Discussion Letter

Analysis of the base substitutions found in the *Xenopus laevis* 5 S RNA pseudogene

Janet Andersen, Nicholas Delihas and Keith Thompson+

Dept. of Microbiology, SUNY Stony Brook, Stony Brook, NY 11794 and *+Biology Dept., Brookhaven National Laboratory, Upton, NY 11793, USA*

Received 5 January 1983

A 5 S RNA pseudogene is associated with the major oocyte 5 S RNA gene of *Xenopus laevis*. *X. borealis* has several oocyte specific 5 S RNA genes. Gene 1 is the dominant 5 S RNA gene. Gene 3 has sometimes been referred to as a pseudogene. We show that the base substitutions in the *X. laevis* 5 S pseudogene are non-random with respect to double and single-stranded regions of the 5 S RNA using the χ^2 test of homogeneity with Yates correction for continuity. In addition, conserved positions of eukaryotic 5 S RNAs are predominantly maintained. *X. borealis* gene 3 is random in mutations.

5 S RNA gene *Pseudogene* *RNA* *Oocyte* Xenopus *mutation*

1. INTRODUCTION

5 S RNA is a small RNA that is about 120 nucleotides in length. It is associated with specific ribosomal proteins in the larger ribosomal subunit in all organisms and is associated with proteins in 7 S and 42 S storage particles in certain amphibians and teleosts [1]. The protein in the 7 S particle is a transcription factor protein [2]. All 5 S RNAs conform to a consensus secondary structure which is essentially the Fox and Woese model for prokaryotic 5 S RNA with extended base pairing [3–6]. This secondary structure has been proposed for eubacterial, archaebacterial, organelle, and eukaryotic 5 S RNA sequences [7,8]. Conserved nucleotide positions and conserved chain lengths between some of these positions were determined [8]. The conserved nucleotides found in greater than 90% of the compared eukaryotic 5 S RNAs are termed common eukaryotic positions. Most of these common positions are concentrated in single-stranded regions and at the ends of helices in the 5 S RNA model (fig.1). The function of these conserved nucleotide positions is not known but they may be of use at several metabolic levels: at the DNA level during transcription, or at the RNA level for maintaining secondary and tertiary structure for the binding of ribosomal and storage proteins including the transcription factor protein, or in protein synthesis itself.

The 5 S ribosomal RNA genes of *Xenopus laevis* and *Xenopus borealis* have been well characterized [9]. Several of the 5 S RNA genes occur as multigene families reiterated hundreds to thousands of times. Both *X. laevis* and *X. borealis* have 5 S ribosomal RNA genes that code for somatic-specific and oocyte-specific 5 S RNAs. *X. laevis* has major and trace oocyte 5 S RNA genes. A 5 S RNA pseudogene is associated in a tandem unit with only the major oocyte 5 S RNA gene. This pseudogene is less than 80 nucleotides downstream from the major oocyte gene [10]. The unit is repeated 24000-times/cell [10]. Representatives of each of these classes of 5 S RNA genes

Published by Elsevier Biomedical Press
00145793/83/0000–0000/$3.00

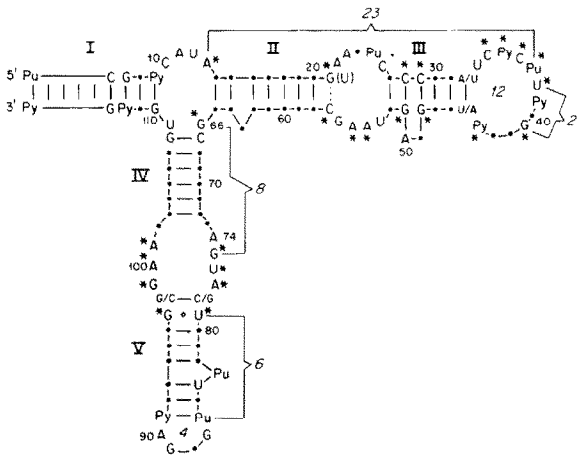

Fig.1. Consensus secondary structural model for eukaryotic 5 S RNAs with common eukaryotic positions and conserved chain lengths between some of these positions highlighted. (*) Common eukaryotic positions that are also 'universal' and found in all classes of 5 S RNAs.

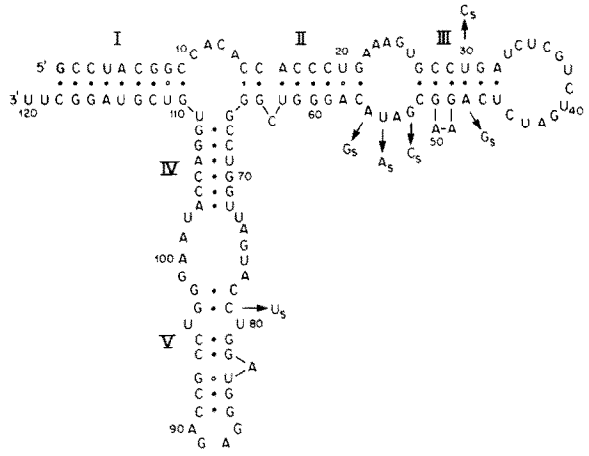

Fig.2. *X. laevis* major oocyte 5 S RNA drawn in the consensus secondary model. The arrows show the base substitutions found in the somatic 5 S RNA of *X. laevis*.

have been sequenced [11,12] and have been shown to be transcriptionally active in the in vitro system involving the microinjection of 5 S DNA into *Xenopus* oocyte nuclei [13,14]. Although detected in vitro, no transcript has been found for the 5 S RNA pseudogene in vivo [13]. *X. borealis* also has several oocyte-specific 5 S RNA genes. Genes 1, 2, and 3 have been sequenced from a cloned region (Xbo1) of the 5 S DNA [15]. Gene 1 is the dominant oocyte 5 S RNA gene. Gene 3 has sometimes been referred to as a pseudogene.

2. ANALYSIS OF *X. LAEVIS* 5 S RNA PSEUDOGENE

Although the extent of sequence difference between 5 S RNA genes 1 and 3 of *X. borealis* is about the same as the difference between the 5 S RNA gene and 5 S pseudogene of *X. laevis* [15], the nature of the base substitutions are very different. In examining the *X. laevis* 5 S RNA pseudogene in light of the common eukaryotic positions some interesting base substitutions appear to be evident.

X. laevis major oocyte 5 S RNA has 6 substitutions from the somatic 5 S RNA (fig.2). Positions 52 (G) and 55 (A) occur in common eukaryotic positions in a single-stranded region. Positions 54 (U) and 79 (C) occur in 'universal' positions; i.e., residues conserved in all classes (eukaryotic, eubacterial, archaebacterial, and organelle) of functional 5 S RNAs [8]. These 4 substituted residues are in the intragenic transcriptional control region [9]. The positions in the 5′-section (positions 52, 54 and 55) of this control region have been shown to effect the competitive transcription strength of the DNA. Mutant somatic 5 S RNA genes having oocyte base substitutions in this region have the reduced transcriptional competition strength of the oocyte-type genes [16].

The *X. laevis* RNA 5 S pseudogene putative transcript has been drawn in the consensus secondary structural model for 5 S RNAs in fig.3. The base pairings found in the consensus model have been left in where base substitutions do not directly disrupt the base pair, even though some of the base pairs would be thermodynamically unstable in the structure as shown [17]. The *X. laevis* 5 S pseudogene has the same 6 base substitutions that the major oocyte 5 S gene has. In addition, it has another 10 base substitutions from positions 1–101. Positions 102–108 form part of the termination signal for transcription [13,18], however these positions are only partially effective for transcriptional termination in vitro. The 10 base substitutions in the region 1–101, as viewed in a putative transcript (fig.3), do not occur in any

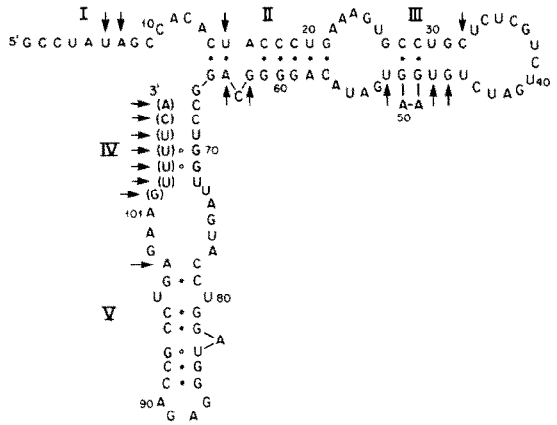

Fig.3. *X. laevis* 5 S pseudogene drawn as a transcript using the consensus secondary structure as a template. The arrows show the base substitutions that occur from the major oocyte 5 S RNA from *X. laevis*. The base pairing found in the consensus model of fig.1 has been left in where base substitutions do not directly disrupt the base pair, even though some of the base pairs would be thermodynamically unstable in the structure as shown.

single-stranded regions as determined by the consensus secondary structural model. The substitutions either disrupt or weaken the secondary structure of the putative transcript. Helices I and IV are totally disrupted. Helix I is disrupted as a consequence of the deletion of residues beyond 101. The base substitutions from positions 102–108 completely disrupt helix IV. It is of interest to note that no base substitutions occur on the 5′-side of helix IV. Helices II and V are partially disrupted at one end by the base substitutions. The pseudogene putative transcript seems to maintain a G–C rich hairpin situated near its 3′-end (positions 79–97). On the DNA, this region constitutes the distal part of the transcriptional control region. Helix III seems to be totally disrupted since the G–C base pairs (positions 28, 29, 48 and 51) flanking the 2 looped-out positions would be thermodynamically unstable [17]. Table 1 shows a summary of the base substitutions of the putative transcript. There are no changes in common eukaryotic positions with the exception of positions 6 and 7 which are part of helix I in the 5 S RNA consensus model (cf. figs.1 and 3). The two conserved positions appear to elude assignment of a function in the RNA since they are not situated at the end of a helix or flanking looped-out positions as are the other conserved positions in double-helical regions.

Table 1

Summary of base substitutions in putative transcripts of 5 S pseudogene and gene 3

	5 S Pseudogene	Gene 3
No. of positions	101	120
in single-strands	44	47
in double-strands	57	73
No. of mutations	10	15
in single-strands	0	5
in double-strands	10	10
Chi square	6.18[a]	0.0
Probability	0.05	
No. of conserved positions	50	56
No. of mutations	10	15
in conserved positions	2	6
in non-conserved positions	8	9
Chi square	2.40	0.0
Probability	0.15	

[a] Significant at the 5% level

In contrast, gene 3 of *X. borealis* has an entirely different quality of base substitutions. Even though it has the oocyte base substitutions of gene 1 at most positions, position 53, a common eukaryotic position in 5 S RNAs, remains a G as in

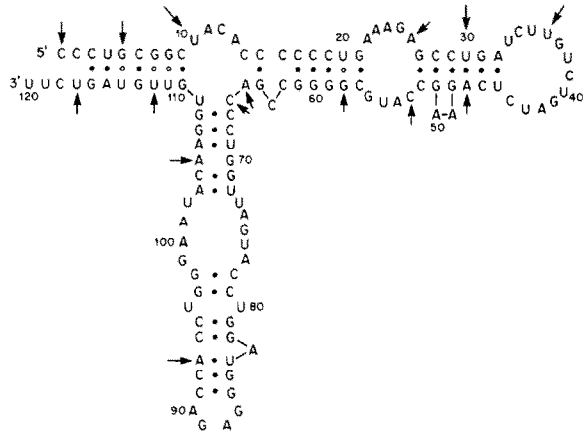

Fig.4. *X. borealis* oocyte 5 S RNA gene 3 drawn as an RNA using the consensus secondary structure. The arrows show the base substitutions that occur from gene 1.

the somatic 5 S gene. The other 15 base substitutions that are dispersed throughout the gene appear in single-stranded and double-stranded regions of its transcript (see fig.4). Five of the substitutions are in common eukaryotic positions. Four of the base substitutions partially disrupt the secondary structure while 6 of the substitutions simply maintain the helices. Despite the large number of base substitutions found in this 5 S RNA gene, the secondary structure of its transcript still conforms to the consensus model for 5 S RNA in fig.1. Two of the 15 base substitutions have been observed in the population of RNAs from *X. borealis* [15]. Therefore we suggest that the product of gene 3 may be an actual ribosomal 5 S RNA in *X. borealis*.

To test the hypothesis that the distribution of the observed mutations in the *X. laevis* 5 S RNA pseudogene and in *X. borealis* 5 S RNA gene 3 is random over the double- and single-stranded regions and the conserved and the non-conserved regions of the secondary structures, we used the chi square test of homogeneity (one degree of freedom) with Yates correction for continuity [19]. Table 1 shows the results. All 10 of the mutations in the 5 S pseudogene were located in the putative double-stranded region ($\chi^2 = 6.63$, $p < 0.05$), providing evidence of a non-random distribution. Only two of the 10 mutations were in conserved positions ($\chi^2 = 2.40$, $p < 0.15$). Although this latter result was not statistically significant due in part to the conservative nature of the corrected chi square test, the 4:1 ratio of mutations in the conserved *vs.* the non-conserved positions is striking. The difference in probability noted here ($p = 0.05$ *vs* $p = 0.15$) results from the two base substitutions mentioned at positions 6 and 7 found in helix I which are common eukaryotic positions.

In contrast, the 15 mutations in *X. borealis* gene 3 were distributed proportionally over the single- and double-stranded regions ($\chi^2 = 0.0$) and over the conserved and non-conserved positions ($\chi^2 = 0.0$), providing no evidence of a non-random distribution.

3. CONCLUSION

What we show is that the base substitutions in the *X. laevis* 5 S pseudogene are non-random with respect to double- and single-stranded regions of the 5 S RNA. Because of the combination of mutations in double-stranded regions and the block deletion at the 3'-end, the 5 S RNA pseudogene transcript can not form the 5 S RNA secondary structure depicted in fig.1. It may form a functional secondary structure unique to itself. It is unlikely that a 5 S RNA pseudogene functions in ribosomal protein binding where the 5 S RNA secondary structure is crucial [20]. The importance of the 5 S RNA secondary structure in 7 S and 42 S particle protein binding is not known, nor is it known whether a 5 S RNA pseudogene transcript could bind to these proteins.

The conservation in the 5 S RNA pseudogene of the common eukaryotic positions that appear in single-stranded regions of a ribosomal 5 S RNA suggests that the pseudogene has a function that employs such positions. It is worth noting again that there are base substitutions in two common eukaryotic positions [6,7] in helix I. These positions appear to be different from the other conserved positions in double-helical regions since they are not situated at the ends of the helix or flanking any looped-out positions in a 5 S ribosomal RNA.

One problem in determining a function of the *X. laevis* 5 S RNA pseudogene is that its transcript has not been found in the oocyte. There are several explanations proposed for the evasiveness of a transcript in the oocytes:

(1) No transcription of the 5 S RNA pseudogene occurs in the oocyte;

(2) The transcripts are not of uniform length and so have not been detected as distinct bands on gels [13];

(3) The transcripts have a short half-life [13];

(4) The RNAs are only transcribed during a brief interval in early oocyte development. It is also possible that a function of the pseudogene may not depend on its being transcribed.

Even though the 5 S RNA pseudogene from *X. laevis* has been considered a relic of evolution due to the number of observed base substitutions, we have shown that the type of mutations are highly specific. This suggests that the 5 S RNA pseudogene has a functional role.

ACKNOWLEDGEMENTS

We thank Daniel Bogenhagen for reading the manuscript and for suggestions. Supported by grant National Institutes of Health GM-20052 and in part by US Dept. of Energy.

REFERENCES

[1] Picard, B. and Wegnez, M. (1979) Proc. Natl. Acad. Sci. USA 76, 241–245.
[2] Pelham, H.R.B. and Brown, D.D. (1980) Proc. Natl. Acad. Sci. USA 77, 4170–4174.
[3] Fox, G.E. and Woese, C.R. (1975) Nature 256, 505–507.
[4] Nishikawa, K. and Takemura, S. (1974) J. Biochem. (Tokyo) 76, 935–947.
[5] Luehrsen, K.R. and Fox, G.E. (1981) Proc. Natl. Acad. Sci. USA 78, 2150–2155.
[6] Studnicka, G.M., Eiserling, F.A. and Lake, J.A. (1981) Nucl. Acids Res. 9, 1885–1904.
[7] De Wachter, R., Chen, M.-W. and Vandenberghe, A. (1982) Biochimie 64, 311–329.
[8] Delihas, N. and Andersen, J. (1982) Nucleic Acids Res. 10, 7323–7344.
[9] Korn, L.J. and Bogenhagen, D.F. (1983) in: Organization and Transcription of the *Xenopus* 5 S Ribosomal RNA Genes (Busch, H. ed) Cell Nucleus vol.12, in press.
[10] Jacq, C., Miller, J.R. and Brownlee, G.G. (1977) Cell 12, 109–120.
[11] Miller, J.R., Cartwright, E.M., Brownlee, G.G., Fedoroff, N.V. and Brown, D.D. (1978) Cell 13, 717–725.
[12] Ford, P.J. and Brown, R.D. (1976) Cell 8, 485–493.
[13] Miller, J.R. and Melton, D.A. (1981) Cell 24, 829–835.
[14] Korn, L.J. and Gurdon, J.B. (1981) Nature 289, 461–465.
[15] Korn, L.J. and Brown, D.D. (1978) Cell 15, 1145–1156.
[16] Wormington, W.M., Bogenhagen, D.F., Jordan, E. and Brown, D.D. (1981) Cell 24, 809–817.
[17] Tinoco, I. jr, Uhlenbeck, O.C. and Levine, M.D. (1971) Nature 230, 362–367; and Tinoco, I. jr (1982); updated parameters (personal communication).
[18] Bogenhagen, D.F. and Brown, D.D. (1981) Cell 24, 261–270.
[19] Yates, F. (1934) in: Contingency Tables Involving Small Numbers and the Chi Square Test, J. Royal Stat. Soc. Suppl. 1, 217–235.
[20] Zimmermann, R.A. (1979) Methods Enzymol. 59, 551–583.