Molecular Microbiology (1990) 4(6), 861–865

MicroReview

Recent advances in peptide chain termination

W. J. Craigen, C. C. Lee and C. T. Caskey*
Institute for Molecular Genetics, Howard Hughes Medical Institute, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA.

Summary

Peptide chain termination occurs when a stop codon is decoded by a release factor. In *Escherichia coli* two codon-specific release factors (RF1 and RF2) direct the termination of protein synthesis, while in eukaryotes a single factor is required. The *E. coli* factors have been purified and their genes isolated. A combination of protein and DNA sequence data reveal that the RFs are structurally similar and that RF2 is encoded in two reading frames. Frame-shifting from one reading frame to the next occurs at a rate of 50%, is regulated by the RF2-specific stop codon UGA, and involves the direct interaction of the RF2 mRNA with the 3′ end of the 16S rRNA. The RF genes are located in two separate operons, with the RF1 gene located at 26.7 min and the RF2 gene at 62.3 min on the chromosome map. Ribosomal binding studies place the RF-binding region at the interface between the ribosomal subunits. A possible mechanism of stop-codon recognition is reviewed.

Introduction

The termination of protein synthesis occurs when an in-frame stop codon enters the ribosomal A site. Unlike sense codons that are decoded by specific tRNAs via RNA-RNA interactions, stop codons are decoded by proteins termed release factors (RFs). Release factors have been purified from both prokaryotic and eukaryotic organisms, and recent progress in defining the structural features of these factors has led to several unexpected results. This review addresses the structural features of RFs, the regulation and genetic organization of the RF genes in bacteria, RF ribosomal binding, and codon recognition by RFs.

Ganoza (1966) first proposed that protein factors are required for polypeptide chain termination, and this was subsequently confirmed by Capecchi (1967). A simple *in vitro* assay for termination activity (Caskey *et al.*, 1968) allowed for the identification of two codon-specific protein factors in *Escherichia coli*. Termination was found to require stop codons, a peptide analogue (f-met), and release factor. In *E. coli*, two codon-specific release factors (RF1 and RF2) are required for termination; RF1 participates in UAA- and UAG-directed termination and RF2 in UAA- and UGA- directed termination (Scolnick, 1968). These factors have a predicted monomer size of 47–50 kiloDaltons (kD), are highly acidic proteins, and lack any nucleic acid or carbohydrate moieties (Ratcliff, 1979). A third factor has been identified in *E. coli* that stimulates the termination reaction and binds guanine nucleotides but is not codon-specific (Milman, 1969). In eukaryotes, a single larger (56 kD) factor that directs termination for all three stop codons has been purified from a variety of sources, including rabbit reticulocytes (Goldstein *et al.*, 1970), guinea-pig liver (Innanen and Nichols, 1973), insect cells (Ilan, 1973), and the brine shrimp *Artemia salina* (Reddington and Tate, 1979), and, likewise, a single 45 kD factor has been shown to direct translational termination in rat liver mitochondria in a codon-specific manner (Lee *et al.*, 1987).

Release factor genes

The genes encoding the *E. coli* RFs have been isolated. RF1 was identified by a genetic screen for anti-suppression using an amber mutant *lacZ* indicator gene weakly suppressed by an amber suppressor tRNA (Weiss *et al.*, 1984). This experiment was based upon the observation of Beaudet and Caskey (1970) that purified RF1 can compete with a suppressor tRNA for the *in vitro* translation of UAG. The gene encoding RF2 was identified by an immune precipitation assay from the Clarke and Carbon pColE1 library (Caskey *et al.*, 1984), relying upon overexpression of the gene in a strain harbouring an RF2-containing plasmid.

The map position of RF1 (designated *prfA*) has recently been reported to be 26.7 min on the *E. coli* chromosome map (Ryden *et al.*, 1986; Lee *et al.*, 1988), the previously reported location of two temperature-sensitive amber suppressors (Davidoff-Abelson and Mindich, 1979; Reyden and Isaksson, 1983). The RF2 gene has been localized to 62.3 min on the map by two independent biochemical techniques: pulse-field gradient gel electrophoresis and hybridization to a contiguous, ordered

Received 18 December, 1989; revised 5 February, 1990. *For correspondence. Tel. (713) 798 4774; Fax (713) 798 6521.

lambda phage *E. coli* library, and P1 phage transduction (Lee *et al*; 1988; Kawakami *et al.*, 1988a). Recent preliminary work by Kawakami *et al.* (1988a) has demonstrated that the *supK* mutation of *Salmonella typhimurium*, a recessive suppressor of UGA mutations, is likely to be the *RF2* gene.

The RF2 gene (designated *prfB*) resides in an operon with *herC*, a gene previously shown to be involved in the suppression of a replication-defective ColE1 plasmid (Kawakami *et al.*, 1988a, 1989). A more recent report (Gampel and Tzagoloff, 1989) suggests that *herC* is the constitutively expressed lysyl-tRNA synthetase. Transcription initiates immediately upstream of the RF2 start codon and proceeds in a counter-clockwise direction (with respect to the chromosome map) through the *prfB* and *herC* genes, producing a 2800 bp transcript that terminates at a rho-independent termination site within 50 bp of the predicted *herC* stop codon. Kawakami *et al.* (1988b) have isolated mutations of the RF2 gene by selecting for suppression of a *lacZ* UGA allele or temperature-sensitive conditional lethality using mutagenized P1 phage. This study showed RF2 to be essential for cell growth and a recessive suppressor of UGA codons when mutagenized.

The preliminary organization of the RF1 operon has recently been reported for both *E. coli* (Verkamp and Chelm, 1989) and *S. typhimurium* (Elliott, 1989). In both species the RF1 gene resides immediately downstream of the *hemA* gene, an enzyme involved in the haem synthesis pathway. The operon appears to encode three transcripts: two species that differ in length by 92 bp at the 5′ end, and a third transcript of unknown function synthesized from the opposite strand. The *hemA* gene terminates with the RF1-specific stop codon UAG, followed 41 bp downstream by the RF1 initiation codon. It has been suggested (Elliott, 1989) that RF1 synthesis may be regulated by the termination rate at this stop codon, with readthrough leading to increased translatability of the RF1 codon region. No data are yet available with regard to this hypothesis.

DNA sequencing of both genes (Craigen *et al.*, 1985) has revealed several interesting features: first, the proteins share a 30% sequence identity and contain many conserved substitutions; second, the distribution of sequence similarity suggests a divergent evolution, with the genes having arisen from a single primordial gene; and third, short regions of sequence similarities exist between the RFs and several ribosomal proteins (Craigen and Caskey, 1987a).

Release factor regulation

DNA and protein sequencing of *prfB* demonstrated that RF2 is encoded in two separate reading frames, requiring a +1 frameshift during translation for synthesis of the entire protein. This observation provides a possible mechanism for the autoregulation of RF2, since the more proximal reading frame is terminated by the RF2-specific stop codon UGA. RF1 is synthesized from a single reading frame based upon the DNA sequence, although attempts to sequence the purified protein to confirm this have so far been unsuccessful. Subsequently, it was shown that this in-frame stop codon within the RF2 coding region is regulatory in an *in vitro* translation system; the addition of purified RF2 to the system produces a marked drop in RF2 synthesis (Craigen and Caskey, 1986). This finding is reminiscent of the autogenous regulation that has been described for several ribosomal protein operons; for example, the L1-mediated regulation of the L11 operon (Thomas and Nomura, 1987), and serves to highlight the diverse regulatory mechanisms that have evolved in bacteria. RF2 employs its codon specificity to provide exquisite specificity for its own regulation, much as particular ribosomal proteins use the specificity of their primary function, e.g. binding ribosomal RNA, to regulate the operons within which they reside.

The rate of frameshifting between the two reading frames has been measured by constructing *lacZ* fusion genes with RF2 DNA fragments lacking or containing the frameshift. In this manner, it was demonstrated that frameshifting into the correct reading frame occurs approximately 30–50% of the time (Craigen and Caskey, 1986; Curran and Yarus, 1988; Weiss *et al.*, 1987), indicating that RF2 has evolved a capacity to thwart the ribosomal termination and reading-frame maintenance functions. This observation has since been extended to several eukaryotic organisms; a variety of eukaryotic viruses, including human immunodeficiency virus-1, bovine leukosis virus, mouse mammary tumour virus, and the coronaviruses, exhibit high-efficiency spontaneous frameshifts, as do certain yeast transposons (reviewed in Craigen and Caskey, 1987b). Although these organisms employ high-efficiency frameshifts, presumably to regulate cellular levels of particular proteins, the mechanism of frameshifting may be considerably different from that of RF2 (e.g. Brierley *et al.*, 1989). Work by Curran and Yarus (1988) has confirmed that the translation of RF2 reflects a simple competition between termination and frameshifting, and, in addition, they identified a short sequence element immediately upstream of the in-frame UGA that is required for high-efficiency frameshifting. This sequence has been shown to interact with the Shine-Dalgarno region of the 16S rRNA (Weiss *et al.*, 1988), providing experimental evidence that the 16S rRNA is involved in maintaining the correct reading frame during translation.

A role for the Shine-Dalgarno region in maintaining the reading frame has previously been suggested by Trifonov

(1987), based upon the distribution of nucleotides within codons. If true, this would extend the function of the 16S rRNA beyond its essential role in translation initiation to chain elongation. Furthermore, work by others suggests it also has a role in termination, specifically stop-codon recognition (see below). It would appear that, consistent with the growing view that RNA provides the enzymatic basis for ribosomal function, initiation, elongation, and termination may all be catalysed by the RNA moiety.

Although the exact mechanism of RF2 frameshifting is still unclear, it appears that three elements are required for high-efficiency frameshifting: a 'shifty' codon at the site of the frameshift, a stop codon adjacent to the shifty codon, and an upstream element that binds the 3′ end of the 16S rRNA. The 'shiftiness' of the frameshift codon is probably due to a string of pyrimidines within which the tRNA can slip, while the upstream element possibly constrains the 16S rRNA in such a way that reading-frame maintenance is perturbed. The fact that the UGA stop codon (or a sense codon that uses a minor tRNA species (Curran and Yarus, 1989)) is required for a high rate of frameshifting suggests that the presence of a stop codon slows the decoding event, thus allowing for the frameshift while the presence of a sense codon (and the more rapid rate of tRNA binding at the A site) hinders frameshifting by decoding in-frame. One possible scenario for the RF2 frameshift is that normally during translation the Shine-Dalgarno region continually scans the mRNA as it moves through the ribosome. When it encounters the upstream element, binding may be so tight as to block this scanning function. If indeed stop-codon recognition involves a second region of the 16S rRNA binding to A site stop-codons, as has been suggested by Murgola *et al.* (1988), perhaps the binding of the Shine-Dalgarno region hinders stop-codon binding to its recognition region, thereby inhibiting RF binding to the ribosome and allowing for the frameshift.

The RFs probably contribute to the regulation of a variety of genes. To date, a single example of RF-mediated regulation (apart from the autoregulation of RF2) has been described: that of the tryptophan operon (Roesser *et al.*, 1989). Taking advantage of the RF2 mutant strains described by Kawakami *et al.* (1988b), these authors showed that a reduced level of RF2 activity leads to an increase in transcription termination at the *trp* operon attentuator via translational termination at an authentic UGA codon within the *trp* leader peptide. The same authors had previously shown that an RF1-defective strain has a similar effect on the basal level of *trp* expression when the leader peptide contains an engineered UGA codon at the same position (Roesser and Yanofsky, 1988). Other examples of RF-mediated regulation are likely to be identified, such as the incorporation of selenocysteine at UGA codons in several prokaryotic and eukaryotic genes (Böck and Stadtman, 1988).

RF ribosomal binding

A number of studies employing a variety of experimental approaches, including antibody inhibition, ribosomal protein reconstitution, protein cross-linking, and antibiotic inhibition, have been published addressing the ribosomal requirements for RF binding and function. For the large subunit, antibodies to L11 and L16 block RF-mediated peptidyl-tRNA hydrolysis, but leave RF ribosomal binding intact (Tate *et al.*, 1975). Since L16 is believed to be an essential part of the peptidyl-transferase centre, this finding supports the view that the hydrolysis reaction is carried out by the peptidyl-transferase centre. Antibodies to the small subunit proteins S3, S4, S5, and S10, which are clustered on the lower head and upper part of the small lobe of the subunit, inhibit RF function, suggesting that the RFs bind in the interface region between the subunits in the area of the large subunit stalk (Tate *et al.*, 1988). By similar antibody inhibition techniques it has been shown that proteins L7/L12, S2, S9, and S11 are also required for RF ribosomal binding, while antibodies to a number of ribosomal proteins appear to block the termination reaction by interfering with subunit association (Tate *et al.*, 1975). RF2 has been cross-linked to the large subunit proteins L2, L7/L12, and L11, and to a lesser extent to the small subunit proteins S6, S17, and S18 (Tate *et al.*, 1975). Tate *et al.* (1973; 1983a) reported that ribosomes containing either EF-G or EF-Tu fail to bind either RF, while ribosome-bound aminoacyl-tRNA alone has little inhibitory effect, suggesting some overlap in the binding domains of the translation factors with a separate site for tRNA binding. The presence of L11 has been shown to have a differential effect on RF1 and RF2 activities, stimulating RF1-specific function while inhibiting that of RF2 (Tate *et al.*, 1983b), suggesting that differences in the binding sites for RF1 and RF2 exist. Furthermore, displacement of the termination codon from the ribosomal A site by spacer nucleotides still allows for codon recognition and termination to take place (Tate *et al.*, 1983a), which suggests a degree of flexibility in RF binding to the ribosome. A recent report by Lang *et al.* (1989) demonstrates that, in the presence of RF2, the stop-codon triplet UGA can be cross-linked to a number of proteins, including S6, S18, L2, L7/12, L10, and L20, further suggesting that codon recognition occurs in the interface region at the base of the large subunit stalk.

Antibiotics have played a useful role in defining both the functional and structural requirements of RF function. Streptomycin and erythromycin inhibit peptide chain termination, and resistance to these agents has been mapped to mutations in S4, suggesting that this protein is also important in termination. Antibiotics that inhibit peptidyl-transferase-mediated peptide bond formation also inhibit RF-mediated peptidyl-tRNA hydrolysis while leaving RF codon recognition intact, again suggesting that hydrolysis

is catalysed by peptidyl-transferase. In contrast, tetracycline, an inhibitor of aminoacyl-tRNA binding, abolishes RF-codon recognition but has no effect on peptidyl-tRNA hydrolysis (Thompkins *et al.*, 1970).

While progress has been made in defining contact points for the RFs on the ribosome, the structural details of these interactions are unclear and crystallographic data will probably be required for their clarification.

Stop-codon recognition

The mechanism of stop-codon recognition by RFs is unknown but is of considerable interest since it may entail protein-RNA recognition rather than a codon-anticodon (RNA-RNA) interaction. Previous data on this point are inconclusive, although more recent evidence supports a role for ribosomal RNA in codon recognition. The proposition that RFs directly bind stop codons is supported by two lines of experimental evidence: first, RFs compete with suppressor tRNAs for translation of termination codons in an *in vitro* translation system (Beaudet and Caskey, 1970); and second, equilibrium dialysis studies (Capecchi and Klein, 1969) have suggested that RFs can bind stop codons in the absence of ribosomes, although the codon specificity in these studies was not absolute and the experimental evidence weak. A more recent study of stop-codon binding by RFs failed to demonstrate stop-codon binding in the absence of ribosomes, but a 10-fold difference in binding between the cognate stop-codon UGA and the non-cognate codon UAG in the presence of ribosomes was seen; these results are consistent with the idea that codon recognition requires the RNA component of the ribosome (Lang *et al.*, 1989).

An alternative mechanism was previously put forward by Shine and Dalgarno (1974), who hypothesized that stop-codon recognition is due to interactions between the 3′-end of the 16S rRNA and the codon, based upon stop-codon complementary sequences in this region. This hypothesis is supported by the demonstration that cleavage of the 3′-end of the 16S rRNA by cloacin DF13 inhibits the partial reaction of codon recognition in the *in vitro* termination assay (Caskey *et al.*, 1977). In addition, the recent observation that the region upstream of the RF2 frameshift site is complementary to the Shine-Dalgarno sequence provides a possible mechanism both for high-frequency frameshifting and for termination in general.

A second line of evidence that suggests a role for the 16S rRNA in stop-codon recognition is the recent demonstration that a UGA-specific nonsense suppressor is, in fact, a single-base deletion of a cytosine in the 16S rRNA at position 1054 in a region containing sequences complementary to UGA stop codons (Murgola *et al.*, 1988). Based upon this observation, a model for codon recognition has been proposed whereby codon recognition occurs by means of RF2 interactions with this region of the 16S rRNA and not the 3′-end. The fact that *Mycoplasma* species use UGA as a sense codon and, likewise, lack this complementary sequence in an otherwise highly conserved region of the 16S rRNA adds support to this putative mechanism of stop-codon recognition. Although no definitive proof of RNA-RNA recognition has yet been provided, the most recent work in this area favours a role for the 16S rRNA, at least for RF2-mediated UGA termination. A recent report by Yano and Yura (1989) describes the suppression of a UGA mutant allele of *rpoH* by ribosomes lacking the S15 protein. Although this suppressor failed to suppress other UGA mutations, this observation does suggest that S15 may be involved in UGA-directed termination.

Although nothing of comparable detail is known about eukaryotic RF-codon recognition, progress is being made. Our laboratory has recently isolated the gene encoding the rabbit RF, and DNA sequence analysis of the cDNA reveals that it contains considerable structural similarity to aminoacyl-tRNA synthetases in general, and to tryptophanyl-tRNA synthetase in particular, yet lacks any discernable similarity to the prokaryotic RFs (Lee *et al.*, in press). This suggests that termination in eukaryotes has evolved in a manner greatly different from that of prokaryotes.

In summary, after a long period when little headway was made in understanding peptide chain termination, the last few years have seen considerable progress in this essential step in protein biosynthesis. Hopefully, future studies may uncover the molecular basis for stop-codon recognition, perhaps the last remaining 'coding problem'.

Acknowledgements

The authors wish to acknowledge the support of the Howard Hughes Medical Institute, and NIH Grant no. GM34438 from the Public Health Service.

References

Beaudet, A.L., and Caskey, C.T. (1970) *Nature* **227:** 38–40.
Böck, A., and Stadtman, T. (1988) *Biofactors* **1:** 245–250.
Brierley, I. *et al.* (1989) *Cell* **57:** 537–547.
Capecchi, M.R. (1967) *Proc Natl Acad Sci USA* **58:** 1144–1151.
Capecchi, M.R., and Klein, H.A. (1969) *Cold Spring Harbor Symp Quant Biol* **34:** 469–477.
Caskey, C.T. *et al.* (1968) *Science* **162:** 135–138.
Caskey, C.T. *et al.* (1977) *J Biol Chem* **252:** 4435–4437.
Caskey, C.T. *et al.* (1984) *J Bacteriol* **158:** 364–368.
Craigen, W.J., and Caskey, C.T. (1986) *Nature* **322:** 273–275.
Craigen, W.J., and Caskey, C.T. (1987a) *Biochimie* **69:** 1031–1041.
Craigen, W.J., and Caskey, C.T. (1987b) *Cell* **50:** 1–2.
Craigen, W.J. *et al* (1985) *Proc Natl Acad Sci USA* **82:** 3616–3620.

Curran, J., and Yarus, M. (1988) *J Mol Biol* **203:** 75–83.
Curran, J., and Yarus, M. (1989) *J Mol Biol* **209:** 65–77.
Davidoff-Abelson, R., and Mindich, L. (1979) *Mol Gen Genet* **159:** 161–169.
Elliott, T. (1989) *J Bacteriol* **171:** 3948–3960.
Gampel, A., and Tzagoloff, A. (1989) *Proc Natl Acad Sci USA* **86:** 6023–6027.
Ganoza, M.C. (1966) *Cold Spring Harbor Symp Quant Biol* **31:** 273–276.
Goldstein, J.L. *et al.* (1970) *Proc Natl Acad Sci USA* **67:** 99–106.
Ilan, J. (1973) *J Mol Biol* **77:** 437–448.
Innanen, V.T., and Nicholls, D.M. (1973) *Biochim Biophys Acta* **324:** 533.
Kawakami, K. *et al.* (1988a) *Proc Natl Acad Sci USA* **85:** 5620–5624.
Kawakami, K. *et al.* (1988b) *J Bacteriol* **170:** 5378–5381.
Kawakami, K. *et al* (1989) *Mol Gen Genet* **219:** 333–340.
Lang, A. *et al.* (1989) *J Biochem* **180:** 547–554.
Lee, C.C. *et al* (1987) *J Biol Chem* **262:** 3548–3552.
Lee, C.C. *et al.* (1988) *J Bacteriol* **170:** 4537–4541.
Milman, G. *et al.* (1969) *Proc Natl Acad Sci USA* **63:** 183–190.
Murgola, E. J. *et al.* (1988) *Proc Natl Acad Sci USA* **85:** 4162.
Ratliff, J.M. (1979) PhD Thesis, Baylor College of Medicine, USA.
Reddington, M.A., and Tate, W.P. (1979) *FEBS Lett* **97:** 335.
Reyden, S.M., and Isaksson, L.A. (1983) *Mol Gen Genet* **193:** 38–45.
Roesser, J.R., and Yanofsky, C. (1988) *J Biol Chem* **263:** 14251–14255.
Roesser, J. *et al.* (1989) *J Biol Chem* **264:** 12284–12288.
Ryden, S.M. *et al.* (1986) *J Bacteriol* **168:** 1066–1069.
Scolnik, E. *et al.* (1968) *Proc Natl Acad Sci USA* **61:** 768–774.
Shine, J., and Dalgarno, L. (1974) *Proc Natl Acad Sci USA* **71:** 1342–1346.
Tate, W.P. *et al.* (1973) *Proc Natl Acad Sci USA* **70:** 2350–2355.
Tate, W.P. *et al.* (1975) *J Mol Biol* **93:** 375–389.
Tate, W.P. *et al.* (1983a) *J Biol Chem* **258:** 10360–10365.
Tate, W.P. *et al.* (1983b) *J Biol Chem* **258:** 12816–12820.
Tate, W.P. *et al.* (1988) *Biochem Intl* **17:** 179–186.
Thomas, M.S., and Nomura, M. (1987) *Nucl Acids Res* **15:** 3085–3096.
Thompkins, R. *et al.* (1970) *Proc Natl Acad Sci USA* **65:** 702–708.
Trifonov, E.N. (1987) *J Mol Biol* **194:** 643–652.
Verkamp, E., and Chelm, B. (1989) *J Bacteriol* **171:** 4728–4735.
Weiss, R. *et al.* (1984) *J Bacteriol* **158:** 362–364.
Weiss, R. *et al.* (1987) *Cold Spring Harbor Symp Quant Biol* **52:** 687–693.
Weiss, R. *et al.* (1988) *EMBO J* **7:** 1503–1507.
Yano, R., and Yura, T. (1989) *J Bacteriol* **171:** 1712–1717.