Sequence of RNA2 of the Helicoverpa armigera stunt virus (Tetraviridae) and bacterial expression of its genes

Terry N. Hanzlik,* Susan J. Dorrian, Karyn N. Johnson, Elizabeth M. Brooks and Karl H. J. Gordon

CSIRO Division of Entomology, Box 1700, Canberra, ACT 2601, Australia

The complete nucleotide sequence of RNA2 of Helicoverpa armigera stunt virus (HaSV), a member of the Tetraviridae, was determined by characterization of cloned cDNA and PCR products and direct sequencing of genomic RNA. The capped, positive sense, single-stranded RNA is 2478 nucleotides in length and has two overlapping open reading frames (ORFs) likely to be cistrons which are situated between terminal non-coding regions of 282 and 168 bases, 5' and 3', respectively. Extensive secondary structure of the RNA strand is indicated, including a tRNA-like structure at the 3' terminus which is the first such structure discerned in an animal virus. The first ORF encodes a 17 kDa PEST protein (p17) of unknown function while the second ORF encodes the 71 kDa coat protein precursor (p71) that is cleaved at an Asn–Phe site into the 64 kDa and 7 kDa coat proteins. The precursor coat protein is 66% identical to that of another tetravirus, the Nudaurelia ω virus, with most of the difference residing in a 165 amino acid region located in the middle of the sequence. Despite the extensive similarity, no serological relationship was observed between the two viruses, suggesting that the dissimilar region is exposed on the capsid exterior. Expression in bacteria of the two RNA2 gene products shows they are likely to be expressed by a leaky scan-through mechanism. Bacterial expression of p71 did not produce virus-like particles while expression of p17 produced large arrays of mostly hollow, hexagonal tube-like structures.

Introduction

The Tetraviridae is a relatively poorly characterized family of viruses: its members have small, single-stranded RNA genomes and have been isolated exclusively from larvae of lepidopteran insects. Little is known about the ecology of these viruses although large scale epizootics that have sharply reduced host populations have been observed (Moore, 1991). Horizontal transmission per os has been demonstrated (Hanzlik et al., 1993) and some form of vertical transmission is suspected (T. N. Hanzlik, unpublished observations). Symptoms of infection range from severe stunting and death to negligible weight loss (Moore, 1991; Reinganum, 1991). As a group, these viruses display a marked tropism for midgut cells of their hosts, a characteristic which may account for the lack of success in growing them in continuous cell lines. In spite of this handicap, the study of tetraviruses continues due to their unique capsid structure and their possible utility for control of agricultural pests.

Structural studies of tetraviruses have centred on the unique T = 4 symmetry (tetra) found in the icosahedral structure of their capsids, the distinguishing feature of the family (Francki et al., 1991). Two members of the group, the Nudaurelia β (NβV) and Nudaurelia ω (NωV) viruses, have been studied in detail by cryoelectron microscopy (Olson et al., 1990) and X-ray crystallography (Cavarelli et al., 1991). Typically 35–40 nm in diameter, the tetravirus capsid was originally thought to have one protein component of 62–66 kDa; however, recent work has shown that the capsid also contains an equimolar amount of a 7–8 kDa protein (Agrawal & Johnson, 1992; Hanzlik et al., 1993). It was also thought that the capsid contained only a monopartite, single-stranded RNA genome of approximately 5.5 kb; however, a subgroup of tetraviruses with bipartite genomes has been characterized. NωV and the Helicoverpa armigera stunt virus (HaSV) were recently shown to have bipartite ssRNA genomes with capped nonpolyadenylated strands of 5.5 kb (RNA1) and 2.5 kb (RNA2) and no 3' blockage (Agrawal & Johnson, 1992; Hanzlik et al., 1993). With the exception of the additional genome component, which represents 30% more genetic information, NωV and HaSV are highly similar to the tetraviruses with monopartite genomes in all other

* Author for correspondence. Fax +61 6 246 4173. e-mail terryh@ento.csiro.au

The nucleotide sequence data reported in this paper have been deposited in GenBank under the accession number L37299.

0001-2855 © 1995 SGM
measured physicochemical characteristics, including T = 4 symmetry of their capsids.

What little is known about the replication strategy of tetraviruses has been surmised from in vitro translation studies and partial sequencing of the genome; most detail is known for the two viruses with bipartite genomes. Hanzlik et al. (1993) reported variable data from translations of RNA1 of HaSV in different systems, while translations of RNA2 showed a major 24 kDa product and a minor 70 kDa product. No time-dependent processing of polyproteins was noted. Agrawal & Johnson (1992) reported the sequence of RNA2 of NoV during the course of their structural studies on the capsid of this virus. Their analysis concluded that this strand contained only one gene for the coat protein’s 70 kDa precursor. The precursor was shown to be cleaved at an Asp/Phe site producing the 62 kDa and 8 kDa proteins which make up the NoV capsid. Similarities were noted between the NoV coat protein and the coat proteins of four viruses of the Nodaviridae, whose members also have insect hosts and bipartite RNA genomes.

Interest in the possible utility of a tetravirus in agriculture stems from the ability of HaSV to infect midgut cells and rapidly retard the growth of its host. H. armigera is a highly damaging, worldwide insect pest for which new control strategies are urgently required. More precise knowledge of the structures that allow HaSV to survive a degradative milieu after ingestion and enter the cells lining its host’s midgut may provide leads for the design of novel, proteinaceous, control factors (Christian et al., 1993). Such factors, after expression in transgenic crop plants, would be able to mimic these abilities of HaSV and deliver deleterious contents into the midgut cells of H. armigera, thus serving to protect the plant.

In testing the above approach to a novel control factor(s) for H. armigera, we sequenced RNA2 of HaSV and found that it contains an additional gene to the one for the viral coat protein. Furthermore, we expressed the two genes found on RNA2 in a bacterial expression system to test their functionality and to help interpret data from in vitro translation experiments on the replication strategy of HaSV. This paper reports the results of that work and, together with a forthcoming paper reporting the sequence of RNA1 (K. H. J. Gordon, K. N. Johnson & T. N. Hanzlik, submitted) forms the first sequence of the complete genome of a member of the Tetraviridae.

Methods

Animals and virus. H. armigera larvae were raised as described by Teakle & Jensen (1985). HaSV particles were purified as described by Hanzlik et al. (1993).

Molecular reagents and procedures. All reagents were purchased from Pharmacia unless otherwise noted and oligo-nucleotides were synthesized on a Pharmacia-LKB Geneassembler. Common procedures for molecular biology were as described by Sambrook et al. (1989) unless otherwise noted. PCR reactions were done with Taq polymerase under conditions recommended by Innis & Gelfand (1990) on a rapid cycling, capillary machine (Corbett Scientific). Computer analyses used programs from the University of Wisconsin Genetics Computer Group (Devereux et al., 1984) and the STAR program for RNA secondary structure described by Abrahams et al. (1990). A computer search using the BLAST algorithm for homology of RNA2 sequences to the combined updated GenBank/EMBL databases was done in December 1993 (Altschul et al., 1990).

Synthesis of cDNA and cloning. HaSV RNA2 was extracted and resolved as described by Hanzlik et al. (1993). The RNA was polyadenylated with poly(A) polymerase and cDNA was synthesized in the presence of d(T)18 oligomer primers, dNTPs and reverse transcriptase (RNase H- Superscript I, BRL) as recommended by the supplier. The cDNA was blunt-ended with T4 DNA polymerase, size-selected for lengths > 1 kb on a spin column (Clontech), and ligated into the EcoRI site of pBluescript SK(–) (Stratagene). The ligation mixture was used to transform E. coli strain JM 109 by electroporation. The cDNA clones phr236 and phr247, with insert sizes of 2.2 kbp and 1.6 kbp respectively, were isolated from a screen of the transformants (Fig. 1a).

Cloned DNA fragments were also generated by PCR from first-strand cDNA synthesis reactions which used HaSV genomic RNA, primers (described below) and avian myeloblastosis virus (AMV) reverse transcriptase under conditions recommended by the supplier (Pharmacia). The cDNA clone PBSr25p was derived from an ampiclon made with primers hr2cdn5 (corresponding to the ultimate 5' 18 base sequence) and hr236R7 (complementary to bases 1653-1669) and subsequently digested with PstI and cloned into the vector pGEM-1 (Promega). A second cDNA clone, phr236p71, was made with an amplicon generated from the primers hr236F3 (bases 875-891) and hr236R7 (complementary to bases 1653-1669), digested with AatII, kinased, and ligated to a 940 bp fragment of phr236 produced by digestion with HindIII, end-filling, and another digestion with AatII.

RNA and DNA sequencing. The cDNA clones and their subclones were sequenced by the method of Sanger et al. (1977) using T7 DNA polymerase (Pharmacia) with both single-stranded and double-stranded DNA templates generated from pBluescript SK(–). At least one of the nucleotide analogues deaza-dGTP, deaza-dATP, deaza-dTTP or dTTP (Pharmacia and US Biochemical) were used in all sequencing reactions to eliminate severe compression problems arising from the viral sequence. On occasion, the sequencing reactions were run on 6% acrylamide gels containing 40% formamide to reduce further the occurrence of compressions.

RNA was sequenced directly using chain terminators (Fichot & Girard, 1990), AMV reverse transcriptase, Moloney murine leukaemia virus (MMLV) transcriptase, and its derivative, Superscript I (BRL). By running the sequence reaction products of these enzymes in parallel, we found that unambiguous sequences in certain regions produced by each enzyme differed and complemented each other to produce a single, interpretable sequence. The sequence of the 5' region was confirmed by sequencing a 1.67 kb PCR fragment amplified from cDNA made with AMV RT. The 5' terminal region of RNA2 was also sequenced by RNase digestion under conditions which conferred base specificity using reagents supplied in a kit (US Biochemical). The ultimate 5' and 3' nucleotides of RNA2 were determined by the methods described by Shirako & Wilson (1993).

Electrophoresis and immunoblotting. NoV and rabbit antiserum against NoV particles were kindly provided by Don Hendry (Rhodes University, Grahamstown, South Africa). SDS-PAGE was done by...
following the procedure of Laemmli (1970), with 12.5% gels and low molecular mass standards (Bio-Rad). Immunoblotting was performed according to Towbin et al. (1979).

**Bacterial expression of HaSV genes.** A protocol described by Studier et al. (1990) and based on T7 RNA polymerase and a derivative of the pET-11 series of expression vectors (Novagen) was used to express the HaSV genes in bacteria (E. coli strain BL21, pLysS). The expression vector pT7T2 was made by ligating a 910 bp fragment from an pET-11 series of expression vectors (Novagen) was used to express the expression plasmids for p71 and p64 were each produced from ligations of which are present in the pET-11 vector.

To express p17, the plasmid pT7T2p17 (Fig. 1c) was constructed by the ligation of an XbaI–BamHI-digested vector and a 460 bp XbaI–BgIII restriction fragment derived from a 500 bp PCR product that was generated from phr236 with the primers hr236F3 and hvpet64N. The sequence of hvPET64N (GGGGATCCCTTACATATGGAGATGCTGGAGTG) containing a T7 promoter, an ORF encoding the large cleavage fragment of the coat protein precursor, and the Shine–Dalgarno sequence was used with hr236F3 to form a 1.4 kb PCR product; this product was digested with a BamHI–XbaI digestion of this product was gel-purified. The C-terminal fragment of p71, primer hvpet64N (GGGGATCCCTTACATATGGAGATGCTGGAGTG) containing a BamHI site adjacent to bases complementary to nucleotides 2290–2309 was used with primer hr236F3 to form a 1.2 kb PCR product; this product was digested with AatII and BamHI and the 1.1 kbp fragment was gel-purified. The two gel-purified fragments were ligated in one step to the vector to form pT7T2p17 (Fig. 1c). To form p64 with a C-terminal fragment different to p71, primer hvpet64N (GGGGATCCCTTACATATGGAGATGCTGGAGTG) containing a BamHI site adjacent to bases complementary to nucleotides 2072–2091 was used with hr236F3 to form a 1.2 kb PCR product; this product was digested with AatII and BamHI and the 0.9 bp fragment was gel-purified and ligated in a single step to the N-terminal fragment of both p71 and p64, primer hvpet64N (GGGGATCCCTTACATATGGAGATGCTGGAGTG) containing a BamHI site adjacent to bases complementary to nucleotides 2290–2309 was used with primer hr236F3 to form a 1.2 kb PCR product; this product was digested with AatII and BamHI and the 1.1 kbp fragment was gel-purified. The two gel-purified fragments were ligated in one step to the vector to form pT7T2p64 (Fig. 1c).

**Electron microscopy.** Bacteria expressing HaSV genes were harvested 2 h after induction with IPTG, pelleted, washed, and fixed in 2.5% glutaraldehyde in phosphate buffer (100 mM-sodium phosphate, pH 7.2, 150 mM-sucrose) for 2 h. After another wash in phosphate buffer, the pellet was incubated with 1% osmium tetroxide in phosphate buffer, the pellet was incubated with 1% osmium tetroxide in phosphate buffer and post-stained in 1% aqueous uranyl acetate for 30 min. The samples were embedded in polyester resin, sectioned at 60 nm, stained with lead citrate, and examined in a Hitachi H-600 transmission electron microscope.

Fig. 1. Genomic map of HaSV RNA2 and corresponding cDNA fragments used in sequencing and expression studies. (a) Sequence of HaSV RNA2 determined by sequencing both strands of two cDNA clones (solid bars) and RNA sequencing (open bars) with primers ( ). In addition, two cDNA fragments generated by PCR (lines ended with ) were cloned and sequenced in both directions in the area indicated by a thick line. (b) Genetic organization and restriction map of RNA2. ORFs $> 5$ kDa are represented as boxes (no such ORFs were found in Frame 2) with those believed not to be expressed during HaSV replication represented by open boxes. Only indirect evidence exists for the expression of ORFp71 (■) while ORFp71 (■) encodes the coat protein's precursor which is cleaved at the indicated site (†). Orientation of prominent stem-loop structures can be formed in the RNA; the bracketed area at the 3' end forms the tRNA-like structure. (c) Two ORFs noted in (b), ORFp17, ORFp71 and the area of ORFp71 encoding the large cleavage fragment of the coat protein precursor, were placed immediately behind the T7 promoter and the Shine–Dalgarno sequence (▲) in bacterial expression vectors. No viral sequences exterior to the indicated ORFs were present.
buffer for 1 h then dehydrated by washing with a series of solutions of ascending concentrations of ethanol to 100%. The pellet was embedded in LR White resin, cured overnight at 60 °C and sectioned. Sections were collected on parlodion coated copper grids and stained with 2% aqueous uranyl acetate and Reynold’s lead citrate before examination with a JEOL 100CX transmission electron microscope.

Results

Sequencing of RNA2 and detection of artifacts during cloning

Our strategy for obtaining the total sequence of RNA2 was to sequence both strands of phr236, which contained a DNA insert making up 90% of the total length of RNA2. Direct RNA sequencing was then used to determine those portions of RNA2 not contained in phr236. The RNA sequencing was then confirmed by sequencing both strands of a cloned PCR fragment, pBSr25p (Fig. 1a) from the area.

Implementing the sequencing strategy, however, proved problematical. In addition to having a high G + C content that regularly produced compressions and secondary structures that led to frequent strong stops, the reverse transcription of RNA2 reproducibly produced a precise deletion in the cDNA product. The deletion was detected when the sequence of phr236 showed two open reading frames (ORFs) coding for proteins of 39 kDa and 32 kDa, which were difficult to reconcile with previous information on the structure of the HaSV capsid proteins (Hanzlik et al., 1993). Genomic RNA was then sequenced directly in the region where the first ORF ended using the three different reverse transcriptases and a primer complementary to bases 1653–1669 of the RNA2 sequence (Fig. 2). The sequence ladders showed that both MMLV reverse transcriptase and its derivative, Superscript I, failed to transcribe 50 bases of contiguous sequence which were present in the AMV RT sequence ladder (bases 1497–1547, Fig. 2). This same 50 bp region was also absent from phr236 and another cDNA clone, phr247 (both made with Superscript I), resulting in the frame-shift and unexpected putative ORFs described above. Computer analysis predicted that the deleted region would likely form a stem–loop structure with a hexanucleotide sequence (CGGCAG) directly repeated and flanking its base (Fig. 3a). This deletion phenomenon, a model for its mechanism, and the effect of the 50 bp region on other reverse transcriptases will be described in more detail elsewhere (K. N. Johnson, K. H. J. Gordon & T. N. Hanzlik, unpublished data). The nucleotide sequence is 63% identical to the N°V RNA2 sequence as measured by the GCG GAP algorithm; however, the N°V sequence is 30 bases shorter which may have certain implications (see Discussion).

Secondary structure of RNA2 and comparison to HaSV RNA1

Two notable secondary structures, one from each terminus, can be discerned from the RNA2 sequence with the Star computer program (Abrahams et al., 1990), in addition to the putative stem–loop structure associated with the deletion noted above. Both structures are associated with the only two discernible areas of significant sequence identity to RNA1 of HaSV (K. H. J. Gordon, K. N. Johnson & T. N. Hanzlik, submitted). The first 48 bases of RNA2 can be folded into a stem–loop structure (Fig. 3b) that has a hexamer sequence, GGUAAA, identical to one located at the analogous position of a stem–loop structure at the 5' terminus of RNA1. The 5' stem–loop structures of each RNA otherwise have little structural or sequence similarity to each other.

The 3' terminal 130 bases of HaSV RNA2 can be folded into a distinct tRNA-like structure which has a valine anticodon. The terminal 108 nucleotides have 79% contiguous sequence identity to the 3' end of RNA1 of HaSV where a similar structure is found (Gordon et al., 1993). The products showed the same pattern on an autoradiograph as reported earlier (Hanzlik et al., 1993; data not shown).
RNA2 sequence of HaSV

Fig. 2. Nucleotide sequence of complete cDNA for HaSV RNA2. Translations of ORFp17 and ORFp71 (upper and lower, respectively) are shown under the nucleotide sequence in the single letter code. Accession number, L37299.
al., 1995). Significantly, where the two sequences differ, and with only three exceptions, the secondary structure is conserved by the complementary base pairing being retained (e.g. G–C becomes A–U or C–G). These structures are similar to those found at the ends of the genomic RNAs in some groups of plant RNA viruses (Fig. 3c). However, the tRNA-like structures of HaSV lack the typical pseudo-knot found in tRNA-like structures of plant viruses and differ from normal tRNA structures in the smaller size of the anticodon loop (normally 7 bases) and the larger size of the variable loop (10 bases greater than largest known loop; Mans et al., 1991; Haenni et al., 1982). As will be discussed in more detail in paper on RNA1 of HaSV (Gordon et al., 1995) this is the first report of distinct tRNA-like structures being discerned in the genome of an animal virus as well as the first report of a viral tRNA-like structure that lacks a pseudo-knot.

**Analysis of ORFp17**

ORFp17 predicts a 157 residue, 16522 Da protein (p17) whose sequence shows no significant homology to other sequences in the databases. Its composition has a high content (49%) of the four amino acids – proline (13%), glutamate (6%), serine (17%) and threonine (13%) – and thus resembles PEST proteins (Rogers et al., 1986) which have distinct regions having a high content of these residues that are flanked by basic residues. However, it must be noted that p17 possesses more than just a limited region with a PEST sequence given its high content of the four amino acids. The protein is not a structural component of the HaSV capsid as there are no proteins of this size extracted from HaSV particles which react with rabbit polyclonal antisera made against bacterially expressed p17 (data not shown). Interestingly, while apparently not an HaSV structural protein, it has the capability to assemble into distinct structures when expressed in bacteria (see below). While ORFp17 begins with the first AUG of RNA2, its context (AUU AUG A) matches poorly with the consensus for initiation of protein synthesis by scanning ribosomes in higher eukaryotes (CA/GCC AUG G) (Kozak, 1989) or invertebrates (A/CAAA/C AUG G) (Cavener & Ray, 1991).

**Analysis of ORFp71**

The second AUG of RNA2 at bases 366–368 initiates an ORF that encodes a predicted 647 amino acid, 70670 Da
protein (p71). Its context (CAGG AUG G) matches well with the consensus for initiation of protein synthesis and the protein is proven to be the precursor of the two protein components of the HaSV capsid. Two regions of the sequence predict peptides that match with partial amino acid sequences from the capsid proteins reported earlier (Hanzlik et al., 1993). Bases 2091–2145 predict the sequence FAAAASAFANMLSSVLSKS (Fig. 2), which is identical with that reported for the N-terminus of the 7 kDa capsid protein, while the prediction of bases 1128–1185 (PTLVDQGF WIGGQYALTPTS, Fig. 2) exactly matches a 19 amino acid sequence from a 20 kDa cyanogen bromide fragment of the N-terminally blocked 63 kDa major capsid protein. These data indicate that the precursor protein is cleaved at an Asn-Phe site located after residue 575, producing two proteins with molecular masses of 63378 Da and 7309 Da. These molecular masses are highly similar to the values determined by SDS–PAGE for the HaSV capsid proteins (Hanzlik et al., 1993). Both termini of the precursor protein are basic in nature, with 11 of the first 43 residues of the N terminus being Arg and the C terminus having eight Arg residues in the last 15. Other than the sequence predict peptides that match with partial protein sequences, the dissimilarity between the two coat proteins resides mainly in a 165 amino acid region in the middle of the sequences and a smaller 49 amino acid region located at the N termini (Fig. 4b). Also, there is a region of 29 residues towards the C terminus which has significantly less identity than found elsewhere yet possesses high similarity when amino acids having similar properties are taken into account. Surprisingly, despite the high identity (>80%) possessed by 58% of the length of the two coat protein sequences, reciprocal immunoblots show no common epitopes which cross-react with antisera raised against particles of either virus (Fig. 5). Furthermore, no cross-reaction was noted with immuno-double diffusion, or with another rabbit antiserum preparation raised against HaSV particles (Hanzlik et al., 1993). This latter technique should allow antibodies specific for antigens with secondary and tertiary structures to react.

Comparison of sequence and antigenicity of the HaSV and No9V coat proteins

When the sequence of the HaSV coat protein precursor was compared to that of No9V, a high degree of overall identity (66%) was found (Fig. 4). The Asn–Phe cleavage site is conserved at the same position (Fig. 4a, b; arrow) along with the regions of high basicity at each terminus. The dissimilarity between the two coat proteins resides mainly in a 165 amino acid region in the middle of the sequences and a smaller 49 amino acid region located at the N termini (Fig. 4b). Also, there is a region of 29 residues towards the C terminus which has significantly less identity than found elsewhere yet possesses high similarity when amino acids having similar properties are taken into account. Surprisingly, despite the high identity (>80%) possessed by 58% of the length of the two coat protein sequences, reciprocal immunoblots show no common epitopes which cross-react with antisera raised against particles of either virus (Fig. 5). Furthermore, no cross-reaction was noted with immuno-double diffusion, or with another rabbit antiserum preparation raised against HaSV particles (Hanzlik et al., 1993). This latter technique should allow antibodies specific for antigens with secondary and tertiary structures to react.

Bacterial expression of HaSV genes and analysis of their products

The discrepancy between the size of p17 and the major product produced from RNA2 in in vitro translation studies (molecular mass 24 kDa) presented an opportunity to gain an insight into the replication strategy of HaSV. Among the possible explanations for the difference were ribosomal frame-shifting, stop codon read-through, polyprotein processing or anomalous migration by p17 in SDS–PAGE. While some of these possibilities were excluded after examination of the RNA2 sequence, we tested the latter possibility by analysis of the sizes and antigenicity of the proteins made from ORFp17 in a bacterial expression system. The construction of the plasmid expressing p17 excluded all possibilities involving gene products larger than 17 kDa. We found that the product of ORFp17 exhibited a phenomenon commonly seen with proline-rich proteins (Pham & Sivasubramanian, 1992; Zeimer et al., 1982) and migrated at an anomalously slow rate on SDS–PAGE, showing an apparent molecular mass identical to that of the in vitro translation product (Fig. 6, lane 4; arrowhead). These observations indicate that, during in vitro translation, the majority of ribosomes start translation at the first AUG and that neither processing nor frame-shifting occurs during the process.

We analysed the products of ORFp71 in a similar manner. The data (Fig. 6a, b) show that the size and antigenicity of the predicted 63 kDa cleavage product of p71 is highly similar to that of the large HaSV capsid protein, further confirming the sequence data. The data also show that no self-processing of p71 occurs in bacteria. Because the structural proteins of capsids of a number of viruses have been shown to self-assemble when their genes have been recombinantly expressed, we used transmission electron microscopy to look for virus-like particles (VLPs) in bacteria expressing p71. In this experiment, we also looked at bacteria expressing p17 as a control for bacteria expressing large amounts of an exogenous protein. Contrary to expectations, we found no ordered structures resembling VLPs in the bacteria expressing p71 (Fig. 7a). However, in bacteria expressing p17, large arrays of hexagonal tube-like structures (Fig. 7b), apparently without specific length, were observed. When viewed in cross-section, the mostly hollow tubes had diameters of 32 nm and possessed distinct electron dense cores 2–3 nm in diameter; the cores were not apparent in other sectioning perspectives (Fig. 7c, d). We confirmed these observations by verifying the plasmid constructions and repeating the experiment. In addition, the structures were seen under the electron microscope to be strongly labelled with immuno-gold particles using polyclonal antisera raised against p17 (data not shown).
Fig. 4. For legend see opposite.
**Fig. 5.** Reciprocal immunoblots of HaSV and Nc0V coat proteins showing lack of cross-reaction. Particles (3 µg protein each) of HaSV (lane 1) and Nc0V (lane 2) and molecular mass markers (lane M) were separated by SDS-PAGE and stained with Coomassie Blue (a) or immunoblotted by probing with antisera against HaSV (b) or Nc0V (c).

**Fig. 6.** Bacterial expression of genes of HaSV RNA2. Bacteria were induced to express genes placed behind the T7 promoter as outlined in Fig. 1(c). Lysates were then examined by SDS-PAGE and stained with Coomassie Blue (a) or immunoblotted by probing with antisera against HaSV (b). Lane M, molecular mass markers of indicated size; lane 1, HaSV control; lane 2, pT7T2p64; lane 3, pT7T2p71; lane 4 pT7T2p17; lane 5, induced bacteria without expression vector. The arrowhead (a, lane 4) points to the product of ORFp17 migrating anomalously at 24 kDa.

**Fig. 4.** Comparison between the amino acid sequences of the coat protein precursors of HaSV and Nc0V. (a) Alignment obtained using the GCG program BESTFIT. Identical residues are indicated by vertical lines. Underlined residues are those believed to make up the eight components (B–I) of the β-barrel of the viral coat protein proposed by Johnson et al. (1994). (b) Graphical representation of distinct regions in the two proteins showing greater and lesser (shaded) similarity with percentage identities between similar and dissimilar regions indicated below and above, respectively. The values in parentheses are percentage similarities for the same region calculated with conservative substitutions. The area indicated by a dotted line shows a region having relatively less identity but having high similarity. The arrow in both (a) and (b) points at the precursor cleavage site.
Fig. 7. Electron micrographs of bacteria expressing genes of HaSV RNA2. Bacteria induced to express p71 and p17 were fixed, sectioned, and examined by transmission electron microscopy. (a) Bacteria expressing p71 show no definite structures resembling VLPs (bar, 0.5 μm). (b) Bacteria expressing p17 show ordered arrays of hexagonal tube-like structures (bar, 0.5 μm). (c) Micrograph of bacteria expressing p17 at higher magnification (bar, 0.2 μm) which shows both a cross-section and a sagittal section of the apparently flexible structures which have diameters of 32 nm, and electron dense cores. (d) Further magnification (bar, 0.2 μm) shows the electron-dense cores more distinctly.
Discussion

This report, along with that for RNA1 of HaSV (Gordon et al., 1995) represents only the second group of insect small RNA viruses whose level of study has advanced to the stage of obtaining a complete sequence of the genome. The Nodaviridae are the only other group of RNA viruses with exclusively insect hosts whose genomes have been so characterized (Hendry, 1990), as well as the only other group of animal RNA viruses with single-stranded, bipartite genomes (1.4 kb and 3.2 kb). Previous reports on both HaSV and NoV have detailed other similarities between tetraviruses and nodaviruses (Agrawal & Johnson, 1992; Hanzlik et al., 1993). This report, however, further distinguishes the two groups by noting the possible presence of two genes on RNA2 and its 3′ tRNA-like structure. Surprisingly, the same two characteristics apparently distinguish HaSV from the other tetravirus, NoV, to which it is clearly closely related. Close examination, however, shows the differences may be artifactual. The method of cloning may have removed the terminal bases of NoV RNA2, which is 30 bases shorter than that of HaSV (Agrawal & Johnson, 1992). Indeed, a partial tRNA structure lacking 30 bases can be formed. The lack of an ORF analogous to p17 may be due to a sequencing deletion between positions 276-364 of the NoV sequence. The latter supposition was supported with recent evidence from in vitro translation of the NoV genome where two protein products similar in size to p17 and p71 were noted (D. Agrawal, personal communication).

RNA2 is likely to possess a complex secondary and tertiary structure as extensive self-complementarity is indicated by both theoretical calculations and experimental observations. The STEM–LOOP algorithm predicts 24 stem–loop structures, where folding due to self-complementarity occurs, that have bond strengths greater than 22 (where A–T = 1 and G–C = 2; Devereux et al., 1984). Experimentally, we observed that a labelled probe specific for the negative-sense strand of RNA2 also hybridized readily to the genomic, positive strand RNA (data not shown). Also, the difficulties that enzymes had in processing RNA2 templates, such as the deletion that occurred during its reverse transcription and the many strong stops observed during its sequencing, are strong indications of extensive secondary structure. As discussed for tRNA-like structures (Haenni et al., 1982) and shown by work for nodaviruses (Zhong et al., 1992) and hepadnaviruses (Pollack & Ganem, 1993), some of the secondary structures are likely to play important roles in the biology of HaSV.

There is as yet no direct evidence for the expression or function of p17 during the replication of HaSV. While p17 is the major product of RNA2 from in vitro translation experiments, no detectable in vivo expression of the protein occurs in insects. It is highly unlikely to be a structural protein as immunoblots of proteins extracted from HaSV particles show no reaction against rabbit polyclonal antisera raised against the bacterially expressed p17. Furthermore, the protein is not expressed in SF9 cells when RNA2 is placed behind the polyhedrin promoter and expressed with a recombinant baculovirus (T. N. Hanzlik, S. J. Dorrian, E. M. Brooks & K. H. J. Gordon, unpublished data). However, the protein may be expressed at low levels by leaky scanning by ribosomes (Kozak, 1989).

Assuming its expression, two functions for p17 can be postulated from its primary structure and the results from its bacterial expression. Its status as a PEST protein hints at a regulatory role; many PEST proteins are tightly regulated proteins such as transcription factors or gene products involved in the cell cycle (Rodgers et al., 1986). However, the striking, organized structures composed of bacterially expressed p17 seen in Fig. 7 suggest an entirely different function, one analogous to that of the movement proteins of plant viruses, which would facilitate movement of HaSV to adjacent midgut cells. The tubular structures seen in Fig. 7 have a pore size roughly the size of the HaSV particle and are reminiscent of the tubes formed by the movement protein of the cauliflower mosaic virus (Perbal et al., 1993).

The analysis by Agrawal & Johnson (1992) of the sequence of the coat protein precursor of NooV revealed a striking degree of similarity to the coat proteins of the nodaviruses which have T = 3 symmetry. The sequence of the HaSV coat protein precursor supports this picture. For example, 24/28 similar residues in the eight β-barrel regions (B–I) noted in the alignment of Agrawal & Johnson are conserved when the HaSV sequence is added to the alignment. Even in the D, E and F β-barrel strands (Johnson et al., 1994) located in the dissimilar region of the two tetraviruses, the conservation among the viruses is retained.

The lack of a serological relationship between the coat proteins of NooV and HaSV despite the high sequence identity outside the middle dissimilar region suggests that the middle region is highly exposed and is solely responsible for the antigenicity of each particle. This picture is supported by preliminary experiments showing that a bacterially expressed protein lacking this region failed to react with the antisera. This finding may modify the conclusions of previous serological work done on the group (Greenwood & Moore, 1981). A serological reaction between two tetraviruses may denote a much closer relationship than is commonly understood from such work. For example, a serological relationship is retained by the two nodaviruses flock house virus and...
boolara virus, even though their primary sequences are only 50% similar (Kaesberg et al., 1990).

The HaSV RNA2 sequence now makes it possible to test the utility of HaSV to agriculture as outlined in the introduction. Together with the RNA1 sequence (K. H. J. Gordon, K. N. Johnson & T. N. Hanzlik, submitted) it extends our knowledge of an intriguing group of small RNA viruses, and shows that they have characteristics in common with plant RNA viruses, such as bipartite genomes and 3' tRNA-like structures. Because of these similarities, it is tempting to speculate on a plant viral origin for tetraviruses whereby, due to the continuous association of the midgut cells of herbivorous insects with infected plant material, chance favoured the adaptation of an ancestral plant virus to the environment of the insect midgut cell. However, the possibility must be examined that evolution occurred the opposite way, i.e. many plant viruses may have had their origins from insect viruses that were progenitors to HaSV, as originally proposed by Goldbach (1986). Previous analyses of the relationships among eukaryotic single-stranded RNA viruses have been distorted by the mass of information available on viruses having angiosperm and vertebrate hosts. These analyses have had only one virus with an exclusive insect host and no sequences at all from RNA viruses with gymnosperm hosts, both groups which appeared earlier than vertebrates or angiosperms in the fossil record. The prospect of some angiosperm viruses originating from insect viruses is supported by the HaSV tRNA-like structure being more closely related to putatively more ancestral tRNAs, the earlier appearance of insects in the fossil record, and various sequence analyses (Goldbach, 1986; Bruenn, 1991). More extensive information on RNA viruses that infect the largest group of animals will allow such questions to be addressed less speculatively.

We thank Eric Hines, Crispin Hull and Leon Court for valuable technical assistance. This work is supported in part by Pacific Seeds Pty Ltd (Australia).

References


(Received 23 August 1994; Accepted 29 November 1994)