Proc. Natl. Acad. Sci. USA
Vol. 80, pp. 3416–3420, June 1983
Genetics

A short 5′ flanking region containing conserved sequences is required for silkworm alanine tRNA gene activity

(upstream control region/RNA polymerase III/transcription)

DRENA LARSON, JANICE BRADFORD-WILCOX, LISA S. YOUNG, AND KAREN U. SPRAGUE

Institute of Molecular Biology and Department of Biology, University of Oregon, Eugene, Oregon 97403

Communicated by George Streisinger, February 28, 1983

ABSTRACT **Using partially deleted genes, we have identified an upstream control signal required for transcription of a *Bombyx mori* (silkworm) $tRNA_2^{Ala}$ gene. The 5′ boundary of this essential region lies between 34 and 11 nucleotides preceding the transcription initiation site. Sequences in the 5′ half of the $tRNA_2^{Ala}$ coding region are also important. Both regions contain oligonucleotides that are conserved among several *Bombyx* genes transcribed by RNA polymerase III, suggesting that these sequences may have a general control function.**

Identification of *cis*-acting elements that affect transcription of eukaryotic genes is essential for understanding how these elements interact with diffusible regulatory molecules to control gene expression. Recent localization of DNA sequences required for transcription by RNA polymerase III has drawn attention to control regions lying inside genes (1–11). In several cases, however, 5′ flanking DNA also influences the efficiency or the position of transcription initiation by RNA polymerase III (3, 12–16). In one instance, 5′ flanking sequences are absolutely required for activity: replacement of ≈5 kilobases (kb) of the normal 5′ flanking DNA with unrelated sequences destroys the activity of a *Bombyx mori* (silkworm) alanine $tRNA_2$ gene in homologous *in vitro* transcription systems (17).

To understand how upstream sites control gene activity, we wished to identify the essential region within the 5′ flanking DNA of this $tRNA_2^{Ala}$ gene. We therefore constructed a series of deletions removing successively larger amounts of flanking sequence. We expected analysis of the resulting mutant templates to yield two kinds of information. First, the position of the external control region relative to the transcription start site should suggest the nature of its role. For example, a control region lying close to a gene could interact directly with RNA polymerase or associated factors at the point of transcription initiation. It would thus be functionally analogous to regulatory sequences adjacent to prokaryotic genes. In contrast, the contribution of a far upstream control region would likely be somewhat different. Distant sites might function as regulators of local DNA superhelicity, as has been suggested for some genes transcribed by eukaryotic RNA polymerase II (reviewed in ref. 18).

In addition to providing positional information, localization of the required region of the $tRNA^{Ala}$ gene might also allow us to recognize the particular oligonucleotides that act as transcriptional control elements. Comparison with similar regions adjacent to other *Bombyx* RNA polymerase III templates could reveal sequence homologies. The distribution of such conserved sequences among different classes of genes would suggest how general their control function might be.

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.

To learn whether the activity of *Bombyx* $tRNA_2^{Ala}$ genes is controlled entirely by upstream elements, or whether intragenic elements are involved as well, we also examined the properties of genes from which variable amounts of coding sequence had been removed. In this paper we show that an essential upstream control region is located within the 34-nucleotide sequence immediately adjacent to the transcription initiation site. In addition, a region downstream from this site, within the 5′ half of the coding region, is important for activity. We believe that certain sequences within these regions are particularly significant because they are found in other genes transcribed by *Bombyx* RNA polymerase III.

MATERIALS AND METHODS

Cloned Genes Used for Deletion Analysis. The $tRNA_2^{Ala}$ gene used in this work was originally isolated as a 13.5-kb *Eco*RI fragment of the *Bombyx* genome (19). The 18-kb pBR322 derivative containing this large fragment is called pBm11. pWT ALA 5′, a 4.8-kb plasmid, is a subclone in which a 437-base-pair (bp) $tRNA_2^{Ala}$ gene-containing *Pvu* II/*Ava* I fragment replaces the *Pvu* II/*Ava* I region of pBR322. The transcriptional activity of this subclone is indistinguishable from that of pBm11. Therefore, it was used as the starting material for the construction of 5′ and 3′ deleted genes and provided an activity standard for comparison with mutant derivatives. pWT ALA 3′ contains essentially the same *Bombyx* DNA as pWT ALA 5′, except that the *Ava* I site has been replaced by an *Eco*RI linker. The resulting 435-bp fragment was inserted between the *Pvu* II and *Eco*RI sites of pBR322, yielding a 2.5-kb plasmid. pWT ALA 3′ also has wild-type transcriptional activity and, because of its structural similarity to the 3′ deletion mutants, it was used as an additional standard.

Construction of Deletion Mutants and DNA Sequence Analysis. The 5′ and 3′ deleted $tRNA_2^{Ala}$ genes were constructed as outlined in Fig. 1, using BAL-31 mutagenesis procedures (20). The endpoints for all 5′ and 3′ deletions were established by DNA sequence analysis (21). The sequences were searched for homologies and dyad symmetries by using the SEQ computer program (22) through the Stanford MOLGEN project of the National Institutes of Health SUMEX-AIM facility.

***In Vitro* Transcription Assays and RNA Sequence Analysis.** Template efficiencies (transcripts per gene per hr) of wild-type and deleted genes were measured by *in vitro* transcription assays using *Bombyx* silk gland and ovary extracts (17). Most assays were conducted at the previously determined optimal DNA concentration of 20 μg/ml. However, to ensure that the effects we observed were not limited to a particular set of reaction conditions, assays were repeated with different extract preparations and at different DNA concentrations (10–30 μg/ml). The

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

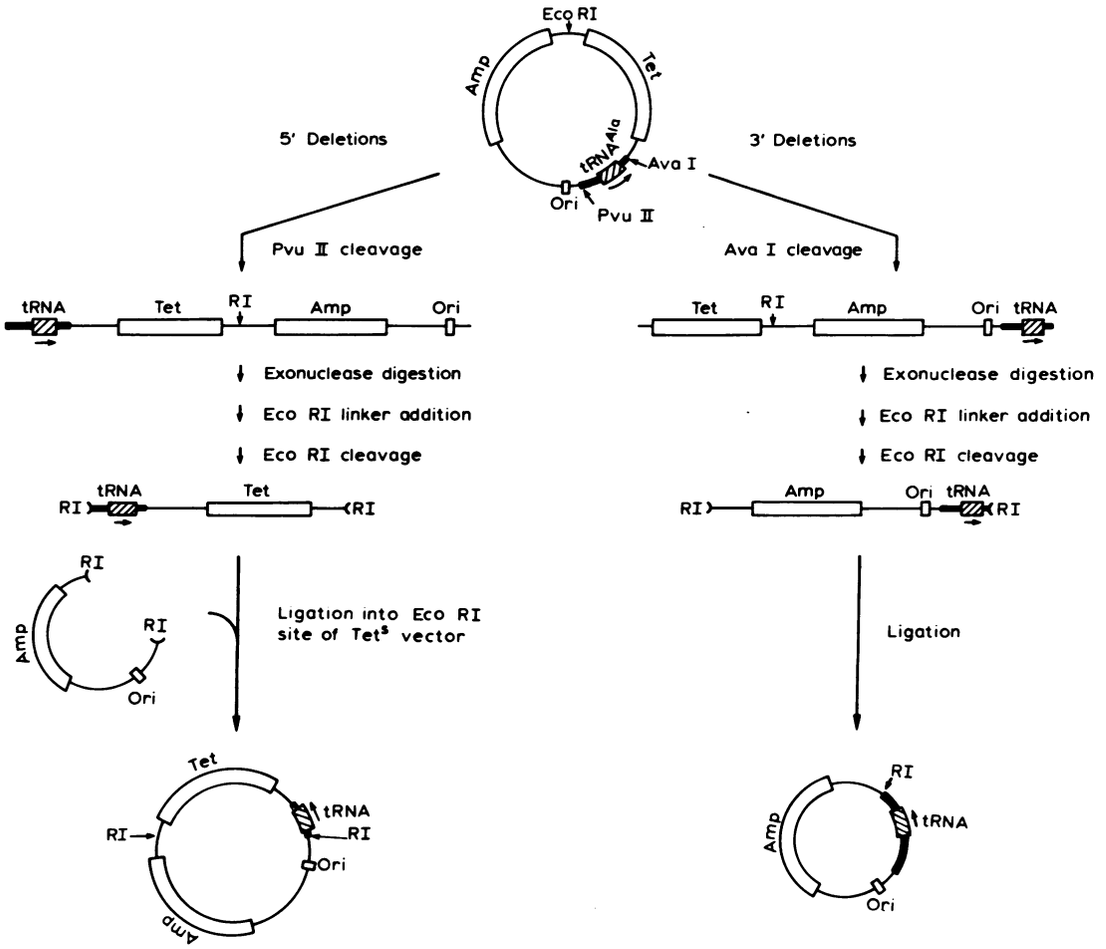

FIG. 1. Construction of 5′ and 3′ deleted $tRNA_2^{Ala}$ genes. The starting material contains a 437-bp fragment of *Bombyx* DNA (heavy line) inserted between the *Pvu* II and *Ava* I sites of pBR322 (thin line). The alanine tRNA gene is represented by a hatched box and its direction of transcription is indicated by an arrow. The pBR322 genes conferring resistance to ampicillin and tetracycline are represented by open boxes. Deletions at the 5′ or 3′ ends of the tRNA gene were generated by performing the indicated manipulations. Clones containing 5′ deleted genes were recovered by ligation of the *Eco*RI fragments into a tetracycline-sensitive vector followed by introduction into *Escherichia coli* (RRI) cells by transformation and selection for tetracycline-resistant colonies. For most 5′ deletions, inserts were obtained in both of the two possible orientations, designated "L" (as shown) or "R." The 3′ deleted fragments were simply circularized by ligation and recovered as ampicillin-resistant plasmids. Ori, origin of replication; RI, *Eco*RI site.

sequences of transcripts from mutant genes and the identities of 5′ terminal nucleotides were confirmed by partial RNA sequence analysis as described (23).

RESULTS

Experimental Strategy and Production of Partially Deleted tRNA Genes. Truncation at a natural *Hin*dIII site has already provided evidence that all or part of an essential $tRNA_2^{Ala}$ gene control element lies within a 5-kb region upstream from −11 (17). By taking advantage of another restriction site (*Pvu* II), we were able to limit the possible upstream boundary of the control signal to the 222 bp immediately adjacent to the gene. That is, a $tRNA_2^{Ala}$ gene retaining only 222 bp of normal 5′ flanking DNA directs transcription as accurately and efficiently as does the same gene with an intact 5-kb flanking region (data not shown). On the 3′ side of this gene, the presence of an *Ava* I site allowed removal of all but 117 bp of flanking sequence. This manipulation also did not affect template activity.

Further delineation of control regions required excision of additional sequences by exonuclease digestion. As diagrammed in Fig. 1, the *Pvu* II/*Ava* I gene-containing fragment inserted in the bacterial plasmid pBR322 was the starting material for this mutagenesis. Our protocol (shown in Fig. 1) took advantage of selectable markers artificially linked to the $tRNA_2^{Ala}$ gene to facilitate recovery of particular mutagenized DNA fragments. To distinguish the mutant phenotypes caused by removal of natural sequences from those due to addition of vector sequences, we wished to vary the vector sequence that replaced normal DNA. Therefore, in constructing the 5′ deleted series we replaced normal DNA with two different sequences by simply varying the orientation of each gene-containing fragment within the plasmid vector. Our choice of possible replacement sequences in the 3′ deletion series, however, was restricted by the necessity of producing transcripts of defined length from genes whose natural termination site had been removed. Thus, all of the *Bombyx* sequences deleted in this series were replaced by a single pBR322 sequence containing a potential alternative RNA polymerase III terminator, the T_5 oligonucleotide at positions 4,324–4,328.

The transcriptional efficiencies of mutant and wild-type templates were compared by measuring the amounts of primary transcript produced from them by using extracts from *Bombyx* silk glands or ovaries. Because processing and degradation are negligible in these extracts (unpublished observations), all of the transcription products are accounted for by the single well-characterized primary transcript (17, 19). To monitor possible

changes in the sites of transcription initiation and termination, the nucleotides at the ends of transcripts from mutant and wild-type genes were compared.

Limits of the 5′ Flanking Signal. As illustrated in Fig. 2*A*, replacement of normal 5′ flanking sequences between positions −222 and −34 with either of two vector sequences does not significantly alter transcriptional activity. Templates deleted up to positions −157, −97, −46, and −34 have 85–100% of wild-type activity (Fig. 3) and direct initiation at the normal site (not shown). In striking contrast, longer deletions that remove the additional 31 nucleotides between −34 and −3 result in complete inactivation of the tRNA gene. These mutants, as well as others with deletions extending even farther into the gene, give no detectable transcripts.

These results localize the upstream boundary of the 5′ flanking control element to the region between −34 and −3. Because the phenotypes of the −3 deletions and the previously reported −11 deletions (17) are indistinguishable, we conclude that this boundary probably lies between −34 and −11. The behavior of templates deleted up to −13 is consistent with this interpretation. The intermediate transcriptional activities of these mutants (15–85% of wild type) could be a consequence of their retention of part but not all of a control region. In keeping with the idea that the transcriptional apparatus may be particularly sensitive to sequences close to −13, the behavior of −13 deletion mutants is strongly influenced by the nature of the vector sequence juxtaposed at this point. Although the site of initiation is not affected, replacement by different vector sequences alters transcriptional activity by 5- to 6-fold (compare −13L and −13R in Figs. 2 and 3). In no other case have we observed an effect of varying these sequences.

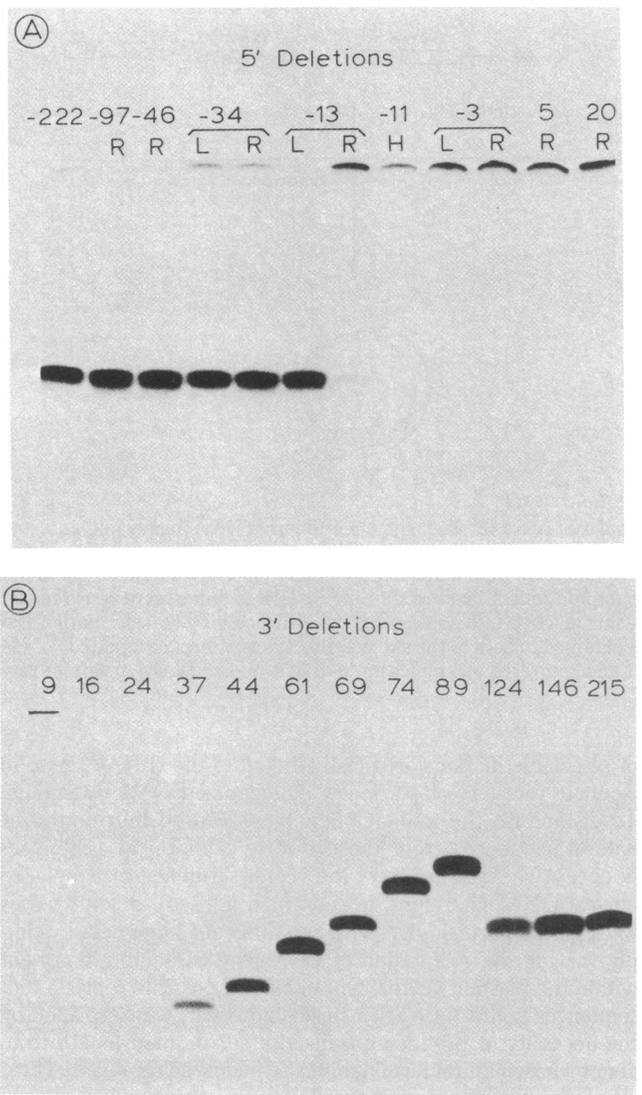

FIG. 2. *In vitro* transcription products of 5′ and 3′ deleted $tRNA_2^{Ala}$ genes. RNA was labeled with [α-^{32}P]GTP during the transcription of mutant genes in extracts from *Bombyx* silk glands. Total reaction mixtures were run on a 10% polyacrylamide gel containing 7 M urea and 0.1% sodium dodecyl sulfate. (*A*) The 5′ deletion mutants. (*B*) The 3′ deletion mutants. Lanes are labeled with the positions of deletion endpoints, numbered relative to the transcription initiation site. Negative numbers indicate positions upstream from the gene. "L" and "R" refer to the orientation of the insert in 5′ deletion mutants. "H" signifies that the previously reported −11 deletion (17) was created by truncation at a natural *Hin*dIII site.

Coding Sequence Requirements. To learn whether essential control elements lie downstream as well as upstream from −11, we measured the template efficiencies of genes deleted from the 3′ direction. Figs. 2*B* and 3 show that the removal of sequences downstream from +61 has little effect on the efficiency of transcription. In contrast, deletion of additional sequences upstream from +61 reduces template activity. Unlike the sharp drop caused by removal of 5′ flanking sequences, the loss of activity associated with progressive deletion of coding sequences is gradual. The transcription efficiency of a gene deleted at position +44 is 60% of the wild-type level, and that of a gene deleted at +16 is still 13%. Complete inactivation finally occurs when only nine nucleotides of the original coding sequence remain.

Because we expected some 3′ deleted genes to encode novel RNAs containing both *Bombyx* and vector sequences, it was important to verify that mutations which appeared to reduce transcriptional efficiency did not act by simply altering the structure—and hence the stability—of the RNA product. Therefore, we examined the transcripts derived from these genes in detail. RNA sequence analysis confirmed the presence of the predicted *Bombyx* and vector sequences in these transcripts, and it also established that none of the mutations alters the site at which initiation occurs. Moreover, long, short, and wild-type-length transcripts are all stable when tested directly by isolation and reincubation in transcription extracts (data not shown).

DISCUSSION

Our analysis of the transcriptional properties of partially deleted $tRNA_2^{Ala}$ genes shows that both 5′ flanking and internal coding sequences probably contribute control functions. It is clear that the upstream boundary of an essential control element lies between nucleotides −34 and −11 preceding the gene. The downstream boundary is less sharply defined, but it appears to enclose sequences within the gene. The coding region close to the transcription initiation site seems particularly crucial.

The 5′ Flanking Control Element. Activity of wild-type $tRNA_2^{Ala}$ genes requires a short, well-defined element just upstream from the transcription initiation site. Alteration of the normal DNA sequence in this region abolishes transcription. The variety of sequence changes here that inactivate the gene suggests that loss of activity is caused by the removal of a positively acting element rather than by the gain of a negatively acting one. As shown in Fig. 4, we have tested four different nonhomologous sequences in this region: two in the −11 deletion mutants described earlier (17) and two others in this work. It is improbable that all of these sequences contain signals capable of turning off the $tRNA_2^{Ala}$ gene. Moreover, genes that

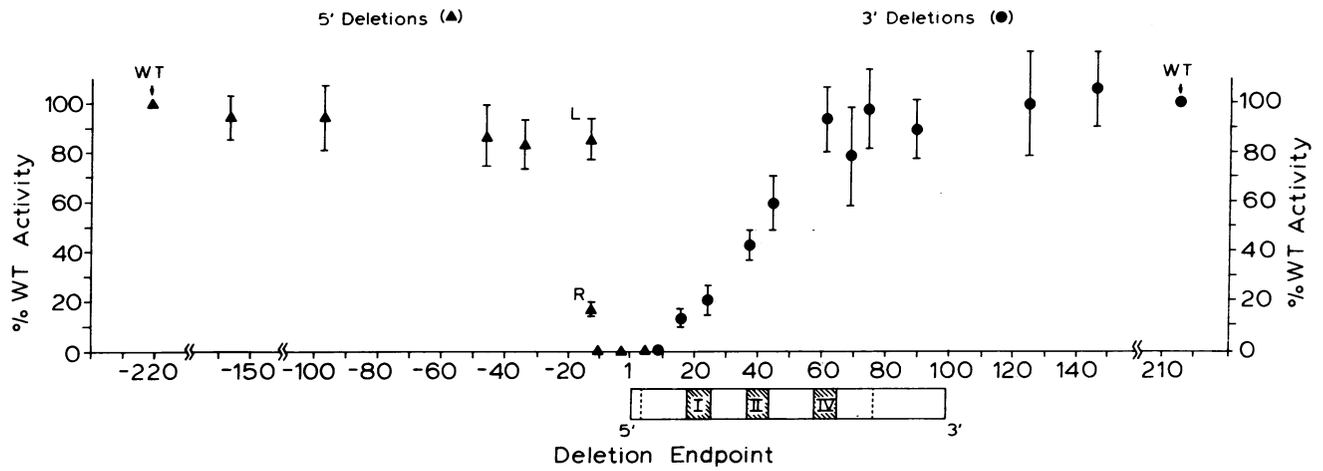

FIG. 3. Transcriptional activity of partially deleted genes. Template activities of wild-type (WT) and partially deleted genes were measured in transcription assays like the ones shown in Fig. 2. The level of transcription from each clone is indicated by a ▲ (5' deletions) or ● (3' deletions) at the position of the deletion endpoint. For reference, the $tRNA_2^{Ala}$ primary transcript is diagrammed below the abscissa. The points at which processing occurs to yield a transcript of the mature length are indicated by broken lines, and loops I, II, and IV of the mature tRNA are shown as hatched boxes. Transcriptional activity values for the mutants are expressed as percentages of the activity of wild-type genes in the same experiment. For 3' deleted genes, values were corrected for variation in the length of the transcripts. The value reported for each mutant represents the mean of five or more independent determinations, using at least two different preparations of supercoiled plasmid DNA. Error bars show the standard deviation of these data. For the −13 deletion mutants, the values for each orientation of the insert are given separately. For all other 5' deletions, L and R values were not significantly different and the value shown represents the mean for both orientations.

have been separated from potentially inhibitory sequences by cleavage with a restriction enzyme are not reactivated (17).

For these reasons, we favor the idea that a particular sequence flanking normal $tRNA_2^{Ala}$ genes provides a positive signal to the RNA polymerase III transcription apparatus. More detailed mutagenesis will be needed to identify the critical sequence precisely, but we have attempted to recognize it by comparing sequences that flank several genes transcribed by *Bombyx* RNA polymerase III. Strikingly, similarities are found only within the region defined by this work as functionally important. Three blocks of sequence homology are located at similar positions within the 30 nucleotides preceding two $tRNA_2^{Ala}$ genes and a 5S RNA gene of *Bombyx* (23). These oligonucleotides, T-A-(C)-T-A-T, A-A-T-T-T-T, and T-T-C are found at positions −29, −20, and −4, respectively, in the wild-type $tRNA_2^{Ala}$ gene analyzed here. Deletions that eliminate or reduce transcriptional activity also remove parts of these sequences (see Fig. 4). The −11 mutants are totally inactive and lack both T-A-(C)-T-A-T and A-A-T-T-T-T. The −13 deletion mutants retain part of the A-A-T-T-T-T sequence and have intermediate activity. In both cases, the tetranucleotide A-A-T-T is fortuitously restored close to its normal location by the *Eco*RI linker used to construct the mutants. We do not know why juxtaposition of different vector sequences in the −13 mutants alters their transcriptional activity. It may be significant, however, that although both mutants contain the A-A-T-T linker-derived sequence, the vector DNA in the more active template provides additional homology (T-A-C) to the T-A-(C)-T-A-T consensus sequence (Fig. 4).

Essential Coding Sequences. To interpret the behavior of 3' deleted genes, we have focused on two major aspects of the results: (*i*) changes just downstream from the transcription initiation site inactivate the gene, but (*ii*) the sequence alterations we have imposed in the 3' half of the coding region have little effect on transcriptional activity. The finding that the $tRNA_2^{Ala}$ gene is very sensitive to sequence changes in the 5' half of the coding region suggests that an intragenic control region may lie just downstream from the transcription initiation site. It does not tell us, however, whether sequences in this part of the gene are actually required for activity, or whether they have simply been replaced in these experiments by an inhibitory vector sequence. We have noted that the sequence from +6 to +18, G-G-G-C-G-T-A-G-(C)-T-C-A-G, within the apparently critical part of the $tRNA_2^{Ala}$ coding region, is also found

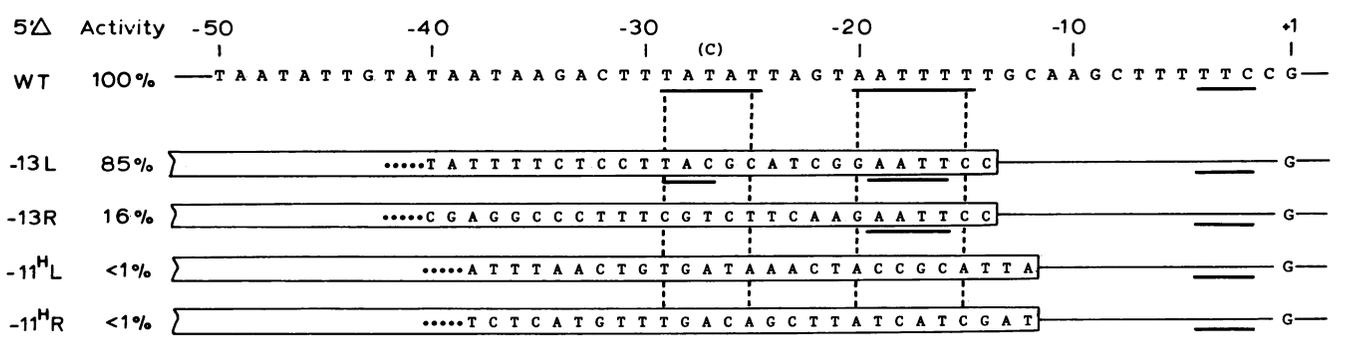

FIG. 4. Comparison of 5' flanking nucleotide sequences of wild-type (WT) and deleted $tRNA_2^{Ala}$ genes. The upper sequence is that of the wild-type gene. Three blocks of sequence homology found in both tRNA and 5S genes of *Bombyx* are underlined. The (C) above the first homology indicates that this sequence sometimes reads T-A-C-T-A-T (23). Deleted tRNA genes are diagrammed below and are aligned with respect to the initiation site (+1). The normal *Bombyx* sequences that are retained in the mutants are represented by thin lines, and replacement sequences supplied by pBR322 are shown inside boxes. The deletion endpoints, insert orientations, and transcriptional activities of the mutants are shown on the left.

inside a *Bombyx* 5S gene (23). The functional significance of this sequence for both of these genes is not yet known.

The fact that replacement of the 3′ half of the $tRNA_2^{Ala}$ coding sequence with vector DNA has little effect on activity suggests that the normal sequences in this region may not be critical for function. It is possible, however, that functionally equivalent sequences provided by vector DNA are responsible for the activity we observe. Others (24) have suggested that deviations from typical secondary or tertiary conformation may be important features of RNA polymerase III control sites. Because a variety of sequences could generate such structures, a systematic testing of nonhomologous replacement sequences will be required to determine whether the 3′ coding region contributes control functions.

Some Control Signals for Transcription of tRNA Genes May Be Species Specific. Our results focus attention on DNA sequences close to the transcription initiation site. At least part of an essential control region is located just upstream from this point, and additional regulatory information may be provided by sequences that are nearby but downstream. The location of the 5′ flanking control element suggests that it is required for direct interaction with the RNA polymerase III complex during transcription initiation. The fact that sequences in this region precede other *Bombyx* genes (23) indicates that the element is likely to have a general role in controlling transcription by *Bombyx* RNA polymerase III. This possibility can be tested by constructing mutant derivatives of a *Bombyx* 5S RNA gene that lack these sequences.

The control regions for *Bombyx* tRNA gene transcription differ from those described for tRNA genes in other organisms in the relative contributions made by flanking and coding sequences. For example, the activity of a *Xenopus* $tRNA_1^{Met}$ gene in homologous transcription extracts requires signals in both the 5′ and 3′ halves of the coding region, but not signals provided by normal 5′ flanking DNA (8). Although transcription is not dependent on the presence of a particular sequence in the 5′ flanking region, sequences with the potential to form Z-DNA can inhibit the expression of this gene *in vitro* (16). Similarly, transcription of *Drosophila* tRNA genes by homologous extracts does not require a single 5′ flanking sequence, but it can be modulated both negatively and positively by removal of normal 5′ flanking sequences (12–14).

Differences among tRNA gene control regions may reflect the use of alternative sets of signals by the RNA polymerase III complexes in different organisms. Comparisons of the activities of tRNA genes in homologous versus heterologous extracts (8, 11, 14) support this idea. For example, the −11 deleted $tRNA_2^{Ala}$ gene that is inactive in *Bombyx* extracts still serves as a template for efficient transcription in *Xenopus* extracts (17). Moreover, a variety of RNA polymerase III templates from other organisms are inactive in *Bombyx* extracts (unpublished data). It appears that tRNA genes contain several regions with the potential for regulatory activity and that the requirements for particular control elements may vary in different organisms. Thus, in *Bombyx*, regulatory information provided by 5′ flanking regions may compensate for a reduction of control functions in regions farther downstream. Variation in the schemes adopted by different organisms may reflect alternative ways of integrating the regulatory elements of individual genes with a more general control network. In this regard it will be interesting to learn whether the conserved sequences just upstream from several yeast tRNA genes are involved in the regulation of tRNA biosynthesis by certain amino acids (25).

We are grateful to Diane Morton, Ira Herskowitz, William Sistrom, and George Sprague for valuable discussions during the course of this work. We also thank Mary Gilland for preparing the figures. This work was supported by National Institutes of Health Research Grants: GM 25388 and HD 00420 to K.U.S., Research Service Award GM 08213 to L.S.Y., and Genetics Training Grant GM 07413 to D.L.

1. Bogenhagen, D. F., Sakonju, S. & Brown, D. D. (1980) *Cell* **19**, 27–35.
2. Sakonju, S., Bogenhagen, D. F. & Brown, D. D. (1980) *Cell* **19**, 13–25.
3. Fowlkes, D. M. & Shenk, T. (1980) *Cell* **22**, 405–413.
4. Guilfoyle, R. & Weinmann, R. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 3378–3382.
5. Fuhrman, S. A., Deininger, P. L., LaPorte, P., Friedmann, T. & Geiduscheck, E. P. (1981) *Nucleic Acids Res.* **9**, 6439–6456.
6. Koski, R. A., Clarkson, S. G., Kurjan, J., Hall, B. D. & Smith, M. (1980) *Cell* **22**, 415–425.
7. Koski, R. A., Allison, D. S., Worthington, M. & Hall, B. D. (1982) *Nucleic Acids Res.* **10**, 8127–8143.
8. Hofstetter, H., Kressman, A. & Birnstiel, M. L. (1981) *Cell* **24**, 573–585.
9. Galli, G., Hofstetter, H. & Birnstiel, M. L. (1981) *Nature (London)* **294**, 626–631.
10. Ciliberto, G., Castagnoli, L., Melton, D. A. & Cortese, R. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 1195–1199.
11. Sharp, S., DeFranco, D., Dingerman, T., Farrell, P. & Söll, D. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 6657–6661.
12. DeFranco, D., Sharp, S. & Söll, D. (1981) *J. Biol. Chem.* **256**, 12424–12429.
13. Sharp, S., Dingermann, T. & Söll, D. (1982) *Nucleic Acids Res.* **10**, 5393–5406.
14. Dingermann, T., Burke, D. J., Sharp, S., Schaak, J. & Söll, D. (1982) *J. Biol. Chem.* **257**, 14738–14744.
15. Wormington, W. M., Bogenhagen, D. F., Jordan, E. & Brown, D. D. (1981) *Cell* **24**, 809–817.
16. Clarkson, S. G. (1982) in *Eukaryotic Genes: Their Structure, Activity and Regulation*, eds. Maclean, N., Gregory, S. P. & Flavell, R. A. (Butterworths, London), in press.
17. Sprague, K. U., Larson, D. & Morton, D. (1980) *Cell* **22**, 171–178.
18. Yaniv, M. (1982) *Nature (London)* **297**, 17–18.
19. Hagenbüchle, O., Larson, D., Hall, G. I. & Sprague, K. U. (1979) *Cell* **18**, 1217–1229.
20. Legerski, R. J., Hodnett, J. L. & Gray, H. B., Jr. (1977) *Nucleic Acids Res.* **5**, 1445–1464.
21. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
22. Brutlag, D. L., Clayton, V., Friedland, P. & Kedes, L. H. *Nucleic Acids Res.* **10**, 279–294.
23. Morton, D. G. & Sprague, K. U. (1982) *Mol. Cell. Biol.* **2**, 1524–1531.
24. Hall, B. D., Clarkson, S. G. & Tocchini-Valentini, G. (1982) *Cell* **29**, 3–5.
25. Del Rey, F. J., Donahue, T. F. & Fink, G. R. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 4138–4142.