Biochem. J. (1984) **224**, 661–666
Printed in Great Britain

Convicilin mRNA from pea (*Pisum sativum* L.) has sequence homology with other legume 7S storage protein mRNA species

Rod CASEY, Claire DOMONEY* and John STANLEY
John Innes Institute, Colney Lane, Norwich NR4 7UH, U.K.

(*Received 16 August/Accepted 2 October 1984*)

1. Nucleotide-sequence analysis of a complementary-DNA clone for convicilin, one of the storage proteins from pea (*Pisum sativum* L.) seeds, shows it to be homologous with the 7S legume seed storage proteins vicilin, conglycinin and phaseolin. 2. Convicilin is more similar to vicilin than to phaseolin or to conglycinin. 3. Significant areas of sequence difference are discussed.

The storage proteins of pea (*Pisum sativum* L.) seeds have been classified, on the basis of sedimentation behaviour and solubility, into 11S (legumin) and 7S groups (Derbyshire *et al.*, 1976). The latter group comprises both vicilin and convicilin. Native vicilin contains a number of polypeptides, ranging from M_r 12000 to 50000 (Gatehouse *et al.*, 1981; Higgins & Spencer, 1981). Convicilin contains only polypeptides of M_r ~70000, although oligomeric proteins containing both M_r ~70000 and ~50000 polypeptides may exist. Vicilin is synthesized as two types of polypeptide, M_r ~50000 and ~47000 (Gatehouse *et al.*, 1981; Higgins & Spencer, 1981); the last is proteolytically cleft at two specific points during seed maturation to give the range of smaller polypeptides found in native vicilin, whereas the M_r ~70000 and M_r ~50000 polypeptides do not seem to be processed in this way (Gatehouse *et al.*, 1981, 1983; Higgins & Spencer, 1981; Domoney & Casey, 1983; Lycett *et al.*, 1983).

Convicilin and vicilin are serologically closely related (Croy *et al.*, 1980), but the mRNA species for the three polypeptides are, however, sufficiently different that a cloned cDNA for any one mRNA will not select either of the other two mRNA species in hybrid-release translation experiments (Domoney & Casey, 1983).

Since no amino-acid-sequence data exist for convicilin, we have sequenced a partial-length cDNA copy of convicilin mRNA and compared its sequence with those of cDNA species for the two vicilin polypeptides and for two other 7S proteins, phaseolin from French bean (*Phaseolus vulgaris*) (Slightom *et al.*, 1983) and conglycinin from soya bean (*Glycine max*) (Schuler *et al.*, 1982*b*).

Abbreviations used: cDNA, complementary DNA.

* To whom correspondence and reprint requests should be addressed.

Materials and methods

The construction and characterization of the plasmid pCD59 has been described elsewhere (Domoney & Casey, 1983). From the restriction map of pCD59 (Domoney & Casey, 1983), internal restriction-endonuclease-*Pst*I and -*Bgl*II sites were utilized for cloning into bacteriophage vector M13 mp9 (see Fig. 1). *Sal*PI (*Pst*I) and *Bgl*II restriction fragments of pCD59 were therefore ligated into *Pst*I- and *Bam*HI-linearized replicative forms of M13 mp9 respectively (Messing & Vieira, 1982). Recombinant phage were identified by the *lac* complementation assay. Bacteriophage isolation, DNA extraction (Messing & Vieira, 1982) and DNA sequence analysis of single-stranded bacteriophage DNA (Sanger *et al.*, 1977) were carried out as described by Duckworth *et al.* (1981) with a synthetic 17-residue oligonucleotide as primer.

Results and discussion

The sequence of the pCD59 insert and the derived partial convicilin amino acid sequence are shown in Fig. 1. With the exception of the first six nucleotides, the entire insert was sequenced on both strands and most of the sequences were determined twice.

The sequence was compared with cDNA sequences for phaseolin, conglycinin and the M_r ~47000 and ~50000 vicilin polypeptides. In these comparisons the numbering of the phaseolin cDNA sequence is as given by Slightom *et al.*

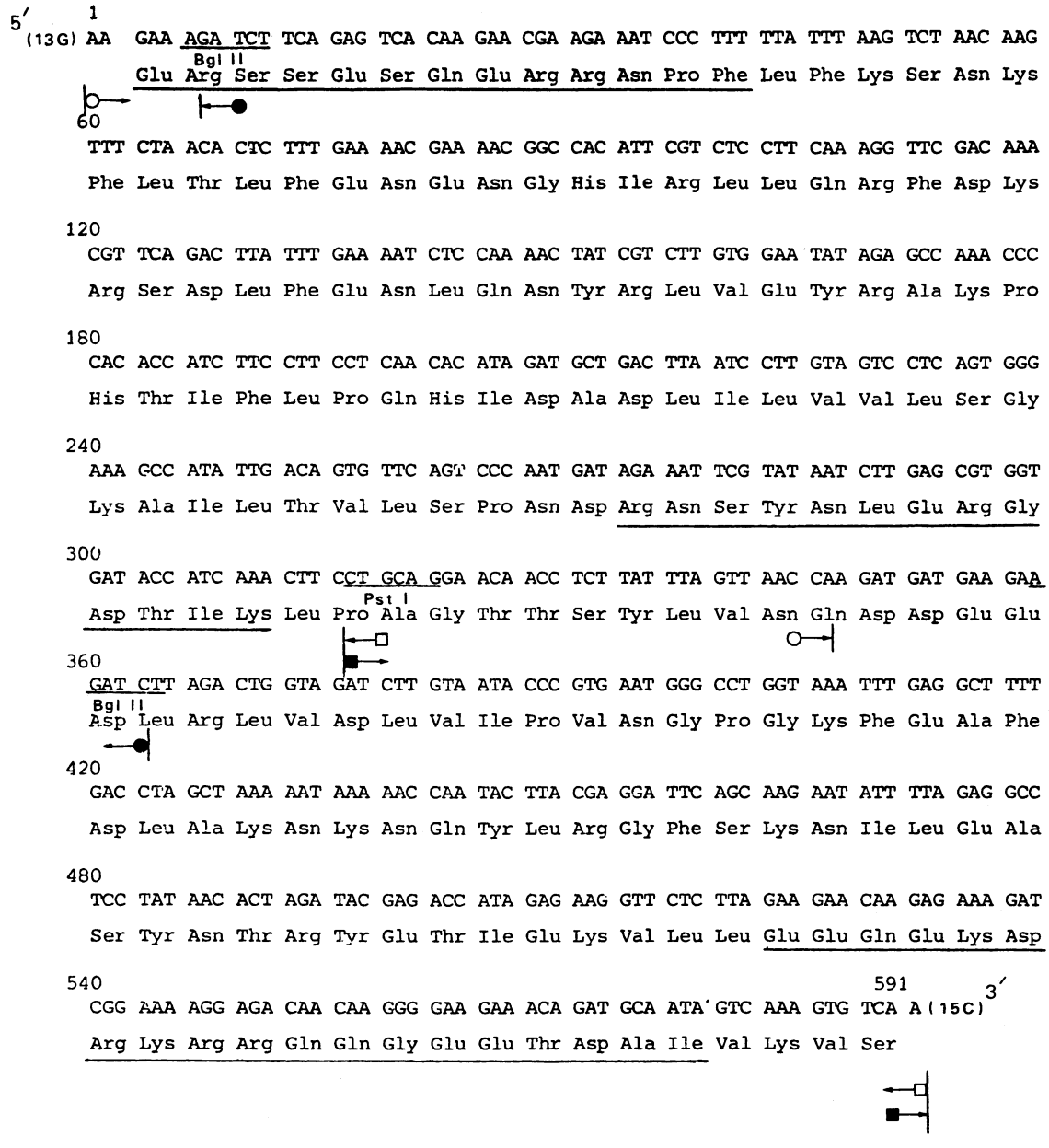

Fig. 1. Nucleotide sequence of the cDNA insert of pCD59 written as a single strand corresponding to the mRNA sense, and the derived partial amino acid sequence of convicilin

The underlining denotes the regions considered in more detail in the present paper. Restriction-endonuclease sites relevant to the strategy adopted for sequencing the cDNA insert are shown beneath the sequence, whereas the extent and direction of sequence determination are indicated beneath the sequence by the arrowed symbols. ○, □, ■, *Pst*I restriction fragments; ●, *Bgl*II restriction fragment.

(1983) for the gene λ 177-4 and the cDNA 31; for the vicilin M_r ~47000 cDNA the numbering starts at the 5′-end of clone pDUB7 (Lycett *et al.*, 1983), for the vicilin M_r ~50000 polypeptide cDNA it starts at the 5′-end of clone pDUB12 (Lycett *et al.*, 1983), and for conglycinin it starts at the 5′-end of clone GMCα' 32 (Schuler *et al.*, 1982*a*,*b*). Only small regions of the vicilin M_r ~50000 polypeptide cDNA and conglycinin cDNA can be compared with pCD59, but exten-

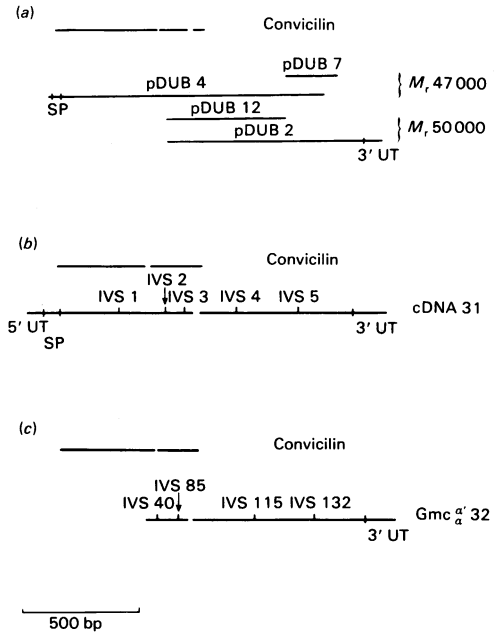

Fig. 2. Regions of the cDNAs for (a) vicilin polypeptides, (b) phaseolin and (c) conglycinin that correspond to pCD59
Abbreviations used: SP, signal peptide; 3′-UT and 5′-UT, 3′ and 5′ untranslated mRNA sequences; IVS1, etc., positions of introns in corresponding genomic sequences; bp, base-pairs. The breaks in the lines indicate regions where there is a deletion relative to the other sequence. For numbering of cDNA clones, see Lycett *et al.* (1983), Schuler *et al.* (1982*b*) and Slightom *et al.* (1983).

sive comparison can be made with phaseolin (the only full-length cDNA) and the M_r ~47000 vicilin polypeptide cDNA (Figs. 2*a* and 2*b*).

The sequence comparison (Fig. 3) clearly shows that convicilin mRNA has extensive regions of homology with phaseolin and the M_r ~47000 vicilin polypeptide mRNA species; similarity to conglycinin and the M_r ~50000 vicilin polypeptide mRNA species was also observed. Where differences in predicted amino acid sequence do occur, they tend to be conservative. There are, however, a number of regions where convicilin differs appreciably from one or more of the other sequences; these are underlined in Figs. 1 and 3 and discussed in more detail below.

Nucleotides 3–41

This region (Fig. 4) includes the signal peptides for the vicilin M_r ~47000 polypeptide (Lycett *et al.*, 1983) and for phaseolin (Murai *et al.*, 1983). Convicilin contains an Arg-Arg sequence that is absent from the other two. It may be significant that convicilin has two polar amino acids in the region corresponding to the (hydrophobic) signal peptide sequences of vicilin and phaseolin; since the convicilin polypeptide is appreciably longer than those of vicilin or phaseolin, it is possible that convicilin does not have a signal peptide in this region, but is extended at the *N*-terminus relative to the others.

Nucleotides 276–311

The sequences of vicilin and convicilin mRNA species are very similar (Fig. 5), but there is only slight resemblance to phaseolin mRNA in this region; phaseolin mRNA contains an extra 18 nucleotides.

Nucleotides 522–578

This region includes the putative $\alpha\beta$ processing site (Fig. 6) (Gatehouse *et al.*, 1983; Lycett *et al.*, 1983), one of the two positions where the M_r ~47000 vicilin polypeptide is thought to be internally hydrolysed. None of convicilin, phaseolin, conglycinin or the vicilin M_r ~50000 polypeptides is processed in this way. The notable features of the comparison are the extensive deletions in this region within the convicilin, phaseolin and conglycinin sequence and the extremely polar nature of the region. Such a region of high polarity would be difficult to accommodate in the interior of a protein and is likely to be on the surface of the native molecule, a supposition that is consistent with the proposal (Lycett *et al.*, 1983) that the sites of proteolysis of vicilin must be near to the surface of the folded protein. Confirmation of this must await the determination of the three-dimensional structure of this region of the vicilin molecule.

The sequence information presented here clearly shows that convicilin is homologous with vicilin, phaseolin and conglycinin and it is likely that it shares common ancestry with them. Although the convicilin sequence only differs to any great extent from the vicilin M_r ~47000 polypeptide sequence in two regions, the overall differences are sufficient to prevent selection of the vicilin mRNA by the convicilin cDNA clone in hybrid-selection experiments (Domoney & Casey, 1983). Convicilin is much more similar to vicilin than to phaseolin and conglycinin over the sequence examined; the data support the possibility that divergence of genes for phaseolin, conglycinin and the pea 7S proteins occurred before that of the convicilin and vicilin genes.

Some of this work was carried out with the aid of a research award from Agrigenetics Corporation, Boulder CO, U.S.A., to whom we are very grateful.

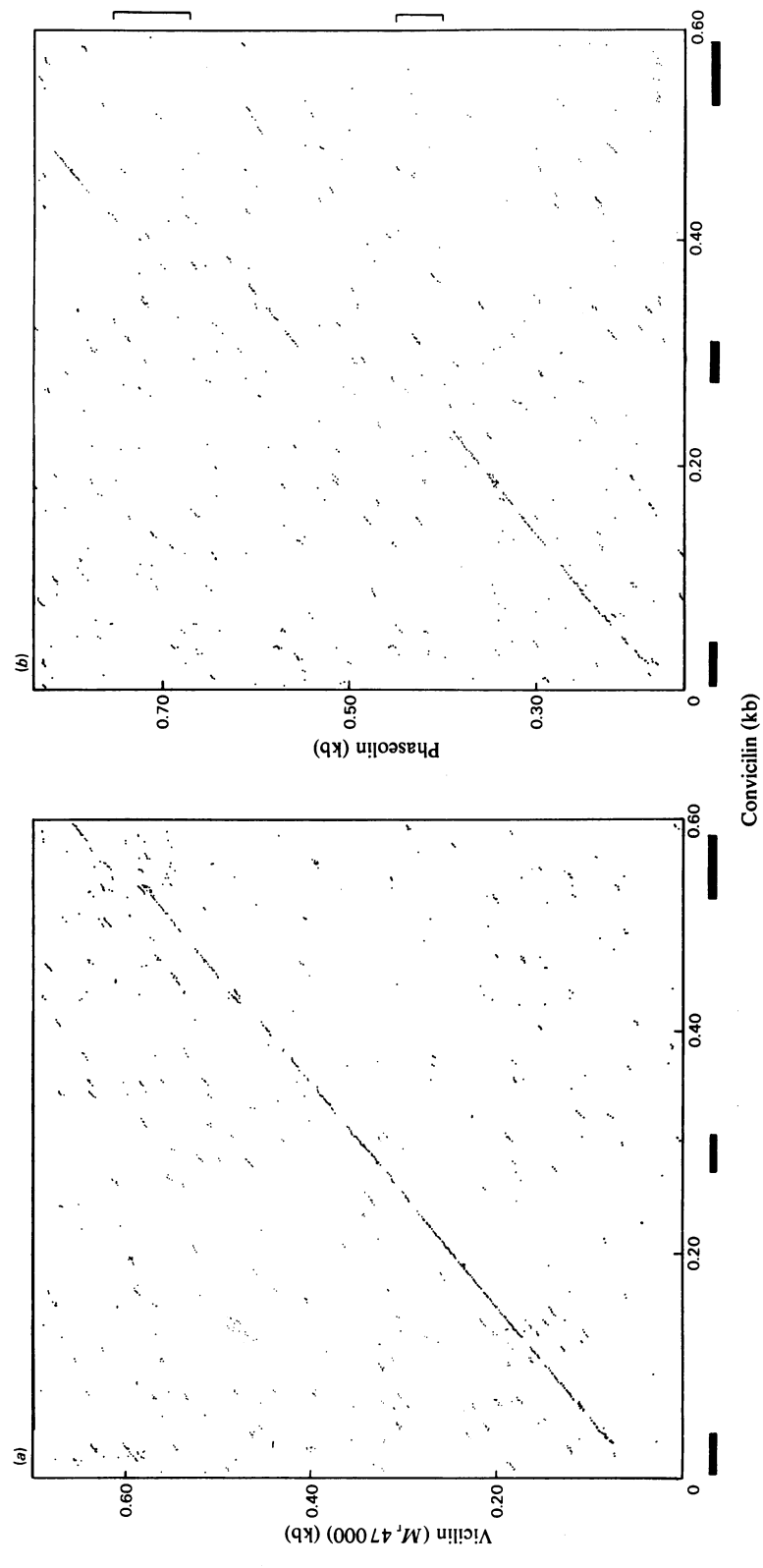

Fig. 3. Homology-matrix comparison of the convicilin cDNA with the vicilin M_r ~47000 polypeptide cDNA and the phaseolin gene λ 177.4
Sequence comparisons were made on a VAX computer by using the DIAGON program (Staden, 1982) with a span of 11 and a score at the 8% level of expectation. The diagonal of the plot represents perfect homology of two sequences; diagonal lines of dots to either side of the centre diagonal represent homologous sets of nucleotides that are aligned by nucleotide deletions or insertions (such as introns). The position of introns in the phaseolin gene are indicated by brackets, and the areas of convicilin sequence discussed in the present paper are indicated by lines beneath the abscissae. Abbreviations used: kb, kilobases.

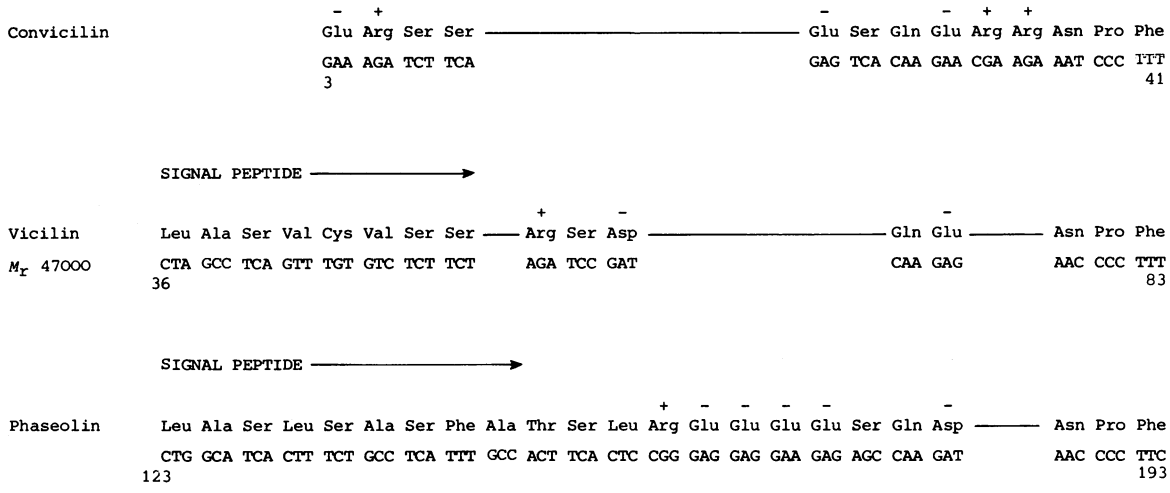

Fig. 4. Sequence differences between convicilin and other storage-protein mRNA species: nucleotides 3–41

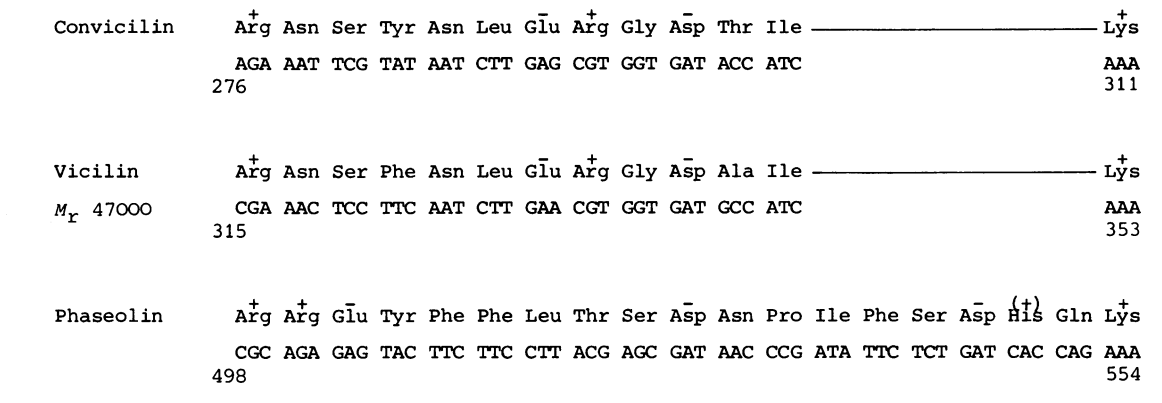

Fig. 5. Sequence differences between convicilin and other storage-protein mRNA species: nucleotides 276–311

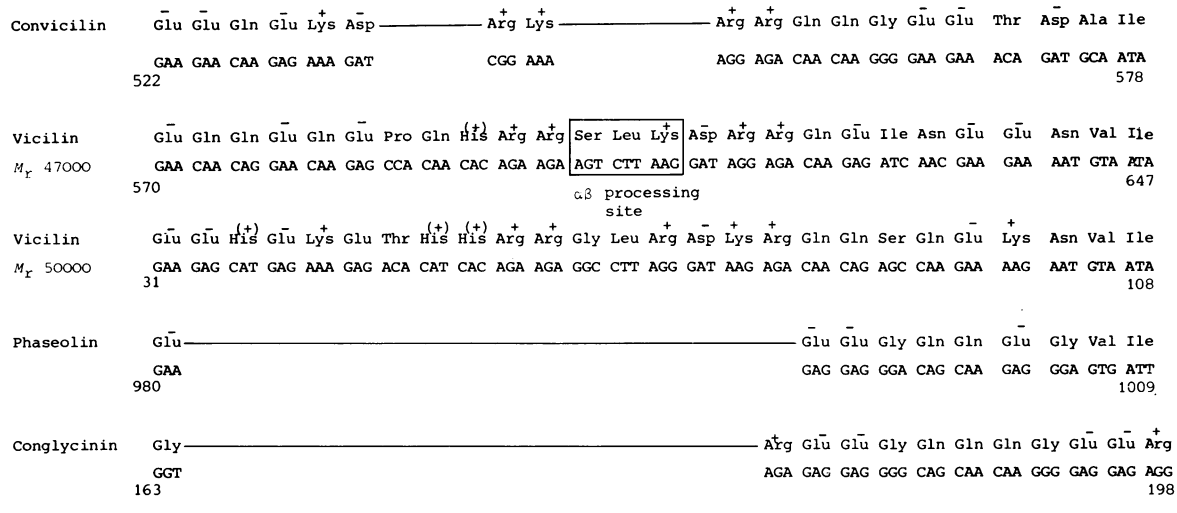

Fig. 6. Sequence differences between convicilin and other storage-protein mRNA species: nucleotides 522–578

References

Croy, R. R. D., Gatehouse, J. A., Tyler, M. & Boulter, D. (1980) *Biochem. J.* **191**, 509–516

Derbyshire, E., Wright, D. J. & Boulter, D. (1976) *Phytochemistry* **15**, 3–24

Domoney, C. & Casey, R. (1983) *Planta* **159**, 446–453

Duckworth, M. L., Gait, M. J., Goelet, P., Hong, G. F., Singh, M. & Titmas, R. C. (1981) *Nucleic Acids Res.* **9**, 1691–1706

Gatehouse, J. A., Croy, R. R. D., Morton, H., Tyler, M. & Boulter, D. (1981) *Eur. J. Biochem.* **118**, 627–633

Gatehouse, J. A., Lycett, G. W., Delauney, A. J., Croy, R. R. D. & Boulter, D. (1983) *Biochem. J.* **212**, 427–432

Higgins, T. J. V. & Spencer, D. (1981) *Plant Physiol.* **67**, 205–211

Lycett, G. W., Delauney, A. J., Gatehouse, J. A., Gilroy, J., Croy, R. R. D. & Boulter, D. (1983) *Nucleic Acids Res.* **11**, 2367–2380

Messing, J. & Vieira, J. (1982) *Gene* **19**, 268–276

Murai, N., Sutton, D. W., Murray, M. G., Slightom, J. L., Merlo, D. J., Reichert, N. A., Sengupta-Gopalan, C., Stock, C. A., Barker, R. F., Kemp, J. D. & Hall, T. C. (1983) *Science* **222**, 476–482

Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467

Schuler, M. A., Schmitt, E. S. & Beachy, R. N. (1982*a*) *Nucleic Acids Res.* **10**, 8225–8244

Schuler, M. A., Ladin, B. F., Pollaco, J. C., Freyer, G. & Beachy, R. N. (1982*b*) *Nucleic Acids Res.* **10**, 8245–8261

Slightom, J. L., Sun, S. M. & Hall, T. C. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 1879–1901

Staden, R. (1982) *Nucleic Acids Res.* **10**, 2951–2961