Conserved 5' flank homologies in dipteran 5S RNA genes that would function on 'A' form DNA

Alice Rubacha, Walton Sumner III, Lizabeth Richter and Kathy Beckingham

Department of Biochemistry, Rice University, P.O. Box 1892, Houston, TX 77251, USA

Received 18 September 1984; Accepted 3 October 1984

ABSTRACT

We have sequenced the 480 base pair (bp) repeating unit of the 5S RNA genes of the Dipteran fly Calliphora erythrocephala and compared this sequence to the three known 5S RNA gene sequences from the Dipteran Genus Drosophila (1,2). A striking series of five perfectly conserved homologies identically positioned within the 5' flanks of all four Dipteran 5S RNA coding regions has thus been identified. The spacing (12-13 bp) between all of these homologies is typical of A form rather than B form DNA. Given that the eukaryotic 5S RNA gene specific initiation factor TFIID (3) is a DNA unwinding protein (4), a role for these Dipteran 5' flank homologies in initiation site selection on 5S RNA genes transiently unwound for transcription is suggested. One of the Dipteran homology blocks is highly conserved in sequence and position in all but one of the eukaryotic 5S RNA gene sequences known to date (17/18 genes). Its sequence (consensus: TATAAG) and position (average center: -26 bp) are highly reminiscent of the polymerase II gene 'TATA' box (5).

INTRODUCTION

To date, intensive analysis of 5S RNA gene transcription has only been performed for species within the Amphibian Genus Xenopus (for reviews, see References 6 and 7). These studies have established the essential role in the initiation of transcription played by a small DNA region internal to the 5S RNA coding sequence (8,9) (the internal control region or ICR) and have led to the purification of a 5S RNA gene specific transcription factor (3) (TFIID) which appears to bind initially to the 3' region of the ICR (10). An interaction between the polymerase III transcription machinery and the sequences 5' to the 5S RNA coding region has also been revealed by these studies (8,11) and a series of conserved residues GAC, AC/GAAG, and AAAAG has been identified at approximately positions -36, -27, and -17 in the 5' flanking regions of five 5S RNA gene types from the Genus Xenopus (12,6).

We have isolated a cloned stretch of DNA from the 5S RNA gene set of the Dipteran fly Calliphora erythrocephala and sequenced collectively the ~11 copies of a 480 bp tandemly repeating unit present in this plas-

mid. We have compared this sequence to the 5S RNA gene sequences of three species from the Genus *Drosophila* (1,2) (a Genus within the same suborder -- Cyclorrapha -- as *C. erythrocephala*). This is thus the first comparison of 5S RNA genes for a closely related group of organisms outside that performed for the Genus *Xenopus* (12). We have particularly searched these Dipteran 5S genes for homologies within the regions implicated in the control of transcription by the *Xenopus* studies. The ICR is identical in all four species as a result of an extremely high level of 5S RNA coding sequence conservation. The organization of the 3' terminus of transcription is also very similar in all four organisms. Our most interesting finding, however, is a set of five homologies present within the 55 residues upstream from the transcription initiation site in all four species. The possible function of these conserved sequence blocks is discussed in relation to current knowledge of insect and amphibian 5S RNA gene transcription. Other than these homologies, the spacer of the *C. erythrocephala* 5S gene repeats shows no similarity to those of the three *Drosophila* species, but contains two partial pseudogenes of the 3' terminal region of the 5S RNA transcription unit.

MATERIALS AND METHODS

1. Isolation of a cloned DNA fragment containing *C. erythrocephala* 5S RNA genes.

A collection of ~4500 colonies containing cloned random shear fragments of *C. erythrocephala* genomic DNA (Karlson strain) cloned into vector pWR77 by a poly(dA): poly(dT) tailing procedure was generated previously (13). Mixed plasmid DNA preparations, each derived from a fraction of these colonies, were prepared for long-term storage. In Southern hybridization experiments, one of these mixed plasmid preparations was shown to contain a single plasmid species which hybridized strongly to a cloned stretch of 5S RNA genes from *D. melanogaster* (the largest Hae II fragment of plasmid 12D1 -- see Reference 14) and carried a *C. erythrocephala* DNA fragment of 7 kb.

Supercoiled DNA from this mixed plasmid preparation was therefore re-transfected into *E. coli* K12 strain HB101 as described previously (13). Three colonies - pLR1, pLR2 and pLR3 - were identified as containing the 5S gene-bearing plasmid by colony hybridization to i) the Hae II fragment of 12D1 and ii) purified, ³²P-labelled, 5S RNA from *C. erythrocephala*.

Analysis of plasmid DNA preparations from each of these three re-isolates of the cloned 5S DNA revealed that, despite re-introduction into a rec A⁻ host, the 5S DNA insert had undergone deletion in all of them. For

the pLR1 and pLR2 isolates, plasmid DNA prepared from individual colonies was found to be a heterogeneous population of molecules varying in size by a unit length of ~500 bp. For both pLR1 and pLR2, the vast majority of plasmid molecules contained 10 or 11 of these ~500 bp units, as determined from the size of the inserted DNA, but a small fraction of molecules with 2-13 such repeats could be detected. The 7 kb fragment of the original plasmid presumably contained 14 such 500 bp repeats. In the pLR3 re-isolate, a drastic deletion event had reduced the cloned 5S DNA fragment to 800 bp, or approximately 1.5 repeating units.

2. Restriction enzyme site mapping and gel electrophoresis.

Restriction enzymes were obtained commercially and used according to suppliers' directions. Digests were analyzed by agarose gel electrophoresis in TA buffer (40 mM Tris-HCl, pH 7.8, 10 mM sodium acetate, 1 mM EDTA) or by polyacrylamide gel electrophoresis in TBE buffer (90 mM Tris-HCl, pH 8.3, 90 mM sodium borate, 2.5 mM EDTA). The complementary strands of the Rsa I fragments of the 5S gene repeating units were separated by gel electrophoresis as described by Maxam and Gilbert (15).

3. DNA radiolabelling and sequencing.

DNA fragments to be sequenced were isolated from polyacrylamide gels by elution (16). 5S gene fragments from an Alu I/Ava II double digest of pLR1 were labelled at the 3' Ava II termini using DNA polymerase I (Klenow fragment) and α -³²P-dGTP as previously described (17). Single strands of the Rsa I fragments of the 5S repeats were labelled with terminal deoxynucleotidyl transferase and α -³²P-ATP essentially as described by Maxam and Gilbert (15). Radiolabelled DNA was sequenced using the chemical cleavage method (15). Additional reactions for thymine (18,19) and cytosine (19) were sometimes included to confirm sequence data.

Computerized DNA sequence analyses were performed using the University of Minnesota Apple II Sequence Analysis Program (Version 2.1) developed by Larson and Messing (20) and supplied by Dr. S. Berget. All computer analyses were done on an Apple II microcomputer in the laboratory of Dr. K. Matthews.

RESULTS

Restriction enzyme digestion studies of pLR1, pLR2 and pLR3 confirmed that the cloned 5S DNA fragment present in all three re-isolates was entirely composed of ~500 bp repeating units and permitted the construction of a restriction enzyme site map of these units (see Fig. 1). The repeating

FIGURE 1. Restriction enzyme site map of the 5S gene repeating units of pLrI and the strategy used to sequence the repeats. Restriction enzyme sites for Rsa I (●), Alu I (▼), Sau 3A (▲), Hha I (◐) and Ava II (◑) are shown. The position of the 5S coding sequence (solid black bar) and direction of 5S RNA transcription (~~~~~→) are indicated. Distances between sites in bp are shown. Arrows and asterisks below the map indicate the start positions, directions and extents of DNA sequencing.

units of the cloned fragment proved sufficiently homogeneous for the determination of their collective DNA sequence and the strategy used to sequence the repeats of pLrI is shown on the repeat site map in Figure 1. As indicated, almost all of the sequence has been determined on both strands and from more than one DNA fragment.

Figure 2 shows the 480 residue sequence of these 5S RNA gene repeats of *C. erythrocephala*. The sequence coding for the mature 5S RNA (residues 1-120) is almost identical to the single 5S RNA coding sequence common to the 5S genes of the three *Drosophila* species for which data is available -- *melanogaster* (1), *simulans* and *teissieri* (2). The only difference in the *C. erythrocephala* sequence is a C → A transversion at the penultimate residue (no. 119) of the RNA. This level of 5S RNA coding sequence conservation is far greater than that found previously for related insect species (21,22). Residue 119 is part of a Hae III site present within the 5S genes of *D. melanogaster*, *D. simulans* and *D. teissieri*, which is thus missing in *C. erythrocephala*. Interestingly, this Hae III site is also absent in the 5S genes of *D. virilis* and *D. nasuta* (2), suggesting that this C → A mutation may be widespread amongst the Diptera. In the secondary structure proposed for the *D. melanogaster* 5S RNA (23), the two 3' terminal residues (119 and 120) of the 5S RNA are outside the helix formed by the 3' and 5' regions of the RNA molecule (helix I). It is unlikely therefore that any disturbance of the RNA secondary structure is produced by this C → A change.

The transcription units of the three *Drosophila* species are unique in

FIGURE 2. Consensus sequence of the 5S ribosomal RNA gene repeating units present in pLRI.

The sequence of the anti-coding strand is shown, the 5S RNA coding region is underlined and nucleotides are numbered from the first residue of the coding sequence. Direct repeats (●—●—●) and dyad symmetries of greater than 7 residues in length and >80% homology (●—●—●) are shown. The termination region of the 5S transcription unit and the two pseudogenes of this region are boxed.

that the thymine (T) cluster required for termination is 11-12 residues downstream from the last nucleotide of the RNA (1,2). Thus a precursor RNA, first detected in heat-shocked *D. melanogaster* cells (24) and subsequently shown to be the natural precursor of the RNA (25), is the primary transcription product. The 3' region of the *C. erythrocephala* transcription unit shows a very similar organization and therefore a precursor RNA is also clearly implicated in this species. The transcription termination region of *C. erythrocephala* is compared to those of the three *Drosophila* species in Figure 3. The 3' RNA precursor region of *C. erythrocephala* differs from that of the *Drosophila* sequences by the expansion and duplication of a tetra-nucleotide sequence (ACAA) present in *D. melanogaster* and *D. simulans*. In the three *Drosophila* sequences, the first two T clusters of the termination signal seen in *C. erythrocephala* are separated by tandem repetitions or variant derivatives of a heptameric sequence, GCTGCCT (see Fig. 3).

FIGURE 3. Comparison of the transcription termination regions within the 5S gene repeats of 4 Cyclorrhaphous Diptera.

Transcription termination sequences for *C. erythrocephala* (Ce), *D. melanogaster* (Dm) (1), *D. simulans* (Ds) and *D. teissieri* (Dt) (2) are shown. Sequences for the anti-coding strands are shown and numbers indicate base position from the first residue of the coding sequence (see Fig. 2). The sequences are aligned so as to emphasize their similarity. The terminal residues of the 5S RNA coding sequence are overlined with a solid line and those of the 3' precursor region with a dashed line. The known point of transcription termination in *D. melanogaster* is marked by a vertical arrow. The C+A transversion at position 119 of the *C. erythrocephala* 5S RNA coding sequence (see text) is marked by *. A heptameric repeating unit or variant thereof, present in the *Drosophila* repeats, has been displaced from the main sequence so as to align the T clusters, common to all 4 sequences (indicated by —).

Our most striking finding, in comparing the *C. erythrocephala* 5S RNA gene sequence to those of the *Drosophila* species, is a set of five homologies (Blocks 1-5, see Fig. 4) upstream from the transcription initiation site in all four sequences. Given that the evolutionary distance between *C. erythrocephala* and the *Drosophila* species is far greater than that between the *Xenopus* species previously examined for 5S gene homologies (26), these homologies are particularly compelling. It should be noted that they are perfectly conserved in the four species and that their positions relative to the transcription initiation site are very precisely maintained in all four species (± 1 residue).

The most interesting feature of these Dipteran 5' flank homology blocks is their center-to-center distances along the DNA helix (see Table 1). Unlike the homologies seen upstream from the *Xenopus* 5S RNA genes (see Introduction), which have a center-to-center distance of ~ 10 bp, these homologies are all positioned with a spacing of 12-13 bp. This suggests strongly that if functional, these sequences would be recognized on DNA with a helical repeat more typical of A form (27) rather than B form DNA (see Discussion).

FIGURE 4. Conserved sequences within the 5' flanking regions of Dipteran 5S RNA genes.

The sequences of 80 residues preceding the 5S RNA coding sequence for *C. erythrocephala* (C.e.), *D. melanogaster* (D.m.), *D. teissieri* (D.t.) and *D. simulans* (D.s.) are shown. Five blocks of perfectly conserved homology can be identified (Blocks 1-5 shown boxed). These blocks are positioned identically (± 1 residue) within the 5' flank for all 4 species (see Table 1). An additional homology which is not conserved in position within the 4 sequences is underlined. Data for *D. melanogaster* from Reference 1; for *D. teissieri* and *D. simulans* from Reference 2.

An additional homology, ACAA(A)C, is also present in the 5' flanks of all four Dipteran 5S RNA genes (see Fig. 4). However, this sequence is more distant from the transcription start site, and, unlike the other five homologies, differs in position relative to the transcription initiation point. Within the *C. erythrocephala* 5S gene spacer, this sequence is part of an 18 bp almost perfect (17/18 residues) duplication of the 3' terminus of transcription (compare residues 125-142 and 408-424, Fig. 2) and lies within what we have termed a partial pseudogene of the transcription unit (see below). This sequence may therefore represent an evolutionary remnant of

TABLE 1. Center-to-center distances (bp) for 5' flank conserved sequences in Dipteran 5S RNA genes

Species	Block 4-5	Block 3-4	Block 2-3	Block 1-2
<i>C. erythrocephala</i>	13.5	13.5	11.5	12.5
<i>D. melanogaster</i>	13.5	12.5	12.5	12.5
<i>D. teissieri</i>	11.5	13.5	12.5	12.5
<i>D. simulans</i>	13.5	12.5	12.5	12.5
Average distance	13.0	13.0	12.25	12.5

[illegible]

events involved in the formation of the 5S RNA gene repeating units.

We found no significant conservation of the Dipteran homology blocks 1-5 in the 5' flanks of the 28 *D. melanogaster* tRNA gene sequences reported to date (28). Blocks 1, 2, 4, and 5 are also not well conserved in the 5' flanks of the 18 further eukaryotic 5S RNA genes for which sequence data is available (29-38) (see Fig. 5). However, block 3, as an expanded consensus, TATAAG, was found to be conserved in both position and sequence within the 5' flanks of all eukaryotic 5S RNA gene sequences determined to date except one from the Syrian hamster, *Mesocricetus auratus* (see Fig. 5). Interestingly, a second class of 5S RNA genes is also present in the genome of the Syrian hamster and it is not known which genes are actually transcribed *in vivo* (37).

Beyond the immediate vicinity of the 5S RNA coding sequence, the spacer of the *C. erythrocephala* 5S gene repeats shows no homology to those of the *Drosophila* species, the only common feature being a very high A-T content (*C. erythrocephala* 79%; *D. melanogaster* 74%). Many repeated sequences and dyad symmetries (see Fig. 2) were located within the *C. erythrocephala* spacer using the computer programs of Larson and Messing (20), but very few examples of tandem duplication of a sequence were found. Thus, the spacer contains only three examples of tandem duplication of a sequence greater than 5 bp in length (see Fig. 2), one of which is the duplication of [T₆A] found within the transcription termination signal. Versions of one of these tandemly duplicated sequences (ACATGCAAA) containing a single base mismatch were also found at three further places within the sequence.

A conspicuous feature of the sequence repetitions within the *C. erythrocephala* 5S repeating unit is that almost none of them are repetitions

FIGURE 5. (see opposite) Homologies to the Dipteran conserved sequences in the 5' flanking regions of other eukaryotic 5S RNA genes. Residues in bold type represent homologies or partial homologies to Blocks 1 to 5 of the Dipteran 5S RNA genes that are located at relatively similar positions (± 10 residues) to the equivalent Dipteran homology blocks. Only the 5S RNA genes of *B. mori* and *L. luteus* show significant conservation of all five homology blocks at approximately the appropriate positions within their 5' flanking regions. In general, this comparison indicates that Blocks 1, 2, 4 and 5 are not conserved outside the Diptera, but shows that Block 3, shown boxed and expanded to a longer consensus, TATAAG, is highly conserved (both in sequence and position) in the 5' flanks of all of these eukaryotic 5S RNA genes except that of the Syrian hamster, *Mesocricetus auratus*. Both the location (center position = -26) of this homology and the sequence itself are highly reminiscent of the "TATA" box of polymerase II genes (5). The asterisks over residues in the *B. mori* 5' flanking sequence indicate 3 oligonucleotides which are conserved in the 5' flanks of 3 polymerase III genes of *B. mori* (35).

FIGURE 6. Organization of the pseudogenes within the spacer of the *C. erythrocephala* 5S gene repeating units.

The positions of the two pseudogenes (indicated by ∇ above hatched boxes) relative to the gene and the termination region of the transcription unit (hatched box) are shown. The locations of the precursor RNA and the T-clusters of the termination signal are also indicated. Numbers refer to residue positions as shown in Figure 2.

of sequences within the region encoding the mature 5S RNA (see Fig. 2). However, sequences highly homologous (~80%) to a ~40-bp region encoding the 3' terminal precursor RNA sequences of the transcription unit and the termination signal itself are found at two positions within the spacer. These duplications are indicated in Figure 2 and their overall organization within the repeat unit is shown in Figure 6. We have termed these duplications partial pseudogenes since they represent abortive duplications of a portion of the functional 5S gene transcription unit.

DISCUSSION

1. The Dipteran 5S gene 5' flank homologies.

Although in *Xenopus* the ICR plays the dominant role in initiation of polymerase III transcription, it has been suggested that in lower eukaryotes the 5' flanking sequences might be the crucial determinants in engaging the polymerase III transcription machinery (39). Thus the native 5' flanking sequences of both a tRNA gene (40) and a 5S RNA gene (41) from the silkworm, *Bombyx mori*, and of a tRNA gene from *D. melanogaster* (42) have been shown to be critical for transcription in their homologous polymerase III transcription systems, but not for activity in extracts prepared from *Xenopus* oocytes. This suggests that the conserved 5' flank homologies we have detected in the Dipteran 5S RNA genes might prove essential for 5S RNA gene transcription by Dipteran polymerase III. However, an analysis of the 5S genes which are active in a transcription system from *D. melanogaster* K_c cells (43) indicates that, in this species, as in *X. laevis*, the interaction of a TFIID-like initiation factor with the ICR ultimately determines whether a 5S RNA gene transcription unit will engage the polymerase III machinery (see Fig. 7). These Dipteran homology blocks are thus more likely to be involved in a secondary interaction -- perhaps functioning to enhance

5S RNA GENE ICR REGION SEQUENCE									
+50		+60		+70		+80		+90	
CCGAAATTAAGCAGCGTCGGGCGCGGT		TTAGTACTTAGATGGGGACCGCTT							<u>D. melanogaster</u>
-A-G-----G-----CT-----C-G-----A-----C-									<u>N. viridescens</u>
-G-GCG-T--G-C--CT-----C-G-----A-----C-									gene 1
-G-GCG-T--AG-----CT-----C-G-----A-----C-									gene 2 <u>X. borealis</u>
-G-GCC-T--G-G-C-AC-CT-----C-G-----A-----AC-									oocyte gene 3
-T-C-G-----CTCTGA--CTC-----AT-G-T--A--AA-A-									<u>S. pombe</u>

FIGURE 7. Comparison of the ICR regions for 5S RNA genes tested in a polymerase III transcription system from D. melanogaster K_c cells (43). The region of the D. melanogaster 5S RNA coding sequence (1) equivalent to the ICR of *Xenopus* (6) is shown and compared to the same region of three 5S RNA gene types tested *in vitro* for activity in a D. melanogaster transcription system. Only residues which differ from the D. melanogaster sequence are shown (see Fig. 5 for data sources). The 5S RNA genes of N. viridescens are active in the D. melanogaster system (43), whereas the S. pombe 5S RNA gene, pYM3, and the X. borealis oocyte-type genes are not (Reference 43 and S. Sharp, personal communication). The ICR comparison shown above demonstrates that the homology of the active N. viridescens 5S RNA gene ICR region to the Dipteran equivalent is considerably higher (particularly in the 5' region) than that of the inactive 5S RNA gene types. None of these three non-Dipteran 5S RNA gene types shows compelling conservation of the Dipteran homology blocks 1-5 in their 5' flanks (see Fig. 5). Thus, as in *Xenopus*, ICR interaction with a TFIIIA-like initiation factor would appear to determine whether a 5S RNA gene type may be transcribed by the D. melanogaster polymerase III machinery.

the efficiency of initiation by precise and rapid orientation of the polymerase to the correct initiation site -- a critical decision for 5S RNA genes where transcription begins at the first residue of the functional RNA sequence.

The center-to-center distances between these Dipteran 5' homology blocks coincide with the helical repeat spacing of A form rather than B form DNA. As transcription commences, the initiation site must at least transiently adopt an A form configuration, since RNA/DNA double helices are always of this type (27). Thus, it can be postulated that these sequences would operate on DNA transiently underwound to an A form helix in preparation for transcription and would function to effect efficient engagement of the polymerase at the correct start position.

Previous studies (4) readily suggest a possible effector source for this transition in the DNA configuration. Reynolds and Gottesfeld (4) have determined that binding of the *Xenopus* initiation factor TFIIIA to the ICR

of X. laevis oocyte-specific 5S RNA genes produces an unwinding of the DNA. Their results are more consistent with overall DNA unwinding as a result of a change in winding angle rather than melting of the DNA. The magnitude of the change observed on binding TFIID to cloned 5S DNA in an otherwise protein-free solution was 0.3 helical turns per ICR region. Although this is not sufficient to account for a transition to A form DNA for a region extending upstream as far as 50 residues 5' to the transcription initiation site, it is possible that in vivo, where nucleosomal packaging (44,45) and other polymerase III specific transcription factors (46) would be involved, more extensive unwinding would be achieved. It is also possible that the range and extent of the helical transition produced by the TFIID-type initiation factor differs significantly from organism to organism. As noted earlier, the 5' flanking homologies of the *Xenopus* genes show a spacing more consistent with recognition on DNA with a helical repeat of 10 bp.

2. Limitation of the Dipteran homology blocks to Dipteran 5S RNA genes.

The absence of any significant conservation of these Dipteran homologies within the 5' flanks of all known tRNA gene sequences of D. melanogaster suggests that, if functional, these homologies interact with a component of the polymerase III transcription system specific for 5S RNA genes. This finding contrasts sharply with data for one 5S RNA and two alanine tRNA genes of another insect, B. mori (35). These three genes have proved to share three highly conserved 5' flank homologies, shown by functional studies to lie within 5' flank regions required for transcription (40,41).

The conservation of homology block 3 of the Dipteran 5S genes in all known functional eukaryotic 5S genes (see Fig. 5) and the strong similarity of this sequence to the 'TATA' box of polymerase II genes (5) argue persuasively that this region plays an active role in 5S gene transcription. Functional significance for this sequence, at least in lower eukaryotes, is also indicated by a recent deletional analysis of 5S RNA gene transcription in B. mori (41). Deletion of the 5' flanking sequences more distal than this sequence block (residues -22 to -28, see Fig. 5) has no effect upon transcription whereas deletions which disrupt or remove this sequence reduce transcriptional activity at least two fold.

Homology blocks 1, 2, 4 and 5 show poor conservation outside the Diptera (see Fig. 5) although partial conservation of all five blocks is seen within the only other insect 5S RNA gene examined -- that of B. mori

(see Fig. 5). Similarly, the two additional blocks of homology present in the *Xenopus* 5S RNA genes (see Introduction) show no conservation outside the Genus. This absence of conservation need not indicate a lesser role for these residues in transcription initiation but might rather reflect a special property of the 5S gene transcriptional machinery shared with the polymerase I system -- that of recognizing a single DNA sequence within the genome. This property could allow rapid co-evolution of some of the transcription initiation site sequences and their binding protein(s), so that, for these two transcription systems, sequence conservation might only be detectable -- as indicated here -- amongst related groups of organisms. DNA sequence comparisons around polymerase I transcription start sites across large evolutionary distances are also disappointing (47) and two polymerase I transcription systems (48,49), like the *B. mori* polymerase III transcription system (35), have proved specific for their homologous genes.

3. Pseudogenes within the spacer of the *C. erythrocephala* 5S RNA genes.

Pseudogenes are found within the 5S RNA gene spacers of *Xenopus laevis*, *Xenopus borealis* (12,50) and *Notophthalmus viridescens* (36). Those identified here in the *C. erythrocephala* 5S genes (see Fig. 6) are most similar to the pseudogene of the *N. viridescens* 5S repeats which is a duplication of the 3' terminal 36 bp of the 5S RNA coding sequence and 2 bp of adjacent spacer DNA. All four examples of 5S RNA pseudogenes show the same transcriptional polarity as their related 5S gene sequence and show sequence homology to the 5S gene sequence proper ranging from 75-100%. A distinct feature of the *C. erythrocephala* pseudogenes is their location relative to the functional 5S RNA coding sequence. They are localized closer to the 5' rather than the 3' terminus of the 5S RNA coding region with the pseudogene showing greatest homology to the gene being only 38 residues from the transcription start site (see Fig. 6). It seems likely that all of these 5S pseudogenes have their origins in the recombinational events involved in the generation and maintenance of the 5S gene arrays.

ACKNOWLEDGEMENTS

We thank Dr. S. Artavanis-Tsakonas for a gift of plasmid 12D1 DNA; Dr. J. D. Procunier for a gift of pure *D. melanogaster* 5S RNA; Dr. A. Hopper for chromatographic purification of 5S RNA from *C. erythrocephala*; and Drs. S. Berget and K. Matthews for use of their computer facilities and programs. We are grateful to Drs. S. Sharp and K. Sprague for helpful discussions and Marie Monroe for her careful preparation of the manuscript. This work was

supported by NIH Grant HD 17688 and grant C-848 from the Welch Foundation.
A.R. is supported by NIH Training Grant GM 07833.

REFERENCES

1. Tschudi, C. and Pirrotta, V. (1980) *Nucleic Acids Res.* 8, 441-451.
2. Samson, M.-L. and Wegnez, M. (1984) *Nucleic Acids Res.* 12, 1003-1014.
3. Engelke, D. R., Ng, S. Y. N., Shastri, B. S. and Roeder, R. G. (1980) *Cell* 19, 717-728.
4. Reynolds, W. F. and Gottesfeld, J. M. (1983) *Proc. Natl. Acad. Sci. U.S.A.* 80, 1862-1866.
5. Corden, J., Waslylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, D. and Chambon, P. (1980) *Science* 209, 1406-1414.
6. Korn, L. J. and Bogenhagen, D. F. (1982) in *The Cell Nucleus*, Busch, H. and Rothblum, L., Eds., Vol. 12, rDNA, Part C, pp. 1-27, Academic Press, New York.
7. Ciliberto, G., Castagnoli, L. and Cortese, R. (1983) in *Current Topics in Developmental Biology*, Vol. 18, pp. 59-87.
8. Sakonju, S., Bogenhagen, D. F. and Brown, D. D. (1980) *Cell* 19, 13-25.
9. Bogenhagen, D. F., Sakonju, S. and Brown, D. D. (1980) *Cell* 19, 27-35.
10. Sakonju, S., Brown, D. D., Engelke, D., Ng, S. Y. N., Shastri, B. S. and Roeder, R. G. (1981) *Cell* 23, 665-669.
11. Wormington, W. M., Bogenhagen, D. F., Jordan, E. and Brown, D. D. (1981) *Cell* 24, 809-817.
12. Korn, L. J. and Brown, D. D. (1978) *Cell* 15, 1145-1156.
13. Beckingham, K. and White, R. (1980) *J. Mol. Biol.* 137, 349-373.
14. Artavanis-Tsakonas, S., Schedl, P., Tschudi, C., Pirrotta, V., Steward, R., and Gehring, W. J. (1977) *Cell* 12, 1057-1067.
15. Maxam, A. M. and Gilbert, W. (1980) in *Methods in Enzymology*, Grossman, L. and Moldave, K. D., Eds., Vol. 65, pp. 499-560, Academic Press, New York.
16. Munjaal, R. P., Chandra, T., Woo, S. L. C., Dedman, J. R., and Means, A. R. (1981) *Proc. Natl. Acad. Sci. U.S.A.* 78, 2330-2334.
17. Smith, V. L. and Beckingham, K. (1984) *Nucleic Acids Res.* 12, 1707-1724.
18. Friedman, T. and Brown, D. M. (1978) *Nucleic Acids Res.* 5, 615-622.
19. Rubin, C. M. and Schmid, C. W. (1980) *Nucleic Acids Res.* 8, 4613-4619.
20. Larson, R. and Messing, J. (1982) *Nucleic Acids Res.* 10, 39-49.
21. Xian-Rong, G., Nicoghossian, K. and Cedergren, R. J. (1982) *Nucleic Acids Res.* 10, 5711-5716.
22. Kawata, Y. and Ishikawa, H. (1982) *Nucleic Acids Res.* 10, 1833-1840.
23. Leuhrsen, K. R. and Fox, G. E. (1981) *Proc. Natl. Acad. Sci. U.S.A.* 78, 2150-2154.
24. Rubin, G. M. and Hogness, D. S. (1975) *Cell* 6, 207-213.
25. Jacq, B., Jourdan, R., and Jordan, B. R. (1977) *J. Mol. Biol.* 117, 785-795.
26. Beverley, S. M. and Wilson, A. C. (1984) *J. Mol. Evol.*, in press.
27. Saenger, W. (1984) *Principles of Nucleic Acid Structure*, pp. 242-281, Springer-Verlag, New York.
28. Gauss, D. H. and Sprinzl, M. (1984) *Nucleic Acids Res.* 12, r59-132.
29. Maxam, A., Tizard, R., Skryabin, K. G. and Gilbert, W. (1977) *Nature* 267, 643-645.
30. Tabata, S. (1980) *Eur. J. Biochem.* 110, 107-114.
31. Tabata, S. (1981) *Nucleic Acids Res.* 9, 6429-6437.
32. Mao, J., Appel, B., Schaack, J., Sharp, S., Yamada, H. and Söll, D. (1982) *Nucleic Acids Res.* 10, 487-500.

33. Selker, E. U., Yanofsky, C., Driftmier, K., Metzenberg, R. L., Alzner-DeWeerd, B., and RajBhandary, U. L. (1981) *Cell* 24, 819-828.
34. Lu, A.-L., Steege, D. A. and Stafford, D. W. (1980) *Nucleic Acids Res.* 8, 1839-1853.
35. Morton, D. G. and Sprague, K. U. (1982) *Mol. Cell. Biol.* 2, 1524-1531.
36. Kay, B. K. and Gall, J. G. (1981) *Nucleic Acids Res.* 9, 6457-6468.
37. Hart, R. P. and Folk, W. R. (1982) *J. Biol. Chem.* 19, 11706-11711.
38. Rafalski, J. A., Wiewiorowski, M. and Söll, D. (1982) *Nucleic Acids Res.* 10, 7635-7642.
39. Ciliberto, G., Rangei, G., Costanzo, F., Dente, L. and Cortese, R. (1983) *Cell* 32, 725-733.
40. Sprague, K. U., Larson, D. and Morton, D. (1980) *Cell* 22, 171-178.
41. Morton, D. G. and Sprague, K. U. (1984) *Proc. Natl. Acad. Sci. U.S.A.*, in press.
42. Sharp, S., Dingermann, T. and Söll, D. (1982) *Nucleic Acids Res.* 10, 5393-5406.
43. Dingermann, T., Sharp, S., Appel, B., DeFranco, D., Mount, S., Heiermann, R., Pongs, O. and Söll, D. (1981) *Nucleic Acids Res.* 9, 3907-3918.
44. Louis, C., Schedl, P., Semel, B. and Worcel, A. (1980) *Cell* 22, 387-392.
45. Gottesfeld, J. M. and Bloomer, S. L. (1980) *Cell* 21, 751-760.
46. Lassar, A. B., Martin, P. L. and Roeder, R. G. (1983) *Science* 222, 740-748.
47. Long, E. O., Rebbert, M. L. and Dawid, I. B. (1981) *Proc. Natl. Acad. Sci. U.S.A.* 78, 1513-1517.
48. Kohorn, B. D. and Rae, P. M. M. (1982) *Proc. Natl. Acad. Sci. U.S.A.* 79, 1501-1505.
49. Grummt, I., Roth, E. and Paule, M. R. (1982) *Nature* 296, 173-174.
50. Jacq, C., Miller, J. R., and Brownlee, G. G. (1977) *Cell* 12, 109-120.