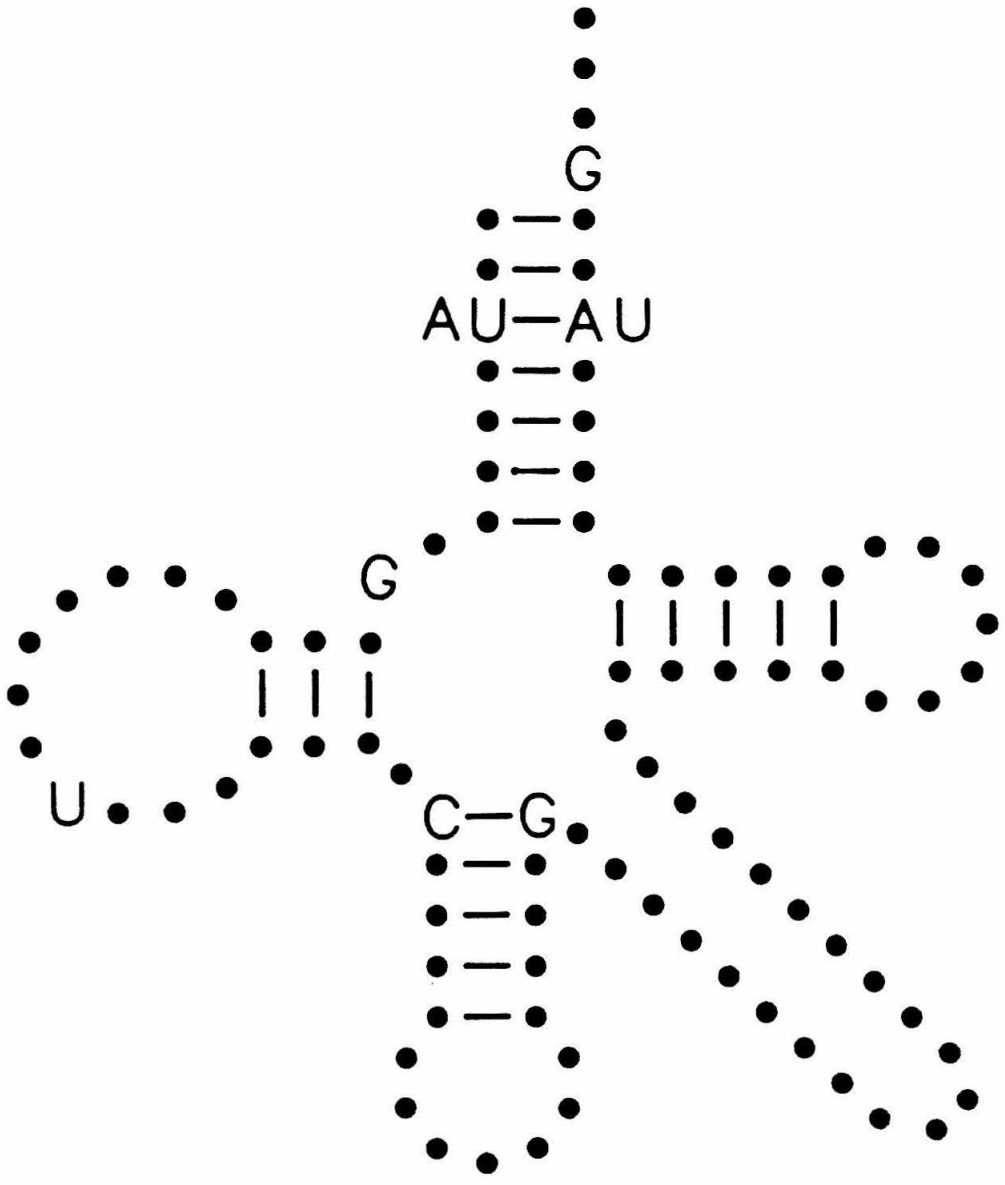


Figure 10

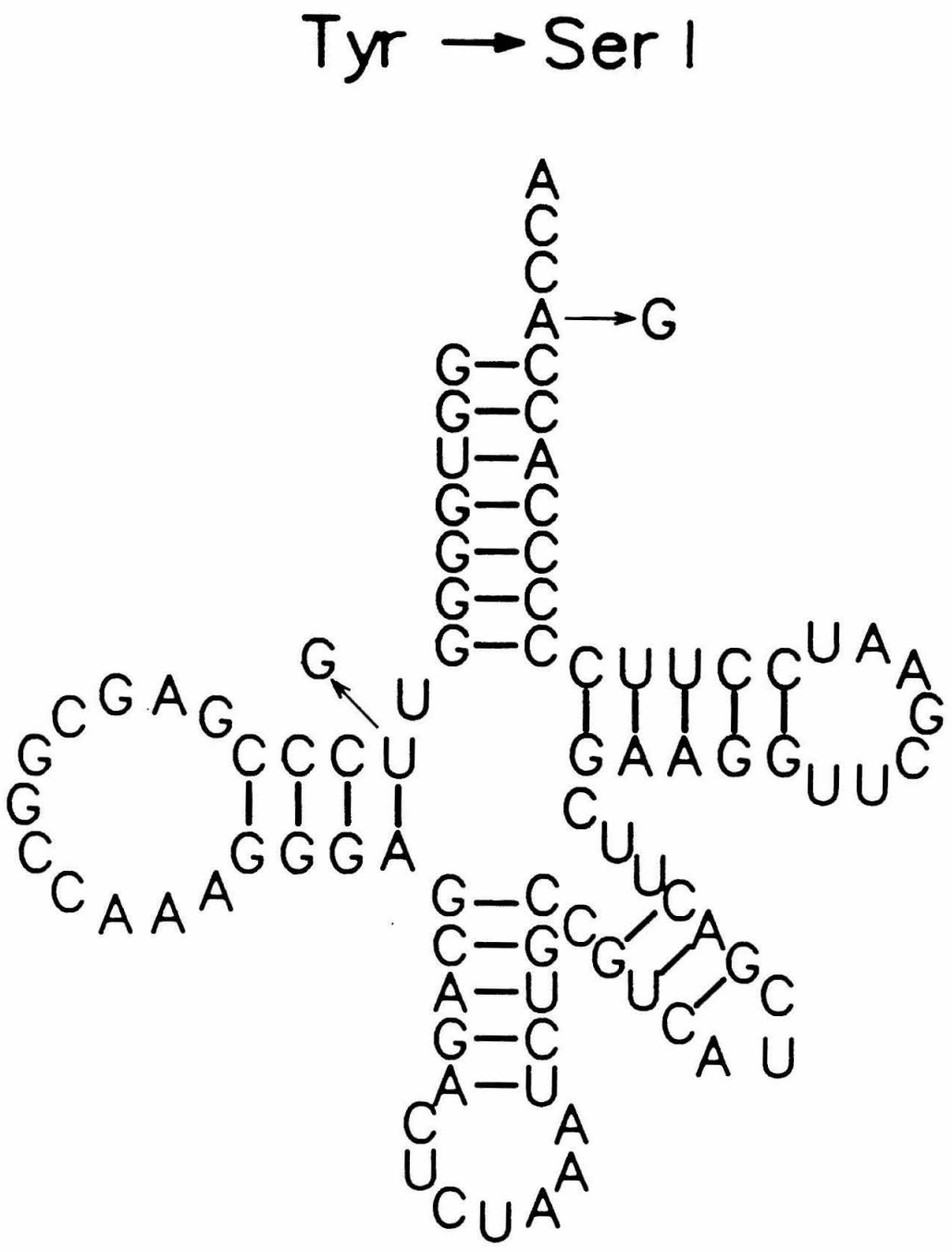


Figure 11

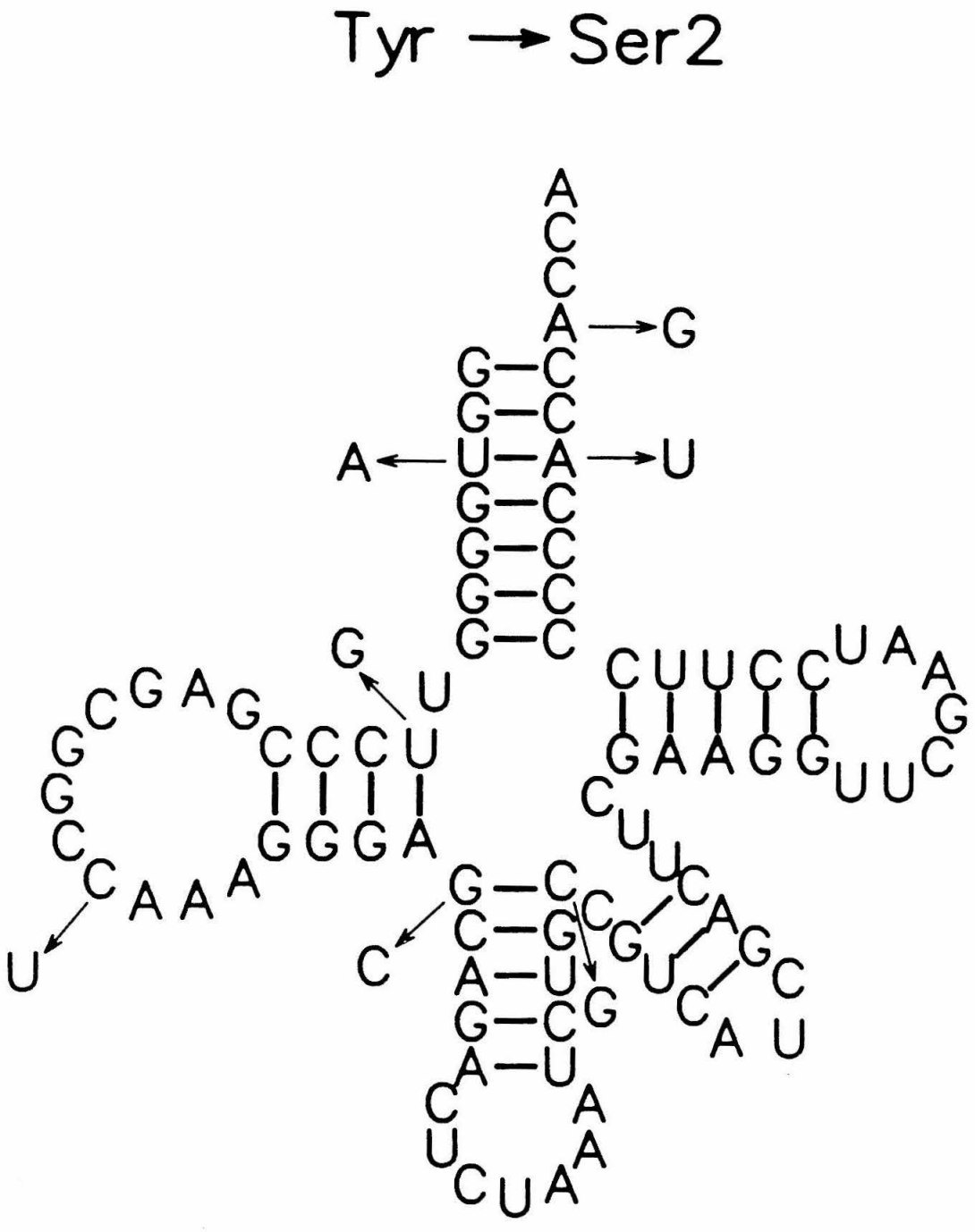


Figure 12

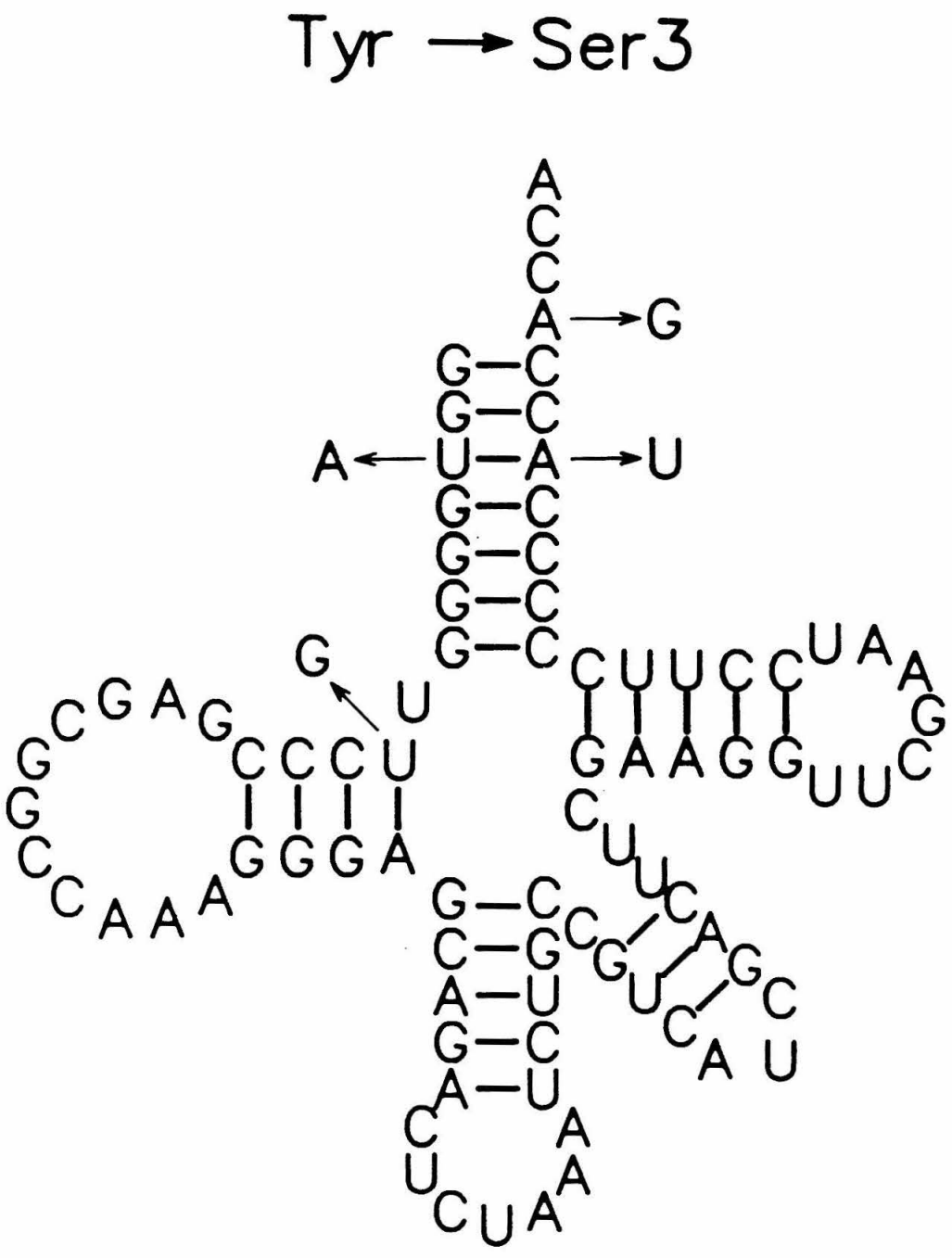


Figure 13

CHAPTER III

Construction of two *Eschericia coli* amber suppressor genes: $tRNA^{Phe}_{CUA}$ and $tRNA^{Cys}_{CUA}$

Jennifer Normanly[1], Jean-Michel Masson[2], Lynn G. Kleina[2], John Abelson[1], and Jeffrey H. Miller[2]

[1]Division of Biology, California Institute of Technology, Pasadena, CA 91125; and
[2]Department of Biology, University of California, Los Angeles, CA 90024

[Proc. Natl. Acad. Sci. USA (1986) 83:6548-6552]

ABSTRACT

Amber suppressor genes corresponding to *Escherichia coli* $tRNA^{Phe}$ and $tRNA^{Cys}$ have been constructed for use in amino acid substitution studies as well as protein engineering. The genes for either $tRNA^{Phe}_{GAA}$ or $tRNA^{Cys}_{GCA}$ both with the anticodons 5' CTA 3' were assembled from four to six oligonucleotides, which were annealed and ligated into a vector. The suppressor genes are expressed constitutively from a synthetic promoter, derived from the promoter sequence of the *E. coli* lipoprotein gene. The $tRNA^{Phe}$ suppressor ($tRNA^{Phe}_{CUA}$) is 54-100% efficient *in vivo*, while the $tRNA^{Cys}$ suppressor ($tRNA^{Cys}_{CUA}$) is 17-50% efficient. To verify that the suppressors insert the predicted amino acids, both genes were used to suppress an amber mutation in a protein coding sequence. N-terminal sequence analysis of the resultant proteins revealed that $tRNA^{Phe}_{CUA}$ and $tRNA^{Cys}_{CUA}$ insert phenylalanine and cysteine, respectively. To demonstrate the potential of these suppressors, $tRNA^{Phe}_{CUA}$ and $tRNA^{Cys}_{CUA}$ have been used to effect amino acid substitutions at specific sites in the *E. coli lac* repressor.

INTRODUCTION

Nonsense suppressor are alleles of tRNA genes altered in the anticodon, resulting in insertion of an amino acid in response to a termination codon. These suppressors have been important tools in bacterial genetics (1) as well as in the study of recognition of tRNA by aminoacyl tRNA synthetase (2). The use of nonsense suppressors also offers exciting possibilities for protein engineering, because numerous altered proteins with known amino acid changes can be generated easily. By this method, genes containing amber, ochre, or opal mutations (resulting in UAG, UAA, or UGA chain-terminating codons, respectively) are expressed in mutant strains that produce suppressor tRNAs capable of inserting an amino acid in response to the nonsense mutation. Many single amino acid substitutions at specified positions in proteins have already been made. For example, nonsense suppression has been used to create close to 400 mutant *lac* repressor proteins, each carrying a known amino acid replacement (3). Unfortunately, the range of amino acid exchanges made possible by nonsense suppression has been limited by the relatively small set of nonsense suppressors available. Until recently, only five different amino acids could

be inserted at an amber codon by suppressor tRNAs. The amber suppressor tRNAs that have been generated *in vivo* (*supD, supE, supF, and supP*), insert serine, glutamine, tyrosine, and leucine, respectively, in response to UAG codons. In addition, Murgola and coworkers have described a glycine-inserting amber suppressor derived from a series of *in vivo* genetic manipulations (4). [Ochre suppressors operating at reduced efficiencies can insert additional amino acids--for instance, *supG* directs the insertion of lysine at UAA and UAG codons (5), but the inefficient suppression greatly limits their utility for generating altered proteins.] With the exception of the glycine suppressor described by Murgola, the existing amber suppressor alleles arose from a single base change, resulting in the anticodon 5' UAG 3' (6-10). Creating new suppressor alleles by standard genetic means is unlikely, as virtually every other tRNA requires more than a single base change to yield a CUA anticodon. However, advances in oligonucleotide synthesis (11) have made possible the *in vitro* construction of new suppressor tRNAs, which permits us to greatly extend the utility of this approach. To expand the available collection of amber suppressors in *E. coli*, we have synthesized the genes for several tRNAs, altering the anticodon to enable recognition of UAG (amber) codons. Here we describe the construction of the genes for two such suppressor tRNAs, which insert phenylalanine and cysteine at high efficiency, and we demonstrate their utility in amino acid substitution studies of the *E. coli lac* repressor.

METHODS

Bacteria, Bacteriophage, Media, and Reagents. *E. coli* strain XAC-1 is F' $lacI_{373}$ $lacZ_{u118am}$ $proB^+$/ F$^-$ $\Delta(lacproB)_{x111}$ *nalA, rif,* $argE_{am}$, *ara*. The F1 phage, IR1 was obtained from E. Meyerowitz. Indicator plates used for transformations were minimal M9 glucose (12) supplemented with ampicillin (100ug/ml) and the indicator dye 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside (40ug/ml). Folic acid and methotrexate were purchased from Calbiochem and ICN, respectively.

Oligonucleotide Synthesis and Purification. Oligonucleotides were synthesized on an Applied Biosystems 380A DNA synthesizer. Three to five A_{260} units of oligonucleotide were purified by electrophoresis through a 20% acrylamide/7M urea gel,

visualized by UV shadowing, excised, and eluted in 0.3 M NaCl, 10mM Tris-HCl, pH 7.4, 1mM EDTA, 1% phenol, at 37°C for 18 hr. The DNA was precipitated and washed with ethanol, dried, and resuspended in distilled H_2O.

Gene Synthesis. Each oligonucleotide (2ug) was phosphorylated in 70mM Tris-HCl, pH 7.6, 10mM $MgCl_2$, 5mM dithiothreitol (DTT), 100uM ATP, with 4 units of polynucleotide kinase for 1 hr at 37°C. The phosphorylated oligonucleotides (80 pmol each) were mixed together in 100mM NaCl, heated to 80°C for 5 min, then allowed to cool to room temperature over a period of 3 hr. The entire mixture was combined with vector pGFIB-1 (previously cleaved with Eco RI and Pst I) to give a mass ratio of 10:1 (insert:vector). Subsequent ligation was carried out in 50mM Tris-HCl, pH 7.6, 10mM $MgCl_2$, 20 mM (DTT), 50mM NaCl, 1mM ATP, bovine serum albumin (50ug/ml), with 1 unit of T4 DNA ligase at 15°C for 12 hr. The ligation mixture was used to transform competent *E. coli* XAC-1 cells. From purified colonies exhibiting the suppressor phenotype on indicator medium, single-stranded DNA was made by using the F1 phage IR1 as helper (13) and sequenced by the chain-termination method of Sanger *et al.* (14).

Protein Purification. Purification of mutant dihydrofolate reductase (DHFR) protein was carried out by a combination of the methods of Baccanari *et al.* (15-17) and is briefly described here. *E. coli* XAC-1 carrying plasmid pDA3-12:Phe was cultured in 6 liters of minimal M9 glucose medium supplemented with ampicillin (50ug/ml) at 37°C with aeration for 18 hr. Cells were harvested and lysed (15) and the crude lysate was extracted with ammonium sulfate to 90% saturation (16). The precipitate was brought to a protein concentration of 20 mg/ml, 3 ml of methotrexate Sepharose was added, and the mixture was dialyzed for 18 hr (16). The resin was batch-washed (17) and the DHFR was eluted with folic acid (17). At this point, the DHFR was 99% pure as judged by silver staining (18) of sodium dodecyl sulfate/polyacrylamide gels (19). To remove folate and nucleic acid, peak fractions were first dialyzed against 50 mM potassium phosphate, pH 8.0, 1 mM DDT, and then applied to a DEAE-Sephacel (Pharmacia) column (1 mg/ml protein) equilibrated with the same buffer. DHFR was eluted in a linear gradient of 0-0.4M KCl in 50mM potassium phosphate, pH 8.0, 1mM DTT. Peak fractions were pooled, dialyzed against water, and lyophilized.

Protein Sequencing. Mutant DHFR protein was sequenced on an Applied Biosystems Model 470A gas-phase protein sequenator (20) by the USC Microchemical Core Laboratory. Mutant β-galactosidase was purified (21) and sequenced by A. Fowler (University of Los Angeles).

RESULTS

Gene Construction. The sequences of *E. coli* $tRNA^{Phe}_{GAA}$ and $tRNA^{Cys}_{GCA}$ with the alterations necessary to convert them to amber suppressor tRNAs are shown in Figure 1. The synthetic genes for $tRNA^{Phe}_{CUA}$ and $tRNA^{Cys}_{CUA}$ are depicted in Figure 2. Each gene is constructed from a set of 4-6 complementary oligonucleotides ranging in size from 23 to 46 mucleotides, with 7-base pair (bp) overlaps at the junctions. Each tRNA gene sequence is flanked at the 5' and 3' ends by Eco RI and Pst I sites, respectively. Since the *E. coli* cellular nucleases that process tRNA precursors recognize conserved structural elements within the tRNA and not surrounding sequences (25), we expected that these constructs would produce fully processed functional tRNAs.

The oligonucleotides were annealed and ligated into the Eco RI-Pst I sites of plasmid pGFIB-1 as described in the Methods section. In this vector (Figure 3), the tRNA gene is expressed constitutively from a strong synthetic promoter, based on the promoter sequence of the *E. coli* lipoprotein gene *lpp* (26). Distal to the promoter and restriction site polylinker is a transcription terminator, which we have synthesized based on the termination sequence of the ribosomal RNA operon *rrnC* (27-28). Plasmid pGFIB-I also carries the F1 intergenic region, which in the presence of F1 helper phage allows the production of single-stranded DNA (13), thus facilitating dideoxy sequencing of the synthetic gene.

To clone active suppressors, the ligation mixtures pGIB:Phe and pGFIB:Cys were transformed into *E. coli* stain XAC-1. This strain carries an F' *lac proB* episome with an amber mutation early in the *lacZ* gene. In an *sup*$^-$ strain, virtually no active β-galactosidase (the *lacZ* gene product) is formed. Suppression of the amber mutation should yield functional β-galactosidase, resulting in blue colonies on medium containing the indicator 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside. When XAC-1 was transformed with pGFIB:Phe or pGFIB:Cys, on average, 60% of the transformants were blue. To confirm

that this phenotype was plasmid encoded and not due to spontaneous reversion of the amber mutation, plasmid DNA was isolated from transformants that gave blue colonies and sequenced to ensure that there were no sequence anomalies in the synthetic construct.

Determination of Suppression Efficiency. We have exploited the properties of a *lacI-Z* fusion system, to measure the efficiency of suppression. In this system, the *lacI* and *Z* genes are fused, resulting in a hybrid protein with full β-galactosidase activity (29). The N-terminal, *I*-encoded portion of the fusion is not required for β-galactosidase activity, although continued transcription and translation through this portion of the fusion is necessary. Thus, the introduction of nonsense mutations that interrupt translation of the *I* portion of the message results in negligible β-galactosidase synthesis, while suppression of these nonsense mutations restores β-galactosidase synthesis. As β-galactosidase activity is independent of the nature of the amino acid being inserted into the *I*-encoded portion of the hybrid protein and dependent only on the degree of transmission through the translation block, the level of β-galactosidase activity is a direct measure of the efficiency of suppression. Previous studies of suppression at different nonsense sites have shown that the efficiency of suppression is dependent on the sequence immediately surrounding the nonsense site; the two bases 3' to the nonsense site are especially important (29, 30). Generally, amber codons followed by an A or G are suppressed more efficiently than those followed by a C or U, although codons followed by the sequence CU are well-suppressed.

We have examined $tRNA^{Phe}_{CUA}$ and $tRNA^{Cys}_{CUA}$ against a set of ambers in the *lacI-Z* fusion by assaying β-galactosidase activity of the suppressed fusion protein (12). Table I shows the relative efficiency of the new suppressors as well as those for other suppressors. $tRNA^{Phe}_{CUA}$ has a suppression efficiency of 54-100%, while the efficiency of $tRNA^{Cys}_{CUA}$ ranges from 17% to 50%, depending upon the context of the amber codon.

Determination of Suppressor Specificity. A single change in the anticodon of *E. coli* $tRNA^{Trp}$ results in mischarging by the glutamine aminoacyl tRNA synthetase (31). Similarly, alteration of the anticodon of *E. coli* $tRNA^{Met(f)}$ results in mischarging by the glutamine synthetase (32-34). Given these precedents, we designed an assay to show directly that the new suppressors actually insert the predicted amino acid. $tRNA^{Phe}_{CUA}$ and

$tRNA^{Cys}_{CUA}$ were each used to suppress an amber mutation in a protein-coding sequence. The resulting altered proteins were subsequently isolated and subjected to N-terminal sequence analysis to determine which amino acid(s) had been inserted in response to the amber codon.

For the determination of the specificity of $tRNA^{Phe}_{CUA}$, we constructed an amber mutation via oligonucleotide-directed mutagenesis, in the 10th residue (valine) of the *E. coli* *fol* gene encoding DHFR (35) (Figure 4). DHFR was chosen because it can be purified in virtually one step by methotrexate affinity chromatography. Furthermore, the structure of DHFR has been well-delineated by X-ray crystallography (39-41). The tenth residue of this protein is not highly conserved throughout phyla and is not in close proximity to the active site. Hence, we predicted that substitutions at this site would not significantly alter the protein's activity or its behavior during purification. We also converted the 11th residue from Asp to Asn. The purpose of this GAT to AAT conversion was to provide a better context for the amber mutation. Subsequently, we constructed a plasmid (pDA3-12:Phe) that carries both the *fol* amber gene under the control of the *tac* promoter (38) and $tRNA^{Phe}_{CUA}$ flanked by the lipoprotein promoter and *rrnC* terminator (Figure 4).

DHFR was isolated from XAC-1 cells harboring this plasmid and subjected to Edman degradation on an automated protein sequenator (20). In Figure 5, the picomolar yields of phenylthiohydantoin-derivatized phenylalanine, valine, asparagine, and aspartic acid are plotted versus the residue number. It should be noted that the wild-type copy of the *fol* gene is also present in the host chromosome. Under control of the *tac* promoter, the plasmid-borne *fol* gene overproduces DHFR approximately 100-fold above the level produced from the chromosomal copy. This should result in minimal levels of Vas-Asp at positions 10 and 11 (compared with Phe-Asn) in the presence of a strong suppressor such as $tRNA^{Phe}_{CUA}$. The major amino acid at position 10 is indeed phenylalanine, and no other amino acids are detected at a significant level, including valine or glutamine. In other studies, we have been able to detect 10% insertion of an amino acid at residue 10 (28). Therefore, to our limits of detection, phenylalanine is the sole amino acid inserted by the new suppressor. Residue 11 is predominantly asparagine, with a trace amount of aspartic acid. This may be the result of deamidation of asparagine or due to wild-type background.

The specificity of $tRNA^{Cys}_{CUA}$ was determined by A. Fowler in a similar fashion (personal communication). The synthetic $tRNA^{Cys}_{CUA}$ gene was used to suppress an

amber mutation located at residue 17 in the *lacZ* gene. The resulting β-galactosidase was purified and subjected to N-terminal sequence analysis. Cysteine is the major amino acid inserted at this site, and again glutamine is not detected at a significant level (data not shown).

Substitutions in the *lac* Repressor. To demonstrate the potential of the new suppressors for structure function studies, we have used them to generate specific amino acid replacements in the *E. coli lac* repressor. The *lac* repressor is especially suited to amino acid substitution studies as its properties are well-understood, its activity is readily assayed, and a large number of amber mutations in the repressor gene, *lacI*, have been generated(42). We have previously described the use of suppressed nonsense mutations to generate altered *lac* repressor molecules (3). Close to 90 different nonsense sites in the *lacI* gene have been described, and the ability to add different amino acids by suppression offers the opportunity to study a heirarchy of amino acid substitutions at each site. Table II shows several positions in the repressor that are particularly sensitive to certain amino acid substitutions. These data can yield useful information regarding the nature of the amino acid being inserted by the new suppressors. For instance, glutamine is the wild-type residue at positions 18 and 248. From Table 2 it is evident that these sites specifically require glutamine among the amino acids examined so far. Exchanges at position 18 affect operator binding, and substitutions at position 248 affect inducer binding (the I^s phenotype). These two sites can be used to determine whether a suppressor is inserting glutamine. Also, the phenylalanine at position 293 is not easily replaced without affecting the activity of the protein. Note that the phenylalanine-inserting suppressor restores the wild-type character, reinforcing the protein chemistry results for this suppressor. In addition, the tyrosines at position 17, 47, and 282, and the tryptophan at position 201 are sensitive to substitution. Only certain hydrophobic amino acids are allowed at positions 201 and 282, and the phenylalanine-inserting suppressor does restore activity in both cases. None of the previous substitutions for tyrosine at position 47 could restore even partial activity. However, as can be seen from Table 2, the phenylalanine-inserting suppressor can substitute reasonably well. Possibly, the ring moiety of the tyrosine is crucial at this position, and the phenylalanine fulfills a similar function. We note that phenylalanine cannot substitute for tyrosine at position 17, a residue that has been implicated in operator binding mediated by hydrogen bonding (43). This would be

expected, because the phenylalanine ring lacks the hydroxyl side chain of tyrosine that could participate in hydrogen bonding.

DISCUSSION

Nonsense suppressors can be generated by altering the anticodon of a tRNA to allow recognition of chain-termination codons. This approach has already met with success in several cases via oligonucleotide-directed mutagenesis (44-47).

We have described the construction of two amber suppressor tRNA genes, $tRNA^{Phe}_{CUA}$ and $tRNA^{Cys}_{CUA}$, which insert the predicted amino acid at a high level of efficiency. By using the approach of total gene synthesis, we have been able to construct these suppressor genes very rapidly. Furthermore, the manner in which the genes are designed, i.e., without their natural 5' and 3' flanking sequences, does not preclude the synthesis of a functional tRNA. We have subsequently assembled the genes for several other suppressor tRNAs in an effort to create as large a "bank" of amber suppressors as possible (unpublished results). One potential obstacle to creating a complete collection of amber suppressors is the possibility that the anticodon of certain tRNAs is crucial for recognition by its cognate aminoacyl tRNA synthetase. We therefore developed an assay that would enable us to determine whether the anticodon had affected the specificity of the tRNA in question, by showing directly which amino acid it inserts into an amber site of a protein coding sequence. From these protein chemistry studies, it is evident that the new suppressors insert only the original amino acids, phenylalanine and cysteine, at least to the limits of detection of the protein sequencing systems. Data from amino acid replacement studies in the *lac* repressor, utilizing $tRNA^{Phe}_{CUA}$ and $tRNA^{Cys}_{CUA}$, support this conclusion. To extend the sensitivity of detection beyond the limits of protein sequencing, genetic systems similar to the one featured in Table 2 can be developed further.

ACKNOWLEDGEMENTS

We thank Lynn Williams of the University of Southern California Microchemical Core Laboratory and Audree Fowler of the University of California at Los Angeles for protein sequencing. The manuscript was improved by comments from V. Bankaitis, M. Clark, E. Grayhack, A. Lustig, E. Phizicky, and S. Ruby. This work was supported in part by a

National Science Foundation grant (DMB-8417353) to J. A. and J. H. M., a National Research Service Award (T32GM07616) from the National Institute of General Medical Sciences to J. N., and a North Atlantic Treaty Organization fellowship and Philippe Foundation fellowship to J.-M. M.

REFERENCES

1. Steege, D. A., Söll, D. 1979. in *Biological Regulation and Development,* (ed. Goldberger, R. F.) Plenum Press, New York Vol. I, pp. 433-85
2. Ozeki, H., Inokuchi, H., Yamao, F., Kodaira, M., Sakano, H., Ikemura, T., Shimura, Y. 1980. in *Transfer RNA: Biological Aspects* (eds. Söll, D., Abelson, J. N, Schimmel, P. R.) Cold Spring Harbor Laboratory Press, New York pp. 341-2
3. Miller, J. H., Coulondre, C., Hofer, M., Schmeissner, U., Sommer, H., Schmitz, A., Lu, P. 1979. *J. Mol. Biol.* 131:191-222
4. Murgola, E. J., Prather, N. E., Pagel, F. T., Mims, B. H., Hijazi, K. A. 1984. *Mol. Gen. Genet.* 193:76-81
5. Gorini, L. 1970. *Ann. Rev. Genet.* 4:107-34
6. Steege, D. A. 1983. *Nucleic Acids Res.* 11:3823-32
7. Yoshimura, M., Inokuchi, H., Ozeki, H. 1984. *J. Mol. Biol.* 177:627-44
8. Thorbjarnardottir, S., Dingermann, T., Rafnar, T., Andreson, O. S., Söll, D., Eggertsson, G. 1985 *J. Bacteriol.* 161:219-22
9. Inokuchi, H., Yamao, F., Sakano, H., Ozeki, H. 1979. *J. Mol. Biol.* 132:649-62
10. Goodman, H. M., Abelson, J., Landy, A., Brenner, S., Smith, J. D. 1968. *Nature* 217:1019-24
11. Khorana, H. G. 1979. *Science* 203:614-25
12. Miller, J. H. 1972. in *Experiments in Molecular Genetics*. Cold Spring Harbor Laboratory Press, New York
13. Dente, L., Cesareni, G. Cortese, R. 1983. *Nucleic Acids Res.* 11:1645-55
14. Sanger, F., Nicklen, S., Coulson, A. R. 1977. *Proc. Natl. Acad. Sci. U.S.A.* 74:5463-7
15. Baccanari, D., Phillips, A., Smith, S., Sinski, D., Burchall, J. 1975. *Biochemistry* 14:5267-73

16. Baccanari, D. P., Averett, D., Briggs, C., Burchall, J. 1977. *Biochemistry* 16:3566-72
17. Baccanari, D. P., Stone, D., Kuyper, L. 1981. *J. Biol. Chem.* 256:1738-47
18. Wray, W., Boulikas, T., Wray, V. P., Hancock, R. 1981. *Anal. Biochem.* 118:197-203
19. Laemmli, U. K. 1970. *Nature* 227:680-5
20. Hewick, R. M., Hunkapiller, M. W., Hood, L. E., Dreyer, W. J. 1981. *J. Biol. Chem.* 256:7990-7
21. Fowler, A. V., Zabin, I. 1983. *J. Biol. Chem.* 258:14354-8
22. Barrell, B. G., Sanger, F. 1969. *Febs. Lett.* 3:275-8
23. Mazzara, G. P., McClain, W. H. 1977. *J. Mol. Biol.* 117:1061-79
24. Deutscher, M. P. 1984. *CRC Crit. Rev. Biochem.* 17:45-71
25. Abelson, J. 1979. *Ann. Rev. Biochem.* 48:1035-69
26. Nakamura, K., Inouye, M. 1979. *Cell* 18:1109-17
27. Young, R. A. 1979. *J. Biol. Chem.* 254:12725-31
28. Normanly, J., Ogden, R. C., Horvath, S. J., Abelson, J. 1986. *Nature* 321:213-19
29. Miller, J. H., Albertini, 1983. *J. Mol. Biol.* 164:59-71
30. Bossi, L. 1983. *J. Mol. Biol.* 164:73-87
31. Yaniv, M., Folk, W. R., Berg, P., Soll, L. 1974. *J. Mol. Biol.* 86:245-68
32. Schulman, L. H., Pelka, H., Susani, M. 1983. *Nucleic Acids Res.* 11:1439-55
33. Schulman, L. H., Pelka, H. 1983. *Proc. Natl. Acad. Sci. U.S.A.* 80:6755-9
34. Schulman, L. H. 1972. *Proc. Natl. Acad. Sci. U.S.A.* 69:3594-7
35. Smith, D. R., Calvo, J. M. 1980. *Nucleic Acids Res.* 8:2255-74
36. Newman, A. J., Ogden, R. C., Abelson, J. A. 1983. *Cell* 35:117-25
37. Villafranca, J. E., Howell, E. E., Voet, D. H., Strobel, M. S., Ogden, R. C., Abelson, J.N., Kraut, J. 1983 *Science* 222:782-8
38. Amann, E., Brosius, J., Ptashne, M. 1983. *Gene* 25:167-78
39. Bolin, J. T., Filman, D. J., Matthews, D. A., Hamlin, R. C., Kraut, J. 1982. *J. Biol. Chem.* 257:13650-62
40. Matthews, D. A., Alden, R. A., Bolin, J. T., Freer, S. T., Hamlin, R., Xuong, N., Kraut, J., Poe, M., Williams, M., Hoogsteen, K. 1977. *Science* 197:452-5

41. Filman, D. J., Bolin, J. T., Matthews, D. A., Kraut, J. 1982. *J. Biol. Chem.* 256:13663;72

42. Miller, J. H. 1978. in *The Operon,* (eds. Miller, J. H., Reznikoff, W. S.) Cold Spring Harbor Laboratory, New York pp. 31-88

43. Steitz, T. A., Wever, I. T., Matthews, J. B. 1982. *Cold Spring Harbor Symp. Quant. Biol.* 47:419-26

44. Capone, J. P., Sharp, P. A., RajBhandary, U. L. 1985. *EMBO J.* 4:213-21

45. Laski, F. A., Belagaje, R., RajBhandary, U. L., Sharp, P. A. 1982.*Proc. Natl. Acad. Sci. U.S.A.* 79:5813-17

46. Laski, F. A., Belagaje, R., Hudziak, R. M., Capecchi, M. R., Palese, P., RajBhandary, U. L., Sharp, P. A. 1984. *EMBO J.* 3:2445-52

47. Hudziak, R. M., Laski, F. A., RajBhandary, U. L., Sharp, P. A., Capecchi, M. R., 1982. *Cell* 31:137-46

Table 1. Suppression of amber mutations in a *lacI-Z* fusion.

Suppressor	β-Galactosidase activity, % of wild type						
	O21c	A30	O28c	A26	A16	O17c	O13c
$tRNA^{Phe}_{CUA}$	79	100	78	100	78	54	84
$tRNA^{Cys}_{CUA}$	34	53	50	51	35	17	35
supD	27	54	32	25	21	6	26
supE	15	26	10	11	10	0.8	11
supF	34	67	62	100	43	11	46
supP	72	70	80	100	62	30	58

All assays were carried out at 37°C and were determined in duplicate (12). Values are given as the percentage of the wild-type fusion in the respective suppressor strains. Values for *supD, supE, supF, and supP* are taken from Miller and Albertini (29).

Table 2. Amino acid substitutions in the *lac* repressor.

Wild-type residue	Replacement by suppression					
	Ser	Gln	Tyr	Leu	Phe	Cys
Tyr-17	+/-	-/+	+	-/+	-/+	-/+
Try-47	-	-	+	-	+/-	-
Tyr-282	-	-	+	+	+	-
Trp-210	-	-	+	-	+	-
Phe-293	+/-	+/-	s	s	+	s
Gln-18	-	+	-	-	-	-/+
Gln-248	s	+	s	s	s	s

+, Normal activity; s, DNA binding activity retained but not inducibility; -, no activity, presumably because of defects in folding, aggregation or DNA binding. Partial activities: +/- indicates a less severe defect that -/+.

FIGURE LEGENDS

Figure 1. RNA sequences of *E. coli* $tRNA^{Phe}_{GAA}$ (22) and $tRNA^{Cys}_{GCA}$ (23) with the base changes that would enable the tRNAs to recognize amber codons.

Figure 2. Synthetic genes for $tRNA^{Phe}_{CUA}$ and $tRNA^{Cys}_{CUA}$. The first and last nucleotides encoding each tRNA (indicated by a dot) are immediately flanked by the cohesive ends of Eco RI and Pst I restriction endonuclease sites, respectively. Arrows indicate the oligonucleotide junctions, and the altered anticodon is underlined. Note that the $tRNA^{Phe}_{CUA}$ gene does not encode the 3'-terminal C and A residues. These residues are added *in vivo* by tRNA nucleotidyltransferase (24).

Figure 3. Plasmid pGFIB-I. This vector is a derivative of pEMBL8+ (13). We have replaced the *lac* promoter with a synthetically constructed promoter (48) based upon the promoter sequence of the *E. coli* lipoprotein gene *lpp* (26). Two complementary oligonucleotides encoding a transcription termination sequence from the *E. coli rrnC* operon (27) have been ligated into the Pst I and Hind III site of the polylinker (28). The oligonucleotides encoding the tRNA genes are ligated into the Eco RI and Pst I sites of the polylinker.

Figure 4. Construction of plasmid pDa3-12. Oligonucleotide-directed mutagenesis (36) was carried out on the *E. coli fol* gene (35) encoded by a 1-kilobase fragment contained in the vector M13mp8 (37). Plasmid pDa3-12:sup tRNA, in this case pDa3-12:Phe, was constructed by a series of subcloning steps. The 247-bp Eco RI-Hind III fragment of pTAC12H (38) containing the *tac* promoter was ligated into the Eco RI-Hind III sites of pBR322 to create pTAC10. Subsequently, the DHFR amber encoded by a 1057-bp Hind III fragment, was subcloned into the Hind III site of pTAC10, which yielded pDa3-12. The 350 bp Pvu II fragment from pGFIB:Phe containing the lipoprotein promoter, $tRNA^{Phe}_{CUA}$, and the *rrnC* terminator was ligated into pDa3-12 to yield pDa3-12:Phe.

Figure 5. The picomolar yields of phenylthiohydantoin-derivatized phenylalanine, valine, asparagine, and aspartic acid have been plotted versus the residue number for DHFR purified from *E. coli* XAC-1 cells carrying pDa3-12:Phe.

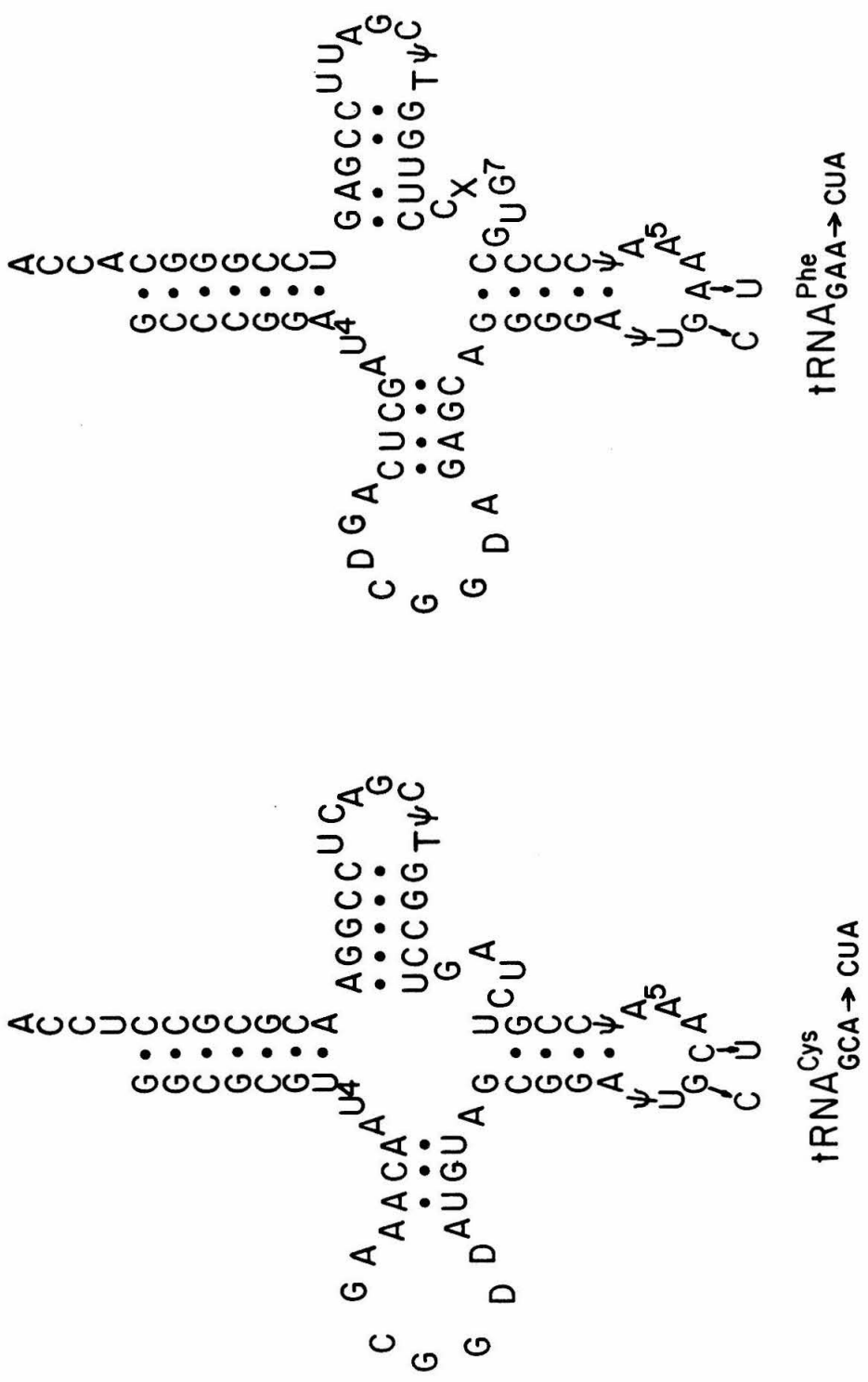


Figure 1

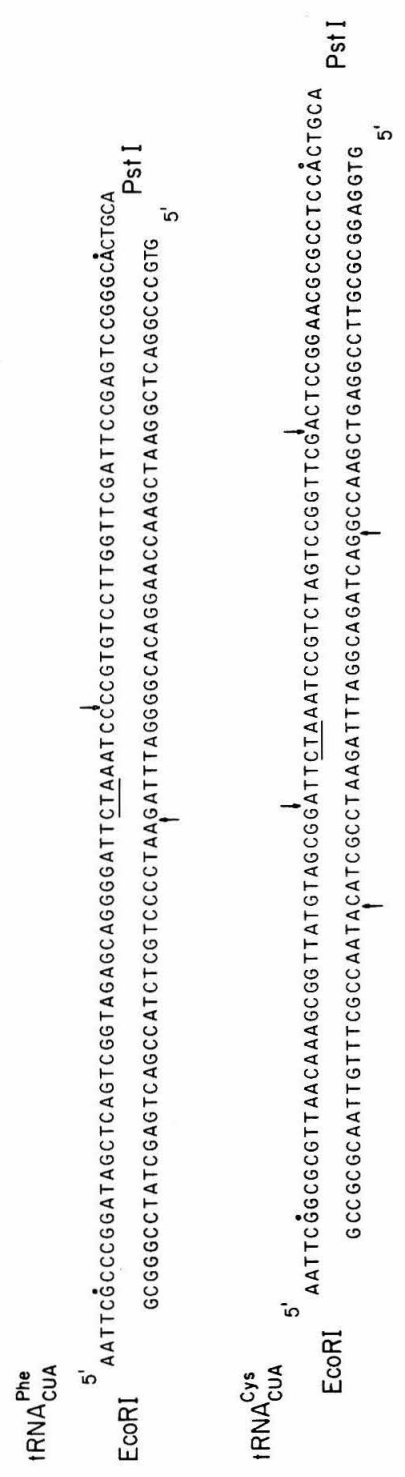

tRNA$^{Phe}_{CUA}$
5'
AATTCGCCCGGATAGCTCAGTCGGTAGAGCAGGGGATTCTAAATCCCCGTGTCCTTGGTTCGATTCCGAGTCCGGGCACTGCA
EcoRI
GCGGGCCTATCGAGTCAGCCATCTCGTCCCCTAAGATTTAGGGGCACAGGAACCAAGCTAAGGCTCAGGCCCGTG
5'
Pst I
tRNA$^{Cys}_{CUA}$
5'
AATTCGGCGCGTTAACAAAGCGGTTATGTAGCGGATTCTAAATCCGTCTAGTCCGGTTCGACTCCGGAACGCGCCTCCACTGCA
EcoRI
GCCGCGCAATTGTTTCGCCAATACATCGCCTAAGATTTAGGCAGATCAGGCCAAGCTGAGGCCTTGCGCGGAGGTG
5'
Pst I

Figure 2

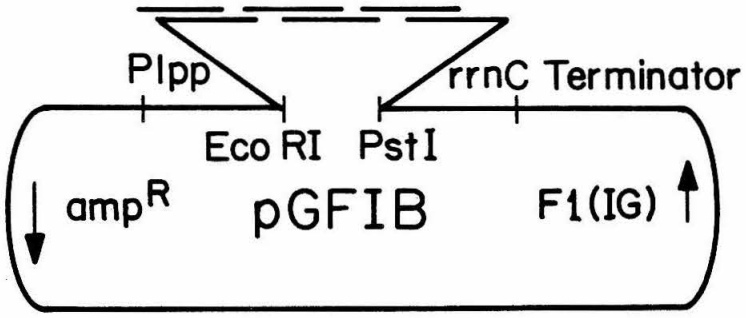


Figure 3

DHFR amber

	2									12
DHFR wt:	Ile	Ser	Leu	Ile	Ala	Ala	Leu	Ala	Val	Asp Arg
	ATC	AGT	CTG	ATT	GCG	GCG	TTA	GCG	GTA	GAT CGC
DHFR am:	...	...	...	...	...	...	...	...	am Asn	...
	...	...	...	...	...	...	...	...	TAG AAT	...

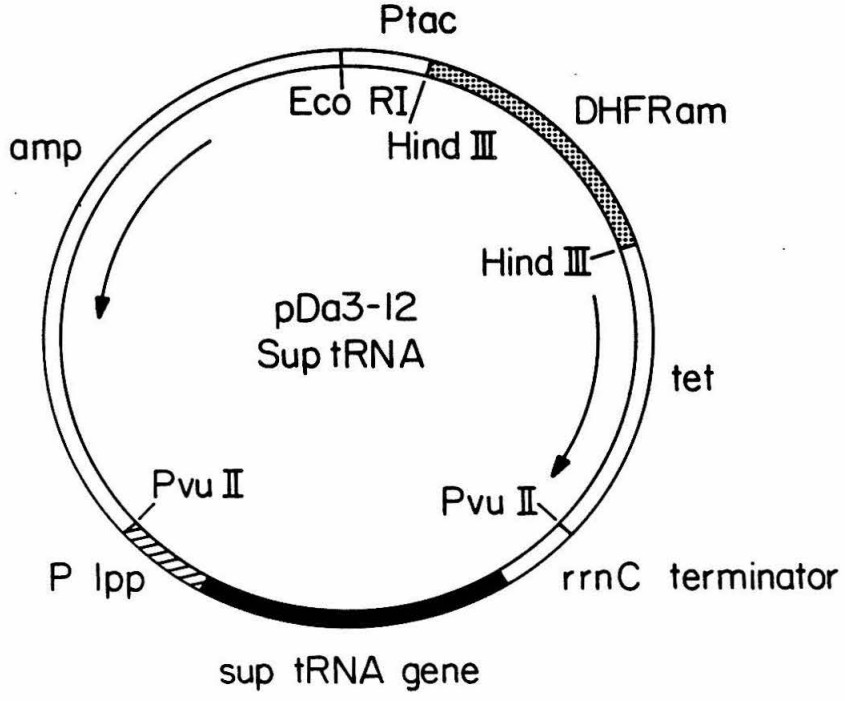


Figure 4

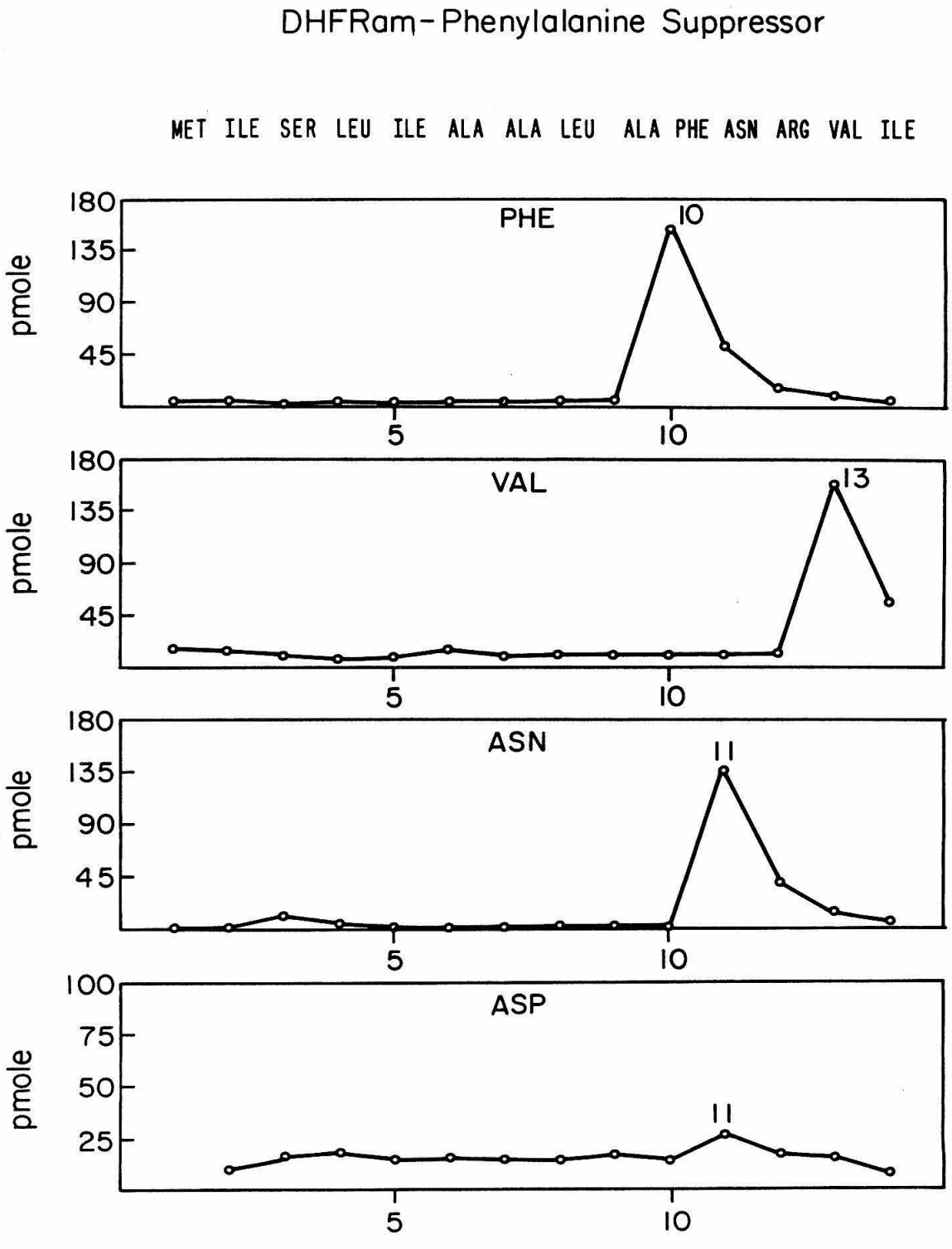

DHFRam-Phenylalanine Suppressor
MET ILE SER LEU ILE ALA ALA LEU ALA PHE ASN ARG VAL ILE
PHE
10
VAL
13
ASN
11
ASP
11
pmole
180
135
90
45
100
75
50
25
5
10

Figure 5

CHAPTER IV

Construction of *Escherichia coli* amber suppressor genes: Optimization of suppressor efficiency and determination of tRNA specificity

The work described herein was carried out in collaboration with Jeffrey H. Miller, Lynn G. Kleina, and Jean-Michel Masson; two other manuscripts detailing parts of this work are in preparation.

ABSTRACT

Using synthetic oligonucleotides, we have constructed a collection of *E. coli* amber suppressor tRNA genes (1, 2). These suppressors vary in efficiency, and we have been able to increase the suppression efficiency for some of the tRNAs via mutagenesis. In order to determine the specificity of these tRNAs, each was used to suppress an amber (UAG) nonsense mutation in the *E. coli* dihydrofolate reductase gene *fol*. The mutant proteins were purified and subjected to N-terminal sequence analysis to determine which amino acid had been inserted by the suppressor tRNAs at the position of the amber codon. The suppressors can be classified into three groups, based upon the protein sequence information. Class I suppressors, $tRNA^{Ala2}_{CUA}$, $tRNA^{GlyU}_{CUA}$, $tRNA^{HisA}_{CUA}$, $tRNA^{Lys}_{CUA}$, and $tRNA^{ProH}_{CUA}$, inserted the predicted amino acid. The Class II suppressors, $tRNA^{GluA}_{CUA}$, $tRNA^{GlyT}_{CUA}$, and $tRNA^{Ile1}_{CUA}$ were either partially or predominantly mischarged by the glutamine aminoacyl tRNA synthetase (AAS). The Class III suppressors, $tRNA^{Arg}_{CUA}$, $tRNA^{AspM}_{CUA}$, $tRNA^{Ile2}_{CUA}$, $tRNA^{Thr2}_{CUA}$, $tRNA^{Met(m)}_{CUA}$ and $tRNA^{Val}_{CUA}$ inserted predominantly lysine.

INTRODUCTION

We have set out to construct a complete collection of *E. coli* amber suppressor tRNA genes for use in studies of protein structure and function. We previously reported the construction of amber suppressor alleles of *E. coli* $tRNA^{Phe}$ and $tRNA^{Cys}$ from synthetic oligonucleotides (3). Subsequently, we have constructed amber suppressor alleles for an additional 15 tRNAs in a similar fashion (1, 2). These tRNAs, each with a CUA anticodon to enable recognition of amber codons, are constitutively expressed from a high copy number plasmid (4, 5). Table 1 lists the suppression efficiencies of the newly generated tRNAs. For a number of the amber suppressor tRNAs the suppression efficiencies were so low that they would not be useful in amino acid substitution studies. We undertook to improve the suppression efficiencies of a number of the tRNAs by several means.

I. Optimization of suppressor efficiency

Different Isoacceptors. We found that in some cases, the amber suppressor efficiencies varied among isoacceptors of a given tRNA species. This was the case for alanine, isoleucine, threonine and valine tRNAs. $tRNA^{Ala1}_{CUA}$, $tRNA^{Ile1}_{CUA}$, $tRNA^{Thr1}_{CUA}$, and $tRNA^{Val1}_{CUA}$ were all non-functional suppressors as evidenced by their very low suppression efficiencies. Simply generating an amber allele of a different isoacceptor for these tRNAs resulted in suppressor tRNAs which were significantly more efficient (Table 1). Subsequent to generating the Ala1 suppressor, we learned that the published sequence, which the tRNA gene was based on, was in question. The fact that the sequence is incorrect would help to explain why the Ala1 suppressor was inefficient. We do not know if the same holds for the other inefficient tRNAs.

The Extended Anticodon Hypothesis. The extended anticodon hypothesis put forth by Yarus (6) states that ". . . the structures of the anticodon loop and the proximal anticodon stem are related to the sequence of the anticodon." Comparison of the anticodon stem and loop sequences from *E. coli* tRNAs reveals that they are related to the 3' nucleotide of the anticodon, referred to as the *cardinal* nucleotide. Put another way, for a given nucleotide at the 3' position of the anticodon, the sequence of the anticodon stem and loop (the extended anticodon) will have a predictable sequence. These findings are applicable to suppression efficiency in the following manner. To generate an amber suppressor allele, it is necessary to alter the anticodon to CUA, by one or more base changes. Yarus noted that for weak amber suppressors the mutations necessary to render an amber suppressor resided in the cardinal nucleotide, while strong amber suppressors did not have an alteration in the cardinal nucleotide. According to the extended anticodon hypothesis, when the cardinal nucleotide is altered as in the case of the weak suppressors, the proximal anticodon stem and loop sequences are no longer related to the cardinal nucleotide, resulting in decreased translational efficiency. To generate the aspartic acid suppressor, the cardinal nucleotide was altered from C to A, and the resultant suppressor was non-functional. Based upon the consensus sequences that were derived for each of the four possible cardinal nucleotides, the extended anticodon of $tRNA^{Asp}_{CUA}$ suppressor deviated at five positions. Using oligonucleotide-directed mutagenesis, we altered those five nucleotides so that the extended anticodon would conform to the consenesus sequence for a cardinal nucleotide of "A." The resultant tRNA, $tRNA^{AspM}_{CUA}$, was tenfold more efficient than the original $tRNA^{Asp}_{CUA}$.

3' Double A. The extended anticodon hypothesis dictates that if the cardinal nucleotide is an A, the two 3' proximal nucleotides (37 and 38) should be A as well. Position 37 of the glutamic acid and histidine suppressors, both fairly weak, was changed to A, resulting in a two- to sixteen-fold increase in suppression efficiency (Table 1). It should be noted that in generating $tRNA^{Ala2}_{CUA}$, the alternate alanine isoacceptor, position 37 was changed to an A as well.

The Hybrid Molecule Approach. While the proline amber suppressor is non-functional (<1% efficient), the phenylalanine amber suppressor is very strong (48-100%). Speculating that the anticodon stem and loop do not play a role in proline AAS recognition, we generatd a tRNA variant (via gene synthesis) with the anticodon stem and loop region of $tRNA^{Phe}_{CUA}$ and the remaining regions of the tRNA from $tRNA^{Pro}$. This mutant, $tRNA^{ProH}_{CUA}$ had a suppression efficiency ranging from 8 to 60%.

In summary, using the above approaches we were able to generate amber suppressor tRNAs for 14 species of tRNA that had reasonably high suppression efficiencies, from 6 to 100%, depending upon the context of the amber codon being suppressed (Table 1). We have not yet been able to generate an efficient suppressor allele for $tRNA^{Asn}$.

II. Determination of tRNA specificity

The anticodon of a tRNA potentially plays a role in the recognition of the tRNA by its cognate aminoacyl tRNA synthetase (AAS). For example, altering the anticodon of the tryptophan or formyl-methionine tRNAs from *E. coli*, results in mischarging by the glutamine AAS (7-10). Chemical modification of the *E. coli* $tRNA^{Arg}$ or yeast $tRNA^{Phe}$ anticodons reduces cognate aminoacylation *in vitro* (11-14). It was possible, then, that the amber suppressor alleles that we had constructed would have altered specificity, i.e., the recognition of these tRNAs by their cognate AAS would be affected by the change introduced into the anticodon. Determining the amino acid specificity of the newly constructed suppressor tRNAs would reveal the extent to which the anticodon plays a role in AAS recognition for each of these tRNAs. Previously, we reported the development of a specificity assay in which a tRNA was used to suppress an amber mutation in the tenth residue of the *E. coli* dihydrofolate reductase (DHFR) gene *fol*. The mutant protein was subsequently purified and partly sequenced to determine which amino acid the suppressor

tRNA had inserted at the amber codon (3, 5). In this paper we present the results of the specificity assays for the l4 new suppressor tRNAs.

METHODS

Bacterial Strains and Media. All strains were derived from *E. coli* XAC which is F$^-$ $\Delta(lacproB)_{xlll}$, *nalA, rif, arg*$E_{am,}$ *ara*. XACRF is XAC with *uar* and tn10 transduced into the chromosome. The episome (*lacI*$_{373}$*lacZ*$_{ull8am}$*pro*B$^+$) is designated "-1" as in XAC-1 and XACRF-1. The episome (*lacI*^{q}*'ZYAtn5*) is designated "9100.1" and is found in XAC/9100.1 and XACRF/9100.1. The episome Zmet3 is (*lacI lacZ*$_{am3}$) and is found in XAC/Zmet3. Minimal M9 media was prepared according to Miller (15) and supplemented with 100ug/ml ampicillin, and 30ug/ml chloramphenicol, both from Sigma. L broth was 10g/l Bactotryptone, 5g/l yeast extract, and 8g/l NaCl.

Plasmid Construction. Plasmid pDA3-12 (3, 5) contains the coding region of the *E. coli* DHFR gene *fol* with an amber mutation at residue ten (*fol*^{am10}), under control of the *tac* promoter, cloned into the Eco RI-Hind III sites of pBR322. pDA3-12:Arg was constructed by subcloning the Pvu II fragment of pGFIB:Arg (1), which carries the tRNA$^{Arg}_{CUA}$ gene expressed from the *lpp* promoter, and followed by the *rrn*C transcription terminator, into the Pvu II site of plasmid pDA3-12 (Figure 1a). pDA3-12:AspM, pDA3-12:Ile2, pDA3-l2:Thr2, and pDA3-12:Val were constructed in a similar fashion. To construct plasmid pDA5YC, a Hind III-Eco RI adaptor was ligated to the Eco RI cohesive end of the Eco RI-Bam HI fragment of pDA3-12, and subcloned into the Hind III-Bam HI sites of plasmid pACYC184 (Figure 1b). Plasmid pDAYQ was constructed by attaching Sal I linkers to the Eco RI fragment from plasmid pJWlacI (16), which carries *lacI*^{q}, and ligating the entire fragment into the Sal I site of pDA5YC (Figure 1b).

All of the suppressor tRNA genes were subcloned into wild type lambda in the following manner. Xba I or Sal I linkers were ligated to the Pvu II fragment of each of the pGIB:suptRNAs (1, 2), which carries the tRNA gene, along with the *lpp* promoter and a transcription terminator. The linkered fragment was then ligated into the single Xba I site of wild type lambda (Figure 1c). Cloning techniques were as described by Davis *et al.* (17). Phage constructs containing only a single copy of the suppressor tRNA gene were used to lysogenize *E. coli* strains.

Purification of DHFR. XAC-1 or XACRF-1 carrying both pGFIB:sup and pDAYQ; XAC/9100.1 or XACRF/9100.1 carrying pGFIB:sup and pDA5YC; XAC/9100.1 or XACRF/9100.1 carrying pDA3-12:suptRNA alone, were all cultured in minimal M9 media supplemented with ampicillin (chloramphenicol was added in the presence of plasmids pDA5YC or pDAYQ) to an A_{600} of 0.5 to 0.7. The cell cultures were diluted 100 fold into fresh minimal media, grown to an A_{600} of 0.4 to 0.6, at which time the *fol* gene was induced with isopropyl thiogalactoside (IPTG) at a final concentration of lmM, for 4 to 18 hr. Alternatively, XAC/9100.1 or XACRF/9100.1, which had been lysogenized with the lambda suppressor constructs and transformed with plasmid pDA3-12, were cultured and induced as above except that L broth supplemented with ampicillin was used instead of minimal medium. The cells were harvested, and DHFR was purified as described by Normanly *et al.* (3, 5) with only one notable change. A MonoQ column (Pharmacia) was used interchangeably with a DEAE-sephacel column (Pharmacia) in the final stage of the purification. Purified DHFR was subjected to N-terminal sequence analysis at the USC Microchemical Core Facility or the Caltech Microchemical Facility.

Purification of β-galactosidase. Srain XAC/Zmet3 carrying pGFIB:Val was cultured in minimal media as described for the purification of DHFR except that the IPTG induction was eliminated. Cells were harvested and frozen at -20°C. Subsequent steps were carried out on ice with stirring. The cell pellet was thawed in 2.14 ml/g cells TEP (l00mM Tris HCl, pH 7.4, l0mM EDTA, lmM phenylmethylsulfonyl fluoride, PMSF). The following additions were made at 30 min. intervals: (1) 0.15 ml/g cells of lysozyme (17mg/ml in TEP); (2) 0.055 ml/g cells of lM $MgCl_2$ and 0.015 ml/g cells of DNaseI (13.2 mg/ml in TEP). The suspension was centrifuged at 12,000 x g for l5 min. at 4°C. The supernatant was saved and the pellet resuspended in 2.14 ml/g cells TEP. Steps l and 2 were repeated and the suspension centrifuged as above. The pellet was washed with 0.92ml/g cells TEP and centrifuged again. All three supernatants were pooled, brought to 40% saturation with ammonium sulfate and stirred on ice for 2 hr. The suspension was centrifuged at 19,000 x g for 30 min. at 4°C. The precipitate was resuspended in lml of TEP, and brought to a concentration of 4 mg/ml with TBSN (l0mM Tris HCl, pH 8.0, l50mM NaCl, 0.2% NP-40). The suspension was equilibrated with lml of Promega Proto Sorb Lac Z column matrix, batch washed with TBSN, and applied to a column. β-galactosidase was eluted from the column with 100mM Na-Carbonate buffer, pH

10.8, according to the manufacturers instructions. Purified beta-galactosidase was dialyzed against distilled water, lyophilized and sequenced at the USC Microchemical Core Facility.

RESULTS

In order to determine the specificity of the amber suppressor tRNAs, we constructed strains which carried both the suppressor tRNA gene to be assayed and *fol*^{am10}. We used three different methods to create these strains. In one case the suppressor tRNA gene and *fol*^{am10} were carried on the same plasmid. In another case, *fol*^{am10} resided on a plasmid, while the suppressor tRNA gene was integrated into the chromosome on a lambda prophage. Last, both the suppressor tRNA gene and *fol*^{am10} were carried on separate, compatible plasmids. We used the *E. coli* strains, XAC and XACRF, which differ in two genetic markers. XACRF was derived from XAC by transduction of *uar*^{ts} (linked to tn10), into the chromosome. The *uar*^{ts} allele is a release factor mutant. Its effect is to enhance suppression efficiency, because the release factor competes poorly with the suppressor tRNA for the nonsense codon. This strain was employed for those suppressor tRNAs with lower efficiencies.

Single Plasmid System. Previously, we described the plasmid pDA3-12, which carries *fol*^{am10} under control of the *tac* promoter (3, 5). The suppressor alleles corresponding to tRNAs Ile2, Val, AspM, Thr2, and Arg (1, 2) were subcloned into the pDA3-12 vector (Figure 1a).

Lambda Prophage and Plasmid Encoded *fol* Amber. We have found the plasmids encoding suppressor tRNAs to be somewhat unstable unless suppression is required for growth. The instability is probably due to the high copy number of the vector (pGFIB-1) and the fact that the suppressor genes are expressed constitutively (1, 4). The *E. coli* strain XAC has an amber mutation in the *ArgE* operon, which will accept a variety of amino acids. Consequently, if the suppressor-carrying strains are maintained on minimal media, the plasmids are stable. Our long term goal is to compile a collection of suppressor tRNA genes, which can be used in any strain background. We speculated that if the suppressor genes were present in lower copy number they would be more stable and would not require the *ArgE* background. Additionally, we wished to determine the effect of lowered copy number on the suppression efficiency of the tRNAs. We cloned 13 of the suppressor tRNAs genes, expressed from a constitutive promoter, *lpp*, (4) and followed by

a transcription terminator, *rrnC* (3, 5), into the single Xba I site of wild type lambda. These phage constructs were integrated into the chromosome of *E. coli* strain XAC and various derivatives, and the suppressor tRNAs were assayed for efficiency (data not shown) using the $LacI_{am}$-*Z* system of Miller (18). It was clear that the copy number of these tRNA genes played an important role in their suppression efficiencies. Of the 13 suppressor-carrying prophage, only those encoding $tRNA^{GlyT}_{CUA}$, $tRNA^{HisA}_{CUA}$ and $tRNA^{Ala2}_{CUA}$ had suppression efficiencies high enough to render the tRNAs useful in amino acid substitution studies (data not shown). These strains were transformed with plasmid pDA3-12, which encodes fol^{am10}, for the specificity assay (Figure 1b).

Compatible Plasmid System. We also developed a system wherein the suppressor tRNA gene resided in pGFIB-1, and fol^{am10} was carried on a compatible plasmid. We subcloned the fol^{am10} gene into the Hind III-Bam HI site of pACYC184 to yield pDA5YC (Figure 1c). Because fol^{am10} is under *tac* promoter control, it is necessary to maintain the plasmid in an $lacI^q$ strain (19). For added flexibility, we subcloned the $lacI^q$ gene into plasmid pDA5YC to yield plasmid pDAYQ (Figure 1c).

Suppressors for tRNAs Ile1, Ile2, Arg, AspM, Val, Thr2, ProH, Met(m), Lys, GluA and GlyU were transformed into either strains (XAC-1 or XACRF-1) carrying plasmid pDAYQ or strains (XAC/9100.1 or XACRF/9100.1) carrying plasmid pDA5YC. The cells were cultured in minimal media, and after induction of the *fol* gene with IPTG, DHFR was purified from these cells by affinity chromatography. N-terminal sequence analysis of the purified proteins revealed which amino acid(s) had been inserted in response to the amber codon by each suppressor tRNA.

Protein Sequence Data. Figure 2 shows the results from the protein sequencing. $tRNA^{Ala2}_{CUA}$, $tRNA^{HisA}_{CUA}$, $tRNA^{Lys}_{CUA}$, and $tRNA^{ProH}_{CUA}$ all inserted the predicted amino acid exclusively (Figure 2a). Interestingly, two different isoacceptors of $tRNA^{Gly}$ behaved differently when their anticodons were altered to CUA. $tRNA^{GlyU}_{CUA}$ inserted exclusively glycine, while $tRNA^{GlyT}_{CUA}$ inserted a mixture of glycine (62%) and glutamine (37%) (Figure 2b). Similar results for the GlyT suppressor have been obtained elsewhere (20). The glutamic acid suppressor $tRNA^{GluA}_{CUA}$ inserted mostly glutamic acid (59%) but was mischarged by the glutamine AAS as well (17%). One isoacceptor of $tRNA^{Ile}$, $tRNA^{Ile1}_{CUA}$, was mischarged by the glutamine AAS (74%) and to some extent by the lysine AAS (16%), while another isoacceptor of Ile, $tRNA^{Ile2}_{CUA}$, was mischarged

exclusively by the lysine AAS (Figure 2b). Five other suppressor alleles, $tRNA^{Arg}_{CUA}$, $tRNA^{AspM}_{CUA}$, $tRNA^{Met(m)}_{CUA}$, $tRNA^{Thr2}_{CUA}$, and $tRNA^{Val}_{CUA}$ were mischarged by the lysine AAS (Figure 2c). In the cases of $tRNA^{Arg}_{CUA}$, $tRNA^{Thr2}_{CUA}$, and $tRNA^{Val}_{CUA}$, some of the predicted amino acid was inserted as well, but not more than 10%. The average background level for any amino acid residue is approximately 3%. Amino acid levels of 5% and above were considered significant. The percent amino acid incorporated into protein may not add up to 100%, because background levels are not included. For example, while proline is the major amino acid inserted by $tRNA^{ProH}_{CUA}$, and no more than 3% of any other amino acid was found, the value given for percent proline incorporated into protein is only 85%.

As stated in the introduction, there is precedent for mischarging by the glutamine AAS when the anticodon is altered; however, mischarging of a tRNA by the lysine AAS was completely unexpected. Since we began this work others have observed mischarging by the lysine AAS (21). The existence of a second lysine AAS in *E. coli*, which is a heat shock protein, may explain the lysine mischarging (22). It occurred to us that the overproduction of DHFR might induce a heat shock response and cause mischarging by the lysine AAS. To rule out this possibility, we decided to assay the specificity of $tRNA^{Val}_{CUA}$ using another protein. We chose a mutant of the *LacZ* gene, with an amber mutation residing in the third residue. This gene is encoded episomally, and β-galactosidase is not overproduced to the extent that DHFR is. pGFIB:Val was transformed into XAC/Zmet3, from which β-galactosidase was purified. As for the *fol* amber, lysine was found to be the predominant amino acid inserted, with no more than 5% valine shown to be incorporated (data not shown).

DISCUSSION

We have generated amber functional suppressor alleles for 14 species of tRNA (Table 1), and determined their amino acid specificities. Tables 2 and 3 summarize the results. Table 2 gives the specificities of amber suppressor alleles generated in this work as well as from previous work by us and others. With the exception of $tRNA^{Asn}$, functional supressor alleles exist for all tRNA species (and in some cases multiple isoacceptors). The specificities of these suppressors are known *in vivo* for all but $tRNA^{Met(f)}_{CUA}$, which has been examined *in vitro* (10). The specificity assays conducted upon these tRNAs reveal that they may be classified in one of three ways (Table 3). For Class I, altering the

anticodon to CUA does not affect cognate AAS recognition *in vivo*. For the other two classes of tRNA, such an alteration results in mischarging by either the glutamine (Class II) or lysine (Class III) AAS. Interestingly, different isoacceptors of a given tRNA species may fall into separate classes, as in the case of $tRNA^{Ile1}_{CUA}$ and $tRNA^{Ile2}_{CUA}$, as well as $tRNA^{GlyU}_{CUA}$ and $tRNA^{GlyT}_{CUA}$.

Our goal to generate a complete bank of amber suppressor tRNA genes has only been partially achieved. The suppressors that fall into class I will be useful in amino acid substitution studies, and to this end the existing collection of useful amber suppressors has been increased from 4 to 11. (A suppressor allele of $tRNA^{GlyU}_{CUA}$ had been generated previously by genetic means but the exact specificity was unknown; 23). The tRNAs in class II and III remain an obstacle. It is not clear what distinguishes one class of tRNA from another; sequence comparison among the three classes does not reveal any uniquely conserved nucleotides. A fortuitous mutation in the $tRNA^{Arg}_{CUA}$ suppressor has resulted in a variant that inserts exclusively arginine (24), although it is not immediately obvious why this is the case. Clearly, if we are to succeed in our goal of generating a complete collection of amber suppressor tRNA genes, each inserting the predicted amino acid, we must define those elements of class II and III tRNAs that dictate cognate AAS recognition.

REFERENCES

1. Masson, J.-M., Kleina, L.G., Normanly, J.H., Miller, J.H., and Abelson, J. 1988 manuscript in preparation.
2. Kleina, L.G., Normanly, J., Masson, J.-M., Miller, J.H., and Abelson, J. 1988. manuscript in preparation.
3. Normanly, J., Masson, J-M., Kleina, L.G., Abelson, J., and Miller, J.H. 1986. *Proc. Nat. Acad. Sci.* 83:6548-6552
4. Masson, J.-M., and Miller, J.H. 1986. *Gene*,47:179-183
5. Normanly, J., Ogden, R.C., Horvath, S.J., and Abelson, J. (1986). *Nature* , 321:213-21
6. Yarus, M. 1982. *Science*, 218:646-652
7. Celis, J.E., Coulondre, C., and Miller, J.H. 1976. *J. Mol. Biol.* 104:729-734
8. Yaniv, M., Folk, W.R., Berg, P., and Soll, L. 1974. *J. Mol. Biol.* 86:245-268
9. Yarus, M., Knowlton, R., Soll, L. 1977. in *Nucleic acid-protein recognition*, pp. 391-408 New York, Academic Press
10. Schulman, L.H., and Pelka, H. 1985. *Biochemistry,* 24:7309-7314

11. Chakraburrty, K. 1975. *Nucleic Acids Res.* 2:1793-1804
12. Bruce, A. G., and Uhlenbeck, O.C. 1982. *Biochemistry*, 21:855-861
13. Bruce, A.G., and Uhlenbeck, O.C. 1982. *Biochemistry*,21:3921-3927
14. Kisseliv, L.L. 1985. in *Progress in Nucleic Acids Research and Molecular Biology* (Cohn, W.E. Moldave, K. eds.) Academic Press, Florida, v.32,237-266
15. Miller, J.H. 1972. In *Experiments in Molecular Genetics* , Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
16. Zumstein, L., and Wang, J.C. 1986. *J. Mol. Biol.* 191:333-40
17. Davis, R.W., Botstein, D., and Roth, J.R. 1982. In *A Manual for Genetic Engineering: Advanced Bacterial Genetics.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
18. Miller, J. H., and Albertini, A. M. 1983. *J. Mol. Biol.* 164:59-71
19. Amann, E., Brosius, J., and Ptashne, M. 1983. *Gene*, 25:167-178
20. McClain, W.H., and Foss, K. 1988. *Science* 240:793-796
21. McClain, W.H., and Foss, K. 1988. *J. Mol. Biol.* 202:697-709
22. Hirshfield, I.N., Bolch, P.L., BanBogelen, R.A., and Neihardt, F.C. 1981. *J. Bacteriol.*146:345-351
23. Murgola, E. J. 1985. *Ann. Rev. Genetics.* 19:57-80
24. McClain, W.H., Foss, K. 1988. *Science* in press.
25. Normanly, J., Abelson, J. manuscript in preparation.
26. Hewick, R.M., Hunkapiller, M.W., Hood, L.E., and Dreyer, W.J. 1981. *J. Biol. Chem.* 256:7990-7997

Table 1. Suppression efficiencies of amber suppressor tRNAs.

Suppressor	Suppression efficiency*
SET 1	
Ala1	<1%
Asp	<1%
Asn	<1%
Arg	11-55%
Cys	17-54%
Glu	1-44%
GlyT	44-67%
GlyU	24-100%
His	1-38%
Ile1	1-10%
Lys	9-29%
Met(m)	5-45%
Phe	48-100%
Pro	<1%
Thr1	<1%
Val1	<1%
SET 2	
Ala2	8-83%
AspM	10%
GluA	8-100%
HisA	16-100%
Ile2	6-67%
ProH	8-60%
Thr2	1-14%
Val2	18-83%

*Suppression efficiencies are a direct measurement of β-galactosidase activity (% of wild type) resulting from suppression of amber mutations in *lacI-Z* fusions (15).

Table 2. Compilation of amber suppressor tRNA specificities.

Suppressor	Specificity (given as percent amino acid inserted at amber codon)		Reference
Ala2	Ala	97%	this work
Lys	Lys	94%	"
ProH	Pro	85%	"
HisA	His	94%	"
GlyT	Gly,Gln	62%,37%	"
GlyU	Gly	90%	"
Ile1	Gln,Lys	74%,16%	"
Met(m)	Lys	97%	"
Ile2	Lys	93%	"
AspM	Lys	81%	"
Arg	Lys,Arg	91%,5%	"
Thr2	Lys,Thr	86%,8%	"
Val	Lys,Val	84%,5%	"
GluA	Glu,Gln	59%,17%	"
	Tyr,Arg	6%,6%	"
Phe	Phe	98%	3
Cys	Cys		3
Leu5, and su^+6	Leu	99%	5
Ser(su^+1)	Ser	92%	5
Gln(su^+2)	Gln†	93%	5
Tyr(su^+3)	Tyr	95%	25
Trp(su^+7)	Gln,Trp		7
Met(f)	Gln*		10

*Demonstrated in vitro only. †Includes Glu as well.

Table 3. tRNAs classified by the effect of a CUA anticodon upon AAS recognition.

Class I No affect upon charging	Class II Glutamine AAS mischarging	Class III Lysine AAS mischarging
Ala2	Ile1	Ile2
GlyU	GlyT	Arg
Cys	Met(f)	Met(m)
Phe	GluA	AspM
ProH	Trp (su^+7)	Thr2
HisA		Val
Lys		
Ser (su^+1)		
Gln (su^+2)		
Tyr (su^+3)		
Leu (su^+6;also synthetic Leu5)		

FIGURE LEGENDS

Figure 1a. The Pvu II fragment encoding the amber suppressor tRNA genes, flanked by the *lpp* promoter (4) and *rrnC* terminator (3), were subcloned into the Pvu II site of plasmid pDA3-12 (3). "suptRNA" in pDA3-12:suptRNA is generic, representing each of the suppressor tRNA genes that have been subcloned into the plasmid. For example pDA3-12 carrying the Arg suppressor is actually referred to as pDA3-12:Arg.

Figure 1b. Xba I or Sac I linkers were ligated to the Pvu II fragment shown in Figure 1a, and subcloned into the single Xba I site of wild type lambda DNA. The lambda constructs were used to lysogenize *E. coli* strains, which were then transformed with plasmid pDA3-12.

Figure 1c. pGFIB:suptRNA is derived from plasmid pGFIB-1 (3). Again "suptRNA" is a generic term. Plasmid pDA5YC encodes the *fol* amber and a chloramphenicol resistance marker. Plasmid pDAYQ differs from pDA5YC in that the *lacI*^{q} gene is inserted into the Sal I site of pDA5YC. The *lacI*^{q} gene overproduces the *lac* repressor, which is necessary for the regulation of the *fol* gene's *tac* promoter. With the *lacI*^{q} gene encoded by the same plasmid that carries the *fol* amber, it is not necessary to use an *E. coli* strain that carries the *lacI*^{q} allele chromosomally.

Figure 2. Protein sequence analysis of DHFR substituted at position ten by the newly generated suppressor tRNAs. Sequence analysis was carried out on an Applied Biosystems model 470A gas phase sequenator (26) by the USC Microchemical Core Laboratory. The picomolar yields of phenylthiohydantoin (PTH) derivatives of various amino acids are plotted against the residue number. The N-terminal sequence (up to 15 residues) of the DHFR mutant is indicated below each plot, and the percent yield of the amino acid at residue ten is indicated to the right of the plot. Only the most significant yields are presented; amino acids not shown can be presumed to be present at background levels. Protein sequence results from DHFR substituted at position ten with a) $tRNA^{Ala2}_{CUA}$, $tRNA^{Lys}_{CUA}$, $tRNA^{ProH}_{CUA}$, and $tRNA^{HisA}_{CUA}$ b) $tRNA^{GlyT}_{CUA}$, $tRNA^{GlyU}_{CUA}$, $tRNA^{GluA}_{CUA}$, and $tRNA^{Ile1}_{CUA}$. *Data not shown. c) $tRNA^{Ile2}_{CUA}$, $tRNA^{Met(m)}_{CUA}$, $tRNA^{Arg}_{CUA}$, $tRNA^{AspM}_{CUA}$, $tRNA^{Thr2}_{CUA}$, and $tRNA^{Val}_{CUA}$.

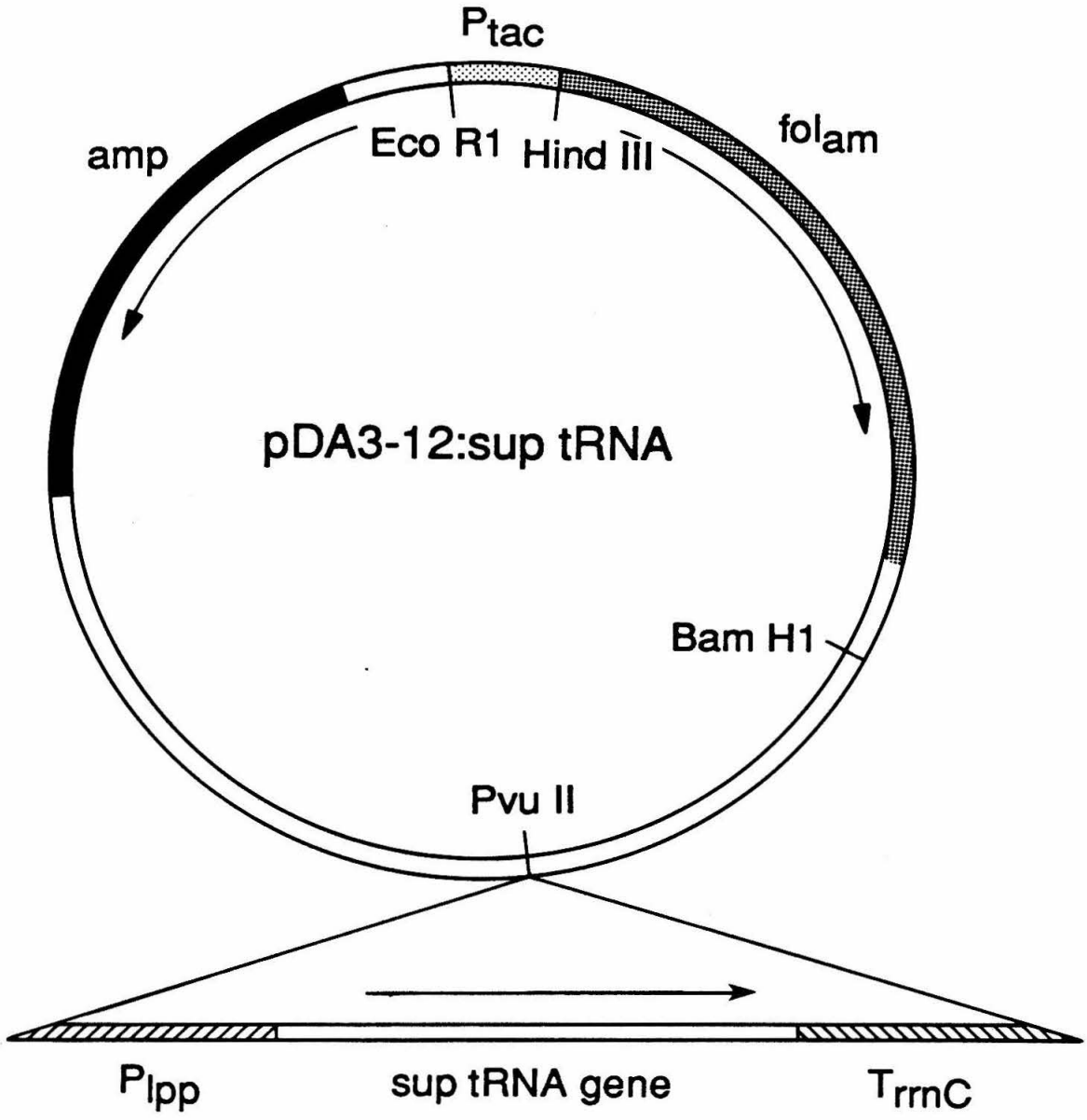

Ptac
amp
Eco R1
Hind III
folam
pDA3-12:sup tRNA
Bam H1
Pvu II
Plpp
sup tRNA gene
TrrnC

Figure 1a

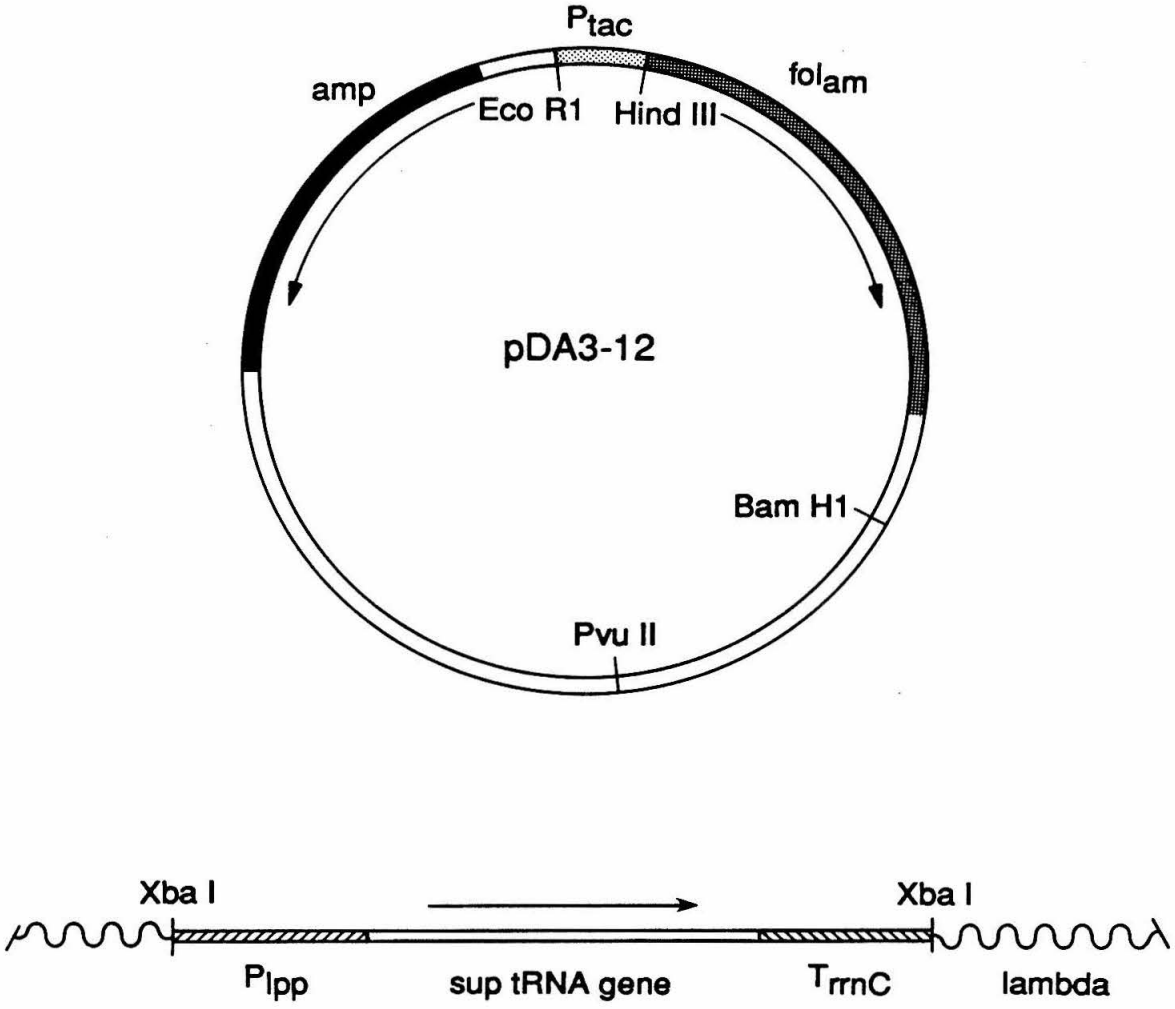

Ptac
amp
Eco R1
Hind III
folam
pDA3-12
Bam H1
Pvu II
Xba I
Xba I
Plpp
sup tRNA gene
TrrnC
lambda

Figure 1b

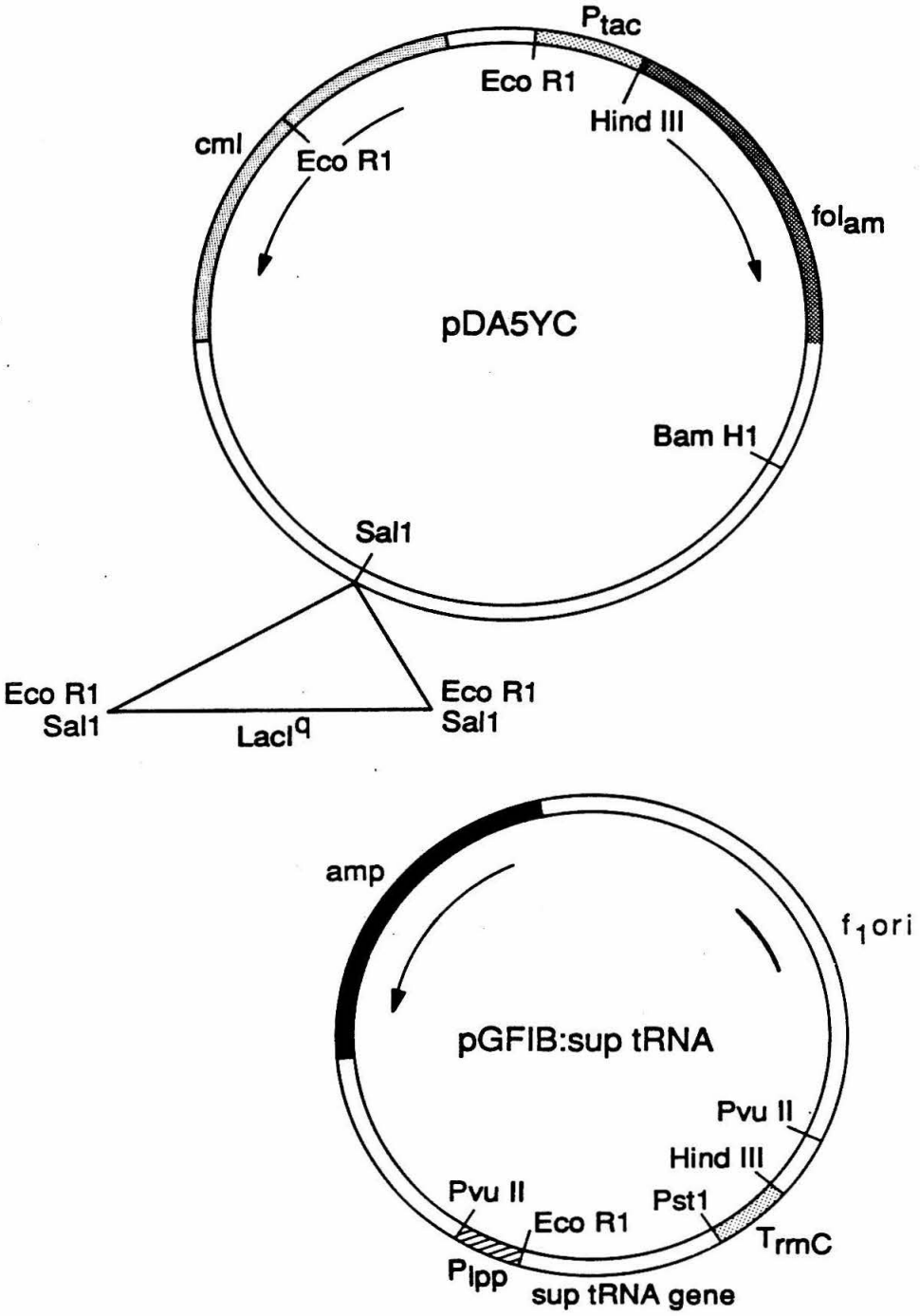

Ptac
Eco R1
Hind III
cml
Eco R1
folam
pDA5YC
Bam H1
Sal1
Eco R1
Sal1
LacIq
Eco R1
Sal1
amp
f1ori
pGFIB:sup tRNA
Pvu II
Hind III
Pvu II
Pst1
Eco R1
TrrnC
Plpp
sup tRNA gene

Figure 1c

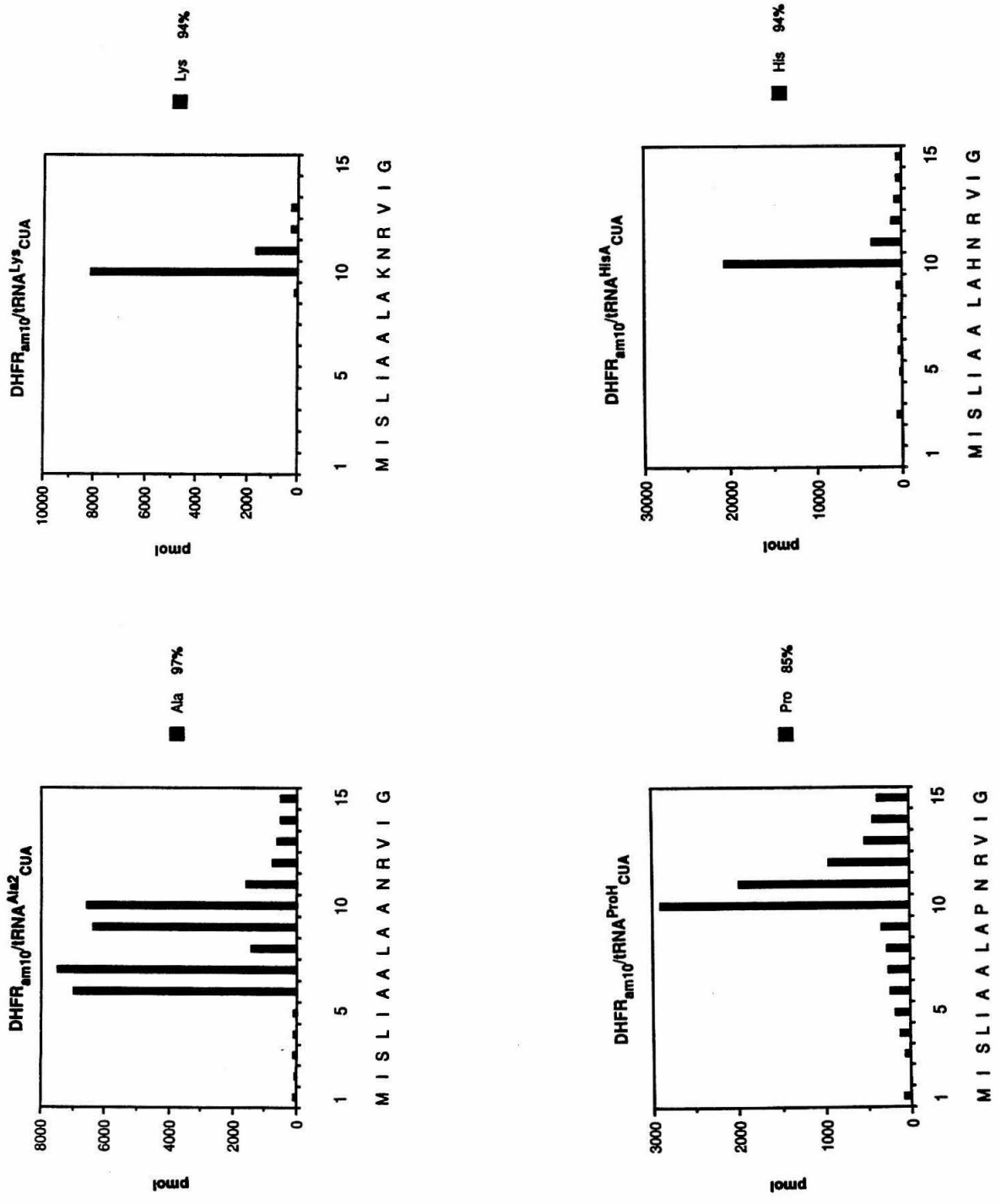

DHFR$_{am10}$/tRNA$^{Lys}_{CUA}$
Lys 94%
pmol
M I S L I A A L A K N R V I G
DHFR$_{am10}$/tRNA$^{HisA}_{CUA}$
His 94%
pmol
M I S L I A A L A H N R V I G
DHFR$_{am10}$/tRNA$^{Ala2}_{CUA}$
Ala 97%
pmol
M I S L I A A L A A N R V I G
DHFR$_{am10}$/tRNA$^{ProH}_{CUA}$
Pro 85%
pmol
M I S L I A A L A P N R V I G

Figure 2a

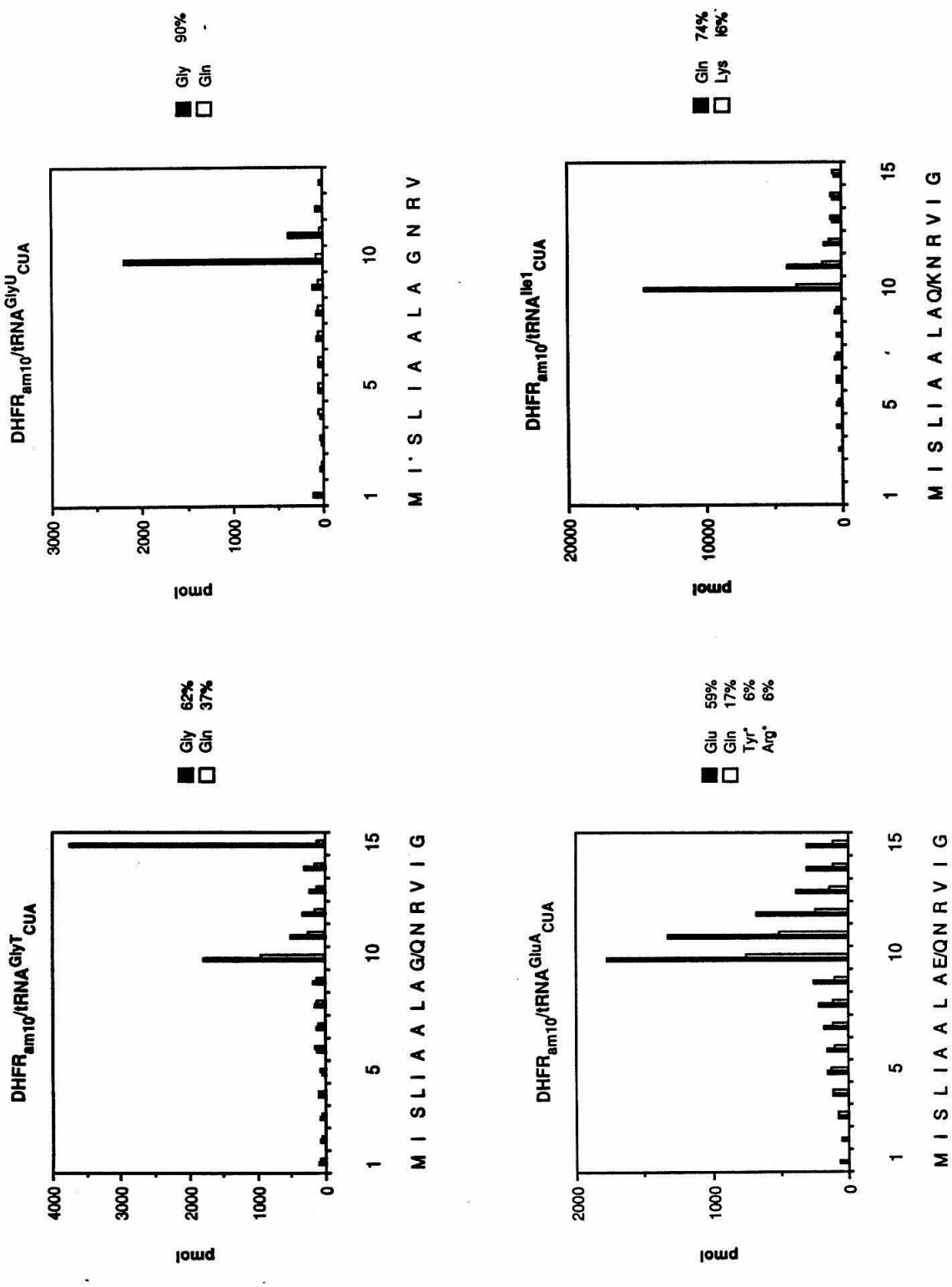

DHFR$_{am10}$/tRNA$^{GlyU}_{CUA}$
Gly 90%
Gln -
pmol
3000
2000
1000
0
1 5 10
M I S L I A A L A G N R V
DHFR$_{am10}$/tRNA$^{Ile1}_{CUA}$
Gln 74%
Lys 16%
pmol
20000
10000
0
1 5 10 15
M I S L I A A L A Q/K N R V I G
DHFR$_{am10}$/tRNA$^{GlyT}_{CUA}$
Gly 62%
Gln 37%
pmol
4000
3000
2000
1000
0
1 5 10 15
M I S L I A A L A G/Q N R V I G
DHFR$_{am10}$/tRNA$^{GluA}_{CUA}$
Glu 59%
Gln 17%
Tyr* 6%
Arg* 6%
pmol
2000
1000
0
1 5 10 15
M I S L I A A L A E/Q N R V I G

Figure 2b

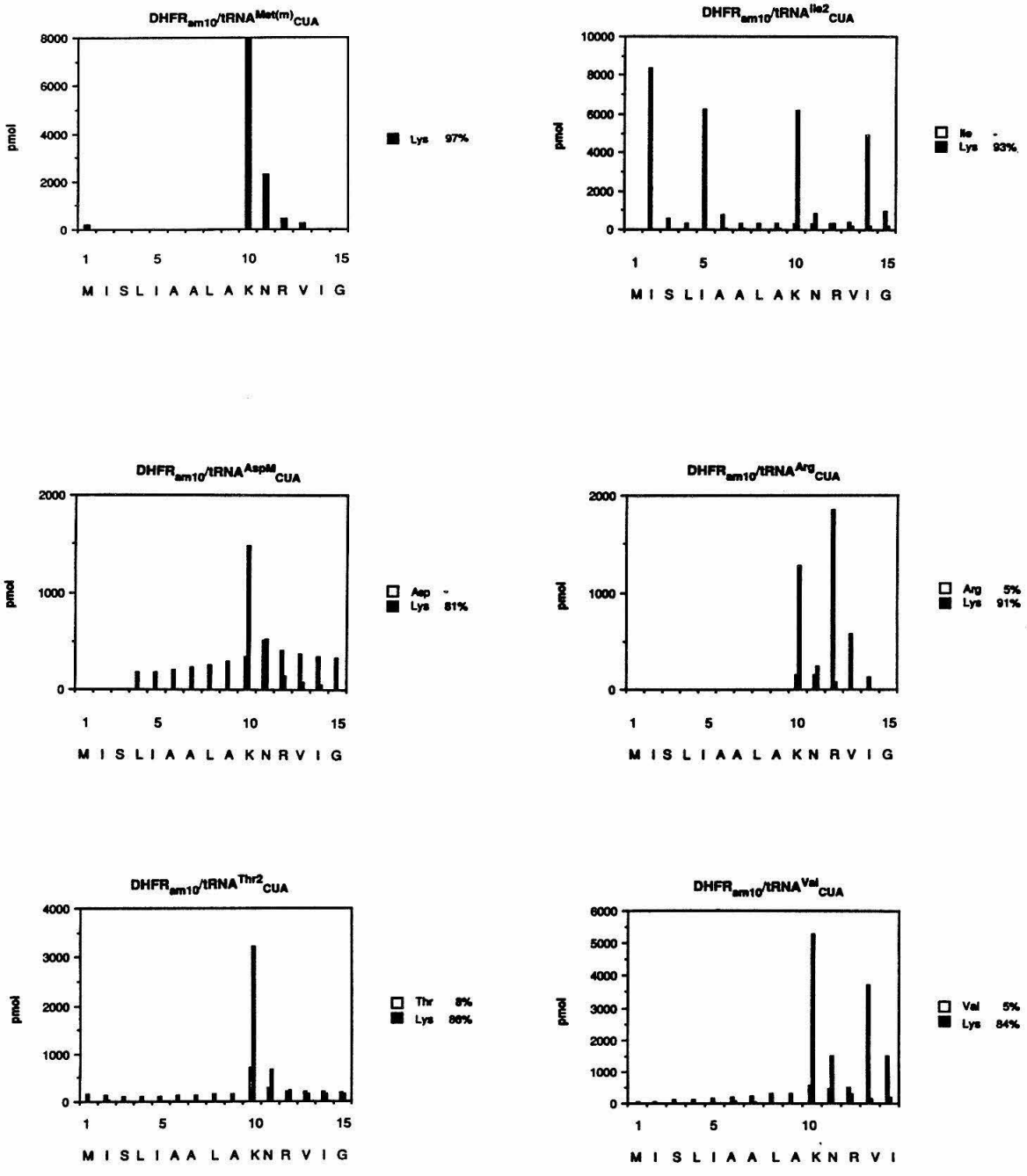

DHFR$_{am10}$/tRNA$^{Met(m)}_{CUA}$
pmol
8000
6000
4000
2000
0
1 5 10 15
M I S L I A A L A K N R V I G
Lys 97%
DHFR$_{am10}$/tRNA$^{Ile2}_{CUA}$
pmol
10000
8000
6000
4000
2000
0
1 5 10 15
M I S L I A A L A K N R V I G
Ile -
Lys 93%
DHFR$_{am10}$/tRNA$^{AspM}_{CUA}$
pmol
2000
1000
0
1 5 10 15
M I S L I A A L A K N R V I G
Asp -
Lys 81%
DHFR$_{am10}$/tRNA$^{Arg}_{CUA}$
pmol
2000
1000
0
1 5 10 15
M I S L I A A L A K N R V I G
Arg 5%
Lys 91%
DHFR$_{am10}$/tRNA$^{Thr2}_{CUA}$
pmol
4000
3000
2000
1000
0
1 5 10 15
M I S L I A A L A K N R V I G
Thr 8%
Lys 86%
DHFR$_{am10}$/tRNA$^{Val}_{CUA}$
pmol
6000
5000
4000
3000
2000
1000
0
1 5 10
M I S L I A A L A K N R V I
Val 5%
Lys 84%

Figure 2c