*RNA* (1998), *4*:1047–1054. Cambridge University Press. Printed in the USA.


# Circular mRNA can direct translation of extremely long repeating-sequence proteins in vivo

**RHONDA PERRIMAN and MANUEL ARES, JR**

Center for Molecular Biology of RNA, Biology Department, Sinsheimer Laboratories, University of California, Santa Cruz, Santa Cruz, California 95064, USA

## ABSTRACT

**Many proteins with unusual structural properties are comprised of multiple repeating amino acid sequences and are often fractious to expression in recombinant systems. To facilitate recombinant production of such proteins for structural and engineering studies, we have produced circular messenger RNAs with infinite open reading frames. We show that a circular mRNA containing a simple green fluorescent protein (GFP) open reading frame can direct GFP expression in *Escherichia coli*. A circular mRNA with an infinite GFP open reading frame produces extremely long protein chains, proving that bacterial ribosomes can internally initiate and repeatedly transit a circular mRNA. Only the monomeric forms of GFP produced from circular mRNA are fluorescent. Analysis of the translation initiation region shows that multiple sequences contribute to maximal translation from circular mRNA. This technology provides a unique means of producing a very long repeating-sequence protein, and may open the way for development of proteinaceous materials with novel properties.**

**Keywords: circular mRNA; long repeating protein; translation**

## INTRODUCTION

All known natural eukaryotic and prokaryotic messenger RNAs are linear. Although circular RNAs exist in nature, they either do not contain open reading frames (Symons, 1992), are products of mis-splicing events (Nigro et al., 1991; Cocquerelle et al., 1992; Capel et al., 1993; Pasman et al., 1996), or their capacity to translate proteins directly from their circular form is not clear (Chan et al., 1988; Kjems & Garrett, 1988; Dalgaard & Garrett, 1992; Dalgaard et al., 1993). Although not common in nature, circular RNA has features that may make it attractive in several biological applications. Circular RNA can accumulate to high levels in vivo (Chan et al., 1988; Harland & Misher, 1988; Ford & Ares, 1994; Puttaraju & Been, 1996), possibly through increased exonuclease resistance. Increased levels or persistence of a desired antisense RNA or ribozyme presented in circular form could improve its efficacy.

Circular RNAs might also direct the synthesis of proteins, thus providing a novel means of identifying sequence requirements for internal translation initiation (e.g., Chen & Sarnow, 1995), and potentially leading to the production of recombinant forms of novel proteinaceous materials. For example, the eukaryotic proteins that comprise spider dragline silk (Prince et al., 1995; Lewis et al., 1996; Fahnestock & Bedzyk, 1997; Fahnestock & Irwin, 1997), cocoon silk (Oshimi & Suzuki, 1977), or mollusc shell framework (Sudo et al., 1997) contain contiguous units of consensus repeat sequence that are difficult to maintain and express in bacteria (Prince et al., 1995; Lewis et al., 1996; Fahnestock & Bedzyk, 1997; Fahnestock & Irwin, 1997; Oshimi & Suzuki, 1977). If circular bacterial mRNAs could be created with a single in-frame repeat unit of the coding sequence, then translating ribosomes might circuit the mRNA numerous times to produce a multiple repeating-sequence protein. Development of this aspect of circular RNA using artificial repeating protein sequences might lead to the production of novel protein-based materials. Identifying and studying RNA sequences capable of initiating internal ribosome entry in both prokaryotes and eukaryotes would also benefit from circular mRNA technology whereby direct internal initiation is isolated from other modes of ribosome entry (Chen & Sarnow, 1995; Iizuka et al., 1995).

To explore the function of circular mRNAs in vivo, we applied circular RNA technology (Puttaraju & Been,

Reprint requests to: M. Ares, Center for Molecular Biology of RNA, Biology Department, Sinsheimer Laboratories, University of California, Santa Cruz, Santa Cruz, California 95064, USA; e-mail: ares@biology.ucsc.edu.

1992; Ford & Ares, 1994) to ask whether circular mRNA translation could occur in *Escherichia coli*. Ribosomes from prokaryotes (Kozak, 1979; Konarska et al., 1981) have been shown previously to bind circular RNAs in vitro. Furthermore, it has been demonstrated that eukaryotic ribosomes can both bind and initiate translation on circular RNAs in vitro (Chen & Sarnow, 1995). Here we show that circular RNAs can function as mRNAs in vivo in *E. coli*, provided they contain the necessary translation initiation signals. A 795-nt circular mRNA with an infinite 30-kDa-encoding open reading frame produces proteins greater than 300 kDa in size, indicating that the prokaryotic ribosomes can transit the circle more than 10 times. These results demonstrate the utility of circular mRNA technology for the expression of very large repeating-sequence polypeptides.

## RESULTS

### Design of genes for expression of circular mRNA

In vivo expression of a rearranged group I intron can produce circular RNA through the splicing activity of the autocatalytic group I RNA elements (Ford & Ares, 1994; Puttaraju & Been, 1996). We have called this arrangement of group I elements an "RNA cyclase ribozyme" because the group I elements render an internal sequence circular by the autocatalytic splicing reactions. A general-purpose plasmid for creating RNA cyclase ribozymes that produce desired circular RNAs was created by inserting multiple cloning sites into the region that becomes circular after the phage T4 td group I intron reactions (Ford & Ares, 1994; see Fig. 1). An open reading frame (ORF) encoding green fluorescent protein (GFP; Siemering et al., 1996) was inserted into this polylinker (Fig. 1A). To ensure that no ribosomes could translate the GFP ORF from unspliced linear RNAs, the natural GFP AUG was mutated to UUA. A sequence designed for translation initiation was placed downstream of the GFP ORF in the transcription unit (Fig. 1A). This sequence contains a "Shine–Dalgarno sequence" (SD; Shine & Dalgarno, 1974; Steitz & Jakes, 1975; Gold, 1988), an AUG codon, and a "downstream box" (DB; Sprengart et al., 1996). Within the RNA circles formed after transcription and the RNA cyclase reactions, the translation initiation sequence is fused in-frame with (and now upstream of) the GFP ORF, to produce a 795-nt circular mRNA transcript from which a ~30-kDa GFP species can be made (plasmid pGFPT, Fig. 1B). In a second construct, the GFP stop codon was mutated to code for tyrosine. This circular RNA has an infinite ORF in the reading frame of the initiating AUG (pGFPTI, Fig. 1C). As a control, we made a GFP construct with the equivalent translation initiation sequences fused upstream of GFP, expressed as a linear RNA species without the group I elements (pTGFP).

### The desired circular RNAs are produced in vivo

Previously, circular RNAs 100 nt or less have been produced in *E. coli* (Ford & Ares, 1994; Puttaraju & Been, 1996). To determine whether larger circular RNAs can be produced in vivo, we introduced the plasmids described above into *E. coli* [BL21]. RNA was isolated and analyzed by northern blot hybridized with a GFP probe (Fig. 2). A 795-nt circular RNA is observed in cells carrying either the pGFPT or pGFPTI plasmids (Fig. 2, lanes 4, 6), which comigrates with the circular RNA obtained in vitro by T7 transcription of pGFPT and the RNA cyclase reaction of the transcripts (Fig. 2, lane 1). Plasmids from which the group I intron 5′ half is deleted do not express circular RNA in vivo (Fig. 2, lanes 5, 7). Consistent with previous observations (Ford & Ares, 1994; Puttaraju & Been, 1996), the circular RNA accumulates to high levels in vivo. Our analysis suggests an average of 180 circular molecules are present per *E. coli* cell (data not shown). In contrast, the linear control GFP transcript is not detectable under these conditions, possibly because it may be heterogeneous in length (Fig. 2, lane 3). We conclude that our strains express the desired 795-nt circular RNAs, and that each circular RNA accumulates to high levels.

### Circular RNA can be translated in vivo

We analyzed *E. coli* strains expressing circular RNAs for GFP expression on western blots probed with an antibody to GFP (Fig. 3A). In cells expressing the circular RNA with the finite GFP ORF, we observe a protein of 30 kDa, consistent with expression of GFP (Fig. 3A, lane 3). On a per cell basis, the efficiency of GFP translation from the circular mRNA is ~20% of that from the equivalent linear mRNA (Fig. 3A, compare lanes 1 and 3), although per RNA molecule the efficiency of circular RNA translation may be far less. No GFP is detected in strains containing vector alone (Fig. 3A, lane 2), or containing plasmid lacking the translation initiation sequence (Fig. 3A, lane 4). A derivative of pGFPT from which the group I intron 5′ half is deleted and that does not express circular RNA (Fig. 2, lane 5) does not produce detectable GFP (Fig. 3A, lane 5), indicating that GFP expression requires RNA circularization. We conclude that GFP expression requires the translation initiation sequence to be upstream of the GFP ORF, an event that occurs only after RNA circularization.

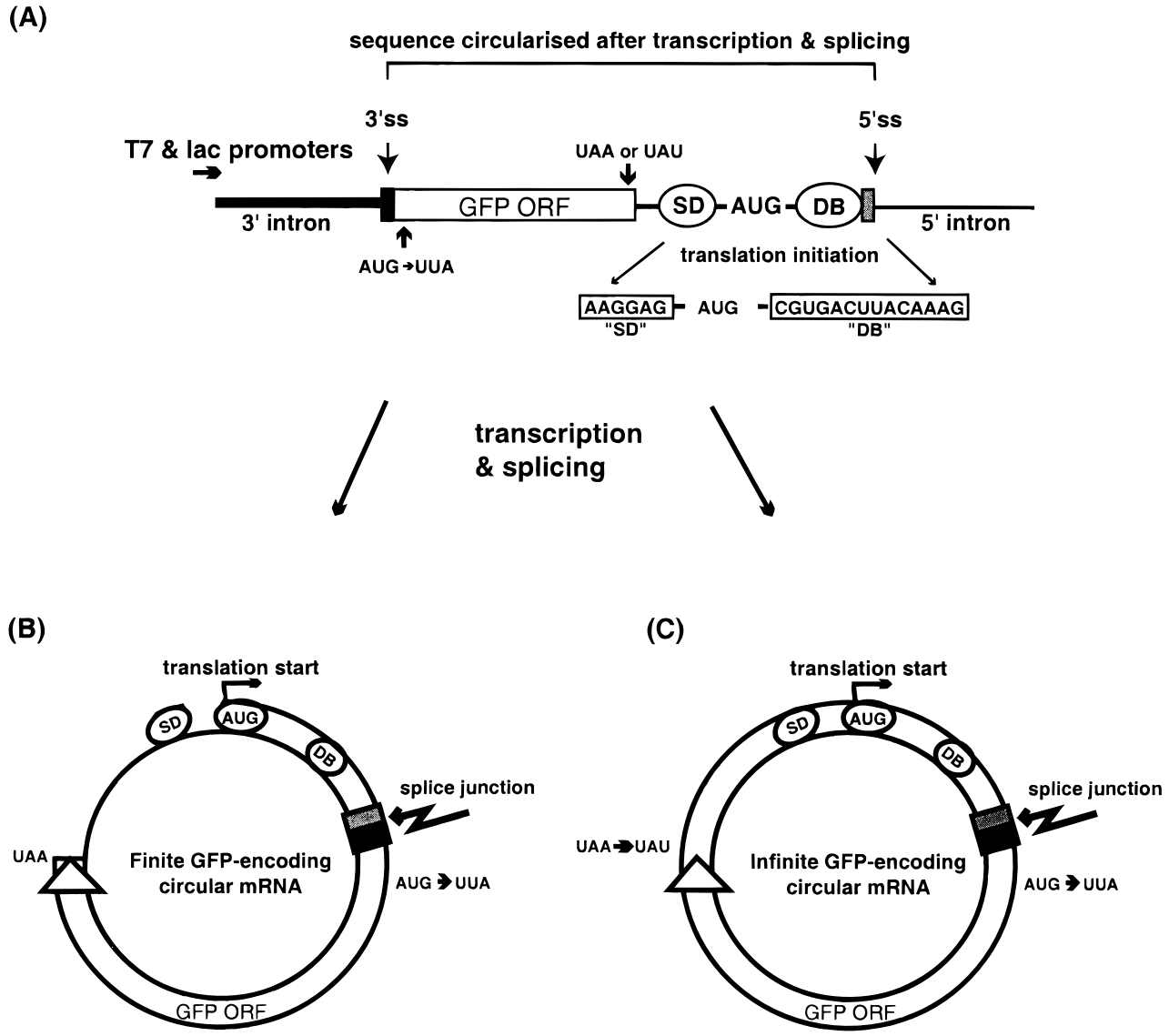


**FIGURE 1.** Plasmid containing rearranged group I intron elements for expression of GFP from circularizing mRNAs in *E. coli.* Transcription and splicing results in circularization of the bracketed sequence, between the 3′ (3′ss) and 5′ (5′ss) splice sites. **A:** Relevant region of pGFPT (finite GFP-encoding) and pGFPTI (infinite GFP-encoding) in which the GFP-ORF and translation initiation region have been inserted. Enlarged, boxed nucleotides are the SD and DB motifs of the translation initiation sequence. pGFPT has UAA termination codon indicated, pGFPTI has UAU at this position. Other abbreviations are: GFP, green fluorescent protein open reading frame; SD, Shine–Dalgarno; AUG, initiating codon; DB, downstream box. **B:** Circularized finite GFP-encoding mRNA created after transcription and splicing from plasmid pGFPT. This mRNA encodes a ~30-kDa GFP species. Jagged arrow indicates fused 3′/5′ splice junction, bent arrow is translation start, UAA is the translation stop signal. **C:** Circularized infinite GFP-encoding mRNA created after transcription and splicing from plasmid pGFPTI. This mRNA is devoid of stop codons in the GFP reading frame. Translation stop signal is replaced with UAU codon (arrow). Translated portion of circular mRNAs is shown as double circle.

### Circular RNA carrying an infinite ORF can produce huge polyproteins in vivo

To rigorously prove that translation initiation can occur on circular RNA in vivo, we analyzed the construct from which all known translation termination signals are removed (Fig. 3B). The circular RNA produced in this strain has the potential to direct synthesis of an infinitely long protein (see Fig. 1). In extracts from this strain, the GFP antibody detects a heterogeneous collection of proteins ranging in size from ~50 to more than 300 kDa (Fig. 3B, lane 3). Control constructs lacking either the engineered translation initiation sequence (Fig. 3B, lane 4) or the 5′ intron half (Fig. 3B, lane 5) do not produce detectable GFP-related proteins. We conclude that *E. coli* ribosomes can bind and circuit the infinite GFP-encoding circular RNA as many as 10 times to produce repeating GFP derived exclusively from the circular mRNA templates in vivo.

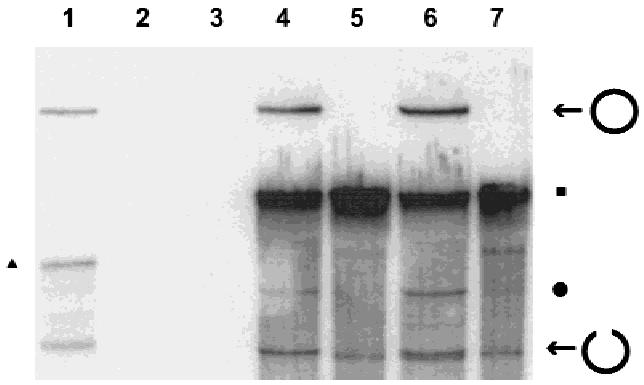


**FIGURE 2.** Expression and accumulation of large circular mRNAs in *E. coli*. Northern blot of in vitro (lane 1) and in vivo (lanes 2–7) RNA from strains expressing GFP-encoding circularizing mRNAs. Lane 1, 10 pg in vitro-transcribed RNA from pGFPT; lane 2, vector control; lane 3, linear GFP mRNA; lane 4, finite GFP-encoding circular mRNA; lane 5, finite GFP-encoding circular mRNA lacking the 5′ half group I element; lane 6, infinite GFP-encoding circular mRNA; lane 7, infinite GFP-encoding circular mRNA lacking the 5′ half group I intron element. Arrows indicate circular and nicked circular GFP-encoding mRNAs; square, in vivo unspliced precursor; triangle, in vitro unspliced precursor; filled circle, splicing intermediate after cleavage at the 5′ splice site. The in vitro unspliced precursor is shorter than the in vivo derivative because it derives from linearized plasmid.

### Maximal circular mRNA translation requires both the Shine–Dalgarno and downstream box motifs

The contributions of the SD and DB motifs to GFP expression from circular mRNAs were analyzed by mutating them either individually or together (Fig. 4). For the finite GFP-encoding RNA, changing the SD sequence abolishes detectable GFP translation (Fig. 4A, lane 2), consistent with the general importance of this sequence in recruiting ribosomes to mRNA in bacteria (Steitz & Jakes, 1975). Changing the DB, but maintaining the SD sequence, reduces GFP expression by ~60% (Fig. 4A, compare lanes 1 and 3). Altering both sequences also eliminates GFP production (Fig. 4A, lane 4).

The same mutations in the infinite GFP-encoding mRNA reduce but do not eliminate repeating GFP translation (Fig. 4B). Altering the SD sequence reduces production of repeating GFP by ~50% (Fig. 4B, compare lanes 2 and 3), whereas mutating the DB, but maintaining the SD sequence reduces translation by ~10% (Fig. 4B, compare lanes 2 and 4). When both sequences are mutated, ~40–50% of original repeating GFP levels are still observed (Fig. 4B, lane 5). To ensure that the reduction in GFP protein is not due to lower in vivo levels of accumulated circular mRNA, RNA from *E. coli* strains expressing the altered circular RNAs was analyzed. The mutations caused no significant reduction in accumulated circular RNAs (data not shown), indicating that the effect of the mutations is on translation. We conclude that the presence of both sequence motifs in the translation initiation sequence of the circular mRNAs produces maximal protein levels, and that the SD sequence contributes more than the DB motif. The apparent effects of the mutations are minimized during expression from an infinite ORF compared with a finite ORF (compare Fig. 4A and B). Because the yield of GFP per translation initiation event is expected to be comparatively higher during translation

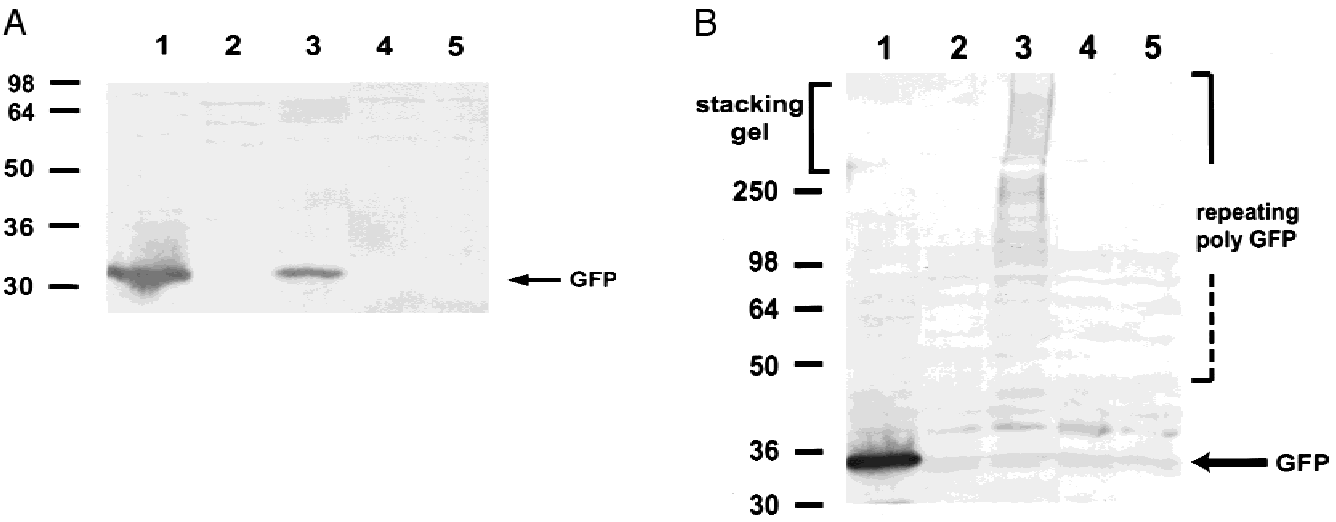


**FIGURE 3.** Translation of GFP from circular mRNAs in *E. coli*. Western blot for GFP expression from strains expressing various finite (**A**) and infinite (**B**) GFP-encoding circular mRNAs, and linear control mRNA. **A:** Lane 1, linear GFP mRNA control; lane 2, vector control; lane 3, finite GFP-encoding circular mRNA; lane 4, finite GFP-encoding circular mRNA without translation initiation sequence; lane 5, finite GFP-encoding lacking the 5′ half group I intron element. Arrow indicates "GFP" protein species. Protein size markers are indicated. **B:** As for A, except that infinite-encoding circular mRNAs are being expressed. Brackets indicate 4% stacking gel and heterogeneous repeating GFP proteins.

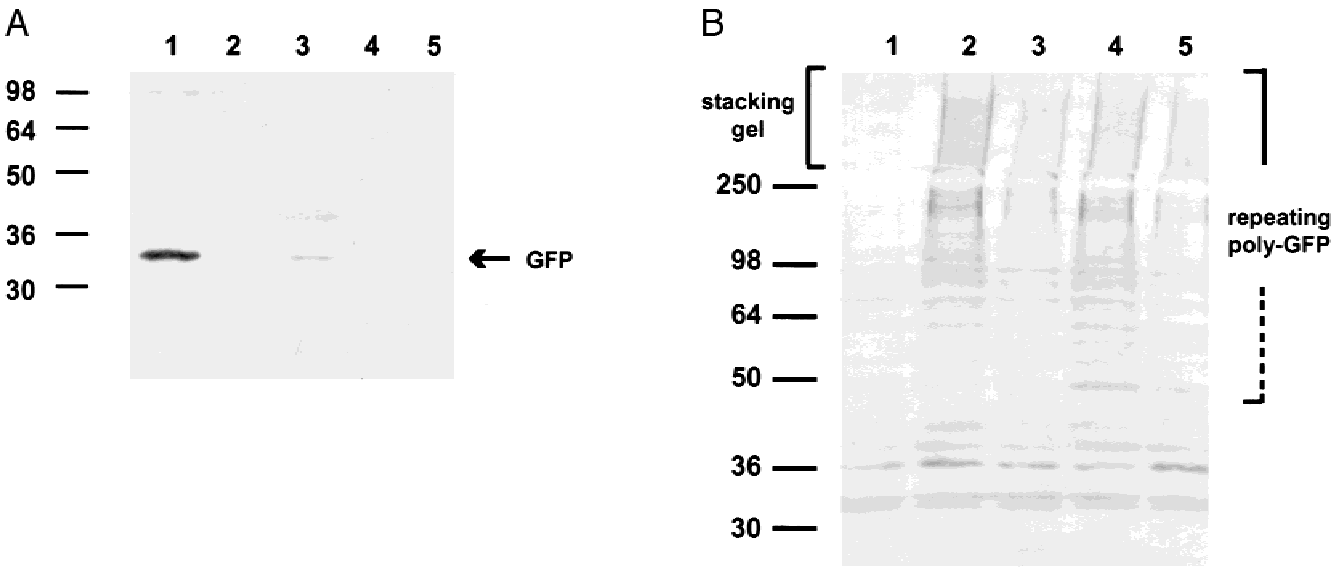


**FIGURE 4.** Relative efficiencies of translation from SD and DB motifs in the translation initiation sequence from strains expressing various finite (**A**) and infinite (**B**) GFP-encoding circular mRNAs. **A:** Lane 1, finite GFP-encoding circular mRNA; lane 2, finite GFP-encoding circular mRNA lacking SD sequence; lane 3, lacking DB sequence; lane 4, lacking both SD and DB sequences; lane 5, vector control. Arrow indicates GFP protein species. **B:** Lane 1, vector control; lane 2, infinite GFP-encoding circular mRNA; lane 3, infinite GFP-encoding circular mRNA lacking SD sequence; lane 4, infinite GFP-encoding circular mRNA lacking DB sequence; lane 5, infinite GFP-encoding circular mRNA lacking both SD and DB sequences. Brackets indicate 4% stacking gel and heterogeneous GFP polyproteins.

of an infinite ORF, these results indicate that initiation is the rate-limiting step in translation from circular mRNAs.

### Monomeric GFP translated from circular mRNA produces an active chromophore in vivo

Correctly folded and modified GFP emits green light upon photoexcitation (Chalfie et al., 1994). GFP fluorescence from *E. coli* strains expressing either the 30-kDa or "infinite" GFP from circular mRNAs was assayed by observing *E. coli* colonies on a UV lightbox, under the light microscope, or by fluorimetry. GFP fluorescence is observed from the 30-kDa-expressing circular mRNA but not from *E. coli* strains expressing "infinite" GFP (data not shown). We conclude that a significant proportion of the 30-kDa protein expressed from the circular mRNA in *E.coli* is present as correctly folded and processed GFP protein.

## DISCUSSION

We have shown that a rearranged group I intron, derived from a phage T4 gene (Ford & Ares, 1994), can be used to synthesize large circular mRNAs in *E. coli.* These circular mRNAs can recruit ribosomes directly via specific engineered sequences and are abundant and stable, but seem inefficiently translated on a per molecule basis when compared with an equivalent linear mRNA. When the circular RNAs contain infinite open reading frames, they can be translated to produce very long repeating polypeptides, indicating that prokaryotic ribosomes can transit the circle multiple times. This technology may aid the production of repeat-sequence recombinant proteins with unusual physical characteristics.

### A repeating protein expression system based on circular mRNAs

We have shown that it is possible, using circular mRNAs, to produce proteins that are much larger than the size of the largest known *E. coli* protein (Pasquali et al., 1996). Most repeating GFP produced in this study is more than 300 kDa and does not enter the separating polyacrylamide gel. To our knowledge, these are the largest recombinant proteins produced in *E. coli.* A lack of both good markers and resolving gels for this size range means we have no good estimate for the true size of our repeated GFP proteins or what the upper limit for the size of a single protein molecule might be. Estimating the amounts of these proteins by western blotting is also of limited accuracy because we are unsure about transfer efficiencies for proteins this large.

What kinds of proteins might be produced using this type of expression system? Proteins similar to those containing repeat peptide motifs (e.g., Oshimi & Suzuki, 1977; Prince et al., 1995; Lewis et al., 1996; Brown, 1997; Fahnestock & Bedzyk, 1997; Fahnestock & Irwin, 1997) could be expressed using an infinite-encoding circular mRNA. One example is the protein compo-

nents in spider dragline silk, spidroin 1 and 2, whose recombinant production is highly desirable for analysis of mechanical and structural properties, but has been problematic due to difficulties maintaining and expressing tandem arrays of repeating sequence in *E. coli* (Prince et al., 1995; Lewis et al., 1996; Fahnestock & Bedzyk, 1997; Fahnestock & Irwin, 1997; Oshimi & Suzuki, 1977). To date, recombinant systems expressing ligated tandem arrays of repeating spidroin 1 consensus motifs in both *E. coli* (Prince et al., 1995; Lewis et al., 1996; Fahnestock & Irwin, 1997) and a methylotrophic yeast (Fahnestock & Bedzyk, 1997) have produced peptides less than 160 kDa. An infinite-encoding circular mRNA expressing a spidroin consensus motif could allow the in vivo translation of much longer spidroin-like proteins. Application of this technology is not limited to proteins with unusual structural properties. Repeating versions of enzymes used in surface coating or column-bound form may retain more activity upon crosslinking because fewer potentially inactivating crosslinks will be needed to covalently attach a very large protein to the solid matrix.

### What limits the length of translation products derived from infinite ORFs?

Infinite-encoding circular mRNAs provide a unique window for studying protein termination in *E. coli*. In our samples, we see a heterogeneous size distribution of repeating GFP, suggesting that translation termination might be occurring at random, or that much larger proteins are being broken by cleavage at many sites, either in vivo or during protein isolation procedures. We believe that translation termination and not multiple sites of initiation is the cause of the heterogenous GFP products because mutagenesis of the engineered initiation region significantly diminishes GFP protein products, demonstrating that this is the major site of ribosome recruitment. This is particularly apparent from the 30-kDa-encoding circular species for which a new initiation event is required for translation of all GFP molecules.

What could be causing termination in the absence of stop codons? One possibility is that frame-shifting by a ribosome could cause it to encounter an out-of-frame stop codon and terminate. Alternatively, a low rate of spontaneous error in codon recognition might direct untimely ribosome release at sense codons. There is also evidence in *E. coli* that sequences other than the stop codon can affect translation termination (Poole et al., 1995). Finally, in those cases where the circular mRNA becomes nicked, ribosomes that encounter the end of the formerly circular RNA would be expected to terminate translation via the 10Sa pathway (Tu et al., 1995). Given these possible mechanisms for termination in the absence of a stop codon, there may be a practical upper limit for the size of proteins *E. coli* can produce. Nonetheless, 300 kDa is a conservative estimate of the size of the repeating GFP produced, and is well beyond the size of the largest natural *E. coli* protein. Indeed, the size limit may equal or surpass the largest known eukaryotic proteins.

### Translation initiation on circular mRNAs is inefficient and rate limiting

The translational efficiency of circular mRNAs is poor when compared with a similar linear RNA species (Fig. 3), and is most likely due to inefficient initiation (Fig. 4). Why might initiation be less efficient? One possibility is that the reaction required to create circular mRNAs limits ribosome recruitment until after transcription and circularization have taken place. In *E. coli*, translation is normally coupled to transcription, and this coupling can contribute to protein expression (Lopez et al., 1994; Iost & Dreyfuss, 1995). One means of improving protein expression from circular mRNA might be the use of *E. coli* strains in which transcription and translation are no longer coupled (Jensen, 1988; Chen et al., 1997). A second possibility is that the lack of a 5′ end reduces translation initiation on the circular mRNAs. Although there is little evidence for a role of 5′ ends of mRNA contributing to translation in *E. coli*, there is also little evidence against it. This may also mean that the Shine–Dalgarno and downstream box used in this study might not be best for exclusively internal initiation. We are interested in identifying other RNA sequence motifs that can enhance direct ribosome recruitment and thus improve circular mRNA translation. Finally, a circular mRNA is unusual because the beginning and the end of the same ORF are close to one another topologically. This may create a unique situation where initiation and termination of identical proteins are occurring simultaneously in the same space. If correct folding of nascent chains is necessary for termination and release, this situation could cause translation to proceed inefficiently.

## MATERIALS AND METHODS

### Cloning

Plasmid pRR1 derives from pEFCΔNde (Ford & Ares, 1994) with multiple cloning sites inserted between the 3′ and 5′ splice sites of the rearranged group I intron. The AUG on the GFP ORF was mutated by site-directed mutagenesis (Kunkel et al., 1987) to UUA to make pGFPNT. This GFP ORF was inserted as a blunt *Bst*B I–*Sac* I fragment into blunt *Nco* I–*Sac* I-digested pRRI to make pRGFP. The translation initiation sequence was inserted by site-directed mutagenesis (Kunkel et al., 1987) into pGFPNT and pRGFP using oligo SDGFP: 5′-ATTGACCTGA GATCGCTTTT TGCTTTGTAA GTCACGTTAG AGCTAGCCAT CTTGTGTCTC CTTGTGC AGA CCTCTCGAGC TCCAT-3′ to create pGFPT and pTGFP. The stop codon on pRGFP and pGFPT was mutated (Kunkel et al., 1987) to UAU to create pGFPI and pGFPTI, respec-

tively. The SD and DB motifs on pGFPT and pGFPTI were mutated (Kunkel et al., 1987) using the oligos SDKO: 5′-GATCGCTTTT TGGAAACATT CAGTGCTTAG AGCTA GCCA-3′, and DBKO: 5′-CCATCTTGT GTGAGGAAGT GCAGACCTC-3′, respectively.

### In vivo RNA analysis

RNA was isolated as described previously (Ford & Ares, 1994), after induction treatments identical to those for protein extraction. In vitro-transcribed RNAs of the finite GFP-encoding circular RNA (Milligan & Uhlenbeck, 1989) were markers for in vivo circular RNAs. One microgram of *E. coli* RNA was separated on denaturing 4% polyacrylamide gels, electroblotted to Nytran nylon filters (Schleicher and Schuell), and immobilized by cross-linking using a Stratalinker (Stratagene). Filters were incubated for 30 min at 42 °C in 10 mL of 5× Denhardts/5× SSC/50% formamide/0.1% SDS/sonicated herring sperm DNA (150 mg/mL) hybridization solution. One million cpm of random-primed (Amersham) [$\alpha$-$^{32}$P] dATP–GFP probe was denatured at 95 °C for 5 min, added to the hybridization buffer, and incubated overnight. Filters were washed at room temperature in 2× SSC/0.1% SDS, then 0.1× SSC/0.1% SDS, and at 60 °C in 0.1× SSC/0.1%SDS, then autoradiographed.

### Protein extraction

One-milliliter cultures of *E. coli* strain [BL21] expressing the various finite GFP-encoding or infinite GFP-encoding circular constructs were used to seed 5 mL of LB to $A_{600} = 0.05$. These were grown at 37 °C to an $A_{600} = 0.4$, at which time 1 mM IPTG was added. Cultures were grown a further 4 h and harvested for protein extraction. For the finite GFP-encoding constructs (Figs. 3A, 4A), cells were pelleted and resuspended in 400 $\mu$L 50 mM HEPES, pH 7.5, and lysed by sonication. The lysates were clarified by centrifugation at 10,000 × *g* for 10 min at 4 °C. Aliquots were frozen and stored at −70 °C. Sonication destroyed the repeating GFP proteins from the infinite-encoding constructs (data not shown). Protein for Figs. 3B and 4B was isolated as follows. Cells were pelleted and resuspended in 8 M urea/0.1 M $NaH_2PO_4$/10 mM Tris, pH 6, frozen on dry ice, heated to 90 °C for 5 min, and clarified by centrifugation at 10,000 × *g* for 10 min at 4 °C. Aliquots were stored as above.

### Western blots

Protein concentrations were determined using Bradford assay kit (Bio-Rad) and separated on 8% (for the infinite constructs) or 12.5% (for the finite constructs) separating/4% stacking polyacrylamide/SDS gels. Proteins were electrotransferred to nitrocellulose (Micron Sep. Inc) in 0.24% Tris/1.13% glycine/20% methanol. Blots were blocked in 3% bovine serum albumen in TBST (15 mM NaCl/100 mM Tris, pH 7.5/0.1% Tween) for 30 min at room temperature. Polyclonal $\alpha$-GFP antibody (Clontech) was added at 1:1,000 dilution in TBST for 2 h at room temp. Blots were washed in TBST and IgG-goat-anti-rabbit secondary antibody (Southern Biotech. Assoc.) was added at 1:1,000 dilution for 1 h at room temperature. Blots were washed in TBST and GFP products were visualized with conjugated alkaline phosphatase and 5-bromo-4-chloro-3-indolyl phosphate [BCIP] and nitroblue tetrazolium chloride [NBT] (Bio-Rad).

## ACKNOWLEDGMENTS

R.P. is supported by American Cancer Society Postdoctoral fellowship and in part by grant GM40478 from the National Institutes of Health. We thank Michelle Haynes Pauling, Marc Spingola, Ken Howe, and Roland Nagel for critical reading of this manuscript, and Haller Igel for technical support.

*Received April 16, 1998; returned for revision June 6, 1998; revised manuscript received June 12, 1998*

## REFERENCES

Brown S. 1997. Metal-recognition by repeating polypeptides. *Nature Biotech 15*:269–272.

Capel B, Swain A, Nicolis S, Hacker A, Walter M, Koopman P, Goodfellow P, Lovell-Badge R. 1993. Circular transcripts of the testis-determining gene *Sry* in adult mouse testis. *Cell 73*:1019–1030.

Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC. 1994. Green fluorescent protein as a marker for gene expression. *Science 263*:802–805.

Chan WKY, Belfort G, Belfort M. 1988. Stability of group I intron RNA in *Escherichia coli* and its potential application in a novel expression vector. *Gene 73*:295–304.

Chen C, Sarnow P. 1995. Initiation of protein synthesis by the eukaryotic translational apparatus on circular RNAs. *Science 268*:415–417.

Chen H, Ferbeyre G, Cedergreen R. 1997. Efficient hammerhead ribozyme and antisense RNA targeting in a slow ribosome *Escherichia coli* mutant. *Nature Biotech 15*:432–435.

Cocquerelle C, Daubersies P, Majerus MA, Kerckaert JP, Baileul B. 1992. Splicing with inverted order of exons occurs proximal to large introns. *EMBO J 11*:1095–1098.

Dalgaard JZ, Garrett RA. 1992. Protein encoding introns from the 23S rRNA-encoding gene form stable circles in the hyperthermophilic archaeon *Pyrobaculum organotrophum*. *Gene 121*:103–110.

Dalgaard JZ, Garrett RA, Belfort M. 1993. A site-specific endonuclease encoded by a typical archaeal intron. *Proc Natl Acad Sci USA 90*:5414–5417.

Fahnestock SR, Bedzyk LA. 1997. Production of synthetic spider dragline silk protein in *Pichia pastoris*. *Appl Microbiol Biotechnol 47*:33–39.

Fahnestock SR, Irwin SL. 1997. Synthetic spider dragline silk proteins and their production in *Escherichia coli*. *Appl Microbiol Biotechnol 47*:23–32.

Ford E, Ares M. 1994. Synthesis of circular RNA in bacteria and yeast using RNA cyclase ribozymes derived from a group I intron of phage T4. *Proc Natl Acad Sci USA 91*:3117–3121.

Gold L. 1988. Posttranscriptional regulatory mechanisms in *Escherichia coli*. *Annu Rev Biochem 57*:199–233.

Harland R, Misher L. 1988. Stability of RNA in developing *Xenopus* embryos and identification of a destabilizing sequence in TFIIA messenger RNA. *Development 102*:837–852.

Iizuka N, Chen C, Yang Q, Johannes G, Sarnow P. 1995. Cap-independent translation and internal initiation of translation in eukaryotic cellular mRNA molecules. *Curr Topics Microbiol Immunol 203*:155–177.

Iost I, Dreyfuss M. 1995. The stability of *Escherichia coli* lacZ mRNA depends upon the simultaneity of its synthesis and translation. *EMBO J 14*:3252–3261.

Jensen KF. 1988. Hyper-regulation of pry gene expression in *Escherichia coli* cells with slow ribosomes. Evidence for RNA polymerase pausing in vivo? *Eur J Biochem 175*:587–593.

Kjems J, Garrett RA. 1988. Novel splicing mechanism for the ribosomal RNA intron in the archaebacterium *Desulfurococcus mobilis*. *Cell 54*:693–703.

Konarska M, Filipowicz W, Domdey H, Gross HJ. 1981. Binding of ribosomes to linear and circular forms of the 5′-terminal leader fragment of tobacco-mosaic-virus RNA. *Eur J Biochem 114*:221–227.

Kozak M. 1979. Inability of circular mRNA to attach to eukaryotic ribosomes. *Nature 280*:82–85.

Kunkel TA, Roberts JD, Zakour RA. 1987. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods Enzymol 154*:367–382.

Lewis RV, Hinman M, Kothakota S, Fournier MJ. 1996. Expression and purification of a spider silk protein: A new strategy for producing repetitive proteins. *Protein Exp & Purif 7*:400–406.

Lopez PJ, Iost I, Dreyfuss M. 1994. The use of a tRNA as a transcriptional reporter: The T7 late promoter is extremely efficient in *Escherichia coli* but its transcripts are poorly expressed. *Nucleic Acids Res 22*:1186–1193.

Milligan JF, Uhlenbeck OC. 1989. Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol 180*:51–62.

Nigro JM, Cho KR, Fearon ER, Kern SE, Ruppert JM, Oliner JD, Kinzler KW, Vogelstein B. 1991. Scrambled exons. *Cell 64*:607–613.

Oshimi Y, Suzuki Y. 1977. Cloning of the silk fibroin gene and its flanking sequences. *Proc Natl Acad Sci USA 74*:5363–5367.

Pasman Z, Been MD, Garcia-Blanco MA. 1996. Exon circularization in mammalian nuclear extracts. *RNA 2*:603–610.

Pasquali C, Frutiger S, Wilkins MR, Hughes GJ, Appel RD, Bairoch A, Schaller D, Sanchez JC, Hochstrasser DF. 1996. Two-dimensional gel electrophoresis of *Escherichia coli* homogenates: The *Escherichia coli* SWISS-2DPAGE database. *Electrophoresis 17*:547–555.

Poole ES, Brown CM, Tate WP. 1995. The identity of the base following the stop codon determines the efficiency of in vivo translational termination in *Escherichia coli*. *EMBO J 14*:151–158.

Prince JT, McGrath KP, DiGirolamo CM, Kaplan DL. 1995. Construction cloning and expression of synthetic genes encoding spider dragline silk. *Biochemistry 34*:10879–10885.

Puttarju M, Been MD. 1992. Group I permuted intron–exon (PIE) sequences self-splice to produce circular exons. *Nucleic Acids Res 20*:5357–5364.

Puttarju M, Been MD. 1996. Circular ribozymes generated in *Escherichia coli* using group I self-splicing permuted intron–exon sequences. *J Biol Chem 271*:26081–26087.

Shine J, Dalgarno L. 1974. The 3′-terminal sequence of *Escherichia coli* 16S ribosomal RNA: Complementarity to nonsense triplets and ribosome binding sites. *Proc Natl Acad Sci USA 71*:1342–1346.

Siemering KR, Golbik R, Sever R, Haseloff J. 1996. Mutations that suppress the thermosensitivity of green fluorescent protein. *Curr Biol 6*:1653–1663.

Sprengart ML, Fuchs E, Porter AG. 1996. The downstream box: An efficient and independent translation initiation signal in *Escherichia coli*. *EMBO J 15*:665–674.

Steitz JA, Jakes K. 1975. How ribosomes select initiator regions in mRNA: Base pair formation between the 3′ terminus of 16S rRNA and the mRNA during initiation of protein synthesis in *Escherichia coli*. *Proc Natl Acad Sci USA 72*:4734–4738.

Sudo S, Fujikawa T, Nagakura T, Ohkubo T, Sakaguchi K, Tanaka M, Nakashima K, Takahashi T. 1997. Structures of mollusc shell framework proteins. *Nature 387*:563–564.

Symons RH. 1992. Small catalytic RNAs. *Annu Rev Biochem 61*:641–671.

Tu GF, Reid GE, Zhang JG, Moritz RL, Simpson RJ. 1995. C-terminal extension of truncated recombinant proteins in *Escherichia coli* with a 10Sa RNA decapeptide. *J Biol Chem 270*:9322–9326.