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. 3b). 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. 3c). Similarly, serum human C-peptide levels responded to tolbutamide, a sulphonylurea derivative known to promote insulin release10 (Fig. 3d). 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 insulinas11 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 two different transgenic mice. Cell-type and tissue-specific expression of human13 and rat12-14 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 line15,16. 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 mice14. 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. Steinert, D. F. & Tager, H. S. in Endocrinology Vol. 2 (eds DeGroot, L. J. et al.) 921-934 (Grune & Stratton, New York, 1979).
2. Hubert-Pestel, A. H. in Endocrinology Vol. 2 (eds DeGroot, L. J. et al.) 951-957 (Grune & Stratton, New York, 1979).
3. Tager, H. S. et al. Nature 281, 122-125 (1979).
4. Walker, M. D., Edlund, T., Boulet, A. M. & Rutter, W. J. Nature 306, 557-561 (1983).
5. Edlund, T., Walker, M. D., Barr, P. J. & Rutter, W. J. Science 236, 912-916 (1987).
6. Maniatis, T., Fritsch, E. F. & Sambrook, J. Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory, New York, 1982).
7. Sobel, R. A., Blanchette, B. W., Bhan, A. B. & Colvin, R. B. J. molec. Biol. 215, 115-122 (1985).
8. Southern, E. M. molec. and cell. Genet. 37, 515 (1984).
9. Lacy, P. E. & Kostianovsky, M. J. molec. BioL 16, 35-39 (1967).
10. Ganda, O. P. et al. Diabetes 24, 154-361 (1975).
11. Fajans, S. S. & Conn, J. W. F. J. Lab. clin. Med. 54, 811 (1959).
12. Edlund, T., Edlund, T., Trabulsi, A. M. & Rutter, W. J. Nature 306, 557-561 (1983).
13. Edlund, T., Walker, M. D., Barr, P. J. & Rutter, W. J. Biochemistry 18, 5294-5399 (1979).
14. Lomalico, P. & Kostianovsky, M. J. molec. BioL 18, 352-357 (1979).
15. Maniatis, T., Fritsch, E. F. & Sambrook, J. Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory, New York, 1982).
16. Walker, M. D., Edlund, T., Boulet, A. M. & Rutter, W. J. Biochemistry 18, 5294-5399 (1979).
17. Sobel, R. A., Blanchette, B. W., Bhan, A. B. & Colvin, R. B. J. Immunol 132, 2402-2403 (1984).
18. Larkins, B. G. Diabetes 22, 351-355 (1973).
the last 200 of which are similar in sequence but not identical in the three RNAs12. Their 3'-proximal 134 nucleotides are noncoding region had been deleted. The 3'-noncoding region we used wild-type RNA1 and RNA2 together with an engineered mutation17 and adenylationl8. The RNA3 mutant, designated m4 , for each of the three BMY RNAs extends for -300 nucleotides, its tRNA-like structure'9 (bases 81-100; see Fig. 1).

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 RNAs12. Their 3' proximal 134 nucleotides are folded into a transfer RNA-like conformation13,14 which is involved in initiation of BMV RNA replication15,16, aminoacylation17 and adenylation18. The RNA3 mutant, designated m4 , had been constructed previously by S, nuclease deletion of a 20-base long, stem-and-loop hairpin corresponding to arm D in its tRNA-like structure19 (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 having the m4-type 3' region in both RNA3 and RNA4 (ref. 21). RNA4 is a subgenomic RNA derived during virus infection having the m4-type 3' region in both RNA3 and RNA4 (ref. 21). RNA4 is a subgenomic RNA derived during virus infection having the m4-type 3' region in both RNA3 and RNA4 (ref. 21). RNA4 is a subgenomic RNA derived during virus infection having the m4-type 3' region in both RNA3 and RNA4 (ref. 21). RNA4 is a subgenomic RNA derived during virus infection having the m4-type 3' region in both RNA3 and RNA4 (ref. 21). RNA4 is a subgenomic RNA derived during virus infection having the m4-type 3' region in both RNA3 and RNA4 (ref. 21). RNA4 is a subgenomic RNA derived during virus infection having the m4-type 3' region in both RNA3 and RNA4 (ref. 21). RNA4 is a subgenomic RNA derived during virus infection having the m4-type 3' region in both RNA3 and RNA4 (ref. 21). RNA4 is a subgenomic RNA derived during virus infection having the m4-type 3' region in both RNA3 and RNA4 (ref. 21). RNA4 is a subgenomic RNA derived during virus infection having the m4-type 3' region in both RNA3 and RNA4 (ref. 21). RNA4 is a subgenomic RNA derived during virus infection having the m4-type 3' region in both RNA3 and RNA4 (ref. 21). RNA4 is a subgenomic RNA derived during virus infection having the m4-type 3' region in both RNA3 and RNA4 (ref. 21). RNA4 is a subgenomic RNA derived during virus infection having the m4-type 3' region in both RNA3 and RNA4 (ref. 21). RNA4 is a subgenomic RNA derived during virus infection having the m4-type 3' region in both RNA3 and RNA4 (ref. 21). RNA4 is a subgenomic RNA derived during virus infection having the m4-type 3' region in both RNA3 and RNA4 (ref. 21). RNA4 is a subgenomic RNA derived during virus infection having the m4-type 3' region in both RNA3 and RNA4 (ref. 21). RNA4 is a subgenomic RNA derived during virus infection having the m4-type 3' region in both RNA3 and RNA4 (ref. 21). RNA4 is a subgenomic RNA derived during virus infection having the m4-type 3' region in both RNA3 and RNA4 (ref. 21). RNA4 is a subgenomic RNA derived during virus infection having the m4-type 3' region in both RNA3 and RNA4 (ref. 21). RNA4 is a subgenomic RNA derived during virus infection having the m4-type 3' region in both RNA3 and RNA4 (ref. 21). RNA4 is a subgenomic RNA derived during virus infection having the m4-type 3' region in both RNA3 and RNA4 (ref. 21). RNA4 is a subgenomic RNA derived during virus infection having the m4-type 3' region in both RNA3 and RNA4 (ref. 21). RNA4 is a subgenomic RNA derived during virus infection having the m4-type 3' region in both RNA3 and RNA4 (ref. 21). RNA4 is a subgenomic RNA derived during virus infection having the m4-type 3' region in both RNA3 and RNA4 (ref. 21). RNA4 is a subgenomic RNA derived during virus infection having the m4-type 3' region in both RNA3 and RNA4 (ref. 21). RNA4 is a subgenomic RNA derived during virus infection having the m4-type 3' region in both RNA3 and RNA4 (ref. 21). RNA4 is a subgenomic RNA derived during virus infection having the m4-type 3' region in both RNA3 and RNA4 (ref. 21). RNA4 is a subgenomic RNA derived during virus infection having the m4-type 3' region in both RNA3 and RNA4 (ref. 21).

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 RNAs 3 or 4. 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-
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' non-coding 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 emerging 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 picornaviruses24, negative-strand viruses5,15 and alpha viruses7.

The recent discoveries of remarkable sequence similarities among encoded proteins of diverse plant RNA viruses3,23 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. 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 Dzianot-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. J. Virol. 41, 66-77 (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 Huijsduijen & 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 protein-resistant and soluble at pH 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 function3,4. Previously, we cloned the RNA from healthy or TMV-infected tobacco as probes. Forty thousand clones were isolated, all of which are almost 100,000 times greater than sucrose on a molar basis.5 The most abundant forms are thaumatin I and II, polypeptides of 207 amino acids each that differ at only five positions6. 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 acids7. 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.

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 10. 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 rainforest shrub. At least five different forms of thaumatin (1, II, III, b and c) can be isolated, all of which are almost 100,000 times sweeter than sucrose on a molar basis. The most abundant forms are thaumatin I and II, polypeptides of 207 amino acids each that differ at only five positions. 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. 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. These potential polyadenylation signals have been found in thaumatin mRNA.

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