STUDIES OF PLANT PATHOGENIC RNAs:

VIRUSES

VIROIDS

SATELLITES

Chris Davies, B.Sc. (Hons.)

Department of Biochemistry,

The University of Adelaide,

ADELAIDE, SOUTH AUSTRALIA.

Thesis submitted to the University of Adelaide in fulfillment of the requirements for the degree of Doctor of Philosophy.

November, 1988.

CONTENTS

STATEMENT

This thesis contains no material which has been previously submitted for an academic record at this or any other University and is the original work of the author, except where due reference is made in the text. I consent to this thesis being made available for photocopying and loan.

Chris Davies

ACKNOWLEDGEMENTS

I wish to thank Professor W.H. Elliott for permission to undertake these studies in the Department of Biochemistry and my supervisor Professor Bob Symons for his help, advice and encouragement. The following people also deserve my gratitude: Dr. R.I.B. Francki for encouraging me to take up academic life and for his continued support; Dr. Jim Haseloff for his humour, help and for reading the relevant parts of this thesis; Drs. Paul Keese, Ali Rezaian, and Tony "turtle" Forster for technical advice, discussions and friendship; Professor Geid Jaspars for advice and great postcards; the many other inmates of the Symon's group past and present, including Peter Rathjen, Alex Jeffries, Candice Sheldon, Jane Visvader; Drs. Jason "lizard" Loveridge, Rhett Swanson and Iain "elbows" Borthwick for their alcoholic capacity and sporting adventures; Dr. Peter Waterhouse for backup supplies of sSCMoV RNAs; Dr. Filip Lim for practical jokes and practical advice; Mr. Dennis Talford for plants; the ladies in the office; all the friendly staff of Bresatec; Dr. Derek Skingle for Chemistry lessons; Brian Denton for workshop skills and fishing stories.

I wish to especially thank Tammy Greatrex for technical assistance, cheek and for her terrific help with diagrams, Lesley Crocker for being herself and for help with photography and figures and Jenny Cassady and Sharon Fruend for technical assistance.

I am indebted to Julie Pocock who typed this thesis for her friendly approach and excellent work.

My most special thanks go to my friends and parents and to Catherine and Anna for their love and steadfastness.

ABBREVIATIONS

AMV	alfalfa mosaic virus
ArMV	arabis mosaic virus
ASBV	avocado sunblotch viroid
BSA	bovine serum albumin
CAT	chloramphenicol acetyltransferase
cDNA	complementary DNA
CMV	cucumber mosaic virus
CP	coat protein
ds	double-stranded
DTT	dithiothreitol
3'E	3' end
5'E	5' end
EDTA	ethylenediaminetetraacetate
3'F	3' fragment
5'F	5' fragment
FPLC	fast protein liquid chromatography
HPLC	high performance liquid chromatography
Kb	kilobase
LGT	low geling temperature
LTSV	lucerne transient streak virus
NTP	nucleoside-5'-triphosphate
PEG	polyethylene glycol
RF	replicative form
RNase	ribonuclease
SCMoV	subterraneun clover mottle virus
SDS	sodium dodecyl sulphate
SNMV	solanum nodiflorum mottle virus
ss	single-stranded
TMV	tobacco mosaic virus
Tris	tris(hybroxymethyl)aminomethane
tRNA	transfer RNA
TRSV	tobacco ringspot virus
TSV	tobacco streak virus
VTMoV	velvet tobacco mottle virus

DEFINITIONS AND NOMENCLATURE

The group of satellite RNAs previously called virusoids (Haseloff *et al.*, 1982) are referred to as satellites not virusoids in this work. Satellite RNAs are denoted by a lower case s prefix e.g. sCMV is the satellite RNA of cucumber mosaic virus. As the satellite and viroid RNAs discussed in this thesis do not appear to have mRNA activity the predominant RNA from is termed the "plus" RNA and its complement is referred to as the "minus" RNA. Viroid and satellite RNAs are numbered using the system given in the relevant original publications.

Clones containing viral, viroid or satellite cDNAs are named according to the following convention. As an illustration the clone named" dimeric Sau (+) ASBV pSP64" refers to a dimeric ASBV insert (tandem repeat) in the pSP64 vector. The site of cloning relative to the ASBV sequence is referred to in this case as the Sau3A I site, and the polarity of the RNA transcript produced by SP6 RNA polymerase is given (in this case plus).

SUMMARY

This thesis contains work aimed at furthering the understanding of the stucture and function of certain plant pathogenic viruses and RNAs. The RNA3 of cucumber mosaic virus, Q strain (Q-CMV), a virus with a plus-sense tripartite RNA genome, has been resequenced to correct errors in previously published data. Analysis of this sequence has resulted in extensive changes to the sequences of the two proteins (the 3a and coat proteins) encoded by RNA3. The relationship between CMV and a number of other plant viruses has been analyzed at both the nucleic acid and protein levels. Evidence is presented that indicates a role for the 3a protein in cell to cell movement of the infectious entity. Regions of the RNA which potentially control synthesis of RNAs 3 and 4 (the subgenomic coat protein messenger) have been determined.

Viroids (e.g. avocado sunblotch viroid (ASBV)) and satellite RNAs (e.g. the satellites of CMV (sCMV), lucerne transient streak virus (sLTSV), subterranean clover mottle virus (sSCMoV) and tobacco ringspot virus (sTRSV)) are small, single-stranded plant pathogenic RNAs. ASBV, sLTSV and sTRSV RNAs have previously been shown to undergo site-specific self-cleavage *in vitro*. In this work the self-cleavage of ASBV, sLTSV and sSCMoV RNAs were investigated.

1) Self-cleavage of ASBV plus and minus RNAs occurs by a double-hammerhead secondary structure, not the previously proposed single-hammerhead structure. This has implications for the proposed rolling-circle replication cycle and also offers a mechanism to prevent futile self-cleavage of monomers.

2) The nature of the structural features which determine whether cleavage occurs by a single- or double-hammerhead structure were analyzed using mutant RNAs.

3) Full-length randomly-mutated sLTSV and sCMV cDNAs were cloned. A number of the sLTSV mutants have been sequenced and the effect of mutation on self-cleavage *in vitro* determined.

(4) Two sSCMoV RNAs (which contain large sequence re-arrangements) were sequenced. Self-cleavage *in vitro* was demonstrated for plus but not minus RNAs. Northern analysis suggests that these RNAs also replicate by a rolling-circle model.

CHAPTER 1

GENERAL INTRODUCTION

1. GENERAL INTRODUCTION

1.1 RNA PLANT VIRUSES

Plant viruses are a diverse group of pathogens consisting of protective protein coats (which sometimes include lipids) surrounding a nucleic acid genome. The virions can be of a variety of shapes including, icosohedral (e.g. cucumber mosaic virus (CMV)), rod-shaped (e.g. tobacco mosaic virus (TMV)) and bacilliform (e.g. alfalfa mosaic virus (AMV)). The nucleic acid genome of these viruses can consist of either double-stranded (ds) or single-stranded (ss), RNA or DNA. By far the greater proportion of known plant viruses have plus-sense ssRNA genomes. The genomes of these viruses may be a single RNA species (i.e. a monopartite genome, e.g. TMV) or segmented (e.g. CMV which has a tripartite genome).

Despite obvious differences in genome structure and particle morphology plus-sense ssRNA viruses appear to be remarkably similar in the types of proteins they encode and in the organization and expression of the genes encoding these proteins. This has led to the proposal that plus sense ssRNA viruses can be placed in only two "supergroups", i.e. "picona-like" viruses and "Sindbis-like" viruses (Goldbach and Wellink, 1988).

1.2 THE TRICORNAVIRIDAE

Of the plant viruses having a divided plus-sense ssRNA genome there are four groups that together have been proposed to form a family of viruses, the Tricornaviridae (van Vloten-Doting *et al.*, 1981). These four groups, the Bromoviruses (type member brome mosaic virus, BMV), the Cucumoviruses (type member CMV) the Ilarviruses (type member tobacco streak virus, TSV) and AMV, have tripartite genomes encapsidated in small polyhedral or bacilliform particles. Part

of the work described in this thesis involves CMV, a virus that has been much used as a model system for a variety of investigations.

CMV shares a number of basic features with other members of the Triconaviridae (Francki, 1985b; Symons, 1985). These include (a) similarly sized RNAs encoding translation products of similar size, (b) an untranslated 3'-terminal region conserved within the RNAs of each virus, (c) 5' M^7G caps on all four RNAs, (d) a dicistronic RNA3, the 3'-proximal cistron encodes the coat protein and is expressed via a subgenomic RNA, RNA4, and (e) only the three largest viral RNAs (labelled 1 to 3 in order of decreasing size) are required for infection. In addition to these overall similarities, comparisons of the nucleic acid and putative protein sequences of AMV, BMV and CMV RNAs 1 and 2 have revealed close evolutionary links within the Triconaviridae and somewhat more distant links with TMV (Cornelissen and Bol, 1984; Haseloff *et al.*, 1984; Rezaian *et al.*, 1984, 1985).

All three RNAs of the Q strain of CMV (Q-CMV) i.e. RNA1 (3389 nucleotides) RNA2 (3035 nucleotides) and RNA3 (2193 nucleotides) have been sequenced (Gould and Symons, 1982; Rezaian *et al.*, 1984, 1985).

1.3 SATELLITE RNAs

In addition to the three genomic RNAs and the subgenomic coat protein messenger other, small, linear ssRNAs are encapsidated in certain CMV isolates. These are satellite RNAs (reviewed by Murant and Mayo, 1982; Francki, 1985a, 1987; Simon, 1988) as they require a helper virus for replication and encapsidation. CMV satellite RNAs have only limited sequence homology with their helper virus and do not appear to encode any proteins (Garcia-Arenal *et al.*, 1987; Collmer and Kaper, 1988). These satellite RNAs may either exacerbate or attentuate CMV symptoms, attenuation being more common. To simplify the nomenclature of satellite RNAs, in this work they will be prefixed by a lower case s, thus the satellite

RNA of the Q-strain of CMV will be denoted as Q-sCMV RNA. Little is known regarding the interaction between satellites and their helper viruses. Linear dsRNAs believed to result from RNA replicative processes have been found for both CMV genomic and satellite RNAs (Kaper and Diaz-Ruiz, 1977; Takanami *et al.*, 1977; Bar-Joseph *et al.*, 1983; Rosner *et al.*, 1983; Collmer and Kaper, 1985) and it is thought that they may share replicative machinery.

A large number of CMV satellite RNAs arising from different isolates have been sequenced and have been shown to be closely related (Kaper *et al.*, 1988a). These RNAs can form structures with a signficant degree of base-pairing (Gordon and Symons, 1983; Garcia-Arenal *et al.*, 1987) but this base-pairing is considerably less than that found in some other small, infectious plant RNAs (e.g. viroids).

Apart from CMV other viruses also support the replication of small satellite RNAs. These viruses include members of the Nepovirus group (viruses with bipartite genomes (Harrison and Murant, 1977)) and putative members of the Sobemovirus groups (viruses with monopartite genomes (Hull, 1988)). Of the Nepovirus satellite RNAs, those of tobacco ringspot virus (TRSV) and arabis mosaic virus (ArMV) are of particular interest here. Both of these RNAs have been sequenced (Buzayan *et al.*, 1986c; Kaper *et al.*, 1988b) and are significantly homologous in some regions. For sTRSV it has been demonstrated that the linear form of the plus RNA is encapsidated (Kiefer *et al.*, 1982) but circular forms exist *in vivo* (Linthorst and Kaper, 1984).

The satellite RNAs discussed above can be divided into three groups on the basis of the RNA forms encapsidated and in plant tissue extracts. These groups are, 1) those satellites (e.g. CMV) that do not appear to have circular form, 2) those satellites (e.g. the satellite of lucerne transient streak virus (sLTSV)) whose RNAs are predominately circular in virions and in plant tissue extracts, and 3) those satellites

(e.g. sTRSV) whose RNAs are predominately linear in virions but are largely circular in plant tissue extracts.

The group of putative Sobemoviruses that contain satellite RNAs includes velvet tobacco mottle virus (VTMoV; Randles *et al.*, 1981) solanum nodiflorum mottle virus (SNMV; Gould and Hatta, 1981) LTSV (Tien-Po *et al.*, 1981) and subterraneun clover mottle virus (SCMoV, Francki *et al.*, 1983b). These viruses are similar to Sobemoviruses in many respects, however they differ in that they contain a unique class of satellite RNAs (Francki *et al.*, 1985; Francki, 1987; Keese and Symons, 1987).

The complete nucleotide sequences of the satellite RNAs of VTMoV, SNMV (Haseloff and Symons, 1982) and LTSV (Keese *et al.*, 1983) are known. These RNAs are single-stranded covalently-closed circular molecules of 300 to 400 nucleotides which possess a high degree of intramolecular base-pairing and lack conserved open reading frames. These RNAs were originally termed virusoids (Haseloff *et al.*, 1982) as it was thought that they were not satellite RNAs but in fact part of the viral genome (Gould *et al.*, 1981). They have subsequently been shown to be true satellite RNAs (Francki *et al.*, 1983a; Jones *et al.*, 1983, Jones and Mayo, 1983, 1984; Paliwal, 1984; Francki *et al.*, 1986) and so will be referred to as such in this work. Thus in line with the nomenclature outlined above the satellite of LTSV, for example, will be termed sLTSV not vLTSV previously.

1.4 VIROIDS

A further group of small plant pathogenic RNAs are the viroids, which are unencapsidated, covalently-closed circular, ssRNAs of between 200 and 400 nucleotides that infect a range of flowering plant species (reviewed by Diener, 1983, 1987; Sanger, 1984; Riesner and Gross, 1985). Like the satellite RNAs mentioned

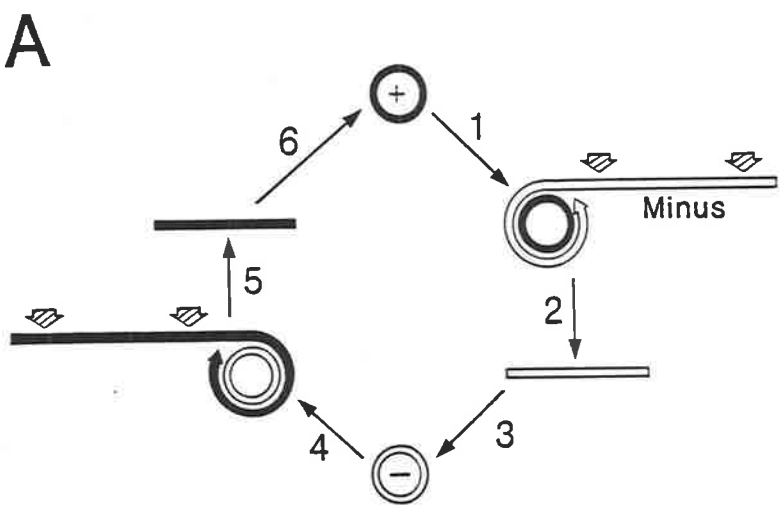
above they appear to lack mRNA activity but unlike these RNAs they are not dependent upon a helper virus for replication. As such they are the only known pathogens that consist entirely of nucleic acid.

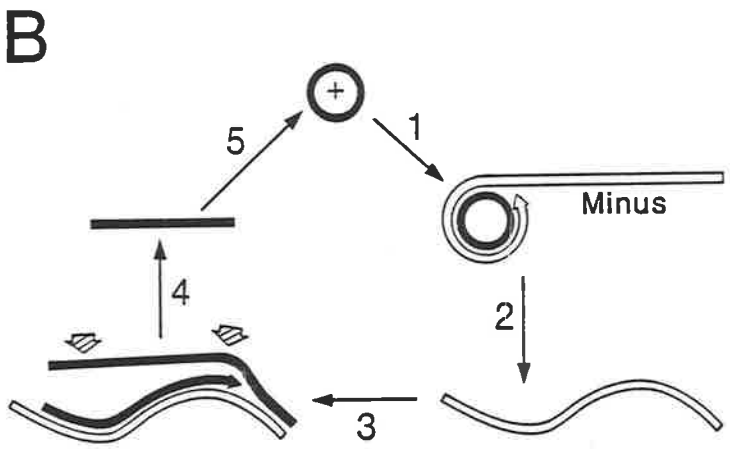
The complete nucleotide sequence of a number of viroids has been determined (Puchta *et al.*, 1988 and references therein). This has enabled them to be classified on the basis of sequence and structural similarity. A number of different classification schemes have been proposed (Rienser and Gross, 1985; Keese and Symons, 1985, 1987; Hashimoto *et al.*, 1987; Koltunow and Rezaian, 1988) but in all the schemes so far devised avocado sunblotch viroid (ASBV) has been placed in a separate group. The features that distinguish ASBV from the other viroids are its A/U rich nucleotide composition, the lack of sequence homology with other viroids and the presence of nucleotide sequences, absent from other viroids, shown to be involved in RNA self-cleavage *in vitro*.

1.5 RNA SELF-CLEAVAGE AND ROLLING-CIRCLE REPLICATION

The circular nature of viroids and the satellite RNAs described above and the presence of longer-than-unit-length forms of these RNAs in infected tissues (and virions for the satellites) suggest that these RNAs are replicated by rolling-circle mechanisms (Branch *et al.*, 1981; Bruening *et al.*, 1982; Owens and Diener, 1982; Kiefer *et al.*, 1982; Branch and Robertson, 1984; Ishikawa *et al.*, 1984; Hutchins *et al.*, 1985; Symons *et al.*, 1985). In these models longer-than-unit-length RNA transcripts are cleaved in a site-specific manner to monomeric units that are then circularized (Fig. 1.1). These circularized RNAs are in turn transcribed, the transcripts are cleaved and then circularized to complete the cycle. In the case of minus sTRSV RNA, circularization may be by reversal of the self-cleavage reaction (Buzayan *et al.*, 1986a,d) rather than by a plant enzyme(s). sSNMV and sVTMoV have a slightly more complex replication cycle as only the plus-strands appear to be

Fig. 1.1 Rolling circle models for the replication of viroid and satellite RNAs as per Forster and Symons (1987a). (A) A circular monomeric plus RNA of sLTSV, sTRSV or ASBV is transcribed (1) to yield a greater-than-unit-length minus RNA that self-cleaves (at the sites indicated by hatched arrows) to yield a monomeric minus RNA (2). This RNA is circularized (3) and transcribed (4) to yield a greater-than-unit-length plus RNA that self-cleaves to give a monomeric plus RNA (5) that is then circularized (6) thus completing the cycle. (B) The model for sSNMV and sTVMoV is the same as for (A) except that no cleavage of the minus RNA occurs. This results in a greater-than-unit-length minus RNA acting as a template for the production of plus RNA transcripts (3).


A
6
1
Minus
5
2
4
3


B
5
1
Minus
4
2
3

capable of cleavage and therefore plus-strand synthesis is from longer-than-unit-length minus templates (Fig. 1.1) (Chu *et al.*, 1983; Forster and Symons, 1987a).

There is considerable understanding of the self-cleavage of some of the satellite and viroid RNAs. Self-cleavage has been demonstrated *in vitro* for the RNAs of ASBV (Hutchins *et al.*, 1986) sTRSV (Buzayan *et al.*, 1986a,b; Prody *et al.*, 1986) and sLTSV (Forster and Symons, 1987a). This reaction is dependent on the presence of divalent metal ions and gives rise to 5'-hydroxyl and 2',3'-cyclic phosphodiester termini (Buzayan *et al.*, 1986a,c; Hutchins *et al.*, 1986; Prody *et al.*, 1986; Forster and Symons, 1987a). So called "hammerhead" secondary structures have been proposed for the active self-cleavage structures of plus and minus ASBV, sLTSV and plus sTRSV, sSNMV, sVTMoV RNAs and a number of conserved nucleotides proximal to the self-cleavage site identified (Hutchins *et al.*, 1986; Forster and Symons, 1987a). Cleavage of the minus RNA of sTRSV appears to be unique as the hammerhead secondary structure and conserved sequences found in the other self-cleaving RNAs are not found in this molecule. Also cleavage of the minus sTRSV RNA is readily reversible *in vitro* (Buzayan *et al.*, 1986a,d) an effect that has not been reported for the other self-cleaving satellite or viroid RNAs which cleave by a hammerhead structure.

Self-cleavage of RNAs by the hammerhead model is not restricted to the plant kingdom. It has recently been demonstrated that transcripts of newt, Notophthalmus viridescens, satellite DNA self-cleave *in vitro* by a structure very similar to that proposed for plus ASBV RNA (Epstein and Gall, 1987; Koizumi *et al.*, 1988).

The rolling-circle model of replication requires that cleavage of greater-than-unit-length transcripts occurs *in vivo*. There are a number of indications that the self-cleavage reactions reported to occur *in vitro* are relevant *in vivo*. First, the sequences required for self-cleavage *in vitro* are conserved in a number of different satellite (and

a viroid) RNAs, and therefore are likely to perform an essential function *in vivo*. Second, sVTMoV and sSNMV RNAs isolated from virions have a 2'-phosphomonoester, 3',5'-phosphodiester linkage at the predicted site of self-cleavage (Kiberstis *et al.*, 1985). This linkage is presumed to arise from ligation of the 5'-phosphate and 2',3'-cyclic phosphodiester terminal groups formed as a consequence of self-cleavage (Kiberstis *et al.*, 1985). Third, self-cleavage of the plus sTRSV RNA *in vitro* (Prody *et al.*, 1986) gives rise to terminal sequences that correspond to those of the linear RNA isolated from virions (Buzayan *et al.*, 1986c). Fourth, the self-cleavage of these RNAs *in vitro* occurs under conditions likely to also be present *in vivo*. Fifth, sTRSV RNAs expressed in transgenic tobacco plants undergo cleavage in a manner that is consistent with the self-cleavage reactions demonstrated *in vitro* (Gerlach *et al.*, 1987).

1.6 AIMS

Despite extensive sequence and structural analysis of the satellite and viroid RNAs, their replication, which involves interaction with the host plant, and in the case of satellite RNAs also the helper virus, remains poorly understood. The work presented in this thesis was aimed at giving further insight into the structure and function of these RNAs. In particular the aims were, (1) to resolve the secondary structure responsible for the self-cleavage of ASBV RNAs and to relate this to their replication, (2) to develop systems for the production and assay of mutant satellite RNAs, (3) to characterize satellite RNA species in addition to those already reported and (4) to further charactertise CMV, the helper virus for sCMV RNA.

CHAPTER 2

MATERIALS AND METHODS

2.1 MATERIALS

2.1.1 Reagents, enzymes and isotopes

Synthetic oligonucleotides (including BamH I-linkers) were supplied by Bresatec (Adelaide) in crude form. General laboratory reagents were of analytical reagent grade. Bovine serum albumin, DNase I (from bovine pancreas), Klenow fragment of DNA polymerase I, SP6 RNA polymerase, T4 DNA ligase, T4 polynucleotide kinase, α-^{32}p-dATP, α-^{32}P-dCTP, α-^{32}P-UTP, γ-^{32}P-ATP (3000 Ci/mmole) and the vector DNAs, pUC19, pSP64, pSP65 were from Bresatec. Ampicillin 3'-CMP *E. coli* tRNA, dNTP's, NTP's, ddNTP's, nuclease S1, RNase A and RNase T were from the Sigma Chemical Co. AMV reverse transcriptase was purchased from Molecular Genetics Inc. RNase U_2 was obtained from Calbiochem, calf-intestinal phosphatase, endopeptidase lysine-C from Boehringer Mannheim, T4 RNA ligase and T4 RNA polymerase from Pharmacia, pGem-1 and -2 vectors, T7 RNA polymerase from Promega Biotech. Restriction endonucleases were obtained from New England Biolabs. Nitrocellulose for hybridisation blots was obtained from Schleicher and Schuell. "Geneclean" kits for DNA extraction from agarose were purchased from Bio-101 (La Jolla). pIBI76 vector and helper phage were obtained from International Biotechnology, Inc. RNase Phy M from Physarum polycephalum was prepared by Dr. J. Haseloff by the method of Donis-Keller (1980).

2.1.2 Bacterial Strains

E. coli GM119: F^-, dam 3, dcm 6, met B1, gal K2, gal T22, lac Y1, tsx 7, supE44.

E. coli JM101: Δ(lac-pro), F^1lacI^q ZΔM15, traD1.

E. coli MC1061: araD139, Δ(ara,leu)7697, ΔlacX74, gal U^-, gal K^-, hsr^-, hsm^+, strA.

2.1.3 Media and Solutions

L-Broth: 1% (w/v) bacto-tryptone (Difco), 0.5% (w/v) yeast extract (Difco), 1% (w/v) NaCl (pH 7.0). When appropriate ampicillin (100 µg/ml) was added.

2 X YT broth: 1.6% (w/v) tryptone, 1% (w/v) yeast extract, 0.5% (w/v) NaCl (pH 7.0).

Unless otherwise stated all other solutions and media were prepared as described by Maniatis *et al.* (1982).

2.2 METHODS

2.2.1 Preparation of M13 RF DNA

Bacteriophage M13 RF DNA was prepared essentially as described by Yanisch-Perron *et al.* (1985). When highly pure DNA was required, partially purified DNAs were run on CsCl gradients (Maniatis *et al.*, 1982).

2.2.2 Preparation of plasmid DNA

Small scale (1 ml) preparations of plasmid DNA were performed as described by Birnboim and Doly (1979) except that the incubations on ice were carried out for 10 minutes only. The final DNA pellets were redissolved in 40-50 µl of 0.1 mM EDTA. Larger scale preparations (500 - 1000 ml) were done by the alkali lysis method (Maniatis *et al.*, 1982). In some preparations the CsCl gradient step was replaced by RNase A digestion and phenol/chloroform extraction.

2.2.3 Agarose gel electrophoresis and extraction of nucleic acids

RNAs and DNAs were run in TAE (40 mM Tris-acetate, 3 mM EDTA, pH 8.0) buffered horizontal agarose gels whose concentration depended on the size of the fragments to be resolved (Maniatis *et al.*, 1982). Samples were prepared for loading by the addition of a 1/10 volume of glycerol load buffer (50%, v/v, glycerol, 0.5 mM EDTA, 0.02% bromophenol blue, 0.02% xylene cyanol). Gels made with low geling temperature (LGT) agarose were run at 4°. Bands were visualized by staining with 0.5 µg/ml ethidium bromide followed by rinsing with water. Small RNA and DNA molecules were extracted by electroelution (Maniatis *et al.*, 1982), or in the case of (low geling temperature) LGT agarose gels by the following method. A maximum of 140 µl of gel slice was melted in a 1.5 ml Eppendorf tube at 65°C, 2 volumes of preheated extraction buffer (50 mM Tris-Hcl (pH 8.0), 0.5 mM EDTA) was added, followed by extraction with 140 µl of Tris-saturated phenol. After brief centrifugation the supernatant was removed, re-extracted with phenol and the nucleic acids precipitated at room temperature for 30 minutes with two volumes isopropanol, 0.4 volumes 5 M ammonium acetate. The precipitate was collected by centrifugation for 15 minutes, washed with 70% ethanol and dried *in vacuo*. For larger DNAs (i.e. >0.5 kb) extraction was performed using a Geneclean kit under conditions described by the manufacturer.

2.2.4 End-filling of 5'-overhangs with Klenow enzyme

DNA with 5'-overhanging ends was blunted using the Klenow fragment of DNA polymerase I as described by Maniatis *et al.*, (1982).

2.2.5 Removal of 3'-overhanging nucleotides with T4 DNA polymerase

The DNA to be treated (1-3 pmol of 3' ends) was incubated in a volume of 30 µl, containing 5 units of T4 DNA polymerase, 70 mM Tris-HCl (pH 7.4), 10 mM $MgCl_2$, 5 mM DTT and 500 mM of each of the four dNTPs, for 20 minutes at 37°C. The reaction was stopped by heating at 70°C for 10 minutes.

2.2.6 DNA vector preparation

Digestion with restriction endonucleases of double-stranded plasmid and M13 DNAs for use as vectors was carried out under the conditions specified in the New England Biolabs catalogue (as were all other restriction endonuclease digestions). Phosphatasing was carried out using calf intestinal alkaline phosphatase essentially as described by Maniatis *et al.*, (1982). The DNA was then purified on an agarose gel and the DNA extracted (a phenol extraction was performed at this stage if not already done as part of the extraction procedure). Following estimation of DNA concentration on an agarose gel the DNA was diluted to a final concentration of 20 or 40 ng/µl in 0.1 mM EDTA prior to use.

2.2.7 DNA ligations

In general, ligation of DNA fragments into vector DNA was carried out as described below. 20 ng of vector DNA was incubated with the desired insert fragment (at a molar ratio of approximately 1:3 (vector:insert)) in 50 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 10 mM DTT and 1 mM ATP with T4 DNA ligase. 0.02 units/µl of T4 DNA ligase was used for sticky-ended fragments and 0.1 units/µl for blunt-ended fragments. To increase the efficiency of ligation of blunt-ended fragments 5% PEG 6000 was included when required.

The specific ligation conditions required for dimer production are outlined in the individual chapters.

2.2.8 Transformation of **E. coli**

In general, when only moderate efficiencies of transformation were required, the simple transformation method of Hanahan (1984) was used. When higher transformation efficiencies were required (for JM101 and related strains) the standard high efficiency transformation protocol of Hanahan (1984) was used except that dimethyl formamide was substituted for dimethyl sulphoxide.

2.2.9 Preparation of single-stranded DNA markers

pUC19 plasmid DNA was digested with Hpa II and end-labelled with α-^{32}P-CTP using the Klenow fragment of DNA polymerase I as described by Maniatis *et al.* (1982). Prior to loading the markers were denatured by heating, at 100°C for three minutes, with an equal volume of formamide loading solution (2.2.11) (containing 50 mM sodium hydroxide) and snap cooled on ice.

2.2.10 **In vitro** synthesis of RNAs from plasmid DNA templates

Plasmid DNA templates were transcribed with either T7 or SP6 RNA polymerases. Linearized templates were prepared by endonuclease restriction under conditions recommended by the manufacturer followed by extraction of the DNA with an equal volume of phenol:chloroform (1:1, v/v) and ethanol precipitation. Non-radioactive transcription reactions containing approximately 0.1 µg/µl of template DNA, 0.5 units/µl RNA polymerase, 40 mM Tris-HCl (pH 7.5), 6 mM $MgCl_2$, 0.1 µg/µl bovine serum albumin, 10 mM DTT and 0.5 mM of each of the NTPs were incubated at 38°C for 45 to 90 minutes. Radioactive transcription reactions (for the

preparation of labelled transcripts for self-cleavage reactions or for use as probes) varied in containing 0.025 mM UTP and between 1 and 2.5 μCi/μl of α-^{32}P-UTP.

2.2.11 Polyacrylamide gel electrophoresis and purification of cDNAs and RNAs

RNAs and DNAs to be purified were added to an equal volume of formamide loading solution (95% formamide, 10 mM EDTA, 0.02% bromophenol blue, 0.02% xylene cyanol FF), heated at 80°C for 30 seconds and snap cooled on ice before loading on a 35 x 20 x 0.05 cm polyacrylamide gel. The gel percentage was usually 5% but varied depending on the size of the bands to be resolved (acrylamide: bisacrylamide = 25:1, w/w). Gels were run in 90 mM Tris-borate (pH 8.3), 2 mM EDTA (TBE). Non-radioactive bands were vizualized by staining in 0.05% toluidine blue followed by destaining in water. Nucleic acids were eluted from gel slices in 0.1% SDS, 1 mM EDTA, 10 mM Tris-HCl (pH 7.5), for 6-16 hours at 37°C and then precipitated by the addition of 3 M sodium acetate (pH 5) to 0.3 M and 2.5 volumes of cold ethanol. After storage on ice for 10-30 minutes the precipitate was collected by centrifugation in an Eppendorf microfuge for 15-30 minutes at 4°C. For very small molecules two volumes of ethanol and one volume of acetone were substituted for the 2.5 volumes of ethanol.

2.2.12 Self-cleavage of RNAs purified by polyacrylamide gel electrophoresis

Typically, gel purified RNAs in 1 mM EDTA were heated at 80°C for 1 minute, snap-cooled on ice for 5 minutes, mixed with an equal volume of ice-cold 10 mM $MgCl_2$, 10 mM Tris-HCl (pH 7.5) and incubated at 25°C for 10 minutes. Reactions were terminated by the addition of an equal volume of loading solution.

2.2.13 5'-^{32}P-labelling of RNA fragments

The RNA to be labelled was mixed with 100 µCi of γ-^{32}P-ATP, dried *in vacuo*, resuspended in 8 µl of water, heated at 80°C for 1 minute and snap cooled on ice. 2 µl of 5x T4 polynucleotide kinase buffer (125 mM Tris-HCl (pH 9.0), 50 mM $MgCl_2$, 50 mM DTT) and 1 µl (5 units) T4 polynucleotide kinase was added and the reaction incubated at 37°C for 30 minutes then stopped by heating at 70°C for 10 minutes.

2.2.14 3'-labelling of RNA fragments

Purified RNAs were labelled at the 3'-end by incubating them with ^{32}P-pCp and T4 RNA ligase. ^{32}P-pCp was prepared in the following manner. 20 µl of γ-^{32}P-ATP was dried and dissolved in 20 µl of 25 mM Tris-HCl (pH 9.0), 5 mM $MgCl_2$, 3 mM DTT, 5 mM 3'-CMP, 0.05 µg/µl BSA and 6 units of T4 polynucleotide kinase. This mix was incubated for 60 minutes at 37°C then the reaction was stopped by heating at 65°C for 10 minutes. To label RNA, 2 µl of the reaction mix, containing 5'-^{32}P-pCp, was incubated for 16 hours at 0°C with approximately 5 picomoles of RNA in a final volume of 20 µl containing 50 mM Hepes-KOH (pH 7.5), 15 mM $MgCl_2$, 3.3 mM DTT, 0.015 mM ATP, 5 ng/µl BSA, 5% (v/v) redistilled dimethyl sulphoxide and 2 units of T4 RNA ligase. 3'-labelled RNAs were then purified by polyacrylamide gel electrophoresis.

2.2.15 Enzymic sequencing of end-labelled RNAs

Partial enzymic hydrolysis methods were used to sequence purified 5'- or 3'-^{32}P-labelled RNAs. Partial digestions were carried out with RNase T_1, RNase U_2 and RNAase Phy M as described by Haseloff and Symons (1981).

In detail, dried aliquots of labelled RNAs, containing 40 µg of tRNA (as carrier), were dissolved in 7 µl sterile water and divided into 5 x 1 µl aliquots. These aliquoted RNAs were then digested under the following conditions in a final reaction volume of 10 µl.

(i) 10 mM sodium citrate (pH 5.0), 7 M Urea, 1 mM EDTA.

(ii) 1 mM $MgCl_2$, 1 µg/µl *E. coli* tRNA in dionized formamide.

(iii) 20 mM sodium citrate (pH 5.0), 7 M Urea, 1 mM EDTA, 10 units RNase T_1.

(iv) 20 mM sodium citrate (pH 5.0), 7 M Urea, 1 mM EDTA, 1 µl RNase Phy M extract.

(v) 20 mM sodium citrate (pH 3.5), 7 M Urea, 1 mM EDTA, 5 units RNase U_2.

Reaction (ii) was heated at 100°C for 90 seconds, the other reactions were incubated at 50° for 20 minutes then 2 µl of formamide load buffer was added to each. Samples were heated for 1 minute at 80°C, snap cooled on ice then run on a 20% polyacrylamide , 7 M Urea gel in TBE buffer.

2.2.16 DNA sequencing by the dideoxy chain termination method

Single-stranded M13 DNA was prepared by the method of Winter and Fields (1980). Sequencing was carried out by the dideoxy chain termination technique (Sanger *et al.*, 1980) essentially as described by Rezaian *et al.* (1985). To resolve difficult areas of sequence the following variations were used. (1) Sequencing using the Klenow fragment of DNA polymerase I was carried out at 50°C instead of at 37°C, (2) reverse transcriptase was used in place of the Klenow fragment of DNA polymerase I and sequencing was carried out at 42°C, (3) 30% formamide was included in the polyacrylamide gel mix to improve the denaturation properties.

2.2.17 Purification of oligonucleotides

The oligonucleotides used for site-directed mutagenesis and as primers for sequencing directly from viral RNAs were purified in the following manner. Approximately 40 µg of the crude oligonucleotide mixture was dried, dissolved in formamide loading buffer, heated for 1 minute at 80°C and snap cooled on ice prior to loading on a 4 cm wide well of a 15% polyacrylamide, 7 M urea gel (2.2.11). Bands were visualized by staining with 0.05% toludine blue followed by washing with sterile water. The slowest migrating band (presumed to be the full-length product of synthesis) was excised and eluted overnight at 37°C in 0.1% SDS, 1 mM EDTA, 10 mM Tris-HCl (pH 7.5) then precipitated by the addition of 3 M sodium acetate (pH 5) to 0.3 M and two volumes of ethanol, one volume of acetone. After storage for 4-16 hours at -20°C the precipitate was recovered by centrifugation in an Eppendorf microfuge for 30 minutes at 4°C.

2.2.18 Oligonucleotide-directed mutagenesis

Oligonucleotide-directed mutagenesis was performed as described by Zoller and Smith (1983). After completion of the mutagenesis reaction aliquots of 0.5, 1 and 2 µl of the mix were transformed into competent JM101 cells.

Screening plaques for mutants was done using the oligonucleotide hybridization procedure of Wood *et al.* (1985) which utilizes tetramethylammonium chloride.

CHAPTER 3

THE NUCLEOTIDE SEQUENCE OF Q-CMV RNA3 AND ITS ANALYSIS

3.1 INTRODUCTION

Previous to this work the nucleotide sequence of the entire genome of Q-CMV had been published (Gould and Symons, 1982; Rezaian *et al.*, 1984, 1985). However a number of irregularities were discovered in the nucleotide sequence of RNA3 by this author. cDNA prepared from Q-CMV RNA3 yielded patterns of restriction endonuclease fragments that were inconsistent with the published sequence. Subsequent sequencing of this cDNA cloned into M13 vectors confirmed that there were anomalies. This may, however, have been due to natural variation in the RNA from which the cDNA was made. As well as this direct evidence, the actual structure of Q-CMV as published, was difficult to rationalize. In the sequence of Gould and Symons (1982) the 3'-terminal region homologous between RNAs 1, 2 and 3, overlapped with the coat protein open reading frame. In RNAs 1 and 2 the open reading frames did not overlap with this region. Even though coat proteins are amongst the least conserved of viral proteins, the overlapping of this coding region with the conserved sequence appeared incongruous. The doubts raised by these observations seemed sufficient to justify re-examination of the RNA3 nucleotide sequence; this work is presented here.

3.2 METHODS

3.2.1 Preparation of cDNA clones

Q-CMV RNA was prepared (Peden and Symons, 1973) and RNAs 3 and 4 were purified by polyacrylamide gel electrophoresis (Gould and Symons, 1977). cDNA clones in phage M13mp9 DNA were prepared in two ways. (i) Q-CMV RNA 3 was primed randomly (Taylor *et al.*, 1976) and double-stranded cDNA was produced as described by Gould and Symons (1982). After digestion with FnuD II, the resultant fragments were purified (2.2.11) then ligated into dephosphorylated Sma I-

digested M13mp9 vector DNA (Messing and Vieira, 1982). (ii) The pBR322 clone, p2Tip2 of Gould and Symons (1982), contains an insert which extends from nucleotide 729 to 1784 of Q-CMV RNA 3. This insert was excised with Taq I, purified by agarose gel electrophoresis (2.2.3), and digested with FnuD II. The resulting fragments were cloned into dephosphorylated Sma I-cut M13mp9 DNA.

3.2.2 Sequencing of RNA and DNA

RNAs 3 and 4, and single-stranded M13 DNA containing cDNA to Q-CMV RNA 3, were sequenced by the chain termination technique (2.2.16). Synthetic oligodeoxynucleotide primers (15-18-mers) were used for sequencing from RNA.

3.2.3 Amino acid sequencing of coat protein

Coat protein was isolated (Francki *et al.*, 1966) from purified virus and was further purified by fast protein liquid chromatography (FPLC) using a Pharmacia Mono S HR 5/5 (cationic exchange) column. The purified protein appeared to be N-terminally blocked as it yielded no sequence data when amino acid sequencing was attempted. To produce peptides suitable for sequencing, the pure coat protein was digested to completion with endoproteinase lysine-C (Boehringer) in accordance with the manufacturer's recommendations. The resultant fragments were further purified by reverse-phase high performance liquid chromatography (HPLC) using a Nova-Pak C18 column (Waters Associates). A number of peptides prepared in this manner were analyzed on a model 470A Applied Biosystems protein sequencer and the resultant PTH amino acids were analyzed on a Waters Associates WISP 710 HPLC system. Chromatographs were produced by a Varian Vistar Data System.

3.2.4 Computer analysis of protein sequences

The DIAGON program (Staden, 1982) was used to analyze similarities between proteins and also nucleic acids at the sequence level. Global amino acid sequence alignment was done using the SEQA program (Needleman and Wunsch, 1970; Kanehisa, 1982). To compare the secondary structures of the putative proteins the predictive program of Novotny and Auffray (1984) was used. The parameters of Kyte and Doolittle (1982) were used for calculating the hydrophobicity profiles and the curves were smoothed 4-fold.

3.3 RESULTS

3.3.1 Strategy used for sequencing Q-CMV RNA 3

Much of Q-CMV RNA 3 was sequenced using M13 clones generated either directly from cDNA made *de novo* or by subcloning the p2Tip2 clone used in the original sequencing of Q-CMV RNA 3 (Gould and Symons, 1982). Some regions of the RNA which were not represented in the library of clones or were located where further confirmation of the sequence was required were resolved by sequencing with synthetic oligonucleotide primers on purified RNA 3 (Fig. 3.1). The entire sequence of the 2197 nucleotides of Q-CMV RNA 3 is given in Fig. 3.2.

Included in Fig. 3.2 are the regions of the coat protein (the underlined amino acids) for which amino acid sequence data were obtained. The amino acid sequence agreed completely with that for the putative coat protein predicted from the RNA sequence and confirms the coat protein reading frame. It is possible however that the N-terminal end of the coat protein is modified *in vivo* as is the case for BMV (Moosic *et al.*, 1983).

Fig. 3.1 Schematic diagram of the strategy used for sequencing Q-CMV RNA3. Deoxynucleotide primers used for dideoxy sequencing by transcription of the RNA are represented as solid triangles. The location of M13 DNA clones of plus (+) and minus (-) orientations are indicated; the vertical bars define the extent of individual clones. The RNA sequence at the 3' terminus has been published (Symons, 1974); no corrections to the original data were required

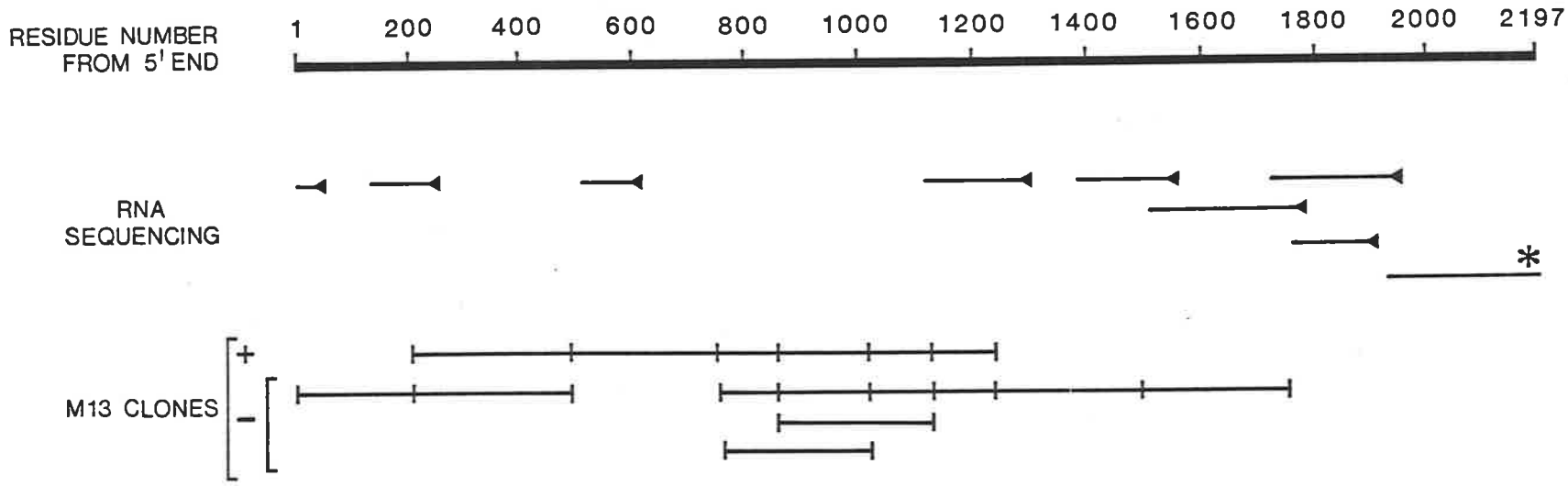
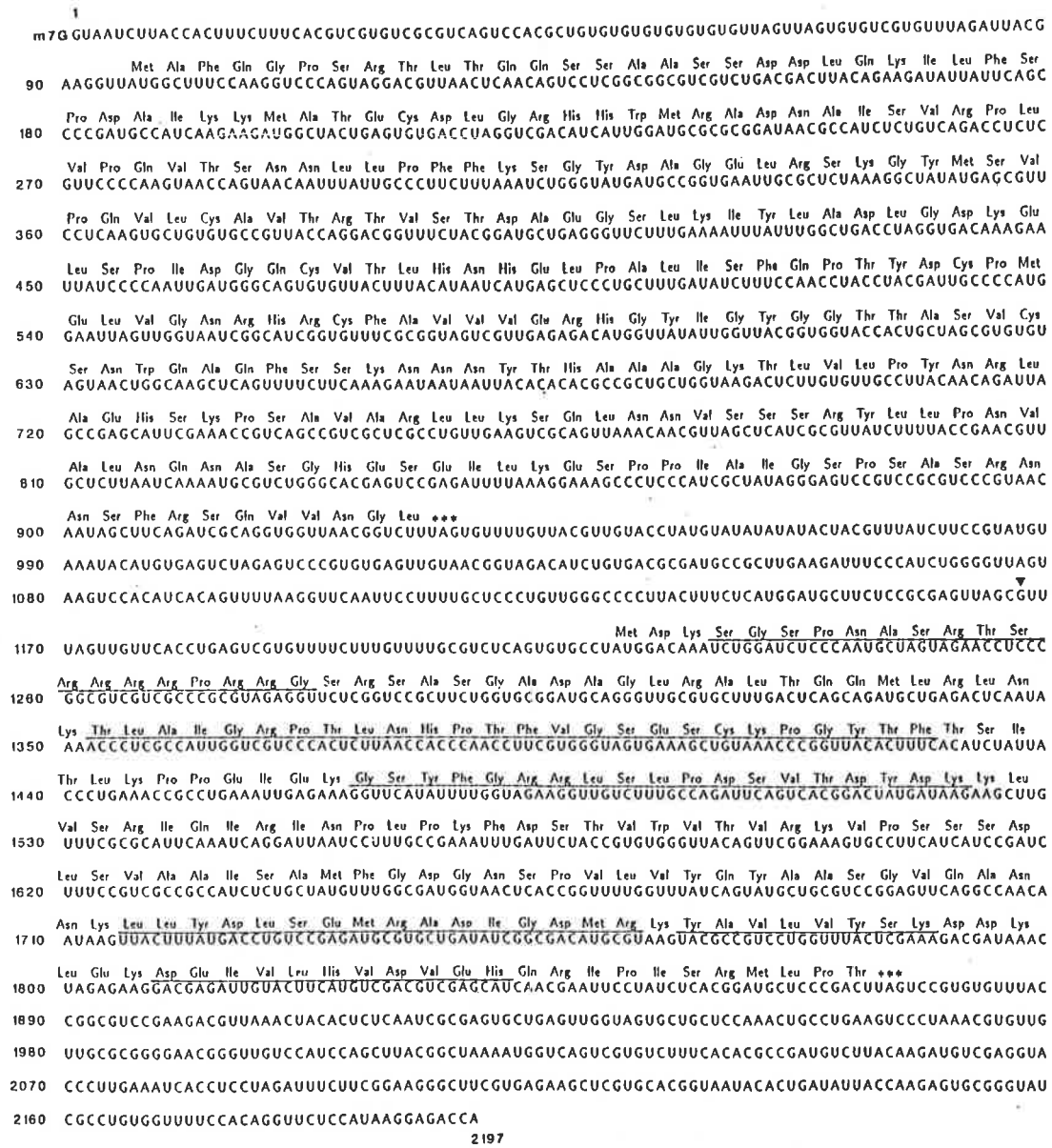

RESIDUE NUMBER
FROM 5' END
1
200
400
600
800
1000
1200
1400
1600
1800
2000
2197
RNA
SEQUENCING
*
M13 CLONES
+
−

Fig 3.2 Nucleotide sequence of Q-CMV RNA3 and the deduced amino acid sequences of the two long open-reading frames. The stop codons of these reading frames are identified as ***. Regions of the coat protein for which amino acid sequence data were obtained are underlined. The G nucleotide at position 1167 (labelled)is the 5'-terminal residue of RNA4.


1
m7G GUAAUCUUACCACUUUCUUUCACGUCGUGUCGCGUCAGUCCACGCUGUGUGUGUGUGUGUGUUAGUUAGUGUGUCGUGUUUAGAUUACG
Met Ala Phe Gln Gly Pro Ser Arg Thr Leu Thr Gln Gln Ser Ser Ala Ala Ser Ser Asp Asp Leu Gln Lys Ile Leu Phe Ser
90 AAGGUUAUGGCUUUCCAAGGUCCCAGUAGGACGUUAACUCAACAGUCCUCGGCGGCGUCGUCUGACGACUUACAGAAGAUAUUAUUCAGC
Pro Asp Ala Ile Lys Lys Met Ala Thr Glu Cys Asp Leu Gly Arg His His Trp Met Arg Ala Asp Asn Ala Ile Ser Val Arg Pro Leu
180 CCCGAUGCCAUCAAGAAGAUGGCUACUGAGUGUGACCUAGGUCGACAUCAUUGGAUGCGCGCGGAUAACGCCAUCUCUGUCAGACCUCUC
Val Pro Gln Val Thr Ser Asn Asn Leu Leu Pro Phe Phe Lys Ser Gly Tyr Asp Ala Gly Glu Leu Arg Ser Lys Gly Tyr Met Ser Val
270 GUUCCCCAAGUAACCAGUAACAAUUUAUUGCCCUUCUUUAAAUCUGGGUAUGAUGCCGGUGAAUUGCGCUCUAAAGGCUAUAUGAGCGUU
Pro Gln Val Leu Cys Ala Val Thr Arg Thr Val Ser Thr Asp Ala Glu Gly Ser Leu Lys Ile Tyr Leu Ala Asp Leu Gly Asp Lys Glu
360 CCUCAAGUGCUGUGUGCCGUUACCAGGACGGUUUCUACGGAUGCUGAGGGUUCUUUGAAAAUUUAUUUGGCUGACCUAGGUGACAAAGAA
Leu Ser Pro Ile Asp Gly Gln Cys Val Thr Leu His Asn His Glu Leu Pro Ala Leu Ile Ser Phe Gln Pro Thr Tyr Asp Cys Pro Met
450 UUAUCCCCAAUUGAUGGGCAGUGUGUUACUUUACAUAAUCAUGAGCUCCCUGCUUUGAUAUCUUUCCAACCUACCUACGAUUGCCCCAUG
Glu Leu Val Gly Asn Arg His Arg Cys Phe Ala Val Val Val Glu Arg His Gly Tyr Ile Gly Tyr Gly Gly Thr Thr Ala Ser Val Cys
540 GAAUUAGUUGGUAAUCGGCAUCGGUGUUUCGCGGUAGUCGUUGAGAGACAUGGUUAUAUUGGUUACGGUGGUACCACUGCUAGCGUGUGU
Ser Asn Trp Gln Ala Gln Phe Ser Ser Lys Asn Asn Asn Tyr Thr His Ala Ala Ala Gly Lys Thr Leu Val Leu Pro Tyr Asn Arg Leu
630 AGUAACUGGCAAGCUCAGUUUUCUUCAAAGAAUAAUAAUUACACACACGCCGCUGCUGGUAAGACUCUUGUGUUGCCUUACAACAGAUUA
Ala Glu His Ser Lys Pro Ser Ala Val Ala Arg Leu Leu Lys Ser Gln Leu Asn Asn Val Ser Ser Ser Arg Tyr Leu Leu Pro Asn Val
720 GCCGAGCAUUCGAAACCGUCAGCCGUCGCUCGCCUGUUGAAGUCGCAGUUAAACAACGUUAGCUCAUCGCGUUAUCUUUUACCGAACGUU
Ala Leu Asn Gln Asn Ala Ser Gly His Glu Ser Glu Ile Leu Lys Glu Ser Pro Pro Ile Ala Ile Gly Ser Pro Ser Ala Ser Arg Asn
810 GCUCUUAAUCAAAAUGCGUCUGGGCACGAGUCCGAGAUUUUAAAGGAAAGCCCUCCCAUCGCUAUAGGGAGUCCGUCCGCGUCCCGUAAC
Asn Ser Phe Arg Ser Gln Val Val Asn Gly Leu ***
900 AAUAGCUUCAGAUCGCAGGUGGUUAACGGUCUUUAGUGUUUUGUUACGUUGUACCUAUGUAUAUAUAUACUACGUUUAUCUUCCGUAUGU
990 AAAUACAUGUGAGUCUAGAGUCCCGUGUGAGUUGUAACGGUAGACAUCUGUGACGCGAUGCCGCUUGAAGAUUUCCCAUCUGGGGUUAGU
1080 AAGUCCACAUCACAGUUUUAAGGUUCAAUUCCUUUUGCUCCCUGUUGGGCCCCUUACUUUCUCAUGGAUGCUUCUCCGCGAGUUAGCGUU
Met Asp Lys Ser Gly Ser Pro Asn Ala Ser Arg Thr Ser
1170 UAGUUGUUCACCUGAGUCGUGUUUUCUUUGUUUUGCGUCUCAGUGUGCCUAUGGACAAAUCUGGAUCUCCCAAUGCUAGUAGAACCUCCC
Arg Arg Arg Arg Pro Arg Arg Gly Ser Arg Ser Ala Ser Gly Ala Asp Ala Gly Leu Arg Ala Leu Thr Gln Gln Met Leu Arg Leu Asn
1260 GGCGUCGUCGCCCCGCUAGAGGUUCUCGGUCCGCUUCUGGUGCGGAUGCAGGGUUGCGUGCUUUGACUCAGCAGAUGCUGAGACUCAAUA
Lys Thr Leu Ala Ile Gly Arg Pro Thr Leu Asn His Pro Thr Phe Val Gly Ser Glu Ser Cys Lys Pro Gly Tyr Thr Phe Thr Ser Ile
1350 AAACCCUCGCCAUUGGUCGUCCCACUCUUAACCACCCAACCUUCGUGGGUAGUGAAAGCUGUAAACCCGGUUACACUUUCACAUCUAUUA
Thr Leu Lys Pro Pro Glu Ile Glu Lys Gly Ser Tyr Phe Gly Arg Arg Leu Ser Leu Pro Asp Ser Val Thr Asp Tyr Asp Lys Lys Leu
1440 CCCUGAAACCGCCUGAAAUUGAGAAAGGUUCAUAUUUUGGUAGAAGGUUGUCUUUGCCAGAUUCAGUCACGGACUAUGAUAAGAAGCUUG
Val Ser Arg Ile Gln Ile Arg Ile Asn Pro Leu Pro Lys Phe Asp Ser Thr Val Trp Val Thr Val Arg Lys Val Pro Ser Ser Ser Asp
1530 UUUCGCGCAUUCAAAUCAGGAUUAAUCCUUUGCCGAAAUUUGAUUCUACCGUGUGGGUUACAGUUCGGAAAGUGCCUUCAUCAUCCGAUC
Leu Ser Val Ala Ala Ile Ser Ala Met Phe Gly Asp Gly Asn Ser Pro Val Leu Val Tyr Gln Tyr Ala Ala Ser Gly Val Gln Ala Asn
1620 UUUCCGUCGCCGCCAUCUCUGCUAUGUUUGGCGAUGGUAACUCACCGGUUUUGGUUUAUCAGUAUGCUGCGUCCGGAGUUCAGGCCAACA
Asn Lys Leu Leu Tyr Asp Leu Ser Glu Met Arg Ala Asp Ile Gly Asp Met Arg Lys Tyr Ala Val Leu Val Tyr Ser Lys Asp Asp Lys
1710 AUAAGUUACUUUAUGACCUGUCCGAGAUGCGUGCUGAUAUCGGCGACAUGCGUAAGUACGCCGUCCUGGUUUACUCGAAAGACGAUAAAC
Leu Glu Lys Asp Glu Ile Val Leu His Val Asp Val Glu His Gln Arg Ile Pro Ile Ser Arg Met Leu Pro Thr ***
1800 UAGAGAAGGACGAGAUUGUACUUCAUGUCGACGUCGAGCAUCAACGAAUUCCUAUCUCACGGAUGCUCCCGACUUAGUCCGUGUGUUUAC
1890 CGGCGUCCGAAGACGUUAAACUACACUCUCAAUCGCGAGUGCUGAGUUGGUAGUGCUGCUCCAAACUGCCUGAAGUCCCUAAACGUGUUG
1980 UUGCGCGGGGAACGGGUUGUCCAUCCAGCUUACGGCUAAAAUGGUCAGUCGUGUCUUUCACACGCCGAUGUCUUACAAGAUGUCGAGGUA
2070 CCCUUGAAAUCACCUCCUAGAUUUCUUCGGAAGGGCUUCGUGAGAAGCUCGUGCACGGUAAUACACUGAUAUUACCAAGAGUGCGGGUAU
2160 CGCCUGUGGUUUUCCACAGGUUCUCCAUAAGGAGACCA
2197

3.3.2 Overall structure of Q-CMV RNAs 3 and 4 and their coding capacity

The overall structures of Q-CMV RNAs 3 and 4 are shown in Fig. 3.3. RNA 3 contains two large open reading frames; the 5'-proximal open reading frame which encodes the 3a protein and the 3'-proximal open reading frame which encodes the coat protein. The coding region (840 nucleotides, including the stop codon) for the 3a protein is preceded by an untranslated region of 95 nucleotides. The putative 3a protein consists of 279 amino acids with a molecular weight of 30,353. An intercistronic region of 284 nucleotides separates the 3a and coat protein cistrons. The coat protein consists of 218 amino acids with a molecular weight of 24,247. The 3'-untranslated region, which contains sequences homologous to Q-CMV RNAs 1 and 2 (Rezaian *et al.*, 1984, 1985) is 321 nucleotides in length. RNA 4, the subgenomic message for the coat protein, consists of 1031 nucleotides located at the 3' end of RNA 3 (Fig. 3.3). The location of the 5' terminus of RNA 4 (Fig 3.2), is as previously reported (Gould and Symons, 1982). The revised nucleotide sequence of 2197 nucleotides (Fig. 3.2) differs from the previously published sequence of 2193 nucleotides (Gould and Symons, 1982) at 45 positions. These changes result in a reduction of the size of the open reading frames which encode the putative 3a and coat proteins. In addition, the predicted amino acid sequences of the 3a and coat proteins differ from those previously predicted in 6 and 31 positions, respectively. The secondary structure models proposed for RNA 3 by Gould and Symons (1982) are not substantially affected by the changes to the nucleotide sequence.

The sizes of Q-CMV RNAs 3 and 4, the length of their untranslated regions, and the sizes of the encoded proteins are similar to those of other members of the Tricornaviridae (Table 3.1). Apart from the open reading frames which encode the 3a and coat proteins, the largest AUG-initiated open reading frames in either the plus- or minus-sense Q-CMV RNA 3 molecules consist of 95 (nucleotides 1515-1802) and 41 (nucleotides 1130-1255) amino acids, respectively (stop codons are included in the

Fig. 3.3 Schematic diagram of the general structure of Q-CMV RNAs 3 and 4. Numbers above the lines refer to the number of nucleotides in each region and for the coding regions include the stop codons. Numbers below the lines refer to the distance in nucleotides from the 5' end of Q-CMV RNA3; those in brackets refer to the distance from the 5' end of RNA4. The coding regions are represented by boxes.

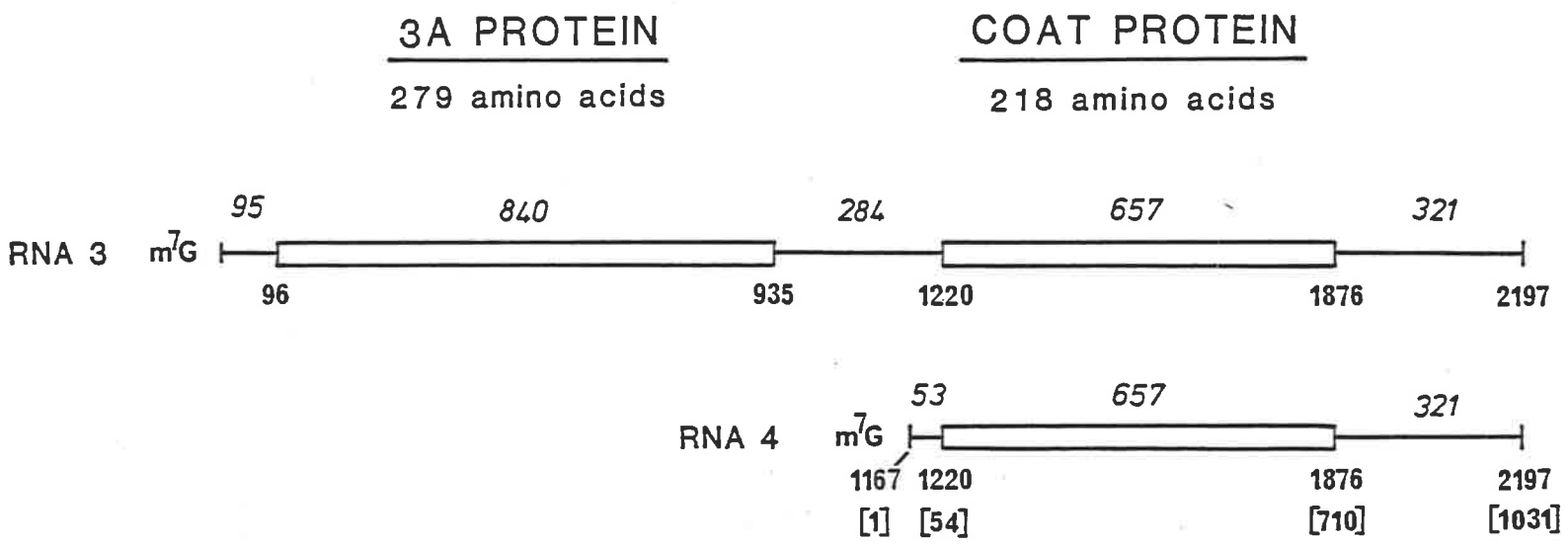

3A PROTEIN
279 amino acids
COAT PROTEIN
218 amino acids
RNA 3
m7G
95
840
284
657
321
96
935
1220
1876
2197
RNA 4
m7G
53
657
321
1167
1220
1876
2197
[1]
[54]
[710]
[1031]

TABLE 3.1

Comparison of the structure and translation products of various plant viral RNAs

VIRAL RNA	LENGTH OF REGION IN NUCLEOTIDES						PREDICTED SIZE OF TRANSLATION PRODUCT		
	TOTAL	5' UNTRANS-LATED	CODING[+]	INTER-CISTRONIC	CODING	3' UNTRANS-LATED	NO. OF AMINO ACIDS	MOL. WT.	REF.
Q-CMV RNA 3	2197	95	840	284	657[#]	321	279	30,353	This work
BMV RNA 3	2111-2117	91	912	241-247	570[#]	297	303	32,480	[a]
AMV RNA 3	2142	345	903	49	666[#]	179	300	32,337	[b]
TSV* RNA 3	2205	210	870	123	714[#]	288	289	31,742	[c]
Q-CMV RNA 4	1031	53	-	-	657	321	218	24,247	This work
D-CMV	N.D.	N.D.	-	-	657	301	218	24,100	[d]
BMV RNA 4	876	9	-	-	570	297	189	20,387	[e]
AMV RNA 4	881	36	-	-	666	179	221	24,380	[f]
TSV* RNA 4	N.D.	N.D.	-	-	714	288	237	26,346	[c]

[+] Includes stop codons [#] Not translated *in vitro* *TSV, tobacco streak virus N.D. = not determined

References: [a] Ahlquist *et al.*, 1981b; [b] Langereis *et al.*, 1986; [c] Cornelissen *et al.*, 1984; [d] Cuozzo *et al.*, 1988; [e] Dasgupta and Kaesberg, 1982; [f] Brederode *et al.*, 1980.

nucleotide numbers). Sequence and structural homology can be demonstrated between the 3a and the coat proteins CMV and BMV (see below) but there are no other long, AUG initiated open reading frames for which such homology can be demonstrated (data not shown). There are, however, two short open reading frames of 36 and 34 amino acids in the plus-sense RNA 3 of Q-CMV and BMV, respectively, which have similar locations (nucleotides 1824 to 1934 for Q-CMV, nucleotides 1792 to 1896 for BMV, stop codons included.) The putative peptides encoded by these open reading frames can be aligned to give 11 out of 33 matches with a single deletion in each sequence. However, the statistical significance of such a match is low and the biological significance (if any) of these short open reading frames is at present unknown.

3.3.3 Codon usage and nucleotide composition

The codon usage for the Q-CMV 3a and coat proteins is given in Table 3.2. The codon usage is non-random in both proteins and is similar to that generally found in plant viral proteins (for example, Q-CMV RNA 1 and 2 translation products (Rezaian *et al.*, 1984, 1985) and the translation products of BMV (Ahlquist *et al.*, 1981b, 1984)). The codon usage of the coat proteins of the Q and D (Cuozzo *et al.*, 1988) strains of CMV are very similar. There is a strong preference in the third position of four-fold degenerate codons for U over A in the Q-CMV 3a protein, however this preference is less evident in the coat protein cistron (Table 3.2b). The nucleotide composition of Q-CMV RNA 3 is also similar to that found in other plant viral RNAs. The 5'-untranslated regions of both RNAs 3 and 4 are high in U but low in C and A (Table 3.3). The 3'-untranslated region has an unbiased nucleotide composition very similar to that of the overall RNA. Presumably this is a reflection of the different, perhaps more relaxed, selective pressures that this region is submitted to compared with the pressures exerted on the coding and leader sequences.

Table 3.2A

Codon usage table for Q-CMV 3A protein(A)

(A)

First Position	Second Position								Third Position
	U		C		A		G		
U	Phe	2	Ser	9	Tyr	5	Cys	5	U
		6		5		4		1	C
	Leu	11		3	+*	0		0	A
		7		6	*	1	Trp	2	G
C	Leu	4	Pro	6	His	7	Arg	2	U
		2		6		2		3	C
		2		1	Gln	7		1	A
		2		3		6		2	G
A	Ile	4	Thr	5	Asn	9	Ser	4	U
		3		4		9		5	C
		3		1	Lys	5	Arg	4	A
	Met	5		3		7		2	G
G	Val	10	Ala	12	Asp	6	Gly	12	U
		3		7		5		1	C
		2		0	Glu	4		0	A
		4		6		7		4	G

+ Asterisks represent stop codons.

TABLE 3.2B

Codon usage table for Q-CMV Coat Protein. Figures in brackets represent codon usage for the D strain of CMV (Cuozzo *et al.*, 1985)

(B)

First Position	Second Position								Third Position
	U		C		A		G		
U	Phe	3(3)	Ser	8(7)	Tyr	5(6)	Cys	1(1)	U
		2(2)		6(6)		3(3)		0(0)	C
	Leu	1(3)		6(4)	*	0(0)	*	0(1)	A
		7		6	*	1	Trp	1(1)	G
C	Leu	5(5)	Pro	(4)	His	2(1)	Arg	7(9)	U
		3(3)		3(2)		1(2)		2(5)	C
		1(3)		2(5)	Gln	2(3)		1(3)	A
		4(2)		5(3)		4(3)		4(0)	G
A	Ile	7(5)	Thr	4(1)	Asn	4(2)	Ser	2(2)	U
		4(2)		5(4)		3(5)		1(0)	C
		0(2)		2(3)	Lys	9(7)	Arg	4(3)	A
	Met	6(4)		1(4)		5(5)		2(1)	G
G	Val	7(6)	Ala	6(5)	Asp	8(8)	Gly	8(5)	U
		5(6)		5(8)		7(8)		2(0)	C
		1(3)		1(2)	Glu	2(2)		2(3)	A
		3(3)		2(5)		5(3)		1(3)	G

+ Asterisks represent stop codons.

TABLE 3.3

Percent nucleotide composition of various regions of Q-CMV RNA 3, including percent nucleotide at third position of four-fold degenerate codons (i.e. those for Ala, Gly, Pro, Thr, Val).

Region	U	C	A	G
Entire RNA	29.9	23.7	22.4	24.0
5'-untranslated (RNA 3)	38.7	18.3	15.0	28.0
3A protein	28.5 (50)*	23.8 (23.4)	24.1 (4.4)	23.7 (22.2)
Intercistronic region	38.7	20.9	18.1	22.3
Coat protein	28.0 (30)	24.9 (26)	23.4 (19.3)	23.6 (24.7)
3'-untranslated	26.8	24.9	22.4	25.9
5'-untranslated (RNA4)	49.0	17.7	7.8	25.5

* Figures in brackets indicate nucleotide frequences at the third position of four-fold degenerate codons.

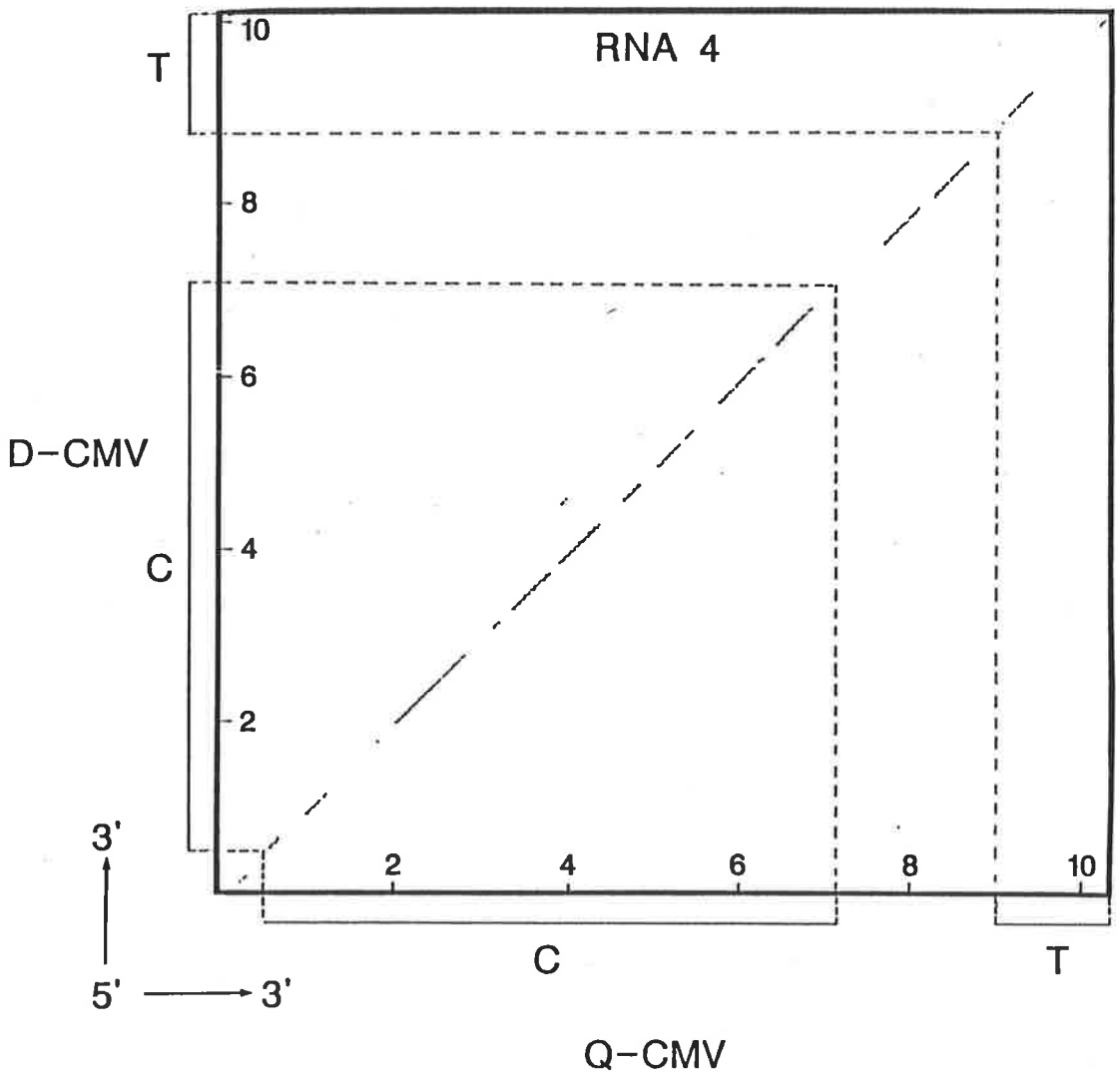
3.3.4 Nucleotide sequence homology between Q-CMV RNAs and other viral RNAs

The four Q-CMV RNAs can be divided into two groups on the basis of sequence homology present in their 5'-untranslated regions. No extensive homology exists between the 5'-untranslated regions of all four RNAs. There is, however, limited homology between the 5'-untranslated regions of RNAs 3 and 4 (RNA 3 ^{70} UGUGUCGUGUUU, RNA 4 ^{1182} UGAGUCGUGUUU) and a high degree of homology between the 5'-untranslated regions of RNAs 1 and 2 (Rezaian *et al.*, 1985). This grouping may reflect the different functional requirements (e.g. control of translation or transcription) of the respective 5'-non-coding regions rather than differing ancestry as all four RNAs share a high degree of 3'-terminal sequence homology (Symons, 1979). The 3'-terminal homologous sequence of 303 nucleotides does not overlap with the coat protein coding region in RNAs 3 and 4 as previously stated (Gould and Symons, 1982), due to the alteration to the predicted size of the coat protein.

It is noteworthy that a sequence related to that found in the 5'-untranslated regions of Q-CMV RNAs 3 and 4 (see above) is also present in the 5'-untranslated region of D-CMV RNA 4 (Cuozzo *et al.*, 1988). This sequence (UGUGCUGUGUUU) occurs at a similar distance from the start codon of the D-CMV coat protein gene as does the related sequence in Q-CMV RNA 4.

When the RNA 4 sequences of Q-CMV and D-CMV are compared by the DIAGON program (Fig. 3.4) it is evident that they are closely related. The closeness of the relationship is emphasized by the high degree of homology in the coat protein coding regions. The level of homology between the 3'-untranslated sequences which are required for the formation of the tRNA-like structures (implicated in RNA replication in BMV (Bujarski *et al.*, 1985; Miller *et al.*, 1986)) is not particularly high. However these sequences can still be folded to yield similar structures (data not

Fig. 3.4 DIAGON comparison of Q-CMV and D-CMV (Cuozzo *et al.*, 1988) RNA sequences. Q-CMV RNA4 (nucleotides, 1-1008). Numbers on axes are nucleotides $X10^2$ from the 5' end the RNA. Note that for D-CMV this sequence is not caomplete at the 5' end. A span length of 15 nucleotides and a threshold score equivalent to a double-matching probability of 10^{-4} were used. The coding regions and the sequences at the 3' end required for formation of the tRNA-like structures are indicated as C and T respectively.


RNA 4
T
C
D-CMV
10
8
6
4
2
2
4
6
8
10
C
T
3'
5'
3'
Q-CMV

shown). There is a high degree of sequence homology in that region of the 3'-untranslated region which extends 5' to the sequences required to form the tRNA-like structure (Fig. 3.4). As for BMV (see below) these sequences may be required for virion assembly and also RNA 3 accumulation *in vivo*.

The RNAs of AMV and TMV share only limited nucleotide sequence homology with those of Q-CMV (Rezaian *et al.*, 1984, 1985, this work). However, considerable homology exists between Q-CMV and BMV RNAs. The most striking is present within the 3'-untranslated region which contains sequences homologous between all Q-CMV and BMV RNAs (Ahlquist, *et al.*, 1981a). The similarity between the 3'-ends of these viral RNAs is even more apparent when secondary structures are considered as they can all be base-paired to form tRNA-like structures. Using the secondary structure model proposed for the tRNA-like structures by Rietveld *et al.* (1983) the tRNA-like structures end at 134 and 133 nucleotides from the 3' termini of BMV and Q-CMV, respectively. It has been shown for BMV that this 134 nucleotide segment contains all the requirements for the initiation of transcription on plus-sense templates *in vitro* (Miller *et al.*, 1986). Recent evidence, however, suggests that sequences more 5' to this segment (up to 200 nucleotides from the 3' terminus) are required for BMV RNA 3 accumulation *in vivo* (French and Ahlquist, 1987). As stated by French and Ahlquist (1987) this is in agreement with the high degree of homology between the 3'-terminal 193 nucleotides of the BMV RNAs. Q-CMV and BMV RNA 3 exhibit considerable sequence and structural homology throughout this region. Interestingly, homology between Q-CMV RNAs (and to a lesser extent BMV RNAs) continues beyond this region for a considerable distance. A possible function for these sequences may be in RNA packaging (French and Ahlquist, 1987).

In addition to the 3'-untranslated region common between all Q-CMV RNAs, the RNA 2 of Q-CMV appears to have a 3'-untranslated segment which has no

parallel in RNAs 1, 3 or 4. In RNAs 1 and 4 (and 3) there is only a short distance (9 and 17 nucleotides, respectively) between the stop codon of the protein and the first of the 3'-terminal homologous nucleotides. The RNA 2 of the recently sequenced Fny strain of CMV (Rizzo and Palukaitis, 1988) appears to have a similar arrangement of sequences. The RNAs 2 of Q-CMV and Fny-CMV exhibit a high degree of nucleotide sequence homology throughout their 3'-untranslated regions, part of which is homologous between all three Q-CMV RNAs and Fny-CMV RNA 2. As for Q-CMV RNA2 there is a region of approximately 100 nucleotides between the stop codon of the open reading frame and the first of the 3'-terminal homologous nucleotides. Significantly the RNAs 2 of Q and Fny-CMV are highly homologous in this extra region. Due to the lack of any additional data, no function can be proposed for this area of Q-CMV RNA 2. In RNA 2 this distance is some 116 nucleotides. In BMV the distance between the stop codon and the homologous region is very similar in all RNAs, there being no extra region in BMV RNA 2.

Q-CMV and BMV RNA 3 also contain an homologous sequence in their intercistronic regions. This sequence, which occurs to the 5' side of the origin of RNA 4 (Table 3.4) and begins at nucleotides 1101 and 1100 in Q-CMV and BMV, respectively, can be aligned to give 13 out of 14 matches. The revised sequence of Q-CMV RNA 3 presented in this work is more homologous to BMV RNA 3 than that reported by Gould and Symons (1982). A closely related sequence has been reported previously in the 5'-untranslated regions of Q-CMV and BMV RNAs 1 and 2 as has a related but less homologous sequence in the complement of the satellite RNA of Q-CMV (Rezaian *et al.*, 1985; French and Ahlquist, 1987).

TABLE 3.4

SEQUENCE HOMOLOGIES IN THE 5′-UNTRANSLATED AND INTERCISTRONIC REGIONS OF Q-CMV AND BMV

																		Reference[b]
5′-untranslated region	CMV RNA 1 (20)[a]	A	C	G	G	U	U	C	A	A	C	C	C	C	U	G	C	C	1
	BMV RNA 1 (15)	G	A	G	G	U	U	C	A	A	U	C	C	C	U	U	G	U	2
	CMV RNA 2 (19)	A	U	G	G	U	U	C	A	A	C	C	C	C	U	G	C	C	3
	BMV RNA 2 (15)	G	A	G	G	U	U	C	A	A	U	C	C	C	U	U	G	U	2
Intercistronic region	CMV RNA 3 (1099)	A	A	G	G	U	U	C	A	A	U	U	C	C	U	U	U	U	This work
	BMV RNA 3 (1098)	U	G	G	G	U	U	C	A	A	U	U	C	C	C	U	U	A	4
Consensus sequence		–	–	G_6	G_6	U_6	U_6	C_6	A_6	A_6	u_4 c_2	c_1 u_2	C_6	C_6	u_5 c_1	u_4 g_2	–	–	

[a] Numbers in parentheses indicate the distance of the first nucleotide from the 5′-end of the RNA.

[b] References: (1) Rezaian *et al.*, 1985; (2) Ahlquist *et al.*, 1984; (3) Rezaian *et al.*, 1984; (4) Ahlquist *et al.*, 1981b.

3.3.5 Amino acid sequence and structural comparison of Q-CMV 3a and coat proteins with those of other Tricornaviridae

Comparison of the coat protein amino acid sequences of CMV Q and D strains reveals a high degree of similarity at the identity level (Fig. 3.5). Apart from the arginine rich region near the N-termini, the ends of the sequences appear less similar than the internal regions. The overall homology is greater than that previously reported for Q and D strains (Cuozzo *et al.*, 1988) due to the revisions to the Q-CMV sequence reported in this work.

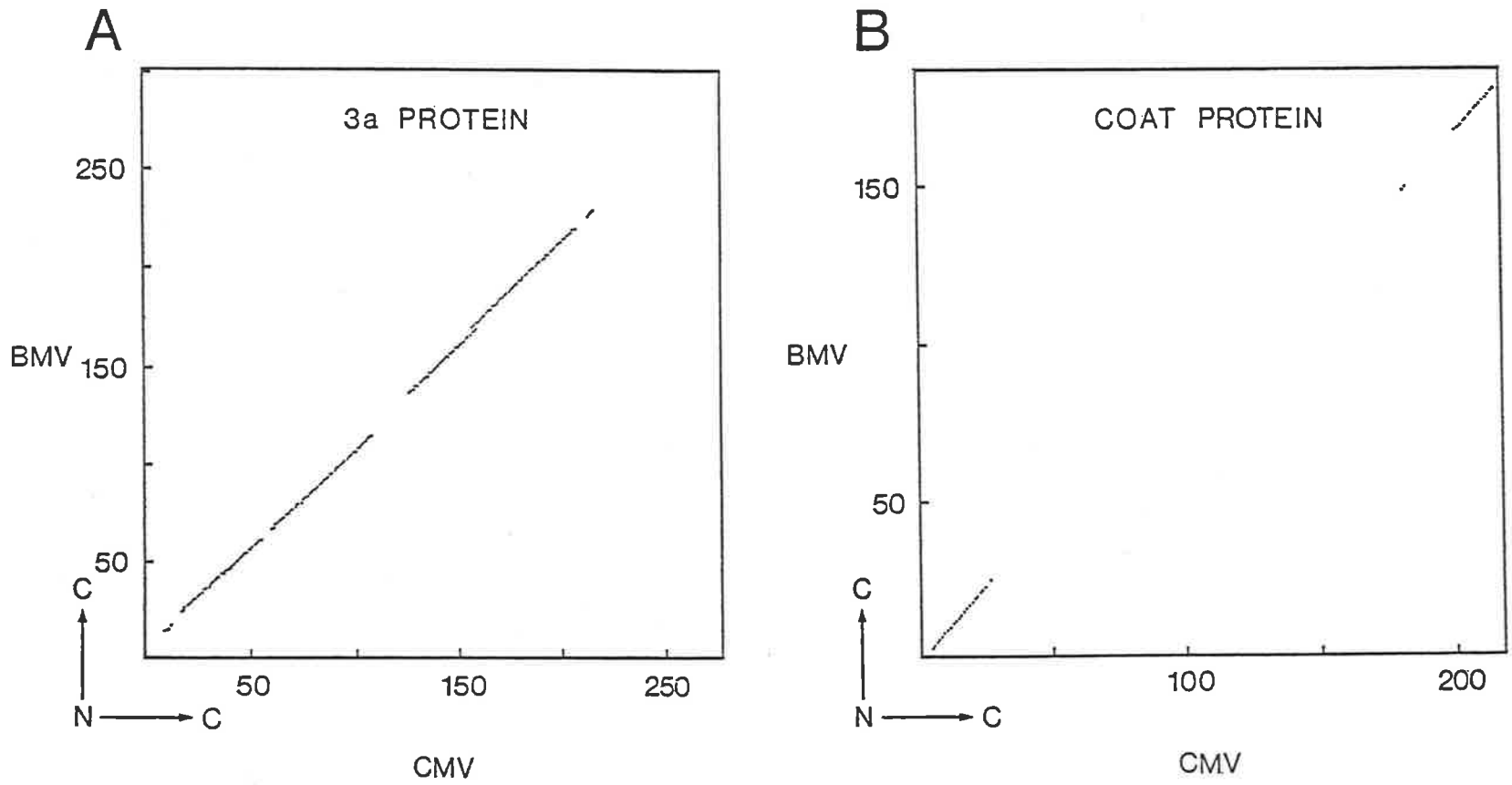
The DIAGON computer program (Staden, 1982) was used to compare the Q-CMV 3a protein with that of BMV and other viral 3a proteins. The proportional algorithm, which utilizes the weighted score matrix of Dayhoff (1969), was used with parameters similar to those used by Haseloff *et al.* (1984) who compared non-structural proteins of AMV, BMV, TMV and Sindbis virus. Fig. 3.6A shows such a comparison between the 3a proteins of Q-CMV and BMV using a span length of 31 and a threshold score corresponding to a score expected for a probability as calculated by the McLachlan (1971) 'double matching probability' of approximately 10^{-4}. No points occur off the diagonal and most of the proteins appear to be homologous at this level of expectation. The fact that the diagonal line consists of segments which are parallel but not co-linear (Fig. 3.6A) suggests (in this case) that either deletions have occurred in the Q-CMV 3a protein or insertions have occurred in the BMV 3a protein. If the score is reduced to correspond to a probability of 10^{-3} the diagonal line extends almost the entire distance virtually uninterrupted (results not shown). Using the same parameters (i.e. a span of 31, score equivalent to a probability of 10^{-4}) only limited homology was detected between the Q-CMV 3a protein and that of AMV, TSV and the 30K protein of TMV (results not shown). Amino acid sequence homology between the 3a proteins of Q-CMV and BMV has been reported by Murthy (1983), who aligned the two sequences to show that a considerable amount of homology

Fig. 3.5 Alignment of CMV-Q and D strain (Cuozzo *et al.*, 1988) coat protein amino acid sequences using the SEQA program. Upper sequence is that of Q-CMV, lower D-CMV. Identical residues are indicated by dots. Penalty for deletion 2.

ALIGNMENT OF Q- AND D-CMV COAT PROTEINS

```
        10         20         30         40         50         60         70
MDKSGSPNASRTSRRRRPRRGSRSA-SGADAGLRALTQQMLRLNKTLAIGRPTLNHPTFVGSESCKPGYTFTSITLKPPE
:::: :  : :  :::::::::::: : :::  : : ::  ::::::: :::: ::::::::: : ::::::::::::
MDKSESTSAGR-NRRRRPRRGSRSAPSSADANFRVLSQQLSRLNKTLAAGRPTINHPTFVGSERCRPGYTFTSITLKPPK
        10         20         30         40         50         60         70
         90        100        110        120        130        140        150
IEKGSYFGRRLSLPDSVTDYDKKLVSRIQIRINPLPKFDSTVWVTVRKVPSSSDLSVAAISAMFGDGNSPVLVYQYAASG
:  ::: : :: :::::::::::::::::::: :::::::::::::::::: :::::::::::: :: ::::::::::::
IDRGSYYGKRLLLPDSVTDYDKKLVSRIQIRVNPLPKFDSTVWVTVRKVPASSDLSVAAISAMFADGASPVLVYQYAASG
         90        100        110        120        130        140        150
        170        180        190        200        210
VQANNKLLYDLSEMRADIGDMRKYAVLVYSKDDKLEKDEIVLHVDVEHQRIPISRMLPT
:::::::::::: :::::::::::::::::::::: :: :: ::::: :::::: :  ::
VQANNKLLYDLSAMRADIGDMRKYAVLVYSKDDALETDELVLHVDIEHQRIPTSGALPV
        170        180        190        200        210
```

Fig. 3.6 DIAGON comparison of Q-CMV and BMV translation products. (A) Comparison of Q-CMV 3a protein with BMV 3a protein (Ahlquist *et al.*, 1981b). (B) Comparison of Q-CMV coat protein with BMV coat protein (Dasgupta and Kaesburg, 1982). The proportional algorithm was used with a span of 31 and a threshold score equivalent to a "double-matching probability" of 10^{-4}. The N-terminal to C-terminal direction is indicated and numbers on the axes refer to the number of amino acids.


A
3a PROTEIN
250
150
50
BMV
C
N
C
50
150
250
CMV
B
COAT PROTEIN
150
50
BMV
C
N
C
100
200
CMV

exists, at the identity level, between these proteins, throughout their length. The changes in the sequence of the 3a protein reported in this paper (i.e. six amino acid substitutions and the foreshortening of the protein by 54 amino acids) do not greatly alter this pattern of homology (Fig. 3.7).

In addition to the sequence homology, similarities in the charge distributions, hydrophobicity profiles and α-helix, β-sheet and β-turn propensities (as computed by the program of Novotny and Auffray (1984)) occur throughout the 3a proteins of Q-CMV and BMV (data not shown). The 3a protein of AMV and the 30K protein of TMV were analyzed in the same manner. Figure 3.8A shows regions of the 3a proteins of AMV and Q-CMV (also present in the BMV 3a protein, results not shown) which have very similar secondary structure prediction profiles. These data suggest that the Q-CMV and BMV 3a proteins share homology at the structural level as well as at the sequence level and that, although the AMV 3a protein has only limited sequence homology with Q-CMV and BMV (Savithri and Murthy, 1983; data not shown), it appears to contain structural elements which are common between the three 3a proteins examined. The 30K protein of TMV although having little sequence homology with the Tricornaviridae 3a proteins (results not shown) also appears to contain at least some of these common structural elements (Fig. 3.8C).

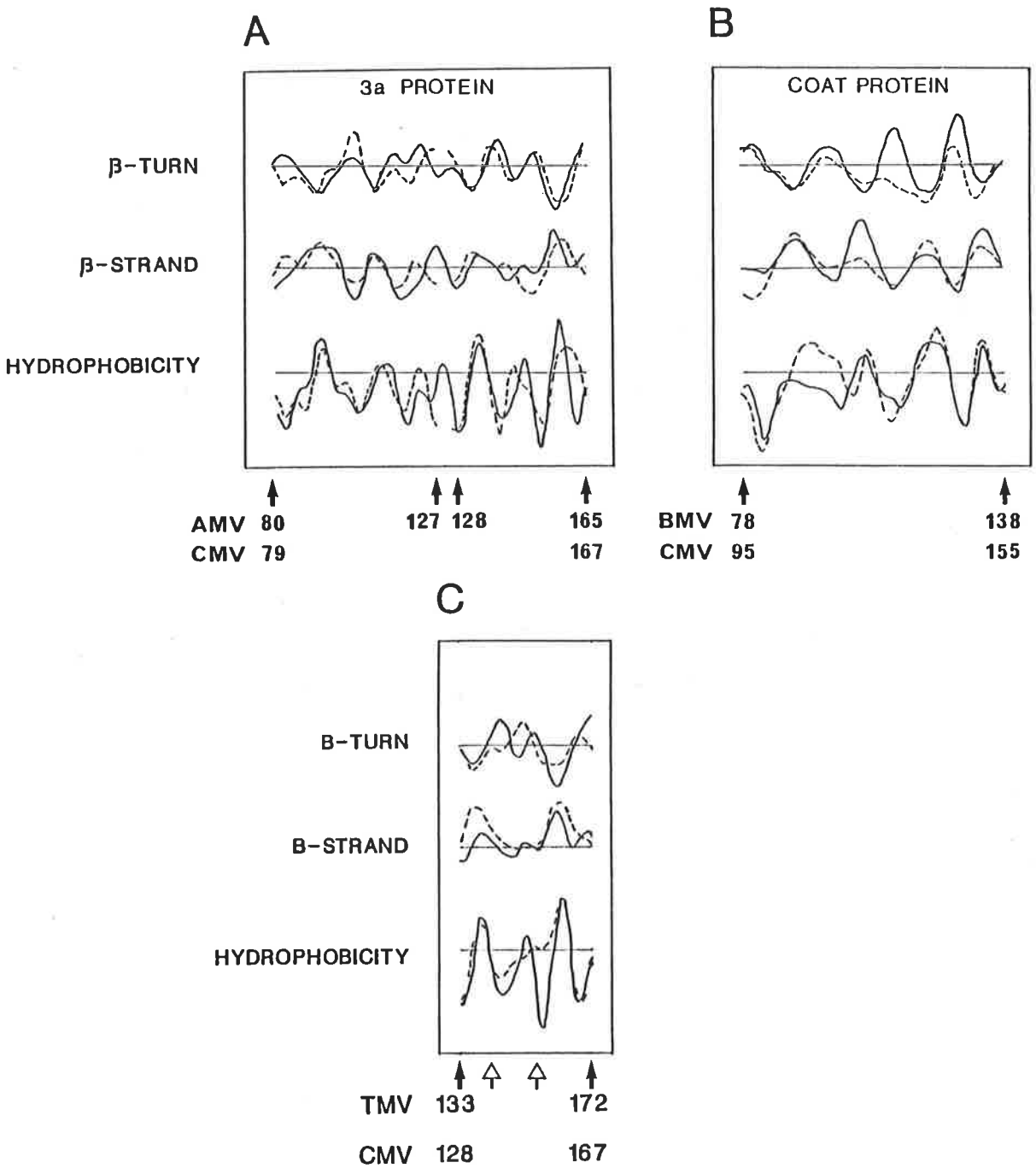
When the coat proteins of Q-CMV and BMV are compared by the DIAGON program using the same parameters as those used in Fig. 3.6A for the 3a proteins (i.e. span 31, probability equivalent to 10^{-4}), homology is apparent in both the N and C-terminal regions (Fig. 3.6B). The homology in the N-terminal region is due to the cluster of basic amino acids which are predicted to form an α-helix by helical 'wheel' analysis (Schiffer and Edmundson, 1967, data not shown). This region is characteristic of a number of coat proteins and is implicated in protein/RNA interactions in the virion (Harrison, 1984). The C-terminal regions can be aligned to give 6 matches out of a 10 residue contiguous sequence beginning at residue 200 for

Fig. 3.7 Alignment of Q-CMV and BMV (Ahlquist *et al.*, 1981b) 3a proteins amino acid sequences using the SEQA program. Upper sequence is that of Q-CMV, lower BMV. Identical residues are indicated by dots. Penalty for deletion 2.

ALIGNMENT OF CMV AND BMV 3a PROTEINS

```
                10         20         30         40         50         60         70
M-----AFQGPSR-TLTQQSSA-ASSD-DLQKILFSPDAIKKMATECDLGRHHWMRADNAISVRPLVPQVTSNNLLPFFK
:      : : :: :       :   ::  :   :::  : :  : :: ::                 :::       :
MSNIVSPFSGSSRTTSDVGKQAGGTSDEKLIESLFSEKAVKEIAAECKLGCYNYLKSNEPRNYIDLVPKSHVSAWLSWAT
        10         20         30         40         50         60         70         80
      80         90        100        110            120        130        140
SGYDAGELRSKGYMSVPQVLCAVTRTVSTDAEGSLKIYLADLG----DKELSPIDGQCVTLHNHELPALISFQPTYDCPM
: :: ::: : : : ::   :   ::       ::    : : :       :  :: :  :     :::::   : :::::
SKYDKGELPSRGFMNVPRIVCFLVRTTDSAESGSITVSLCDSGKAARAGVLEAIDNQEATIQLSALPALIALTPSYDCPM
        90        100        110        120        130        140        150        160
             160        170        180        190        200        210        220
E-LVG--NRHRCFAVVVERHGYIGYGGTTASVCSNWQAQFSSKNNNYTHAAAGKTLVLPYNRLAEHSKPSAVARLLKSQL
:   :   : :::       :  :  :  :    ::: :  : :::          : :  ::    : :
EVIGGDSGRNRCFGIATQLSGVVGTTGSVAVTHAYWQANFKAKPNNYKLHGPATIMVMPFDRLRQLDKKSLKNYIRGISN
       170        180        190        200        210        220        230        240
  230        240        250        260        270
NNVSSSRYL-LPNVALNQNASGHESEILKESPPIAIGSPSASRNNSFRSQVVNGL--------
  :     :  :      : :          :::          :   :  :
QSVDHGYLLGRPLQSVDQVAQEDLLVEESESPSALGRGVKDSKSVSASSVAGLPVSSPTLRIK
       250        260        270        280        290        300
```

Fig. 3.8 Secondary structure prediction dones by the program of Novotny and Auffray (1984). Curves from top to bottom are β-turn propensity, β-strand propensity and hydrophobicity profile (the parameters of Kyte and Doolittle (1982) are used). The numbers below the profiles refer to distance in amino acids from the N-termini of the respective proteins. (A) Comparison of regions of the 3a protein of AMV (Langereis *et al.*, 1986) (broken line) with the 3a protein of Q-CMV) (solid line). Note that a gap has been inserted in the AMV sequence to aid the alignment of the two protein sequences. (B) Comparison of a region of the coat proteins of BMV (broken line) and Q-CMV (solid line). (C) Comparison of a region of the 30K protein of TMV (Takamatsu *et al.*, 1983) (broken line) and the 3a protein of Q-CMV (solid line). The profile is aligned below that of AMV vs CMV to allow comparison. The asterisked numbers refer to the location of two amino acids thought to be required for activity of the TMV 30K protein (Ohno *et al.*, 1983; Zimmern and Hunter, 1983).


A
3a PROTEIN
β-TURN
β-STRAND
HYDROPHOBICITY
AMV 80 127 128 165
CMV 79 167
B
COAT PROTEIN
BMV 78 138
CMV 95 155
C
B-TURN
B-STRAND
HYDROPHOBICITY
TMV 133 172
CMV 128 167

Q-CMV and 168 for BMV (BMV, VVHLEVEHVR; CMV, VLHVDVEHQR). The functional reason, if any, for the similarity in the C-terminal region is at present unknown.

Using the secondary structure prediction program of Novotny and Auffray (1984), potential structural homology was found between the coat proteins of Q-CMV and BMV. A region of approximately 60 amino acids, located in the central region of each protein, was shown to have similar hydrophobicity, β-sheet and β-turn propensity profiles (Fig. 3.8B).

No homology, at either the sequence or structural levels, was found between the Q-CMV and AMV or TMV coat proteins, apart from the predominance in the AMV coat protein of basic residues in the N-terminal region. This lack of homology may be a result of both AMV and TMV having capsids whose morphologies are different from those of Q-CMV and BMV. In addition the AMV coat protein, has a function in the recognition of template RNA by the RNA replicase (Houwing and Jaspars, 1978).

3.4 DISCUSSION

In revising the sequence of Q-CMV RNA 3 it was important that the sequence obtained was that of the Q-CMV strain originally sequenced by Gould and Symons (1982). The following steps were taken to ensure that this was so: (1) much of the sequence (nucleotides 729-1784) was obtained by sequencing the same cDNA used by these workers (i.e. clone p2Tip2): (2) the RNA used for generating the M13 clones prepared in this study and for sequencing using synthetic primers was prepared from plants inoculated with the original preserved inoculum of Gould and Symons (1982).

3.4.1 Sequences and mechanisms likely to be involved in CMV RNA 3 and 4 replication

The function of the short sequence conserved in the 5'-untranslated and intercistronic regions of the genomal RNAs of Q-CMV and BMV (Table 4) is not known, but it may be a cis-acting sequence involved in the initiation of plus-strand RNA synthesis and is probably required for the efficient accumulation of BMV RNAs *in vivo* (French and Ahlquist, 1987). Its different relative location in RNAs 1 and 2 (in the 5' leader region) and RNA 3 (in the intercistronic region 5' to the start of RNA 4) does not preclude this possibility (French and Ahlquist, 1987). The absence of this sequence from RNA 4 is in accordance with such a proposal as for BMV at least, RNA 4 does not appear to replicate autonomously but is produced by internal initiation on minus-strand RNA 3 (Miller *et al.*, 1985). Evidence concerning the autonomous replication of CMV RNA 4 is inconclusive (Hull and Maule, 1985). In addition, as this sequence is not present in RNA 4 and yet RNA 4 is readily encapsidated and stable, it is unlikely that this sequence acts as a recognition signal for packaging or is involved in RNA stability.

The presence of a related but less homologous sequence in the complement of Q-sCMV RNA may be important in helper virus: satellite RNA interaction as suggested by Rezaian *et al.*, (1985). The conserved sequences contained in the plus-sense Q-CMV RNAs 1, 2 and 3 and the related sequence present in the minus-sense satellite RNA could compete with each other for a particular function. The alternative, suggested by Rezaian *et al.*, (1985), is that the complementary sequence in the plus-sense satellite RNA binds directly to the sequence contained in the plus-sense genomal RNAs. Recent *in vitro* investigations (Rezaian and Symons, 1986) did not find evidence supporting the latter alternative but hybridization of two regions of the satellite RNA to a sequence in the coat protein cistron was detected. Further

extensive studies will be required to define the interaction between the host plant, the satellite RNA and the helper virus.

For BMV RNAs 3 and 4 considerable data concerning the sequences required for RNA replication has been acquired (Ahlquist *et al.*, 1984; Dreher *et al.*, 1984; Bujaski *et al.*, 1985, 1986; Miller *et al.*, 1986; French and Ahlquist, 1987, 1988; Marsh *et al.*, 1987; Dreher and Hall, 1988). Despite the lack of direct data for CMV, but because of the many similarities between CMV and BMV, it is interesting to summarize the possible implications of the BMV data for CMV. There is little sequence homology between the 5'-untranslated regions of Q-CMV and BMV RNAs 3 but for BMV this region has been shown to be required for efficient RNA 3 accumulation *in vivo* (French and Ahlquist, 1987) and this region may reasonably be expected to perform a similar function in Q-CMV RNA 3. For BMV the tRNA-like structure at the 3' end of RNA 3 has been shown to be sufficient for minus-strand initiation *in vitro* (Ahlquist *et al.*, 1984; Miller *et al.*, 1986; Dreher and Hall, 1988) and sequences beyond this are required for RNA 3 accumulation *in vivo* (French and Ahlquist, 1987). Based on nucleotide and secondary structure homologies (see Results) a similar situation is highly probable for CMV.

French and Ahlquist (1988) have shown that nearly the entire intercistronic region of BMV has been implicated in some way as required for either genomic (RNA 3) or sub-genomic (RNA 4) RNA replication. The intercistronic regions of CMV and BMV RNAs 3 share an homologous sequence, also found at the 5' ends of RNAs 1 and 2 (see Table 3.4) which in BMV, as discussed above, has been shown to be in a region required for efficient *in vivo* accumulation of RNA 3 (French and Ahlquist, 1987). The poly (A) tract present in the BMV intercistronic region (but absent in Q-CMV) has been shown to be required for the efficient accumulation of BMV RNA 4 (but not RNA 3) *in vivo* (French and Ahlquist, 1988). If as suggested (Marsh and Hall, 1987) this sequence is required to allow replicase access to promoter

sequences, then the U rich regions in Q-CMV may perform a similar function. The so-called "core promoter" sequences (and their repeats) proposed as important for BMV RNA 4 production (Marsh *et al.*, 1987; Marsh and Hall, 1987; French and Ahlquist, 1988) are not well represented in Q-CMV RNA 4. This, however, does not preclude the presence of sequences that may perform similar functions in CMV. It has been postulated (Marsh *et al.*, 1987; French and Ahlquist, 1988) that sequences around and upstream from the BMV RNA 4 site of initiation are homologous with alphavirus promoters. However, for Q-CMV this homology seems poor. Homology between 5' and intercistronic sequences have also been proposed as related to the internal control regions found in various eukaryotic tRNAs (Marsh and Hall, 1987). Again although these suggestions are interesting the rather poor homology found in Q-CMV RNA 3 sequences prevents any firm conclusions.

In the discussions concerning sequences which may be important for RNA 4 production in CMV it is important to remember that there is no direct evidence for CMV RNA 4 being produced by internal initiation. That it does is an assumption based on analogy with what occurs in BMV (Miller *et al.*, 1985).

3.4.2 Probable role of the Q-CMV 3a protein is in cell to cell movement

Apart from the coat protein, there is little direct evidence concerning the functions of the Q-CMV encoded proteins. On the basis of sequence and structural homology, the 3a proteins of Q-CMV, BMV and AMV and the 30K protein of TMV are likely to share similar functions *in vivo*. The function of the 3a proteins of Q-CMV and BMV has not been determined but a role in the cell to cell movement of the infectious entity during infection (see Atabekov and Dorokhov, 1984, for review) is indicated for both the TMV 30K protein (Ohno *et al.*, 1983; Zimmern and Hunter, 1983; Deom *et al.*, 1987; Meshi, *et al.*, 1987) and the 3a protein of AMV (Huisman *et al.*, 1986; Stussi-Garaud *et al.*, 1987).

Further evidence points to the possible functional equivalence of the TMV 30K protein and the BMV 3a protein. For example, TMV is able to complement the spread of BMV in plants that are normally non-hosts for BMV (Taliansky *et al.*, 1982). It is interesting to note that the region exhibiting structural homology in AMV, BMV, CMV and TMV (Fig 3.8A,C) contains the amino acids (whose location is indicated on Fig. 3.8B) that when substituted affect function of the TMV (30K) movement protein *in vivo* (Ohno *et al.*, 1983; Zimmern and Hunter, 1983).

At present there is little evidence regarding the actual mechanism of action of these movement proteins but for AMV and TMV they appear to be closely associated with the cell walls (Godefroy-Colburn *et al.*, 1986; Stussi-Garaud *et al.*, 1987; Tomenius *et al.*, 1987; Moser *et al.*, 1988). These observations are in agreement with the suggestion that these movement proteins act by modifying the cell wall (or associated components) to allow movement of the infectious entity between cells (Atabekov and Dorokhov, 1984).

3.4.3 The 3a and coat proteins may be involved in host range specificity and symptom induction

The foregoing discussion indicates the possible influence of movement proteins on host range and symptom induction. The role of coat proteins in these two processes is less clear and in the past they have been thought to play a relatively minor role (Van Loon, 1987). However it has recently been unequivocally demonstrated that the coat protein of TMV is involved in symptom expression and the modulation of the host response (Saito *et al.*, 1987; Dawson *et al.*, 1988). The investigation of the factors involved in symptom expression will be a difficult task due to the large number of possible interactions between the various plant and viral proteins and nucleic acids. It should not be forgotten in these considerations that the

viral nucleic acid sequence itself may play some part. The obvious precedent for this is the influence of sCMV RNAs on symptom expression (for review of satellite RNAs, see Francki, 1985).

3.4.4 Divergence of the 3a (30K) and coat proteins of the related viruses: AMV, BMV, CMV, TMV

When considering the relationships between the members of the Tricornaviridae family of viruses (and TMV) it is apparent that the 3a (30K in the case of TMV) and coat proteins are considerably more divergent than the RNA 1 and RNA 2 translation products (and the corresponding 126K and 183K translation products of TMV). The sequence conservation between the C-terminal ends of the Q-CMV and BMV coat proteins (Fig. 3.6B), although of unknown function, emphasizes the closeness of the relationship between these two viruses. The lack of sequence homology in the remainder of these two coat proteins (excepting the N-terminal region) is not surprising as other coat proteins have been shown to be structurally very similar but to contain little sequence homology (Rossman *et al.*, 1983; Harrison, 1984). The divergence of the amino acid sequence of the apparently functionally related 3a (and 30K) proteins would suggest that differences have arisen due to either host selection, the emergence of different secondary roles, or the tolerance of these proteins to mutation due to the relatively relaxed nature of their structural/functional requirements.

CHAPTER 4

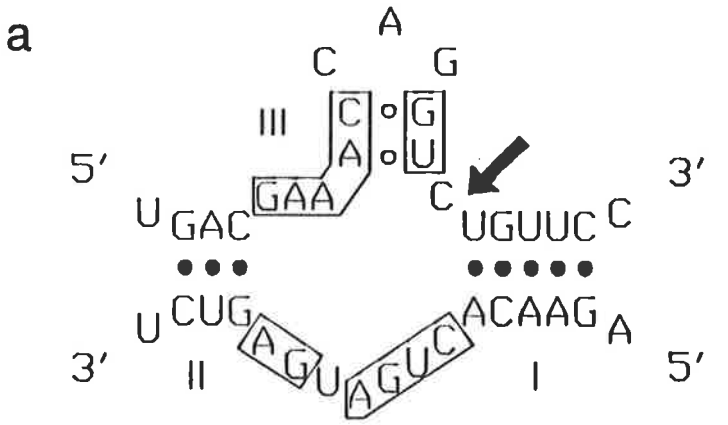
THE ACTIVE STRUCTURE INVOLVED IN PLUS ASBV RNA AND SELF-CLEAVAGE IS A DOUBLE-HAMMERHEAD

4.1 INTRODUCTION

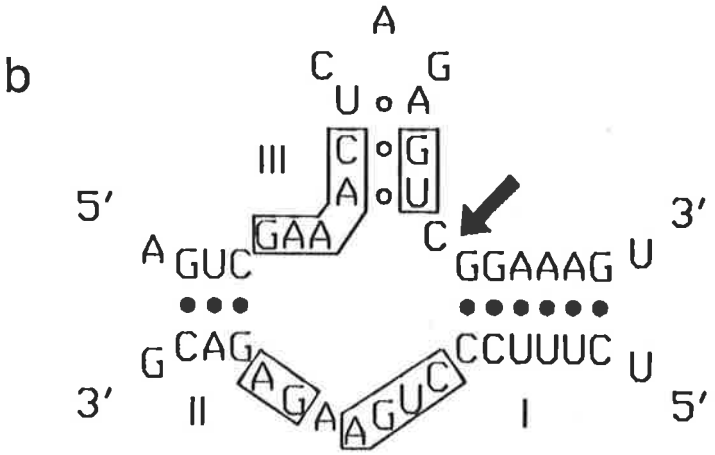
As stated in Chapter 2 certain plant viral RNAs, both satellites and viroids, and transcripts of a newt satellite DNA have been shown to undergo site-specific self-cleavage *in vitro*. Sequence comparisons between some of these RNAs (Buzayan *et al.*, 1986c, Hutchins *et al.*, 1986) allowed the prediction of hammerhead-shaped secondary structures thought to be involved in the cleavage of ASBV RNAs (Hutchins *et al.*, 1986). These structures consist of three base-paired regions joined by single-stranded regions (Fig. 4.1a,b) most of which are conserved between all the self-cleaving RNAs presented. For the plus and minus ASBV RNAs they can be formed by only minor alterations to the proposed rod-like native secondary structure (Steger *et al.*, 1984) (Fig. 4.2).

Similar structures have been proposed for the plant virus satellite RNAs (Forster and Symons, 1987a) (Fig. 4.3) and the newt self-cleaving RNA (Epstein and Gall, 1987) (Fig. 4.1c). For the newt transcript, plus and minus ASBV, sLTSV RNAs and plus sTRSV RNAs it has been shown that the hammerhead sequences are necessary and sufficient for cleavage (Forster and Symons, 1987b; Forster *et al.*, 1987; Uhlenbeck, 1987; Haseloff and Gerlach, 1988; Koizumi *et al.*, 1988). Comparison of the hammerhead structures proposed for ASBV and newt RNAs (Fig. 4.1), with the consensus sequence for the satellite RNAs (Rig. 4.3) revealed an obvious difference in their structures. The stem IIIs present in the satellite RNAs appear to be potentially much more stable than those proposed for either the ASBV or newt RNAs. The stability of the stem IIIs of ASBV has been questioned by other authors (Hutchins *et al.*, 1986; Forster and Symons, 1987b) and the structure proposed for the newt RNA (Epstein and Gall, 1987) appears to be extremely unstable. The degree of base-pairing in these stems is obviously limited (two or three base-pairs) and the loops are small (two or three nucleotides). Thus these self-cleaving RNAs appear to fall into two classes; those which have theoretically stable stem IIIs (the

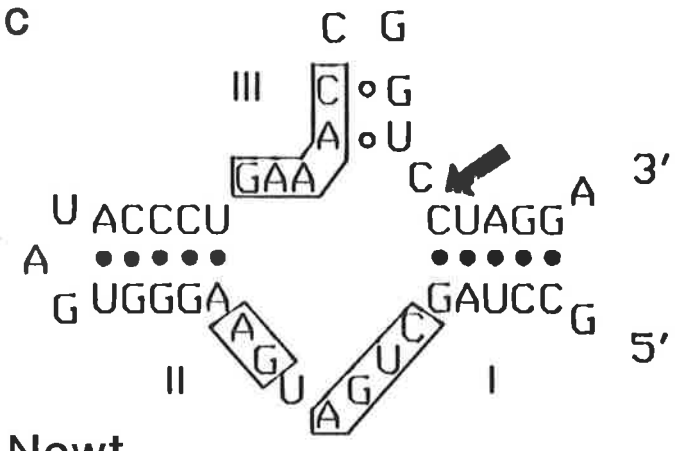
Fig. 4.1 Single-hammerhead models for the active self-cleavage structures of plus and minus ASBV RNAs (a and b) (Hutchins *et al.*, 1986) and the self-cleaving newt RNA (c) (Epstein and Gall, 1987). Boxed residues are those conserved in these three RNAs and in the self-cleaving satellite RNAs (listed in Fig. 4.3). The site of self-cleavage in each RNA is indicated by an arrow, the stems are numbered 1-111. Solid dots indicate base-pairing, hollow circles indicate previously proposed base-pairs that may not form.



ASBV +



ASBV −



Newt

Fig. 4.2 The proposed secondary structure of monomeric plus ASBV RNA (Steger *et al.*, 1984) with the location of the plus and minus self-cleavage domains indicated by boxes and lines respectively. The thick arrows indicate the sites of self-cleavage and the Sau3A I site used in cloning is indicated.

PLUS ASBV

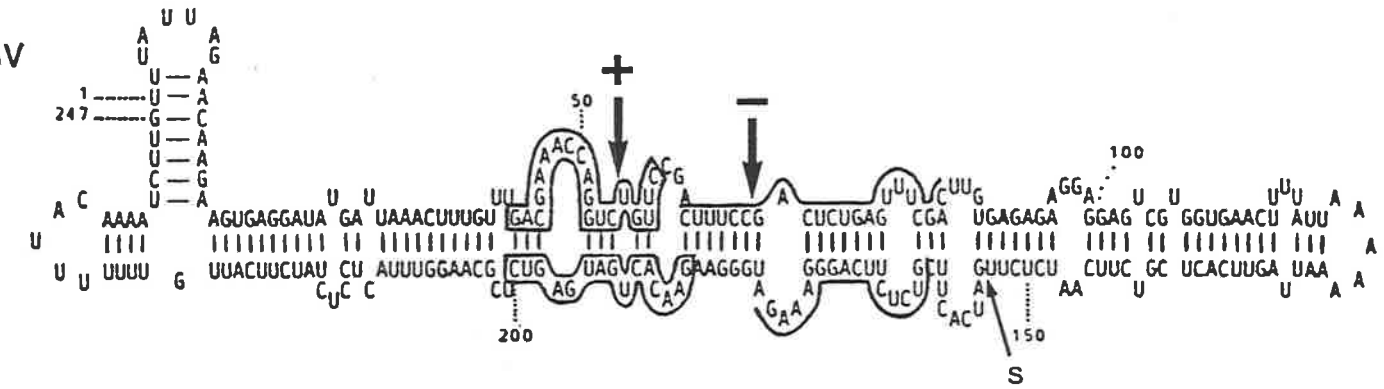
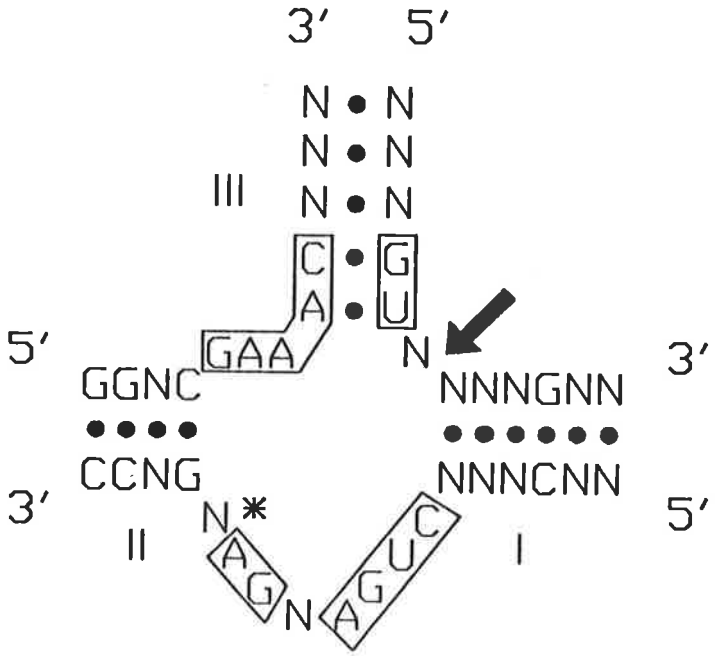


Fig. 4.3 The consensus single-hammerhead structure derived from seven plant satellite RNAs (Forster and Symons, 1987a) i.e. plus and minus sLTSV-A, sLTSV-N (New Zealand strain) and the plus RNAs of sTRSV, sSNMV and sVTMoV. Nucleotides conserved between the satellite RNAs are boxed, the arrow indicates the site of self-cleavage and the asterisked nucleotide is unique to the sequence of the plus sLTSV RNA structure where it is a U. The stems are numbered 1-111.

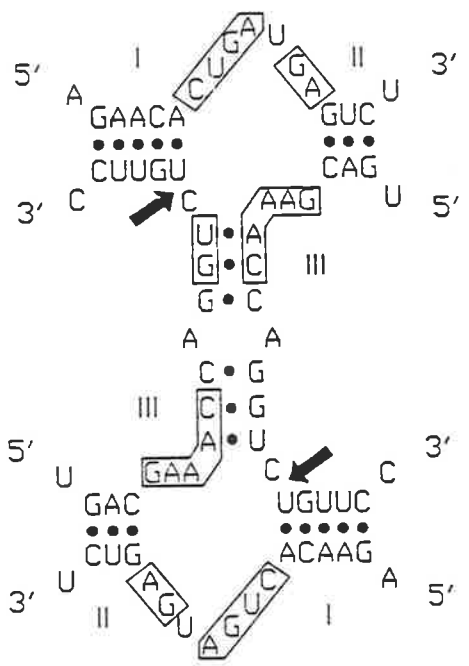

3′
5′
N • N
N • N
III
N • N
C
A
G
U
5′
GAA
N
3′
GGNC
NNNGNN
CCNG
NNNCNN
3′
5′
N*
II
I
A
G
N
A
G
U
C

satellite RNAs) and those that have potentially unstable stem IIIs (ASBV and newt RNAs). It should be noted that the stability of the satellite RNA stem IIIs may be enhanced by additional base-pairing in the remainder of the RNA molecules, a process that cannot occur in the ASBV and newt RNAs.

Even though these self-cleaving RNAs can be placed in two different groups on the basis of stem III stability, the considerable amount of sequence conservation between all the RNAs argues that they may well cleave by similar secondary and tertiary structures. It is possible to form secondary structures for these RNAs which contain stable stem IIIs by folding pairs of these sequences into double-hammerhead structures. These structures contain two hammerheads which have theoretically more stable stem IIIs (Fig. 4.4) and which therefore more closely fit the secondary structure model developed for the satellite RNAs (Fig. 4.3). Deletion analysis of the plus-sense sLTSV RNA has shown that only the two lowermost base pairs of stem III (see Fig. 4.3) are required for self-cleavage (Forster and Symons, 1987b). If this criterion is extended to the double-hammerhead structure (Fig. 4.4) it would seem that either the top or bottom halves of these structures alone are sufficient for cleavage at a single site. Thus, a consensus hammerhead structure (Fig. 4.5) can be drawn for all the RNAs in Figs. 4.1,4.3. Plus and minus sLTSV RNAs (containing only one set of the sequences required for self-cleavage) which, when drawn as single-hammerheads, contain stable stem IIIs have been shown to self-cleave efficiently during transcription (Forster and Symons, 1987a). The requirement for a stable stem III for self-cleavage may account for the efficient cleavage of dimeric ASBV RNAs (Hutchins *et al.*, 1986) (which contain two sets of sequences required for self-cleavage and could therefore form the double-hammerhead structure with stable stem IIIs intramolecularly) and for the lack of cleavage of monomeric ASBV RNAs (P. Rathjen, S. McNamara, A. Jeffries; unpublished data) (which contain only one set of sequences required for self-cleavage and therefore can only form stable stem IIIs intermolecularly). This statement implies that self-cleavage of monomeric ASBV and

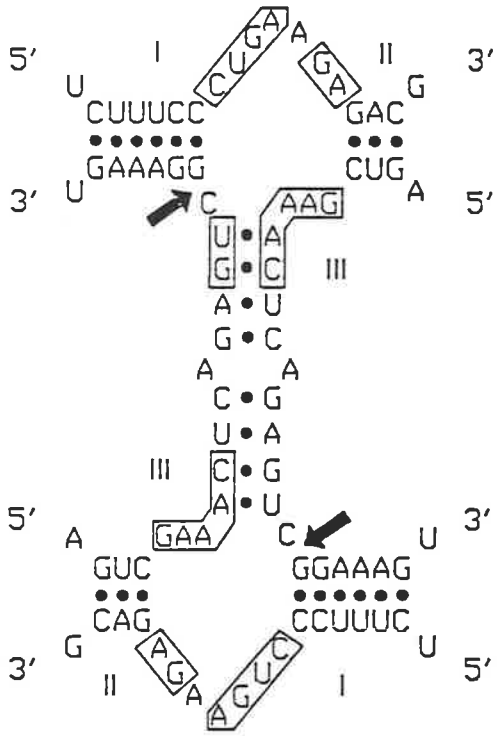
Fig. 4.4 Double-hammerhead secondary structures for the active self-cleavage structures of plus and minus ASBV RNAs and the newt RNA. The structures are formed by the alignment of two identical single-hammerheads such that the sequences of the apparently weak stem IIIs cooperate to form a stable structure.

a



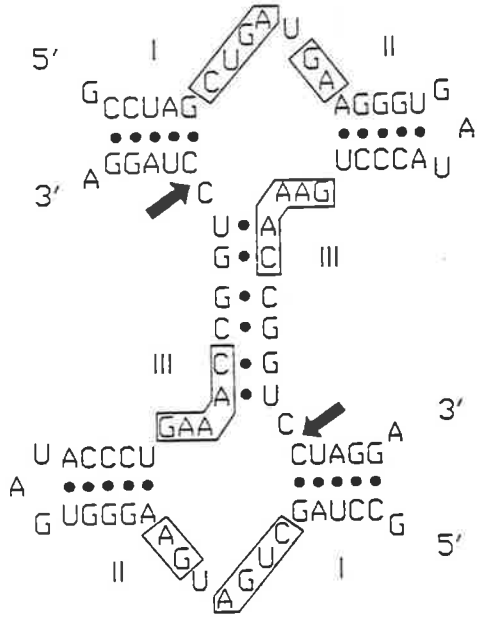
ASBV +

b



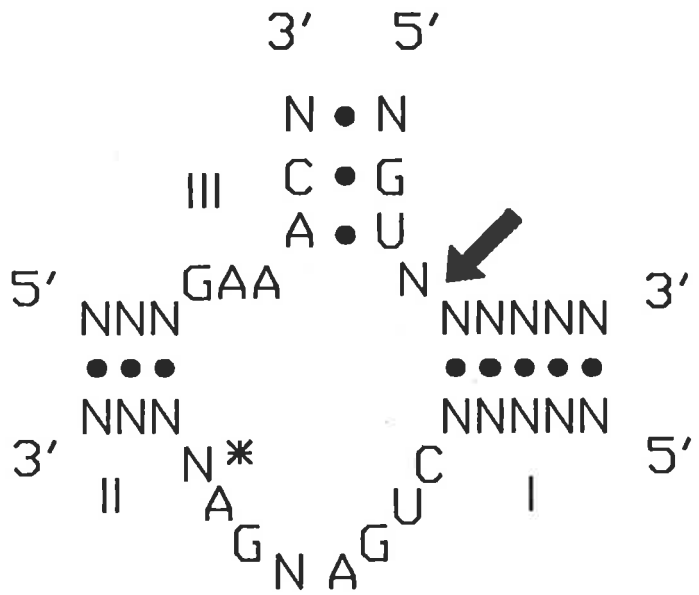
ASBV −

c



Newt

Fig. 4.5 Consensus single-hammerhead structures for ASBV, newt and the self-cleaving plant satellite RNAs listed in Fig. 4.3. Non-conserved nucleotides are denoted as N, the stems are numbered 1 to 111 and the site of self-cleaving is indicated by an arrow. The asterisked nucleotide is unique to plus sLTSV RNA where it is a U.


3′ 5′
N • N
III C • G
A • U
5′ GAA N 3′
NNN NNNNN
NNN NNNNN
3′ N* C 5′
II A U I
G
N A G

newt RNAs should be a concentration dependent event due to the requirement for intermolecular association to form the double-hammerhead structure capable of cleavage. This reaction should conform to the second order rate equation. Therefore there are at least two possible ways in which to test the double-hammerhead model. One way, a quantitative approach, would be to determine whether monomeric RNAs cleave in a concentration dependent manner (Forster *et al.*, 1988; and Discussion). The other approach, which would give a qualitative answer, would be to test whether two or more self-cleavage domains on a multimeric RNA transcript have to interact in order to achieve cleavage. The latter approach was investigated in this work.

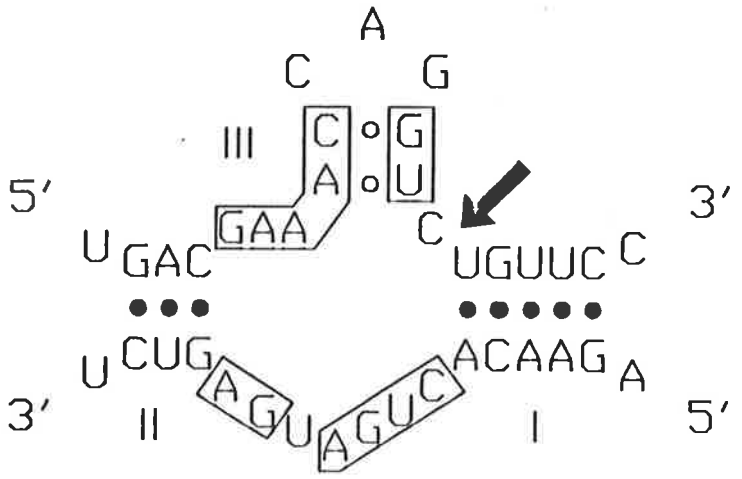
4.2 METHODS

4.2.1 Preparation of monomeric ASBV clones mutated in the sequences required for plus RNA self-cleavage

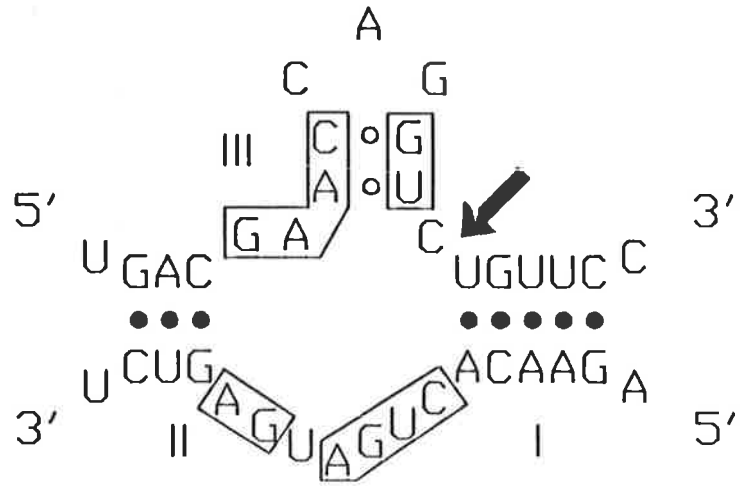
The M13mp93 clone of Barker *et al.* (1985) containing a full-length monomeric ASBV insert in the BamH I site was used as the starting material for mutagenesis. Oligonucleotide-directed mutagenesis using synthetic oligonucleotides was performed as described in the Methods chapter. Two different oligonucleotides were used for mutagenesis (and as probes for hybridization screening) i.e. "AAAAAC", 5'-GAACAGACCTGGTTTTTGTCAAACAAAGTTTA; "GAAC", 5'-CGGAACAGACCTGGTTCGTCAAACAAAGTT, the changes these oligonucleotides induced in the plus ASBV sequence are shown in Fig. 4.6. Positive clones detected by hybridization screening were sequenced by the dideoxy method to confirm that the desired mutations were present (Fig. 4.7).

Fig. 4.6 Single-hammerhead structures for plus ASBV RNAs having wild-type (a) and mutant (b,c) sequences. The "GAAC" mutant (b) is a deletion of an A residue from the conserved GAAAC sequence. The "AAAAAC" mutant (c) is a deletion of a G and the insertion of an A nucleotide in the GAAAC sequence. Nucleotides conserved between RNAs that self-cleave by the hammerhead model (see Fig. 4.5) are boxed, arrows represent the site of self-cleavage.

a


A
C
G
III
C
A
G
U
5′
3′
GAA
U
GAC
C
UGUUC
C
U
CUG
ACAAG
A
3′
II
A
G
U
A
G
U
C
I
5′

b


A
C
G
III
C
A
G
U
5′
3′
G A
U
GAC
C
UGUUC
C
U
CUG
ACAAG
A
3′
II
A
G
U
A
G
U
C
I
5′

c

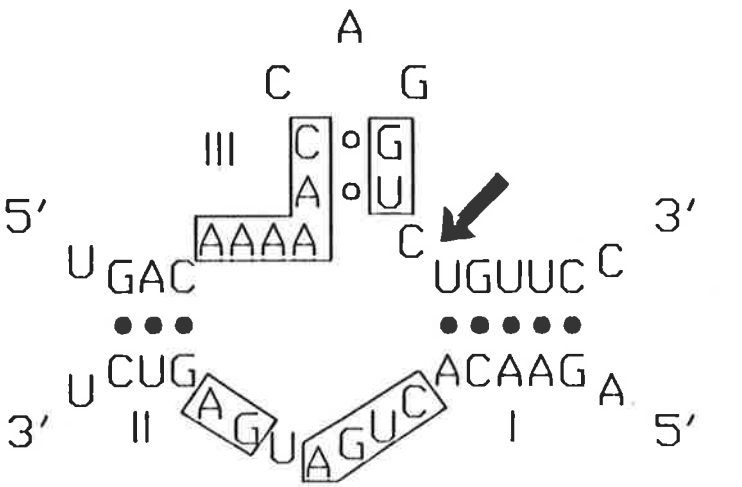
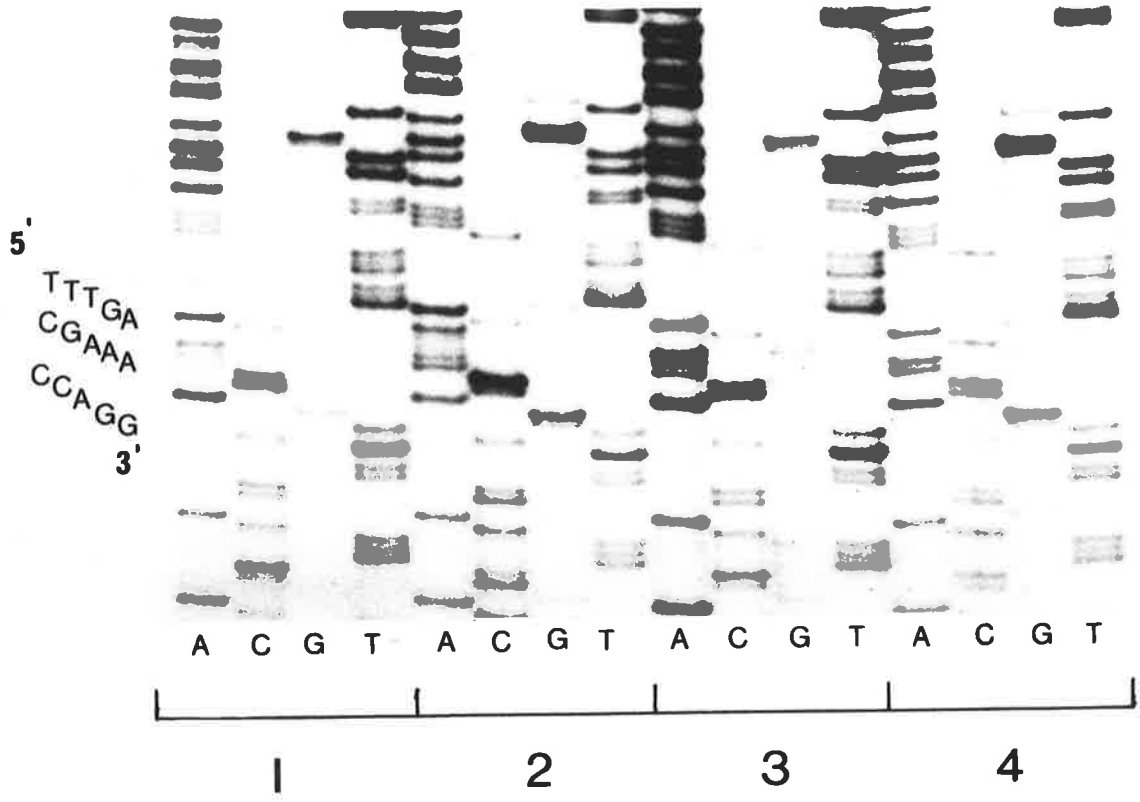

A
C
G
III
C
A
G
U
5′
3′
AAAA
U
GAC
C
UGUUC
C
U
CUG
ACAAG
A
3′
II
A
G
U
A
G
U
C
I
5′

Fig. 4.7 Dideoxy sequencing of wild-type and mutant ASBV M13 DNAs. 1 and 3 are the wild-type sequence, 2 is the sequnce of the "AAAAAC" mutant, and 4 is the sequence of the "GAAC" mutant. A part of the wild-type sequence is indicated.


5'
TTTGA
CGAAA
CCAGG
3'
A C G T A C G T A C G T A C G T
1
2
3
4

4.2.2 Construction of pSP64 trimeric plus ASBV clones

Trimeric ASBV clones, either wild-type or mutant in the second (middle) monomeric unit were prepared in the following manner. Double-stranded DNA of the mutant ("AAAAAC" and "GAAC") and wild-type monomeric ASBV clones in M13mp93 were prepared and digested with Sau3A I. These digests were run on an agarose gel and the monomeric sized bands (247 nucleotides for wild-type) were excised and eluted. These DNA fragments were ligated into a vector consisting of the dimeric ASBV pSP64 clone of Hutchins *et al.* (1986) which had been linearized at the junction of the two monomers with Bcl I and dephosphorylated. A schematic outline of clone preparation is given in Fig. 4.8. The ligated DNA was then transformed into competent MC1061 cells. DNA from the resulting ampicillin resistant colonies was prepared by the small scale method and screened for inserts of the correct (trimeric) size by a double digestion with Hind III and EcoR I followed by agarose gel electrophoresis. Clones containing inserts of the expected size were linearized with EcoR I and transcribed with SP6 RNA polymerase (using ^{32}P-UTP). Transcripts were analyzed by polyacrylamide gel electrophoresis and the gels autoradiographed. The inserts from clones yielding the expected transcription patterns were excised with Hind III and EcoR I, purified, and ligated into Hind III/EcoR I-cut M13mp18 and M13mp19 vectors. These clones were then sequenced to confirm the nature of the inserts.

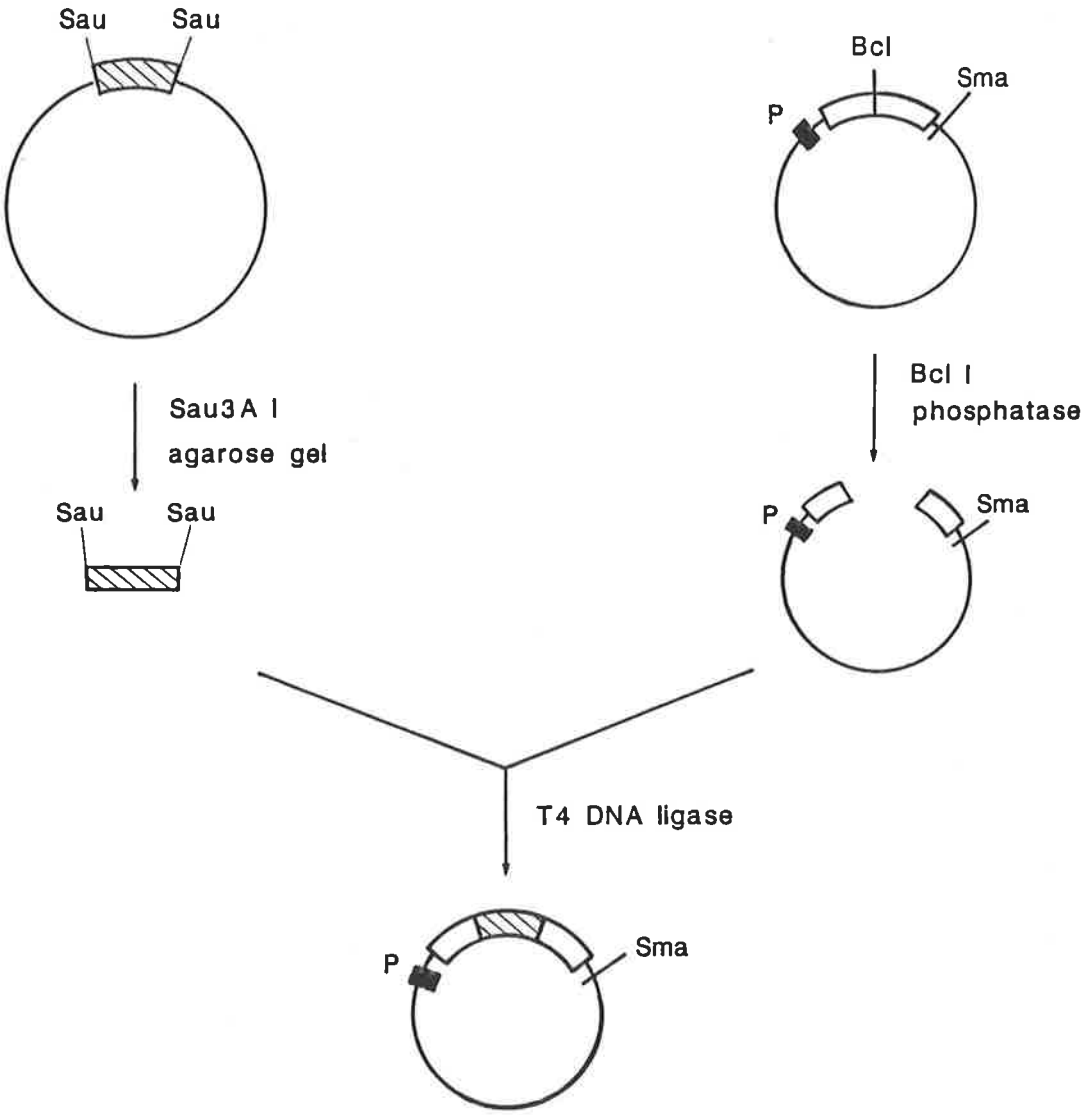
4.2.3 Construction of dimeric (+) ASBV pSP64 clones mutant in the sequences required for plus RNA self-cleavage

A schematic outline of the preparation of ASBV clones, containing a mutation in neither (m/m), both (m^{Δ}/m^{Δ}) or either (m^{Δ}/m, m/m^{Δ}) of the two monomeric units making up the dimeric insert, is given in Figs. 4.9, 4.10. The same Sau3A I monomeric fragments (both wild-type and the "GAAC" mutant) as used in the

Fig. 4.8 Outline of the method used for the construction of trimeric plus ASBV clones in pSP64. Construction of trimers mutant in middle monomer is shown. The monomeric ASBV insert (in this case mutant) is shown as a hatched box, open boxes represent wild-type ASBV sequences, solid boxes represent the SP6 RNA polymerase promoter.

WILD-TYPE AND MUTANT ASBV MONOMERS IN M13mp93

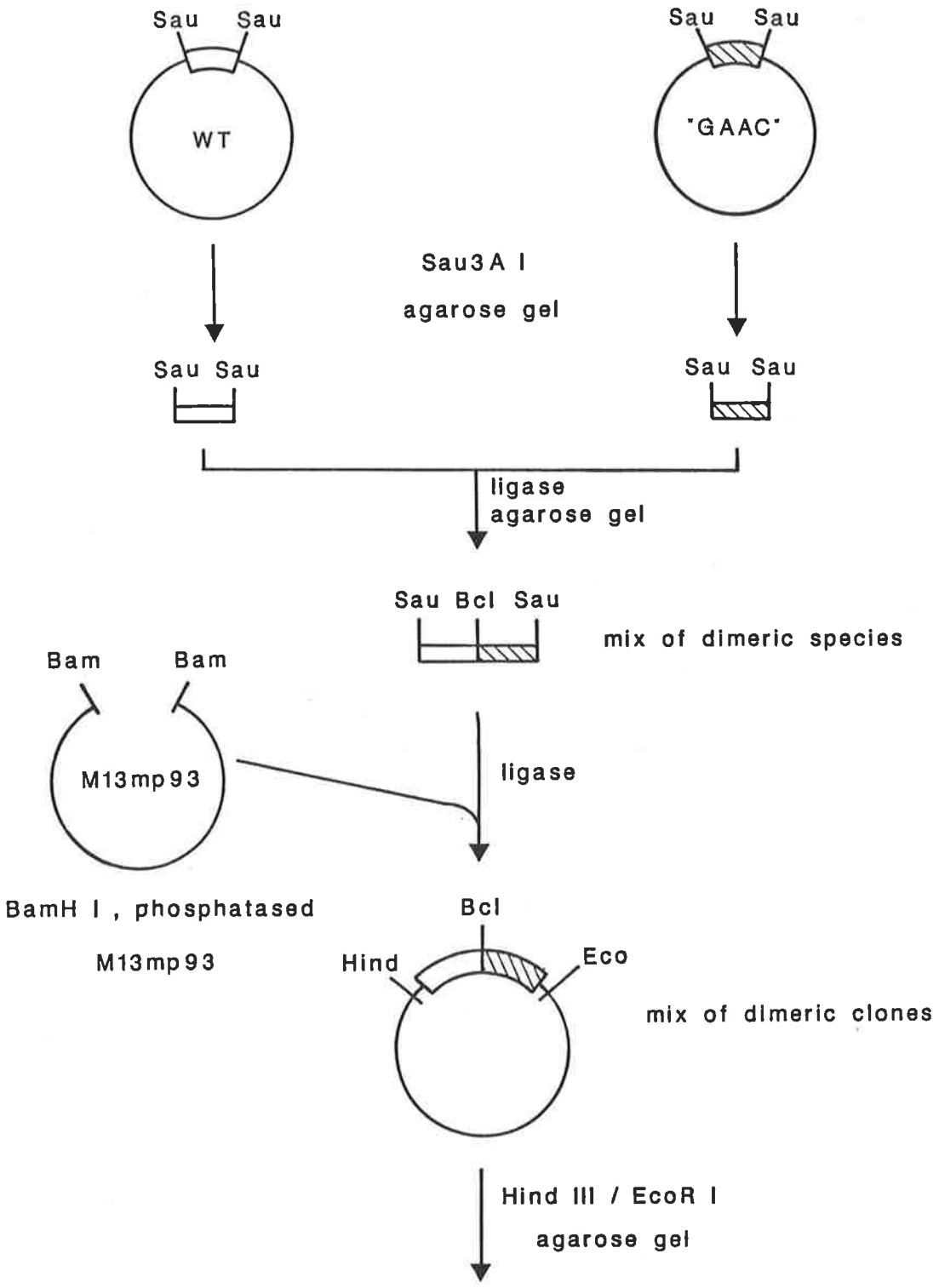
DIMERIC + ASBV pSP64 CLONE



TRIMERIC + ASBV pSP64 clones:wild-type or mutant in middle monomer

Fig. 4.9 Outline of the method used in the construction of mutant and wild-type dimeric (+) ASBV clones in pSP64. Mutant ASBV sequences are represented by hatched boxes, open boxes represent wild-type ASBV sequences, closed boxes represent the SP6 RNA polymerase promoter.

M13mp93 clones : wild-type & "GAAC" mutant monomer inserts



Continued over

Fig.4.9 continued

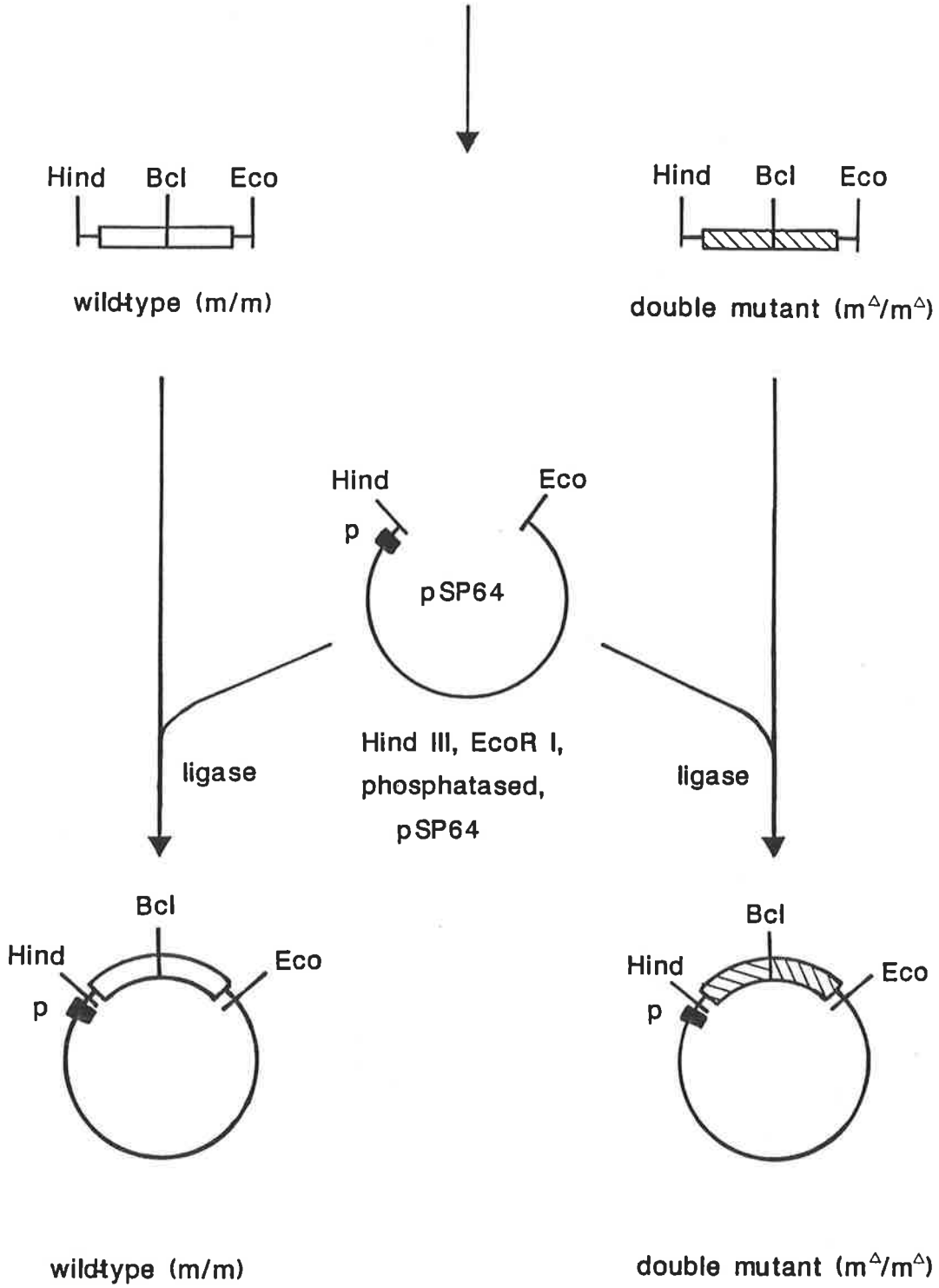
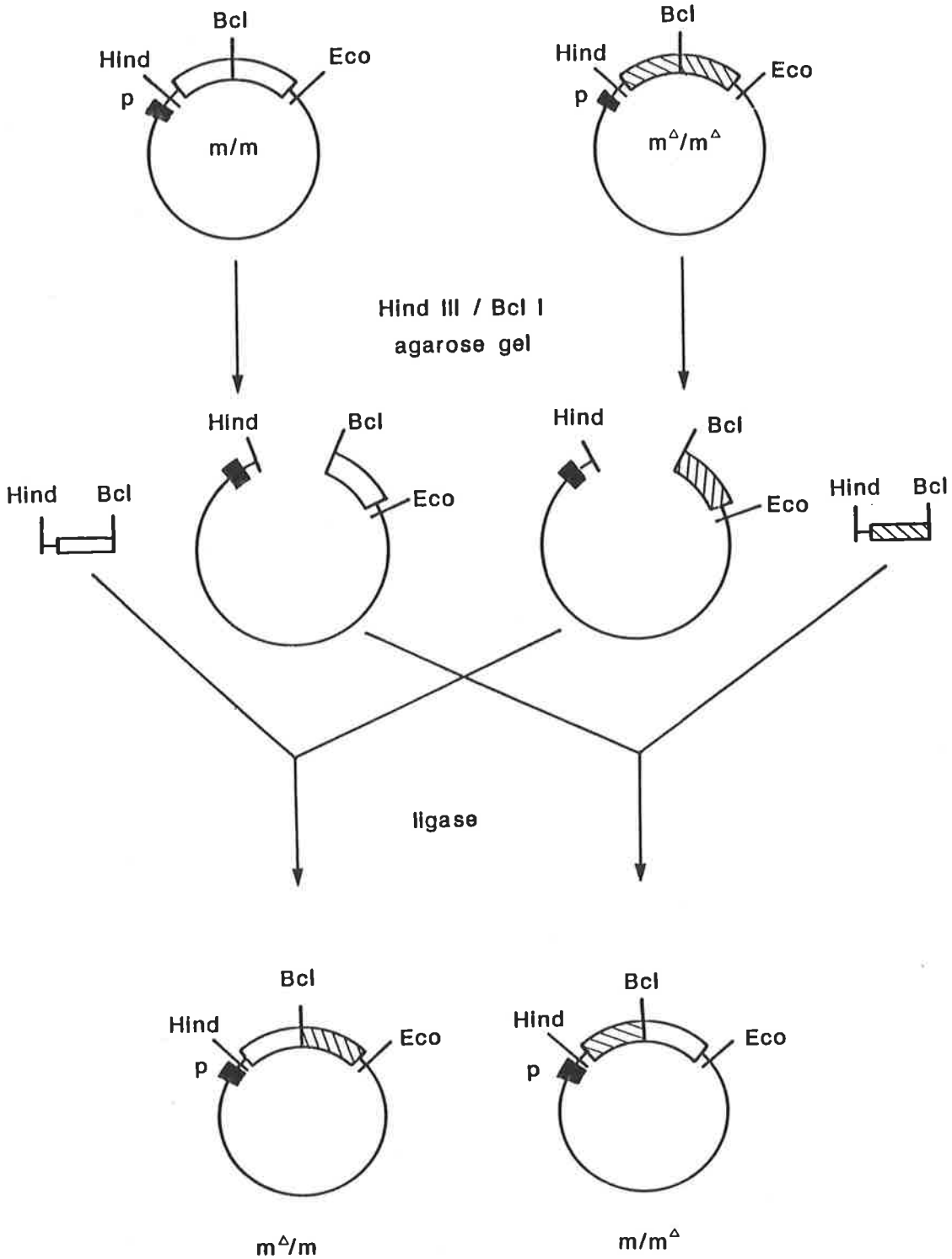


Fig. 4.10 Outline of the method used in the construction of dimeric (+) ASBV pSP64 clones mutant in one or other (m^{Δ}/m, m/m^{Δ}) of the monomeric units comprising the dimeric insert. Mutant ASBV sequences are represented by hatched boxes, open boxes represent wild-type ASBV sequences, closed boxes represent the SP6 RNA polymerase promoter.


Bcl
Hind
Eco
p
m/m
Bcl
Hind
Eco
p
m^Δ/m^Δ
Hind III / Bcl I
agarose gel
Hind
Bcl
Eco
Hind
Bcl
Hind
Bcl
Eco
Hind
Bcl
ligase
Bcl
Hind
Eco
p
m^Δ/m
Bcl
Hind
Eco
p
m/m^Δ

preparation of the trimeric clones were prepared and ligated at a ratio of 1:1 to form dimeric fragments. 100 ng/µl of each of mutant and wild-type monomers were ligated for 5 minutes at 16°C in 50 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 10 mM DTT, 5 mM ATP, 0.1 U/µl T4 DNA ligase. The reaction was stopped by the addition of EDTA to 20 mM followed by heating to 70°C for 10 minutes. This mix was run on an agarose gel and the dimer-sized band excised and eluted. This fragment was ligated into BamH I digested, dephosphorylated, M13mp93 vector and transformed into JM101. Double-stranded M13 DNA was prepared and assayed for the presence of dimer-sized inserts by agarose gel electrophoresis of Hind III/EcoR I digests. Positive clones were sequenced. By chance only m^{Δ}/m and m/m^{Δ} clones were detected and so to save searching through large numbers of clones (there being many possible permutations) the m^{Δ}/m and m/m^{Δ} clones were prepared as follows. The inserts from the m^{Δ}/m^{Δ} and m/m clones were excised with a double digest of Hind III and EcoR I and ligated into dephosphorylated Hind III/EcoR I-cut pSP64 vector. The resultant clones were grown in GM119 (a methylase free strain) and DNA prepared. Both clones were then digested with Bcl I and Hind III, the resultant fragments were purified by agarose gel electrophoresis and the purified fragments ligated as shown in Fig. 4.10 so that m^{Δ}/m and m/m^{Δ} clones were produced. To confirm their identity the inserts from all four dimeric pSP64 clones were subcloned into M13mp18 and M13mp19 and sequenced.

4.3 RESULTS

4.3.1 Mutant multimeric ASBV RNAs: testing of the double-hammerhead model

Tandem trimeric and dimeric clones of ASBV in the transcription vector pSP64 were made such that upon transcription with SP6 RNA polymerase, plus-sense RNAs were produced (see Methods). In these clones the conserved GAAAC sequences, present in the hammerhead structures (Fig. 4.5) were either wild-type (i.e.

GAAAC) or altered by oligonucleotide-directed mutagenesis. This sequence is conserved in all self-cleaving RNAs proposed to cleave by the hammerhead model (Forster and Symons, 1987a) and therefore is believed to be essential for the self-cleavage mechanism.

If the mutation proved effective (in that it prevented the GAAAC sequence from performing its role in self-cleavage) and if self-cleavage of these RNAs occurs by the single-hammerhead structure given in Fig. 4.1a, then mutation of a GAAAC sequence should only affect cleavage at the site less than 10 nucleotides 3' to it and not the other self-cleavage site(s) located much further away on the RNA strand. If the double-hammerhead model is correct then mutation of a GAAAC sequence will not affect the self-cleavage site located within only a few nucleotides but will affect the site(s) distant on the RNA sequence.

4.3.2 Choice of mutations in the ASBV RNAs

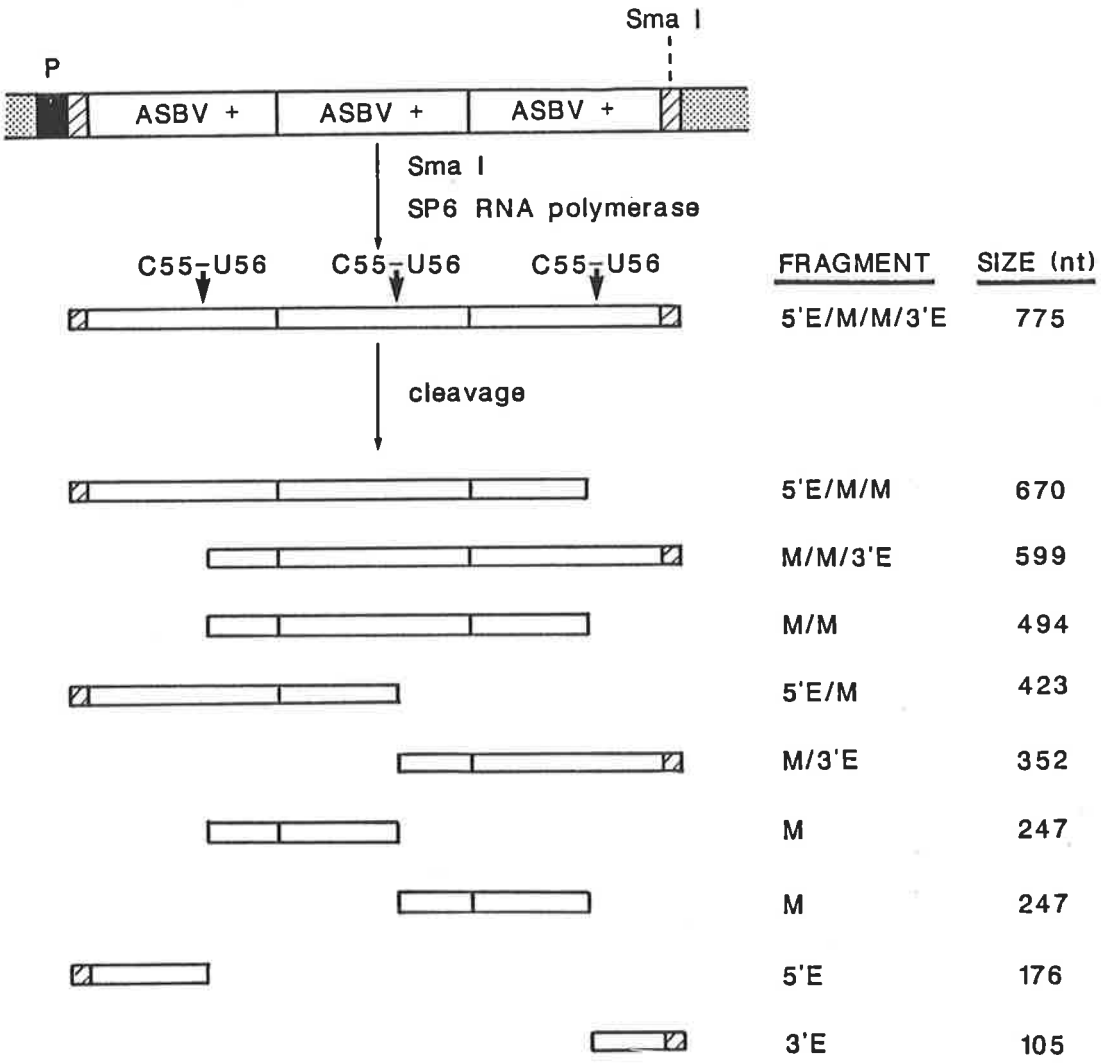
Oligonucleotide-directed mutagenesis was used to create mutations in an ASBV monomeric cDNA insert in M13mp93. Two different mutations were made; the location of these mutations is shown on the plus ASBV sequence drawn as a single-hammerhead (Fig. 4.6). In both mutants the same conserved GAAAC sequence was altered. The choice of which nucleotides to mutate was a difficult one as it was desirable to totally nullify the function of this sequence in the self-cleavage structure without drastically affecting the wild-type secondary structure. Too much disturbance of the structure may prevent the detection of any *cis* interaction either by not allowing enough of the wild-type structure to form or by creating an alternative, thermodynamically more stable conformation. Obviously it was important not to disrupt any potential base-pairing in the stem IIIs. The GAAAC sequence was chosen as the candidate for mutation because it was thought to be essential for self-cleavage (see above) and because the position of a particular GAAAC sequence relative to the

actual site of self-cleavage is different in the single and double-hammerhead models (compare Figures 4.1a and 4.4a). The deletions made were to remove an A nucleotide (GAAAC to GAAC, Fig. 4.6b) and to remove the G and insert two A nucleotides (GAAAC to AAAAAC, Fig. 4.6c). These two different mutations were tried as there was at the time no data regarding the sequence requirements of this area apart from its total conservation.

4.3.3 Trimeric plus ASBV RNAs may fold sequentially and cleave by the double-hammerhead structure

Initially it was decided to prepare trimeric clones, one of which was wild-type for all three GAAAC sequences and others that were mutated in the second (middle) of the GAAAC sequences (Fig. 4.8). There were two reasons for doing this. First, this approach was technically expedient as it did not require the formation of dimeric inserts but merely the insertion of a Sau3A I monomeric fragment, excised from the M13 clones, into the dimeric Sau (+) ASBV clone in pSP64 (Hutchins *et al.*, 1986) which had been linearized with Bcl I (see Methods). Thus it was hoped that this would allow the double-hammerhead concept to be tested in an economical fashion. This approach also provided an assay for the effectiveness of the mutations in the GAAAC sequences, as the very low levels of cleavage in the wild-type monomers (see below and Chapters 5, 6) makes use of monomers for this purpose impractical. Second, this approach could indicate (and in fact relies on) sequential folding of greater-than-unit-length transcripts as they are produced. This may give some insight into the *in vivo* replicative processes. If interaction between two self-cleavage domains by the double-hammerhead structure is required for self-cleavage and sequential folding from the 5'-end of the RNA occurs during transcription one would expect the wild-type RNA not to cleave at the 3'- proximal self-cleavage site and in the mutants cleavage would only occur at the second (middle) site (Fig. 4.11).

Fig. 4.11 Diagram of the wild-type trimeric Sau (+) ASBV template and the RNAs produced by its transcription with SP6 RNA polymerase. The DNA template is shown at the top with the full-length transcription product and the predicted self-cleavage fragments below. The name and size (in nucleotides) of the various RNAs are given at the right. The Sma I site at which the template DNA was linearized is indicated. The open boxes represent ASBV sequences, numbered after Symons (1981), hatched boxes represent M13mp93 derived sequences, the solid box represents the SP6 RNA polymerase promoter and the stippled boxes are pSP64 sequences. The self-cleavage sites are indicated by thick arrows.


Sma I
P
ASBV +
ASBV +
ASBV +
Sma I
SP6 RNA polymerase
C55-U56
C55-U56
C55-U56
cleavage
FRAGMENT
SIZE (nt)
5'E/M/M/3'E
775
5'E/M/M
670
M/M/3'E
599
M/M
494
5'E/M
423
M/3'E
352
M
247
M
247
5'E
176
3'E
105

The wild-type trimeric template used for transcription is shown in Fig. 4.11 with the sizes of all the possible fragments arising from self-cleavage of the full-length plus RNA transcript. Fig. 4.12 is an autoradiograph of the transcription products of both wild-type and mutant templates. Transcription of the dimeric Sau (+) ASBV pSP64 clone of Hutchins *et al.*, (1986) (for details of this clone see Fig. 4.13a) gives rise to plus-sense RNA transcripts (Fig. 4.12 lane 2, non-linearized; lane 3, linearized with Sma I). These are included as reference markers as these RNAs have been extensively characterized (Hutchins *et al.*, 1986). The disappearance of the RNAs of the size expected for the 3'-end (3'E) fragments in runoff transcriptions of both dimeric (lane 2) and trimeric templates (lanes 4, 5) confirms their identity.

If the premises, 1) folding of the RNA transcript occurs in a strictly sequential manner as it is produced (and does not refold) and 2) cleavage occurs by the double-hammerhead model, are valid, then the following results could be expected upon transcription of the trimeric clones. Cleavage at the 3'-proximal self-cleavage site which gives rise to the 3'E fragment (Fig. 4.11) should not occur in both the wild-type and mutant transcripts, as in theory it should not be able to participate in the formation of a double-hammerhead structure as it would have no partner. This is because both of the other self-cleavage domains predicted to be involved in the formation of a double-hammerhead structure are synthesized prior to synthesis of the 3' monomer. In addition, in the mutant, if the first two monomers (starting from the 5' end) form a double-hammerhead then cleavage at the 5'-proximal self-cleavage site will also be prevented. This should result in a decrease in the amount of monomer (M) and an absence of the dimeric and 5'-end fragments. As a corollary to this the amount of 5'E/M fragment should increase due to prevention of cleavage at the 5'-proximal self-cleavage site.

A comparison of the transcripts of the Sma I linearized trimeric wild-type and mutant templates (Fig. 4.12, lanes 6 and 7 respectively) reveals that as predicted by

Fig. 4.12 Autoradiograph of 5% polyacrylamide, 7M urea gel, showing plus RNA transcripts of dimeric and trimeric ASBV templates (Fig. 4.11, 4.13). Lane 1; ^{32}P-labelled Hpa II digested pUC19 DNA markers with the sizes (in nucleotides) of the various bands indicated. Lane 2; Transcripts of non-linearized wild-type dimeric ASBV pSP64 template (Fig. 4.13a). Lane 3; As for lane 2 but template linearized with Sma I. Lane 4; Transcripts of non-linearized wild-type trimeric ASBV pSP64 template (Fig. 4.11). Lane 5; Transcripts of non-linearized mutant trimeric ASBV pSP64 template. Lane 6; As for lane 4 but template linearized with Sma I. Lane 7; As for lane 5 but template linearized with Sma I. The fragments are labelled as shown in Fig. 4.11 and 4.13a.

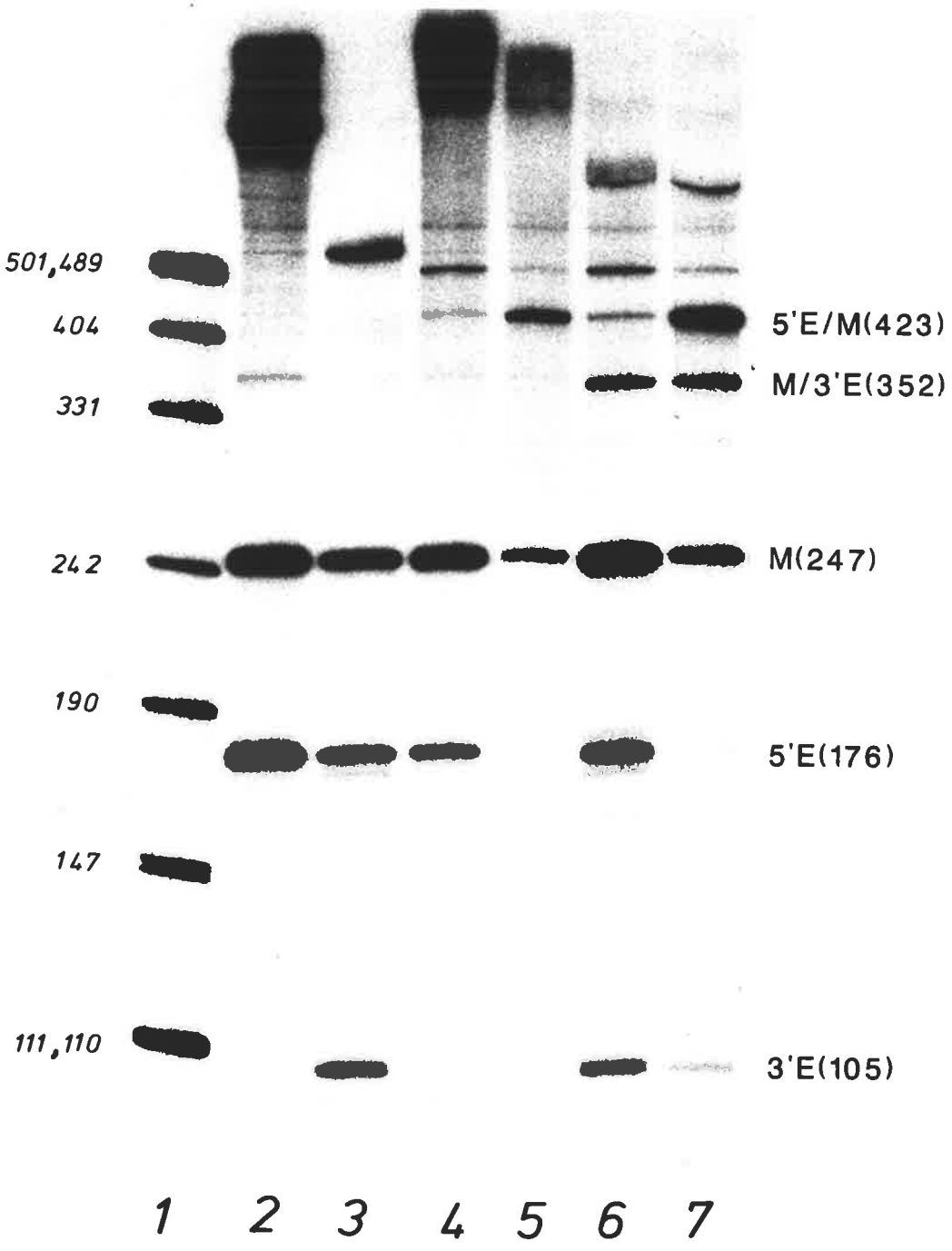
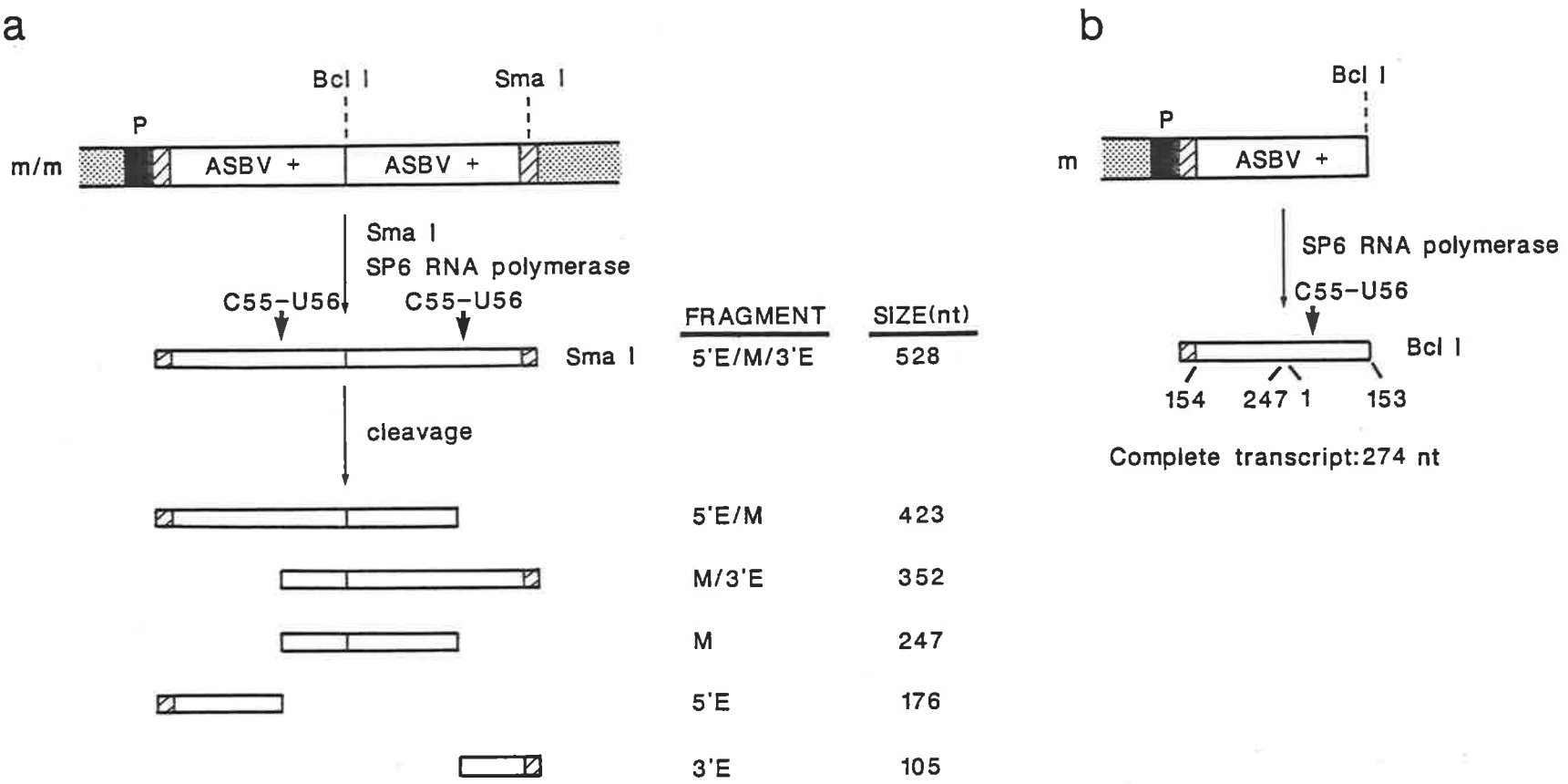

501,489
404
331
242
190
147
111,110
5'E/M(423)
M/3'E(352)
M(247)
5'E(176)
3'E(105)
1 2 3 4 5 6 7

Fig. 4.13 Diagram of wild-type Sau (+) ASBV pSP64 templates and the RNAs produced by SP6 RNA polymerase transcription. The dimeric clone was linearized with either Sma 1 (a; m/m) or Bcl I (b; m). Hatched boxes represent M13mp93 derived sequences, stippled boxes represent pSP64 vector sequences, open boxes represent ASBV sequences and the SP6 RNA polymerase promoter is represented by a solid box. Self-cleavage sites are indicated by thick arrows; the ASBV sequence is numbered after Symons (1981). The names and sizes (in nucleotides) of the various fragments are given.


a
Bcl I
Sma I
P
m/m
ASBV +
ASBV +
Sma I
SP6 RNA polymerase
C55–U56
C55–U56
Sma I
cleavage
FRAGMENT
SIZE(nt)
5'E/M/3'E
528
5'E/M
423
M/3'E
352
M
247
5'E
176
3'E
105
b
Bcl I
P
m
ASBV +
SP6 RNA polymerase
C55–U56
Bcl I
154
247
1
153
Complete transcript:274 nt

the double-hammerhead model, mutation of the middle GAAAC sequence does not affect the self-cleavage site nearest to it (i.e. just 3' to it) but eliminates cleavage at another site distant to the mutated GAAAC on the linear RNA molecule. This is shown by the absence of the 5'E fragment, the decrease in the amount of monomer and an increase in the amount of 5'E/M fragment in the transcription products of the mutant template (compare lanes 6, 7). Although the amount of dimeric RNA may have been reduced in the mutant, identification and quantification of the higher molecular weight bands was difficult. The AAAAAC mutant is not shown on this gel but it had precisely the same effect as the GAAC mutation. As the GAAC mutant was later shown to completely prevent self-cleavage (see below) it is reasonable to assume that this is also the case of the AAAAAC mutant.

These results appear to indicate that the double-hammerhead model is valid and that the two mutations made were effective but there are obvious inconsistencies which do not allow total confidence. The 3'E fragment present in the transcripts of linearized trimeric templates (Fig. 4.12, lanes 6, 7) should not be present if sequential folding of the transcript is absolute and if melting and refolding of the RNAs does not occur to a significant extent. Thus although predictions concerning the 5'-proximal cleavage site seem to be fulfilled, for the 3'-proximal self-cleavage site some other interaction is occurring to enable cleavage at this site. For this reason a less complicated experiment to test the double-hammerhead model, using dimeric clones was performed.

4.3.4 Mutant dimeric ASBV RNAs demonstrate the validity of the double-hammerhead model for the plus-sense sequence

To overcome the problems associated with experiments involving trimeric templates an experiment using dimeric clones was undertaken. This involved the production of four different dimeric clones, i.e. clones that contained either, two wild-

type GAAAC sequences (m/m) or two mutant GAAAC sequences (m^{Δ}/m^{Δ}) and clones that contained one wild-type sequence (m^{Δ}/m, m/m^{Δ}) (see Methods for details of production). The wild-type dimeric (+) ASBV pSP64 template and the RNAs produced by SP6 RNA polymerase transcription of this template linearized with either Sma I (m/m) or Bcl I (m) are given in Fig. 4.13. This clone was the same as that used by Hutchins *et al.*, (1986) (see Materials this chapter for further details). Upon transcription the Sma I linearized template (m/m) yields a full-length product (5'E/M/3'E) which contains two self-cleavage sites. This RNA cleaves efficiently at both sites to yield a monomeric RNA (M) of 247 nucleotides (Fig. 4.14, lane 2) and the two end fragments, 5'E and 3'E. When the dimeric template was truncated with Bcl I and transcribed with SP6 RNA polymerase a plus monomer-sized RNA (M*) was produced. This RNA underwent negligible self-cleavage during transcription (Fig. 4.14, lane 1) even though it contained one complete set of the sequences required for self-cleavage (according to the single-hammerhead model). This result supports the double-hammerhead model because according to this model this monomeric RNA can only form an active self-cleavage structure (i.e. a structure with stable stem IIIs, Fig. 4.4a) if two monomers become associated during the transcription reaction to form a double-hammerhead. The intermolecular formation of a double-hammerhead structure is expected to be much less rapid than formation of the equivalent structure within a single dimeric molecule. For comparison a schematic representation of the dimeric RNA folded as a double-hammerhead is compared with the same RNA folded to contain two single-hammerheads in Fig. 4.15.

The mutated dimeric ASBV templates described above yielded mutant plus RNA transcripts that provided convincing evidence for the double-hammerhead model. When both GAAACs were mutated, cleavage at both sites was abolished (Fig. 4.14, lane 3), demonstrating the importance of this sequence for cleavage. When only the 5'-proximal GAAAC (labelled A in Fig. 4.15) was mutated, cleavage at self-cleavage site one (SC-1 in Fig. 4.15), which is spatially removed from the

Fig. 4.14 Autoradiograph of 5% polyacrylamide, 7M urea gel showing RNA transcripts of Sau plus ASBV templates (Fig. 4.13). Lane 1; Transcript (*) of wild-type dimeric ASBV template truncated with Bcl I (m, see Fig. 4.13b). Lane 2; As for lane 1 but template linearized with Sma I (m/m, see Fig. 4.13a). Lane 3; As for lane 2 but both template GAAAC sequences (A and B) (see Fig. 4.14, 4.15) mutated to GAAC (m^{Δ}/m^{Δ}). Lane 4; As for lane 3 but only GAAAC (A) mutated (m^{Δ}/m). Lane 5; As for lane 3 but only GAAAC (B) mutated (m/m^{Δ}). Labels down margin represent RNA fragments, either full-length (5'E/M/3'E) for Sma I digested template, * for Bcl I digested template or fragments arising from self-cleavage (see Fig. 4.13).

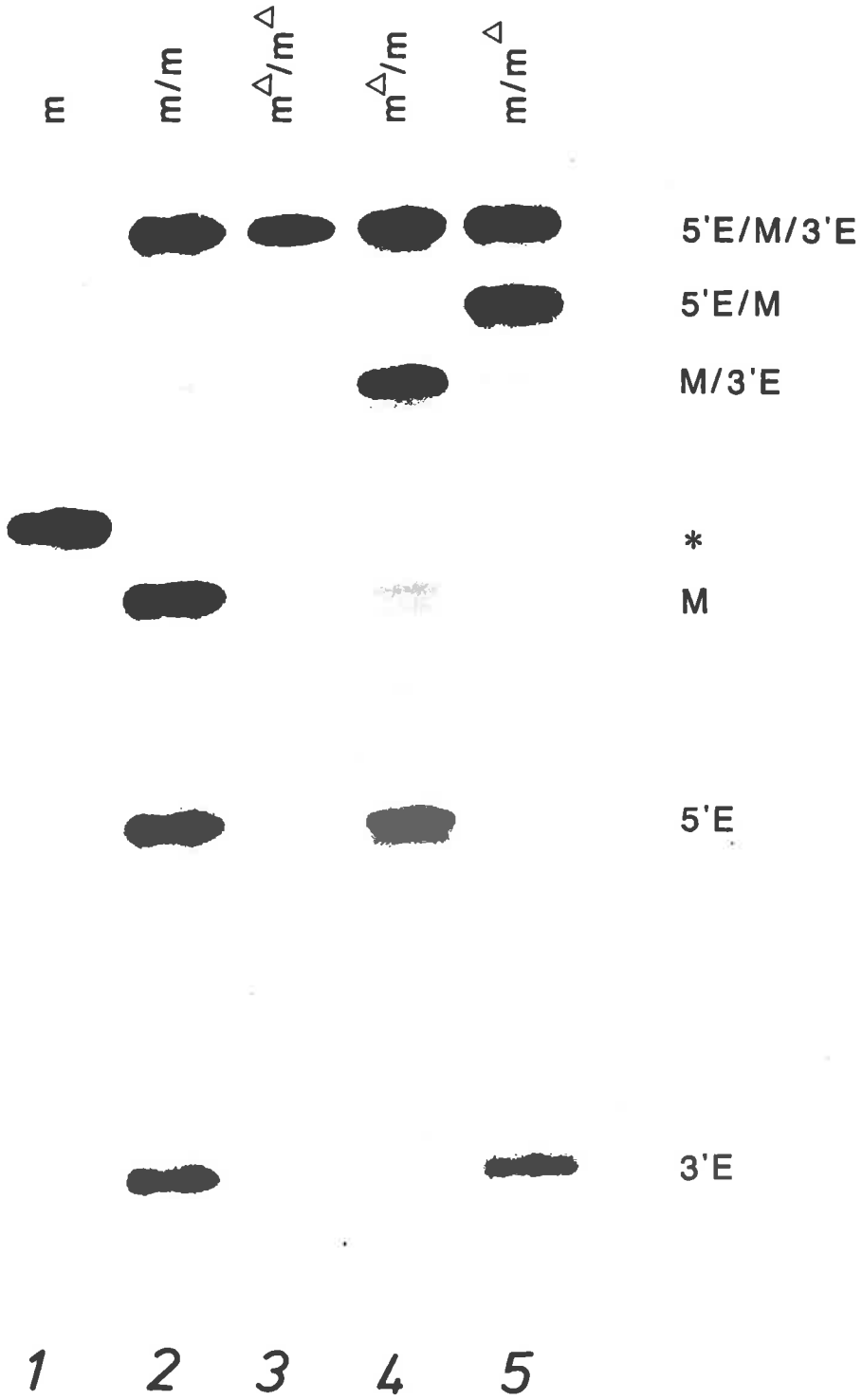
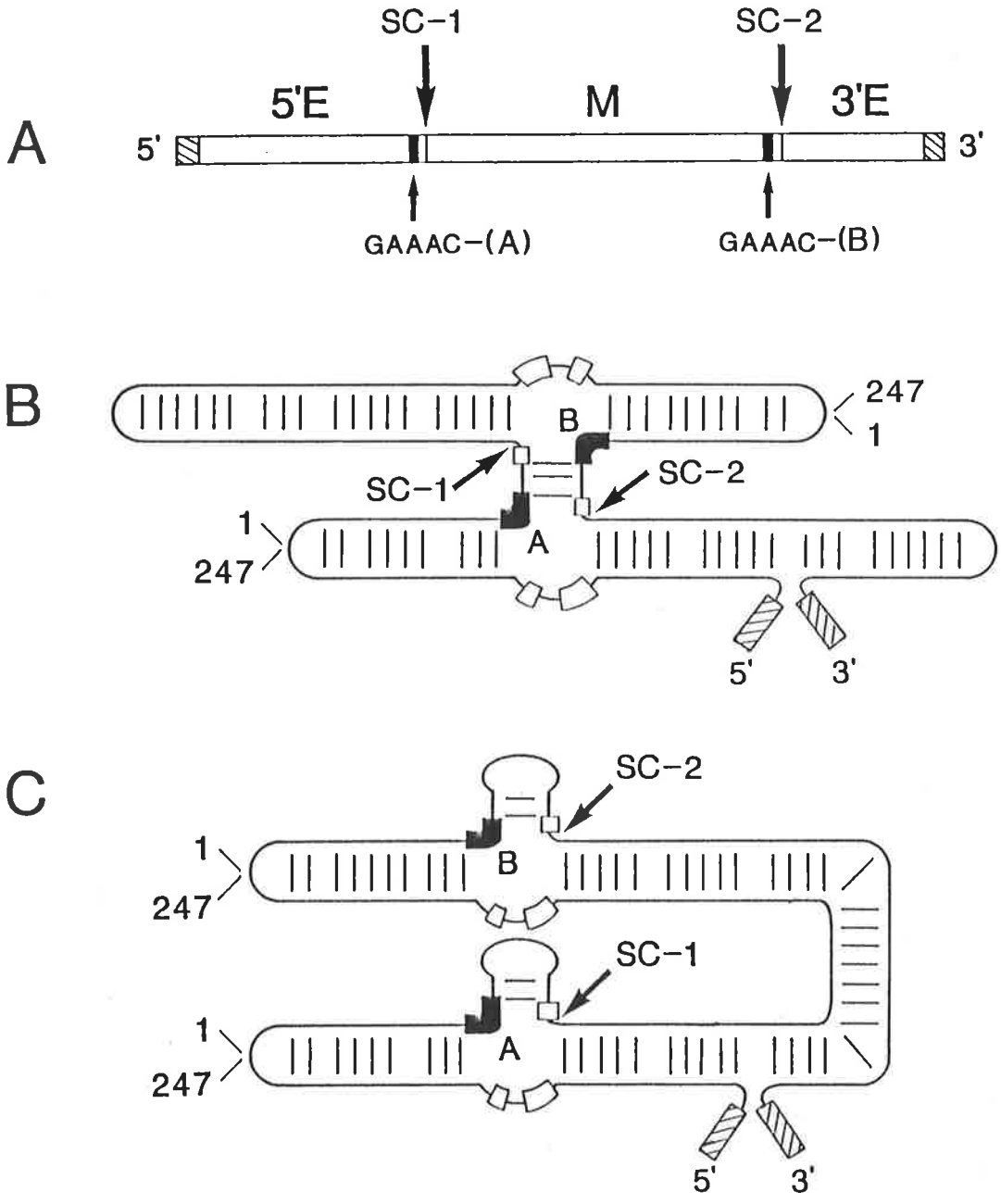

E
m/m
m^Δ/m^Δ
m^Δ/m
m/m^Δ
5'E/M/3'E
5'E/M
M/3'E
*
M
5'E
3'E
1
2
3
4
5

Fig. 4.15 Schematic representation of dimeric plus ASBV RNA transcript unfolded and folded in single- and double-hammerhead conformations (a), Unfolded SP6 RNA polymerase transcript of dimeric plus ASBV clone (see Fig. 4.13). Self-cleavage sites, labelled SC-1 and SC-2, are indicated by arrows; hatched boxes, vector sequences; closed boxes, GAAAC sequences (see Fig. 4.1a) labelled A and B indicated by arrows. (b), Schematic diagram of dimeric plus ASBV RNA drawn containing a double-hammerhead. Labelled as for Fig. 4.15a except that the open boxes represent conserved nucleotides (see Fig. 4.5); base-pairing is represented by lines between the RNA strands. Sequence numbered after Symons (1981). (c), Schematic diagram of dimeric plus ASBV RNA transcript drawn to contain two single-hammerheads. Labelled as in Fig. 4.15b.


A
SC-1
SC-2
5'E
M
3'E
5'
3'
GAAAC-(A)
GAAAC-(B)
B
247
1
B
SC-1
SC-2
1
247
A
5'
3'
C
SC-2
1
247
B
SC-1
1
247
A
5'
3'

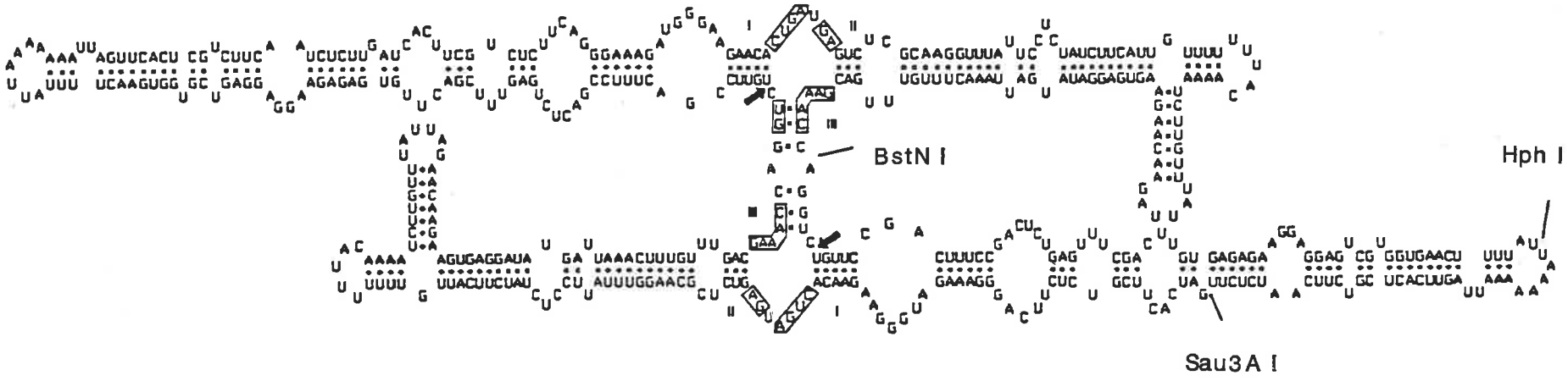
mutated GAAAC in the double-hammerhead model, was unaffected whereas cleavage at self-cleavage site two (SC-2) was abolished as indicated by the presence of only the 5'E (176 nucleotides) and M/3'E (352 nucleotides) fragments (Fig. 4.14, lane 4). In the single-hammerhead model the reverse would be expected (compare Fig. 4.15b and c). Similarly, mutation of the 3'-proximal GAAAC (labelled B in Fig. 4.15) abolished cleavage at SC-1 but did not affect cleavage at SC-2 as indicated by the presence of only the 5'E/M (423 nucleotides) and 3'E (105 nucleotides) fragments (Fig. 4.14, lane 5). Hence, these results provide strong evidence for the double-hammerhead model for plus ASBV. The complete dimeric plus ASBV RNA sequence folded to contain the double-hammerhead structure is given in Fig. 4.16.

4.4 DISCUSSION

4.4.1 The conserved GAAAC sequence is required for self-cleavage

Transcription of clones mutated in the conserved GAAAC sequence that is found in all RNAs proposed to cleave by a hammerhead type secondary structure (Fig. 4.5) showed that this sequence is required for self-cleavage. Some apparently less drastic changes to this sequence in the newt RNA have been shown to only partially prevent self-cleavage (Koizumi *et al.*, 1988). What role this sequence plays in the formation of the active structure is unknown but it has been noted by Sampson *et al.* (1987) that in tRNA it is usually the semi-conserved or conserved single-stranded regions that participate in tertiary structure interactions and so the conserved nucleotides in single-strand regions of the hammerhead structure may perform a similar function. At least a part of the GAAAC sequence appears to be in a single-strand region and it is therefore likely to be involved in tertiary interactions crucial to the formation of a functional self-cleavage structure. These tertiary interactions apparently bring the functional groups required for the reaction into close proximity thus allowing self-cleavage to occur.

Fig. 4.16 Predicted secondary structure of dimeric plus ASBV RNA folded to contain a double-hammerhead. The self-cleavage sites are indicated by thick arrows; boxed nucleotides, conserved sequences (see Fig. 4.5); stems are numbered 1 to 111. Various restriction endonuclease cleavage sites are marked. Numbering after Symons (1981).


BstN I
Hph I
Sau3A I

The use of trimeric ASBV clones was important in the development of experiments to determine the active structure for plus ASBV RNA self-cleavage. Not only did these results suggest that the mutations in the GAAAC sequence were effective in preventing self-cleavage and that the interaction of more than one self-cleavage domain was occurring during transcription (a fact more clearly demonstrated in mutant dimeric RNAs, see below) but they also showed that, at least to some degree, sequential folding of these RNAs occurs during transcription. These experiments involving the self-cleavage of RNAs during transcription, where the RNAs are produced and folded sequentially, may more closely resemble the *in vivo* situation than studies using RNAs purified from transcription reactions.

The main aim of the experiments outlined in this chapter was to identify the structure responsible for self-cleavage in plus ASBV RNAs. Although single-hammerhead structures for a number of self-cleaving RNAs have been proposed on the basis of sequence and structural homologies the stem IIIs of the newt and ASBV RNAs are much weaker than those found in the satellite RNAs (compare Figs. 4.1 and 4.3). Two possible reasons for this anomaly are that 1) the parameters used for the prediction of stem stabilities in RNA are not relevant in this case; for example peculiar primary sequence related effects or tertiary interactions could be stabilizing the apparently weak stems, and 2) these stems are indeed weak, and that the single-hammerhead model is not adequate and that some other mechanism is in operation. The proposal that plus ASBV RNAs form a double-hammerhead in which two self-cleavage structures come together to form a structure which has stable stem IIIs has resolved this dilemma.

Apart from the evidence arising from the trimeric and dimeric ASBV RNAs studies using short RNAs generated by RNA polymerase transcription of synthetic oligonucleotide templates have also supported the double-hammerhead model (Forster *et al.*, 1988). In these experiments it was shown that the self-cleavage of an

RNA, containing only those sequences of the newt RNA thought to be sufficient for cleavage, was concentration dependent in a manner expected for a bimolecular reaction. The stem III of the newt-like RNA used in these experiments (Fig. 4.1c) is theoretically even less stable than its counterpart in the plus ASBV single-hammerhead structure. Taken together these results offer strong support for the cleavage of molecules with theoretically labile stem IIIs occurring by the proposed double-hammerhead model.

Demonstration of the double-hammerhead self-cleavage structure answers the previously puzzling question of the lack of cleavage in monomeric RNA. This model suggests that the cleavage of monomers should be concentration dependent. The lack of cleavage of monomeric transcripts is therefore probably due to their relatively low concentration under the transcription reaction conditions used.

The double-hammerhead model is also relevant to the work of Koizumi *et al.* (1988) who utilized two short RNAs with newt-like self-cleavage sequences that can act in *trans* to undergo cleavage. Their results obtained for mutations in the various conserved sequences should be viewed in terms of the double- not the single-hammerhead model (this will be discussed more fully in Chapter 5).

CHAPTER 5

DOUBLE-HAMMERHEAD SELF-CLEAVAGE OF MINUS ASBV RNAs

5.1 INTRODUCTION

In the previous chapter it was demonstrated that self-cleavage of plus ASBV RNAs probably occurs by double-hammerhead structures involving the association of two self-cleavage domains. In the structures proposed (Fig. 4.4a,b) the strong stem IIIs apparently required for formation of the active self-cleavage structures are created by base-pairing between the sequences which form the weak stem IIIs in the single-hammerhead model (Fig. 4.1a,b).

Demonstration that plus ASBV RNAs are likely to cleave by a double-hammerhead structure still left the question of the mode of minus ASBV RNA self-cleavage unresolved. This is because, of the three RNAs that have theoretically labile stem IIIs when drawn as single-hammerheads (i.e. newt, plus and minus ASBV RNAs, Fig. 4.1), the minus ASBV structure would appear to be potentially more stable. To allow the double-hammerhead model to be universally acceptable it was therefore important to demonstrate that it pertained to the self-cleavage of both plus and minus ASBV RNAs. If this could not be demonstrated then the requirement for double-hammerhead formation of plus ASBV RNAs *in vitro* would rightfully be viewed as an interesting result but one that was less likely to be relevant *in vivo*. For example it could be proposed that the labile stem IIIs are stabilized by cellular proteins (a suggestion that cannot be ruled out but for which there is no evidence). This question of general applicability was resolved by the demonstration that minus ASBV RNA self-cleavage, like plus ASBV RNA self-cleavage occurred by the double-hammerhead model.

5.2 METHODS

5.2.1 Preparation of mutant monomeric ASBV clones in M13mp93; mutant in the sequence required for minus RNA self-cleavage

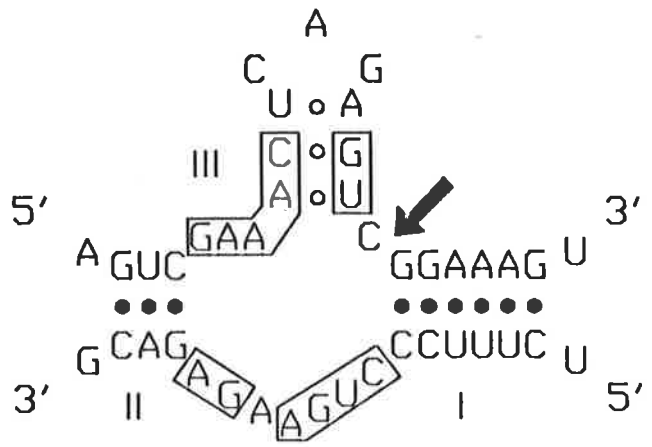
An M13mp93 clone containing a full-length monomeric ASBV insert in the BamH I site (Barker *et al.*, 1985) was the starting material for mutagenesis. Oligonucleotide-directed mutagenesis of this clone was performed as described in Chapter 2. The following oligonucleotide was used for mutagenesis (and as a probe for screening) "GAAC", 5'-TCTCTCACAAGTCGAACTCAGAGTCGGAAA, the change induced in the wild-type minus ASBV sequence is shown in Fig. 5.1. Positive clones detected by hybridization screening were sequenced by the dideoxy method (see Methods).

5.2.2 Construction of dimeric ASBV pGem-I clones mutant in the minus ASBV self-cleavage domain

A schematic outline of clone preparation in given in Fig. 5.2. Clones were prepared in which either both (m^{Δ}/m^{Δ}), none (m/m) or one (m^{Δ}/m, m/m^{Δ}) of the monomeric units making up the dimeric insert were mutant. Double-stranded DNA of the mutant ("GAAC") M13 clone was prepared and the insert excised with Sau3A I. This monomer was then ligated into the dephosphorylated Bcl I linearized, dimeric ASBV pSP64 clone (see Chapter 4) to produce trimers that were mutant in the second (middle) monomer. After transformation into competent MC1061 cells, the DNA from ampicillin resistant colonies were tested for trimer-sized inserts by agarose gel electrophoresis of the Hind III, EcoR I excised inserts. These trimers were assumed to be head to tail. This procedure using trimeric clones was designed to ensure that digestion with BstN I would produce an exact 1:1 ratio of wild-type:mutant monomers for ligation to form dimers. It also circumvented the need to

Fig. 5.1 Single-hammerhead models for wild-type (a) (Hutchins *et al.*, 1986) and mutant (b) minus ASBV RNAs. The boxed residues are conserved (see Fig. 4.5); arrows indicate the site of self-cleavage and stems are numbered 1 to 111. In (b) the conserved GAAAC sequence has been mutated to GAAC.

a


A
C G
U o A
III
C o G
A o U
5′
3′
A GUC
GAA
C GGAAAG U
G CAG
CCUUUC U
3′
II
AG A AGUC
I
5′

b

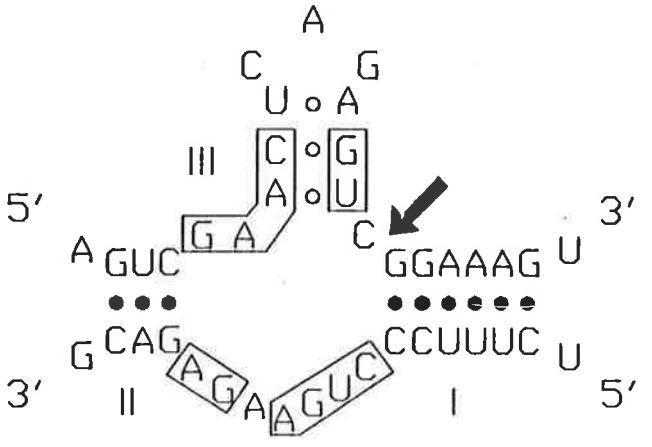
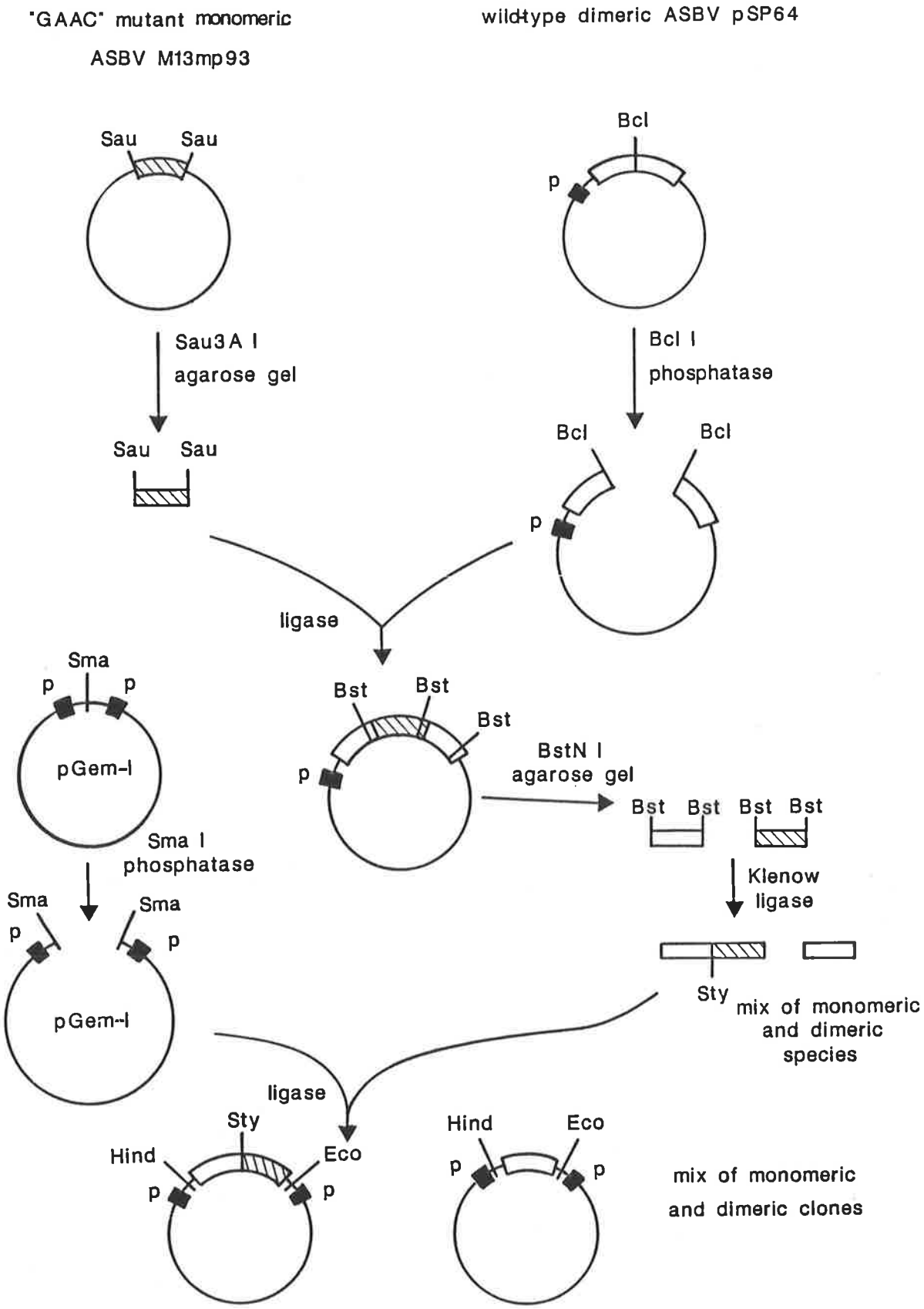

A
C G
U o A
III
C o G
A o U
5′
3′
A GUC
G A
C GGAAAG U
G CAG
CCUUUC U
3′
II
AG A AGUC
I
5′

Fig. 5.2 Diagram of the method used in the preparation of monomeric and dimeric Bst N (-) ASBV pGem-1 clones. Hatched boxes, mutant ASBV sequences; open boxes, wild-type ASBV sequences; closed boxes, SP6 and T7 RNA polymerase promoters. The Sty I site generated at the junction of the two monomers is indicated.


"GAAC" mutant monomeric
ASBV M13mp93
wild-type dimeric ASBV pSP64
Sau
Sau
Bcl
p
Sau3A I
agarose gel
Bcl I
phosphatase
Sau
Sau
Bcl
Bcl
p
ligase
Sma
p
p
pGem-I
Bst
Bst
Bst
p
BstN I
agarose gel
Bst
Bst
Bst
Bst
Sma I
phosphatase
Klenow
ligase
Sma
Sma
p
p
pGem-I
Sty
mix of monomeric
and dimeric
species
ligase
Sty
Hind
Eco
p
p
Hind
Eco
p
p
mix of monomeric
and dimeric clones

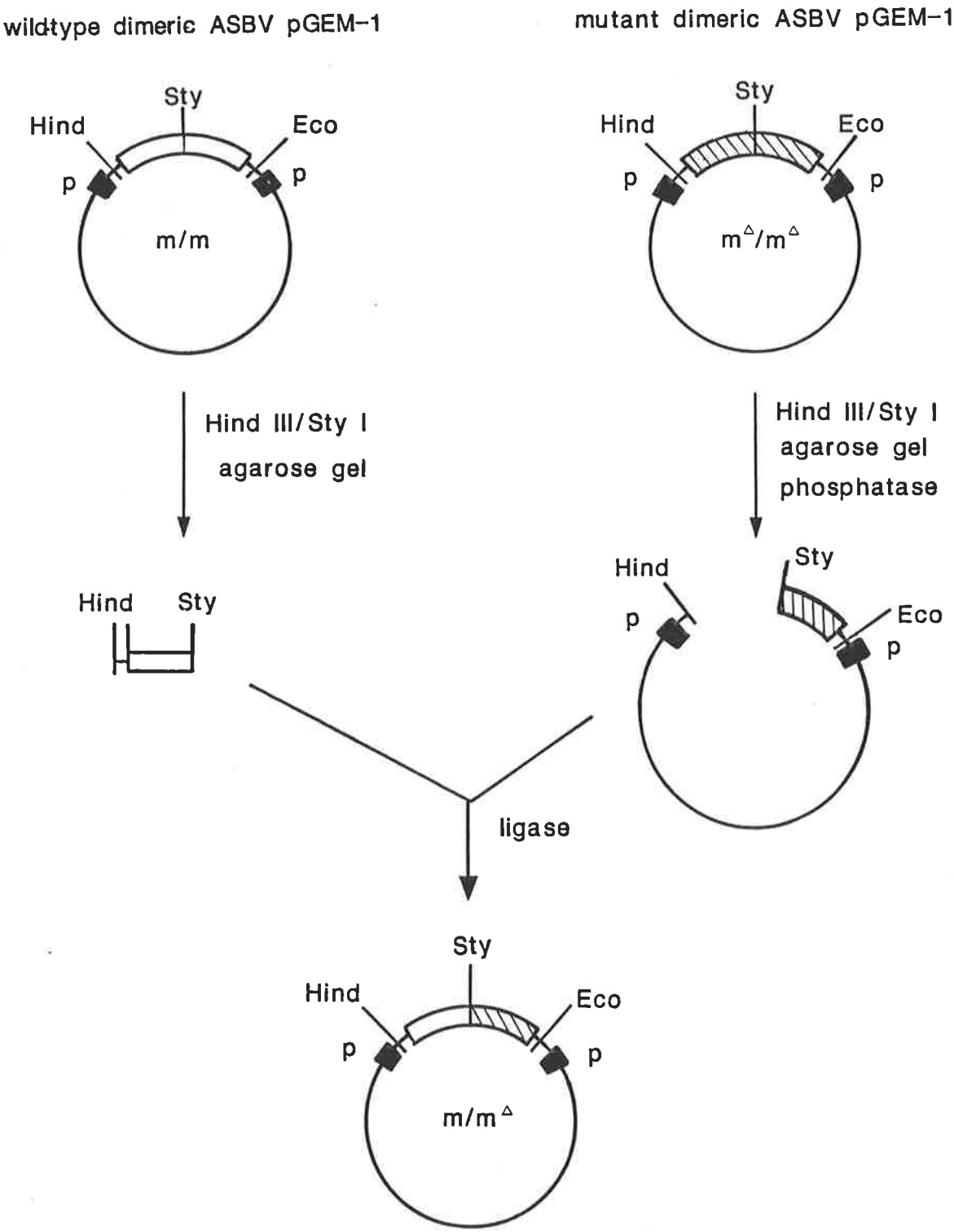
produce dimeric Sau3A I mutant dimers which would be required to obtain the mutant BstN I monomeric fragment necessary for dimer production. A suitable trimeric clone was digested with BstN I and the monomer-sized fragments purified by agarose gel electrophoresis. BstN I digestion produces a single nucleotide 5'-overhang and hence these fragments were end-filled with the Klenow fragment of DNA polymerase I. After end-filling the DNA was ethanol precipitated and ligated to form linear dimeric molecules. As this was a ligation of blunt ends the reaction was expected to be much less efficient than that performed in Chapter 4 for the sticky, Sau3A I generated ends. Therefore conditions that gave the best yield of dimers were determined empirically by varying the time of ligation and ligase concentration. The conditions determined were the same as those used for preparation of Sau3A I dimers (i.e. high concentrations of monomeric fragments were required to promote intermolecular ligation rather than intramolecular ligation (circularization) except that the reactions were carried out for 70 minutes and 0.2 U/μl of T4 DNA ligase was used. The ligation mix (containing dimers) was then ligated into dephosphorylated Sma I digested, pGem-I vector. The dimer-sized fragments were not purified by agarose gel electrophoresis, to reduce losses and because clones with dimeric and monomeric inserts were sought. The DNA mix was transformed into competent MC1061 cells. DNA from ampicillin resistant colonies was prepared by the rapid small-scale method, digested with Hind III and EcoR I and assayed for inserts by agarose gel electrophoresis. Clones containing monomer- and dimer-sized inserts were transcribed with SP6 and T7 RNA polymerases after digestion with either EcoR I or Hind III respectively. The inserts were excised, from clones which gave the appropriate transcription patterns, with Hind III and EcoR I and subcloned into dephosphorylated Hind III/EcoR I digested, M13mp18 (for monomers and dimers) and M13mp19 (for dimers only). These clones were sequenced by the dideoxy method. Both orientations of wild-type monomers but only the m/m, m^{Δ}/m^{Δ} and m^{Δ}/m dimeric clones were obtained. Rather than search through a large number of clones for the m/m^{Δ} clone an alternative approach was used. This approach was

made possible by the generation of a unique Sty I site, at the junction of the two BstN I monomers, resulting from the end-filling of BstN I ends during the cloning procedure. m/m and m^{Δ}/m^{Δ} clones were digested with Hind III and Sty I and the purified wild-type monomer (arising from the m/m clone) was ligated into the purified, dephosphorylated larger fragment from the m^{Δ}/m^{Δ} digest (which contained the mutant monomer) so that a m/m^{Δ} clone was formed (Fig. 5.3). The insert from this clone was subcloned into bacteriophage M13 DNA and sequenced as described above.

5.2.3 Preparation of monomeric and dimeric Hph I ASBV clones

A wild-type Hph I monomeric clone was prepared by excising the insert from the pSP64 dimeric ASBV clone (see Chapter 4) with Hind III and Sma I. This insert containing fragment was purified by agarose gel electrophoresis, digested with Hph I and the monomer-sized fragment purified by agarose gel electrophoresis. Preparation of Hph I monomers was done in this way because digestion of the pSP64 vector with Hph I yields a number of fragments similar in size to the required monomer. The monomer-sized fragment was ligated to form dimers under conditions similar to those used previously (5.2.2). To allow ligation into the Sma I site of pGem-I the 3'-overhang resulting from Hph I digestion was removed with T4 DNA polymerase. The resultant mixture was ligated into dephosphorylated Sma I digested, pGem-I vector and transformed into competent MC1061 cells. Plasmid DNAs, from ampicillin resistant colonies, were prepared by the small scale method and the size of the inserts assayed by agarose gel electrophoresis of Hind III/EcoR I digests of these clones. A monomeric and a dimeric clone (used in Chapter 6) were selected and sequenced by the dideoxy method after subcloning of their inserts into M13mp18 and M13mp19, Hind III/EcoR I digested vectors.

Fig. 5.3 Diagram of the construction of a dimeric BstN ASBV clone (m/m^{Δ}) mutant in the second monomer. Hatched boxes, mutant ASBV sequences; open boxes, wild-type ASBV sequences; closed boxes, SP6 and T7 RNA polymerase promoters.


wildtype dimeric ASBV pGEM-1
mutant dimeric ASBV pGEM-1
Sty
Hind
Eco
p
p
m/m
m△/m△
Hind III/Sty I
agarose gel
Hind III/Sty I
agarose gel
phosphatase
Hind
Sty
Hind
Sty
Eco
p
p
ligase
Sty
Hind
Eco
p
p
m/m△

5.3 RESULTS

5.3.1 Cleavage of monomeric and dimeric minus ASBV RNAs

As mutant dimeric ASBV clones had proved to be effective in resolving the structure required for self-cleavage in the plus sense RNA a similar approach was used to investigate the minus ASBV RNA. It was not possible to use the dimeric Sau (-) ASBV pSP65 clone of Hutchins *et al.* (1986) as the two self-cleavage sites in transcripts produced from this template cleave to different extents (Hutchins *et al.*, 1986; and Fig. 5.4). This template and its expected transcription products are shown in Fig. 5.5a. This clone contains the same Sau3A I dimeric cDNA insert as that in the dimeric Sau (+) ASBV clone in pSP64 (see Chapter 4) but it was inserted into the pSP65 vector so that transcription with SP6 RNA polymerase produces a minus-sense RNA transcript. The 3'-proximal self-cleavage site of this transcript cleaves poorly in comparison to the 5'-proximal site as indicated by the large amount of the M/3'E fragment (430 nucleotides) (Fig. 5.4, lane 2). In contrast, the dimeric Sau (+) ASBV clone in the pSP64 vector (Hutchins *et al.*, 1986) yields a plus-sense RNA transcript that cleaves equally well at both sites during transcription (Fig. 5.4, lane 4). It was considered necessary to construct a clone which produced minus-sense transcripts capable of cleaving equally well at both sites to ensure clearly interpretable results. It was thought that whatever was interfering with self-cleavage at the 3'-proximal site was probably due to the location of the cloning site. Thus probable causes for the lack of cleavage at the 3'-proximal site were either that the vector sequences were interfering by preventing formation of the correct secondary structure or that the site of cloning of cDNA insert was dictating an order of sequence synthesis that was unsuitable for cleavage to occur equally well at both sites. Hence clones were prepared by ligating two blunted BstN I monomeric fragments into the Sma I site of pGem-I vector (see 5.2.2) to yield a clone that produced a minus-sense RNA upon transcription with SP6 RNA polymerase. Transcripts of this template cleaved

Fig. 5.4 Autoradiograph of 5% polyacrylamide, 7M urea gel showing RNA transcripts of wild-type dimeric ASBV templates. Lane 1; ^{32}P-labelled Hpa II digested pUC19 DNA markers with the sizes (in nucleotides) indicated. Lane 2; Minus RNA transcripts of dimeric Sau (-) ASBV pSP65 template (Hutchins *et al.*, 1986; Fig. 5.5a). Lane 3; Minus RNA transcripts of dimeric BstN ASBV pGem-1 clone (Fig. 5.6). Lane 4; Plus RNA transcripts of dimeric Sau (+) ASBV pSP64 clone (Hutchins *et al.*, 1986; Fig. 4.13a). The complete transcripts 5'E/M/3'E and the resultant self-cleavage fragments are marked.

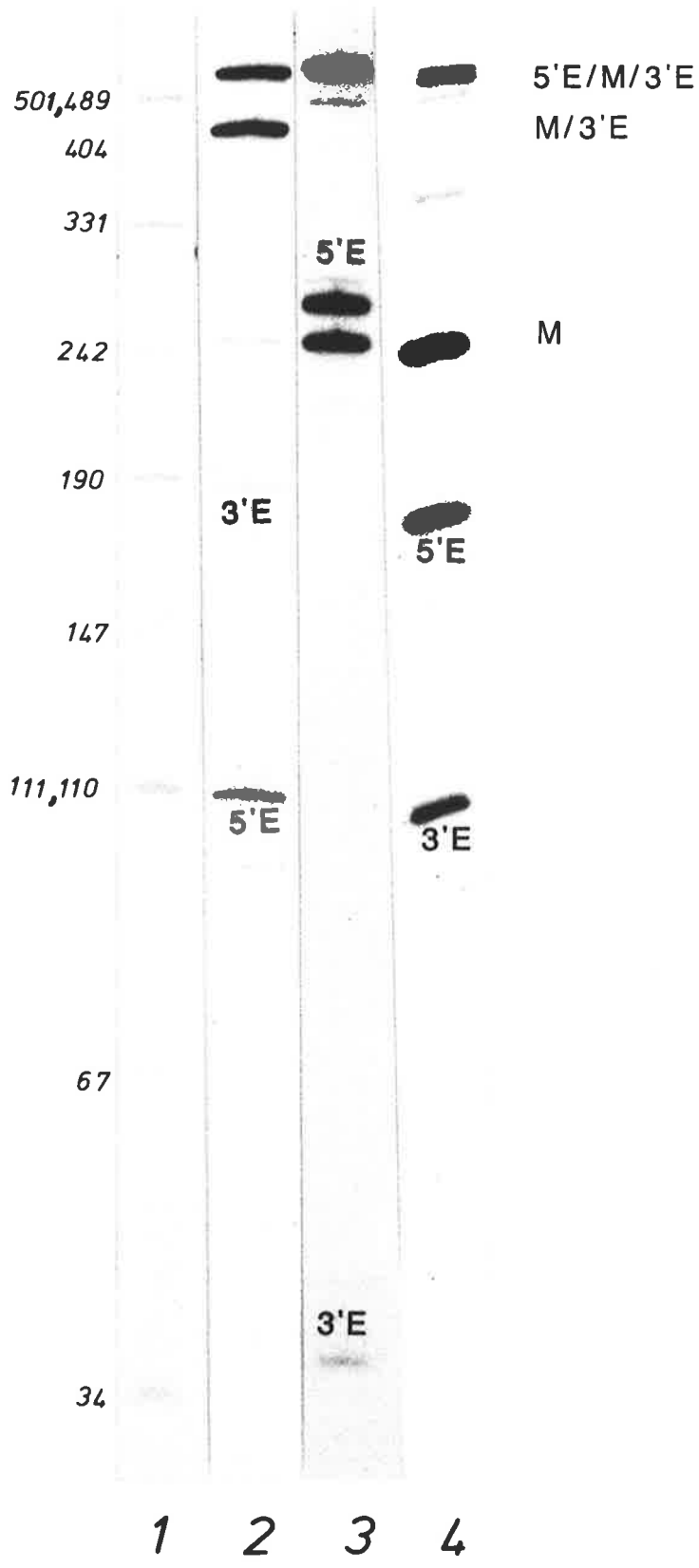
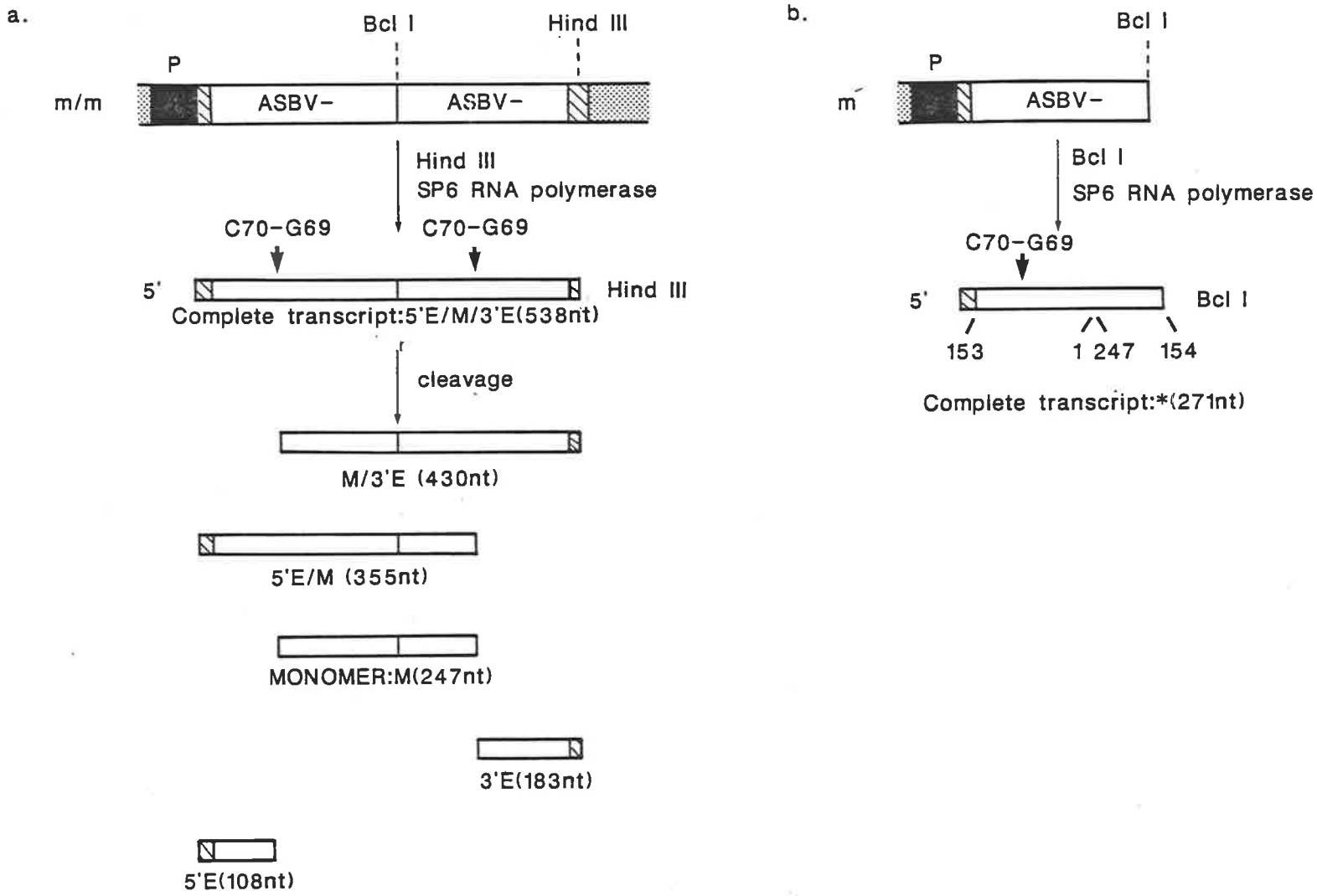

501,489
404
331
242
190
147
111,110
67
34
5'E/M/3'E
M/3'E
5'E
M
3'E
5'E
5'E
3'E
3'E
1
2
3
4

Fig. 5.5 Diagram of wild-type Sau (-) ASBV pSP65 templates and the RNAs produced by SP6 RNA polymerase transcription. The dimeric clone was linearized with either Hind III (a; m/m) or Bcl I (b; m). Hatched boxes, M13mp93 derived sequences; stippled boxes, pSP65 vector sequences; open boxes, ASBV sequences; closed boxes, SP6 RNA polymerase promoter. The numbering of the plus ASBV sequence (Symons, 1981) is retained for the minus sequence. The self-cleavage sites are indicated by thick arrows.


a.
Bcl I
Hind III
P
m/m
ASBV-
ASBV-
Hind III
SP6 RNA polymerase
C70-G69
C70-G69
5'
Hind III
Complete transcript:5'E/M/3'E(538nt)
cleavage
M/3'E (430nt)
5'E/M (355nt)
MONOMER:M(247nt)
3'E(183nt)
5'E(108nt)
b.
Bcl I
P
m
ASBV-
Bcl I
SP6 RNA polymerase
C70-G69
5'
Bcl I
153
1 247
154
Complete transcript:*(271nt)

efficiently at both sites (Fig. 5.4, lane 3) and so BstN I ASBV clones were considered suitable for further experimentation.

Additional monomeric clones were also prepared. The various monomeric and dimeric templates and their respective full-length RNA transcripts are given in Figs. 5.6, 5.7. RNA transcripts of the BstN I and Hph I monomeric templates (Fig. 5.7a,b) cleave very poorly (Fig. 5.8, lanes 3, 5) despite containing a complete set of sequences required for self-cleavage (a self-cleavage domain). In addition monomeric RNAs produced from truncated dimeric templates, i.e. the dimeric Sau3A I clone linearized with Bcl I (Fig. 5.5b), and the dimeric BstN I clone linearized with Sty I (Fig. 5.6b), which in contrast to the monomeric clones mentioned above do not have extra, vector derived sequences, at their 3' ends, also cleave to a negligible extent (Fig. 5.8, lanes 4, 6). The fact that these monomeric RNAs have different start sites (relative to the ASBV sequence) and two out of the four RNAs have no extra sequences at their 3' ends provides evidence that the site of cloning and the presence of extra nucleotides are not the reasons for the lack of cleavage in these molecules. As shown for monomeric plus ASBV RNAs these results support the double-hammerhead model which predicts that cleavage of monomers occurs by a concentration dependent, bimolecular reaction.

5.3.2 Dimeric minus ASBV RNAs cleave by a double-hammerhead

Strong evidence in support of the double-hammerhead model was provided by transcribing dimeric BstN I templates. As in the case of the Sau3A I ASBV clones used in the previous chapter in the investigation of plus RNA self-cleavage, four clones were prepared, i.e. wild-type (m/m), mutant at both sites (m^{Δ}/m^{Δ}), mutant in one or other of the monomers (m^{Δ}/m and m/m^{Δ}). The mutation created by site-directed mutagenesis was such that in minus-sense RNA transcripts the conserved GAAAC sequence present in the minus self-cleavage structure was mutated to GAAC

Fig. 5.6 Diagram of wild-type BstN (-) ASBV pGem-1 templates and the RNAs produced by SP6 RNA polymerase transcription. The dimeric clone was linearized with either EcoR I (a; m/m) or Sty I (b; m). Hatched boxes, pGem-1 sequences; open boxes, ASBV sequences; closed boxes, SP6 and T7 RNA polymerase promoters. Cleavage sites are indicated by thick arrows. The fragment names and their size (in nucleotides) are given. The numbering of the plus ASBV sequence (Symons, 1981) is retained for the minus RNA sequence.

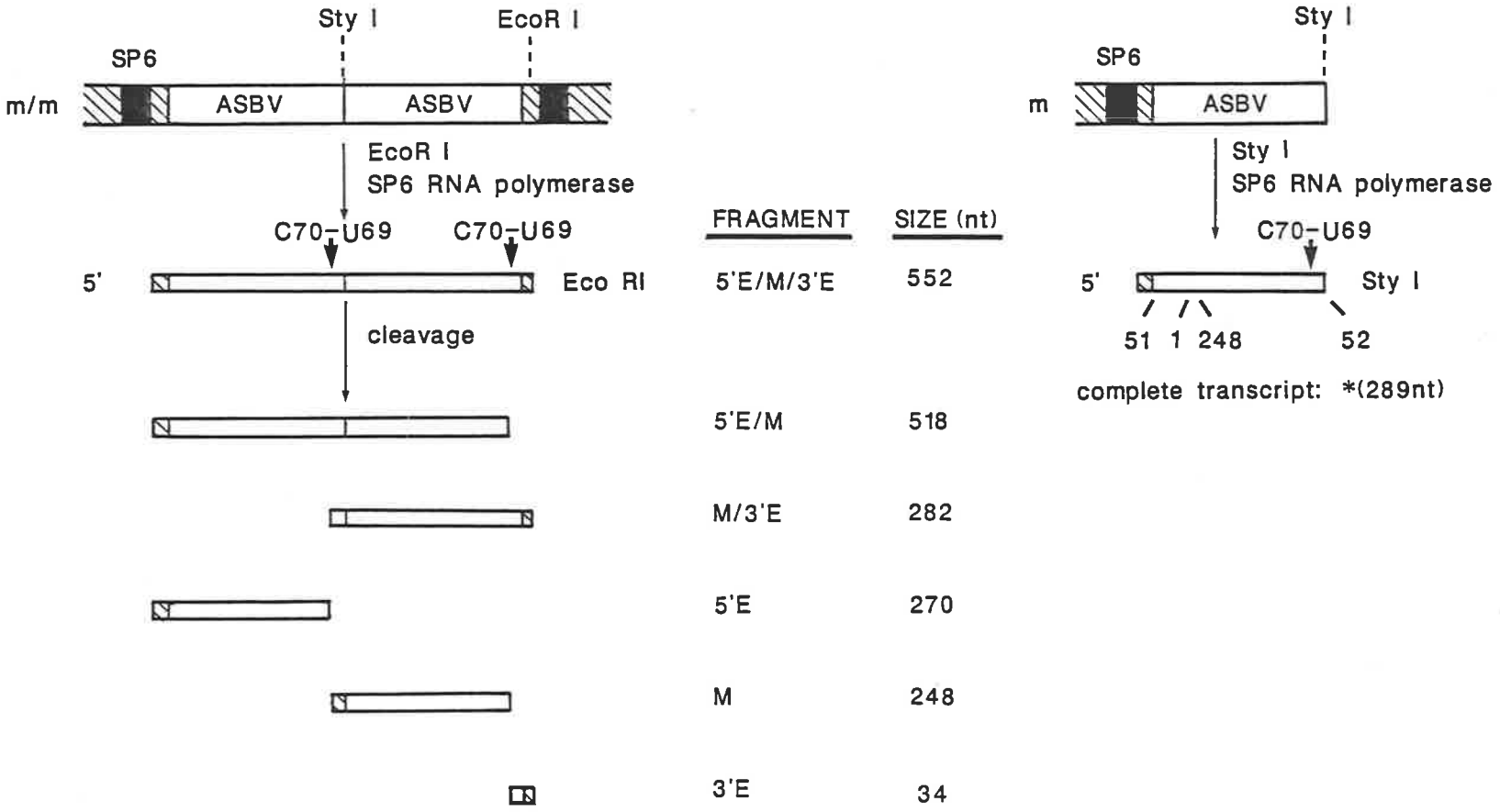
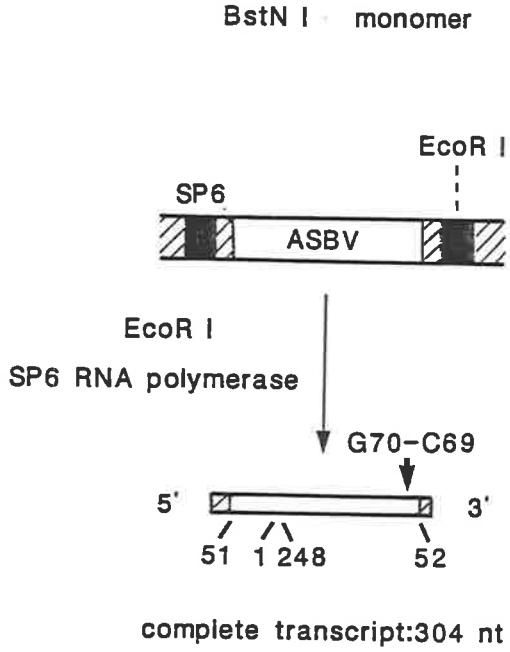

Sty I
EcoR I
SP6
m/m
ASBV
ASBV
EcoR I
SP6 RNA polymerase
C70–U69
C70–U69
5'
Eco RI
cleavage
FRAGMENT
SIZE (nt)
5'E/M/3'E
552
5'E/M
518
M/3'E
282
5'E
270
M
248
3'E
34
Sty I
SP6
m
ASBV
Sty I
SP6 RNA polymerase
C70–U69
5'
Sty I
51 1 248
52
complete transcript: *(289nt)

Fig. 5.7 Diagram of monomeric ASBV pGem-1 templates and their minus RNA transcripts. (a) monomeric BstN I ASBV template transcribed with SP6 RNA polymerase. (b) monomeric Hph I ASBV template transcribed with T7 RNA polymerase. Hatched boxes, pGem-1 sequences; open boxes, ASBV sequences; closed boxes, SP6 and T7 RNA polymerase promoters. Cleavage sites are indicated by thick arrows. The numbering of the plus ASBV sequences (Symons, 1981) is retained for the minus RNA sequence.


BstN I monomer
EcoR I
SP6
ASBV
EcoR I
SP6 RNA polymerase
G70-C69
5'
3'
51 1 248 52
complete transcript:304 nt

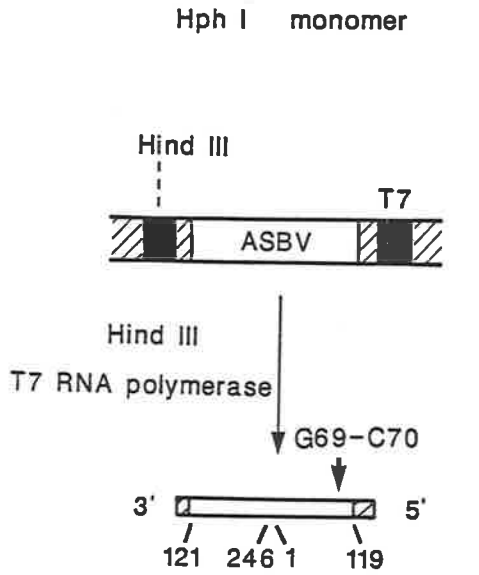
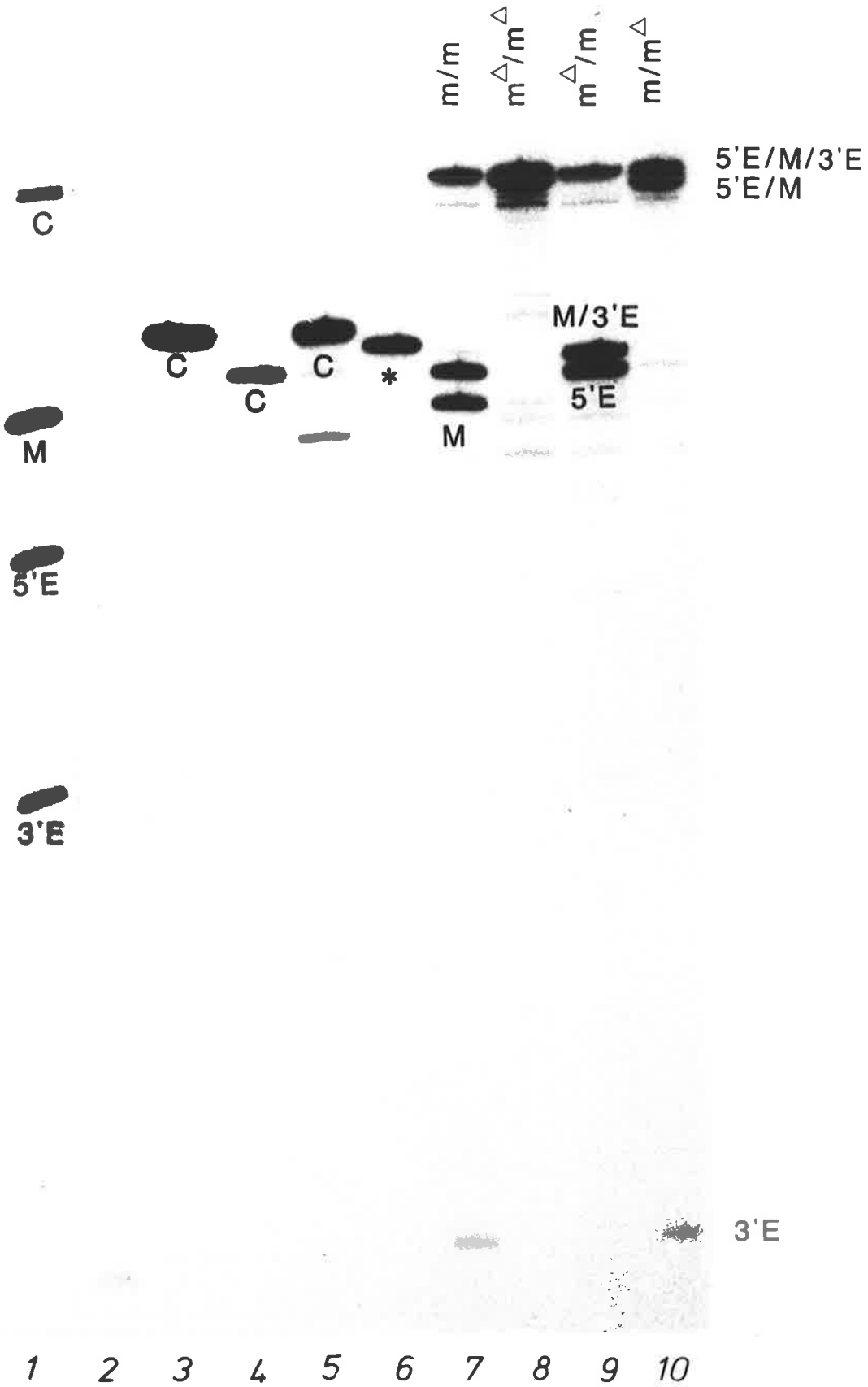

Hph I monomer
Hind III
T7
ASBV
Hind III
T7 RNA polymerase
G69-C70
3'
5'
121 246 1 119
complete transcript:305 nt

Fig. 5.8 Autoradiograph of 5% polyacrylamide, 7M urea gel showing the minus RNAs resulting from transcription of various monomeric and dimeric ASBV templates. Lane 1; Plus RNA transcripts of dimeric Sau ASBV pSP64 template linearized with Sma I (Fig. 4.13a). Lane 2; ^{32}P-labelled Hpa II digested marker DNAs (for band sizes see Fig. 5.4). Lane 3; Minus RNA transcripts of monomeric BstN ASBV template (Fig. 5.7a). Lane 4; Monomeric minus RNA transcripts of dimeric Sau (-) ASBV pSP65 template linearized with Bcl I (Fig. 5.5b). Lane 5; Minus RNA transcripts of monomeric Hph ASBV pGem-1 template (Fig. 5.7b). Lane 6; Minus RNA transcripts of dimeric BstN ASBV template truncated with Sty I (m; Fig. 5.6b). Lane 7; as for lane 6 but template linearized with EcoR I (m,/m; Fig. 5.6a). Lane 8; as for lane 7 but both template GAAAC sequences (A and B) (see Figs. 5.1a, 5.9) mutated to GAAC (m^{Δ}/m^{Δ}). Lane 9; as for lane 7 but only GAAAC (A) mutated (m^{Δ}/m). Lane 10; As for lane 7 but only GAAAC (B) mutated (m/m^{Δ}). Labels down margins represent the various RNA fragments. The complete transcripts in lane 1, 3, 4 and 5 are labelled C; the complete transcripts in lanes 6 and lanes 7-10 are labelled * and 5'E/M/3'E respectively. Other fragments arising from self-cleavage are as given in Fig. 5.6.

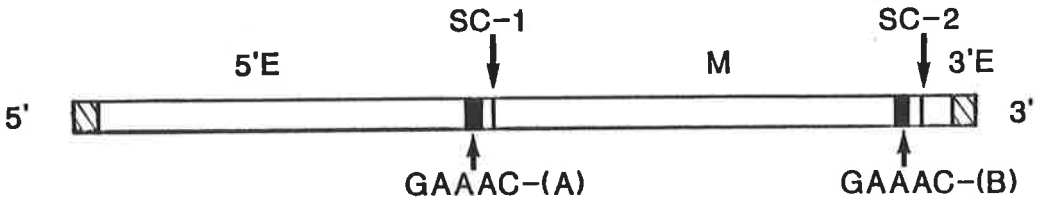

m/m
m△/m△
m△/m
m/m△
5'E/M/3'E
5'E/M
C
C
C
C
*
M/3'E
5'E
M
M
5'E
3'E
3'E
1
2
3
4
5
6
7
8
9
10

(Fig. 5.1). Mutation of the GAAAC sequence to GAAC had proved effective in the case of the plus-sense ASBV RNAs (see Chapter 4). The folded dimeric minus ASBV RNA, containing the proposed double-hammerhead, is schematically represented in Fig. 5.9b; for comparison, this RNA folded to contain two single-hammerhead structures is given in Fig. 5.9c. The complete minus-sense transcript (5'E/M/3'E, 552 nucleotides) of the dimeric BstN I clone wild-type for both GAAAC sequences, cleaves at both sites to produce monomer (M, 248 nucleotides; the extra nucleotide is due to end-filling of the BstN I ends during the cloning procedure) and the 5'- and 3'-end fragments (5'E, 270 nucleotides; 3'E, 34 nucleotides) (Fig. 5.8, lane 7). Mutation of both GAAAC sequences results in no cleavage at either site (Fig. 5.8, lane 8) demonstrating that, as in plus ASBV RNA, this sequence is important for self-cleavage. When only the 5'-proximal GAAAC (labelled A in Fig. 5.9) was mutated, cleavage at self-cleavage site two (SC-2 in Fig. 5.9) was prevented whilst cleavage at self-cleavage site one (SC-1), which is spatially removed from the mutated GAAAC in the double-hammerhead model (but not in the single-hammerhead model) was not affected as indicated by the presence of only the 5'E and M/3'E (282 nucleotide) fragments (Fig. 5.8, lane 9). The reverse results would be predicted from the single-hammerhead model (compare Figs. 5.9b and c). When the 3'-proximal GAAAC sequence (labelled B in Fig. 5.9) was mutated, cleavage at SC-1 but not at SC-2 was abolished, as shown by the presence of only the 5'E/M (517 nucleotides) and 3'E fragments (Fig. 5.8, lane 10).

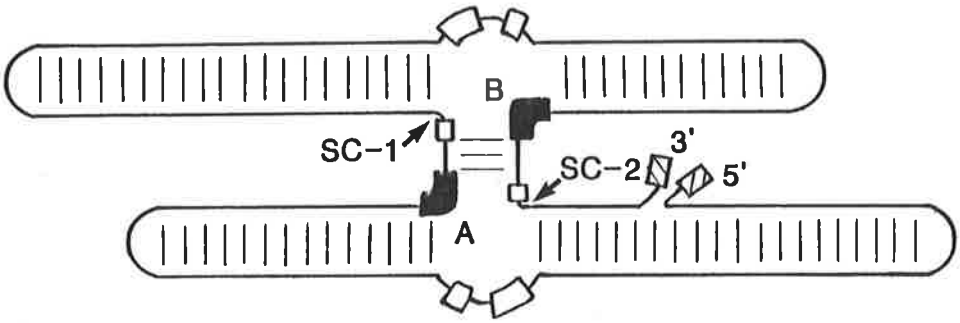
These results provide strong evidence in favour of the double-hammerhead model. The complete dimeric minus ASBV RNA sequence folded to contain the double-hammerhead structure is given in Fig. 5.10.

Fig. 5.9 Schematic representation of dimeric minus ASBV RNA transcript unfolded and folded in single- and double-hammerhead conformations. (a), Unfolded minus ASBV RNA transcript of dimeric BstN ASBV pGem-1 template (Fig. 5.6a). Self-cleavage sites, labelled SC-1 and SC-2, are indicated by arrows; hatched boxes, vector sequences; closed boxes, GAAAC sequences (see Fig. 5.1a) labelled A and B, indicated by arrows. (b), Schematic diagram of dimeric minus ASBV RNA drawn containing a double-hammerhead. Labelled as for Fig. 5.9a except for open boxes representing conserved nucleotides (see Fig. 5.1a); base-pairing is represented by lines between RNA strands. The numbering for the plus ASBV RNA (Symons, 1981) is retained for the minus ASBV RNA. (c) Schematic diagram of dimeric minus ASBV RNA drawn to contain two single-hammerheads. Labelled as in Fig. 5.9b.

a



b



c

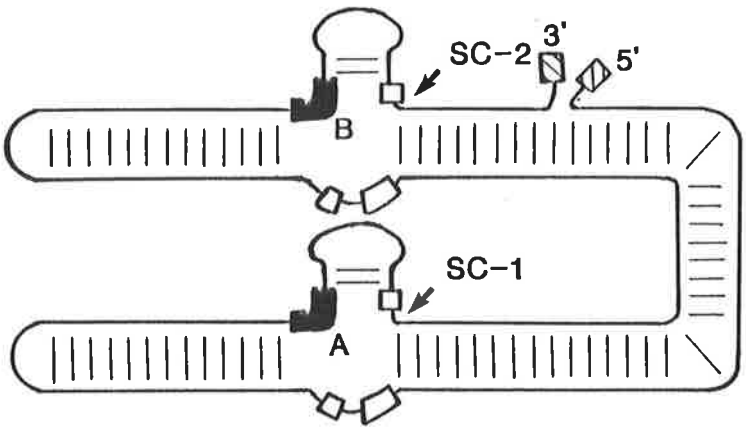
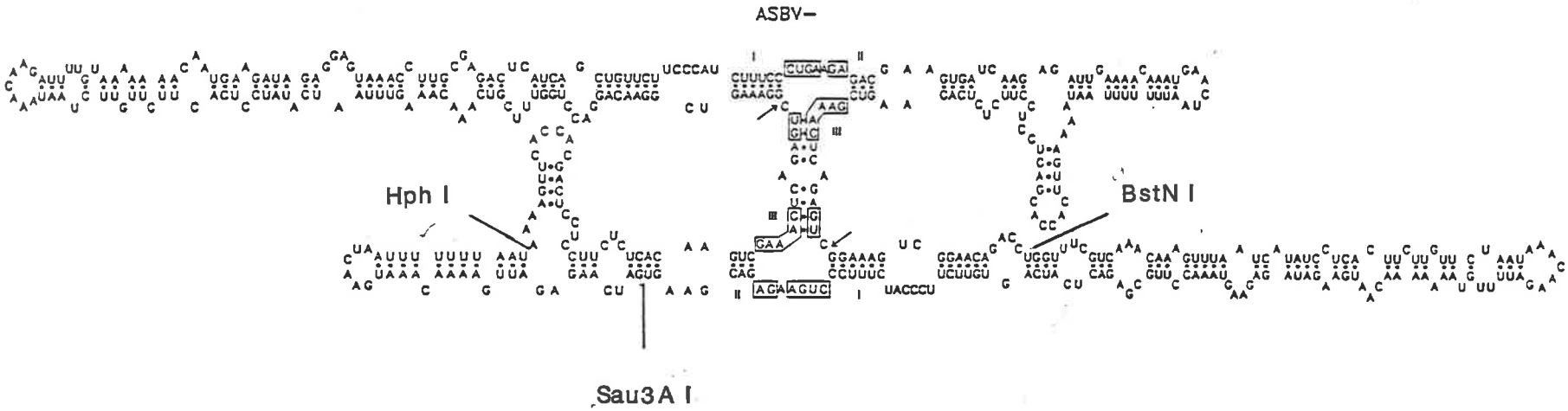


Fig. 5.10 Predicted secondary structure of dimeric minus ASBV RNA folded to contain a double-hammerhead. The self-cleavage sites are indicated by thick arrows; boxed nucleotides, conserved sequences (see Fig. 5.1a); stems are numbered 1 to 111. Various restriction endonuclease cleavage sites are marked. The numbering for plus ASBV (Symons, 1981) is retained for minus ASBV RNA.


ASBV–
Hph I
BstN I
Sau3A I

5.4 DISCUSSION

5.4.1 Monomeric and mutant dimeric RNAs demonstrate that minus ASBV RNA self- cleavage occurs by a double-hammerhead structure

As was demonstrated in the previous chapter for the monomeric plus ASBV RNA produced by transcription of a dimeric Sau3A I template truncated with Bcl I (Fig. 4.14, lane 1), the monomeric minus ASBV RNAs prepared in this chapter also did not cleave to any significant extent under similar conditions (Fig. 5.8). The variety of monomeric RNAs tested for self-cleavage (i.e. with or without vector derived sequences at the 3'-end and beginning at different positions on the ASBV RNA) argues against the non-cleavage of these RNAs being a result of the manner in which they were produced but rather, suggests that the lack of cleavage is due to some property of the ASBV sequences themselves. This is in agreement with the double-hammerhead model, demonstrated in the previous chapter for plus ASBV RNAs. Even though the minus ASBV RNA when drawn as a single-hammerhead has a potentially stronger stem III than that found in plus ASBV, it too was shown likely to form a double-hammerhead self-cleavage structure by the use of dimeric clones in which the conserved GAAAC sequence was mutated. As with plus ASBV RNAs this mutation in minus ASBV abolished cleavage demonstrating that this sequence is conserved due to its essential function in self-cleavage.

5.4.2 A unified model for ASBV and newt RNA self-cleavage

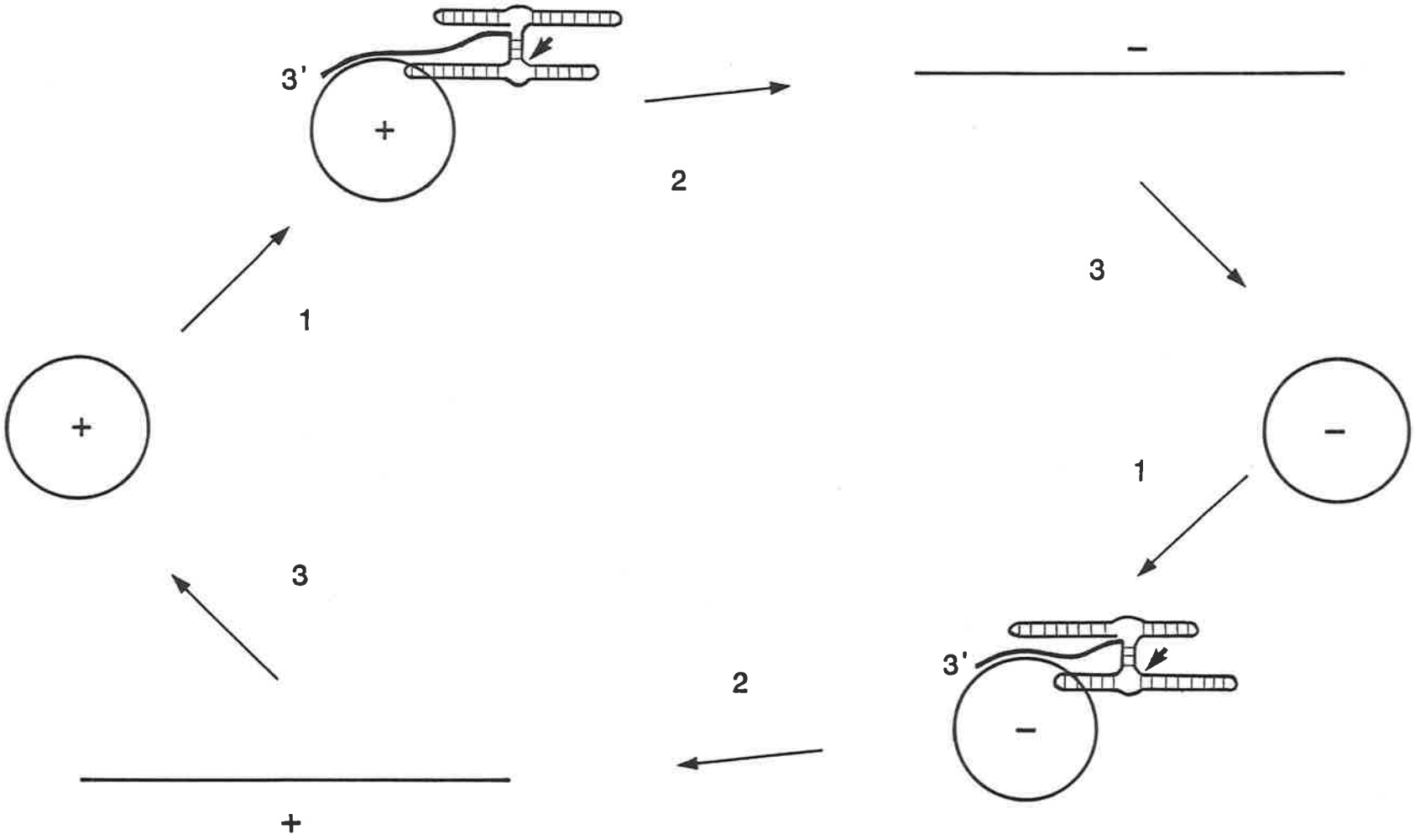
The demonstration that both plus and minus ASBV RNAs cleave by a double-hammerhead structure has unified the model for their self-cleavage and has indicated that a similar model can be assigned to the newt self-cleaving RNA. It was crucial that the double-hammerhead model was found valid for both plus and minus sequences *in vitro*. If the minus ASBV RNA which contains a stem III which is

theoretically more stable than that found in plus ASBV RNA was able to cleave to a significant extent by the single-hammerhead structure *in vitro* it would have made extrapolation of the *in vitro* data to the *in vivo* situation highly questionable. In this regard it should be remembered that the *in vivo* situation may be considerably more complex than that found *in vitro* due to the presence of other cell constituents (in particular nucleic acids and proteins) that may, for example, either aid or hinder secondary and tertiary structure formation.

5.4.3 *In vivo* implications of the double-hammerhead model for ASBV

Full-length RNAs, both monomers and dimers, can be considered to more closely resemble possible events *in vivo* than the short RNAs containing the bare minimum sequence required for self-cleavage used by some workers. The ASBV sequences outside those required purely for self-cleavage may play a role *in vivo* as has been suggested for sLTSV (Forster and Symons, 1987a). The double-hammerhead model provides a potential mechanism for the prevention of futile self-cleavage of circularized monomeric RNAs. ASBV is proposed to replicate by a rolling-circle model in which circular monomeric RNAs, formed by the self-ligation of monomers generated by self-cleavage, act as templates for transcription (Fig 5.11). The replication cycle presented in Fig. 5.11 is a modification of one previously proposed (Forster and Symons, 1987a). Although separated in this diagram, it would appear that transcription and self-cleavage are tightly coupled *in vivo*. Self-cleavage of both plus and minus RNAs is shown as occurring by double-hammerhead structures. The sites of initiation of transcription of ASBV RNAs *in vivo* are not known but after the first self-cleavage, folding of the RNAs must occur from the site of cleavage. This is the folding pattern displayed in this figure. It is obviously detrimental to have a cycle in which monomers produced by self-cleavage are continuously circularized and linearized. The double-hammerhead model may prevent such cycling by requiring the association of two RNA molecules to effect

Fig. 5.11 Proposed rolling-circle replication cycle for ASBV RNAs. The cycle consists of three, twice repeated events. 1; RNA transcription from a circular template, 2; self-cleavage of greater-than-unit-length RNAs by double-hammerhead structures and, 3; circularization of the cleaved monomers. After cleavage at the sites indicated by the thick arrows, folding begins again at these points to form the subsequent double-hammerhead structure.


3'
+
2
−
3
−
1
3'
−
2
+
3
+
1

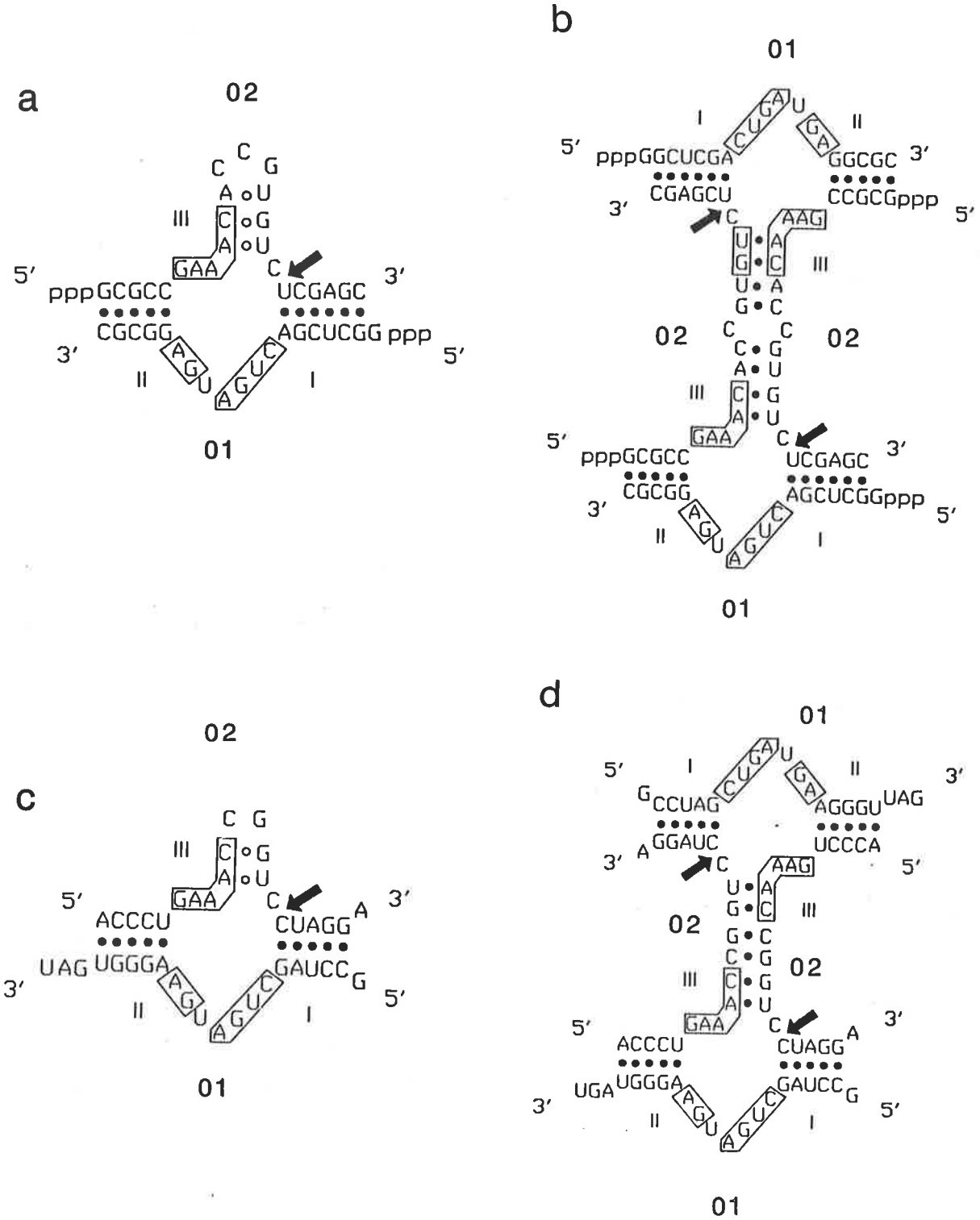
cleavage. Of the RNAs that cleave by some form of hammerhead structure, ASBV RNAs are peculiar in that unlike the others they are not encapsidated and the 2'-phosphomonoester group (which would prevent self-cleavage) resulting from the ligation of self-cleaved RNAs is apparently removed *in vivo*, making them vulnerable to futile cycling. Apart from the proposed role in the prevention of futile cycling in ASBV the double-hammerhead model provides a mechanism whereby the release of excised ASBV monomers can occur by the dissociation of the base-pairing in the two stem IIIs (Fig. 5.10).

5.4.4 The double-hammerhead model has important implications for other *in vitro* and *in vivo* studies

The double-hammerhead model may have important implications for the enzyme/substrate complex of Uhlenbeck (1987) which used minus ASBV-like sequences whereby a 19-mer (01) RNA was shown to be capable of catalytically cleaving a 24-mer (02) RNA, and for the similar system of Koizumi *et al.* (1988) (previously mentioned in the Chapter 4 Discussion) which used two 21-mer RNAs containing the newt self-cleavage sequences. These RNAs may undergo cleavage by a double-hammerhead structure formed by the association of four RNA molecules (Fig. 5.12b,d) rather than by the previously proposed single-hammerhead structures (Fig. 5.12c,d) composed of two RNA molecules. It is also possible that an active self-cleavage structure containing a stable stem III may be formed as an incomplete double-hammerhead by the association of only one catalytic RNA and two substrate RNAs (Fig. 5.12b,d), only one of which would be cleaved.

The double-hammerhead model is especially relevant to future efforts to elucidate the tertiary structure responsible for self-cleavage. Attempts to use short RNAs containing ASBV-like or newt-like sequences (i.e. those lacking stable stem IIIs) for x-ray crystallographic or NMR studies may be complicated by the

Fig. 5.12 Secondary structure models for the active self-cleavage structures of two short minus ASBV-like and newt RNAs. (a), Single-hammerhead model, proposed by Uhlenbeck (1987), for minus ASBV-like RNAs that together undergo self-cleavage *in vitro* (Uhlenbeck, 1987). (b), Double-hammerhead model for the RNAs described in (a). (c), Single-hammerhead model, proposed by Koizumi *et al.* (1988) for two self-cleaving RNAs. (d), Double-hammerhead model for the RNAs described in (c). Self-cleavage in the double-hammerhead structures (b,d) may be able to occur at one site by the association of one 01 and two 02 RNAs. Nucleotides conserved in all hammerhead self-cleavage structures (Fig. 4.5) are boxed; stems are numbered to 1 to 111; sites of self-cleavage are indicated by thick arrows.


a
02
01
5′
3′
pppGCGCC
CGCGG
UCGAGC
AGCUCGG ppp
GAA
b
01
02
02
01
5′
3′
pppGCGCC
CGCGG
UCGAGC
AGCUCGGpppp
GAA
c
02
01
5′
3′
ACCCU
UAG UGGGA
CUAGG A
GAUCC G
GAA
d
01
02
02
01
5′
3′
GCCUAG
GGAUC
AGGGU UAG
UCCCA
ACCCU
UGA UGGGA
CUAGG A
GAUCC G
GAA

requirement for intermolecular interaction and the presence of two self-cleavage sites per unit structure. Better substrates for such experiments may be structures, such as the sLTSV hammerhead, that contain stable stem IIIs (Fig. 4.3).

CHAPTER 6

THE IMPORTANCE OF STEM III IN ASBV RNA SELF-CLEAVAGE

6.1 INTRODUCTION

It was demonstrated in the previous two chapters that both plus and minus ASBV RNAs cleave by the association of two self-cleavage domains. In the double-hammerhead secondary structures proposed to be responsible for self-cleavage, stable stem IIIs (lacking in the single-hammerhead models for these RNAs (Fig. 4.1)) are produced by base-pairing between two self-cleavage domains (Fig. 4.4). The sequences that combine to form the stable stems are those sequences responsible for the weak stem IIIs of the single-hammerhead models. These results provided a consensus structure that accounted for all known self-cleaving RNAs of this type (Fig. 4.5). They did not however provide any direct data concerning the stem III sequences.

The aim of the work in this chapter was to determine what changes were required to the plus ASBV stem III sequences to allow self-cleavage by the single-hammerhead structure (Fig. 4.1a). If, as seemed the most probable explanation, the lack of self-cleavage of monomeric ASBV RNAs during transcription was due entirely to the stem III sequences it should be possible to alter these RNAs so that they cleave readily as single-hammerheads. To this end a number of plus ASBV RNAs were prepared that contained a range of alterations to the stem III sequences.

6.2 METHODS

6.2.1 Preparation of a monomeric Sau3A I ASBV clone mutant in the loop of stem III ("CAAG" mutant)

The insert from the wild-type dimeric BstN I ASBV clone previously prepared (5.2.2) was isolated by Hind III/EcoR I digestion and agarose gel electrophoresis. The purified insert was digested with Sau3A I and the monomeric sized band purified

by agarose gel electrophoresis. This DNA was then ligated into dephosphorylated BamH I digested pGem-2 vector and transformed into MC1061. DNA was prepared from ampicillin resistant colonies by the small scale method and assayed for the presence of inserts by restriction endonuclease mapping and by transcription with SP6 and T7 RNA polymerases. The inserts from clones which appeared to be as required were excised with Hind III/EcoR I, purified and ligated into dephosphorylated, Hind III/EcoR I digested M13mp18 and then sequenced by the dideoxy method.

6.2.2 Preparation of additional templates for transcription

Other clones used in this study for the preparation of templates for RNA polymerase transcription were prepared in Chapters 4 and 5. RNA transcripts different from those generated in these previous chapters were produced by using the alternative RNA polymerase promoter (either SP6 or T7) or by truncation of the template with different enzymes.

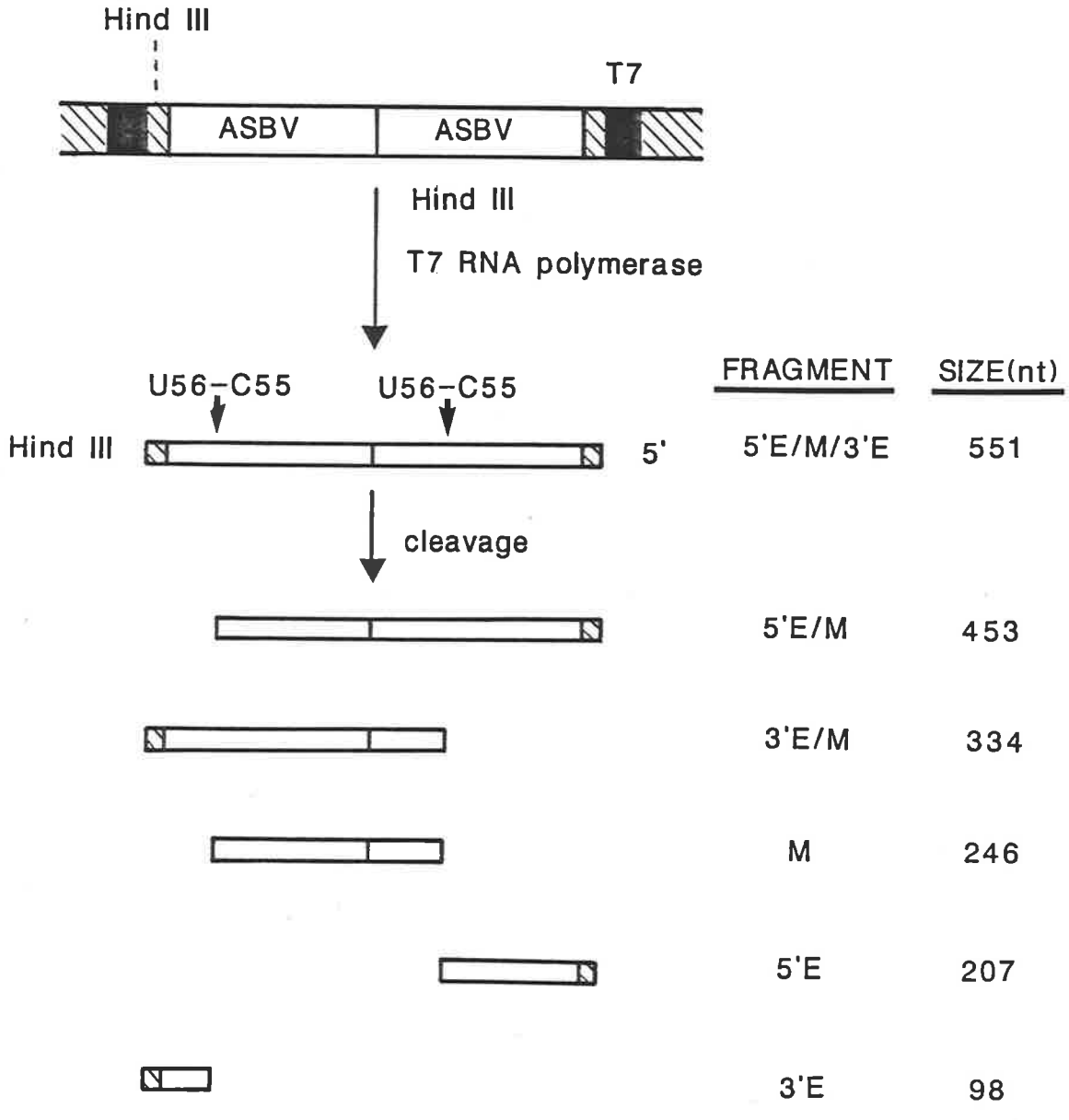
6.3 RESULTS

As an important corollary to the investigation of the double-hammerhead model using dimeric ASBV RNA transcripts a study was undertaken to demonstrate that it was indeed the stem III structure that is crucial to the self-cleavage of ASBV RNAs. If this was so then it should be possible, by altering the potentially unstable stem IIIs, by either altering the loop or by increasing the potential for base-pairing in the stem, to overcome the inhibition of self-cleavage in monomeric ASBV RNAs.

6.3.1 The site of cloning can influence self-cleavage of plus ASBV RNAs

Dimeric plus ASBV RNAs produced by transcription of a dimeric template cloned in the Hph I site (Fig. 6.1) cleaved well at both sites, as shown by the

Fig. 6.1 Diagram of the dimeric Hph ASBV pGem-1 template (cloned prepared in Chapter 5; 5.2.3) and the plus RNAs produced by T7 RNA polymerase transcription. Hatched boxes, vector sequences; open boxes, ASBV sequences; closed boxes, SP6 and T7 RNA polymerase promoters. The sites of self-cleavage are indicated by thick arrows. The names and sizes (in nucleotides) of the various self-cleavage fragments are given. The sequence is numbered after Symons (1981).


Hind III
T7
ASBV
ASBV
Hind III
T7 RNA polymerase
U56-C55
U56-C55
Hind III
5'
cleavage
FRAGMENT
SIZE(nt)
5'E/M/3'E
551
5'E/M
453
3'E/M
334
M
246
5'E
207
3'E
98

monomer (246 nucleotides) 5'E (207 nucleotides) and 3'E (98 nucleotides) fragments (Fig. 6.2, lane 2). Transcription of a monomeric clone in the Hph I site (Fig. 6.3a) produced a full-length RNA (302 nucleotides) that underwent negligible cleavage during transcription (Fig. 6.1, lane 5). These results mirror those obtained with Sma I and Bcl I linearized dimeric Sau3A I (+) ASBV templates (Fig. 6.2, lanes 1,4; Fig. 4.13). This is in agreement with the double-hammerhead model which predicts efficient cleavage of dimeric but not monomeric RNAs under transcription conditions. Interestingly a different outcome ensued when dimeric BstN I (+) ASBV templates were transcribed.

One of the reasons for choosing the BstN I site in ASBV as the cloning site for the production of dimeric ASBV clones for templates for minus RNA synthesis (see Chapter 5) was that it facilitated mutation of the stem III in full-length plus-sense clones. This is because the BstN I recognition site occurs in the stem III loop of the plus RNA self-cleavage sequence when drawn as a single-hammerhead (Fig. 6.4a). By transcribing the wild-type dimeric BstN I ASBV clone prepared in pGEM-1 vector (see Chapter 5) with T7 RNA polymerase instead of SP6 RNA polymerase it was possible to produce a plus-sense RNA transcript. This templates and the expected fragments arising from its transcription are given in Fig. 6.5a. This RNA contained two sets of sequences required for self-cleavage one of which, because of the end-filling of the BstN I site during preparation of the insert (see Methods, Chapter 5), contained a mutation in the loop in stem III. The end-filling meant that an extra A nucleotide is present in this loop (when the RNA is drawn as a single-hammerhead) (Fig. 6.4b). Also as a result of the cloning site the sequences required for self-cleavage were arranged in an unusual manner; the distribution of regions around the conserved sequences required for self-cleavage is shown in Fig. 6.6a. Upon transcription the dimeric BstN I ASBV template yielded a plus RNA that cleaved at both sites as shown by the presence of the expected fragments (Fig. 6.2, lane 3). Schematic representations of this DNA containing a double-hammerhead and

Fig. 6.2 Autoradiograph of 5% polyacrylamide, 7M urea gel showing plus RNA transcripts of various monomeric and dimeric ASBV templates. Lane 1; Plus RNA transcripts of dimeric Sau (+) ASBV pSP64 clone linearized with Sma I (m/m) (Fig. 4.13a). Lane 2; Plus RNA transcripts of dimeric Hph ASBV pGem-1 clone transcribed with T7 RNA polymerase (Fig. 6.1). Lane 3; Plus RNA transcripts of the dimeric BstN ASBV pGem-1 template (as in Fig. 5.6) linearized with Hind III and transcribed with T7 RNA polymerase. Lane 4; Plus RNA transcripts of dimeric Sau (+) pSP64 template truncated with Bcl I (Fig. 4.13b). Lane 5; Plus RNA transcripts of the monomeric Hph ASBV pGem-1 template (Fig. 6.3a) transcribed with SP6 RNA polymerase. Lane 6; Plus RNA transcripts of the monomeric BstN ASBV pGem-1 template (6.3b1) linearized with Hind III and transcribed with T7 RNA polymerase. Lane 7; As for lane 6 but ASBV insert in opposite orientation, therefore a plus RNA transcript was produced by linearization with EcoR I and transcription with SP6 RNA polymerase (Fig. 6.3b2). Lane 8; As for lane 3 but the dimeric template was truncated with Sty I to produce an RNA containing a monomeric plus ASBV sequence. Lane 9; Plus RNA transcripts of mutant monomeric Sau (+) ASBV template. In this template the CAG sequence in the stem III loop was mutated to CAAG (see Fig. 6.4b). Complete transcripts are labelled C, the 5' end (5'E), monomer (m) and 3'E end (3'E) fragments arising from self-cleavage of the complete transcripts are indicated. Lane 10; ^{32}P-labelled Hpa II digested pUC19 marker DNA fragments, sizes are indicated in nucleotides.

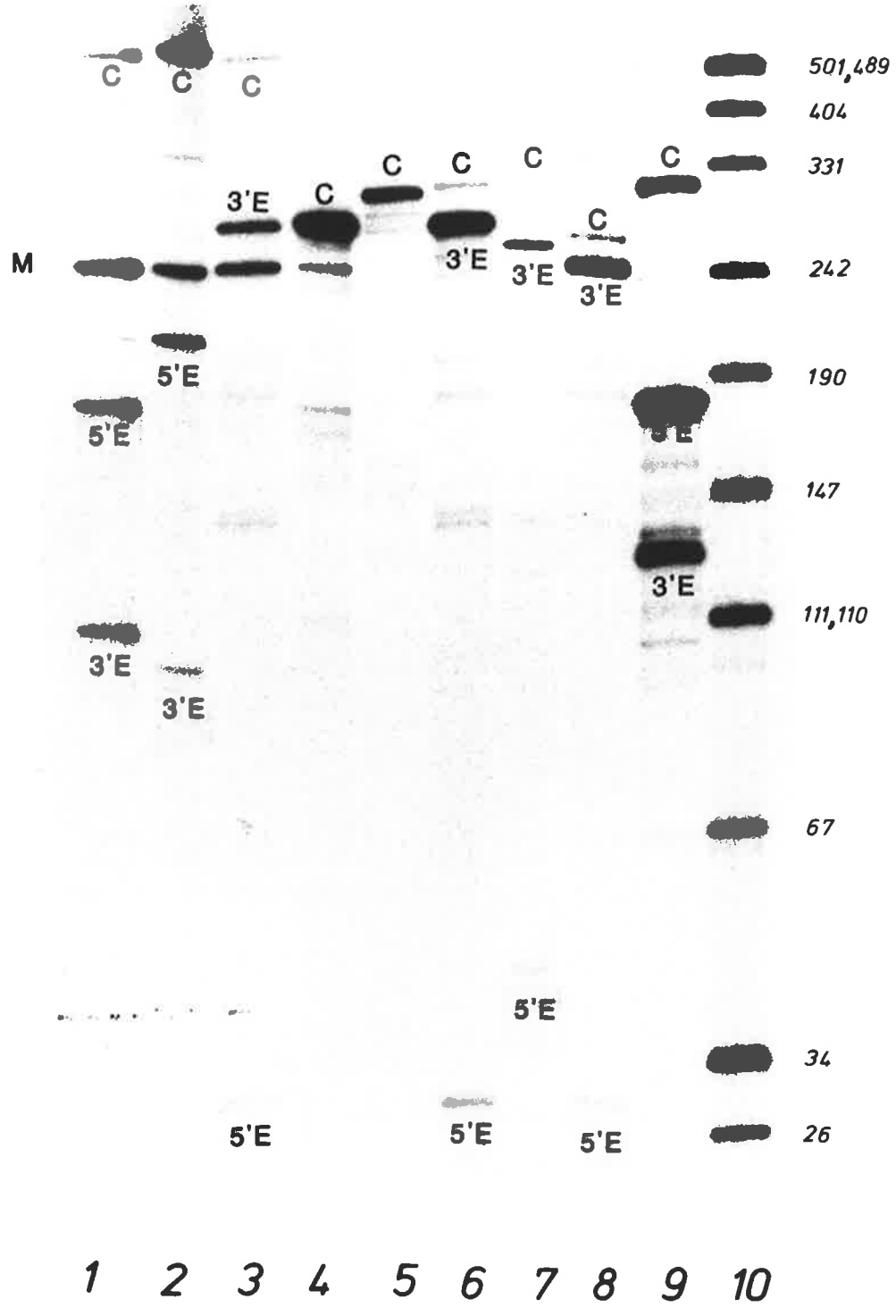
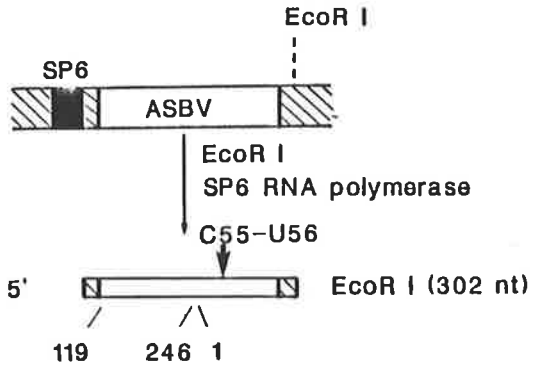
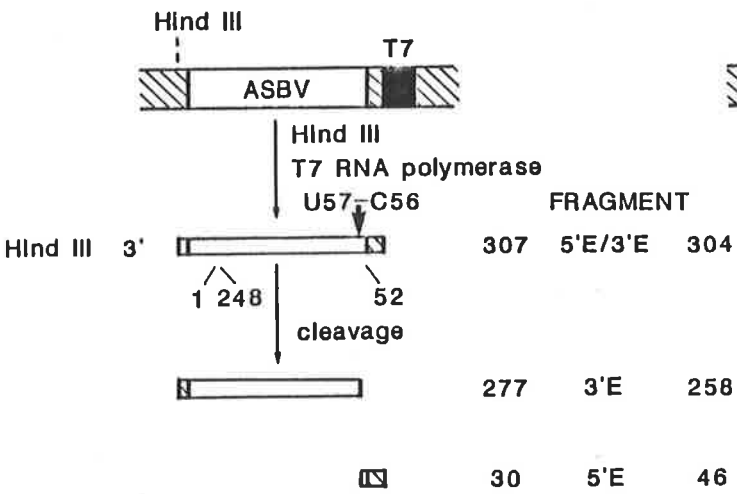

C
C
C
C
C
C
C
C
C
3'E
3'E
3'E
3'E
M
5'E
5'E
3'E
3'E
3'E
5'E
5'E
5'E
501,489
404
331
242
190
147
111,110
67
34
26
1
2
3
4
5
6
7
8
9
10

Fig. 6.3 Monomeric ASBV templates and the plus RNAs resulting from their transcription. (a), Monomeric Hph ASBV pGem-1 template and RNA produced by SP6 RNA polymerase transcription. (b1), Monomeric BstN (-) ASBV pGem-1 template and plus RNAs produced by T7 RNA polymerase transcription. (b2), Monomeric BstN (+) ASBV pGem-1 template and plus RNAs produced by SP6 RNA polymerase transcription. Hatched boxes, vector sequences; closed boxes, SP6 or T7 RNA polymerase promoters; open boxes, ASBV sequence. The self-cleavage sites are indicated by thick arrows. Numbering after Symons (1981).

a


EcoR I
SP6
ASBV
EcoR I
SP6 RNA polymerase
C55-U56
5'
EcoR I (302 nt)
119
246
1

b1


Hind III
T7
ASBV
Hind III
T7 RNA polymerase
U57-C56
FRAGMENT
Hind III
3'
307
5'E/3'E
304
1
248
52
cleavage
277
3'E
258
30
5'E
46

b2

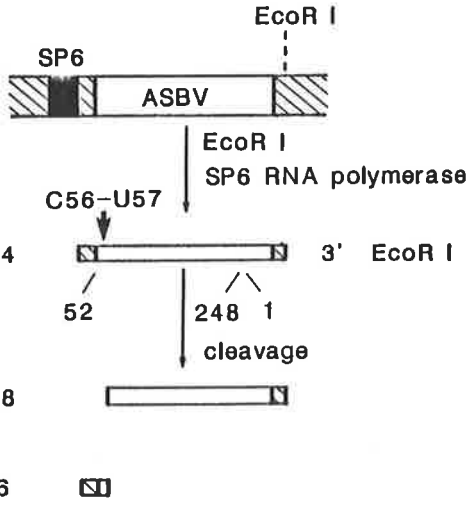

EcoR I
SP6
ASBV
EcoR I
SP6 RNA polymerase
C56-U57
3'
EcoR I
52
248
1
cleavage

Fig. 6.4 Single-hammerhead secondary structures for plus (Hutchins *et al.*, 1986), mutant plus and minus (Hutchins *et al.*, 1986) ASBV RNAs. (a), Plus ASBV RNA; note that the BstN I recognition sequence (CCAGG) is present in the stem III loop. (b), Mutant ASBV RNA; an extra A has been inserted into the stem III loop converting the BstN I site to a Sty I site (CCAAGG). (c), Minus ASBV RNA. Boxed residues are those conserved in RNAs that cleave by the hammerhead model (listed in Fig. 4.5). The self-cleavage sites are indicated by thick arrows.

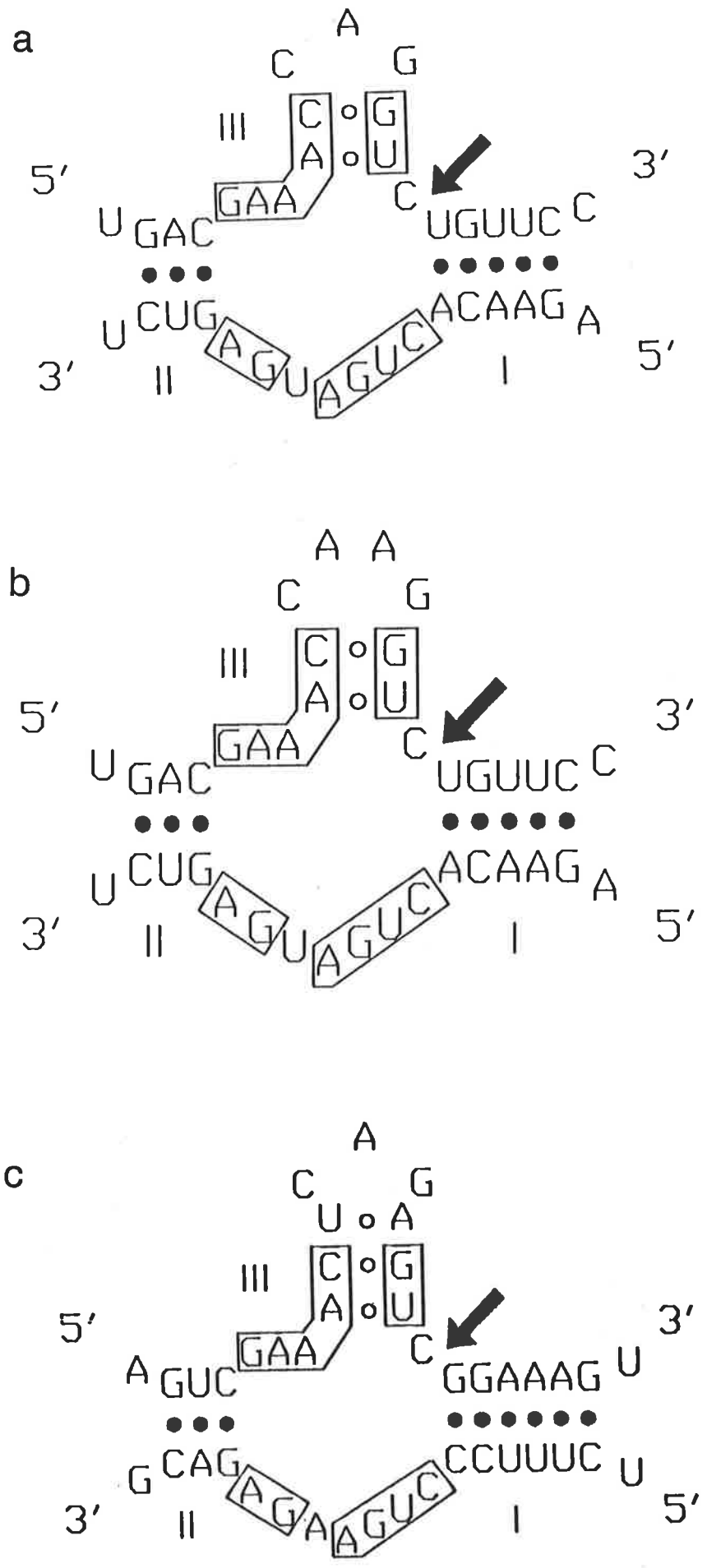

a
A
C
G
III
C
A
G
U
5′
3′
GAA
C
U GAC
UGUUC C
U CUG
ACAAG A
3′
II
A G U
A G U C
I
5′
b
A A
C
G
III
C
A
G
U
5′
3′
GAA
C
U GAC
UGUUC C
U CUG
ACAAG A
3′
II
A G U
A G U C
I
5′
c
A
C
G
U A
III
C
A
G
U
5′
3′
GAA
C
A GUC
GGAAAG U
G CAG
CCUUUC U
3′
II
A G A
A G U C
I
5′

Fig. 6.5 Diagram of dimeric BstN ASBV pGem-1 template linearized with Hind III (m/m, a) and Sty I (m, b) and the plus RNA transcripts produced by transcription with T7 RNA polymerase. Hatched boxes, vector sequences; open boxes, ASBV sequences; closed boxes, RNA polymerase promoter sequences. Self-cleavage sites are indicated by thick arrows. Numbering after Symons (1981). The names and sizes (in nucleotides) of the RNA fragments are given.

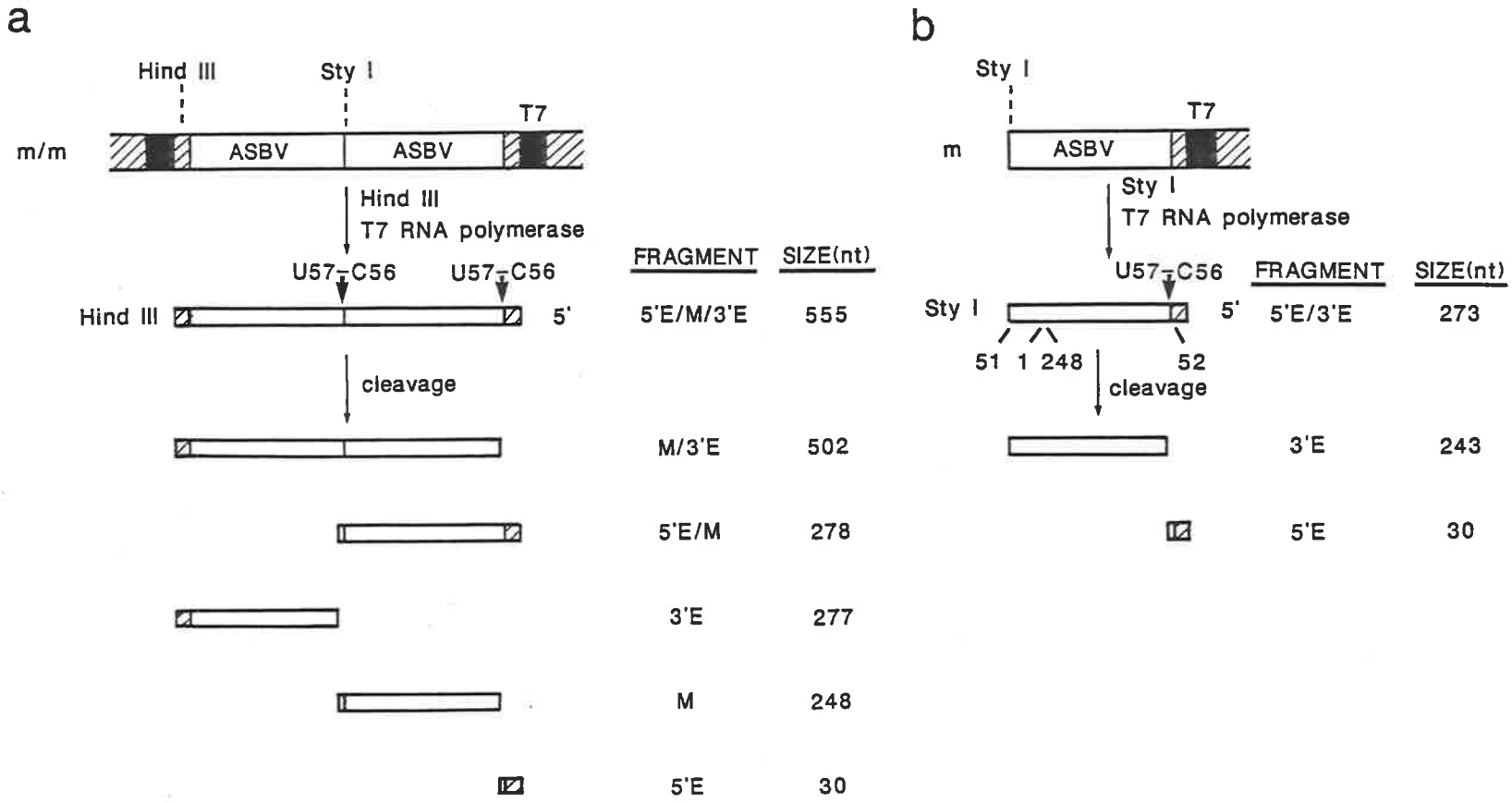
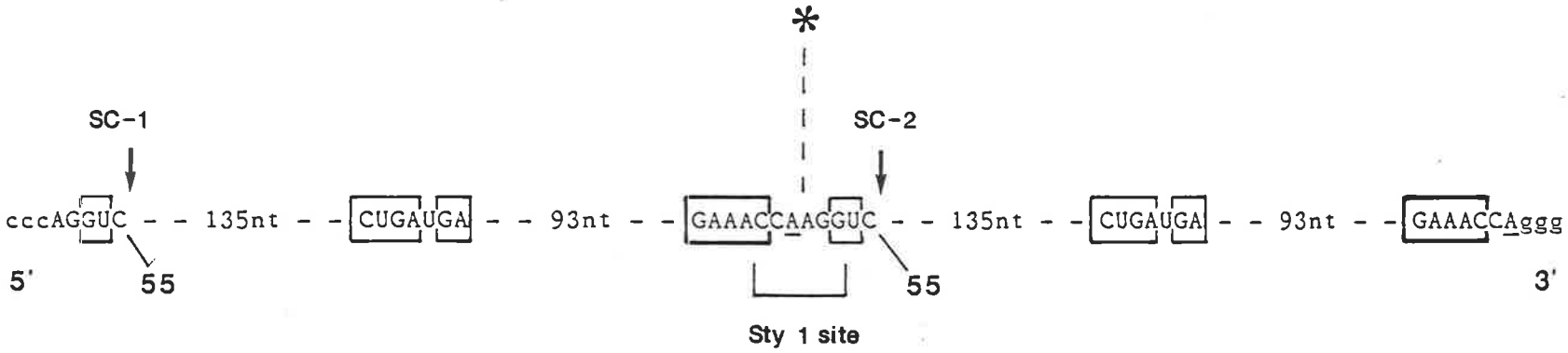
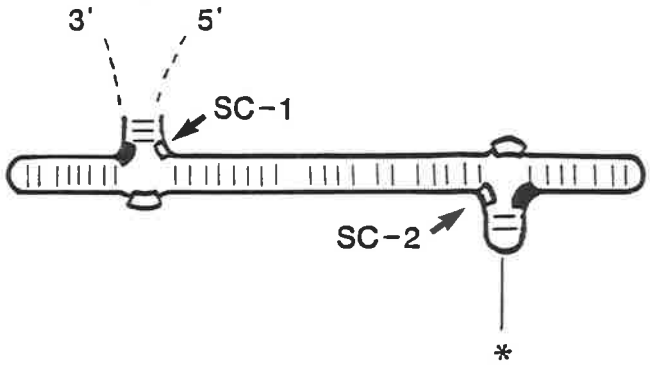
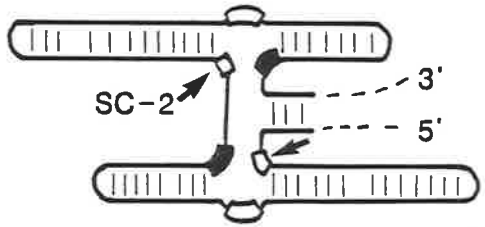

a
Hind III
Sty I
T7
m/m
ASBV
ASBV
Hind III
T7 RNA polymerase
U57–C56
U57–C56
Hind III
5'
cleavage
FRAGMENT
SIZE(nt)
5'E/M/3'E
555
M/3'E
502
5'E/M
278
3'E
277
M
248
5'E
30
b
Sty I
T7
m
ASBV
Sty I
T7 RNA polymerase
U57–C56
Sty I
5'
51 1 248
52
cleavage
FRAGMENT
SIZE(nt)
5'E/3'E
273
3'E
243
5'E
30

Fig. 6.6 The sequence and possible secondary structures of dimeric plus ASBV RNA. (a), The arrangement of sequences required for self-cleavage in dimeric BstN plus RNA. The boxed residues are those conserved between RNAs that cleave by the hammerhead model. To aid orientation the GAAAC sequence is boxed with heavier lines. Upper case letters, ASBV sequences; lower case letters, vector derived sequences. The extra A nucleotides resulting from the cloning procedure are underlined. Self-cleavage sites 1 and 2 are marked with thick arrows. The distances in nucleotides between the blocks of sequences are indicated. The junction between the two monomeric units which comprise the dimer is asterisked. Numbering after Symons (1981). (b), Secondary structure of the dimeric BstN plus RNA drawn to contain a double-hammerhead. The vector sequences are dashed; boxes, conserved sequences found in all RNAs that self-cleave by the hammerhead model; closed boxes, conserved GAAAC sequence. Self-cleavage sites 1 and 2 are indicated by thick arrows. (c), As above but the dimeric RNA folded to contain two single-hammerheads.


SC-1
SC-2
*
cccAGGUC - - 135nt - - CUGAUGA - - 93nt - - GAAACCAAGGUC - - 135nt - - CUGAUGA - - 93nt - - GAAACCAggg
5'
55
55
3'
Sty 1 site


3'
5'
SC-1
SC-2
*


SC-2
3'
5'

containing two single hammerheads are given in Fig. 6.6b,c. At first glance, cleavage at both sites in this RNA may seem surprising as one may expect the essential stem IIIs in the double-hammerhead structure to be disrupted by the "tails" of vector sequences. However the fact that the insert of this clone was blunt-ligated into the Sma I site of the vector is very important. It is possible to draw a double-hammerhead structure for this dimeric RNA (Figs. 6.6b, 6.7) which may cleave in the same fashion as that previously proposed for the Sau3A I (+) and BstN I (-) RNAs (Fig. 4.4b,c). This is because complementarity exists between the vector sequences surrounding the insert, due to the palindromic nature of the Sma I recognition site, which may allow an active double-hammerhead self-cleavage structure to form. It is also possible that the two self-cleavage domains may form single-hammerheads (Fig. 6.6c). The single-hammerhead structure formed by a coming together of the 5'- and 3'-end sequences may be expected to cleave efficiently as it contains a stem III that is theoretically quite stable due to the co-operation of the vector derived sequences described above (Fig. 6.8a). The other self-cleavage site in this dimeric RNA may be active because the extra A nucleotide in the CAG loop of stem III (Fig. 6.4b) may increase the stability of this stem sufficiently to allow cleavage. To determine whether cleavage by these two single-hammerhead self-cleavage structures was possible the dimeric clone was dissected to produce monomeric clones that contained one or other of the two structures shown in Figs. 6.4b, 6.8. A monomeric insert with blunted BstN I ends was ligated into the Sma I site of pGem-I (see Methods). As the insert was blunt-ended both orientation were selected. Transcription of these clones gave rise to plus RNAs that cleaved efficiently to yield 5'E and 3'E products (Fig. 6.2, lanes 6, 7) whose sizes agreed with those expected from the predicted cleavage site (Fig. 6.3b). Transcripts from clones in both orientations cleaved efficiently and thus the orientation of the vector derived sequences contributing to stem III did not appear to have a significant effect (Fig. 6.8).

Fig. 6.7 Double-hammerhead secondary structure for dimeric BstN plus ASBV RNA showing how this structure may form despite the presence of potentially disruptive vector sequences. Boxed residues are those conserved between all RNAs that cleave by the hammerhead model. Vector sequences are in lower case or represented by dashed lines. Parts of the sequence not detailed are represented by lines. Stems are numbered 1 to 111, self-cleavage sites are indicated with thick arrows.

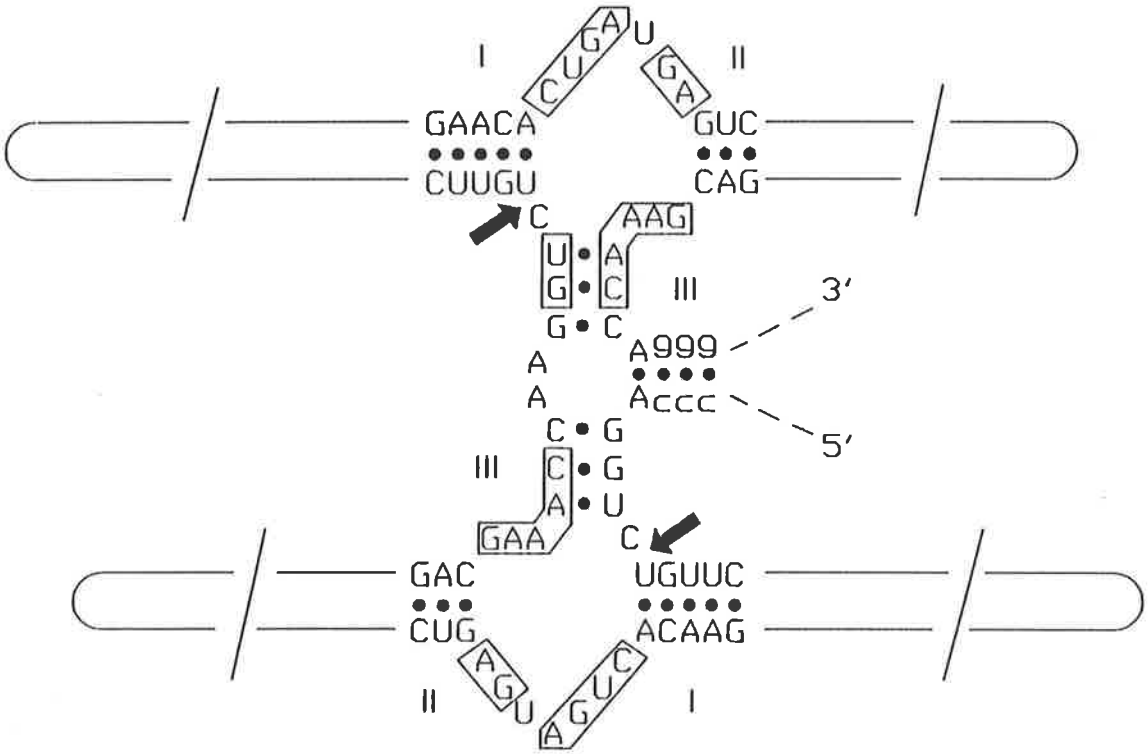
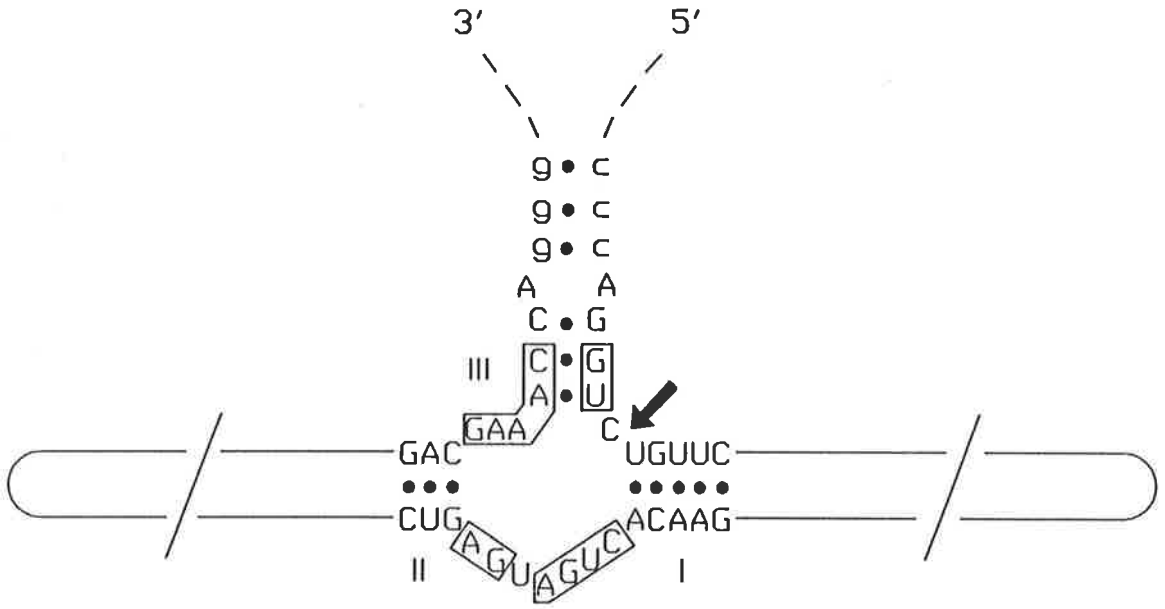
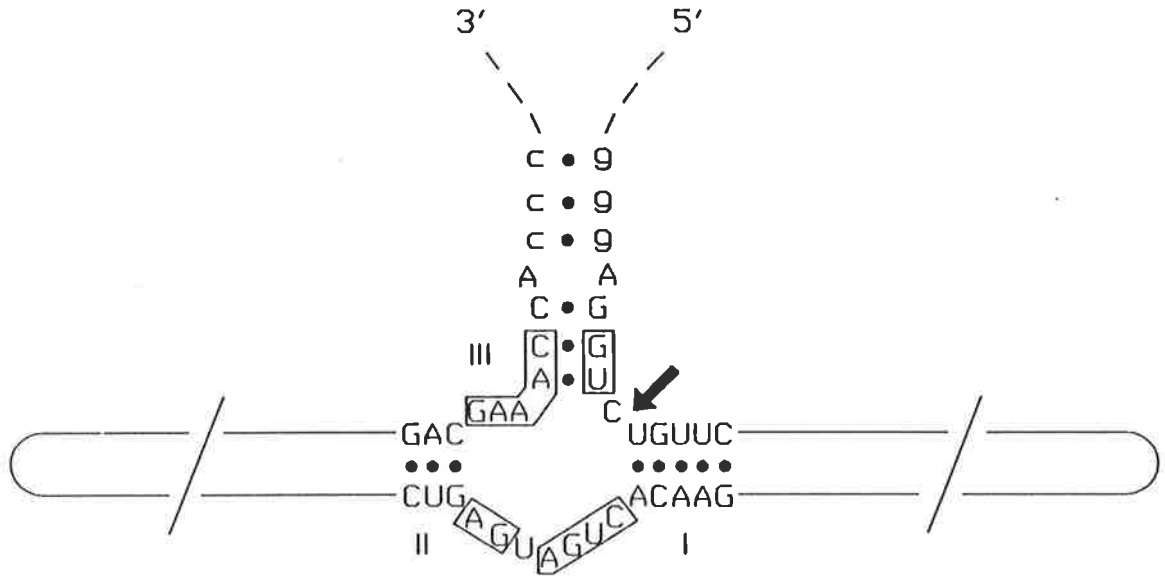

I
II
III
GAACA
CUUGU
GUC
CAG
AAG
3'
5'
A999
ACCC
GAA
GAC
CUG
UGUUC
ACAAG

Fig. 6.8 Single-hammerhead secondary structures for monomeric BstN plus ASBV RNAs derived from inserts in both orientations in pGem-1. (a), RNA transcript of the template shown in Fig. 6.3b2. (b), RNA transcript of the template shown in Fig. 6.3b1. Boxed nucleotides are those conserved between all RNAs that cleave by the hammerhead model (see Fig. 4.5 for details), stems are numbered 1 to 111, self-cleavage sites indicated by thick arrows, sequences not shown indicated by lines. The vector derived nucleotides that contribute to stem III are in lower case, the remaining vector sequences are dashed.

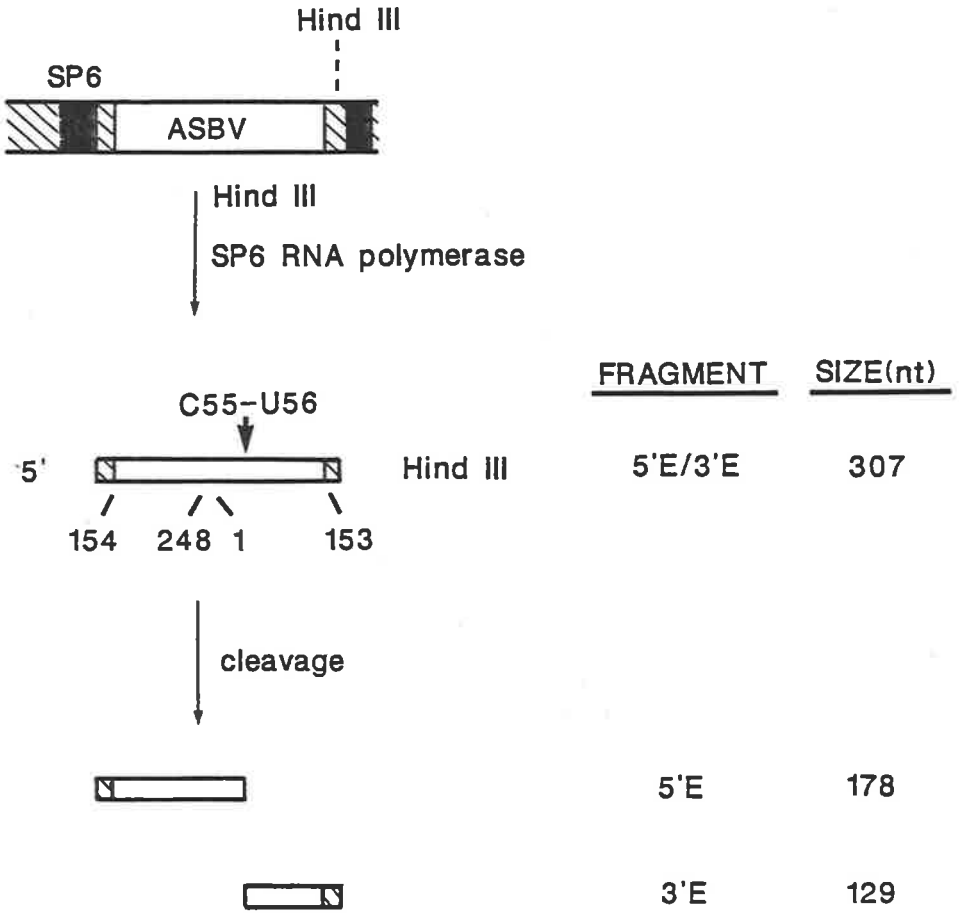

3′
5′
g • c
g • c
g • c
A A
C • G
III
C
A
G
U
GAA
C
GAC
UGUUC
CUG
ACAAG
II
A G U
A G U C
I


3′
5′
c • g
c • g
c • g
A A
C • G
III
C
A
G
U
GAA
C
GAC
UGUUC
CUG
ACAAG
II
A G U
A G U C
I

6.3.2 A single nucleotide insertion greatly effects monomeric plus ASBV RNA self-cleavage

By removing the Sau3A I fragment from the dimeric BstN I clone and inserting it into the BamH I site of pGem-2 vector a template was created that contained the CAAG sequence (an A insertion) in the plus ASTV stem III loop but, unlike the monomers described above, the sequences required for self-cleavage were not located at the ends of the insert, i.e. adjacent to the vector sequences. Thus the RNA produced by transcription of this clone after linearization with Hind III is the same as that from the dimeric Sau3A I (+) ASBV clone truncated with Bcl I, which cleaves to a negligible extent (Fig. 6.2, lane 4), except for the additional A nucleotide and the presence of vector sequences at the 3'-end. This template, its plus RNA transcript and the expected self-cleavage fragments are given in Fig. 6.9. The monomeric RNA containing the additional A residue in the stem III loop cleaved efficiently during transcription (Fig. 6.2, lane 9), a fact interpreted as indicative of cleavage by the single-hammerhead structure. The single-hammerhead structure believed to be responsible for this self-cleavage is given in Fig. 6.4b. The mutation in the stem III loop may be expected to increase the stability of stem III and thus increase the ability of this RNA to cleave by a single-hammerhead structure but should not increase the probability of self-cleavage by an intermolecular double-hammerhead structure as it does not increase the stability of this structure (Fig. 6.10). It has been demonstrated by other workers that transcripts of the wild-type monomeric Sau3A I (+) ASBV clone undergo significant levels of self-cleavage only under extreme conditions, i.e. high RNA concentration, high Mg^{2+} concentration, long incubation period and elevated temperature (P. Rathjen, pers. comm.).

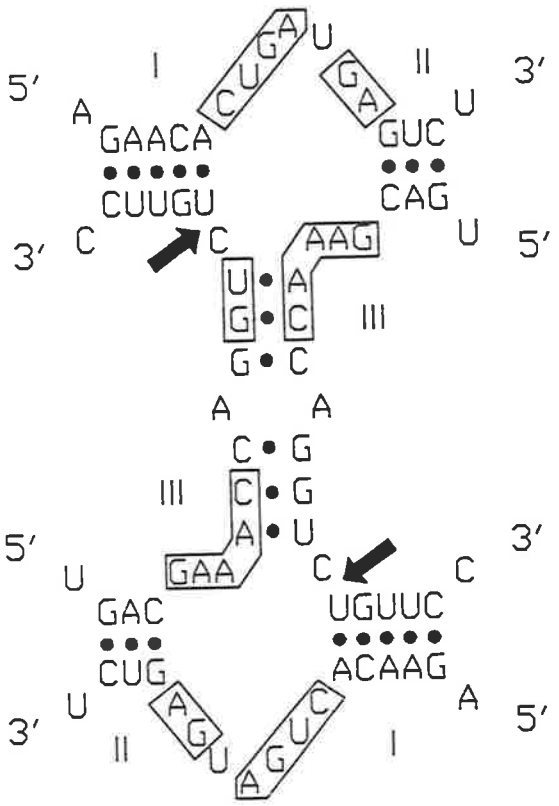
Fig. 6.9 Diagram of the mutant monomeric Sau ASBV pGem-2 template and the plus RNAs produced from it by SP6 RNA polymerase transcription. This template (see 6.2.1 for preparation) has an extra A nucleotide in the stem III loop sequence i.e. CAAG not CAG (Fig. 6.4). Hatched boxes, vector sequences; solid boxes, SP6 and T7 RNA polymerase promoters; open boxes, ASBV sequences. Self-cleavage site is indicated by a thick arrow. Numbering after Symons (1981).



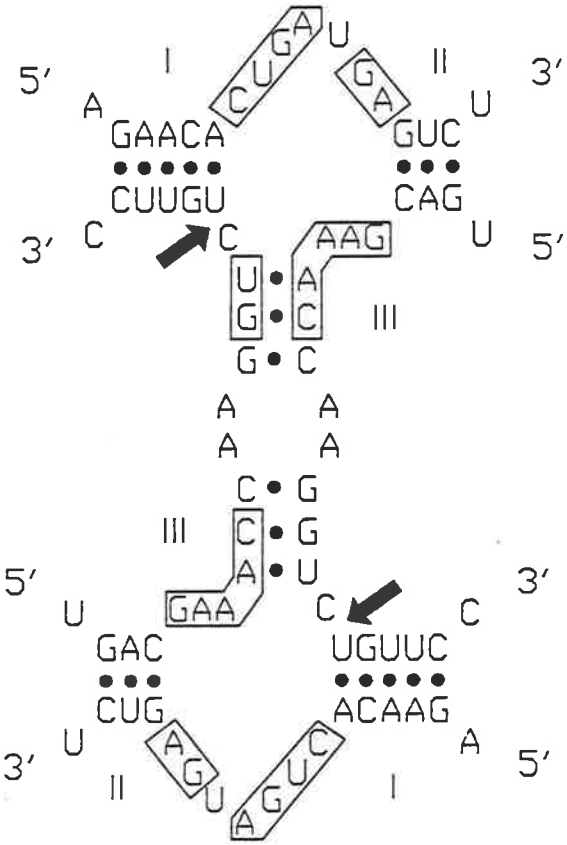
FRAGMENT	SIZE(nt)
5'E/3'E	307
5'E	178
3'E	129

Fig. 6.10 Double-hammerhead secondary structures for wild-type and mutant plus ASBV RNAs. (a), Double-hammerhead formed by the association of two wild-type Sau ASBV RNAs. (b), Double-hammerhead formed by the association of two Sau ASBV RNAs mutant in the stem III loop. In these (indentical) RNAs an extra A residue has been inserted into the CAG loop sequence (Fig. 6.4b). Boxed nucleotides, conserved in RNAs that cleave by the hammerhead model; stems are numbered 1 to 111, self-cleavage sites are indicated by thick arrows.

a


5′
3′
I
II
III
GAACA
CUUGU
GUC
CAG
AAG
GAA
GAC
CUG
UGUUC
ACAAG

b


5′
3′
I
II
III
GAACA
CUUGU
GUC
CAG
AAG
GAA
GAC
CUG
UGUUC
ACAAG

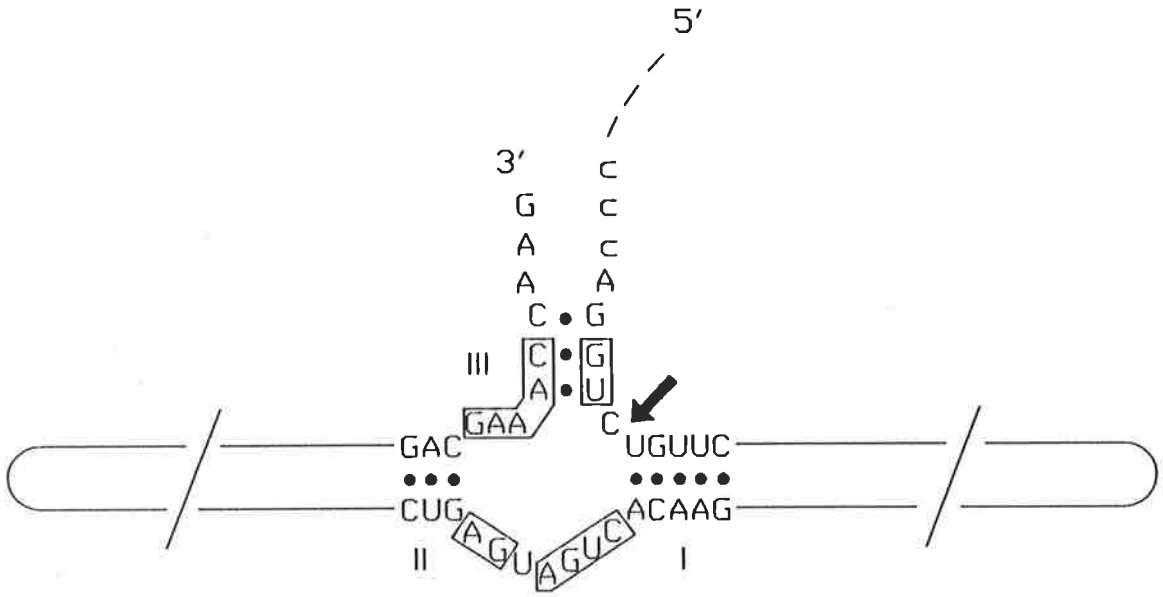
6.3.3 Self-cleavage of monomeric plus ASBV RNAs is promoted by the opening of stem III

Apart from altering the number sequence of bases in the loop of stem III or increasing the number of base-pairs in this stem the only other way to stabilize stem III is to relieve the stress in the loop by opening it out. This can be done simply by linearizing the dimeric BstN I clone between the two monomeric units at the Sty I site (Figs. 6.5b, 6.6) created by the blunting and ligation of the two BstN I monomers. The BstN I recognition sequence CC↓AGG was changed to the Sty I recognition sequence C↓CAAGG by this process during cloning. The full-length RNA produced from this vector undergoes very efficient self-cleavage during transcription (Fig. 6.2, lane 8). The structure proposed to be responsible for this self-cleavage reaction is given in Fig. 6.11 . Unfortunately, due to the nature of the sequences, the loop is not merely opened up but an extra G:C basepair is added to stem III.

6.4 DISCUSSION

In previous chapters (4, 5) it was shown that dimeric plus and minus ASBV RNAs cleave by the interaction of two self-cleavage domains probably by the proposed double-hammerhead model. All evidence suggests that (at least *in vitro*) the single-hammerhead self-cleavage structures previously proposed for plus and minus ASBV RNAs (Fig. 4.1) cannot cleave because they contain stem IIIs that do not allow an active conformation to be formed. Self-cleavage in sLTSV RNAs and plus sTRSV RNA occurs efficiently in RNAs that contain a single self-cleavage domain (Buzayan *et al.*, 1986b; Forster and Symons, 1987a; Forster *et al.*, 1987) presumably by a single-hammerhead. These RNAs contain stable stem IIIs and therefore there is no reason to suggest that these RNAs may cleave as double-hammerheads. If it is indeed the stem IIIs that prevent ASBV RNAs from cleaving as single-hammerheads then, by

Fig. 6.11 Single-hammerhead secondary structure for the monomeric plus ASBV RNA resulting from transcription of the dimeric BstN ASBV template truncated with Sty I (Fig. 6.5b). Boxed nucleotides, conserved in RNAs that cleave by the hammerhead model; stems numbered 1 to 111; self-cleavage site is indicated by an arrow; ASBV sequence not detailed is shown as lines. Vector derived nucleotides are in lower case or shown as a dashed line.


5′
3′
c
G c
A c
A A
C G
III
C
A
G
U
GAA
C
GAC
UGUUC
CUG
ACAAG
II
A G U
A G U C
I

altering the nature of this stem, it should be possible to convert these RNAs so that they will cleave as a single-hammerhead.

6.4.1 Cleavage of monomeric ASBV RNAs with alterations to stem III

Data presented in this chapter shows that the only monomeric ASBV RNAs that cleave efficiently during transcription are those which contain alterations in the stem III sequence. These results therefore offer further evidence in favour of the double-hammerhead model. Plus monomeric ASBV RNAs can be made to cleave efficiently during transcription (and hence presumably as single-hammerheads) by each of the following means:

(1) Opening the stem III loop and increasing the size of this stem by an additional base-pair (Fig. 6.11).

(2) As in (1) but with a greater increase in the number of base-pairs in stem III (Fig. 6.8).

(3) Adding an additional A nucleotide to the stem III loop (Fig. 6.4b).

Each of these changes would be expected to increase the stability of the hairpin; (1) and (2) quite obviously by relieving the strain created by the small loop and increasing the base-pairing, (3) probably by reducing the strain due to the small size (three nucleotides) of the loop and perhaps also by a more subtle influence due to the actual sequence generated in the loop.

The efficient cleavage of the plus ASBV RNA with an additional A nucleotide in the stem III loop, by a single-hammerhead, is contrasted by the lack of cleavage of the wild-type RNA. This appears somewhat anomalous as the stability of both stem IIIs (as calculated using the values of Freier *et al.*, 1986) are similarly low. However the effect of the loop sequences on hairpin stability has not been studied extensively

and this influence is therefore hard to estimate. There are two recent papers (Senior *et al.*, 1988; Tuerk *et al.*, 1988) which have shown that the loop sequence can play an important role in stem stability but the data in these papers is not specific for this example. It seems that the CAAG sequence in the mutant plus ASBV RNA is a sequence that exerts a stabilizing effect on the stem sufficient to allow self-cleavage by a a single-hammerhead. Therefore the CAG (wild-type) sequence may be conserved in ASBV RNAs as it destabilizes the stem III in monomeric RNAs, thus forcing cleavage to occur by a double-hammerhead structure. This may be a selective advantage as it could prevent the futile cleavage of circularized RNAs. It is interesting to note that the ASBV-like RNAs used by Uhlenbeck (1987) to demonstrate RNA catalysis contained a sequence which if drawn as a single-hammerhead had CCG not the wild-type CAG in the loop (Fig. 5.12a). This apparently minor change may allow some cleavage to occur by the single-hammerhead rather than the double-hammerhead model.

In these discussions it should be noted that the secondary structure models do not provide a mechanism for self-cleavage; it is the as yet unknown tertiary structure which determines whether cleavage will or will not occur.

6.4.2 The arrangement of ASBV self-cleavage sequences

The use of full-length monomeric and dimeric RNAs has been important in the development of structural models for ASBV self-cleavage as it has focused attention on the entire RNA sequence rather than just on those isolated sequences sufficient and necessary for cleavage. One obvious feature of the ASBV sequences is that the sequences required for self-cleavage are arranged in two blocks widely separated on the monomeric molecule (Fig. 4.2). The arrangement of the sequences required for self-cleavage in ASBV in two distinct blocks has implications for the

evolution of these molecules. The modular nature of the self-cleavage sequences may allow more flexibility in regard to recombinational events.

CHAPTER 7

PREPARATION AND ANALYSIS OF MUTANT LTSV AND CMV SATELLITE RNAs

7.1 INTRODUCTION

Although the nucleotide sequence of many satellite RNAs are known the functions of these sequences are generally poorly understood. One exception to this is the sequence involved in the self-cleavage of certain satellite RNAs (see Chapter 1). One approach that could be used to gain a better understanding of the functions of the satellite RNA is to generate mutants and then assay them for their effects on particular processes. This approach was used in this chapter to investigate two plant viral satellite RNAs; sLTSV and sCMV. The aim of this work was not so much to do an extensive investigation of these RNAs but rather to develop a suitable technique for mutagenesis and thereby prepare clones suitable for future studies.

sLTSV and sCMV seemed good choices for these studies for the following reasons. Both helper viruses have been at least partially characterized, for Q-CMV the entire nucleotide sequence is known (Rezaian *et al.*, 1984, 1985; this work). Both of these satellite RNAs can be grown in plant hosts that are easily propagated and from which virions and nucleic acids can be readily extracted. These two RNAs should also provide interesting contrasts as they have both similar properties (e.g. they are both small satellite RNAs) and dissimilar properties (e.g. unlike sCMV, sLTSV undergoes self-cleavage *in vitro* and is presumed to replicate by a rolling-circle mechanism (Forster and Symons, 1987a)). Also, not only should mutants of these two satellite RNAs prove valuable for *in vivo* studies, such as testing their infectivity and effect on symptomatology, they should also lend themselves to certain *in vitro* assays. For example sLTSV mutants could be tested for their ability to self-cleave. For sCMV mutants it may be possible to test them for their ability to act as templates for minus RNA synthesis using a template dependent RNA polymerase system similar to those prepared for the related viruses, AMV (Houwing and Jaspars, 1986) and BMV (Miller and Hall, 1983).

7.2 METHODS

7.2.1 Preparation of a full-length CMV satellite RNA clone

Kinased Q-sCMV ds cDNA was prepared from purified, RNA by Dr. Ali Rezaian using synthetic oligonucleotides as primers for first and second strand synthesis. This DNA was heated at 70°C for 10 minutes to inactivate any remaining kinase activity then ligated into dephosphorylated, Sma I digested pGem-I vector. The ligated DNA was transformed into competent MC1061 cells. Following overnight growth on L-broth/ampicillin plates 24 colonies were picked and DNAs were prepared by the small scale method. The size of any inserts was assayed by Hind III/EcoR I digestion followed by agarose gel electrophoresis. Clones containing what appeared to be full-length inserts were mapped using restriction endonucleases and their inserts were excised by Hind III/EcoR I digestion. Insert containing fragments were purified by agarose gel electrophoresis, eluted and ligated into dephosphorylated, Hind III/EcoR I digested M13mp18 and 19 vectors. The resultant clones were then sequenced and a large scale DNA preparation of the required full-length clone in the pGem-I vector was performed.

7.2.2. sLTSV-A full-length clone

The full-length sLTSV-A (the Australian isolate of sLTSV) clone prepared by D. Mitchell was used in this study. This clone was prepared in the following way. An M13mp18 clone containing a full-length monomeric sLTSV-A insert in the Sal I site was digested with Hind III/EcoR I and the insert containing fragment was subcloned into the pSP64 vector. This clone yielded plus-sense RNAs when transcribed with SP6 RNA polymerase. The M13mp18 clone containing the monomeric insert had been prepared by D. Mitchell from partial-length clones (Keese *et al.*, 1983).

7.2.3 Preparation of pIBI76 vector without a BamH I recognition sequence

Purified pIBI76 DNA was digested to completion with BamH I. The linearized vector was purified on an agarose gel, the DNA extracted and end-filled with the Klenow fragment of DNA polymerase I enzyme. This DNA was then extracted with an equal volume of phenol:chloroform (1:1, v/v) and ethanol precipitated. The resultant DNA pellet was dissolved, circularized with T4 Ligase and transformed into competent JM101 cells. Plasmid DNA from ampicillin resistant colonies was prepared by the small scale method and the polylinker sequence checked by digestion with BamH I, EcoR I and Sal I enzymes and agarose gel electrophoresis. Single-stranded DNA of the required clones was prepared for sequencing (using the appropriate helper phage) as described by the vector manufacturer and dideoxy sequencing was attempted as described in Methods. Despite a number of attempts no completely satisfactory results were obtained using this procedure and so, to save time, inserts were prepared by Hind III, EcoR I digestion and subcloned into M13mp18 vector. These clones were sequenced and the pIBI76 clone with an inactivated BamH I site was designated BamH I Δ pIBI76.

7.2.4 Preparation of full-length sCMV and sLTSV-A clones in BamH I Δ pIBI76 vector

The insert from the monomeric sLTSV-A clone in pSP64 (7.2.2) was excised with Sal I and the insert from the sCMV clone in pGem-I (7.2.1) was excised by Sma I, BamH I. Both inserts were purified by agarose gel electrophoresis. The sLTSV-A cDNA insert was ligated into dephosphorylated, Sal I digested, BamH I Δ pIBI76 vector. The sCMV cDNA insert was ligated into dephosphorylated Sma I digested, BamH I Δ, pIBI76 vector. Both ligation reactions were transformed into competent JM101 cells, DNA was prepared from ampicillin resistant colonies by the small scale method and the inserts assayed by restriction mapping. The Hind III,

EcoR I fragments containing the appropriate inserts were subcloned into M13mp18 and 19 and sequenced by the dideoxy chain termination technique.

7.2.5 Preparation of sCMV and sLTSV clones mutated by Bam-linker insertion

sLTSV and sCMV mutant clones were prepared using a protocol kindly provided by Dr. J. Haseloff, with only minor modifications. As this method has not been published it will be described in full.

Purified DNAs of the BamH I Δ pIBI76 sLTSV and sCMV clones were partially digested with DNase I in the presence of manganese ions. Optimal conditions were defined by a time course experiment. Thus 25 µg of each plasmid was digested for 3 minutes at 37°C in a final volume of 35 µl in 10 mM Tris-HCl pH 7.0, 15 mM $MnCl_2$, 0.35 U/µl DNase I. The reactions were stopped by the addition of 10 µl of ice-cold 1M EDTA and aliquots were assayed on an agarose gel to ensure that the reactions had proceeded satisfactorily. The DNAs were then phenol/chloroform extracted and ethanol precipitated. The resultant dried DNA were resuspended in 60 µl water and incubated in a final volume of 100 µl containing 50 mM Tris-HCl pH 8.3, 50 mM KCl, 10 mM $MgCl_2$, 10 mM DTT, 0.02 U/µl T4 DNA polymerase, 0.2 U/µl Klenow fragment of DNA Polymerase I for 20 minutes at 37°C. A mix of these two enzymes was used as the preferred T4 DNA polymerase enzyme was in short supply. Following this all four NTPs were added to a final concentration of 250 mM and the reactions were incubated for a further 45 minutes at 37°C then stopped by heating for 10 minutes at 70°C. The linearized plasmid fragments were purified by agarose gel electrophoresis and resuspended in a final volume of 30 µl of 0.1 mM EDTA. Phosphorylated BamH I linkers (CGGAUCCG) were then ligated onto the blunted ends of these fragments by incubating them in a final volume of 60 µl in the presence of 80 ng/µl phosphorylated BamH I linkers, 50 mM Tris-HCl pH 7.5, 5 mM $MgCl_2$, 10 mM DTT, 1 mM ATP, 5% PEG 6000,

0.4 U/µl T4 DNA ligase for 16 hours at 18°C. These DNAs were ethanol precipitated, redissolved in x1 high salt restriction enzyme buffer and digested in a final volume of 20 µl with 20 units of BamH I enzyme for 2 hours at 37°C. Linearized DNAs were separated from the non-linkered (non-digested) DNAs on an agarose gel and then circularized with T4 DNA ligase. Ligations were carried out for 4 hours at 20°C in a final volume of 60 µl containing 50 mM Tris-HCl pH 7.5, 5 mM $MgCl_2$, 10 mM DTT, 1 mM ATP, 0.1 U/µl T4 DNA ligase. The reaction mix was then diluted with an equal volume of water and 6 x 20 µl samples were transformed into competent JM101 cells prepared by the modified Hanahan high efficiency protocol. The many ampicillin resistant colonies were pooled for each of the sLTSV and sCMV clone preparations by washing each agar plate with 2 x 2 ml of 2x YT media. These suspensions were than added to 500 ml of 2x YT media containing 100 µg/ml ampicillin and grown overnight at 37°C. DNA was prepared by the large scale method and a portion of each of the clone preparations was digested with Hind III and EcoR I enzymes to excise the potentially mutated inserts which were then purified by agarose gel electrophoresis. The inserts were then ligated into dephosphorylated Hind III/EcoR I digested, pGem-2 vector and transformed into competent MC1061 cells. This step is needed to enable insertion of the potentially mutated inserts into non-mutated vectors. The presence of mutated inserts in these clones was assayed by analysis of the fragments produced by a triple digestion using Hind III, EcoR I and BamH I. The desired clones were those that produced a large vector sized band and two smaller fragments whose summed size was roughly equal to that of the original insert. Only sLTSV clones were extensively characterized. sLTSV clones yielding the desired pattern of bands were digested with either Hind III or EcoR I and transcribed with SP6 and T7 RNA polymerases respectively in the presence of α-^{32}P-UTP. The inserts from clones yielding the expected pattern of RNA bands (which included any bands resulting from self-cleavage) were excised with Hind III and EcoR I, purified by agarose gel electrophoresis and ligated into

dephosphorylated Hind III/EcoR I digested, M13mp18 and 19 vectors. Those clones were then sequenced by the dideoxy chain termination technique.

7.3 RESULTS

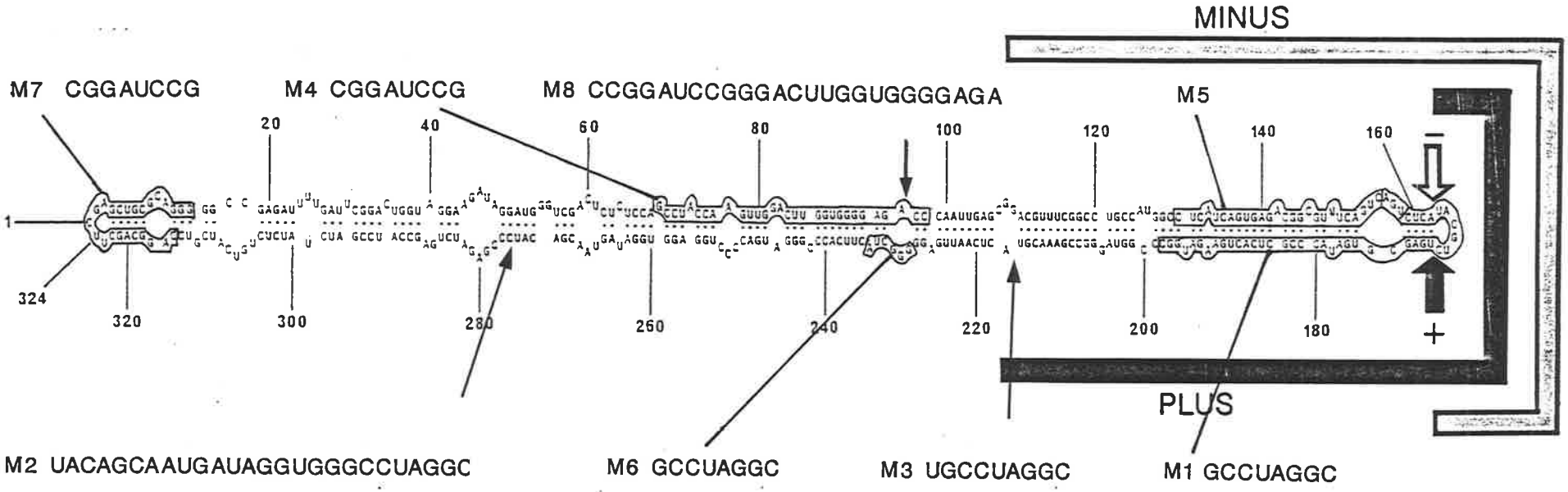
7.3.1 A full-length clone of Q-sCMV

cDNA to Q-sCMV was cloned into an RNA transcription vector and M13 vectors. Sequencing a number of clones with potentially full-length inserts enabled the isolation of a clone having a sequence very similar to the Q-sCMV originally described by Gordon and Symons (1983). This insert was full-length (337 nucleotides) and had a sequence which differed from the previously published sequence in only three positions. The differences in sequence were a U to A substitution at nucleotide 96, an insertion of an extra C after nucleotide 138, and a C to U substitution at nucleotide 254 (numbers refer to the 337 nucleotide sequence presented in this work).

7.3.2 Full-length sLTSV-A clones mutated by Bam-linker mutagenesis

Of the sLTSV-A mutants prepared by Bam-linker mutagenesis eight were sequenced fully and their ability to self-cleave *in vitro* was examined. The sequences of eight mutants are shown in Fig. 7.1. The type of alteration made to the wild-type sequence depends on the positioning of the nicks made by the DNase I enzyme. If the nicks in the opposite strands were adjacent (or nearly so) then the Bam-linker was inserted with few other changes (e.g. mutants M3, M6). If the two nicks were staggered then either insertions (e.g. mutants M2, M8) or deletions (e.g. mutants M1, M5) were made depending on whether a 3'- or 5'-overhang was produced by DNase I.

Fig. 7.1 Secondary structure model of plus sLTSV-A RNA (Forster and Symons, 1987a) including the sequence changes in the eight characterized mutants (7.2.5). Three of the mutants (M2, M3, M8) are simple insertions, the inserted sequence is given in large letters and the site of insertion is marked by a thin arrow. One of the mutants (M5) is a simple deletion, the deleted sequence is boxed. The remainder of the mutations (M1, M4, M6, M7) are deletions of sLTSV sequences accompanied by insertions of the BamH I linkers. The deleted sequences are boxed and the site of the insertion of the Bam-linker indicated. The sequences required for plus RNA self-cleavage are indicated by a heavily shaded box, the sequences required for minus RNA self-cleavage are indicated by a lightly shaded box. Numbering after Keese *et al.* (1983).


MINUS
M7 CGGAUCCG
M4 CGGAUCCG
M8 CCGGAUCCGGGACUUGGUGGGGAGA
M5
1
20
40
60
80
100
120
140
160
324
320
300
280
260
240
220
200
180
PLUS
M2 UACAGCAAUGAUAGGUGGGCCUAGGC
M6 GCCUAGGC
M3 UGCCUAGGC
M1 GCCUAGGC

7.3.3 Self-cleavage of sLTSV-A mutant RNAs during transcription

Transcripts of both polarities were produced from mutant and wild-type sLTSV-A clones by digestion with Hind III and transcription with SP6 RNA polymerase, for the plus RNAs and EcoR I and T7 RNA polymerase for minus RNAs. The resultant products were run on a denaturing polyacrylamide gel (Fig. 7.2 for plus RNAs, Fig. 7.3 for minus RNAs). The wild-type templates and expected plus and minus transcripts are given in Fig. 7.4. As previously demonstrated for partial-length sLTSV-A transcripts (Forster and Symons, 1987a) both plus and minus wild-type full-length monomeric RNAs (Figs. 7.2, 7.3, lanes 2) cleaved specifically at a single site (to produce 5' and 3' fragments the sizes of which were as expected for the previously demonstrated site of self-cleavage). No cleavage occurred in either the plus or minus RNA transcripts of the M1 mutant clone (Figs. 7.2, 7.3, lanes 3). This is in accordance with previous observations as sequences required for self-cleavage in both plus and minus sLTSV-A RNAs (Forster and Symons, 1987a,b; Forster *et al.*, 1987) are deleted in this clone (Fig. 7.1). The minus RNA transcript of clone M5 does not undergo self-cleavage (Fig. 7.3, lane 7). This is again due to the deletion of some of the sequences required for self-cleavage (Fig. 7.1). In contrast the plus RNA transcript of the M5 clone cleaves at the predicted site (Fig. 7.2, lane 7) albeit not as efficiently as the wild-type RNA. The M3 mutant is the only other clone which has a mutation in sequences closely adjacent to those required for self-cleavage (Fig. 7.1). The plus RNA from this clone has the Bam-linker (plus an additional U nucleotide) inserted into the stem III of the secondary structure proposed to be responsible for plus RNA self-cleavage (Forster and Symons, 1987a,b). Despite this alteration this RNA cleaves to roughly the same extent as the wild-type RNA during transcription (Fig. 7.2, lane 5).

Fig. 7.2 Autoradiograph of 5% polyacrylamide , 7M urea gel showing plus RNA transcripts of wild-type and mutant monomeric sLTSV-A templates (see Figs. 7.1, 7.4a). Lane 1; ^{32}P-labelled Hpa II digested pUC19 DNA as markers, with band sizes given in nucleotides. Lane 2; Wild-type template. Lane 3; M1 mutant template. Lane 4; M2 mutant template. Lane 5; M3 mutant template. Lane 6; M4 mutant template. Lane 7; M5 mutant template. Lane 8; M6 mutant template. Lane 9; M7 mutant template. Lane 10; M8 mutant template. The full-length transcripts (5'F/3'F) and the RNAs produced by self-cleavage (5'F, 3'F) are indicated.

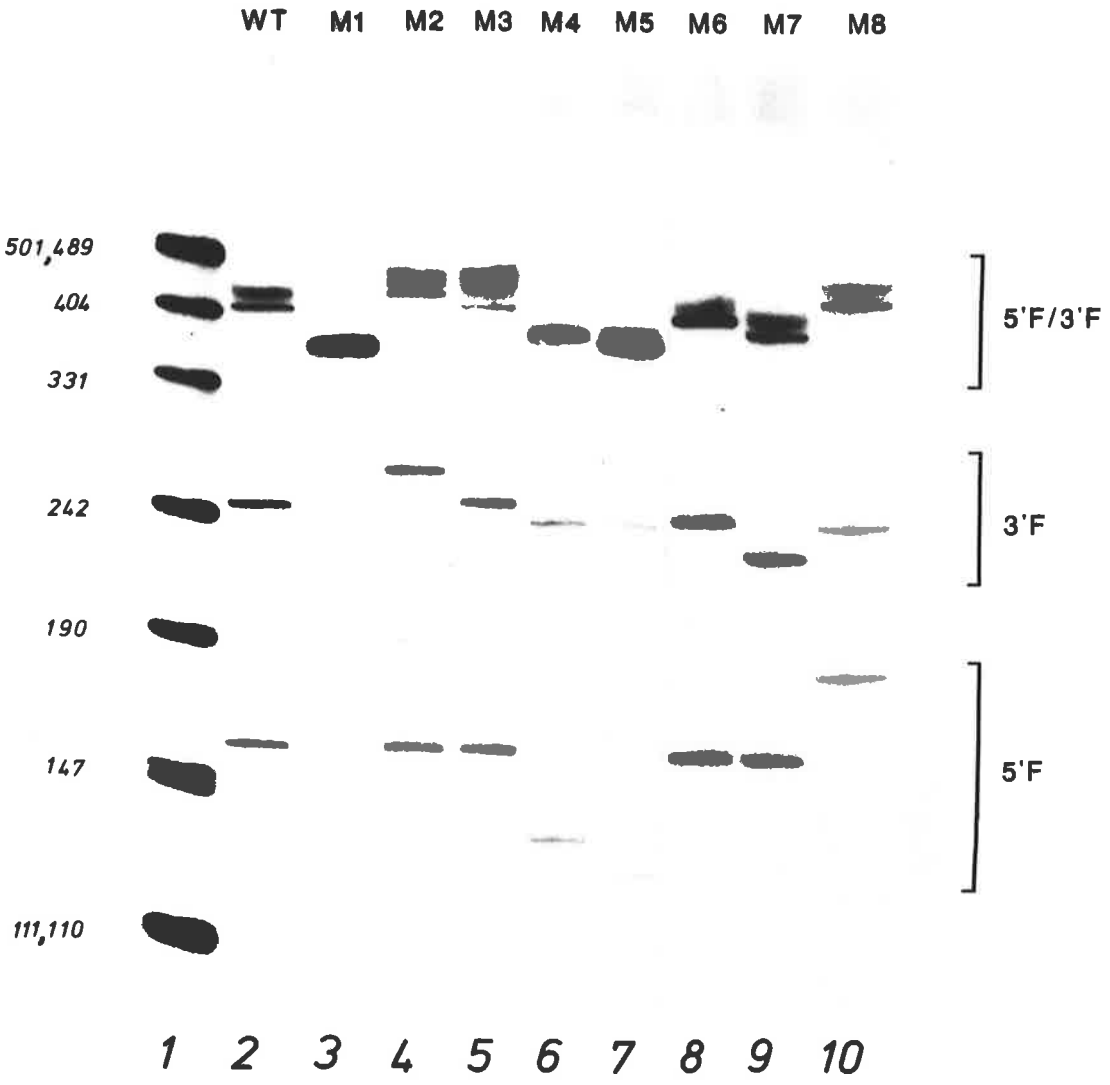

WT
M1
M2
M3
M4
M5
M6
M7
M8
501,489
404
331
242
190
147
111,110
5'F/3'F
3'F
5'F
1
2
3
4
5
6
7
8
9
10

Fig. 7.3 Autoradiograph of 5% polyacrylamide, 7M urea gel showing minus RNA transcripts of wild-type and mutant monomeric sLTSV-A templates (see Fig. 7.1, 7.4b). Lane 1; ^{32}P-labelled Hpa II digested pUC19 DNA as markers, sizes given in nucleotides. Lane 2; Wild-type template. Lane 3; M1 mutant template. Lane 4; M2 mutant template. Lane 5; M3 mutant template. Lane 6; M4 mutant template. Lane 7; M5 mutant template. Lane 8; M6 mutant template. Lane 9; M7 mutant template. Lane 10; M8 mutant template. The full-length transcripts (5'F/3'F) and the RNAs produced by self-cleavage (5'F, 3'F) are indicated.

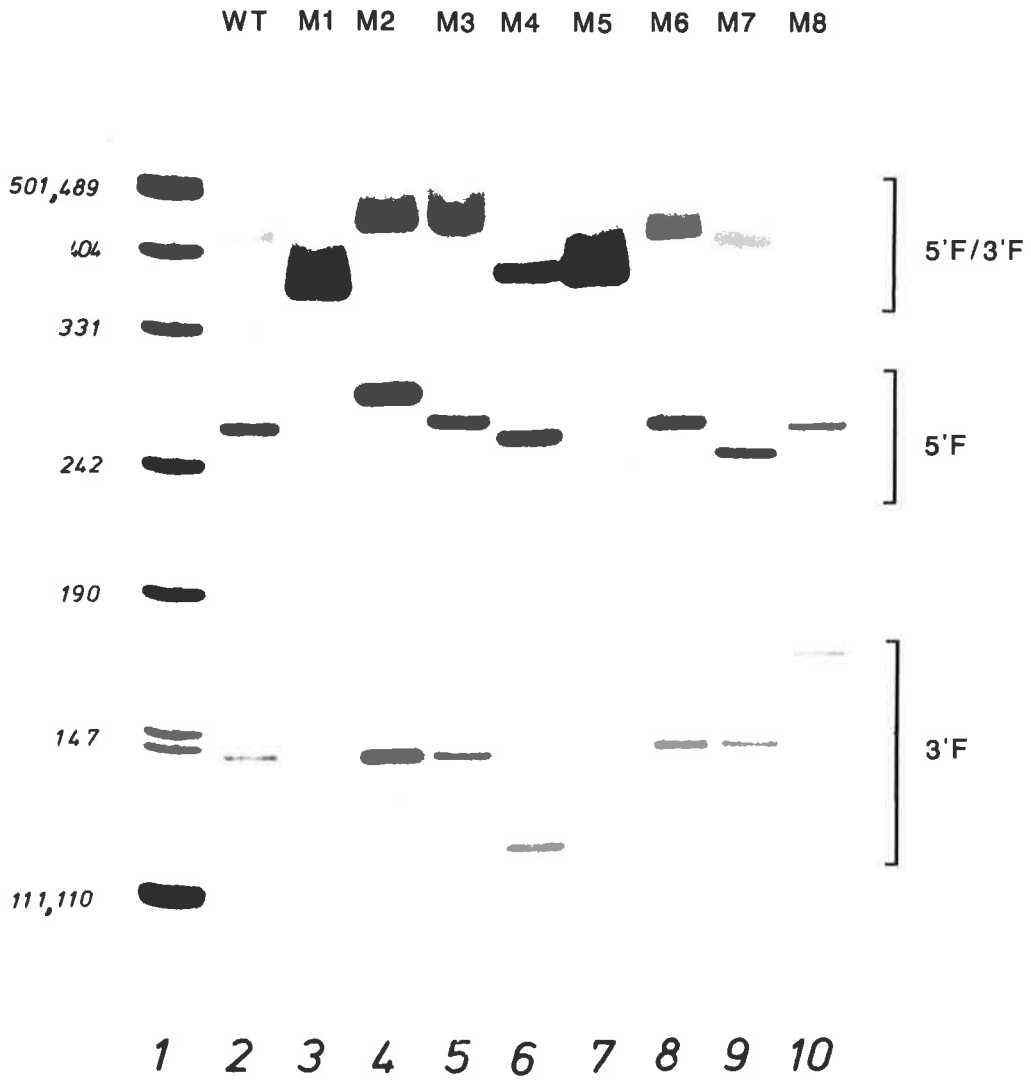
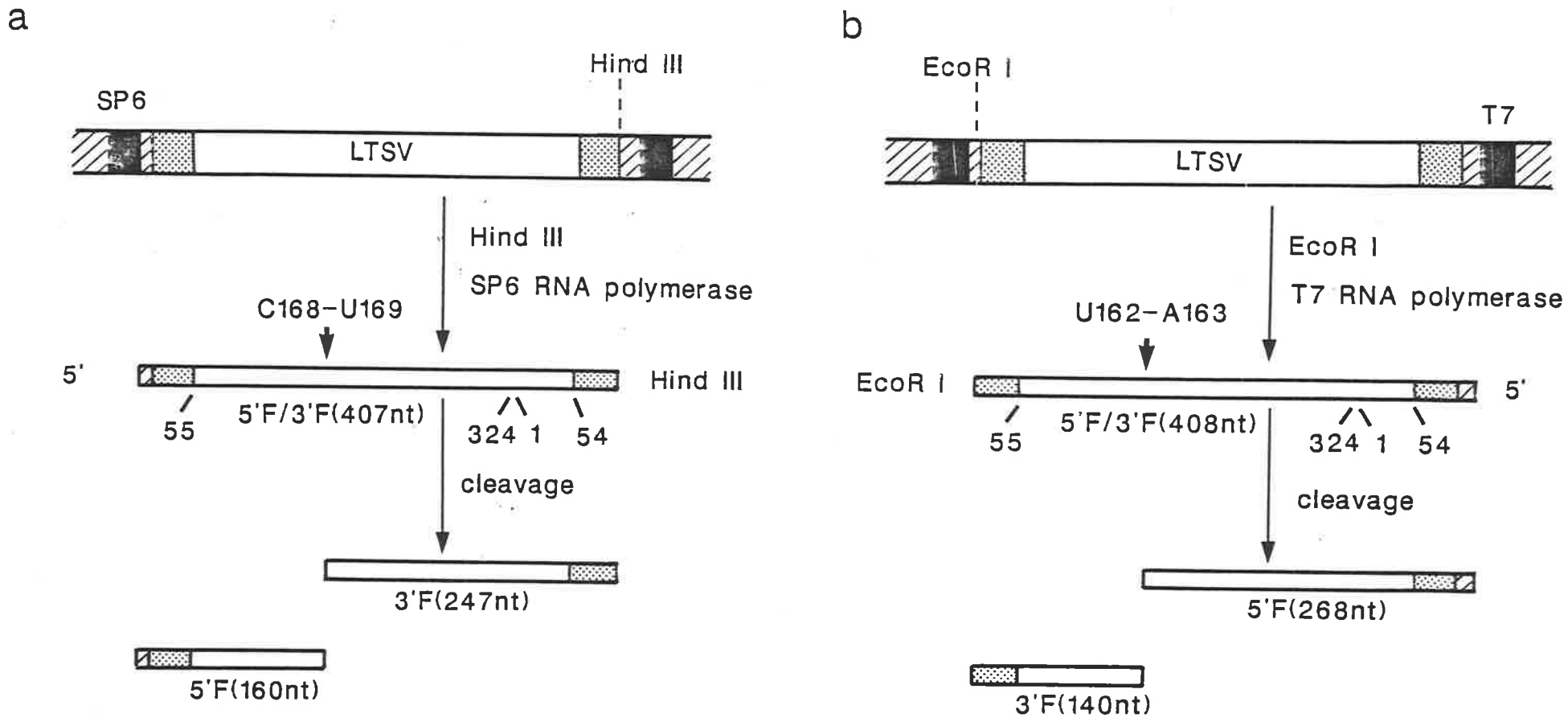

WT
M1
M2
M3
M4
M5
M6
M7
M8
501,489
404
331
242
190
147
111,110
5'F/3'F
5'F
3'F
1
2
3
4
5
6
7
8
9
10

Fig. 7.4 Wild-type monomeric sLTSV templates and the RNAs resulting from their transcription. (a), Template linearized with Hind III and transcribed with SP6 RNA polymerase to yield plus RNAs. (b), Template linearized with EcoR I and transcribed with T7 RNA polymerase. Hatched boxes, pGem-2 vector sequences; stippled boxes, pIBI76 vector sequences; closed boxes, SP6 and T7 RNA polymerase promoters; open boxes, sLTSV sequences. Full-length transcripts (5'F/3'F) and RNA fragments resulting from self-cleavage (5'F, 3'F) are shown. Self-cleavage sites are indicated by thick arrows. The plus sequence is numbered after Keese *et al.* (1983). The numbering for the plus sequence is retained for the minus RNA.


a
Hind III
SP6
LTSV
Hind III
SP6 RNA polymerase
C168–U169
5'
Hind III
55
5'F/3'F(407nt)
324 1 54
cleavage
3'F(247nt)
5'F(160nt)
b
EcoR I
T7
LTSV
EcoR I
T7 RNA polymerase
U162–A163
EcoR I
5'
55
5'F/3'F(408nt)
324 1 54
cleavage
5'F(268nt)
3'F(140nt)

7.3.4 Self-cleavage of full-length sLTSV-A RNAs *In vitro*

Full-length RNA transcripts of the wild-type and mutant clones were prepared and incubated, with and without prior heating then snap cooling, in 5 mM $MgCl_2$ for 10 minutes at 25°C, as described in Methods. Transcripts from those RNAs (both plus and minus) which cleaved during transcription (Figs. 7.2, 7.3) did not cleave unless first treated by heating and snap cooling. After heating and snap cooling, those RNAs that cleaved during transcription self-cleaved to yield products of the expected size and cleaved to approximately the same extent as they did during transcription (results not shown).

7.4 DISCUSSION

7.4.1 The full-length Q-sCMV clone is a sequence variant of the original isolate

The full-length Q-sCMV clone prepared in this study contained three differences compared to the originally sequenced Q-sCMV RNA (Gordon and Symons, 1982). All three changes occurred in positions which have been shown to be variable in at least one or more different CMV satellite RNA isolates (Kaper *et al.*, 1988a). Sequence comparisons will not be discussed in any detail here as an exhaustive comparison of 14 different CMV satellite RNAs has recently been published (Kaper *et al.*, 1988a). However comparison of the sequence of this clone with another that has been shown to be capable of acting as a template for the production of infectious RNAs (Kurath and Palukaitis, 1988) indicates that transcripts of this clone should be infectious *in vivo*. Infectivity trials using transcripts of this clone were attempted but problems with sCMV contamination of the helper virus stocks have thus far prevented any meaningful experiments. Full-length infectious clones of the genomal CMV RNAs are currently being prepared in this laboratory and it was decided to wait upon their production before further infectivity experiments are

attempted and extensive characterization of the Bam-linkered Q-sCMV clones undertaken.

7.4.2 Bam-linker mutagenesis

The technique of Bam-linker mutagenesis described in Methods allowed the production of clones having mutations which depend on the site and manner of DNase I cleavage. It was considered desirable to refine a method which, as nearly as was practicable, produced randomly distributed mutations of various types. This is because, apart from the sequences involved in RNA self-cleavage, there is little or no information regarding the function of the remainder of the sequences in plant satellite RNAs. The clones most suitable for *in vivo* studies may be those that do not contain insertions (consisting of repeated sequences), as the repeated sequences may be removed by homologous recombination thus allowing the wild-type sequence (or something akin to it) to be reconstituted. A study of this phenomenon in itself however may prove interesting.

7.4.3 Self-cleavage of mutant and wild-type sLTSV-A RNAs

One feature of the sLTSV-A mutant clones that can be assayed *in vitro* as a preliminary to infectivity studies is the self-cleavage of the plus and minus RNAs. The sizes of the fragments arising from self-cleavage of the wild-type full-length plus and minus RNAs used in this study agree with those expected if cleavage occurred at the previously demonstrated site of self-cleavage in partial-length sLTSV-A RNAs (Forster and Symons, 1987a). Forster and Symons (1987a) have shown that transcripts of partial-length clones cleave efficiently during transcription but cleavage of the complete RNA (in the presence of Mg^{2+} ions), purified by polyacrylamide gel electrophoresis of the transcription mixture, occurred only after prior heating and snap cooling. The complete (purified) transcripts of the wild-type full-length templates

used in this study also required heating and snap cooling prior to self-cleavage. This demonstrates that this requirement is not an artifact of transcripts of partial-length clones. An equilibrium between active and inactive structures has been proposed by Forster and Symons (1987a) (Fig. 7.5). During transcription the equilibrium is somewhat to the right, i.e. towards the active form. After transcription the equilibrium for the complete transcript appears to be almost totally towards the inactive form and self-cleavage of this RNA can therefore only occur if the structure is altered to the active form by heating and snap cooling. In the plus RNA it is therefore the sequence (or part of it) between nucleotides 220 (approximately) and 3 (remembering that the clone used by Forster and Symons (1987a) was only of partial length) that acts to push the equilibrium towards the inactive structure. This is despite the fact that the sequences proposed to base-pair with those required for plus RNA self-cleavage (Fig. 7.5) are synthesized prior to the self-cleavage sequences themselves and are therefore present in RNAs that cleave during transcription. With this in mind it should be possible to produce sLTSV-A RNAs that do not cleave (or cleave less well) during transcription by initiating synthesis from a different location. For example if initiation of the plus monomeric RNA began at around nucleotide 220, one would expect the native (inactive) structure to form. This demonstrates the importance of the site of cloning in experiments such as these and may also reflect a restriction on the site of initiation of RNA synthesis *in vivo*. It is difficult to speculate however on the *in vivo* situation as other factors, such as proteins, may influence the formation and stability of the RNA secondary structures.

None of the mutations to the vLTSV-A sequence circumvented the requirement for heating and snap cooling of purified, complete plus or minus transcripts prior to cleavage. Those large mutations that might be expected to disrupt the inhibiting secondary structures (e.g. M2, M4, M5, M8, Fig. 7.1) obviously still allow them to form. For example mutant M5 in the plus RNA is a deletion of some of the sequences involved in base-pairing with the self-cleavage sequences in the

Fig. 7.5 Secondary structure models showing the proposed structural switching of the sLTSV RNA self-cleavage domains (Forster and Symons, 1987a) with the sequnces of three mutants (M1, M3, M5) superimposed. The heavily shaded boxes delineate the sequences that base-pair to form the three stems in the active hammerhead structure. The site of plus RNA self-cleavage is indicated by a heavily shaded arrow. Stems are numbered 1 to 111. The lightly shaded boxes delineate the complementary sequences that base-pair in the minus RNA to form an active hammerhead structure. Because of their almost total complementarity the sequences involved in minus RNA self-cleavage can form structures very similar to the plus RNA (not shown). Numbering after Keese *et al.* (1983).

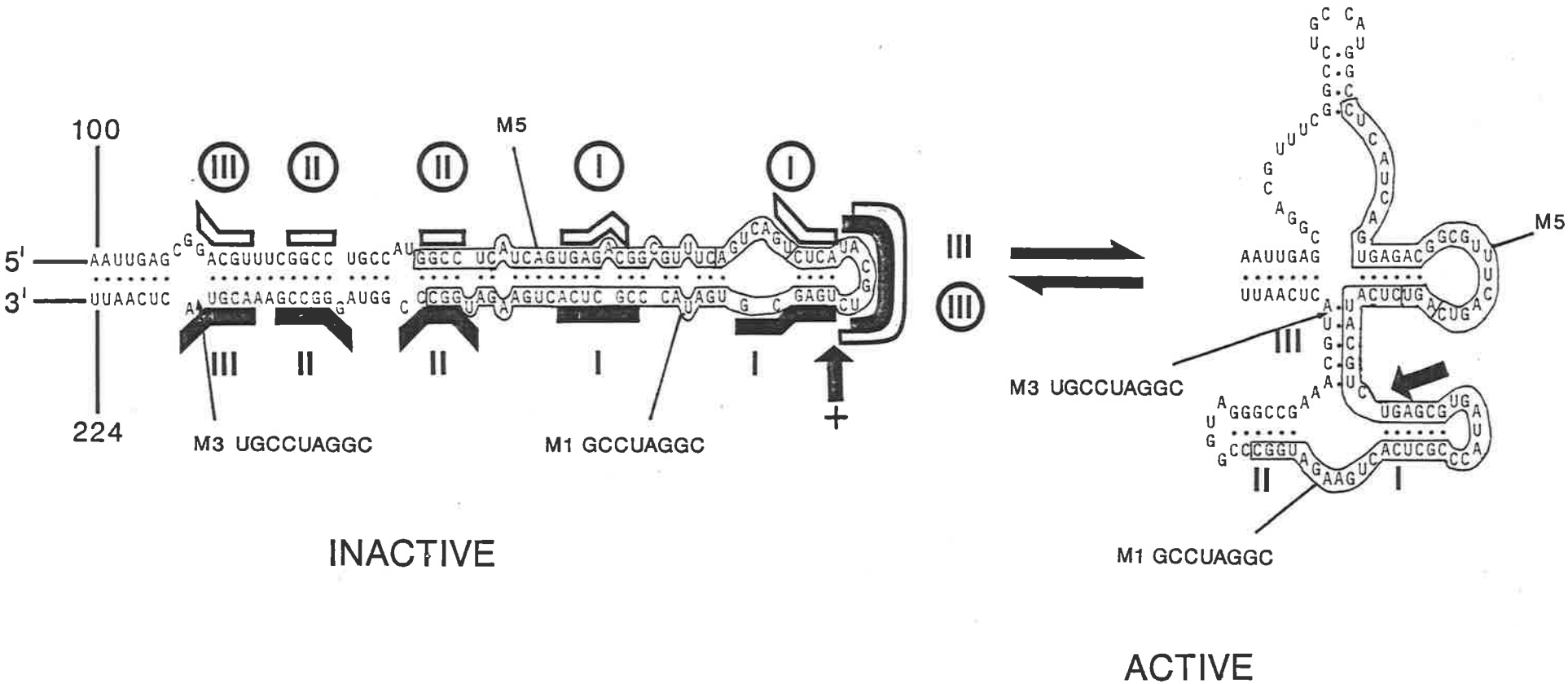
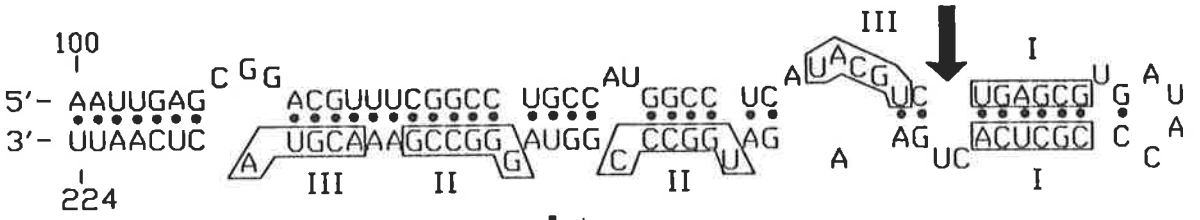
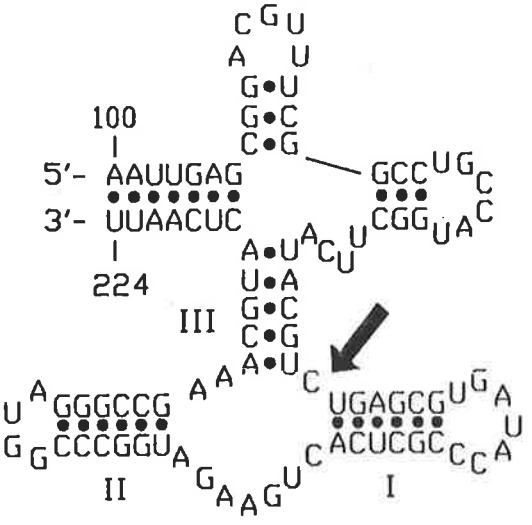

100
5'
3'
224
AAUUGAG
UUAACUC
M5
M3 UGCCUAGGC
M1 GCCUAGGC
INACTIVE
M3 UGCCUAGGC
M5
M1 GCCUAGGC
ACTIVE

Fig. 7.6 Secondary strucIture models showing the proposed structural switching of the M5 mutant plus sLTSV RNA (Figs. 7.1, 7.5) self-cleavage domain. Labelled as in Fig. 7.5 but only the sequences involved in plus RNAS self-cleavage are indicated (boxed).

INACTIVE





ACTIVE

inactive structure (Fig. 7.5). However a rod-like structure that inhibits formation of the active self-cleavage can still be formed in this mutant RNA (Fig. 7.6a). Heating and snap cooling of the complete plus RNA of this mutant in its inactive structure still allows self-cleavage, probably by the structure given in Fig. 7.6b.

Forster and Symons (1987a) proposed that the inability of sLTSV-A RNAs to cleave after transcription may be necessary to prevent futile cleavage of the circularized monomers believed to be templates for rolling-circle replication. However Kiberstis *et al.* (1985) showed that a 2'-phosphomonoester group is present in stoichiometric amounts at the self-cleavage site of plus sVTMoV and sSNMV RNAs. This group would almost certainly prevent self-cleavage without having to invoke the proposed (Forster and Symons, 1987a) control by structural switches.

It is noteworthy that the M5 mutant (Fig. 7.1) reduces the level of self-cleavage in the plus RNA without preventing it. This is another demonstration that sequences other than those shown to be necessary and sufficient for self-cleavage (Forster and Symons, 1987b) can effect the reactions efficiency and emphasizes the importance of considering phenomena such as self-cleavage in the context of the entire molecule.

Infectivity of the sLTSV-A mutants was not reported in this study but preliminary results suggest that Sal I excised cDNA inserts of the monomeric wild-type clone are infectious (results not shown). Infectivity studies were complicated by the fact that they had to be performed using sowbane mosaic virus (SoMV) as the helper virus due to the lack of a satellite-free LTSV stock. Although SoMV supports the replication of sLTSV it is a heterologous system and so the validity of any results arising from experiments using this combination may be questionable. As yet attempts to propagate a truly satellite free LTSV strain in this laboratory have not been successful. Partial nucleotide sequence data for the LTSV helper virus has been

obtained (P. Keese, pers. comm.) and it is hoped that its completion and the development of a satellite-free virus stock will encourage further experiments. Even with this, an alternative to whole plants as the assay system may have to be developed as problems of extensive recombination have been found in mutant RNAs in such systems (J. Haseloff, pers. comm.)

The M5 mutant (Fig. 7.1) could prove an interesting candidate for infectivity studies. If this clone is infectious the mutation may convert the replication cycle from one in which both plus and minus RNAs arise from transcription of circular monomeric templates to one in which the template for plus RNA transcription is longer-than-unit-length. This is the type of replication cycle previously proposed for sSNMV and sVTMoV RNAs (Fig. 1.1) (Chu *et al.*, 1983; Forster and Symons, 1987a) both of which have a hammerhead type self-cleavage domain in the plus RNA only and appear to produce only greater-than-unit-length minus species *in vivo*.

CHAPTER 8

THE SEQUENCES, SELF-CLEAVAGE AND REPLICATION OF SCMoV SATELLITE RNAs

8.1 INTRODUCTION

SCMoV is a member of a group of putative Sobemo viruses that encapsidate circular satellite RNAs. Four isolates of SCMoV have been described (Francki *et al.*, 1983b). The virions from all isolates were found to be of the same appearance as judged by electron microscopy and the coat proteins were indistinguishable from each other by serological tests. All isolates contained a linear ssRNA of approximately 4.5 Kb (the helper virus RNA). However whilst the helper virus RNAs of the four isolates appeared to be similar the encapsidated circular RNA components varied. Each isolate contained either one or both of two satellite RNAs, termed s(388) SCMoV and s(332) SCMoV on the basis of their sizes as determined in this work. This nomenclature replaces that of Francki *et al.* (1983b) who referred to these RNAs as 2 and 2' respectively. It appears therefore that either there is a single virus which can support the replication of two different satellite RNAs or that there are two very similar viruses each capable of supporting a different satellite RNA. In this work the homogeneity of the helper virus genomes was investigated and the sequences of the satellite RNAs determined. The analysis of the SCMoV helper virus RNAs and the sequencing of the sSCMoV RNAs was a collaborative effort with Dr. J. Haseloff. Additional studies relate some regions of the satellite RNAs to their replicative cycle.

8.2 METHODS

8.2.1 Virus and viral RNA purification

SCMoV isolates were kindly provided by Dr. R. Francki. Dried leaf inocula were ground in water and mechanically inoculated to young clover (Trifolium subterraneum L. cv. Dinninup or cv. Mt. Barker) or pea plants (Pisum sativum L. cv. Greenfeast). Approximately two weeks after inoculation infected leaves were harvested and either total nucleic acids extracted by the sodium perchlorate method of

Rezaian and Krake (1987) or the virus was purified (Francki *et al.*, 1983b) and the nucleic acids extracted using phenol/SDS (Gould, 1981). Uninfected material was also extracted by these two methods. The various required RNAs were purified by polyacrylamide electrophoresis essentially as described by Gould (1981).

8.2.2 Synthesis and restriction endonuclease cleavage of SCMoV ds cDNA

Double-stranded cDNA was produced from purified SCMoV RNAs as described in Chapter 3 for CMV RNA3. These cDNAs were digested with various restriction endonucleases and fractionated on a 6% polyacrylamide gel containing 2M urea.

8.2.3 Sequence determination of sSCMoV RNAs

The nucleotide sequences of the satellite RNAs from all four SCMoV isolates were determined using both RNA and cDNA sequencing methods. Linear RNA fragments of purified sSCMoV RNAs were generated by partial digestion under non-denaturing conditions essentially as described by Haseloff and Symons (1981). The resulting linear fragments were ^{32}P-labelled at either the 5'- or 3'-ends, fractionated by polyacrylamide gel electrophoresis, eluted and sequenced by the partial enzymic cleavage method. The resulting RNA sequence was used to design two synthetic oligonucleotides which were complementary to regions conserved in both sSCMoV RNAs (5' TTCAAAATCCGCGAGGAGGG 3' and 5' CGAAAGTGAGGTGGGGCC 3'). Double-stranded cDNAs were synthesized from circular RNAs using the oligonucleotides as primers essentially as described by Gould and Symons (1982). The ends of these RNAs were trimmed with S1 nuclease and ligated into bacteriophage M13mp19 and 19 vectors. These clones were then sequenced by the chain termination technique. Overlapping sequences obtained from RNA and cDNA

sequence analysis were assembled to give the complete nucleotide sequences of each sSCMoV RNA.

8.2.4 Preparation of s(322) SCMoV clones in pGem transcription vectors

An s(322) SCMoV clone in pSP64 containing a greater-than-unit-length cDNA insert was prepared by Dr. J. Haseloff. The insert from this clone was excised by Hind III/EcoR I digestion and purified by agarose gel electrophoresis. The insert containing fragment was then ligated into dephosphorylated, Hind III/EcoR I digested M13mp18 and 19 vectors (for sequencing) and into pGem-2 vector (for RNA synthesis). The M13 DNAs were transformed into competent JM101 cells and the pGem DNAs into MC1061. The M13 clones were sequenced by the dideoxy chain termination technique. The pGem-2 clones were assayed for the presence and orientation of the sSCMoV insert by restriction endonuclease mapping and transcription with SP6 and T7 RNA polymerases. Some of the purified Hind III/EcoR I insert prepared from the pSP64 clone was digested with Alu I and Taq I enzymes to allow isolation of the subset of sequences thought to be required for RNA self-cleavage. The small Alu I/Taq I fragment was isolated from an agarose gel and end-filled using the Klenow fragment of DNA polymerase I. This blunt-ended fragment was then ligated into dephosphorylated, Sma I digested, pGem-1 vector and transformed into competent MC1061 cells. DNAs prepared by the small scale method from ampicillin resistant colonies were assayed for the presence of inserts of the correct size by assaying Hind III/EcoR I digests on an agarose gel. Clones containing correctly sized inserts were digested with Hind III/EcoR I, the insert containing fragments were purified by agarose gel electrophoresis and then ligated into dephosphorylated Hind III/EcoR I digested M13mp18 for sequencing by the chain termination technique.

8.2.5 Identification of the site of self-cleavage in s(322)SCMoV RNA

A pGem-1 clone containing the Alu I/Taq I fragment in the Sma I site (see above) was digested with Hind III and transcribed with T7 RNA polymerase in the presence of all four (unlabelled) NTP's. The plus-sense RNA transcripts produced by this reaction were purified on a denaturing polyacrylamide gel and the 109 nucleotide 3'-fragment (Fig. 8.5) resulting from self-cleavage was excised and eluted. This RNA was ^{32}P-labelled at the 5' end and sequenced by the enzymic sequencing method.

8.2.6 Blot hybridization analysis of sSCMoV RNAs

Nucleic acid samples were denatured with 1 M glyoxal (McMaster and Carmichael, 1977) in the absence of dimethyl sulphoxide and run on 2.0% agarose vertical gels in 10 mM sodium phosphate (pH 6.5) at 30 mA. Nucleic acids were transferred to nitrocellulose by blotting for 16 hr. and baked *in vacuo* at 80° for 2 hr. (Thomas, 1980). The nitrocellulose filters were prehybridized, hybridized and washed as described by Thomas (1980) except for the following modifications. Filters were washed for 10 minutes at 90°C in distilled water after baking, prehybridization and hybridization were done at 55°C and 68°C respectively and after hybridization the filters were washed at 55°C. To prevent any cross-contamination between nucleic acids from virions and plant extracts, the baked filters were cut into strips that contained either of the two types of sample (including the appropriate healthy control). These strips were prehybridized and hybridized in separate bags.

8.2.7 Preparation of ^{32}P-labelled RNA probes for hybridization blot analysis

The pGem-2 clone containing the greater-than-unit-length s(322) SCMoV insert (see above) was digested with EcoR I and transcribed (in the presence of

α-^{32}P-UTP) with T7 RNA polymerase to produce minus RNA transcripts suitable for the detection of plus sequences. A probe for minus sequences was produced by digesting with Hind III and transcribing with SP6 RNA polymerase. The transcription mixes were run on a 5% polyacrylamide, 7 M urea gel and the appropriate bands (two in the case of the plus RNA) were excised and eluted. After ethanol precipitation and drying the labelled RNAs were dissolved in 0.1 mM EDTA and used as probes.

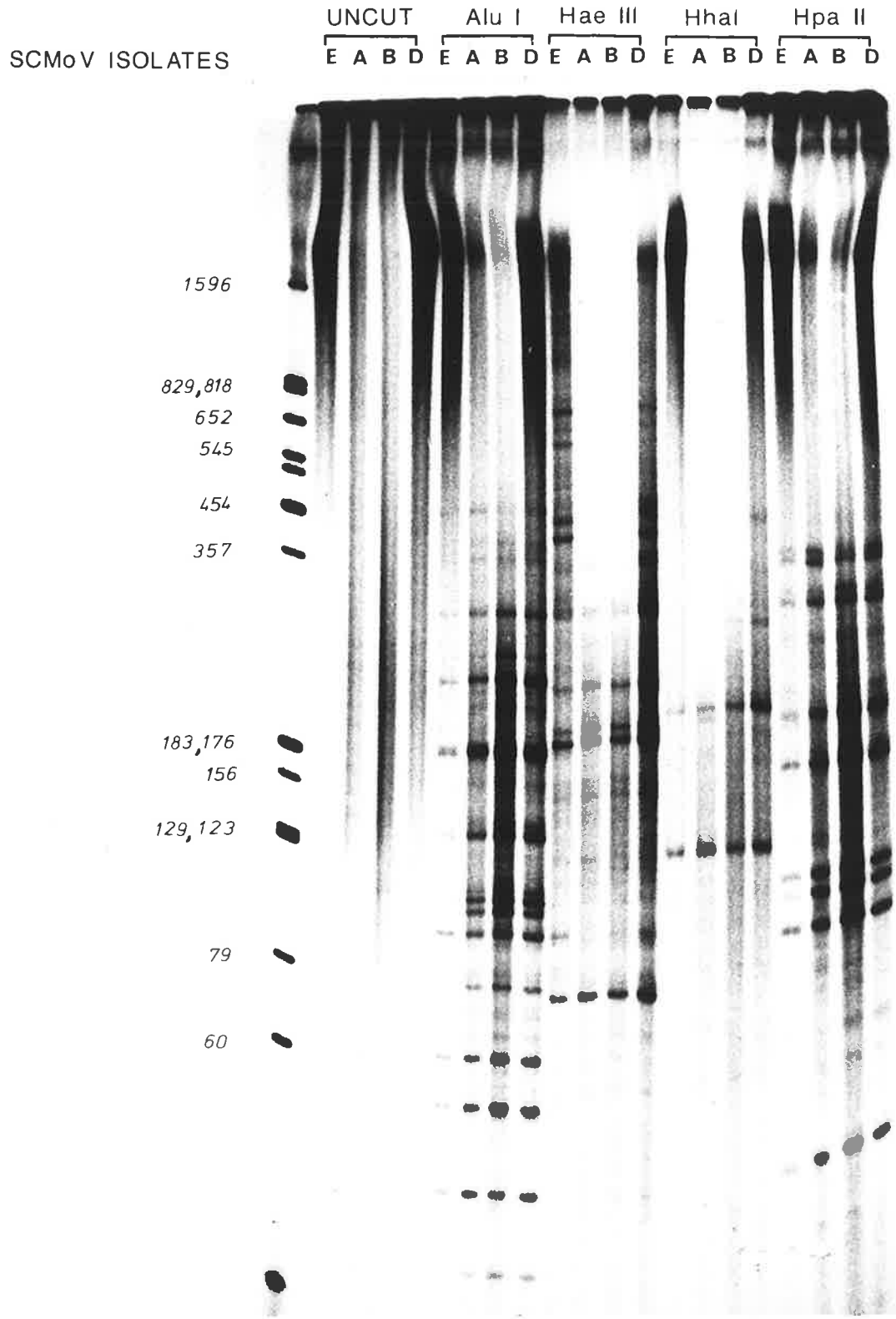
8.3 RESULTS

The four SCMoV isolates described by Francki *et al.* (1983b) i.e isolates A, B, D, and E were investigated in this study. SCMoV-A contains both s(388)SCMoV and s(332)SCMoV RNAs with the helper virus RNA. In contrast, SCMoV-E contains only the helper virus RNA and s(388)SCMoV RNAs while SCMoV-B and SCMoV-D contain only the helper virus and s(332)SCMoV RNAs. Interestingly, while isolates A, D and E were obtained from the field, SCMoV-B was produced from SCMoV-A by passage through single lesions on pea leaves. Both the helper virus and satellite RNAs were analyzed for all four isolates.

8.3.1 Analysis of SCMoV viral RNA sequences

In order to analyze the sequence relationships between the viral RNAs of the four isolates, double-stranded cDNAs were transcribed from purified RNAs using random oligonucleotide primers, and then digested with various restriction endonucleases, i.e. Alu , Hae III, Hha I, Hpa II and Acc I. Polyacrylamide gel electrophoresis of digested double-stranded cDNAs resulted in gel patterns which reflected the arrangement of restriction endonuclease recognition sequences on the original RNA molecules. As shown in Figure 8.1, all four isolated viral RNA species gave rise to indistinguishable patterns of double-stranded restriction endonuclease

Fig. 8.1 Restriction endonuclease digests of SCMoV viral RNA ds cDNAs. Double-stranded cDNA was prepared from purified viral RNAs from four SCMoV isolates i.e. SCMoV-E, (E); SCMoV-A (A); SCMoV-B (B) and SCMoV-D (D). The ds cDNAs were not digested or were digested with Alu I, Hae III, Hha I or Hpa II and fractionated on a 6% polyacrylamide, 2M urea gel. The sizes of ^{32}P-labelled Hpa II digested M13mp8 RF DNA (prepared by the method used for Hpa II digested pUC19 DNA markers (2.2.9)) are indicated in the left-hand margin.

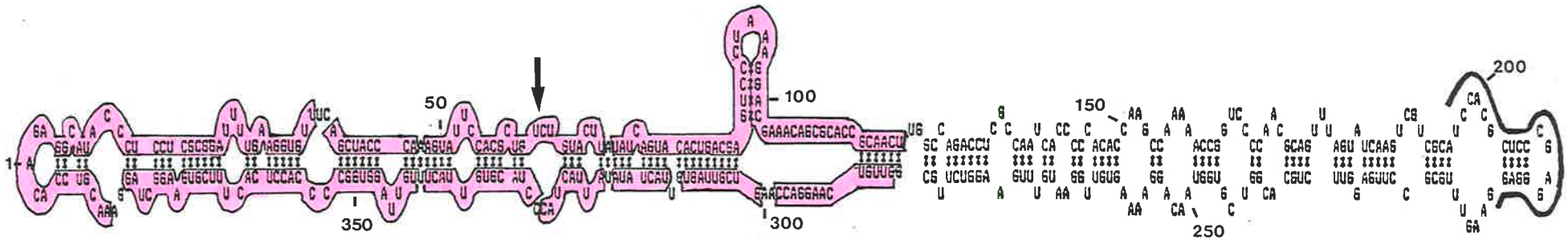

SCMoV ISOLATES
UNCUT
Alu I
Hae III
HhaI
Hpa II
E A B D E A B D E A B D E A B D E A B D
1596
829,818
652
545
454
357
183,176
156
129,123
79
60

fragments for the four nucleases shown. Similar results were obtained with Acc I (results not shown). While the synthesized RNAs may not be wholly representative of the respective RNA species and small nucleotide sequence differences between the RNAs may exist which do not affect the size or number of observed cDNA restriction endonuclease fragments, the results suggest that the four isolated viral RNA species are certainly closely related in nucleotide sequence, and may be identical.

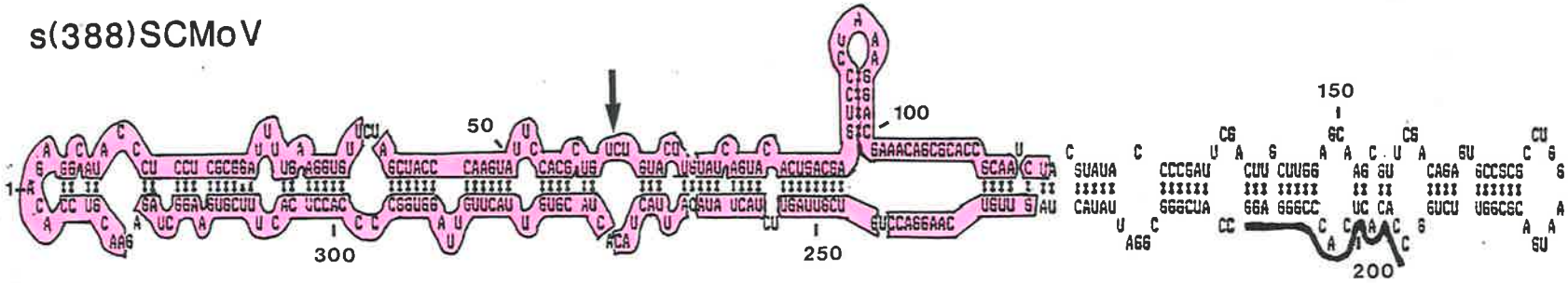
8.3.2 Sequence determination of s(388)SCMoV and s(332)SCMoV RNAs

The nucleotide sequences of the satellite RNAs from all four isolates of SCMoV were determined using both RNA and cDNA sequencing methods. No sequence differences were observed between the s(388)SCMoV RNAs obtained from SCMoV-A and SCMoV-E, or between the s(332)SCMoV RNAs obtained from SCMoV-A, SCMoV-B and SCMoV-D. The s(388)SCMoV and s(332)SCMoV RNA species are 388 and 332 nucleotides in size respectively. Secondary structure models for these molecules are shown in Figure 8.2. Both RNAs may base-pair intramolecularly to form helical rod-like structures similar to those of the other circular satellite RNAs and viroids. A number of alternative base-pairings of the sequences around the central region (containing the vertical hairpin structure) are possible, only one of which is shown. The two sSCMoV RNAs share a region of approximately 220 nucleotides of almost complete, contiguous sequence homology. Strikingly the conserved sequences are located so as to form the entire base-paired left-hand portions of the proposed native structures (Fig. 8.2). Consequently it is the different lengths of the largely nonconserved right-hand portions of these molecules which account for their difference in size. The right-hand portions of these molecules do contain some homologous sequences. For example, sequences from these regions can be aligned to give 14 out of 18 matches (shown by the lines above the sequence in Fig. 8.2) and both can be base-paired to form part of a stem-loop in which the

Fig. 8.2 Proposed secondary structures for s(388)SCMoV and s(332)SCMoV RNAs. Nucleotides conserved between both RNAs at similar positions in the secondary structures are boxed. The site of self-cleavage is indicated by an arrow. An 18 nucleotide sequence of which 14 nucleotides are conserved in both RNAs is indicated by a thick line. These sequences may form similar local secondary structures (Fig. 8.3).



s(388)SCMoV



s(322)SCMoV

conserved sequences are in an identical position (most of them being unpaired) (Fig. 8.3).

8.3.3 Self-cleavage of s(332)SCMoV RNA

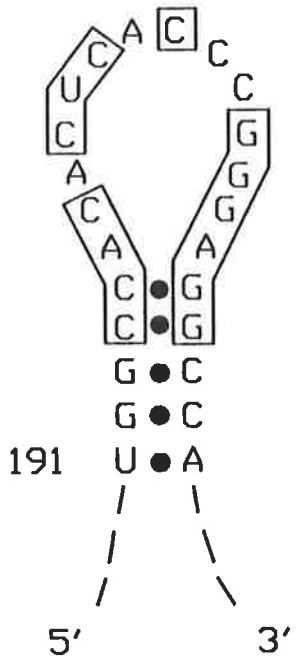
Linearized pGEM clones containing s(332)SCMoV cDNA inserts of both partial- and greater-than-unit-length were transcribed to give both plus- and minus-sense RNAs. The templates used for these transcriptions and the expected products are given in Figure 8.4. The plus but not the minus RNAs underwent self-cleavage during transcription (Fig. 8.5). The fragments produced by self-cleavage were of sizes commensurate with the site of self-cleavage predicted from previously proposed secondary structure models of other self-cleaving RNAs (Hutchins *et al.*, 1986; Forster and Symons, 1987a; also see Discussion). Enzymic sequencing of the partial-length plus-sense RNA transcript showed that the phosphodiester linkage between nucleotides C 63 and U 64 had been cleaved. Cleavage during transcription was so extensive that little or none of the uncleaved transcripts were visible (Fig. 8.5, lanes 3 and 5).

8.3.4 Blot hybridization analysis of sSCMoV RNAs

SCMoV RNAs extracted from both virions and whole plant tissues were subjected to blot hybridization analysis using probes for both plus and minus sequences. Little or no hybridizing material was detected in any of the tracks in which nucleic acids from uninfected tissue had been run (Fig. 8.6, lanes 1,3,5,7). Probing of RNAs extracted from purified virions for plus-sense sSCMoV sequences revealed the presence of an oligomeric series of bands (Fig. 8.6, lane 2). In long exposures a band corresponding to seven times the monomeric RNA unit length was visible. In RNAs extracted directly from whole tissue only monomeric and a lesser amount of dimeric plus RNAs were discernible (Fig. 8.6, lane 6).

Fig. 8.3 Secondary structure models for part of the s(332)SCMoV and s(388)SCMoV RNAs. These sequences are from the largely non-conserved right-hand portions of the proposed secondary structures (Fig. 8.2). Boxed residues indicate sequences conserved in the two RNAs.

s(332)SCMoV



s(388)SCMoV

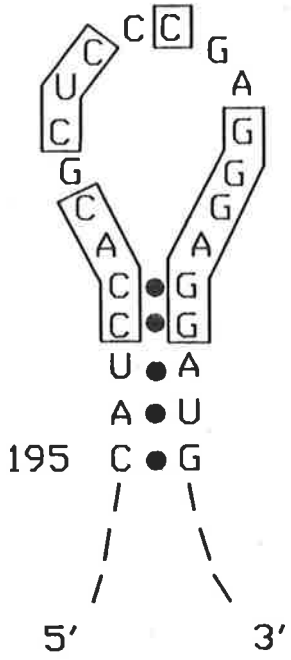


Fig. 8.4 Full- and partial-length s(332)SCMoV templates (8.2.4) and the RNAs resulting from their transcription with either SP6 or T7 RNA polymerases. (a), Full-length s(332)SCMoV pGem-2 template and the plus RNAs produced by SP6 RNA polymerase transcription. (b), Full-length s(332)SCMoV pGem-2 template and the minus RNA produced by T7 RNA polymerase transcription. (c), Partial-length s(332)SCMoV pGem-1 template (containing the Alu I/Taq I fragment from the full-length s(332)SCMoV clone) and the plus RNAs produced by T7 RNA polymerase transcription. (d), Partial-length s(332)SCMoV pGem-1 template as in (c) but linearized with EcoR I and transcribed with SP6 RNA polymerase to produce a minus RNA transcript. Hatched boxes, vector sequences; open boxes, sSCMoV sequences; closed boxes, SP6 and T7 RNA polymerase promoters. The self-cleavage sites are indicated by thick arrows. For the plus RNAs the complete transcripts are labelled 5'F/3'F and the self-cleavage products 5'F and 3'F. For minus RNAs the complete transcripts are labelled c and the minus RNAs are numbered according to the plus sequence.

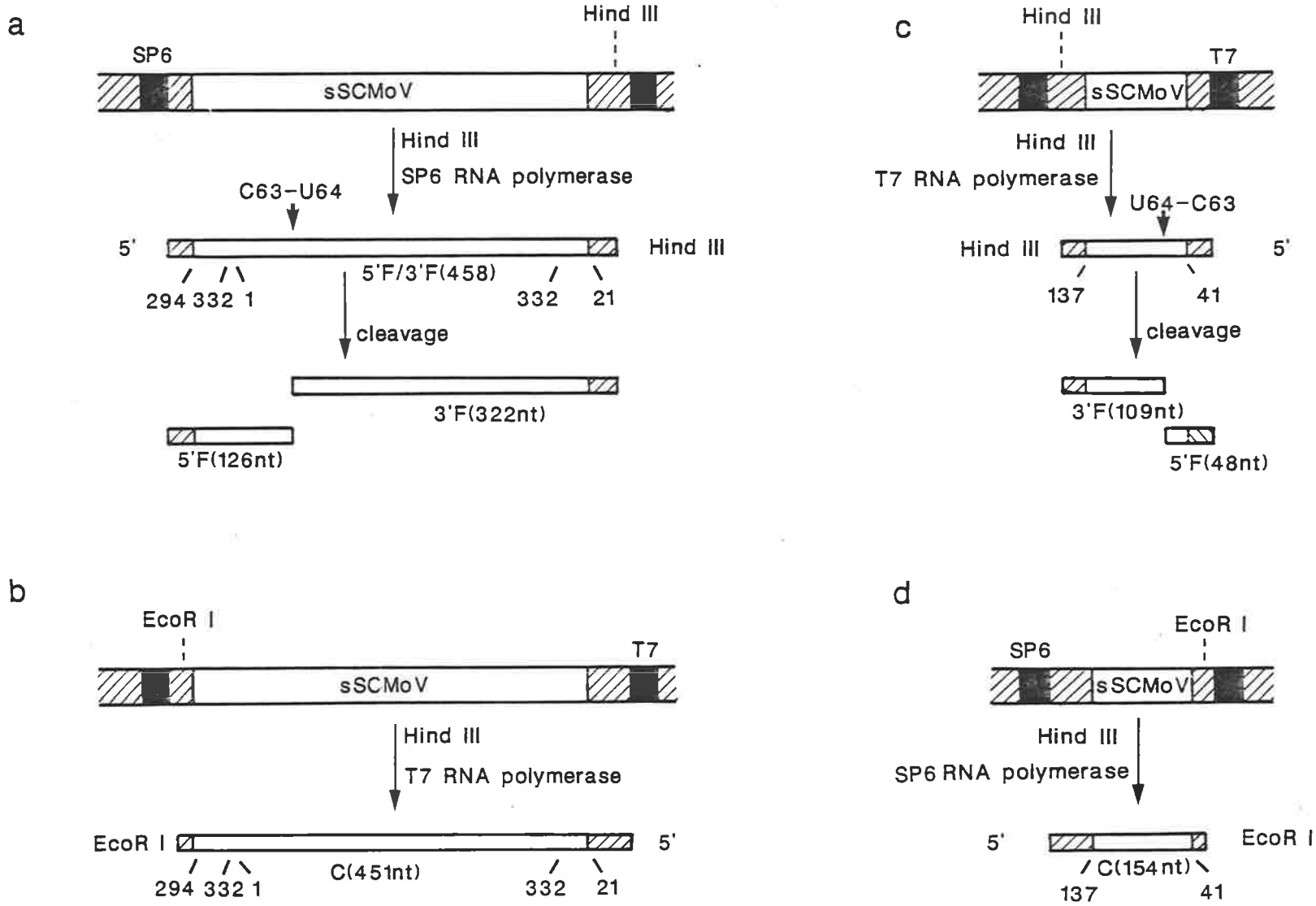
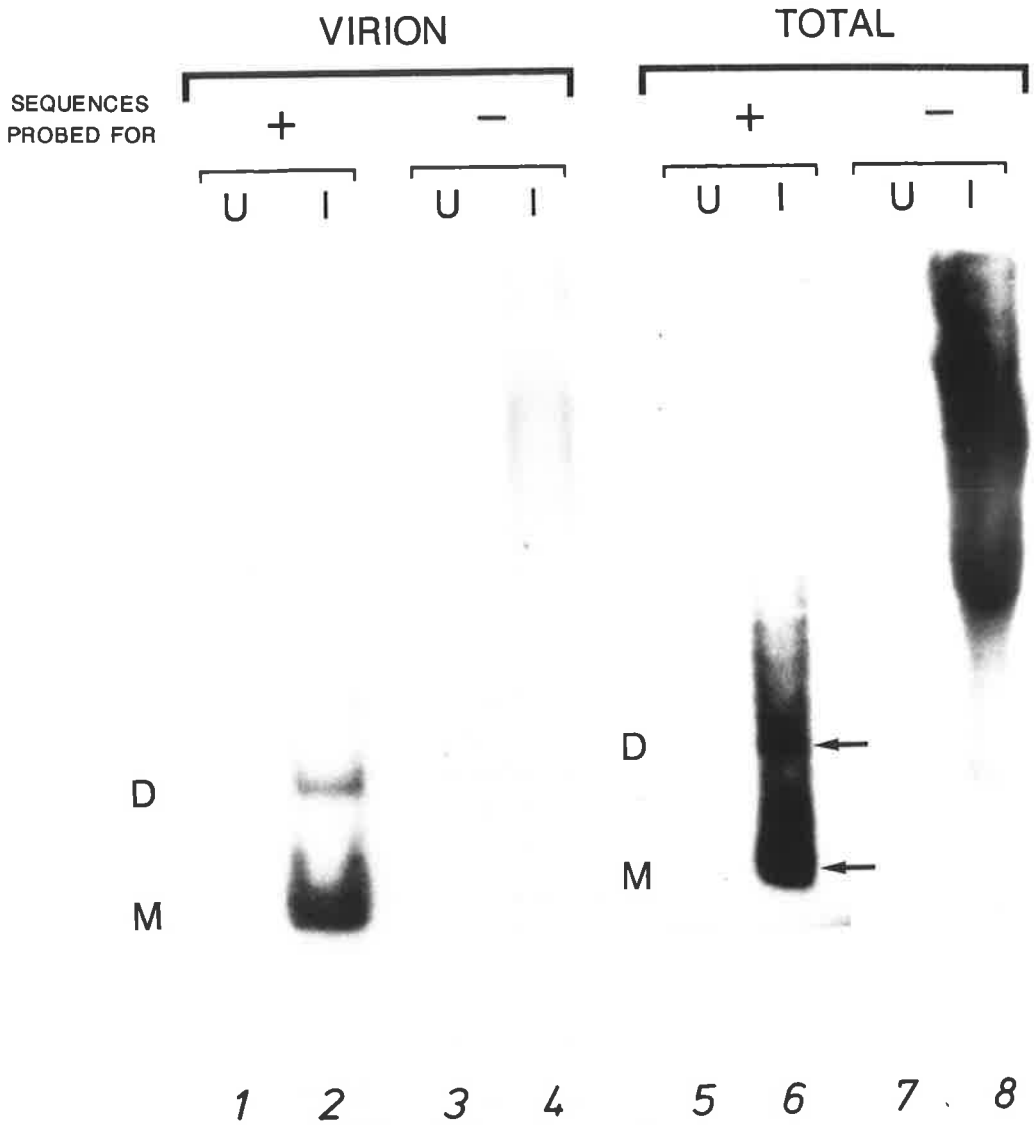

a
Hind III
SP6
sSCMoV
Hind III
SP6 RNA polymerase
C63–U64
5'
Hind III
5'F/3'F(458)
294 332 1
332 21
cleavage
3'F(322nt)
5'F(126nt)
c
Hind III
T7
sSCMoV
Hind III
T7 RNA polymerase
U64–C63
Hind III
5'
137
41
cleavage
3'F(109nt)
5'F(48nt)
b
EcoR I
T7
sSCMoV
Hind III
T7 RNA polymerase
EcoR I
5'
C(451nt)
294 332 1
332 21
d
EcoR I
SP6
sSCMoV
Hind III
SP6 RNA polymerase
5'
EcoR I
C(154nt)
137
41

Fig. 8.5 Autoradiograph of 5% polyacrylamide, 7M urea gel showing plus and minus RNA transcripts of full- and partial-length s(332)SCMoV templates. Lane 1; ^{32}P-labelled Hpa II digested pUC19 DNA markers, sizes in nucleotides are given at the left-hand margin. Lane 2; Minus RNA transcript of full-length s(332)SCMoV template (Fig. 8.4b). Lane 3; Plus RNA transcripts of full-length s(332)SCMoV template (Fig. 8.4a). Lane 4; Minus RNA transcript of partial-length s(332)SCMoV template (Fig. 8.4d). Lane 5; Plus RNA transcript of partial-length s(332)SCMoV template (Fig. 8.4c).


FULL
PARTIAL
− + − +
501,489
404
331
242
190
147
111,110
67
34
C
3'F
C
5'F
3'F
5'F
1 2 3 4 5

Fig. 8.6 Blot hybridization analysis of s(332)SCMoV RNAs. Nucleic acids from uninfected (U) and infected (I) Pisum sativum L. tissue were extracted either directly (total) or from purified virus (virion) (8.2.1). These nucleic acids were analyzed by blot hybridization (8.2.6) using ^{32}P-labelled RNA probes for the detection of plus or minus s(332)SCMoV sequences. Bands corresponding to monomeric (M) and dimeric (D) RNAs are indicated in the margins. Note that the monomeric and dimeric plus RNAs from the total nucleic acids extract (lane 6) appear to have higher molecular weights that those present in purified virus extracts (lane 2). This is because the samples from total extracts contained viscous materials that slowed migration of nucleic acids (and marker dyes). Autoradiography was carried out as follows: Lanes 1 and 2, 20 minutes; Lanes 5 and 6, 2 hr., -80°C, with intensifying screen; Lanes 3, 4, 7 and 8, 18 hr., -80°C, with intensifying screen.


VIRION
TOTAL
SEQUENCES PROBED FOR
+
−
+
−
U
I
U
I
U
I
U
I
D
M
D
M
1
2
3
4
5
6
7
8

Minus RNAs were present at much lower levels than plus RNAs. An extremely low level of high molecular weight material was detected in RNA extracted from purified virions probed for minus RNA sequences (Fig. 8.6, lane 4). Greater amounts of high molecular weight material were detected in total plant nucleic acid extracts (Fig. 8.6, lane 8). Most of this hybridizing material was contained in a smear above an ill-defined band of approximately 1.3 Kb. Monomer- or dimer-sized bands were not detected but a faint band was present whose size was between that of the monomeric and dimeric plus RNAs.

8.4 DISCUSSION

8.4.1 Relationship between the SCMoV isolates

The four SCMoV isolates used in this work have been shown to be indistinguishable by serology electron microscopy (Francki *et al.*, 1983), and by restriction endonuclease cleavage of cDNAs made to the helper virus RNA species (Fig. 8.1). The virus isolates therefore apparently differ only in containing one or both of the satellite RNA species. No variation within s(388)SCMoV or s(332)SCMoV species was detected between isolates. It therefore seems likely that these satellite RNAs must be functionally equivalent.

8.4.2 Sequence homology between s(388)SCMoV and s(332)SCMoV

The predicted secondary structures of the two sSCMoV RNAs show remarkable conservation of the left-hand portions of the molecules (Fig. 8.2). Presumably this pattern of homology has arisen from RNA recombination. Such intermolecular rearrangements have been proposed for viroids (Keese and Symons, 1985), a number of other infectious plant RNAs (Haseloff *et al.*, 1984; Bujarski and Kaesberg, 1986; Simon and Howell, 1986; Buzayan *et al.*, 1987; Robinson *et al.*,

1987; Kaper *et al.*, 1988a) and in plant viral defective interfering RNAs (Anzola *et al.*, 1987; Hillman *et al.*, 1987). It is probable that the common sequences and structures in the sSCMoV RNAs mirror the apparent interchangeable functions of these molecules. Thus the conserved left-hand portions may contain those sequences required for replication by helper virus or plant host components. Some of the conserved sequences contained in this region (Fig. 8.2) were shown to be responsible for the *in vitro* self-cleavage of s(332)SCMoV RNAs (Fig. 8.5), a reaction thought to be an integral part of the replication cycle. This will be discussed in detail below. The constraints on the largely nonconserved right-hand end sequences may only be general, such as, they must be able to form a base-paired structure and must be within certain size limits. However the limited homologies present in this area may indicate the presence of signals common to both RNAs. The sequencing of further sSCMoV isolates with a wider geographical distribution may be a worthwhile exercise as it would be interesting to see if other variations to the right-hand regions are found or if only these two alternatives exist.

Like the sSCMoV RNAs, the nucleotide sequences of the satellites of TRSV (Buzayan *et al.*, 1986c) and arabis mosaic virus (ArMV) (Kaper *et al.*, 1988a) share considerable homology, albeit not as extensive as that observed between the sSCMoV RNAs. Also like the sSCMoV RNAs these homologous regions include those sequences responsible for plus RNA self-cleavage. It is also probable that at least some of the sequences conserved between sTRSV and sArMV are required for minus RNA self-cleavage. However, although having extensive regions of sequence in common, sTRSV and sArMV are apparently not able to share helper viruses (Kaper *et al.*, 1988a) and therefore the homolgous regions contain at best only part of the sequences required for replication. This is in contrast to the sSCMoV RNAs where the homologous regions probably contain all of the sequences required for replication.

8.4.3 Sequence homology between sSCMoV, sVTMoV and sLTSV

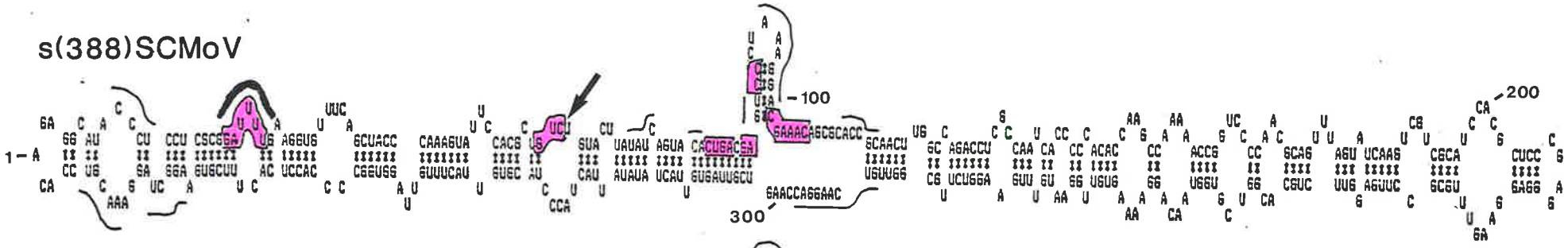
The sequences and predicted secondary structures of sSCMoV, sVTMoV (Haseloff and Symons, 1982) and sLTSV-A (Keese *et al.*, 1983) are shown in Figure 8.7, and common sequences indicated. sSNMV is not shown due to its high degree of sequence homology with sVTMoV. There are a number of notable features. First, the sequence GAUUUU is present in all RNAs in a similar position on the secondary structures (approximately located at residues 20-25 in all cases and indicated by a thick line in Fig. 8.7). the conservation of this sequence suggests that it may play some role which is common to the replication of these RNAs. Second, sSCMoV and sVTMoV RNAs share conserved sequences (indicated by thin lines in Fig. 8.7) in addition to those directly involved in self-cleavage. Some of these sequences occur at the left-hand end of the base-paired secondary structures while others occur in the central region (to the right of those sequences involved in self-cleavage). These sequences may be involved in functions specific to these RNAs however they are unlikely to be essential for replication as LTSV can support the replication of sLTSV, sSCMoV and sSNMV (Jones and Mayo, 1983; Jones *et al.*, 1983; Keese *et al.*, 1983). Third, there are sequences conserved between all the RNAs presented (boxed in Fig. 8.7) that have been shown to be or proposed to be involved in the self-cleavage of these RNAs.

8.4.4 sSCMoV RNA self-cleavage and replication

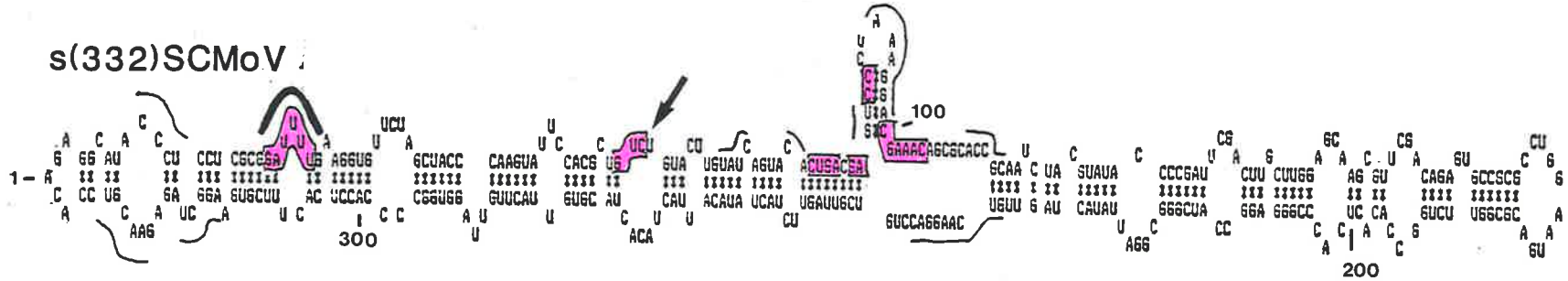
Plus-sense transcripts from s(332)SCMoV clones were shown to self-cleave in a site-specific manner during transcription *in vitro*. The RNA containing only a subset of s(332)SCMoV sequences cleaved as efficiently as the RNA containing the entire sequence, demonstrating that this sequence alone is sufficient for cleavage. This region in s(332)SCMoV contains sequences conserved between the

Fig. 8.7 Sequence and structural homology between s(332)SMCoV, s(388)SCMoV, sVTMoV and sLTSV-A RNAs. Due to the high degree of homology between sSNMV and sVTMoV RNAs only the sVTMoV sequence has been presented. Residues common between all four RNAs are boxed (these include those sequences that make up the hammerhead self-cleavage structures). The conserved GAUUUU sequences are highlighted by thick lines. Sequences conserved between sSCMoV and sVTMoV RNAs only, are indicated by thin lines (some of these sequences are those involved in formation of the hammerhead structures but are not conserved in sLTSV). The self-cleavage sites are indicated by arrows.

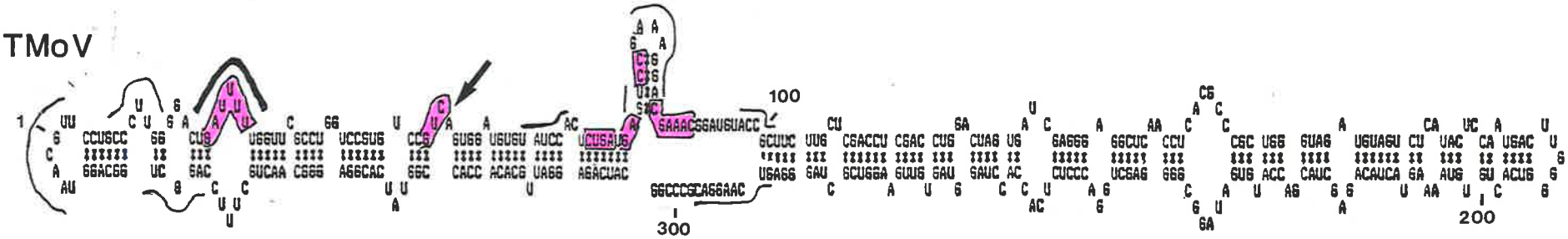
s(388)SCMoV



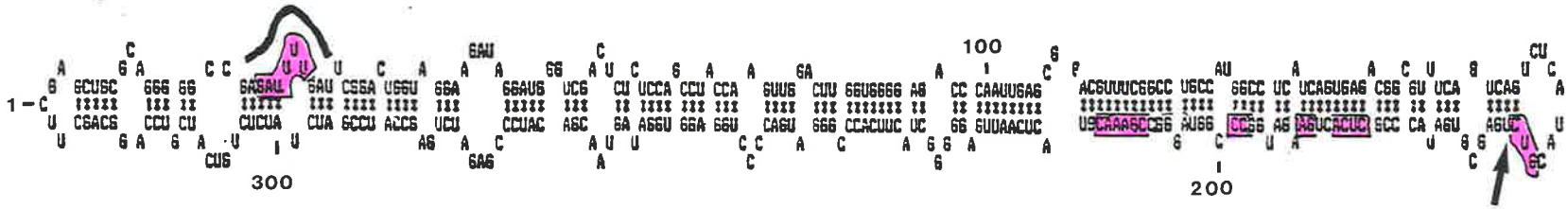
s(332)SCMoV



sVTMoV



sLTSV-A

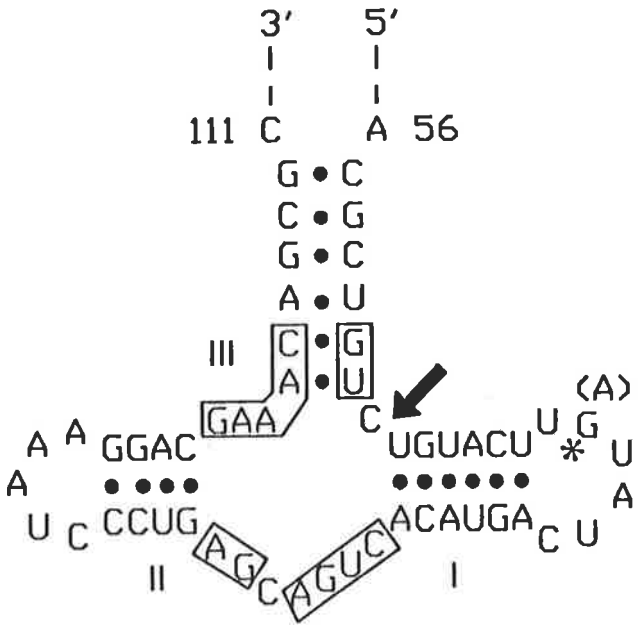


s(332)SCMoV and s(388)SCMoV RNAs that can be folded into secondary structures (Fig. 8.8) similar to those proposed for the active self-cleavage structures of other RNAs (Hutchins *et al.*, 1986; Forster and Symons, 1987a). The site of cleavage predicted from this structure agrees with the demonstrated site of cleavage. The high efficiency of cleavage of this RNA may be related to the very strong base-pairing in stem III (Fig. 8.8). No cleavage of the minus RNAs was observed under the conditions used and the minus RNAs do not contain a self-cleavage domain like that found in plus sSCMoV RNAs and the RNAs of other self-cleaving RNAs (including minus sTSRV RNA) (Fig. 4.5).

The cleavage of only plus sSCMoV RNAs resembles the situation found in sVTMoV (and sSNMV) where the plus but not the minus sequences contain hammerhead self-cleavage domains (a small RNA containing the plus sVTMoV self-cleavage domain has been shown to cleave *in vitro* (S. McNamara, unpublished data)). It would therefore seem likely that the minus sSCMoV RNAs do not self-cleave, unless there is an as yet undiscovered self-cleavage mechanism that operates under conditions different to those used in this work. This proposed lack of cleavage is supported by the finding that the predominant minus sVTMoV, sSNMV and sSCMoV RNAs detected in nucleic acids extracted from virions or infected plants are high molecular weight forms (Chu *et al.*, 1983; Hutchins *et al.*, 1986; this work).

The detection of minus RNAs is made difficult by the preponderance of plus RNAs in these samples. The problems associated with the detection of minus RNAs has been adequately discussed by other workers (Branch and Roberston, 1984; Hutchins *et al.*, 1985). Forster and Symons (1987a) have proposed a modified rolling circle replication model for sVTMoV and sSNMV RNAs (Fig. 1.1) which accounts for the apparent lack of cleavage in minus-sense RNAs. In this model, transcription from the monomeric plus-sense circular RNA template gives rise to a greater-than-unit-length minus RNA that does not undergo cleavage. This RNA then acts as a

Fig. 8.8 Proposed single-hammerhead self-cleavage structure for plus s(332)SCMoV RNA. Boxed resiudes are those conserved in all RNAs that cleave by the hammerhead model (see Fig. 4.5). The site of self-cleavage is indicated by an arrow. Stems are numbered 1-111. The asterisked nucleotide is an G in s(332)SCMoV but a A in s(388)SCMoV.


3′
5′
111
C
A
56
G • C
C • G
G • C
A • U
III
C • G
A • U
GAA
C
(A)
A A GGAC
UGUACU U G
*
U
A
A • • • •
• • • • • •
A
U C CCUG
ACAUGA C U
II
AGC
AGUC
I

s(322)SCMoV +

template for plus RNA synthesis. This plus RNA is then cleaved and circularized, thus completing the cycle. In the light of the results presented here this model could also be applied to sSCMoV RNA replication.

REFERENCES

REFERENCES

Ahlquist, P., Bujarski, J.J., Kaesberg, P. and Hall, T.C. (1984). *Plant Mol. Biol.* **3**, 37-44.

Ahlquist, P., Dasgupta, R. and Kaesberg, P. (1981a). *Cell*, **23**, 183-189.

Ahlquist, P., Dasgupta, R. and Kaesberg, P. (1984). *J. Mol. Biol.* **172**, 369-383.

Ahlquist, P., Luckow, V. and Kaesberg, P. (1981b). *J. Mol. Biol.* **153**, 23-38.

Anzola, J.V., Xu, Z., Asmaizu, T. and Nuss, D.L. (1987). *Proc. Natl. Acad. Sci. USA* **84**, 8301-8305.

Atabekov, J.G. and Dorokhov, Yu. L. (1984). *Adv. Virus Res.* **29**, 313-364.

Bar-Joseph, M., Rosner, A., Moscovitz, M. and Hull, R. (1983). *J. Virol. Methods* 6, 1-8.

Barker, J.M., McInnes, J.L., Murphy, P.J. and Symons, R.H. (1985). *J. Virol Meth.* **10**, 87-98.

Birnboim, H.C. and Doly, J. (1979). *Nucleic Acids Res.* **7**, 1513-1523.

Branch, A.D. and Robertson, H.D. (1984). *Science* **223**, 450-455.

Branch, A.D., Robertson, H.D. and Dickson, E. (1981). *Proc. Natl. Acad. Sci. USA* **78**, 6381-6385.

Brederode, F. Th., Koper-Zwarthoff, E.C. and Bol, J.F. (1980). *Nucleic Acids Res.* 8, 2213-2223.

Bruening, G., Gould, A.R., Murphy, P.J. and Symons, R.H. (1982). *FEBS Lett.* **148**, 71-78.

Bujarski, J.J., Ahlquist, P., Hall, T.C., Dreher, T.W. and Kaesberg, P. (1986). *The EMBO J.* **5**, 1769-1774.

Bujarski, J.J., Dreher, T.W. and Hall, T.C. (1985). *Proc. Natl. Acad. Sci. USA* **82**, 5636-5640.

Bujarski, J.J. and Kaesberg, P. (1986). *Nature* **321**, 528-531.

Buzayan, J.M., Gerlach, W.L. and Bruening, G. (1986a). *Nature* **323**, 349-353.

Buzayan, J.M., Gerlach, W.L. and Bruening, G. (1986b). *Proc. Natl. Acad. Sci. USA* **83**, 8859-8862.

Buzayan, J.M., Gerlach, W.L., Bruening, G., Keese, P.K. and Gould, A.R. (1986c). *Virology* **151**, 186-199.

Buzayan, J.M., Hampel, A. and Bruening, G. (1986d). *Nucleic Acds Res.* **14**, 9729-9743.

Buzayan, J.M., McNinch, J.S., Schneider, J.R. and Breuning, G. (1987). *Virology* **160**, 95-99.

Chu, P.W.G., Francki, R.I.B. and Randles, J.W. (1983). *Virology* **126**, 480-492.

Collmer, C.W. and Kaper, J.M. (1985). *Virology* **145**, 249-259.

Collmer, C.W. and Kaper, J.M. (1988). *Virology* **163**, 293-298.

Cornelissen, B.J.C. and Bol, J.F. (1984). *Plant Mol. Biol.* **3**, 379-384.

Cornelissen, J.C., Janssen, H., Zuidema, D. and Bol, J.F. (1984). *Nucleic Acids Res.* **12**, 2427-2437.

Cuozzo, M., O'Connell, K.M., Kaniewski, W., Fang, R., Chua, N. and Turner, N.E. (1988). *Biotechnology* **6**, 549-557.

Dasgupta, R. and Kaesberg, P. (1982). *Nucleic Acids Res.* **10**, 703-713.

Dawson, W.O., Bubrick, P. and Grantham, G.L. (1988). *Phytopathol.* **78**, 783-789.

Dayhoff, M.O., Eck, R.V. and Park, C.M. (1969). A model of evolutionary change in proteins. In *"Atlas of Protein Sequence and Structure, 1969"* (M.O. Dayhoff, ed.) pp. 75-83. National Biomedical Research Foundation, Silver Spring.

Deom, C.M., Oliver, M.J. and Beachy, R.N. (1987). *Science* **237**, 389-394.

Diener, T.O. (1983). *Adv. Virus Res.* **28**, 241-283.

Diener, T.O., ed. (1987). *"The Viroids"* Plenum Press, New York.

Donis-Keller, H. (1980) *Nucleic Acids Res.* **8**, 3133-3142.

Dreher, J.W., Bujarski, J.J. and Hall, T.C. (1984). *Nature* **311**, 171-175.

Dreher, T.W. and Hall, T.C. (1988). *J. Mol. Biol.* **201**, 31-40.

Epstein, L.M. and Gall, J.G. (1987). *Cell* **48**, 535-543.

Forster, A.C., Davies, C., Sheldon, C.C., Jeffries, A.C. and Symons, R.H. (1988). *Nature* **334**, 265-267.

Forster, A.C., Jeffries, A.C., Sheldon, C.C. and Symons, R.H. (1987). *Cold Spring Harbor Symp. Quant. Biol.* **52**, 249-259.

Forster, A.C. and Symons, R.H. (1987a). *Cell* **49**, 211-220.

Forster, A.C. and Symons, R.H. (1987b). *Cell* **50**, 9-16.

Francki, R.I.B. (1985a). *Ann. Rev. Microbiol.* **39**, 151-174.

Francki, R.I.B. (1985b). The viruses and their taxonomy. In *"The Plant Viruses"*, Vol. 1 (R.I.B. Francki, ed.) pp. 1-18, Plenum Press, New York.

Francki, R.I.B. (1987). Encapsidated viroid-like RNA. In *"The Viroids"* (T.O. Diener, ed.) pp. 205-218, Plenum Press, New York.

Francki, R.I.B., Chu, P.W.G. and Keese, P.K. (1983a). The satellite nature of a viroid-like RNA from lucerne transient streak virus. In *"A Current Communications in Molecular Biology: Plant Infectious Agents, Viruses, Viroids, Virusoids and Satellites"* (H.D. Robertson, S.H. Howell, M. Zaitlin and R. Malmberg, eds.) pp. 175-180, Cold Spring Harbor Laboratory.

Francki, R.I.B., Grivell, C.J. and Gibb, K.S. (1986). *Virology* **148**, 381-384.

Francki, R.I.B., Randles, J.W., Chambers, T.C. and Wilson, S.B. (1966). *Virology* **28**, 729-741.

Francki, R.I.B., Randles, J.W., Chu, P.W.G., Rohozinski, J. and Hatta, T. (1985). Viroid-like RNAs incorporated in conventional virus capsids. In *"Subviral Pathogens of Plants and Animals: Viroids and Prions"* (K. Maramorosch and J.J. McKelvey, eds.), pp. 265-297, Academic Press, Orlando, Florida.

Francki, R.I.B., Randles, J.W., Hatta, T., Davies, C. and Chu, P.W.G. (1983b). *Plant Pathology* **32**, 47-59.

Freier, S.M., Kierzek, R., Jaeger, J.A., Sugimoto, N., Caruthers, M.H., Neilson, T. and Turner, D.H. (1986). *Proc. Natl. Acad. Sci. USA* **83**, 9373-9377.

French, R. and Ahlquist, P. (1987). *J. Virol.* **61**, 1457-1465.

French, R. and Ahlquist, P. (1988). *J. Virol.* **62**, 2411-2420.

Garcia-Arenal, F., Zaitlin, M. and Palukaitis, P. (1987). *Virology* **158**, 339-347.

Gerlach, W.L., Llewellyn, D. and Haseloff, J. (1987). *Nature* **328**, 802-805.

Godefroy-Colburn, T., Gagey, M.J., Berna, A. and Stussi-Garaud, C. (1986). *J. Gen. Virol.* **67**, 2233-2239.

Goldbach, R. and Wellink, J. (1988). *Intervirology* **29**, 260-267.

Gordon, K.H.J. and Symons, R.H. (1983). *Nucleic Acids Res.* **11**, 947-960.

Gould, A.R. (1981). *Virology* **108**, 123-133.

Gould, A.R., Francki, R.I.B. and Randles, J.W. (1981). *Virology* **110**, 420-426.

Gould, A.R. and Hatta, T. (1981). *Virology* **109**, 137-147.

Gould, A.R. and Symons, R.H. (1977). *Nucleic Acids Res.* **4**, 3787-3802.

Gould, A.R. and Symons, R.H. (1982). *Eur. J. Biochem.* **126**, 217-226.

Hanahan, D. (1984). Techniques for transformation of *E. coli* . In *"DNA Cloning, Vol. 1, A Practical Approach"* (D.M. Glover, ed.) pp. 109-135, IRL Press, Oxford.

Harrison, S.C. (1984). *Trends Biochem. Sci.* **9**, 345-351.

Harrison, B.D. and Murant, A.F. (1977). CMI/AAB "Descriptions of Plant Viruses" **185**, 4 pp.

Haseloff, J. and Gerlach, W.L. (1988). *Nature* **334**, 585-591.

Haseloff, J., Goelet, P., Zimmern, D., Ahlquist, P., Dasgupta, R. and Kaesberg, P. (1984). *Proc. Natl. Acad. Sci. USA* **81**, 4358-4362.

Haseloff, J., Mohamed, N.A. and Symons, R.H. (1982). *Nature* **299**, 316-321.

Haseloff, J. and Symons, R.H. (1981). *Nucleic Acids Res.* **9**, 2741-2752.

Haseloff, J. and Symons, R.H. (1982). *Nucleic Acids Res.* **10**, 3681-3691.

Hashimoto, J. and Koganezawa, H. (1987). *Nucleic Acids Res.* **15**, 7045-7052.

Hillman, B.I., Carrington, J.C. and Morris, T.J. (1987). *Cell* **51**, 427-433.

Houwing, C.J. and Jaspars, E.M.J. (1978). *Biochemistry* **17**, 2927-2933.

Houwing, C.J. and Jaspars, E.M.J. (1986). *FEBS Lett.* **209**, 284-288.

Huisman, M.J., Sarachu, A.N., Ablas, F., Broxterman, H.J.G., Van Vloten-Doting, L. and Bol. J.F. (1986). *Virology* **154**, 401-404.

Hull, R. (1988). The sobemovirus group. In *"The Plant Viruses"*, Vol. 3 (R. Koenig, ed.) pp. 113-146, Plenum Press, New York.

Hull, R. and Maule, A.J. (1985). Virus multiplication. In *"The Plant Viruses"*, Vol. 1 (Francki, R.I.B., ed.) pp. 83-115, Plenum Press, New York.

Hutchins, C.J., Keese, P.K., Visvader, J.E., Rathjen, P.D., McInnes, J.L. and Symons, R.H. (1985). *Plant Mol. Biol.* **4**, 293-304.

Hutchins, C.J., Rathjen, P.D., Forster, A.C. and Symons, R.H. (1986). *Nucleic Acids Res.* **14**, 3627-3640.

Ishikawa, M., Meshi, T., Ohno, T., Okada, Y., Sano, T., Veda, I. and Shikata, E. (1984). *Mol. Gen. Genet.* **196**, 421-428.

Jones, A.T., Mayo, M.A. and Duncan, G.H. (1983). *J. Gen. Virol.* **64**, 1167-1173.

Jones, A.T. and Mayo, M.A. (1983). *J. Gen. Virol.* **64**, 1771-1774.

Jones, A.T. and Mayo, M.A. (1984). *J. Gen. Virol.* **65**, 1713-1721.

Kanehisa, M.I. (1982). *Nucleic Acids Res.* **10**, 183-196.

Kaper, J.M. and Diaz-Ruiz, J.R. (1977). *Virology* **80**, 214-217.

Kaper, J.M., Tousignant, M.E. and Steen, M.T. (1988a). *Virology* **163**, 284-292.

Kaper, J.M., Tousignant, M.E. and Steger, G. (1988b). *Biochem. Biophys. Res. Comm.* **154**, 318-325.

Keese, P.K., Bruening, G. and Symons, R.H. (1983). *FEBS Lett.* **159**, 185-190.

Keese, P.K. and Symons, R.H. (1985). *Proc. Natl. Acad. Sci. USA* **82**, 4582-4586.

Keese, P.K. and Symons, R.H. (1987). The structure of viroids and virusoids. In *"Viroids and Viroid-like Pathogens"* (J.S. Semancik, ed.) pp. 1-47, CRC Press, Boca Raton, Florida.

Kiberstis, P.A., Haseloff, J. and Zimmern, D. (1985). *The EMBO J.* **4**, 817-822.

Kiefer, M.C., Daubert, S.D., Schneider, I.R. and Bruening, G. (1982). *Virology* **121**, 262-273.

Koizumi, M., Iwai, S. and Ohtsuka, E. (1988). *FEBS Lett.* **228**, 228-230.

Koltunow, A.M. and Rezaian, M.A. (1988). Submitted.

Kurath, G. and Palukaitis, P. (1987). *Virology* **159**, 199-208.

Kyte, T. and Doolittle, R.F. (1982). *J. Mol. Biol.* **157**, 105-132.

Langereis, K., Mugnier, M., Cornelissen, B.J.C., Pinck, L. and Bol, J.F. (1986). *Virology* **154**, 409-414.

Linhorst, H.J.M. and Kaper, J.M. (1984). *Virology* **137**, 206-210.

Maniatis, T.A., Fritsch, E.F. and Sambrook J. (1982). Molecular Cloning: A Laboratory Manual (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).

Marsh, L.E., Dreher, T.W. and Hall, T.C. (1987). *UCLA Symp. Mol. Cell. Biol.* **54**, 327-336.

Marsh, L.E. and Hall, T.C. (1987). *Cold Spring Harbor Symp. Quant. Biol.* **52**, 331-341.

McLachlan, A.D. (1971). *J. Mol. Biol.* **61**, 409-424.

McMaster, G.K. and Carmichael, G. (1977). *Proc. Natl. Acad. Sci. USA* **74**, 4835-4838.

Meshi, T., Watanabe, Y., Saito, T., Sugimoto, A., Maeda, T. and Okada, Y. (1987). *The EMBO J.* **6**, 2557-2563.

Messing, J. and Vieira, J. (1982). *Gene* **19**, 269-276.

Miller, W.A., Bujarski, J.J., Dreher, T.W. and Hall, T.C. (1986). *J. Mol. Biol.* **187**, 537-546.

Miller, W.A., Dreher, T.W. and Hall, T.C. (1985). *Nature* **313**, 68-70.

Miller, W.A. and Hall, T.C. (1983). *Virology* **125**, 236-241.

Moosic, J.P., McKean, D.J., Shih, D.S. and Kaesberg, P. (1983). *Virology* **129**, 517-520.

Moser, O., Gagey, M.J., Godefroy-Colburn, T., Stussi-Garaud, C., Ellwart-Tschurtz, M., Nitschko, H. and Mundry, K.W. (1988). *J. Gen. Virol.* **69**, 1367-1373.

Murant, A.J. and Mayo, M.A. (1982). *Ann. Rev. Phytopathol.* **20**, 49-70.

Murthy, M.R.N. (1983). *J. Mol. Biol.* **168**, 469-475.

Needleman, S.B. and Wunsch, C.D. (1970). *J. Mol. Biol.* **48**, 443-453.

Novotny, J. and Auffray, C. (1984). *Nucleic Acids Res.* **12**, 243-255.

Ohno, T., Takamatsu, N., Meshi, T., Okada, Y., Nishiguchi, M. and Kiho, Y. (1983). *Virology* **131**, 255-258.

Owens, R.A. and Diener, T.O. (1982). *Proc. Natl. Acad. Sci. USA* **79**, 113-117.

Paliwal, Y.C. (1984). *Can. J. Plant Pathol.* **6**, 93-97.

Peden, K.W.C. and Symons, R.H. (1973). *Virology* **53**, 487-492.

Prody, G.A., Bakos, J.T., Buzayan, J.M., Schneider, I.R. and Bruening, G. (1986). *Science* **231**, 1577-1580.

Puchta, H., Ramm, K. and Sanger, H.L. (1988). *Nucleic Acids Res.* **16**, 4197-4216.

Randles, J.W., Davies, C., Hatta, T., Gould, A.R. and Frankci, R.I.B. (1981). *Virology* **108**, 111-122.

Rezaian, M.A. and Krake, L.R. (1987). *J. Virol.* **17**, 277-285.

Rezaian, M.A. and Symons, R.H. (1986). *Nucleic Acids Res.* **14**, 3229-3239.

Rezaian, M.A., Williams, R.H.V., Gordon, K.H.J., Gould, A.R. and Symons, R.H. (1984). *Eur. J. Biochem.* **143**, 277-284.

Rezaian, M.A., Williams, R.H.V. and Symons, R.H. (1985). *Eur. J. Biochem.* **150**, 331-339.

Riesner, D. and Gross, H.J. (1985). *Ann. Rev. Biochem.* **54**, 531-564.

Rietveld, K., Pleij, C.W.A. and Bosch, L. (1983). *The EMBO J.* **2**, 1079-1085.

Rizzo, T.M. and Palukaitis, P. (1988). *J. Gen. Virol.* **69**, 1777-1787.

Rosner, A., Bar-Joseph, M., Moscovitz, M. and Mevarech, M. (1983). *Phytopathol.* **73**, 699-702.

Robinson, D.J., Hamilton, W.D.O., Harrison, B.D. and Baulcombe, D.C. (1987). *J. Gen. Virol.* **68**, 2551-2561.

Rossman, M.G., Abad-Zapatero, C., Murthy, M.R.N., Liljas, L., Jones, T.A. and Strandberg, B. (1983). *J. Mol. Biol.* **165**, 711-736.

Saito, T., Meshi, T., Takamatsu, N. and Okada, Y. (1987). *Proc. Natl. Acad. Sci. USA* **84**, 6074-6077.

Sampson, J.R., Sullivan, F.X., Behlen, L.S., Di Renzo, A.B. and Uhlenbeck, O.C. (1987). *Cold Spring Harbor Symp. Quant. Biol.* **52**, 267-275.

Sanger, H.L. (1984). Minimal infectious agents: The viroids. In *"The Microbe 1984: Part I Viruses"* (B.W.J. Mahy and J.R. Pattison, eds.) (Society for General Microbiology Symposium 36: Cambridge University Press), pp. 281-334.

Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. (1980). *J. Mol. Biol.* **143**, 161-178.

Savithri, H.S. and Murthy, M.R.N. (1983). *J. Biosci.* **5**, 183-187.

Schiffer, M. and Edmundson, A.B. (1967). *Biophys. J.* **7**, 121-135.

Senior, M.M., Jones, R.A. and Breslauer, K.J. (1988). *Proc. Natl. Acad. Sci. USA* **85**, 6242-6246.

Simon, A.E. and Howell, S.H. (1986). *The EMBO J.* **5**, 3423-3428.

Simon, A.E. (1988). *Plant Molecular Biology Reporter* **6**, 240-252.

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

Steger, G., Hofmann, H., Fortsch, J., Gross, H.J., Randles, J.W., Sanger, H.L. and Riesner, D. (1984). *J. Biomol. Struct. Dyn.***2**, 543-571.

Stussi-Garaud, C., Garaud, J., Berna, A. and Godefroy-Colburn, T. (1987). *J. Gen. Virol.* **68**, 1779-1784.

Symons, R.H. (1979). *Nucleic Acids Res.* **7**, 825-837.

Symons, R.H. (1981). *Nucleic Acids Res.* **9**, 6527-6537.

Symons, R.H. (1985). Viral Genome Structure. In *"The Plant Viruses"*, Vol. 1 (R.I.B. Francki, ed.) pp. 57-81, Plenum Press, New York.

Symons, R.H., Haseloff, J., Visvader, J.E., Keese, P.K., Murphy, P.J., Gill, D.S., Gordon, K.H.J. and Bruening, G. (1985). On the mechanism of replication of viroids, virusoids and satellite RNAs. In *"Subviral Pathogens of Plant and Animals: Viroids and Prions"* (K. Maramorosch and J.J. McKelvey, eds.) pp. 235-263, Academic Press, Orlando, Florida.

Takanami, Y., Kubo, S. and Imaizumi, S. (1977). *Virology* **80**, 376-389.

Taliansky, M.E., Malyshenko, S.J., Pshennikova, E.S. and Atabekov, J.G. (1982). *Virology* **122**, 327-331.

Taylor, J.M., Illmensee, R. and Summers, J. (1976). *Biochim. Biophys. Acta* **422**, 324-330.

Tien-Po, Davies, C., Hatta, T. and Francki, R.I.B. (1981). *FEBS Lett.* **132**, 353-356.

Thomas, P.S. (1980). *Proc. Natl. Acad. Sci. USA* **77**, 5201-5205.

Tinoco, I., Uhlenbeck, O.C. and Levine, M.D. (1971). *Nature* **230**, 362-367.

Tomenius, K., Clapham, D. and Meshi, T. (1987). *Virology* **160**, 363-371.

Tuerk, C., Gaus, P., Thermes, C., Groebe, D.R., Gayle, M., Guild, N., Stromo, G., D'Aubenton-Carafa, Y., Uhlenbeck, O.C., Tinoco, I., Brody, E.N. and Gold, L.C. (1988). *Proc. Natl. Acad. Sci. USA* **85**, 1364-1368.

van Loon, L.C. (1987). *Adv. Virus Res.* **33**, 205-255.

van Vloten-Doting, L., Francki, R.I.B., Fulton, R.W., Kaper, J.M. and Lane, L.C. (1981). *Intervirology* **15**, 198-203.

Uhlenbeck, O.C. (1987). *Nature* **328**, 596-600.

Wood, W.J. Gritschier, J., Lasky, L.A. and Lawn, R.M. (1985). *Proc. Natl. Acad. Sci. USA* **82**, 1585-1588.

Yanisch-Perron, C., Vieira, J. and Messing J. (1985). *Gene* **33**, 103-119.

Zimmern, D. and Hunter, T. (1983). *The EMBO J.* **2**, 1893-1900.

Zoller, M.J. and Smith, M. (1983). *Meth. Enzym.* **100**, 468-500.