Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 3616–3620, June 1985
Biochemistry

Bacterial peptide chain release factors: Conserved primary structure and possible frameshift regulation of release factor 2

(protein synthesis/DNA sequence/readthrough)

WILLIAM J. CRAIGEN*†, RICHARD G. COOK*‡, WARREN P. TATE§, AND C. THOMAS CASKEY*†¶

*Howard Hughes Medical Institute and Departments of †Biochemistry, of ‡Microbiology and Immunology, and of ¶Medicine, Baylor College of Medicine, Houston, TX 77030; and §Department of Biochemistry, University of Otago, Dunedin, New Zealand

Communicated by Joan Steitz, February 11, 1985

ABSTRACT *Escherichia coli* **peptide chain release factors are proteins that direct the termination of translation in response to specific peptide chain termination codons. The mechanisms of codon recognition and peptidyl-tRNA hydrolysis are unknown. We have characterized the genes encoding release factor 1 (RF-1) and release factor 2 (RF-2) to study the structure–function relationships of the proteins and their regulation in the bacterium. In this report, we present the gene structure of RF-1 and RF-2, and a partial peptide sequence of RF-2. RF-1 and RF-2 are highly homologous in their primary structure. In addition, an in-frame premature opal (UGA) termination codon is located within the RF-2 coding region at amino acid position 26. This region of the protein was sequenced by automated Edman degradation to confirm the predicted reading frame, and a second independent isolate of the RF-2 gene was identified and sequenced to confirm the DNA sequence. These results imply that a frameshift occurs prior to the premature termination codon, thus allowing for translation of RF-2 to be completed. This may represent a mechanism of translational control of RF-2 expression. An alternative possible means of translational regulation is discussed.**

Polypeptide chain termination is the final step in protein biosynthesis resulting in peptidyl tRNA hydrolysis to release the nascent polypeptide from the ribosome. In bacteria, two codon-specific protein factors participate in the termination event. Release factor 1 (RF-1) recognizes the termination codons UAG and UAA, and release factor 2 (RF-2) recognizes UGA and UAA. The details of termination codon recognition and hydrolysis of peptidyl tRNA are unresolved.

Three activities associated with RF-1 and RF-2 can be examined: ribosome binding, codon recognition, and peptidyl tRNA hydrolysis (1). The functional similarities of RF-1 and RF-2 suggest that they may have structural similarities. Indeed, antibody binding studies have demonstrated several shared structural determinants (2). Earlier attempts at structural studies were thwarted by the small quantities of purified RFs available and by insensitive peptide sequencing techniques. The genes encoding RF-1 and RF-2 have recently been isolated (3, 4), and in this report we have characterized these genes by DNA sequencing. We have also sequenced the amino-terminal region of the RF-2 protein to confirm the correct reading frame. We show that the RF-1 and RF-2 proteins maintain remarkable sequence conservation. In addition, these structural studies suggest that the genes may be under translational regulation.

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.

METHODS

Materials. Restriction enzymes, polymerase I (Klenow fragment), T4 polynucleotide kinase, T4 DNA ligase, 17-mer oligonucleotide sequencing primer, and bacterial alkaline phosphatase were purchased from either Bethesda Research Laboratories or New England Biolabs and used according to the supplier's recommendations. Radionucleotides were purchased from Amersham.

DNA Sequencing. DNA sequencing was performed as described by Maxam and Gilbert (5) or Sanger *et al.* (6), as modified by Biggin *et al.* (7). Sequence analysis was carried out using the Messing program for an Apple computer (R. Larson and J. Messing, University of Minnesota, Apple II sequence analysis programs).

Amino Acid Sequencing. Approximately 200–400 pmol of the purified RF-2 protein (as described in ref. 8) was sequenced on the Applied Biosystems model 470 A protein sequencer using the 02NRUN program with trifluoroacetic acid conversion to phenylthiohydantoin (PTH) derivatives. PTH amino acids were identified by high-performance liquid chromatography (HPLC) on a Waters Nova Pac C18 column using a Waters HPLC system. Solvent A was 84% sodium acetate (33 mM, pH 5.0)/16% acetonitrile; solvent B was 60% isopropanol/40% water. The column was maintained at 40°C and PTH amino acids were eluted at a flow rate of 1 ml/min using the following gradient: 0–0.5 min, 0% solvent B; 0.5–3.5 min, convex gradient to 35% solvent B (Waters' curve 4); 3.5–12 min, isocratic elution at 35% solvent B. All PTH amino acids were eluted from the column by 12 min at 35% solvent B. The amino-terminal sequence was determined twice.

Protein Labeling and Immune Precipitation. Protein labeling and immune precipitation were performed by the method previously described (3).

Construction of an *Escherichia coli* λ Bank. *E. coli* genomic DNA from the strain RR1 was isolated by the cleared lysate method (9). The DNA was partially digested with *Sau*3A and fragments between 15 and 20 kilobases (kb) were isolated from low melting agarose. The fragments were ligated into the *Bam*HI site of cut and dephosphorylated Charon 28 (10). The average insert size was 10–15 kb. The phage was grown on the *E. coli* host strain LE392 and ≈5000 plaques were screened with the RF-2 sequence by the method of Benton and Davis (11). Positive clones were plaque purified (12), and *Hin*dIII fragments were subcloned into pUC9 (13). Further subcloning of the 5′ region of RF-2 into the dideoxy sequencing vector M13 (mp9) (14) was accomplished by using an *Acc* I restriction site found within the RF-2 gene and the *Acc* I site within pUC9.

Abbreviations: RF, release factor; kb, kilobase(s); bp, base pair(s).

RESULTS

RF-1 and RF-2 Gene Structure. DNA sequencing by the method of Maxam and Gilbert was conducted on the 1.7-kb insert (Fig. 1) of the recombinant plasmid pRF1-1 (3). A single open reading frame of 969 bases, encoding a protein of 35,911 daltons, was identified. No other reading frame of the appropriate size was observed. Our prediction that this sequence corresponded to the RF-1 coding sequence was confirmed by expression studies.

We subcloned a 1.3-kb fragment, using restriction sites 52 base pairs (bp) 5′ to the predicted initiation AUG and 250 bp 3′ to the termination codon (Fig. 2), downstream from the *lac* promoter of the plasmid pUC9. The new recombinant, pRF1AP, was shown to contain the RF-1 coding sequence by an immune precipitation assay, as described (3). A protein of the correct molecular size was produced at high levels (Fig. 2), indicating the cloned RF-1 gene was still intact. We were unable to confirm the predicted amino-terminal sequence of purified RF-1 by peptide sequencing, despite attempts to deformylate the amino-terminal methionine by acid treatment of the protein.

Fig. 3 depicts the nucleic acid sequence of RF-1. The region ≈65 bp 5′ to the translation start site contains possible transcription control sequences homologous to the canonical −10 box and −35 box described for *E. coli* promoters (15). The exact transcription initiation site has not been determined. A possible Shine–Dalgarno sequence (A-G-G-G-U) is present 12 bp upstream from the initiator methionine in the RF-1 gene. The termination codon of the RF-1 gene is UGA, an RF-2 specific codon.

The RF-2 gene, contained on an insert of 2.7 kb within the recombinant plasmid pRF2-1 (3), was likewise sequenced by the strategy outlined in Fig. 1. An open reading frame was identified in the same manner as for the RF-1 gene. The open reading frame DNA sequence identified a coding sequence for RF-2 that predicted a protein of 30,056 daltons with an initiator methionine at position 610 in Fig. 3. Amino acid sequencing of the amino-terminal 44 amino acids of the purified RF-2 protein did not confirm the predicted initiator methionine, but rather identified a methionine 206 bp upstream as the initiator methionine. As described below, a single base frameshift in the DNA sequence allows extension of the open reading frame to 1019 bp, which would encode a protein of 38,404 daltons. This latter molecular size is more compatible with earlier studies (16).

To eliminate the possibility of a cloning artifact, the plasmid from which the DNA sequence was derived was

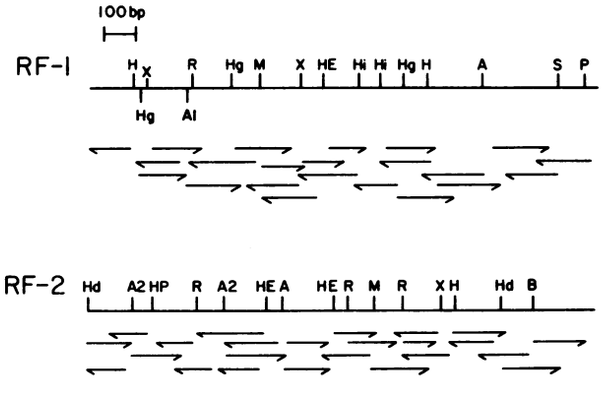

FIG. 1. Sequencing strategy for RF-1 and RF-2. Arrows represent direction and extent of each sequence; all fragments were labeled on the 5′ end. Only the restriction sites used in sequencing have been included. A, *Acc* I; A1, *Ava* I; A2, *Ava* II; H, *Hin*cII; Hg, *Hga* I; HE, *Hae* III; HP, *Hpa* II; Hd, *Hin*dIII; Hi, *Hin*fI; M, *Mbo* II; P, *Pvu* II; R, *Rsa* I; S, *Sac* II; X, *Xho* II.

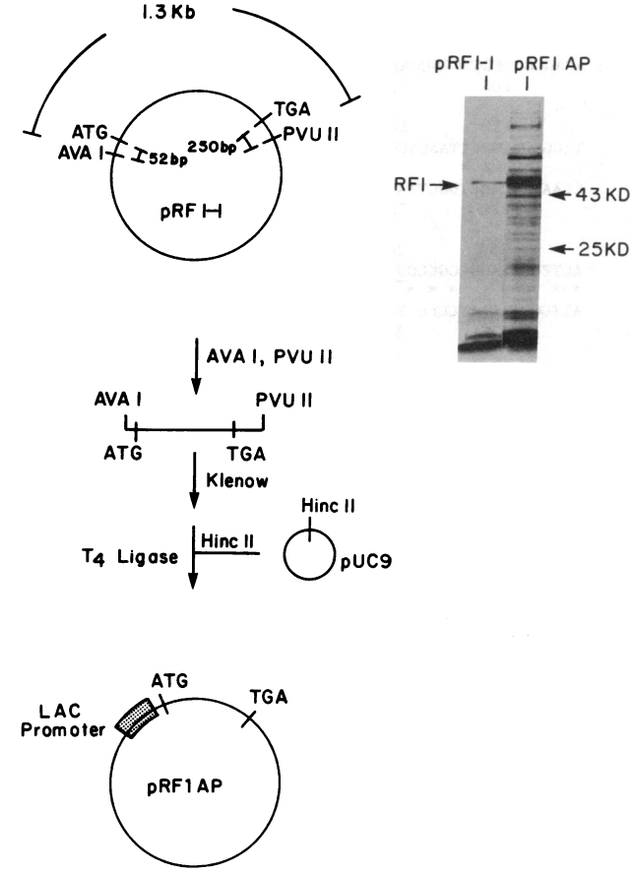

FIG. 2. Construction and assay of RF1AP. (*Left*) Strategy for inserting the RF-1 gene adjacent to the *lac* promoter of pUC9. Predicted translational start and stop codons are indicated. The 1.3-kb *Ava* I/*Pvu* II fragment containing the RF-1 gene was isolated from low melting agarose and blunt-ended with *Pol* I, Klenow fragment. The resulting fragment was inserted into the *Hin*cII site of pUC9, and the correct orientation was identified by restriction mapping. (*Right*) Immune precipitation of [^{35}S]methionine-labeled RF-1, showing that the inserted fragment contains the entire RF-1 coding region.

assayed for the production of RF-2 by immune precipitation of [^{35}S]methionine-labeled protein. A protein of the correct molecular size was produced (data not shown). In addition, we have determined the sequence of the region of the plasmid between the authentic site of peptide chain initiation and the predicted site of initiation based on the open reading frame, by both the dideoxy and Maxam–Gilbert techniques (Fig. 4). A single in-frame premature UGA termination codon and adjacent (+1) frameshift is contained within the RF-2 coding sequence. To confirm this finding, we obtained an independent RF-2 clone from a Charon 28 λ bank constructed from the *E. coli* strain RR1. The 5′ region of this second isolate was inserted into the M13 vector mp9 and sequenced by the Sanger method (Fig. 4). These sequence studies confirmed the occurrence of the premature UGA termination codon initially detected using the Maxam–Gilbert method. A summary of RF-2 peptide and nucleic acid sequence in this region is given in Fig. 5.

Fig. 3 depicts the nucleic acid sequence of RF-2. The RF-2 gene contains sequences ≈55 bp 5′ to the initiator methionine that are homologous to the −10 box *E. coli* promoter sequence. As with the RF-1 gene, a possible ribosomal binding site is present in the RF-2 gene upstream from the initiator methionine; but in neither gene is the region exactly complementary to the 3′ end of the 16S RNA. The termina-

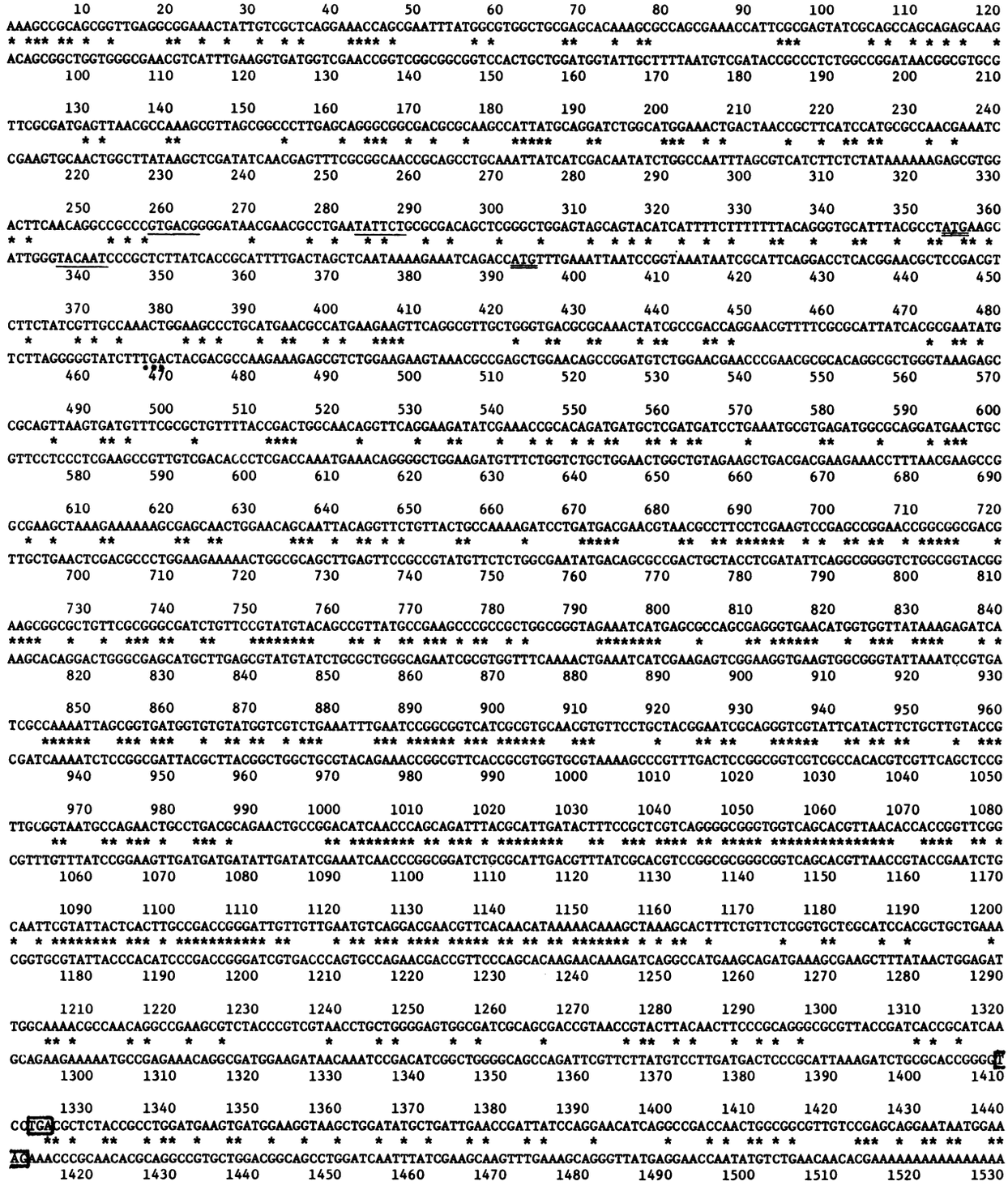

FIG. 3. DNA sequence of RF-1 and RF-2. Upper line represents RF-1 and lower line represents RF-2. The sequences are aligned to maximize homology and asterisks denote positions of identity. The extra thymine nucleotide is included in the RF-2 sequence. Possible control sequences are underlined, and the translation start sites are underlined with two lines. The termination codons of each gene are boxed and the in-frame premature UGA termination codon is indicated by three dots.

tion codon of the RF-2 is UAG, an RF-1-specific codon. A comparison of the DNA sequences reveals a 41% sequence homology between RF-1 and RF-2 (Fig. 3). The DNA homology is greatest in the 3′ end of the genes and corresponds with the amino acid homology.

Amino Acid Sequence Comparison of RF-1 and RF-2. RF-1 is composed of 323 amino acids and RF-2 is composed of 339 amino acids, as predicted by the nucleic acid sequences. Fig. 6 depicts the two amino acid sequences aligned for comparison, where the first methionine of RF-1 has been aligned with serine-18 of RF-2 to maximize homology.

There is a 31% amino acid sequence identity between RF-1 and RF-2. If conserved amino acid substitutions are included in this calculation, the homology increases to 51%. Regional homology is much greater; between residues 231 and 284, the homology is 77%; between residues 130 and 284, it is 57%. These values increase to 89% and 77%, respectively, if conserved amino acid substitutions are taken into account.

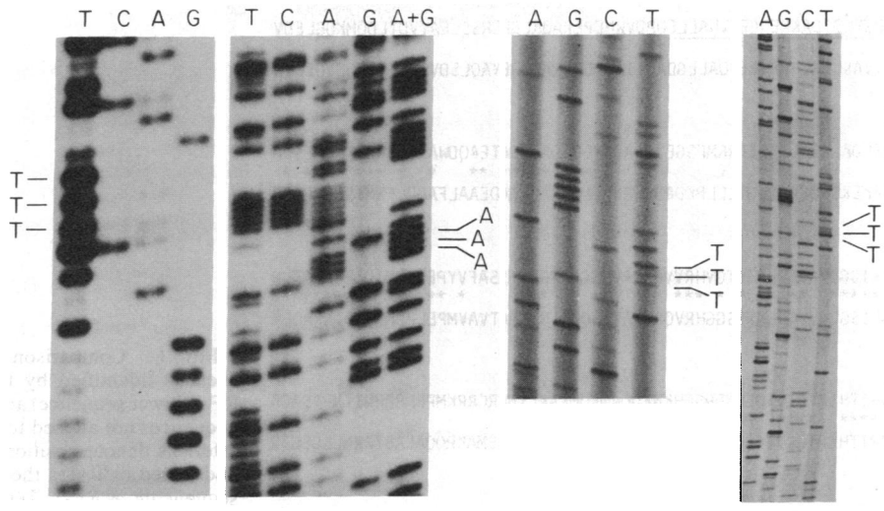

FIG. 4. DNA sequence of the region of RF-2 gene that contains the premature UGA termination codon. The relevant feature is the presence of three thymidine nucleotides. The two panels on the left side of the figure are Maxam–Gilbert sequences of both strands from the original RF-2 isolate pRF2-1. The two panels on the right side are dideoxy sequencing gels of the same region, to eliminate the possibility of a sequencing artifact. The panel farthest right is of a second independent RF-2 isolate that shows the same sequence (see text).

The alignment can be increased slightly by introducing a one-residue gap in the amino-terminal end of RF-2 at position 80. It would appear that RF-1 and RF-2 are quite conserved in their carboxyl sequence and less so in their amino-terminal sequence.

DISCUSSION

In the present study, the structures of the genes encoding the *E. coli* peptide chain RF-1 and RF-2 have been determined. In addition, purified RF-2 protein has been partially sequenced twice to confirm the assignment of the amino-terminal methionine.

The RF-1 and RF-2 gene sequences predict protein sizes of 35,911 daltons and 38,404 daltons, respectively. This is in contrast to the estimated molecular sizes of ≈47,000 and ≈48,000 daltons, respectively, as measured by their electrophoretic mobility on polyacrylamide gels (16). Sedimentation equilibrium studies of RF-2 estimated the molecular size to be ≈35,000 daltons (16), a finding consistent with the data presented here. The anomalous mobility on the polyacrylamide gel system can possibly be accounted for by the acidic nature of the two RFs; RF-1 has a pI of 5.6 and RF-2 has a pI of 5.1 (16).

A comparison of the amino acid sequences of both RFs demonstrates the extensive homology that exists between them. This homology is presumably a reflection of the shared functional features of the RFs; namely, both must bind to the ribosome, recognize the termination codon, and interact with the peptidyl transferase center to cause hydrolysis of the nascent peptide chain. The most obvious divergence of function is in the exclusive codon specificity that each RF maintains. Structural divergence has been more difficult to assess, but available evidence indicates that the ribosomal binding domains show definite dissimilarity. This difference in ribosomal binding has recently been studied in detail by Tate *et al.* (17), by measuring the termination activity of each RF with reconstituted ribosomes. The ribosomal protein L11 differentially influenced the RF-1 and RF-2 codon-dependent, but not the independent, termination reaction. This finding suggests that ribosomal factors influence RF-1 and RF-2 codon-dependent recognition. Since both factors recognize UAA it would appear that the L11 effect is related to structural differences of RF-1 and RF-2 influencing ribosomal interactions, and not specifically codon recognition.

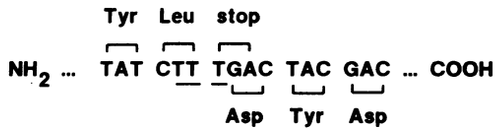

FIG. 5. Summary of the DNA and peptide sequences of the region of RF-2 containing the in-frame termination codon. Amino acid sequencing confirms the presence of the amino acids Tyr, Leu, Asp, Tyr, and Asp on either side of the termination codon. The three thymidine nucleotides that are indicated in Fig. 4 are underlined.

While nucleotide sequence recognition is predictably a complex mechanism, the obvious amino acid sequence differences of RF-1 and RF-2 reside in the amino-terminal region. This amino-terminal region has several repeating amino acid doublets that are found in both RFs; a leucine-glutamic acid doublet and arginine-glutamic acid doublet that are both reiterated a number of times and reflect repeating DNA sequences. The significance of this is unclear.

Earlier studies indicated that the peptidyl transferase activity of the ribosome has the capacity to catalyze peptidyl tRNA hydrolysis in the presence of RF or certain antibiotics (18). Numerous efforts to identify covalent intermediates of the nascent peptides with RF have failed. *N*-Ethylmaleimide alkylation of RF-2, however, markedly diminished its ability to participate in peptidyl tRNA cleavage without influencing its codon recognition or ribosomal binding properties (unpublished data).

RF-2 has four cysteine residues (positions 128, 274, 300, and 336); RF-1 has three cysteine residues (positions 50, 200, and 256). Cysteine-274 and cysteine-256 are in identical positions (as aligned by amino acid homology) in the two genes, and both are found within highly conserved regions, implying that this amino acid may have a role in the peptidyl tRNA hydrolysis function. Knowledge of the RF-1 and RF-2 protein primary structures now permit targeted *in vitro* mutation of the cloned genes for determining the functional regions and residues.

Several structural features of the RFs suggest a translational regulation of RF activity. A simple extension of the finding that RF-2 terminates with the codon UAG and RF-1 terminates with the codon UGA is that each RF can effect the translation of the other RF transcript. Earlier studies have shown that suppressor tRNAs and RF compete for peptide-

```
MFEINPVNNRIQDLTERSDVLRGYLDYDAKKERLEEVNAELEQPDVWNEPERAQALGKERSSLEAVVDTLDQMKQGLEDV
                                                                        *
                    MKPSIVAKLEALHERHEEVQALLGDAQTIADQERFRALSREYAQLSDVSRCFTDWQQVQEDIE

SGLLELAVEADDEETFNEAVAELDALEEKLAQLEFRRMFSGEYDSADCYLDIQAGSGGTEAQDWASMLERMYLRWAESRG
     *      *    *          *    **            *      *    ** ** **    *  * *** * ** *
TAQMMLDDPEMREMAQDELREAKEKSEQLEQQLQVLLLPKDPDDERNAFLEVRAGTGGDEAALFAGDLFRMYSRYAEARR

FKTEIIEESEGEVAGIKSVTIKISGDYAYGWLRTETGVHRVVRKSPFDSGGRRHTSFSSAFVYPEVDDDIDIEINPADLR
   **   ****  * *    *****  ** *  * * *** *      * ** ***      * **  *      *******
WRVEIMSASEGEHGGYKEIIAKISGDGVYGRLKFESGGHRVQRVPATESQGRIHTSACTVAVMPELPDAELPDINPADLR

IDVYRTSGAGGQHVNRTESAVRITHIPTGIVTQCQNDRSQHKNKDQAMKQMKAKLYNWRCRRKMPRNRRWKITNPTSAGA
**  * ********* * ** **** *****  **  *******  *     *                            *
IDTFRSSGAGGQHVNTTGSAIRITHLPTGIVVECQDERSQHKNKAKALSVLGARIHAAEMAKRQQAEASTRRNLLGSGDR

ARFVLMSLMTPALKICAPG(TAG)

SDRNRTYNFPQGALPITAST(TGA)
```

FIG. 6. Comparison of the amino acid sequences (identified by the single-letter code) of RF-1 (lower sequence) and RF-2 (upper sequence). Sequences are aligned to maximize homology and asterisks denote positions of identity. The region underlined indicates the extent of the amino acid sequencing of RF-2. Termination codons used by RF-1 and RF-2 are indicated at the end of each protein. A dot is above the position in RF-2 that contains the premature termination codon.

chain termination codon translation (19). By using a peptide-chain termination codon recognized by the other RF, the cell avoids readthrough of an RF by the suppressor tRNA with which the RF competes. Thus, for example, RF-1 synthesis would be unaltered by an amino acid $tRNA^{UAG}$, avoiding an imbalance of peptide-chain termination codon recognition molecules and the implicit effects on cellular proteins terminated by UAG.

It is also possible that RF-2 autoregulates its production by the in-frame UGA premature termination codon found within the early coding region of RF-2. We speculate that when adequate RF-2 is present, peptide-chain termination would occur at this site, thus limiting its translational production. Since spontaneous termination in the absence of termination factors should not occur (20) when levels of RF-2 are low, a readthrough/frameshift mechanism would permit increased expression of the RF-2 protein. This second possibility for regulation (premature termination) would imply autogenous regulation of RF-2. These proposed translational regulatory mechanisms are experimentally testable.

The in-frame UGA termination codon within RF-2 is preceded by the leucine tRNA codon CUU. It is noteworthy that leucine $tRNA^{CUN}$ has been reported as a "shifty tRNA" (21), with an ability to translocate 4 bases (22). The sequence C-U-U-U-G-(A) should be the site of the frameshift because peptide sequencing confirms the presence of a leucine at position 25. Runs of uracil have been identified as possible regions of frameshifting in the yeast mitochondrial oxi-1 gene (23), gene 10 of phage T7 (24), and the *Salmonella trpE* gene (25). It has been postulated that a frameshift event regulates the expression of the lysis protein of the RNA phage MS2 (26).

The distribution of amino acid homology within the proteins argues for a single ancestral RF protein, with the existent genes arising from a gene duplication event and subsequent evolutionary divergence (27). This is interesting in light of the fact that there is only a single mammalian RF that encompasses the functions of both RF-1 and RF-2 (28).

With the knowledge of the primary structure of the RFs, elucidation of the functional domains can now be attempted. In addition, the question of how RF synthesis is regulated will also be addressed.

W.J.C. was supported by the Welch Foundation and the Baylor College of Medicine Medical Scientist Training Program 5 T32 GM07330-08; W.P.T. by the Medical Research Council (N.Z.).

1. Caskey, C. T. & Campbell, J. M. (1979) in *Nonsense Mutations and tRNA Suppressors*, eds. Celis, J. E. & Smith, J. D. (Academic, New York), pp. 81–98.
2. Ratliff, J. M. & Caskey, C. T. (1977) *Arch. Biochem. Biophys.* **181,** 671–677.
3. Caskey, C. T., Forrester, W. C., Tate, W. & Ward, C. (1984) *J. Bacteriol.* **158,** 365–368.
4. Weiss, R. B., Murphy, J. P. & Gallant, J. A. (1984) *J. Bacteriol.* **158,** 362–364.
5. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560.
6. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
7. Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 3963–3965.
8. Caskey, C. T., Scolnick, E., Tompkins, R., Milman, G. & Goldstein, J. (1971) *Methods Enzymol.* **20,** 367–375.
9. Katz, L., Kingsburgy, D. T. & Helinski, D. R. (1973) *J. Bacteriol.* **114,** 577–591.
10. Rimm, D. L., Horness, D., Kucera, J. & Blattner, F. R. (1980) *Gene* **12,** 301–309.
11. Benton, W. D. & Davis, R. W. (1977) *Science* **196,** 180–182.
12. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York).
13. Vieira, J. & Messing, J. (1982) *Gene* **19,** 259–268.
14. Messing, J. & Vieira, J. (1982) *Gene* **19,** 269–276.
15. Hawley, D. K. & McClure, W. R. (1983) *Nucleic Acids Res.* **11,** 2237–2255.
16. Campbell, J. M. (1978) Dissertation (Baylor College of Medicine, Houston, TX).
17. Tate, W. P., Schulze, H. & Nierhaus, K. H. (1983) *J. Biol. Chem.* **258,** 12816–12820.
18. Scolnick, E., Milman, G., Rosman, M. & Caskey, C. T. (1970) *Nature (London)* **225,** 152–154.
19. Beaudet, A. L. & Caskey, C. T. (1970) *Nature (London)* **227,** 38–40.
20. Bretscher, M. (1968) *J. Mol. Biol.* **34,** 131–136.
21. Weiss, R. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 5797–5801.
22. Weiss, R. & Gallant, J. (1983) *Nature (London)* **302,** 389–393.
23. Fox, T. D. & Weiss-Brummer, B. (1980) *Nature (London)* **288,** 60–63.
24. Dunn, J. J. & Studier, F. W. (1983) *J. Mol. Biol.* **166,** 477–535.
25. Atkins, J. F., Nichols, B. P. & Thompson, S. (1983) *EMBO J.* **2,** 1345–1350.
26. Kastelein, R. A., Remaut, E., Fiers, W. & Van Duin, J. (1982) *Nature (London)* **295,** 35–40.
27. Dayhoff, M. O. & Barker, W. C. (1972) in *Atlas of Protein Sequence and Structure*, ed. Dayhoff, M.O. (Natl. Biomed. Res. Found., Washington, DC), Vol. 5.
28. Goldstein, J. L., Beaudet, A. L. & Caskey, C. T. (1970) *Proc. Natl. Acad. Sci. USA* **67,** 99–106.