Chemistry & Biology, Vol. 9, 237–244, February, 2002, ©2002 Elsevier Science Ltd. All rights reserved. PII S1074-5521(02)00094-7

Exploring the Limits of Codon and Anticodon Size

J. Christopher Anderson,[1] Thomas J. Magliery,[2,5] and Peter G. Schultz[1,3,4]
[1]Department of Chemistry
The Scripps Research Institute
10550 North Torrey Pines Road
La Jolla, California 92037
[2]Department of Chemistry
University of California
Berkeley, California 94720
[3]Genomics Institute
of the Novartis Research Foundation
3115 Merryfield Row
Suite 200
San Diego, California 92121

Summary

We previously employed a combinatorial approach to identify the most efficient suppressors of four-base codons in *E. coli*. We have now examined the suppression of two-, three-, four-, five-, and six-base codons with tRNAs containing 6–10 nt in their anticodon loops. We found that the *E. coli* translational machinery tolerates codons of 3–5 bases and that tRNAs with 6–10 nt anticodon loops can suppress these codons. However, *N*-length codons were found to prefer *N* + 4-length anticodon loops. Additionally, sequence preferences, including the requirement of Watson-Crick complementarity to the codon, were evident in the loops. These selections have yielded efficient suppressors of four-base and five-base codons for our ongoing efforts to expand the genetic code. They also highlight some of the parameters that underlie the fidelity of frame maintenance.

Introduction

The canonical genetic code of triplet nucleotides is a nearly universal feature of life, and its universality is considered key evidence for the common ancestry of all known life [1, 2]. Nevertheless, exceptions to this code, including codon reassignment in mitochondria and ciliates, coding of a "twenty-first" amino acid (selenocysteine or Sec), and programmed translational frameshifting as a regulatory mechanism in viruses and bacteria [2–4], have received attention lately. These phenomena fall into two broad and perhaps overlapping categories: those that occur because of slipping or hopping of "normal" tRNAs at the ribosome and those that occur because of the presence of an unusual tRNA species, such as the $tRNA^{Sec}$ that decodes the stop codon UGA [5], the $tRNA^{Gln}$ from ciliates that decodes the stop codon UAG [6], or various prolyl, glycyl, alanyl, and seryl tRNA species that have been discovered or engineered with extended anticodon loops [7–9].

Slipping and hopping, or translational bypassing, are mediated by the commonly occurring set of tRNAs and usually reach appreciable levels only in the presence of other "stimulatory" elements that cause translational pausing. Such elements include mRNA secondary structure, underused codons for which the cognate tRNA is in low abundance ("hungry" codons), and upstream Shine-Delgarno-like sequences. One prominent example of this kind of recoding is the self-regulating release factor 2 (RF2) gene of bacteria; in this case, a low concentration of RF2, which mediates termination at UGA, results in tRNA slippage past an in-frame UGA codon. This slippage of $tRNA^{Leu}$ at the sequence CUU UGA is also dependent upon an upstream sequence that binds to 16S rRNA [4]. However, even without these secondary signals, "leakiness" in frame maintenance has been observed [10]. Using a library approach, we analyzed the propensity of a +1 frameshift at all possible 4 nt sequences in the absence of other stimulatory elements. The resulting sequences largely corresponded to those (such as AAAA) for which re-pairing of the tRNA was possible in the +1 frame, with little dependence upon the abundance of the tRNA [9]. This mechanism of +1 and −1 shifts at "slippery" sequences is well established [11].

Other sorts of high-level translational recoding involve special tRNA species, such as nonsense suppressors (which often have a canonical anticodon to UAG or UGA stop codons) or +1 frameshift suppressors, which also typically decode canonical four-base codon sequences with an extended (8 nt) anticodon loop. There may be cases in which more than one mechanism for recoding is involved, as in the example of an anticodon insertion that appears to render the tRNA inactive and results in decoding with slippage by a normal near-cognate tRNA at CCCC codons [12]. Nevertheless, most of the known four-base codon suppressors exhibit a Watson-Crick canonical four-base anticodon [7]. Besides naturally occurring prolyl and glycyl suppressors from yeast and *Salmonella*, efficient four-base codon suppressors have been engineered by using $tRNA^{Ala}$, $tRNA^{Leu}$, and $tRNA^{Ser}$ as scaffolds. Moore et al. found that $tRNA^{Leu}$(UCUA) efficiently decodes UAGA codons [13]. Using a library approach to examine all possible four-base codons, we isolated efficient suppressors of the four-base codons AGGA, UAGA, CCCU, and CUAG. These $tRNA^{Ser}$ derivatives had 8 or 9 nt anticodon loops and consistently exhibited a Watson-Crick complementary four-base anticodon. The four-base codons typically corresponded to underrepresented three-base codons (such as AGG, the least used codon in *E. coli*) plus one additional nucleotide [9].

Our primary interest in identifying four-base codon suppressors has been to augment the number of ways to uniquely specify amino acids for in vitro and in vivo genetic-code expansion. For example, we have recently engineered a mutant of *Methanococcus jannaschii* tyrosyl-tRNA synthetase that acylates a modified $tRNA^{Tyr}$ (CUA) with *O*-methyltyrosine. This acylation results in site-

[4]Correspondence: schultz@scripps.edu
[5]Present address: Department of Molecular Biophysics and Biochemistry, Yale University, P.O. Box 208114, New Haven, Connecticut 06520.

specific insertion of this unnatural amino acid in response to *amber* stop codons [14]. We are in the process of developing other tRNA/synthetase "orthogonal" pairs that will be used in selections for the insertion of other unnatural amino acids in *E. coli*. Among these are $tRNA^{Leu}$ derivatives from archaebacteria that are being engineered to suppress four-base codons (J.C.A. and P.G.S., unpublished data). This prompted us to examine whether efficient suppression of five-base codons is also possible with tRNAs bearing 8, 9, or 10 nt anticodon loops.

This combinatorial approach to the discovery of four- and five-base codon suppressors can also be used to address the question of what limits exist on codon and anticodon size in *E. coli*. In particular, we have employed a library approach to identify suppressors of two-, three-, four-, five- and six-base codons with tRNAs containing 6–10 nt anticodon loops. Our results indicate that the translational apparatus permits decoding of three-, four-, or five-base codons and that each codon type favors tRNAs of discrete sizes. Thus, there are limits on both codon size, presumably governed at the ribosome, and tRNA anticodon loop size, corresponding to codon length. This supports the notion that, within limits, tRNA is the "molecular ruler" that measures out codon size during translation.

Results

Strategy: Library Construction

A combinatorial approach involving two types of libraries was used to investigate the suppression of codons of various sizes with tRNAs containing anticodon loops of various sizes (Figure 1). The first type of library (the codon libraries) was a reporter constructed by the replacement of a codon in the gene for β-lactamase with two, four, five, or six randomized nucleotides. These effectively encode β-lactamase frameshift mutants that contain a 1 nt deletion (requiring a –1 frameshift to produce the full-length protein), a 1 nt insertion (requiring a +1 frameshift), a 2 nt insertion (requiring a +2 frameshift), or an amino acid insertion (requiring a +3 frameshift to produce β-lactamase of the proper length), respectively. These libraries were constructed at the codon corresponding to Ser70, the catalytic serine of β-lactamase. Additionally, libraries were constructed with four, five, or six random nucleotides at the codon corresponding to Ser124, which is known to be permissive to most amino acids (except possibly proline, a property we previously verified [9]). These libraries were expressed in the context of a derivative of pBR322 with the natural promoter on this plasmid. When cells containing these reporters are grown on media containing ampicillin, therefore, the concentration of ampicillin at which the cells can survive reflects the amount of suppression of the randomized codon.

The second type of library (the suppressor anticodon libraries) was constructed with a pACYC184-derived plasmid. High-level transcription of the tRNAs was driven by the *lpp* promoter with a terminator from the *rrnC* gene. The libraries consisted of $tRNA_2^{Ser}$ derivatives in which the 7 nt anticodon loop was replaced with six, seven, eight, nine, or ten random nucleotides. The seryl tRNA was chosen as a scaffold since seryl-tRNA synthetase (SerRS) is known to acylate mutants of $tRNA^{Ser}$ with changes in the anticodon loop. The enzyme does not contact the anticodon loop; instead, SerRS recognizes the long variable loop of $tRNA^{Ser}$ and nucleotides in the acceptor stem [15, 16]. This should minimize the effects of differential aminoacylation resulting from different anticodon sequences, so that survival on ampicillin is related only to the degree of suppression by the tRNA (i.e., the codon-anticodon interaction). Moreover, since these tRNAs deliver a serine, and serine is required at Ser70 in β-lactamase, the engineered tRNA should be responsible for suppression at the randomized Ser70 codons.

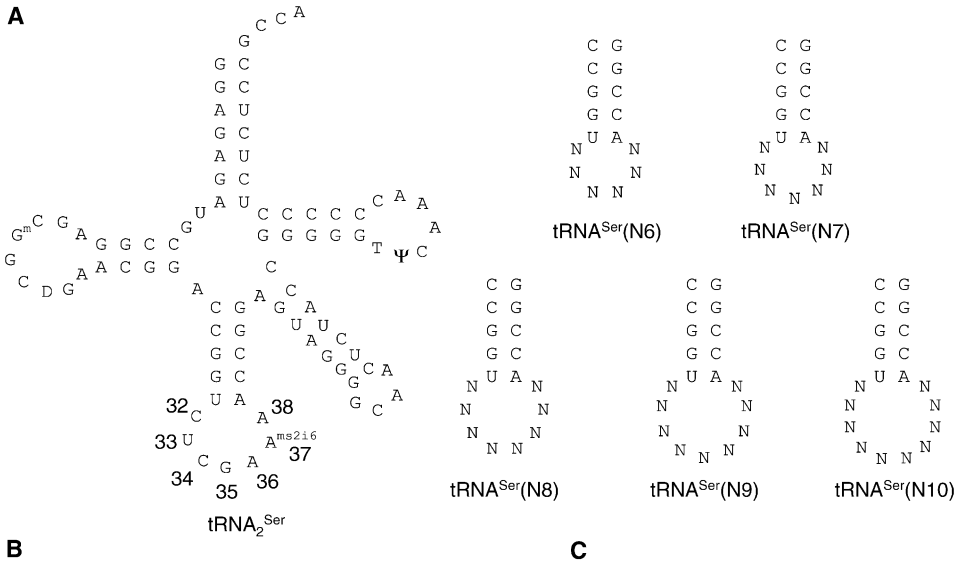
Sequencing of clones from each of the naive libraries of reporters and suppressors showed them to contain mainly codons and anticodon loops of the proper size and to have no evident sequence bias at the randomized positions. Moreover, when cells containing reporter constructs at the permissive Ser124 site were grown on 100 μg/ml ampicillin, the level of contamination by ampicillin-resistant cells was found to be very low (10^{-5}–10^{-3}). This shows that the amount of +1, +2, or +3 frameshifting is quite low in the absence of external suppressors. Also, the Ser70(N6) and Ser124(N6) reporter libraries, whose members are in frame but contain an amino acid insertion at the indicated codon, do not contain active clones. Therefore, amino acid insertions are not tolerated immediately before or after Ser70 or Ser124 in β-lactamase. In fact, when clones from the S70(N6) and S124(N6) libraries were subjected to selection at lower levels of ampicillin (10–50 μg/ml), the surviving clones invariably contained β-lactamase genes with only three nucleotides at the appropriate sites (and only corresponded to Ser at Ser70).

Suppression experiments were carried out by the creation of competent cells of an *E. coli* strain transformed with the appropriate reporter library. These competent reporter strains were then transformed with one of the $tRNA^{Ser}$ libraries, and the resulting cotransformed cells were grown on media containing various ampicillin concentrations. The tRNA and reporter plasmids were then separated as previously described, and the tRNA and β-lactamase genes were sequenced. Since the Ser124 site is permissive to changes in the amino acid, only the Ser70 site (which requires a serine or cysteine) can be used to examine missense and nonsense suppressors (i.e., suppressors of three-base codons). Therefore, suppression was initially examined at least at Ser70; some codons were also examined at Ser124 to confirm that the suppression effects were not entirely context dependent.

Suppression of Two-Base Codons

In the case of the Ser70(N2) library, very few clones were observed at any level of ampicillin with tRNAs containing six or seven nucleotides in their anticodon loops (the natural size for a tRNA or one smaller). Invariably, when the reporters from these selectants were sequenced, they were found to contain three-nucleotide codons at the Ser70 position. We previously observed a "deletion" phenomenon in our β-lactamase Ser70(N4) libraries which we attributed to inefficiency in the coupling and

A



B

β-lactamase Ser70 site and libraries

… Arg Phe Pro Met Met **Ser** Thr Phe Lys Val …
… CGT TTT CCA ATG ATG **AGC** ACT TTT AAA GTT …

… Arg Phe Pro Met Met Nnn Leu Leu Lys Phe …
… CGT TTT CCA ATG ATG **NN**A CTT TTA AAG TTC …

… CGT TTT CCA ATG ATG **NNN** ACT TTT AAA GTT …
… CGT TTT CCA ATG ATG **NNN** **N**AC TTT TAA AGT …
… CGT TTT CCA ATG ATG **NNN** **NN**A CTT TTA AAG …
… CGT TTT CCA ATG ATG **NNN** **NNN** ACT TTT AAA …

β-lactamase Ser124 site and S124(N5) library

… Arg Glu Leu Cys **Ser** Ala Ala Ile Thr Met …
… AGA GAA TTA TGT **AGT** GCC GCC ATA ACC ATG …

… Arg Glu Leu Cys Nnn Nnn Pro Pro End Pro …
… AGA GAA TTA TGT **NNN** **NN**G CCG CCA TAA CCA …

C

Nucleotides in anticodon loop	Decoding pattern
6 nt	NNN 5'~NNNNNN~3'
7 nt (wild-type)	NNN 5'~NNNNNNN~3' 32 33 34 35 36 37 38
8 nt	NNNN 5'~NNNNNNNN~3' NNN
9 nt	NNNNN 5'~NNNNNNNNN~3' NNNN NNN
10 nt	NNNNN 5'~NNNNNNNNNN~3'

Figure 1. Libraries Employed and the Patterns of Decoding Observed in This Study

(A) Anticodon loop libraries were derived from a seryl tRNA of *E. coli*.

(B) Replacement of Ser70 or Ser124 sites in β-lactamase with two, four, or five bases results in abortive translation unless the frameshift is suppressed. Replacement of the highly conserved Ser70 site with three or six bases results in a nonfunctional enzyme unless the missense mutation is decoded as a single serine residue.

(C) The observed alignment of the putative anticodon within the anticodon loop is shown for codons of different sizes. The numbering scheme for the anticodon loop is indicated on the 7 nt loop.

capping reactions during synthesis of the oligonucleotides used to construct these libraries [9]. Perhaps this "insertion" phenomenon is due to incomplete protection of the phosphoramidites, resulting in coupling of more than one deoxyribonucleotide per step. Whatever the reason, it is clear that tRNAs derived from $tRNA_2^{Ser}$ with six or seven nucleotides are incapable of suppressing −1 frameshifts by a direct reading mechanism within the limits of our assay.

Suppression of Three-Base Codons

Because of the fact that there are six Ser codons and two Cys codons, the amount of background from a Ser70(N3) library would make selection of suppressors impractical. Therefore, we instead constructed several missense or nonsense mutants of β-lactamase at Ser70 and selected for suppressors from $tRNA^{Ser}$(N7) and $tRNA^{Ser}$(N6) libraries (Table 1). These sense codons, AGG, AGA, and CGG, are among the least used in the *E. coli* genome. The suppression level by the best tRNA suppressor isolated for each three-base codon was quantified by the nitrocefin chromogenic assay described previously [9].

Each of the suppressors, with either 6 nt or 7 nt anticodon loops, had a readily identifiable Watson-Crick complementary anticodon. In the efficient 7 nt anticodon loop suppressors, there were invariably two nucleotides on either side of the anticodon; in the efficient 6 nt anticodon loop suppressors, there was 1 nucleotide 5′ to the anticodon and two nucleotides 3′ to it. Although preference for a U 5′ to the anticodon and an A 3′ to the anticodon was observed, it was not universal. Interestingly, the most efficient 7 nt anticodon loop suppressor tRNA was for the *amber* stop codon, for which many suppressors, including a seryl *amber* suppressor, are known in

Table 1. Suppressors of Selected Three-Base Codons with 6 nt and 7 nt Anticodon Loops

Codon	7 nt Anticodon Sequence	7 nt Anticodon Suppression	6 nt Anticodon Sequence	6 nt Anticodon Suppression
AGA	CU**UCU**AC	13.4%	**CUCU**UG	2.8%
AGG	UU**CCU**AG	30.2%	**ACCU**AA	10.2%
CGG	GG**CCG**GC	29.2%	**ACCG**AA	18.9%
UAG	UU**CUA**AC	68.4%	none	none
UGA	A**UUCA**AA	18.6%	none	none

The location of the anticodon within the loop is noted. Suppression is measured by the nitrocefin assay and is given as a percentage of the positive control, wild-type β-lactamase. No 6 nt anticodon loop suppressors for UAG or UGA were identified.

E. coli. In general, suppression of three-base codons was poorer with the 6 nt anticodon loop tRNAs than with the 7 nt anticodon loop tRNAs.

Suppression of Four-Base Codons

We previously demonstrated that a variety of seryl-tRNA derivatives with 8 or 9 nt loops were capable of suppressing +1 frameshifts (four-base codons) at moderate to high levels (up to 1,500 μg/ml ampicillin) [9]. However, when we selected for suppressors of four-base codons from tRNAs containing 7 nt anticodon loops (the natural length), we were unable to isolate suppressors at any reasonable level of ampicillin (as low as 5 μg/ml). Here again, the few selectants were found to contain a three-base codon at the Ser70 position, presumably because of the "deletion" phenomenon we previously reported after testing our N4 reporter libraries.

Suppression of Five-Base Codons

Since it seemed clear that the best suppressors of three-base codons contain 7 nt anticodon loops and those of four-base codons contain 8 nt anticodon loops, we first tested the ability of tRNAs with 9 nt anticodon loops to suppress five-base codons. Here, the number of possible tRNA/reporter combinations is quite large (3×10^8), and so we did not anticipate being able to identify all suppressor/codon combinations. Upon crossing of the $tRNA^{Ser}$(N9) library with the Ser70(N5) reporter library, we identified ten different five-base codons at 300 μg/ml ampicillin (Table 2). We also crossed this tRNA library against the five-base codon reporter library at Ser124 by using 100 μg/ml ampicillin to confirm that the suppression phenomenon was not context dependent. A similar set of codons was seen here (data not shown). Two of these codons, CCACU and CUAGU, were selected at 1,000 μg/ml ampicillin at Ser70.

Since we could not cover all sequence space in these experiments, and since we previously found AGGN codons to be the best-suppressed four-base codons, we constructed a β-lactamase reporter library of the type Ser70(AGGNN) and crossed this against the $tRNA^{Ser}$(N9) library. The codon AGGAC was selected at 100 μg/ml ampicillin, and AGGAU was selected at 100 and 300 μg/ml ampicillin. As we previously observed with four-base codon suppression, the last nucleotides of the codon can have a profound effect on the suppression level, so that the four AGGAN codons are not suppressed equally well.

Upon sequencing the tRNA suppressors of some of these five-base codons, we discovered that all of the tRNAs exhibited clear Watson-Crick complementary anticodons, except for AGGAU, whose GUCCU anticodon has a G•U wobble pair in the last position of the codon. Also, in each case, the five-base anticodon was flanked by two nucleotides on each side, which is the same pattern exhibited naturally by three-base anticodons and by the engineered four-base anticodons of our previous study. Interestingly, a single suppressor of CCAUC was found to contain a 10 nt anticodon loop. In this case, a canonical anticodon was observed with three nucleotides on the 5′ side of the anticodon. In every case but one, there was a U 5′ to the anticodon and an A 3′ to the anticodon, a bias seen in natural *E. coli* tRNAs as well. The observed suppression efficiencies of these tRNAs, as measured by the amount of the chromogenic substrate nitrocefin turned over by a known quantity of cells, were on the order of 1%–12% of the amount of turnover by β-lactamase with an AGC (Ser) codon at S70. These suppressors are therefore poorer than known three-base or four-base codon suppressors. It is also of note that, although the AGGNN codons were not identified in the more general selection, they were among the most efficient suppressors.

Table 2. Suppressers of Five-Base Codons with 9 nt and 10 nt Anticodon Loops

Codon	Usage	Anticodon loop	Suppression
AGGAC	AGG	CU**GUCCU**AA	5.0%
AGGAU[a]	Arg 0.12%	CU**GUCCU**AA	11.3%
CCAAU	CCA	CU**AUUGG**AC	4.4%
CCACC	Pro	UU**GGUGG**AA	1.6%
CCACU	0.84%	CU**AGUGG**AC	7.4%
CCAUC		GU**GAUCC**AA	8.0%
CCAUC[b]		UUU**GAUGG**AG	5.6%
CCCUC	CCC Pro 0.55%	CU**GAGGG**UC	3.8%
CGGUC	CGG Arg 0.54%	UU**GACCG**AC	4.5%
CUACC	CUA	GU**GGUAG**AA	7.4%
CUACU	Leu	UU**AGUAG**AU	11.2%
CUAGC	0.39%	CU**GCUAG**AA	8.5%
CUAGU		UU**ACUAG**AC	12.0%

Codon usage is expressed as a percentage of all codons in *E. coli*. The position of the anticodon is indicated within the loop. Suppression was measured by the nitrocefin assay. The AGGNN codons were identified in the selection with 9 nt anticodon tRNAs and the Ser70(AGGNN) library.

[a] There is a U•G wobble pair to the last base of the codon here.

[b] A 10 nt anticodon loop was isolated here.

Table 3. Crossreactivity of Suppressors Elicited Against Codons Based on AGG

	Anticodon Loop			
Ser70 Codon	**ACCU**AA	UU**CCU**AG	CU**UCCU**AG	CU**GUCCU**AA
AG	–	–	–	–
AGG	10.4%	30.2%	25.0%	2.6%
AGGA	–	–	11.2%	1.4%
AGGAC	–	–	–	5.7%

The 6 nt and 7 nt suppressors were identified with the Ser70(AGG) reporter. The 8 nt suppressor was previously shown to be an efficient suppressor of AGGA. The 9 nt suppressor was identified from selection against the Ser70(AGGNN) reporter library. Suppression was quantified with the nitrocefin assay. No activity was observed when the $tRNA^{Ser}$(CUA) *amber* suppressor was used with these codons. Ser70 must be a Ser or Cys for β-lactamase activity, and AGG codes for Arg.

Suppression of Six-Base Codons

We attempted to suppress six-base codons from the Ser70(N6) library with tRNAs containing 10 nt anticodon loops. All of the selectants exhibited four-base or five-base codons at the Ser70 site. Since sequence space for this library is even larger than with the five-base codon experiment (4×10^9), we again constructed a smaller reporter library of the type Ser70(AGGNNN). However, upon crossing this reporter library with the $tRNA^{Ser}$(N10) library, we found that all the selectants still had four-base or five-base codons at Ser70. This result suggests that suppression of six-base codons is highly inefficient but that tRNAs with 10 nt anticodon loops may be able to weakly suppress four-base or five-base codons (as seen in one case above). It is also possible that these tRNAs exhibit the same "deletion" phenomenon we have observed in the reporter libraries, affording tRNAs with 8 or 9 nt anticodon loops.

Crossreactivity of Suppressors with Codons of Different Lengths

Using the nitrocefin assay to quantify suppression efficiency, we assessed the crossreactivity of codons based on AGG with the best suppressors of those codons (Table 3). None of the tRNAs mediated appreciable suppression of Ser70(AG). However, tRNAs with 6, 7, 8, or 9 nt anticodon loops were able to suppress AGG at Ser70, although the 7 nt anticodon loop tRNA was most efficient. The four-base codon AGGA could be suppressed by 8 nt and 9 nt anticodon loop tRNAs, and the five-base codon AGGAC could only be suppressed by its 9 nt anticodon loop tRNA.

Leakiness in Frame Maintenance: +2 Frameshifting

We previously showed that a low level of +1 frameshifting occurs in the absence of four-base codon suppressors, and this frameshifting tended to be highest at codons where the anticodon could re-pair in the +1 frame (i.e., slip) [9]. When we examined five-base codon reporter clones that could survive on a modest amount of ampicillin (25 μg/ml), the following sequences were found at Ser124: CACUA, CUCUA, UAGAC, UAGCU, UAGGC, and UCUUA. Ser124 was used here because it is permissive to amino acid changes and we were interested in events that do not involve our seryl suppressors. Since the base that follows Ser124 is a G, all of the five-base codons at which +2 shifting occurs are of the type UAGNN or (NNNUA)G; that is, they all have an in-frame *amber* stop codon.

Discussion

What Do Three-Base, Four-Base, and Five-Base Codon Suppressors Share?

Nearly all of the efficient suppressors that we have isolated for three-base, four-base and five-base codons act at rare codons or the *amber* stop codon, the least used stop codon in *E. coli* (7.6% of stops). Most of the efficiently suppressed codons are based on AGG (AGG, AGGA, and AGGAC; AGG constitutes 0.12% of *E. coli* codons), CCC (CCCU, CCCCU, and CCCUC; 0.55%), CUA (CUAG, CUAGC, and CUACU; 0.39%), and UAG (UAG, UAGA, and UAGGC). (Note that if all codons were used equally, they would have 1.6% usage. Usage data are from the Kyoto Encyclopedia of Genes and Genomics [17]). Notably, however, not all rare codons elicited suppressors. For example, we were able to construct suppressors of AGA (0.21% usage, the second-least-used codon), but they were weaker suppressors than UAG, AGG, and CGG suppressors, and no AGAN or AGANN suppressors were found. Also, not all related codons could be suppressed equally; for most codons, only a subset of the (NNN)N or (NNN)NN series could be suppressed for a given (NNN) rare codon. For example, AGGA was the most efficiently suppressed of the AGGN series, and only AGGAU and AGGAC were identified from the Ser70(AGGNN) library. Similarly, all of the five-base codons that were efficiently suppressed by 9 nt anticodon tRNAs have a pyrimidine in the last position of the codon. This may suggest increased difficulty of presenting an mRNA structure that allows pairing with five bases. Therefore, other factors besides the rarity of the codon affect its ability to be suppressed.

All of the efficient suppressors of the identified codons were found to contain an anticodon that allowed Watson-Crick pairing with all the bases of the codon. This phenomenon was observed for three-base, four-base, and five-base codons and clearly implies a direct-reading mechanism for the most efficient suppressors.

The position of the anticodon within the anticodon loop followed a clear pattern (Figure 1). Typically, a codon of *N* bases is best suppressed with an anticodon loop of $N + 4$ bases, wherein two bases are 5′ to the anticodon and two bases are 3′. For undersized anticodon loops (as with the 6 nt loop for suppression of three-base codons), the additional pair is made on the 5′ side of the anticodon; for oversized anticodon loops (as with a 9 nt loop suppression of four-base and three-base codons), the unpaired nucleotides are on the 5′ side of the anticodon. Another way of looking at this is that the anticodon loop nearly always has two nucleotides 3′ to the anticodon, at least for efficient suppressors. This is probably the reason for the extremely high bias for an A37 (or at least purine 37) in all the tRNAs in *E. coli* and nearly all those elicited in our selections; it serves as a

structural "punctuation mark" to demarcate the edge of the anticodon. This requirement is even more stringent than the bias for a U33 (or at least pyrimidine 33), although this is also prominent in tRNA suppressors of all sizes. In anticodon loops of all sizes, the identity of the nucleotides at positions 32 and 38 (the beginning and end of the loop) was anomalous and varied with the codon. The only bias we observed in these positions was against Watson-Crick pairing, especially against C•G or G•C pairs.

Codon Size: Limitations

One of the fundamental questions that underlies the universal genetic code is why all life relies on 20 encoded amino acids, and indeed why the particular twenty were selected. It is often noted that with four bases, only 16 amino acids could be encoded with two-base codons; however, it seems unlikely that this is the reason for the three-base codon since it is well known that proteins can fold into native-like structures with fewer than 20 amino acids [18]. Another possibility is that the redundancy of the genetic code performs a valuable service to organisms with imperfect replication, transcription, and translation systems. Moreover, it is evident that a three-base Watson-Crick interaction is more stable than a two-base interaction, on simple thermodynamic grounds. Whatever reason underlies nature's choice of three-base codons, it is clear that the ribosome, presumably optimized for the three-base codon, is incapable of using two-base codons, at least with tRNAs bearing 6 nt anticodon loops.

At the other end of the spectrum, it is equally clear that four-base and even five-base codons are not excluded by the ribosome or other translational machinery. In fact, at least some four-base codons (AGGN, UAGN) can be read with very high fidelity and nearly as efficiently as three-base codons. If one accepts that the redundancy of 64 codons for 20 amino acids aids fidelity in translation, then why not 256 four-base codons, perhaps with 80 amino acids? It seems likely that part of the reason is that (1) the 20 amino acid repertoire is sufficient for every function that life exhibits, and (2) there is a 33% additional cost in material and time to replicate, transcribe and perhaps translate a signal composed of four-base codons instead of three-base codons. Another part of the answer may be hinted at by the fact that six-base codons are incompatible with the translational machinery and that even five-base codons cannot be suppressed with high efficiency. It may simply be that there are diminishing returns for increasingly long codon-anticodon interactions because of the conformational restrictions of the anticodon loop and the mRNA at the ribosome. Also, the crossreactivity of larger tRNAs is a clear limitation to frame maintenance with larger codons. Whereas 7 nt anticodon loop tRNAs (the natural size) do not read two-base or four-base codons to any appreciable degree, 9 nt anticodon loop tRNAs are capable of reading three-base and four-base codons in addition to five-base codons. Of course, it is difficult to distinguish this from a bias caused by the ribosome against larger codons since the ribosome has not evolved for this purpose.

What we can conclude from our data is that the contemporaneous translational machinery is in fact most efficient at using three-base codons and 7 nt anticodon tRNAs. However, four-base and even five-base codons can be processed by the ribosome with tRNAs containing up to at least 10 nt in their anticodon loops.

Anticodon Size: Limitations

For the codon sizes that are accessible to the translational machinery, there appear to be limits on the sizes of the tRNA anticodons that can decode each type. For example, three-base codons (such as AGG) can be suppressed by tRNAs with six, seven, eight, or nine nucleotides in their anticodon loops, although the interaction with 7 nt anticodon loop tRNAs is favored. Larger codons, however, require larger anticodon loops; four-base codons require 8 or 9 nt loops, and five-base codons require 9 or 10 nt loops. Part of the reason for this is probably steric; in order to "measure out" a larger codon, the anticodon loop probably needs to occlude more of the mRNA. Additionally, all of the efficient suppressors that we have discovered have Watson-Crick base pairing for all of the nucleotides of the codon, even for five-base codons. Thus, the larger anticodon loop is also probably required to present the larger anticodon in such a fashion that it can favorably interact with the mRNA. Structural characterization of the interaction of tRNAs with extended anticodon loops, with and without RNAs bearing complementary codons, could aid greatly in elucidating the mechanism of decoding for various types of codons and anticodons.

It should also be noted that a tRNA anticodon loop of a given size can suppress codons of various lengths. For example, it is known that NCUA anticodons (in 8 nt anticodon loops) can decode either UAGN four-base codons or UAG three-base codons [13, 19]. Our data show that the same 8 nt anticodon loop tRNAs can decode AGG and AGGA and that the same 9 nt anticodon loop tRNA can decode AGG, AGGA, and AGGAC. These tRNAs do seem to have steric preferences in terms of suppression efficiency; 9 nt anticodon loops for five-base codons, 8 nt loops for four-base codons, and 7 nt anticodon loops for three-base codons. We believe that this preference is partially due to an idea proposed by Curran and Yarus [19], namely that the ability to make the maximum number of Watson-Crick pairs between the codon and anticodon leads to the highest suppression efficiency.

Although we did not find any in this study, 7 nt anticodon loop tRNAs that promote +1 frameshifting have been identified elsewhere [20]. In some cases, though, mutations in the D-loop were required for one to see any appreciable level of frameshifting. Since we observed both 6 nt anticodon loops capable of suppressing three-base codons and 8 nt anticodon loops capable of suppressing five-base codons, we suspect that sequence outside of the anticodon loop influences four-base codon suppression by 7 nt anticodon loop tRNAs. It may simply be that our $tRNA^{Ser}$ scaffold is incompatible with this phenomenon, perhaps in D-loop sequence. It may also be that the sensitive phenotypic tests employed to detect these kinds of suppressors in

earlier studies are below the level of detection by our ampicillin selection.

Frameshifting Mechanisms: Slippage Versus Reading

Slippage mechanisms play a role in the inherent fidelity limit of translation. We have observed +1 frameshifting at low levels mediated by the natural set of tRNAs. This frameshifting is apparently stimulated by the favorability of re-pairing in the +1 frame. However, +2 frameshifting is different and only occurred at appreciable levels when an *amber* stop codon was in frame. We believe that the translational pause afforded by the termination signal is what allows this unlikely event to occur. It is perhaps not surprising that +2 frameshifting would be more difficult and rare than +1 frameshifting, and it likewise stands to reason that a more significant pause would be required for the stimulation of this event than a +1 slip. We also tried to identify 8 nt anticodon loop tRNAs that could mediate +2 frameshifting (i.e., read a five-base codon), and many of the identified codons were simply "shifty" by themselves (i.e., contained an in-frame *amber* stop codon). In a single case, we identified a suppressor of AGGGA with an ACUCCUAU anticodon loop. Presumably, this tRNA also functions by a +1 slip; however, because it already reads a four-base codon, the result is a +2 frameshift.

Of course, the slippage mechanism is well-known for certain −1 and +1 frameshifts as well as for more exotic translational bypassing events [4]. In order for these events to occur at an appreciable level, however, other stimulatory elements that cause translational pausing are required (such as mRNA secondary structure that impedes translocation through the ribosome). In fact, our data support a model in which high-level subversion of frame maintenance requires one of two things: a tRNA that is able to make a Watson-Crick pair to an extended codon or, absent this, secondary elements that cause translational pausing. This essentially means that there are two ways to defeat normal frame maintenance, a thermodynamic one and a kinetic one. When a larger codon can be stably recognized, the ribosome is capable of translocating a larger piece of mRNA than it typically does. Alternately, when the normal, rapid events that lead to frame maintenance are slowed, other processes occur on the time-scale necessary to compete. Another way of saying this is that fidelity in frame maintenance is maintained both by thermodynamic "proofreading" (the codon-anticodon interaction) and kinetic proofreading.

Significance

We have demonstrated that the *E. coli* ribosome is capable of using codons of 3 to 5 bases and that the tRNAs that decode these are most efficient with a Watson-Crick complementary anticodon, 2 nt on either side of the anticodon in the loop and U33 and A37 nucleotides. These results highlight a thermodynamic process in determining frame; namely, that efficient Watson-Crick pairing to an *N*-length anticodon is the molecular ruler that measures out *N*-length codons. This also sheds light on the limits of translational fidelity and the molecular evolution of life. Importantly, we have isolated efficient suppressors of four-base and five-base codons, and these suppressors provide unique ways to encode unnatural amino acids in living cells by using "orthogonal" tRNA/synthetase pairs. We are in the process of engineering pairs that decode four-base codons, and we are using these in selections for novel amino acid specificity.

Experimental Procedures

Strains, Plasmids, and Materials

Subcloning was carried out in *E. coli* strains DH10B (Gibco Life Technologies) and TOP10 (Invitrogen), and selections were carried out in TOP10 (which is *rpsL*). PCR reactions were carried out according to standard protocols with Taq (Promega) or Pfu (Stratagene) polymerases. Standard protocols were employed for subcloning with restriction enzymes (NEB) and T4 DNA ligase (NEB or Roche).

Construction of Libraries

β-lactamase reporter libraries were constructed as previously described [9] by cloning a cassette derived from synthetic oligonucleotides into the vectors pBRBstPstXmaKO or pBRS70KO for Ser124 and Ser70 libraries, respectively. The tRNA anticodon loop libraries were also constructed as previously described, by using synthetic oligonucleotides to generate a cassette for cloning into pACGFP.

Selection and Sequencing of Selectants

Selections were carried out with TOP10 *E. coli* cells cotransformed with a reporter library and tRNA library, as previously described [9]. Cells were then subjected to growth on media containing various concentration of ampicillin (typically 5–1,000 μg/ml). For sequencing of reporters and tRNAs, a plasmid separation protocol was carried out before sequencing as described previously [9].

Nitrocefin Assay for Suppression Efficiency

The chromogenic assay for β-lactamase activity was carried out in a 96-well format. Cultures were inoculated from a saturated overnight starter culture and grown to $OD_{600} = 0.5$. The plate was centrifuged, and the cells were resuspended in 500 μl of 50 mM sodium phosphate buffer (pH 7) with 0.5% DMSO and 1 mg/ml deoxycholate. After 2 hr of agitation to obtain a periplasmic lysate, the solution was cleared by centrifugation, and 10 μl of the supernatant was added to 90 μl of 1 mM nitrocefin in the same sodium phosphate/DMSO buffer. The change in absorbance at 486 nm was monitored for 15 min, and the rate of reaction was deduced from the slope. The rates are reported as a percentage of the rate from cells containing β-lactamase with Ser codon at position 70 (i.e., the wild-type sequence).

Acknowledgments

J.C.A. and T.J.M. are National Science Foundation predoctoral fellows. This work was supported by grants from the Department of the Army, the Office of Naval Research, and the Skaggs Research Fund.

Received: July 30, 2001
Revised: November 8, 2001
Accepted: November 13, 2001

References

1. Fox, T.D. (1987). Natural variation in the genetic code. Annu. Rev. Genet. *21*, 67–91.
2. Lehman, N. (2001). Molecular evolution: please release me, genetic code. Curr. Biol. *11*, R63–R66.
3. Farabaugh, P.J. (1996). Programmed translational frameshifting. Annu. Rev. Genet. *30*, 507–528.
4. Gesteland, R.F., and Atkins, J.F. (1996). Recoding—dynamic

reprogramming of translation. Annu. Rev. Biochem. *65*, 741–768.
5. Stadtman, T.C. (1996). Selenocysteine. Annu. Rev. Biochem. *65*, 83–100.
6. Luzopone, C.A., Knight, R.D., and Landweber, L.R. (2001). The molecular basis of nuclear genetic code change in ciliates. Curr. Biol. *11*, 65–74.
7. Atkins, J.F., Weiss, R.B., Thompson, S., and Gesteland, R.F. (1991). Towards a genetic dissection of the basis of triplet decoding, and its natural subversion: programmed reading frame shifts and hops. Annu. Rev. Genet. *25*, 201–228.
8. Kramer, G., Kudlicki, W., and Hardesty, B. (1998). In vitro engineering using synthetic tRNAs with altered anticodons including four-nucleotide anticodons. Methods Mol. Biol. *77*, 105–116.
9. Magliery, T.J., Anderson, J.C., and Schultz, P.G. (2001). Expanding the genetic code: selection of efficient suppressors of four-base codons and identification of "shifty" four-base codons with a library approach in Escherichia coli. J. Mol. Biol. *307*, 755–769.
10. Atkins, J.F., Elseviers, D., and Gorini, L. (1972). Low activity of β-galactosidase in frameshift mutants of *Escherichia coli*. Proc. Natl. Acad. Sci. USA *69*, 1192–1195.
11. Atkins, J.F., Weiss, R.B., and Gesteland, R.F. (1990). Ribosome gymnastics—degree of difficulty 9.5, style 10.0. Cell *62*, 413–423.
12. Qian, Q., Li, J.N., Zhao, H., Hagervall, T.G., Farabaugh, P.J., and Björk, G.R. (1998). A new model for phenotypic suppression of frameshift mutations by mutant tRNAs. Mol. Cell *1*, 471–482.
13. Moore, B., Persson, B.C., Nelson, C.C., Gesteland, R.F., and Atkins, J.F. (2000). Quadruplet codons: implications for code expansion and the specification of translation step size. J. Mol. Biol. *298*, 195–209.
14. Wang, L., Brock, A., Herberich, B., and Schultz, P.G. (2001). Expanding the genetic code of Escherichia coli. Science *292*, 498–500.
15. Biou, V., Yaremchuk, A., Tukalo, M., and Cusack, S. (1994). The 2.9 Å crystal structure of T. thermophilus seryl-tRNA synthetase complexed with tRNA(Ser). Science *263*, 1404–1410.
16. Price, S., Cusack, S., Borel, F., Berthet-Colominas, C., and Leberman, R. (1993). Crystallization of the seryl-tRNA synthetase:tRNAS(ser) complex of Escherichia coli. FEBS Lett. *324*, 167–170.
17. Kyoto Encyclopedia of Genes and Genomics. Release 18.0, April, 2001 (http://www.genome.ad.jp/kegg).
18. Davidson, A.R., and Sauer, R.T. (1994). Folded proteins occur frequently in libraries of random amino acid sequences. Proc. Natl. Acad. Sci. USA *91*, 2146–2150.
19. Curran, J.F., and Yarus, M. (1987). Reading frame selection and transfer RNA anticodon loop stacking. Science *238*, 1545–1550.
20. Tuohy, T.M., Thompson, S., Gesteland, R.F., and Atkins, J.F. (1992). Seven, eight and nine-membered anticodon loop mutants of tRNA(2Arg) which cause +1 frameshifting. Tolerance of DHU arm and other secondary mutations. J. Mol. Biol. *228*, 1042–1054.