Organization of ribosomal DNA in yellow lupine (*Lupinus luteus*) and sequence of the 5.8 S RNA gene

Jan Antoni Rafalski*†, Maciej Wiewiórowski* and Dieter Söll††

*Institute of Bioorganic Chemistry, Polish Academy of Sciences, PL-61704 Poznan, Noskowskiego 12, Poland, and †Department of Molecular Biophysics and Biochemistry, Yale University, PO Box 6666, New Haven, CT 06511, USA

Received 4 October 1982; revision received 30 December 1982

The large ribosomal repeat of the *Lupinus luteus* genome is 8.8 kb in length as revealed by Southern blotting and hybridization to 5.8 S rRNA and cloned rDNA fragments of *L. luteus* and *Schizosaccharomyces pombe*. It codes for 18 S, 5.8 S and 28 S rRNAs. A cloned 3.7 kb *Eco*R1 subfragment of the rDNA repeat, coding for 5.8 S rRNA and a large portion of the 28 S rRNA, has been sequenced in the region of 5.8 S rRNA gene. The predicted rRNA sequence is homologous to broad bean 5.8 S rRNA and can be arranged in the generalized model of 5.8 S RNA secondary structure.

Lupinus luteus *RNA gene* *5.8 S RNA Plant gene* *Ribosomal repeat DNA sequence* *Ribosomal DNA*

1. INTRODUCTION

The nuclear genes coding for 16 S, 5.8 S and 28 S rRNA of higher eukaryotes are arranged as tandem repeats, each repeat coding for all 3 rRNAs. The 3 genes are transcribed as a single ribosomal precursor which is subsequently processed in several steps. The ribosomal repeat varies in length between 8000 and 44000 bp and consists of both the coding sequences and non-transcribed spacer. The nucleotide sequences of rRNA genes from yeast have been determined [1,2]. In higher plants, the organization of the large ribosomal repeat has been studied in detail in wheat, where the major repeat is 8.8 kb in length, and barley where two large rDNA repeats are found (9 kb and 9.9 kb) [3]. Repeat lengths and restriction maps have also been reported for soybean [4,5], radish [6], rye [7], rice [8], broad bean [9] and carrot [10].

† To whom correspondence should be addressed

Abbreviations: kb, kilobases; rRNA, ribosomal RNA; 1 × SSC buffer, 0.15 M NaCl, 0.015 M sodium citrate (pH 7.0)

Wheat [11] and broad bean [12] 5.8 S rRNA sequences are known. The broad bean sequence does not, however, fit the generalized 5.8 S rRNA secondary structure model [13,14]. Further rRNA sequences from higher plants are required to study homology and phylogenic relationships [15].

Yellow lupine, a legume of agricultural importance [16], has been a subject of numerous biochemical studies [17–21]. The mitochondrial ribosomal RNA genes of this organism have been mapped [22], and the organization and nucleotide sequences of *Lupinus* 5 S RNA genes are known [23]. Here we describe the organization of the large ribosomal repeat and its DNA sequence in the 5.8 S RNA gene region.

2. MATERIALS AND METHODS

2.1. *Preparation of DNA hybridization probes*

DNA restriction fragments were labelled by either *E. coli* DNA polymerase I Klenow fragment-catalyzed repair of staggered ends [24], or using T4 DNA polymerase [25]. Appropriate DNA fragments were purified by low gelling temperature agarose gel electrophoresis [26]. After

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

autoradiography gel pieces were cut out, DNA denatured by boiling 10 min in 0.1 × SSC and used for hybridization directly, without elution.

2.2. *DNA–DNA hybridization conditions*

DNA probe was hybridized to DNA immobilized on nitrocellulose filters in 2 × SSC, 0.02% polyvinylpyrrolidone 360000, 0.02% Ficoll 400000, 0.02% bovine serum albumin, 0.5% SDS, 0.1% sodium pyrophosphate (pH 7) at 65°C for 24–48 h. Filters were washed several times in a large volume of 0.5 × SSC at 60°C. Other materials and methods were described in [23].

3. RESULTS

3.1. *Construction and analysis of recombinant phages and plasmids*

L. luteus genomic DNA was digested with *Eco*R1, separated on a 0.7% agarose gel, transferred to nitrocellulose filter and hybridized with 5′-^{32}P-labelled 5.8 S rRNA from *L. luteus* (fig.1) or *S. pombe* (not shown). Only one band of about 3.7 kb appears on autoradiograms (fig.1A,E). Extremely overexposed autoradiograms show several additional very weak bands. *Eco*R1-restricted *L. luteus* DNA was therefore used to construct a genomic library in phage λ607 [27]. The library was screened [28] with 5′-^{32}P-labelled *S. pombe* 5.8 S rRNA.

Two positive clones were selected for further analysis. One of them (AR12) contained an insert of the major 3.7 kb size class, another (AR6) a fragment of DNA of about 4.7 kb that gave a weaker hybridization signal. For further analysis the DNA fragments were recloned into pBR325 (plasmids pAR6 and pAR12). Computer analysis [29] of broad bean 5.8 S rRNA sequence [12], presumed to be similar to *L. luteus* sequence, revealed a *Cla*1 restriction site (nucleotides 37–42 of the RNA sequence) and an *Eco*RV site (nucleotides 20–25 of the RNA sequence). The 3.7 kb fragment of clone pAR12 was indeed cut with both *Cla*1 and *Eco*RV, which allowed to orient and position the 5.8 S gene within the fragment. The 4.7 kb fragment of clone pAR6 was cut only with *Cla*1. More detailed restriction maps were obtained using both unlabelled and terminally-labelled fragments. Southern blots, probed with various labelled restriction fragments

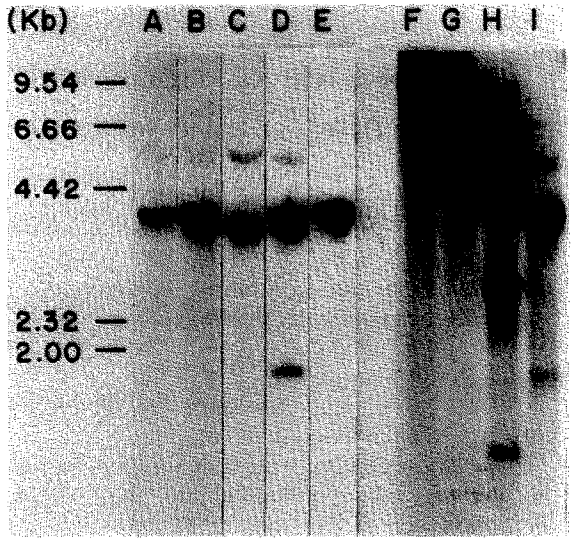

Fig.1. (A–E) Genomic blots of *L. luteus* DNA, digested with *Eco*R1 and probed with: (A) *L. luteus* 5.8 S r [^{32}P] RNA; (B) ^{32}P-labelled 2.9 kb *Cla*1–*Eco*R1 fragment of plasmid pYM 32 [15] containing *S. pombe* 25 S rRNA gene with exception of its 3′-end; (C) ^{32}P-labelled 6.1 kb *Eco*R1–*Eco*R1 fragment of plasmid pYM 32 [15] containing *S. pombe* 17 S rRNA gene with exception of its 3′-end; (D) ^{32}P-labelled plasmid pYM 32 containing complete *S. pombe* ribosomal repeat [15]; (E) *L. luteus* 5.8 S r [^{32}P] RNA – an overexposed autoradiogram to show absence of additional bands; (F–I) Genomic blots of *L. luteus* DNA probed with ^{32}P-labelled plasmid pAR12. Restriction enzymes used were: (F) *Bgl*2; (G) *Bcl*1; (H) *Bam*H1; (I) *Eco*R1.

of plasmid pYM32 [15] allowed to position the 18 S and 28 S genes on the restriction map of clones pAR6 and pAR12.

3.2. *Mapping of ribosomal repeat*

To provide a restriction map of the whole ribosomal repeat, genomic *L. luteus* DNA was digested with several restriction nucleases, separated on agarose gels, blotted and probed with labelled fragments of pYM32, pAR12 and pAR6 DNAs. Endonucleases *Cla*1 and *Hin*d3 do not cut *L. luteus* genomic DNA within the ribosomal repeat. Endonucleases *Bcl*1 and *Bgl*2 cut once within the ribosomal repeat defining its size as 8.8 kb (fig.1F,G), although some rDNA remained uncut by *Bgl*2 (not shown). Digests with these enzymes in combination with *Eco*R1 and *Bam*H1 allowed to construct the map of the complete

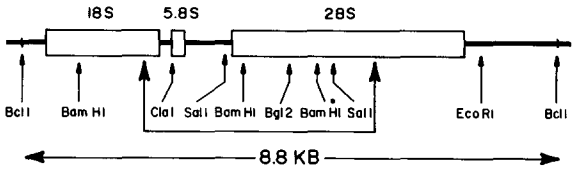

Fig.2. Restriction map of *L. luteus* large ribosomal repeat. Regions coding for 18 S, 5.8 S and 28 S rRNA are indicated based on hybridization and sequence (in the case of 5.8 S) data. The *Bam*H1 site indicated with * is absent in some repeats. Two arrows connected by a bracket indicate the *Eco*R1–*Eco*R1 fragment present in pAR12.

repeat (fig.2). *Eco*RI cuts 3-times within the ribosomal repeat, giving fragments of about 3.7 kb, 3.4 kb, 1.7 kb (fig.1D). The 3.7 kb fragment is the one cloned in pAR12 (fig.1I). It contains 5.8 S rRNA gene (fig.1A,E) and part of the 28 S rRNA gene (fig.1B). The 3.4 kb fragment contains a large part of the 18 S rRNA gene (fig.1C). The 1.7 kb fragment contains 3′-part of the 28 S rRNA gene (fig.1D and fig.2). *Bam*H1 also cuts 3-times within the ribosomal repeat, giving fragments of about 5 kb, 2.6 kb, 1.2 kb (fig.1H). One of the *Bam*H1 sites was resistant to cleavage in about 60–70% of the ribosomal repeats, as evidenced by the appearance of a band at about 6.2 kb (fig.1H and fig.2). A small proportion of other endonuclease sites was resistant to enzyme cutting, as evidenced by the presence of some partial digestion products even at saturating enzyme concentrations (fig.1).

3.3. *Sequence determination of the 5.8 S rRNA gene*

Recombinant plasmids were partially sequenced from their *Cla*1 sites within the 5.8 S rRNA gene, using both 5′-end and 3′-end labelling of DNA fragments to obtain overlaps. Fig.3 shows the resulting sequence of pAR12. The ends of the *L. luteus* 5.8 S rRNA molecule were determined, to enable exact positioning of the coding sequences within the DNA stretch required. The 5′-end of the 5.8 S rRNA was heterogenous: 85% pC (corresponding to the sequence of the clone pAR12), 15% A (corresponding to the sequence of the clone pAR6). The 3′-end of the rRNA was determined to be CAC_{OH}. Most 5.8 S rRNA molecules terminated with C_{164} but some were one or two nucleotides shorter. It is not clear whether this is due to nuclease action in the organism. Clone pAR6 did not contain the complete 5.8 S gene sequence. It terminated with nucleotide 82 of the

```
        10         20         30         40         50         60         70
GTCCCACCCTCTGCGGTGTCCTCCTGGCCTAATAACAAAACCCCGCGCCGAACGCGCCAAGGAAATTGAA
CAGGGTGGGAGACGCCACAGGAGGACCGGATTATTGTTTTGGGGCGCGGCTTGCGCGGTTCCTTTAACTT

        80         90        100        110        120        130        140
ATCGTTTAGTTCGCCCCCGCCGGCCCGGAGACGGTGCTCGTGCGGGCGGCGTTGCGACACGCTTATCCTA
TAGCAAATCAAGCGGGGGCGGCCGGGCCTCTGCCACGAGCACGCCCGCCGCAACGCTGTGCGAATAGGAT

       150        160        170        180        190        200        210
AAGACTCTCGGCAACGGATATCTCGGCTCTTGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGT
TTCTGAGAGCCGTTGCCTATAGAGCCGAGAACGTAGCTACTTCTTGCATCGCTTTACGCTATGAACCACA

       220        230        240        250        260        270        280
GAATTGCAGAATCCCGTGAACCATCGAGTCTTTGAACGCAAGTTGCGCCCGAAGCCATTAGGCCGAGGGC
CTTAACGTCTTAGGGCACTTGGTAGCTCAGAAACTTGCGTTCAACGCGGGCTTCGGTAATCCGGCTCCCG

       290        300        310        320        330        340
ACGCCTGCCTGGGTGTTGCACATCGTTGCCCCCGTGCCTTGGCCACGTGCAGGCACGAAACGGGGC-3'
TGCGGACGGACCCACAACGTGTAGCAACGGGGGCACGGAACCGGTGCACGTCCGTGCTTTGCCCCG-5'
```

Fig.3. Nucleotide sequence of clone pAR12 in the region of 5.8 S rRNA gene. The 5.8 S rRNA coding sequence is boxed. Direct repeats are indicated with arrows. The *Cla*1 restriction site used for sequencing is underlined.

RNA sequence, with the *Eco*R1 site (which does not occur in pAR12 due to a base change).

4. DISCUSSION

4.1. *rDNA organization*

In higher plants, the length of the large ribosomal repeat coding for 18 S, 28 S and 5.8 S rRNA varies between about 7.8 and 11.5 kb [3–10]. In many instances size heterogeneity was observed [3]. In *L. luteus*, unlike in several other species, *Eco*R1 cuts more than once within the ribosomal repeat. *Cla*1 and *Hin*d3 do not cut *L. luteus* genomic DNA; this is presumably due to methylation, since these restriction sites are present in cloned fragments. Endonucleases *Bcl*1 and *Bgl*2, the recognition sequences of which do not include CG base sequence, which is highly methylated in plants [3], define the length of repeat as 8.8 kb. No repeat length heterogeneity was observed. A similar situation was found in soybean [4,5]. The restriction map of the whole repeat, deduced from Southern hybridization data using *Schizosaccharomyces pombe* clone pYM32, and *L. luteus* clones pAR12 and pAR6, is similar to the published restriction maps of other eukaryotes (e.g. [3]). One of two *Bam*H1 sites within the 28 S rRNA coding region appears only in about 30–40% of the ribosomal repeats. Similar observation was made in carrot [10]. Although clone pAR12 has both *Bam*H1 sites in the 28 S rDNA region, we cannot differentiate between two possible explanations for this effect: methylation and sequence heterogeneity. It is pertinent to note that we have observed extensive methylation of the *L. luteus* 5 S gene DNA [23]. Cross-hybridization between *L. luteus* and *S. pombe* ribosomal DNA containing 28 S and 18 S RNA genes indicates significant sequence homology. Similar cross-hybridization was observed between *S. pombe* 5.8 S rRNA and *L. luteus* 5.8 S DNA.

4.2. *5.8 S rDNA sequence*

We have sequenced completely the 5.8 S rRNA coding region of the *L. luteus* rDNA repeat from clone pAR12 (fig.3). It is the first known 5.8 S gene sequence from higher plants. As observed before in other systems, the GC content of the coding sequence (53.6%) is lower than that of the spacer (64.8%). The coding sequence is very similar to broad bean 5.8 S rRNA – the difference being 8 base changes and 3 base deletions (6.7% total difference).

The 3′-end of the 5.8 S rRNA coding sequence is preceded and followed by two pairs of short direct repeats; GTTGC (proximal to the 3′-end), TGCCT (distal to the 3′-end of 5.8 S rDNA). These sequences might constitute signals for processing the RNA transcript. Similar repeats have been found in other 5.8 S rRNA genes [30]. Compared to *L. luteus*, broad bean 5.8 S rRNA has a deletion of two nucleotides within the sequence corresponding to proximal repeat (GTTGC ⟶ GTC) [12]. It would be interesting to ascertain whether the extragenic part of the repeat also has an appropriate sequence change. No similar repeats were found at the 5′-end of the 5.8 S gene.

The sequence of the 5.8 S gene provided by clone pAR6 is incomplete, as the clone extends only to the *Eco*R1 site at position 82 of the RNA. The available sequence is in 18.4% different from the clone pAR12 sequence and only 8% different from *S. pombe* 5.8 S rRNA gene [15]. We cannot exclude the possibility that clone (AR6) DNA is fungal in origin, although care was taken to exclude such contamination during seed germination. The genomic blots probed with labelled clone pAR6 show that the abundance of the 4.7 kb *Eco*R1 fragment that gave rise to that clone in *L. luteus* DNA is extremely low.

4.3. *5.8 S rRNA secondary structure*

It has been recently pointed out [15] that broad bean 5.8 S rRNA sequence does not conform to the secondary structure model of Nazar [13] or Pace [14] and that more plant sequences have to be known, before this question can be studied in detail. We have examined the *L. luteus* 5.8 S rRNA sequence as deduced from the DNA sequence and found that it could be arranged in a structure similar to the Pace model ([14]; see also [15] and fig.4). The continuity of the helix *e* is, however, disrupted; 7 out of 8 base pairs remain intact, but with an unpaired stretch of 3 nucleotides in the middle. Loop VI consists of 4 nucleotides. Unpaired 5′-terminus is 31 nucleotides long and the unpaired 3′-terminus 21 nucleotides long. Most of the changes, in relation to *S. pombe* 5.8 S RNA occur in the helix *e* and loop VI. This may suggest a somewhat different

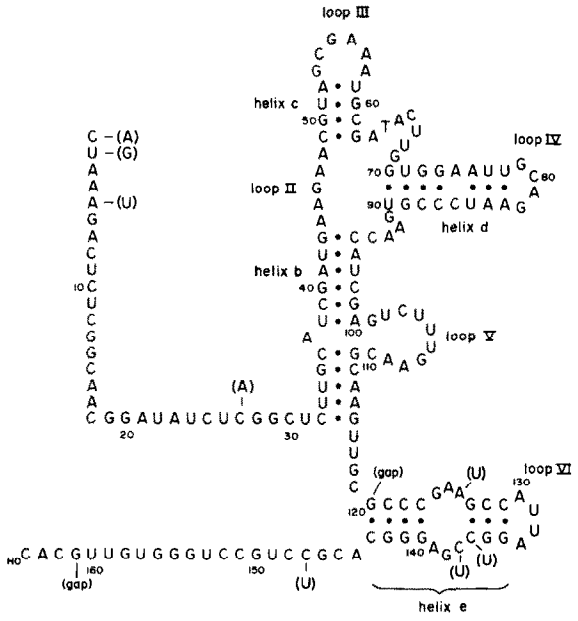

Fig.4. *L. luteus* 5.8 S rRNA as deduced from the DNA sequence and RNA sequencing data on the 5′- and 3′-terminus, arranged in the standard secondary structure model [14,15]. Differences with broad bean RNA sequence [12] are indicated in parentheses.

arrangement of the 5.8 S–28 S complex in plants, since helix *e* is proposed to play an important role in its stabilization [14]. In broad bean, two of the G–C base pairs proposed here for the *e* helix are not possible, due to a base change and a base deletion. Clearly, more 5.8 S rRNA and rDNA sequences are needed before any generalizations can be made.

ACKNOWLEDGEMENTS

We thank Mr J. Schaack for providing *S. pombe* 5.8 S rRNA and plasmid pYM32 and Ms B. Bekasiak for supplying crude *L. luteus* RNA. This work was supported by the joint US-Polish project 7F7F072P and by a grant from the National Institutes of Health.

REFERENCES

[1] Veldman, G.M., Klootwijk, J., de Regt, V.C.H.F., Planta, R.J., Branlant, Ch., Krol, A. and Ebel, J.P. (1981) Nucl. Acids Res. 9, 6935–6952 and references cited therein.

[2] Georgiev, O.I., Nikolaev, N., Hadjiolov, A.A., Skryabin, K.G., Zakharyev, V.M. and Bayev, A.A. (1981) Nucl. Acids Res. 9, 6953–6958 and references cited therein.

[3] Gerlach, W.L. and Bedbrook, J.R. (1979) Nucl. Acids Res. 7, 1869–1885.

[4] Varsanyi-Breiner, A., Gusella, J.F., Keys, C., Housman, D.E., Sullivan, D., Brisson, N. and Verma, D.P.S. (1979) Gene 7, 317–334.

[5] Friedrich, H., Hemleben, V., Meagher, R.B. and Key, J.L. (1979) Planta 146, 467–473.

[6] Delseny, M., Aspart, L., Cooke, R., Grellet, F. and Penon, P. (1979) Biochem. Biophys. Res. Commun. 91, 540–547.

[7] Appels, R., Gerlach, W.L., Dennis, E.S., Swift, H. and Peacock, W.J. (1980) Chromosoma (Berlin) 78, 293–311.

[8] Oono, K. and Sugiura, M. (1980) Chromosoma (Berlin) 76, 85–89.

[9] Yakura, K. and Tanifuji, S. (1981) Plant and Cell Physiol. 22, 1105–1111.

[10] Kato, A., Yakura, K. and Tanifuji, S. (1982) Plant and Cell Physiol. 23, 151–154.

[11] Nazar, R.N. and Matheson, A.T. (1978) J. Biol. Chem. 253, 5464–5469.

[12] Tanaka, Y., Dyer, T.A. and Brownlee, G.G. (1980) Nucl. Acid Res. 8, 1259–1272.

[13] Nazar, R.N., Sitz, T.O. and Busch, H. (1975) J. Biol. Chem. 250, 8591–8597.

[14] Pace, N.R., Walker, T.A. and Schroeder, E. (1977) Biochemistry 16, 5321–5328.

[15] Schaack, J., Mao, J. and Söll, D. (1982) Nucl. Acid Res. 10, 2851–2864.

[16] Gill, N.T. and Vear, K.C. (1966) Agricultural Botany, Gerald Duckworth, London.

[17] Jakubowski, H. and Pawełkiewicz, J. (1975) Eur. J. Biochem. 52, 301–310.

[18] Jakubowski, H. (1980) Biochemistry 19, 5071–5078.

[19] Joachimiak, A., Barciszewski, J., Twardowski, T., Barciszewska, M. and Wiewiórowski, M. (1978) FEBS Lett. 93, 51–54.

[20] Rafalski, A.J., Barciszewski, J., Gulewicz, K., Twardowski, T. and Keith, G. (1977) Acta Biochim. Polon. 24, 301–318.

[21] Rognes, S.E. (1980) Phytochemistry 19, 2287–2294.

[22] Bartnik, E., Augustyniak, H., Borsuk, P. and Stępień, P.P. (1981) Communication at the 'Third Meeting on the Control of Gene Expression', Reinhardsbrunn.

[23] Rafalski, J.A., Wiewiórowski, M. and Söll, D. (1983) Nucl. Acid Res. 10, in press.

[24] Goodman, H.M. (1980) Methods Enzymol. 65, 63–64.

[25] Challberg, M.C. and Englund, P.T. (1980) Methods Enzymol. 65, 39–43.
[26] Langridge, J., Langridge, P. and Berquist, P.L. (1980) Anal. Biochem. 103, 264–271.
[27] Murray, N.E., Brammar, W.J. and Murray, K. (1977) Molec. Gen. Genet. 150, 53–61.
[28] Benton, W.D. and Davis, R.W. (1977) Science 196, 180–182.
[29] Sege, R.D., Söll, D., Ruddle, F.H. and Queen, C. (1981) Nucl. Acid Res. 9, 437–444.
[30] Goldman, W.E., Bowman, L.H., Steinmetz, D., Goldberg, G.I. and Schlessinger, D. (1983) (Comm. from L.H. Bowman) Cell, in press.