The EMBO Journal vol.4 no.10 pp.2431–2438, 1985

Promoter analysis of a soybean nuclear gene coding for nodulin-23, a nodule-specific polypeptide involved in symbiosis with *Rhizobium*

S.L.Wong and D.P.S.Verma

Centre for Plant Molecular Biology, Department of Biology, McGill University, 1205 Docteur Penfield Avenue, Montreal H3A 1B1, Canada

Communicated by D.C.Burke

Soybean nodulin-23 gene, induced in nodules formed due to symbiosis with *Rhizobium*, was found to contain multiple sequences capable of acting as eucaryotic and procaryotic promoters. The transcription start sites of this gene were localized by S1 nuclease mapping, primer extension with nodule mRNA and *in vitro* run-off transcription analysis. A major transcription start site was observed by S1 mapping; however, the primer extension revealed a second start site. Sequence analysis showed the presence of tandemly arranged eucaryotic promoter sequences, each corresponding to one of the observed transcription start sites. *In vitro* transcription analysis with HeLa cell extract demonstrated the presence of another α-amanitin-sensitive transcript ('anti-sense') with its direction opposite to that of the nodulin mRNA. The 5′ ends for 80 nucleotides of nodulin and 'anti-sense' messages are complementary to each other. In addition, three tandemly arranged, functional procaryotic promoters (assayed in *Escherichia coli*) are found to be located within the eucaryotic promoter regions of this gene. One of the procaryotic transcription start sites is similar to that of the major *in vivo* transcript in nodules. The potential significance for the presence of several promoter-like sequences in this gene is discussed.

Key words: nodulin-23 gene/*in vitro* transcription/*Glycine max*/symbiotic nitrogen fixation/eucaryotic and procaryotic promoters

Introduction

The symbiotic association of *Rhizobium* spp. with leguminous plants results in the formation of a highly specialized organ, the root nodule, which provides the appropriate biochemical environment for the reduction of dinitrogen. The formation and maintenance of the functional nodules depend on a coordinated expression of both plant and bacterial genes (for reviews, see Beringer *et al.*, 1980; Verma and Long, 1983). Both genetic and biochemical evidence (Caldwell and Vest, 1977; Verma and Nadler, 1984) indicates that the host plant plays an important role in the nodulation process. In several legumes (soybean, pea and alfalfa), a group of nodule-specific plant proteins, nodulins, are expressed (Legocki and Verma, 1980; Bisseling *et al.*, 1983; Lang-Unnasch and Ausubel, 1985). However, the exact role of nodulins, except that of nodulin-35 which is involved in ureide biosynthesis (Bergmann *et al.*, 1983), is unknown. Analyses of nodulin genes may allow us not only to understand the functional contribution of plant genes in symbiotic nitrogen fixation, but also the regulation of induction of these genes. Towards this end, we determined the fine structure of the 5′ region of a nodulin gene of soybean.

We report a complex organization of both eucaryotic and procaryotic promoter clusters in the soybean nuclear gene coding for nodulin-23. The genomic clone for nodulin-23 was isolated (Mauro *et al.*, 1985) by screening a genomic library in bacteriophage lambda with the cDNA clone, pNodA-25, as probe (Fuller *et al.*, 1983). The transcription start sites for this gene were precisely located by three different approaches: (i) determination of the size of the α-amanitin sensitive, run-off transcripts *in vitro*; (ii) S1 nuclease mapping with both *in vivo* nodule mRNA and *in vitro* transcribed RNA and (iii) primer extension analysis using nodule mRNA. We demonstrated the presence of two (a major and a minor) transcription start sites for this nodulin gene. *In vitro* transcription analyses in a heterologous system revealed the presence of another α-amanitin-sensitive promoter which is located within the 5′ end of nodulin genes and is transcribed in the opposite direction ('anti-sense') relative to that of the nodulin promoters. Sequence analyses at the 5′ end flanking region of this gene also revealed the presence of a tandemly arranged cluster of procaryotic promoters. *In vivo* analysis of these procaryotic promoters demonstrated that they are functional in *Escherichia coli*. The potential regulatory roles for the presence of procaryotic regulatory sequences in this region of a nodulin gene are discussed in relation to the symbiotic nature of the tissue in which this gene is induced.

Results

Figure 1 shows the restriction map of a *Hin*dIII-*Eco*RI fragment which carries the 5′ end region of the soybean nodulin-23 gene. This fragment was subcloned into pBR322 and was used for analyzing the promoter region of the nodulin-23 gene. The complete nucleotide sequence of this region, and the primary structure of nodulin-23 gene is presented elsewhere (Mauro *et al.*, 1985).

Localization of the in vivo transcription start site(s)

By comparing the pattern of restriction enzyme sites on the cDNA clone (pNodA-25) of the nodulin-23 gene with that on the genomic clone (data not shown), the promoter region for this gene was expected to be within a 1.3-kb *Hin*dIII-*Xho*I fragment. This fragment (Figure 1) with the *Xho*I site labelled was used for *Exo*VII and S1 nuclease mapping resulting in the protected fragments shown in Figure 2. Although the species protected from *Exo*VII are slightly longer (i.e., 6–8 pairs longer as determined by high resolution mapping, data not shown) than the S1 nuclease-protected species (see Bonahug *et al.*, 1982), the size is very similar and is ~550 nucleotides long. This indicates that there is no intron sequence between the transcription start site and the *Xho*I site, and that the transcription start site is located ~550 nucleotides upstream from the labelled *Xho*I site (see Mauro *et al.*, 1985). The two minor S1-protected fragments appear to be due to a nodulin-23-related gene transcript (see below).

To determine the transcription start site at the nucleotide level, a *Hin*fI-*Dde*I fragment of 350 bp was used for high resolution S1 nuclease mapping following the end-labelling of the *Dde*I site (Figure 1). The results in Figure 3 (lane 9) show a major and a very minor band separated from each other by 24 nucleotides.

© IRL Press Limited, Oxford, England.

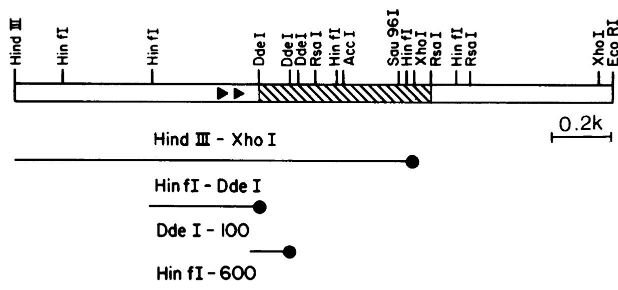

Fig. 1. Restriction map of the 2-kb *Hind*II-*Eco*RI fragment carrying the 5′ end flanking region of the soybean nodulin-23 gene. Arrowheads indicate the location and transcription direction for the two tandemly arranged eucyarotic promoters and hatched area indicates the coding region of first exon (see Mauro *et al.*, 1985). Fragments 1 and 2 are the *Hind*III-*Xho*I-1300 and *Hin*fI-*Dde*I-350 fragments for low resolution *Exo*VII-S1 nuclease mappings and high resolution S1 nuclease mapping, respectively. Fragment 3 is the *Dde*I-100 fragment for primer extension analysis. Fragment 4 is the *Hin*fI-600 fragment for *in vitro* transcription.

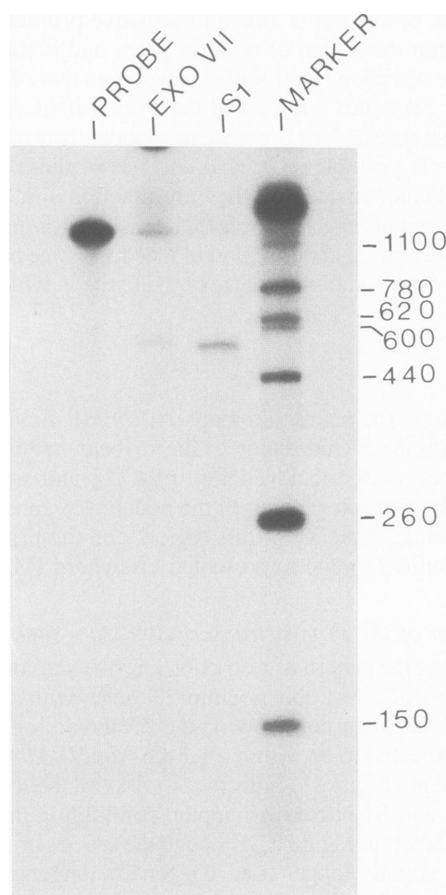

Fig. 2. Localization of the *in vivo* transcription start site by *Exo*VII and S1 nuclease mapping. A *Hind*III-*Xho*I-1300 fragment, with the *Xho*I site labelled, was used as a probe for *Exo*VII and S1 nuclease mapping. **Lane 1** shows the size of the probe. **Lanes 2** and **3** show the protected fragment generated by *Exo*VII and S1 treatment. **Lane 4** shows the *Hind*III-digested phage 29 as a size marker.

To substantiate further these transcription start sites, primer extension analysis was performed with a 100 nucleotides long, single-stranded *Dde*I fragment (Figure 1). Two primer extended fragments in lane 3 with a size of 184 and 160 nucleotides, respectively, correspond to the major and minor transcription start sites observed by S1 nuclease mapping (lane 9). However, a major primer-extended fragment of 140 nucleotides in length which does not correspond to any S1-protected fragment was also observed. The absence of any thermostable secondary structures at the 5′ end of the nodulin-23 mRNA, as predicted from the DNA sequence data (Mauro *et al.*, 1985), excludes the possibility that this fragment is an artifact of the primer extension experiment due to secondary structure.

As reported previously (Fuller *et al.*, 1983), two different cDNA sequences (NodA-25 and NodA-15) were found to have extensive homology with each other. Southern blot analysis of genomic DNA and Northern blot of nodule mRNA, using these clones as probes, suggest the existence of a family of at least two related NodA genes (Fuller and Verma, 1984). Thus, this extra band (140 nucleotides long) appears to be the extension product with the transcript from the other member of the NodA gene family. The relative intensity of this fragment indicates that the transcript from the nodulin-23-related gene is much more abundant than that of the nodulin-23 gene. This is consistent with the result of quantitative dot-blot analysis which shows that they are, in fact, present in a ratio of 30:1 in 21-day nodules (Fuller and Verma, 1984). These observations imply that the primer has almost perfect homology with the transcript from these two related genes. However, the 5′ end region upstream of the primer sequence for these two genes is clearly not completely homologous since S1 nuclease mapping does not show any protected fragment of this size. Preliminary analysis of a nodule-specific cDNA, pNodA-15 (Fuller *et al.*, 1983) sequences showed stretches of homology with the nodulin-23 gene (unpublished data).

In vitro transcription of the nodulin-23 gene

The *in vitro* transcription system of the HeLa cell extract has been successfully applied to analyze the transcription of plant genes. At least at low resolution level, the HeLa cell extract has shown to initiate transcription from one of the cloned zein genes at the same position as that of the *in vivo* mRNA in maize (Langridge and Feix, 1983). Thus, we used the HeLa cell transcription system to determine whether the major transcription start site on the nodulin-23 gene can be functional *in vitro* with a 600-bp *Hin*fI fragment (Figure 1) as template. The *in vitro*-transcribed RNA was used for S1 nuclease mapping. The result shown in Figure 4 indicates that the *in vitro* transcript initiates at exactly the same site as the major *in vivo* transcript in 21-day nodules.

Presence of an α-amanitin-sensitive transcript promoter in vitro

During the *in vitro* transcription analysis with the *Hin*fI-600 fragment as template, two α-amanitin-sensitive RNA run-off transcripts were observed (Figure 5). One of them is ~260 nucleotides long, which corresponds to the transcript derived from one of the nodulin-23 promoters, while the other is 420 nucleotides long derived from an unknown promoter on this fragment. This could have resulted by transcription, either in the same direction as the nodulin promoter or in the opposite direction. In any event, this transcript must overlap with the nodulin message since the template is only 600 bp long (see sequence analysis, Mauro *et al.*, 1985). We determined the location and transcription direction for this unknown promoter by an approach summarized in Figure 5A. Keeping one arm of the *Hin*fI-600 fragment constant, while increasing the length of the template at either arm and observing for the corresponding changes in the transcript length allowed us to determine the transcription direction for this promoter and thus the promoter location. With the *Hin*fI-600 frag-

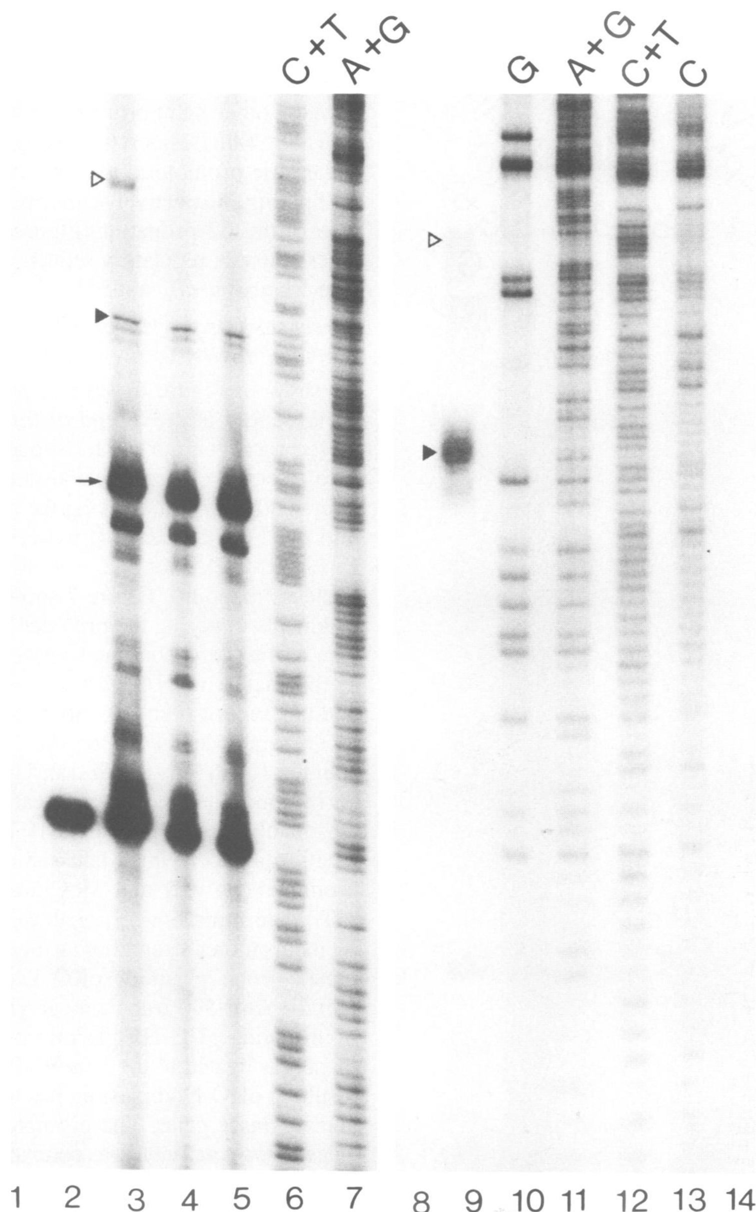

Fig. 3. Primer extension and S1 nuclease mapping of the *in vivo* nodulin-23 transcripts. The *Dde*I-100 fragment was used as a probe for the primer extension analysis. Primer alone and the primer-extended fragments (hybridized with 21-day nodule RNA at 70°, 65° and 60°C) are shown in **lanes 2–5**. Open and closed arrowheads (▷and ▶) indicate the upper and lower transcription start sites for the nodulin-23 gene. The arrow (→) shows the transcription site from a nodulin-23-related gene (see text). The *Hinf*I-*Dde*I-350 fragment was used as a probe for S1 nuclease mapping with RNA from 21-day-old nodules (**lane 9**). *E. coli* tRNA hybridized to the probe shows no protection as indicated in **lane 8. Lanes 6–7** and **10–13** show the sequence ladder of the *Hinf*I-*Dde*I fragment generated by the Maxam-Gilbert reaction. The ladder in **lanes 6** and **7** is a size marker only. **Lanes 1** and **14** are blank.

ment as template, the two transcripts are 260 and 420 nucleotides long (Figure 5B, lane 3). With an increase in length at the right arm of the template of 150 nucleotides, one of the transcripts (nodulin message) increases in size as expected from 260 to 410 nucleotides while the other transcript remains the same (lane 5). When the third template (mainly composed of a *Hinf*I-600 fragment with an extension of 400 bp at the left arm) is used, the corresponding increase in transcript length to 820 nucleotides from this unknown promoter (lane 7) confirms its transcription direction and location. As shown in Figure 5A, this promoter is transcribed in the opposite direction ('anti-sense') relative to that of the nodulin promoter and has the 5′ end of its transcript overlapping for ~80 nucleotides with that of the nodulin message. The α-aminitin sensitivity at 1 μg/ml (lanes 2, 4 and 8) indicates that these two promoters are transcribed by RNA polymerase II (Roeder, 1976). The addition of wheat germ RNA polymerase II to the *in vitro* transcription system stimulates the transcription from both promoters (data not shown). Attempts to isolate 'anti-sense' transcript *in vivo* from either uninfected roots or nodule tissues were unsuccessful, indicating that this transcript if initiated *in vivo* is not released and hence does not accumulate (see Discussion).

Regulatory sequence at the 5′ end of the nodulin-23 gene

Sequence analysis of the 5′ end region of the nodulin-23 gene revealed two tandemly arranged eucaryotic promoter-like sequences (E1 and E2) for the two corresponding transcription start sites, as determined by S1 nuclease mapping and primer extension (Figure 6). For the downstream promoter (E2), a 'CAAT' box with the sequence of CAATTT at position −73 and a

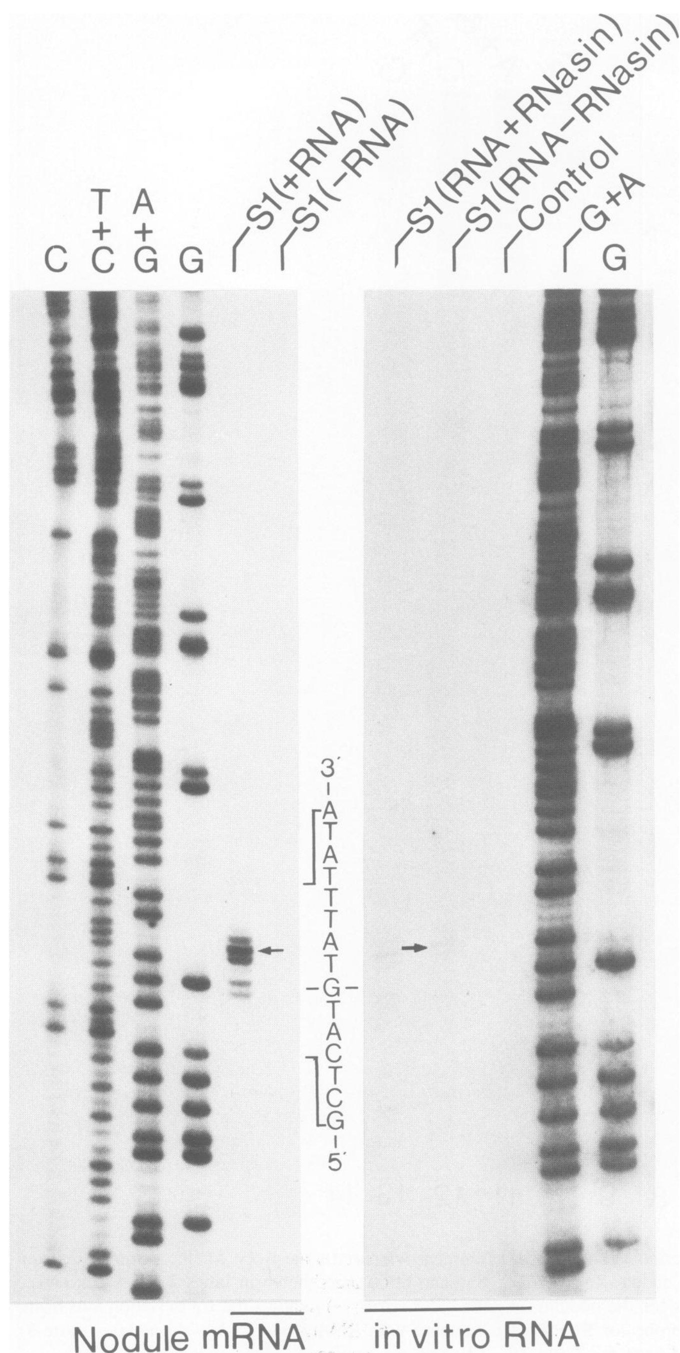

Fig. 4. S1 nuclease mapping of *in vitro*-transcribed RNA from the nodulin-23 promoter. A *Hinf*I-600 fragment shown in Figure 1 was used as template for *in vitro* transcription with the HeLa cell crude extract. On the left, it shows the S1 nuclease mapping with and without *in vivo* 21-day-old nodule polysomal RNA. On the right, it shows the S1 nuclease mapping with the *in vitro*-transcribed RNA synthesized in the presence or absence of RNasin. Arrows indicate the corresponding transcription start site observed *in vivo* and *in vitro*. Endogenous RNA derived from the HeLA cell crude extract (in the absence of template during transcription reaction) was used as control for S1 nuclease mapping.

'TATA' box with the sequence of TTATAAAA at position −26 relative to downstream transcription start site (+1) was observed. Both these consensus sequences and their location are in good agreement with the known consensus sequence for animal promoters (Breathnach and Chambon, 1981), as shown in Table 1B. For the upstream promoter (E1), a 'AGGA' box sequence of AATTGAAT at position −81 and a 'TATA' box of TTATATA at position −24 were observed (Figure 6). This set of 'AGGA' box and 'TATA' box sequences is very similar to that observed in many other plant promoters (Messing *et al.*, 1983), as shown in Table 1A. Besides the presence of two tandemly arranged eucaryotic promoters, two inverted repeat sequences of 31 and 21 bp long, respectively, are observed around the 'TATA' box area of the E2 promoter (Figure 6). In addition, there may be other potential regulatory sequences further upstream in this gene (see Mauro *et al.*, 1985).

Presence of procarytic promoter sequence within the eucaryotic promoter region of the nodulin-23 gene

As show in Figure 6, several procaryotic promoter sequences are observed at the 5′ end of this gene. To determine whether these procaryotic promoter sequences can be functional *in vivo* in a procaryotic cell, we tested their activity in *E. coli*. A *Hinf*I-600 fragment carrying the promoter sequence was inserted into pUC9. The resulting pU-HF plasmid was transferred into *E. coli* JM101. Total *E. coli* RNA was then prepared for S1 nuclease mapping. Figure 7 shows two clusters of S1-protected fragments implying the presence of two transcription start sites. One of the corresponding transcription start sites is very similar to the major mRNA start site observed *in vivo* in the 21-day nodule. Figure 6 shows three sets of putative procaryotic promoter sequences in this region. Two of them, U1 and U2, accounted for the upstream transcription start site. The third one corresponds to the downstream site. They have not only reasonably good −35 and −10 sequences, but also the proper length of spacer which is essential for a functional procaryotic promoter (Hawley and McClure, 1983).

To determine how efficient this procaryotic promoter cluster is, its promoter strength was measured by using an *E. coli* promoter probe plasmid, pKO-1 (McKenney *et al.*, 1981). A *HInd*III-*Rsa*I-530 fragment carrying the promoters was prepared by digesting pUC-HF plasmid by *Hind*III and *Rsa*I. This fragment was ligated to the *Hind*III/*Sma*I-digested pKO-1 so that the resulting pKO-Nod plasmid has the promoter facing towards the galactokinase gene. The promoter strength, as reflected by the galactokinase activity, was compared with *E. coli* carrying pKO-1 plasmid alone or pKG-1800 carrying a native *gal* promoter in front of the galactokinase gene (serving as a control for this assay system). The promoter strength for the procaryotic promoter clusters on the nodulin gene has a value of 198 galactokinase units which is comparable with the promoter strength of *ara*C in *E. coli* (see McKenney *et al.*, 1981).

Discussion

Nodulin-23 represents two closely related polypeptides (23 500, 24 500) synthesized in soybean nodule tissue (see Mauro *et al.*, 1985; Fuller *et al.*, 1983). Expression of nodulin-23 is specific to nodule tissue and is under developmental control (Fuller *et al.*, 1983; Fuller and Verma, 1984). Its mRNA 1100 nucleotides in length, is detectable in the infection zone as early as 5 days after *Rhizobium* infection and reaches the maximum point in ~11 days (Fuller and Verma, 1984). Nucleotide sequence analysis and fine structural mapping of the nodulin-23 gene revealed a complex organization of both eucaryotic and procaryotic-type promoter clusters at the 5′ end flanking region.

Transcription start sites

Two nodulin-23 specific mRNAs with their 5′ ends separated from each other by 24 nucleotides were detected by high resolu-

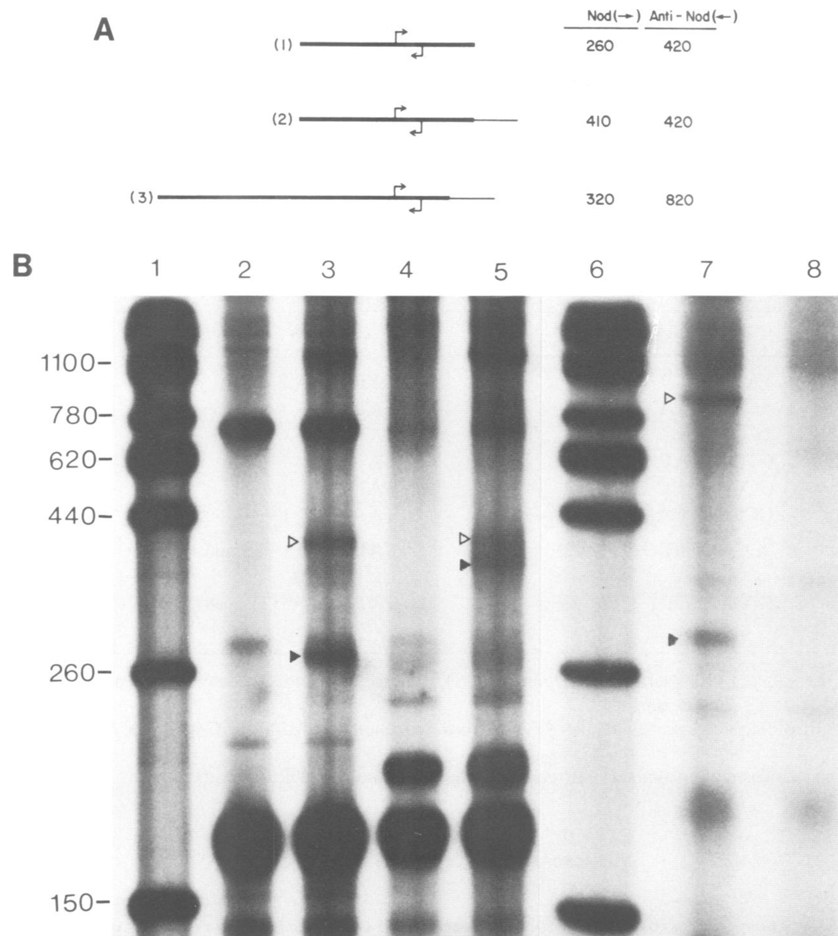

Fig. 5. (**A**) Transcriptional mapping of promoters at the 5′ end region of the nodulin-23 gene. Solid bars represent the DNA templates used for *in vitro* transcription with HeLa cell extract. Fragment 1 is a *Hinf*I-600 fragment (Figure 1). Fragment 2 is ~750 bp long and is composed of the same *Hinf*I-600 fragment as fragment 1 except that the right arm of the *Hinf*I-600 fragment is extended by 150 bp into the *Pvu*II site of the β-galactosidase gene on the pUC-HF plasmic. Fragment 3 is ~1100 bp long and is composed of a *Hind-Rsa*I-950 fragment derived from the *Hind*III-*Eco*RI-2000 fragment of the nodulin-23 gene and a 150-bp fragment derived from pUC-*Rsa*I plasmid. This 150-bp portion at the right arm of fragment 3 is the same as that at the right arm of fragment 2 and is represented by a thin line. Arrows above and below the templates show the location and transcription direction of nodulin (→) and anti-nodulin (←) promoters, respectively. The table on the right lists the size of nodulin (→) and anti-nodulin (←) transcripts derived from each template.
(**B**) Autoradiogram showing the run-off transcripts from fragments 1, 2 and 3 as template in the absence (**lanes 3, 5, 7**) and presence (**lanes 2, 4, 8**) of α-amanitin at a final concentration of 1 μg/ml. Open and closed arrow heads indicate the α-amanitin-sensitive transcripts from the nodulin and anti-nodulin promoters, respectively. The *Hind*III-digested phage 29 was used as a size marker (**lanes 1** and **6**). **Lanes 2** and **3** correspond to fragment 1, **lanes 4** and **5** to fragment 2 and **lanes 7** and **8** to fragment 3.

tion S1 nuclease mapping and primer extension analysis. The shorter transcript represents the major form of nodulin-23 mRNA in 21-day nodules. Sequence analysis at the 5′ end region reveals the presence of a plant-like eucaryotic promoter (E1) and an animal-like promoter (E2) which are located upstream to each of the transcription start sites, respectively. Two possible mechanisms can account for the production of mRNA with different 5′ ends. They can be transcribed from different promoters or the shorter transcript is generated through processing of the longer one (Grossland *et al.*, 1984). Although the processing mechanism cannot be completely ruled out, *in vitro* transcription of nodulin-23 gene with HeLa cell extract generated the main transcript from the downstream promoter (E2).

It is not uncommon for a single gene to have two or more promoters. In procaryotes, the *Bacillus subtilis* sporulation gene (*spo VG*) (Johnson *et al.*, 1983), the *Anabaena* glutamine synthetase gene (Tumer *et al.*, 1983) and the *E. coli* rRNA and galactose operons (Sarmientos and Caskel, 1983; Musso *et al.*, 1977) are all known to have dual promoters. In eucaryotes, two or more promoters occur in the α-amylase gene of mouse (Schilbler *et al.*, 1983), the alcohol dehydrogenase gene of *Drosophila* (Benyajati *et al.*, 1983) and the zein gene of maize (Langridge and Feix, 1983).

Multiple promoters in a single gene may provide a mechanism to achieve a differential expression under different physiological conditions. For example, one of the *Anabaena* glutamine synthetase promoters is turned on during growth when ammonia is available. The other one is activated only under nitrogen-fixation conditions (Tumer *et al.*, 1983). Furthermore, transcripts with different non-coding lengths at their 5′ end may potentially affect their stability or translation efficiency (Wood *et al.*, 1984). It is apparent that in nodulin-23 gene either both promoters are not used with equal efficiency or the longer transcript is more unstable.

Eucaryotic promoter sequence at the 5′ end region of the nodulin-23 gene

The promoter sequences of plant genes are more diverse than in animal genes and can basically be divided into two groups. Group I, as listed in Table 1B, has two consensus sequences

Table I.

A. Comparison of upstream promoter sequence (E1) in the nodulin 5′ end region and the consensus 'AGGA' box sequences found in plant promoters

	'AGGA' box	'TATA' box	References
Nodulin (E1)	GAATTGAAT (−81)	TTTATATACAAT (−24)	This work
Consensus sequence	${}^{C}_{T}A_{2-5}{}^{T}_{G}NGA_{2-3}{}^{T}_{C}$ (−70 −90)	$T^{C}_{G}TATA^{T}_{A}A_{1-3}T$ (−29 −33)	Messing *et al.*, (1983)

B. Comparison of animal-like promoter sequence (E2) in the nodulin 5′ end region and other plant promoter systems with consensus animal promoter sequence

	'CAAT' box	'TATA' box	References
Nodulin (E2), soybean	TACAATTT (−73)	TATAAAA (−26)	This work
Zein (E1), maize	ACCAATTA (−92)	TATAATG (−28)	Langridge and Feix (1983)
Legumin, pea	TCCAATTC (−88)	TATAAT (−31)	Lycett *et al.* (1984)
Nopalin synthase	GTCACTAT (−73)	CATAAT (−23)	Depicker *et al.*, (1982)
Alcohol dehydrogenase, maize	GCCAAACC (−94)	TATAAAT (−35)	Dennis *et al.* (1984)
RuBP carboxylase (SS), wheat	GCCAACCA (−89)	TATATAT (−34)	Broglie *et al.* (1983)
Plant consensus	${}^{G}_{T}CCAAT^{T}_{C}N$ (−73 −92)	TATAAAT (−23 and −35)	
Animal consensus	$G^{C}_{T}CAATCT$ (−70 −90)	$TATA^{A}_{T}A^{A}_{T}$ (−23 −34)	Breathnach and Chambon (1981)

All the transcription start site(s) for the selected promoters are determined by S1 nuclease mapping. The numbers in parentheses are the distances between the center of the consensus sequence and the transcription start site.
The promotor of zein gene (E1) and RuBP carboxylase has another 'CAAT' box sequence of TACAATAG and TCCAACCA at nucleotide −66 and −65, respectively.
E1 and E2 represents the upstream and downstream promoter; SS represents the small subunit of RuBP carboxylase.

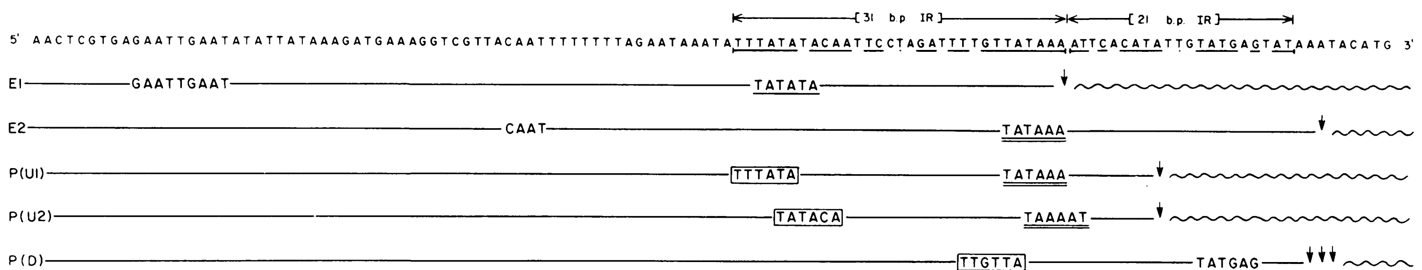

Fig. 6. Organization of eucaryotic and procaryotic promoters within the 5′ flanking region of nodulin-23 gene. Arrows show the transcription start sites as defined by S1 nuclease mapping and primer extension analysis. E1 represents the upstream eucaryotic promoter sequence showing the putative 'AGGA' box and 'TATA' box sequences. E2 represents downstream promoter with an animal-like promoter sequence showing the putative 'CAAT' box and 'TATA' box sequences. P(U1) and P(U2) show the procaryotic promoter sequence (putative −35 and −10 regions) corresponding to the upstream transcription start site defined by S1 nuclease mapping. P(D) represents the downstream procaryotic promoter sequence. Two regions of dyad symmetry overlapping the promoter region are indicated above the sequences. Dots represent the center of symmetry for each inverted repeat sequence.

similar to that of animal genes. Group II (Table 1A as reported by Messing), has the so-called 'AGGA' box sequence at position −70 to −100 rather than the regular 'CAAT' box (Messing *et al.*, 1983).

Examination of the 5′ end flanking sequence of the nodulin-23 gene shows a good match for both the 'CAAT' box and 'TATA' box sequences at nucleotides −73 and −26 (numbering from the downstream transcription start site) for the downstream promoter (Figure 6, E2) while putative 'AGGA' box and 'TATA' box sequences are located at nucleotide −81 and −24, respectively (numbering from the upstream transcription start site), for the upstream promoter (Figure 6, E1). S1 nuclease mapping of the *in vitro*-transcribed RNA demonstrates that the downstream promoter, having more similarity to animal promoters, can be recognized and accurately transcribed by the HeLa cell transcription system.

Presence of sequences capable of producing an 'anti-sense' transcript in vitro

In vitro transcription of the 5′ end region of the nodulin-23 gene demonstrated the presence of sequences which can act as an 'anti-sense' promoter giving rise to an α-amanitin-sensitive transcript whose 5′ end is complementary to the nodulin sense transcript for ~80 nucleotides. The arrangement of the sense and anti-sense promoters in the nodulin-23 gene could provide a mechanism to regulate the expression of this gene during the nodulation process through RNA duplex formation. However, no anti-nodulin transcript was obtained *in vivo*, indicating that if such a transcript is initiated its expression is either too low to detect or it is not released. Ample examples are available in procaryotes to illustrate that an anti-sense message indeed plays a role in regulating gene expression (Tomizawa *et al.*, 1981; Simons and Kleckner, 1983; Mizuno *et al.*, 1984). There are indications that such a regula-

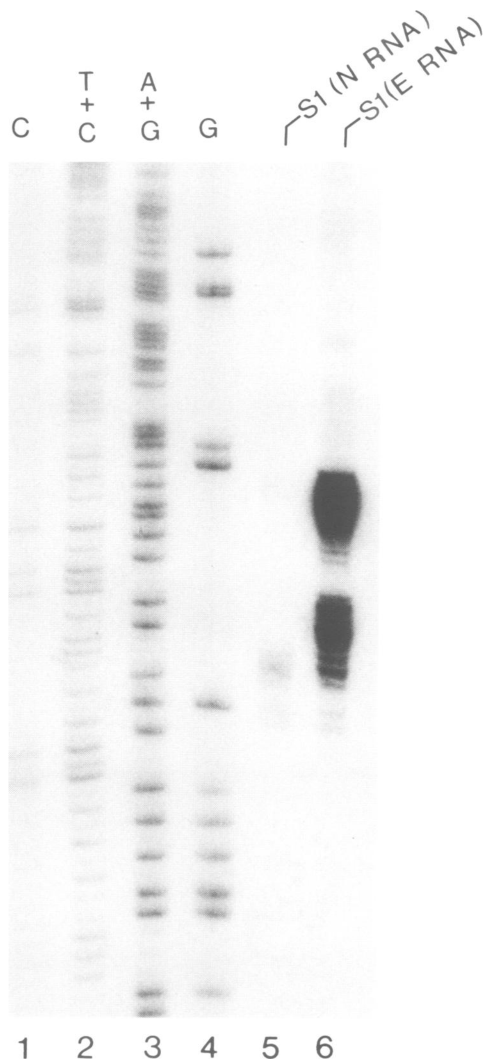

Fig. 7. Localization of potential transcription start sites within the procaryotic promoter region of the nodulin-23 gene. A *Hinf*I-*Dde*I-350 fragment with the *Dde*I site labelled was used as a probe for S1 nuclease mapping with *in vivo* 21-day-old nodule message (**lane 5**) and *in vivo* total RNA from *E. coli* carrying the pUC-HF plasmid (**lane 6**). **Lanes 1 – 4** show that the sequence ladder of the same *Hinf*I-*Dde*I-350 fragment *in vivo* RNA from *E. coli* carrying the pUC9 plasmid was used for controlled hybridization.

tion can also work in eucaryotes (Izant and Weintraub, 1984). Further analyses of the possible transcription of an anti-sense message *in vivo* is essential in understanding the regulation of expression of this gene in nodules.

Procarotic promoter clusters in the 5′ end of the nodulin-23 gene

The presence of a functional, tandemly arranged cluster of procaryotic sequences within the regulatory region of the nodulin-23 gene was demonstrated by S1 nuclease mapping of *in vivo E. coli* RNA. The 5′ end flanking region of the nopaline synthase gene on a Ti-plasmid is the other system known to have the simultaneous presence of functional eucaryotic and procaryotic promoters (Depicker *et al.*, 1982; Herrera-Estrella *et al.*, 1983). One common feature of these two systems is that both nodulation and crown gall formation heavily involve plant-bacterial interactions. Whether any *Rhizobium* components help in the transcription process of this nodulin gene is unknown, but many bacterial proteins are secreted inside the nodule cell (unpublished observations). Such an interaction would evidently bring a nodulin gene under *Rhizobium* control during symbiotic association.

Materials and methods

Materials

Restriction enzymes, DNase I, calf intestine alkaline phosphatase, T4 polynucleotide kinase and S1 nuclease were from Boehringer. *Exo*VII, RNasin, Klenow fragment, Vanadyl ribonucleotide complex were from Bethesda Research Laboratories. RNase and tRNA were from Sigma. Wheat germ RNA polymerase II was from P-L Biochemicals. Reverse transcriptase was from Life Science Inc.

Plasmid construction and isolation

pUC-HF was constructed by ligating a blunt-ended *Hinf*I-600 fragment into the *Sma*I site of pUC9 with the nodulin promoters reading toward the *Eco*RI site on the plasmid. pUC-*Rsa*I was constructed by ligating a *Hind*III-*Rsa*I-950 fragment into the *Hind*III/*Sma*I digested pUC9. pKO-Nod was prepared by ligating a *Hind*III and *Rsa*I fragment into the *Hind*III/*Sma*I-digested pKO-1 (McKenney *et al.*, 1981). All plasmids were purified by the CsCl-EtBr step gradient method (Garger *et al.*, 1983).

DNA sequence analysis

The DNA sequence was determined by the method of Maxam and Gilbert (1980).

In vitro transcription

The HeLa cell transcription system was purchased from New England Nuclear and was used as described by the manufacturer. The transcripts were analyzed on an 8 M urea-6% acrylamide gel (Maniatis and Efstratiadis, 1980). The DNA templates used for transcription analysis are shown in Figure 1.

Isolation of in vivo and in vitro RNA transcripts for S1 nuclease mapping

Total polysome RNA from 21-day nodules was prepared as described (Verma *et al.*, 1974). The *in vitro*-transcribed RNA for S1 mapping was prepared as described above except that the [α-^{32}P]GTP was replaced by the unlabelled GTP at a final concentration of 10 μM. The DNA template was removed by digestion with RNase-free DNase I. The isolation of total RNA from *E. coli* JM101 carrying either pUC9 or its derivatives was performed as described (Gilman and Chamberlin, 1983).

ExoVIII/S1 nuclease mapping

S1 nuclease mapping was performed as described by Berk and Sharp (1977) with the modification of Weaver and Wiessmann (1979). *Exo*VII nuclease mapping was according to the method of Rose and Botstein (1983) using the 5′ end-labelled fragments shown in Figure 1.

Primer extension with reverse transcriptase

For the primer extension experiment, a 100-bp long *Dde*I fragment (Figure 1) was used as a primer. It was labelled at the 5′ end and strand-separated as described by Maxam and Gilbert (1980). The labelled primer (20 000 c.p.m.) was mixed with 100 μg of nodule polysomal RNA and hybridized in 10 μl of hybridization buffer (40 mM Pipes, pH 6.4, 0.4 M NaCl, 1 mM EDTA) for 5 h, at 60, 65 and 70°C respectively. The primer extension was performed as described by Luse *et al.* (1981) with 5 units of AMV reverse transcriptase at 42°C for 1 h.

Acknowledgements

We would like to thank Truyen Nguyen and Vincent Mauro for communicating the sequence information, Peter Kunstner and Eric Olson for their comments and Yvette Mark for preparing the manuscript. This study was supported by Operating and Strategic research grants from the Natural Sciences and Engineering Research Council of Canada and the Ministry of Education, Quebec.

References

Benyajati,C., Spoeral,N., Haymerle,H. and Ashburner,M. (1983) *Cell*, **33**, 125-133.

Bergmann,H., Preddie,E. and Verma,D.P.S. (1983) *EMBO J.*, **2**, 2333-2339.

Berk,A.J. and Sharp,P.A. (1977) *Cell*, **12**, 721-732.

Beringer,J.E., Brewin,N.J. and Johnson,A.W.B. (1980) *Heredity*, **45**, 161.

Bisseling,T., Been,C., Klugkist,J., Nadler,K. and Van Kammen,A. (1983) *EMBO J.*, **2**, 961-966.

Bonahug,P.F., Farabagh,P.J. and Fink,G.R. (1982) *Gene*, **18**, 47-59.

Breathnach,R. and Chambon,P. (1981) *Annu. Rev. Biochem.*, **50**, 349-383.

Broglie,R., Coruzzi,G., Lamppa,G., Keith,B. and Chua,N.H. (1983) *Bio/Technology*, **1**, 55-61.

Caldwell,B.E. and Vest,H.G. (1977) in Hardy,R.W.F. and Silver,W.S. (eds.), *A Treatise on Dinitrogen Fixation*, Wiley, NY, pp. 557-576.

Crossland,L.D., Rodermal,S.R. and Bogorad,L. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 4060-4064.
Dennis,E.S., Gerlach,W.L., Pryor,A.J., Bennetzen,J.L., Inglis,A., Llewellyn,D., Sachs,M.M,. Ferl,R.J. and Peacock,W.J. (1984) *Nucleic Acids Res.*, **12**, 3983-4000.
Depicker,A., Stachel,S., Dhaese,P., Zambryski,P. and Goodman,H.M. (1982) *J. Mol. Appl. Genet.*, **1**, 561-574.
Fuller,F. and Verma,D.P.S. (1984) *Plant Mol. Biol.*, **3**, 21-28.
Fuller,F., Kunstner,P.W., Nguyen,T. and Verma,D.P.S. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 2594-2598.
Garger,S.J., Griffin,O.M .and Grill,L.K. (1983) *Biochem. Biophys. Res. Commun.*, **117**, 835-842.
Gilmann,M.Z. and Chamberlin,M.J. (1983) *Cell*, **35**, 285-293.
Hawley,D.K. and McClure,W.R. (1983) *Nucleic Acids Res.*, **11**, 2237-2249.
Herrera-Estrella,L., Block,M.D., Messens,E., Hernalsteens,J.P., Montagu,M.V. and Schell,J. (1983) *EMBO J.*, **2**, 987-995.
Izant,J.G. and Weintraub,H. (1984) *Cell*, **36**, 1007-1015.
Johnson,W.C., Moran,C.P.,Jr. and Losick,R. (1983) *Nature*, **302**, 800-804.
Lang-Unnasch,M. and Ausubel,F.M. (1985) *Plant Physiol.*, **77**, 833-839.
Langridge,P. and Feix,G. (1983) *Cell*, **34**, 1015-1022.
Legocki,R.P. and Verma,D.P.S. (1980) *Cell*, **20**, 153-163.
Luse,D.S., Haynes,J.R. Van Leeuwen,D., Schon,E.A., Cleary,M.L., Shapiro,S.G., Lingrel,J.B. and Roeder,R.G. (1981) *Nucleic Acids Res.*, **9**, 4339-4354.
Lycett,G.W., Croy,R.R.D., Shirsat,A.H. and Boulter,D. (1984) *Nucleic Acids Res.*, **12**, 4493-4506.
Maniatis,T. and Efstratiadis,A. (1980) *Methods Enzymol.*, **65**, 299-305.
Mauro,V., Nguyen,T., Katinakis,P. and Verma,D.P.S. (1985) *Nucleic Acids Res.*, **13**, 239-249.
Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499-560.
McKenney,K., Shimatake,H., Court,D., Schmeissner,U. and Rosenberg,M. (1981) in Chirikjaian,J.G. and Papas,T.S. (eds.), *Gene Amplification and Analysis*, Vol. 2, Elsevier, North Holland, pp. 384-415.
Messing,J., Geraghty,D., Heidecker,G., Hu,N., Kridl,J. and Rubenstein,I. (1983) in Kosuge,T., Meredity,C.P. and Hollaender,A. (eds.), *Genetic Engineering of Plants*, Plenum Press, NY, pp. 211-227.
Mizuno,T., Chou,M.Y. and Inouye,M. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 1966-1970.
Musso,R.E., DiLauro,R., Adhya,S. and Crombruggle,B. (1977) *Cell*, **12**, 847-857.
Roeder,R.G. (1976) in Losick,R. and Chamberlin,M. (eds.), *RNA Polymerase*, Cold Spring Harbor Laboratory Press, NY, pp. 285-239.
Rose,M. and Botstein,D. (1983) *J. Mol. Biol.*, **170**, 883-904.
Sarmientos,P. and Askel,M. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 7010-7013.
Sehilbler,U., Hagenbuchle,O., Wallauer,P.K. and Pittet,A.C. (1983) *Cell*, **33**, 501-508.
Simons,R.W. and Kleckner,N. (1983) *Cell*, **34**, 683-691.
Tomizawa,J.I., Itoh,T., Zelzer,G. and Som,T. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 1421-1425.
Tumer,N.E., Robinson,S.J. and Haselkorn,R. (1983) *Nature*, **306**, 337-342.
Verma,D.P.S. and Long,S. (1983) in Jeon,K. (ed.), *International Review of Cytology*, Suppl. 14, Academic Press, NY, pp. 211-245.
Verma,D.P.S., Nash,D.T. and Schulman,H.M. (1974) *Nature*, **251**, 74-77.
Weaver,R.F. and Weissman,C. (1979) *Nucleic Acids Res.*, **7**, 1175-1193.
Wood,C.R., Boss,M.A., Patel,T.P. and Emtage,J.S. (1984) *Nucleic Acids Res.*, **12**, 3937-3950.

Received on 19 March 1985; revised on 24 June 1985