The EMBO Journal vol.5 no.11 pp.2759–2762, 1986

Splicing of an intervening sequence from hybrid cauliflower mosaic viral RNA

Barbara Hohn, Ervin Balázs, Doris Rüegg and Thomas Hohn

Friedrich Miescher-Institut, P.O. Box 2543, CH-4002 Basel, Switzerland

Communicated by B.Hohn

A DNA sequence encompassing intervening sequence 2 of the soybean leghemoglobin gene was inserted into the cauliflower mosaic virus (CaMV) genome. Upon passage through host plants progeny virus slowly accumulated that had lost the intron, apparently by exact splicing of the RNA form of the genome at the exon/intron borders. These findings suggest that reverse transcription from (spliced) RNA to DNA is a normal step in CaMV replication. S1 nuclease mapping revealed a mixed population of unspliced and spliced viral RNAs. The inefficiency of the splicing process is discussed.
Key words: cauliflower mosaic virus/intron/leghemoglobin/retroid virus/splicing

Introduction

Cauliflower mosaic virus (CaMV) is a plant virus with double-stranded relaxed circular DNA (for reviews, see Hohn and Hohn, 1982; Howell, 1982; Hull, 1984; Dixon and Hohn, 1985). The icosahedral virus particles accumulate in typical inclusion bodies within the cytoplasm, whereas a supercoiled form of the genome is found as a minichromosome in the nucleus. The genome is 8 kb long and contains 6–8 open reading frames (ORFs) in a densely packed form. All of these are located on the plus strand. Two transcripts have been characterized, one coding for the inclusion body matrix protein, the other encompassing the total genome plus a 180-base terminal repeat. The latter has been suggested to be a replicative intermediate that becomes replicated to the DNA form of the genome by reverse transcription (Hull and Covey, 1983; Pfeiffer and Hohn, 1983; Guilley *et al.*, 1983; Robertson *et al.*, 1983; Hohn *et al.*, 1985; Takatsuji *et al.*, 1986).

Although this model is based on a wealth of circumstantial evidence, direct proof is still missing. However, certain predictions could be made, e.g. that intervening sequences are unstable in a CaMV population: the cycle would depend on the presence of genomic RNA, which upon passage of the nuclear–cytoplasmic barrier should lose its intervening sequences. A direct test of this assumption is presented in this paper.

Results and Discussion

Intervening sequence 2 of the soybean leghemoglobin gene (Brisson and Verma, 1982) was chosen for insertion into the gene II region of CaMV because a restriction fragment encompassing it could be used to replace most of the structural information of gene II without violating the 'rules' CaMV dictates onto inserts (Sieg and Gronenborn, 1982; Dixon and Hohn, 1984). The 143-bp fragment is small enough and its insertion changes the architecture of reading frames only to an extent which might be acceptable by the virus (Figure 1a). The new gene II region, which replaces the dispensable aphid transmission factor, now contains two coding regions separated by 57 bp. Such an arrangement resembles that of ORF VII–ORF I, which are separated by 63 bp (Guilley *et al.*, 1982). After splicing, a 'pseudogene II' would be created, coding for a 27 amino acid long peptide and retaining the original intergenic distances to gene I and gene III (Figure 1b). Thus the artificially constructed CaMV version as well as the expected spliced one should not be under selective pressure for rearrangement, which might otherwise distort the ratios of the molecules.

CaMV recombinants with the insert in the sense orientation with respect to CaMV transcription (Figure 1a), in the reverse orientation, and those lacking the insert altogether were cleaved at the cloning site and inoculated onto turnips. The kinetics of symptom appearance is shown in Table I. The sense clone yielded symptoms at about the normal time post-inoculation (a wild-type clone of this virus isolate gives rise to local chlorotic lesions in ~14 days and to a systemic infection consisting of vein clearing of newly formed leaves in ~21 days post-infection). The clones of the original virus vector, a non-infectious construction, or of the hybrid viral DNA with the insert in the anti-sense orientation did not yield symptoms during the usual observation time of 2 months (samples 1 and 3, Table I). Systemic infection was observed at a very late stage only, at which time DNA rearrangements of originally lethal constructions frequently are observed (Dixon and Hohn, 1984). A virus sample isolated from late infection with the intron-sense clone was used, in a cloned version, as a control for the infection kinetics of the intron-containing clones. It was found to produce symptoms with the same kinetics as the original intron-sense containing clone (Table I, samples 2 and 4).

The electrophoretic mobilities of DNA fragments from the region of the CaMV that contained the intervening sequence were tested. The *Cla*I–*Bam*HI fragment from the genome of the hybrid virus (Figure 1a) was used as a probe for the size determination of the corresponding viral segment representative during the infection cycle. Figure 2 shows, for two infected plants, the electrophoretic mobilities of the DNA fragment encompassing the intervening sequence as a function of time after inoculation. At the earliest time point analysed, a fragment of the original size, as well as a fragment ~100 bp smaller, was found. The larger fragment persisted, in decreasing amounts, even up to 5 weeks after inoculation, after which time it was replaced totally by the smaller fragment. A very similar behaviour was found for two additional independently inoculated plants (data not shown). Thus there appeared to be no strong pressure to delete these sequences. The absence of a strong bias in favor of deletion therefore seems to justify the experimental design.

To be able to draw conclusions on the mechanism of the DNA excision, the DNA fragments that had undergone the deletion during the infectious cycle were sequenced. Viral DNAs from late infections were cloned and sequenced in the 'gene II' region. The sequence of the original intron-sense clone used as inoculum agreed with the one predicted from CaMV and leghemoglobin sequences (Figure 1b, upper part). The sequences obtained from

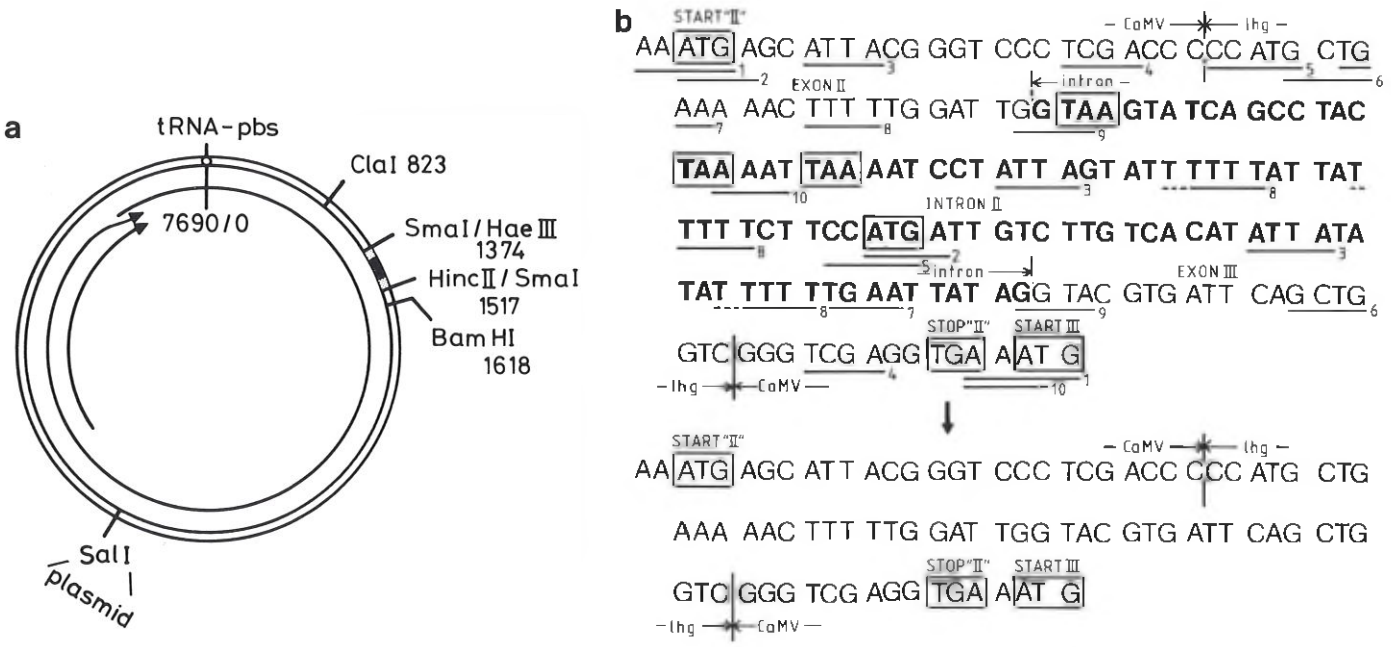

Fig. 1. (a) Simplified map of the intron-containing hybrid of CaMV. ▭ CaMV DNA, ▩ leghemoglobin DNA (■ intron, ▩ exon). → RNA with direction of transcription. tRNA-pbs = tRNA-primer binding site. The plasmid was constructed by inserting the 143 bp long *Hae*III–*Hin*cII fragment, encompassing intervening sequence 2 of the soybean leghemoglobin gene I (Brisson and Verma, 1982), into the *Sma*I site that had been introduced into CaMV strain CM4184 deleted for gene I by B.Gronenborn (personal communication). Only the sense orientation clone is shown. Only relevant restriction sites are marked. (b) Sequence of the chimaeric gene II containing the intervening sequence in the sense orientation. Thick letters are intron nucleotides; boxed triplets indicate translation signals. Sequences underlined (pairs, in one case four members) show homologies. They were selected using the following rules: (i) they hve to be between ORF I and III, (ii) the new 'gene II' should code for one protein only (or no protein at all, for underlined sequence No. 1), (iii) only in-phase deletions are permissible. (iv) the minimal size of the deletion is 50 bp, (v) not more than 50 bp are allowed as the intergenic space to gene III, (vi) only homologies from 4 bp onwards and without mismatch are considered. On the lower part the sequence of the deleted version of the chimaeric gene II is drawn.

three independent clones, each from an independent inoculation, were identical and are shown on the lower part of Figure 1b.

Comparison of the original and deleted sequences implies that splicing of the soybean leghemoglobin gene intervening sequence No. 2 has been carried out precisely, according to the splice signal rules. Alternative modes of excision of the intervening sequence cannot rigorously be ruled out, however: at the splice junction there is a 4 bp long direct duplication which could have served as substrate for homologous DNA recombination (Figure 1b). Mechanisms of this kind have previously been implicated in removing sequences contradicting 'the CaMV rules' (Sieg and Gornenborn, 1982: Dixon and Hohn, 1984). However, there are many other such direct repeat sequences which could have been used by a homologous recombination mechanism (see Figure 1b) and obviously were ignored or led to low virulence viruses. In addition, the removal of the intervening sequence in CaMV occurred much faster and under much less selective pressure than recombinational deletion (Dixon and Hohn, 1984) and, in contrast to it, it occurred at one site only.

The sequence data, together with the considerations discussed above, thus make it rather likely that a *bona fide* splicing event, recognizing the splice sequences as in the original organism soybean, had taken place to remove the intervening sequences that had been artificially inserted into the CaMV genome. The fact that CaMV sequences had undergone the same screening for spliceable sequences as do normal cellular mRNA molecules implies a DNA – RNA – DNA replication step.

A replication mechanism involving reverse transcription predicts that natural viral RNA is not spliced or, if it is, that the viral mRNA and the viral template RNA molecules are functionally (by compartmentalization?) separated. Indeed, there is

Table I. Infectivity of CaMV constructs

Sample number	Description of inoculum	Symptoms (days after inoculation)			
		14	20	53	105
1	CaMV parent[a]	–	–	–	+
2	intron clone, sense	(+)	+	+	+
3	intron clone, anti-sense	–	–	–	+
4	intron clone, sense, spliced version	(+)	+	+	+

4–8 plants were inoculated with 5 μg each of *Sal*I-cleaved CaMV clone DNA. (+) local symptoms, + systemic infection.
[a]This is the *Sma*I vector used for the construction (see Figure 1a), a non-viable mutation.

no evidence for splicing in the CaMV genome. An exception seems to be a Japanese isolate of CaMV: in plants infected with this strain a subpopulation of molecules deleted for a specific sequence has been described, for which after analysis of mutations a splicing mechanism was proposed (Hirochika *et al.*, 1985).

The very slow accumulation of the spliced form of CaMV DNA (Figure 2) could be explained in two ways. The persistence of the unspliced version is due to a DNA – DNA replication step in CaMV multiplication which is normally hidden by the DNA – RNA – DNA pathway and which, by unknown selective or counter-selective forces, became unmasked in the present experiment or, alternatively, the splicing process in this case is inefficient. To distinguish between these possibilities, polyadenylated RNA was isolated from infected plants at various times after inoculation and assayed for spliced and unspliced forms of CaMV RNA by S1 mapping using as a probe a 5'-labelled CaMV DNA restriction fragment covering the intron (Figure

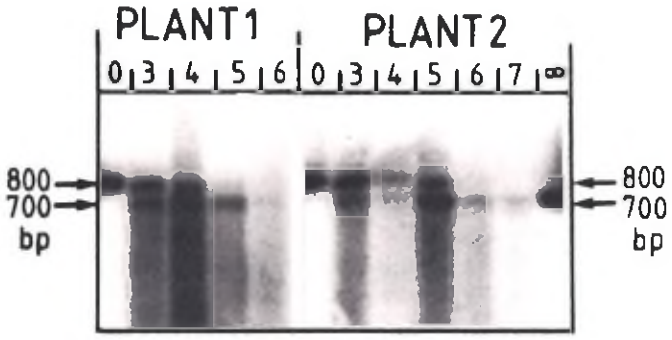

Fig. 2. Kinetics of intron removal. Plants (*Brassica rapa*) were infected with the CaMV intron-sense clone after digestion with the cloning enzyme *Sal*I. At periodic intervals CaMV DNA was extracted from one leaf each (Gardner and Shepherd, 1980). DNA was digested with *Cla*I and *Bam*HI, electrophoretically separated and transferred to nitrocellulose. Hybridization was with nick-translated *Cla*I – *Bam*HI fragment originating from the intron clone. Time points are 3, 4, 5, 6, 7 weeks after infection. Lanes ○ and ∞ contain inoculum DNA and DNA from one of the sequenced intron-less clones, digested with the same enzymes as total plant DNA.

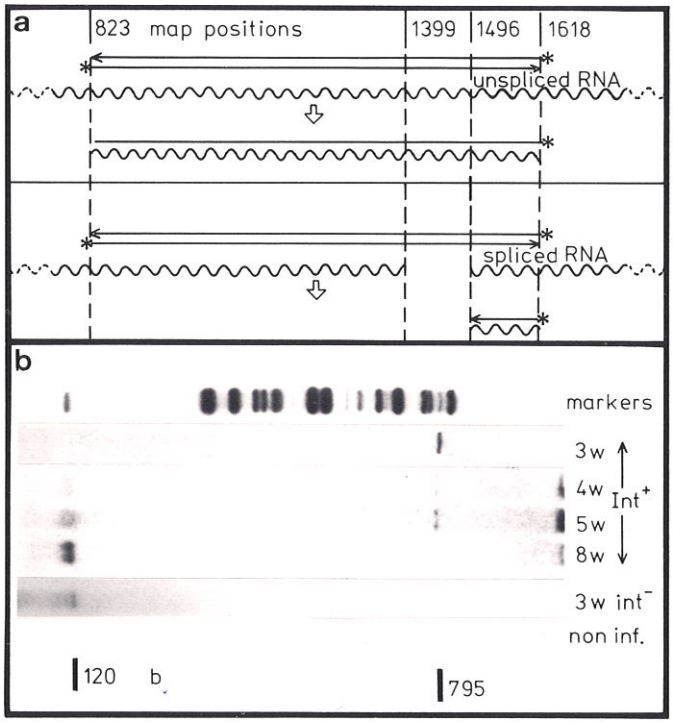

Fig. 3. (a) S1 nuclease mapping of CaMV RNA with and without intron. Overview of the experiment showing the *Cla*I (823)/*Bam*HI (1818) 5'-labelled DNA restriction fragment of the CaMV intron clone, unspliced and spliced RNA and the DNA/RNA hybrid fragments expected after annealing and S1 nuclease digestion. **(b)** First line: bacteriophage M13 RF DNA, restricted with *Hin*fI, as size markers. Following lines: experiments with RNA isolated from plants infected with the intron clone 3, 4, 5 and 8 weeks after inoculation; from plants infected with the intron-less clone 3 weeks after inoculation; and from uninfected plants.

3). For unspliced RNA a DNA/RNA fragment of the size of the original restriction fragment (795 bp) should be found and for the spliced version a 120-bp fragment (Figure 3a). Plants infected with the intron-containing construct do in fact show mainly unspliced RNA early after inoculation and a mixture of spliced and unspliced molecules at later times (Figure 3b). Control experiments performed with an intronless inoculum showed only the short DNA – RNA hybrid and no signals if RNA from uninfected plants was used. This result clearly favours the 'inefficient splicing' explanation for the slow accumulation of intronless CaMV DNA in infected cells.

Inefficient splicing could have several causes, none of which deserves priority at the present stage. (i) The splice signals of the intervening sequences of a *Leguminoseae* gene are not efficiently recognized in the CaMV host, a *Crucifereae* plant. (ii) The gene which served as intervening sequence donor, namely the soybean leghemoglobin gene, is a tightly regulated one, depending on the stimuli exerted by the nitrogen-fixing symbiotic organism (Brisson and Verma, 1982). Splicing in the natural environment might be developmentally regulated as well. CaMV-infected brassica leaves might be a poor substitute for the roots, where leghemoglobin is usually expressed. (iii) Viral RNA might have its special rules: the very long, possibly polycistronic molecule (Dixon and Hohn, 1984) might have special means to avoid the splicing machinery by compartmentalization, by other means of special resistance or by simply exhausting the splicing machinery. (iv) The soybean leghemoglobin gene has three intervening sequences, the efficiency of splicing of which might depend on a complete gene structure. This might not be available any more in the truncated CaMV – soybean leghemoglobin construction. In addition, the sequence information contained in the second intron plus adjacent sequences might not be sufficient for its efficient removal, although data on the splicing mechanism and conserved intron sequences make this rather unlikely (Ruskin *et al.*, 1984; Padgett *et al.*, 1984; Keller and Noon, 1984).

The CaMV reverse transcription mode of replication has its counterparts in animal retrovirus replication. In these animal viruses, artificially introduced DNA (Sorge and Hughes, 1982; Shimotohno and Temin, 1982) as well as biologically incorporated genomic DNA (Bishop, 1981) is spliced out. Also there, artificially introduced splice sequences are removed rather inefficiently (Sorge and Hughes, 1982; Shimotohno and Temin, 1982). In the animal retroviral case mechanisms must exist that ensure the proper separation and use of genomic and mRNA since it is known that viral messengers have to be spliced. The fact that CaMV RNA is spliced so slowly might imply that also in this case a protection mechanism may exist for genomic RNA. Whether it might have to be protected from specific splicing of viral messenger molecules (if that exists at all) or of the general screen of RNAs for spliceable sequences remains even more speculative.

Whatever the reason for the splicing event being so slow, constructs such as the presently used ones serve as models for analysis of plant splice signals. The results from the experiment suggest that genomic sequences cloned in CaMV as a vector should not contain introns or else they could be lost. On the other hand, this procedure might be used to recover intronless genomic sequences.

Materials and methods

CaMV strain CM 4184 with a *Sma*I linker replacing ORF II was provided by B.Gronenborn, and the soybean leghemoglobin clone by D.P.S.Verma. Constructions were performed according to the protocols described by Maniatis *et al.* (1982). DNA homologies were screened for by using the computer program of Queen and Korn (1984). Sequencing was done from cloned viral DNAs using the Maxam and Gilbert (1980) method.

Isolation of polyadenylated RNA

Total nucleic acids were extracted from CaMV-infected and uninfected control leaves by a modification of the method of Howell and Hull (1978): 40 g of leaf tissue were quickly frozen in liquid nitrogen and immediately ground in a Waring blender with a mixture of 80 ml of GPS buffer (0.2 M glycine, HCl, 0.1 M Na_2HPO_4, 0.6 M NaCl, 1% diethylpyrocarbonate, pH 9.6) and 100 ml phenol.

After low-speed centrifugation the aqueous phase was collected and extracted twice with phenol and three times with ether. Nucleic acids were ethanol precipitated and resuspended in 25 ml of ice-cold 3 M Na acetate (pH 6). The high mol. wt RNA, which does not dissolve in this buffer, was collected by centrifugation and washed twice with 3 M Na acetate. The RNA was dissolved in water, ethanol precipitated and re-dissolved. The solution was made 10 mM in $MgCl_2$ and incubated for 40 min at 20°C with 50 μg/ml DNase (RNase-free, Miles Company). Samples were again phenol extracted and ethanol precipitated. Polyadenylated RNA was separated from non-polyadenylated RNA by two cycles of chromatography on oligo(dT)-cellulose (Pharmacia) and then concentrated by ethanol precipitation.

S1 nuclease mapping of CaMV RNA

This was performed as described by Weaver and Weissmann (1979). The relevant *ClaI/Bam*HI restriction fragment (Figure 3) was isolated from electrophoresis gels and 5′-labelled with γ-ATP and polynucleotide kinase. Fragment and polyadenylated RNA were dissolved in 30 μl of hybridization buffer (0.4 M NaCl, 0.04 M Pipes, 1 mM EDTA, pH 6.4, 80% formamide), incubated for 5 min at 72°C and then rapidly transferred to a water bath of 48°C. After 4 h the mixture was diluted 1:10 with S1 buffer (0.25 M NaCl, 0.001 M $ZnSO_4$, 0.3 M Na acetate, pH 5) and treated with 500 units of S1 nuclease at 37°C for 45 min. After phenol and $CHCl_3$ extractions, ethanol precipitation and heat denaturation the samples were loaded onto 6% acrylamide sequencing gels, electrophoresed and analysed by autoradiography.

Acknowledgements

We appreciate having received a CaMV clone from B.Gronenborn and a soybean leghemoglobin clone from D.P.S.Verma. Discussions with N.Brisson and J.Penswick contributed to our work. We are grateful to colleagues at the FMI for discussions on the manuscript.

References

Bishop,J.M. (1981) *Cell*, **23**, 5–6.
Brisson,N. and Verma,D.P.S. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 4055–4059.
Dixon,L. and Hohn,T. (1984) *EMBO J.*, **3**, 2731–2736.
Dixon,L. and Hohn,T. (1985) In Becker,Y. (ed.), *Recombinant DNA Research and Viruses*. Kluwer-Nijhoff Publishing, Boston, MA, pp. 247–275.
Gardner,R.C. and Shepherd,R.J. (1980) *Virology*, **106**, 159–161.
Guilley,H., Dudley,R.K., Jonard,G., Balazs,E. and Richards,K. (1982) *Cell*, **30**, 763–773.
Guilley,H., Richards,K.E. and Jonard,G. (1983) *EMBO J.*, **2**, 277–282.
Hirochika,H., Takatsuji,H., Ubasawa,A. and Ikeda,J.-E. (1985) *EMBO J.*, **4**, 1673–1680.
Hohn,B. and Hohn,T. (1982) In Kahl,G. and Schell,J. (eds), *Molecular Biology of Plant Tumors*. Academic Press, NY, Vol. **99**, pp. 549–560.
Hohn,T., Hohn,B. and Pfeiffer,P. (1985) *Trends Biochem. Sci.*, **10**, 205–209.
Howell,S.H. (1982) *Annu. Rev. Plant Physiol.*, **33**, 609–650.
Howell,S.H. and Hull,R. (1978) *Virology* **86**, 468–481.
Hull,R. (1984) *CMI/AAB Descriptions of Plant Viruses*, **295**.
Hull,R. and Covey,S.N. (1983) *Trends Biochem. Sci.*, **8**, 119–121.
Keller,E.B. and Noon,W.A. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 7417–7420.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, NY.
Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499–560.
Padgett,R.A., Konarska,M.M., Grabowski,P.J., Hardy,S.F. and Sharp,P.A. (1984) *Science*, **225**, 898–903.
Pfeiffer,P. and Hohn,T. (1983) *Cell*, **33**, 781–789.
Queen,C. and Korn,L.J. (1984) *Nucleic Acids Res.*, **12**, 429–436.
Robertson,H.D., Howell,S.H., Zaikin,M. and Maliuberg,R.L. (eds) (1983) *Plant Infectious Agents*. Cold Spring Harbor Laboratory Press, NY.
Ruskin,B., Krainer,A.R., Maniatis,T. and Green,M.R. (1984) *Cell*, **38**, 317–331.
Shimotohno,K. and Temin,H.M. (1982) *Nature*, **299**, 265–268.
Sieg,K., Gronenborn,B., Ciferri,O. and Dure,L. (eds) (1982) *N.A.T.O. Advanced Studies Institute/F.E.B.S. Advanced Course on Structure and Function of Plant Genomes*. Porto Portese, Italy, p. 154.
Sorge,J. and Hughes,S.H. (1982) *J. Mol. Appl. Genet.*, **1**, 547–559.
Takatsuji,H., Hirochika,H., Fukushi,T. and Ikeda,J.-E. (1986) *Nature*, **319**, 240–243.
Weaver,R.F. and Weissmann,C. (1979) *Nucleic Acids Res.*, **7**, 1175–1193.

Received on 18 March 1985; revised on 18 August 1986