Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 4587–4591, July 1985
Biochemistry

Initiator tRNA may recognize more than the initiation codon in mRNA: A model for translational initiation

($tRNA_i$ structure/gene starts/initiation mutants/evolution of the genetic code)

M. C. GANOZA, P. MARLIERE, E. C. KOFOID, AND B. G. LOUIS

Banting and Best Department of Medical Research and University of Toronto, Toronto, ON, Canada

Communicated by J. Tuzo Wilson, April 22, 1985

ABSTRACT **A special methionine tRNA ($tRNA_i$) is universally required to initiate translation. Amongst species a $tRNA_i$ structural conservation is most apparent in the anticodon and T arms of the molecule but extends into the variable loop and the 3′ strand of the D stem. This suggested that they could share a similar ancestral or current function in initiation of translation. We report that the sequence of bases neighboring the translational start codons of many eubacterial genes are complementary not only to the extended anticodon but also to the D and T loops of $tRNA_i$. Study of the coding properties of $tRNA_i$ and of mutations that affect translation suggests that the translational start domain can be a mosaic of signals complementary to the loops of $tRNA_i$. The hypothesis of multiple loop recognition suggests that unusual triplets can start prokaryotic and mitochondrial genes and predicts the occurrence of other reading frames. Furthermore, it suggests a unifying model for chain initiation based on RNA contacts and displacements.**

An unresolved and perhaps universal issue about the genetic code is the ambiguity of the initiator codon AUG (and occasionally UUG and GUG), which not only specifies initiation of a polypeptide but also codes for internal amino acids. These codons also occur randomly, out of phase, in mRNAs. A unique secondary or tertiary structure of the messenger could specify initiation, but such a structure has not been observed in most mRNAs examined (1, 2). In prokaryotes, the specificity of initiation relies partially on recognition of regions complementary to 16S rRNA (the Shine–Dalgarno or SD region) (3). Sequence analysis and biochemical and genetic data support this hypothesis (1–4). However, (*i*) other nonstart regions in mRNAs also complement the 16S rRNA (2, 5), (*ii*) no simple correlation exists between the degree of complementarity with the 16S rRNA and the efficiency of translation (2), (*iii*) at least 10 normal *Escherichia coli* proteins lack this region in the prestart sequence (2, 6–11), and (*iv*) most eukaryotic mRNAs lack such signals and instead frequently initiate synthesis at the AUG closest to the 5′ terminus (12). A 5′-terminal 7-methylguanine is essential for the efficient initiation of many of these mRNAs, but, as with the SD region, it is not always sufficient or necessary (12). Certain eukaryotic mRNAs have long leader sequences (see ref. 13 for examples) where the translation apparatus must recognize the start site amidst many nonstart and out-of-phase homonymous triplets.

We have, therefore, searched for the one reactant for initiation that has both constancy through evolution and flexibility in recognizing the variety of observed start signals. We propose it to be initiator tRNA ($tRNA_i$; frequently referred to as $tRNA_f$) and suggest that a variety of interactions between it and mRNA may underlie much of the specificity of initiation.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Evolutionary Conservation and Special Features of $tRNA_i$ Structure

A special methionyl-tRNA is universally required to initiate protein synthesis (14). Phylogenetic studies and heterologous aminoacylation, formylation, and translation experiments indicate that initiators are the most highly conserved family of tRNAs (15–17). tRNAs, from eubacteria (including chloroplasts) have remarkable sequence invariance (18, 19).

The hypothesis that the anticodon and T arms emerged from an internal duplication in an ancestral tRNA molecule has been proposed by Woese (20).

This high degree of $tRNA_i$ sequence conservation suggests that initiation of protein synthesis may exhibit similar features in eubacteria and eukaryotes.

Features of $tRNA_i$ other than the anticodon loop could be important in discriminating initiation signals. For example, there are subtle differences in tertiary structure of the unpaired regions of *E. coli* $tRNA_i$ and yeast $tRNA^{Phe}$ (21, 22). Other features distinguishing initiators from elongators are as follows: (*i*) The absence of modified bases in the anticodon loop that could permit additional pairing. (*ii*) The D loop contains only one dihydrouridine, leaving seven nucleotides around the CCU region available for pairing. These have been shown to bind complementary oligonucleotides (23). (*iii*) T and Ψ residues pair as uridine, allowing potential pairing of seven bases in the T loop.

The natural tertiary link between the T and D loops can be disrupted by the codon–anticodon association, making the loops available for intermolecular contacts (24). Given these properties of $tRNA_i$, we asked whether signals complementary to the D and T loops as well as the anticodon loop neighbor translational start codons.

Potential Recognition of D, T, and A Loops by mRNAs

If regions of $tRNA_i$ other than the anticodon are recognized during initiation, one expects the complementary sequences to neighbor the translational start codon. To determine this, the sequences of nucleotides of 268 genes with 34 nucleotides on each side of the start codon were computer-searched for regions able to pair with D, T, or extended anticodon (A) loops of $tRNA_i$ or with the SD region of rRNA.

Pairings with five or more complementary bases in mRNA are compiled in Fig. 1 as a function of position along the mRNA sequence. We find that the regions surrounding the initiation codons of the prokaryotic genes examined are rich in potential sites of interaction with the A, D, and T loops of $tRNA_i$. In addition, the start codon itself is frequently embedded in stretches of larger sequence homology to the

Abbreviation: SD, Shine–Dalgarno.

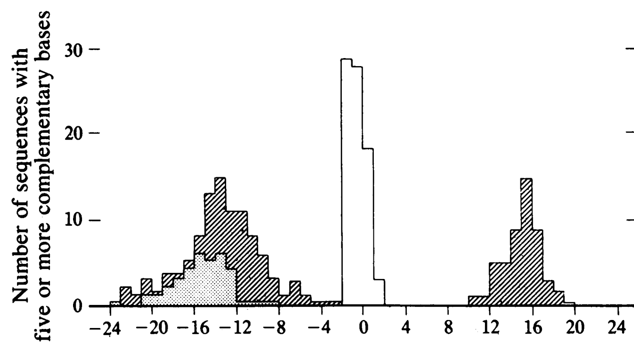

FIG. 1. Histogram showing the location of signals in start sequences complementary to $tRNA_i$. The sequences of 268 different prokaryotic genes, each consisting of the start AUG at position 0 and 34 nucleotides on either side, were searched for signals complementary to five or more nucleotides of the extended anticodon (open bars), T signals (stippled bars), and D signals (hatched bars) of $tRNA_i$ as described in the text.

anticodon loop. We name the sequences complementary to the A (3′-AAUACUC-5′), D (3′-ADGGUCCGA-5′) (D, 5,6-dihydrouridine), or T (3′-UAAACΨT-5′) loops of $tRNA_i$ in the vicinity of a start codon A, D, or T signals. The 3′ or 5′ prefix refers to their position relative to the start codon. Position numbers are 5′ to 3′, with the central codon beginning at 0.

There is a concentration of A signals at the start position. The 5′ T signals peak around position −14, and D signals occur predominantly at positions −13 and +15 (Fig. 1).

Occurrences of the various signals at positions on the mRNA consistent topologically with an interaction with $tRNA_i$ (D or T loop at positions −25 to −7 or +7 to +25 and A signal, including the anticodon) were counted. The totals were compared by the χ^2 test with expected values that were calculated from the proportions of nucleotides in the library on the null hypothesis of random nucleotide distribution: $E = PT$, in which E is the expected number of matches in the total T of sequences, and P is the probability of matching at least one of the signals of length n or more somewhere in the range (k, l) of permitted start positions for one sequence. To calculate P we first obtained values for P_i, $i = 1$ to n, the probability of finding a pairing nucleotide at a given location. These P_i values were calculated on the assumption of random distribution of the nucleotides in the library. Given the P_i, we calculated

$$P = 1 - \left[\prod_{j=1}^{m}\left(1 - \prod_{i=1}^{n} P_i\right)\right]^{l-k+1},$$

in which m is the number of sequences in the library.

The data were compared to 268 71-nucleotide prokaryote sequences that were centered on matching in-frame AUG or GUG codons in the coding portions of different mRNAs. A similar comparison was made of 119 eukaryote start domains and 119 matching internal sequences.

Prokaryote Signals: SD Interactions

Many prokaryote gene starts had SD signals ($P = 3.08 \times 10^{-8}$ for five base pairs or more). By contrast (Table 1), the prokaryotic internal sequences displayed a marked avoidance of the SD signal on the 5′ side of internal AUG codons. This avoidance was absolute for pairings of six nucleotides or more, and the numbers of length − 5 ($P = 3.75 \times 10^{-4}$) and length − 4 ($P = 1.31 \times 10^{-6}$) matches were well below those predicted under the random distribution hypothesis. Nine of 268 start sequences that lacked a SD signal or 37 that had this signal in more than one position of the 71-base-pair window had, in each case, a D, T, or A signal. Thus, signals complementary to $tRNA_i$ could be ancillary to the SD signal.

A Signal. Of the 268 prokaryotic start sequences examined, 214 contained 4 or more bases complementary to the 5′-extended anticodon ($P = 4.41 \times 10^{-7}$) (Table 1). Of the 268 starts 239 were AUGs, and 72 of these were extended to UUAUG or UAUGA. There were 207 nonstart AUGs in positions other than 0 in this same library, and of these only 41 were extended to UUAUG or UAUGA ($P = 3 \times 10^{-4}$) (data not shown). There was a nearly random distribution of bases neighboring internal AUG (Table 1). It is thus quite possible that prokaryotic ribosomes start preferentially at UAUG or UUAUG. Coding data for $tRNA_i$, discussed below, support this suggestion. Transcripts of phage λ and phage 434 *cI* that lack the ribosome binding site retain substantial complementarity to the anticodon loop. This could account for the residual translation of the messages (2, 6, 9).

D Signal. In prokaryotes, the 5′ and 3′ D signals were found in both the start and the internal sequences. D signals peaked at −13 and +15 (cf. Fig. 1) in the start sequences but were scattered uniformly in the internal sequences. This discrete distribution was observed even though there were no more D signals in the relevant areas (7 through 24) of the start regions than would be expected if chance governed the nucleotide distribution.

T Signal. The T signal, on the other hand, was significantly more common in the prestart region ($P = 3.04 \times 10^{-7}$ for six bases) than in the internal coding region ($P = 0.84$ for six bases) (Table 1). Though relatively scarce, several very strong (seven- to eight-base-pair) T signals were observed at position 12 of start sequences. A rough correlation was observed between the abundance of gene products and the length of the T signal—e.g., *tsf, rplJ, rplL*, and the *tufA* and *tufB* cistrons are coordinately transcribed (25), but only the last three messages have a long T signal. The coat protein cistrons are the best expressed RNA phage genes (1) and are the richest in T loop complementarity. A sequence complementarity of the phage Qβ coat protein and the T loop was noted by Gupta *et al.* (26).

Given these results, an identical distribution and statistical analysis of 119 eukaryotic sequences was carried out. These preliminary data are summarized below.

Eukaryote Signals: SD Interactions

Polypyrimidine tracts abound in the 5′ prestart region of the eukaryotic genes examined. Some of these tracts have been assigned as points of potential contact with the 3′-UAGGAAGGCGU-5′ region of 18S rRNA (27), but they do not occur with statistical significance in 119 genes examined (data not shown and ref. 28).

A Signal. A significant number of four-base-pair A signals were observed. Seventy percent of the sequences started with AUGG, the tetranucleotide complementary to the 5′ extended anticodon of $tRNA_i$, whereas only half the predicted number of spurious A signals appeared within internal sequences ($P = 3 \times 10^{-5}$ for four nucleotides or more). Unlike the case in prokaryotes, the 5′ A signal does not occur with significant frequency. We note that a base modification of the eukaryotic $tRNA_i$ (A to *N*-[(9-β-D-ribofuranosylpurine-6-yl)carbamoyl]threonine) located 3′ to the anticodon may restrict the latter from base pairing (29), precluding an A signal.

T Signal. The T signal does not seem significant in the eukaryotic sequences examined. Again, the T loop of vertebrate $tRNA_i$ contains 1-methyladenosine rather than A. The modification would weaken the base-pairing interactions.

Table 1. Distribution of signals in prokaryotes

Sequences	Length of signals		A	5′ D	5′ T	SD	3′ D	3′ T
Start	4	Exp.	158	213	125	91.1	213	125
		Obs.	214	220	135	171	225	128
		P	4.41×10^{-7}	0.294	0.201	4.38×10^{-9}	0.0675	0.682
	5	Exp.	69.9	85.2	34.6	21.9	85.2	34.6
		Obs.	72	83	44	61	88	38
		P	0.595	0.768	0.0821	3.08×10^{-8}	0.717	0.540
	6	Exp.	15.1	24.1	8.90	3.64	24.1	8.90
		Obs.	22	2	30	20	18	8
		P	0.0650	0.513	3.04×10^{-7}	3.56×10^{-8}	0.189	0.758
Internal	4	Exp.	157	223	123	114	223	123
		Obs.	178	239	126	63	240	135
		P	0.0102	0.00884	0.693	1.31×10^{-6}	0.00569	0.129
	5	Exp.	71.8	100	33.8	30.8	100	33.8
		Obs.	62	111	31	11	93	26
		P	0.172	0.171	0.617	3.75×10^{-4}	0.636	0.149
	6	Exp.	15.4	31.1	8.44	5.60	31.1	8.44
		Obs.	13	40	9	0	31	6
		P	0.543	0.0854	0.840	0.0161	0.984	0.602

mRNA sequences of 71 nucleotides, 268 centered on initiator codons (Start) and 268 centered on internal codons of the same sequence (Internal), were examined with the aid of a computer program. The program identified regions capable of pairing the extended anticodon, the D loop, the T loop, and the SD sequence. Pairing between G and U was allowed. Complementarity of contiguous pairs was sought in parts of the sequence, the locations of which were specified relative to the 5′ nucleotide of the control codon with position numbers increasing from 5′ to 3′. The probability of finding such a complementarity within the specified range, on the assumption of a random distribution of nucleotides about the central codon, appears for each signal in *P* rows. These probabilities take into account the different frequencies of occurrence of the nucleotides in start and internal sequences. The number of sequences expected (Exp.) to possess the sought-for complementarity is compared with observed values (Obs.) by the χ^2 test. Probabilities below 1% are considered to give reasonable support to rejection of the null hypothesis. For start sequences, 17,959 nucleotides were searched, of which 5344 were A, 3731 were C, 4687 were U, and 4197 were G. For internal sequences, 18,957 nucleotides were searched, of which 4724 were A, 4442 were C, 4763 were U, and 5028 were G.

D Signal. For the 119 eukaryote starts, the 5′ D signal was the only one found in significant numbers ($P = 1.11 \times 10^{-4}$ for six nucleotides or more).

The fact that eukaryotic start domains conserve D and 3′ A signals suggests that these sequences may underlie a primitive mechanism for the initiation of protein synthesis. Alternatively, these signals could transiently interact with the $tRNA_i$ during the earliest stages of translation, favoring the association of the active ribosome complex. Observations outlined below support this hypothesis.

Coding Properties of $tRNA_i$

In prokaryotes, AUG, GUG, UUG (2, 30–35), and, in one gene, AUU (36) function as initiator codons *in vivo*. *In vitro* both GUG and AUG code for fMet-tRNA. Surprisingly, $tRNA_i$ binds to ribosomes with poly(U), but not poly(A), and the complex so formed can be chased into fMet-$(Phe)_n$ (37). The poly(U) coding can be explained by interactions of the T loop with UUU in the mRNA. Remarkably, the isolated anticodon of $tRNA_i$ binds to both AUG and GUG, while fragments of $tRNA_i$ containing the T and anticodon arms bind UUG (38). Thus, the T signal in the fragment could have paired with UUG acting as an anticodon.

Jay *et al.* reported that prokaryotic initiation complexes programmed with d(AATTCTAGGATTTAATCCATG) and d(AATTCTAGGATTTAATC) stimulated binding of fMet-tRNA equally well (39). This suggests that fMet-tRNA can bind through the T signal (overlined) contained in both polymers, even when the ATG (underlined) is changed to ATC. In addition, the T loop is the only $tRNA_i$ region able to bind to AUU, the start of the initiation factor 3 gene (36). Furthermore, T loop recognition could account for anomalous coding properties of $tRNA_i$ if, in fact, the T loop acts as an anticodon.

UAUG promotes binding of the *E. coli* $tRNA_i$ more efficiently than any other tetranucleotide and preferentially stimulates dipeptide synthesis directed by heptanucleotides bearing a 3′ coding region (40–43). Furthermore, UUAUG is the most efficient pentanucleotide (40). UAUG or UUAUG could pair with the CAUAA anticodon of $tRNA_i$. Thus, an extended anticodon could be used for gene start designation (40–44).

Table 2 shows that a number of prokaryotic genes begin with UUG and one starts with AUU (30–36). In each case, four or more bases complement the T loop of $tRNA_i$. Quite

Table 2. Unusual initiation triplets predicted by the model

Species	T signal (3′-5′)	Predicted start triplet	Genes with observed start triplet
Prokaryotes	UAAACΨT	UUG	*lacA*, NDH, S20, *lacI* restart proteins, T4 *rIIB* restart proteins
		AUU	Initiation factor 3
Beef and *Drosophila* mitochondria	CAUAUUU	UAU	
		AUA	Bovine *URF-2, URF-3, URF-5*
		UAA	*Drosophila* cytochrome oxidase subunit I
Human mitochondria	CAUA–UU	UAU	
		AUA	*URF-1, URF-3, URF-5*
		UAA	
Yeast mitochondria	UAAACUU	AUU	*URF-6L*

T stands for 5-methyluridine. The references for the observed start triplets are *lacA* (30), diphosphopyridine nucleotide dehydrogenase (NDH) (31), ribosomal protein S20 (32), *lacI* restart proteins (45), phage T4 *rIIB* restart protein (35), initiation factor 3 (36), and mitochondrial genes (46–48) (*URF*, unassigned reading frame).

remarkably, several mitochondrial genes begin with AUA, UAA, or AUU (46–48), and each exhibits up to six bases of complementarity with the T loop of the mitochondrial tRNA$_i$ (see Table 2 for examples).

Prokaryotic Mutants Defective in A, D, and T Signals

A number of mutations whose effects on translation have remained unexplained can be rationalized in terms of the proposal that signals complementary to tRNA$_i$ modulate initiation. Several that alter the initial AUG abolish translation as expected (2, 49–51). However, at least four mutants of this type do not shut off protein synthesis completely: Qβ coat protein (41), *trpE* (49), T4 *rIIB* (52, 53) (see Table 3), and *lacZ* (54, 55). In these the AUG codon is changed to AUA, resulting in lower but significant initiation at the correct site, consistent with the use of an extended anticodon (40–44).

Recent experiments by Munson *et al.* (54) suggest that, depending on the base-pairing potential of the translational start domain, when two AUGs occur in the 5′ prestart region, the one able to pair with the extended anticodon of tRNA$_i$ is the preferred start codon.

A second set of mutations affecting the SD interactions affect protein synthesis as expected. The sequences of some of these have been summarized by Gold *et al.* (ref. 2; see also refs. 50 and 51). Since the SD and D signals have bases in common, mutations altering the SD region can overlap potential D signals. In the little tumor antigen gene, *t*, of simian virus 40 in plasmid pTR436 (Table 3) a two-base substitution from GAAA to GCCA increases expression (2). Note that the underlined sequence can pair with the 16S rRNA while the overlined sequence cannot; CCA is part of the D signal (Table 3). Similarly, in the *lacZ* gene, mutants affecting the D signal CAG alter *lacZ* expression (54).

Other mutations between the SD domain and AUG (2, 55) have remained unexplained—e.g., changes in the *his* leader peptide of *Salmonella typhimurium* that alter the AUUU T signal decrease expression (Table 3). Also, longer transcripts of *galE*, containing AUUUC, translate 4 times better than shorter transcripts lacking this sequence (56). The GUUU sequence (57), a T signal, when inserted 5′ to the simian virus 40 *t* gene enhanced its expression in prokaryotic cells 10-fold.

Table 3. Effect of mutations in D, T, and A signals on the efficiency of protein synthesis

Gene	Sequence	Expression, %	Ref.
	tRNA$_i$ A loop 3′-CAAUACUCG-5′		
trpE	ACAAUGAAA	100	49
	ACAAU*A*AAA	26	
Qβ coat protein	CAUGG	100	41
	CAUG*A*	280	
	CAU*A*G	10	
	CAU*AA*	33	
T4 *rIIB*	AUUAUGUA	100	51
	AUUAU*A*UA	30	
	tRNA$_i$ D loop 3′-ADGGUCCGA-5′		
t (pTR436)	GAAAGAUGGAUA		2
	G*CC*AGAUGGAUA	↑	
	tRNA$_i$ T loop 3′-UAAACΨT-5′		
his leader	AUUUUUAUGACAC		2
	AU*G*UUUAUGACAC	↓	
	AU*GC*UUAUGACAC	↓	
T4 *rIIB* restart X	UUUUGAAAAAUGCUGA		51

The efficiency of translation is given in percent or as ↑ for increased or ↓ for decreased protein synthesis. The latter symbols represent a minimum of 4-fold changes in expression. Base changes are indicated by italics. Regions of D or T loop complementarity and the start codon are underlined.

Striking mutations occur when premature termination results in reinitiation of translation. In the T4 *rIIB* protein, 5 of 11 functional restart sites began with UUG (51) and invariably contained regions of extended complementarity with the T loop. Two restarts, X (Table 3) and 360, had no SD signal but had T and D signals, respectively. Three of the remaining restart sites contained four bases that could be SD or D signals. No single mutant of this class obliterated a T signal (2, 51). In view of the above, these results best fit the idea that a T signal can be a critical determinant of initiation.

It is tempting to fit the sequences of many of these mutants into differing arbitrary potential secondary structures. However, alternative structures can usually be drawn, and even if these were responsible, one expects ribosomes to disrupt secondary structures to initiate synthesis (1, 2).

Implications for the Start of Protein Synthesis

We propose that the code to initiate protein synthesis is more flexible than previously thought. This proposal is based on the ability of mRNA to complement one of the three tRNA$_i$ loops in addition to or instead of the SD 16S rRNA region. The prestart regions of most messengers are, in fact, mosaics of sequences complementary to 16S rRNA and one or more tRNA$_i$ loops (Fig. 1, Table 1).

The D and T signals peak at about 14 bases from the initiator codon. Approximately 50 Å span the D or T loop and the anticodon in the tRNA$_i$ molecules (21, 22). Thus, functional tRNA$_i$ mRNA contacts involving D and T loops are sterically possible.

In prokaryotes, signals are accommodated 3′ or 5′ to the start codon, which is consistent with the finding that initiation efficiency can be affected by bases 5′ or 3′ to the initiator (41, 44). If the signals are functional, then the 30S subunit could support a tRNA$_i$ scan for a 3′ or 5′ signal prior to peptide bond synthesis. Data for a scanning mechanism by 40S particles are documented (12). Accordingly, in both eukaryotes and prokaryotes codon–anticodon association could be lengthened or absent altogether. Interactions of 16S rRNA could also occur (58).

If these RNA–RNA contacts confer specificity, then there could be a mechanism for promoting or displacing them so that translation can continue. Reported interactions of the 5′ end of the tRNA$_i$ with the 23S rRNA (59) or the 16S rRNA (58), or of the T arm of tRNA$_i$ with 16S rRNA (60), could play such a role.

We propose that the association of fMet-tRNA$_i$, 30S particle, and mRNAs may involve interactions with signals present in the latter, leading to many configurations, each dependent on the sequence of the message. The ability of such aggregates to reorganize into the proper configuration may determine the efficiency of initiation. The conversion could occur upon association of the ribosomal subunits, with 23S RNA displacing the mRNA from the lateral arms of fMet-tRNA$_i$. An overall increase in pairing stability could drive such rearrangements. Thus, entering elongation would require energy. Consistent with this is the fact that peptide bond synthesis cannot occur when initiation complexes are formed with nonhydrolyzable analogues of GTP (61). An additional function of initiation factors and ribosomal proteins could be to promote or destabilize RNA–RNA contacts. This could explain the necessity for initiation factor-2-catalyzed GTP hydrolysis in activation of ribosome-bound tRNA$_i$ (62). Alternatively, an early mRNA–tRNA$_i$ interaction may occur independent of any proteins.

The model could be tested by analyzing the effects of judicious tRNA$_i$ modifications on the specificity and efficiency of protein chain initiation.

Depending on the extent to which the T loop of $tRNA_i$ can pair UUG, AUU, or UUU, the hypothesis of loop recognition predicts that the anomalous coding properties of $tRNA_i$ are used to start the synthesis of perhaps many as-yet-undiscovered genes (see Table 2 for discovered examples). Alternatively, the conservation of such RNA pairing schemes in diverse phyla could reflect an ancient mechanism of translational initiation that utilized flexible RNA–RNA interactions to ensure the preservation of the reading frame of genes amidst adaptive evolutionary constraints.

Note Added in Proof. An alternative model for the AUU start in initiation factor 3 has been proposed by Gold et al. (63).

We thank Dr. Michel Kabay for help with statistical analysis of the data and many colleagues, particularly Drs. A. J. Becker and B. Davis, who offered helpful discussions and criticism of the manuscript. We are also grateful to Drs. U. RajBhandary, S. Bayley, S. Nishimura, Y. Kuchino, and J. Rabinowitz for making tRNA sequences available to us. This work was supported by a Medical Research Council grant to M.C.G.

1. Steitz, J. A. (1979) in *Biological Regulation and Development*, ed. Goldberger, R. F. (Plenum, New York), Vol. 21, pp. 349–399.
2. Gold, L., Pribnow, D., Schneider, T., Shinedling, S., Singer, B. S. & Stormo, G. (1981) *Annu. Rev. Microbiol.* **35**, 365–403.
3. Shine, J. & Dalgarno, I. (1975) *Nature (London)* **254**, 34–38.
4. Neilson, T., Kofoid, E. C. & Ganoza, M. C. (1980) *Nucleic Acids Res. Symp. Ser.* **7**, 313–323.
5. Taniguchi, T. & Weissman, C. (1979) *J. Mol. Biol.* **128**, 481–500.
6. Pirotta, V. (1979) *Nucleic Acids Res.* **6**, 1495–1508.
7. Beck, E., Sommer, R., Auerswald, E. A., Kurz, C., Zink, B., Osterburg, G., Schaller, H., Sugimoto, K., Sugisaki, H., Okamoto, T. & Takanami, M. (1978) *Nucleic Acids Res.* **5**, 4495–4503.
8. Ineichen, K., Shepherd, J. C. & Bickel, T. A. (1981) *Nucleic Acids Res.* **9**, 4639–4653.
9. Ptashne, M., Backman, K., Humayun, M. Z., Jeffrey, A., Maurer, R., Meyer, B. & Sauer, T. (1976) *Science* **194**, 156–161.
10. Godson, G. N., Barrell, B. G., Staden, R. & Fiddes, J. C. (1978) *Nature (London)* **276**, 236–247.
11. Zurawski, G., Gunsalus, R. P., Brown, K. D. & Yanofsky, C. (1981) *J. Mol. Biol.* **145**, 47–73.
12. Kozak, M. (1981) *Curr. Top. Microbiol. Immunol.* **93**, 81–123.
13. Kozak, M. (1984) *Nucleic Acids Res.* **12**, 857–872.
14. Gauss, D. H. & Sprinzl, M. (1981) *Nucleic Acids Res.* **9**, r1–r23.
15. Kuchino, Y., Ihara, M., Yabusaki, Y. & Nishimura, S. (1982) *Nature (London)* **298**, 684–685.
16. Wrede, P., Woo, N. H. & Rich, A. (1979) *Nature (London)* **76**, 3289–3293.
17. Cedergren, R. J., Sankoff, D., Larue, B. & Grosjean, H. (1981) *Crit. Rev. Biochem.* **11**, 35–104.
18. Larue, B., Cedergren, R. J., Sankoff, D. & Grosjean, H. (1979) *J. Mol. Evol.* **14**, 287–300.
19. Marliere, P. (1983) Dissertation (Pasteur Institute, Paris).
20. Woese, C. R. (1980) in *Ribosomes*, eds. Chambliss, G., Craven, G. R., Davies, J., Davis, K., Kahan, L. & Nomura, M. (University Park Press, Baltimore), pp. 357–373.
21. Woo, N. H., Roe, B. A. & Rich, A. (1980) *Nature (London)* **286**, 346–351.
22. Schevitz, R. W., Podjarny, A. D., Krisnamachari, N., Hughes, J. J. & Sigler, P. B. (1979) *Nature (London)* **278**, 188–190.
23. Uhlenbeck, O. C. (1972) *J. Mol. Biol.* **65**, 25–41.
24. Rich, A. & RajBhandary, U. L. (1976) *Annu. Rev. Biochem.* **45**, 805–860.
25. Nomura, M., Morgan, E. A. & Jaskunas, S. R. (1977) *Annu. Rev. Genet.* **11**, 297–347.
26. Gupta, S. L., Chen, J., Schaefer, L., Lengyel, P. & Weissman, S. M. (1970) *Biochem. Biophys. Res. Commun.* **39**, 883–888.
27. Hagenbuchle, O., Santer, M. & Steitz, J. A. (1978) *Cell* **13**, 551–563.
28. DeWachter, R. (1979) *Nucleic Acids Res.* **7**, 2045–2054.
29. Sherman, F., McKnight, G. & Stewart, J. W. (1980) *Biochim. Biophys. Acta* **609**, 343–346.
30. Buchel, B. E., Gronenborn, B. & Muller-Hill, B. (1980) *Nature (London)* **283**, 541–545.
31. Young, I. G., Rogers, B. L., Campbell, H. D., Jaworawski, A. & Shaw, D. C. (1981) *Eur. J. Biochem.* **116**, 165–170.
32. Mackie, G. A. (1981) *J. Biol. Chem.* **256**, 8177–8182.
33. Roy, A., Haziza, C. & Danchin, A. (1983) *EMBO J.* **2**, 791–797.
34. McLaughlin, J. R., Murray, C. L. & Rabinowitz, J. C. (1983) *J. Biol. Chem.* **256**, 11283–11291.
35. Pribnow, D., Sigurdson, D. C., Gold, L., Singer, B. S. & Napoli, C. (1981) *J. Mol. Biol.* **149**, 337–376.
36. Sacerdot, C., Fayet, G., Dessen, P., Springer, M., Plumbridge, J. A., Grunberg-Manago, M. & Blanquet, B. L. (1982) *EMBO J.* **1**, 311–315.
37. Van Der Laken, K., Baker, Steeneveld, H. & Van Knippenberg, P. (1979) *FEBS Lett.* **100**, 230–234.
38. Dube, S. K., Rudland, P. S., Clarck, B. F. C. & Marcker, K. A. (1969) *Cold Spring Harbor Symp. Quant. Biol.* **34**, 161–166.
39. Jay, E., Seth, A. K. & Jay, G. (1980) *J. Biol. Chem.* **255**, 3809–3812.
40. Ganoza, M. C., Fraser, A. R. & Neilson, T. (1978) *Biochemistry* **17**, 2769–2775.
41. Taniguchi, T. & Weissman, C. (1978) *J. Mol. Biol.* **118**, 533–565.
42. Ganoza, M. C., Sullivan, P., Cunningham, C. C., Kofoid, E. C. & Neilson, T. (1982) *J. Biol. Chem.* **257**, 8228–8232.
43. Eckhardt, H. & Luhrmann, R. (1981) *Biochemistry* **20**, 2075–2080.
44. Manderschied, U., Bertram, S. & Gassen, H. G. (1978) *FEBS Lett.* **90**, 162–166.
45. Steege, D. A. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 4163–4167.
46. Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H. L., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, R. P. H., Smith, A. J. H., Standen, R. & Young, I. G. (1981) *Nature (London)* **290**, 457–465.
47. de Bruijn, M. H. L. (1983) *Nature (London)* **304**, 234–240.
48. Bibb, M. J., Van Etten, R. A., Wright, C. T., Walberg, M. W. & Clayton, D. A. (1981) *Cell* **26**, 167–180.
49. Zurawski, G., Elseviers, D., Stauffer, G. U. & Yanofsky, C. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 5988–5992.
50. Dunn, J. J., Buzash-Pollert, E. & Studier, F. W. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 2741–2745.
51. Napoli, C., Gold, L. & Singer-Swebilius, B. (1981) *J. Mol. Biol.* **149**, 433–449.
52. Belin, D., Hedgpeth, J., Selzer, G. B. & Epstein, R. H. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 700–704.
53. Singer, B. S. & Gold, L. (1976) *J. Mol. Biol.* **103**, 627–646.
54. Munson, L. M., Stormo, G. D., Niece, R. L. & Reznikoff, W. S. (1984) *J. Mol. Biol.* **177**, 663–683.
55. Hui, A., Hayflick, J., Dinkelspiel, K. & de Boer, H. A. (1984) *EMBO J.* **3**, 623–629.
56. Queen, C. L. & Rosenberg, M. (1980) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **39**, 810 (abstr.).
57. Jay, E., Seth, K. A., Rommens, J., Sood, A. & Jay, G. (1982) *Nucleic Acids Res.* **10**, 6319–6323.
58. Thompson, J. F. & Hearst, J. E. (1983) *Cell* **33**, 19–24.
59. Dahlberg, J. E., Kintner, C. & Lund, E. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 1071–1075.
60. Cunningham, C. & Ganoza, M. C. (1984) *Mol. Biol. Rep.* **10**, 115–121.
61. Lucas-Lenard, J., Tao, P. & Haenni, A. L. (1969) *Cold Spring Harbor Symp. Quant. Biol.* **34**, 455–462.
62. Thach, S. S. & Thach, R. E. (1971) *Proc. Natl. Acad. Sci. USA* **68**, 1791–1795.
63. Gold, L., Stormo, G. & Saunders, R. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 7061–7065.