Molecular cloning and nucleotide sequencing of the coat protein gene of citrus tristeza virus

Mary E. Sekiya,† Susan D. Lawrence, Michael McCaffery and Kenneth Cline*

Department of Fruit Crops, University of Florida, Gainesville, Florida 32611, U.S.A.

Citrus tristeza virus (CTV) contains approximately 20 000 bases of positive-sense ssRNA, encapsidated by a coat protein of approximately 25 000 M_r that has previously been reported to consist of at least two size variants, cp1 and cp2. In the present study, a cDNA library of the T36 isolate of CTV was prepared in a protein expression vector and screened with a polyclonal antibody against the coat protein. Five immuno-positive clones produced proteins in Escherichia coli that reacted with monoclonal as well as polyclonal antibodies to the CTV coat protein. Nucleotide sequence analysis of a region common to the five clones revealed the presence of a 669 nucleotide open reading frame flanked by numerous in-frame termination codons. The encoded protein has a predicted M_r of 24 909 and an amino acid composition consistent with that previously reported for the CTV coat protein. Comparison of the predicted amino acid sequence of the coat protein with the amino-terminal sequences of cp1 and cp2 indicated that these coat protein species arise from the same primary translation product, as a result of post-translational proteolysis at sites approximately 12 to 15 and 26 amino acids from the amino terminus respectively. These results are the first reported cloning and sequencing of a CTV gene and provide evidence that CTV may be translated using subgenomic RNA.

Introduction

Citrus tristeza virus (CTV) is economically the most important citrus virus world-wide (Bar-Joseph et al., 1989); it is responsible for several different disease syndromes in citrus plants, including 'stem pitting' and 'quick decline', in which phloem necrosis frequently results in death of the tree. CTV is normally transmitted by aphids or by grafting and is limited to the phloem of infected trees. CTV is the largest member of the clustervirus group, which is characterized by a long flexuous rod morphology (Bar-Joseph et al., 1979). The virion consists of a single-stranded mRNA, about 20 000 bases in length, encapsidated by a single type of coat protein (Lee & Calvert, 1987; Lee et al., 1988), which has reported M_r of 26 000 (Nagel et al., 1982; Lee et al., 1988) or 27 000 (Dulieu & Bar-Joseph, 1990) from in vitro translation studies. However, when purified CTV is analysed, two polypeptides can be detected (Gonsalves et al., 1978; Lee & Calvert, 1987); the larger protein (cp1) has a reported M_r of 23 000 and is estimated to be present in amounts fivefold greater than that of the smaller protein (cp2), which has an M_r of 21 000 (Lee et al., 1988).

Both forms of the protein react with CTV-specific antisera and have similar amino acid compositions and proteolytic peptide maps (Lee et al., 1988) and it has been hypothesized that cp2 is a post-translational proteolytic product of cp1 (Lee & Calvert, 1987; Lee et al., 1988). There are no other known proteins encoded by this virus, although the size of the genome suggests a coding capacity for between 10 and 20 proteins of moderate size. Indeed, translation of CTV RNA in vitro has yielded products with M_r's of 65 000, 50 000 and 33 000 in addition to the coat protein (Nagel et al., 1982; Lee et al., 1988).

The aim of the present work was to clone and sequence the coat protein gene of CTV. Cloning of the coat protein gene of the severe Florida isolate, T36, was undertaken in order to use this gene ultimately to engineer viral resistance into citrus plants (Powell Abel et al., 1986). Transgenic plants expressing coat protein genes of virus pathogens show delayed disease development or are resistant to infection upon inoculation with the respective virus. This form of resistance has been shown for tobacco mosaic virus, alfalfa mosaic virus, tobacco rattle virus, cucumber mosaic virus, potato virus X and potyviruses (Powell Abel et al., 1986; Loesch-Fries et al., 1987; VanDun et al., 1987; VanDun & Bol, 1988; Tumer et al., 1987; Cuozzo et al., 1988; Hemenway et al., 1988; Hoekema et al., 1989; Stark & Beachy, 1989). The wide
applicability of this method offers promise for coat protein-mediated protection of citrus plants.

An additional goal of this project was to gain some insight into the genome organization of CTV by cloning the open reading frame (ORF) of the only known protein product of the virus. At present, little is known about either the genome organization or translation strategy of closteroviruses, although subgenomic dsRNAs can be isolated from infected tissues (Dods & Bar-Joseph, 1983; Dods et al., 1987; Dulieu & Bar-Joseph, 1990), which are often indicative of subgenomic RNA transcripts.

Here, we report the cDNA cloning and nucleotide sequencing of the coat protein gene of CTV isolate T36. The coat protein is encoded by a 669 nucleotide ORF that is flanked by numerous in-frame termination codons. This ORF was identified by several criteria, including alignment with peptide sequences obtained from the amino terminus of isolated coat protein. Our results indicate that CTV proteins are produced by translation of individual cistrons rather than by autocatalytic processing of a polyprotein. However, our analyses also show that the amino terminus of the CTV coat protein is processed both in the plant and during isolation of the virus.

Methods

Materials. Etnrog citron (Citrus medica var ethrog Engl.), C. hystrix var boholensis Wester and C. excelsa Wester, inoculated with CTV isolate T36, were the generous gift of Dr Steve Garnsey (USDA, Orlando, Florida, U.S.A.). Polyclonal antiserum to the SDS-denatured CTV coat protein (Gonsalves et al., 1978; Garnsey et al., 1979) was generously provided by Dr Ernest Hiebert. A monoclonal antibody (MAb) to the coat protein, MCA 13, was generously provided by Drs Thomas Permar and Steve Garnsey (Permar et al., 1988). The cDNA clones pSBD8 and p6A6, from CTV isolate T36, were the generous gift of Drs Lee Calvert, Richard Lee and Ernest Hiebert. Lambda Zap II, Escherichia coli strain XI1 Blue and the Exo III/mung bean deletion kit (pBluescript SK-) by the addition of helper phage R408 (manufacturer's instructions). Plasmid DNA was isolated from the clones by the alkaline lysis miniprep procedure (Maniatis et al., 1982). The cDNA inserts were mapped with restriction enzymes as previously described (Maniatis et al., 1982) and, because RNA is predicted to be approximately 5% of the virus particle by weight (Bar-Joseph et al., 1979), 40 µg CTV RNA from 30 g of starting tissue is a high yield.

Synthesis of cDNA. Complementary DNA was prepared to viral RNA by the procedure of Gubler & Hoffman (1983) with a cDNA synthesis kit, using random primers [[pd(N)₆] for first-strand synthesis. The cDNA was methylated using EcoRI methylase, blunt-ended with T4 DNA polymerase and ligated to EcoRI linkers as described (Maniatis et al., 1982). The resulting cDNA was fractionated on a 0.7 x 25 cm Sephacryl S-500 column and cDNAs of greater than 350 bp in length were selected. The cDNA was then digested into Lambda Zap II EcoRI-restricted arms (Huynh et al., 1984; Davis et al., 1986).

Immunoscreening the cDNA library. The resulting bacteriophage libraries were immunoscreened (Huynh et al., 1984; Mierendorf et al., 1987; Snyder et al., 1987) with CTV-specific polyclonal antibody, essentially as described by Huynh et al. (1984). Primary antiserum was preadsorbed with cell extract from E. coli (Mierendorf et al., 1987); the secondary antibody, alkaline phosphatase-conjugated goat anti-rabbit IgG, was used at a dilution of 1/1000 from a stock concentration of 100 µg/ml.

Analysis of coat protein-positive clones. Plaque-purified, immuno-positive Lambda Zap II clones were converted to plasmid clones (pBluescript SK+) by the addition of helper phage R408 (manufacturer's instructions). Plasmid DNA was isolated from the clones by the alkaline lysis miniprep procedure (Maniatis et al., 1982). The cDNA inserts were mapped with restriction enzymes as previously described (Maniatis et al., 1982) and were also used as probes in Southern hybridizations. DNA fragments were radiolabelled with [³²P]dCTP and a random-priming labelling kit, and Southern hybridizations were performed as described (Maniatis et al., 1982, Church & Gilbert, 1984). Double-stranded DNA sequencing was performed by the dideoxynucleotide chain termination method of Sanger et al. (1977) and sequence analysis was performed on a DuPont Genesis 2000 DNA Analysis System in the Interdisciplinary Center for Biotechnology Research (ICBR) core facility at the University of Florida, and also using Sequenase and a Gibco BRL S2 electrophoresis unit; primers were supplied by the ICBR DNA synthesis core facility. Sequence reactions were conducted using clones pCPS5 and pCP8, and subclones of pCP17.2 (see Fig. 1 for description) generated with an Exo III/mung bean nuclease deletion kit, according to the manufacturer's instructions. The nucleotide sequence was analysed using the University of Wisconsin Genetics Computer Group Sequence Analysis software available through the ICBR Biological Computing core facility.

Analysis of coat protein products produced from cloned genes in E. coli. DH5α or XL1 Blue E. coli, harbouring plasmids from immunopositive clones, were grown in 10 ml overnight cultures. An aliquot (1 ml) of the overnight culture was inoculated into 10 ml LB medium (Maniatis et al., 1982) and isopropyl β-D-thiogalactoside was added to a concentration of 10 mM. After 4 to 5 h at 37 °C, bacteria were harvested and lysed by a cycle of freezing and thawing and treatment with SDS-PAGE
Fig. 1. cDNA clones containing the coding region for the CTV coat protein and structure of the cloned region. Five cDNA clones containing coding sequences for the coat protein were identified by immunoscreening of a cDNA library prepared in Lambda Zap II. The cloned inserts were aligned by Southern analysis and restriction enzyme mapping. The structure of the cloned region is shown at the top of the figure. Restriction enzyme sites are B, BamHI; Bs, BstEII; P, PstI; Pv, PvuII; X, XbaI. The region common to all clones was sequenced and an ORF encoding the coat protein is shown by the large arrow.

Sample buffer. Samples were analysed by SDS-PAGE and immunoblotting with antibody to the coat protein essentially as described (Harlow & Lane, 1988), using an alkaline phosphatase-linked secondary antibody and 5-bromo-4-chloro-3-indolyt phosphate/nitro-blue tetrazolium.

Rapid extraction of infected citrus tissue for immunoreplica analysis. Stem pieces were excised from plants infected with T36 CTV and were immediately frozen in liquid nitrogen and stored at -80 °C until use. Frozen tissue was pulverized with a mortar and pestle in the presence of liquid nitrogen, boiling SDS-PAGE sample buffer was added to the tissue powder at a ratio of 4:1 (v/w), the mixture was immediately homogenized with a preheated Polytron and then further heated at 100 °C for 2 min. Particulate material was removed by two sequential centrifugations in a microcentrifuge. For comparison with this rapid extraction, tissue powder was allowed to warm to room temperature for 90 min before the addition of SDS-PAGE sample buffer. Samples were diluted fivefold before immunoblotting analysis.

Protein isolation and amino acid analysis. Coat protein for aminoterminal microsequencing was obtained from the enriched viral preparation described above. Virus was dissolved in SDS sample buffer, subjected to SDS-PAGE on 12.5% gels and then electroblotted onto a poly (vinylidene fluoride) (PVDF) membrane using 10 mM-2-(N-morpholino)ethanesulphonic acid pH 6.0, 20% methanol transfer buffer. The PVDF membrane was stained in 0.1% Coomassie Brilliant Blue R in 50% methanol for 5 min, destained in 50% methanol/10% acetic acid for 3 min, washed in distilled deionized water and allowed to air dry. The appropriate bands were cut from the membrane and used for amino acid composition and sequence analyses. All protein analyses were performed at the University of Florida ICBR Protein Chemistry core facility. Peptide sequencing was conducted on an ABI Model 470A Gas Phase Protein Sequencer with Model 120A On-Line PTH Analyzer. Amino acid composition analysis was performed on a Beckman 6300 Amino Acid Analyzer.

Results

A cDNA library of highly enriched viral RNA of the T-36 isolate of CTV was prepared in Lambda Zap II (see Methods) and screened with polyclonal antisera to CTV coat protein; 15 immunopositive plaques were obtained from a total of approximately 8000 recombinant bacteriophage clones screened. Of these, five immunopositive bacteriophages were converted into plasmid clones, designated pCP8, pCP5, pCP17.3, pCP17.1 and pCP17.2. The cDNA inserts of these clones cross-
hybridized by Southern analysis and were aligned by restriction enzyme mapping (Fig. 1). Southern analysis confirmed that the region of homology of the five clones was between the BamHI site of pCP8 and the PstI site of pCP17.2; these two clones span approximately 3 kb of the CTV genome (Fig. 1). Previously, Calvert (1987) prepared two large overlapping cDNA clones which he designated p5D8 and p6A6, that together contain approximately 14 kb of the T36 CTV genome. The region cloned in the present study is apparently a different region of the genome because none of the five clones hybridized with the previously cloned region (data not shown).

All five clones produced immunoreactive proteins in E. coli, as determined by immunodecoration of Western blots (Fig. 2). As expected, E. coli harbouring only the

plasmid vector was immunonegative. The major immunoreactive proteins from clones pCP5, -8, -17.1 and -17.3 migrated during SDS–PAGE similarly to the larger coat protein variant present in infected citrus tissue, approximately at the location of cp 1 (however, see Discussion below), whereas the immunoreactive protein produced by clone pCP17.2 was approximately 2K larger (see Discussion). The absolute amount of immunoreactive protein varied with the experiment; in the experiment shown in Fig. 2, the protein produced by clone pCP17.2 was approximately 2K larger (see Discussion) than the protein produced by clone pCP17.3, which migrated similarly to the major immunoreactive protein from clone pCP5, -8, -17.1 and -17.3.

Nucleotide sequencing of the region common to all clones revealed an ORF of 669 bases flanked by numerous in-frame termination codons (Fig. 3). The predicted amino acid sequence is shown above the nucleotide sequence and the location of in-frame termination codons is designated by an asterisk.
Fig. 5. A larger coat protein species is present in the plant. Stem segments from C. hystrix (lanes 1 and 2), C. medica (lanes 3 and 4) and C. excelsa (lanes 5 and 6) infected with T36 CTV were frozen with liquid nitrogen, powdered and immediately extracted with boiling SDS-PAGE buffer. The extract was then subjected to SDS-PAGE and immunoblotted with a polyclonal antibody to the coat protein (lanes 1, 3 and 5). A portion of each powdered sample was allowed to warm to room temperature prior to SDS denaturation (lanes 2, 4 and 6). Lane 7, purified T36 CTV.

In order to resolve the relationship between the ORF, cpl and cp2, amino-terminal microsequencing of the coat protein variants from purified virus was undertaken. Cp2 yielded an unambiguous sequence of 12 residues (Fig. 4a) which matches perfectly a region of the ORF starting at residue 27. Cpl yielded ambiguous sequence data indicative of three different amino termini which could result if there were heterogeneity at the nucleotide level. However, we obtained nucleic acid sequence data for the amino-terminal region of several clones, pCP5, pCP8 and pCP17.2, and did not observe heterogeneity. Rather, the peptide microsequencing data are more easily explained if the different amino termini arose from the coat protein by imprecise (staggered) post-translational proteolysis; high-resolution SDS gels support this interpretation because cpl is seen to consist of at least two bands (not shown). Furthermore, it is possible to account for the sequencing data if one assumes that cpl contains the three polypeptides that originate from the region approximately 11 to 15 amino acids upstream of the perfect match between cp2 and the ORF as shown in Fig. 4. The amino acids predicted from sequence analysis of a mixture of these three polypeptides (c) exactly match the actual data (b). The range in the $M_r$ of the three polypeptides would be no greater than 400, thereby explaining the inability to obtain a clean separation of individual polypeptides. These results strongly suggest that cpl, as well as cp2, is a degradation product that arises from the primary coat protein gene product, probably by post-translational proteolysis.

We attempted to determine whether cpl and cp2 originate in the plant or during isolation procedures by rapidly extracting CTV-infected plant tissue into boiling SDS buffer and then analysing the extracted coat protein by immunoblotting (Fig. 5). Tissue extracted in this manner displays at least two immunoreactive bands; the most prominent band migrates with an $M_r$ slightly greater (approximately 900) than that of cp1 and may represent the intact coat protein, whereas the second band migrates similarly to cp2. When the powdered tissue was allowed to warm to room temperature prior to addition of the SDS buffer, a band of the same $M_r$ as cpl appeared. This was most apparent in the CTV-infected C. medica and C. excelsa samples. The fact that the most predominant coat protein species in the plant was larger than cpl suggests that the intact coat protein is present in vivo; the appearance of a band comigrating with cp1 upon incubation of powdered tissue at room temperature suggests that proteolysis of the coat protein can occur during isolation procedures. Nevertheless, the fact that a substantial amount of protein comigrating with cp2 was present in the rapidly extracted preparation suggests that processing can also occur in the plant.

Discussion

The coat protein gene of CTV has been cloned using a protein expression vector and identified as an ORF of 669 bases. Based on amino acid composition analysis (Calvert, 1987; Lee et al., 1988), the coat protein gene was estimated to consist of approximately 600 bases. Therefore the actual size of the ORF agrees with the predicted value. We believe that we have cloned the true coat protein gene because antigenic coding regions were identified by immunoscreening with polyclonal antisera and were subsequently confirmed using a MAb against the coat protein. Furthermore, amino acid sequences predicted from the ORF are in agreement with N-terminal sequences from the CTV coat protein itself.

The cloning vector used in these studies, Lambda Zap II, is designed to produce fusion proteins between the $\alpha$ peptide of $\beta$-galactosidase and the cloned in-frame coding region (Short et al., 1988). However, analysis of the proteins produced in E. coli and nucleotide sequence analysis of clones pCP5, pCP8 and pCP17.2 suggest that only pCP17.2 produced an amino-terminal fusion protein. The 5' end of pCP17.2 lacks 15 bases of the coat protein gene and in E. coli produces an immunoreactive protein that is larger than cp1. This would be expected from a fusion between the $\alpha$ peptide and the coat protein. Conversely, pCP8, pCP5 and, presumably, pCP17.3 contain the entire coat protein coding sequence and produce immunoreactive proteins with approximately the same $M_r$ as the authentic coat protein. As several termination codons precede the initiation codon of the coat protein gene, these proteins could not be translated as amino-terminal fusion proteins. Apparently, a functional ribosome-binding site is present near the start of the coat protein gene because it is clear from Fig. 2 and
other similar analyses that significant amounts of immunoreactive proteins are produced in these clones.

Clone pCP17.1 routinely produced two immunoreactive protein bands in E. coli; the major band migrated similarly to the bands observed with pCP5, pCP8 and pCP17.3, whereas a minor band migrated at the location of the fusion protein produced by pCP17.2. It is not clear why this clone produces these reactive species. Our estimate of the size of this clone and the alignment of the clone by the location of the BstI site suggest that it contains the initiation codon but may be too small to contain the entire coding sequence. This would suggest that the immunoreactive protein of pCP17.1 is a carboxy-terminal fusion protein, terminating in vector sequences 3' to the insert. Partial proteolysis of the resulting fusion protein may explain the appearance of the immunoreactive bands in this sample. Their apparent comigration with the coat protein band and the fusion protein produced by pCP17.2 may be fortuitous. Nucleotide sequence analysis is required to resolve this apparent anomaly.

Complementary DNA libraries to CTV RNA have been constructed previously (Rosner et al., 1983; Calvert, 1987). The first library, cloned from an Israeli isolate of CTV, yielded cDNA probes capable of differentiating CTV isolates (Rosner & Bar-Joseph, 1984; Rosner et al., 1986); two cDNA libraries of Florida isolates have been prepared and the library of the severe isolate T36 yielded a partial (70%) map of the viral genome (Calvert, 1987). In neither of the two previous studies were protein coding sequences identified. This is the first report, to the best of our knowledge, of cDNA cloning and nucleotide sequencing of the coat protein gene of CTV.

Questions regarding the genome organization of CTV still remain unresolved because the cloned coat protein region was not homologous to the previously cloned region of CTV representing approximately 70% of the genome. Determining the position of the coat protein gene in the genome is important and should now be possible utilizing clones encompassing the coat protein gene, two clones consisting of 14000 bases of the CTV genome (Calvert, 1987) and the cDNA library prepared in this study. The work described here does provide an important insight into the translation strategy of CTV. Previously, immunoprecipitation studies of in vitro translation products suggested that the translation strategy of the closteroviruses is similar to that of the potyvirus group, in which a large polyprotein undergoes autocatalytic cleavage to yield the individual viral proteins (Dougherty & Hiebert, 1985; Dougherty & Carrington, 1988). For example, antibody to the coat protein immunoprecipitated a large (approximately 200K) polypeptide in addition to the 26K coat protein product from in vitro translation products of CTV RNA (Lee & Calvert, 1987). Similarly, a large (250K) polypeptide was immunoprecipitated from translation products of RNA from beet yellows virus (BYV) by an antibody to BYV coat protein (Karasev et al., 1989). However, the results of the present study indicate that the CTV coat protein is not translated as part of a polyprotein. Nucleotide sequences of the region upstream of the ORF of the coat protein gene have revealed many in-frame termination codons, indicating that translation must initiate from the coat protein gene. This conclusion is further supported by the recent work of Dulieu et al. (1990), who reported that the CTV coat protein could be translated from an 800 bp dsRNA. It remains to be determined whether the other CTV genes are translated from internal cistrons. The recent cloning and sequencing of another closterovirus, apple chlorotic leaf spot virus, has identified three putative ORFs, indicating that this virus is also polycistronic (German et al., 1990).

Finally, the experiments reported here indicate that the coat protein recovered from purified virus lacks amino-terminal peptides apparently present in the plant (Fig. 5). This conclusion is consistent with the observations of Guerri et al. (1990), who recently reported that semi-purified preparations of CTV have a major coat protein band that is larger than that of cpl of highly purified virus. They also suggested that cpl is a degradation product that is produced during purification. In their study, a MAb (3DF1) reacted with the major coat protein, but not with cpl or cp2. Our analyses suggest that MAb 3DF1 recognizes an epitope on the amino terminus of the intact coat protein. The fact that such proteolysis is possible during isolation has important consequences for studies in which isolated virus is used for biological analyses. Future efforts should explore ways in which post-extraction degradation of the coat protein can be avoided.

Our efforts are focused upon using this cloned coat protein gene to determine whether coat protein-mediated protection will protect citrus plants against CTV infection. Recent advances in the genetic transformation of citrus plants makes this experiment possible (Hidaka et al., 1990; Moore et al., 1989; Vardi et al., 1990).

The authors thank Dr Steve Garsney and Dr Ernest Hiebert for excellent advice, helpful discussions and for providing many of the materials essential for these studies. We thank Drs Lee Calvert, Richard Lee and Ernest Hiebert for the gift of clones P5D8 and P6A6 and Dr Richard Lee for advice and assistance in virus isolation. We thank Ernest Hiebert, Steve Garsney and Gloria Moore for critical reading of the manuscript and Dr Nancy Denslow and Benne Parten at the University of Florida ICBR Protein Chemistry Core for amino acid composition and sequence analysis. We thank Dr Ernesto Almira of the ICBR DNA Sequencing Core for sequence analysis, Dr Vivien Chang at the ICBR Biological Computing Core for helpful advice and Jace Dinehart of the ICBR DNA Synthesis Core for preparation of
primers. Support for the core facilities was provided by the Division of Sponsored Research and the ICRR, and that for the Protein Chemistry facility was additionally provided by an NIH Shared Instrument Grant, an NSF grant, the Department of Biochemistry and Molecular Biology, and the College of Medicine. The DNA Synthesis Facility was additionally supported by the Department of Immunology and Medical Microbiology and an NSF grant. This work was supported in part by a grant from the Florida High Technology and Industry Council to K.-C. and USDA-ARS/IFAS Cooperative Agreement 58-43YK-0-008. Florida Agricultural Experiment Station Journal Series No. R01107.

References


(Received 1 October 1990; Accepted 11 February 1991)