the return to basal glucose levels were similar for the transgenic and control animals. In addition, intravenous (i.v.) administration of glucagon increased serum glucose levels by ~50% within 15 min in both transgenic and control mice. Taken together, these results strongly suggest that serum glucose levels were appropriately modulated in the transgenic mice. The weights of the transgenic mice, growth rates, feeding behaviour, reproductive capability and longevity appeared normal.

The role of human insulin in the regulation of blood glucose levels in transgenic mice was investigated by performing a glucose tolerance test on transgenic and control mice (Fig. 3*b*). No human C-peptide was detected in the sera of fasting transgenic mice, but within 10 min of i.p. administration of glucose, human C-peptide appeared in the serum, and peak levels were attained within ~20 min. By 45 min post-glucose, human C-peptide levels fell to values approaching the pre-stimulation or basal level. This pattern of human C-peptide expression correlates closely with the glucose tolerance curves presented above, and suggests that serum human insulin levels were being appropriately regulated by glucose. The control mice did not express any detectable human C-peptide, indicating that the human gene must have been the source of the human C-peptide in the transgenic animals.

Insulin is regulated by several other factors, including amino acids and certain pharmacological agents. An i.v. amino-acid infusion test was performed on fasting transgenic and control mice and human C-peptide levels in the serum were determined. Peak human C-peptide levels were seen within 5 min of amino-acid infusion and declined gradually over the next 40 min (Fig. 3*c*). Similarly, serum human C-peptide levels responded to tolbutamide, a sulphonylurea derivative known to promote insulin release[10] (Fig. 3*d*). Within 20 min of i.v. tolbutamide administration, serum human C-peptide levels peaked, then decreased rapidly over the next 10 min. Tolbutamide has been used clinically to diagnose insulinomas[11] because in normal subjects serum insulin (or C-peptide) levels rapidly return to normal from their tolbutamide-induced peak, but in insulinoma patients elevated insulin levels persist. That the transgenic mice quickly regained basal serum human C-peptide levels supports the conclusion that their insulin expression was tightly regulated.

We have demonstrated that the human insulin gene is expressed in the pancreas of transgenic mice. Cell-type- and tissue-specific expression of human[12] and rat[12–14] insulin genes has been documented in two other laboratories. A 230-base-pair (bp) region (from −103 bp to −333 bp with respect to the transcriptional start site) of the rat insulin I promoter was reported to be sufficient to allow tissue-specific expression of insulin/chloramphenicol acetyltransferase fusion genes in a hamster pancreatic cell line[12,13]. Similarly, a rat insulin II/simian virus 40 large-T antigen fusion gene has been reported to cause the development of islet cell tumours in transgenic mice[14]. As both of these studies used fusion genes, the regulation of circulating human insulin could not be studied.

Serum insulin levels are regulated by glucose, amino acids, proteins and drugs such as the sulphonylurea derivatives. The human insulin gene in these transgenic mice is regulated appropriately by all of these agents, and serum glucose homeostasis is normal. These transgenic animals can therefore now be used to study several critical aspects of the physiological regulation of insulin gene expression, including the mechanisms controlling serum insulin and total β-cell insulin levels. Because at least one additional insulin gene is being expressed in the transgenic mice and total insulin RNA and protein levels are approximately the same as in control mice, the question of dosage compensation can be investigated. Moreover, our tolbutamide results indicate that drugs thought to affect human insulin metabolism can now be tested in an *in vivo* animal system. In a more general sense, the *in vivo* effects of various pharmacological agents on human gene expression and protein function can therefore be evaluated in a non-human setting.

Finally, it is noteworthy that a 12.5-kb DNA fragment contains sufficient information for the appropriate physiological regulation of insulin levels in these transgenic mice. The organism's ability to modulate foreign DNA sequences and proteins on a minute to minute basis clearly has important implications for both molecular biology and gene therapy.

We thank Dr Tom Wagner for instruction in microinjection, Victoria Roman for technical assistance and Patrick Mattoon for animal care. The human C-peptide assay was the gift of Dr H. H. Schoene and Behringwerke, AG. This work was supported by grants from Hoechst AG and the NIH (AM-07055).

Received 17 December 1985; accepted 18 March 1986.

1. Steiner, D. F. & Tager, H. S. in *Endocrinology* Vol. 2 (eds DeGroot, L. J. *et al.*) 921–934 (Grune & Stratton, New York, 1979).
2. Rubenstein, A. H. in *Endocrinology* Vol. 2 (eds DeGroot, L. J. *et al.*) 951–957 (Grune & Stratton, New York, 1979).
3. Tager, H. *et al. Nature* **281**, 122–125 (1979).
4. Bell, G. I. *et al. Nature* **284**, 26–32 (1980).
5. Lally, P. A. & Chirgwin, J. M. *Cytogenet. cell. Genet.* **37**, 515 (1984).
6. Cordell, B. *et al. Cell* **18**, 533–543 (1979).
7. Lomedico, P. *et al. Cell* **18**, 545–558 (1979).
8. Moskalewski, S. *Gen. comp. Endocr.* **5**, 342–353 (1965).
9. Lacy, P. E. & Kostianovsky, M. *Diabetes* **16**, 35–39 (1967).
10. Ganda, O. P. *et al. Diabetes* **24**, 354–361 (1975).
11. Fajans, S. S. & Conn, J. W. *J. Lab. clin. Med.* **54**, 811 (1959).
12. Walker, M. D., Edlund, T., Boulet, A. M. & Rutter, W. J. *Nature* **306**, 557–561 (1983).
13. Edlund, T., Walker, M. D., Barr, P. J. & Rutter, W. J. *Science* **230**, 912–916 (1985).
14. Hanahan, D. *Nature* **315**, 115–122 (1985).
15. Bell, G. I. *et al. Nature* **282**, 525–527 (1979).
16. Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
17. Wagner, T. E. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 6376–6380 (1981).
18. Rogol, A. D. *et al. Pediat. Res.* **19**, 489–492 (1985).
19. Southern, E. M. *J. molec. Biol.* **98**, 503–517 (1975).
20. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. *Biochemistry* **18**, 5294–5299 (1979).
21. Sobel, R. A., Blanchette, B. W., Bhan, A. B. & Colvin, R. B. *J. Immun.* **132**, 2402–2403 (1984).
22. Larkins, R. G. *Diabetes* **22**, 251–255 (1973).

Genetic recombination between RNA components of a multipartite plant virus

Jozef J. Bujarski & Paul Kaesberg

Biophysics Laboratory and Department of Biochemistry, University of Wisconsin-Madison, Madison, Wisconsin 53705, USA

Genetic recombination of DNA is one of the fundamental mechanisms underlying the evolution of DNA-based organisms and results in their diversity and adaptability. The importance of the role of recombination is far less evident for the RNA-based genomes that occur in most plant viruses and in many animal viruses. RNA recombination has been shown to promote the evolutionary variation of picornaviruses[1–4], it is involved in the creation of defective interfering (DI) RNAs of positive- and negative-strand viruses[5–9] and is implicated in the synthesis of the messenger RNAs of influenza virus[10] and coronavirus[11]. However, RNA recombination has not been found to date in viruses that infect plants. In fact, the lack of DI RNAs and the inability to demonstrate recombination in mixedly infected plants has been regarded as evidence that plants do not support recombination of viral RNAs. Here we provide the first molecular evidence for recombination of plant viral RNA. For brome mosaic virus (BMV), a plus-stranded, tripartite-genome virus of monocots, we show that a deletion in the 3′ end region of a single BMV RNA genomic component can be repaired during the development of infection by recombination with the homologous region of either of the two remaining wild-type BMV RNA components. This result clearly shows that plant viruses have available powerful recombinatory mechanisms that previously were thought to exist only in animal hosts, thus they are able to adapt and diversify in a manner comparable to animal viruses. Moreover, our observation suggests an increased versatility of viruses for use as vectors in introducing new genes into plants.

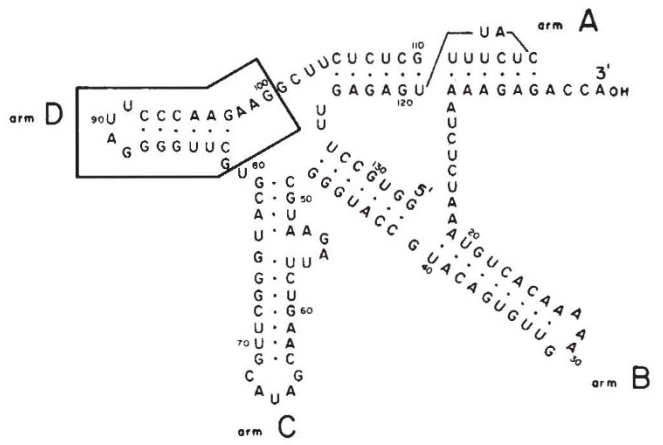

Fig. 1 Schematic representation of the postulated conformation of the tRNA-like structure at the 3′ end of BMV RNA3 (after Rietveld *et al.*[14]). The open box indicates the location of the 20-nucleotide deletion (deletion m4) so that the entire D arm is absent from the resulting structure. This deletion, located between nucleotides 81 and 100, was originally introduced into a cDNA clone comprising the last 200 bases of BMV RNA3 as described in ref. 19 and subsequently transferred into the full-length cDNA clone of BMV RNA3 as described in ref. 21.

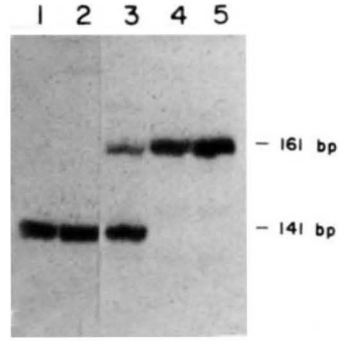

Fig. 2 Analysis by polyacrylamide gel electrophoresis for the presence of the m4 deletion in a 3′-terminal fragment released from progeny BMV RNA3 by limited digestion with RNase T_1. Lanes 5 and 1, standard T_1 fragments obtained from wild-type (161 base pairs (bp) long[23]) or m4-type (141 bp) transcript RNA3, respectively. Lanes 2, 3 and 4, T_1 fragments from progeny BMV RNA3 isolated from the virus at 3, 10 and 16 days post-inoculation, respectively.

Methods. Barley seedlings were inoculated with a mixture of transcribed wild-type BMV RNAs 1 and 2 and the m4 RNA3 following the procedure described in ref. 20. Progeny viral RNA was isolated from virus extracted from 0.1 g of leaves of a single infected plant, as described elsewhere[27]. The BMV RNA3 component was separated by electrophoresis on a low-melting point 1% agarose gel and a 3′-terminal fragment was released from the purified RNA3 by limited digestion with RNase T_1 as described elsewhere[23]. To generate standard 161- and 141-bp T_1 fragments, both wild-type and m4 RNA3 were synthesized by transcription *in vitro* as described in ref. 28, and the fragments were released from transcripts with RNase T_1. After labelling of the 3′ ends with ^{32}P-pCp and T_4 RNA ligase[29], the released RNA fragments were analysed on a 6% polyacrylamide/7 M urea sequencing gel and their positions were determined by autoradiography.

To study recombination between the individual BMV RNAs, we used wild-type RNA1 and RNA2 together with an engineered BMV RNA3 in which a sequence located in its 3′-proximal noncoding region had been deleted. The 3′-noncoding region for each of the three BMV RNAs extends for ~300 nucleotides, the last 200 of which are similar in sequence but not identical in the three RNAs[12]. Their 3′-proximal 134 nucleotides are folded into a transfer RNA-like conformation[13,14], which is involved in initiation of BMV RNA replication[15,16], aminoacylation[17] and adenylation[18]. The RNA3 mutant, designated m4, had been constructed previously by S_1 nuclease deletion of a 20-base long, stem-and-loop hairpin corresponding to arm D in its tRNA-like structure[19] (bases 81–100; see Fig. 1).

RNAs 1 and 2, transcribed from cDNA clones of wild type (Madison M1 strain)[20], together with transcribed m4-type RNA3, are infectious to barley plants and progeny virus initially has the m4-type 3′ region in both RNA3 and RNA4 (ref. 21). RNA4 is a subgenomic RNA derived during virus infection from the 5′-proximal one-third of negative-strand RNA3 (ref. 22). However, m4 RNA3 and RNA4 accumulated less well than their wild-type counterparts, probably because of modified activity of the m4-type 3′ region, as determined *in vitro*[19,21]. We therefore decided to test the possibility that, after a prolonged infection, the original m4 RNA3 might be outcompeted by a revertant form of RNA3 which had regained the deleted stem-and-loop region as a result of recombination with wild-type RNA1 or RNA2.

Time-course analysis of progeny viral RNA3 species from successively appearing leaves of plants inoculated with transcribed wild-type RNAs 1 and 2 and m4 RNA3 revealed a stepwise decrease of the m4 RNA3 over days 3–16 after inoculation (Fig. 2). The presence of the m4 deletion near the 3′ end in RNAs 3 and 4 was determined by observation of a readily identifiable, 3′-terminal fragment released by partial digestion with ribonuclease T_1 (ref. 23). The disappearance with time of the 141-base-long m4-type fragment expected for m4 RNA3, and the parallel appearance of a 161-base fragment characteristic of wild-type RNA3 (compare lanes 2 and 4 of Fig. 2) indicated a gradual increase of an RNA3 variant that was more similar to wild-type RNA3.

To delineate the emergent RNA more precisely, we examined progeny RNA3 and RNA4 isolated from three individual barley plants on day 15 post-inoculation. In plant I we found that the variant RNA3 and RNA4 each migrated more slowly on agarose gel electrophoresis than m4-type and also wild-type RNAs 3 and 4 (see Fig. 3, lanes 1 and 2), indicating that they are larger than both m4-type and wild-type RNAs 3 and 4. There was no evidence of m4-type or wild-type RNA3 or RNA4. This pattern was unaltered after two passages of the virus through other barley plants. Sequence analysis of RNA3 from the emergent virus (designated mutant A; see Fig. 4 legend) indicated that the region corresponding to the last 267 bases of wild-type RNA3 had been replaced with the last 307 bases of wild-type RNA2. Thus, mutant A RNA3 had the deleted stem-and-loop region restored plus 39 more bases than wild-type RNA3 and 59 more bases than its m4 RNA3 progenitor.

In plant II, viral RNA3 and RNA4 co-migrated with wild-type BMV RNA3 (Fig. 3, lanes 4 and 5). Electrophoresis on a higher-resolution gel showed that the RNA3 was composed of at least two bands, the lower band containing most of the material (Fig. 3, lane 6, bands 3a and 3b). In plant III, gel electrophoresis revealed two bands each in the regions of RNA3 and RNA4 (compare lanes 7 and 10 of Fig. 3); a similar pattern was present after the first passage of the virus (lane 8), whereas after the second passage the upper band was absent (lane 9). Analysis on a higher resolution gel indicated that the lower band in the region of RNA3, present as a single band in lane 9, in fact consisted of at least two closely migrating bands (lane 11, bands 3c and 3d).

To identify the species present in RNA3 preparations obtained from plants II and III, a complementary DNA library was created and individual clones were sequenced in their 3′-noncoding regions, as described in Fig. 4 legend. Three sequences, corresponding to mutants designated B, C and D, were found in both libraries. In addition, a fourth sequence, corresponding to mutant E, was identified in the library of plant III. While four clones of mutant B, three clones of mutant C, eight of mutant D and one of mutant E were identified, cDNA clones corresponding to neither m4-type RNA3 nor the strictly wild-

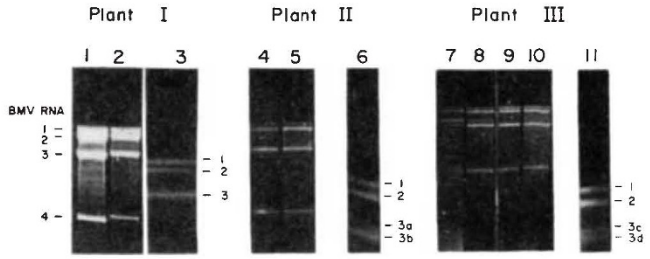

Fig. 3 Electrophoretic analysis of progeny BMV RNA extracted from three independently inoculated barley plants. Barley leaves were infected with transcribed wild-type RNAs 1 and 2 and m4 RNA3. Plants I and II: BMV RNA was extracted from leaves showing symptoms of infection, on day 15 after inoculation, and was analysed on 1.2% agarose gels (either 10 cm long, lanes 2 and 5; or 30 cm long, lanes 3 and 6). Plant III: viral RNA was obtained from leaves 15 days after inoculation with transcribed BMV RNAs (lane 7) as well as on day 15 after the first and second passages through barley (lanes 8 and 9, respectively; 10-cm gels). Lane 11, RNA from lane 9 analysed on a 30-cm gel. Lanes 1, 4 and 10, wild-type BMV RNA.

type 3′ end sequence of BMV RNA3 were found. This last observation tends to exclude the possibility that a causal infection with wild-type BMV RNA3 contaminated the experiments.

The sequences of RNA3 in mutants B, C and D revealed recombination between RNA3 and RNA1 or RNA2 at a site in their homologous 3′ end region. As these RNAs differ in only a small number of bases in this region (see Fig. 4 legend), the exact site of putative recombination could not be specified. We can conclude, however, that mutant B is a combination of RNA3 and RNA1 somewhere between bases 130 and 101, mutant C combines the RNA3 and RNA1 sequences somewhere between bases 175 and 131, and mutant D combines RNA3 and RNA2 somewhere between bases 205 and 101. Similar to mutant A, mutant E had its deleted stem-and-loop region restored and was larger than wild-type RNA3. Sequence analysis showed that the region corresponding to the last 206 bases of the wild-type RNA3 had been replaced by 215 3′-terminal nucleotides from wild-type RNA2 (see Fig. 4).

The structure of these five mutants suggests strongly that one or more recombination events have occurred between the 3′-noncoding regions (within or outside the homologous sequences) of individual BMV RNAs, resulting in the restoration of the deleted hairpin in progeny RNA3. Clearly, that stem-and-loop region is not essential for infectivity but its presence nevertheless confers a selective advantage. In mutants A and E recombination outside the closely homologous region resulted in emergent noncoding regions larger than wild-type RNA3, indicating that a degree of structural variability is permissible there. In mutants B, C and D the recombination event(s) occurred within the homologous region and resulted in sequences of wild-type lengths only.

The results described here demonstrate that recombination between viral RNAs occurs in plant cells at a relatively high frequency and at various positions on the RNA. As with animal RNA viruses, the mechanism of recombination between the BMV sequences is unknown. Two diverse mechanisms have been suggested for RNA genomes: (1) Enzymatic cutting and re-ligation. (2) A 'copy-choice', during replication, between adventitiously proximal templates. The latter mechanism has been favoured for recombination in picornaviruses[24], negative-strand viruses[5,10] and alpha viruses[7].

The recent discoveries of remarkable sequence similarities among encoded proteins of diverse plant RNA viruses[25,26] suggest that recombination may be a common phenomenon during synthesis of plant viruses. Like other processes of natural mutagenesis, recombination between plant viruses may render the progeny more adaptable to new environments or hosts.

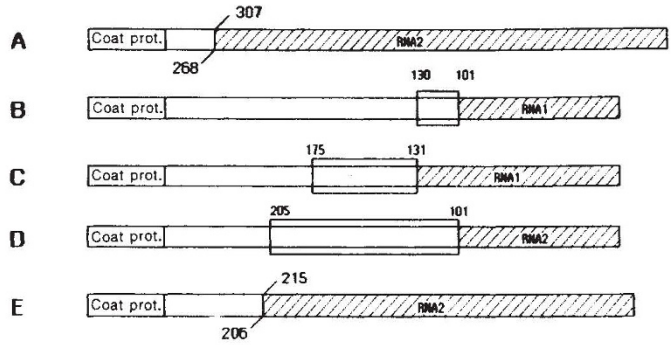

Fig. 4 Schematic representation of BMV RNA3 recombinants identified by sequence analysis of BMV RNA3 preparations obtained from plants I–III. Mutant A was obtained from plant I; mutants B–E were identified in plants II and III. The figure shows the 3′ region of RNA3 for the five mutants downstream to the coat protein open reading frames (ORFs). The open regions adjacent to the ORFs represent wild-type BMV RNA3 sequences. The cross-hatching indicates regions identical to wild-type RNA1 or RNA2, possessing the stem-and-loop region of arm D, and also differing from wild-type RNA3 at the following base positions: RNA1 has distinctive bases at positions 101, 44 and 43 while RNA2 has distinctive bases at positions 101 and 44 and has a point deletion corresponding to base 74 of RNA3 (refs 13, 20). The boxes enclose regions in which the recombination event occurred, assuming a recombination mechanism involving a single event. Sequences of wild-type RNAs 1, 2 and 3 are identical in these regions.
Methods. Mutant A RNA3 was sequenced directly by a chain-temination (dideoxy) technique with RNA-dependent DNA polymerase (reverse transcriptase[30]. Briefly, 2 μg of RNA3, obtained after fractionation on a sucrose gradient[31], was hybridized with 1 μg of either the 15-mer d(GAGATTTTCTCTGGT), complementary to nucleotide positions 1–15, or the 15-mer d(TCTCACAGATCCTCG), complementary to nucleotides 129–143, and these primers were extended with 10 units of reverse transcriptase (Life Science) in the presence of dideoxy nucleoside triphosphates. RNA3 preparations obtained from plants II and III (mutants B–E) were sequenced as cDNA clones in an M13mp19 vector by a chain-termination (dideoxy) technique with a synthetic 17-mer primer (BioLabs) and DNA polymerase large fragment (Klenow)[32,33]. Two cDNA libraries were created for BMV RNA3 preparations obtained either from primary infection in plant II or after a second passage of the virus in plant III. The first cDNA strand was synthesized with reverse transcriptase as described elsewhere[34], using the d(TAGAGATTTTCTCTGGT) primer complementary to the last 17 nucleotides of BMV RNA3. After purification on a 1×20 cm Sepharose CL2B column, the second strand was obtained by means of a replacement reaction with RNase H, *Escherichia coli* DNA polymerase and *E. coli* DNA ligase[35]. After treatment with T_4 DNA polymerase (blunt-ending), the resulting double-stranded cDNA preparation was cloned into a *Sma*I restriction site of M13amp19 vector and the sequence of the corresponding single-stranded M13 DNA templates was determined by a chain-termination procedure using the Klenow fragment of *E. coli* DNA polymerase and a synthetic 17-mer (BioLabs) as primer.

Conversely, RNA recombination between plant viral RNAs may be important for the preservation and stabilization of crucial functional sequences of multipartite genome viruses. Because it facilitates the association of unrelated genes, recombination also represents an attractive candidate for the mechanism of origin of viruses and for the insertion of new genes into existing viruses.

We thank Ola Dzianott-Bujarska for technical assistance, and G. Duke, P. Shaklee and P. Ahlquist for valuable advice with respect to cDNA cloning. This work was supported by NIH research grants AI01466 and AI15342 and Career Award AI21942.

Received 21 January; accepted 18 March 1986.

1. Cooper, P. D. *Virology* **35**, 584–596 (1968).
2. King, A. M. Q., McCahon, D., Slade, W. R. & Newman, J. W. I. *J. Virol.* **41**, 66–77 (1982).

3. Tolskaya, E. A., Romanova, L. A., Kolesnikova, M. S. & Agol, V. I. *Virology* **124**, 121-132 (1983).
4. Emini, E. A., Leibowitz, J., Diamond, D. C., Bonin, J. & Wimmer, E. *Virology* **137**, 74-85 (1984).
5. Lazzarini, R. A., Keene, J. D. & Schubert, M. *Cell* **26**, 145-154 (1981).
6. Jennings, P. A., Finch, J. T., Winter, G. & Robertson, J. S. *Cell* **34**, 619-627 (1983).
7. Monroe, S. S. & Schlesinger, S. *J. Virol.* **49**, 865-872 (1984).
8. Stark, C. & Kennedy, S. I. T. *Virology* **89**, 285-299 (1978).
9. Lai, M. M. C. *et al. J. Virol.* **56**, 449-456 (1985).
10. Plotch, S. J., Bouloy, M., Ulmanen, I. & Krug, R. M. *Cell* **23**, 847-858 (1981).
11. Baric, R. S., Stohlman, S. A., Razavi, M. K. & Lai, M. M. C. *Virus Res.* **3**, 19-33 (1985).
12. Ahlquist, P., Dasgupta, R. & Kaesberg, P. *J. molec. Biol.* **172**, 369-383 (1984).
13. Ahlquist, P., Dasgupta, R. & Kaesberg, P. *Cell* **23**, 183-189 (1981).
14. Rietveld, K., Pleij, W. A. & Bosch, L. *EMBO J.* **2**, 1079-1985 (1983).
15. Ahlquist, P., Bujarski, J. J., Kaesberg, P. & Hall, T. C. *Pl. molec. Biol.* **3**, 37-44 (1984).
16. Miller, W. A., Bujarski, J. J., Dreher, T. W. & Hall, T. C. *J. molec. Biol.* **187**, 537-546 (1986).
17. Hall, T. C. *Int. Rev. Cytol.* **60**, 1-26 (1979).
18. Joshi, R. L., Joshi, S., Chapeville, F. & Haenni, A. L. *EMBO J.* **2**, 1123-1127 (1983).
19. Bujarski, J. J., Dreher, T. W. & Hall, T. C. *Proc. natn. Acad. Sci. U.S.A.* **82**, 5636-5640 (1985).
20. Ahlquist, P., French, R., Janda, M. & Loesch-Fries, L. S. *Proc. natn. Acad. Sci. U.S.A.* **81**, 7066-7070 (1984).
21. Bujarski, J. J., Ahlquist, P., Hall, T. C., Dreher, T. & Kaesberg, P. *EMBO J.* (submitted).
22. Miller, W. A., Dreher, T. W. & Hall, T. C. *Nature* **313**, 68-70 (1985).
23. Bastin, M., Dasgupta, R., Hall, T. C. & Kaesberg, P. *J. molec. Biol.* **103**, 737-745 (1976).
24. Cooper, P. D., Steiner-Pryor, A., Scotti, P. D. & DeLong, D. J. *Gen. Virol.* **23**, 41-49 (1974).
25. Haseloff, J. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 4358-4362 (1984).
26. Cornelissen, B. & Bol, J. *Pl. molec. Biol.* **3**, 379-384 (1984).
27. Shih, D. S., Lane, L. C. & Kaesberg, P. *J. molec. Biol.* **64**, 353-362 (1972).
28. Ahlquist, P. & Janda, M. *Molec. cell. Biol.* **4**, 2876-2882 (1984).
29. England, T. & Uhlenbeck, O. *Nature* **275**, 560-561 (1978).
30. Zimmern, D. & Kaesberg, P. *Proc. natn. Acad. Sci. U.S.A.* **75**, 4257-4261 (1978).
31. Kiberstis, P. A., Loesch-Fries, L. S. & Hall, T. C. *Virology* **112**, 804-808 (1981).
32. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
33. Biggin, M. D., Gibson, T. J. & Hong, G. F. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3963-3965 (1983).
34. Fields, S. & Winter, G. *Cell* **28**, 303-313 (1982).
35. Okayama, H. & Berg, P. *Molec. cell. Biol.* **2**, 161-170 (1982).

A tobacco mosaic virus-induced tobacco protein is homologous to the sweet-tasting protein thaumatin

Ben J. C. Cornelissen, Rob A. M. Hooft van Huijsduijnen & John F. Bol

Department of Biochemistry, State University of Leiden, PO Box 9505, Leiden, The Netherlands

Infection of tobacco plants with tobacco mosaic virus (TMV) results in an increase in the activities of several enzymes and induces the *de novo* synthesis of about 10 proteins that are protease-resistant and soluble at *p*H 3. These proteins accumulate in the intercellular leaf space[1,2]. The appearance of pathogenesis-related (PR) proteins is closely associated with the phenomenon of 'systemic acquired resistance' and it has been suggested that such proteins have an antiviral function[3,4]. Previously, we cloned complementary DNAs to the messenger RNAs for the three smallest PR proteins, PR-1a, -1b and -1c, and these clones were used to show that there is an increase of more than 100-fold in the concentration of PR-1 mRNAs following TMV infection of tobacco[5,6]. Here, we describe the cDNA cloning of another mRNA whose synthesis is induced by TMV infection. Sequencing of the cDNA showed that the encoded protein is highly homologous to thaumatin, the intensely sweet-tasting protein from the fruits of the monocot *Thaumatococcus daniellii* Benth, a West African rainforest shrub. The conservation of a gene encoding a thaumatin-like protein in tobacco suggests that the encoded protein may have a more general function than that of being sweet-tasting.

Poly(A)-containing RNA from TMV-infected Samsun NN tobacco was used as a template to construct a cDNA library; 1,400 transformants were screened by a differential hybridization procedure, using ^{32}P-labelled DNA complementary to poly(A) RNA from healthy or TMV-infected tobacco as probes. Forty transformants were found to correspond to TMV-induced mRNAs and an analysis of their inserts showed that they could be divided into six clusters, with cross-hybridization occurring within but not between the clusters. Details of the cloning and selection procedure will be given elsewhere (R.A.M.H. *et al.*, in preparation).

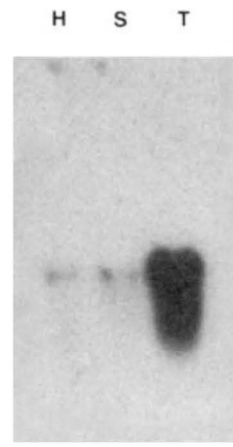

Fig. 1 Northern blot loaded with poly(A) RNA from healthy tobacco (H), tobacco sprayed with salicylic acid (S) and tobacco infected with TMV (T). The blot was probed with ^{32}P-labelled PROB 12. Samsun NN tobacco plants were sprayed with salicylic acid (5 mM, adjusted to *p*H 7) on three successive days before RNA extraction. RNA was extracted from TMV-infected tobacco plants 7 days after inoculation. The method of extraction, Northern blotting and the cloning of PROB 12 are described in ref. 6.

One cluster was represented by a single clone, PROB 12, which was used to probe the Northern blot shown in Fig. 1. The gel used to make the blot was loaded with poly(A) RNA from healthy tobacco (lane H), tobacco sprayed with salicylic acid (lane S), and tobacco infected with TMV (lane T). Salicylic acid is known to induce the synthesis of several PR proteins, notably proteins 1, 2 and N[2]. The mRNA corresponding to PROB 12 occurs at a low level in healthy tobacco, is not induced by treatment with salicylic acid, but is strongly induced by TMV infection. This mRNA is estimated to be 1,000–1,100 nucleotides long. Sequencing studies showed that the insert in PROB 12 is 845 base pairs (bp) long. As no poly(A) tract was found in the cDNA, the 3′ end of the mRNA is probably not represented in the clone. Figure 2 shows that the insert contains an open reading frame for a protein of 226 amino acids, flanked by 5′- and 3′-noncoding regions of 3 and 164 nucleotides, respectively.

The nucleotide sequence of the insert of PROB 12 was compared, by computer analysis, with plant sequences stored in a databank (Genbank). Extensive homology was found with the mRNA for thaumatin, the sweet-tasting protein that occurs in the arils of *Thaumatococcus daniellii* Benth, a West African shrub[7]. At least five different forms of thaumatin (I, II, III, b and c) can be isolated, all of which are almost 100,000 times sweeter than sucrose on a molar basis[8]. The most abundant forms are thaumatin I and II, polypeptides of 207 amino acids each that differ at only five positions[9]. The mature protein is derived from preprothaumatin by removal of an amino-terminal signal peptide of 22 amino acids and an acidic carboxy-terminal extension of 6 amino acids[7]. In Fig. 3, preprothaumatin II is aligned with protein encoded by PROB 12; the amino-acid sequence homology between the two proteins is 65%. In addition, there are several conserved amino-acid changes; when these are taken into account, the homology is more than 70%. The Ala–Ala sequence, which represents the site of cleavage of the signal peptide, is conserved in both proteins. However, except for their hydrophobic nature, there is little similarity between the signal peptides of the two proteins, and the tobacco protein lacks the C-terminal extension of the thaumatin precursor. Because of its homology to thaumatin, we provisionally refer to the PROB 12-encoded protein as a thaumatin-like (or TL) protein of tobacco.

The alignment shown in Fig. 3 suggests that PROB 12 contains the complete coding region for the precursor of the TL protein. A possible polyadenylation signal in TL mRNA (underlined in Fig. 1) is located 80 bp upstream from the 3′ end of the cDNA insert. There may be additional polyadenylation signals in the TL mRNA sequence that are not represented in PROB 12. Three potential polyadenylation signals have been found in thaumatin mRNA[7].

The relative molecular mass (M_r) of the mature TL protein of tobacco is 21,596. It is unknown whether the TL protein