Proc. Natl. Acad. Sci. USA
Vol. 79, pp. 1195–1199, February 1982
Genetics

Promoter of a eukaryotic $tRNA^{Pro}$ gene is composed of three noncontiguous regions

(transcription/recombinant DNA/nuclear microinjection)

GENNARO CILIBERTO*, LUISA CASTAGNOLI, DOUGLAS A. MELTON†, AND RICCARDO CORTESE

European Molecular Biology Laboratory, Postfach 102209, 6900 Heidelberg, Federal Republic of Germany

Communicated by John C. Kendrew, October 5, 1981

ABSTRACT **The 71-base-pair coding sequence of the $tRNA^{Pro}$ gene from *Caenorhabditis elegans* contains all of the information required for transcription and processing in the injected oocytes. Several subclones of the DNA coding for the $tRNA^{Pro}$ were constructed, carrying deletions or insertions, or both. Their transcriptional properties lead to the hypothesis that the $tRNA^{Pro}$ gene promoter is composed of three discontinuous regions within the coding sequence.**

In previous studies we have characterized steps in the synthesis of mature tRNA molecules in eukaryotic cells by injecting cloned DNAs coding for tRNA into *Xenopus laevis* oocytes (1–4). These studies have shown that tRNA genes are transcribed into precursor molecules, which are converted into tRNAs through a series of nucleolytic cleavages and nucleotide modifications.

In order to gain further insight into the process of tRNA biosynthesis and to determine how cells regulate the expression of these genes, it will be instructive to identify the transcriptional signals present on the DNA sequence and recognized by the RNA polymerase III. The work of Brown and his colleagues (5, 6) has revealed that the relatively small 5S RNA gene contains a 30-base-pair region within the coding sequence that is essential for initiation of transcription by RNA polymerase III. Also for tRNA genes (7–10), it appears probable that their promoters are not located in the 5′ flanking region, even though this region may play a regulatory role in some cases (9, 10). It is unclear, however, whether they have an internal control region analogous to that of the 5S RNA genes. To establish this, we constructed deletion and insertion mutants, altering sequences within the coding regions of the $tRNA^{Pro}$ gene of *Caenorhabditis elegans*. This allowed the identification of three separated regions within the coding sequence that are important for transcription of the $tRNA^{Pro}$ gene.

MATERIALS AND METHODS

Chemicals and Enzymes. Isopropyl-D-thiogalactopyranoside and 5-bromo-4-chloro-3-indolyl-D-galactoside were purchased from Sigma, and recombinant DNA linkers were purchased from Collaborative Research (Waltham, MA). T4 DNA ligase, polynucleotide kinase, and endonucleases *Eco*RI and *Hae* III were gifts of V. Pirrotta. Endonucleases *Hin*dIII and *Sma* I and DNA polymerase large fragment Klenow enzyme were purchased from Boehringer. Endonucleases *Hin*fI and *Bam*HI and nucleases S1 and *BAL* 31 were purchased from Bethesda Research Laboratories. Exonuclease *Xma* was a gift of M. Zabeau. All radioactive compounds were purchased from Amersham Buchler, Braunschweig.

Bacterial Strains, Plasmids, and Cloning Vehicles. *Escherichia coli* K-12 (strain 71-18) was used for transformation with phage *M13* (strain mp2) replicative form I and all of its recombinant derivatives (4, 11, 12). *E. coli* K-12 (strain K514) was used for transformation with pBR322 plasmid and its recombinant derivatives. Protocols for transformation and preparation of double-stranded DNA were as described (1, 4).

Polyacrylamide Gel Electrophoresis. DNA restriction fragments and RNA transcripts were separated electrophoretically on Tris glycine/polyacrylamide gels and Tris borate/EDTA gels, respectively, as described (1). DNA fragments were purified as described (1, 4).

Microinjections. Microinjections into the nucleus of *X. laevis* oocytes were performed as described (1) by using a 50-nl solution of DNA (100 μg/ml) and radioactive RNA precursors ([α-^{32}P]GTP; 350 Ci/mmol, 10 mCi/ml; 1 Ci = 3.7×10^{10} becquerels). Typically, a set of 20 oocytes was injected for each sample and incubated for 5 hr before RNA extraction. For a quantitative analysis of the extent of transcription, bands were cut from the gel and their Cerenkov radioactivity was measured. Calculations were made by normalizing to total trichloroacetic acid-precipitable material and to high M_r RNA (visible at the top of 10% polyacrylamide gels). These two normalizations always give concordant results.

DNA Sequence Determinations. The sequences of the mp2 derivatives were determined by using the dideoxy chain termination method of Sanger *et al.* (13), with the single-stranded DNA as template (14). pBR322 derivatives were sequenced by the method of Maxam and Gilbert (15).

Autoradiography. X-ray Kodak films (X5) were used occasionally with preflashing as described by Laskey and Mills (16).

RESULTS

Construction of Subclones Carrying Deletions of the Flanking Regions and Their Transcriptional Analysis. We have established that a plasmid (Mcet1) containing a DNA segment 263 base pairs in length that codes for the $tRNA^{Pro}$ from *C. elegans* was sufficient to direct the synthesis of mature $tRNA^{Pro}$ when microinjected into the nucleus of *X. laevis* oocytes (1, 4). In Fig. 1*A* is a schematic representation of the 263-base-pair nematode DNA segment coding for $tRNA^{Pro}$. It can be divided into three domains: the 5′ flanking region (domain 1), the structural domain or coding sequence (domain 2), and the 3′ flanking region (domain 3). Two conveniently located *Hae* III sites (unpublished sequence) allow the physical separation of these three domains

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.

* Permanent address: I Istituto di Chimica Biologica, II Medical School, University of Naples, Naples, Italy.
† On leave of absence from MRC Laboratory of Molecular Biology, Hills Road, Cambridge, CB2 2QH, England. Present address: Dept. of Biochemistry and Molecular Biology, Harvard University, Cambridge, MA 02138.

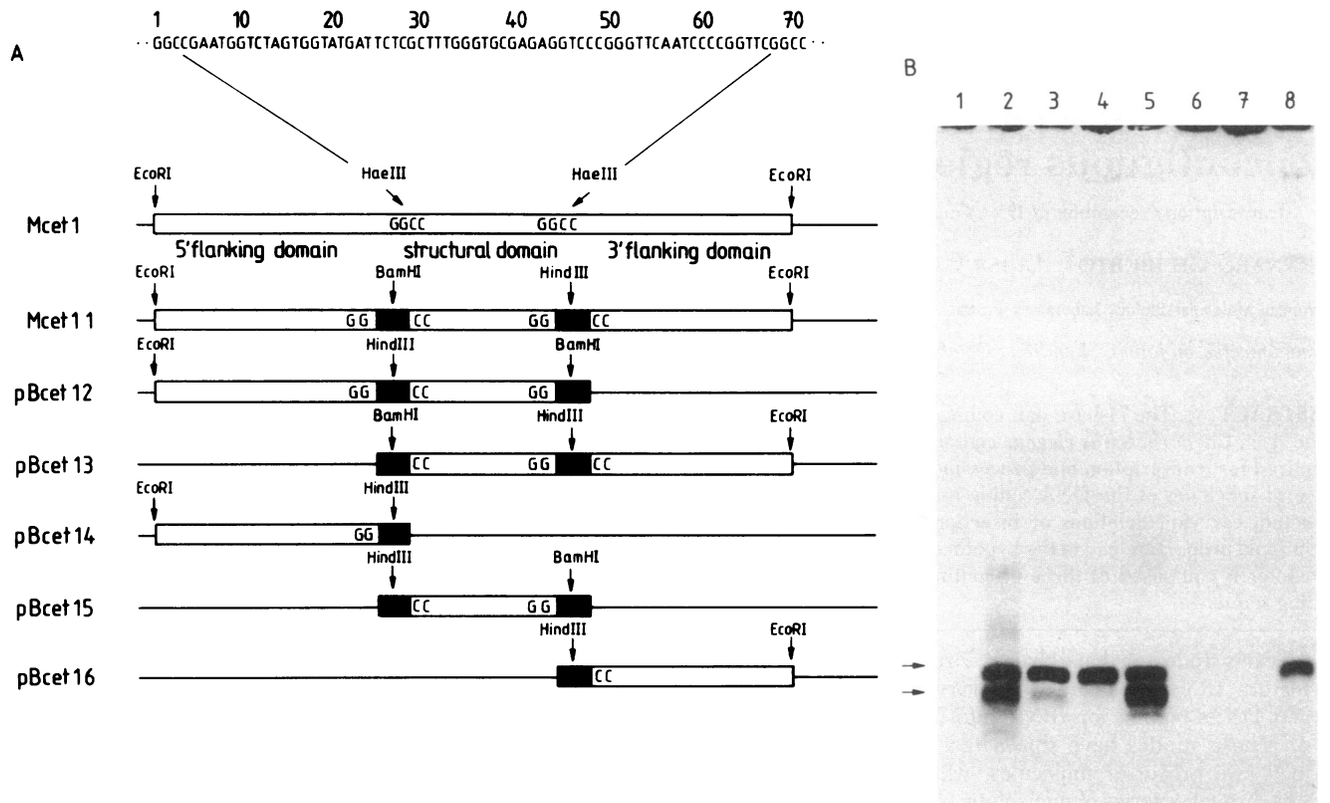

FIG. 1. (*A*) Diagram of the relevant features of Mcet1 DNA and its subclones. Only the 263-base-pair *Eco*RI–*Eco*RI insert is shown. The DNA sequence is that of the coding region (structural core). Plasmids beginning with M are mp2 (from phage M13) recombinants, those beginning with pB are pBR322 recombinants. ——, Vector DNA (either mp2 or pBR322); ▭, segments derived from nematode DNA; ▬, *Bam*HI (C-C-G-G-A-T-C-C-G-G) and *Hin*dIII (C-C-A-A-G-C-T-T-G-G) linker molecules. (*B*) Polyacrylamide gel electrophoresis fractionation of RNA transcribed in the nucleus of DNA-injected *X. laevis* oocytes. Slots: 1, no DNA; 2, Mcet1; 3, Mcet11; 4, pBcet12; 5, pBcet13; 6, pBcet14; 7, pBcet16; 8, pBcet15. X-ray Kodak films (X5) were used.

that were subcloned in various combinations into mp2 or pBR322 vectors, with linkers whenever necessary (Fig. 1*A*).

Double-stranded DNAs of the various subclones were injected into the nucleus of *X. laevis* oocytes, and the resulting transcripts were analyzed on 10% (wt/vol) polyacrylamide gel electrophoresis (Fig. 1*B*). Plasmid Mcet1, containing the original 263-base-pair fragment, yields two RNA species (indicated by arrows in Fig. 1*B*, slot 2). Two-dimensional peptide map analysis (data not shown) demonstrates that they have the same RNA sequence, corresponding to $tRNA^{Pro}$. A preliminary modified-base analysis shows the presence in the upper band of a modified A residue, which is absent in the lower band. The relative intensity of the two bands varied from experiment to experiment, and we attribute it to a variable extent of post-transcriptional modification. Identical $tRNA^{Pro}$ (as shown by two-dimensional peptide mapping, data not shown) is obtained with (*i*) pBcet13, which carries a deletion of the 5′ flanking domain (slot 5); (*ii*) pBcet12, with a deletion of the 3′ flanking domain (slot 4); (*iii*) pBcet15, with a deletion of both the 5′ and 3′ flanking domains (slot 8); and (*iv*) Mcet11, which contains all three domains separated by *Bam*HI linkers between the 5′ flanking domain and the structural domain and by *Hin*dIII linkers between the structural domain and the 3′ flanking domain. Note that the insertion of both *Bam*HI and *Hin*dIII linkers restores the G-G at the 5′ end and the C-C at the 3′ end, which are deleted by the *Hae* III cleavage. Therefore, the insertion of these linkers does not change the sequence of the structural domain. No transcript is observed when pBcet14 or pBcet16 are injected (slot 6 and 7). These plasmids contain the 5′ flanking domain and the 3′ flanking domain, respectively.

These results are all consistent with the interpretation that the $tRNA^{Pro}$ gene contains all of the information for transcription and processing within the domain 2, the structural domain of the gene.

***In Vitro* Construction of Mutants with Deletions in the Structural Domain.** Having established that the structural domain of $tRNA^{Pro}$ contains all of the information for transcription and processing, we constructed a set of mutant genes carrying deletions of parts of the coding region. The *Hin*fI site at position 22 and the *Sma* I (*Xma* I) site at position 47, indicated by arrows in Fig. 2, enclose three DNA segments A, B, and C within the structural domain. Exploiting these sites we constructed the following deletion mutants: pBcet17, carrying a deletion of segment C; Mcet18, carrying a deletion of segment A; pBcet19, carrying a deletion of segment B, with a *Hin*dIII linker (10 base pairs) joining segment A to C; pBcet191 (derivative of pBcet19), carrying a substitution for segment B—a DNA fragment of identical length; pBcet22, carrying a four-base-pair deletion (C-48 through G-51) at the B/C junction; pBcet23, carrying a seven-base-pair deletion (T-19 through T-25) at the 3′ end of segment A.

Transcriptional analysis of these mutants is shown in Fig. 3. pBcet17 and Mcet18 (Fig. 3*A*, slots 3 and 4) do not give any detectable product. pBcet23 (slot 5) directs transcription of RNA species with about 50% of the efficiency of the intact gene (slot 1). Two-dimensional peptide mapping (data not shown) indicates that this RNA contains the sequence of $tRNA^{Pro}$ (but with the absence of oligonucleotides U-A-U-G and A-U-U-C-U-C-G, corresponding to the deleted sequences) and, in addition,

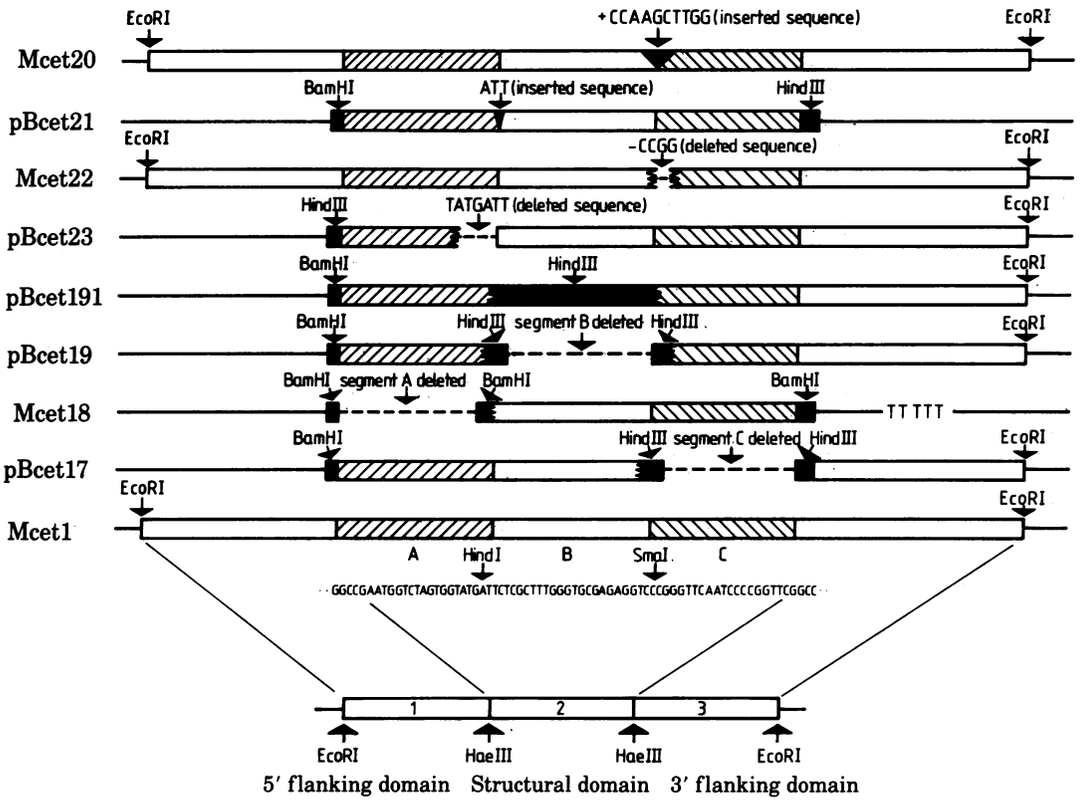

FIG. 2. Structure of insertion and deletion mutants of Mcet1. These plasmids were constructed as follows. pBcet17: fragments *Bam*HI–*Sma* I from pBcet11 and *Hin*dIII–*Eco*RI from pBcet16 were joined by means of *Hin*dIII linkers and then cloned into pBR322 DNA. Mcet18: fragment *Hin*fI–*Bam*HI from pBcet15, after addition of *Bam*HI linkers to the filled-in *Hin*fI end, was cloned in mp2 vector. pBcet19: fragments *Bam*HI–*Hin*fI from pBcet13 and *Sma* I–*Eco*RI from Mcet1 were joined (after fill-in of the *Hin*fI end) by *Hin*dIII linker and cloned in pBR322. pBcet191: DNA from pBcet19 was digested with *Hin*dIII, filled in, and blunt-end ligated to dephosphorylated *Hin*dIII linkers. pBcet23: pBcet15 was digested with *Hin*fI, followed by S1 nuclease treatment, and digestion with *Hin*dIII and *Bam*HI. The desired fragments were purified and then joined by blunt-end ligation. The resulting fragment was filled in at the *Bam*HI site, blunt-end ligated to domain 3, and finally cloned in pBR322. Mcet22: Mcet1 DNA was digested with *Xma* I, followed by S1 treatment and blunt-end ligation. pBcet21: DNA from Mcet11 was digested with *Hin*fI, filled in, and digested with *Hin*dIII and *Bam*HI. The desired fragments were joined by blunt-end ligation and cloned into pBR322. Mcet20: DNA from Mcet1 was digested with *Sma* I and blunt-end ligated in the presence of *Hin*dIII linkers. ——, DNA of plasmid vector; ----, deleted regions; □, nematode sequences; ■, linker molecules of the sequence specified in the legend to Fig. 1. (Note that in the case of pBcet19, Mcet18, and pBcet17, the complete linker sequence is represented by the two halves of the black box separated by a dotted line, used here merely to allow the alignment of homologous regions in all clones.); ▲, insertions in an otherwise intact gene.

extranucleotides derived both at the 5′ and at the 3′ end. Apparently this deletion interferes with the size-reducing maturation reactions of a precursor molecule. Mcet22 (slot 6) is severely impaired in transcription, with an efficiency of about 1%. pBcet19 and pBcet191 (Fig. 3*B*, slots 2 and 3, respectively) direct transcription of RNA species with an efficiency of 10% and 20%, respectively. These results indicate that the DNA sequences essential for promotion of transcription are spread all along the structural domain, and they are not contiguous.

***In Vitro* Construction of Insertion Mutants Within the Structural Domain.** An effective way of mutating DNA sequences is to interrupt them by the insertion of extra nucleotides. We constructed two insertion mutants: Mcet20, carrying a 10-base-pair insertion at the junction between segments B and C, and pBcet21, carrying a 3-base-pair insertion at the junction between segments A and B (see Fig. 2). Transcriptional analysis of these mutants is shown in Fig. 3*C*. The results indicate that both types of insertion do not significantly affect the efficiency of transcription. Two-dimensional peptide mapping of the transcriptional products of Mcet20 and pBcet21 (data not shown) shows that the fastest migrating RNAs in both cases have mature 5′ and 3′ ends and that their sequence differs from $tRNA^{Pro}$ only for the presence of the inserted bases. The other slower migrating RNAs contain the $tRNA^{Pro}$ sequence but also additional nucleotides. This suggests that they are accumulating tRNA precursors and that there is a reduced efficiency of the processing reactions as a consequence of the altered tRNA sequence.

Deletion Mutants at the 5′ and at the 3′ Ends of the Structural Domain. To better define the external boundaries of the regions essential for promotion of transcription in $tRNA^{Pro}$, we started the construction of a series of deletions at the 5′ and at the 3′ end of the structural domain. The construction and the structure of these mutants is described in Fig. 4 *A* and *B*. Double-stranded DNAs from these clones were microinjected in the nucleus of *X. laevis* oocytes to define their transcriptional properties (Fig. 4 *C* and *D*).

Deletions up to seven or eight base pairs from the 5′ end (plasmids B5′Δ8 and B5′Δ9) do not significantly reduce the extent of transcription, whereas deletions of 15 or 16 nucleotides (B5′Δ16 and B5′Δ17) decrease the transcription efficiency to 5%. The transcripts obtained from B5′Δ8 and from B5′Δ9 appear heterogeneous; we do not yet know whether this is due to heterogeneity of the starting point or to processing reactions.

At the 3′ end, deletions of 9 or 11 base pairs do not reduce significantly transcription (plasmids B3′Δ62 and B3′Δ60). Transcripts from both mutants are about 110 bases long (according to size markers not shown). This is in agreement with the hypothesis that both transcripts are terminated at the first available run of Ts (four or more); this is at position T-162 through

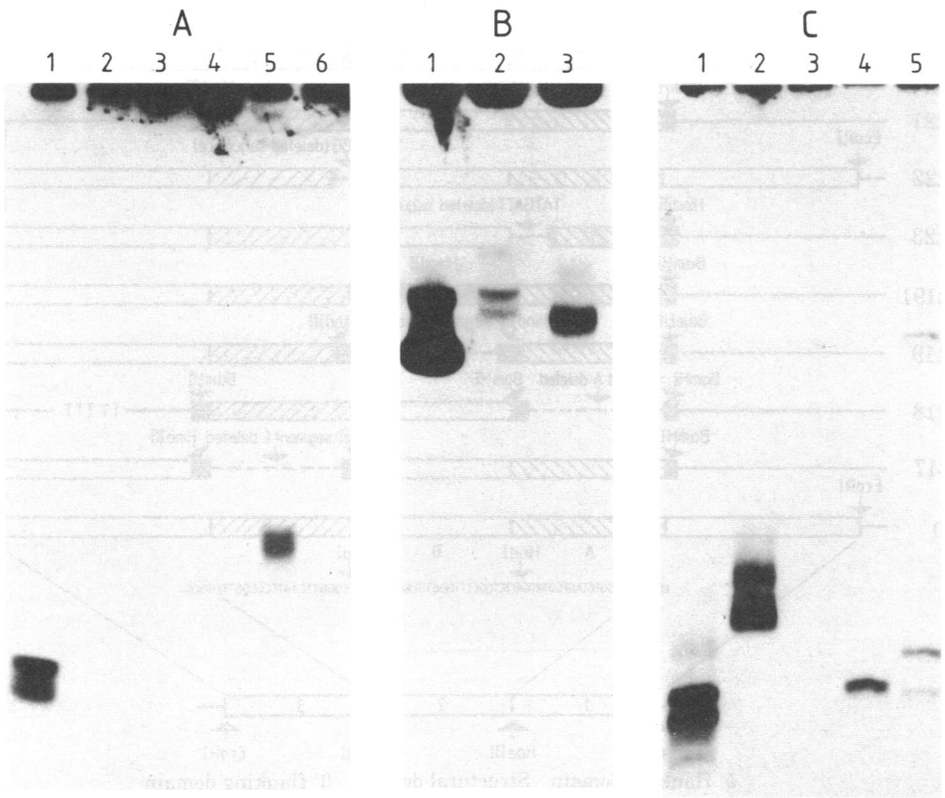

FIG. 3. Polyacrylamide gel electrophoresis fractionation of RNA transcribed in the nucleus of DNA-injected *X. laevis* oocytes. (*A*) Slots: 1, Mcet1; 2, no DNA; 3, pBcet17; 4, Mcet18; 5, pBcet23; 6, Mcet22. (*B*) Slots: 1, pBcet13; 2, pBcet19; 3, pB191. (*C*) Slots: 1, Mcet1; 2, Mcet20; 3, no DNA; 4, pBcet12; 5, pBcet21.

T-165 of the nematode sequence for B3′Δ62 (unpublished data) and at position T-4328 through T-4324 in the flanking pBR322 sequence for B3′Δ60.

We conclude that the regions essential for transcription do not start before nucleotide +9 on the 5′ side and do not extend beyond base +60 on the 3′ side of the structural gene.

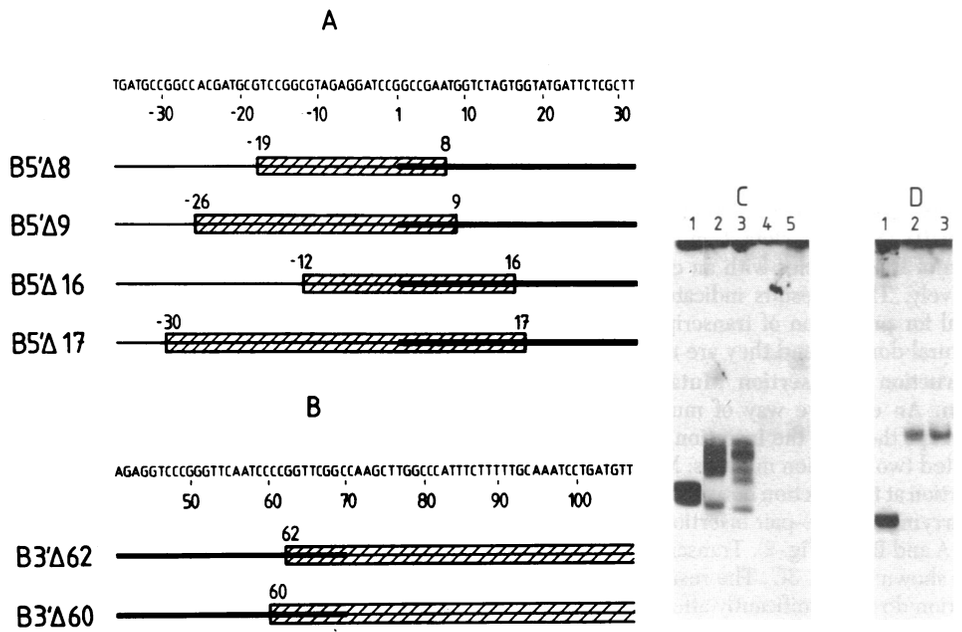

FIG. 4. (*A* and *B*) Construction of deletions at the 5′ and 3′ ends of the structural domain of the $tRNA^{Pro}$ gene. pBcet13 was digested either with *Bam*HI (deletions at the 5′ end) or with *Hin*dIII (deletions at the 3′ end) and incubated with nuclease *Bal* 31 for a time sufficient to delete about 20–40 nucleotides. (*A*) Structure of mutants at the 5′ end. (*B*) Structure of mutants at the 3′ end. Heavy lines and bold type and numbers are used for the structural domain. Thin lines are used for flanking sequences. ▨, Deleted nucleotides. (*C* and *D*) Polyacrylamide gel electrophoresis fractionation of RNA transcribed in the nucleus of DNA-injected *X. laevis* oocytes. (*C*) Slots: 1, pBcet13; 2, B5′Δ8; 3, B5′Δ9; 4, B5′Δ16; 5, B5′Δ17. (*D*) Slots: 1, pBcet15; 2, B3′Δ62; 3, B3′Δ60. pBcet13 and pBcet15 were used as wild-type controls for 5′- and 3′-end deletion mutants, respectively.

FIG. 5. The $tRNA^{Pro}$ gene. Sequences essential for transcription are boxed.

DISCUSSION

The most important result of our work is the discovery that the $tRNA^{Pro}$ gene "promoter," defined here as the sum of all the sequences important for efficient promotion of transcription, is constituted by three separated sequences in the structural domain.

Region a is contained in segment A. The importance of this region is shown by the complete lack of transcription in the mutant carrying a total deletion of sequence A (Mcet18). However, it does not include the sequence 19–24, T-A-T-G-A-T, and the sequence 1–8, G-G-C-C-G-A-A-T, which can be deleted in mutants pBcet23 and B5′Δ9, respectively, without much impairment of transcription. In conclusion, region a must be contained within the sequence 9–18, G-G-T-C-T-A-G-T-G-G.

Region c is contained in segment C. The importance of this region, is shown by complete lack of transcription in the mutant carrying a deletion of segment C (pBcet19). The essential region is bounded at its 5′ end by the nucleotide C-49. This is shown by the fact that the insertion of 10 base pairs between nucleotides C-49 and G-50 (insertion mutant Mcet20) does not significantly affect transcription. Region c must include the two nucleotides G-50 and G-51. This is shown by the fact that deletion of C-48 through G-51 (Mcet22) results in a decrease of transcription to 1/100th, whereas deletion of segment B including C-48 and C-49 (pBcet19) results in a decrease to only 1/10th. On its 3′ side, region c does not include the sequence C-61 through C-71, which can be deleted without impairment of transcription in mutant B3′Δ60. Region c must be contained within the sequence 50–60, G-G-G-T-T-C-A-A-T-T-C. Our nucleotide C_{55} corresponds to the invariant C_{56} that Koski *et al.* (17) have proved essential for transcription.

The presence of two separated essential regions in approximately the same location has been clearly established also in the case of a $tRNA^{Met}$ gene from *X. laevis* by Hofstetter *et al.* (18). Our investigations lead to the conclusion that another region (b), within segment B, contains information important, though not essential, for transcription. This is revealed by a decrease of transcription to 1/10th in a mutant carrying a deletion of segment B (pBcet19). When segment B is replaced by a segment of identical length but composed of *Hin*dIII linker sequences (pBcet191), there is a two-fold increase in transcription with respect to the shorter segment (pBcet19), but the overall activity is still decreased to 1/5th that obtained with the intact gene. Region b is delimited at its 5′ end by the sequence G-A-T-T, which separates it from region a and can be destroyed without significantly altering transcription (pBcet21). On the other side, at the 3′ end, it does not extend beyond nucleotide C-49. Hofstetter *et al.* (18) concluded that the sequence between the two essential regions, equivalent to our regions a and c, has only a "spacer" function. However, it must be stressed that because of the difference of the sequences of DNA coding for the $tRNA^{Pro}$ and $tRNA^{Met}$ genes, the restriction enzyme dissection and the construction of subclones were necessarily different in the two cases. Therefore, a precise comparison between our result and their result is impossible. It may be argued that the decrease to 1/5th in the amount of transcript in the case of pBcet191 may be due to instability of the transcript. We think that this is unlikely because any RNA molecule so far injected in *X. laevis* oocytes has been proven to be stable (2, 19). Further support to the hypothesis of the existence of a transcriptionally important region within segment B derives from a study of single-base-pair substitution mutants of the same gene (20). The same decrease of transcription to 1/5th obtained with plasmid pBcet191 is actually observed with the single-point mutation G-44 to A-44 (20).

A diagrammatic representation of the regions essential for transcription, which we have deduced from our results, is shown in Fig. 5. Region b is still to be further defined.

It has been known for many years that all tRNA molecules, and therefore all tRNA genes, have invariant nucleosides in their sequence (21). It has been proposed that precisely these invariant regions in the tRNA genes are the sites recognized by the transcriptional machinery (17). However, our mutant pBcet23, carrying a deletion that includes the invariant A21, is not significantly impaired in transcription. This shows that not all invariant bases are essential for transcription. Moreover, the deletion of mutant Mcet22 does not involve any invariant base, and yet it has a profound effect on transcription. It is clear, however, that region c is highly conserved in all tRNA genes (21). One may speculate that it interacts with a common element of the transcriptional machinery, whereas region a, which is more variable both in size and in sequence (21), may be responsible for a more specific control of transcription, which is different in the various tRNA genes.

We are grateful to Professor Francesco Blasi and Dr. Vincenzo Pirrotta for their careful criticism of the manuscript. G.C. was supported by a Deutscher Akademischer Austauschdienst Fellowship, D.M. by a short-term European Molecular Biology Organization Fellowship, and L.C. by a Cenci-Bolognetti Fellowship.

1. Cortese, R., Melton, D. A., Tranquilla, T. & Smith, J. D. (1978) *Nucleic Acids Res.* **5**, 4593–4611.
2. Melton, D. A. & Cortese, R. (1979) *Cell* **18**, 1165–1172.
3. Melton, D. A., De Robertis, E. M. & Cortese, R. (1980) *Nature (London)* **284**, 143–148.
4. Cortese, R., Harland, R. & Melton, D. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 4147–4151.
5. Sakonju, S., Bogenhagen, D. F. & Brown, D. D. (1980) *Cell* **19**, 13–25.
6. Bogenhagen, D. F., Sakonju, S. & Brown, D. D. (1980) *Cell* **19**, 27–35.
7. Telford, J. L., Kressmann, A., Koski, R. A., Grosschedl, R., Muller, F., Clarkson, S. G. & Birnstiel, M. L. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 2590–2594.
8. Garber, R. L. & Gage, L. P. (1979) *Cell* **18**, 817–828.
9. Sprague, K. U., Larson, D. & Morton, D. (1980) *Cell* **22**, 171–178.
10. De Franco, D., Schmidt, O. & Soll, D. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 3365–3368.
11. Gronenborn, B. & Messing, J. (1978) *Nature (London)* **272**, 375–377.
12. Rothstein, R. J., Lau, L. F., Bahl, C. P., Narang, S. A. & Wu, R. (1980) *Methods Enzymol.* **68**, 98–109.
13. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
14. Schreier, P. H. & Cortese, R. (1979) *J. Mol. Biol.* **129**, 169–172.
15. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
16. Laskey, R. A. & Mills, A. D. (1975) *Eur. J. Biochem.* **56**, 335–341.
17. Koski, R. A., Clarkson, S. G., Kurjan, J., Hall, B. D. & Smith, M. (1980) *Cell* **22**, 415–425.
18. Hofstetter, H., Kressman, A. & Birnstiel, M. L. (1981) *Cell* **24**, 573–585.
19. Miller, J. R. & Melton, D. A. (1981) *Cell* **24**, 829–835.
20. Ciampi, S., Melton, D. A. & Cortese, R. (1982) *Proc. Natl. Acad. Sci. USA*, in press.
21. Gauss, D. H. & Sprinzl, M. (1981) *Nucleic Acids Res.* **9**, r1/r23.