JOURNAL OF BACTERIOLOGY, June 1988, p. 2860–2865
0021-9193/88/062860-06$02.00/0

Vol. 170, No. 6

Evidence for Endonucleolytic Cleavages in Decay of *lacZ* and *lacI* mRNAs

MAKAM N. SUBBARAO† AND DAVID KENNELL*

Department of Microbiology and Immunology, Box 8093, Washington University School of Medicine, St. Louis, Missouri 63110

Received 20 November 1987/Accepted 14 March 1988

S1 nuclease mapping revealed *lacZ* mRNA molecules whose 5′ and 3′ ends were internal to the transcription start and consistent with cleavages at pyrimidine-adenosine bonds 20 to 50 nucleotides apart. With the net 5′-to-3′ direction known, *lacZ* mRNA is probably degraded by sequential cleavages of naked mRNA at vulnerable sites exposed by transit of the last translating ribosome.

The biochemistry of mRNA breakdown has continued to be a difficult problem to solve, although key aspects of the process have been known for some time (for a review, see reference 12). The three contiguous messages (*lacZ*, *lacY*, and *lacA*) from the lactose (*lac*) operon are made as a single polycistronic mRNA in *E. coli*. However, *lac* mRNA molecules with 5′ and 3′ ends that map in the interior of *lacY* mRNA are sufficiently more abundant than average in the decaying population for isolation and sequencing (4). Those "internal" ends are separated by 30 to 50 nucleotides, with a strong preference for A at 5′ ends and for U at 3′ ends (4). We have proposed a model for mRNA degradation in which primary cleavages, usually between pyrimidine-adenosine (Pyr-A) bonds, release large oligonucleotides that are then degraded to mononucleotides by secondary RNase activities (4). Those studies could not establish a temporal sequence or direction of such cleavages. While some messages can be inactivated at their 3′ ends (8, 9, 15, 17), the only message whose breakdown is known to proceed in a specific direction is the *lacZ* mRNA, which was shown to be degraded from the 5′ to the 3′ end (2). The presence of molecules of this mRNA with specific internal ends would be consistent with cleavages following the last translating ribosome down the inactivated message (4, 12).

Molecules with internal ends were detected by S1 nuclease mapping as described previously (4), except that 100 rather than 300 U of S1 nuclease was used. It has been shown that the ends are not generated by S1 cleavages of the DNA-RNA duplex (4). The large 789-base-pair (bp) *Hin*cII DNA fragment (Fig. 1) was 5′-end labeled to detect specific 5′ ends within the first 479 nucleotides of *lacZ* RNA. Molecules with the original 5′ transcription end were at least 100 times more abundant than any molecules with a specific end inside this region; the latter could only be observed with longer film exposure times (Fig. 2). A ladder of molecular weight markers generated a smooth curve of log nucleotide length versus migration distance, and the 5′ ends could be mapped generally to within one or two nucleotides.

Other DNA probes (Fig. 1) were used for comparison (Fig. 2). We were especially interested in the early region of the mRNA, and Fig. 2C concentrates on the first 100 nucleotides. In general, the results were very reproducible and there was agreement between probes with respect to positions and relative intensities of the major 5′-ended species (Fig. 2).

We also looked for *lacZ* molecules with 3′ ends inside the mRNA, using [3′-^{32}P]DNA labeled by filling a staggered end in the Klenow fragment reaction. It was not possible to look for 3′ ends in the most proximal region, since the 3′-^{32}P-labeled end of DNA must be covered by the 5′ end of the RNA in order to protect the ^{32}P. The nearest suitable end site was a *Hin*fI target 54 bp downstream from the *lacZ* start (Fig. 1). The mapping showed molecules with different 3′ ends throughout the region covered by the 338-bp probe (Fig. 3A).

The distal region of the first 479 bp of *lacZ* mRNA also contains molecules with 5′ or 3′ ends (Fig. 3B and C). Two broad 3′ bands were especially strong. It became apparent with shorter exposures that one of them was really two bands at 463 and 466 nucleotides and that the other contained multiple bands in the region of 416 to 423 nucleotides. 5′-End labeling of this 202-bp DNA fragment or the 789-bp DNA showed strong bands in the same 420-nucleotide region. This region contains a continuous stretch of Pyr-A and U-U bonds, which were shown by sequence analysis of *lacY* mRNA to be especially vulnerable to cleavage (4), suggesting above average cleavages in this 420-nucleotide region. Abundant 3′ ends could also arise from transcription pausing or termination. Richardson and Ruteshouser (16) observed 3′-ended molecules generated during in vitro transcription of this region of *lac* DNA in the presence of Rho protein. A major end was in the 466-nucleotide region. While it represents a small fraction of the in vivo *lac* mRNA seen here (<5% from densitometer tracings), it is possible that some termination or pausing occurs in vivo to account for the 463-nucleotide bands even in nonpolar strains.

The constitutive *lacI* mRNA also contained 5′-ended molecules in its early region (data not shown). This is the third mRNA we have seen that includes molecules with specific internal ends that are sufficiently abundant to stand out in the population. Their spacing was similar to that seen in *lacZ* and *lacY* (4) mRNAs. The positions of these ends was not random as was shown by their reproducibility in different strains and under different growth conditions. The same band pattern was seen in cells induced or uninduced for *lac* operon expression and also in uninduced cells with a mutation in the *lac* promoter.

The observed 5′ ends could conceivably result from weak transcription initiations within the *lac* genes. However, they are not present in cells uninduced for *lac*, and generally they

* Corresponding author.
† Present address: Department of Clinical Pathology, The Oregon Health Sciences University, Portland, OR 97201.

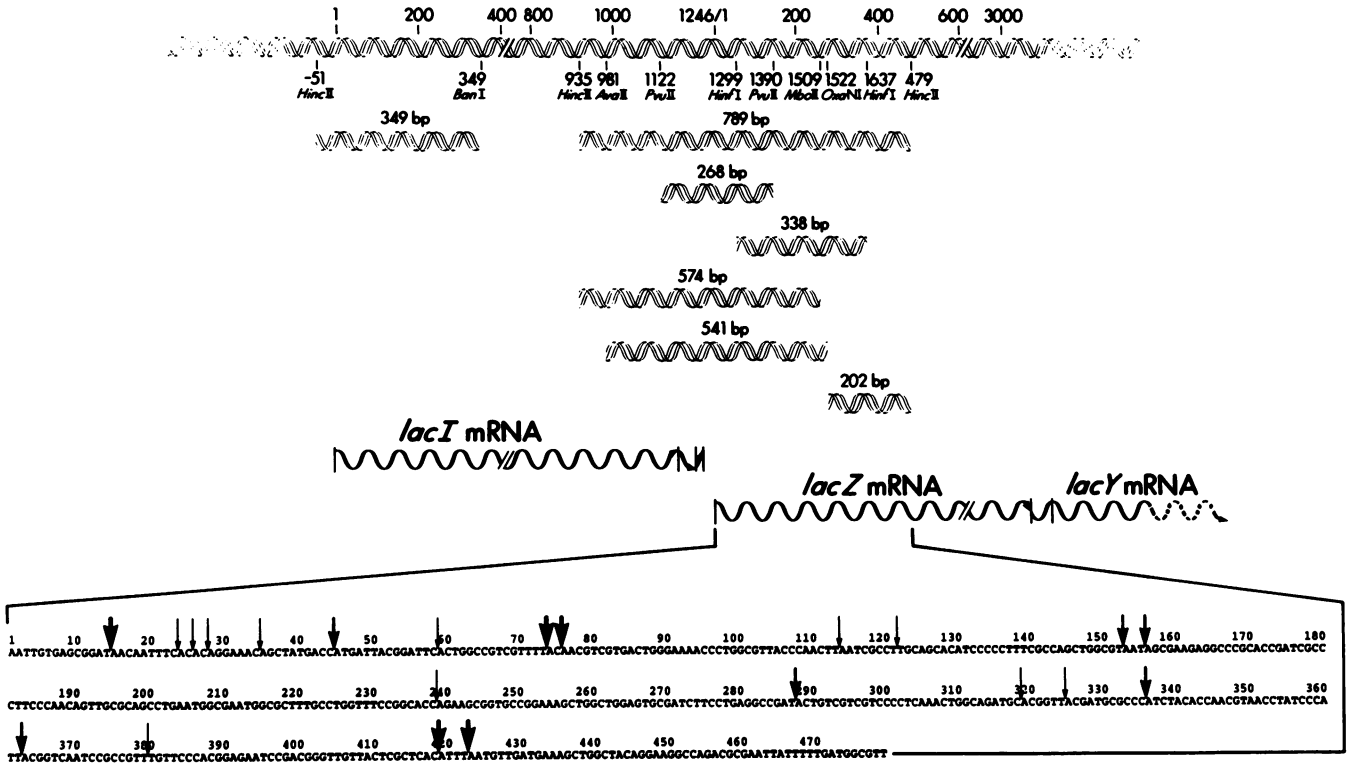

FIG. 1. *lacI* gene, part of the *lacZ* gene, and the intervening control region of the *lac* operon. The numbering system above the topmost DNA has +1 at the left end for the start of the *lacI* mRNA (20) and gives the number of base pairs down the DNA up to 1246, which is the major start of *lacZ* mRNA in vitro (14) and in vivo (3) and the start of the new numbering system of +1 for *lacZ* mRNA. The numbers below that DNA show the restriction enzyme sites that were the targets for generation of the DNA fragments shown below and used in the experiments in this study. The *lacI* mRNA is shown with three proposed termination sites (6), although more recent work suggests that some in vivo *lacI* transcripts may read into the *lacZ* gene (19). The *lacZ* gene ends at position 3063 (7, 10) and *lacY* starts at position 3117 (1) (vertical lines), but almost all of the *lacZ* transcripts continue into *lacY* in vivo with this strain and growth conditions (4). The sequence of the first 479 nucleotides of *lacZ* mRNA is shown. The arrows above the sequence show the approximate cleavages for some of the most reproducible 5′-ended RNA molecules (see text). The relative intensities of the arrows indicate the relative intensities of the S1 nuclease gel bands, which in turn reflect the half-lives for cleavage at the site and for survival of the resulting 5′ end.

are not proceded by the usual −10 and −35 consensus sequences for initiation. They could also conceivably arise by 5′ exonucleolytic activity, but there is no evidence for a 5′ exo-RNase in *E. coli*. Thus, we think that cleavage is much more likely. Similarly, the 3′ ends could result from transcription pausing or termination, as well as from cleavage. However, it seems unlikely that such a large proportion of polymerases would be pausing or terminating at any instant. About half of the completed *lacZ* molecules are being degraded, and degradation starts at the 5′ end (2); the combination of such extensive premature termination with degradation would not be consistent with about one-third of completed *lacZ* mRNA molecules being full length (11, 13).

We were not successful in efforts to sequence internal ends of *lacZ* mRNA, except in the CA-rich region at 20 to 25 nucleotides (3). One reason for this difficulty is that the *lacZ* mRNA is degraded twice as slowly as *lacY* mRNA (4), and the intensity of cleavage ends should be related to rate of degradation. The original transcription initiation 5′ ends with AAUUG or AUUG accounted for more than 90% of the 5′ ends on the 479 nucleotides (3). The *lacY* ends appeared to be generated by cleavages that usually occurred between Pyr-A residues (4). Although there is less certainty from S1 nuclease mapping, these *lacZ* ends were also probably derived from Pyr-A cleavages. Secondary-structure analyses of the *lacY* region suggest that such cleavages only occur in single strands (4). Different representative lengths of *lacZ* RNA, all starting at the 5′ end, were analyzed by computer (21) for most likely structures (Fig. 4). The most proximal secondary 5′ end was observed with consistency at position 15 on the 574-, 541-, and 268-bp DNAs. (This end was too close to the full-length molecule to observe as a discrete band with the 789-bp DNA.) The U-A bond between nucleotides 14 and 15 was in a single-stranded loop whose single-strandedness was remarkably persistent in molecules that ranged in length from those containing only the nontranslated leader region to ones which would have the last ribosome more than 200 nucleotides downstream (Fig. 4).

The next band seen consistently with all probes mapped at about 25 nucleotides from the start. This position is within three consecutive pairs of CA residues and was the only internal 5′ end identified by sequencing (3). The secondary-structure analyses did not show this region to be in single strands as often as the more proximal region containing nucleotides 14 and 15, which might account for it being weaker. At the same time, the 5′ end from cleavage at the 14-15 bond would have been hard to detect in the sequence analysis, since the RNase T1 digestion to liberate oligonucleotides would have given a 16-mer, which is large for the separation steps used (5). All DNA probes showed a very strong band at about residue 75 (Fig. 2). This may be a major cleavage site. The most stable secondary structures for the

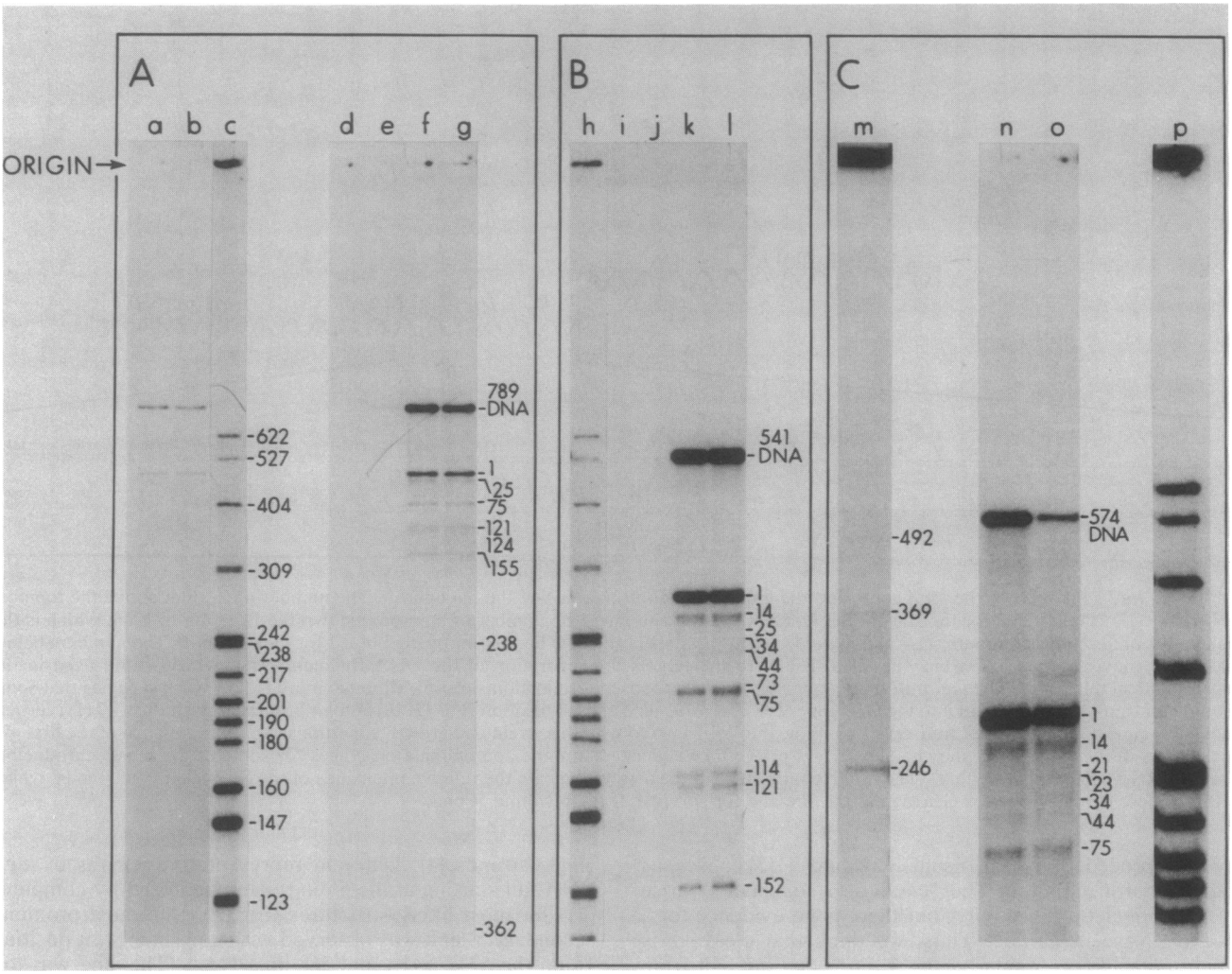

FIG. 2. Specific 5′-ended molecules of *lacZ* mRNA. *E. coli* AT2508 (HfrC *proC metB*) was grown at 37°C in inorganic salts-casein hydrolysate-glycerol (4) with a doubling time of 50 min and harvested onto ice after 9 min of *lac* operon induction (close to steady state) (4). RNA was purified and hybridized to a specific [5′-^{32}P]DNA probe (4) (see Fig. 1). (A) The 789-bp *Hin*cII-*Hin*cII fragment. (B) The 541-bp *Ava*II-*Oxa*N1 fragment. (C) The 574-bp *Hin*cII-*Mbo*II fragment. The hybrids were treated with S1 nuclease to remove single-stranded ends, denatured and then subjected to electrophoresis through a 8% polyacrylamide gel containing 8 M urea. Lanes a and b, Duplicates showing the 789-nucleotide full-length DNA band and the 479-nucleotide DNA protected by the full-length mRNA (which has the original 5′ end and has ≥479 nucleotides of *lacZ* mRNA) on a film exposed for only 22 h. Lanes f and g, Same gel lanes on a film exposed for 7 days, which allows visualization of the molecules with 5′ ends downstream. Lanes d and e, Control hybridizations; S1 reactions with the DNA alone and DNA plus yeast RNA, respectively. Lanes k and l, Duplicates for the hybrids to 541-nucleotide DNA. Lanes i and j, Controls; DNA alone and DNA plus yeast RNA, respectively, with a 7-day film exposures. Lanes c, h, and p, ^{32}P-labeled fragments of *Mst*I-digested pBR322 plasmid used as molecular weight standards (New England BioLabs). Lane m, ^{32}P fragments of the ladder from multiples of 123 bp (Bethesda Research Laboratories). The numbers adjacent to the bands in the lanes for hybrids give the distance of the 5′ end from the start of *lacZ* mRNA (numbering system in Fig. 1). Note that the actual length of [5′-^{32}P]DNA is 479 nucleotides minus the value shown; e.g., the 25-nucleotide band was calculated from a band corresponding to 454 nucleotides. Not shown is a band at 422 from the start of *lacZ* nucleotides in the 789-nucleotide DNA lanes.

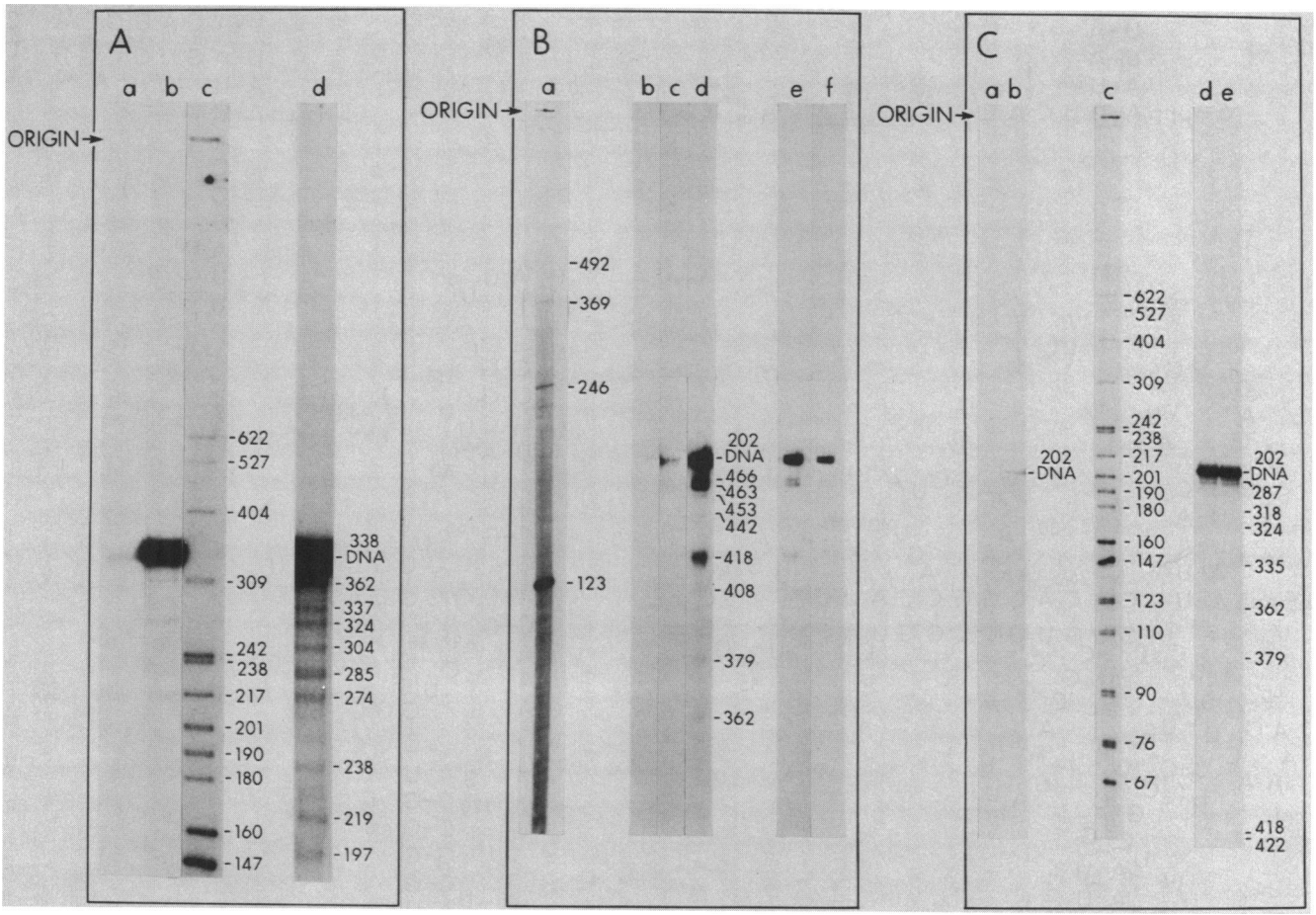

FIG. 3. Specific 3′-ended molecules of *lacZ* mRNA. (A) The 338-bp *Hin*fI-*Hin*fI DNA shown in Fig. 1 was end filled with [α′-^{32}P]dATP in the Klenow fragment of the DNA polymerase reaction and used to map 3′ mRNA ends. Lanes: a, DNA reacted without RNA; b, DNA reacted to RNA with a short film exposure (22 h); c, *Mst*I DNA markers; d, longer exposure (68 h) of the same gel lane as in lane b, with the positions of 3′ ends (numbering as in Fig. 1) for specific bands shown. Each position is calculated from the size of the DNA giving a band plus 53 nucleotides, since the ^{32}P end of the DNA is 53 nucleotides downstream from the start of transcription. (B) The 202-bp *Oxa*NI-*Hin*cII fragment from the distal end of the 789-bp DNA (Fig. 1) was 3′-end labeled with [α-^{32}P]dTTP. Lanes: a, multiples of 123-bp molecular weight standards; b, [3′-^{32}P]DNA reacted without RNA; c, yeast RNA control; d, bands resulting from RNA-DNA hybrids (film exposure, about 1 week). Film from the same gel as in lane d was exposed for 16 h in lane e and for 6 h in lane f. (C) The 5′-labeled bands from the same 202-bp DNA. Lanes: a, [5′-^{32}P]DNA reacted without RNA; b, [5′-^{32}P]DNA reacted with yeast RNA; c, *Mst*I DNA markers; d and e, bands resulting from RNA hybrids. The numbers adjacent to the hybrid lanes are based on the numbering system shown in Fig. 1.

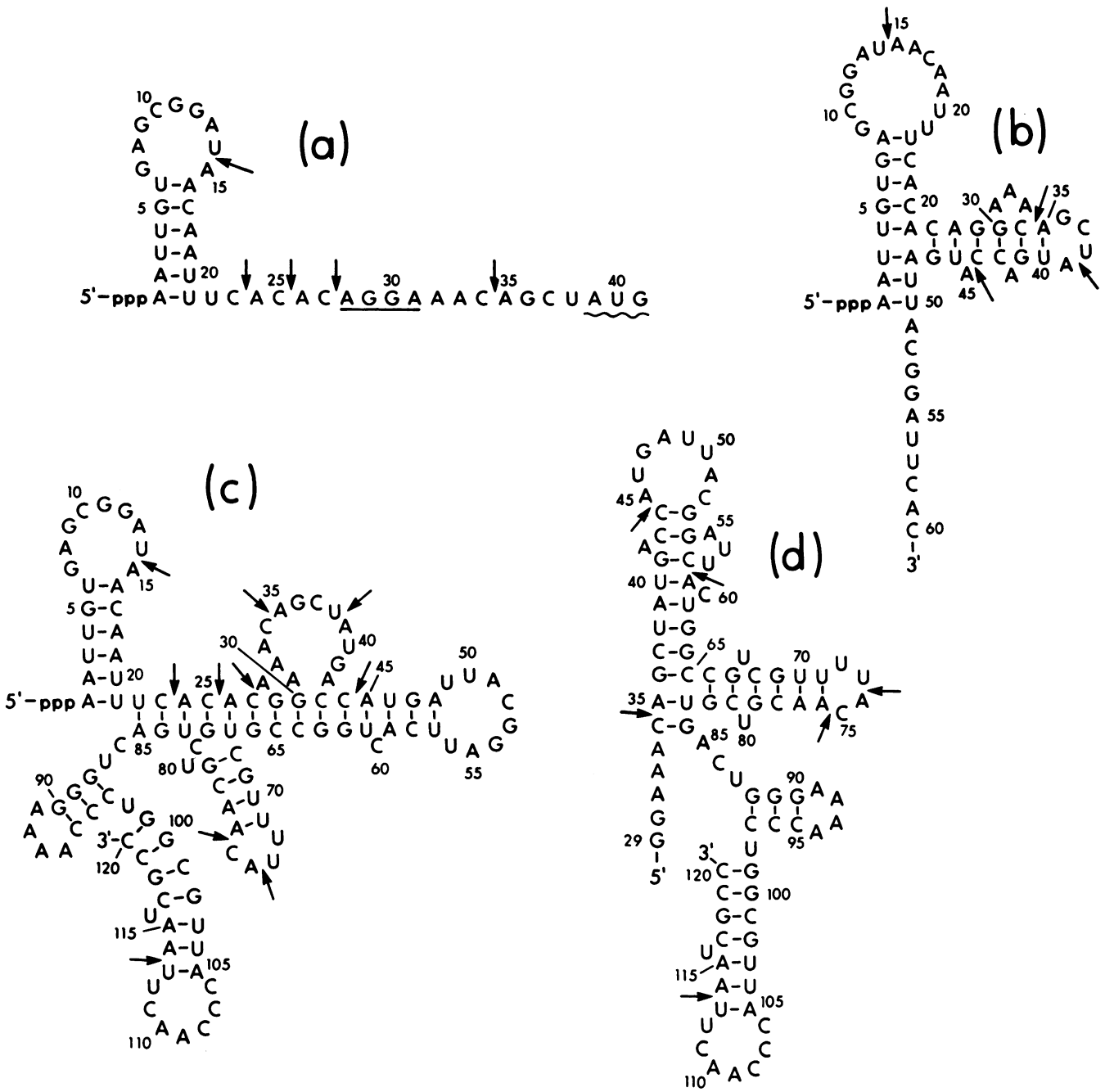

FIG. 4. The most stable structures, derived by computer analyses (21), of selected lengths of *lacZ* mRNA (Fig. 1). (a) The first 41 nucleotides of *lacZ* mRNA to include the translation initiation AUG (total free energy at 25°C, −1.6 kcal mol^{-1}). (b) The first 60 nucleotides (−7.5 kcal mol^{-1}). (c) The first 120 nucleotides (−33.0 kcal mol^{-1}). (d) The first 120 nucleotides minus the first 28 nucleotides (−26.4 kcal mol^{-1}). The structure in panel a shows the ribosome recognition sequence (——) and translation initiation AUG (〰). The arrows show possible cleavage sites based on the results in this paper, using [5′-^{32}P]DNA probes. The structure for the first 180 nucleotides (not shown) showed the site at positions 14 and 15 in a single strand, and the structure from position 21 to 120 was very similar to that segment from the first 120 nucleotides; e.g., the major cleavages at positions 73 and 74 and positions 75 and 76 were in the same single-stranded loop region.

first 120 (Fig. 2C) or 180 (data not shown), nucleotides or for the first 120 (or 180) nucleotides minus the ca. 30 leader nucleotides, all gave the UA at positions 73 and 74 and CA at positions 75 and 76 in a single-stranded loop region (Fig. 4). As with the region containing nucleotides 14 and 15, positions 73 and 74 the persistence of single strandedness with a variety of RNA 3′ endpoints could account for its above average vulnerability.

These results confirm the earlier observations (4) of molecules in an mRNA population with 5′ or 3′ ends that are internal to the full-length species and are sufficiently abundant to stand out as discrete sizes in the total population.

These molecules very likely are generated by cleavages, but a rigorous proof awaits other experiments. *lacZ* mRNA is degraded in a net 5′-to-3′ direction (2), and following the initial inactivating event at the 5′ end, the ensuing wave of degradation keeps up with the last translating ribosome (12, 18). It follows that these cleavages must release oligonucleotide fragments by chopping off the 5′ end of the mRNA that has become exposed with the transit of the last ribosome.

Undergraduate student Camilia Rivera participated in these experiments.

This work was supported by Public Health Service research grant GM34127 from the National Institutes of Health.

LITERATURE CITED

1. **Büchel, D. E., B. Gronenborn, and B. Müller-Hill.** 1980. Sequence of the lactose permease gene. Nature (London) **283**:541–545.
2. **Cannistraro, V. J., and D. Kennell.** 1985. Evidence that the 5′ end of *lac* mRNA starts to decay as soon as it is synthesized. J. Bacteriol. **161**:820–822.
3. **Cannistraro, V. J., and D. Kennell.** 1985. The 5′ ends of *Escherichia coli lac* mRNA. J. Mol. Biol. **182**:241–248.
4. **Cannistraro, V. J., M. N. Subbarao, and D. Kennell.** 1986. Specific endonucleolytic cleavage sites for decay of *Escherichia coli* mRNA. J. Mol. Biol. **192**:257–274.
5. **Cannistraro, V. J., B. M. Wice, and D. E. Kennell.** 1985. Isolating and sequencing the predominant 5′-ends of specific mRNA in cells. II. End-labeling and sequencing. J. Biochem. Biophys. Methods **11**:163–175.
6. **Cone, K. C., M. A. Sellitti, and D. A. Steege.** 1983. *lac* repressor mRNA transcription terminates *in vivo* in the *lac* control region. J. Biol. Chem. **258**:11296–11304.
7. **Fowler, A. F., and I. Zabin.** 1978. Amino acid sequence of β-galactosidase. XI. Peptide ordering procedures and the complete sequence. J. Biol. Chem. **253**:5521–5525.
8. **Gottesman, M., A. B. Oppenheim, and D. Court.** 1982. Retroregulation: control of gene expression from sites distal to the gene. Cell **29**:727–728.
9. **Hayashi, M. N., M. Hayashi, and U. R. Müller.** 1983. Role for the J-F intercistronic region of bacteriophages QX174 and G4 in stability of mRNA. J. Virol. **48**:186–196.
10. **Kalnins, A., K. Otto, U. Rüther, and B. Müller-Hill.** 1983. Sequence of the *lacZ* gene of *Escherichia coli*. EMBO J. **2**: 593–597.
11. **Kennell, D., and H. Riezman.** 1977. Transcription and translation initiation frequencies of the *E. coli lac* operon. J. Mol. Biol. **114**:1–21.
12. **Kennell, D. E.** 1986. The instability of messenger RNA in bacteria, p. 101–142. *In* W. S. Reznikoff and L. Gold (ed.), Maximizing gene expression. Butterworths, Stoneham, Mass.
13. **Lim, L. W., and D. Kennell.** 1979. Models for decay of *Escherichia coli lac* messenger RNA and evidence for inactivating cleavages between its messages. (Appendix: Mathematical models for decay of linear macro-molecules.) J. Mol. Biol. **135**:369–390.
14. **Maizels, N. M.** 1973. The nucleotide sequence of the lactose messenger ribonucleic acid transcribed from the UV5 promoter mutant of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **70**: 3585–3589.
15. **Newbury, S. F., N. H. Smith, E. C. Robinson, I. D. Hiles, and C. F. Higgins.** 1987. Stabilization of translationally active mRNA by prokaryotic REP sequences. Cell **48**:197–210.
16. **Richardson, J. P., and E. C. Ruteshouser.** 1986. rho factor-dependent transcription termination: interference by a mutant rho. J. Mol. Biol. **189**:413–419.
17. **Schindler, D., and H. Ecols.** 1981. Retroregulation of the *int* gene of bacteriophage λ: control of translation completion. Proc. Natl. Acad. Sci. USA **78**:4475–4480.
18. **Schneider, E., M. Blundell, and D. Kennell.** 1978. Translation initiation and mRNA decay. Mol. Gen. Genet. **160**:121–129.
19. **Sellitti, M. A., P. A. Pavco, and D. A. Steege.** 1987. *lac* repressor blocks *in vivo* transcription of *lac* control region DNA. Proc. Natl. Acad. Sci. USA **84**:3199–3203.
20. **Steege, D. A.** 1977. 5′-Terminal nucleotide sequence of *Escherichia coli* lactose repressor mRNA: features of translational initiation and reinitiation sites. Proc. Natl. Acad. Sci. USA **74**:4163–4167.
21. **Zuker, M., and P. Stiegler.** 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxilliary information. Nucleic Acids Res. **9**:133–148.