covered easily from the hydrolysate by adsorption on cellulosic material, which can be blended with the solids added at the hydrolysis stage. Enzymes on the solid residue left after hydrolysis can be desorbed by *p*H adjustment (*9*) and the extract used to supplement the fresh enzyme. Beta-glucosidase, a key component of cellulase which splits cellobiose to glucose, is not very stable and cannot be recovered by either of the methods just presented. However, low-cost beta-glucosidase can be obtained by a separate fermentation. There are indications that about 50 percent of the cellulase can be recycled, and the effect on costs should be dramatic.

## Conclusion

Biomass refining is approaching commercialization. The first factories are likely to emphasize high-value products such as lignin, molasses, and paper pulp rather than ethanol. With products from all the major fractions of biomass very little waste will be generated, so these factories will have few problems in meeting environmental standards. Establishment of refining technologies will open the way for factories producing alcohol fuels. An excellent return on investment is likely, and guarantees and tax incentives may not be needed to stimulate the construction of factories.

### References and Notes


1. Periodic reports on projects supported by the Department of Energy have appeared in the *Alcohol Fuels Program Technical Review*, available by contacting L. Douglas, Solar Energy Research Institute, Golden, Colo. 80401. Reviews can be found in H. R. Bungay, *Energy, the Biomass Options* (Wiley, New York, 1981); *Adv. Biochem. Eng.* **20**, 1 (1981); and in H. W. Blanch and C. R. Wilke, *Rev. Chem. Eng.* **1**, 71 (1982).
2. G. C. Avgerinos and D. I. C. Wang, *Annu. Rep. Ferment. Process* **4**, 165 (1980).
3. E. A. Johnson *et al.*, *Appl. Environ. Microbiol.* **43**, 1125 (1982).
4. *Alcohol Fuels Program Technical Review* (Winter 1981), p. 43.
5. *Ibid.*, p. 5.
6. *Ibid.*, p. 51.
7. J. Krieger, *Chem. Eng. News* **60**, 38 (31 May 1982).
8. *Alcohol Fuels Program Technical Review* (Winter 1981), p. 35.
9. A. P. Sinitsyn *et al.*, *Appl. Biochem. Biotechnol.*, in press.


# Translational Efficiency of Transfer RNA's: Uses of an Extended Anticodon


Michael Yarus



A transfer RNA (tRNA) efficiently translates its codon (or codons) despite the apparent fact that its principal device for doing so, the anticodon trinucleotide, seems to function weakly and inaccurately when not a part of a tRNA molecule. We therefore expect that the detailed architecture of tRNA's will account for their superiority. But while some of the important contributions of the tRNA structure are known (*1*, *2*), there has been no unifying scheme that connects tRNA structure with translational performance. The extended anticodon hypothesis is such a scheme. It suggests that the structure of the anticodon loop and the proximal anticodon stem are related to the sequence of the anticodon. Thus, it is as if the anticodon itself were extended into the nearby structure. This extension can therefore be used to specifically enhance the selected, accurate pairing with a cognate codon.

*Summary.* Transfer RNA's are probably very strongly selected for translational efficiency. In this article, the argument is presented that the coding performance of the triplet anticodon is enhanced by selection of a matching anticodon loop and stem sequence. The anticodon plus these nearby sequence features (the extended anticodon) therefore contains more coding information than the anticodon alone and can perform more efficiently and accurately at the ribosome. This idea successfully accounts for the relative efficiencies of many transfer RNA's.



The author is a professor of biology in the Department of Molecular, Cellular and Developmental Biology, Campus Box 347, University of Colorado, Boulder 80309.


The tRNA's (as well as the rest of the translational apparatus) of free-living, fast-growing microorganisms are probably highly selected for translational efficiency. More than half the dry weight of a bacterium can be protein, and the rate at which that mass can be generated is quite critical to evolutionary success. Even a tiny difference in growth rate can quickly be amplified by a period of exponential growth into a decisive advantage in numbers. It is therefore plausible to begin by treating the tRNA's functional in *Escherichia coli* as a set selected for good (if not optimal) translational efficiency.

## Order in Natural Sequences

In Fig. 1, the sequences of 42 wild-type tRNA's functional in *E. coli* are abstracted and grouped so as to allow comparison of their anticodon loop and stem regions. The sequences are presented linearly, beginning at the first (5′) anticodon loop nucleotide, running 3′ through the anticodon loop, and extending through the five 3′ nucleotides of the anticodon helix (see the drawing in Table 1). Thus, the figure allows one to visualize the entire loop, and the helix by inference.

The sequences are grouped so that translators with the same 3′ anticodon nucleotide (same 5′ or "first" codon nucleotide) are written together; 3′ pyrimidines are at the top of the figure, 3′ U or A is on the left, 3′ C or G is on the right of the groupings (*3*). This helps to make the design of the sequences most evident, because all sequences having the same 3′ anticodon nucleotide also have a similar set of anticodon region sequence features. The constrained features include the sequence of the 3′ nucleotides in the anticodon loop, and some of the base pairs of the anticodon stem (Table 1). A comparable set of shared features does not appear when sequences are grouped by first [wobble; (*4*)] or second (middle) anticodon positions.

These relationships are summarized



0036-8075/82/1112-0646$01.00/0

by a set of "rules" in Table 1, which may be compared with the tRNA sequences in Fig. 1. The rules of Table 1 say that, if a certain sort of nucleotide is at the 3′ anticodon position, then a proximal nucleotide or nucleotide pair will have a particular nature. A bar indicates that no strong preference related to coding exists. Because of the apparent primary role of the 3′ anticodon nucleotide, I refer to it as the *cardinal nucleotide* of the anticodon. I call the anticodon, together with its related nearby sequence features, the *extended anticodon*.

### Validity for Other Organisms

The uniformities of the sequences in Fig. 1, which are codified in Table 1, are quite rigorous for *E. coli* tRNA's. The data are also valid, with a few exceptions, in other organisms, including the cytoplasmic fraction of higher eukaryotes. However, these regularities are detectable, but not strong at all, in the tRNA's of archaebacteria and even less marked for eukaryotic organelles. The organelles may have alternative rules not yet discerned, or may not be selected primarily for translational efficiency.

There is an interesting point to be made about phylogenetic breadth from Fig. 1 itself. The T-bacteriophage tRNA's clearly show most of the same regularities as *E. coli* tRNA's, but they also contain a number of exceptions associated with an unusual concentration of nucleotide mispairs in the anticodon stem. These variations may reflect the fact that the T4 tRNA's are dispensable genes in laboratory strains of *E. coli*, and therefore may not be as strongly selected for efficiency (*5*, *6*) as are the tRNA's of the host cell.

It is possible to take a list of tRNA sequences and, without looking at the cardinal anticodon nucleotide, deduce it in the overwhelming majority of cases by inspection of the anticodon stem and 3-prime loop. This exercise shows that the extended anticodon contains sufficient information to aid translational accuracy. This is the more impressive, since not all constraint on the anticodon local sequence has been listed in Table 1, just the part that appears as the nearly invariant occurrence of certain nucleotides. Additional information exists as biases in nucleotide sequence less strong than those shown, as other restricted groups of alternative nucleotides (*7*), and some information may not have been detected because too few sequences are available.

### Relation Between Structure and Function

There is a notable contrast between the two halves of the anticodon loop, comparing the sequence 5′ to the anticodon to that 3′ to it. The two conserved pyrimidines on the 5′ side of the loop (Table 1) are stereotyped by comparison with the variable nucleotides of the extended anticodon. The 5′ nucleotide of the loop (Table 1) may be either C or U, although this is only weakly related to the anticodon. The next nucleotide is one of the most strongly conserved features of the tRNA sequence, the universal U which precedes the anticodon. Thus the sequences on the 3′ side of the loop and in the stem are strongly related to the anticodon, but the 5′ side of the loop is almost constant.

Although all the nucleotides proximal to the anticodon are probably important to translational efficiency, the cardinal nucleotide, the nucleotides just 3′ to it, and some of the stem nucleotide pairs are constrained as if they were a coding unit. The 5′ part of the loop, in contrast, is apparently shaped to perform some common function, whose nature is only weakly related to the coding nucleotides. The anticodon loop is therefore composed of two regions of rather different character, whose translational roles may be expected to be distinct.

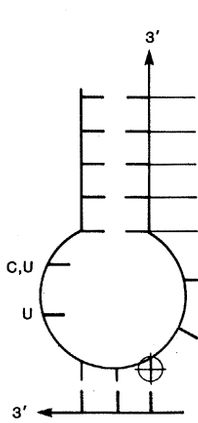
What is the origin and purpose of the nonrandomness of the anticodon loop region structure? The extended anticodon hypothesis proposes that the sequence regularities attendant to a given anticodon are evolved to optimize its translational function.

| | 5′ LOOP | A.C. | 3′ LOOP (I II) | 3′ HELIX (III IV V VI VII) |
|---|---|---|---|---|
| ASN | CU | QUU | t6 AA | ψCCGU̅ |
| ILE1 | CU | GAU | t6 AA | GGGUG |
| ILE2 | CU | NAU | t6 AA | ψCGCU |
| LYS | CU | mams UUU | t6 AA | ψCAAU |
| $MET_m$ | CU | ac4 CAU | t6 AA | ψGAUG |
| SER3 | s2 CU | GCU | t6 AA | GGGAG |
| THR | UU | GGU | mt6 AA | GGGUG |
| T4 ARG | CU | NCU | AA | GψU̅U̅G |
| T4 ILE | CU | NAU | * AA | GGGAA̅ |
| T4 THR | CU | * UGU | * AA | ψGAGG |
| T5 ASP | UU | GUU | AA | GψU̅GA |
| INITIATOR $MET_f$ | CU | CAU | AA | CCCGA |
| CYS | ψU | GCA | msi AA | ψCCGU |
| LEU5 | ψU | * AAA | msi AA | ψCCCU |
| PHE | ψU | GAA | msi AA | ψCCCC |
| SER1 | CU | o5 UGA | msi AA | ACCGG |
| TRP | CU m | CCA | msi AA | ACCGG |
| TYR | CU m | QUA | msi AA | ψCUGC |
| T4 LEU | CU | * UAA | msi AA | ψGCUG |
| T4 SER | CU m | NGA | msi AA | ACCGG |

| | 5′ LOOP | A.C. | 3′ LOOP (I II) | 3′ HELIX (III IV V VI VII) |
|---|---|---|---|---|
| ALA | UU | o5 UGC | AC | GCAGG |
| ASP | CU | QUC | m2 AC | GCAGG |
| GLU1 | CU | mams UUC | m2 AC | GGCGG |
| GLU2 | CU | mams UUC | m2 AC | GGCGG |
| GLY1 | UU | CCC | AA | GCUCU |
| GLY2 | CU | * UCC | AA | GCUGA |
| GLY3 | UU | * GCC | AA | GGUCG |
| VAL1 | CU | o5 UAC | m6 AA | GGAGG |
| VAL2 | UU | GAC | AU | GGUGG |
| T4 GLY | CU | mams * UCC | AA | ψCUGA |
| ARG1 | s2 CU | ICG | m2 AA | CCGAG |
| ARG2 | CU | ICG | m2 AA | CCGAG |
| GLN1 | UU m | s2 UUG | m2 Aψ | ACCGG |
| GLN2 | UU m | CUG | m2 Aψ | ψCCGG |
| HIS1 | UU | QUG | m2 Aψ | ψCCGG |
| LEU1 | UU | CAG | * Gψ | GψU̅AG |
| LEU2 | UU | GAG | * Gψ | GGUAG |
| T4 GLN | CU | * UUG | m2 AC | ψGCUA |
| T4 PRO | UU m | * UGG | m1 GA | ψCAGG |
| T5 GLN | CU | UUG | AA | GψU̅CGA |
| T5 HIS | UU | GUG | AA | ψCAGG |
| T5 LEU | CU | UAG | m6 AU | ψU̅CCGU |
| T5 PRO | UU | UGG | m1 GG | CGGUG |

Fig. 1. The sequences of *E. coli* tRNA's. Only the anticodon loops and the 3′ side of the anticodon stems are shown. Roman numerals correspond to the positions in Table 1. Abbreviations may be defined by, and references found in (*55*). A superscript bar on a helix nucleotide implies that nucleotide is not involved in an AU or GC pair.

Table 1. Sequence organization around the anticodon in tRNA's. Across the top is the nucleotide at the cardinal (3′) position of the anticodon, and, in the body of the table is the corresponding local sequence feature, as shown in the diagram of the anticodon stem and loop at the left. A dash implies that there is no evident strong constraint. Pu means the nucleotides A or G, Py means U or C, and not G means U, C, or A (*55*). Entries in parentheses appear in *E. coli* tRNA, but are not strongly supported by other sequences.




| | Cardinal Nucleotide A | U | G | C |
|---|---|---|---|---|
| VIII | — | — | PyPu | — |
| VI | (GC or CG) | — | PyPu | GC or CG |
| V | Pu Py | — | — | — |
| IV | GC | GC or CG | GC or CG | GC or CG |
| III | AΨ or UA | AΨ or CG | — | CG |
| II | A | A | not G (A or Ψ) | not G (A or C) |
| I | $ms^2i^6A$ | $t^6A$ | $m^2A,G^*$ | $A,m^2A,$ $m^6A$ |

In the case of one particular anticodon loop nucleotide, a similar idea and data confirming it have been known for some time. Nishimura (*1*) has reviewed the structural constraints on the nucleotide 3′ to the cardinal nucleotide (nucleotide I, Table 1). The nature of its unique posttranscriptional modifications, which change to match a mutational change in the cardinal nucleotide (*1, 8*), and the requirement for the modification for optimal function in translation, are well established (*1, 2, 9, 10*). The evidence concerning this nucleotide therefore supports both (i) the existence of an extended anticodon and (ii) its special relation to the cardinal nucleotide. This kind of specific structural requirement is applied to most of the nearby stem and loop by the extended anticodon hypothesis.

## Translational Efficiency

The idea of the extended anticodon has been implicit in the well-known fact that, for example, amber (UAG) suppressor tRNA's have different efficiencies. Because all amber suppressor tRNA's share the same codon-anticodon pairing, the differences must reside in the rest of the tRNA structure. The extended anticodon hypothesis assigns this additional influence to a location in the molecule.

There are seven classes of amino acid acceptor tRNA's that ought to easily mutate to amber (UAG) or ochre (UAA) suppressors; that is, by a single anticodon nucleotide change. Yet only a subset of these have been found in intensive hunts in *E. coli* and yeast (*11*). Of those found, there is a class of easily isolated and efficient suppressor tRNA's, and a class of weak ones, sometimes requiring a deeper search. These two classes are distributed in a very particular way in Table 2. The strong or obvious set (Leu, Ser, Tyr, Trp) are all in the top line, the weak and obscure in the third vertical column (Gln, Lys, Glu).

In fact, the *E. coli* amber suppressing mutant of $tRNA^{Trp}$ (tryptophan tRNA) was not easily isolated because there is only one *E. coli* gene for $tRNA^{Trp}$ and the suppressor ($Su^+7$), therefore, could be isolated only in a partial diploid (*12*) where it would not make a lethal deletion in the cell's coding capacity. It is nevertheless strong, being among the most efficient known suppressor tRNA's (*12*). It therefore exemplifies the pattern rather than an exception. Both the *E. coli* serine suppressor and the amber suppressor mutant of T4 phage $tRNA^{Ser}$ (serine tRNA) fit the pattern cited; both are efficient and prevent termination of the majority of growing peptide chains at UAG (*13, 14*). They also have nearly identical anticodon stem-and-loop sequences (Fig. 1).

The effective difference between the strong and weak suppressors, which is obvious in Table 2, is that the weak ones require changes at the cardinal position (first position of the codon) and therefore risk creating a tRNA that contains a conflict between the cardinal nucleotide and its adjacent stem-and-loop sequence. The strong suppressors are those formed by mutation at the two other positions, which therefore cannot violate these rules of anticodon stem-and-loop organization. The weaker ones do violate these principles, as follows.

## Detailed Analysis of the Weak Suppressors

The amber suppressor $Su^+2$ is an anticodon mutation of $tRNA^{Gln}$ (glutamine tRNA) (*15, 16*; original codon CAG). It is also the least efficient of the extensively studied amber suppressors (*12*). The mutated tRNA is changed at the cardinal position so as to conflict with the natural pattern at position II (Table 1). Bradley *et al.* (*17*) have studied the efficiency of this suppressor genetically and we have studied it by tRNA gene construction (*18*). It appears that it is weak solely because of the sequence proximal to the anticodon (*17, 19*). That is, when the anticodon-proximal sequence is corrected to be in accord with Table 1, the resulting tRNA is an efficient amber suppressor despite the unchanged sequence throughout the rest of the tRNA.

The same low efficiency is predicted (Table 1 and Fig. 1) if T4 phage $tRNA^{Gln}$ mutates to a termination suppressor. The ochre suppressor so derived (T4 $psu^{+2}$) is, in fact, extremely weak; its action is not detectable unless the termination codon to be suppressed resides in a gene whose product is needed in small amounts (*20*).

The lysine-inserting tRNA mutant, $Su^+5$, has been detected in *E. coli* as an ochre suppressor (*21*); that is, because of the wobble of the $tRNA^{Lys}$ (lysine tRNA) mutated in the cardinal position, it reads both UAA and UAG (original codon AAA and AAG). The sequence of the known *E. coli* $tRNA^{Lys}$ (*22, 23*) if mutated to an ochre suppressor, would conflict with Table 1 at position V. Indeed $Su^+5$ is a weak suppressor, being less efficient than the ochre suppressor $Su^+4$ [which is $tRNA^{Tyr}$ (tyrosine tRNA) mutated in the noncardinal first anticodon position] by a factor of up to eightfold, depending on the marker studied (*21*). The difficulty is probably not with the insertion of lysine itself, since $Su^+5$ $tRNA^{Lys}$ was never superior at any of 12 sites tested, and both immunologically detected protein and enzyme activity were low and similarly decreased, when compared to $Su^+4$ $tRNA^{Tyr}$.

The case of the potential $tRNA^{Glu}$ (glutamic acid tRNA) *E. coli* ochre suppressor (original codons GAA and GAG; it might again wobble to read UAA and UAG) is perhaps the most interesting because it has not been possible to isolate this suppressor at all in *E. coli*, even by extensive mutagenesis of a cloned $tRNA_2^{Glu}$ gene (*23*). In fact, the required mutant $tRNA_2^{Glu}$ sequence, altered from C to A at the cardinal anticodon site, would differ from the sequence required at positions II, III, and IV (of Table 1 and Fig. 1). Its grossly conflicting internal structure presumably makes this suppressor too inefficient to be detected.

Taken together, these data support the concept of an extended anticodon. They also confirm that the relation between

the cardinal nucleotide and the extended anticodon has functional meaning because the weak suppressors are just those mutated at the cardinal position (Table 2). Even more particularly, these results comprise an argument for the importance of the nucleotide labeled II, and the nucleotide pairs labeled III, IV, and V (Table 1) in translational efficiency.

| First position (5' end) | Second position | | | | Third position (3' end) |
|---|---|---|---|---|---|
| | U | C | A | G | |
| U | Leu<br>Leu | Ser<br>Ser | Tyr<br>Tyr<br>Ochre<br>Amber | Trp | U<br>C<br>A<br>G |
| C | | | Gln<br>Gln | | U<br>C<br>A<br>G |
| A | | | Lys<br>Lys | | U<br>C<br>A<br>G |
| G | | | Glu<br>Glu | | U<br>C<br>A<br>G |

Table 2. A modified coding table. The origins of two classes of termination (amber or ochre) suppressors by single mutation in the anticodon are shown. An efficient, ribosome-insensitive class comes from the barred region, and an inefficient class which is expected to be restricted by streptomycin-resistant ribosomes comes from mutations in the dotted region.

### Translational Accuracy

A second apparent use of the extended anticodon relates not only to translational efficiency but also to the intrinsic accuracy of translation. The accuracy of tRNA selection is directly responsive to the state of small ribosomal subunit protein S12 (*24*). This protein, mutated to confer streptomycin resistance, makes the ribosome more accurate both in vivo (*25*) and in vitro (*26*). One expression of this enhanced accuracy is that the drug-resistant ribosome reduces the efficiency of certain tRNA's mutated to act as termination or missense suppressors (*27*; reviewed in *28*), with only small simultaneous effects on the action of normal tRNA's. The concept of a "ribosomal screen" (*28*), which filtered out the suppressors, was devised to summarize these observations. The "ribosomal screen" as defined by mutational analysis also controls the normal accuracy of translation (*25, 26*). The relevant point here is that the relation between the cardinal nucleotide and the extended anticodon seems to be the molecular character which is under observation by the accuracy-determining portion of the ribosome. The critical data, again, arise because tRNA suppressors have been changed by mutation, and therefore may not conform to the pattern of natural sequences (Table 1). Those which do conform to the natural pattern are relatively slightly affected by the ribosome's accuracy function; those which deviate from the pattern are strongly rejected.

Consider the amber supressors $Su^+1$ (Ser), $Su^+2$ (Gln), $Su^+3$ (Tyr), and $Su^+7$ (Gln). $Su^+1$, $Su^+3$, and $Su^+7$ have been cited just above as efficient suppressors, mutated at noncardinal sites (the first or second anticodon positions). They are also the least affected by streptomycin-resistant ribosomes (*29*). More severely affected by far is $Su^+2$, which, for that reason, is the termination suppressor usually used in experiments on the streptomycin resistance effect. As also discussed just above, $Su^+2$ is mutated at the cardinal (3') anticodon position, and the mutation does produce a tRNA in which the mutated position is in conflict with Table 1, position II. This is as expected if the translational apparatus determines accuracy by criteria embodied in Table 1.

Missense suppressors are also affected by ribosomal mutations. When missense suppression in an essential gene is required, a mutant cell sometimes grows much more slowly when a strong streptomycin-resistance allele is introduced (*30*), as a result of the less efficient production of the suppressed product (*30*). Figure 2 shows three missense suppressors, whose genes have been sequenced and which were characterized by Biswas and Gorini (*30*); two are unhindered by a more rigorous ribosome ($su^+58$ and $su^+36$), but one is greatly reduced in activity ($su^+78$). All three act as predicted from Table 1. The unaffected two still conform to the rules; *glyV* $su^+58$ is mutated at the noncardinal middle position (Fig. 2). More interestingly, *glyT* $su^+36$ is altered at the cardinal position; however, the C to U anticodon change which produces the suppressor (a $tRNA^{Gly}$ which works at Asp codons) happens to be consistent with the surrounding sequence (Fig. 2 and Table 1). The relative immunity of this tRNA to a rigorous ribosome is therefore explicable. The strongly ribosome-sensitive $su^+78$ suppressor is critically altered; its sequence conflicts with the pattern in Table 1 at positions III and IV, which comprise the loopward terminus of the anticodon stem. This argument extends these ideas about the function of tRNA to translation of normal codons as well as termination codons.

In summary, the susceptibility of both nonsense and missense suppressors to accuracy-controlling ribosomal mutations is well accounted for if some part of the translation apparatus compares positions II, III, and IV of Table 1 with the codon-anticodon sequence itself. Pairings in which these two regions are mismatched are discriminated against. This at first seems a sophisticated calculation for the ribosomal A site to make, but it can nevertheless be conceived of simply (below).

This analysis also leads to a critical prediction; the susceptibility of a synthetic or composite tRNA gene product (*18*) to a streptomycin-resistant ribosome should be determined by the anticodon stem and loop region and not by the rest of the tRNA sequence. We have confirmed that this is true, by moving anticodon region sequences from tRNA to tRNA using a tRNA gene construction method (*18, 31*).

### Efficiency of Missense Supressors

Missense suppressor tRNA's also vary in efficiency. That is, they vary in the proportion of peptide chains bearing the suppressor's amino acid, rather than the amino acid inserted by the unmutated tRNA's which normally translate the codon at the missense site (*32*). They are generally of very low efficiency compared to termination suppressors, but there is one which is strikingly more efficient than is usual (*33*). One would, in principle, like to compare this spectrum with the predictions of the extended anticodon hypothesis.

But, on the basis of current information, this does not seem practical. The termination suppressors all compete with the same cellular peptide chain termination pathway, but different mis-

sense suppressors compete with different sets of natural tRNA's. Each natural competitor potentially has a characteristic cellular abundance and an individual translational efficiency. One might meet this difficulty by comparing pairs of missense suppressor tRNA's which translate the same codon. However, such existing pairs of suppressors have different anticodon sequences. Therefore, the effects of different anticodons is confounded with the effect of the anticodon's extension. Further, no available pair includes two fully sequenced tRNA's (*34*).

In addition, the usable data concerns suppressors which are rendered difficult to aminoacylate by the suppressor mutation (*35*). An increase in aminoacyl-tRNA synthetase activity in vivo, managed genetically, increases suppressor efficiencies (*32*). Therefore, the efficiency data are biased by aminoacylation levels.

Further, data often refer to only one genetic context, such as a single missense mutant in the trpA gene. The potentially strong effect of the suppressed codon cannot be extracted, as it can for termination suppressors which have been examined in many genes and positions.

Finally, there is an interesting intrinsic uncertainty in the study of missense suppressors. At some point, as suppressor efficiency rises, efficiency itself will be altered by the substitution of a missense amino acid at many sites in the proteins of the translational apparatus. We do not know where this effect begins. In view of these reservations, the data on missense suppression efficiency will not be evaluated here.

## Other Genetic Data on tRNA Efficiency

Several other individual results confirm the role of the noncoding anticodon loop and extended anticodon nucleotides. The *E. coli* $Su^+2$ has two pseudouridine residues in its extended anticodon region at positions II and III in Table 1 (*16*). When these are not modified in the tRNA, but remain as U's due to a mutation in a cellular enzyme (the *hisT* product), the translational efficiency of the suppressor is decreased (*33*).

Similarly, tRNA mutants have been selected which overcome the low translational efficiency of *E. coli* $Su^+2$. These mutants, which not only increase efficiency but also reduce susceptibility to a streptomycin-resistant ribosome, are at position II (Table 1) (*17*), in the region implicated here.

Furthermore, when we compared (*18*) two tRNA genes constructed so as to differ only at four nucleotides, the two of the terminal base pair of the anticodon stem and the two topmost nucleotides of the anticodon loop, these tRNA's differed in translational efficiency (*18*). This independently indicates that translational efficiency can be determined by the structure near, but not at, or a part of, the anticodon itself.

The extensive set of suppressor-inactivating mutations of Kurjan *et al.* (*36*) includes the 5′ anticodon loop nucleotide, and the nucleotide 3′ to the anticodon, despite the fact that these mutant genes are transcribed normally (*37*) and are not obviously altered in coding or tertiary structure (*1*).

A last example involves the inefficient glutamine-inserting T4 ochre suppressor, $psu^+2$, which has appeared in the discussion above. When this tRNA is further weakened (in fact, inactivated) by a cellular mutation which removes a modification at the first (wobble) position of the anticodon, it can be partially reactivated by a second host-cell mutation which does not restore the modification, but instead sensitizes the ribosome to streptomycin (*38*). This reinforces the evidence that the anticodon loop is under surveillance by the ribosome, and extends the range of effects to the wobble position. Presumably, effects 5′ of the cardinal position may be propagated to it, and the whole codon-anticodon region thereby may be kept under surveillance.

Taken as a group, these results suggest that nucleotides ranging from the 5′ anticodon (wobble) position to the upper loop and nearby anticodon stem region strongly influence a tRNA's translational efficiency and susceptibility to the accuracy function of the ribosome. That is, they further support the concept of an extended anticodon.

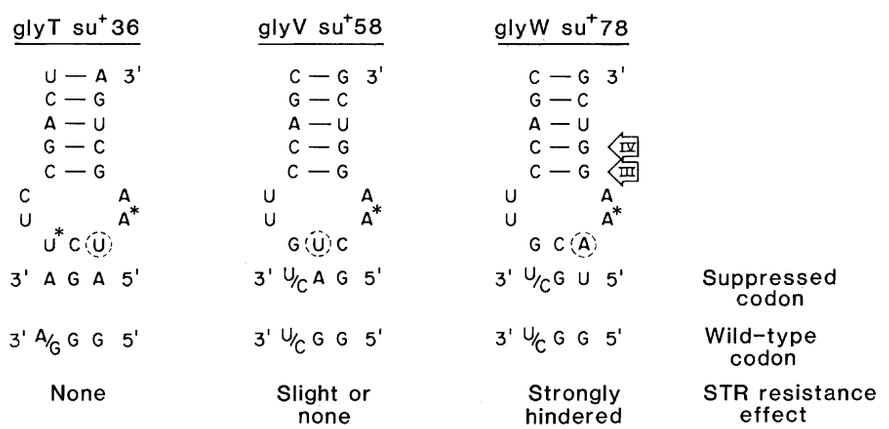



Fig. 2. The effect of anticodon region structure on missense tRNA's. The topward diagrams show anticodon loops of three missense suppressors [$su^+36$ (*8*); $su^+58$: sequence deduced from identification of the tRNA gene (*42*); $su^+78$ (*42*)] whose response on streptomycin-resistant ribosomes is known. Below each is the codon it has mutated to read, and its original codon. The lowest section of the figure shows the response of each suppressor when asked to function in a streptomycin-resistant cell (*30*). A dashed circle surrounds the mutated anticodon nucleotide. Arrows containing roman numerals refer to positions in disagreement with Table 1.

## Mechanisms of Extended Anticodon Function

In the argument thus far I have not identified the part of the translational apparatus which surveys the extended anticodon. In principle, aminoacylation, transfer factor interaction, nucleotide modification, or strictly ribosomal function might be affected, singly or in combination. I believe there is, nevertheless, substantive reason to prefer the hypothesis that the extended anticodon acts during the effective occupation of the ribosomal A site.

1) This is the simplest explanation; it explains all observations as efficiency at a single locus. The variety of tRNA's which perform in accord with prediction argues strongly against explanations which would require an unproven similarity between different cellular elements, for example, the different aminoacyl-tRNA synthetases.

2) The coherence between the several effects on translational efficiency for amber suppressors and the effects on susceptibility to a streptomycin-resistant ribosome suggests that efficiency and streptomycin resistance effects are the same phenomenon. Since the latter is definitely ribosomal, the former is likely to be ribosomal also.

3) The extended anticodon region is connected to ribosomal events by the inefficiency of $Su^+2$ in the *hisT* mutant cell, which does not modify U's in the anticodon loop region (*36*). Suppression can be restored in this case by a muta-

tional change in the message context. This is nicely consistent with a defect in the unmodified tRNA which is predominantly ribosomal, although it does not prove a defect at this step.

4) The anticodon loop (position II) mutations that overcome the streptomycin resistance constraint on $Su^+2$ (*17*) imply an anticodon loop-ribosomal interaction.

The weak, ochre-suppressing $tRNA^{Gln}$ derivative, T4 $psu^+2$, is not weak because of the particular amino acid it inserts, and the mature tRNA is abundant in the phage-infected cell. Unless T4 $psu^+2$ is poorly acylated, this suggests that it is defective at the ribosome (*20*).

One is also led to the ribosome by physical evidence. An important constraint comes from the experiments of Grosjean *et al.* (*2, 39*), who showed that the codon-anticodon interaction is strengthened by incorporating the anticodon into a hairpin loop. In contrast, the specificity of codon-anticodon interaction is not improved. In fact, potential new mistakes [pyrimidine-pyrimidine mispairs (*39*)] appear. Thus, the superior coding capabilities of tRNA's seem to require an interaction with the ribosome. The elevated accuracy may not generally pertain to conditions inside the ribosome, but only to the functioning A site; neither the P site at equilibrium nor the A site filled nonenzymatically appear to have an enhanced intrinsic accuracy (*40*). On the whole, while the information in the extended anticodon may ultimately be found to be used in other ways, a role at the ribosomal A site is very likely.

How, in molecular terms, can the A site compare the parts of the extended anticodon? I offer two possibilities for which there is suggestive evidence. These two are not mutually exclusive, but both are of interest because they have different implications for ribosomal events.

In the first possibility, the extended anticodon is viewed as the recognition site for modifying enzymes that make the modification at position I in Table 1 (*1*). The active sites of these enzymes presumably make the modification only when the cardinal nucleotide and extended anticodon are consistent. Because this modification, in some cases, is already known to be required for efficient tRNA function (*1, 2, 8, 9*), the origin of the rules in Table 1 might be explained. The cardinal nucleotide (the 3′ anticodon nucleotide) is also accommodated in an economical way, because the particular modification at position I (Table 1) has long been known to be highly correlated with the particular nucleotide in the cardinal position (*1, 7*).

The modification hypothesis neatly accounts for the logic of the extended anticodon, but it meets difficulties where its predictions can be compared with structural data. Inefficient tRNA's and tRNA's strongly rejected by streptomycin-resistant ribosomes should have unmodified anticodon loops acquired as a result of mutation; this prediction is usually not borne out.

The inefficient, ribosome-sensitive *E. coli* suppressor $tRNA^{Gln}$ $Su^+2$ is apparently still modified in the anticodon loop, even though modification is partial because the heavily ultraviolet-irradiated bacteria used to label the tRNA for sequence studies are defective in modification (*15, 16*). The weak T4 phage $tRNA^{Gln}$ suppressor, $psu^+2$, mutated, like the bacterial $Su^+2$ in the cardinal anticodon nucleotide, does not alter its modification at any position after mutation (in *E. coli* B), including the nucleotides in the anticodon loop (*41*). The glycine missense suppressor $su^+78$, which is very sensitive to streptomycin-resistant ribosomes, is appropriately modified (*42*).

A second plausible mechanism can be introduced in terms of a related example. Initiator tRNA ($tRNA_f^{Met}$, formyl methionine tRNA) has a unique ability to occupy the translation initiation site on the 30*S* ribosome, even in the absence of initiation factors. This property is not shared by any elongator tRNA (even $tRNA_m^{Met}$, methionyl $tRNA^{Met}$) (*43, 44*). The $tRNA_f^{Met}$ is therefore optimized for coding at initiation in the same way that the tRNA's discussed so far are optimized for coding during peptide chain elongation.

Initiator tRNA's also characteristically resemble each other but differ from elongator tRNA's in the region that I have called the extended anticodon. As Wrede *et al.* (*45*) have pointed out, the terminus of the anticodon stem of $tRNA_f^{Met}$ is uniquely three successive GC pairs in initiators, with the 5′ nucleotides all G's. As might be expected for a tRNA intended to occupy the P site instead of the A site, this sequence is different from that required of natural elongator tRNA's having the same cardinal nucleotide (compare Table 1).

Under S1 nuclease digestion conditions in solution, the anticodon loop digestion pattern of several initiators is similar, but all are different from elongator tRNA's (*45*). Therefore, initiators probably have a characteristic and particular anticodon loop conformation in solution. Even more striking is the result of crystallographic study of $tRNA_f^{Met}$ (*46, 47*); the three-dimensional disposition of the anticodon loop differs from that of elongator tRNA's. Therefore, a unique distinctive sequence feature in the extended anticodon region is well correlated with a distinctive three-dimensional structure for the anticodon loop region and with a distinctive coding potential. It is tempting to suppose that we can explain the operation of the extended anticodon by generalizing this structural effect to the ribosomal A site, and to all tRNA's.

In this generalized case, the constraint on the anticodon extension (Table 1) can be thought of as an accommodation to codon-anticodon helices which, for different tRNA's, will have different sequences and therefore potentially different pitch, stacking, and tilt. The local stem and loop sequences vary so as to compensate the codon-anticodon differences. The result of this covariation is that the message-tRNA combination always appears the same to the ribosomal A site, which is optimized for one nucleic acid conformation. The extended anticodon therefore solves the problem of designing an accurate site when 60 different substrates (the codon-anticodon complexes) must be accommodated. The ribosome enhances translational accuracy by rejecting message-tRNA complexes which do not fit its entire A site, including not only the codon-anticodon portion, but also an adjacent template surface for the anticodon loop and stem. The more accurate, streptomycin-resistant ribosome is presumed to have a more tightly fitting template with which to compare the entire extended anticodon-message complex.

### The Hirsh UGA Suppressor

There is one definite indication that there may be coding determinants in tRNA outside the region considered here, and this therefore requires comment [see (*48*)]. The Hirsh suppressor is a mutated $tRNA^{Trp}$ UGG translator which reads UGA. This tRNA has apparently acquired the ability to wobble to read UGA because of a D-helix mutation, without obvious change in the anticodon region (*49*). Thus, it may be necessary to expand the notion of the extended anticodon to this nearby site. However, it is not yet clear whether this expansion is necessary.

First, the Hirsh UGA suppressor is unique. Despite many suppressor mutant searches, no other such example has been found. This is particularly striking because the Hirsh-type coding change

in, for example, an amber suppressor, if it can occur, would produce a recognizable new type of ochre suppressor, which wobbles to read UAG or UAA. These have never been detected.

Second, the UGG or UGA coding proclivity preexists in the unmutated $tRNA^{Trp}$—in vitro (*50*) and in vivo (*51*)—and so is not created by the Hirsh mutation.

Third, the result of the mutation does not act as a normal *E. coli* translator, because it apparently can decode the unusual set of codons UGN, which has any base in the wobble position (*52*).

Fourth, a modified anticodon nucleotide has recently been discovered in *E. coli* which, if formed in $tRNA^{Trp}$, potentially explains the Hirsh coding change as an alteration of the anticodon itself (*53, 54*).

Thus, there is no doubt of the value of the Hirsh UGA suppressor as a provocative example. It may be too narrow a basis, however, for a general theory of coding.

## The Puzzle of Ineffective tRNA's

The existence of very weak termination suppressors like $Su^+2$ might be taken as an indication that natural tRNA's vary greatly in their translational efficiency. That is, a weak suppressor might arise as a mutation in an ineffective tRNA. This potentially conflicts with the plausible idea that most *E. coli* tRNA's are optimized for efficiency.

However, this potential complexity does not arise in the extended anticodon hypothesis, which suggests that weak suppressors become defective as a result of the suppressor mutation itself. The idea that translation is a relatively homogeneous process is therefore preserved.

## The Basic Notions

There are two essential ideas in this article. First, the functional anticodon sequence is a part of a larger structure, the extended anticodon. Second, the extended anticodon is organized predominantly to suit the properties of a cardinal nucleotide (or to suit the nucleotide pair of which it is a member).

When the cardinal nucleotide is embedded in a suitable extended anticodon, and further embedded in an otherwise normal tRNA molecule, the anticodon performs as an efficient and accurate translational device.

These are ideas and not yet facts. They are nevertheless considerable ideas because they successfully organize a large body of data that was previously in disarray. Modern synthetic and recombinant methods provide the opportunity to test the extended anticodon by intervening directly in the sequence of the tRNA gene. Thereby we can distinguish the significant aspects of tRNA structure and assign quantitative importance to the various parts of the extended anticodon, which undoubtedly differ in this respect.

If the extended anticodon hypothesis is correct in any part, then RNA structures may be shaped with unexpected subtlety to do a particular job. I expect that other functions carried out by RNA's will benefit from the same subtle enhancement, whatever its ultimate mechanism.

### References and Notes


1. S. Nishimura, in *Transfer RNA: Structure, Properties, and Recognition* (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1979), p. 59.
2. H. Grosjean, S. deHenau, D. Crothers, *J. Mol. Biol.* **103**, 499 (1976).
3. Abbreviations for the nucleotide bases are A, adenine; T, thymine; U, uracil; G, guanine; C, cytosine. Those for the amino acid residues are; Ala, alanine; Asn, asparagine; Asp, aspartic acid; Arg, arginine; Cys, cysteine; Glu, glutamic acid; Gln, glutamine; Gly, glycine; His, histidine; Ile, isoleucine; Leu, leucine; Lys, lysine; Met, methionine; Phe, phenylalanine; Pro, proline; Ser, serine; Thr, threonine; Trp, tryptophan; Tyr, tyrosine; Val, valine. Nonacylated tRNA's are designated $tRNA^{Ala}$, for example; m, methyl; f, formyl; see also (*55*).
4. F. H. C. Crick, *J. Mol. Biol.* **19**, 548 (1966).
5. W. McClain, C. Guthrie, B. Barrell, *Proc. Natl. Acad. Sci. U.S.A.* **69**, 3703 (1972).
6. J. Wilson, J. Kim, J. Abelson, *J. Mol. Biol.* **71**, 547 (1972).
7. M. Yarus, *Annu. Rev. Biochem.* **38**, 841 (1969).
8. J. Roberts and J. Carbon, *J. Biol. Chem.* **250**, 5530 (1975).
9. J. Miller, Z. Hussain, M. Schweizer, *Nucleic Acids Res.* **3**, 1185 (1976).
10. M. Gefter and R. Russel, *J. Mol. Biol.* **39**, 145 (1969).
11. A. M. Korner, S. I. Feinstein, S. Altman, Eds., in *Transfer RNA* (MIT Press, Cambridge, Mass., 1978), p. 105.
12. L. Soll and P. Berg, *Proc. Natl. Acad. Sci. U.S.A.* **63**, 392 (1969).
13. W. McClain, C. Guthrie, B. Barrell, *J. Mol. Biol.* **81**, 157 (1973).
14. J. H. Wilson and S. Kells, *ibid.* **69**, 39 (1972).
15. H. Inokuchi, F. Yamao, H. Sakano, H. Ozeki, *ibid.* **132**, 649 (1979).
16. H. Inokuchi, M. Kodaira, F. Yamao, H. Ozeki, *ibid.*, p. 663.
17. D. Bradley, J. Park, L. Soll, *J. Bacteriol.* **145**, 704 (1981).
18. M. Yarus, C. McMillan III, S. W. Cline, D. Bradley, M. Snyder, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 5092 (1980).
19. M. Yarus and S. W. Cline, unpublished data.
20. M. M. Comer, C. Guthrie, W. H. McClain, *J. Mol. Biol.* **90**, 665 (1974).
21. E. Gallucci and A. Garen, *ibid.* **15**, 193 (1966).
22. L. Chakraburtty, A. Steinschneider, R. Case, A. Mehler, *Nucleic Acids Res.* **2**, 2069 (1975).
23. H. Ozeki *et al.*, in *Transfer RNA: Biological Aspects* (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1980), p. 341.
24. M. Ozaki, S. Mizushima, M. Nomura, *Nature (London)* **222**, 333 (1969).
25. P. Strigini and E. Brickman, *J. Mol. Biol.* **75**, 659 (1973).
26. J. Davies, W. Gilbert, L. Gorini, *Proc. Natl. Acad. Sci. U.S.A.* **51**, 883 (1964).
27. D. Apirion and D. Schlessinger, *ibid.* **58**, 506 (1967).
28. L. Gorini, in *Ribosomes* (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1974), p. 791.
29. M. Kuwano, M. Ishizawa, H. Endo, *J. Mol. Biol.* **33**, 513 (1968); L. Gorini, *Nature New Biol.* **234**, 261 (1971); L. Soll, thesis, Stanford University (1971).
30. D. Biswas and L. Gorini, *J. Mol. Biol.* **64**, 119 (1972).
31. D. Bradley, M. Yarus, L. Soll, unpublished data.
32. C. W. Hill, C. Squires, J. Carbon, *J. Mol. Biol.* **52**, 557 (1970).
33. L. Bossi and J. R. Roth, *Nature (London)* **286**, 123 (1980).
34. C. W. Hill, *Cell* **6**, 419 (1975).
35. J. Carbon, C. Squires, C. W. Hill, *J. Mol. Biol.* **52**, 571 (1970).
36. J. Kurjan, B. D. Hall, S. Gillam, M. Smith, *Cell* **20**, 701 (1980).
37. R. Koski, S. G. Clarkson, J. Kurjan, B. D. Hall, M. Smith, *ibid.* **22**, 415 (1980).
38. D. Colby, P. Schedl, C. Guthrie, *ibid.* **9**, 449 (1976).
39. H. Grosjean, S. deHenau, D. Crothers, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 610 (1978).
40. M. Peters and M. Yarus, *J. Mol. Biol.* **134**, 471 (1979).
41. J. Seidman, M. Comer, W. McClain, *ibid.* **90**, 677 (1974).
42. C. Squires and J. Carbon, *Nature (London) New Biol.* **233**, 274 (1971); J. Carbon and E. Fleck, *J. Mol. Biol.* **85**, 371 (1974).
43. P. S. Rudland, W. Whybrow, K. A. Marcher, B. F. C. Clark, *Nature (London)* **222**, 750 (1969).
44. H. U. Petersen, A. Danchin, M. Grunberg-Manago, *Biochemistry* **15**, 1357 (1976).
45. P. Wrede, N. Woo, A. Rich, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 3289 (1979).
46. R. Schevitz, A. Podjarny, N. Krishnamachari, J. Hughes, P. Sigler, J. Sussman, *Nature (London)* **278**, 188 (1979).
47. N. Woo, B. Roe, A. Rich, *ibid.* **286**, 346 (1980).
48. C. G. Kurland, in *Nonsense Mutations and tRNA Suppressors*, J. E. Celis and J. S. Smith, Eds. (Academic Press, New York, 1979), p. 97.
49. D. Hirsh, *J. Mol. Biol.* **56**, 439 (1971).
50. _____ and L. Gold, *ibid.* **58**, 459 (1971).
51. S. Lastick and M. Yarus, unpublished data.
52. R. H. Buckingham and C. G. Kurland, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5496 (1977).
53. F. Harada and S. Nishimura, *Biochemistry* **13**, 300 (1974).
54. Y. Kuchino, S. Watanabe, F. Harada, S. Nishimura, *ibid.* **19**, 2085 (1980).
55. D. Gauss and M. Sprinzl, *Nucleic Acids Res.* **9**, 41 (1981).


56. I thank Doug Bradley, Linda Breeden, Steve Cline, Steve Eisenberg, Stan Lastick, Laurel Raftery and Bob Thompson for help in clarifying earlier versions of this manuscript. The preparation of this article was supported by USPHS research grant GM25627.