Volume 10 Number 6 1982 Nucleic Acids Research

Nucleotide sequence of an exceptionally long 5.8S ribosomal RNA from Crithidia fasciculata

Murray N.Schnare and Michael W.Gray

Department of Biochemistry, Faculty of Medicine, Dalhousie University, Halifax, Nova Scotia, Canada B3H 4H7

Received 28 October 1981; Revised 1 February 1982; Accepted 12 February 1982

ABSTRACT

In *Crithidia fasciculata*, a trypanosomatid protozoan, the large ribosomal subunit contains five small RNA species (e, f, g, i, j) in addition to 5S rRNA [Gray, M.W. (1981) *Mol. Cell. Biol. 1*, 347-357]. The complete primary sequence of species i is shown here to be pAACGUGUmCGCGAUGGAUGACUUGGCUUCCUAUCUCGUUGA... AGAmACGCAGUAAAGUGCGAUAAGUGGUAΨCAAUUGmCAGAAUCAUUCAAUUACCGAAUCUUUGAACGCAAACGG... CGCAUGGGAGAAGCUCUUUUGAGUCAUCCCCGUGCAUGCCAUAUUCUCCAmGUGUCGAA(C)$_{OH}$. This sequence establishes that species i is a 5.8S rRNA, despite its exceptional length (171-172 nucleotides). The extra nucleotides in *C. fasciculata* 5.8S rRNA are located in a region whose primary sequence and length are highly variable among 5.8S rRNAs, but which is capable of forming a stable hairpin loop structure (the "G+C-rich hairpin"). The sequence of *C. fasciculata* 5.8S rRNA is no more closely related to that of another protozoan, *Acanthamoeba castellanii*, than it is to representative 5.8S rRNA sequences from the other eukaryotic kingdoms, emphasizing the deep phylogenetic divisions that seem to exist within the Kingdom Protista.

INTRODUCTION

In *Crithidia fasciculata*, a trypanosomatid protozoan, the large ribosomal subunit contains five small RNA species besides 5S rRNA [1,2]. One of these, species i, appears to be the analogue of 5.8S rRNA [3,4] in other eukaryotes, but it migrates distinctly more slowly than conventional 5.8S rRNAs during electrophoresis in non-denaturing polyacrylamide gels [1]. We present here the complete primary sequence of species i, which verifies that it is indeed a 5.8S rRNA homologue, although it is exceptionally large. We discuss the location and possible origin of the additional nucleotides in *Crithidia* 5.8S rRNA and comment on the evolutionary position of this molecule.

METHODS

Preparation of Unlabelled *C. fasciculata* 5.8S rRNA

C. fasciculata total cellular RNA was fractionated by selective precipi-

© IRL Press Limited, 1 Falconberg Court, London W1V 5FG, U.K.

0305-1048/82/1006-2085$2.00/0

tation from 3 M NaCl [1]. The NaCl-insoluble RNA was dissolved at 10 mg/ml in the loading buffer of Peattie [5], heated to 50°C for 5 min, and resolved in a 0.3 cm, 10% polyacrylamide slab gel containing a 2.5% stacker gel [6]. Small rRNAs [1] were visualized by ultraviolet shadowing and 5.8S rRNA (species i) was recovered by electrophoretic elution, as described [6].

Preparation and Sequencing of ^{32}P-End-Labelled 5.8S rRNA

End-labelling of 5.8S rRNA was carried out as described by Peattie (3'-; [5]) and Donis-Keller *et al.* (5'-; [7]), respectively [6]. The end-labelled RNA was electrophoresed for 4.75 h at 1500 V in thin (0.05 cm) 6% polyacrylamide gels. Under these conditions, the 5.8S rRNA appeared in autoradiograms as two discrete bands, which were separately recovered. Terminal nucleotide analysis (5'- and 3'-) was performed [6] and sequence analysis was carried out by chemical [5] and enzymatic [7,8] partial degradation methods.

Modified Nucleotide Analysis

C. fasciculata was cultured in the presence of $[^{32}P]P_i$ [9] and uniformly-^{32}P-labelled 5.8S rRNA was prepared from total cellular RNA, as above. The Np + Nm-Np + pNp products of alkaline (10 µl 1 M NaOH, room temperature, 90 h) or RNase T_2 (Sankyo; 0.5 unit in 10 µl 10 mM NH_4COOCH_3 (pH 4.5), 37°C, 16 h) hydrolysis and the pN + pNm products of snake venom phosphodiesterase hydrolysis [6] were subjected, together with unlabelled markers, to thin-layer chromatography (Merck glass-backed cellulose tlc plates containing fluorescent indicator; isobutyric acid/0.5 M NH_4OH, 5/3, in the first dimension, propan-2-ol/conc. HCl/H_2O, 70/15/15, in the second; [10]). In some experiments, an Nm-Np fraction was isolated from alkaline or T_2 hydrolysates by chromatography on small columns of DEAE-cellulose [11,12]. This fraction was subjected to two-dimensional tlc either directly or following dephosphorylation. In the latter case, first dimension solvent was butan-1-ol/isobutyric acid/conc. NH_4OH/H_2O, 75/37.5/2.5/25 [13], 3X, followed in the second dimension by propan-2-ol/conc. HCl/H_2O, 70/15/15, 1X. Individual Nm-N's were eluted from the tlc plate in 200 µl 0.6 M NH_4OH, lyophilized, and hydrolyzed with snake venom phosphodiesterase. The resulting $[^{32}P]$pN's were identified by two-dimensional tlc [6].

RESULTS

End Groups

Hydrolysis of 5'-end-labelled *C. fasciculata* 5.8S rRNA with snake venom phosphodiesterase released most (>95%) of the radioactivity as pA, establish-

ing the 5'-terminal residue as A. Hydrolysis of uniformly-^{32}P-labelled 5.8S rRNA with either alkali or T_2 RNase released pAp (as the sole pNp derivative), showing that the 5'-terminus is phosphorylated *in vivo*.

Alkaline hydrolysis of unresolved 3'-end-labelled 5.8S rRNA liberated both [^{32}P]Cp (63% of the total radioactivity) and [^{32}P]Ap (34%), indicating 3'-terminal heterogeneity. End group analysis of the two species of 5.8S rRNA that could be resolved in 6% polyacrylamide gels showed that the longer species ended with C and the shorter with A.

Modified Nucleoside Constituents

Various analyses indicated the presence of Am, Gm, Um and Ψ, at a level of 2, 1, 0.8, and 1 mol, respectively, per mol of *C. fasciculata* 5.8S rRNA. Ψp, Am-Ap, Am-Gp, Gm-Cp and Um-Cp were identified among the products of either T_2 RNase hydrolysis or alkaline hydrolysis of uniformly-^{32}P-labelled 5.8S rRNA. The Nm-Np's were also recovered as a separate fraction and were identified by co-chromatography with authentic Nm-Np markers or with Nm-N markers following dephosphorylation. Each Nm-N was further characterized by venom phosphodiesterase hydrolysis and identification of the resulting [^{32}P]pN. Venom phosphodiesterase hydrolysis of uniformly-^{32}P-labelled 5.8S rRNA yielded pAm, pUm and pΨ (pGm co-migrates with pC in the tlc system used and so was not detected in this analysis).

Primary Sequence

The sequence of *C. fasciculata* 5.8S rRNA is shown in Fig. 1. Virtually the entire sequence was obtained in each of three different ways: chemical degradation of 3'-end-labelled RNA (Fig. 2 and other data not shown due to space limitations); enzymic degradation of 3'-end-labelled RNA (not shown); and enzymic degradation of 5'-end-labelled RNA (not shown). Very extensive overlapping provided confirmation of the sequence. Chemical sequencing revealed that the two separated 3'-end-labelled species were identical except for an extra 3'-terminal C in the slower-migrating one. Residues 70, 82, 86

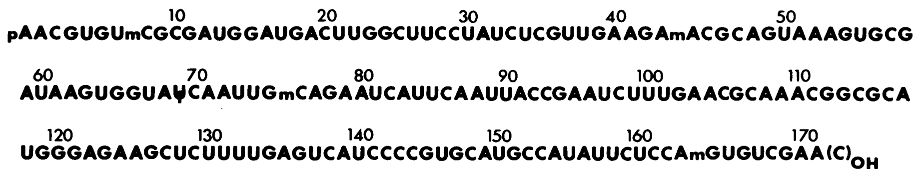

FIGURE 1. Primary sequence of *C. fasciculata* 5.8S rRNA. The bracketed C residue at the 3'-end denotes heterogeneity at this terminus (see text).

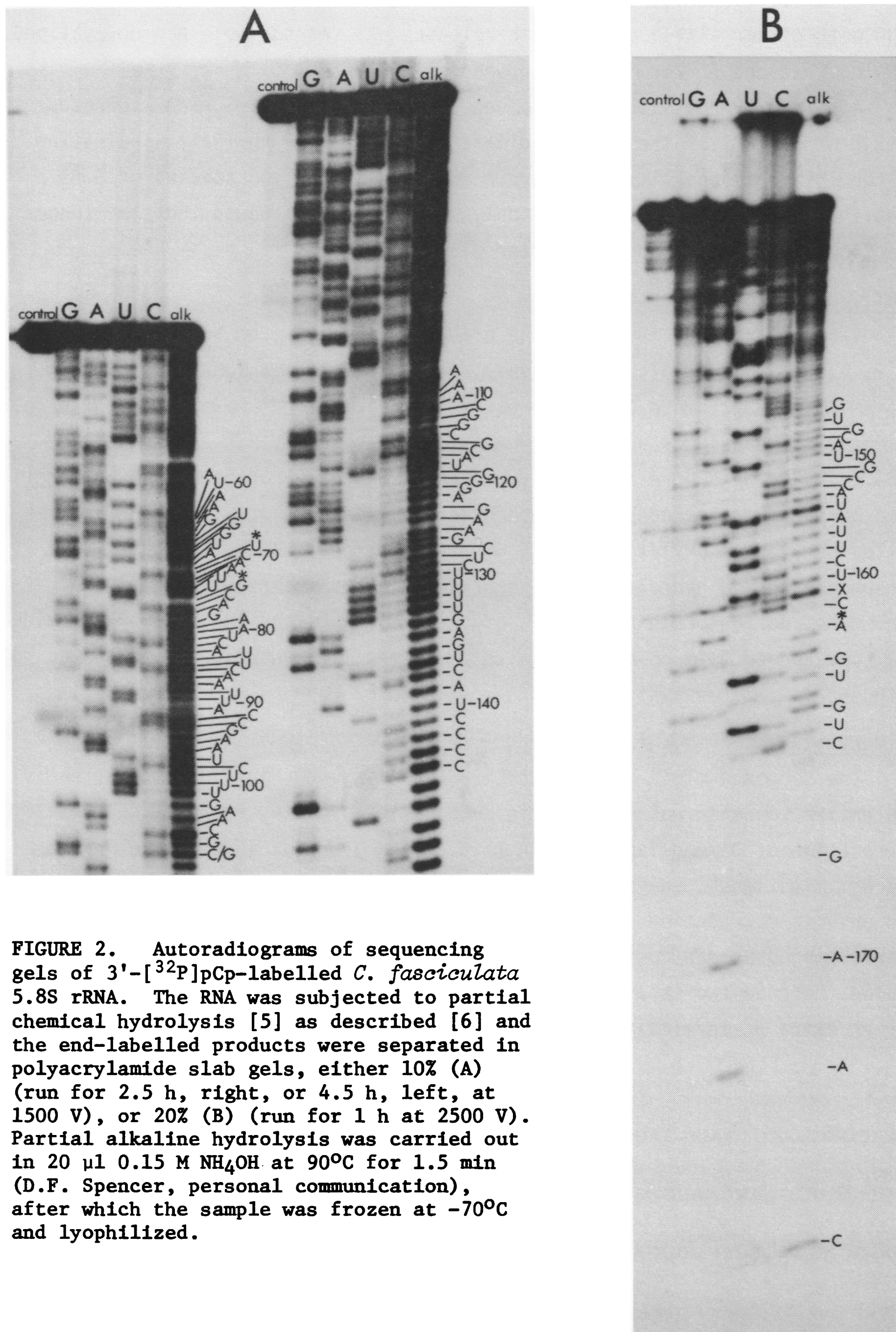

FIGURE 2. Autoradiograms of sequencing gels of 3'-[^{32}P]pCp-labelled *C. fasciculata* 5.8S rRNA. The RNA was subjected to partial chemical hydrolysis [5] as described [6] and the end-labelled products were separated in polyacrylamide slab gels, either 10% (A) (run for 2.5 h, right, or 4.5 h, left, at 1500 V), or 20% (B) (run for 1 h at 2500 V). Partial alkaline hydrolysis was carried out in 20 µl 0.15 M NH_4OH at 90°C for 1.5 min (D.F. Spencer, personal communication), after which the sample was frozen at -70°C and lyophilized.

and 107 appeared reproducibly as both C and G in chemical sequencing gels (Fig. 2); however, susceptibility to cleavage by pancreatic RNase indicated they are all C's.

Anomalies in the chemical and alkaline ladders served to localize the modified nucleoside constituents identified in hydrolysates of uniformly-^{32}P-labelled 5.8S rRNA. Residue 69 appeared as a U in enzyme gels but gave a blank in chemical sequencing gels; also, the alkali band at this position was much fainter than normal (Fig. 2). These are all features diagnostic of Ψ residues [5,6], and we therefore place the single Ψ residue in *C. fasciculata* 5.8S rRNA at position 69. Residues 43, 75 and 163 had no corresponding bands in the alkali ladder (Fig. 2), indicating the presence of alkali-stable (presumably $O^{2'}$-methylnucleoside) residues at these positions. Considering the residues 3' to these positions and the results of Nm-Np analysis, we place Am at positions 43 (Am-Ap) and 163 (Am-Gp) and Gm at position 75 (Gm-Cp). Residue U7 (confirmed as such by chemical sequencing gels) gave a much weaker band in the alkali ladder than any of the neighboring residues. Since phosphodiesterase hydrolysis of uniformly-^{32}P-labelled 5.8S rRNA gave ∿0.8 mol pUm/mol 5.8S rRNA, and RNase T_2 digestion yielded ∿0.8 mol Um-Cp/mol, we infer the presence of Um at position 7 and conclude that *C. fasciculata* 5.8S rRNA is incompletely (80%) $O^{2'}$-methylated at this position.

DISCUSSION

Alignment with published 5.8S rRNA sequences (Fig. 3) clearly shows that *C. fasciculata* species i is a 5.8S rRNA, despite its exceptional length (171-172 nucleotides). Particularly notable regions of homology are two moderately long stretches (residues 38-46 and 102-108) that appear to be conserved in all 5.8S rRNAs; the former contains the G-A-A-C tetranucleotide that has been proposed to interact with the common G-T-Ψ-C sequence in tRNA [19,20]. The conserved G-G-A-U sequence that may be involved in the interaction of 5.8S rRNA with the large subunit (25S-28S) rRNA [21] is also present in *Crithidia* 5.8S rRNA (positions 14-17).

Quantitative evaluation of sequence identity (excluding the G+C-rich hairpin) indicates that *C. fasciculata* 5.8S rRNA is no more closely related to the 5.8S rRNA of another protozoan, *Acanthamoeba castellanii* (87 identities) than it is to representative 5.8S sequences from the other three eukaryotic kingdoms (82-94 identities). Thus, comparisons of 5.8S as well as 5S [22] sequences suggest that intrakingdom diversity among the protists is at least as great as interkingdom diversity among eukaryotes in general, rein-

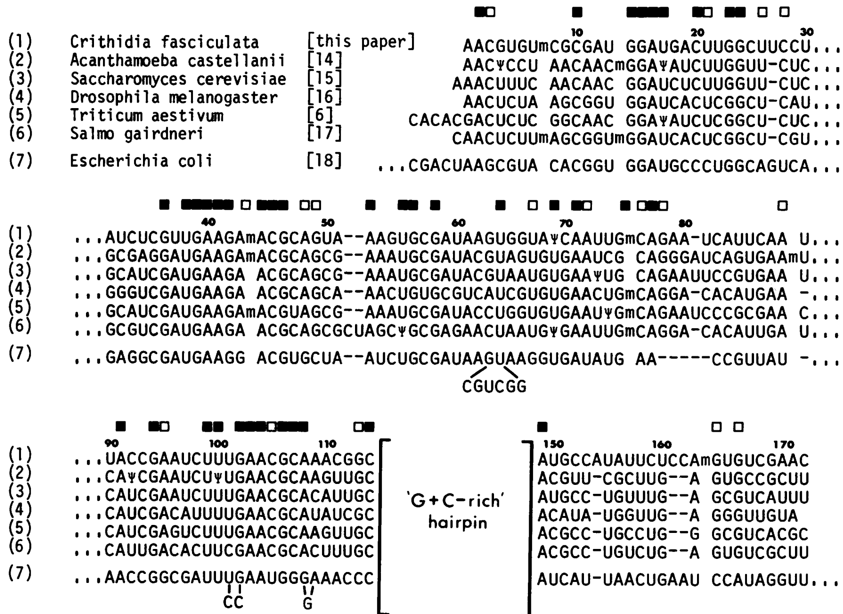

FIGURE 3. Alignment of 5.8S rRNA sequences, arranged for maximal homology in pairwise comparisons with a minimum number of assumed additions and deletions. Also included is the 5'-terminal sequence of *E. coli* 23S rRNA (residues 8-171). The symbols above each position denote residues totally conserved in the six 5.8S sequences <u>and</u> in the *E. coli* 23S sequence (■) and residues conserved in the six 5.8S sequences but <u>not</u> in the *E. coli* sequence (□). The highly variable "G+C-rich hairpin" (see Fig. 4) is not included in this alignment.

forcing the view [23] that the Kingdom Protista is evolutionarily the most ancient and phylogenetically the most diverse of the eukaryotic kingdoms.

Fig. 4 illustrates that neither length nor primary sequence is very highly conserved in the G+C-rich hairpin region of diverse 5.8S rRNAs, although the potential for a substantial degree of base pairing is preserved. The extra nucleotides in *Crithidia* 5.8S rRNA are located in this region: its length is 34 nucleotides in *Crithidia* but only 22-27 nucleotides in other eukaryotes. It is noteworthy that in *D. melanogaster* [16] and *S. coprophila* [24], the 5.8S rRNA consists of two fragments (5.8Sa and 5.8Sb) that are joined non-covalently at the G+C-rich hairpin. This arrangement results from the excision (without subsequent splicing) of a short transcribed spacer that separates the 5.8Sa and 5.8Sb regions in the primary 5.8S transcript. Since the extra nucleotides in *Crithidia* 5.8S rRNA occur in the same region as the

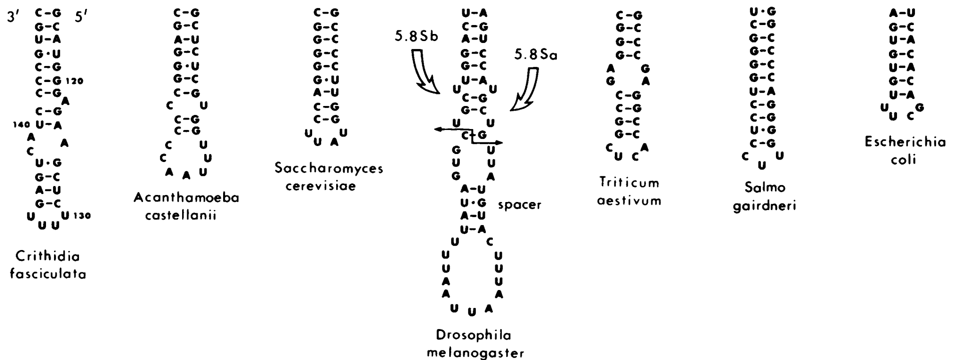

FIGURE 4. Potential secondary structure in the "G+C-rich hairpin" region of the 5.8S rRNA sequences listed in Fig. 3, and in the analogous region of *E. coli* 23S rRNA. The arrows in the *Drosophila* 5.8S sequence indicate the cleavages that occur during 5.8S rRNA maturation and which result in removal of the transcribed spacer region [16]; 5.8Sa and 5.8Sb refer to the 5'- and 3'-terminal pieces, respectively, which remain joined together by base pairing in the G+C-rich hairpin region. In the case of wheat, initial ambiguities in the sequence of this region [6] have recently been resolved [25], and the revised version of the G+C-rich hairpin is shown above.

Drosophila and *Sciara* spacers, it is possible they represent vestiges of a transcribed but unexcised spacer. However, the putative *Crithidia* spacer is not obviously homologous with either the *Drosophila* or *Sciara* spacers, being neither as long nor as A+U-rich.

Our alignment of 5.8S rRNAs (Fig. 3) also includes the 5'-end of *E. coli* 23S rRNA, which Nazar [26] has recently postulated to be the structural analogue in prokaryotes of 5.8S rRNA. The alignment confirms that there is extensive homology between the 5'-end of *E. coli* 23S rRNA and a range of eukaryotic 5.8S rRNAs, although quantitatively the degree of homology (64-70 identities in pairwise comparisons) is significantly less than between any two 5.8S rRNAs (82-111 identities). Interestingly, a stable hairpin loop can be formed with that part of the *E. coli* 23S rRNA sequence that is analogous to the G+C-rich hairpin region of 5.8S rRNA (Fig. 4). These and additional observations made by Walker [27] fully support the suggestion made by Nazar [26] that in prokaryotic ribosomes, a 5.8S-like structure is present at the 5'-end of the large subunit (23S) rRNA.

ACKNOWLEDGMENTS

We thank A.G. Wildeman and R.N. Nazar for permission to cite sequence

data prior to publication. We are grateful to R.M. MacKay for a gift of unlabelled *C. fasciculata* 5.8S rRNA. This work was supported by a grant from the Medical Research Council of Canada to M.W.G.

REFERENCES

[1] Gray, M.W. (1979) *Can. J. Biochem. 57*, 914-926.
[2] Gray, M.W. (1981) *Mol. Cell. Biol. 1*, 347-357.
[3] Erdmann, V.A. (1976) *Progr. Nucleic Acids Res. Mol. Biol. 18*, 45-90.
[4] Erdmann, V.A. (1981) *Nucleic Acids Res. 9*, r25-r42.
[5] Peattie, D.A. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 1760-1764.
[6] MacKay, R.M., Spencer, D.F., Doolittle, W.F. and Gray, M.W. (1980) *Eur. J. Biochem. 112*, 561-576.
[7] Donis-Keller, H., Maxam, A.M. and Gilbert, W. (1977) *Nucleic Acids Res. 4*, 2527-2538.
[8] Simoncsits, A., Brownlee, G.G., Brown, R.S., Rubin, J.R. and Guilley, H. (1977) *Nature 269*, 833-836.
[9] MacKay, R.M., Gray, M.W. and Doolittle, W.F. (1980) *Nucleic Acids Res. 8*, 4911-4917.
[10] Nishimura, S., Harada, F., Narushima, U. and Seno, T. (1967) *Biochim. Biophys. Acta 142*, 133-148.
[11] Singh, H. and Lane, B.G. (1964) *Can. J. Biochem. 42*, 1011-1021.
[12] Lane, B.G. (1965) *Biochemistry 4*, 212-219.
[13] Rogg, H., Brambilla, R., Keith, G. and Staehelin, M. (1976) *Nucleic Acids Res. 3*, 285-295.
[14] MacKay, R.M. and Doolittle, W.F. (1981) *Nucleic Acids Res. 9*, 3321-3334.
[15] Rubin, G.M. (1973) *J. Mol. Biol. 248*, 3860-3875.
[16] Pavlakis, G.N., Jordan, B.R., Wurst, R.M. and Vournakis, J.N. (1979) *Nucleic Acids Res. 7*, 2213-2238.
[17] Nazar, R.N. and Roy, K.L. (1978) *J. Biol. Chem. 253*, 395-399.
[18] Branlant, C., Krol, A., Machatt, M.A. and Ebel, J.P. (1979) *FEBS Lett. 107*, 177-181.
[19] Nishikawa, K. and Takemura, S. (1974) *FEBS Lett. 40*, 106-109.
[20] Wrede, P. and Erdmann, V.A. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 2706-2709.
[21] Nazar, R.N. and Sitz, T.O. (1980) *FEBS Lett. 115*, 71-76.
[22] MacKay, R.M., Spencer, D.F., Schnare, M.N., Doolittle, W.F. and Gray, M.W. (1982) *Can. J. Biochem.*, in press
[23] Whittaker, R.H. (1969) *Science 163*, 150-160.
[24] Jordan, B.R., Latil-Damotte, M. and Jourdan, R. (1980) *Nucleic Acids Res. 9*, 3565-3573.
[25] Wildeman, A.G. and Nazar, R.N. (1981) *Eur. J. Biochem.*, in press.
[26] Nazar, R.N. (1980) *FEBS Lett. 119*, 212-214.
[27] Walker, W.F. (1981) *FEBS Lett. 126*, 150-151.