THE MOLECULAR BIOLOGY OF ERYSIMUM LATENT TYMOVIRUS

A thesis submitted for the degree of Doctor of Philosophy

of the Australian National University

PATTANA SRIFAH

Molecular Evolution and Systematics Group

Research School of Biological Sciences

The Australian National University

Canberra, A.C.T., Australia

August, 1991

TABLE OF CONTENTS

STATEMENT (v)

ACKNOWLEDGEMENTS (vi)

SUMMARY (viii)

CHAPTER 1 - INTRODUCTION 1

ERYSIMUM LATENT TYMOVIRUS 1

The biology of ELV 1

Properties of ELV Particles 3

The ELV Genome and Virion Protein 3

THE TYMOVIRUSES 5

General Biology 5

Genomic Sequences of Viruses 7

Tymovirus Gene Expression 11

Evolution of Tymoviruses 13

OBJECTIVES OF THE WORK REPORTED IN THIS THESIS 18

CHAPTER 2 - GENERAL MATERIALS AND METHODS 19

MATERIALS AND SOURCES 19

Enzymes and Proteins 19

Nucleic Acids and Nucleotides 19

Other Reagents 21

GENOMIC RNA PREPARATION 22

Virus Propagation and Purification 22

Viral RNA Extraction 22

SYNTHESIS OF dsDNA ENCODING THE VIRAL RNA GENOMES 23

Phosphorylation of Oligodeoxyribonucleotide Primers 23

Polyadenylation at 3'-end of Viral RNA 24

Preparation of a dsDNA Copy of the Genomic RNA 25

cDNA Synthesis Using Specific ELV Primers 26

Determination of 5'-terminus by Primer Extension 27

Polymerase Chain Reaction 28

PLASMID AND BACTERIOPHAGE CLONING 29

Preparation of Vectors for Cloning 29

Preparation of DNA for Insertion into Vectors 30

Ligation of DNA Fragments into Vectors 30

Transformation 31

PLASMID AND BACTERIOPHAGE PREPARATION 32

Bacteriophage ssDNA Preparation 33

ssDNA Preparation from dsDNA Plasmid Vectors 33

Preparation of Large Amounts of dsDNA 34

HYBRIDIZATION FOR TRANSFORMANT IDENTIFICATION 35

Radioactively-labelled Probe Preparation by Nick Translation 35

Colony Hybridization 36

Dot-blot Hybridization Procedure 37

DNA SEQUENCING 38

Denaturated dsDNA Preparation for DNA Sequencing 38

DNA Sequencing Protocols 39

PRODUCTION OF A dsDNA ENCODING THE ELV GENOME 40

Construction of Full-length ELV DNA 42

In vitro Transcription and Capping 44

Biological Assay by Inoculation of Chinese Cabbage Plants 44

Detection of Viral RNA by Dot-blot Hybridization 45

Identification of Virion Coat Protein Gene by PCR 46

COMPUTER ANALYSIS 46

CHAPTER 3 - PRIMARY STRUCTURE OF THE ELV GENOME 48

INTRODUCTION 48

SEQUENCE OF THE ELV GENOME 48

FEATURES OF THE ELV GENOME 50

CHAPTER 4 - COMPARISONS OF ELV AND OTHER TYMOVIRUSES; GENOMIC SEQUENCES AND ENCODED PROTEINS 52

THE BASIC FEATURES 52

GENOMIC REGIONS 53

The 5' Non-coding Region 53

Proteins Encoded by Tymoviruses 53

Sequence of the 3' Non-coding Region 58

TYMOVIRUS EVOLUTION 60

VIRAL SIGNALS FOR HOST SPECIFICITY 61

CHAPTER 5 - RELATIONSHIPS OF TYMOVIRUSES AND OTHER VIRUSES 63

INTRODUCTION 63

VIRAL REPLICASE PROTEINS 64

The A Segments 66

The B Segments 66

The C Segments 66

Comparitive Phylogenetics of the RP motifs 67

THE 'SINDBIS-LIKE' VIRUSES, A SUPERGROUP BASED ON REPLICASE SIMILARITIES; CORRELATIONS WITH OTHER FEATURES 68

Genome Organization of Sindbis-like Viruses 69

5' - and 3' -terminal Non-coding Regions 71

Initiation and Termination Codons in Tymovirus Genes 71

Comparisons of Possible Subgenomic Promoters in the Genomes of Tymoviruses and Other Viruses 73

CHAPTER 6 - CONSTRUCTION OF CLONED DNA ENCODING THE ELV GENOME 76

INTRODUCTION 76

CLONE CONSTRUCTION 76

Construction and Selection of Full-size cDNA Clones 76

Bioassay of Transcripts from the ELV DNA Clones 77

CHAPTER 7 - DISCUSSION and CONCLUSIONS 79

GENOMIC ELV SEQUENCE 79

The Terminal Untranslated Regions 79

Internal Control Regions 81

Encoded Proteins 82

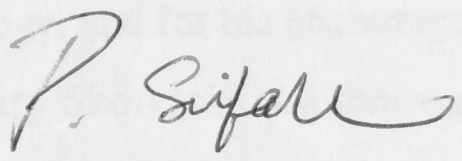
RELATIONSHIPS WITH OTHER VIRUSES
BASED ON SEQUENCE COMPARISONS 84

INFECTIOUS RNA TRANSCRIPTS FROM CLONES OF ELV 85

REFERENCES 89

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 acknowledged.



Pattana Srifah

ACKNOWLEDGEMENTS

I wish to express my gratitute to my supervisor, Dr. Adrian Gibbs, for generously providing me with the opportunity to work in the Molecular Evolution and Systematics Group, for his friendship and continued support, especially, during the writing of this thesis and his guidance and encouragement throughout the course of this work. My sincere thanks go to my co-supervisor Dr. Paul Keese for his help with the theory and practice of all the molecular biology, and for his encouragement and helpfulness. In addition, I wish to thank Dr. Mary Skotnicki, another co-supervisor for all of her advice.

The work described in this thesis was supported by the following people; Anne Mackenzie, Marjo Torronen and Paul Keese for their technical assistance and advice; Adrian Gibbs, David Sandilands, Jack Palmer, Shouwei Ding, Georg Weiller, and John Armstrong for their computing advice; James Whitehead for drawing all the complicated figures; and for this willing help I am truly grateful.

Thanks also go to Australian International Development Assistance Bureau for the AIDAB-ACIAR Fellowship Award who provided financial support throughout the course of this study.

I am grateful to all my best friends, including Sunee Budsayavith, Karen Crabtree, Julie Glover, Jennie Gibbs, Sakkarin Niyomsilpa, Kala Paul, Suchirat Sakuanrungsiri-kul, Elizabeth Smith, Nualnoi - Rungrueng Threerat and Ken Wessen, whose kindness provided not only knowledge but also happiness, while I was studying in Australia. Special thanks due to my family, my friends in Thailand, especially Ms. Patchanee Leeladee, for caring, supporting and understanding me.

SUMMARY

The ssRNA genome of erysimum latent tymovirus (ELV) is 6034 nucleotides in length and closely resembles other tymoviruses in its organization. The genome has small non-coding regions at both ends and these together comprise about 3% of the genome. The coding region consists of three large open reading frames (ORFs) encoding proteins with M_r values of 48,596 (49K), 193,806 (194K) and 21,482 (21.5K) with the first starting seven nucleotides before the second and overlapping it. These three proteins are homologous with the overlapping, replicase and virion proteins (OP, RP and VP, respectively) already reported for other tymoviruses. Although the sequence relationships show that ELV is a tymovirus, it is the most distinct of those that have been sequenced so far, and its VP has an extra unique segment of eleven amino acid residues near its N-terminus. Nonetheless the serological relationships of the particles of different tymoviruses significantly correlate with the sequence relationships of their VPs.

Computer analyses failed to find any obvious and specific similarities between the three ELV and TYMV proteins that could explain why their host ranges and symptoms are so similar, and differ, in this respect, from ononis yellow mosaic, kennedya yellow mosaic and eggplant mosaic tymoviruses. It seems that protein signals encoding those particular biological functions are probably subtle.

The non-coding nucleotides at the 3'-terminus of the ELV genome is only 78 nucleotide long and, unlike that region of most other tymoviruses, is unable to form a complete tRNA-like structure able to be aminoacylated with valine.

The amino acid sequence of the RPs (206K proteins) of five tymoviruses were compared in detail with those of the corresponding proteins of alfalfa mosaic, apple chlorotic leaf spot, beet necrotic yellow vein, narcissus mosaic, potato M, potato X, Sindbis, tobacco mosaic and white clover mosaic viruses; these viruses of the Sindbis-like group had been shown to have nucleotide binding and replicase motifs, and also possible methyltransferase motifs, that are clearly related to those of tymoviruses. It was shown that the five tymovirus RPs are most closely related to those of apple chlorotic leaf spot closterovirus, potato M carlavirus, narcissus mosaic, potato X and white clover mosaic potexviruses. These viruses differ considerably in gene

organization, base composition and particle morphology, and comparisons with other groups suggest that genetic recombination has been an important feature in the origins of these virus groups.

Plasmids containing DNA segments encoding the complete ELV genome were made. pELV35 and pELV81 were constructed in the plasmid Bluescript SK(+) adjacent to the bacteriophage T7 RNA polymerase promoter. When transcribed in the presence of the cap homologue, m^7GpppG, RNAs were obtained with the same sequence as the ELV genome, but with one extra guanine residues at the 5'-end and five extra nucleotides at the 3'-end. Chinese cabbage plants manually inoculated with such transcripts showed mild chlorotic spotting and mottling two weeks later. These symptoms were different from the characteristic and severe symptoms of ELV. They were only produced by transcripts made in the presence of the cap homologue, m^7GpppG. No virus particles were detected in plants inoculated with the transcripts, but small amounts of ELV RNA was detected by nucleic acid hybridization, and the VP gene was detected after PCR amplification.

CHAPTER 1 - INTRODUCTION

ERYSIMUM LATENT TYMOVIRUS

The Biology of ELV

Erysimum latent tymovirus (ELV), first described by Shukla and Schmelzer (1972), was isolated in November 1969 from symptomless *Erysimum helveticum* (Jacq.) DC. collected from the Plantation of Zentralinstitut für Genetik und Kulturpflanzenforschung, Gatersleben in East Germany. It has only been recorded in East Germany where it causes diseases of wild and crop brassicas. It has been isolated from naturally infected plants of the perennial species *Erysimum perovskianum, E. pulchellum, E. sylvestre,* and *Barbarea vulgaris,* which were symptomless and *Arabis ludoviciana, E. crepidifolium* and *Fibigia clypeata* which showed mosaic symptoms that could have been caused by simultaneous infection with other viruses (Shukla *et al.*, 1972, 1975).

ELV has a narrow experimental host range, mostly infecting brassicas but also a few species of the Cucurbitaceae, Labiatae, Leguminosae, Resedaceae, and Solanaceae. ELV is readily transmitted experimentally by sap inoculation, and is transmitted in nature by four species of *Phyllotreta sp.* (flea-beetles) in a semipersistent manner; the virus can be acquired in less than 10 min, can infect plants in 30 to 60 min, and persists in insects for up to 3 days. The beetle vectors of ELV themselves have very restricted host ranges and are poorly dispersed, and this may be the main reason for the limited distribution of ELV. ELV is not known to be transmitted through seeds (Shukla *et al.*, 1975; Proeseler and Schmelzer, 1977).

ELV, like other tymoviruses, seems only to infect dicotyledonous plants (Table 1) (Shukla *et al.*, 1972; Guy *et al.*, 1984; P. Srifah unpublished data). Guy *et al.* (1984) reported a broad correlation between the host ranges and taxonomy of ten tymoviruses including belladonna mottle virus (BMV), dulcamara mottle virus (DMV), clitoria yellow vein virus (CYVV), cacao yellow mosaic virus (CoYMV), ELV, eggplant mosaic virus (EMV), kennedya yellow mosaic virus (KYMV), ononis yellow mosaic virus (OYMV), scrophularia mottle virus (ScMV), and turnip yellow mosaic virus (TYMV). They used the taxonomic scheme proposed by Young and Watson (1970) in

Table 1. Plants susceptible and insusceptible to infection by erysimum latent tymovirus (ELV).

Susceptible plants	Insusceptible plants
Amaranthaceae:	
Celosia argentea L.	
Apocynaceae:	
Catharanthus roseus (=*Vinca rosea* L.)	
Aizoaceae:	
-	*Dorotheanthus bellidiformis* (Burm. f.) N.E.Br.
Bignoniaceae:	
-	*Catalpa bignonioides* Walt.
Boraginaceae:	
-	*Myosotis sylvatica* (Ehrh.) Hoffm.
Brassicaceae:	
Aethionema arabicum (L.) Andrz., *A. recurvum* Hausskn. et Bornm., *Alyssum alyssoides* L., *Arabis hirsuta* (L.) Scop., *A. ludoviciana*, *Arabidopsis thaliana* L., *Aubrieta deltoides* (L.) DC., *Brassica chinensis* L., *B. juncea* (L.) Czern. et Coss., *B. napus var.* napobrassica (L.) Rchb., *B. nigra* (L.) Koch, *B. oleracea* var. *botrytis* (L.)Alef., *B. oleracea* var. *gongyloides* L., *B. pekinensis* (Lovr). Rupr., *Bunias erucago* L., *Camelina sativa* (L.) Crantz, *Capsella bursa-pastoris* (L.) Medik., *Conringia orientalis* (L.) Dum., *Eruca vesicaria* (L.) Cav. em. Thell., *Erysimum allionii* Hort., *E. barbarea* L., *E. crepidifolium* , *E. helveticum* (Jacq.) DC., *E. perovskianum* Fisch. & Mey, *E. pulchellum* Gay, *E. sylvestre* Scop., *Fibigia clypeata*, *Hesperis matronalis* L., *Iberis umbellata* L., *Isatis tinctoria* L., *Lepidium sativum* L., *Lobularia maritima* (L.) Desv., *Lunaria annua* L., *Malcolmia africana* (L.) R. Br., *M. bicolor* Boiss. et Heldr., *M. flexuosa* Sibt. et Sm., *M. maritima* R. Br., *Matthiola bicornis* DC., *Sinapis alba* L., *S. arvensis* L.	*Aethionema grandiflorum* Boiss. et Reyn., *Alyssum saxatile* L., *Arabis caucasica* Willd., *Brassica oleracea* var. *acephala* (DC.) Alef., *B. oleracea* var. *capitata* (L.) Alef., *B. oleracea* var. *gemmifera* DC., *Cherianthus cheiri* L., *Fibigia clypeata* Medik., *Iberis amara* L., *I. sempervirens* L., *Malcolmia littorea* R. Br., *Matthiola incana* (L.) R. Br., *Raphanus raphanistrum* L.
Campanulaceae:	
-	*Lobelia erinus* L.
Caryophyllaceae:	
Stellaria media (L.) Vill.	*Gypsophila elegans* Bieb., *Myosotis sylvaticus* Hoffm.
Chenopodiaceae:	
-	*Beta vulgaris* L., *Chenopodium amaranticolor* Coste et Reyn., *C. foetidum* Schard., *C. murale* L., *C. quinoa* Willd., *Spinacia oleracea* L.
Convolvulaceae:	
-	*Ipomoea purpurea* (L.) Roth.

Susceptible plants	Insusceptible plants
Asteraceae:	*Bellis perennis* L., *Dimorphotheca amaratiaca* DC., *Helianthus annuus* L., *Lactuca sativa* L., *Podolepis robusta* (Maiden et Betche) J.H.Willis
Cucurbitaceae: *Cucumis sativus* L.	
Euphorbiaceae: -	*Euphorbia marginata* Pursh.
Gesneriaceae: -	*Sinningia speciosa* (Lodd.) Hiern.
Hydrophyllaceae: -	*Nemophila menziesii* Hook. et Arn.
Labiatae: *Ocimum basilicum* L.	*Salvia splendens* F. Sellow
Fabaceae: *Trigonella coerulea* (L.) Ser.	*Arachis hypogaea* L., *Cassia artemisiodes* Galdich., *Phaseolus vulgaris* L. cv. Plena and Pinto, *Pisum sativum* L., *Vicia faba* L., *Vigna sinensis* (L.) Savi ex Hassk.
Malvaceae: -	*Gossypium hirsutum* L.
Myrtaceae: -	*Eucalyptus cloeziana* F. Muell
Onagraceae: -	*Godetia amoena* (Lehn) G. Don.
Papaveraceae: -	*Papaver nudicaule* L.
Plantaginaceae: -	*Plantago major* L.
Polemoniaceae: -	*Phlox drummondii* Hook.
Polygonaceae: -	*Fagopyrum esculentum* Moench
Pontederiaceae: -	*Leptosiphon* sp.
Ranunculaceae: -	*Delphinium consolida* L., *D. hybridum* Hort.

Susceptible plants	Insusceptible plants
Resedaceae:	
Reseda odorata L.	
Rosaceae:	
-	*Fragaria vesca* L.
Scrophulariaceae:	
-	*Antirrhinum majus* L., *Collinsia heterophylla* Buist ex Grah., *Erinus alpinus* L.
Solanaceae:	
Nicotiana glutinosa L., *N. megalosiphon* Heurck et Muell., *Solanum integrifolium* Poir.	*Nicotiana clevelandii* Gray, *N. tabacum* L., *Petunia hybrida* Hort. ex Vilm.
Tropaeolaceae:	
-	*Tropaeolum majus* L.
Umbelliferae:	
-	*Amni majus* L., *Anthriscus cerefolium (L.) Hoffm.*, *Apium graveolens* var. *dulce* (Mill.) Pers., *Pastinaca sativa* L., *Petroselinum crispum* (Mill.) Nym. ex A.W.Hill
Valerianaceae:	
-	*Valerianella locusta* (L.) Betcke
Violaceae:	
Viola cornuta L.	

Best Assay Species known are:

B. napus var.*napobrassica* (L.) Rchb. (local lesion host).

which, at the highest level, there are just two groups of dicotyledons; the 'crassinucelli' and the 'tenuinucelli'. Guy *et al.* (1984) noted that most tymoviruses found naturally in crassinucellate plants (e.g. Brassicacae and Leguminosae) were more closely related to others from crassinucellate plants than to tymoviruses from tenuinucellate plants (e.g. Solanaceae), and *vice versa.* This grouping also significantly correlated with the experimental host preferences of the viruses, and so Guy *et al.* (1984) suggested that there are two main groups of tymoviruses, which they named the crassi-tymoviruses and the tenui-tymoviruses. Thus ELV, like turnip yellow mosaic (TYMV), was classified as a crassi-tymovirus. This grouping is confirmed by the experimental host range studies of ELV (Table 1) which has been found to infect 44 out of 79 (55%) crassinucellate species, but only 5 out of 27 (19%) tenuinucellate species. These figures are similar to those obtained in the analysis of the taxonomy of tymoviruses and their naturally infected hosts by Guy *et al.* (1984), who, using a chi-square test, obtained a score of $\chi^2 = 5.7$, $P < 0.025$. It showed that the natural hosts of crassinucellate species were significantly more susceptible than were species from the tenuinucellate species.

Shukla and Schmelzer (1973c, d) reported that environmental conditions influence symptom expression and the accumulation of ELV particles in infected hosts. Infected *B. chinensis* (L.) show the most severe systemic symptoms and the greatest concentration of particles, 0.5 mg/g of plant tissue, 14 days after inoculation when grown under high intensity light (1200 lux) at 22°C and a relative humidity of 70 percent. Whereas, in lower light intensities (75 or 300 lux) and lower or higher temperatures (18°C or 25°C), the symptoms are milder or absent (at 32°C) and the virus particles present at lower concentrations. The virus concentration decreases 21 days after inoculation (Shukla and Schmelzer, 1973d).

ELV also induces, in infected cells, cytological changes that are characteristic of tymovirus infections (Lesemann, 1977). Vesicles bounded by double membranes are found in the peripheries of the chloroplasts, and masses of lightly staining empty particles occur in the nuclei (Shukla *et al.*, 1980). As the disease develops, the chloroplasts become rounded and clumped, and form large internal vacuoles which can be easily seen using a light microscope (Matthews and Sarkar, 1976).

Properties of ELV Particles

In sap, particles of ELV have a thermal inactivation point between 76° C to 78°C, a dilution end point up to 1:500,000 and are stable at room temperature for more than 21 days (Shukla and Schmelzer, 1972).

ELV has icosahedral particles with a diameter of about 27 nm. When centrifuged the particles sediment as two components. The so-called 'bottom' component (BC) comprises 'full' nucleoprotein particles, which exclude negative stains from their interior and sediment at approximately 113S. By contrast the 'top' component (TC) comprises 'empty' protein shells and these are penetrated by negative stain and sediment at approximately 59S (Shukla and Gough, 1980). The BC particles are infectious, and contain both the genomic RNA and subgenomic RNA (the messenger RNA of the virion protein) (P. Srifah unpublished data). Whereas the TC particles are not infectious (Shukla *et al.*, 1973a) but contain host tRNA, and in this respect are similar to those of some other tymoviruses including BMV, DMV, CoYMV, EMV, KYMV, OYMV and wild cucumber mosaic (WCMV) (Bouley *et al.*, 1976; Geneveaux *et al.*, 1976; Blok *et al.*, 1987a).

The ELV Genome and Virion Protein

The genome of ELV consists of a single species of linear positive sense, single-stranded RNA molecules. These have a relative molecular mass of about 2x10^{6} and comprise 32% of the nucleoprotein particle weight, as estimated by particle composition and structure (Gough *et al.*, 1982). The genome has a base composition of 15.6% G, 23.9% A, 26.0% U, 34.5% C, and thus has the unusually large cytosine content and small guanine content (Shukla *et al.*, 1980), that is characteristic of tymoviruses.

A single species of virion protein (VP) comprises about 68% of the weight of ELV particles. An electron microscope analysis of ELV particles reveals that, like other tymoviruses, they have 32 surface protuberances, indicating that they have a shell of 180 identical protein subunits arranged in a T=3 icosahedral surface lattice clustered into 20 hexamers and 12 pentamers (Colman *et al.*, 1980).

The ELV VP has been shown to have an estimated M_r of 21.7K by SDS polyacrylamide gel electrophoresis. FITMOL analysis (Gibbs and Knowles, 1977) of its amino acid composition (Table 2; Shukla *et al.*, 1980 and Srifah *et al.*, 1990)

indicates that it has about 208 amino acid residues, which is somewhat larger than the VP of other tymoviruses. Recently, Srifah *et al.*(1990) reported the nucleotide sequence of ELV VP gene and its encoded protein. They found that the ELV VP is of 202 amino acid residues, and this was confirmed by partial peptide analysis.

Table 2. Composition of the virion proteins of ELV and other tymoviruses (Shukla *et al.*, 1980).

	ELV	ELV	EMV	OYMV	TYMV
Molecular weight		21,483	19,769	20,449	20,152
Amino acid	**Moles %**		**Number of residues**		
Alanine (Ala, A)	7.02	14(14)*	25	9	14
Cysteine (Cys, C)	0	0(0)	3	2	4
Aspartic acid (Asp, D)	5.53	11(5)	5	7	7
Glutamic acid (Glu, E)	10.20	21(6)	4	5	6
Phenylalanine (Phe, F)	3.50	7(7)	6	5	6
Glycine (Gly, G)	8.40	17(10)	7	8	8
Histidine (His, H)	2.04	4(4)	3	2	3
Isoleucine (Ile, I)	3.80	8(8)	13	15	16
Lysine (Lys, K)	3.49	7(7)	7	5	7
Leucine (Leu, L)	10.88	22(22)	19	21	16
Methionine (Met, M)	1.44	3(4)	3	3	4
Asparagine (Asn, N)	-	-(6)	10	7	4
Proline (Pro, P)	10.34	21(20)	17	21	20
Glutamine (Gln, Q)	-	-(15)	9	11	8
Arginine (Arg, R)	2.61	5(5)	4	7	3
Serine (Ser, S)	13.80	28(28)	20	34	17
Threonine (Thr, T)	10.00	20(21)	21	11	27
Valine (Val, V)	8.05	16(16)	7	15	14
Tryptophan (Trp, W)	0.48	1(1)	1	1	2
Tyrosine (Tyr, Y)	1.50	3(3)	4	3	3
'Stop'		(1)	(1)	(1)	(1)
Asx (Asp or Asn, B)	5.53	(5+6)			
Glx (Glu or Gln, Z)	10.2	(6+15)			
Total		**208(202)**	**189**	**193**	**190**

* The numbers of amino acids in brackets were deduced from the nucleotide sequence of viral genomic RNA of ELV (Srifah *et al.*, 1990)

The amino acid sequence of ELV VP has been compared with the sequences of the VPs of four other tymoviruses, namely EMV-Trin (Osorio-Keese *et al.*, 1989),

KYMV-JB (Ding *et al.*, 1990d), OYMV-Tin (Ding *et al.*, 1989), and TYMV-CL (Keese *et al.*, 1989). It was found that the sequence of ELV VP is the most distinct of all five tymovirus VPs, and, unexpectedly, it is larger than the VPs of other tymoviruses because it has an additional N-terminal 11 amino acids. The sequence relationships of the five VPs correlate well with relationships assessed from their amino acid compositions (Paul *et al.*, 1980), and with their serological relationships (Shukla and Schmelzer, 1972; Koenig, 1976; Shukla and Gough, 1980) ELV shows only a very distant serological relationship with Andean potato latent (APLV) and OYMV. Nevertheless, it is clear that ELV is a definitive member of the tymovirus group (Shukla and Gough, 1980).

THE TYMOVIRUSES

General Biology

Turnip yellow mosaic virus, type member of the tymoviruses, is one of the most fully characterized plant viruses with small isometric particles. It was first reported over 40 years ago by Markham and Smith (1946).

Since then twenty or so viruses (Table 3) have been described with closely similar properties, especially of their virions, and these comprise the turnip yellow mosaic virus (**tymovirus**) group (Harrison *et al.*, 1971).

Turnip yellow mosaic virus (TYMV), because of its early isolation, ease of handling and the stability and high concentration of its particles, has become one of the most intensively studied plant viruses (Matthews, 1970).

One of the most interesting features of the virus, discovered in the 1940s, is that it produces both intact nucleoprotein particles and empty protein shells with no RNA (Markham *et al*, 1948 and Matthews, 1960). These two types of particles were compared and it was found that (Matthews, 1988):

> *(i) TYMV nucleoprotein particles are infectious, whereas the protein particles are not (Markham and Smith, 1949). This was the first clear indication for any virus that the nucleic acid is essential for infectivity;*
>
> *(ii) the RNA is located centrally within a coat of protein, a feature later found to be common to all small isometric virus particles;*
>
> *(iii) the empty protein shells are stable in strong salt solutions, indicating for the first time that protein-protein hydrophobic interactions are important for the stability of the particles of some viruses.*

Table 3. Members of the Tymovirus Group and their Hosts.

Virus	Host plant families	Reference
Andean potato latent virus (APLV)	Solanaceae	Gibbs *et al.*, 1966
Belladonna mottle virus (BMV)	Solanaceae	Paul, 1971
Cacao yellow mosaic virus (CoYMV)	Sterculiaceae	Brunt, 1970
Clitoria yellow vein virus (CYVV)	Leguminosae	Bock and Guthrie, 1977
Desmodium yellow mottle virus (DYMV)	Fabaceae	Fribourg *et al.*, 1976
Dulcamara mottle virus (DMV)	Solanaceae	Gibbs *et al.*, 1966
Eggplant mosaic virus (EMV)	Solanaceae	Gibbs and Harrison, 1973
Erysimum latent virus (ELV)	Brassicacae	Shukla and Gough, 1980
Kennedya yellow mosaic virus (KYMV)	Fabaceae	Gibbs, 1978
Okra mosaic virus (OkMV)	Malvaceae	Givord and Koenig, 1974
Ononis yellow mosaic virus (OYMV)	Fabaceae	Gibbs *et al.*, 1966
Passiflora yellow mosaic virus (PaYMV)	Passifloraceae	Crestani *et al.*, 1984
Physalis mosaic virus (PhMV)	Solanaceae	Peter and Derks, 1974
Plantago mottle virus (PlMV)	Plantaginaceae	Granett, 1973
Scrophularia mottle virus (ScMV)	Scrophulariaceae	Bercks, 1973
Tomato white necrosis virus (TWNV)	Solanaceae	Barradas, 1983
Turnip yellow mosaic virus (TYMV)	Brassicacae	Matthews, 1970
Voandzeia necrotic mosaic virus (VNMV)	Fabaceae	Fauquet *et al.*, 1984
Wild cucumber mosaic virus (WCMV)	Cucurbitaceae	van Regenmortel, 1972

Another likely member of the group is poinsettia mosaic (PoiMV) (Koenig and Fulton, 1986).

All tymoviruses, like ELV, produce two types of icosahedral particles about 25-30nm in diameter. First, there are the infectious nucleoprotein particles. These contain the monopartite RNA genome of M_r $2x10^6$, which constitutes 35% of their particle weight, and which is of messenger sense and therefore infectious when deproteinized. The particles also sometimes contain minor amounts of a subgenomic mRNA that encodes the virion protein and has relative molecular mass of about $0.25x10^6$ (Kaper

and Steere , 1959; Matthews, 1970; Mellema *et al.*, 1979; Keeling *et al.*, 1979). Particles of the second type are empty protein shells. They are not infectious, but contain small amounts of subgenomic RNA in those of EMV, OkMV, and WCMV (Markham and Smith, 1949; Pleij *et al.*, 1976; Klein *et al.*, 1976; Ricard *et al.*, 1977; Szybiak *et al.*, 1978) and host tRNAs in BMV, DMV, ELV, EMV, OYMV and WCMV (Blok *et al.*, 1987).

The RNA-free particles of tymoviruses seem to be as stable as those which contain the genomic RNA. This indicates that protein-protein interactions, rather than protein-RNA interactions, are important in stabilizing them (Markham *et al.*, 1948). By contrast the particles of many other plant viruses, that have isometric particles of a similar size and a RNA genome, rely mostly on RNA-protein interactions, and are unstable in concentrated salt solutions (Kaper, 1975).

Tymovirus genomes have a large (32-42%) cytosine and small (15-17%) guanine content. Both the RNA genomes and the subgenomic RNAs of tymoviruses have been shown to be capped at their 5'-termini with m^7GpppGp (Klein *et al.*, 1976; Pleij *et al.*, 1976; Ahlquist and Janda, 1984). This 5'-cap structure has been shown to facilitate the binding of eukaryotic cellular mRNAs to ribosomes for translation (Karatzas *et al.*, 1990) and to involve in the stability of RNA molecules for higher translation efficiency (Furuichi *et al.*, 1977; Shimotohmo *et al.*, 1977).

Genomic Sequences of Viruses

Recently the complete genomic sequences of several tymoviruses have been determined. They are:

TYMV - the type strain (Morch *et al.*, 1988);
- the Club Lake (CL) isolate from Australia (Keese *et al.*, 1989);
- the Blue Lake (BL) isolate from Australia (A.D.Meek unpublished data);

EMV - the Trinidad (Trin) isolate (Osorio-Keese *et al.*, 1989);

KYMV - the Jervis Bay (JB) isolate (Ding *et al.*, 1990d); and

OYMV - the Tintagel (Tin) isolate (Ding *et al.*, 1989).

a) Genomic Structure

The genomic sequences confirm previous estimates of the size and composition of tymovirus genomes. All are about 6,300 nucleotides in length, and the large cytosine and small guanine content, previously determined by direct chemical methods, is found to be uniformly distributed throughout the genomes. Their encoded proteins have a high proportion of amino acids, like proline, serine and threonine, that have cytosine-rich codons, but this is not the sole reason for their large cytosine content as they have a preponderance of cytosines in the third codon position of the major open reading frames (ORFs).

b) Tymovirus ORFs

There are three main open-reading frames (ORFs) of significant length that have similar placement relative to one another in every tymovirus genomic sequence (Figure 1).

Two 5'-proximal overlapping ORFs begin at 7 nucleotides apart. The larger ORF, which starts at the second AUG, covers more than 85% of the genome. The protein it encodes is about M_r 206K in size, and is probably a replicase protein (RP) as, in its C-terminal part, there are two sequence motifs common to RNA replicases. These are the nucleotide binding fold with the consensus sequence Gly-X-X-Gly-X-Gly-Lys[Ser/Thr] (GxxGxGK[S/T]) and the RNA-dependent polymerase with the consensus sequence Gly-Asp-Asp (GDD) (Kamer and Argos, 1984; Argos, 1988). These motifs are found in RNA replicases of several plant and animal viruses, and bacteriophages (Kamer and Argos, 1984; Goldbach, 1987; Goldbach and Wellink, 1988; Gorbalenya *et al.*, 1988a, b). The smaller 5' ORF overlaps the RP ORF and is about one third the size of the RP ORF, although this varies in different tymoviruses from M_r 66K to 82K.

The third ORF is 3'-proximal and encodes the virion protein (VP), which is of about M_r 20K, and is expressed via a subgenomic mRNA.

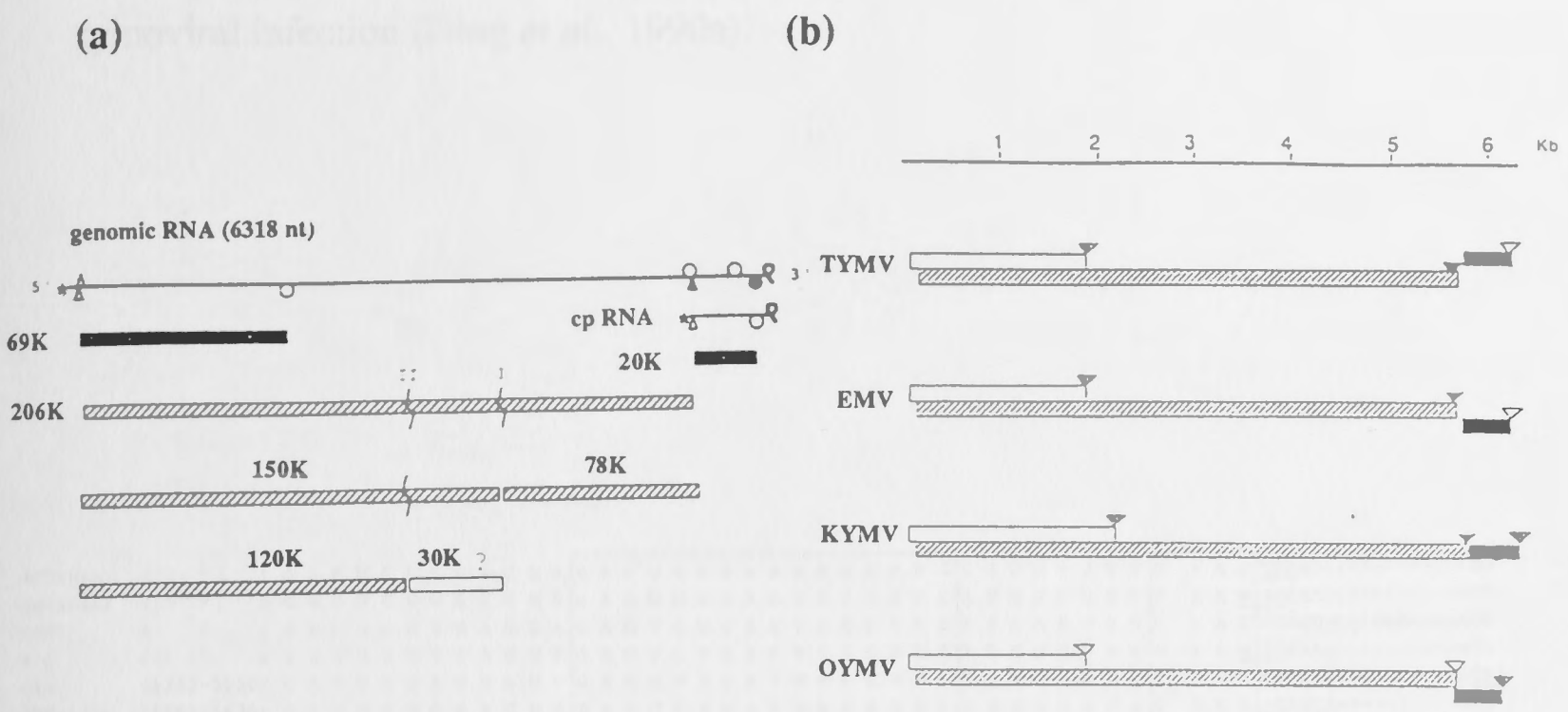


Figure 1. (a) A map of the TYMV genome and the proteins it encodes. Solid lines represent the viral genomic and subgenomic RNAs, and are marked with the positions of the initiation (Δ) and termination (O) codons, and the cap (*) and tRNA-like (ℓ) structures at their 5'- and 3'-ends, respectively. Boxes represent the ORFs, and these are hatched to indicate different reading frames, and the possible post-translational cleavage sites (ϟ) (from Morch *et al.*, 1989).

(b) Maps of four reported tymovirus genomes. Their three main ORFs are represented as open bars; white = OP, shadowed = RP, and black = VP. The triangles represent termination codons, white = UAA, shadowed = UGA, and black = UAG (from Ding, 1989).

c) The Untranslated Regions

The non-coding regions at the termini and between the ORFs constitute only about 3% of the genome.

i) 5'-terminus: The 5'-untranslated leader sequences of the genomic and subgenomic RNAs of TYMV contain a smaller proportion of G than other parts of the genome; the 5'-genomic leader is about 95 nucleotides long (Briand *et al.*, 1978) and that of the subgenomic RNA is only 19 nucleotide long (Guilley and Briand, 1978).

ii) Internal regions: There is, in EMV and TYMV, a small untranslated region between the RP and VP ORFs, whereas, in KYMV and OYMV, the RP and VP ORFs link and overlap, respectively. The RP/VP junction of the genome is also notable because it contains two conserved regions, the tymobox (5'-GAGUCUGAAU UCGUUC-3') and the initiation box (5'-CAAU-3') (Ding *et al.*, 1990a), which are probably concerned with controlling transcription from the negative (complementary) genomic strand to produce the VP mRNA (Figure 2). The tymobox sequence has been shown to be valuable as a general tymovirus primer for VP cloning, as a specific probe

for identifying tymoviruses, and as a potential target site for ribozymes to control tymoviral infection (Ding *et al.*, 1990a).

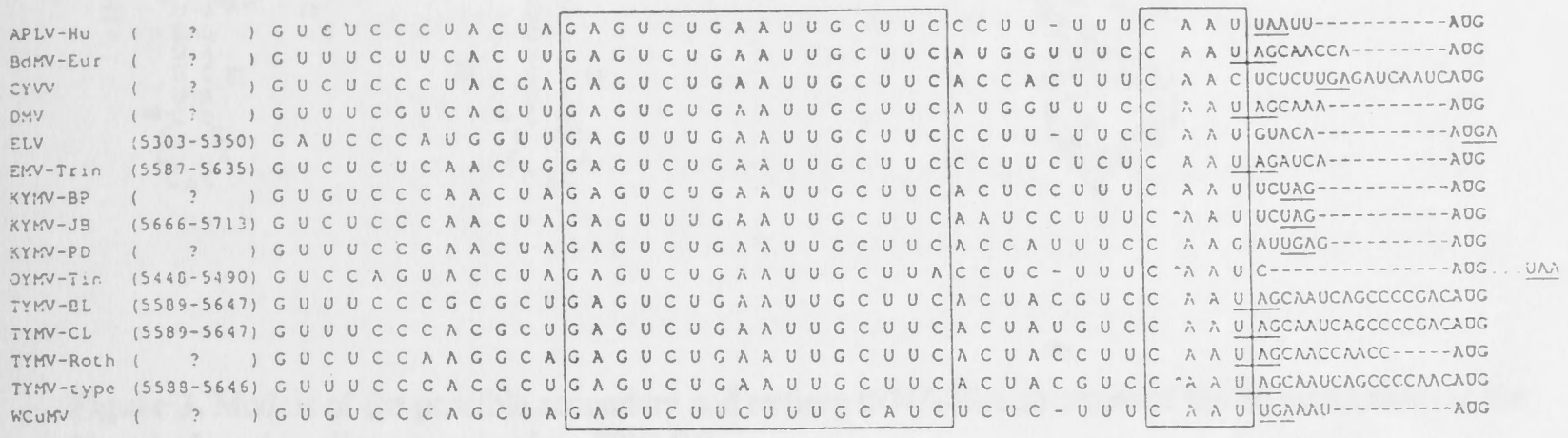


Figure 2. Aligned nucleotide sequences of tymoviral genomic RNAs in the region surrounding the initiation site of subgenomic RNA transcription. Two conserved sequences, the tymobox and the initiation box, are marked (from Ding *et al.*, 1990a).

iii) 3'-terminus: The 3'-termini of several tymoviruses have been shown to be specifically aminoacylated by valyl-t-RNA synthetase. These terminal sequences can be folded into energetically stable tRNA-like structures (Pinck *et al.*, 1970; Yot *et al.*, 1970; Giége *et al.*, 1978; Joshi *et al.*, 1978; Joshi *et al.*, 1982). Models of the secondary and tertiary structure of these 3'-termini have been derived from sequence comparisons, chemical modifications, direct enzymatic and chemical probing analysis, and by computer analysis (Pleij *et al.*, 1985; Pleij, 1990). The most likely structure, which can be represented in various ways (Figure 1), is similar to the classical clover-leaf of tRNA bent into an L-shape by additional basepairing (Figure 3)(Brian *et al.*, 1977; Dumas *et al.*, 1987; Florentz *et al.*, 1982; Joshi *et al.*, 1983 a, b; Rietveld *et al.*, 1982; van Belkum *et al.*, 1987).

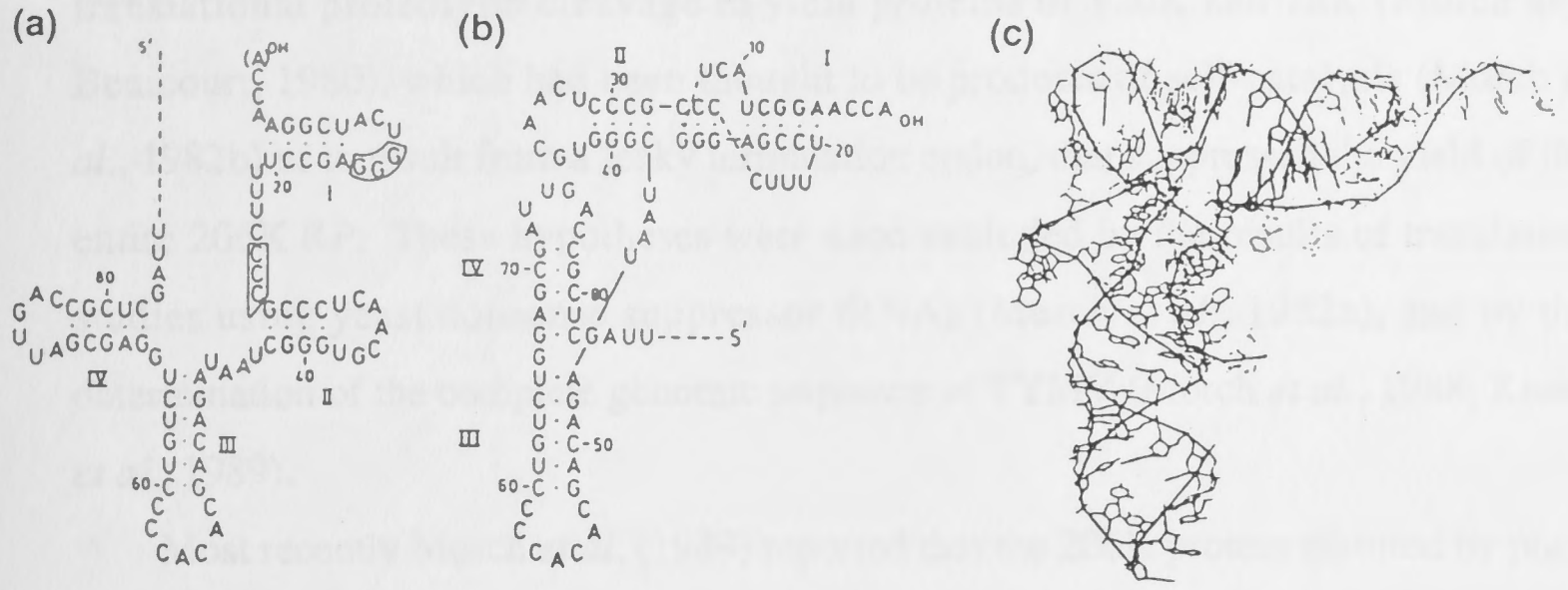


Figure 3. Models of the possible secondary and tertiary tRNA-like structure of the 86 nucleotides at the 3'-end of turnip yellow mosaic virus (TYMV):

(a) a model resembling the classical clover-leaf pattern of tRNA. The boxed regions indicate the pseudoknot-forming tertiary interaction. Numbering is from the 3'-end. Roman numbers indicate hairpins;

(b) L-arrangement showing the co-axial stacking of the triple G-C interaction between stem I and stem II, thus creating the equivalent of the aminoacyl acceptor domain of tRNA;

(c) A three-dimensional model built using computer graphics. The thick solid line represents the sugar-phosphate backbone (from Pleij, 1990).

This tRNA-like structure is found to be a common feature of at least nine reported tymovirus genomes namely APLV, CYVV, CoYMV, EMV, KYMV, OkMV, OYMV, TYMV, and WCMV (Joshi *et al.*, 1983a; van Belkum *et al.*, 1987). All of them have, at the 3'-terminus, the nucleotides -CC$(A)_{OH}$, which is the specific substrate sequence for valyl-t-RNA synthetase, and they have a -ACA(C)- sequence in the anticodon-like loop; these structural features are probably responsible for the specific aminoacylation of these tymovirus genomes by valine (van Belkum *et al.*, 1987).

Tymovirus Gene Expression

There have been several reports of *in vitro* translation studies of the proteins obtained by translating tymovirus genomes using ribosomes prepared from reticulocyte lysates or wheat germ extracts.

Benicourt *et al.* (1978) showed that, in the reticulocyte system, total TYMV RNA was efficiently translated into two large proteins of M_r195K and 150K as well as the smaller M_r 20K VP. The 195K (or 206K RP) protein was shown to undergo post-translational proteolytic cleavage to yield proteins of 120K and 78K (Morch and Benicourt, 1980), which had been thought to be products of self-catalysis (Morch *et al.*, 1982b) or to result from a leaky termination codon, that suppressed the yield of the entire 206K RP. These hypotheses were soon excluded by the results of translation studies using yeast nonsense suppressor tRNAs (Morch *et al.*, 1982a), and by the determination of the complete genomic sequence of TYMV (Morch *et al.*, 1988; Keese *et al.*, 1989).

Most recently Morch *et al.* (1989) reported that the 206K protein matured by post-translational cleavage to yield a 150k N-terminal and a 78K C-terminal fragment and then presumably the 150K fragment cleaved at the secondary sites to yield 120K and 30K proteins (Figure 1).

The 20K protein or VP is synthesised via a subgenomic RNA that corresponds to the 3'-terminal 694 nucleotides of the genomic RNA (Klein *et al.*,1976; Pleij *et al.*, 1976; Ricard *et al.*, 1977; Guilley and Briand, 1978). In addition, *in vitro* translation studies have revealed that the affinity of subgenomic RNA for the translation machinery is greater than that of the genomic RNA (Benicourt and Haenni, 1978). This may account for the efficient production of VP at the late stages of infection. This strategy for the enhanced expression of the VP gene is similar to that of many other viruses in the supergroup of the Sindbis-like plant RNA viruses (Goldbach, 1987). Further work has provided evidence that the TYMV subgenomic VP RNA is synthesised *in vivo* by transcription starting at an internal site on the complementary RNA (minus) genomic strand. This has been confirmed by the isolation of double-stranded RNAs of VP mRNA length from TYMV-infected plants (Gargouri *et al.*, 1989).

Until recently there was little information on the non-virion proteins of tymoviruses, although it has always been presumed that they were involved in genome replication. Preparations of RNA-dependent RNA polymerase (RNA replicase) isolated from infected tissue appear to contain two major subunits, a virus-encoded 120K protein and a host-encoded 45K protein (Mouchés *et al.*, 1984; Joshi *et al.*, 1986; Candresse *et al.*,1986). Early reports of TYMV genome replication by Bové *et al.*

(1965) and Ralph and Clark (1966) showed that the chloroplasts of infected Chinese cabbage (*Brassica campestris* spp. *pekinensis* L.) are associated with viral replication, and that the vesicles within the chloroplast membranes are most probably involved. Double-stranded RNA and the replication complex are associated with the chloroplast membranes (Bovė, 1972; Lafleche *et al.*, 1972; Garnier *et al.* 1980). Furthermore, Hatta and Matthews (1976) reported that TYMV particles probably assembled near the necks of the chloroplast vesicles as endoplasmic reticulum was seen, by electron microscopy, in the cytoplasm overlying the vesicles at an early stage of infection, and was then later replaced by accumulating virion protein. Lastly, the 150K protein was detected in TYMV-infected protoplasts of Chinese cabbage (Candresse *et al.* 1987), which were shown to be excellent plant material for studying cytological modifications in TYMV-infected cells, and virus multiplication, as well as for testing for viral proteins produced during infection by TYMV (Lafleche *et al.*, 1972; Garnier *et al.*, 1980; Cohen *et al.*, 1981).

Weiland and Dreher (1989) have successfully constructed clones of the entire TYMV genome in the plasmid pUC8, and this, with the T7 promoter for transcription *in vitro*, yielded infectious full-length RNA. The expression of the non-virion protein genes of two overlapping open reading frames of TYMV (1.9kb and 5.5kb ORFs encoding for 69K and 206K proteins) were also investigated by site-directed mutagenesis of their initiation codons. Results from experiments with mutant RNAs translated *in vitro* indicated that full length transcripts, in which the initiation codon of the 5.5kb ORF had been changed, were unable to replicate in protoplasts, whereas those with changes in the initiation codon of the 1.9kb ORF replicate although poorly.

Evolution of Tymoviruses

Various criteria may be used to assess the relationships of tymoviruses, and those between tymoviruses and other viruses.

Gibbs *et al.* (1966) suggested that there are two clusters or subgroups of tymoviruses, and that these can be distinguished by serological tests and their base composition. Koenig and Givord (1974) and Koenig (1976) showed that the relationships are more complex; they form a "loop structure" or network that resembles a tennis racquet. Koenig's classification largely correlates with the genome composition of the viruses and with the taxonomy of their natural hosts proposed by

Dale *et al.* (1975) and by Guy *et al.* (1984), who classified them into the crassi-tymovirus and tenui-tymovirus subgroups, previously described. These classifications mostly also correlate well with others based on the amino acid compositions of the virion proteins (Paul *et al.*, 1980), on cDNA-RNA hybridization and base ratio (Blok *et al.*, 1987a), and on the amino acid sequences of some tymovirus virion proteins (Ding, 1989; Srifah *et al.*, 1990). However, there seems at present to be no single criterion which provides an unequivocal classification, and their taxonomy seems less definite than that of tobamoviruses (Gibbs, 1980).

(a) **(b)**

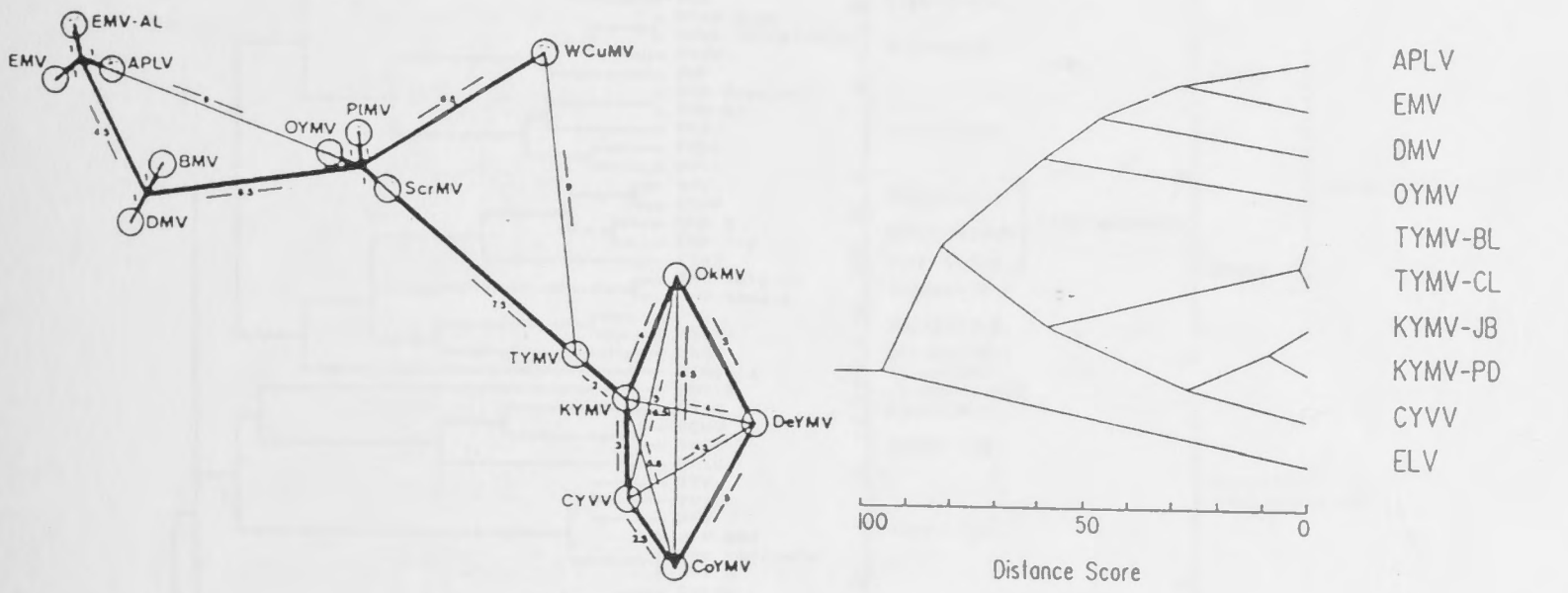


Figure 4. (a) Koenig's serological classification of tymoviruses based on the of average serological differentiation indices (SDI) in reciprocal tests (from Koenig, 1976).

(b) The phylogenetic relationships of ten tymovirus VPs. The tree was calculated by the UPGMA method from the distance score obtained from the progressive alignment of Feng and Doolittle (1987) (from Ding, 1989).

The presence of conserved amino acid sequences in the probable replicase proteins of RNA plant viruses is now well documented (Haseloff *et al.*, 1984; Kamer and Argos, 1984; Ahlquist *et al.*, 1985; Argos *et al.*, 1988; Goldbach and Wellink, 1988; Gorbalenya *et al.*,1989a, b). These conserved sequences have been found in the proteins of animal and bacterial viruses as well as in cellular proteins. The two best known sequences are the ones likely to be associated with, firstly, the active site of viral RNA polymerase (Kamer and Argos, 1984; Argos, 1988), which is characterized

by four conserved blocks of Dxxx[F/Y]D; [S/T]GxxxTxxxN[S/T]; GDD; and GxxxxxxK extending over about 100 amino acids. The other one, thought to be associated with a nucleotide-binding or helicase activity, has up to six conserved blocks, one of which contains the sequence of GxxGxGK[S/T] (Hodgman, 1988; Gorbalenya *et al.*, 1988a, b; 1989a, b).

These shared sequence motifs, together with similarities of gene organization and expression, provide a classification (Figure 5) which places the replicases of animal and plant viruses into a supergroup of 'alpha-like' or 'Sindbis-like' replicases (Ahlquist *et al.*, 1985; Candresse, 1990; Cornelissen and Bol, 1984; Goldbach, 1987; Haseloff *et al.*, 1984; Poch *et al.*, 1989).

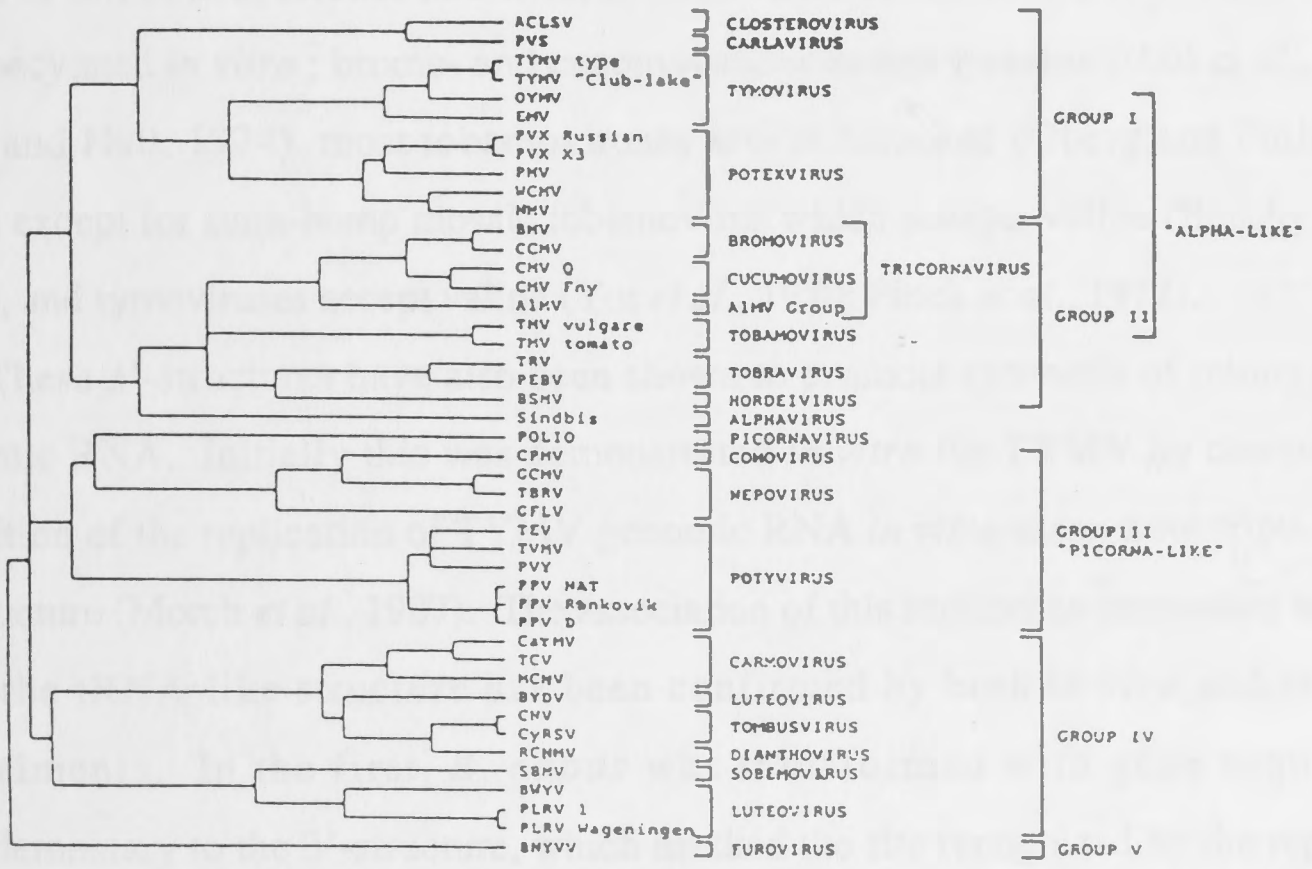


Figure 5. Hierarchical clustering of plant RNA viruses using the conserved 'polymerase' signature (from Candresse *et al.*, 1990). The dendrogram represents the relatedness of the sequences: the degree of homology between sequences is inversely proportional to the length of the branches and their linkages reflect the order obtained during the multiple alignment process and thus the increasing divergence of the sequences.

ACLSV=apple chlorotic leaf spot virus; AlMV=alfafa mosaic virus; BMV=brome mosaic virus; BSMV=barley stripe mosaic virus; BWYV=beet western yellows virus; BYDV=barley yellow dwarf virus; CarMV=carnation mottle virus; CCMV=cowpea chlorotic mottle virus; CMV=cucumber mosaic virus; CNV=cucumber necrosis virus; CPMV=cowpea mosaic virus; CyRSV=cymbidium ringspot virus; EMV=eggplant mosaic virus; GCMV=Hungarian grapevine chrome mosaic virus; GFLV=grapevine fern leaf virus; MCMV=maize chlorotic mottle virus; NMV=narcissus mosaic virus; OYMV=ononis yellow mosaic virus; PEBV=pea early browning virus; PLRV=potato leafroll virus; PMV=papaya mosaic virus; POLIO=poliovirus (type 1); PPV=plum plox virus; PVS=potato virus S; PVX=potato virus X; PVY=potato virus Y; RCNMV=red clover necrotic mosaic virus; SBMV=southern bean mosaic virus; Sindbis=Sindbis virus; TBRV=tomato black ring virus; TCV=turnip crinkle virus; TEV=tobacco etch virus; TMV=tobacco mosaic virus; TRV=tobacco rattle

virus; TVMV=tobacco vein mottling virus; TYMV=turnip yellow mosaic virus; WCMV=white clover mosaic virus (from Candresse *et al.*, 1990).

These are also are classified into supergroup A by Habili and Symons (1989), and a supergroup of 'picorna-like' or 'polio-like' replicases (Franssen *et al.*, 1984; Argos, 1988; Goldbach and Wellink, 1988; Poch *et al.*, 1989). The tymovirus replicases, like those of the tricorna-, tobamo-, tobra-, potex- and some other virus groups (Candresse *et al.*, 1990) are of the alpha-like supergroup, and are unexpectedly close to the potexviruses (Ding, 1989 and Rozanov *et al.*, 1990).

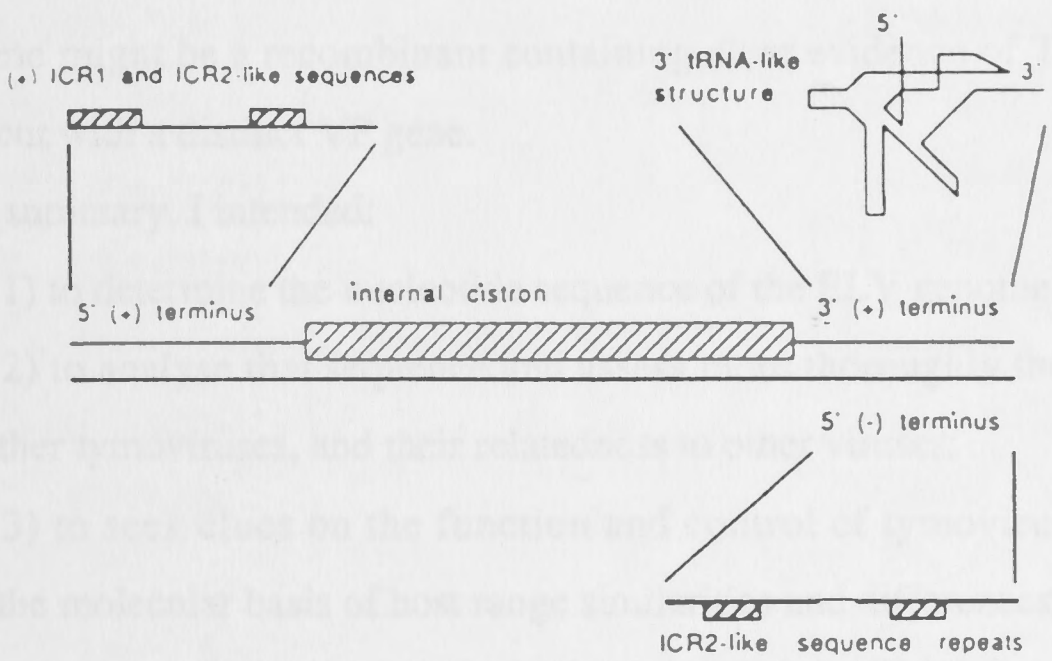
The RNA pseudoknotted structures, which are found at the 3'-end of a number of animal and plant virus genomes, may also be used to provide evolutionary clues. These viral -CCA_{OH} or -CC_{OH} termini can serve as substrates for nucleotidyltransferases (Joshi *et al.*, 1983a; Dreher and Hall, 1988b; van Belkum *et al.*, 1987) and can be aminoacylated *in vitro* ; bromo- and cucumoviruses accept tyrosine (Hall *et al.*, 1972; Kohl and Hall, 1974), most tobamoviruses accept histidine (Oberg and Philipson, 1972) except for sunn-hemp mosaic tobamovirus which accepts valine (Beachy *et al.*, 1976), and tymoviruses accept valine (Yot *et al.*, 1970; Pinch *et al.*, 1972).

These 3'-structures have also been shown to promote synthesis of minus strand genomic RNA. Initially this was demonstrated *in vitro* for TYMV by competitive inhibition of the replication of TYMV genomic RNA *in vitro* using transcripts of the 3'-structure (Morch *et al.*, 1987). The association of this replication promoting activity with the tRNA-like structure has been confirmed by both *in vivo* and *in vitro* experiments. In the first, *B. napus* was transformed with gene sequences complementary to the 3'-structure, which masked the site recognized by the replicase initiating minus strand synthesis and slowed infection (Cellier *et al.*, 1990). In the other experiments, site-directed mutagenesis of the pseudoknot region was found to affect the stability of the tRNA-like structure. Some of the mutants had a greatly decreased ability to be aminoacylated with valine (Mans *et al.*, 1990), even ones in which a single G-C basepair had been replaced by an A-U or the non Watson-Crick G-A or G-U basepairs; such experiments showed that pseudoknot structure is only marginally stable and that its formation lowers the free energy by only a few kcal/mol.

Recent work on bromo-, cucumo-, tobamo- and tymoviruses has shown that the 5'-termini of their genomic (+) and complementary (-) strands have regions of sequence similarity, and that these regions resemble the specific promoters of tRNA

genes that interact with RNA polymerase III and are known as consensus internal control regions ICR1 and ICR2 (Fowlkes and Shenk, 1980; Marsh and Hall, 1987; Marsh *et al.*, 1989; Allison *et al.*, 1989). In particular, the most striking resemblance was found in ICR2, 5'-GGUUCGANUCC-3', which is almost a palindrome and contains the TψC loop of tRNAs and viral analogues. In order to examine the role of these sequences in viral replication, Pogue *et al.* (1990) used site directed mutagenesis to show the influence of the ICR2 sequence in promoting (+)-strand synthesis. The fact that sequences resembling ICR1 and ICR2 are found in the 5'- and 3'-terminal (tRNA-like) regions of these viruses may imply a common evolutionary relationship with tRNA genes (Marsh *et al.*, 1989).

(a)



(b)

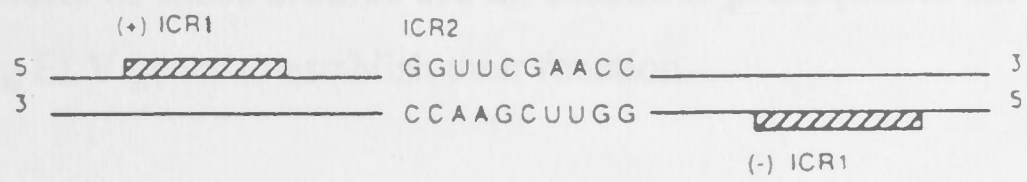


Figure 6. ICR motifs in BMV RNA and a model ancestral genome;

(a) Diagram showing locations of the ICR-like sequences at the 5'(+) and (-) strand termini and the 3'(+) strand tRNA-like terminus on BMV RNA 1 and 2;

(b) Model for an ancestral RNA genome showing ICR1 motifs at the ends of the (+) and (-) strands and a shared internal palidromic ICR2 (from Marsh *et al.*, 1989).

OBJECTIVES OF THE WORK REPORTED IN THIS THESIS

Early work by Schmelzer and Shukla established the basic characteristics of erysimum latent virus and showed unequivocally that it is a tymovirus. My objectives have been to extend the analysis of this virus to the molecular genetic level and allow comparison with the reported genomic sequences of other viruses. In particular I wished to compare ELV and TYMV genomic sequences because they not only share general tymovirus characters, but seemed especially similar biologically (i.e. they share most of their hosts, and cause closely similar symptoms), despite the fact that serological tests with their virions indicate that they are only distantly related. Thus I wished to determine the molecular basis of this anomaly because, for example, the ELV genome might be a recombinant containing clear evidence of TYMV genomic sequences but with a distinct VP gene.

So, in summary, I intended:

1) to determine the nucleotide sequence of the ELV genome;

2) to analyse that sequence and assess more thoroughly the relatedness of ELV and other tymoviruses, and their relatedness to other viruses;

3) to seek clues on the function and control of tymovirus genes, and in particular, the molecular basis of host range similarities and differences;

4) to construct a full-length dsDNA clone of the ELV genome and produce infectious viral molecules transcribed *in vitro*.

The results of these studies are an essential prerequisite for modifying and manipulating ELV genes to establish their function.

CHAPTER 2 - GENERAL MATERIALS AND METHODS

MATERIALS AND SOURCES

Enzymes and proteins

Restriction endonucleases:

Most restriction endonucleases were purchased from New England Biolabs (USA) and Boehringer-Mannheim (FRG).

Polymerases:

E. coli DNA polymerase I	Boehringer-Mannheim, FRG
E. coli DNA polymerase I, large fragment	Pharmacia, USA
T_4 DNA polymerase	Boehringer-Mannheim, FRG
Sequenase^{TM} DNA polymerase	USB, USA
Taq DNA polymerase	Perkin Elmer Cetus, USA
AMV reverse transcriptase	USB, USA
Terminal deoxynucleotidyl transferase	BRL, USA
T7 RNA polymerase	Pharmacia, USA
Poly (A) polymerase	Bresatec, Australia

Nucleases:

DNase I, RNase free	Promega, USA
RNase A	Boehringer-Mannheim, FRG
RNase H	Boehringer-Mannheim, FRG

Others:

T4 DNA ligase	New England Biolabs, USA
Alkaline phosphatase, Calf intestinal	Promega, USA
T4 polynucleotide kinase	New England Biolabs, USA
RNasin	Promega, USA
Proteinase K	Boehringer-Mannheim, FRG

Nucleic Acids and Nucleotides

Vectors:

M13mp18 RF DNA*	Bresatec, Australia
M13mp19 RF DNA*	Bresatec, Australia

pBluescript SK(+)	Stratagene, USA
pGEM-7Zf (+) and (-)	Promega, USA
pTZ18U and pTZ19U DNA*	Gifts from Dr. K. Reed

(* Vector dsDNAs used in this thesis were kindly prepared by A. Mackenzie)

Primers and Oligodeoxynucleotides:

m^7G(5')ppp(5')G mRNA cap homologue	New England Biolabs, USA
M13 universal 17-mer primer	Bresatec, Australia
pPUC/M13 reverse primer	Promega, USA
dT_8dG primer	P.L.Biochemicals,
Synthetic random hexamers	Bresatec, Australia
Synthetic ELV primers; PS2,PS3,PS7,PS8, PS9, PS10 and TALL-comp	Protein/DNA facility, ANU

Nucleotides:

2'-Deoxyadenosine 5'-triphosphate (dATP)	Pharmacia, Sweden
2'-Deoxycytidine 5'-triphosphate (dCTP)	"
2'-Deoxyguanosine 5'-triphosphate (dGTP)	"
2'-Deoxythymidine 5'-triphosphate (dTTP)	"
2',3'-Dideoxyadenosine 5'-triphosphate (ddATP)	"
2',3'-Dideoxycytidine 5'-triphosphate (ddCTP)	"
2',3'-Dideoxyguanosine 5'-triphosphate (ddGTP)	"
2',3'-Dideoxythymidine 5'-triphosphate (ddTTP)	"
Adenosine 5'-triphosphate (ATP)	"
Cytidine 5'-triphosphate (CTP)	"
Guanosine 5'-triphosphate (GTP)	"
Uridine 5'-triphosphate (UTP)	"

Radioisotopes:

[α-^{32}P]dATP (3,000 Ci/mmol)	Amersham, U.K.
[α-^{32}P]dCTP (3,000 Ci/mmol)	"
[γ-^{32}P]ATP (>5,000 Ci/mmol)	"
[α-^{32}P]GTP (3,000 Ci/mmol)	"

Other reagents

Acrylamide	Sigma Chemicals, USA
N,N'-methylene-bisacrylamide	"
Dithiothreitol (DTT)	"
Kanamycin sulphate	"
Ethidium bromide	"
Spermidine	"
5-Bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal)	"
Isopropyl-β-D-thiogalactopyranoside (IPTG)	"
Ampicillin (sodium salt)	Beecham Veterinary Products, Australia
Di-sodium ethylene-diamine-tetra-acetic acid (EDTA)	AJAX Chemicals,Australia
Polyethylene glycol 6000	BDH Chemicals, Australia
Sodium dodecyl sulphate (SDS)	Bio-Rad Laboratories,USA
N,N,N',N'-tetra-methyl-ethylene-diamine (TEMED)	Sigma Chemicals, USA
Urea (enzyme grade)	BRL, USA

GENOMIC RNA PREPARATION

Virus Propagation and Purification

The ELV isolate used in these studies was kindly provided by Dr D.D.Shukla. It was originally collected from *Erysimum helveticum* (Jacq.) DC. at Gatersleben, GDR in 1969 (Shukla and Schmelzer, 1972). The virus was propagated and maintained in *Brassica campestris* spp. *pekinensis* L. Two weeks after inoculation, the virus particles were extracted and purified using a modification of the method described by Shukla *et al.* (1973a), with an extraction buffer of 100mM Na_2HPO_4 and 50mM ascorbic acid, pH 7.0 (2 ml/g tissue). The filtered homogenate was shaken with an equal volume of a 1:1 mixture of *n*-butanol and chloroform and then constantly stirred for 30 minutes. The resulting emulsion was separated into its components by low-speed centrifugation (Sorvall RC-5C, GSA rotor, 5,000 rpm for 15 min at 4^oC). The virus particles in the clear yellow supernatant were concentrated by high-speed centrifugation (Beckman, Ti 60, 35,000 rpm for 120 min at 4^oC). The pellet was resuspended in 1x SSC buffer (150mM NaCl, 15mM Na_3-citrate). After a low-speed clarification spin, the suspension was layered onto a 10-40% sucrose density gradient (in 1x SSC buffer) and centrifuged for 3 hours at 26,000 rpm in a Beckman SW27 rotor at 4^oC. The two light-scattering bands, top and bottom viral components, were collected from the gradient separately. After diluting with 1x SSC buffer, the viral particles were concentrated by high-speed centrifugation, resuspended in 1x SSC buffer including 1mM sodium azide and then stored at 4^oC.

Viral RNA Extraction

The RNA used in early experiments to determine the nucleotide sequences of ELV genome was kindly supplied by A. Mackenzie. In later experiments, viral genomic RNA and subgenomic RNA of ELV were extracted separately from purified bottom and top component particles using the proteinase K method described by Both and Air (1979). 1 ml of purified virus (1 mg/ml) was mixed with 4 ml RNA extraction buffer (10mM Tris-HCl [pH7.4], 10mM KCl, 1.5mM $MgCl_2$, 0.2% sodium dodecyl sulphate), then 4mg proteinase K (Boehringer-Mannheim) was added and the mixture was incubated at 56^oC for 15 min. The mixture was adjusted to 150mM NaCl and an

equal volume of TE-saturated phenol (TE buffer: 10mM Tris-HCl [pH8.0], 1mM EDTA) was added. After occasional mixing during a further 5 minutes incubation at 56°C, 5 ml of chloroform was added. The solution was mixed well and the phases were separated by centrifuging at 4,000 rpm in a Sorvall RC-5C, SS-34 rotor, at 5°C for 10 min. The RNA in the aqueous phase was precipitated by adding 2 volumes of cold ethanol (-20°C) and 0.1 volume of 3M sodium acetate, pH 5.5. The mixture was mixed well and kept on ice for 10 min. The RNA precipitate was recovered by centrifuging in a Sorvall rotor at 12,000 rpm at 4°C for 15 min. The pellet was resuspended in TE and the phenol was removed by twice reprecipitating with ethanol, before being washed with 70% ethanol. The nucleic acid pellet was dried briefly in a vacuum desiccator, resuspended in water and finally stored in small aliquots at -20°C.

SYNTHESIS OF DOUBLE-STRANDED cDNA TO VIRAL RNA GENOMES

The procedures for the synthesis of dsDNA encoding the ELV genome were basically as described by Gubler and Hoffman (1983).

Phosphorylating of Oligodeoxyribonucleotide Primers

Most synthetic oligodeoxyribonucleotide primers, such as dT_8dG, were phosphorylated by T_4 polynucleotide kinase to catalyze the transfer of the γ-phosphate of ATP to the terminal 5'-hydroxyl group.

dT_8dG (2 μg)(P.L.Biochemicals) was phosphorylated at its 5'-terminus using 2 units of T_4 polynucleotide kinase (New England Biolabs) in a reaction mixture containing 50mM Tris-HCl (pH 9.0), 10mM $MgCl_2$, 5mM dithiothreitol (DTT), and 1mM rATP in a 20 μl reaction volume.

Thus the reaction mixture was prepared as :

	2 μl dT_8dG (1 μg/μl)
	2 μl 10x kinase buffer (500mM Tris-HCl (pH 9.0), 100mM $MgCl_2$, 50mM dithiothreitol (DTT),
	2 μl 10 mM rATP
	1 μl T4 polynucleotide kinase (2 U/μl)
	13 μl sterile water
total	20 μl

The mixture was incubated at 37°C for 30 minutes and then stopped by heating at 80°C for 5 min. The phosphorylated primer was stored at -20°C and 2 µl was used for each priming reaction.

Polyadenylation at 3'-end of Viral RNA

The 3'terminus of the viral RNA was polyadenylated using *E. coli* polyadenylation polymerase (poly(A)polymerase) before priming with dT_8dG. 10-20 µg viral RNA in 30 µl sterile water was denatured by heating at 56°C for 2 minutes, snap-cooled on ice, and then polyadenylated at 37°C for 30 min in 100 µl of a reaction mixture containing 50mM Tris-HCl (pH 8.0), 10mM $MgCl_2$, 250mM NaCl, 1mM DTT, 2.5mM $MnCl_2$, 1mM rATP, and 2.5 units poly(A)polymerase (Bresatec).
Thus the reaction mixture was prepared as:

	30 µl viral RNA
	50 µl 2x reaction buffer (100mM Tris-HCl (pH 8.0), 20mM $MgCl_2$, 500mM NaCl, 2mM DTT, 5mM $MnCl_2$)
	10 µl 10mM rATP
	10 µl poly(A)polymerase (0.25 U/µl)
total	100 µl

The reaction was terminated by shaking with an equal volume of cold phenol/chloroform (ratio = 1:1), mixing with a Vortex mixer for 5 seconds, and then the mixture was separated by centrifugating in an Eppendorf centrifuge at 12,000 rpm for 2 min at room temperature. The aqueous phase was transferred into a sterile Eppendorf tube. The RNA in the aqueous phase was precipitated by adding 0.05 volume of 4 M NaOAc (pH 5.5), and 2-2.5 volumes of cold ethanol (-20°C) and then kept at -20°C for 10 min. After centrifuging at 12,000 rpm for 5 min at 4°C, the pellet was quickly rinsed with cold 70% ethanol and dried under vacuum for 10 min. The polyadenylated RNA was dissolved in sterile distilled water and kept at -20°C.

Preparation of a Double-stranded DNA Copy of the Genomic RNA

a) First-strand cDNA Synthesis

The 10 µg polyadenylated RNA together with 2 µl of phosphorylated dT_8dG (for virion protein gene synthesis) or with synthetic random oligodeoxynucleotide hexamers (Bresatec)(for the whole viral genome synthesis) was heated to 80°C for 2

min (or 90°C for 1 min) and quickly chilled on ice. This mixture was then used as a template to synthesize cDNA in a reaction volume of 40 µl containing 50mM Tris-HCl (pH 8.3), 8mM $MgCl_2$, 50mM KCl, 10mM DTT, 0.5mM each of dGTP, dATP, dTTP and dCTP, 3 µl [α-^{32}P]dATP (10 mCi/ml), and about 20-22 units AMV reverse transcriptase (USB).

Thus the reaction mixture was prepared as:

	22 µl RNA and the primer
	8 µl 5x reaction buffer (250mM Tris-HCl (pH 8.3), 40mM $MgCl_2$, 250mM KCl)
	4 µl 0.1M DTT
	2 µl 10mM dNTPs
	3 µl [α-^{32}P]dATP
	1 µl AMV reverse transcriptase (22 U/µl)
total	40 µl

The first-strand cDNA was synthesized at 42°C for 90 min. The reaction was then terminated and the nucleic acid (RNA-DNA hybrid) was purified by the phenol/chloroform extraction and ethanol precipitation as described previously.

b) Second-strand cDNA Synthesis. Two alternative methods were used.

i) Procedure I: The pellet of the RNA-DNA hybrid was resuspended in 71 µl sterile distilled water and then added to a mixture of 10mM Tris-HCl (pH7.4), 5mM $MgCl_2$, 10mM $(NH_4)_2SO_4$, 100mM KCl, 20 µg/ml BSA, 40µM dNTPs, 1 unit of RNase H (Boehringer-Mannheim), 20 units of DNA polymerase I (holoenzyme, (Boehringer-Mannheim), 1 unit of T_4 DNA ligase (New England Biolabs).

Thus the reaction mixture was prepared as :

	71 µl RNA-DNA hybrid
	25 µl 4x reaction buffer (40mM Tris-HCl (pH7.4), 20mM $MgCl_2$, 40mM $(NH_4)_2SO_4$, 400mM KCl, 80µg/ml BSA, 160µM dNTPs,
	2 µl DNA polymerase I (10 U/µl)
	1 µl RNase H (1 U/µl)
	1 µl T_4 DNA ligase (1 U/µl)
total	100 µl

The mixture was incubated at room temperature for 2 hr or overnight, then the reaction was terminated and the resulting DNA was purified by phenol/chloroform extraction and ethanol precipitation as described above. The DNA pellet was dissolved in 20 µl sterile distilled water and stored at -20°C.

ii) Procedure II: The reaction mixture from the first strand synthesis was directly added to the second strand synthesis buffer and enzymes. The second strand reaction buffer contained 20mM Tris-HCl (pH 8.0), 5mM $MgCl_2$, 10mM $(NH_4)_2SO_4$, 100mM KCl, 20 µg/ml BSA, 2 units of RNase H (Boerhinger-Mannheim), 20 units of DNA polymerase I (holoenzyme, (Boerhinger-Mannheim), 2 unit of T_4 DNA ligase (New England Biolabs).

Thus the reaction mixture was prepared as :

	40 µl the first-strand synthesis reaction mixture
	100 µl 2x buffer (40mM Tris-HCl (pH 8.0), 10mM $MgCl_2$, 20mM $(NH_4)_2SO_4$, 200mM KCl, 40 µg/ml BSA)
	1 µl DNA polymerase I (20 U/µl)
	2 µl RNase H (1 U/µl)
	2 µl T_4 DNA ligase (1 U/µl)
	55 µl sterile distilled water
total	200 µl

The resulting dsDNA was purified by the phenol/chloroform extraction and ethanol precipitation as described previously. The DNA pellet was resuspended in 20 µl sterile distilled water and stored at -20°C.

cDNA Synthesis Using Specific ELV Primers

The sequences of the ELV cDNA genomic fragments were assembled by computer analysis into several contiguous large sequences. To complete the cloning of the entire genome, two oligodeoxynucleotide primers were synthesized. The first, PS2 primer, 5'-CTGGAAAGGATCCTGGAAGG-3'(complementary to nucleotides 2035-2056), was used to obtain part of the sequence missing from the clones, and the other, PS3 primer, 5'-CCGGAGACTTGAATGCCC-3'(complementary to nucleotides 261-279), to determine the 5'-terminal sequence.

dsDNA was synthesized using PS2-5'CTGGAAAGGATCCTGGAAGG 3' as a first strand primer in the AMV reverse transcriptase system and then Procedure II, as described above, was used to synthesize the second strand. The resulting DNA was hydrolysed with the appropriate restriction endonclease, *Nae* I (nucleotides 1933-1938), end-filled and ligated into the *Sma* I site of M13 vector. The protocols for digestion and cloning will be described in the next section.

Determination of 5'-terminus by Primer Extension

The 5'-terminal nucleotide of the ELV genomic RNA was determined using methods described by Gustafson *et al.* (1987) and Veidt *et al.*(1988). The cDNA complementary to the 5'-terminal portion of viral genomic RNA was synthesized using the phosphorylated oligodeoxynucleotide primer: PS3-5'CCGGAGACTTGAATGC-CC 3'(nucleotides 261-279), AMV reverse transcriptase and dNTPs and as in the procedure previously described. The RNA-DNA hybrid was treated with 1 ng RNase for 10 min at room temperature. Then the synthesized cDNA, together with an equal volume of double-dye formamide loading buffer (95% formamide, 10 mM EDTA, 0.02% xylene cyanol and bromophenol blue), was boiled for 5 min before separation from other components by electrophoresis in a 5 % polyacrylamide gel containing 7 M urea in 1x TBE buffer (90mM Tris-borate [pH 8.3], 1 mM EDTA). The eluted cDNA fragment (see page 30 for elution method) was dA-tailed using 20 units of terminal deoxynucleotidyl transferase (BRL) in a 'tailing' mixture (100mM potassium cacodylate [pH 7.2], 2mM $CoCl_2$, 0.2mM DTT, and 50µM dATP).

Thus the reaction mixture was prepared as:

	30 µl cDNA
	10 µl 5X tailing buffer (500mM potassium cacodylate [pH 7.2], 10mM $CoCl_2$, 1mM DTT)
	2.5 µl 1mM dATP
	1.5 µl TdT (14 U/µl)
	6 µl sterile distilled water
total	50 µl

The mixture was incubated at 37°C for 30 min. The reaction was stopped by shaking with phenol/chloroform and the DNA in the aqueous phase was precipitated by addition of sodium acetate and ethanol as described above. The resulting DNA was

amplified by the *Taq* DNA polymerase chain reaction (PCR) procedure I (as described below) using PS3-5'CCGGAGACTTGAATGCCC 3' and 5'TTAAGC$(T)_{15}$ 3' as primers.

Polymerase Chain Reaction (Saiki *et al.*, 1988).

a) Procedure I

The ssDNA products of the first-strand cDNA synthesis were amplified by PCR using *Taq* DNA polymerase. The resulting DNA was dissolved in 80 µl of distilled water and denatured by heating to 94° C together with 1 µl each of two primers (complementary to the 3'and 5'-termini) for 2 min. A reaction mixture containing 2 µl 10mM dNTPs, 0.5 µl *Taq* polymerase (2.5 U/µl), 1 µl [α^{-32}P]dATP, 10 µl 10X PCR buffer (500mM Tris-HCl [pH 8.3], 60mM $MgCl_2$, 10mM DTT, 400mM KCl) was added and volume adjusted to 100 µl by adding distilled water. The mixture was overlaid with 100 µl paraffin oil, and subjected to 30 temperature cycles using a IHB 2024 programmable heating block (Cherlyn Electronic Ltd); 94°C for 30 sec to disassociate double-strands, 37°C for 30 sec for annealing, 70°C for 2 min. 5 µl of the reaction mixture was analysed in a 1-2% agarose gel and the DNA fraction found using a UV transilluminator after staining with EtBr. The amplified material was used for restriction enzyme hydrolysis, cloning, and sequencing after purification by phenol/chloroform extraction and ethanol precipitation.

b) Procedure II

For this work, the PCR method was used for amplifing genes either from ssDNA or dsDNA, for example the inserted genes in bacteriophage or plasmid vectors, as described by Güssow and Clarkson (1989). The amplification was done in 20 µl aliquots of a reaction mixture which comprised 10mM Tris-HCl (pH 8.3 at 25°C), 50mM KCl, 1.5mM $MgCl_2$, 0.1 mg/ml gelatine, 0.25mM dNTPs, about 100 ng of each oligonucleotide primer, 2.5 U *Taq* DNA polymerase (Perkin Elmer Cetus), and the template 0.2 µg DNA. The mixture was overlaid with 100 µl paraffin oil, and subjected to 30 temperature cycles using a IHB 2024 programmable heating block (Cherlyn Electronic Ltd); 94°C for 1 min to disassociate double-strands, 55°C for 1 min for annealing, 72°C for 2 min (or 3 min for inserts larger than 2.5 kb) for strand synthesis and finally 5 min at 55°C. 5 µl of the reaction mixture was analysed in a 1-

2% agarose gel and the DNA fraction found using a UV transilluminator after staining with EtBr. The amplified material was used for restriction enzyme hydrolysis, cloning, and sequencing.

c) Procedure III

This was a one-step amplification of transcripts in total RNA using the PCR method described by Goblet *et al.*(1989). The genomic RNA of 0.5 µg ELV was mixed with 100 ng of each of two primers in 30µl sterile water. The mixture was heated at 65°C for 15 min and then cooled on ice. The annealed RNA was amplified in a reaction mixture of 20µl 2.5x PCR buffer (168mM Tris-HCl (pH 8.8 at 25°C), 41.5mM $(NH_4)_2SO_4$, 16.8mM $MgCl_2$, 25mM β-mercaptoethanol, 0.42 mg/ml BSA), 22 units AMV reverse transcriptase, 1 unit *Taq* DNA polymerase and 0.5mM dNTPs, and overlaid with paraffin oil. After annealing at 42°C for 15 min, the mixture was subjected to 40 temperature cycles: 92°C for 1 sec, 55°C for 2 min, 72°C for 2 min. In the last cycle the mixture was kept at 72°C for 5 min to ensure that all DNA fragments were completed. 5 µl of each sample was electrophoresed in an agarose gel.

PLASMID AND BACTERIOPHAGE CLONING

The vectors I used most frequently were bacteriophages M13mp18 RF and M13mp19 RF and plasmids pGEM7Zf, pTZ18U and pTZ19U as well as Bluescript SK (+). Both the cDNA and vectors were hydrolysed with appropriate restriction endonucleases to generate compatible ends for cloning. Either a single restriction enzyme or two enzymes was used to prepare cDNA fragments, the vector was treated with calf intestinal alkaline phosphatase (CIP) to remove 5'phosphate groups and thus prevent recircularization of the vector during ligation. The cloning procedures used were essentially as described by Maniatis *et al.* (1982).

Preparation of Vectors for Cloning

The M13 RF or pTZ vectors (1 µg) were hydrolysed by the six base restriction endonucleases *Acc* I and *Sma* I in 20 µl reaction volumes of the appropriate buffer and incubated for 60 min in the conditions recommended by the supplier. To dephosphorylate the 5'-end of the linearised vectors with CIP, 0.1-0.4 unit of alkaline

phosphatase (Promega) was added directly to the mixture, together with 5 µl of 10x CIP buffer (500mM Tris-HCl [pH 9.0], 10mM $MgCl_2$, 1mM $ZnCl_2$, and 10mM spermidine) and then the volume was adjusted to 50 µl with sterile water. dsDNA fragments with asymmetrical 'sticky' ends were incubated at 37°C for 30-60 min, whereas those with symmetrical 'blunt' ends were incubated for a further 15 min at 56°C. The vector DNA was purified by phenol/chloroform extraction and ethanol precipitation as described previously. The DNA pellet was resuspended in 50 µl TE buffer (10mM Tris-HCl, 1mM EDTA, pH 8.0) to give a final DNA concentration of about 20 ng/µl.

Preparation of DNA for Insertion into Vectors

The dsDNA synthesised by the methods previously described was hydrolysed using a number of four base restriction endonucleases; *Hae* III, *Hinf* I, *Msp* I, *Rsa* I, *Sau*3A I, and *Taq* I. The DNA fragments produced by *Hinf* I and *Sau*3A I hydrolysis have 5'-overhangs that were 'end-filled' by adding 0.5 U of Klenow large fragment of *E. coli* DNA polymerase I (Boehringer-Mannheim) and 0.5mM dNTPs directly to the reaction mix after hydrolysis and it was further incubated at 37°C for 15 min, before 0.5 volume of formamide, double-dye loading buffer was added. The fragments were purified and fractionated by electrophoresis in 5% polyacrylamide gel in 1x TBE buffer. The individual DNA species (size range between 150-650 basepairs) were detected by autoradiography, and were excised and eluted from the gel slices by soaking for 6 to 12 hours in 0.5 ml of extraction buffer (10mM Tris-HCl [pH 7.4], 1% SDS), and the dsDNA was then precipitated in sodium acetate and ethanol.

Ligation of DNA Fragments into Vectors

Usually equimolar amounts of vector and insert were used in ligation reactions. Thus for a 3.0kb plasmid and a 0.5kb DNA insert fragment 1.0 µg vector was ligated with 0.167 µg insert DNA. I ligated together the *Sma* I-linearised vectors with the end-filled or blunt-ended dsDNA fragments resulting from hydrolysis with *Hae* III, *Hinf* I, *Sau*3A I, and *Rsa* I endonucleases. The *Msp* I and *Taq* I hydrolysed fragments were ligated into *Acc* I-linearised vector molecules.

The ligations were done in 10 µl mixture containing 50 mM Tris-HCl (pH 7.4), 10mM $MgCl_2$, 10mM DTT, 0.5mM rATP, 1 unit of T4 DNA ligase.

Thus the reaction mixture was prepared as:

1 µl vector (20 ng/µl)

1 µl 10x ligase buffer (500mM Tris-HCl [pH 7.4],

100mM $MgCl_2$)

1 µl 0.1M DTT

1 µl 10mM rATP

1 µl T_4 DNA ligase (1 U/µl, New England Biolabs)

1 µl DNA fragments (3.5 ng/µl)

the mixture was adjusted to a volume of 10 µl by adding sterile distilled water and incubated either at 4°C overnight, at 15°C for 4-6 h, or at 25°C for 1 h.

Transformation

Many different methods have been devised to transform M13 bacteriophage and pTZ plasmid vectors into *Escherichia coli*. In Procedure I described below, *E. coli* K12 strain JM101 was regularly and efficiently transformed by M13, whereas procedure II using $CaCl_2$-treated cells was best for transforming by pTZ plasmids.

a) Preparation of Competent Cells by Hanahan's Method (1983)

50 µl of *E. coli* JM101 frozen in glycerol were inoculated into 35 ml of 2x YT medium (16 g Bactotryptone, 10 g yeast extract, 5 g NaCl per litre). The cells were aerated by shaking in an incubator at 37°C and 250 rpm until the cell suspension reached an optical density of 0.3-0.6 in 600 nm wavelength light (about 3-4 h), they were then placed into a 40 ml sterile centrifuge tube and left on ice for 10 min. The cells were centrifuged for 5 min at 3,000g at 4°C, resuspended gently in 10 ml cold TFB (10mM potassium methyl sulphoxide [pH 6.2], 45mM $MnCl_2$, 10mM $CaCl_2$, 3mM hydroxyamine cobalt chloride), and then incubated on ice for 5 min. After centrifuging as above, the cells were resuspended in 3.5 ml cold TFB and kept on ice for 5 min before adding 130 µl cold deionised dimethyl formamide (DMFO). The mixture was left on ice a further 5 min, then 7 µl 2-mercaptoethanol was added and the cells were left on ice another 5 min before finally adding 150 µl DMFO. The cells were further kept on ice for 5 min until used.

b) Preparation of Competent Cells by $CaCl_2$ Treatment.

E. coli JM101 were grown and pelleted as described in the first step of procedure I. Then cells were gently resuspended in 4 ml of cold 50mM $CaCl_2$. After 30 min incubating on ice, the cells were competent for transformation.

c) Procedure for Transfecting Bacterial Cells with Bacteriophage DNA.

5 µl of the M13 ligation reaction mixture was added to 150 µl of JM101 competent cell suspension. The mixture was kept on ice for at least 15 min before being heat-shocked at 42°C for 2 min and then was immediately added to 3 ml of warm 2x YT containing 0.7% agar, together with 20 µl of 0.1M isopropyl βl-D-thio-galactopyranoside (IPTG) and 10 µl of 4% 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside (X-gal). After mixing well, the cell supension mixture was quickly spread onto minimal medium plates (15g bacto-agar, 10.5g K_2HPO_4, 4.5g KH_2PO_4, 1g $(NH_4)_2SO_4$, 0.5g Na_3citrate.$2H_2O$ per litre, supplemented after autoclaving with 0.8 ml of 1M $MgSO_4$, 0.5 ml of 1% Thiamine-HCl and 10 ml of 20% (w/v) glucose) and then incubated at 37°C overnight to select transformants .

d) Procedure for Transforming Bacterial Cells with Plasmid DNA.

100 µl of competent cells and 5 µl of a ligation mixture containing plasmids were mixed and heat-shocked as described above. After incubating on ice for 5 min, the cell suspension was diluted with 0.8 ml of LB (5 g yeast extract, 10 g tryptone, 5 g NaCl per litre) and shaken at 37°C for 45 - 60 minutes. Then 100 µl of this suspension plus 20 µl of 0.1 M IPTG and 30 µl of 4% X-gal was plated onto LB agar plates containing 80 µg/ml ampicillin to select the transformants. The plates were incubated overnight at 37°C.

PLASMID AND BACTERIOPHAGE PREPARATION

Vectors, M13 and pTZ, both contain the gene for the α-peptide of β-galactosidase and thereby permit blue/white colour selection of recombinants. Whereas clones containing DNA inserted in the polylinker site of the vector will not produce active β-

galactosidase. and thus the bacterial colonies of these clones will be colorless or, for bateriophages, clear areas of reduced bacterial growth.

The single- or double-stranded DNA cloned in this way was isolated and purified in the following ways.

Bacteriophage Single-stranded DNA Preparation

A single colourless plaque was sampled using a sterile toothpick and inoculated into 2 ml of 2x YT liquid medium containing either 0.1 volume of overnight *E. coli* JM101 culture, or fresh *E. coli* JM101 cells grown for 60 min. The infected bacterial cell culture was incubated at 37°C for 5-6 hours with shaking (250 rpm). Then the cells were removed by centrifuging twice in an Eppendorf centrifuge at top speed (about 15,000 g) for 5 minutes. The bacteriophage particles in the supernatant were precipitated by adding 250 µl of a mixture of 25% polyethylene glycol (PEG, M_r 8,000) and 2.5M NaCl to 1.25 ml of the supernatant and then the mixture was kept on ice for at least 30 min. The bacteriophage particles were concentrated by centrifuging for 5 min and resuspended in 100 µl of TE. To extract single-stranded genomic DNA, 50 µl of TE-saturated phenol was added to the viral suspension. The mixture was then vortexed twice for 10 seconds, and left on ice for 5 min before being vortexed again. The phases were separated by centrifuging for 2 min and the viral DNA in the aqueous phase was precipitated using ethanol and sodium acetate. The DNA pellet was resuspended in 30 µl of TE.

Single-stranded DNA Preparation from Double-stranded Plasmid Vectors

2 ml of 2x YT containing 80 µg/ml ampicillin was inoculated with a single colony of JM101 harbouring the pTZ plasmid clone followed by 1 µl of helper bacteriophage M13K07 (7.5 x 10^{10} pfu/ml). After incubating at 37°C for at least one hour, kanamycin was added to a concentration of 70 µg/ml. The incubation was continued overnight. The single-stranded plasmid DNA was purified by the same method as the bacteriophage ssDNA preparation described above.

Double-stranded DNA Plasmid Preparation on a Small Scale

Two methods, described below, were used to prepare plasmids of pGEM7Zf, pTZ18U, pTZ19U and Bluescript SK(+) vectors.

a) Procedure I: Gentle Lysis Method (Holmes and Quigley, 1981).

Plasmid DNA can be isolated by this method, which I used most frequently, in under 2 hours. It produces DNA that is readily hydrolysed by restriction enzymes, and which can be ligated and retransformed.

A single transformant white colony was sampled using a sterile toothpick and inoculated into 2 ml of LB containing 80 µg/ml ampicillin and grown at 37°C to the middle of the logarithmic phase of growth (about 4 hours) or overnight with aeration by shaking. Then the bacterial cells were pelleted by centrifuging for 1 min in Eppendorf tubes. After resuspension in 200 µl of STET buffer (8% sucrose, 5% Triton X-100, 50mM EDTA and 50mM Tris-HCl [pH 8.0]), the cells were lysed by adding 15 µl of a 10 mg/ml stock of freshly prepared lysozyme (in 0.25M Tris-HCl [pH 8.0], or TE [pH 8.0]) and placing in a boiling water bath for 45-60 seconds. The degraded cells were immmediately centrifuged in an Eppendorf centrifuge for 10 min. The plasmid DNA in the supernatant was precipitated by adding an equal volume of cold isopropanol and incubated at -20°C for 5 min. The DNA was sedimented, washed and dried as described previously. The pellet was resuspended in 50 µl of sterile distilled water, incubated at 65°C for 10 min and stored at -20°C.

b) Procedure II: Alkaline Lysis Method

The transformed bacteria from a single colony were grown and sedimented in the same way as in procedure I. The pellet of cells was resuspended in 200 µl of GTE (50mM glucose, 25mM Tris-HCl [pH 8.0], 10mM EDTA) before adding 400 µl of freshly made lysis solution (0.2M NaOH, 1% SDS). After incubating on ice for 5 min, 200 µl of an ice-cold solution of 3M potassium acetate/ 2M acetic acid was mixed with the lysed cell suspension and the mixture incubated a further 15 min on ice. The chromosomal DNA and bacterial debris were removed by centrifuging for 5 min in an Eppendorf centrifuge. The plasmid DNA in the supernatant was precipitated by adding 0.6 volume (480 µl) of isopropanol). After centrifugation and washing with 95% ethanol, the DNA pellet was dried and resuspended in 50 µl TE.

Preparation of Large Amounts of Double-stranded DNA

The procedure for preparing large amounts of plasmid DNA was the same in principle as the alkaline lysis method described above. 500 ml LB was inoculated with 15 ml of a cell suspension which had been grown overnight from a single colony. After scaling up the alkaline lysis method, the DNA was further purified by centrifuging it to equilibrium in caesium chloride-ethidium bromide density gradients. The plasmid DNA obtained by precipitating with isopropanol was prepared for centrifuging by resuspending it in 14.4 ml of TE and 246 µl of 2M Tris base, then 15 g CsCl and 1.5 ml ethidium bromide (10 mg/ml) was added. The mixture was then put into six quick-seal centrifuge tubes (Beckman, size 13x32mm) and centrifuged at 100,000 rpm for 5-12 hours at 18-20°C (Beckman, TLA 100.3 rotor, TL-100 ultracentrifuge). The lower band, which was of closed circular plasmid DNA, was collected. The ethidium bromide was removed by shaking four times with an equal volume of 5M NaCl-saturated isopropanol, until the aqueous solution was no longer pink. The preparation was then dialysed against TE at 4°C overnight, and the DNA precipitated in 2.5 volumes 95% ethanol and 0.5 volume 7.5 M ammonium acetate. After centrifugation, the pellet was washed with 75% ethanol, allowed to dry and resuspended in TE.

HYBRIDIZATION FOR TRANSFORMANT IDENTIFICATION

To determine which recombinant clones contained viral insert DNA, dot-blot hybridization tests were done with radioactively-labelled probes using the method described by Maniatis (1982).

Radioactively-labelled Probe Preparation by Nick Translation

dsDNA complementary to the genomic RNA of ELV, was synthesized by PCR method similar to that previously described using synthetic oligodeoxyribonucleotide primers. dsDNA was nicked by adding 5 µl of pancreatic DNase I (200 ng/ml) in a mixture containing 50mM Tris-HCl (pH 7.5), 7.5mM Mg-acetate, 6.6mM DTT, 0.1 µg BSA and 3µM dNTPs. In addition, *E. coli* DNA polymerase I was used for end-filling the 'recessed' 3'-termini of nick translated DNA in the presence of DNase I.

Thus the reaction mixture was prepared as:

20.0 µl dsDNA

3.5 µl 10x buffer (0.5M Tris [pH 7.5], 75mM Mg-acetate)

2.0 µl 0.1M DTT

3.5 µl BSA (1 mg/ml)

1.0 µl 0.1mM dNTPs

5.0 µl DNase I (200 ng/ml)

The mixture was incubated at 14°C for 15 min, then 70°C for 5 min and finally cooled on ice or cold water. [α-^{32}P]dCTP and [α-^{32}P]dATP were incorporated into the nicked DNA by adding 1 µl of each radioactive nucleotide and 1 µl of *E. coli* polymerase I (holoenzyme, Boehringer-Mannheim)(5 U/50 µl reaction mixture). The mixture was incubated at 14°C for 30-45 min, and the reaction was then stopped by adding 4 µl of 10x stop buffer (125mM EDTA and 5% SDS), then with 3 µl of 4M HCl, incubated for 10 min at room temperature and finally neutralized with 4 µl of 4M NaOH and further incubated for 10 min. The labelled cDNA was fractionated in a Sephadex G-50 column and fractions, each of 200 µl, were collected.

Colony Hybridization

a) Lysis of Transformant Colonies and Binding of their DNA to Membrane Filters

Colonies carrying recombinant plasmids were transferred, using toothpicks, onto LB agar plates containing 80 µg/ml ampicillin (two plates for each colony). Small streaks 2-3 mm in length, or dots of inoculum arranged in a grid pattern, were replicated on the two plates. Up to 50 colonies were transferred onto a single 90mm plate. The plates were inverted and incubated at 37°C overnight. The bacteria were transfered to BA 85 nitrocellulose membrane filters (pore size 0.45 µm; Schleicher & Schuell) by laying the membrane, with sterile blunt-ended forceps, onto the surface of the culture plate and, when the filter was thoroughly wet, it was peeled from the plate. The bacteria were lysed by laying the transfer membrane for 9 min, colony side up, on the surface of 3MM paper soaked with 0.5M NaOH, and then the alkali neutralized by transferring the filter to the second sheet of 3MM paper soaked with 1M Tris-HCl (pH 7.4) for 5 min repeating this step. The lysed cells were washed by transferring the filter to a fourth 3MM paper, this one soaked with 1.5M NaCl and 0.5M Tris-HCl (pH 7.4) for 5 min. The filter, colony side up, was laid on a sheet of dry 3MM paper and

allowed to dry at room temperature for at least 30 min. Then the filter was sandwiched between two sheets of 3MM paper and baked at 80°C for 2 hours in a vacuum oven.

b) Hybridization Protocols.

i) Prehydrization: The baked filter was floated on 6x SSC (52.60g NaCl and 26.46g Na_3-citrate per litre) until it was thoroughly wetted and then submerged for 2 min. The filter was transferred to a hybridization bottle (HB-OV-BS, Hybaid) or bag containing 10-15 ml of prehybridization buffer (50mM HEPES, 3x SSC, 100 µg/ml denatured salmon sperm DNA, 0.1% SDS and 1x Denhardt's solution which comprised 0.02% each of Ficoll, polyvinylpyrrolidone and bovine serum albumin). The filter was incubated for 1-2 hours at 65°C with gentely shaking or incubated in the Hybaid mini hybridization oven in the same conditions.

ii) Hybridization: After excess solution was removed from the hybridization bag or bottle, it was filled with 10-15 ml warm hybridization buffer. The ^{32}P-labelled dsDNA probe was denatured by adding an equal volume of deionised formamide and heating for 3-5 min to 100°C, chilled rapidly on ice, and then added to the hybridization buffer. The filter and probe were incubated together at 65°C for 2-16 hours. Then, the filter was washed in a large volume of 2x SSC at room temperature for 5 min and then transferred to 2x SSC solution and gently agitated for 15 min. The filter was further washed with 1x SSC and 0.1% SDS and finally rinsed with 0.1x SSC. At no stage during the washing procedure was the filter allowed to dry. After washing the filter was air-dried at room temperature and exposed to X-ray film (Kodak XAR-5 or equivalent) for 12-14 hours at -70°C with an intensifying screen. Bacterial colonies corresponding to those that hybridized with the probe were collected from the replicate plates.

Dot-blot Hybridization Procedure

The dot-blot hybrization procedures were used to detect DNA or RNA directly. However, the buffer used for DNA-RNA hybridization was different from the DNA prehybrization buffer; it was 40% deionised formamide, 3x SSC, 100 µg/ml denatured salmon sperm DNA, 50mM HEPES buffer (pH7.0), and 5x Denhardt's solution. For DNA-RNA hybridization, the nitrocellulose membrane was pre-wetted with water, then with 2x SSC and finally allowed to air dry before prehybrization.

To detect M13 ssDNA, 1µl of DNA was spotted onto the membrane and then allowed to dry in the air. The nucleic acid was bound to the membrane by baking it at 80°C for 2 hours and the ssDNA detected using the same hybridization protocols as described above.

DNA SEQUENCING

Single-stranded DNA, obtained from recombinant M13 clones, or generated from pTZ plasmids in the presence of M13KO7 helper phage, was the most suitable for sequencing by the dideoxynucleotide chain termination method (Sanger *et al.*, 1977, 1980). This usually involved *in vitro* synthesis of a DNA strand by a modified T7 DNA polymerase using a ssDNA template as descibed by Tabor and Richardson (1987) in the Sequenase^{TM} system of United States Biochemical Corporation. However, double-stranded DNA was also sequenced effectively using the Sequenase^{TM} system.

Denaturated Double-strand DNA Preparation for DNA Sequencing

The alkaline-denaturation method worked well when 3-5 µg of RNA-free plasmid DNA was used as the template, but not so well with smaller amounts. The purified DNA (50 µl in TE), prepared using the small scale alkaline lysis method previously described, was hydrolysed with 10 µg/ml of RNase A (Sigma) at 37°C for 30 min. The DNA fragments were precipitated by adding 30 µl of a mixture of 20% PEG 6,000 and 2.5M NaCl, and keeping the mixture on ice for an hour. The DNA was pelleted and washed with ethanol. The pellet by centrifugation was dissolved in 70 µl of water or TE. 35 µl of the resulting dsDNA solution was denatured in 0.2M NaOH and 2mM EDTA at 37°C for 30 min, the mixture was neutralized by adding 0.1 volume of 3M sodium acetate (pH4.5-5.5) and the DNA precipitated with 2.5 volume of ethanol. After centrifugation and washing the pellet with 70% ethanol, it was dissolved in 7 µl of distilled water and 2 µl of Sequenase reaction buffer and 1 µl of primer were immediately added.

DNA Sequencing Protocols

The kit of United States Biochemical Corporation for DNA sequencing with Sequenase^{TM} was used in this study with the recommended protocol.

a) Annealing Template and Primer

7 µl of single-stranded DNA solution was added to the same amount of reaction buffer and primer as for the double-strand DNA annealing mixture as described above. Then the capped tubes of both ss-and dsDNA were warmed to 65°C for 2 min, placed at room temperature on the bench to cool slowly in heating block over a period of about 30 min. When the temperature was below 30°C, the tubes were placed on ice. The annealed template was used within about 4 hours.

b) Labelling Reaction

The annealed template-primer DNAs were added to the labelling reaction mixture which comprised:

10.0 µl template-primer DNA
1.0 µl 0.1M DTT
2.0 µl diluted labelling mix (1:5 in water)
0.5 µl [α-^{35}S]dATP
2.0 µl diluted Sequenase (1:8 in TE)

These were mixed thoroughly and incubated for 2-5 min at room temperature. When, in the subsequent sequence analyses, band compressions were found, the dITP labelling mix was substituted for the dGTP mix.

c) Termination Reaction

2.5 µl of ddGTP, ddATP, ddTTP and ddCTP termination mix was added to the four set of tubes labelled G, A, T, and C as appropriate. These sets of four tubes were pre-warmed at 37°C for at least 1 min, then 3.5 µl of the labelled DNA mixture was added in each tube of the set, mixed well, centrifuged briefly and kept at 37°C for 3-5 min. The reactions were terminated by adding 4 µl of 'stop solution', mixed and stored on ice.

d) Denaturing Gel Electrophoresis for Sequence Determination

The sample mixtures were heated to 75°C for 2-4 min, immediately before loading onto a 6% polyacrylamide/ 2x TBE/ 7M Urea (0.25mm thick) gel. After electrophoresis at 30mA for about 3 hr, when the dye marker had migrated 35 cm, the sequencing gel was dried on to Whatman 3MM filter paper in a vacuum dryer at 80°C for an hour before being autoradiographed using Kodak RP or Kodak XAR-5 film.

e) Determining the Nucleotide Sequence Adjacent to that Part which Hybridizes with the Primer

It was found that adding Mn^{2+} to normal (Mg^{2+})sequencing reaction mixtures decreased the average length of DNA synthesized and intensified the bands corresponding to nucleotides close to the primer. Thus, 1 µl of Mn buffer was added to the normal sequencing reaction mixture in the labelling step before adding Sequenase. No other changes were found to be necessary.

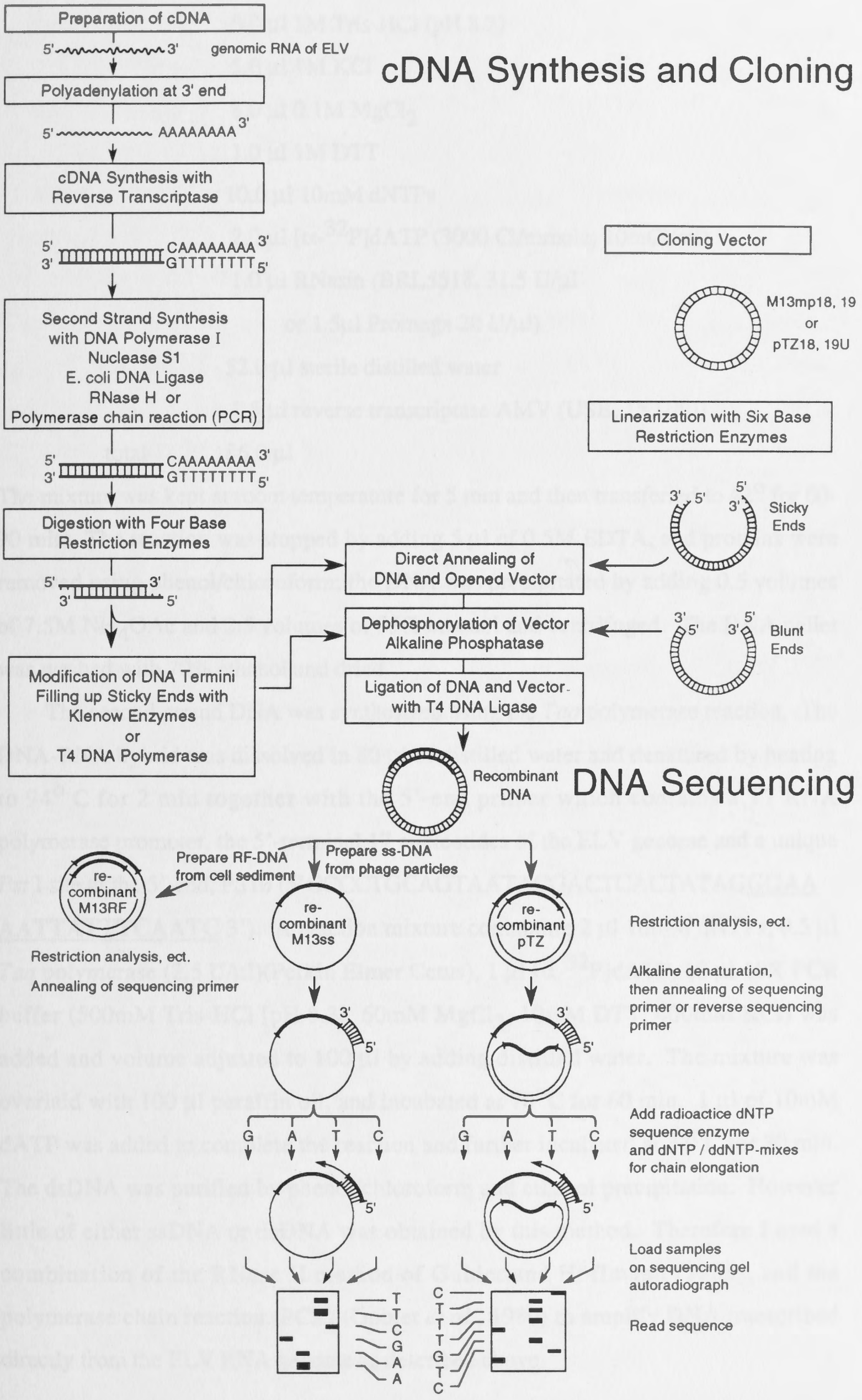
PRODUCTION OF A dsDNA ENCODING THE ELV GENOME

dsDNA clones encoding the genome of ELV were synthesized by various methods.

First attempts were made by the well-tried method of Gall *et al.* (1988), who used methyl mercuric hydroxide treated RNA to synthesize a full-length DNA copy of the genome of grapevine chrome mosaic virus. This method has been used successfully to prepare full-length DNA clones of TYMV genomes (Anne Mackenzie, personal communication). cDNA to the ELV genome was synthesized in two ways. First about 15 µg of viral genomic RNA was used together with 7.5 µg of the 3'-end PS9 primer, 5'CCCGGATCC<u>TGGGGTGTCCCTTTACC</u> 3', which corresponds to the last 17 nucleotides of the 3'-terminus of ELV and also produces an extra unique *BamH* I site. In the other method polyadenylated ELV RNA and dT_8dG primer was used. Each treated in the same way, first they were denatured in 20mM methyl mercuric hydroxide by preparing a mixture of 7 µl of 40mM CH_3HgOH, 1.5 µl of ELV RNA solution (10 µg/µl), 2 µl primer (3.5 µg/µl) and 3.5 µl of sterile distilled water. The reaction mixture was incubated at room temperature for 10 min, in a fume hood, and to it was added the reverse transcriptase mixture consisting of:

Figure 7. Schematic diagram of the synthesis and cloning of double-stranded DNA to the ELV RNA genome and DNA sequencing of the clones. These procedures were as described by Gubler and Hoffman (1983) for dsDNA synthesis; Goblet *et al.* (1989) and Güssow and Clarkson (1989) for PCR; Maniatis *et al.* (1982) for cloning method and Sanger *et al.* (1977) for DNA sequencing.

cDNA Synthesis and Cloning



DNA Sequencing

	5.0 µl 1M Tris-HCl (pH 8.3)
	5.0 µl 1M KCl
	8.0 µl 0.1M $MgCl_2$
	1.0 µl 1M DTT
	10.0 µl 10mM dNTPs
	2.0 µl [α-^{32}P]dATP (3000 Ci/mmole; 10mCi/ml)
	1.0 µl RNasin (BRL5518, 31.5 U/µl
	or 1.5µl Promega 20 U/µl)
	52.0 µl sterile distilled water
	1.5 µl reverse transcriptase AMV (USB, 15 U/µl)
total	86.0 µl

The mixture was kept at room temperature for 5 min and then transferred to 42° for 60-90 min. The reaction was stopped by adding 5 µl of 0.5M EDTA, and proteins were removed using phenol/chloroform; the DNA was precipitated by adding 0.5 volumes of 7.5M NH_4OAc and 2.5 volumes of 95% ethanol and centrifuged. The DNA pellet was washed with 70% ethanol and dried.

The second-strand DNA was synthesized using the *Taq* polymerase reaction. The DNA-RNA hybrid was dissolved in 80 µl of distilled water and denatured by heating to 94° C for 2 min together with the 5'-end primer which contains a T7 RNA polymerase promoter, the 5'-terminal 17 nucleotides of the ELV genome and a unique *Pst* I site at the 5'-end, PS10 (5'CCCCTGCAGTAATACGACTCACTATAG**GGAA AATTATTTCAATC** 3'). A reaction mixture containing 2 µl 10mM dNTPs, 0.5 µl *Taq* polymerase (2.5 U/µl)(Perkin Elmer Cetus), 1 µl [α-^{32}P]dATP, 10 µl 10X PCR buffer (500mM Tris-HCl [pH 8.3], 60mM $MgCl_2$, 10mM DTT, 400mM KCl) was added and volume adjusted to 100 µl by adding distilled water. The mixture was overlaid with 100 µl paraffin oil, and incubated at 70°C for 60 min. 1 µl of 10mM dATP was added to complete the reaction and further incubated at 37°C for 30 min. The dsDNA was purified by phenol/chloroform and ethanol precipitation. However little of either ssDNA or dsDNA was obtained by this method. Therefore I used a combination of the RNase H method of Gubler and Hoffmann (1983), and the polymerase chain reaction (PCR) (Goblet *et al.*, 1989) to amplify DNA transcribed directly from the ELV RNA genome as described above.

An outline of the successful cloning procedure is shown in Figure 37 and Figure 38, and will be described in more detail here. For first-strand DNA synthesis, the ELV genomic RNA was denatured by heating to 94° C together with two oligonucleotide primers (dT_8dG and synthetic random oligonucleotide hexamers) and the primed ELV RNA was then used as a template to synthesize cDNA in a reaction volume of 40 µl containing 50mM Tris-HCl (pH 8.3), 8mM $MgCl_2$, 50mM KCl, 10mM DTT, 0.5mM each of dGTP, dATP, dTTP and dCTP, 2 µl [α-^{32}P]dATP (10 mCi/ml), and about 20-22 units AMV reverse transcriptase (USB). The second-strand DNA was synthesized by the RNase H method as previously described. The resulting dsDNA was hydrolysed with *Sph* I and then ligated into the *Sph* I site and *Sma* I / *Sph* I sites of pTZ18U and pTZ19U vectors. The resulting clones were called pELV25 (nucleotides 2142-3600), pELV37 (nucleotides 1-2142) and pELV40 (nucleotides 2461-4689). The part of the genome between nucleotides 4186-5494 was obtained by priming genomic RNA with PS7 primer (5'AGAGGGAATTCGAACTGGAA 3'; complementary to nucleotides 5484-5504) for the first-strand DNA synthesis using the method described above, the second-strand DNA synthesis using *Taq* polymerase reaction and the PS10 as previously described. The resulting DNA fragments were hydrolysed with *EcoR* I / *Xba* I and were cloned into the same restriction enzyme sites of pGEM7Zf vectors to give clone pELV2 (nucleotides 4186-5494).

Construction of Full-length ELV DNA

The 5'-end of the genome was constructed in the following way (Figure 37a). An *EcoR* I / *Sph* I fragment of pELV37 was ligated to a *Sph* I / *Kpn* I fragment of pELV25 and then cloned in the *EcoR* I / *Kpn* I site of the pTZ19U vector. The resulting clones were called pELV20 (nucleotides 1-3600). Next, the pELV20 was hydrolysed with *EcoR* I and the resulting fragment was inserted into the same restriction enzyme site of pELV40 to produce pELV31 (nucleotides 1-4689).

A 5'-substitution of pELV31 (Figure 37b) was prepared using the primer PS10 containing T7 RNA polymerase promoter. pELV31 was hydrolysed with *Kpn* I and converted to a blunt-ended molecule by adding 1 µl of T_4 DNA polymerase (1 U/µl) (Boehringer-Mannheim) and 0.5mM dNTPs directly to the reaction and further incubated at 37°C for 30 min after hydrolysis. The fragments were purified by phenol/chloroform and ethanol as described above and then hydrolysed with *Apa* I.

The dsDNA fragment corresponding to nucleotides 1 to 2056 of the ELV genome was prepared using PCR method of Güssow and Clarkson (1989) as described above with pELV37 as a template and PS2 (complementary to nucleotides 2037-2056) and PS10 (nucleotides 1-17) as primers. The resulting dsDNA was hydrolysed with *Apa* I and ligated to the blunt-end / *Apa* I site of pELV31. This substituted pELV31 clone was called pELV38 (nucleotides 1-4689).

The 3'-end of the genome was constructed in a similar way to the 5'-end (Figure 37c). A clone encoding the 3'-end was obtained by ligating two products of the PCR method of Güssow and Clarkson (1989) using the M13mp18 clones originally obtained for sequencing of ELV genome; these were ELV9 (nucleotide 5373-5768) and ELV24 (nucleotide 5745-6034). M13 primer together with pPUC/M13 reverse primers (Bresatec and Promega) were used. The two fragment products were digested with *BstE* II / *Sph* I and cloned in the *Sph* I site of pTZ19U. The resulting clone was called pELV5 (nucleotides 5373 to 6034). Clones of pELV5 and pELV2 (nucleotides 5484-5504) were hydrolysed with *EcoR* I / *Xba* I and ligated into the *Xba* I site of the pGEM7Zf vector to give clone pELV3. The pELV3 clone was substituted at the 3'-end of the ELV sequence (Figure 37d) using the PCR method of Güssow and Clarkson (1989) and two primers; PS8-5'ATGGAAAGCTTGCCAAACTC 3'(nucleotides 4350-4370) and PS9 (nucleotides 6019-6034). The DNA products were cloned in *EcoR* V site of pBluescript SK(+) and the resulting clone was called pELV18 (nucleotides 4350-6034).

Finally two full-length clones of ELV, pELV35 and pELV81, were derived from the *Pst* I / *EcoR* V fragment of pELV31 ligated into the *Pst* I / *EcoR* V sites of pELV18 in pBluescript SK(+) (Figure 37e). Thus clones encoding ELV were obtained with T7 RNA polymerase promoters at both termini together with a *Pst*I site at the 5'-end and *BamH* I sites at both ends; the T7 RNA polymerase promoter at the 3'-end is part of the pBluescript SK(+) vector.

The cloning methods involved restriction endonuclease digestion, selection of nucleic acid fragments by agarose-gel electrophoresis, ligation, plasmid preparation, and DNA sequencing for gene identification as described above.

In vitro Transcription and Capping

The cap homologue, m^7GpppGp was directly incorporated into the 5'-end of RNA during the transcription reaction to yield capped RNA. 10µg each of the DNA clones of pELV35 and pELV81 for use as templates was linearised with *Pst* I / *BamH* I and transcribed with T7 RNA polymerase, as advised by Pharmacia, the supplier of the polymerase (Melton *et al.*, 1984). The DNAs were incubated in a 50 µl reaction mixture containing transcription buffer (40mM Tris-HCl [pH 7.5], 6mM $MgCl_2$, 2mM spermidine, and 10mM NaCl), 10mM DTT, 0.5 mg/ml BSA, 0.5mM ATP, CTP, UTP, 0.25mM GTP, 0.5mM m^7GpppGp (New England Biolabs), 1 Unit/µl of RNasin (Promega), and 150 units of T7 RNA polymerase (Pharmacia).

Thus the reaction mixture was prepared by mixing:

15.0 µl linearized DNA

10.0 µl 5x transcription buffer (200mM Tris-HCl [pH 7.5], 30mM $MgCl_2$, 10mM spermidine, and 50mM NaCl)

5.0 µl 0.1M DTT

2.5 µl BSA (1mg/ml)

2.5 µl RNasin ribonuclease inhibitor

5.0 µl 5mM ATP, CTP, UTP, 2.5mM GTP

5.0 µl 5mM m^7GpppGp

T7 RNA polymerase 3-10 U /µl of reaction mixture.

The mixture was incubated at 37°C for one hour, then an extra 1 U/µl of enzyme was added to boost the reaction and it was further incubated for 2 hours. The template DNA was removed by treating with 0.1 U/µl of deoxyribonuclease (RQ1 RNase-free DNase I, Promega) at 37°C for 30 min, then shaken with phenol/ chloroform and the RNA was precipitated with 3M ammonium acetate and ethanol. The pellets were resuspended in 50 µl of saline-phosphate buffer (PBS) and analysed by electrophoresis in a 4% polyacrylamide gel containing 7M urea and 1xTBE.

Biological Assay by Inoculation of Chinese cabbage plants

The ELV RNA transcipts were directly inoculated to Chinese cabbage (*B. campestris* sp. *pekinensis* L.) and swede (*B. napus* var.*napobrassica* (L.) Rchb.) which were kept in the dark for two days before inoculation. The plants were dusted with

carborundum powder and mechanically inoculated with either 40 µl of *in vitro* transcription mixes or, as controls, infected sap or 10µg of ELV RNA or viral particles in the same PBS buffer. The inoculated plants were kept in an insect-proof glasshouse at *c*.25° and examined daily for symptoms.

Detection of Viral RNA by Dot-blot Hybridization

ELV RNA extracted from infected leaves or sap from systemically infected leaves were tested by dot-blot analysis. Radioactively-labelled dsDNA was synthesized by nick translation of the PCR product of amplified DNA directly from ELV clones using PS9 and TALL-comp primers. This nick-labelled DNA was used for RNA detection as in the method described previously.

a) RNA extraction

The method of Verwoerd *et al.* (1989) was used to extract RNAs from infected tissues. Fresh infected leaf tissue was put in Eppendorf tubes, frozen quickly in liquid nitrogen and stored at -80°C until ready for testing. After grinding with a small steel bar (precooled in liquid nitrogen), the frozen leaf powder was mixed with 500 µl/tube of a heated (80°) extraction buffer which consisted of phenol and a buffer mixture of 0.1M LiCl, 100mM Tris-HCl (pH 8.0), 10mM EDTA and 1% SDS (1:1). The mixture was shaken for 30 seconds, then 250µl chloroform-isoamylalcohol(24:1) was added, shaken and finally centrifuged for 5 min. The aqueous phase was mixed with one volume of 4M LiCl overnight and the precipitated RNA was pelleted by centrifugation. The RNAs were precipitated with 0.1 volumes of 3M NaOAc (pH 5.2) and 2 volumes of cold ethanol. The RNA pellet was dissolved in water.

b) Dot-blot hybridization protocol

Samples were directly spotted onto nitrocellulose membranes (1 µl/spot), baked for 2 hours at 80°, and then the membranes were hybridized with a nucleotide probe complementary to ELV. These probes were produced, as described above, by nick translation from a dsDNA copy of the ELV genome. The hybridization method of Maniatis *et al.* (1982) was used and described above.

Identification of virion coat protein gene by PCR

RNA extracted from infected Chinese cabbage was directly amplified by PCR (Güssow and Clarkson, 1989). The VP gene was detected by the PCR method, as described above, using the PS9 and TALL-comp primers; the former, which primes the 3'-end of the genome has the sequence 5'CCCGGATCCTGGGGTGTCCCTTTACC 3' and the latter, which has the sequence 5'CCCTCGAGT[C/A]TGAATTGCTTC 3' primes the tymobox (Ding *et al.* 1990a). DNA products were analysed and fractionated by gel electrophoresis using the extracts of healthy plant as controls.

COMPUTER ANALYSIS

The nucleic acid and deduced amino acid sequences were analyzed using the SEQ library programs of the Research School of Biological Sciences, Australian National University and the University of Wisconsin Genetic Computing Group (UWGCG) Version 6 programs in a VAX computer.

The programs used included:

1. **The Staden set of 'shotgun' sequencing programs (Staden, 1982):** These record and compile the sequences obtained from sets of overlapping or complementary recombinant clones;

2. **MAPPLOT:** A GCG program that displays the positions of all of the cleavage sites for a chosen restriction endonuclease (Schroeder and Blattner, 1982);

3. **ISOELECTRIC:** A GCG program that plots the positive, negative and net charges of a protein of known composition as a function of pH;

4. **SQUIGGLES:** A GCG program that predicts the secondary structure of single-stranded nucleic acids and calculates their minimum free energy based on published values of stacking and loop destabilizing energies (Zuker and Stiegler, 1981);

5. **DOTPLOT:** A GCG program that compares pairs of sequences by dot-plots using the method of Maizel and Lenk (1981) or the very rapid algorithm of Sussman (Unger *et al.*,1986);

6. **GAP and LINEUP:** A GCG program that optimally aligns sequences by the method of Needleman and Wunsch (1970) and Sellers (1974) and allows editing of multiple alignments;

7. **FASTA and TFASTA:** A GCG program that compares a novel sequence with sets of known sequences in databases (Pearson and Lipman, 1988);

8. **FETCH**: A GCG program that copies data files from GCG database libraries into an active directory;

9. **PRETTY:** A GCG program that displays multiple sequence alignments and calculates a consensus sequence based on any of several different consensus functions;

10. **HETZYG:** A SEQ program that calculates the 'variability' at each position of a set of aligned sequences.

11. **Progressive alignment:** A program that progressively aligns a set of sequences (Feng and Doolittle, 1987);

12. **NJTREE:** A program which calculates a binary dendrogram describing the relationships of a set of objects from a matrix of their dissimilarities (such as the FJD distances (Feng *et al.*, 1985) calculated by the progressive alignment method), by the 'neighbour-joining' method of Saito and Nei (1987) as implememted by Studier and Keppler (1988);

13. **TDRAW:** A program which displays dendrograms calculated by NJTREE (J.W.H. Ferguson, personal communication);

14. **SEQCORR:** SEQ program to assess the correlation between amino acid or nucleotide sequence differences in a set of aligned sequences, and known differences in the properties of the organisms providing the sequences (i.e. host ranges).

CHAPTER 3 - PRIMARY STRUCTURE OF THE ELV GENOME

INTRODUCTION

The single-stranded RNA genome of ELV was sequenced as described in Chapter 2. ELV RNA, isolated from infected Chinese cabbage (*Brassica campestris* spp. *pekinensis* L.), was transcribed into double-stranded DNA and cloned in appropriate vectors for propagation and isolation using six different four-base restriction endonucleases *Dde* I, *Hae* III, *Hinf* I, *Msp* I, *Sau*3A I, and *Taq* I (Figure 8). cDNA obtained from the recombinant M13 clones or pTZ plasmids was sequenced using the chain termination method of Sanger *et al.*(1977) and the Sequenase^{TM} system. The nucleotide sequences of ELV cDNA clones were compiled and analysed using various computer programmes.

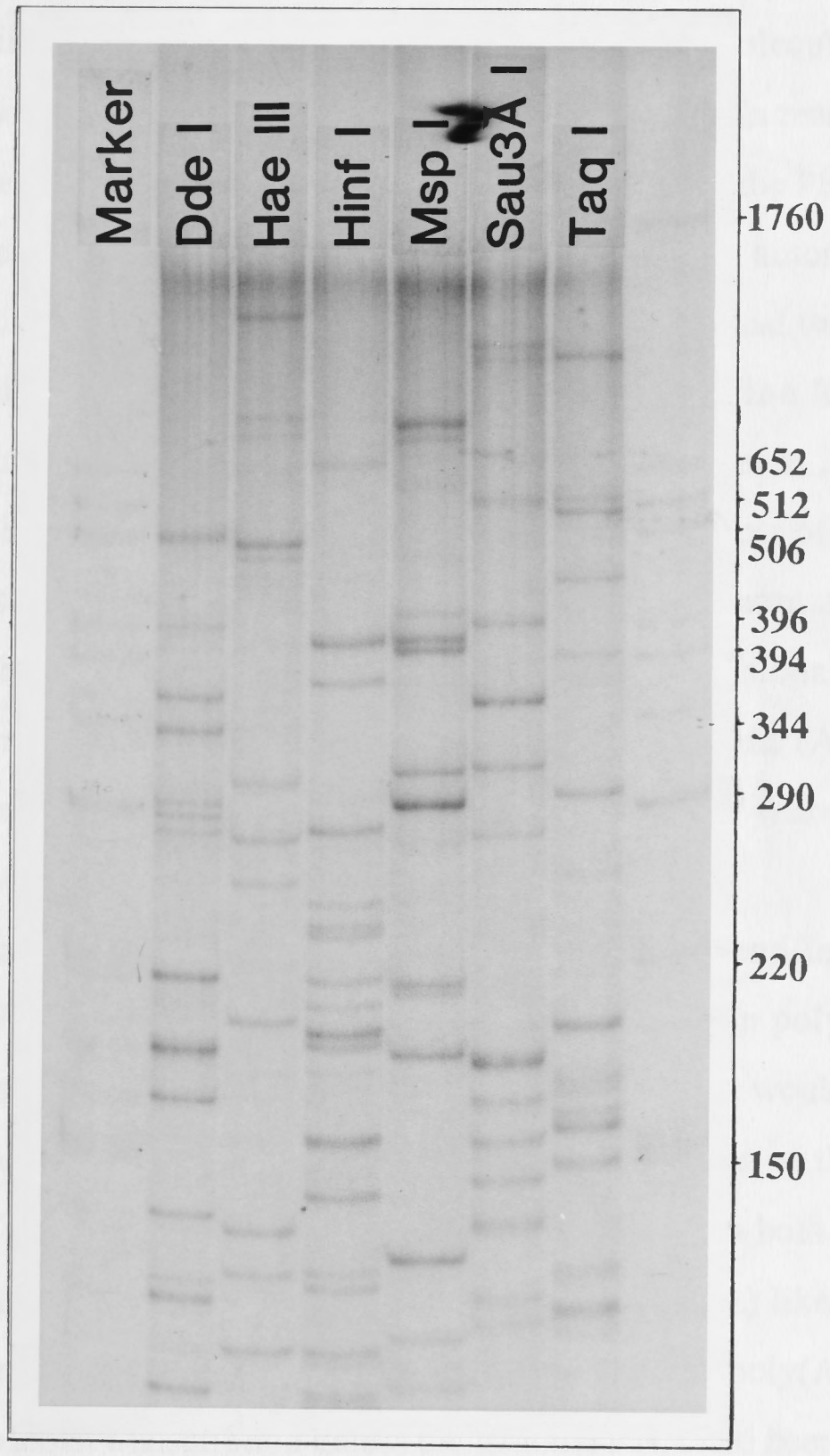
SEQUENCE OF THE ELV GENOME

The viral inserts in 276 randomly-selected M13 clones, mainly from *Hae* III and *Msp* I fragments of cDNA, were sequenced using the dideoxy chain termination method. Clones that were between 150 and 750 nucleotides in length were chosen as those up to 500 nucleotides long could be fully sequenced; longer ones were sequenced from both ends.

A total databank of *c.* 40,000 nucleotides was obtained and used to compile the genomic sequence of ELV using the Staden (1982) and SEQ programs for shotgun sequencing. Most of the sequence was obtained from independent clones which formed an overlapping set in both orientations and extended over about 95% of the genome. Figure 10 shows the positions and lengths of the sequences that contributed to the consensus; many of those subsequences were obtained from several different clones.

No clones were obtained that included nucleotides 1950-2056, therefore the primer PS2, 5'-CTGGAAAGGATCCTGGAAGG-3'(complementary to nucleotides 2035-2056), was used as a first-strand primer in the RNase H method (see Chapter 2)

Figure 8. Autoradiogram showing a restriction map of double-stranded DNA copy of genomic ELV. The dsDNA was hydrolysed by the four base restriction endonucleases *Dde* I, *Hae* III, *Hinf* I, *Msp* I, *Sau*3A I and *Taq* I and the DNA fragments were analysed by electrophoresis in 5% (W/V) polyarylamide gel. Standard molecular weights pBR325 / *Hinf* I markers are 1766, 652, 512, 506, 396, 394, 344, 290, 220, and 150 base pairs.


Marker
Dde I
Hae III
Hinf I
Msp I
Sau3A I
Taq I
1760
652
512
506
396
394
344
290
220
150

to synthesize a dsDNA *Nae*I fragment of approximately 110 bp, that spanned this region. This was sequenced in the same way as the other clones.

The 5'-terminus of the genome was determined from viral RNA using a specific primer, PS3, with the sequence 5'-CCGGAGACTTGAATGCCC-3'. This was complementary to nucleotides 261-276, and was used to synthesize a DNA complementary to the 5'-terminus of the genome, and poly dA was added to its 3'-end using polynucleotide terminal transferase. Then a DNA molecule encoding the 5'-terminal 276 nucleotides was produced by the polymerase chain reaction (PCR) using a primer with the sequence 5'-TTAAGC$(T)_{15}$-3' together with the PS3 primer.

Ten clones were obtained from the PCR product and the autoradiographs of their sequencing gels showed that, at its 5'-terminus, the genome had two guanines (Figure 9), one of which may correspond to the 7'-methylguanosine 'cap' structure that probably occurs in all tymoviruses (Klein *et al.*, 1976; Pleij *et al.*, 1976 and Ahlquist and Janda, 1984); it has been reported that reverse transcripts of capped RNAs can extend one base beyond the 5'-most nucleotide linked in the normal manner (Ahlquist and Janda, 1984; Gupta and Kingsbury, 1984), and the additional base added to the reverse transcript has, in one case, been identified as cytosine (Ahlquist and Janda, 1984). Thus it is likely that the sequence at the 5'-terminus of ELV is m^7GpppGAAAAUU.

Clones encoding the 3'-terminus of the genome were identified because, in order to obtain cDNA complementary to the genome, it had been polyadenylated. Five separate clones with a 3'-terminal sequence that this treatment would produce (-CAAAAAAAA) were found (Figure 9); the 5'-portions of those strands were homologous. The sequence of each clone was determined, in both orientations. Thus the 3'-terminal nucleotide of the ELV genome is -C or -C(A) like other tymoviruses (van Belkum *et al.*, 1987). It is unlikely that it has a longer poly(A) tail as the primer dT_8G failed to assist transcription unless the genome RNA had been polyadenylated.

The complete sequence of ELV genome is 6034 nucleotides in length (Figure 11). Each nucleotide was determined unambiguously in each orientation at least twice, making use of at least two independently-isolated clones, to avoid possible artifacts arising from transcription errors using AMV reverse transcriptase. Only at nucleotide position 5004 was sequence heterogeneity, A and G, found.

Figure 9. Autoradiograms of M13 dideoxy sequencing reactions using *Taq* DNA polymerase and either [α-^{32}P]dATP (C) or [α-^{35}S]dATP (A) and (B). DNA products were analysed by electrophoresis in 6% polyacrylamide / 7M urea gel. In all gels the tracks are in sets of four with dideoxy analogs added in the order G, A, T and C from left to right.

(A) DNA sequence of an ELV clone that includes the 5'-terminus; two guanines (arrowed) are probably those in the cap structure, ^{7}GpppGAAAAUU-.

(B) DNA sequence of an ELV clone representative of another 270 M13 clones.

(C) DNA sequence of an ELV clone at 3'-terminus with nucleotide sequences of -CCCC(A_8)

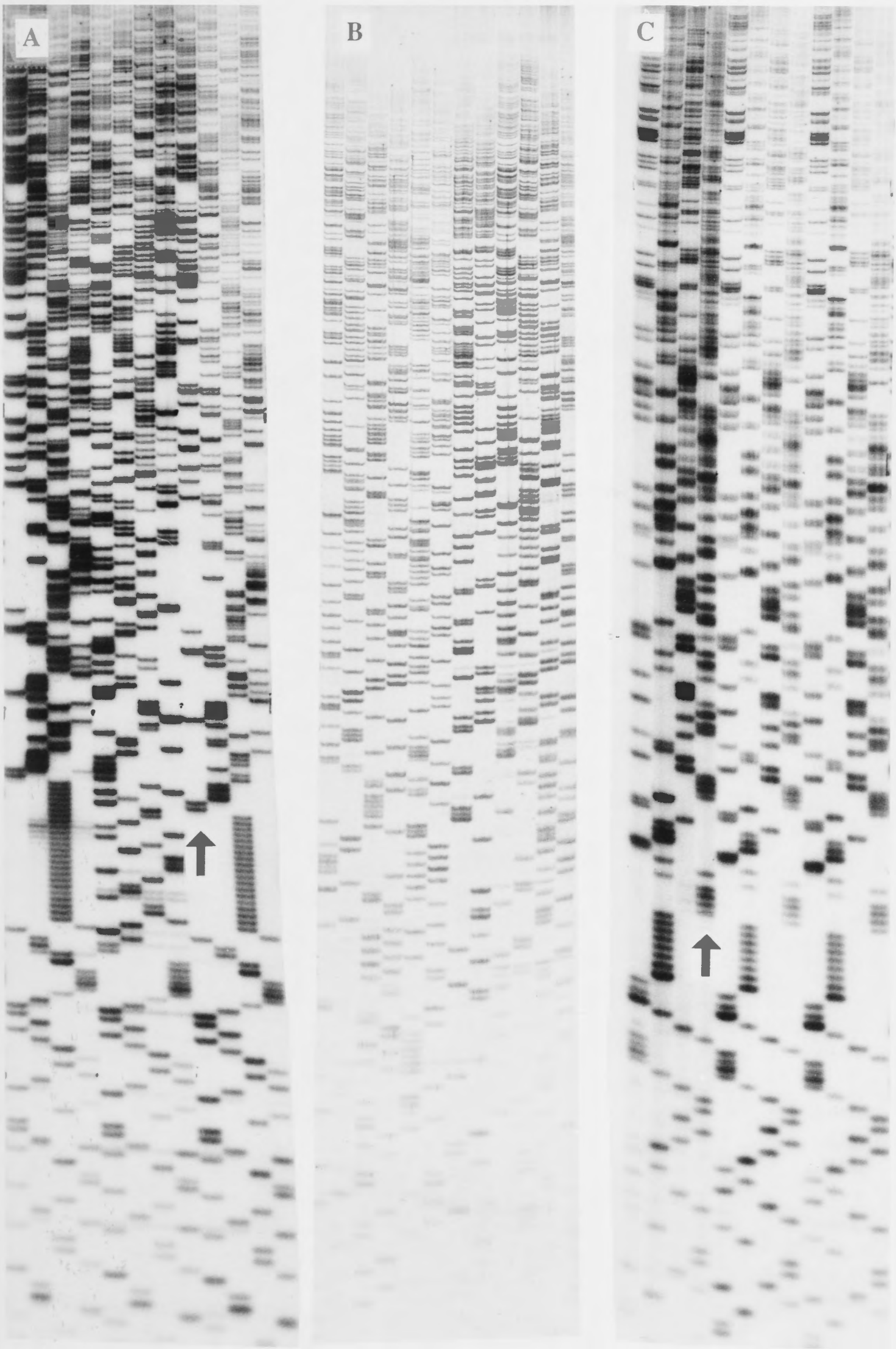
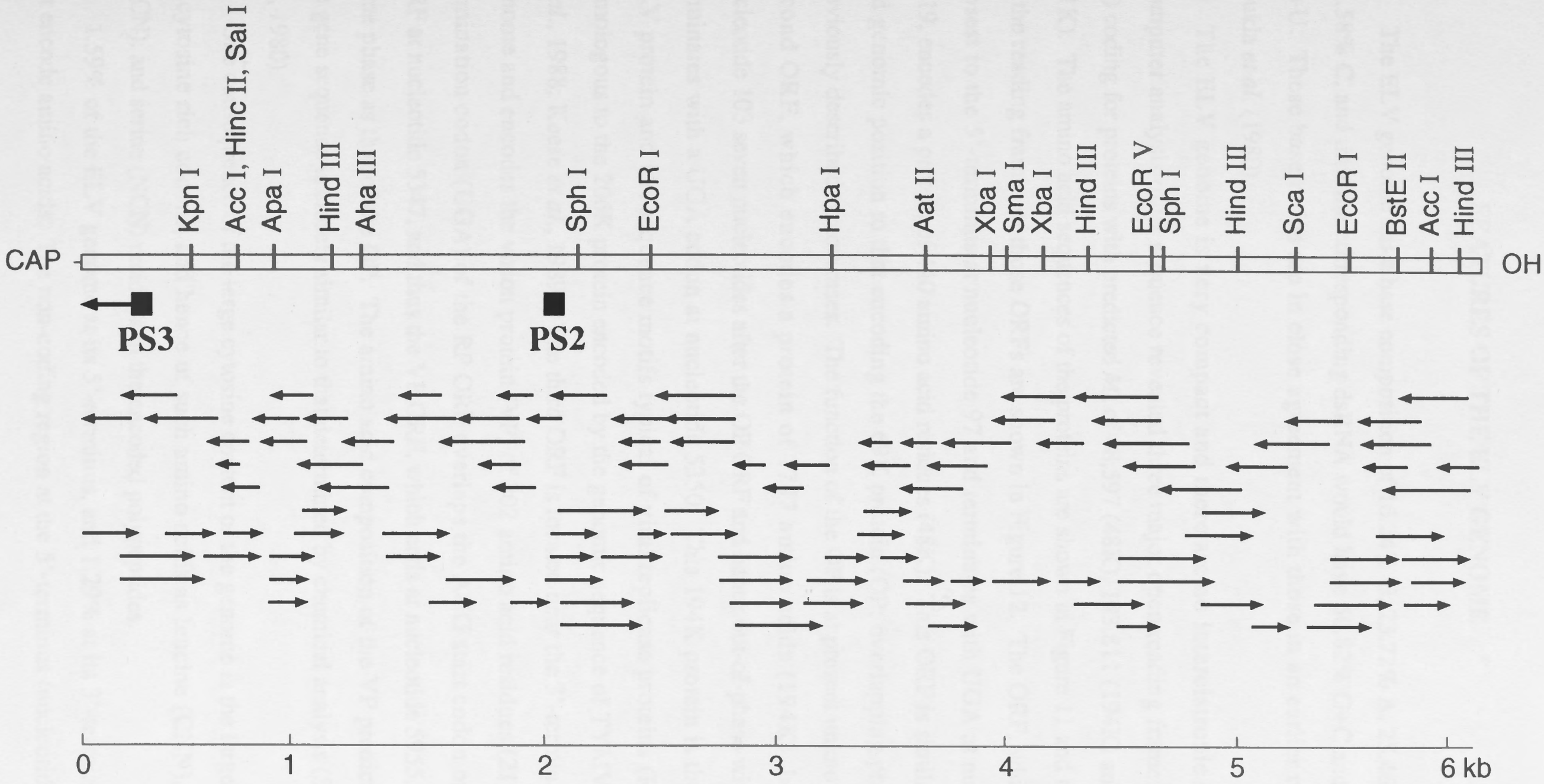

A
B
C

Figure 10. Map of the major overlapping cDNA inserts of clones used to determine the complete sequence of ELV genomic RNA. A composite map of some restriction sites corresponding to cDNA sequence is shown in the upper rectangle. The arrows indicate the regions, orientations and length of nucleotide sequences obtained from each clone. Some of the arrows represent more than one clone. The filled squares indicate the positions of the sequence used to prepare synthetic oligodeoxyribonucleotides in order to prime specific cDNA synthesis, and the 5'-terminal primer used to sequence the 5'-terminal region.


CAP
Kpn I
Acc I, Hinc II, Sal I
Apa I
Hind III
Aha III
Sph I
EcoR I
Hpa I
Aat II
Xba I
Sma I
Xba I
Hind III
EcoR V
Sph I
Hind III
Sca I
EcoR I
BstE II
Acc I
Hind III
OH
PS3
PS2
0
1
2
3
4
5
6 kb

FEATURES OF THE ELV GENOME

The ELV genome has a base composition of 16.24% G, 23.72% A, 25.46% U and 34.58% C, and thus the corresponding dsRNA would have 50.82% G+C and 49.18% A+U. These base ratios are in close agreement with those in an earlier report by Shukla *et al.* (1980).

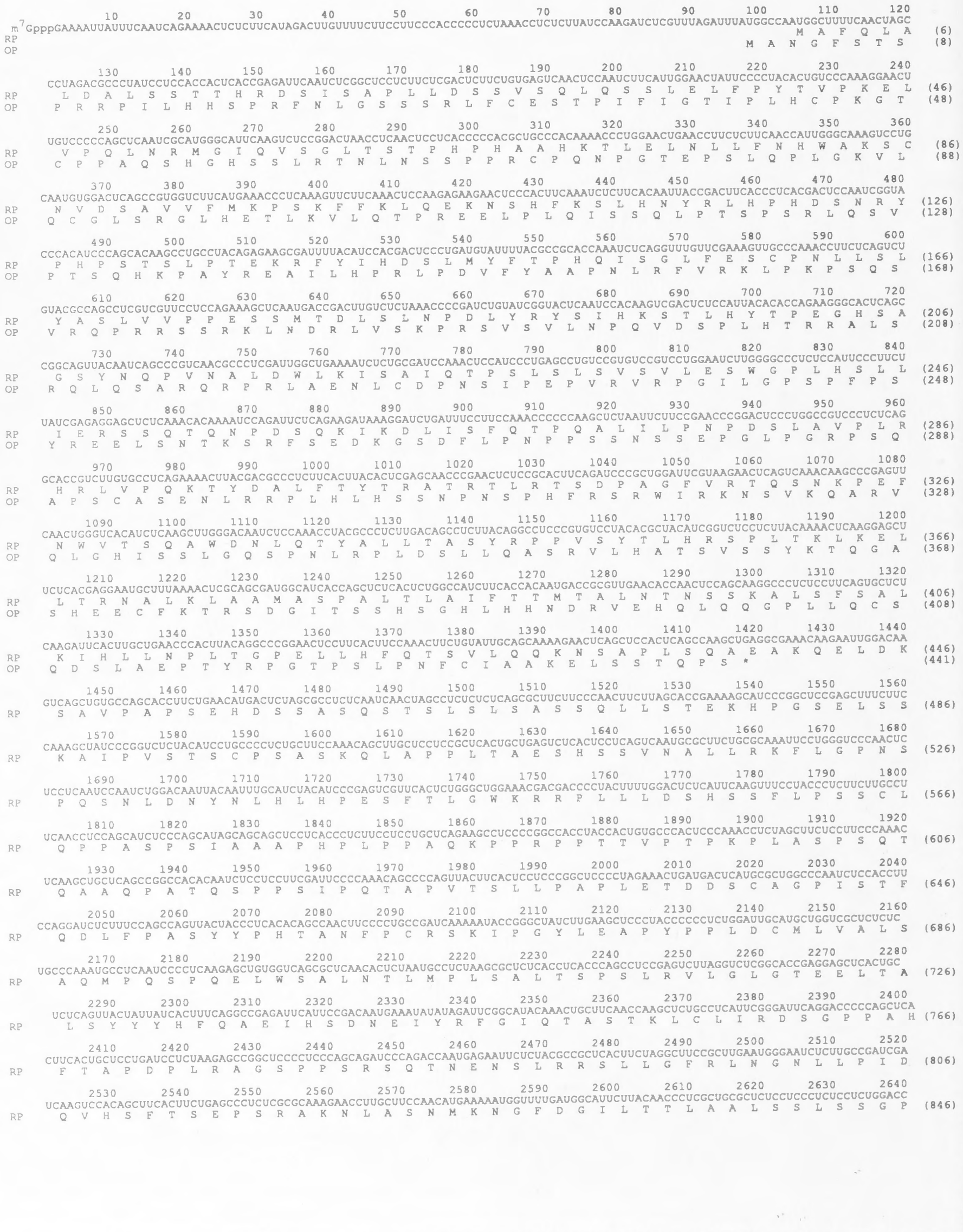
The ELV genome is very compact and there are no intercistronic regions. Computer analysis of its sequence revealed three major open reading frames (Figure 12) coding for proteins with predicted M_r of 48,597 (48K), 193,811 (194K) and 21,483 (21K). The amino acid sequences of the proteins are shown in Figure 11 and the phase of the reading frames of those ORFs are shown in Figure 12. The ORF, which starts closest to the 5'-terminus at nucleotide 97 and terminates with UGA at nucleotide 1419, encodes a protein of 440 amino acid residues (48K). This ORF is similar in size and genomic position to that encoding the 69K protein (OP; overlapping protein) of previously described tymoviruses. The function of the OP is at present unknown. The second ORF, which encodes a protein of 1747 amino acids (194K), begins at nucleotide 103 seven nucleotides after the OP ORF and hence out-of-phase with it, and terminates with a UGA codon at nucleotide 5350. This 194K protein is the largest ELV protein and has sequence motifs typical of viral replicase proteins (RP); it is homologous to the 206K protein encoded by the genomic sequence of TYMV (Morch *et al.*, 1988; Keese *et al.*, 1989). The third ORF is located near the 3'-terminus of the genome and encodes the virion protein (VP) of 202 amino acid residues (21K). The termination codon (UGA) of the RP ORF overlaps the AUG start codon of the VP ORF at nucleotide 5347, and thus the VP ORF, which ends at nucleotide 5955, is in the same phase as that of the OP. The amino acid composition of the VP predicted from the gene sequence is closely similar to that determined by chemical analysis (Shukla *et al.*, 1980).

One consequence of the large cytosine content of the genome is the large number of cytosine rich codons, and hence of such amino acids as leucine (CNN), proline (CCN), and serine (NCN) residues in the encoded polypeptides.

1.59% of the ELV genome at its 5'-terminus, and 1.29% at its 3'-terminus, does not encode amino acids. The non-coding region at the 5'-terminus (nucleotides 1-96)

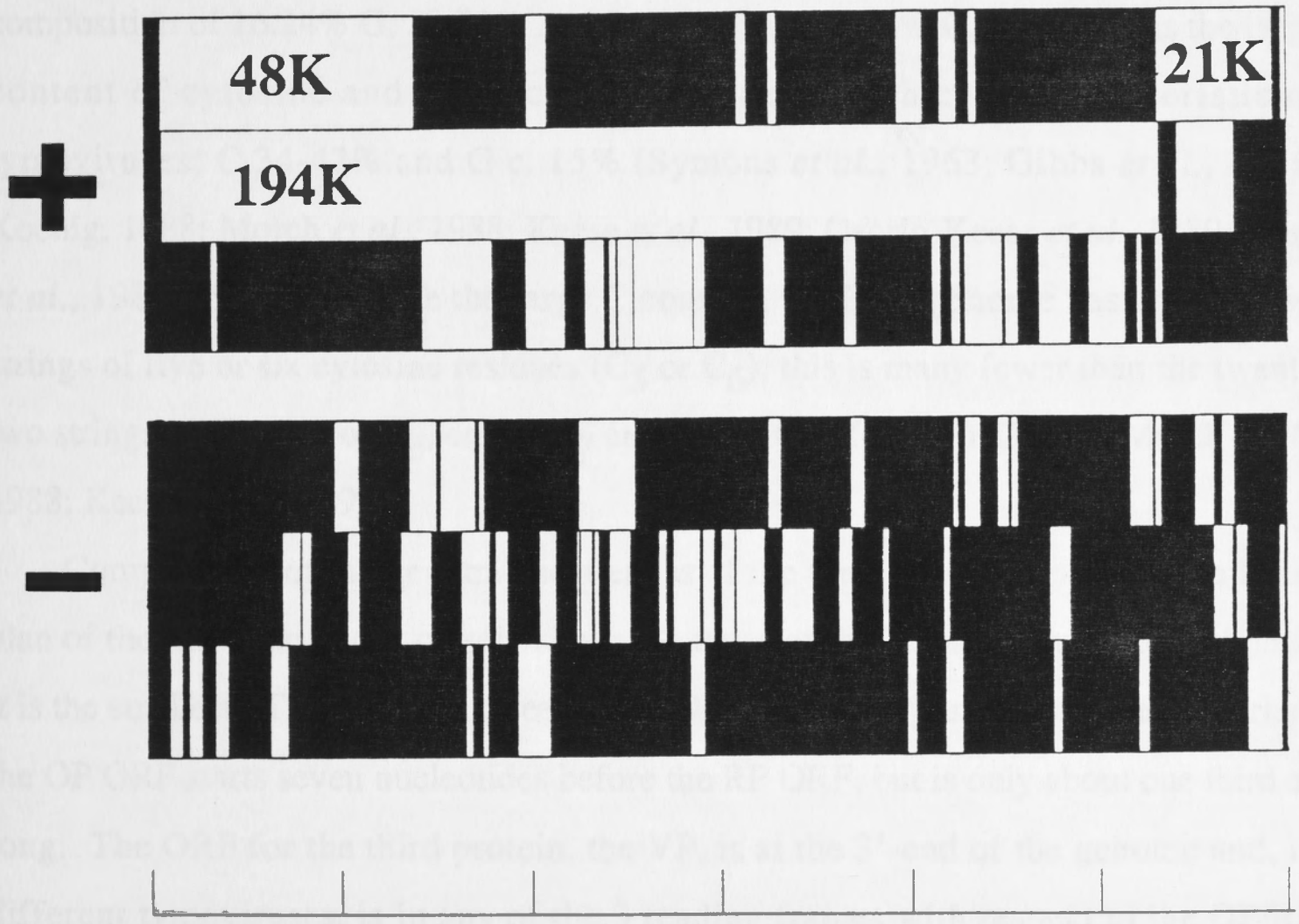
has a base composition of 8.3% G, 26.0% A, 37.5% U and 28.1% C, whereas that at the 3'-terminus (nucleotides 5956 to 6034) has a composition of 20.5% G, 21.8% A, 24.4% U and 33.3% C.

Figure 11. The complete nucleotide sequence of ELV RNA together with the amino acid sequences of polypeptides encoded by the three main open reading frames referred to in the text as replicase protein (RP), overlapping protein (OP) and virion protein (VP) are shown. Termination codons are indicated by asterisks. Two motifs, that of the NTP-binding fold GCGK[T/S] and the RNA polymerase sequence GDD, are boxed.



```
      2650      2660      2670      2680      2690      2700      2710      2720      2730      2740      2750      2760
CUCACCCCGCGAUCGCAUCUUCACUCUUGAUGGCAUCUGCGAUUUUGCUCUACCCAAAACUGUUGACCUUAUUCACCUGUCCGGGUUUGCCGGUUGUGGCAAAACUCAUCCAAUUCAACA
RP   S  P  R  D  R  I  F  T  L  D  G  I  C  D  F  A  L  P  K  T  V  D  L  I  H  L  S  G  F  A  G  C  G  K  T  H  P  I  Q  Q  (886)

      2770      2780      2790      2800      2810      2820      2830      2840      2850      2860      2870      2880
GCUUCUCAAAACCCCUCACUUCCACAACUUCAGGGUGGUCACUCCCACCACCAACCUUCGAUCAGAGUGGAAAUCAGACAUGGCUCUCCCAGCACACCACAACUGGAGAUUUUCAACUUG
RP   L  L  K  T  P  H  F  H  N  F  R  V  V  T  P  T  T  N  L  R  S  E  W  K  S  D  M  A  L  P  A  H  H  N  W  R  F  S  T  W  (926)

      2890      2900      2910      2920      2930      2940      2950      2960      2970      2980      2990      3000
GGAAUCAGCCCUCUUAAAACAUGCCGAAAUACUCGUCAUUGACGAAAUUUACAAACUUCCUCGAGGAUACCUAGACCUUUCUCUCAUAGCUGAUCCCACUGUCAAGCUCGUCAUCCUCCU
RP   E  S  A  L  L  K  H  A  E  I  L  V  I  D  E  I  Y  K  L  P  R  G  Y  L  D  L  S  L  I  A  D  P  T  V  K  L  V  I  L  L  (966)

      3010      3020      3030      3040      3050      3060      3070      3080      3090      3100      3110      3120
CGGAGACCCCCUUCAAGGCGAGUAUCACUCAACUUCUGCUCACUCCUCCAACUUGCGGCUCAGCUCAGAAAUCCCAAGGUUGCUCCCUUUCAUCGACUACUACUGUUACUGGAGCUACCG
RP   G  D  P  L  Q  G  E  Y  H  S  T  S  A  H  S  S  N  L  R  L  S  S  E  I  P  R  L  L  P  F  I  D  Y  Y  C  Y  W  S  Y  R  (1006)

      3130      3140      3150      3160      3170      3180      3190      3200      3210      3220      3230      3240
CGUUCCAAAAUGUGUCGCAAAGCUUUUCUCCCUGCCAUGUUUCAACCCCUCUGAAGGCUUCAUAAAGACUACUCUGGACUUUUUCCCCAGCGCUAACAAUUUGGUUAACAGCCACUCAGU
RP   V  P  K  C  V  A  K  L  F  S  L  P  C  F  N  P  S  E  G  F  I  K  T  T  L  D  F  F  P  S  A  N  N  L  V  N  S  H  S  V  (1046)

      3250      3260      3270      3280      3290      3300      3310      3320      3330      3340      3350      3360
AGUCCACAUCUCUGAAGCUUGCGGCUGGAAUGCUGUCACAAUAUCUUCGAGUCAAGGAUGCACCUUCUCCGACCCUGCUUUCAUCCACCUUGAUCGUAACACAGCUCUUCUGUCUCCCUC
RP   V  H  I  S  E  A  C  G  W  N  A  V  T  I  S  S  S  Q  G  C  T  F  S  D  P  A  F  I  H  L  D  R  N  T  A  L  L  S  P  S  (1086)

      3370      3380      3390      3400      3410      3420      3430      3440      3450      3460      3470      3480
AAACUGCCUCGUAGCUCUCACCAGAUCCCGCUCUGGAGUUUACUUCAAAGGCGAUUUCACAUUUCUGAGCUCUCUCUCCGGUUCCUCUCGAAUGUUCUCGUUAGCUUACUCUGGCCAGCC
RP   N  C  L  V  A  L  T  R  S  R  S  G  V  Y  F  K  G  D  F  T  F  L  S  S  L  S  G  S  S  R  M  F  S  L  A  Y  S  G  Q  P  (1126)

      3490      3500      3510      3520      3530      3540      3550      3560      3570      3580      3590      3600
GAUUCACCUUCCAGACUUCUUUCCAGAGAUUGUCUUCCAACUCAACAUGAUAACUGCCCCUCUGACAAAAAGAAGUUCCAGCUUUCGAUCCGGGUUCCAACCCAACAUCUCUUCUGCCCC
RP   I  H  L  P  D  F  F  P  E  I  V  F  Q  L  N  M  I  T  A  P  L  T  K  R  S  S  S  F  R  S  G  F  Q  P  N  I  S  S  A  P  (1166)

      3610      3620      3630      3640      3650      3660      3670      3680      3690      3700      3710      3720
AAAAAUCCCGGCCCCUCCAAAUCUUCCCUGCCCCCCUCAUAUCCCAACCAAUUACUCGAAAGACGUCAUUGUGAACAAUCAAGCUCUCUAUGGUGAGUCCCUCGAAAGACGUUUGUCUGU
RP   K  I  P  A  P  P  N  L  P  C  P  P  H  I  P  T  N  Y  S  K  D  V  I  V  N  N  Q  A  L  Y  G  E  S  L  E  R  R  L  S  V  (1206)

      3730      3740      3750      3760      3770      3780      3790      3800      3810      3820      3830      3840
UCUGCAUCUGCCCCCCACUCGCAUGACUCUUCAUUCUGACAUCAAUAUCACCGCACCUUCAAGCAGCAGUUUCCAGCCCUCAGACGAGCCAGUUCCUUCUGACCACACUGCUGUCUACCC
RP   L  H  L  P  P  T  R  M  T  L  H  S  D  I  N  I  T  A  P  S  S  S  S  F  Q  P  S  D  E  P  V  P  S  D  H  T  A  V  Y  P  (1246)

      3850      3860      3870      3880      3890      3900      3910      3920      3930      3940      3950      3960
UGGCUUUGACUUCUUCACUCUCGCCGCUCACUUUCUUCCAGCUCACGACCCAGAAGUCAAAGAAAUUGAACUCAAAGACCAAACCAGCCAACAAUUUCCCUGGCUGAAUCUAGACUUCCA
RP   G  F  D  F  F  T  L  A  A  H  F  L  P  A  H  D  P  E  V  K  E  I  E  L  K  D  Q  T  S  Q  Q  F  P  W  L  N  L  D  F  H  (1286)

      3970      3980      3990      4000      4010      4020      4030      4040      4050      4060      4070      4080
CAUAUCAUGCCAGACUUCUUCUUUGAUCUCUGCAAGGCACCAACCCGGGUCAGACUCCACUCUCCUUCCAGCUUCACUCCACAAAAGACUCCGAUUUCGCCCCACAGCAGCUCCAUACCA
RP   I  S  C  Q  T  S  S  L  I  S  A  R  H  Q  P  G  S  D  S  T  L  L  P  A  S  L  H  K  R  L  R  F  R  P  T  A  A  P  Y  Q  (1326)

      4090      4100      4110      4120      4130      4140      4150      4160      4170      4180      4190      4200
AAUCACUCCUUCUGACUCUUUCCUAGGCAAUUGUCUCUACCGCUCCUGGUGUCAAGUUUACCGUCGCGACCCAAAUGUGCGAUUGCCCUUCAACGAAGCUCUUUUUCUAGAAUGCAUUGC
RP   I  T  P  S  D  S  F  L  G  N  C  L  Y  R  S  W  C  Q  V  Y  R  R  D  P  N  V  R  L  P  F  N  E  A  L  F  L  E  C  I  A  (1366)

      4210      4220      4230      4240      4250      4260      4270      4280      4290      4300      4310      4320
UGUCAACGACUACGCCCAGCUGAGCUCAAAAACUCAAGCCACUAUUGUAGCCAAUGCAUCUCGAUCAGACCCUGACUGGCGACAUACUUUUGUAAAAAAUCUUCGCUAAGUCACAACACAA
RP   V  N  D  Y  A  Q  L  S  S  K  T  Q  A  T  I  V  A  N  A  S  R  S  D  P  D  W  R  H  T  F  V  K  I  F  A  K  S  Q  H  K  (1406)

      4330      4340      4350      4360      4370      4380      4390      4400      4410      4420      4430      4440
AGUCAAUGAUGGCUCCAUUUUUGGCCCAUGGAAAGCUUGCCAAACUCUUGCUCUCAUGCAUGAUUAUGUCAUUCUCACCCUGGGUCCGGUUAAGAAAUAUCAACGACUGUUCGACCAGCU
RP   V  N  D  G  S  I  F  G  P  W  K  A  C  Q  T  L  A  L  M  H  D  Y  V  I  L  T  L  G  P  V  K  K  Y  Q  R  L  F  D  Q  L  (1446)

      4450      4460      4470      4480      4490      4500      4510      4520      4530-     4540      4550      4560
GGAGCGACCUUCUCACAUCUACUACCAUGCAGGGAACACUCCUCAUGACCUUAGGCGCUGGUGCUCCAAACACUUAGAAACCUCUCACUGCACCACCAAUGAUUAUACAGCCUUCGACCA
RP   E  R  P  S  H  I  Y  Y  H  A  G  N  T  P  H  D  L  R  R  W  C  S  K  H  L  E  T  S  H  C  T  T  N  D  Y  T  A  F  D  Q  (1486)

      4570      4580      4590      4600      4610      4620      4630      4640      4650      4660      4670      4680
GAGCCAGCACGGUGAAGCCGUCGUGUUUGAAGUCCUGAAGAUGAGACGACUUUCCAUUCCGGAAAAUCUGAUAUCUCUGCACGUCCAUCUCAAAACCAACGUGGAAACCCAGUUCGGACC
RP   S  Q  H  G  E  A  V  V  F  E  V  L  K  M  R  R  L  S  I  P  E  N  L  I  S  L  H  V  H  L  K  T  N  V  E  T  Q  F  G  P  (1526)

      4690      4700      4710      4720      4730      4740      4750      4760      4770      4780      4790      4800
UCUCACUUGCAUGCGACUCACUGGCGAGCCCGGCACCUAUGACGACAACACUGACUACAACUUGGCUGUUCUCAAUCUCCAGUACGACCUUCGCAAAACACCGACCCUUGUCUCAGGCGA
RP   L  T  C  M  R  L  T  G  E  P  G  T  Y  D  D  N  T  D  Y  N  L  A  V  L  N  L  Q  Y  D  L  R  K  T  P  T  L  V  S  G  D  (1566)

      4810      4820      4830      4840      4850      4860      4870      4880      4890      4900      4910      4920
UGACUCCUACUUAUCCGGAACCCUUUCCCCAAGAUCCAACUGGCCUUUCGUCAAAGAGCUCCUCCACCUCCGCUUAAAACCGAGCAGUCUGAUCGACGGCCUCUUCUGCGGUUACUACUU
RP   D  S  Y  L  S  G  T  L  S  P  R  S  N  W  P  F  V  K  E  L  L  H  L  R  L  K  P  S  S  L  I  D  G  L  F  C  G  Y  Y  L  (1606)

      4930      4940      4950      4960      4970      4980      4990      5000      5010      5020      5030      5040
AGGACCCCAAGGCUGCAUCCGCAAUCCACUUGCCCUUUUCGCUAAACUCAUGAUAGCCGAAGACGAUGGCUCCGCCUUUGACAAGCUUCCUUCCUACCUCACCGAGUUCUCGAUUGGACA
RP   G  P  Q  G  C  I  R  N  P  L  A  L  F  A  K  L  M  I  A  E  D  D  G  S  A  F  D  K  L  P  S  Y  L  T  E  F  S  I  G  H  (1646)

      5050      5060      5070      5080      5090      5100      5110      5120      5130      5140      5150      5160
UGGUCUUGGAGAUUCUCUCUGGCAGCUUCUCCCUUCUGAUCUUGUGCUCUAUCAGAGUGCGUGUUUCGAUUACUUCUGCCGCAAAGCCACACGCUCUCAGAAAAUCCUACUGCAACCAGG
RP   G  L  G  D  S  L  W  Q  L  L  P  S  D  L  V  L  Y  Q  S  A  C  F  D  Y  F  C  R  K  A  T  R  S  Q  K  I  L  L  Q  P  G  (1706)

      5170      5180      5190      5200      5210      5220      5230      5240      5250      5260      5270      5280
CCUUGUAGAUCAGGAAACUUUAGACAAGAUUGCUUUAUCGGCCAAGUUCAUAUCUCGUCCUUUCUACUCAAUGCUUUCAAGCCAUGCUCGCAGCCUGAUUAGUACUAAAAUUCAAACUCGA
RP   L  V  D  Q  E  T  L  D  K  I  A  L  S  A  K  F  I  S  R  P  F  Y  S  M  L  S  S  H  A  R  S  L  I  S  T  K  F  K  L  D  (1746)

      5290      5300      5310      5320      5330      5340      5350      5360      5370      5380      5390      5400
CUCUUCCCUCACCACUCUCCAAGAUCCCAUGGUUGAGUUUGAAUUGCUUCCCUUUUCCAAUGUACAAUGACUUCCUCUCGAUCCACUGAACUCAUCGAGACUCAAGAAGACGUUGUUCGU
RP   S  S  L  T  T  L  Q  D  P  M  V  E  F  E  L  L  P  F  S  N  V  Q  *                                                (1769)
VP                                                                  M  T  S  S  R  S  T  E  L  I  E  T  Q  E  D  V  V  R  (18)

      5410      5420      5430      5440      5450      5460      5470      5480      5490      5500      5510      5520
AAACUCUCCCAGAAAACUCCAGUUUCCAACUUCUCAGCUCUUCCACUUCCAAAUGGACCUCAAGCCCCAACUCAGCUCCAGCCGUUCCAGUUCGAAUUCCCUCUCCCGGCUGGCCAAGAAV
VP   K  L  S  Q  K  T  P  V  S  N  F  S  A  L  P  L  P  N  G  P  Q  A  P  T  Q  L  Q  P  F  Q  F  E  F  P  L  P  A  G  Q  E  (58)

      5530      5540      5550      5560      5570      5580      5590      5600      5610      5620      5630      5640
GGUUCUGUCACCCUGCCUCUCGCCACCUUCCCCAAAAUGGCAACCUUCCUUUCCCGACACCGUCGAGCCCAACUCACUCAACUUCACGCUGUCGUCUCUCCCAGUGCUGUAUCCAUUGGG
VP   G  S  V  T  L  P  L  A  T  F  P  K  M  A  T  F  L  S  R  H  R  R  A  Q  L  T  Q  L  H  A  V  V  S  P  S  A  V  S  I  G  (98)

      5650      5660      5670      5680      5690      5700      5710      5720      5730      5740      5750      5760
CACCCCCUCACCGUCCAACUCAUCUGGGUCCCCGCCAGCUCAACAACCACUUCCUCCCAAAUACUGGGCACUUACGGCGGUCAACAAAUCUCGGUCGGUGGUCAGGUAACCAACUCCUCU
VP   H  P  L  T  V  Q  L  I  W  V  P  A  S  S  T  T  T  S  S  Q  I  L  G  T  Y  G  G  Q  Q  I  S  V  G  G  Q  V  T  N  S  S  (138)

      5770      5780      5790      5800      5810      5820      5830      5840      5850      5860      5870      5880
CCAGCUAAAGUCUCUGCUAACCUCCUCAUGAUGAACCCACACAUCAAAGACUCUACAAGCUACACCGACACCCCAAACUCCUCGUCUACAGCACCCCCGCUGUGCCAGACGACAAACUC
VP   P  A  K  V  S  A  N  L  L  M  M  N  P  H  I  K  D  S  T  S  Y  T  D  T  P  K  L  L  V  Y  S  T  P  A  V  P  D  D  K  L  (178)

      5890      5900      5910      5920      5930      5940      5950      5960      5970      5980      5990      6000
ACAACCAGCUCAGCCUCCAUCAUUGUCUUUGGUGAAGUUCUGCUCAGCUCCCCUCAACUCAAUCCUUCAGCUUAAGCAAUAGCGACCUUAAGCUUCUUUCUCCAGCGUGUCACCACGUUU
VP   T  T  S  S  A  S  I  I  V  F  G  E  V  L  L  S  S  P  Q  L  N  P  S  A  *                                          (203)

      6010      6020      6030
CGCUCCUGAAUGUCCCCAGGUAAAGGGACACCCC(A)OH
```

Figure 12. Gene organization of both plus- and minus-strand of ELV RNA showing the open reading frames of different triplet codon phases. The ORFs (white boxes) start with the AUG initiation codon and terminate with either UAA, UAG, or UGA codons. The calculated molecular weight in kilodaltons (K) of three major polypeptides: OP(48K), RP(194K) and VP(21K) encoded by each ORF in plus-strand RNA are shown.



Unit scale 1000

CHAPTER 4 - COMPARISONS OF ELV AND OTHER TYMOVIRUSES; GENOMIC SEQUENCES AND ENCODED PROTEINS

THE BASIC FEATURES

The complete genomic sequences of five tymoviruses are known, namely ELV, EMV-Trin (Osorio-Keese *et al.*, 1989), KYMV-JB (Ding *et al.*, 1990d), OYMV-Tin (Ding *et al.*, 1989), TYMV-Type (Morch *et al.*, 1988) and TYMV-CL (Keese *et al.*, 1989). That of ELV is 6034 nucleotides in length (Figure 11). It has a base composition of 16.24% G, 23.72% A, 25.46% U and 34.58% C, and thus has the large content of cytosine and small content of guanine which is characteristic of tymoviruses; C 34-42% and G *c.* 15% (Symons *et al.*, 1963; Gibbs *et al.*, 1966; Koenig, 1988; Morch *et al.*, 1988; Keese *et al.*, 1989; Osorio-Keese *et al.*, 1989; Ding *et al.*, 1989, 1990). Despite the large C content, the ELV genome has only twelve strings of five or six cytosine residues (C_5 or C_6); this is many fewer than the twenty two strings of C_5, five of C_6, one of C_7 and one of C_{11} found in TYMV (Morch *et al.*, 1988; Keese *et al.*, 1989).

Comparisons of the genomic sequences of the tymoviruses show that the basic plan of the ELV genome is closely similar to those of the others (Figure 13), although it is the smallest. They all have three ORFs (OP, RP and VP) and two of these overlap; the OP ORF starts seven nucleotides before the RP ORF, but is only about one third as long. The ORF for the third protein, the VP, is at the 3'-end of the genome and, in different tymoviruses, is in any of the 3 reading frames with respect to the OP/RP ORFs; ELV and TYMV-CL have VP ORFs in the same phase as their OP ORFs, KYMV-JB has its VP ORF in the same phase as its RP ORF, whereas, EMV-Trin and OYMV-Tin have their VP ORF in the third reading frame.

All components of the genomes, the ORFs and untranslated regions, vary in length. For example, among the ORFs, the OPs are the most variable, whereas, of the intergenic regions, that between the RP and VP genes varies from +14 to -20 nucleotides in length; it is fourteen and four nucleotides long in TYMV and EMV

respectively, whereas in ELV and KYMV-JB there is no intergenic region and in OYMV-Tin the RP gene overlaps that of the VP by twenty nucleotides (Table 4).

GENOMIC REGIONS

Comparisons of the different regions of the genomes of the tymoviruses were made using various techniques. These analyses are reported here, starting at the 5'-termini of the genomes; analyses that involved comparisons with non-tymoviral sequences (e.g. promoter regions) are discussed in Chapter 5.

The 5'non-coding region

The 5'non-coding regions of tymoviral genomes have a smaller guanine content (8.3-12.9%) than the remainder of the genome (15.1-16.8%), and a slightly smaller cytosine content (24.4-30.2% compared with 34.6-39.4%).

The sequences of the 5'-terminal regions of ELV, TYMV-CL, KYMV-JB, EMV-Trin and OYMV-Tin were examined using the SQUIGGLES program of the GCG package and also by eye. They formed structures with 2, 1, 2, 3, and 4 stem-loops (Figure 14a, b), and these had free energies of ΔG -14.5, -13.5, -12.9, -14.6 and -23.0 kcal/mol, respectively (Zuker and Stiegler, 1981). It has been reported previously that the two stem-loops of OYMV have similar sequences and may have arisen by duplication (Ding *et al.*, 1989). The two stem-loops of ELV may also have a similar history as the base-paired parts of the first loop (nucleotides 17-45) and the second loop (nucleotides 52-106) have similar sequences. However there is no obvious shared feature, neither of a sequence motif in these regions, which might act as a signal for ribosomal binding to the 5' side of the initiation codon AUG of the first ORF, nor of a pseudoknot structure which is required for efficient ribosomal frameshifting signal at the junction of two overlapping openreading frames.

Proteins Encoded by Tymoviruses

Tymovirus genomes encode three proteins. The OP ranges in size from M_r 48.6-82.1K, the RP from 193.8-206.5K and the VP from 19.5-21.5K respectively; a ratio of size differences of, approximately, 3:9:1 with the OPs varying the most (Figure 13).

Figure 13. Comparison of the genomic organizations of ELV and the other four tymoviruses. The three ORFs are drawn as rectangles with the similarity of the regions of amino acid sequence of the gene products indicated by shading. Also indicated are the nucleotide positions of the start, read-through and stop codons; and the size (molecular weight) of each polyprotein, the cap structure, VP = virion protein, N = NTP-binding site and P = RNA polymerase site. The conserved regions correspond to the regions aligned by the progessive alignment method of Feng and Doolittle (1987) and shown in Figure 19.

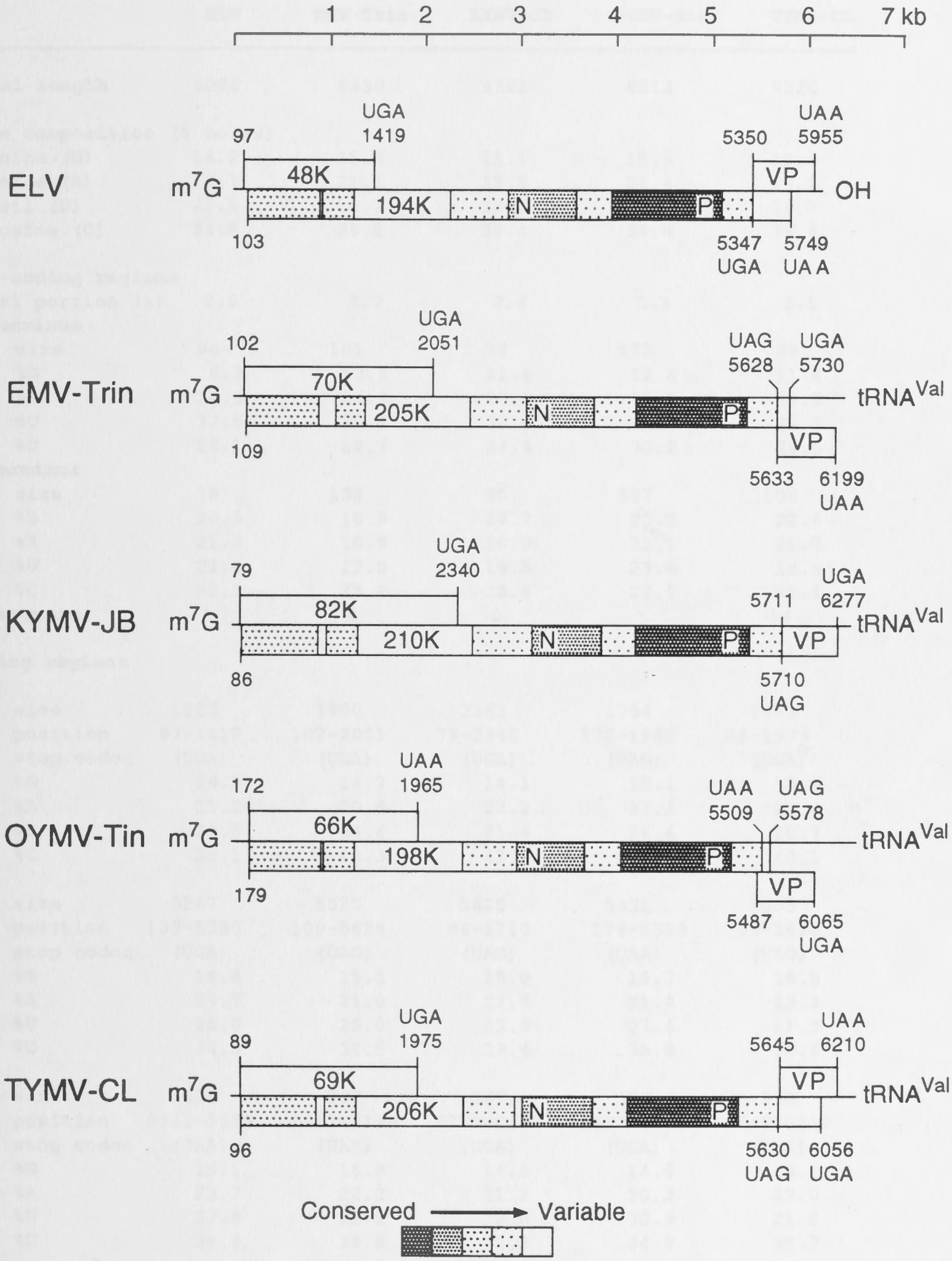

1
2
3
4
5
6
7 kb
ELV
m7G
97
UGA
1419
48K
194K
N
P
VP
UAA
5350
5955
OH
103
5347
UGA
5749
UAA
EMV-Trin
m7G
102
UGA
2051
70K
205K
N
P
UAG
5628
UGA
5730
tRNAVal
VP
109
5633
6199
UAA
KYMV-JB
m7G
79
UGA
2340
82K
210K
N
P
VP
5711
UGA
6277
tRNAVal
86
5710
UAG
OYMV-Tin
m7G
172
UAA
1965
66K
198K
N
P
UAA
5509
UAG
5578
tRNAVal
VP
179
5487
6065
UGA
TYMV-CL
m7G
89
UGA
1975
69K
206K
N
P
5645
UAA
6210
VP
tRNAVal
96
5630
UAG
6056
UGA
Conserved
Variable

Table 4. Nucleotide composition of tymovirus-genomic RNA.

	ELV	EMV-Trin	KYMV-JB	OYMV-Tin	TYMV-CL
Total length	6034	6330	6362	6212	6320
Base composition (% moles)					
Guanine (G)	16.2	15.6	15.1	15.6	16.8
Adenine (A)	23.7	21.1	22.5	21.4	23.1
Uracil (U)	25.5	24.7	23.0	28.1	21.7
Cytosine (C)	34.6	38.5	39.4	34.8	38.4
Non-coding regions					
Total portion (%)	2.9	3.7	2.6	5.1	3.1
5'terminus					
size	96	101	78	172	88
%G	8.3	12.9	12.8	12.8	11.4
%A	26.0	26.7	30.8	15.1	29.5
%U	37.5	30.7	32.0	41.9	29.5
%C	28.1	29.7	24.4	30.2	29.5
3'terminus					
size	79	132	85	147	106
%G	20.5	18.9	24.7	25.2	22.6
%A	21.8	15.9	20.0	21.1	21.7
%U	24.4	22.0	16.5	23.8	24.5
%C	33.3	43.2	38.8	29.9	31.1
Intergenic	*	4	0	*	14
Coding regions					
OP					
size	1323	1950	2262	1794	1887
position	97-1419	102-2051	79-2340	172-1965	89-1975
stop codon	(UGA)	(UGA)	(UGA)	(UAG)	(UGA)
%G	16.2	14.7	14.1	15.1	17.6
%A	25.2	20.6	22.2	21.2	21.6
%U	23.5	24.4	21.4	26.6	20.7
%C	35.1	40.3	42.3	37.1	40.1
RP					
size	5247	5520	5625	5331	5535
position	103-5350	109-5628	86-5710	179-5509	96-5630
stop codon	(UGA)	(UAG)	(UAG)	(UAA)	(UAG)
%G	16.4	15.5	15.0	15.7	16.8
%A	23.7	21.0	22.6	21.9	23.3
%U	25.6	25.0	23.0	27.4	21.5
%C	34.3	38.5	39.4	35.0	38.6
VP					
size	606	567	567	579	570
position	5347-5955	5633-6199	5711-6277	5487-6065	5645-6210
stop codon	(UAA)	(UAA)	(UGA)	(UGA)	(UAA)
%G	15.1	16.8	14.5	14.0	16.1
%A	23.7	22.2	21.2	20.2	23.0
%U	22.8	22.2	22.0	30.9	21.6
%C	38.4	38.8	42.3	34.9	39.3

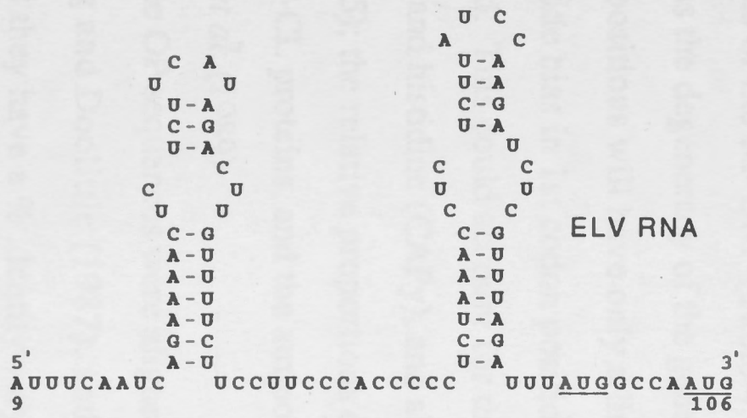
* RP overlaps with VP.

Figure 14. Sequence alignment of the 5'-termini of five tymoviral genomes and their potential secondary structures. These sequences are those of ELV, EMV-Trin, KYMV-JB, OYMV-Tin and TYMV-CL. The structures in (A) were computed and drawn by the SQUIGGLES program and those in (B) by close inspection and by analogy.

Sequence alignment of non-coding 5'termini of five tymoviral RNAs.

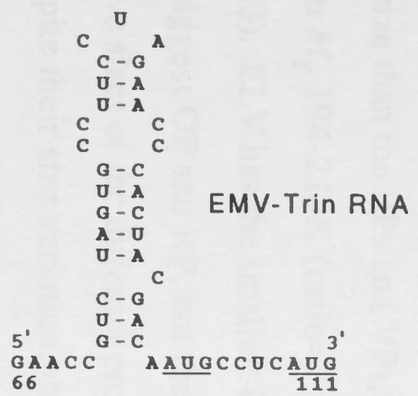
```
                                                
```

B



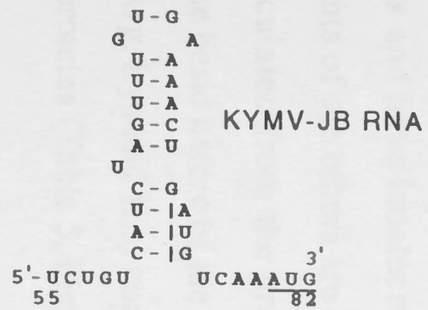
ELV RNA

ΔG = -5.3 kcal/mol



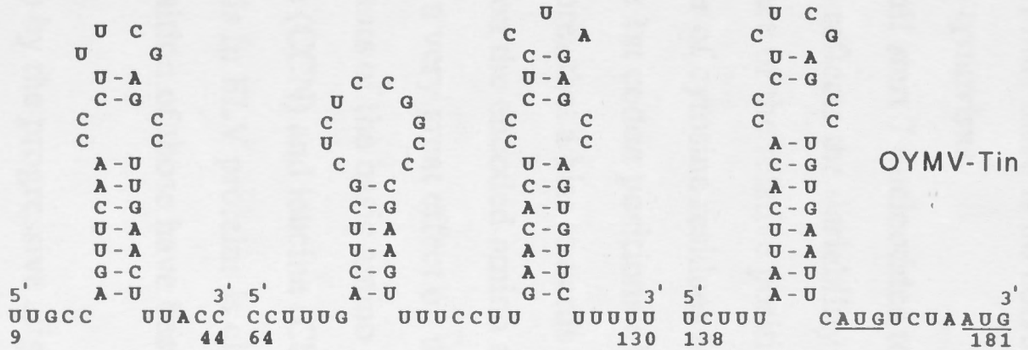
EMV-Trin RNA

ΔG = -4.8 kcal/mol



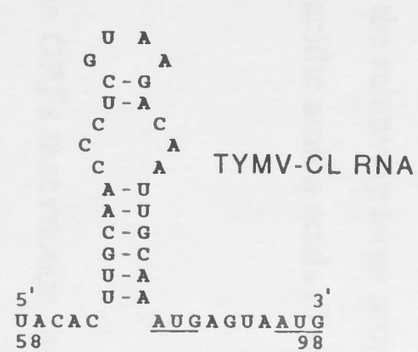
KYMV-JB RNA

ΔG = -3.2 kcal/mol



OYMV-Tin RNA

ΔG = -4.4 kcal/mol ΔG = -7.0 kcal/mol ΔG = -5.3 kcal/mol



TYMV-CL RNA

ΔG = -3.7 kcal/mol

The sizes, compositions, net charges and predicted isoelectric points of the 3 sets of proteins are summarized in Table 5. It can be seen that the OPs are the most basic of the proteins and have isoelectric points in the range pH 10.9-11.9 whereas the isoelectric points of the others are in the more usual pH range of 5.6-9.6. Isoelectric points are calculated from the proportion and type of charged amino acids in the protein, thus the basic nature of the OPs reflects the relatively large proportion of basic amino acids (Arg, Lys and His) compared with acidic amino acids (Tyr, Cys, Glu and Asp) in those proteins (Table 5, Figure 15).

a) The Overlapping Proteins (OP)

As noted above, the tymoviral OPs, and the ORFs that encode them, vary much more in size than the RPs and VPs; the OPs range from M_r 48.5-82K (ratio=1:1.7), the RPs from M_r 194-210K (ratio=1:1.1), and VPs from M_r 19.6-21.5K (ratio=1:1.1) (Figure 13). ELV has the smallest OP and RP, but the biggest VP, whereas KYMV-JB has the biggest OP and RP, but the smallest VP, thus there is no correlation between the relative sizes of the different proteins of each tymovirus.

Despite their size variation, the OP ORFs all start 7 nucleotides to the 5'-side of the start of the RP ORFs, thus the size variation reflects the variability in the position of their termination codons. A major consequence of the relative positions of the start codons of the OP and RP is that the large number of cytosine residues found in the 3rd positions of the RP ORF (Table 4, 6) are in the 1st codon positions of all OP ORFs. Whereas the degeneracy of the genetic code ensures that a bias towards cytosine in 3rd codon positions will have only a limited effect on the encoded amino acids, a similar nucleotide bias in 1st codon positions will have a very great effect on the amino acids encoded. This could account for the large amounts of the basic amino acids, arginine (CGN) and histidine (CAPy), and also of proline (CCN) and leucine (CUN) in the OPs (Table 5); the relative proportions of amino acids in ELV proteins is close to those of TYMV-CL proteins, and the amino acid composition of those have been discussed by Keese *et al.* (1989)

The OP sequences were aligned (Figure 16) by the progressive alignment method of Feng and Doolittle (1987), and pairwise comparisons of the aligned sequences showed they have a % identity varying from 25.7% to 33.5%. Most of the conserved residues are found in the N-terminal region (from 1 to 380 residues), and the C-

Table 5. Amino acid composition of tymovirus-encoded proteins (OP, RP, and VP).

	ELV			EMV-Trin			KYMV-JB			OYMV-Tin			TYMV-CL		
	OP	RP	VP	OP	RP	VP	OP	RP	VP	OP	RP	VP	OP	RP	VP
Amino acid residues	440	1747	202	649	1839	188	750	2262	188	597	1776	192	627	1844	189
Molecular weight	48596	193806	21482	70250	204730	19769	82130	210027	19601	66073	198126	20466	69097	206508	20152
Isoelectric point	11.2	7.9	9.1	10.9	8.4	8.3	11.9	9.6	9.1	11.7	7.9	7.3	11.9	7.9	5.6
Net charge	28	6	1	35	19	2	52	30	-2	46	8	0	34	4	-3
Amino acid composition															
A = Ala	18	121	14	39	112	**25**	37	105	20	18	106	9	28	115	14
C = Cys	10	29	0	21	34	3	8	22	3	19	35	2	2	22	4
D = Asp	10	80	5	25	65	5	17	75	5	13	65	7	29	90	7
E = Glu	18	59	6	8	55	4	20	54	3	14	65	5	17	61	6
F = Phe	12	82	7	24	85	6	21	100	3	25	98	5	14	82	6
G = Gly	21	61	10	16	57	7	25	62	8	23	68	8	30	61	8
H = His	20	68	4	34	93	3	39	90	4	40	83	2	35	93	3
I = Ile	12	68	8	13	84	13	16	80	15	14	91	15	20	85	16
K = Lys	15	72	7	17	60	7	24	77	5	19	72	5	11	61	7
L = Leu	**50**	**235**	**22**	**75**	**261**	19	**82**	**248**	**21**	**67**	**220**	**21**	64	**229**	16
M = Met	1	21	4	2	19	3	1	28	3	6	28	3	3	25	4
N = Asn	20	71	6	15	74	10	20	80	8	18	68	7	13	62	4
P = Pro	**55**	**157**	20	**108**	**204**	17	**161**	**214**	16	**82**	**171**	**21**	**121**	**194**	**20**
Q = Gln	27	76	15	26	78	9	40	92	8	30	60	11	23	80	8
R = Arg	41	73	5	51	79	4	65	82	1	54	66	7	**69**	94	3
S = Ser	**68**	**219**	**28**	**108**	**209**	**20**	**101**	**199**	**23**	**114**	**206**	**34**	**74**	**198**	**17**
T = Thr	19	116	**21**	40	120	**21**	43	123	**25**	18	109	11	44	140	**27**
V = Val	17	66	16	23	77	7	21	74	10	19	86	15	21	74	14
W = Trp	1	19	1	0	20	1	3	24	1	3	25	1	3	25	2
Y = Tyr	5	55	3	4	53	4	6	45	6	1	54	3	6	53	3
Mole percent of the charge group															
H+K+R (basic)	17.273	12.192	7.921	15.716	12.616	7.447	17.067	13.287	9.042	18.928	12.444	7.292	18.341	13.449	6.878
D+C+E+Y (acid)	9.773	12.758	6.931	8.937	11.256	8.511	6.801	10.459	5.319	7.874	12.332	8.855	8.612	12.256	10.581

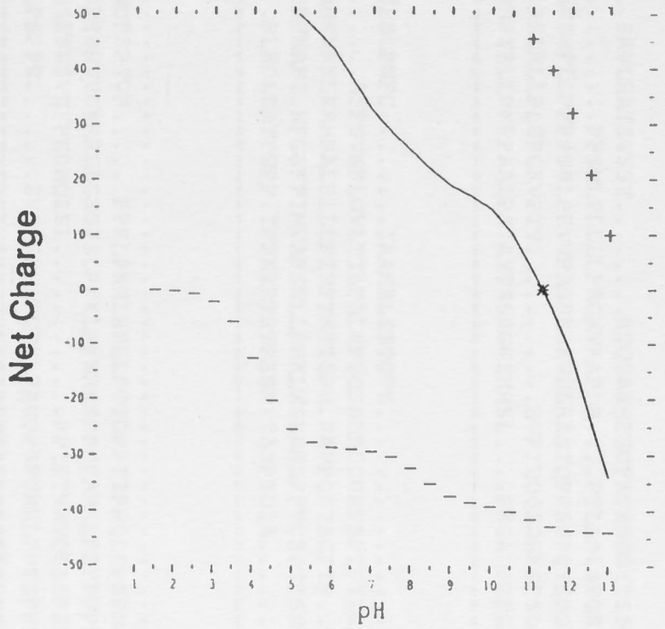
Table 6. Codon usage of five tymoviral genomes in the term of open reading frames.

	ELV			EMV-Tin			KYMV-JB			OYMV-Trin			TYMV-CL		
Position	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
Whole genome															
G	308	419	254	406	277	306	286	424	248	261	421	291	268	360	433
A	419	468	543	442	547	348	353	511	569	330	475	526	553	426	483
T	533	468	535	469	584	513	434	432	597	626	520	602	551	373	446
C	752	656	679	794	702	943	1048	754	706	854	655	651	735	948	744
OP															
G	84	85	45	112	108	66	121	117	80	88	119	63	133	129	69
A	90	115	127	119	130	152	137	167	197	104	135	140	121	131	155
T	96	91	122	143	136	195	128	141	215	151	130	195	90	124	175
C	170	148	146	275	275	236	367	328	261	254	213	199	284	244	229
RP															
G	388	227	248	366	213	275	370	219	254	390	227	219	401	233	296
A	412	481	350	395	478	286	437	513	318	421	467	275	431	500	343
T	389	472	480	404	526	446	382	530	384	409	523	531	371	495	325
C	559	568	670	674	622	832	685	612	918	556	559	751	641	616	880
VP															
G	51	22	19	48	18	29	46	16	19	44	20	16	49	18	25
A	52	46	44	58	42	24	59	39	21	44	40	32	59	38	32
T	33	57	48	32	48	45	35	52	37	43	59	76	33	55	34
C	66	77	91	50	80	90	48	81	111	61	73	68	48	78	98

Figure 15. Graphs comparing the potential isoelectric points (I.E.P.) of the OP, RP and VP encoded by the ELV genome. The isoelectric point, is calculated from amino acid composition; the numbers of amino acids relevant to the calculation are given below each graph.

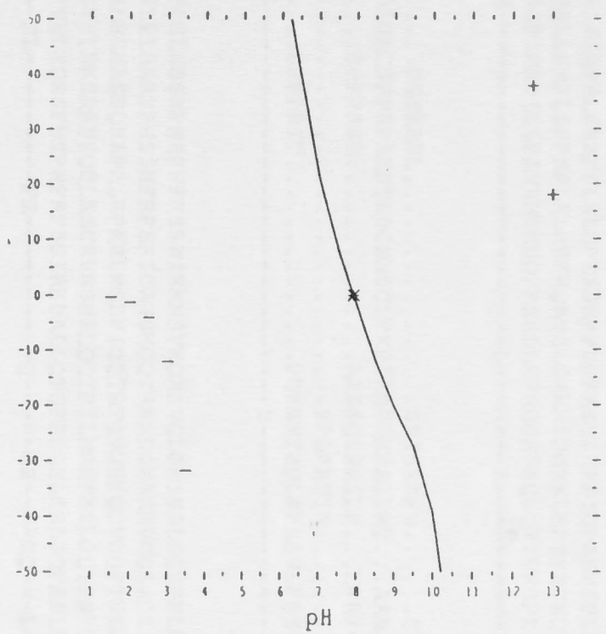
Erysimum latent tymovirus

OP I.E.P. 11.28



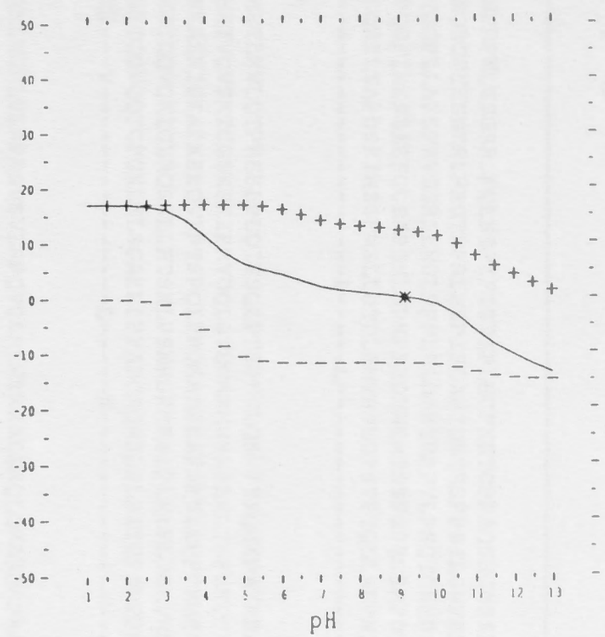
R73 K72 H68 Y55 C29 E59 D80

RP I.E.P. 7.90



R41 K15 H20 Y5 C10 E18 D10

VP I.E.P. 9.16



R5 K7 H4 Y3 E6 D5

Figure 16. Alignment of encoded overlapping proteins of five tymoviruses: ELV, EMV-Trin, KYMV-JB, OYMV-Tin and TYMV-CL by the progressive alignment method of Feng and Doolittle (1987) and PRETTY program.

```
                1                                                                80
ELV       48K MANGFSTSPRRPILHHSPRFNLGSSSRLFCESTPIFIGTIPLHCPKGTCPPAQSHGHSSLRTNLNSSPPRCPQNPGTEPS
EMV-Trin  70K MPHGLSVCSRSSQLNYSQRCFYKSNSELRRGTSPRLSIPISLAPSQRSRSPPSILGHPELRPRSHSPPPPNPQNSRDFSP
KYMV-JB   82K MSNGFPTSSRCSSVDFTQGPLLAPSPRVGSRLLNGLPPNLPLDGPTRPPTLPSQIGNPDQLLRKFPPPPPSAQDAGNPPT
OYMV-Tin  66K MSNGLRTSFERISLFHPQGFISESSAEFCCSAPSNLSPKFPMDHWERTSSFSPRCRHPDLRVRLQSSPPRGPQSDRNLPS
TYMV-CL   69K MSNGLPISIGRPCTHDSQRSLSAPDSRIHSGFNSLLDTDLPMVHSEGTSTPTQLLRHPNIWFGNLPPPPRRPQDNRDFSP
Consensus     M--GL------------------------------L-----L------------------------PP---Q-------

                81                                                               160
ELV       48K LQPLGKVLQCGLSRGLHETLKVLQTPREELPLQISSQLPTSPSRLQSVPTSQHKPAYREAILHPRLPDVFYAAPNLRFVR
EMV-Trin  70K VQSLACSRSPAFNCDVHETVQVSKTCGSKPKIPRVDQLSTHCRRHHSLPLHLTHFSKQFNLLHARCSDVLFSSSDRRSLH
KYMV-JB   82K LHPLDASLHPAFICPLHEASKIHEAPKEEQVLPTSPQLPPDSHRFSAFSFNLSPPPKHPFRVHARCSDVLPTRTDPSSVS
OYMV-Tin  66K LQPLELHGDRPGLCDVHETLQVQKIGLRQSKLLRTSKLPSHRGRFRSIPLHLHLSSEVRNRFHARCPDVLQPLSDTRPLH
TYMV-CL   69K LHPLVFPGHHSQLRHVHETQQVQQTCPGKLKLSGAEELPPAPQRQHSLPLHITRPSRFPHHFHARRPDVLPSVPDHGPVL
Consensus     ---L------------HE---V---------------L-----R------------------H-R--DVL----------

                161                                                              240
ELV       48K KLPKPSQSVRQPRRSSRKLNDRLVSKPRSVSVLNPQVDSPLHTRRALSRQLQSARQRPRLAENLCDPNSIPEPVRVRPGI
EMV-Trin  70K SVSRTRDPVLQSHSASRVSFHRSLSLPRDLHLQDLRSDSPLHPGESPLRLVQSAPPSPILAEDFLHPLAFPRFVCDQAGI
KYMV-JB   82K PSSAVDQPLLLSCDPTRIPLHPPLPHARPLHLHPERSNPPLHPRGTLRRFIQPAHNSPLLAENQLHPLPKSQPLHHNSGI
OYMV-Tin  66K SVPLSSEAPLQSSCSPRKLLHRPLSPPKPLHLHNFRQHSSLCSRRSPCRKLRPTPRCNQLAQAQQHPLPSSKPLSLQAQI
TYMV-CL   69K TETKPRTSVRQPRSATRGPSFRPILLPKVVHVHDDPPHSSLRPRGSRSRQLQPTVRRPLLAPNQFHSPRQPPPLSDDPGI
Consensus     ----------------R----------------------L-------R----------LA-----------------I

                241                                                              320
ELV       48K LGPSPF..............PSYREEL..............SNTKSRFS......EDKGSDFLPNPPSSNSSEPGLP
EMV-Trin  70K LGPSPLHIDPARPTTKALSLCTPPRPAKSTSPCNNSQLPKPTAASDKPAILRT...AAASTQPGLLPNSRLRRTATSHLS
KYMV-JB   82K LGTSPFHTDTKR........SPPPRSE..............AASSQSSSILP...LPRPGNRPGVLPGPKVRGAPSSNLP
OYMV-Tin  66K LGPCPLPPHNKR.............................SPSPAVIR.......KTAGILPHPKLPPSSRGHLP
TYMV-CL   69K LGPRPLAPHSTR........DPPPRIT..............PGPSNTHDLRPLSVLPRTSPRRGLLPNPRRHRTSTGHIP
Consensus     LG--PL-----------------------------------S-------------------LP------------

                321                                                              400
ELV       48K GRPSQAPSCASENLRRPLHLHSSNPNSPHFRSRWIRKNSVKQARVQLGHISSLGQSPNLRPLD.................
EMV-Trin  70K APTSPPPASANKRLQRSLHLHSRSPHSSHFRPSRICANSKQQTRARLGHSKRVGQSADL.............VCQCPPP..
KYMV-JB   82K LPTPQAPPRARERLQRSLHLHASRQNAPRLRPRRLRSDPIQQTRTQVGHPQCVGQPTNLRAAQLPSPTKRGVPRSPQPTP
OYMV-Tin  66K SSTSSSSPRSNRGVRCSVHLHKSRSNSQDLRSCRVRSNSLQQTPILMGHFKSLGQSPNLRSFERPRPTR...........
TYMV-CL   69K PTTTSRPTGPPSRLQRPVHLYQSSPHTPNFRPSSIRKDALLQTGPRLGHLERLGQPANLRTSERSPPTKRRLPRSSEPNR
Consensus     ------------------HL---------FR----------Q-----GH----GQ---L--------------------

                401                                                              480
ELV       48K ....SLLQA.........SRVLHATSVSSY........KTQGASHEECFKTRSDGITSSHSGHLHHNDR.VEHQLQQGPL
EMV-Trin  70K .........................PPSMLPLLLLPRGKVKAPLR....PTLASLSFGSHP.........IPYHVTSSPP
KYMV-JB   82K ENEAVLLTALAPPRRHSRSWPLLPFPPSSLPEVVFASSKGKVHLRLSTQPPSSPQTSSSPPPPSPRTDSGIQTPLAS.PP
OYMV-Tin  66K .................RSLRLLPLSPQKVPTV...........HVPTHQQSGHKGPSLPRPHSP....SRQTHHARLPH
TYMV-CL   69K LPKPLPEATLAPSYRHRRPYPLLPNPPAALPSIAYTSSRGKIHHSL....PKGALPKEGAPPPPRRLPS.........P
Consensus     --------------------------------------------------------------------------------

                481                                                              560
ELV       48K LQCSQDSLAEPTYRPGTPSLPNFC........IAAKELSSTQPS.....................................
EMV-Trin  70K LIQFQHPFPPPS.........ATFSVSPLGVLTTAFALNPTQSAERSCDPFSPTPTLGHKTTSLSRLPLPPPHSPHSAQ
KYMV-JB   82K SKRKEKSLPHPSHQPPSHSKRNLRRHSALPLLLPIHPTKTTQPH.PAVPQPTAGPTP..HPPPTKKIPLHPPKS....QE
OYMV-Tin  66K SKR..VSLPN.SVLHHDRPKRPI.HFGSFPIMVAPSHLLPRKLWSRASSPPTCSPTTSNHGHPEEALRFLPK......NL
TYMV-CL   69K APRPQL.............PLRDLGRTPGFP.TPPKTPTRTPESRITASPTDIA.........PLDSDPVLSVRTEVHAPE
Consensus     --------------------------------------------------------------------------------

                561                                                              640
ELV       48K .....................................................................................
EMV-Trin  70K DRASALATDVSNSETKNCPSPTVP......PPFLPNHLHPLLPGTDPPTTPRQLSPSPSSLSLRTFLDSAVISCDSS...
KYMV-JB   82K RHPSPPPDVFHDCQPSSPTSHVVGYRRLLGSGISLPFKLAFWRRRSPNPARHLPPPPPPRKLHSELDSNSLSRSSSFSLS
OYMV-Tin  66K PQHCQMALMENYCSHFSSPSSSVSFPEDHQSSL.............PPISTRWVQCSSPSFSLQSFLVIGDIPCPISFPLS
TYMV-CL   69K RRTFMDPEALRSALASLPSPPR.......SVGIIHTAPQTVLPANPPSPTRHLPPTSPPWILQSPVGEDAIVDSEDDSIS
Consensus     --------------------------------------------------------------------------------
```

```
                641                                                                    720
ELV       48K ..........................................................................
EMV-Trin  70K ...PVLPPSPSPSSH....SSSSFQSCTSPPRFPC.........................YPPPPSALDLLFSSTPG
KYMV-JB   82K PLCSLPPSPSPSEQLPTVSAPPTPSPIAPTPLPGSRHPSPIRPTNPRRTTPCSSNHRTQTREFQPKQPQPKLLGRFKPPT
OYMV-Tin  66K S.....PQSHSSESLRGDSPPSSHLPSSPSSACSGDSFASCSSFGPSNPTSASS.......ALGGNHFNFSFFS......
TYMV-CL   69K SFHS.HDFDSPSGPLRSQSPSRFRLHL........RSPSTSSGIEPWSPASYDYGSAPDTD...................
Consensus     ------------------------------------------------------------------------------

                721                                                        784
ELV       48K ..............................................................
EMV-Trin  70K TDPFPPPDTPPPHKRTIALERLRLRQCATPFHSCDDVC.........................
KYMV-JB   82K EEFLQRQSPRSQQTYAYLLKYYTPQPPSPIR.IDSLTLSERWAAELFRSPSPSRSYQHHPFSSP
OYMV-Tin  66K ..............................................................
TYMV-CL   69K ..............................................................
Consensus     ------------------------------------------------------------
```

---

* Progressive alignment: Feng and Doolittle, 1987)
Gap penalty = 8
Length with gaps = 781
Actual sequence lengths: 750, 627, 649, 597, 440,

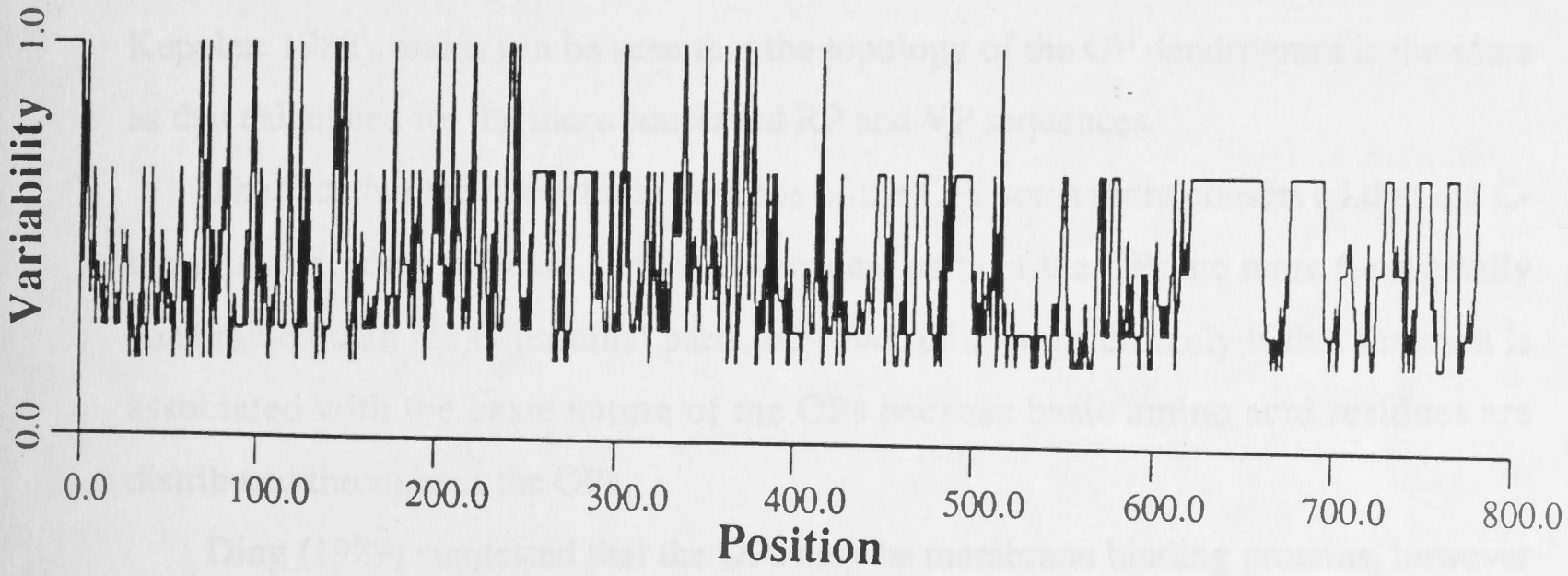
Percent identity (based on aligned regions):

	KYMV	TYMV	EMV	OYMV
TYMV	32.10			
EMV	28.34	28.97		
OYMV	26.49	28.10	33.50	
ELV	25.74	28.86	27.69	29.93

Distance scores for phylogenetic tree:

	EMV	OYMV	KYMV	TYMV
OYMV	117.55			
KYMV	138.99	156.67		
TYMV	130.07	140.00	118.52	
ELV	166.85	154.21	183.65	155.73

Figure 17. Graph showing the variability at each position of the aligned amino acid sequences of the OPs of five tymoviruses; ELV, EMV-Trin, KYMV-JB, OYMV-Tin and TYMV-CL.


Variability
1.0
0.0
0.0
100.0
200.0
300.0
400.0
500.0
600.0
700.0
800.0
Position

terminal sequences show little similarity. This was confirmed by estimating the variation at each position of the aligned proteins by the HETZYG program, which calculates for each position a variability index, that is analogous to the classical 'heterozygosity index', using the formula:

$$\text{Variability} = {}^{1}_{21}\Sigma\ P^2$$

where P is the proportion of each of the 21 amino acids at that position; a gap is considered as the 21st amino acid. Figure 17 is a graph of the variability index for the OPs. It can be seen that the sequences vary throughout their length, but particularly in their N-terminal half.

The progressive alignment program also estimates a 'distance score for phylogenetic tree' between each pair of sequences (Feng *et al.*, 1985). These 'FJD distances' (Feng *et al.*, 1985) estimates were then used to calculate a dendrogram (Figure 25) by the neighbour-joining method (Saito and Nei, 1987; Studier and Keppler, 1988), and it can be seen that the topology of the OP dendrogram is the same as that calculated for the more conserved RP and VP sequences.

The fact that the N-terminal portions of the OPs seem more conserved that the C-terminal parts may indicate that the N-terminal parts of the OPs are more functionally constrained than the C-terminal parts, however this seems unlikely if that function is associated with the basic nature of the OPs because basic amino acid residues are distributed throughout the OPs.

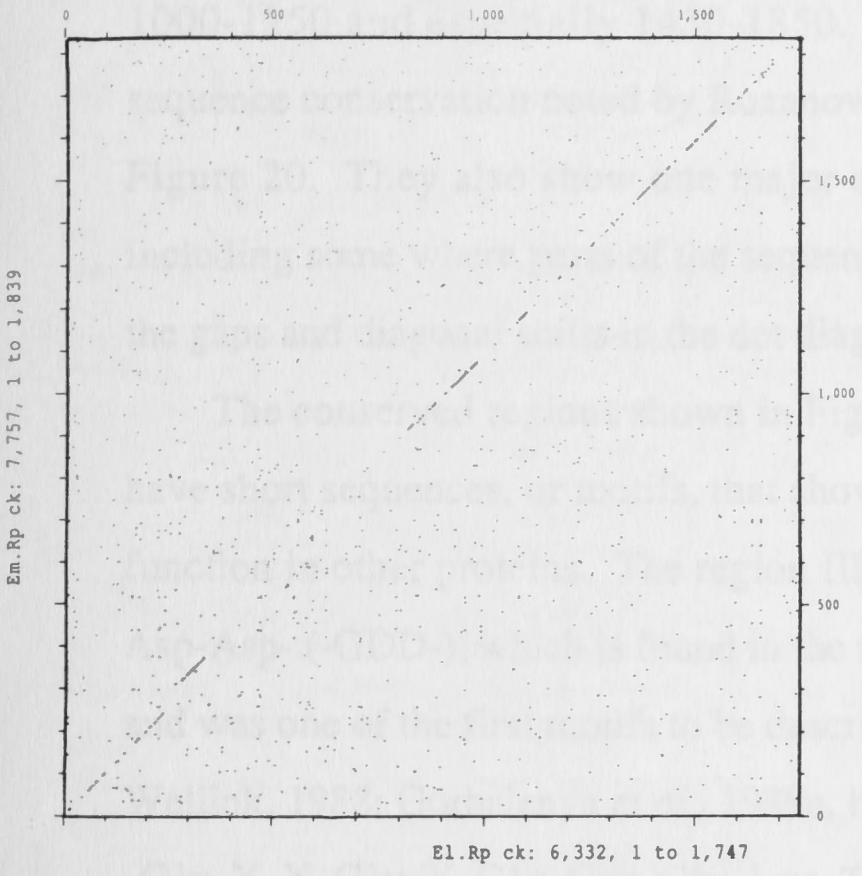
Ding (1989) suggested that the OPs may be membrane binding proteins, however further analysis of their hydrophobicity fails to reveal any consistent pattern of hydrophobicity (data not shown). It seems more likely that, because of their high isoelectric points, they have a histone-like function.

b) The Replicase Proteins (RP)

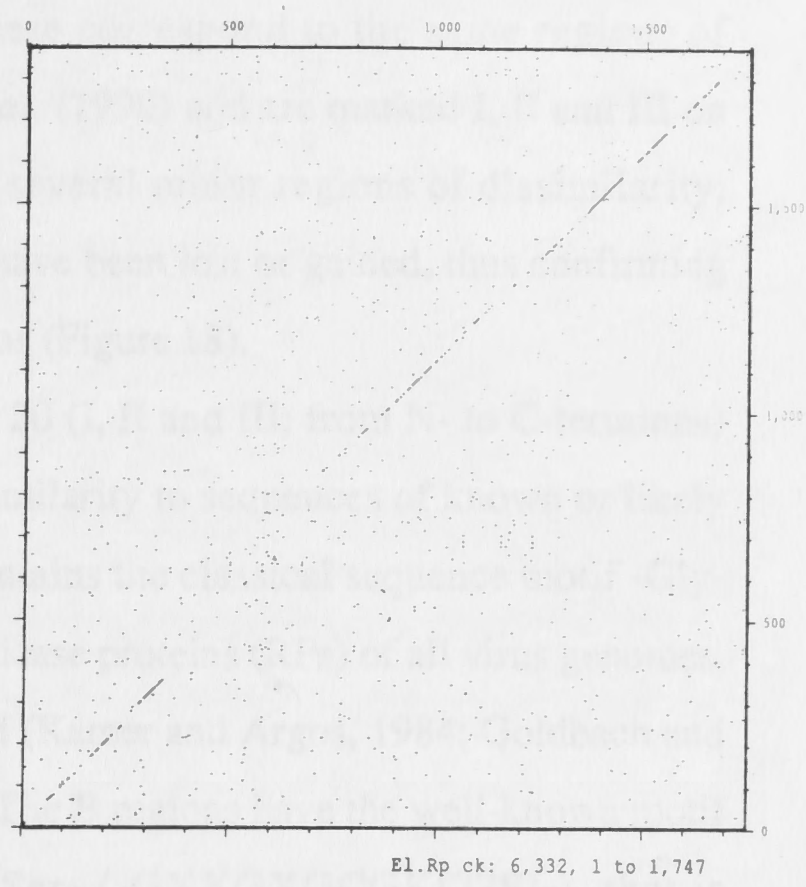
The largest ORFs of the genomes of the tymoviruses, ELV, EMV-Tin, KYMV-JB, OYMV-Trin and TYMV-CL, have closely similar sequences (identities about 50-56%) and encode proteins of M_r 194K, 205K, 210K, 198K and 206K, respectively. The amino acid sequences of these proteins are also clearly homologous as is shown in dot diagrams (Figure 18).

Figure 18. Dot diagram comparisons of the ELV replicase protein with those of four other tymoviruses. The ELV RP is located on the ordinate, and the RPs of EMV-Trin (A), KYMV-JB (B), OYMV-Tin (C) and TYMV-CL (D) are located on the abscissae. A dot is placed on the plot whenever a window of three amino acids matches exactly in the two sequences; calculated by DOTPLOT.

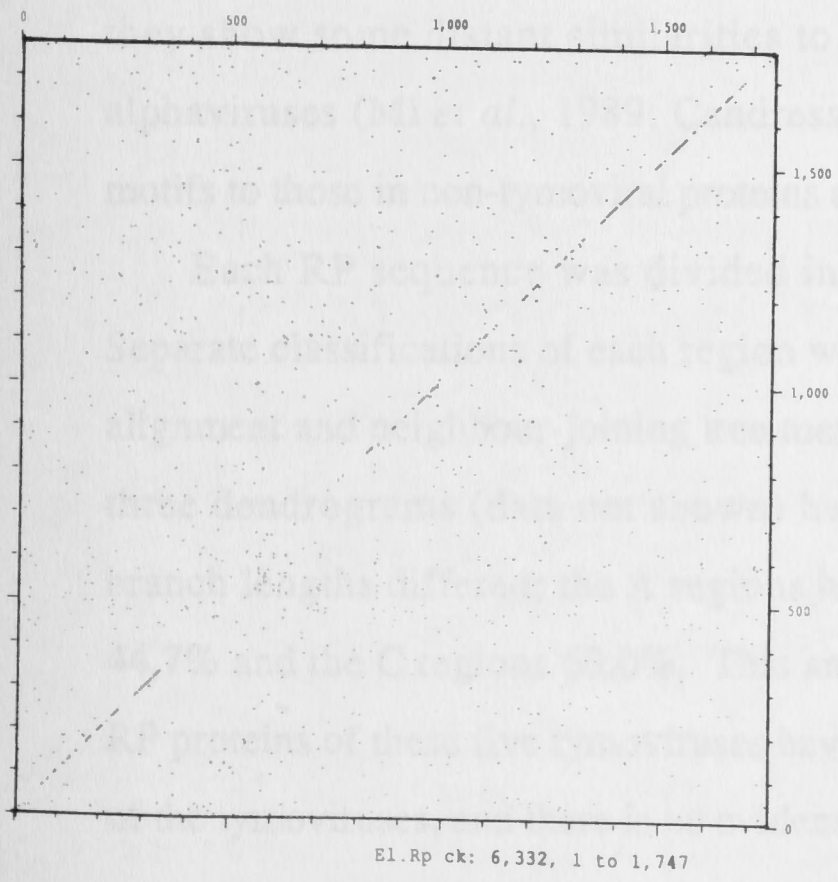
A


0
500
1,000
1,500
Em.Rp ck: 7,757, 1 to 1,839
El.Rp ck: 6,332, 1 to 1,747

B


0
500
1,000
1,500
Jb.Rp ck: 8,952, 1 to 1,874
El.Rp ck: 6,332, 1 to 1,747

C


0
500
1,000
1,500
Oy.Rp ck: 2,690, 1 to 1,776
El.Rp ck: 6,332, 1 to 1,747

D

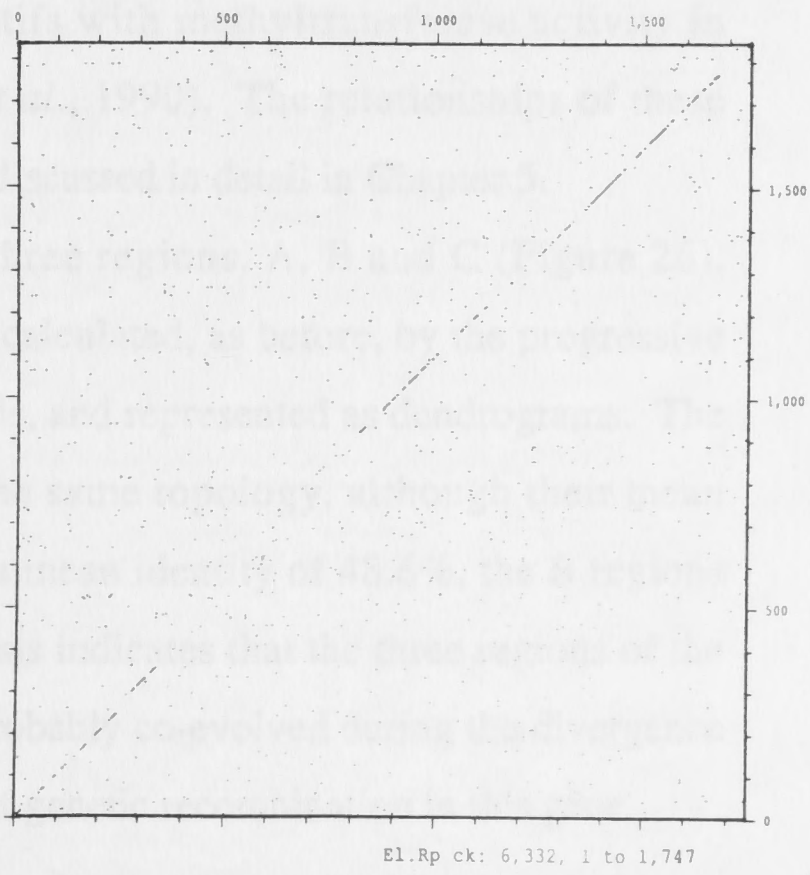

0
500
1,000
1,500
Cl.Rp ck: 7,651, 1 to 1,844
El.Rp ck: 6,332, 1 to 1,747

Figure 19 shows the sequences aligned by the progressive alignment method, together with a graph of the variability indices (Figure 20) for each position along the sequence. These figures confirm the dot diagram analyses, and show that there are three major regions of close sequence similarity, which are around positions 1-450, 1000-1250 and especially 1450-1850. These correspond to the same regions of sequence conservation noted by Rozanov *et al.* (1990) and are marked I, II and III on Figure 20. They also show one major and several minor regions of dissimilarity, including some where parts of the sequence have been lost or gained, thus confirming the gaps and diagonal shifts in the dot diagrams (Figure 18).

The conserved regions shown in Figure 20 (I, II and III; from N- to C-terminus) have short sequences, or motifs, that show similarity to sequences of known or likely function in other proteins. The region III contains the classical sequence motif -Gly-Asp-Asp- (-GDD-), which is found in the replicase proteins (RPs) of all virus genomes, and was one of the first motifs to be described (Kamer and Argos, 1984; Goldbach and Wellink, 1988; Gorbalenya *et al.*, 1989a, b). The B regions have the well-known motif -Gly-X-X-Gly-X-Gly-Cys-Gly-Lys-Thr/Ser- (-GXXGXGCGK[T/S]-), that is characteristic of many NTP-utilizing enzymes, and is found in some, but not all, viral RPs (Gorbalenya *et al.*, 1989b). Of the three conserved regions, the region I shows least similarity although those of different tymoviruses are unequivocally related, and they show some distant similarities to motifs with methyltransferase activity in alphaviruses (Mi *et al.*, 1989; Candresse *et al.*, 1990). The relationships of these motifs to those in non-tymoviral proteins are discussed in detail in Chapter 5.

Each RP sequence was divided into three regions, A, B and C (Figure 26). Separate classifications of each region were calculated, as before, by the progressive alignment and neighbour-joining tree methods, and represented as dendrograms. The three dendrograms (data not shown) had the same topology, although their mean branch lengths differed; the A regions had a mean identity of 48.6%, the B regions 44.7% and the C regions 60.0%. This analysis indicates that the three regions of the RP proteins of these five tymoviruses have probably co-evolved during the divergence of the tymoviruses, and there is no evidence of genetic recombination in this gene.

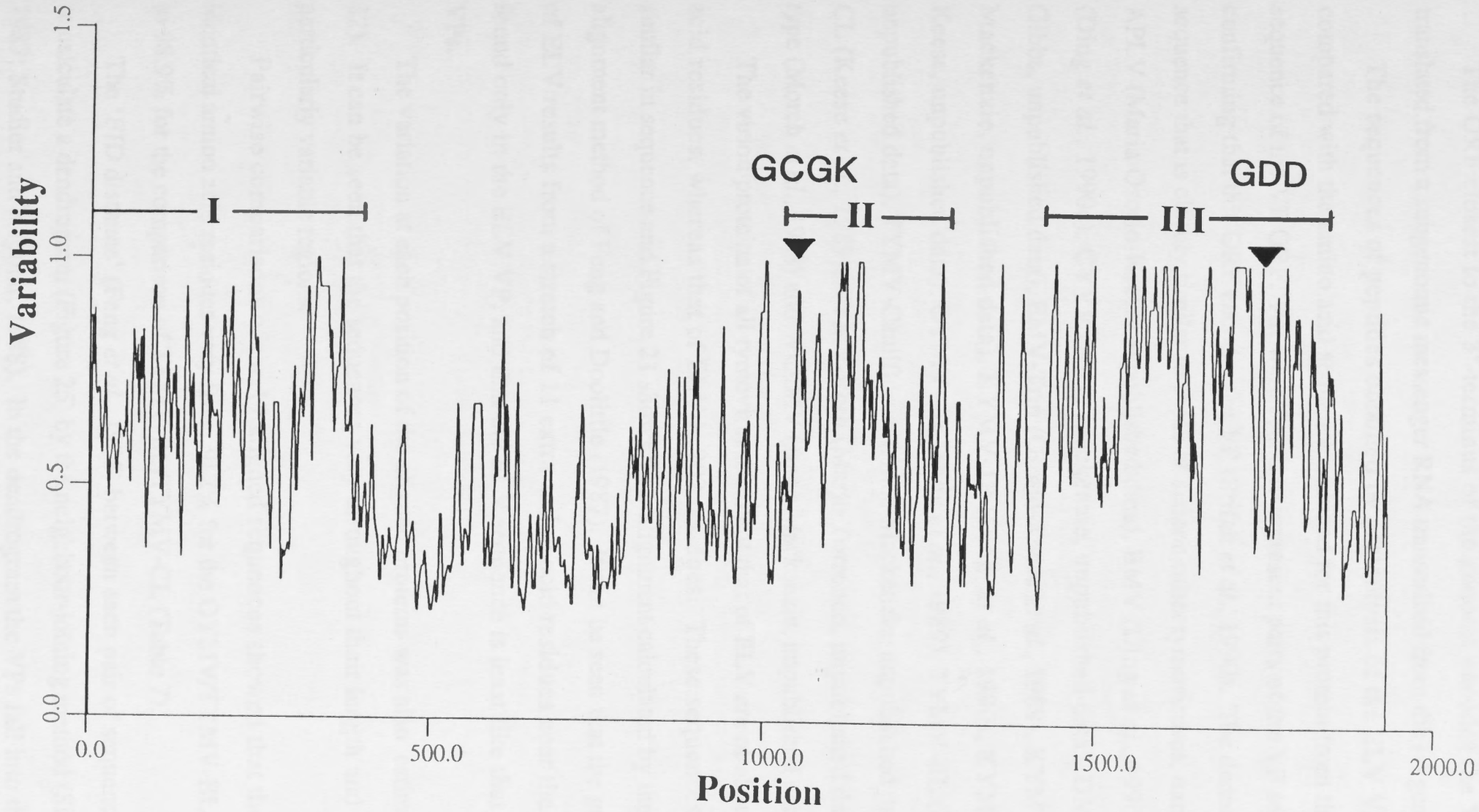
Figure 19. Alignment of encoded replicase proteins of five tymoviruses, ELV, EMV-Trin, KYMV-JB, OYMV-Tin and TYMV-CL, by the progressive alignment method of Feng and Doolittle (1987) and PRETTY program.

```
                1                                                                80
ELV       194K MAFQLALDALSSTTHRDSISAPLLDSSVSQLQSSLELFPYTVPKELVPQLNRMGIQVSGLTSTPHPHAAHKTLELNLLFN
EMV-Trin  205K MAFQSALEALNSTTHRDASTNPILNSVVEPLRDSLSLYPWLLPKEAVPHLLSWGIPNSGLGVTPHPHPIHKTVETFLLFN
KYMV-JB   210K MAFQLALDALASTSHKDPSLHPVLESVHDSLTDSLQTYPWMVPQDLQPFLLKSGIPINSFGSSPHPHPAHKTLETHLLFT
OYMV-Tin  198K MAFELALNALASSTHKDSSLNPVLNSAVQPLQTSLQNFPWIIGKEHLPFLLAAGIPTSGFGCNPHPHAVHKVIETFLLFN
TYMV-CL   206K MAFQLALDALAPTTHRDPSLHPILESTVDSIRSSIQTYPWSIPKELLPLLNSYGIPTSGLGTSHHPHAAHKTIETFLLCT
Consensus      MAF--AL-AL----H-D----P-L-S-------S---YPW----E--P-L---GI----L----HPH--HK--E--LL--

               81                                                               160
ELV       194K HWAKSCNVDSAVVFMKPSKFFKLQEKNSHFKSLHNYRLHPHDSNRYPHPSTSLPTEKRFYIHDSLMYFTPHQISGLFESC
EMV-Trin  205K HWHALARLPSTVMFMKPSKFQKLAALNPKFQELINFRLTAADTTRYPSTSLTFPSNSICFMHDALMYFSPAQIVDLFTQS
KYMV-JB   210K HWMHLCTQPSSVLFMKPQKFMKLQRKNKFFQHLHNYRLTPTDSVRFPSTSPHLPNTPFVFMHDALMYYQPEQILHLFHQV
OYMV-Tin  198K HWSFMATVQASVMFMKPSKFKKLASVNPNFSELVNYRLTAADSVRYPSTSTSLPKYEIVFMHDALMYFNPSQILDLFIQC
TYMV-CL   206K HWSFQATTPSSVMFMKPSKFNKLAQVNSNFRELKNYRLHPNDSTRYPFTSPDLPVFPTIFMHDALMYYHPSQIMDLFLQK
Consensus      HW---------V-FMKP-KF-KL---N--F--L-NYRL---D--RYP--S--LP-----F-HD-LMYF-P-QI--LF---

               161                                                              240
ELV       194K PNLLSLYASLVVPPESSMTDLSLNPDLYRYSIHKSTLHYTPEGHSAGSYNQPVNALDWLKISAIQTPSLSLSVSVLESWG
EMV-Trin  205K PALETLYCSLIVPPESHFTDLSLFPEIYTYKISGQTLHYIPENHHSGSYNQPLQALSWLKISSILSPSLALSVTKLESWG
KYMV-JB   210K PQLTNLFCSLVTPPESHFTHLSLMPDLYTFTLKGQTLHYTPEGHSAGSYNQPITALSWLKINSILSPNLNLSITILESWG
OYMV-Tin  198K PSLQRLHCSLVVPPESSFTDLSLHPNLYTYTISGNTLHYVPEGHHAGSYDQPLDAISWLKLNSIHSPHLNLSVSKLESWG
TYMV-CL   206K PNLERLYASLVVPPEAHLSDQSFFPKLYTYTTTRHTLHYVPEGHEAGSYNQPSDAHSWLRINSIRLGNHHLSVTILESWG
Consensus      P-L--L--SLV-PPE--F---SL-P--Y-Y-----TLHY-PE-H--GSY-QP--A--WL----I------LSV--LESWG

               241                                                              320
ELV       194K PLHSLLIERSSQ.............TQN.......................PDSQKIKDLISFQTPQALILPNPDSLAVPLR
EMV-Trin  205K PVHSILIQRGLPPKPSLSARPPVLPNQPPRATTPNSQNQLLHQTSQLFFELQQPQLSLVSFRIPDCVELPQATFLRQPLR
KYMV-JB   210K PLHSILIQRGLPL.....PDPKLLVRS.............LPP....FSRSPDPETDLVSFQVPKSVELPQATFLSQPLR
OYMV-Tin  198K PVHSLLITRGLPH..........LPSS.....................EKQQ.....VSFHIPNCLPLPEATFLHQPLR
TYMV-CL   206K PVHSLLIQRGTPP.....PDPSLQAPS.............TPMASDLFRSYQEPRLDVVSFRIPDAIELPQATFLQQPLR
Consensus      P-HS-LI-R-------------------------------------------------VSF--P----LP----L--PLR

               321                                                              400
ELV       194K HRLVPQKTYDALFTYTRATRTLRTSDPAGFVRTQSNKPEFNWVTSQAWDNLQTYALLTASYRPPVSYTLHRSPLTKLKEL
EMV-Trin  205K HRLVPTSVYNALFTYTRAVRTLRTSDPAGFVRTQSNKPEHAWVTPNAWDNLQTLS.VNAPHRPQVCYHFFSSPVARLKLH
KYMV-JB   210K HRLVPESVYNALFTYTRAVRTLRVSDPAGFVRTQSNKPEHKWVTPSAWDNLQTFALLNCPLRPNVVYHVLLNPLQKMKLY
OYMV-Tin  198K HRLVPTEVYDALFTYTRAVRTLRTSDPAGFVRTHSNKPQYSWVTSRAWDNLQTYALLNAPVRPVVLFDFFLSPLKKFQLF
TYMV-CL   206K DRLVPRAVYNALFTYTRAVRTLRTSDPAAFVRMHSSKPDHDWVTSNAWDNLQTFALLNVPLRPNVVYHVLQSPIASLALY
Consensus      -RLVP---Y-ALFTYTRA-RTLR-SDPA-FVR--S-KP---WVT--AWDNLQT--------RP-V-Y-----P---L---

               401                                                              480
ELV       194K LTRNALKLAAMASPALT.LAIFTTMTALNTNSSKALSFSALKIHLLNPLTGPE...........LLHFQT..........
EMV-Trin  205K FAQHWRAYLLALTPFLTTSPLLLPLFNFNTPFP.LPRLLSLFRRSV...SSP.R..........LLHSILP.........
KYMV-JB   210K FSQHWRRLGVIAAPGLFCLSLLLRSQKWSLPLP.KAKSISVFRRNLLLPPKPHRPPLLPHPEQMLQEFKLPWHRPPPKGK
OYMV-Tin  198K MSQHINSLVIKALPFLGLIPPLVKLTTLGFPIPSVSHFQILFFTMI...GPSGQ..........FILEAFP.........
TYMV-CL   206K LRQHWRRLTATAVPILSFLTLLQRFLPLPIPLA.EVKSITAFRRELYRKKAPHHP.........LDVFHLQQH.......
Consensus      -------------P-L-----L-----L-----------------------------------L--------------

               481                                                              560
ELV       194K ............SVLQQKNSAPLSQAEAKQELDKSAVPAPSEHDSSASQSTSLSLSASSQLLSTEKHPGSELSSKAIPVS
EMV-Trin  205K ............SQLRGAAIPNRPLPLWVTKL.HHFLDSHS......LLPTP.PIRPRIEL.....QRLPLMSLIPKPK.
KYMV-JB   210K RNPFLTLLINLLHIPREICAGIRRYPSYYQSIQPKPLNPIQQFRNQLLAQLH.TLPLPKKFLSILPRARKDIPLLPRMSF
OYMV-Tin  198K ............SMLH.PAISYLESCGLVPRL.PPPV.AQQ......LQTMG.ILRKRSAF.....SLKTSLSTVKWPSW
TYMV-CL   206K ............LRNHHSAISAVRPASPPHQRL.PHALQKAALL...LLRPIS.PLLTATPFFRSEQKSMLPNAELSW.TL
Consensus      ------------------------------------------------------------F------------------

               561                                                              640
ELV       194K TSCPSASKQLAPPLTAE.SHSSVNALLRKFLGPNSPQSNLDNYNLHLHPESFTLGWKRRPLLLDSHSSFLP.........
EMV-Trin  205K IVLPL.....LSLL.....LSSPTIYIH.FFQAQTPQQLHDNYHLHLHPSRFELSWTLQSYHVTQAQSFLPLLLPAPTQA
KYMV-JB   210K TTVNLPLQPPMWLAIGASLVPELAFLLSWLSGDVDLQTQHDIYHHHLHPENFTLSWTRTPYLALAPSPFLPYAHSPLPPL
OYMV-Tin  198K KITAL.....ISAL.....PAALSVFLK.TISPLSLQSLHDGYNAHLHPSPFNLSWSLETFHVQSPSPFLPLSLTPPSPS
TYMV-CL   206K KRFALPWQASLVLLS....LSESSVLLHKLFSPPTLQAQHDTYHRHLHPGSYSLQWERTPLSIPRTTAFLPFTPTTSTAP
Consensus      -------------------------L----------Q---D-Y--HLHP--F-L-W------------FLP---------
```

```
                641                                                                                                  720
ELV       194K ...SSCLQPPASPSIAAAPHP.............LPPAQKP......PRPPT.............................
EMV-Trin  205K QASNPAPRPPAFHAIPLPPQPSTSSSPPLQEPTLSPHLIHPP....LTREPSPLNGCACDSALLPSTAAMTSAEHPTPLN
KYMV-JB   210K PVNSSPLFPP..PPPLPPSQPPLSQGPATQAPSAQPTPGEPL....LAPPTTELKP.ESSNPNNPNPSSSAGSNPPPKSS
OYMV-Tin  198K E..EIALPPPIFR.VPPPLPAQETPSPP......APALVPPT....QPPQPQPWEEII..STFLSS...LNASS......
TYMV-CL   206K PDHSEASLPPAFASTSVPRPPPVASSLGAQPPTTTAAPPTPIEPTQRAHQNSDLTL.ESSTPIEPPPPPIQSSDIPPSAP
Consensus       --------PP------------------------------P-------------------------------------

                721                                                                                  800
ELV       194K ....TVPTPKPLASPSQTQAA.QPATQSPPS.....................IPQTAPVTSLLP.............APL
EMV-Trin  205K ...PPTPSPTPDVPPPDSPGNPSLLKQVPPEANLHPIHNPDLPSSTTLPSGALTLVPAKTPSIYAN.PTPPSSH.PFTPL
KYMV-JB   210K ...SSDNPPAPNKPTPTSSSTTPPSPNLPLQFG..SIHSPFLS.DGQLNYSALPPPQDPTNTTLSLLPEPKPPTEVQSPL
OYMV-Tin  198K ..HKPSPSSAPLESPSPIPESFEVLAEAPQHQ..RSINE..........CGALNQLPLPVPSPLQN.STNPKPPFPSSDL
TYMV-CL   206K VLFPEINSPHRFSPKLPTTPDFEPTRTSPPPST..SHQDSTDP.ADPLMGSHLLHHSLPAPPTHPL....QSSQLLPAPL
Consensus       ----------------------------P-------------------------------------------------L

                801                                                                                  880
ELV       194K ETDDSCAGPISTFQDLFPASYYPHTANFPCRSKIPGYLEAPYPPLDCMLVALSAQMPQSPQELWSALNTLMPLSALTSPS
EMV-Trin  205K ADDPTAVGPCLPFHVLHPADYFPLSAEFLTRTRHVPPSSLSHPKLNCLLTCFSELSGHSESDLWLSLQSILPDSQLQNPE
KYMV-JB   210K MADPTCVGPAVSFSSLYPRDFFPNTASFLTRLRLSPPTPLPMPKNNCLLTAVAPSLHINPHRLWTSLQEVLPDSLLSNSE
OYMV-Tin  198K LSDMSCTGPVVEFETIFPAEYHMSNGSFPTRLRGHPRSSAPFPQKHCLLTAVASQLSYTEHQLWEFLCDMLPDSLLTNSE
TYMV-CL   206K TNDPTAIGPVLPFEELHPRRYPENTATFLTRLRSLPSNHLPQPTLNCLLSAVSDQTKVSEDHLWESLQTILPDSQLRNEE
Consensus       --D----GP---F----P--Y------F--R-----------P---CLL-------------LW--L---LP-S-L----

                881                                                                                  960
ELV       194K LRVLGLGTEELTALSYYYHFQAEIHSDNEIYRFGIQTASTKLCLIRDSGPPAHFTAPDPLRAGSPPSRSQTNENSLRRSL
EMV-Trin  205K VSTLGLSTDILTALCFIYHSSVTLHAPSGVYHYGIASSSTVYVIHYQPGPPPHFSLSPRLAASAPRCNPT.N.SRLVRQA
KYMV-JB   210K IDSVGMSTDLLTALSHLFNFQAVVHSERGDILFGLQSAKTVIHIYHTNGPPAHYSPPPKIIGSNSPPSSQ.Q.HPLEQAA
OYMV-Tin  198K VENFGLSTDHLTCLSYRLHFECIIHTSHSTIPYGIKKASTVIQISYIDGPPKHFKAFIKLAAAAPGSNPS.K.SNLVRAA
TYMV-CL   206K INSLGLSTEHLTALAHLYNFQATIYSDRGPILFGPSDTIKRIDITHTTGPPSHFSPGKRLLGSQPSAKGH.PSDSLIRAM
Consensus       ----GL-TD-LT-L---Y--------------FG--------------GPP-HF--------------------L----

                961                                                                                 1040
ELV       194K LGFRLNGNLLPIDQVHSFTSEPSRAKNLASNMKNGFDGILTTLAALSSLSSGPSPRDRIFTLDGICDFALPKTVDLIHLS
EMV-Trin  205K LRFKLNGEFLPFTQAYAHESSITHAKNLISNMKNGFDGIMSSLTDSSK...GPSPREKLTTLDSLIDVAAPREVSLIHIA
KYMV-JB   210K LRFKYQGSHLPFSSFHSFTTSVQHAKNLISNMKNGFDGVMSTIEPSIRHQPGHSPREKFIALDAMIDLARPKTVSMFHLA
OYMV-Tin  198K LRFKYNDAFLPFWDAHQHTISVPHAKNLISNMKNGFDGITSQLSGPSN....KSPKMKLLELDATIDVSFPRKCDVIHIA
TYMV-CL   206K KSFKVSGNYLPFSEAHNHPTSISHAKNLVSNMKNGFDGILSLLDVSTGQRTGPTPKDAIIQIDHYLDTNPGKTTPVVHFA
Consensus       --F------LP-------------AKNL-SNMKNGFDGI--------------P-------D---D----------H--

                1041                                                                                1120
ELV       194K GFAGCGKTHPIQQLLKTPHFHNFRVVTPTTNLRSEWKSDMALPAHHNWRFSTWESALLKHAEILVIDEIYKLPRGYLDLS
EMV-Trin  205K GFAGCGKTHPIQKLLQTSPFHDFRISCPTNELRSEWKRDMQPTAENVWRFSTWESSLLKHSEILVIDEIYKLPRGYLDLS
KYMV-JB   210K GFAGCGKTKPLQSLLSTRPFHSFRVSTPTTELRNEWKKDMNLPASQAFRFCTWESSLLKQTKILVIDEIYKLPRGYLDLC
OYMV-Tin  198K GFPGCGKSHPIQKLLQTPAFRHFRLSVPTNELRSEWKRDLNLPESEVWRLCTWETALFKSSNILVVDEIYKLPRGYLDLI
TYMV-CL   206K GFAGCGKTYPIQQLLKTKLFKDFRVSCPTTELRTEWKTAMELHGSQSWRFNTWESSILKSSRILVIDEIYKMPRGYLDLS
Consensus       GF-GCGK--P-Q-LL-T--F--FR---PT--LR-EWK--M-------WRF-TWE---LK---ILVIDEIYKLPRGYLDL-

                1121                                                                                1200
ELV       194K LIADPTVKL.ILLGDPLQGEYHSTSAHSSNLRLSSEIPRLLPFIDYYCYWSYRVPKCVAKLFSLPCFNPSEGF....IKT
EMV-Trin  205K ILADPTLSLVIILGDPLQGEYHSTSPHSSNHFLPSEVHRFKSYIDCYCFWSHRIPKQIASLFGVVCHNTNEGF....VRA
KYMV-JB   210K ILADPCLELVIILGDPLQGEYHSTSPHSSNHQLQSETTRLLPFIDHYCWWTYRVPSHIADLFSVPSFNRSEGHYQMAVRT
OYMV-Tin  198K LLADPSIQLVIMLGDPLQGEYHSSHPSSSNSRLESETTRLSKYIDCYCWWTYRCPKAVADLFGVKTFNSNEGF....IRA
TYMV-CL   206K ILADPALELVIILGDPLQGEYHSQSKDSSNHRLPSETLRLLPYIDMYCWWSYRIPQCIARLFQIHSFNAWQGIIG.SVST
Consensus       --ADP---L-I-LGDPLQGEYHS----SSN--L-SE--RL--YID-YCWW--R-P--IA-LF-----N---G-----V--

                1201                                                                                1280
ELV       194K TLDFFPSANNLVNSHSVVHISEACGWNAVTISSSQGCTFSDPAFIHLDRNTALLSPSNCLVALTRSRSGVYFKGDFTFLS
EMV-Trin  205K LTSHPPNSKNLTNATNTALSLQQMGHHAITISARR.VTFTEAHTILLDRHTNLLSPNNCLVALTRSRTGVYFVGNLHLAS
KYMV-JB   210K ADSYTPGHFNLVNSVATANAVIQLGFPATTISASQGVTHHNRVTILLDKHSRLLSPSNTLVALTRSTVGVEFLGDIGSLS
OYMV-Tin  198K VLSHPPNLPNLVNSIATANTMQSLGHHALTISSSQGMTYSDPVTVLLDRHSLLITPQTALVALTRSRSGIYFIGSMYTAS
TYMV-CL   206K PQDQSPV...LTNSHASSLTFNSLGYRSCTISSSQGLTFCDPAIIVLDNYTKWLSSANGLVALTRSRSGVQFMGPSSYVG
Consensus       -----P----L-N-----------G----TIS-----T------I-LD-----------LVALTRS--GV-F-G------

                1281                                                                                1360
ELV       194K SLSGSSRMFSLAYSGQPIHLPDFFPEIVFQLNMITAPLTKRSSSFRSGF...QPNISSAPKIPAPPNLPCPPHIPTNYSK
EMV-Trin  205K NSFGTNYMFSQALCQGTIDLNNVFPHIMPHLPKMYEPIRSRSNRFVAGSLNFRPTTNSRLLSSLTKPTHLPPHIPTNHSL
KYMV-JB   210K GTNNSSDMFSRAIYRQPINLSSSFPRIFHLLPLLNKPISRRSTRLIGSH...SPIFHNPRLTNI....HLPPHIPTSYSQ
OYMV-Tin  198K GSAGTSYMFSCALTGLPVDMMSAFP.LFHTLPLIHEPIRSRRHRLVAGHT...PSLHVPPSNKWPHRLHLPPHIPTSHSK
TYMV-CL   206K GTNGSSAMFSDAFNNSLIIMDRYFPSLFPQLKLITSPLTTRSPKLNGAT...PSASPTHRSPNF....HLPPHIPLSYDR
Consensus       -------MFS-A-----I-L---FP-----L-----P---R---L------------------------PPHIP-----
```

```
                1361                                                                                                1440
ELV       194K  DVIVNNQALYGESLERRLSVLHLPPTRMTLHSDINITAPSSSSFQPSDEPVPSDHTAVYPGFDFFTLAAHFLPAHDPEVK
EMV-Trin  205K  DVLVSNPVLLGETLDPRLEVLHLPPTRLPLHLDLLPTVPSSSSFSSVDHLFPTPISPAICGYTFENLAAFFLPAHDPDLK
KYMV-JB   210K  DFVVSNPIFQGQ.ADPRLDTHFLPPTRLPLQSELLPAQLSQTTKPTDSFTNNTPFTPVYPGENFENLAAFFLPAHDPELK
OYMV-Tin  198K  DVILAHGIVASNAPERRLTTLHLPPTRLPLHFDLESCNPSTVSTSSTSN.SEVPFTHAFLGESFEELAAHFLPAHDPDLK
TYMV-CL   206K  DFVTVNPTLPDQGPETRLDTHFLPPSRLPLHFD.LPPAITPPPISTSVDPPQAKASPVYPGEFFDSLAAFFLPAHDPSTR
Consensus       D-------------E-RL----LPP-RL-L--D--------------------------G--F--LAA-FLPAHDP---

                1441                                                                                                1520
ELV       194K  EIELKDQTSQQFPWLNLDFHISCQTSSLISARHQPGSDSTLLPASLHKRLRFRPTAAPYQITPSDSFLGNCLYRSWCQVY
EMV-Trin  205K  EVLINDQKSNQFPYLDAPFELSCQPSSLLAPIHKPASDPTLLPGSIKKRLRFRASSSPYSITPSDQLLGQHLFSSLCLAY
KYMV-JB   210K  EVTRRDQTSAQFPWFDRPFSLSCQPSSLIAAKHSPSQDPTLLPFSIPKRLRFRKSDNPHVLSAIDVLLGNQLFFNLCKAY
OYMV-Tin  198K  EVTVGDQTSQQFPYLDQPYTLSCQPSSLLAASHKPASDPTLLIFSISKRLRFRASSSPYAFTPNDLILGHLLYTNWCKAF
TYMV-CL   206K  EVLHKDQSSNQFPWFDRPFSLSCQPSSLISAKHAPNHDPTLLPASINKRLRFRPSEAPHQITADDVVLGLQLFHSLCRAY
Consensus       EV---DQ-S-QFPWL---F--SCQ-SSL----H-P--D-TLL--S--KRLRFR----P------D--LG--LF---C--Y

                1521                                                                                                1600
ELV       194K  RRDPNVRLPFNEALFLECIAVNDYAQLSSKTQATIVANASRSDPDWRHTFVKIFAKSQHKVNDGSIFGPWKACQTLALMH
EMV-Trin  205K  GRNPNSVLPFQPELFSECICINDYAQLSSKTQATIVANHQRSDPDWRLTAVRIFAKAQHKVNDASIFSGWKACQTLALMH
KYMV-JB   210K  RRNPTHVGPFNPALFAECIALNDYAQLSSKTQATLVANHSRSDPDWRHTAVKIFAKSQHKVNDASIFGNWKACQTLALMH
OYMV-Tin  198K  GRCPNSTIPFNPALFAECICLNEYAQLSSKTQATIVSNASRSDPDWRYTVVRIFAKSQHKVNDGSIFGSWKACQTLALMH
TYMV-CL   206K  SRQPNITVPFNPELFAECISLNEYAQLSSKTQSTIVANASRSDPDWRHTTVKIFAKAQHKVNDGSIFGSWKACQTLALMH
Consensus       -R-P----PF---LF-ECI--NDYAQLSSKTQ-T-V-N--RSDPDWR-T-V-IFAK-QHKVND-SIF--WKACQTLALMH

                1601                                                                                                1680
ELV       194K  DYVILTLGPVKKYQRLFDQLERPSHIYYHAGNTPHDLRRWCSKHLE.TSHCTTNDYTAFDQSQHGEAVVFEVLKMRRLSI
EMV-Trin  205K  GYIILVLGPVKKYQRIFDSKDRPPHIYYHCGKTPSQLSQWCQTHLS.GSSYIANDYTAFDQSQHGEAVVLECLKMRRLSI
KYMV-JB   210K  DFVILSLGPVKKYQRIFDALDRPPHLYTHCGKSPADLSAWCQTHLT.GQIKLTNDYTAFDQSQHGESVILEALKMKRLSI
OYMV-Tin  198K  DFVILTLGPVKKYQRIIDHYDRPNFIYTHCGKTPSELSAWSHSFLK.GDAYICNDYTSFDQSQHGEAVIFESLKMHRVGI
TYMV-CL   206K  DYVILVLGPVKKYQRIFDNVDRPSHIYSHCGKTPNQLRDWCQEHLTHSTPKIANDYTAFDQSQHGESVVLEALKMKRLNI
Consensus       -YVIL-LGPVKKYQR--D--DRP---Y-H-G--P--L--W----L--------NDYT-FDQSQHGE-VVLE-LKM-R--I

                1681                                                                                                1760
ELV       194K  PENLISLHVHLKTNVETQFGPLTCMRLTGEPGTYDDNTDYNLAVLNLQYDLRKTPTLVSGDDSYLSGTLSPRSNWPFVKE
EMV-Trin  205K  PDSLIQLHSHLKCSVDTQFGPLTCMRLTGEPGTYDDNSDYNLAVIYSQYSLNGHPILISGDDSVLCGTPPPSPLWPTLKK
KYMV-JB   210K  PSHLIQLHVHLKTNVATQFGPLTCMRLTGEPGTYDDNSDYNLAVIHSQFDMKDIPVMVSGDDSLIDRQPPLAQSWEATKR
OYMV-Tin  198K  PRHLIDLHIYLKTNVSTQFGPLTCMRLTGEPGTYDDNTDYNLAVIFSQYVISDHPIMVSGDDSVICGHPPINPNWPAVEK
TYMV-CL   206K  PSHLIQLHVHLKTNVSTQFGPLTCMRLTGEPGTYDDNTDYNLAVIYSQYDVGSCPIMVSGDDSLIDHPLPTRHDWPSVLK
Consensus       P--LI-LH--LK--V-TQFGPLTCMRLTGEPGTYDDN-DYNLAV---QY-----P-MVSGDDS----------W-----

                1761                                                                                                1840
ELV       194K  LLHLRLKPSSLIDGLFCGYYLGPQGCIRNPLALFAKLMIAEDDGSAFDKLPSYLTEFSIGHGLGDSLWQLLPSDLVLYQS
EMV-Trin  205K  MLHLRFKIERTSHPLFCGYYVSPHGAARNPYALFAKLMICVDDKSLHDKKLSYLSEFSTGHLAGDLVTSILPSHLLPYQS
KYMV-JB   210K  LLHLRFKTEKTTHPLFCGYYTGSAGAIRNPLALFSKLMIAIDDEAIHDRRLSYLTEFSTGHQLGDALWTLLPESTQIYQS
OYMV-Tin  198K  LLHLRFKTEETSLPLFCGYYVGPTGCCRNPFALFAKLMISYDKGNLFETLPSYLYEFSIGHRLGDVVRLLFPDHLLKYYS
TYMV-CL   206K  RLHLRFKLELTSHPLFCGYYVGPAGCIRNPLALFCKLMIAVDDDALDDRRLSYLTEFTTGHLLGESLWHLLPETHVQYQS
Consensus       -LHLRFK-------LFCGYY----G--RNPLALF-KLMI--D-----D---SYL-EF--GH--GD-----LP-----Y-S

                1841                                                                                                1920
ELV       194K  ACFDYFCRKATRSQKILLQPGLVDQETLDKIALSAKFISRPFYSMLSSHARSLISTKFKLDSSLTTLQDPMVEFELL.PF
EMV-Trin  205K  AVHDFFCRNCTPAEKILLSLDPIPESKILQLILKVRWASQAFFSYLPQKARELLVARSSLPSLYSNPKVSQLESELL.PF
KYMV-JB   210K  ACFDYFCRHSPPHEKALLSSFELPDSVISKISSSTKWLSKNAFYALPSKIRKAVIASRHSSSFPENPDVSQLEFELL.QS
OYMV-Tin  198K  ACWDLFCRKCTASQKLILSFEPIPPSFFSKLASTSRWVSKVLFSDLPTKIRDMLISSSKLPSYHQDPRVQYLESELLTSF
TYMV-CL   206K  ACFDFFCRRCPKHEKMLLDDSTPTLSLLERITSSPRWLTKNAMYLLPAKLRLAITSLSQTQSFPESIEVSHAESELL.HY
Consensus       A--D-FCR------K--L------------------W--------L----R----------S----------E-ELL---

                1921
ELV       194K  SNVQ....
EMV-Trin  205K  SQ......
KYMV-JB   210K  FQF.....
OYMV-Tin  198K  NHGRLSTN
TYMV-CL   206K  VQ......
Consensus       --------
```

Figure 20. Graph showing the variability of the aligned amino acid sequences of the RPs of five tymoviruses, ELV, EMV-Trin, KYMV-JB, OYMV-Tin and TYMV-CL. The highly conserved regions (I, II and III) were analysed separately; region I possibly containing methyltransferase motif; region II containing GCGK[T/S] conserved sequence of a NTP-binding motif whereas region III possesses the GDD signal of RNA polymerase.


Variability
1.5
1.0
0.5
0.0
I
II
III
GCGK
GDD
0.0
500.0
1000.0
1500.0
2000.0
Position

c) The Virion Proteins (VP)

The ORF closest to the 3'-terminus of the genome encodes the VP, which is translated from a subgenomic messenger RNA transcribed from the (-) genomic strand.

The sequences of peptides obtained by proteolysis of the ELV VP have been compared with the amino acid sequence deduced for this protein from the nucleotide sequence of the VP ORF. All were found to represent parts of the VP sequence, thus confirming that this ORF encodes the VP (Srifah *et al.*, 1990). The deduced VP has a sequence that is closely similar to those of sixteen other tymoviruses, namely those of APLV (Maria Osorio-Keese, unpublished data), BMV (Ding *et al.*, 1990b), CoYMV (Ding *et al.*, 1990c), CYVV (Anne Mackenzie, unpublished data), DMV (Jennifer Gibbs, unpublished data), EMV-Trin (Osorio-Keese *et al.*, 1989), KYMV-BP (Anne Mackenzie, unpublished data), KYMV-JB (Ding *et al.*, 1990), KYMV-PD (Paul Keese, unpublished data), OYMV-Tin (Ding *et al.*, 1989), TYMV-BL (Drew Meek, unpublished data), TYMV-Cauliflower (Anne Mackenzie, unpublished data), TYMV-CL (Keese *et al.*, 1989), TYMV-Roth (Marjo Torronen, unpublished data), TYMV-type (Morch *et al.*, 1988) and WCuMV (Anne Mackenzie, unpublished data).

The virion proteins of all tymoviruses except that of ELV are of about 189 amino acid residues, whereas that of ELV is 202 residues. These sequences are closely similar in sequence and Figure 21 shows their alignment calculated by the progressive alignment method of Feng and Doolittle (1987). It can be seen that the greater length of ELV results from a stretch of 11 extra amino acid residues near the N-terminus found only in the ELV VP, and that the ELV sequence is least like that of the other VPs.

The variation at each position of the aligned proteins was also estimated (Figure 22). It can be seen that the sequences vary throughout their length and there are no particularly variable regions.

Pairwise comparisons of the 17 aligned sequences showed that the number of identical amino acid residues vary from 30.2% for the OYMV/TYMV-BL comparison to 98.9% for the comparison of TYMV-BL/TYMV-CL (Table 7).

The 'FJD distance' (Feng *et al.*, 1985) between each pair of sequences was used to calculate a dendrogram (Figure 25) by the neighbour-joining method (Saito and Nei, 1987; Studier and Keppler, 1988). In the dendrogram the VPs fall into three clusters

Figure 21. Alignment of amino acid sequences of seventeen tymoviral coat proteins by the progressive alignment method of Feng and Doolittle (1987) and PRETTY program.

```
                1                                                                80
ELV         21K MTSSRSTELIETQEDVVRKLSQKTPVSNFSALPLPNGPQAPTQLQPFQFEFPLPAGQEGSVTLPLATFPKMATFLSRHRR
TYMV-BL     20K MEIDKEL.............APQDRTVTVATVLPIAPGPSPFTIKQPFQSEVLFAGTKDAEASLTIANIDSVSTLTTFYRH
TYMV-CL     20K MEIDKEL.............APQDRTVTVATVLPTVPGPSPFTIKQPFQSEVLFAGTKDAEASLTIANIDSVSTLTTFYRH
TYMV-TYPE   20K MEIDKEL.............APQDRTVTVATVLPAVPGPSPLTIKQPEQSEVLFAGTKDAEASLTIANIDSVSTLTTFYRH
TYMV-CAUL   20K MDIDRQT.............APQDQVITIPSLLNPPPGSQPPIIRQPFQCEILNAGTKDATASITVASLANVSNLTALYRH
TYMV-RO     20K MDIDRQT.............APQDQVITIPSLLNPPPGSQPPIIRQPFQCRDLNAGTKDATASITVASLANVSNLTALYRH
KYMV-BP     20K MATNHVL.............ASQLPVNTKSSEIPLQSGTPPPTIVYPFQIELASLGTADASDAISIASNSLLASVTTLYRH
KYMV-JB     20K MATNHVL.............ASQLPVNTKSSEIPLQSGTPPPTIIYPFQIEMASLGTADASDSISIASNSLLASVTTLYRH
KYMV-PD     20K MASNHIL.............ATQTPVNTKSSEIPLEPGKTPPTITYPFQIELASLGTADASDAISIASNSSLASVTTLYRH
CoYMV       20K MSSDLIL.............APQPLINTKASELPSQSGSPPPEIVYPFQFTIASLGVEPTADFVSIAAQAAITAYTSLYRH
CYVV        20K MSTDVIV.............SAQPLINTKASEIPLQPGSPPPSITYPFQITIASLGVAATSDAVSISAQSSLASFTSLYRH
BMV         20K MDESKIV.............TVKQPSISAPGFTLSAPDGEQAGSIRQIFQFEATSVGVYETLAQVNLSSSDSLAKLTSGYRR
DMV         20K MDESKIV.............TVKQPSISAPGFTLSAPEGEQAGSIRQIFQFEATSVGVVETLAQVSLSASESLAKLTAGYRR
APLV        20K MEDQTPA.............VSKQPSINAPGYNLPPPSSQLSSSFELPFQFQATTFGAAETAAQVSLSSSTVISTIAKNFRH
EMV-Trin    20K MEDTAII.............RSPQPSINAPGFHLPPTDSQQSSAIELPFQFQATTFGATETAAQISLASANAITKLASLYRH
OYMV-Tin    20K MEDSQPI.............KVRQPSISAPGTHLSPNPGQQSPSMVVPFQVSVSDLGVSEVSAQITLSSDPTLAQLTSIYRM
WCMV        20K MTTDTSV.............SAKQPSINAPGHVLSIPDTPRSPSFPFKFQIHVANFGPKEVSSQISLSSCPELLRLTSLFRH
Consensus       M--------------------Q-----------------------FQ----------------------------R-

                81                                                              160
ELV         21K AQLTQLHAVVSPSAVSIGHPLTVQLIWVPASSTTTSSQILGTYGGQQISVGGQVTNSSPAKVSANLLMMNPHIKDSTSYT
TYMV-BL     20K ASLESLWVTIHPTLQAPAFPTTVGVCWVPANSPVTPTQITKTYGGQIFCIGGAINTLSPLIVKCPLEMMNPRVKDSIQYL
TYMV-CL     20K ASLESLWVTIHPTLQAPAFPTTVGVCWVPANSPVTPTQITKTYGGQIFCIGGAINTLSPLIVKCPLEMMNPRVKDSIQYL
TYMV_TYPE   20K ASLESLWVTIHPTLQAPAFPTTVGVCWVPANSPVTPAQITKTYGGQIFCIGGAINTLSPLIVKCPLEMMNPRVKDSIQYL
TYMV-CAUL   20K ASLDELWLTISPTHLAPAFPTQVRVCWVSANSPITDLQITNTYGGQVFCVGGSLNSHSPLEVKCPLHMMNRRVKDSVQYL
TYMV-RO     20K ASLDELWLTISPTHLAPAFPTQVRVCWVSANSPITDLQITNTYGGQIFCVGGSLNSHSPLEVKCPLHMMNRRVKDSVQYL
KYMV-BP     20K AKLTSLKATIHPTGQAPAYPTTVALAWVPYNSTATSAQILSVYGGQMFCIGGSINSLSPIDVPCNLTNVNPVIKDSVTYS
KYMV-JB     20K AKLTSLKATIHPTGQAPAFPTTVALAWVPYNSTATSAQILSVYGGQMFCIGGSINSLSPIDVPCNLTNVNPIIKDSVTYS
KYMV-PD     20K AKLTSLKATIHPTGAAPAYPTTVALAWVPYNSSATAAEVLSVFGGQMFCIGGSINSLSPIDIPCNLTNVNPIIKDSVTYS
CoYMV       20K AILTDLQAIIHPNGYAPAFPTSVALAWVPYNSTATAAKILDVFGGQEFCVGGSINSTSPIIVPCPLTNINPIIKDSVTYT
CYVV        20K AILQDLHATIHPSASAPAFPTSVALAWVPYNSTATASEILNVYGGQEFCIGGAINSSKPIIIPCPLTNINPVIKDSVTYL
BMV         20K AKVVELFLTITPTQRAIDCPVTVDVAWVPANSTAPPSKILSIYGGQRFLIGGPITTSQIIRVPCNLQSVNAMIKDSTIYT
DMV         20K AKLVELFLTITPTQLAIDNPVTVDVVWVPANSTATPSKILSVYGGQRFLIGGTLTTSQVIRVPCNLQSVNAMIKDSTIYT
APLV        20K AKLIQCHAIITPTYLAIANPITINLVWVPDSSTAKPSEILNVYGGTSFTFGGAFCSTKSIIVPLPMNSVNLMLKDSALYT
EMV-Trin    20K VRLTQCAATITPTAAAIANPLTVNIVWVSDNSTAKPTEILNVFGGSSYTFGGALNATKPLTIPLPMNSVNCMLKDSVLYT
OYMV-Tin    20K ASIVECEAVLFPNSTSSKNPVHCDLIWVPSNSSASPKTILQTYGGNRFTVGGPITSNQIISFPLRLDSVNPIIKDSVLYL
WCMV        20K ARLLSAKAIITPFDGVVSLPITVDLAWVSANSPASPTDILKIYGGSSYTFGGAIMSTRPIELPLPINSVNDMLKDSVSYL
Consensus       ----------P-------P------WV---S----------GG-----GG----------------N---KDS--Y-

                161                                             205
ELV         21K DTPKLLVYSTPAVPDDKLTTSSASIIVFGEVLLSSPQLNPSA...
TYMV-BL     20K DSPKLLISITA..QPTAPPASTCIITVSGTLSMHSPLITDTST..
TYMV-CL     20K DSPKLLISITA..QPTAPPASTCIITVSGTLSMHSPLITDTST..
TYMV_TYPE   20K DSPKLLISITA..QPTAPPASTCIITVSGTLSMHSPLITDTST..
TYMV-CAUL   20K DSPKLVLSVTA..QPTPPTASTCIITVSGILSMHSPLITDTST..
TYMV-RO     20K DSPKLVLSVTA..QPTPPTASTCIITVSGILSMHSPLITDTST..
KYMV-BP     20K DTPKLLLYSIA..QTTAPTLATCSVTITGTLTLHSPLLQATA...
KYMV-JB     20K DTPKLLLYSIA..QETAPTLATCSVTITGTLTLHSPLLQATA...
KYMV-PD     20K DTPKLLMYSTA..QSTAPTTATCSLTISGTLTLHSPLLQASS...
CoYMV       20K DTPKLLIYSTA...PSYSTSATCTLTIRGKVRLHSPLLSSSSS..
CYVV        20K DTPKLLVYSTA...PAYSTSPTCTLTITGKVKLHSPLLSSSS...
BMV         20K DSPKLLVYAPI..AKGSPKTPSATVQIGGQILLSAPLLQAL....
DMV         20K DSPKLLVYSPV..AKGSPKTPSATVQIAGQILLSAPLLQAL....
APLV        20K DGPKLLAYSPA..PATPSKSPVATIQISGKILLSSPLLQAS....
EMV-Trin    20K DCPKLLAYSAA..PSSPSKTPTATIQIHGKLRLSSPLLQAN....
OYMV-Tin    20K DSPRLLAFSPA..PPETQSIPSASLLIRGKLRLSSILVQPLLTSS
WCMV        20K DTPKLLVFSPA..PAKATSVTLASPQISGEVLCSSQLLQAL....
Consensus       D-P-L----------------------G---------------
```

```
* Progressive alignment: Feng and Doolittle (1987)
  Gap penalty = 8
  Length of sequences with gaps = 205
  Actual sequence lengths: 189, 189, 189, 189, 189, 188, 188, 188, 188, 187, 188,
                           188, 188, 188, 192, 188, 202,
```

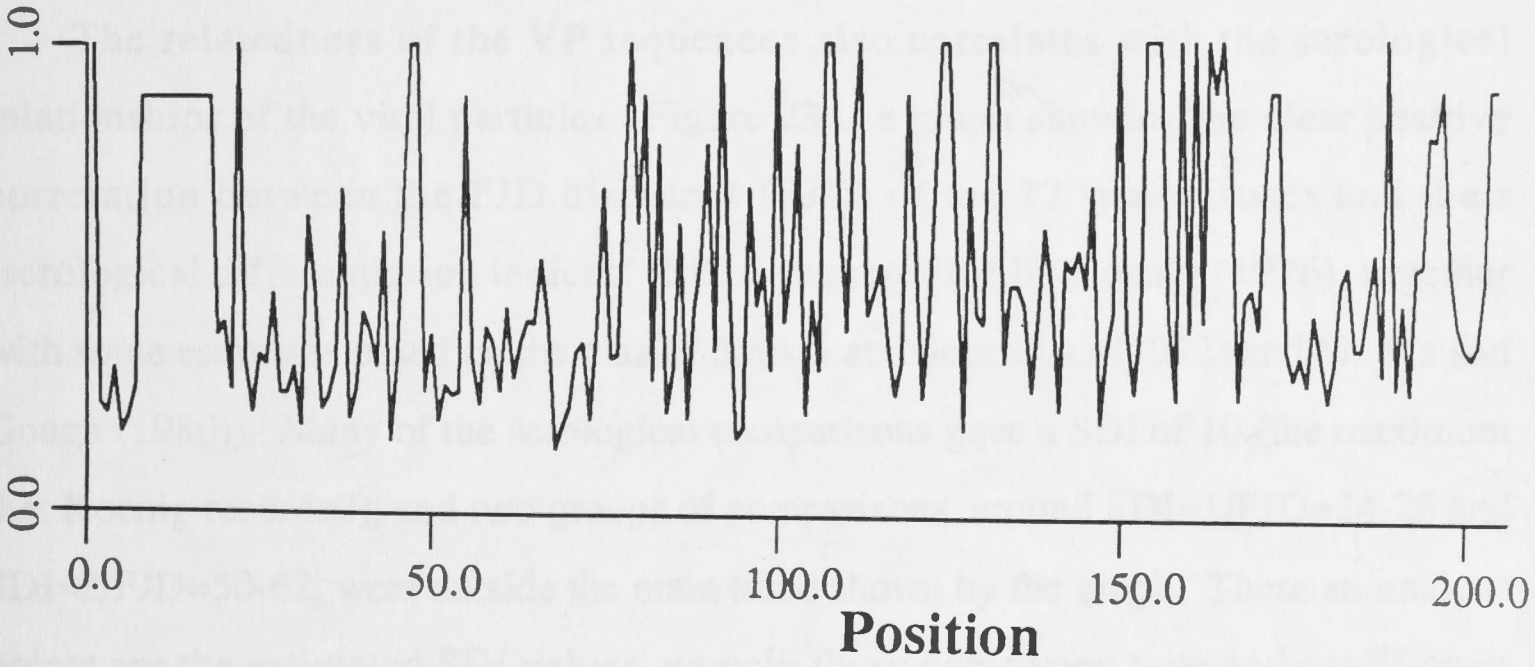
Table 8. Percent identity (based on aligned regions) of seventeen tymoviral coat proteins.

	TYBL	TYCL	TYTY	TYCA	TYRO	KYBP	KYJB	KYPD	COYV	CYVV	BMV	DMV	APLV	EMV	OYMV	WCuMV
TYCL	98.94															
TYTY	97.88	98.41														
TYCA	67.72	67.72	67.72													
TYRO	67.72	67.72	67.72	98.41												
KYBP	50.00	50.00	50.53	46.81	46.28											
KYJB	51.06	51.06	51.60	47.87	47.34	96.81										
KYPD	49.47	49.47	50.00	45.74	45.21	86.17	85.11									
COYV	44.15	44.15	44.68	44.15	43.62	62.57	63.10	63.64								
CYVV	43.85	43.85	43.85	42.25	41.71	64.71	64.17	66.31	76.47							
BMV	34.76	34.76	35.29	33.16	32.62	40.11	40.11	39.04	37.63	41.94						
DMV	34.76	34.76	35.29	35.29	34.76	41.18	41.18	40.11	39.25	43.55	90.43					
APLV	30.48	30.48	30.48	29.41	29.41	36.90	36.90	35.83	41.40	43.01	47.34	50.53				
EMV	32.62	32.62	32.09	34.22	34.22	37.97	37.97	40.64	44.62	41.94	45.21	47.34	68.09			
OYMV	30.16	30.16	30.16	30.69	30.69	34.57	34.57	35.11	36.17	37.43	46.81	45.21	42.55	42.55		
WCuMV	32.62	32.09	31.55	31.02	31.02	39.04	39.04	37.97	36.02	39.78	43.09	41.49	50.00	47.87	45.21	
ELV	33.51	33.51	33.51	32.98	32.45	37.77	37.77	35.64	35.83	36.90	37.23	38.30	32.98	31.38	32.80	31.91

Distance scores for phylogenetic tree

TYCL	0.70															
TYTY	1.46	0.86														
TYCA	25.23	25.10	24.42													
TYRO	26.37	26.41	25.77	1.31												
KYBP	53.25	54.39	53.28	59.92	62.29											
KYJB	52.72	53.87	52.77	59.83	61.69	1.09										
KYPD	51.14	52.07	50.74	57.87	60.18	7.53	8.34									
COYV	62.90	62.41	62.26	67.16	69.25	32.71	32.55	31.09								
CYVV	57.35	58.38	57.67	63.19	65.18	27.53	27.65	24.52	14.55							
BMV	87.41	87.50	86.95	86.97	89.31	70.40	70.84	69.43	78.74	73.32						
DMV	85.68	85.74	85.60	81.78	82.96	67.71	68.51	66.83	75.39	69.84	6.25					
APLV	101.23	100.21	100.70	101.15	102.35	77.60	79.21	76.98	76.63	73.43	52.53	46.40				
EMV	90.18	89.00	89.58	90.83	89.43	70.60	71.84	66.01	67.69	67.01	52.27	47.22	27.97			
WCuMV	94.82	97.83	98.38	99.45	100.97	72.98	74.31	73.91	85.14	73.94	67.73	65.66	52.02	54.16		
OYMV	102.21	100.07	101.45	99.48	101.51	78.63	79.11	76.03	77.92	72.20	63.09	62.22	60.96	55.60	65.90	
ELV	99.26	102.00	102.04	101.87	105.36	86.09	87.62	86.43	90.23	90.11	88.33	82.07	95.08	99.56	105.49	102.27

Figure 22. Graph showing the variability in each position of the aligned amino acid sequences of the virion proteins of five tymoviruses; ELV, EMV-Trin, KYMV-JB, OYMV-Tin and TYMV-CL.


Variability
1.0
0.0
0.0
50.0
100.0
150.0
200.0
Position

with the ELV VP as an 'out-group'. The clusters mostly correlate with the host range preferences of the viruses; the VPs of the brassica-infecting tymoviruses form one cluster, all those of legume-infecting tymoviruses, except OYMV, form a second cluster together with CoYMV, and the third cluster is of the VPs of tymoviruses that infect solanaceous plants together with OYMV and WCuMV. There is an indication that the atypical VPs of the second and third clusters (i.e. CoYMV in the second, and WCuMV and OYMV in the third) have perhaps changed during evolution more than the other VPs of each cluster (i.e. they are on longer branches of the dendrogram). This might indicate that these two clusters have primarily co-evolved with leguminous and solanaceous hosts, respectively, but that these clusters have been the sources of new viral species that have adapted to plant species of other families and, as a result their VPs have undergone increased evolutionary change.

The relatedness of the VP sequences also correlates with the serological relationships of the viral particles. Figure 23 is a graph showing the clear positive correlation between the FJD distances (FJD) of the 17 tymoviruses and their 'serological differentiation indices' (SDIs) (as reported by Koenig (1976), together with some estimates based on the data of Shukla and Schmelzer (1972) and Shukla and Gough (1980)). Many of the serological comparisons gave a SDI of 10 (the maximum that Koenig recorded), and two groups of comparisons, around SDI=1/FJD=24-26 and SDI=2/FJD=50-62, were outside the main trend shown by the graph. These anomalous points are the estimated SDI values, namely those comparing type and cauliflower strains of TYMV and those comparing TYMV and KYMV isolates, respectively. There was a very significant statistical correlation between the SDI and FJD values (correlation coefficient = 0.853; $p < 0.001$), despite the predictable increase in variability at larger values of SDI and FJD; the comparisons involving estimated SDI values were omitted from this calculation.

Sequence of the 3' non-coding region

The 3' non-coding regions have an above average guanine content (18.9-25.2%) and their cytosine content is similar to that of the remainder of the genome (Table 4). This composition probably reflects the fact that the 3' regions form stable stem-loop structures.

Figure 23. Graph showing the correlation between the 'FJD distances' of the seventeen tymovirus virion proteins and their serological differentiation indices (SDIs).

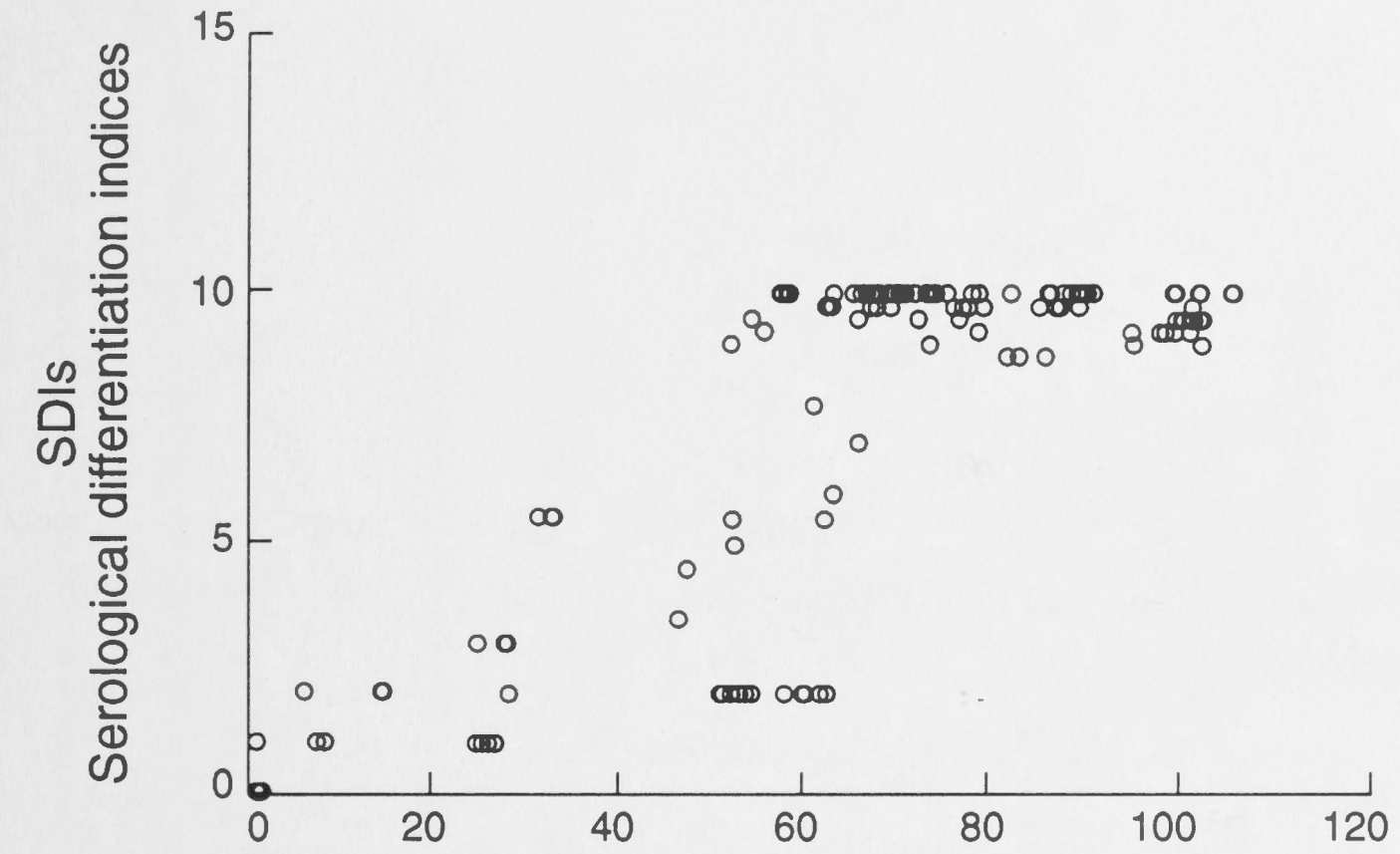
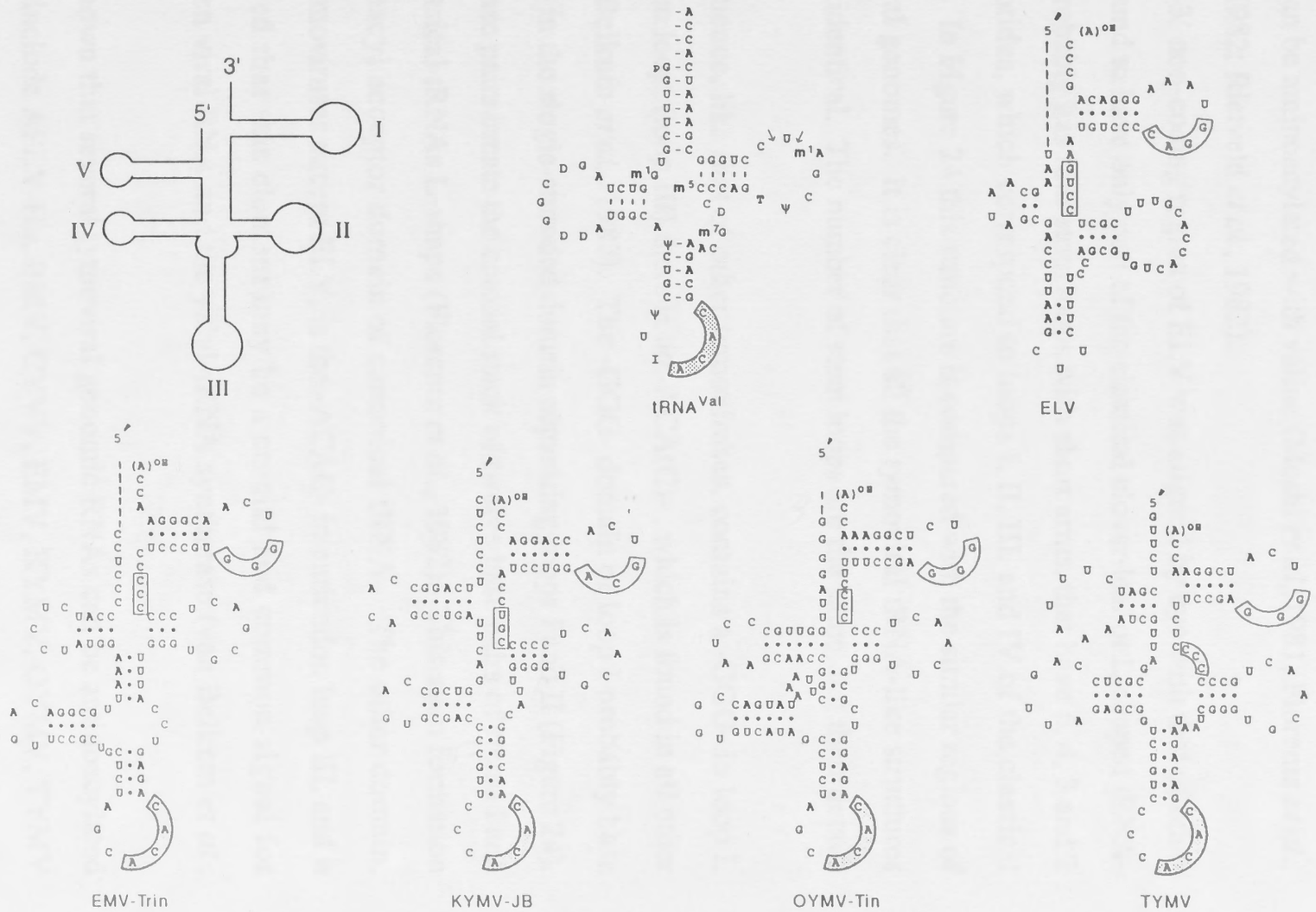

SDIs
Serological differentiation indices
15
10
5
0
0
20
40
60
80
100
120
FJD distance

Figure 24. Sequences of the 3'-termini of five tymoviral genomes and their potential tRNA-like secondary structures together with likely structure of yeast $tRNA^{Val}$ (Bonnet *et al.*, 1974) are presented.

Sequences of the non-coding 3'-termini of six tymoviral RNAs.

```
DMV                                                UGAUUUCUUCACACUUCAGAGCCCUGUCAAUAAACCAAAUUAAAAAAAAAAAAAACAAAAAAAAG poly(A)
ELV                                 UAAGCAAUAGCGACCUUAAG--CUUCUUUC--UCCAGCGUGUCACCACGUUUCGCUCCUGAAUGUCCCCAGGUAAAGGGACACCCC(A)
EMV-Trin CCCCCAUCUCCUAUGGUAAUUGCGGACAGU--UCCGCUCCCUCUAGCACACAGAGGUCCAUUUGGGUGCGACUCCCCCCCC-----UCCCGUGGGUC-AACGGGAACC(A)
KYMV-JB  UCAGGCCACAGCCUGAUUCAGGUCGCCAGU--GCGACACCCGUUCCCACACAACGGG-CAUUGGGGUGCAACUCCCCCGUCCA---UCCUGGACGUC--ACCAGGACC(A)
OYMV-Tin CCUAAGCAACCCGCUUAAAUAUGACCGGU--GUCAUAGCCUCUAGCACACAGAGGUCGCG-UGGGUUCAACUCCCCCCCU----UUUCCGAGGGUA--UCGGAAACC(A)
TYMV-CL   CGAUCUUUAAAAUCGUUAGCUCGCCAGUUAGCGAGGUCUGUCCCCACACGACAGA-UAAUCGGGUGCAACUCCCGCCCCU---CUUCCGAGGGUC--AUCGGAACC(A)
           V         V        IV        IV     III         III       II         II            I            I
```

Secondary structures of non-coding 3'termini of five tymoviruses.



The 3' non-coding region of ELV is 78 nucleotides in length (nucleotides 5956-6034). The SQUIGGLES program failed to predict a secondary structure for the sequence, that resembled the tRNA-like structures found at the 3'-termini of other tymoviral genomes. These structures resemble that of the $tRNA^{Val}$ from yeast (Giége *et al.*, 1978), and can be aminoacylated with valine (Meshi *et al.*,1981; Florentz *et al.*, 1982; Joshi *et al.*, 1982; Rietveld *et al.*, 1982).

Therefore the 3' non-coding region of ELV was aligned by eye with that of other tymoviruses and found to have only part of the classical clover-leaf or L-shaped tRNA-like structure. It probably has four stem loops with short arms, that have 6, 4, 3 and 2 base-paired nucleotides, which correspond to loops I, II, III, and IV of the classical clover-leaf pattern. In Figure 24 this structure is compared with the similar regions of four other tymoviral genomes. It is clear that all the tymoviral tRNA-like structures are similar but not identical. The number of stem loops are the same, but their details differ.

The ELV sequence, like that of other tymoviruses, contains a -GGG- in loop I, but, in the anticodon loop, (loop III) there is no -ACA(C)- , which is found in all other tymoviruses (van Belkum *et al.*, 1987). The -GGG- domain of loop I probably base pairs with a -CCC- in the single-stranded domain separating loops I and II (Figure 24). These three G-C base pairs create the coaxial stack of twelve base pairs of stems I and II found in the classical tRNAs L-shape (Florentz *et al.*, 1982). This stem formation defines the aminoacyl acceptor domain of canonical tRNA. The other domain, conserved in all tymoviruses except ELV, is the -ACAC- in anticodon loop III, and it has been suggested that this element may be a crucial and common signal for interaction between viral RNA and the valyl-tRNA synthetase (van Belkum *et al.*, 1987).

It has been shown that several tymoviral genomic RNAs can be aminoacylated with valine, these include APLV-Hu, BMV, CYVV, EMV, KYMV, OYMV, TYMV and WCuMV (Rietveld *et al.* (1982). The site of aminoacylation is probably the 3' tRNA-like terminus, and sequences of this sort have been found in the 3' terminal genomic sequences of APLV (Maria Osorio-Keese, unpublished data), BMV (Ding *et al.*, 1990b), CoYMV (Ding *et al.*, 1990c), CYVV (Anne Mackenzie, unpublished data), EMV-Trin (Osorio-Keese *et al.*, 1989), KYMV-BP (Anne Mackenzie, unpublished data), KYMV-JB (Ding *et al.*, 1990), KYMV-PD (Paul Keese,

unpublished data), OYMV-Trin (Ding *et al.*, 1989), TYMV-BL (Drew Meek, unpublished data), TYMV-CL (Keese *et al.*, 1989), TYMV-type (Morch *et al.*, 1988), TYMV-cauliflower (Anne Mackenzie, unpublished data), TYMV-Roth (Marjo Torronen, unpublished data), and WCuMV (Anne Mackenzie unpublished data). This observation suggests that some tymovirus genomes may not form tRNA-like structures, and confirms the suggestion of van Belkum *et al.* (1987) that ELV possibly lacks a tRNA-like structure. In addition it has been found (Jennifer Gibbs; unpublished data) that the dulcamara mottle tymovirus genome possesses a poly(A) tail and the associated signal sequence (-AAUAAA-) at its 3' terminus.

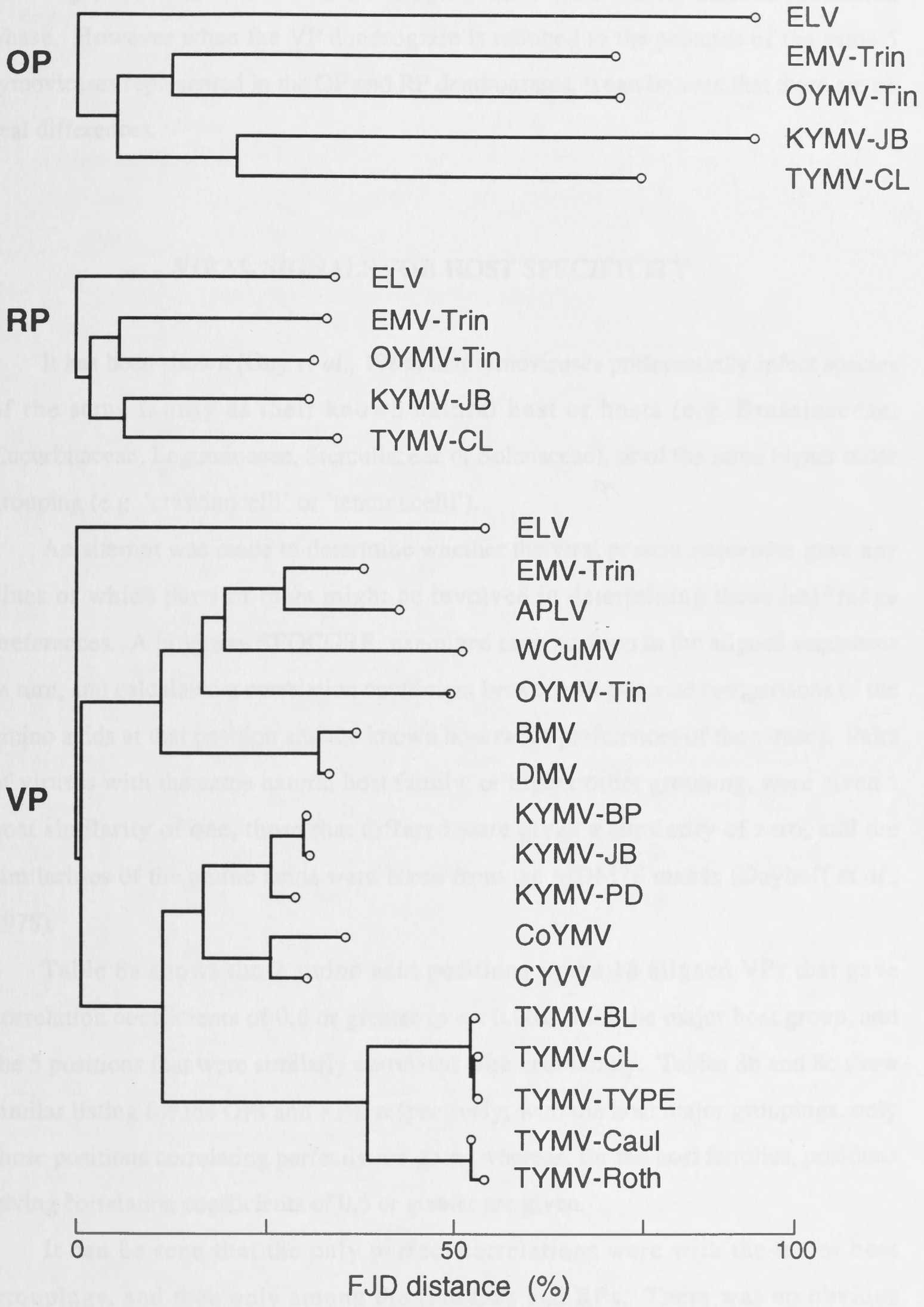
TYMOVIRUS EVOLUTION

The comparisons of various features of tymoviral genomes provide clues on the evolution of this group of viruses. All tymoviruses share characters such as the size and basic plan of their genome, its composition and the number, genomic arrangement and sequence of the proteins it encodes. It is likely that these characters were inherited from the progenitor of the group.

There are other characters that vary between different species of the group. These include details of the sequences in different regions of the genome and the length of the intergenic regions. These differences have probably arisen since the primary radiation of the group, either by random evolutionary changes or as a result of adaptation. Surprisingly there are also major differences in the 3'-terminal regions, which in most tymoviruses are able to form a tRNA-like structure that can be amino-acylated, but which is only partially present in ELV, and which, in DMV, has apparently been replaced by a poly(A) tail like that found in, for example, comoviruses of plants and picornaviruses of animals.

The relationships of the different conserved parts of tymovirus genomes (i.e. the OP, RP-A, RP-B, RP-C and VP) are all closely similar. Figure 25 shows dendrograms representing the relationships of the amino acid sequences of the known tymoviral OPs, RPs and VPs. It can be seen that the topologies of these dendrograms are closely similar, with ELV as the most distinct species of the group. The dendrograms differ in total branch lengths reflecting the fact that the OPs have changed most and the RPs least during evolution of the group. Superficially, the dendrograms seem also to differ

Figure 25. Dendrogram showing the hierarchical clustering of five tymoviruses based on their amino acid sequence similarity in three different proteins (OP, RP and VP); the sequences were aligned by the progressive alignment program of Feng and Doolittle (1987) and the FJD distances (Feng *et al.*, 1985) used to calculate a dendrogram by the neighbour-joining program of Saitou and Nei (1987).


OP
ELV
EMV-Trin
OYMV-Tin
KYMV-JB
TYMV-CL
RP
ELV
EMV-Trin
OYMV-Tin
KYMV-JB
TYMV-CL
VP
ELV
EMV-Trin
APLV
WCuMV
OYMV-Tin
BMV
DMV
KYMV-BP
KYMV-JB
KYMV-PD
CoYMV
CYVV
TYMV-BL
TYMV-CL
TYMV-TYPE
TYMV-Caul
TYMV-Roth
0
50
100
FJD distance (%)

in the relative lengths of the 'speciation phase' and of the branches leading to individual species; in the OP and RP dendrograms the former covers one third of the dendrogram, whereas in the VP dendrogram there is no clearly defined speciation phase. However when the VP dendrogram is reduced to the proteins of the same 5 tymoviruses represented in the OP and RP dendrograms, it can be seen that there are no real differences.

VIRAL SIGNALS FOR HOST SPECIFICITY

It has been shown (Guy *et al.*, 1984) that tymoviruses preferentially infect species of the same family as their known natural host or hosts (e.g. Brassicaceae, Cucurbitaceae, Leguminosae, Sterculiaceae or Solanaceae), or of the same higher order grouping (e.g. 'crassinucelli' or 'tenuinucelli').

An attempt was made to determine whether the viral protein sequences gave any clues of which parts of them might be involved in determining those host range preferences. A program, SEQCORR, examined each position in the aligned sequences in turn, and calculated a correlation coefficient between all pairwise comparisons of the amino acids at that position and the known host range preferences of the viruses. Pairs of viruses with the same natural host family, or higher order grouping, were given a host similarity of one, those that differed were given a similarity of zero, and the similarities of the amino acids were taken from the MDM78 matrix (Dayhoff *et al.*, 1978).

Table 8a shows the 7 amino acid positions in the 18 aligned VPs that gave correlation coefficients of 0.6 or greater ($p << 0.001$) with the major host group, and the 5 positions that were similarly correlated with host family. Tables 8b and 8c show similar listing for the OPs and RPs, respectively; with the host major groupings, only those positions correlating perfectly are given whereas, for the host families, positions giving correlation coefficients of 0.6 or greater are given.

It can be seen that the only perfect correlations were with the major host groupings, and then only among the five OPs and RPs. There was no obvious grouping of the positions that showed significant similarities, and the number of positions that correlated was broadly related to the size and conservation of the

Table 8. Correlations between the host species taxonomy and amino acids at individual positions in the aligned proteins of tymoviruses.

A) Virion Proteins.

	Host	Amino acid position							Host	Amino acid position				
	Crass/Ten[1]	30	60	77	99	129	158	179	Family	26	68	148	149	181
APLV	**T**	Y	A	N	N	T	L	K	**S**	N	S	S	V	P
BMV	**T**	F	L	G	C	L	I	K	**S**	S	S	S	V	P
CoYMV	**C**	S	A	L	F	C	T	T	**St**	N	Q	N	I	A
CYVV	**C**	S	S	L	F	C	T	T	**L**	N	Q	N	I	P
DMV	**T**	F	L	G	N	L	I	K	**S**	S	S	S	V	P
ELV	**C**	S	S	R	H	S	S	T	**B**	V	F	M	M	S
EMV	**T**	F	A	L	N	T	L	K	**S**	N	A	S	V	P
KYMV-BP	**C**	S	S	L	Y	C	T	T	**L**	N	N	N	V	A
KYMV-JB	**C**	S	S	L	F	C	T	T	**L**	N	N	N	V	A
KYMV-PD	**C**	S	S	L	Y	C	T	T	**L**	N	N	N	V	A
OYMV-Tin	**C**	T	S	I	N	T	L	S	**L**	S	D	S	V	P
TYMV-BL	**C**	T	E	F	F	C	Q	P	**B**	V	I	M	M	S
TYMV-Caul	**C**	S	T	L	F	C	Q	T	**B**	I	L	M	M	S
TYMV-CL	**C**	T	E	F	F	C	Q	P	**B**	V	I	M	M	S
TYMV-Ro	**C**	S	T	L	F	C	Q	T	**B**	I	L	M	M	S
TYMV-TYPE	**C**	T	E	F	F	C	Q	P	**B**	V	I	M	M	S
WCuMV	**C**	H	S	L	L	T	S	S	**Cu**	N	C	S	V	T

1 Only those positions that gave correlations coefficients ≥ 0.6 ($P<0.001$) are listed. For details see text.

B) Overlapping Proteins.

Host		Amino acid position												
Crass/Ten[2]		8	41	81	83	119	204	233	307	341	354	376	378	388
ELV	**C**	S	P	L	P	P	R	P	P	S	R	N	R	R
EMV-Trin	**T**	C	S	V	S	S	G	F	S	R	C	D	V	S
KYMV-JB	**C**	S	P	L	P	P	R	P	P	S	R	N	R	R
OYMV-Tin	**C**	S	P	L	P	P	R	P	P	S	R	N	R	R
TYMV-CL	**C**	S	P	L	P	P	R	P	P	S	R	N	R	R

Family	35	39	53	172	180	326	335	342	389	444
ELV **B**	I	I	Q	P	N	S	P	N	R	H
EMV-Trin **S**	R	P	S	S	F	A	S	S	S	V
KYMV-JB **L**	G	N	S	S	L	P	S	R	R	L
OYMV-Tin **L**	N	K	P	S	L	P	S	R	R	L
TYMV-CL **B**	L	D	Q	P	S	T	P	S	R	H

2 Only those position that gave perfect correlations are shown.

C) Replicase Proteins.

Host		Amino acid position																					
Crass/Ten[2]		5	11	36	41	116	123	133	139	171	187	203	206	375	376	377	408	412	585	677	835	845	1076
ELV	C	L	S	E	V	Y	S	L	F	V	L	G	A	A	L	L	L	A	L	P	P	A	L
EMV-Trin	T	S	N	S	L	F	T	F	C	I	I	N	S	S	-	V	Y	L	I	L	S	C	P
KYMV-JB	C	L	A	Q	V	Y	S	L	V	V	L	G	A	A	L	L	L	A	L	P	P	A	L
OYMV-Tin	C	L	A	Q	I	Y	S	L	V	V	L	G	A	A	L	L	L	A	L	P	P	A	L
TYMV-CL	C	L	A	Q	I	Y	S	L	I	V	L	G	A	A	L	L	L	A	L	P	P	A	L

	1154	1166	1181	1208	1228	1229	1280	1465	1525	1554	1582	1595	1687	1688	1732	1776	1817	1824	1836
ELV	L	Y	F	S	S	Q	S	A	N	S	G	D	T	N	V	G	L	L	C
EMV-Trin	F	H	N	A	R	R	N	P	Q	Q	S	G	C	S	I	S	A	I	V
KYMV-JB	L	Y	F	S	S	Q	S	A	N	S	G	D	T	N	V	G	L	L	C
OYMV-Tin	L	Y	F	S	S	Q	S	A	N	S	G	D	T	N	V	G	L	L	C
TYMV-CL	L	Y	F	S	S	Q	S	A	N	S	G	D	T	N	V	G	L	L	C

Family[1]	51	60	119	158	391	402	410	440	498	505	579	662	670	795	809	847	863	883	959	962	1165	1171
ELV	B	N	L	H	E	R	T	A	A	K	A	L	I	P	E	D	S	T	E	L	N	S	C
EMV-Trin	S	L	L	T	T	S	A	L	S	P	H	Q	T	E	A	V	S	S	D	L	E	S	Q
KYMV-JB	L	L	F	T	H	L	S	V	V	C	P	S	L	A	M	S	A	E	D	Y	S	T	H
OYMV-Tin	L	L	F	T	I	L	S	I	I	C	P	S	L	T	L	T	A	D	D	Y	A	T	A
TYMV-CL	B	N	L	H	L	Q	R	A	A	I	S	F	V	P	T	E	S	T	E	V	N	S	C

	1174	1209	1237	1245	1297	1312	1331	1333	1409	1478	1481	1494	1509	1596	1632	1663	1820	1901
ELV	K	H	A	T	F	L	S	A	H	A	H	Q	S	Y	R	V	S	L
EMV-Trin	S	T	H	T	V	I	N	R	I	G	K	S	S	Y	S	V	L	P
KYMV-JB	D	V	V	S	S	I	H	P	F	F	P	V	N	F	S	I	A	P
OYMV-Tin	D	I	V	S	A	I	H	P	F	F	S	A	N	F	S	I	V	P
TYMV-CL	R	H	A	T	Y	L	S	T	A	A	N	Q	S	Y	R	V	S	I

different proteins (i.é. was greatest in the RPs). These results give no definite clues on which amino acids, if any, encode host preferences, but do indicate which residues might be targeted in an analysis of such preferences by site directed mutagenesis of cloned viral genomes.

CHAPTER 5 - RELATIONSHIPS OF TYMOVIRUSES AND OTHER VIRUSES

INTRODUCTION

The structure and organization of the particles, and the genomes, of representatives of about 20 plant viral groups have, so far, been studied. They differ in many significant ways. The particles are of various shapes, structures and sizes, and the genomes are similarly diverse, some are segmented, others unipartite, and their genes are expressed using various translation strategies.

Despite these differences, computer-assisted sequence comparisons have revealed the clear similarity of parts of the non-virion proteins of these genomes. Many of these proteins have at least one region with the characteristic motifs of nucleotide-binding proteins (Hodgman, 1988; Gorbalenya *et al.*, 1988b, Gorbalenya *et al.*, 1989a, b) and all have a region with the RNA polymerase motifs (Argos *et al.*,1984; Franssen *et al.*, 1984; Haseloff *et al.*,1984; Kamer and Argos, 1984; Ahlquist *et al.*,1985; Argos, 1988; Goldbach and Wellink, 1988; Morozov *et al.*, 1990; Skryabin *et al.*, 1988b; Gorbalenya *et al.*,1989a, b).

Comparisons of these sequences motifs place many plant viruses into two supergroups. These supergroups have been named the 'picorna-like' and 'Sindbis-like' viruses after the animal-infecting viruses they contain (Goldbach, 1986; 1987; Goldbach and Wellink, 1988). A taxonomic study by Poch *et al.*(1989) and Candresse *et al.* (1990) of these motifs confirmed these groupings and showed that the tymoviruses are members of the 'Sindbis-like' supergroup. Candresse *et al.* (1990) also showed that the 'Sindbis-like' viruses consistently formed two clusters; cluster I being the tymo-, carla-, clostero-, and potexviruses, and cluster II the alfalfa mosaic, alpha-, bromo-, cucumo-, furo-, hordei-, tobamo- and tobraviruses.

The 'Sindbis-like' supergroup is a very diverse group of plant and animal viruses (Figure 31), and its evolution has involved both point mutational changes, and also the recombination of viral genes. The importance of recombination in the evolution of RNA genomes has only recently been realized, although it was described for viruses with DNA genomes some time ago, and named modular evolution by Botstein (1980)

to explain the results of his studies of phages of gram-negative bacteria. One clear example of natural viral recombination in viruses within RNA genomes is that described by Robinson *et al.* (1987), who showed that some tobravirus isolates are recombinants of tobacco rattle virus (TRV) and pea early browning virus (PEBV).

This chapter reports analyses of the relationships of the 'Sindbis-like' viruses based on the amino acid sequences of their RPs, and then proceeds to determine whether the RP classification correlates with other characters of the viruses such as the type of structure at the 3'-terminus of their genomes, on their genomic signals for the initiation and termination of translation and mRNA production, and on their genome composition and organization. The initial detailed comparisons involved the five fully sequenced tymoviruses and the following 'Sindbis-like' viruses:

- alfalfa mosaic virus (AlMV; Cornelissen *et al.*, 1983);
- apple chlorotic leaf spot closterovirus (ACLSV; German *et al.*, 1990);
- beet necrotic yellow vein furovirus (BNYVV; Bouzoubaa *et al.*, 1987);
- narcissus mosaic potexvirus (NMV; Zuidema *et al.*, 1989);
- potato M carlavirus (PVM; Zavriev *et al.*, 1991);
- potato X potexvirus (PVX; Huisman *et al.*, 1988);
- Sindbis alphavirus (Sindbis; Strauss *et al.*, 1984);
- tobacco mosaic tobamovirus (TMV; Goelet *et al.*, 1982);
- white clover mosaic potexvirus (WClMV; Forster *et al.*, 1988).

VIRAL REPLICASE PROTEINS

The viral replicase sequences examined in this study were obtained from the Genbank, EMBL and NBRF databases. The proteins studied were the 216.5K protein of ACLSV, those encoded by the RNA1 and RNA2 of AlMV, the 237K protein of BNYVV, the 165K proteins of potexviruses, the 223K protein of PVM, the nsp1-nsp2-nsp4 proteins of Sindbis virus, the 126K protein of tobacco mosaic virus and the five known 206K tymoviral RPs. For ease of analysis, the RP sequences were divided into three parts chosen to coincide with the three segments of the tymoviral replicases discussed in Chapter 3. First the tymovirus sequence segments were prealigned with the other 'Sindbis-like' RPs using the GAP program of the GCG package (Devereux *et al.*, 1984). Each of the three sets of replicase segments was then aligned definitively

using the 'progressive alignment' method of Feng and Doolittle (1987). The resulting alignments were arranged for display using the GCG LINEUP and PRETTY programs, and the latter gave the consensus sequence. The resulting segment alignments were named the A segment, which spans the N-terminal third of the RP, the B segment, which is from the end of the A segment to the start of the C-terminal quarter, and the C segment, which is the C-terminal quarter. The A segment in Sindbis alphavirus has RNA methyltransferase activity (Mi *et al.*, 1989), the B segment contains the motifs of a NTP-binding protein, and the C segment the motifs of a RNA-dependent RNA polymerase.

The C-terminal two-thirds of the RPs of the 'Sindbis-like' viruses are clearly homologous (Figure 26), especially around the regions which include motifs of the nucleotide-binding protein (Hodgman, 1988; Gorbalenya *et al.*, 1988b) and the RNA-dependent RNA polymerase (Kamer and Argos, 1984; Argos, 1988). The N-terminal regions are less similar and although the tymo-, potex-, clostero- and carlavirus sequences can be readily aligned, those of the other 'Sindbis-like' viruses show a much smaller amount of homology in this region (Rozanov *et al.*, 1990; Morozov *et al.*, 1990; Goldbach and Wellink, 1988). Nonetheless some recognisable elements of the methyltransferase found in Sindbis alphavirus (Mi *et al.*, 1989) are present in the others.

Quantitative estimates of the similarity of the aligned RPs and their A, B and C segments are given in Table 9, which gives the percentage of identical residues in pairwise comparisons of the different proteins or segments (Feng and Doolittle, 1987). The greatest identity is, as expected, between different species of the same virus group; 48% (EMV-Trin/TYMV-CL) to 54% (EMV-Trin/OYMV-Tin or TYMV-CL/KYMV-JB) in tymoviruses, and between 44% (NMV/PVX) to 53% (NMV/WClMV) in potexviruses, compared with only 10-23% of identical residues between the tymoviral RPs and those of Sindbis-like viruses. Comparisons of the tymoviral sequences with those of the other 'Sindbis-like' viruses showed that the most conserved regions of the RPs are those around the RNA polymerase motifs (range 12-22% of identical residues), and the least conserved around the possible methyltransferase motifs (8-19%). Furthermore, in tymoviruses and potexviruses, the methyltransferase and nucleotide-binding regions are separated by non-homologous sequences of different lengths.

Figure 26. Alignment of the encoded replicase proteins of five tymoviruses and other Sindbis-like viruses (virus names as in the text and Table 9).

```
                1                                                                80
194K ELV        .....................MA..FQLALD............ALS...STTHRDSISAPLLDSSVSQLQSSLELFPYTV
205K EMV        .....................MA..FQSALE............ALN...STTHRDASTNPILNSVVEPLRDSLSLYPWLL
210K KYMV       .....................MA..FQLALD............ALA...STSHKDPSLHPVLESVHDSLTDSLQTYPWMV
198K OYMV       .....................MA..FELALN............ALA...SSTHKDSSLNPVLNSAVQPLQTSLQNFPWII
206K TYMV       .....................MA..FQLALD............ALA...PTTHRDPSLHPILESTVDSIRSSIQTYPWSI
186K NMV        .....................MA.KVRAALE............RIR...DPSVQTALSEAAYTHVRPVLKESLVNCPYAL
166K PVX        .....................MA.KVREVYQ............SFT...DSTTKTLIQDEAYRNIRPIMEKHKLANPYAQ
147K WClMV      .....................MA.KVRAALD............RIT...DPSVKAVLNEEAYSHIRPVLRESLTNNPYAI
216K ACLSV      .....................MAFSYRTPQE............ELLSRLPQSQQEVISGFQYERIQKEEEKKVENFSFYL
223K PVM        .....................MAVTYRTPME............DIVNCFEPATQAVIANSAATLYKNFEEQHCQYFNYYL
1a   AlMV*      .......MNADAQST.....DASLSMREPLSHA..........SIQEMLRRVVEKQAADD.TTAIGKVFSEAGRAYAQDA
237K BNYVV      ...MADSFGFTPMEVLLFGGESVQLLTSDMPIDVQWGFVYSTRCYALWKDDLIHLNPLLKYSQRIAKRWEKLVSGFVGPV
NSP SINDBIS*    MEKPVVNVDVDPQSPFVVQLQKSFPQFEVVAQQVTPNDHANARAFSHLASKLIELEVPTT.ATILDIGSAPARRMFSEHQ
126K TMV        .....................MAYTQTATTS..............ALLDTVRGNNSLVN.DLAKRRLYDTAVEEFN...
  Consensus     ---------------------MA--F---L--------------L-------------------I---L------FPY--

                81                                                               160
194K ELV        P........KELVPQLNRMGIQVSGLTSTP...HPHAAHKTLELNLLFNHWAKSCNVDSAV..VFMKPSKFFKLQEKNS.
205K EMV        P........KEAVPHLLSWGIPNSGLGVTP...HPHPIHKTVETFLLFNHWHALARLPSTV..MFMKPSKFQKLAALNP.
210K KYMV       P........QDLQPFLLKSGIPINSFGSSP...HPHPAHKTLETHLLFTHWMHLCTQPSSV..LFMKPQKFMKLQRKNK.
198K OYMV       G........KEHLPFLLAAGIPTSGFGCNP...HPHAVHKVIETFLLFNHWSFMATVQASV..MFMKPSKFKKLASVNP.
206K TYMV       P........KELLPLLNSYGIPTSGLGTSH...HPHAAHKTIETFLLCTHWSFQATTPSSV..MFMKPSKFNKLAQVNS.
186K NMV        S........NDEADCLESFGITVNPYANQT...HTHAACKVIENRMLEIVGMHLPKHSCTM..LFLKRSKLRYMRRAAI.
166K PVX        T........VEAANDLEGFGIATNPYSIEL...HTHAAAKTIENKLLEVLGSILPQEPVTF..MFLKPRKLNYMRRNPR.
147K WClMV      A........PDAADTLEKYGIATNPFAVKV...HSHGAVKSIENTLLERVGFNLPKEPCIF..LFLKRSKLRYLRRGPS.
216K ACLSV      P........EKTREWFTKSGVYLSPFAYVN...HSHPGCKTLENHLLFNVVASYISKYSYVACLSIKSNKMSKMERLGP.
223K PVM        S........PLAKRKLSMAGIYLSPYSAVV...HSHPVCKTLENYILYSVLPSYINSSFYF..VGIKERKL.QLLKSKC.
1a   AlMV*      L........PSDKGEVLKISFSLDATQQNI....LRANFPGR.RTVF..SNSSSSSHCFA........AAHRLLETDFV.
237K BNYVV      P....LDKLLSLLAKLMRYCVNMGVSVQEI...YLSDAIVSS.SYML...HVSRSAGCVSFSWLYAKLSMFASCGKFWVG
NSP SINDBIS*    YHCVCPMRSPEDPDRMMKYASKLAEKACKITNKNLHEKIKDL.RTVLDTPDAETPSLCFHNDVTCNMRAEYSVMQDVYIN
126K TMV        .........ARDRRPKVNFSKVISEEQTLI....ATRAYPEF.QITF..YNTQNAVHSLA........GGLRSLELEYL.
  Consensus     ----------E----L--YGI----F-------H-H-A-K--E--LL---------------LFMK--KF--L------

                161                                                              240
194K ELV        ................HFKSLHNYRLHPHDSN.RY.....PHPST.......................SLPTEKRFYIH
205K EMV        ................KFQELINFRLTAADTT.RY.....PSTSL.......................TFPSNSICFMH
210K KYMV       ................FFQHLHNYRLTPTDSV.RF.....PSTSP.......................HLPNTPFVFMH
198K OYMV       ................NFSELVNYRLTAADSV.RY.....PSTST.......................SLPKYEIVFMH
206K TYMV       ................NFRELKNYRLHPNDST.RY.....PFTSP.......................DLPVFPTIFMH
186K NMV        ................LKDVFLNKDVEPKDLF.RY.....DRDTITSR....................LQDIDTKIAYMS
166K PVX        ................IKDIFHNVAIEPRDVA.RY.....PKETIIDK....................VTEITTETAYIS
147K WClMV      ................NKDIFINLAIEPRDLQ.RY.....EEDTLVES....................WTRITTRYAYIS
216K ACLSV      ..............NSVKTYDILNRLVTAKDKA.RYG....PLAKPERS....................PCPKKTNIFIH
223K PVM        ..............KNLDSVQVVNRYVTSADRM.RYTNDFVPYGSYEHECLVHKGVGLDNEALRGLVGPLRRHKAKNLFFH
1a   AlMV*      ................YRC.FGNTVDSIIDLGGNF.......VSHM.......................KVKPHNVHCCC
237K BNYVV      SS.....HHTAANMIEGSRAVNGPDVAISEMVEAFHLE....VKSS.......................LVVTVSLTPRE
NSP SINDBIS*    APGTIYHQAMKGVRTLYWIGFDTTQFMFSAMAGSYPAYNTNWADEK.......................VLEARNIGLCS
126K TMV        ................MMQIPYGSLTY...DIGGNF.......ASHL.......................FKGRAYVHCCM
  Consensus     ----------------------N------D---RY-----P--S-------------------------------F--

                241                                                              320
194K ELV        DSLMYFTPHQISGLFESCPNLLSLYASLVVPPESSMT.DLSLNPDLYRYSIH.................KSTLHYTPEGHS
205K EMV        DALMYFSPAQIVDLFTQSPALETLYCSLIVPPESHFT.DLSLFPEIYTYKIS.................GQTLHYIPENHH
210K KYMV       DALMYYQPEQILHLFHQVPQLTNLFCSLVTPPESHFT.HLSLMPDLYTFTLK.................GQTLHYTPEGHS
198K OYMV       DALMYFNPSQILDLFIQCPSLQRLHCSLVVPPESSFT.DLSLHPNLYTYTIS.................GNTLHYVPEGHH
206K TYMV       DALMYYHPSQIMDLFLQKPNLERLYASLVVPPEAHLS.DQSFFPKLYTYTTT.................RHTLHYVPEGHE
186K NMV        DTLHFMSRREIVQLFEDSPKLETLLATVVLPVEALHK.RTSLYPSLYSINYS.................AKGFEYIPGNHG
166K PVX        DTVHSLDPSYIVETFQNCPKLQTLYATLGLPVEAAFK.MESTHPNIYSLKYF.................GDGFQYIPGNHG
147K WClMV      DTLHFFTRKMLADLFFHNPALDVLYATLVLPPEALHK.HPSIEPDLYTINYN.................FNGFQYIPGNHG
216K ACLSV      DEIHYWSRDQLET.FLQVHRPKNLWATLVFPPEILAGYKSSVLPFLYQFEIH.................GKDLVYMPDGVR
223K PVM        DELHYWSSKVLID.FLDVMRPDKLLGTVVYPPELLFKPTRSLNEWCYTYDIV.................GDTLMFFPDGVQ
1a   AlMV*      PILDARDGAKLT.ERILSLKS....YVRKHPEIVGEA.DYCMDT......FQ.........KCSRR...ADYAF.AIHST
237K BNYVV      KKILEREPGFVPLYKQKSRAPRNHPVLAALREVMRQEYSASCNILNTKLKTLVVGAASREVNCYSSNPSVHY.YFANKDS
NSP SINDBIS*    TKLSEGRTGKLSIMRKKELKPGSRVYFSVGSTLYPEH.RASLQSWHLPSVFHLNGKQSYTCRCDTVVSCEGYVVKKITIS
126K TMV        PNLDVRD.....IMRHEGQKDSIELYLSR...LERGG.KTVPN.......FQ.........K.......EAFDRYAEIPE
  Consensus     D-L-YF----I--LF---P-L--LY--LV-PPE-------SL-P--Y-Y---------------------L-Y-PE-H-
```

```
                321                                                                                                  400
194K ELV        AGSYNQPVNALDWLKISAI....QTPS.....................LSLSVSVLESWGPLHSLLIERSSQ...........
205K EMV        SGSYNQPLQALSWLKISSI....LSPS.....................LALSVTKLESWGPVHSILIQRGLPPKPSLSARPPV
210K KYMV       AGSYNQPITALSWLKINSI....LSPN.....................LNLSITILESWGPLHSILIQRGLPL.....PDPKL
198K OYMV       AGSYDQPLDAISWLKLNSI....HSPH.....................LNLSVSKLESWGPVHSLLITRGLPH...........
206K TYMV       AGSYNQPSDAHSWLRINSI....RLGN.....................HHLSVTILESWGPVHSLLIQRGTPP.....PDPSL
186K NMV        GGCYFHPYTTLEWLKVRMI....NAED.........FHKLSTGF.ILTFQLVESLGANHLFIVQKAK...........L
166K PVX        GGAYHHEFAHLQWLKVGKI....KWRDPKDSFLGHLNYTTEQVEMHTVTVQLQESFAANHLYCIRRGD...........L
147K WClMV      GGSYSHEFKQLEWLKVGHL....KSPE.................L.SLTFQMIESIGANHLFMITRGI...........K
216K ACLSV      SESYTQPLENGFLLSSSSIIIKNRVTG.................VEIRYQVSLVYSLGSHHLFHIYPAEDL...MKEEVRR
223K PVM        SEGYQQPLKGGYLLGARSL....KLPD.................GTV.YMVDVLCSKFPHHLISITKGEAA...APTH.RA
1a   AlMV*      SDLDVGELACSLDQKGVMKFICTMMVD...........ADMLIHNEGEIPNFNVRWEIDRKKDLIHFDFMDEP....ILG
237K BNYVV      KDLVRTTLE..LLHSALA.TKYRNMES...........GERELMNNLKGCGYIVKRSVENAIYEVVSDKDVAE....VLR
NSP SINDBIS*    PGITGETVGYAVTHNSEGFLLC.KVTD...........TVKGERVSFPVCTYIPATICDQMTGIMATDISPDDAQKLLVG
126K TMV        DAVCHNTFQ.TMRHQPMQ..QSGRVYA...........IALHSIYDIPADEFGAAL....LRKNVHTCYAA.......FH
  Consensus     -G-Y--PL----WLK---I-------------------------------V---ESWG--H---I---------------

                401                                                                                                  480
194K ELV        .TQN.....................PDSQKIKDLISFQTPQALILPNPDSLAVPLR.HRLVPQ....KTYDALF..TYT
205K EMV        LPNQPPRATTPNSQNQLLHQTSQLFFELQQPQLSLVSFRIPDCVELPQATFLRQPLR.HRLVPT....SVYNALF..TYT
210K KYMV       LVRSLP.............P....FSRSPDPETDLVSFQVPKSVELPQATFLSQPLR.HRLVPE....SVYNALF..TYT
198K OYMV       LPSS.....................EKQQ.....VSFHIPNCLPLPEATFLHQPLR.HRLVPT....EVYDALF..TYT
206K TYMV       QAPSTP.............MASDLFRSYQEPRLDVVSFRIPDAIELPQATFLQQPLR.DRLVPR....AVYNALF..TYT
186K NMV        LTPQ................................MRTFCR.DSLVTLPQVFCPAAMNANRPLSK....TKAMQML..LYC
166K PVX        LTPE................................VRTFGQPDRYVIPPQIFLPKVHNCKKPILK....KTMMQLF..LYV
147K WClMV      ITPR................................VRTFTK.DSYVLFPQIFHPRNLNPSKPFPK....VKAMQLF..TYV
216K ACLSV      FGPYD..............LFDVGSLFVKPVRVPIQDFPLS..VFKKIFIYMSSLKKPDVQSAV....AKLRQL....SD
223K PVM        FGPFE..............AVASEALKATLSPDYPCAFPVSYEVVNKIYRYLRTLKKPDEQSAI....AKLSQIIAEPSG
1a   AlMV*      YSHRF..............SL..........LKHYLTYNAVDLGHAAYRIERKQDFGGVMVID.....LTYS..L..GFV
237K BNYVV      YAQTV..............AS....TKKEAKRKPNTGKRKMVMSEATRRTIELHELSRIVAEE.....KKIPNHF..HFD
NSP SINDBIS*    LNQRI..............VINGRTNRNTNTMQNYL.LPIIAQGFSKWAKERKDDLDNEKMLGTRERKLTYGCLW..AFR
126K TMV        FSENL..............LL..........EDSYVNLDEINACFSRDGDKLTFSFASESTLNYC...HSYSNIL..KYV
  Consensus     L-----------------------------------F------------FL---L--------------Y--LF---Y-

                481                                                                                                  560
194K ELV        RATRTLRT.........SDPAGFVRT....QSNKPEFNWVTS.....................................
205K EMV        RAVRTLRT.........SDPAGFVRT....QSNKPEHAWVTP.....................................
210K KYMV       RAVRTLRV.........SDPAGFVRT....QSNKPEHKWVTP.....................................
198K OYMV       RAVRTLRT.........SDPAGFVRT....HSNKPQYSWVTS.....................................
206K TYMV       RAVRTLRT.........SDPAAFVRM....HSSKPDHDWVTS.....................................
186K NMV        KSVKQVTE.........RDIYAKIRQ....IIPTSELELYDP.....................................
166K PVX        RTVKVAKN.........CDIFAKVRQ....LIKSSDLDKYSA.....................................
147K WClMV      KSVKNPTE.........RDIYAKIRQ....LIKTSELSDYHP.....................................
216K ACLSV      ADISIESVF.........MIQEFASR....IEKNGVESW.................SC............SF.........
223K PVM        REIDFVECFARLVIHNSSMCATIMPE....QLKEFMGNWLGKMPSVLARRFSSVRAVCVNKFIRGLKPYSFTLRLNEITW
1a   AlMV*      PKMPHS.............NGRSC..............AWYN.....................................
237K BNYVV      ESDFAS.V.......GNFTQLVCEDV....GYNFSVDAWLHL.....................................
NSP SINDBIS*    TKKVHSFY.......RPPGTQTCVKVPASFSAFPMSSVWTTS.....................................
126K TMV        CK...TYF.......PASNREVYMKE....FLVTRVNTWFCK.....................................
  Consensus     --V---------------D---FVR-------------W---------------------------------------

                561                                                                                                  640
194K ELV        .QAWDNLQTYALLTASYRPPVSYT.........LHRSPLTKLKELLTRNA......LKLAAMASPALT.LAIFTTMTALN
205K EMV        .NAWDNLQTLS.VNAPHRPQVCYH.........FFSSPVARLKLHFAQHW......RAYLLALTPFLTTSPLLLPLFNFN
210K KYMV       .SAWDNLQTFALLNCPLRPNVVYH.........VLLNPLQKMKLYFSQHW......RRLGVIAAPGLFCLSLLLRSQKWS
198K OYMV       .RAWDNLQTYALLNAPVRPVVLFD.........FFLSPLKKFQLFMSQHI......NSLVIKALPFLGLIPPLVKLTTLG
206K TYMV       .NAWDNLQTFALLNVPLRPNVVYH.........VLQSPIASLALYLRQHW......RRLTATAVPILSFLTLLQRFLPLP
186K NMV        .DEIVHLANYFFFVSSLDSITCYE.........DLLS.....SDIWNRLT......RPLRTAVRKFVELFKGKQDFDKLL
166K PVX        .VELVYLVSYMEFLADLQATTCFS.........DTLS.....GGLLTKTL......APVRAWIQEKKMQLFGLEDYAKLV
147K WClMV      .DEIVHIVNYFVFISKLDSINSYS.........DILS.....LPIWSKAL......LPIKTKITQLWEKLTGARAFNQLL
216K ACLSV      ...WGCMKDWFFDKLPYREVLEKI.........GLADDFT.RRLMKIKPL......AFDIHTTDRPLTV..RMIIDQIWE
223K PVM        WNIWENSYAWFFDTDAEVDVPEKLDSLFMGEGAGLVAHITSRPYVGTVPLADREWNALLCMDSQKLLHAMRRMFMRGAWG
1a   AlMV*      ......RVKGQMVVHTVNEGYYHH.........SYQTAVR..RKVLVD.........KKVLTRVTE.........VAFRQFR
237K BNYVV      .FEVT.GAQTAVGYMALPNELL.........FEHYP..ISDYYDYWEG.........VEKHGSLGGITISPLRNGQVVGMP
NSP SINDBIS*    .LPMSLRQKLKLALQPKKEEKLLQ.........VSEELVMEAKAAFED.........AQEEARAEKLREALPPLVADKGIE
126K TMV        .FS...RIDTFLLYKGVAHKSV...........DSEQFYTAMEDAWHY.........KKTLAMCNS.........ERILLE
  Consensus     ------L--Y-LL---------Y-----------L---------------------------------L--L-----L-
```

```
                641                                                                                                720
194K ELV        TNSSKALSFSALKIH.............................LLNPLTGPELLHFQ..TSVLQQKNSAPLS.......
205K EMV        TPFP.LPRLLSLFRR.............................SVSSP.RLLHSILP...SQLRGAAIPNRP.......
210K KYMV       LPLPKAKSISVFRRN.............................LLLPPKPHRPPLLPHPEQMLQEFKLPWHRPPPKGKR
198K OYMV       FPIPSVSHFQILFFT.............................MIGPSGQFILEAFP...SMLH.PAISYLE.......
206K TYMV       IPLAEVKSITAFRRE.............................LYRKKAPHHP.........LDVFHLQQH........
186K NMV        IALKWQP..FSYSLE.............................PLDFTAYFVSRRVR...TLAKMEDISWH........
166K PVX        KAVDFHPVDFSFKVE.............................TWDFR..FHPLQAW...KAFRPREVS..........
147K WClMV      DALQWKT..FTYSLE.............................VVDFS...............................
216K ACLSV      ERLSSFDNISNIVFY.............GRR.........EW....L..NNGVLPKVK...KKGLAKLIPGREV.......
223K PVM        AHMCVISREFLLKYVEARLKSSCLIAKARRRGQHKEKLEAWEVLGLKSSDALFRAMTYLCNARLEPMFSESGL.......
1a   AlMV*      PNADAHSAIQSIAT...............................MLSSSTNHTIIGGVTLIS.GKPLSPD.D........
237K BNYVV      TGVFQPVHFDKTSA...............................GLGIPGSKMGTAERVICHMSDGLGNGYN........
NSP SINDBIS*    AAAEVVCEVEGLQA...............................DIGAALVETPRGHVRIIPQANDRMIG.Q........
126K TMV        DSSSVNYWFPKMRD...............................MVIVPLFDISLETSKRTRKEVLVSKDFV........
  Consensus     ----------L-------------------------------------------------------L---------------

                721                                                                                                800
194K ELV        ..............QAEAKQELDKSA..VPAPSEHDSSASQSTSLSLSASSQLLSTEKHPGSELSSKAIPV....STSCPS
205K EMV        ................LPLWVTKLH...HFLDSHSLLPTPPIRPR..IELQRLPL....MSLIPKPK.IV..........
210K KYMV       NPFLTLLINLLHIPREICAGIRRYPSYYQSIQPKPLNPIQQFRNQLLAQLHTLPLPKKFLSILPRARKDIPLLPRMSFTT
198K OYMV       ................SCGLVPRLP...PPV.AQQLQTMGILRKR..SAFSLKTS....LSTVKWPSWKI..........
206K TYMV       ..........LRNHHSAISAVRPASPPHQRL.PHALQKAALL...LLRPISPLLTATPFFRSEQKSMLPNAELSW.TLKR
186K NMV        ................QAAELARRLESEPDLLSWEDLCSKPIADPLASVSTPPLATLTSPAISESRVTDTS..........
166K PVX        ................DVEEMESLFSDGDLLD....CFTRM.........PAYAVNAEEDLATIRKTPEM..........
147K WClMV      ..........................................................TAPSQRDCFMEDE...........
216K ACLSV      ..................DSHNYPREIYSD....LLSSTSIWRSYDDDFRHRSASPLVILRSNRAYSEA..........
223K PVM        ............RFFLTRGRNNLYGLTNYTEGKRAVTGVQNLWSNVVHEVSTKRHKGMIRLEKARVTEQP..........
1a   AlMV*      ................YIPVATTIYYRVKKL..YN.AIPEMLSLLDKEERS.........TDAVLKGSEGP...........
237K BNYVV      ................HVKSDWQTLLKHPILSSSKYNFAVEVDLTGRYGCLATFRLTRVTGVKYVARTIK...........
NSP SINDBIS*    ................YIVVSPNSVLKNAKLAPAH.PLADQVKIITHSGRSGRYAVEPYDAKVLMPAGGA...........
126K TMV        ................FTVLNHIRTYQAKAL.....TYANVLSFVE.SIRS.........RVIINGVTAR...........
  Consensus     ------------------------------------------------L-------------------------------

                801                                                                                                880
194K ELV        ASKQLAPPLTAES....HSSVNALLRKFLGPNSPQSNLDNYNLHLHPESFTLGWKRRPLLL.......DSHSSFLP....
205K EMV        ..LPLLSLL........LSSPTIYIHFFQA.QTPQQLHDNYHLHLHPSRFELSWTLQSYHV.......TQAQSFLPLLLP
210K KYMV       VNLPLQPPMWLAIGASLVPELAFLLSWLSGDVDLQTQHDIYHHHLHPENFTLSWTRTPYLA.......LAPSPFLPYAHS
198K OYMV       ..TALISAL........PAALSVFLKTISP.LSLQSLHDGYNAHLHPSPFNLSWSLETFHV.......QSPSPFLPLSLT
206K TYMV       FALPWQASLVLLS....LSESSVLLHKLFSPPTLQAQHDTYHRHLHPGSYSLQWERTPLSI.......PRTTAFLPFTPT
186K NMV        .SPKLTN..........ISVESILHPTSSNHATIAESLNPKFCGWTRGDFHHMIVNQPSDR.......LKGCRSWSYTTS
166K PVX        .DVGQEV..........KEPAGDRNQYSNPAETFLNKLHRKHS.........REVKHQAVKK.......AKOP.........
147K WClMV      ...RLET..........DTLEDEVSQNANN..........................NKPTS.....................
216K ACLSV      AKFSSNC.................LSLCAAPCDEVIARTPFELNHR..REKKELSLKCLDF.......HIK.KMKVK.NV
223K PVM        RSEFASCVLEPEVWRDVEAALDIELGEVACACNARFVQGVVLSNQAGLNVREQVAGASVGL.......YTKDRSNLKWGN
1a   AlMV*      ..M.W.............YSGPTFLS.ALDKVNVPG........D.F.VAKALLS.............LPKRDLKSLFSR
237K BNYVV      ..LRPED...........RYVRVLDLLHIVRSIRLKGHAGLKEPYQYFPVYKREVD..............TTVSYCFSIAE
NSP SINDBIS*    ..VPWPEFLALSESATLVYNEREFVNRKLYHIAMHGPAKNTEEEQ.YKVTKAELAETEYVFDVDKKRCVKKEEASGLVLS
126K TMV        ..SEW.............DVDKSLLQSLSMTFYLHT........K.LAVLKDDLL.............ISKFSLG....S
  Consensus     ----------------------LL-----------------------F----------------------F------

                881 A                  B                                                                         960
194K ELV        ......SSCLQP............PA.SPSIAAAPHPLPPAQK..PPRPPTTVPT.............PKPLASPSQTQAA
205K EMV        APTQAQASNPAPRPPAFHAIPL.PP.QPSTSSSPPLQEPTLSPHLIHPPLTREPSPLNGCACDSALLPSTAAMTSAEHPT
210K KYMV       PLPPLPVNSSPLFP...PPPPLPPS.QPPLSQGPATQAPSAQPTPGEPLLAPPTTELKPESSNPN.NPNPSSSAGSNPPP
198K OYMV       PPSPSEEIALPP........PIFRV.PPPL...PAQETPS....PPAPALVPPT........QPP.QPQPWEEIISTFLS
206K TYMV       TSTAPPDHSEASLPPAFASTSV.PR.PPPVASSLGAQPPTTTAAPPTPIEPTQRAHQNSDLTLESSTPIE.PPPPPIQSS
186K NMV        PDVDLMLENLLLTP.....IPWESR.LSDI..LLTLNTPANACCIQILDCSAASAWVDWSVQPIKTFPVAFVGMGETTLT
166K PVX        ..........................R.LAEI.....................................................QESMR
147K WClMV      ..........................LQNI.....................................................EEAVK
216K ACLSV      LELEVKLRERNTR.ISLSKKGVKKA.GRSRMIPVHLLKPTCGEGNGKPEERNKEEAKIPMNEGTSKEEKGSEPPHSEVKR
223K PVM        SELLSNGWGRSLS.VWMEINSVSQKFDVAVRLSYSKETQMNVLLPSLDGIERGAGATVVNLRKCGAFIVRCARGWRLALA
1a   AlMV*      SATSHSERTPVRDE..SPIRCTDGVFYPI..RMLLKCLGSDKFESV..TITDPRSNTETTVDLYQSFQKKI.ETVFSFIL
237K BNYVV      KSLTVQNIANFIR.HHIGGVSLVNKELVSAWRLNPQLVPSFAYAVYFYVVNLRG........ELDGMLQKLMKKGITWAD
NSP SINDBIS*    GELTNPPYHELALE...GLKTRPAVPYKVETIGVIGTPGSGKSAIIKSTVTARDLVTSGKKENCREIEADVLRLRGMQIT
126K TMV        KTVCQHVWDEIS.......LAFGNAFPSVKERLLNRKLIRVAGDAL..EIRVPDLYVTFHDRLVTEYKASV.DMPALDIR
  Consensus     ----------------------------------------p-----------------------------------------
```

```
                961                                                                                                1040
194K ELV        QPATQSPPSIPQTAPVTS.....................................LLP.....................
205K EMV        PLNPPTPSPTPDVPPPDS.PGNPSLLKQVPPEANLHPIHNPDLPSSTTLPSGALTLV...PAKTPSIYA.NPTPPSSHPF
210K KYMV       KSSSSDNPPAPNKPTPTS.SSTTPPSPNLPLQF...GSIHSPFLSDGQLNYSALPPPQDPTNTTLSLLP.EPKPPTEVQ.
198K OYMV       SLNASSHKPSPSSAPLES.PSPIPESFEVLAEA...PQHQRSINECGALNQLPLPVP.....SPLQNST.NPKPPF.PS.
206K TYMV       DIPPSAPVLFPEINSPHRFSPKLPTTPDFEPTR..TSPPPSTSHQDSTDPADPLMGSHLLHHSLPAPPT.HPLQSSQLLP
186K NMV        FEDDSTLPLKEGEFVFFP.............................PEWLARHKYQIKTGSNLHLC.ATFLVLDTTL
166K PVX        AEGEAEPNETSGGMGAIP..........................................................
147K WClMV      NNPD........................................................................
216K ACLSV      EGVRLDEQHISEPLLSFKLDDFVGREKLCSAGLIKTVGNDYLLLARQIECMPLSQLRGKKAAYFCLDFP.MVYFHDKVSY
223K PVM        WMDHICLEVMANVAYGHECYMRSWGTMDVVVFLKRATVSEQVTFESAQEVGPIEGKSDSGAPGVGVNLDLGGVVGSEYPA
1a   AlMV*      GKIDGPSPLISDPVYFQSL..EDVY....YAEW.....HQGNAIDASNYARTLLDDIR....KQKEESLK.AKAKE....
237K BNYVV      RLKANVSAFLRDMVDPISFLWTWLFERRLVDQIFQDGTDVFYQMDRACVDEKALRLNDH.IKITRDFLPADTLLPEGWSL
NSP SINDBIS*    SKTVDSVMLNGCHKAVEVLYVDEAFACHAGALL.....ALIAIVRPRKKVVLCGDPMQCGFFNMMQLKVHFNHPEKDICT
126K TMV        KKMEETEVMYNALSELSVLRESDKFDVDVFSQM.....CQSLEVDPMTAAKVIVAVMS....NESGLTLTFERPTEANVA
  Consensus     ------------------------------------------------------------------------------

                1041                                                                               1120
194K ELV        APLETDDSCAGPISTFQDLFPASYYPHTANF...PCRSKIPGYLEAPYPPLDCMLVALSAQM......PQSPQELWSALN
205K EMV        TPLADDPTAVGPCLPFHVLHPADYFPLSAEF...LTRTRHVPPSSLSHPKLNCLLTCFSELS......GHSESDLWLSLQ
210K KYMV       SPLMADPTCVGPAVSFSSLYPRDFFPNTASF...LTRLRLSPPTPLPMPKNNCLLTAVAPSL......HINPHRLWTSLQ
198K OYMV       SDLLSDMSCTGPVVEFETIFPAEYHMSNGSF...PTRLRGHPRSSAPFPQKHCLLTAVASQL......SYTEHQLWEFLC
206K TYMV       APLTNDPTAIGPVLPFEELHPRRYPENTATF...LTRLRSLPSNHLPQPTLNCLLSAVSDQT......KVSEDHLWESLQ
186K NMV        LGETLANNCIEPEPLRPTASRKSCQPTPSVD...KSGADSPPEKQIVTPIVDAAVLSCRPKL......QDVTK.....PT
166K PVX        ..................SNAELPGTSDA...RQELTLPTTKPVPARWEDASFTDCSVEE......EQVRL.....LG
147K WClMV      ............................................................................
216K ACLSV      PTFEATGEIRHVMMKARSKWGIDFNSALIQVYNDGCRLPLHSDNEECYDDDGILTINVVGDAKFHTTCHDEVIDLKQGNE
223K PVM        NGAERYKRVSGPGDGCCCWHSFAYLVGMHH....MELKRLCTSHVFENAALNVELEQCKASGAFVTHAAILATALRLRAE
1a   AlMV*      ..VED..AQKLNRAILQV..HAYLEAHPDGG...KIEG.LGLSSQFIAKIPELAIP.............TPKPLPEFEK.N
237K BNYVV      DDWEKAPDSLKTLSAAASL.PVECGAVNCVGKSFKSVRTLLPPSVVTSPVEQFFKSGGKFRD.....DAEFAELLSAHYR
NSP SINDBIS*    KTFYKYISRRCTQPVTAIVSTLHYDGKMKTTNPCKKNIEIDITGATKPKPGDIILTCFRGWV....KQLQIDYPGHEVMT
126K TMV        LALQD..QEKASEGALVVTSREVEEPSMKGS...MARGELQLAG.LAGDHPESSYS............KNEEIESLEQFH
  Consensus     --L--------------------Y-------------------------------L------------------L---L-

                1121                                                                               1200
194K ELV        TLMPLS....ALT.SPSLRVLGLGTEELTAL..SYYYHFQAEIH.....................SDNEIYRF.........
205K EMV        SILPDS....QLQ.NPEVSTLGLSTDILTAL..CFIYHSSVTLH.....................APSGVYHY.........
210K KYMV       EVLPDS....LLS.NSEIDSVGMSTDLLTAL..SHLFNFQAVVH.....................SERGDILF.........
198K OYMV       DMLPDS....LLT.NSEVENFGLSTDHLTCL..SYRLHFECIIH.....................TSHSTIPY.........
206K TYMV       TILPDS....QLR.NEEINSLGLSTEHLTAL..AHLYNFQATIY.....................SDRGPILF.........
186K NMV        KVLMAA....ETS.DSVADSLPWASWVNLLQ..KHGFKGNQQQI.....................AQDGQLII.........
166K PVX        EEAVKT....AT..QQVIEGLPWKHWIPQLN..AVGFKALLIQR.....................DRSGTMIM.........
147K WClMV      .....................LPWAPWLLILQ..AHNADCTQKQY.....................DPENNLIL.........
216K ACLSV      ILMPAGYQKKNRH.AVEVASEGRTSVTLRVH..KRDFSFESKLRFIKGKFDCLFVSVAEIIHKKPEEIMMFIPHIMDRCV
223K PVM        IRVHNAGTGRVHRFAPKQKNMALDLWLESEHYEPQVLRNGCVIE.............SVAQALGTRNADIL.........
1a   AlMV*      AETGEI.....LRINPHSDAILEA...IDYL.KSTSANSIITLN.....................KLGDHCQW.........
237K BNYVV      WQMDNSFCACQVCAALTGKTGSQVVECRWKDESMYTFSMSQTEV.....................DDFRNEIKA.........
NSP SINDBIS*    AAASQG.....LTRKGVYAVRQKV..NENPL.YAITSEHVNVLL.....................TRTEDRLVW.........
126K TMV        MATADS.....LIRKQMSSIVYTGPIKVQQM.KNFIDSLVASLS.....................AAVSNLVKI.........
  Consensus     --L--------L--------L---------L-----F---------------------------------F---------

                1201                                                                               1280
194K ELV        .....GIQTA.........................STKLCLIRDSGPPAHFTAP....DPLRAGSPPSRSQTNENSLRRSLL
205K EMV        .....GIASS.........................STVYVIHYQPGPPPHFSLS....PRL.AASAPRCN.PTNSRLVRQAL
210K KYMV       .....GLQSA.........................KTVIHIYHTNGPPAHYSPP....PKI.IGSNSPPS.SQQHPLEQAAL
198K OYMV       .....GIKKA.........................STVIQISYIDGPPKHFKAF....IKL.AAAAPGSN.PSKSNLVRAAL
206K TYMV       .....GPSDT.........................IKRIDITHTTGPPSHFSPG....KRL.LGSQPSAKGHPSDSLIRAMK
186K NMV        .....PISD.........................IRKL.......PDIPFPEE....VP..............ETLRETLK
166K PVX        .....PITEM........................VSGF.......GKRGLPEG....TP..............KELARELL
147K WClMV      .....PIQE.........................INTL.......PKHQHPD.....IP..............TDLLTLLT
216K ACLSV      SNRGCSLDDAKAICEKYEIKIECEGDCGLVECGTSGLSIGRMLLRGNHFSVASVRRSSMDSLANSSKEIKSDGVLDHVTF
223K PVM        .....AVVEERC.....................CEEVVESVQAGLGLNLHHVEIVLQCFDIVGHCNLGDKEITLNAGGKMPF
1a   AlMV*      ......TTKG.........................LD..VVWAGDDKRRAFIPKK...NTW............VGPTARSYPL
237K BNYVV      ....QSIEKGNRF.....................GELLIGVHQKIPTQAFEVSV..RLEYVKGGPGTGKSFLIRSLADPIR
NSP SINDBIS*    ......KTLQ.........................GDPWIKQPTNIPKGNFQATI...EDWEAEHKGIIAAINSPTPRANPF
126K TMV        ......LKDT.........................AA..IDLETRQKFGVLDVAS...RKW..........LIKPTAKSHAW
  Consensus     ------I-----------------------------L-------P---F-------------------------L----L
```

```
                1281                                                                1360
194K ELV        GFRLNGNLL.PIDQVHSFTSEPSRAKNLASNMKNGF.....DGILTTLAA......LSSLSSGPSPRDRIFTLDGICDFA
205K EMV        RFKLNGEFL.PFTQAYAHESSITHAKNLISNMKNGF.....DGIMSSLTD......S..SKGPSPREKLTTLDSLIDVA
210K KYMV       RFKYQGSHL.PFSSFHSFTTSVQHAKNLISNMKNGF.....DGVMSTIEP......SIRHQPGHSPREKFIALDAMIDLA
198K OYMV       RFKYNDAFL.PFWDAHQHTISVPHAKNLISNMKNGF.....DGITSQLSG......P....SNKSPKMKLLELDATIDVS
206K TYMV       SFKVSGNYL.PFSEAHNHPTSISHAKNLVSNMKNGF.....DGILSLLDV......STGQRTGPTPKDAIIQIDHYLDTN
186K NMV        NIKRFPVEI.TM........QHKRAGSYDSDIKNNR.....TGKLLS.................QMDNKWKAAFAYKLQQE
166K PVX        VMNRSPATI.PL........DLLRARDYGSDVKNKR.....IGAITK.................TQATSWGEYLTGKIESL
147K WClMV      KLHREPTTV.PL........DNHRARAYGSDVKNLR.....IGALLK.................KQSKDWLASFALKTENI
216K ACLSV      NFHKRLKLVEPDLTNADIKVDSSRAGKLLKSLMDGM.....TGIVSHNSTHEGWRMIKGINSTSEMRSFMNMVRGQIEEP
223K PVM        CFDISDEHM.SFCGRRKDPICKLVSGALHGKMFAES.....ALLDLENCGLKIDFEPNWNRAGMLADSMYQGATGVLGSA
1a   AlMV*      AKYERAMSKDGY.VTLRWDGEVLDANCVRS..LSQY.......EIVFVDQ.....SCVFASAEAIIPSL.EKAL.GLEAH
237K BNYVV      DLVVAPFIKLRSDYQNQRVGDELLSWDFHTPHKALDVTGKQIIFVDEFTAYDWRLLAVLAYRNHAHTIYLVGDEQQTGIQ
NSP SINDBIS*    SCKTNVCWAKAL.EPILATAGIVLTGCQWSELFPQFADDKPHSAIYALDV.....ICIKFFGMDLTSGLFSKQSIPLTYH
126K TMV        GVVETHARKYHV.ALLEYDEQGV.VTC......DDW.......RRVAVSS.....ESVVYSDMAKLRTL.RRLLRNGEPH
  Consensus     -F--------PF------------A--L-S-MKN-F------G----------------------------L-------

                1361                                                                1440
194K ELV        LP..........KTVDLI..HLSGFAGCGK.THPIQQLLKT..PHFHNFRVVTPTTNLRS.....EWKSDMAL......
205K EMV        AP..........REVSLI..HIAGFAGCGK.THPIQKLLQT..SPFHDFRISCPTNELRS.....EWKRDMQP......
210K KYMV       RP..........KTVSMF..HLAGFAGCGK.TKPLQSLLST..RPFHSFRVSTPTTELRN.....EWKKDMNL......
198K OYMV       FP..........RKCDVI..HIAGFPGCGK.SHPIQKLLQT..PAFRHFRLSVPTNELRS.....EWKRDLNL......
206K TYMV       PG..........KTTPVV..HFAGFAGCGK.TYPIQQLLKT..KLFKDFRVSCPTTELRT.....EWKTAMEL......
186K NMV        .D..........RKVCGT..IIHGCGGSGKGSFAIQEWMRSLKEDQSVVTVVTPTVLLRN.....DWQTKL........
166K PVX        TE..........RKVATC..VIHGAGGSGK.SHAIQKALREIG.KGSDITVVLPTNELRL.....DWSKKV........
147K WClMV      .E..........RQVLMS..VIHGAGGSGK.SHAIQTWMRSLNRRDRHVTIILPTTDLRN.....DWTTKV........
216K ACLSV      KSDLFDKVQELNFMKVKIY..GIFGFAGSGK.SHAIQNLIQTEFKGSQGIMVICPRRFLAK.....DWSEK.GV......
223K PVM        LFNNKRNMREKFVRNVSLSLHAIVGTFGSGK.STLFKNLLKY..GAGKSLDFVSPRRALAE.....DFKRTVGMNERGGR
1a   AlMV*      FS..............VTIV....DGVAGCGK.TTNIKQIARSSGRDVDLILTSNRSSADELK....ETIDCSP..LTKLH
237K BNYVV      EG..........RGEGIS..ILNKIDLSKVSTHVPIMNFRNPVHDVKVLNYLFGSRMVPMSSVEKGFSFGDIKEFSSLS
NSP SINDBIS*    PA..........DSARPVA..HWDNSPGTRK.YGYDHAIAAELSRRFPVFQLAGKGTQLDLQTGRTRVISAQHNLVPVNR
126K TMV        VS..........SAKVVLV....DGVPGCGK.T...KEILSRVNFDEDLILVPGKQAAEMIR....RRANSSGIIVATKD
  Consensus     ---------------V--------G--G-GK----IQ--L---------F-V--PT--LR------EW------------

                1441                                                                1520
194K ELV        ..PAHHNWRFSTWESALLK....H.AE..ILVIDEIYKLPR.......GYLDLSLI...ADPTV...K..L.ILLGDP..
205K EMV        ..TAENVWRFSTWESSLLK....H.SE..ILVIDEIYKLPR.......GYLDLSIL...ADPTL...S..LVIILGDP..
210K KYMV       ..PASQAFRFCTWESSLLK....Q.TK..ILVIDEIYKLPR.......GYLDLCIL...ADPCL...E..LVIILGDP..
198K OYMV       ..PESEVWRLCTWETALFK....S.SN..ILVVDEIYKLPR.......GYLDLILL...ADPSI...Q..LVIMLGDP..
206K TYMV       ..HGSQSWRFNTWESSILK....S.SR..ILVIDEIYKMPR.......GYLDLSIL...ADPAL...E..LVIILGDP..
186K NMV        ..PILPADVFKTFEKSVIQ....PCNP..ILVFDDYTKLPP.......GLIESVVM...HHQNV...V..FIILTGDN..
166K PVX        ..PNTEPYMFKTYEKALIG....GTGS..IVIFDDYSKLPP.......GYIEALVC...LYSKI...K..LIILTGDS..
147K WClMV      ..PNLEQANFKTFEKALCQ....PCGK..IIVFDDYSKLPQ.......GYIEAFLA...INQNV...I..LAILTGDS..
216K ACLSV      ..DEKDIKTFESALKSDVK....G.KR..LFILDEISLLPK.......GFTDLLMLKMHMEGIL...KKSTIVCIGDP..
223K PVM        AKAGQENWRVTTLETFLARVEFLTEGQ..VVILDEMQLYPP.......GYFDLVVS..MLKVDV...R...LFLVGDP..
1a   AlMV*      YIRT..CDSYLMS.ASAVK...AQ..R...LIFDECFLQHA.......GLVYAAATLAGCSEVIGFGDTEQIPFVSRN..
237K BNYVV      NIPDTKIIHYSDETGEHMMPDYVRGVSKTTVRANQGSTYDNVVLPVLPSDLKLINSAELNLVALSRHRNKLTILLDNDGM
NSP SINDBIS*    NLPHALVPEYKEKQPGPVKKFLNQFKHHSVLVVSEEKIEAPRKRIEWIAPIGIAGADKNYNLAFGFPPQARYDLVFIN..
126K TMV        NVKT..VDSFMMNFGKSTR...CQFKR...LFIDEGLMLHT.......GCVNFLVAMSLCEIAYVYGDTQQIPYINRV..
  Consensus     --P------F-TWE----K----------ILV-DE--KLP--------GY-DL-----------------LIIL-GD---

                1521                                                                1600
194K ELV        LQGEYHSTSAHSSNLRLSS....EIPRLLPFIDY.Y.CYWSYRVPK.CVAKLFSLPCFN.PSEG.F...IKTTLDFFP..
205K EMV        LQGEYHSTSPHSSNHFLPS....EVHRFKSYIDC.Y.CFWSHRIPK.QIASLFGVVCHN.TNEG.F...VRALTSHPP..
210K KYMV       LQGEYHSTSPHSSNHQLQS....ETTRLLPFIDH.Y.CWWTYRVPS.HIADLFSVPSFN.RSEGHYQMAVRTADSYTP..
198K OYMV       LQGEYHSSHPSSSNSRLES....ETTRLSKYIDC.Y.CWWTYRCPK.AVADLFGVKTFN.SNEG.F...IRAVLSHPP..
206K TYMV       LQGEYHSQSKDSSNHRLPS....ETLRLLPYIDM.Y.CWWSYRIPQ.CIARLFQIHSFN.AWQG....IIGSVSTPQD..
186K NMV        RQSVYHETNPEAYIAALPE....AVEIFSPYCEF.Y.LNATHRNVK.DLANKLGV..YS.EREGKLKVNFASHHLKAS..
166K PVX        RQSVYHETAEDASIRHLGP....ATEYFSKYCRY.Y.LNATHRNKK.DLANMLGV..YS.ERTGVTEISMSAEFLEG...
147K WClMV      KQSFHHESNEDAYTATLEP....SINTYQPFCRY.Y.LNITHRNKP.DLANKLGV..YS.CSSGTTSFTMSSQALKG...
216K ACLSV      LQAGYFCPKDDNYLSR.EG....EIKRLFKGGVN.Y.KWYSYRINK.FIAKKLAIETMN.DFIGIDEQSSIYKDMPSA..
223K PVM        AQSDYDSEKDRLVLGAMEE....NMSVVLGAREYNY.KVRSHRFLNCNFIGRLPCEINK.DDCTIDEPHIMRMHLENLLD
1a   AlMV*      PSFVFRHH..KL.TGKVER....KLITWRSPADATY.CLEKYFYKN.KKPVKTNSRVLR.SIEVVPINSPVSVERNT...
237K BNYVV      NIGAVLKGMLEGVPEELER....RDYIVGMYLGLHLPIKKEFFFPESEFAKSFRLMVAKYEAFVPYDSDLPTLVSQG...
NSP SINDBIS*    IGTKYRNHHFQQCEDHAATLKTLSRSALNCLNPGGT.LVVKSYGYA.DRNSEDVVTALA.RKFVRVSAARPDCVSSN...
126K TMV        SGFPYPAHFAKLEVDEVET....RRTTLRCPADVTH.YLNRRYEGF.VMSTSSVKKSVS.QEMVGGAAVINPISKPL...
  Consensus     -Q--YH----------L----------L--Y-D--Y--W---R------A--F-V--F-----G----------------
```

```
                1601                                                                                                1680
194K ELV        .....SANNLVNSHSVVHI.SEACGWNA........VTISSSQGCTFSDPAFIHLDRNTAL....LSPSNCLVALTRSRS
205K EMV        .....NSKNLTNATNTALS.LQQMGHHA........ITISARR.VTFTEAHTILLDRHTNL....LSPNNCLVALTRSRT
210K KYMV       .....GHFNLVNSVATANA.VIQLGFPA........TTISASQGVTHHNRVTILLDKHSRL....LSPSNTLVALTRSTV
198K OYMV       .....NLPNLVNSIATANT.MQSLGHHA........LTISSSQGMTYSDPVTVLLDRHSLL....ITPQTALVALTRSRS
206K TYMV       .....QSPVLTNSHASSLT.FNSLGYRS........CTISSSQGLTFCDPAIIVLDNYTKW....LSSANGLVALTRSRS
186K NMV        .....RIPMLVPSTMKRNA.MFDMGHHS........MTYAGCQGLT.APKIQIL..NHTQF....CSERVLYTCLSRAVD
166K PVX        ......IPTLVPSDEKRKLYMGTGRNDR........FTYAGCQGLT.NPKVQIVLDHNTQV....CSANVMYTALSRATD
147K WClMV      ......MPILSPSIMKKTA.LGEMGQKS........MTYAGCQGLT.TKAVQILLDTNTPL....CSSNVIYTALSRAVD
216K ACLSV      .....HHFMEKKGNHIEVILVASMVEKELYSNYGNVMTFGESQGLTF.NCGVIVLSEEAKL....CSDAHIMVAITRFRR
223K PVM        VAEEYKSVVLVSSFDEKMVVCAHLPEAK.......VLTFGESTGLTF.MHGTIYISAVSER....TNERRWITALRRFRF
1a   AlMV*      .....NALYLCHTQAEKAVLKAQTHLKG....CDNIFTTHEAQGKTFDNVYFCRLTRTSTS....LATGRDP......IN
237K BNYVV      .....DVVVLDIARVENDINDAFDCADF....YNLVSRPNNCLVVAISECLGVTLEKLDNLMQANAVTLDKYHAWLSKKS
NSP SINDBIS*    .....TEMYLIFRQLDNSRTRQFTPHHL....NCVISSVYEGTRDGVGAAPSYRTKRENIADCQEEAVVNAANPLGRPGE
126K TMV        .....HGKILTFTQSDKEALLS....RG....YSDVHTVHEVQGETYSDVSLVRLTPTPVS....IIAGDSPHVLVALSR
  Consensus     ---------L--S-----------------------T----QG-TF-----I-LD----L-----S------AL-R---

                1681                                                                                                1760
194K ELV        GVYFK......GDFTFLSSLSGSSRMFSLAYSGQPIHLPDFFPEIVFQLNMITAPLTKRSSSFRSG...FQPNISSAPKI
205K EMV        GVYFV......GNLHLASNSFGTNYMFSQALCQGTIDLNNVFPHIMPHLPKMYEPIRSRSNRFVAGSLNFRPTTNSRLLS
210K KYMV       GVEFL......GDIGSLSGTNNSSDMFSRAIYRQPINLSSSFPRIFHLLPLLNKPISRRSTRLIGSH...SPIFHNP...
198K OYMV       GIYFI......GSMYTASGSAGTSYMFSCALTGLPVDMMSAFP.LFHTLPLIHEPIRSRRHRLVAGH...TPSLHVPPSN
206K TYMV       GVQFM......GPSSYVGGTNGSSAMFSDAFNNSLIIMDRYFPSLFPQLKLITSPLTTRSPKLNGA....TPSASPT...
186K NMV        RIHFI.............NTGPTTGDYWAKLEST.....................................................
166K PVX        RIHFV.............NTSANSSASWEKLDSTP.....................................................
147K WClMV      HIHFI.............NTGPNSTDFWEKLDSTP.....................................................
216K ACLSV      GFCFA............LGSKGSKEDYMRSMKSGL.............LQRICSGVGASKEFILGS.............
223K PVM        NLCFVNC....SGMDYQQLAGRYKGRVRSKFLCKT.AIPDDLNSMLPGQALFKSEYPRLIGKDEGVR...EEKLAGDPWL
1a   AlMV*      GPC..................NGLVALSRH..KKTFKY....FTIAHDSDDV.....IYNACR...D....AGNTDDSILA
237K BNYVV      PSTWQ........DCRMFADALKVSMYVKVLSDKPYDLTYEVDGAGSSVTLYLTGKESDGHFIAAP....LSSSLSTNER
NSP SINDBIS*    GVCRAIYKRWPTSFTDSATETGTARMTVCLGKKVIHAVGPDFRKHPEAEALKLLQNAYHAVADLVN....EHNIKSVAIP
126K TMV        HTCSLKY........YTVVMDPLVSIIRDL.EKLSSYLLDMYKVDAGTQQLQ.IDSVFKGSNLFVA....APKTGDISDM
  Consensus     GVCF--------------------MF---L----------F--------L---------------------------

                B  1  C                                                                                             1840
194K ELV        PAPPNLPCP.............PHIPTNYSKD.....VIVNNQALYGESLERRLSV.....................
205K EMV        SLTKPTHLP.............PHIPTNHSLD.....VLVSNPVLLGETLDPRLEV.....................
210K KYMV       .RLTNIHLP.............PHIPTSYSQD.....FVVSNPIFQGQ.ADPRLDT.....................
198K OYMV       KWPHRLHLP.............PHIPTSHSKD.....VILAHGIVASNAPERRLTT.....................
206K TYMV       HRSPNFHLP.............PHIPLSYDRD.....FVTVNPTLPDQGPETRLDT.....................
186K NMV        .........................PYLKA.....FI.......DTYRDEKTEV.....................
166K PVX        .........................YLKT.....FL.......SVVREQALRE.....................
147K WClMV      .........................YLKT.....FL.......DCVREERMNE.....................
216K ACLSV      .SSVNLILS.............EKDIAKGAG.....ID.......EMDREARLEG.....................
223K PVM        KTMINLYQA.....................PEVE.....IA.......E.EPEVVMQE.....................
2a   AlMV*      RSYNHNF.......PITLLSEFRASDNAKPLDI.....VEIIPDVSPTKPYEAVISG.....................NDWM
237K BNYVV      ESGDNSKKP.ADDSDTFDAANLFADKGVSSADMEAFCAYLEKTLMATIMKYDLSLQSWANVVDDTDDFYQINISEFRQST
NSP SINDBIS*    LLSTGIYAAHLQKKSVLQNQLTEPTLERNVLERIHAPVLDTSKEEQLKLRYQMMPTE......ANKSRYQSRKVENQKAI
126K TMV        QFYYDKC.....................SDMQF.....Y...........YDKCLPG.....................N...
  Consensus     ------------------------------------F-----------E--L-----------------------

                1841                                                                                                1920
194K ELV        ..LHLPPTRMTLHSDINITAPSSSSFQPSDEPVPSDH...TAVYPGFDFFTLAAH.....FLPAHDPEVK.....EIELK
205K EMV        ..LHLPPTRLPLHLDLLPTVPSSSSFSSVDHLFPTPI...SPAICGYTFENLAAF.....FLPAHDPDLK.....EVLIN
210K KYMV       ..HFLPPTRLPLQSELLPAQLSQTTKPTDSFTNNTPF...TPVYPGENFENLAAF.....FLPAHDPELK.....EVTRR
198K OYMV       ..LHLPPTRLPLHFDLESCNPSTVSTSSTSN.SEVPF...THAFLGESFEELAAH.....FLPAHDPDLK.....EVTVG
206K TYMV       ..HFLPPSRLPLHFD.LPPAITPPPISTSVDPPQAKA...SPVYPGEFFDSLAAF.....FLPAHDPSTR.....EVLHK
186K NMV        ..YNSQPA.........SAEPTEP......EAPATH.............FP.TAPK.......PLLEPLVE.....KLTDK
166K PVX        ..Y..EPA.........EAEPIQE......PEPQTH.............MCVENEE.......SVLEEYKE.....ELLEK
147K WClMV      ..I..VAV.........EEPPAPV......PAPTTH.............FPKVNPT.......TVIESYVH.....DLPEK
216K ACLSV      ..DVWLKSMIYLGKRYHMVEPLGQVIKLTDDAIKCHIPVCSSQTLGPELDNIQAR.......EYREFKGK.....NGWSN
223K PVM        ..E.WFRT........HL..PRDEL.....ESVRAQW............VHKILAK.......EYREVRMG.....DMVSE
2a   AlMV*      TLGRIIPTTPVPTI...RDVFFSG.LSRHGSPEVI....QNALDEFLPLHHSIDDKYFQEWVETSDKSLD....VDPCRI
237K BNYVV      CFGKLLSALEVLKVDVSRKRFISDWLCKNLENKQFRWRWSSSVASASSAGSNVDDDFVNMAGGKTDANADPADVLRQSFM
NSP SINDBIS*    TTERLLSGLRLYNSATDQPECYKITYPKPLYSSSVPANYSDPQFAVAVCNNYLHENY...PTVASYQITD....EYDAYL
126K TMV        ...........STM...MNNF...........DAV....TMRLTDI.....................SLN....VKDCIL
  Consensus     ------P-------------P---------------------------F----------------D---------E----
```

```
                1921                                                                2000
194K ELV        DQTSQQFPWLNLDFHISCQTSSLISARHQPGSDSTLLPASLHKRLRFRPTAAPYQITPSDSFLGNCLYRSWCQVYRRDP.
205K EMV        DQKSNQFPYLDAPFELSCQPSSLLAPIHKPASDPTLLPGSIKKRLRFRASSSPYSITPSDQLLGQHLFSSLCLAYGRNP.
210K KYMV       DQTSAQFPWFDRPFSLSCQPSSLIAAKHSPSQDPTLLPFSIPKRLRFRKSDNPHVLSAIDVLLGNQLFFNLCKAYRRNP.
198K OYMV       DQTSQQFPYLDQPYTLSCQPSSLLAASHKPASDPTLLIFSISKRLRFRASSSPYAFTPNDLILGHLLYTNWCKAFGRCP.
206K TYMV       DQSSNQFPWFDRPFSLSCQPSSLISAKHAPNHDPTLLPASINKRLRFRPSEAPHQITADDVVLGLQLFHSLCRAYSRQP.
186K NMV        AAREIFSPAFGHSNAIQTEDSVVQLFQHQQAKDETLYWATIDTR.AISTPEANLREFNMKRDIGDILFMNYAKLMCLPP.
166K PVX        FDREIHSESHGHSNCVQTEDTTIQLFSHQQAKDETLLWATIHARLKTSNQEANFREFLSKKDIGDVLFLNYQKAMGLPK.
147K WClMV      HGREIFSETHGHSNAIQTDNPVVQLFPHQQAKDETLYWATIEARLQCTSSEENLKEFHLKHDIGDILFLNYKQAMNLPQ.
216K ACLSV      QFREEAGPNWKFPYKVNQAMSYAAVYPRHKMDDDLTFLAAIKKRLRFDNVANNYAKFKAAESRGKYLTKIFLKHVPIKC.
223K PVM        QFTHDHTKQLGAKQLTNAAERFETIYPRHRASDTVTFLMAVKKRLSFSNPGKEKGNLFHAASYGKALLSEFLKRVPLKP.
2a   AlMV*     DLS.VFNNWQSSENC.YEP.RFKTGALSTR.KGTQTEALLAIKKRNMN..VPNLGQIYDVNSVANSVVNKLLTTVIDPD.
237K BNYVV      DYASEFVPILIAESPIFMP.LVEPEPILSKCMVPEFDAFLLIKEFDLDNGADEYQCAYLNESVANRIGDKFVSGVLDTDI
NSP SINDBIS*   DMVDGTVACLDTAT..FCPAKLRSYPKKHEYRAPNIRSAVPSAMQNTLQNVLIAATKRNCNVTQMRELPTLDSATFNVEC
126K TMV        DMSKSVAAPKDQIKP.LIP.MVRTAAEMPR.QTGLLENLVAMIKRNFN..APELSGIIDIENTASLVVDKFFDSYLLKE.
  Consensus     D------P-L-----------------H----D-TLL---I-KRL-F------Y---------G--LF--L---------

                2001                                                                2080
194K ELV        ..................NVRLPFNEALFLEC......IAVNDY............AQLSSKTQATI.........VANA
205K EMV        ..................NSVLPFQPELFSEC......ICINDY............AQLSSKTQATI.........VANH
210K KYMV       ..................THVGPFNPALFAEC......IALNDY............AQLSSKTQATL.........VANH
198K OYMV       ..................NSTIPFNPALFAEC......ICLNEY............AQLSSKTQATI.........VSNA
206K TYMV       ..................NITVPFNPELFAEC......ISLNEY............AQLSSKTQSTI.........VANA
186K NMV        ..................E.PVPFEERLWKIS......ADEVRN............TYI.SKPIGNL.........VNAA
166K PVX        ..................E.RIPFSQEVWEAC......AHEVQS............KYL.SKSKCNL.........INGT
147K WClMV      ..................D.PIPFNPDLWTLC......KQEIEN............TYL.KKSAAAL.........VNAA
216K ACLSV      ..................G.R...DQRLLDQC......RQEFEE............TKL.SKSAATI.........GAHH
223K PVM        ..................N.H...NVRFMEEA......LWNFEE............KKL.SKSAATI.........ENHS
2a   AlMV*     ..................KLCMFPD...FISE.......GEVSYF............QDYIVGKNPDP...........EL
237K BNYVV      ISPLNLRGHPIAENVKYHSMCVAPAQIYFKRNQWQELQVQQARYLFRKVRNSPSSTQDSVARMVAQLFVSDCLVPNVADT
NSP SINDBIS*   FR..............KYACNDEYWEEFARKP....IRITTEFV............TAYVARLKGPK.........AAAL
126K TMV        ..................KRKPNKNVSLFSRE.......SLNRWL............EKQEQVTIGQL...........AD
  Consensus     ----------------------PF---LF--C----------EY--------------L-SK--A----------V---

                2081                                                                2160
194K ELV        SRSDPDWRHTF.........................VKIFAKSQ.HKVNDGSIFGPWKACQTLALMHDYVILTLGPV.KK
205K EMV        QRSDPDWRLTA.........................VRIFAKAQ.HKVNDASIFSGWKACQTLALMHGYIILVLGPV.KK
210K KYMV       SRSDPDWRHTA.........................VKIFAKSQ.HKVNDASIFGNWKACQTLALMHDFVILSLGPV.KK
198K OYMV       SRSDPDWRYTV.........................VRIFAKSQ.HKVNDGSIFGSWKACQTLALMHDFVILTLGPV.KK
206K TYMV       SRSDPDWRHTT.........................VKIFAKAQ.HKVNDGSIFGSWKACQTLALMHDYVILVLGPV.KK
186K NMV        SRQSPDFPKNK.........................IALFLKSQWVKKTEKLGTLKVKPGQTIASFMQETVMLYGTM.AR
166K PVX        VRQSPDFDENK.........................IMVFLKSQWVTKVEKLGLPKIKPGQTIAAFYQQTVMLFGTM.AR
147K WClMV      TRQSPDFDSHA.........................IALFLKSQWVKKTEKIGCLKIKAGQTIAAFMQQTVMIYGTM.AR
216K ACLSV      QRSDSDWPLDK.........................IFLFMKSQLCTKFEK.RFTEAKAGQTLACFPHKILVEFSPW.CR
223K PVM        GRSCRDWPTDV.........................AQIFSKSQLCTKFDN.RFRVAKAAQSIVCFQHAVLCRFAPY.MR
2a   AlMV*     YSDPLGVRSMDS........................YKHMIKSVLKPVEDNSLHLERPMPATITYHDKDIVMSSSPIFLA
237K BNYVV      FSASNLWRIMDKAMHDMVAKNYQGQMEEEFTRNAKLYRFQLKDIEKPLKDPETDLAKAG.QGILAWSKEAHVKFMVAFRV
NSP SINDBIS*   FAKTYNLVPLQEVPMDR...................FVMDMKRDVKVTPGTKHTEERPKVQVIQAAEPLATAYLCGIHRE
126K TMV        F.DFVDLPAVDQ.........................YRHMIKAQPKQKLDTSIQTEYPALQTIVYHSKKINAIFGPLFSE
  Consensus     -R--PDW-----------------------------V--F-KSQ-----D---F--WKA-QTIA-----VIL-LGP----

                2161                                                                2240
194K ELV        YQRLFDQLERPSHIYYHAGNTPHDLRRWCS...KHLE.TSHCTTNDYTAFDQSQHGEAVVFEVLKMRRLSIPENLI....
205K EMV        YQRIFDSKDRPPHIYYHCGKTPSQLSQWCQ...THLS.GSSYIANDYTAFDQSQHGEAVVLECLKMRRLSIPDSLI....
210K KYMV       YQRIFDALDRPPHLYTHCGKSPADLSAWCQ...THLT.GQIKLTNDYTAFDQSQHGESVILEALKMKRLSIPSHLI....
198K OYMV       YQRIIDHYDRPNFIYTHCGKTPSELSAWSH...SFLK.GDAYICNDYTSFDQSQHGEAVIFESLKMHRVGIPRHLI....
206K TYMV       YQRIFDNVDRPSHIYSHCGKTPNQLRDWCQ...EHLTHSTPKIANDYTAFDQSQHGESVVLEALKMKRLNIPSHLI....
186K NMV        YLRKMRRRFQPDNIFI.CETPPEDLDKFIK...SQWDFSRPAHTNDFTAFDQSQDGAMLQFEVIKAKFFNIPAEMI....
166K PVX        YMRWFRQAFQPKEVFINCETTPEDMSVWAL...NNWNFSRPSLANDYTAFDQSQDGAMLQFEVLKAKHHCIPEEII....
147K WClMV      YMRKFRNQYCPRKIFVNCETTPADFNSFIL...DEWNFNRTCFSNDFTAFDQSQDGSILQFEVIKAKFHNIPEDII....
216K ACLSV      YTEKVLTANLPDNYYIHQRKNFSELEDFA....RRFSNGSICVESDYTAFDVSQDHTILAFEVELLRHFGWDDRVL....
223K PVM        YIEMKVHEVLPKNYYIHSGKGLEELDAWV....KKGKFDRICTESDYEAFDASQDEFIMAFELELMKYLRLPSDLI....
2a   AlMV*     AAARLMLILRDK.....ITIPSGKFHQLFSIDAEAFDASFHFKEIDFSKFDKSQNELHHLIQERFLKYLGIPNEFL.TLW
237K BNYVV      LNDLLLKSLNSN.VVYDNTMSETEFVGKINAAMNTVPDSAINGVIDAAACDSGQGVFTQLIERHIYAALGISDFFL..DW
NSP SINDBIS*   LVRRLTAVLLPN.IHTLFDMSAEDFDAIIA...EHFKQGDPVLETDIASFDKSQDDAMALTGLMILEDLGVDQPLL.DLI
126K TMV        LTRQLLDSVDSSRFLFFTRKTPAQIEDFFG.DLDS.HVPMDVLELDISKYDKSQNEFHCAVEYEIWRRLGF.EDFLGEVW
  Consensus     Y-R-F-----P--IY------P-DL--W-------L--------NDYTAFDQSQ-G----FE--KM--L-IP--LI----
```

```
                2241                                                                                                2320
194K ELV        ...SLHVHLKTNVETQFGPLTCMRLTGEPGTYDDNTDYNLAVLNLQYDL.RKTPT..LVSGDDSYLSGTLSPRSNWPFVK
205K EMV        ...QLHSHLKCSVDTQFGPLTCMRLTGEPGTYDDNSDYNLAVIYSQYSL.NGHPI..LISGDDSVLCGTPPPSPLWPTLK
210K KYMV       ...QLHVHLKTNVATQFGPLTCMRLTGEPGTYDDNSDYNLAVIHSQFDM.KDIPV..MVSGDDSLIDRQPPLAQSWEATK
198K OYMV       ...DLHIYLKTNVSTQFGPLTCMRLTGEPGTYDDNTDYNLAVIFSQYVI.SDHPI..MVSGDDSVICGHPPINPNWPAVE
206K TYMV       ...QLHVHLKTNVSTQFGPLTCMRLTGEPGTYDDNTDYNLAVIYSQYDV.GSCPI..MVSGDDSLIDHPLPTRHDWPSVL
186K NMV        ...EGYIYIKLNAAIFLGTLGIMRLSGEGPTFDANTECSIAYNATRFHITDDTAQ..VYAGDDMALDRVSIEKDSFNRLE
166K PVX        ...QAYIDIKTNAQIFLGTLSIMRLTGEGPTFDANTECNIAYTHTKFDIPAGSAQ..VYAGDDSALDCVPEVKHSFHRLE
147K WClMV      ...EGYIQIKTHAKIFLGTLSIMRLSGEGPTFDANTEANIAYTHTKFNIPCDAAQ..VYAGDDMSIDYVASVKPSFNMIE
216K ACLSV      ...QSYIKMKCTLGCRLGGFAIMRFTGEFSTFLFNTLANMVFTFCRYEVPDGTPI..CFAGDDMCALRNLREIDTHEFIL
223K PVM        ...EDYKFIKTSLGSKLGNFAIMRFSGEASTFLFNTLANMLFTFMRYNIRGDEFI..CFAGDDMCASRRLQPTKKFAHFL
2a    AlMV*     FNAHRKSRISDSKNGVFFNVDFQRRTGDALTYLGNTIVTLACLCHVYDLMDPNVKFVVASGDDSLI.GT..VEELPRDQE
237K BNYVV      YFSFREKYVMQSRY.VRAHMSYVKTSGEPGTLLGNTILMGAMLNAMLRGTGPFCM..AMKGDDG....FKRQANLKINDQ
NSP SINDBIS*    ECAFGEISSTHLPTGTRFKFGAMMKSGMFLTLFVNTVLNVVIASRVLEERLKTSRCAAFIGDDNIIHGVVSDKEMAERCA
126K TMV        KQGHRKTTLKDYTAGIKTCIWYQRKSGDVTTFIGNTVIIAACLASMLP.MEKIIK.GAFCGDDSLLY.FPKGCEFPDVQH
  Consensus     ------I--K------FG-L-CMRLTGE--T-D-NTD-NLA-----Y-------------GDD------------W----

                2321                                                                                                2400
194K ELV        ELLHLRLKPSSLID.....GL...FCGYYLG.....PQGCIRNPLALFAKLMIAEDDGSAFDKLPSYLTEFSIGHGLGDS
205K EMV        KMLHLRFKIERTSH.....PL...FCGYYVS.....PHGAARNPYALFAKLMICVDDKSLHDKKLSYLSEFSTGHLAGDL
210K KYMV       RLLHLRFKTEKTTH.....PL...FCGYYTG.....SAGAIRNPLALFSKLMIAIDDEAIHDRRLSYLTEFSTGHQLGDA
198K OYMV       KLLHLRFKTEETSL.....PL...FCGYYVG.....PTGCCRNPFALFAKLMISYDKGNLFETLPSYLYEFSIGHRLGDV
206K TYMV       KRLHLRFKLELTSH.....PL...FCGYYVG.....PAGCIRNPLALFCKLMIAVDDDALDDRRLSYLTEFTTGHLLGES
186K NMV        KQLKLTSKPMFPKQVKGDYAE...FCGWVMT.....PAGIIKHSLKMHASIQLQKKINNIKESARSYALDLRYAYKLGDE
166K PVX        DKLLLKSKPVITQQKKGSWPE...FCGWLIT.....PKGVMKDPIKLHVSLKLAEAKGELKKCQDSYEIDLSYAYDHKDS
147K WClMV      HLMKLKGKPVFNTQTQGDFAE...FCGWTIS.....PKGIIKKPEKMNMSIELQKNINKFHEVKRSYALDHAFAYQLGDE
216K ACLSV      SKLSLKAKVNRTKV.....PM...FCGWRLC.....CDGLIKEPCLIYERLQVAIENGRLMDVIDSYFLEFSFAYKLGER
223K PVM        DKLKLKAKVQFVQFV..NKPT...FCGWHLC.....PDGIYKKPQLVLERMCIAKEMNNLSNCIDNYAIEVAYAYKLGEK
2a    AlMV*     FLFTTLFNLEAK.FPH.NQPF...ICSKFLIT..MPTTSGGKVVLPIPNPLKLLIRLGSKKVNADIFDEWYQSWIDIIGG
237K BNYVV      ML.KLIKKETVLDFKL.DLNVPITFCGYALSNGHLFPSVSRKLTKIAAHRFREYKHFCEYQES...LRDWIKNLPKDPAV
NSP SINDBIS*    TWLNMEVKIIDAVIGE.RPPY...FCGGFIL.....QDSVTSTACRVADPLKRLFKLGK.................PLPAD
126K TMV        SA.NLMWNFEAKLFKK.QYGY...FCGRYVI.......HHDRGCIVYYDPLKLISKLGAKHIK.....DW.....EHLEE
  Consensus     --L-L--K-----------PL---FCGYY-------P-G----P--LF--L--------L-D---SY--EF-----LGD-

                2401                                                                                                2480
194K ELV        LWQLLPSDLVLYQSACFDYFCRKA.TRSQKILLQPGLVDQETLDKIALSAKFISRPFYSMLSSHARSLISTKFKLDSSLT
205K EMV        VTSILPSHLLPYQSAVHDFFCRNC.TPAEKILLSLDPIPESKILQLILKVRWASQAFFSYLPQKARELLVARSSLPSLYS
210K KYMV       LWTLLPESTQIYQSACFDYFCRHS.PPHEKALLSSFELPDSVISKISSSTKWLSKNAFYALPSKIRKAVIASRHSSSFPE
198K OYMV       VRLLFPDHLLKYYSACWDLFCRKC.TASQKLILSFEPIPPSFFSKLASTSRWVSKVLFSDLPTKIRDMLISSSKLPSYHQ
206K TYMV       LWHLLPETHVQYQSACFDFFCRRC.PKHEKMLLDDSTPTLSLLERITSSPRWLTKNAMYLLPAKLRLAITSLSQTQSFPE
186K NMV        LQEHLNEVEA....DYHQQSVRDMHLLHQQDVLNKGSASPPHVFEKTADANTAGSSKTHKRNALKKKKQTRIAEILPSPD
166K PVX        LHDLFDEKQC....QAHTLTCR.....................................TLIKSGRGTV..SLP...
147K WClMV      LHELYNESEA....EHHQLATR.....................................SLILAGQATALDILD...
216K ACLSV      LYSHLEIEQL....NYHQVLTRFF.................................IRNKHLLRGDSRHNISELEWLS
223K PVM        AVNRMDEEEV...AAFYNCV.RII.................................VRNKHLIRSDVKQVFEVL....
2a    Almv*     FNDHH.VIRCVAAMTAHRYLRRP......SLYLEAALESLG....KIFAGKTLCKECL..FNEKHESNVKIKPRRVKKSH
237K BNYVV      YADFLECNASLSCRNVDDVQRWL......DAIISVSRIGRE....Q.FMMMFPIREVFMSLPPVEDSLGELSSTKVAVSI
NSP SINDBIS*    DEQDEDRRRALLDETKAWFRVGI......TGTLAVAVTTR.......YEVDN.ITPVLLALRTFAQSKRAFQAIRGEIKH
126K TMV        FRRSL.CDVAVSLNNCAYYTQLD......DAVWEVHKTAPP....GSFVYKSLVK..Y..LSDK....VLFRSLFIDGSS
  Consensus     L---L----------C--Y--R----------L-----------------------F--L-----------------

                                              C
                2481                                2511
194K ELV        TLQDPMVEFELL.PFSNVQ.............
205K EMV        NPKVSQLESELL.PFSQ...............
210K KYMV       NPDVSQLEFELL.QSFQF..............
198K OYMV       DPRVQYLESELLTSFNHGRLSTN.........
206K TYMV       SIEVSHAESELL.HYVQ...............
186K NMV        ATGLSSLPFRFF....................
166K PVX        ..RLRNFL........................
147K WClMV      .YGLRDLK........................
216K ACLSV      DEDGDNDKGSQIEDRRRGYSMCWGEKLQNLF
223K PVM        ................................
2a    Almv*     SDARSRARRA......................
237K BNYVV      GDNVSNVVRKVARVDMKKF.............
NSP SINDBIS*    LYGGPK..........................
126K TMV        C...............................
  Consensus     --------------------------------
```

---

* AlMV 1a = RNA1 was aligned in A and B segments.
AlMV 2a = RNA2 was aligned in C segment.
Sindbis virus NSP = nsP1-nsP2 (amino acid residues 1-1463)
and nsP4 (residues 1907-2513)

Table 9. Percent of identical residues in the aligned replicase proteins encoded by five tymoviruses and other Sindbis-like viruses. Acronyms of tymoviruses as in text. Those of other viruses are ACLSV: apple chlorotic leaf spot closterovirus, AlMV: alfalfa mosaic virus, BNYVV: beet necrotic yellow vein furovirus, NMV: narcissus mosaic potexvirus, PVM: potato M carlavirus, PVX: potato X potexvirus, Sindbis alphavirus, TMV: tobacco mosaic tobamovirus, and WClMV: white clover mosaic potexvirus.

	Domain	ELV	EMV	KYMV	OYMV	TYMV	NMV	PVX	WClMV	ACLSV	PVM	AlMV	BNYVV	Sindbis
EMV	A	45.00												
	B	45.02												
	C	58.88												
	total	**49.63**												
KYMV	A	44.89	50.44											
	B	45.42	45.66											
	C	56.77	61.80											
	total	**49.03**	**52.63**											
OYMV	A	46.74	53.85	50.18										
	B	44.03	47.77	47.45										
	C	56.67	63.20	59.68										
	total	**49.15**	**54.94**	**52.44**										
TYMV	A	57.92	60.04	52.80	57.92									
	B	43.42	49.64	43.83	49.26									
	C	44.93	41.96	67.55	40.53									
	total	**48.76**	**50.55**	**54.73**	**49.15**									
NMV	A	18.76	20.22	20.18	17.67	23.31								
	B	18.02	17.54	17.50	17.92	23.04								
	C	22.97	19.50	23.21	21.66	21.25								
	total	**19.92**	**19.09**	**20.30**	**19.08**	**22.53**								
PVX	A	18.96	20.04	17.19	18.86	25.59	37.02							
	B	21.67	20.54	19.20	21.62	20.44	41.48							
	C	25.43	23.72	25.70	25.70	18.24	54.12							
	total	**22.02**	**21.43**	**20.70**	**22.06**	**21.42**	**44.21**							
WClMV	A	22.27	22.20	23.04	21.91	23.31	52.15	43.76						
	B	23.77	21.58	21.56	23.43	23.04	49.87	50.94						
	C	23.57	21.02	24.69	23.83	21.25	58.61	55.04						
	total	**23.20**	**21.60**	**23.09**	**23.06**	**22.53**	**53.54**	**49.91**						
ACLSV	A	17.76	17.14	18.82	16.96	21.94	15.96	14.26	18.71					
	B	20.21	16.25	18.15	17.64	16.23	13.79	18.08	19.46					
	C	21.86	24.08	20.97	22.07	19.02	25.05	26.39	25.37					
	total	**19.94**	**19.16**	**19.13**	**18.89**	**19.06**	**18.27**	**19.58**	**21.18**					
PVM	A	15.74	15.75	14.09	17.18	21.67	15.52	15.26	18.04	29.95				
	B	13.93	12.28	11.83	14.06	13.89	13.44	17.23	16.49	14.08				
	C	22.80	20.22	20.69	23.49	14.31	23.71	25.11	25.00	43.25				
	total	**17.49**	**16.08**	**15.53**	**18.24**	**16.62**	**17.56**	**19.20**	**19.84**	**29.09**				
AlMV	A	06.67	08.23	05.92	07.11	12.55	07.54	08.47	09.84	08.30	07.11			
	B	12.21	11.23	10.45	11.80	06.26	10.37	11.60	10.42	10.12	09.24			
	C	12.76	11.43	11.83	10.69	10.66	11.81	12.80	10.31	11.39	13.77			
	total	**10.55**	**10.30**	**09.40**	**09.87**	**09.82**	**09.91**	**10.96**	**10.19**	**09.94**	**15.39**			
BNYVV	A	07.43	07.10	07.29	07.29	11.15	06.57	07.14	07.45	09.04	07.45	09.94		
	B	10.28	10.34	09.61	10.22	07.45	08.92	09.96	10.30	07.60	08.24	08.10		
	C	12.93	12.98	11.52	12.17	08.52	10.74	10.15	10.16	12.33	12.01	17.35		
	total	**10.21**	**10.14**	**09.47**	**09.78**	**09.04**	**08.74**	**09.08**	**09.30**	**09.66**	**09.23**	**11.80**		
Sindbis	A	09.14	08.01	08.64	08.91	11.57	07.27	08.86	08.10	09.21	07.18	17.04	10.39	
	B	09.98	08.12	07.90	10.12	07.77	08.10	06.92	08.36	06.13	07.23	22.95	05.46	
	C	10.44	10.82	11.60	12.17	08.06	11.29	12.33	11.16	12.42	11.53	26.29	14.12	
	total	**09.85**	**08.98**	**09.38**	**10.40**	**09.13**	**08.89**	**09.37**	**09.21**	**09.25**	**08.65**	**22.09**	**09.99**	
TMV	A	09.17	05.74	08.19	07.17	16.74	09.03	05.92	06.91	07.16	08.65	19.02	08.54	12.76
	B	09.29	09.17	09.59	10.14	08.17	08.93	09.57	12.56	10.00	07.32	13.73	06.94	13.07
	C	15.23	16.34	17.66	16.11	10.28	13.54	12.14	08.62	15.86	13.48	13.43	19.10	17.14
	total	**11.23**	**10.42**	**11.81**	**11.14**	**11.73**	**10.50**	**09.21**	**09.36**	**11.00**	**09.82**	**15.39**	**11.53**	**14.32**

| | Domain | ELV | EMV | KYMV | OYMV | TYMV | NMV | PVX | WClMV | ACLSV | PVM | AlMV | BNYVV | Sindbis |

The A Segments

The possible methyltransferase domain was first identified in the Sindbis virus RP (Mi *et al.*, 1989). Candresse *et al.* (1990) reported that the most characteristic signature of the RNA methyl transferase is the consensus sequence Hxxxxx <19-22aa> xxGx <11-13aa> xHxxxxxxxxxDxxRx <22-53aa> xCxxxxxxx. However the 'Sindbis-like' viruses of subgroup I (i.e. the tymo-, potex-, carla- and closteroviruses) seem only to have traces of the two motifs closest to the N-terminus. These traces are found in the ELV RP from around amino acid residue 80 to about 440 (Figure 27).

The B Segments

The second conserved segment is located between residues 870 and 1101 in the central part of the ELV RP, and possesses the well-known sequence GxxGxGKS/T, that has been shown to be associated with nucleotide binding activity (Hodgman, 1988; Gorbalenya *et al.*, 1988b; Gorbalenya *et al.*, 1989a, b). It can be seen that when those portions of the 'Sindbis-like' RPs that contain this sequence are aligned, eight conserved blocks of sequence are found. These are xxxxxGxxGxGK[S/T]xxxx[I/f]xx xxx <4-14aa> xxxx[P/d/s]xxx[L/a][R/a/k] <21-37aa> xxxxxD[E/d]xxxxxx[G/r]xxxx <6-10aa> xxxxxxxG[D/g]xxQxxx <8-25aa> xxx[R/e]xxxxx <1-18aa> x[R/y]x <43-60aa> xxxxTxxxx[Q/r/t]GxTxxxxxxxx[L/i/q] <9-17aa> xxxx[A/c/l/v][L/i]x[R/a] xxxxxxxx (Figure 28).

The C Segments

The third set of motifs are those of the viral RNA polymerase located at the C-terminal end of the RPs. Those of the ELV RP are around amino acid residues 1475 to 1620 and consist of at least five conserved blocks xxxxDxxx[F/y/c]DxSQxxxxxxxxx <23-29aa> xxxxxxxxxxxxGxxxTxxxN[T/s]xxxxxxxxxxxx <5-6aa> xxxxxGDDxxxxxx <8aa> xxxxxxxxxxKxx <1-6aa> xxx[F/i]C[G/s]xxxxxx (Figure 29). However Poch *et al.* (1989) found a larger region of the RPs of viruses with plus-strand, minus-strand, and double-stranded RNA genomes was conserved and this region contained at least four motifs.

Figure 27. The motifs at N-terminus of the RPs of tymoviruses and of the other corresponding proteins of Sindbis-like viruses. The consensus line indicates the residues which most commonly occur in that position. These motifs are possibly of a methyltransferase (Mi *et al.*, 1989). The domains of the methyltransferase identified by Candresse *et al.* (1990) are the lower set of sequences. Numbers in the gaps between motifs gives the numbers of amino acids within each gap.

194K ELV	49	LNRMGIQVSGLT	<4aa>	HPHAAHKTLELNLL	<13aa>	V..VFMKPSKFFKL
205K EMV	49	LLSWGIPNSGLG	<4aa>	HPHPIHKTVETFLL	<13aa>	V..MFMKPSKFQKL
210K KYMV	49	LLKSGIPINSFL	<4aa>	HPHPAHKTLETHLL	<13aa>	V..LFMKPQKFMKL
198K OYMV	49	LLAAGIPTSGFG	<4aa>	HPHAVHKVIETFLL	<13aa>	V..MFMKPSKFKKL
206K TYMV	49	LNSYGIPTSGLG	<4aa>	HPHAAHKTIETFLL	<13aa>	V..MFMKPSKFNKL
186K NMV	50	LESFGITVNPYA	<4aa>	HTHAACKVIENRML	<13aa>	M..LFLKRSKLRYM
166K PVX	50	LEGFGIATNPYS	<4aa>	HTHAAAKTIENKLL	<13aa>	F..MFLKPRKLNYM
147K WClMV	50	LEKYGIATNPFA	<4aa>	HSHGAVKSIENTLL	<13aa>	F..LFLKRSKLRYL
216K ACLSV	54	FTKSGVYLSPFA	<4aa>	HSHPGCKTLENHLL	<13aa>	VACLSIKSNKMSKM
223K PVM	54	LSMAGIYLSPYS	<4aa>	HSHPVCKTLENYIL	<13aa>	F..VGIKERKL.QL
Consensus		**L---GI------**		**H-H---K--E--LL**		**----F-K--K----**
		F V		**I M**		**G S**

Methyltransferase	1a	AlMV	99	HCFAAA	<19aa>	IDLGG	<12aa>
consensus	237K	BNYVV	219	HPVLAA	<22aa>	LVVGA	<13aa>
(Candresse et al. 1990)	nsP1	SINDBIS	38	HANARA	<19aa>	LDIGS	<11aa>
	126K	TMV	81	HSLAGG	<19aa>	YDIGG	<12aa>

<9aa>	LHNYRLHPHDSNRY.P	<15aa>	DSLMYFTPHQISGLFESCPNLLSLYASLVVPPESSMT.DLSLNPDLYR
<9aa>	LINFRLTAADTTRY.P	<15aa>	DALMYFSPAQIVDLFTQSPALETLYCSLIVPPESHFT.DLSLFPEIYT
<9aa>	LHNYRLTPTDSVRF.P	<15aa>	DALMYYQPEQILHLFHQVPQLTNLFCSLVTPPESHFT.HLSLMPDLYT
<9aa>	LVNYRLTAADSVRY.P	<15aa>	DALMYFNPSQILDLFIQCPSLQRLHCSLVVPPESSFT.DLSLHPNLYT
<9aa>	LKNYRLHPNDSTRY.P	<15aa>	DALMYYHPSQIMDLFLQKPNLERLYASLVVPPEAHLS.DQSFFPKLYT
<9aa>	FLNKDVEPKDLFRY.D	<19aa>	DTLHFMSRREIVQLFEDSPKLETLLATVVLPVEALHK.RTSLYPSLYS
<9aa>	FHNVAIEPRDVARY.P	<19aa>	DTVHSLDPSYIVETFQNCPKLQTLYATLGLPVEAAFK.MESTHPNLYS
<9aa>	FINLAIEPRDLQRY.E	<19aa>	DTLHFFTRKMLADLFFHNPALDVLYATLVLPPEALHK.HPSIEPDLYT
<12aa>	ILNRLVTAKDKARYGP	<18aa>	DEIHYWSRDQLET.FLQVHRPKNLWATLVFPPEILAGYKSSVLPFLYQ
<12aa>	VVNRYVTSADRMRYTN	<39aa>	DELHYWSSKVLID.FLDVMRPDKLLGTVVYPPELLFKPTRSLNEWCYT
	--N------D--RY--		**D----------LF--------L---L--PPE-------S--P--Y-**
	F		**T V V E**

1a	AlMV	VHCCCPILDARDGARL	<22aa>	YCMDTFQKC	189
237K	BNYVV	VHYYFANKDSKDLVRT	<28aa>	GCGYIVKRS	319
nsP1	SINDBIS	YHCVCPMRSPEDPDRM	<39aa>	LCFHNDVTC	145
126K	TMV	VHCCMPNLDVRDIMRH	<39aa>	VCHNTFQTM	188

<7aa>	LHYTPEGHSAGSYNQPVNALDWLKISAI	<11aa>	LESWGPLHSLLI
<7aa>	LHYIPENHHSGSYNQPLQALSWLKISSI	<11aa>	LESWGPVHSILI
<7aa>	LHYTPEGHSAGSYNQPITALSWLKINSI	<11aa>	LESWGPLHSILI
<7aa>	LHYVPEGHHAGSYDQPLDAISWLKLNSI	<11aa>	LESWGPVHSLLI
<7aa>	LHYVPEGHEAGSYNQPSDAHSWLRINSI	<11aa>	LESWGPVHSLLI
<7aa>	FEYIPGNHGGGCYFHPYTTLEWLKVRMI	<18aa>	VESLGANHLFIV
<7aa>	FQYIPGNHGGGAYHHEFAHLQWLKVGKI	<29aa>	QESFAANHLYCI
<7aa>	FQYIPGNHGGGSYSHEFKQLEWLKVGHL	<11aa>	IESIGANHLFMI
<7aa>	LVYMPDGVRSESYTQPLENGFLLSSSSI	<17aa>	VYSLGSHHLFHI
<7aa>	LMFFPDGVQSEGYQQPLKGGYLLGARSL	<12aa>	LCSKFPHHLISI
	--Y-P-------Y---------L----I		**--S----H---I**
	F L		**V**

Figure 28. Eight amino acid sequence motifs in the possible NTP-binding part of the RPs of five tymoviruses and of the corresponding region of Sindbis-like viruses. This domain is similar to that of the nucleic acid helicase of *E. coli* (rec B) (Habili and Symons (1989). The number of amino acids in the gap between the motifs is given. The consensus indicates residues which occur most commonly.

194K	ELV	868	LIHLSGFAGCGK.THP	I	QQLLK	<7aa>	RVVTPTTNLR		
205K	EMV	959	LIHIAGFAGCGK.THP	I	QKLLQ	<7aa>	RISCPTNELR		
210K	KYMV	996	MFHLAGFAGCGK.TKP	I	QSLLS	<7aa>	RVSTPTTELR		
198K	OYMV	893	VIHIAGFPGCGK.SHP	I	QKLLQ	<7aa>	RLSVPTNELR		
206K	TYMV	970	VVHFAGFAGCGK.TYP	I	QQLLK	<7aa>	RVSCPTTELR		
186K	NMV	862	GTIIHGCGGSGKGSFA	I	QEWMR	<9aa>	TVVTPTVLLR		
166K	PVX	729	TCVIHGAGGSGK.SHA	I	QKALR	<8aa>	TVVLPTNELR		
147K	WClMV	564	MSVIHGAGGSGK.SHA	I	QTWMR	<9aa>	TIILPTTDLR		
216K	ACLSV	1053	IYGIFGFAGSGK.SHA	I	QNLIQ	<9aa>	MVICPRRFLA		
223K	PVM	1160	LHAIVGTFGSGK.STL	F	KNLLK	<7aa>	DFVSPRRALA		
1a	AlMV	832	TIIVDGVAGCGK.TTN	I	KQIAR	<12aa>	SNRSSADELK		
237K	BNYVV	936	LEYVKGGPGTGK.SFL	I	RSLAD	<4aa>	LVVAPFIKLR		
NSP2	SINDBIS	720	TIGVIGTPGSGK.SAI	I	KSTVT	<14aa>	EIEADVLRLR		
126K	TMV	830	VVLVDGVPGCGK.TKE	I	LSRVN	<4aa>	LILVPGKQAA		
Consensus			-----G--G-GK----I-----	**F**			----P---LR **D AA** **S K**		
<28aa>	EILVI	DE	IYKLPRGYLDL	<7aa>	VKL.ILL	GDPLQGEY	<14aa>	EIPRLLPFI	
<28aa>	EILVI	DE	IYKLPRGYLDL	<7aa>	LSLVIIL	GDPLQGEY	<14aa>	EVHRFKSYI	
<28aa>	KILVI	DE	IYKLPRGYLDL	<7aa>	LELVIIL	GDPLQGEY	<14aa>	ETTRLLPFI	
<28aa>	NILVV	DE	IYKLPRGYLDL	<7aa>	IQLVIML	GDPLQGEY	<14aa>	ETTRLSKYI	
<28aa>	RILVI	DE	IYKMPRGYLDL	<7aa>	LELVIIL	GDPLQGEY	<14aa>	ETLRLLPYI	
<27aa>	PILVF	DD	YTKLPPGLIES	<7aa>	VVFIILT	GDNRQSVY	<20aa>	PYCEFYLNA	
<27aa>	SIVIF	DD	YSKLPPGYIEA	<7aa>	IKLIILT	GDSRQSVY	<20aa>	KYCRYYLNA	
<27aa>	KIIVF	DD	YSKLPQGYIEA	<7aa>	VILAILT	GDSKQSFH	<20aa>	PFCRYYLNI	
<27aa>	RLFIL	DE	ISLLPKGFTDL	<10aa>	LKTIVCI	GGPLQAGY	<13aa>	EIKRLFKGG	
<41aa>	QVVIL	DE	MQLYPPGYFDL	<8aa>	VR.LFLV	GDPAQSDY	<8aa>	LGAREYNYK	
<28aa>	QRLIF	DE	CFLQHAGLVYA	<6aa>	CSEVIGF	GDTEQIPF	<22aa>	ITWRSPADA	
<29aa>	QIIFV	DE	FTAYDWRLLAV	<6aa>	AHTIYLV	GDEQQTGI	<17aa>	MNFRNPVHD	
<21aa>	EVLYV	DE	AFACHAGALLA	<8aa>	RKKVVIC	GDPMQCGF	<24aa>	ISRRCTQPV	
<37aa>	R.LFI	DE	GLMLHTGCVNF	<6aa>	CEIAYVY	GDTQQIPY	<25aa>	TTLRCPADV	
		-----DE------G---- **D R**			-------GD--Q--- **G**			---R----- **E**	
<7aa>	YRV	<47aa>	WNAVTISSSQGCTFSDPAFIHLD				<9aa>	NCLVALTRSRSGVYFK	
<7aa>	HRI	<47aa>	HHAITISARR.VTFTEAHTILLD				<9aa>	NCLVALTRSRTGVYFV	
<7aa>	YRV	<51aa>	FPATTISASQGVTHHNRVTILLD				<9aa>	NTLVALTRSTVGVEFL	
<7aa>	YRC	<47aa>	HHALTISSSQGMTYSDPVTILLD				<9aa>	TALVALTRSRSGIYFI	
<7aa>	YRI	<46aa>	YRSCTISSSQGLTFCDPAIIVLD				<9aa>	NGLVALTRSRSGVQFM	
<1aa>	HRN	<49aa>	HHSMTYAGCQGLT.APKIQI.L.				<9aa>	VLYTCLSRAVDRIHFI	
<1aa>	HRN	<48aa>	NDRFTYAGCQGLT.NPKVQIVLD				<9aa>	VMYTALSRATDRIHFV	
<1aa>	HRN	<47aa>	QKSMTYAGCQGLT.TKAVQILLD				<9aa>	VIYTALSRAVDHIHFI	
<7aa>	YRI	<60aa>	GNVMTFGESQGLTF.NCGVIVLS				<9aa>	HIMVAITRFRRGFCFA	
<3aa>	HRF	<61aa>	AKVLTFGESTGLTF.MHGTIYIS				<9aa>	RWITALRRFRFNLCFV	
<18aa>	SRV	<43aa>	DNIFTTHEAQGKTFDNVYFCRLT				<17aa>	NGLVALSRHKKTFKYF	
<9aa>	SRN	<46aa>	VSKTTVRANQGSTYDNVVLPVLP				<10aa>	LNLVALSRHNKLTILQ	
<7aa>	HYD	<46aa>	HEVMTAAASQGLTRKGVYARVQK				<12aa>	HVNKLLTRTEDRLVWK	
<5aa>	RRY	<51aa>	SDVHTVHEVQGETYSDVSLVRLT				<9aa>	DSPHVLVALSRHTCSL	
		-R- **Y**		----T----QG-T--------L- **R I** **T Q**				----AL-R-------- **CI A** **L** **V**	

Figure 29. Five amino acid sequence motifs in the RNA polymerase region of the RPs of five tymoviruses and of the corresponding region of Sindbis-like viruses. This domain is located near the C-terminus of the RPs. The number of amino acids in the gap between the motifs is given. The consensus indicates residues which occur most commonly.

194K	ELV	1474	CTTNDYTAFDQSQHGEAVVFEVL	<23aa>	TQFGPLTCMRLTGEPGTYDDNTDYNLAVLNLQY
205K	EMV	1568	YIANDYTAFDQSQHGEAVVLECL	<23aa>	TQFGPLTCMRLTGEPGTYDDNSDYNLAVIYSQY
210K	KYMV	1603	KLTNDYTAFDQSQHGESVILEAL	<23aa>	TQFGPLTCMRLTGEPGTYDDNSDYNLAVIHSQF
198K	OYMV	1498	YICNDYTSFDQSQHGEAVIFESL	<23aa>	TQFGPLTCMRLTGEPGTYDDNTDYNLAVIFSQY
206K	TYMV	1573	KIANDYTAFDQSQHGESVVLEAL	<23aa>	TQFGPLTCMRLTGEPGTYDDNTDYNLAVIYSQY
186K	NMV	1373	AHTNDFTAFDQSQDGAMLQFEVI	<23aa>	IFLGTLGIMRLSGEGPTFDANTECSIAYNATRF
147K	WClMV	1072	SLANDYTAFDQSQDGAMLQFEVL	<23aa>	IFLGTLSIMRLTGEGPTFDANTECNIAYTHTKF
166K	PVX	1237	CFSNDFTAFDQSQDGSILQFEVI	<23aa>	IFLGTLSIMRLSGEGPTFDANTEANIAYTHTKF
216K	ACLSV	1634	CVESDYTAFDVSQDHTILAFEVE	<23aa>	CRLGGFAIMRFTGEFSTFLFNTLANMVFTFCRY
223K	PVM	1750	CTESDYEAFDASQDEFIMAFELE	<23aa>	SKLGNFAIMRFSGEASTFLFNTLANMLFTFMRY
2a	AlMV	525	FKEIDFSKFDKSQNELHHLIQER	<29aa>	GVFFNVDFQRRTGDALTYLGNTIVTLACLCHVY
237K	BNYVV	1836	NGVIDAAACDSSQGVFTQLIERH	<28aa>	.VRAHMSYVKTSGEPGTLLGNTILMGAMLNAML
nsP4	SINDBIS	2267	VLETDIASFDKSQDDAMALTGLM	<29aa>	GTRFKFGAMMKSGMFLTLFVNTVLNVVIASRVL
126K	TMV	1381	VLELDISKYDKSQNEFHCAVEYE	<29aa>	GIKTCIWYQRKSGDVTTFIGNTVIIAACLASML
Consensus			**----D----D-SQ----------**		**-----------G---T---N-----------**

<5aa>	PTLVSGDDSYLSGT	<8aa>	FVKELLHLRLKPS	<1aa>	GL.FCGYYLGPQ
<5aa>	PILISGDDSVLCGT	<8aa>	TLKKMLHLRFKIE	<1aa>	PL.FCGYYVSPH
<5aa>	PVMVSGDDSLIDRQ	<8aa>	ATKRLLHLRFKTE	<1aa>	PL.FCGYYTGSA
<5aa>	PIMVSGDDSVICGH	<8aa>	AVEKLLHLRFKTE	<1aa>	PL.FCGYYVGPT
<5aa>	PIMVSGDDSLIDHP	<8aa>	SVLKRLHLRFKLE	<1aa>	PL.FCGYYVGPA
<6aa>	AQVYAGDDMALDRV	<8aa>	RLEKQLKLTSKPM	<6aa>	AE.FCGWVMTPA
<6aa>	AQVYAGDDSALDCV	<8aa>	RLEDKLLLKSKPV	<6aa>	PE.FCGWLITPK
<6aa>	AQVYAGDDMSIDYV	<8aa>	MIEHLMKLKGKPV	<6aa>	AE.FCGWTISPK
<6aa>	PICFAGDDMCALRN	<8aa>	FILSKLSLKAKVN	<1aa>	PM.FCGWRLCCD
<6aa>	FICFAGDDMCASRR	<8aa>	HFLDKLKLKAKVQ	<4aa>	PT.FCGWHLCPD
<6aa>	FVVASGDDSLI.GT	<6aa>	DQEFLFTTLFKLE	<5aa>	PF.ICSKFLITM
<6aa>	CMAMKGDDGFKRQA	<8aa>	INDQMLKLIKKET	<5aa>	PITFCGYALSNG
<6aa>	CAAFIGDDNIIHGV	<8aa>	RCATWLNMEVKII	<5aa>	PY.FCGGFILQD
<5aa>	KGAFCGDDSLLYFP	<8aa>	VQHSA.NLMWKFE	<5aa>	GY.FCGRYVIHH
	-----GDD------		**----------K--**		**---FCG------**
					I S

These alignments resemble, confirm and refine those previously reported (Skryabin *et al.*, 1988; Habili and Symons, 1989; Morozov *et al.*, 1989; Bruenn, 1991; Candresse *et al.*, 1990; German *et al.*, 1990; Poch, *et al.*, 1989). They emphasize the usefulness of the conserved motifs of viral RPs for assessing their relationships, and evolution. The RNA replicase with its very conserved -GDD- motif is shown to be particularly useful for assessing relationships because:

(i) the RNA polymerase signature is the most conserved and is found in all viruses;

(ii) the N-terminal RNA methyltransferase motif seems to be the least conserved, and is thus of limited value for examining the most distant relationships;

(iii) the nucleotide-binding signature is not found in the RPs of all viruses with RNA genomes; it is not found in luteo-, tombus-, diantho-, sobemo- and carmovirus genomes. Furthermore some viral genomes, including those of carla-, potex-, furo-, and hordeiviruses, possess more than one nucleotide-binding domain.

(iv) the parts of the RNA replication protein that are conserved in different tymoviruses, are also those that are found in other viruses.

Comparitive Phylogenetics of the RP motifs

One way to seek evidence of genetic recombination in proteins, especially those with several recognisable and distinct domains, is to classify those parts separately and check whether the resulting classifications are congruent. Therefore each of the three sets of replicase segments (A, B and C) were classified using the 'FJD distance' between each pair of aligned segments (Feng and Doolittle, 1987) to calculate dendrograms by the neighbour-joining method of Saitou and Nei (1987). The dendrograms (Figure 30) are closely similar but not identical; the groupings and the relationships within each group are mostly identical, but there are some differences in the more distant relationships, for example, the relationship of the three segments of BNYVV to those of the other viruses varies, but all agree that, of the viruses studied, BNYVV is the most distant (Candresse *et al.*, 1990). At the moment there is no published method for assessing the statistical significance of the branch topology of neighbour-joining trees, and the observed differences between trees may not be significant. However, it is clear that these analyses give no evidence of recent genetic recombination between viruses of different groups of the 'Sindbis-like' supergroup.

Furthermore, the close similarity between the dendrograms of the A, B and C segments shows that, although there is little obvious methyltransferase signature in the A segments, they contain a phylogenetic signal that is closely similar to that in the B and C segments.

The three segment classifications were combined by calculating mean 'FJD distances' for each pairwise comparison weighted according to the aligned number of amino acids in each segment. These distances gave the combined RP dendrogram shown in Figure 30, which confirms the clustering of 'Sindbis-like' viruses reported by Candresse *et al.* (1990), and this was used in further analyses of the evolution of the 'Sindbis-like' viruses described below.

THE 'SINDBIS-LIKE' VIRUSES, A SUPERGROUP BASED ON REPLICASE SIMILARITIES; CORRELATIONS WITH OTHER FEATURES.

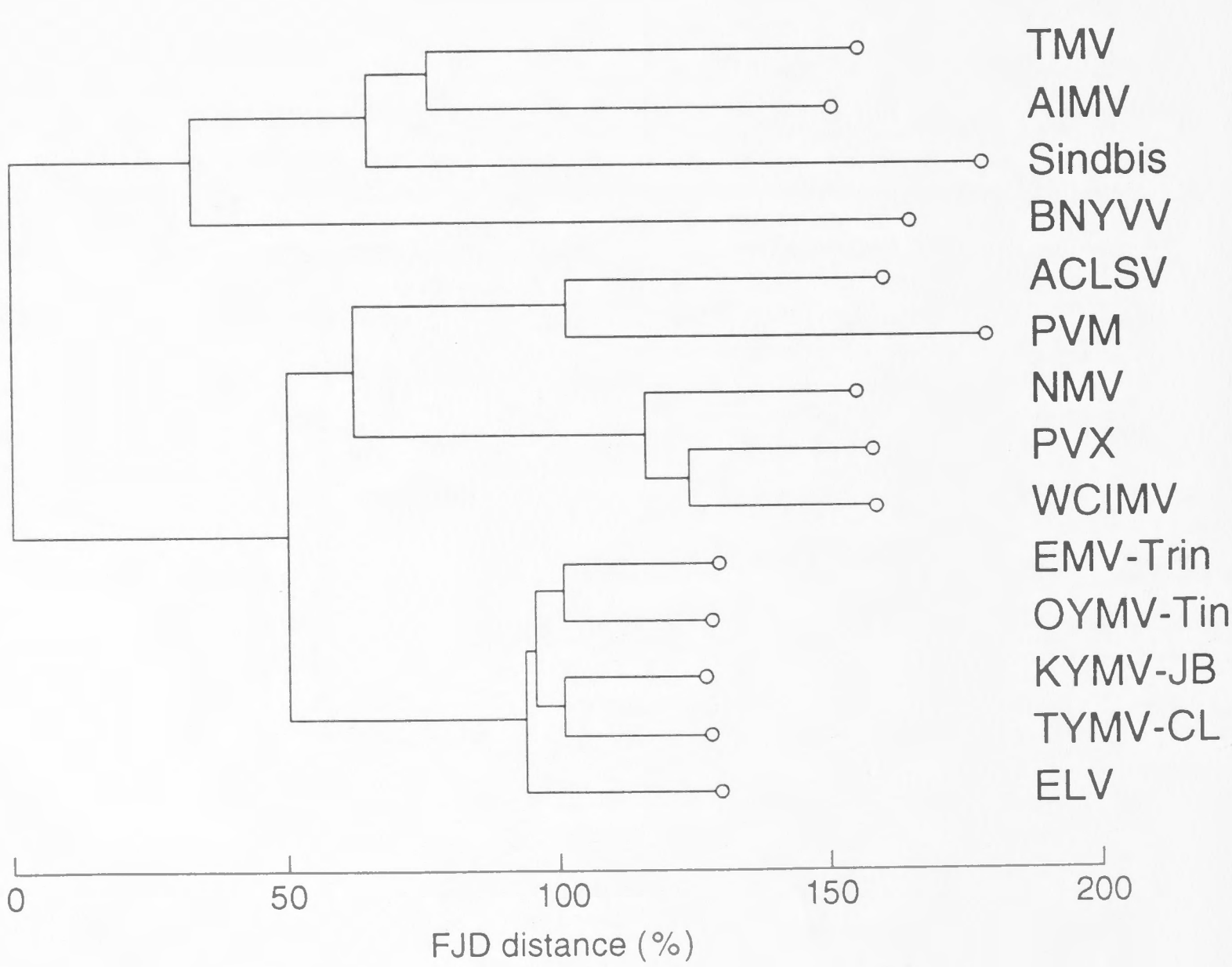
Comparisons of the sequences of the RPs of the tymoviruses and other 'Sindbis-like' viruses have provided a phylogenetic framework to examine and compare some of the other features of these viruses.

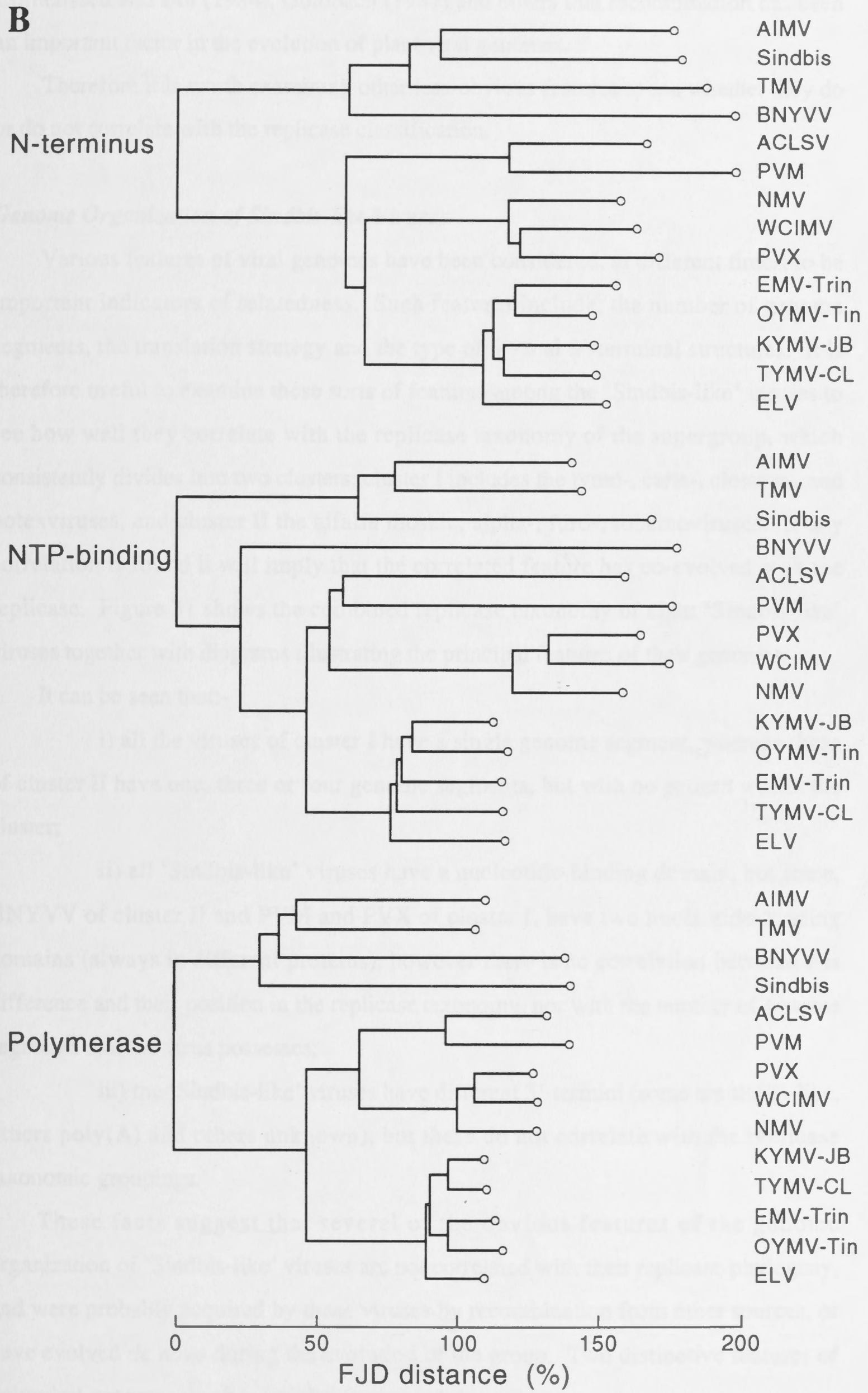
Within the 'Sindbis-like' viruses, the carla-, clostero-, potex- and tymoviruses all have positive-sense ssRNA monopartite genomes ranging from 6-8.5 kb in size. However the morphology of their particles are quite different, those of the tymoviruses are isometric, whereas those of the carla-, clostero- and potexviruses are filamentous. This difference probably reflects a fundamental difference in the virion proteins of these viruses; those of the tymoviruses probably have the familiar 8-stranded β-barrel structure (A.J.Gibbs, unpublished data; Argos, 1981) that is found in most proteins of small isometric virions, whereas the filamentous virions of the other viruses probably have a 4-stranded α-helical structure, similar to that of the tobacco mosaic virus virion protein.

Thus it is likely that the secondary division of the 'Sindbis-like' viruses into tymoviruses on the one hand, and the potex-, carla- and closteroviruses on the other, coincided with a major genetic recombinational event between the replicase protein and virion protein genes from different sources. This confirms the suggestion by

Figure 30. (A) Dendrogram showing the relationship of the RPs of fourteen Sindbis-like viruses. (B) Dendrograms showing the relationships of the A, B and C parts (the N-terminal part, and those containing the NTP-binding domain and RNA polymerase domain respectively). These proteins are the RPs of the tymoviruses, the 186K protein of NMV, the 166K protein of PVX, the 147K protein of WClMV, the 216K protein of ACLSV, the 223K protein of PVM, the 1a and 2a proteins of AlMV, the 237K protein of BNYVV, the NSP protein of Sindbis virus, and the 126K protein of TMV. The neighbour-joining method was used to calculate the dendrogram from FJD distances (Feng *et al.*, 1985) given by the progressive alignment program of Feng and Doolittle (1987).

A




B
N-terminus
AIMV
Sindbis
TMV
BNYVV
ACLSV
PVM
NMV
WCIMV
PVX
EMV-Trin
OYMV-Tin
KYMV-JB
TYMV-CL
ELV
NTP-binding
AIMV
TMV
Sindbis
BNYVV
ACLSV
PVM
PVX
WCIMV
NMV
KYMV-JB
OYMV-Tin
EMV-Trin
TYMV-CL
ELV
Polymerase
AIMV
TMV
BNYVV
Sindbis
ACLSV
PVM
PVX
WCIMV
NMV
KYMV-JB
TYMV-CL
EMV-Trin
OYMV-Tin
ELV
0
50
100
150
200
FJD distance (%)

Cornelissen and Bol (1984), Goldbach (1987) and others that recombination has been an important factor in the evolution of plant viral genomes.

Therefore it is worth examining other less-obvious features to see whether they do or do not correlate with the replicase classification.

Genome Organization of Sindbis-like Viruses

Various features of viral genomes have been considered, at different times, to be important indicators of relatedness. Such features include, the number of genome segments, the translation strategy and the type of 5'- and 3'-terminal structures. It is therefore useful to examine these sorts of features among the 'Sindbis-like' viruses to see how well they correlate with the replicase taxonomy of the supergroup, which consistently divides into two clusters; cluster I includes the tymo-, carla-, clostero-, and potexviruses, and cluster II the alfalfa mosaic, alpha-, furo-, tobamoviruses. If any correlation is found it will imply that the correlated feature has co-evolved with the replicase. Figure 31 shows the combined replicase taxonomy of eight 'Sindbis-like' viruses together with diagrams illustrating the principal features of their genomes.

It can be seen that:-

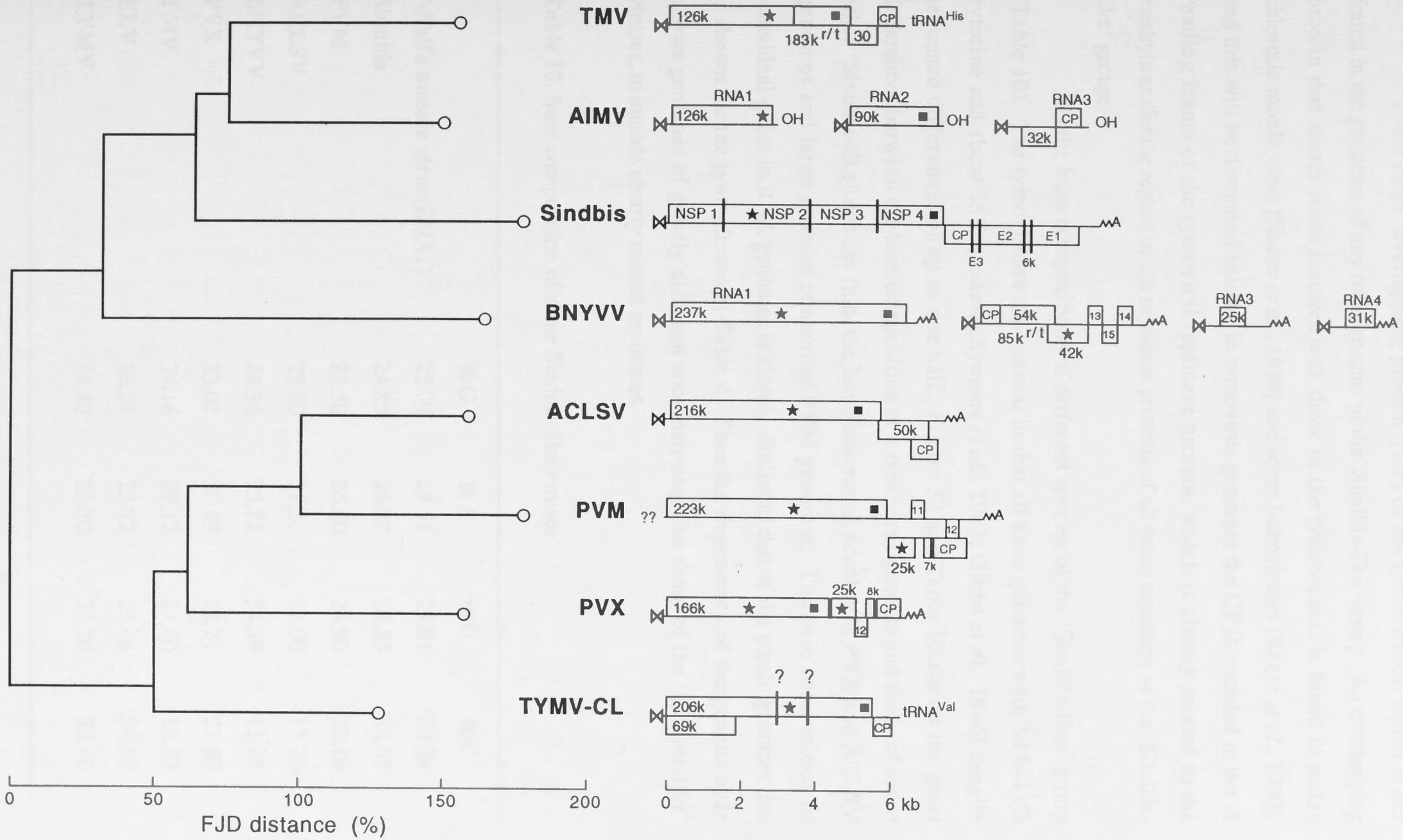
i) all the viruses of cluster I have a single genome segment, whereas those of cluster II have one, three or four genome segments, but with no pattern within the cluster;

ii) all 'Sindbis-like' viruses have a nucleotide-binding domain, but some, BNYVV of cluster II and PVM and PVX of cluster I, have two nucleotide-binding domains (always in different proteins), however there is no correlation between this difference and their position in the replicase taxonomy, nor with the number of genome segments that the virus possesses;

iii) the 'Sindbis-like' viruses have different 3'-termini (some are tRNA-like, others poly(A) and others unknown), but these do not correlate with the replicase taxonomic groupings.

These facts suggest that several of the obvious features of the genome organization of 'Sindbis-like' viruses are not correlated with their replicase phylogeny, and were probably acquired by those viruses by recombination from other sources, or have evolved *de novo* during the evolution of the group. Two distinctive features of tymoviral genomes that have probably evolved *de novo* are:-

Figure 31. A dendrogram of the RPs of eight Sindbis-like viruses based on sequence similarities and, superimposed, outlines of their genome organizations. The dendrogram was calculated like that in Figure 30A. Coding regions in the genomes are indicated as rectangles; ★ = NTP-binding site; ■ = RNA polymerase site; | = proteolytic cleavage site; ? = data uncertain; ⋈ = cap structure, and r / t = read-through.


TMV
126k
183k r/t
30
CP
tRNA His
AlMV
RNA1
126k
OH
RNA2
90k
OH
RNA3
CP
32k
OH
Sindbis
NSP 1
NSP 2
NSP 3
NSP 4
CP
E2
E1
E3
6k
BNYVV
RNA1
237k
CP
54k
85k r/t
42k
13
14
15
RNA3
25k
RNA4
31k
ACLSV
216k
50k
CP
PVM
??
223k
11
12
25k
CP
7k
PVX
166k
25k
8k
CP
12
TYMV-CL
? ?
206k
69k
CP
tRNA Val
0
50
100
150
200
FJD distance (%)
0
2
4
6 kb

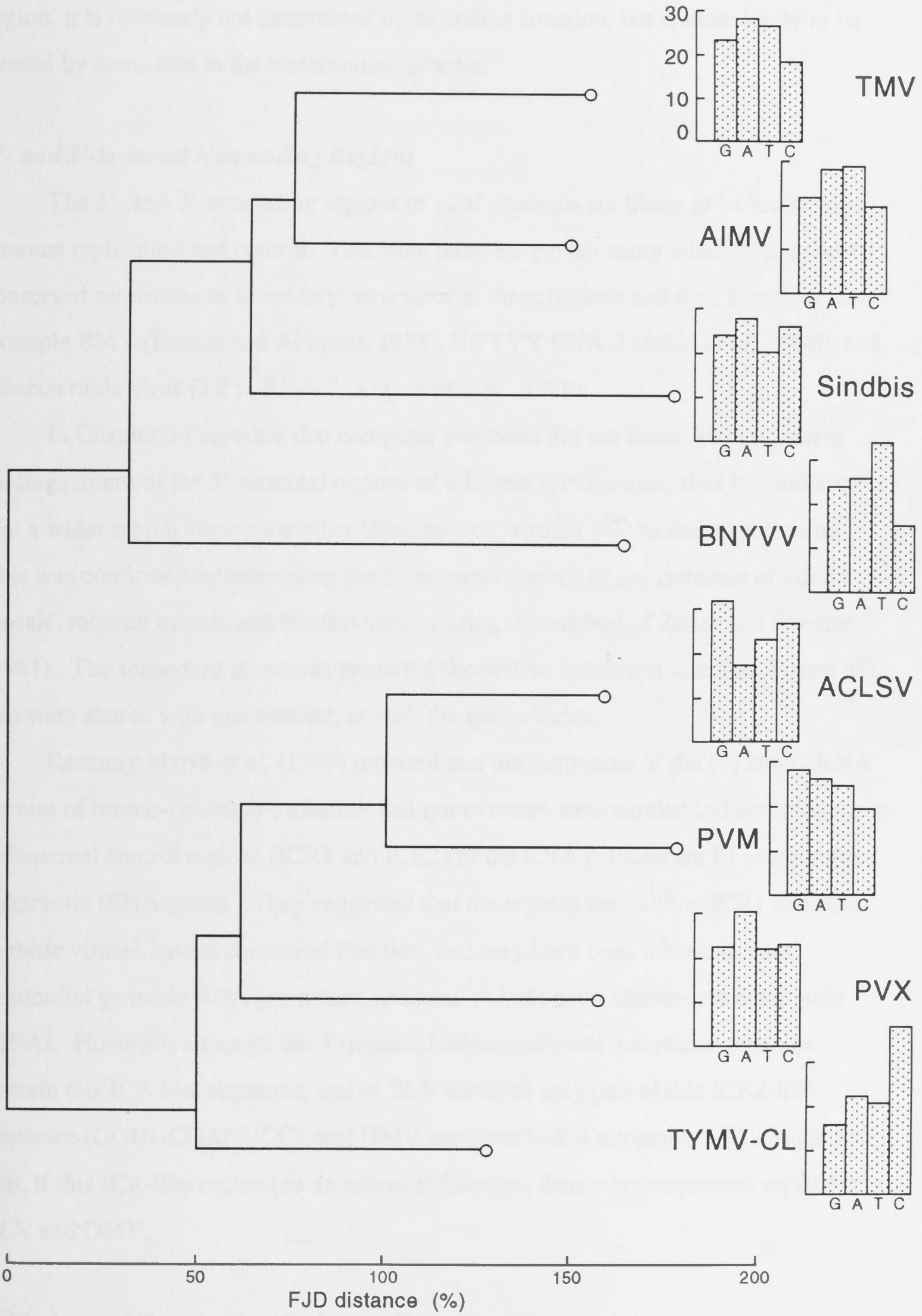
i) the large 'overlapping protein' (OP) of the tymoviruses, which is not found in the genomes of any other species of the 'Sindbis-like' group. An overlapping protein that shares some features with those of the tymoviruses is found in maize chlorotic mottle virus (Nutter *et al.*, 1989) and some luteoviruses (Keese *et al.*, 1990), and this will be discussed below. In tymovirus genomes the OP is encoded in the -1 reading frame of the tymoviral replicase protein, which is clearly related to the methyltransferase region of the replicase proteins of all other members of the 'Sindbis-like' group;

ii) the base compositions of different species of the 'Sindbis-like' group (Table 10). The tymoviruses are unusual in that all have genomes with 34%-42% cytosine and about 15% guanine (Symons *et al.*, 1963; Gibbs *et al.*, 1966) despite sequence differences of up to one half. Figure 32 and Table 10 shows the great differences between the base composition of a tymovirus genome and those of seven other 'Sindbis-like' viruses (n.b. the large content of adenine in PVX and ACLSV genomes and large guanine content of PVM genomes. The base composition of individual genes in RNA genomes is closely similar to that of the whole genome; this is shown for the tymoviruses in Table 4. Thus the degeneracy of the genetic code allows genomes of greatly different compositions, like those of the 'Sindbis-like' viruses, to encode clearly related replicases.

Table 10. Base composition of some Sindbis-like viruses.

	%G	%A	%U	%C
Alfalfa mosaic virus(RNA1)	22.09	28.51	29.01	20.39
Sindbis	24.85	28.27	20.83	26.65
PVM	28.50	26.50	24.90	20.10
ACLSV	23.80	31.50	27.00	17.70
BNYVV	24.36	25.81	34.39	15.45
PVX	23.01	30.67	22.71	23.60
TMV	24.16	29.12	27.60	19.12
ELV	16.24	23.72	25.46	34.58
TYMV	16.80	23.10	21.70	38.40

Figure 32. A Dendrogram of the RPs of eight different Sindbis-like viruses based on sequence similarities (see Figure 30A) and, superimposed, histograms showing the base composition of the genomes of those Sindbis-like viruses.


30
20
10
0
G A T C
TMV
AIMV
Sindbis
BNYVV
ACLSV
PVM
PVX
TYMV-CL
0
50
100
150
200
FJD distance (%)

The reason for the large cytosine content of tymovirus genomes is unknown, but as this bias is found in all parts of the genome except the 5'-terminal non-coding region, it is obviously not determined by its coding function, but is more likely to be caused by some bias in the transcription process.

5'- and 3'-terminal Non-coding Regions

The 5'- and 3'-noncoding regions of viral genomes are likely to be involved in genome replication and control. Therefore there have been many attempts to identify conserved sequences or secondary structures in these regions and their functions; for example BMV (French and Ahlquist, 1988), BNYVV RNA-3 (Jubin *et al.*, 1990), and tobacco rattle virus (TRV, RNA-2, Angenent *et al.*, 1989).

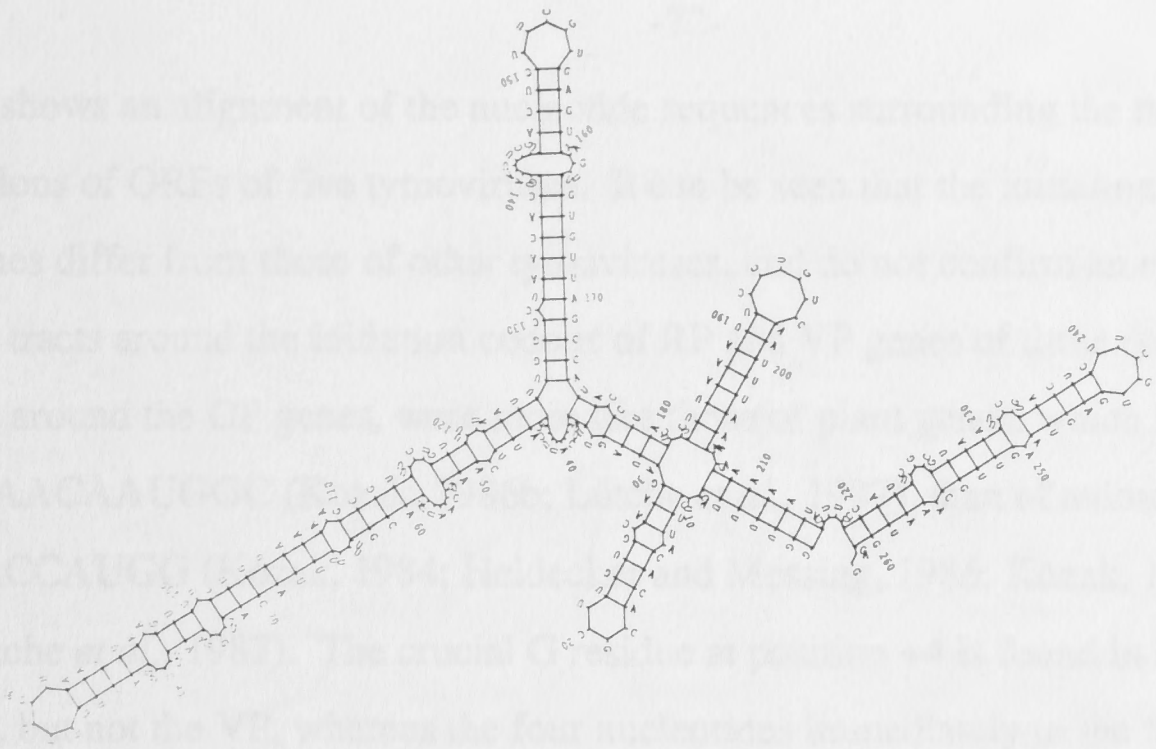
In Chapter 3 I reported that computer programs did not detect any consistent folding pattern of the 5'-terminal regions of different tymoviruses, thus it is unlikely that a wider search among the other 'Sindbis-like' viruses will be more successful. This was confirmed by examining the 5'-terminal regions of the genomes of alfalfa mosaic, tobacco mosaic and Sindbis viruses using the method of Zuker and Stiegler (1981). The secondary structures predicted showed no consistent features (Figure 33), that were shared with one another, or with the tymoviruses.

Recently Marsh *et al.* (1989) reported that the sequences of the (+) and (-)RNA termini of bromo-, cucumo-, tobamo- and tymoviruses were similar and homologous to the internal control regions (ICR1 and ICR2) of the RNA polymerase III promoters of eukaryotic tRNA genes. They suggested that the regions resembling ICR1 and ICR2 in those viruses have a conserved function, and may have been inherited from primordial genomic RNA promoters, ancestral to both plant viruses and eukaryotic tRNAs. However, although the 3'-terminal regions of most tymovirus genomes contain this ICR-like sequence, that of ELV contains only part of this ICR2-like sequence (GGUUCGANUCC), and DMV genomes lack it altogether. This suggests that, if this ICR-like region has an essential function, then other sequences replace it in ELV and DMV.

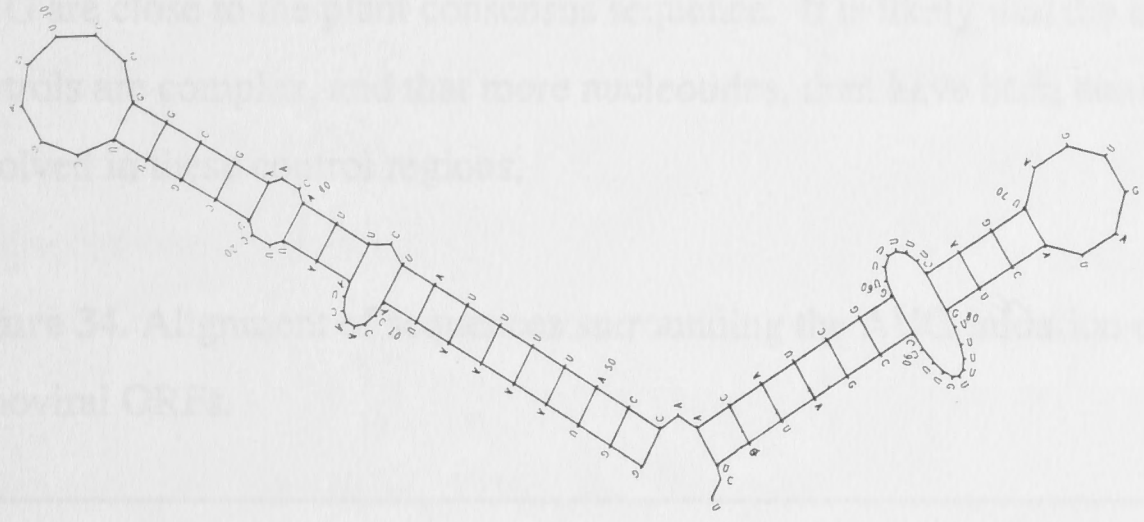
Initiation and Termination Codons in Tymovirus Genes.

There are tracts of conserved sequences around the initiation codons of mRNAs. These probably regulate the frequency with which the mRNAs are translated. Figure

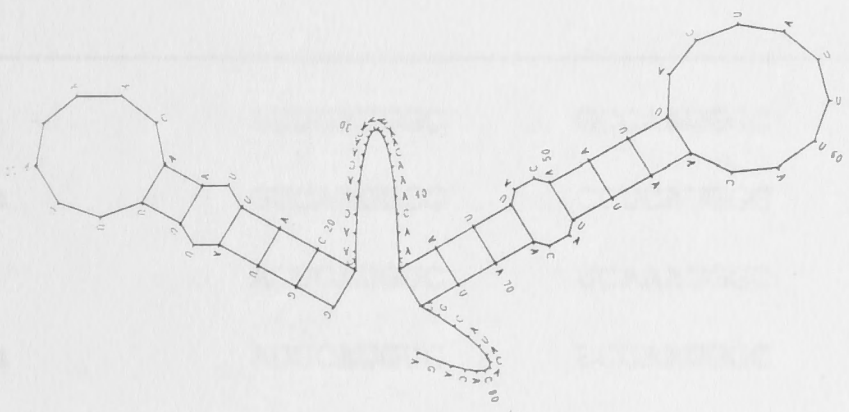
Figure 33. The predicted secondary structures of 5'non-coding regions of alfalfa mosaic virus (AlMV RNA3; Barker *et al.* , 1983), brome mosaic virus (BMV RNA3; Ahlquist *et al.*, 1984), tobacco mosaic virus (TMV; Goelet *et al.*, 1982) and white clover moasic virus (WClMV; Forster *et al.*, 1988) for comparison with those of tymoviruses (Figure 14).



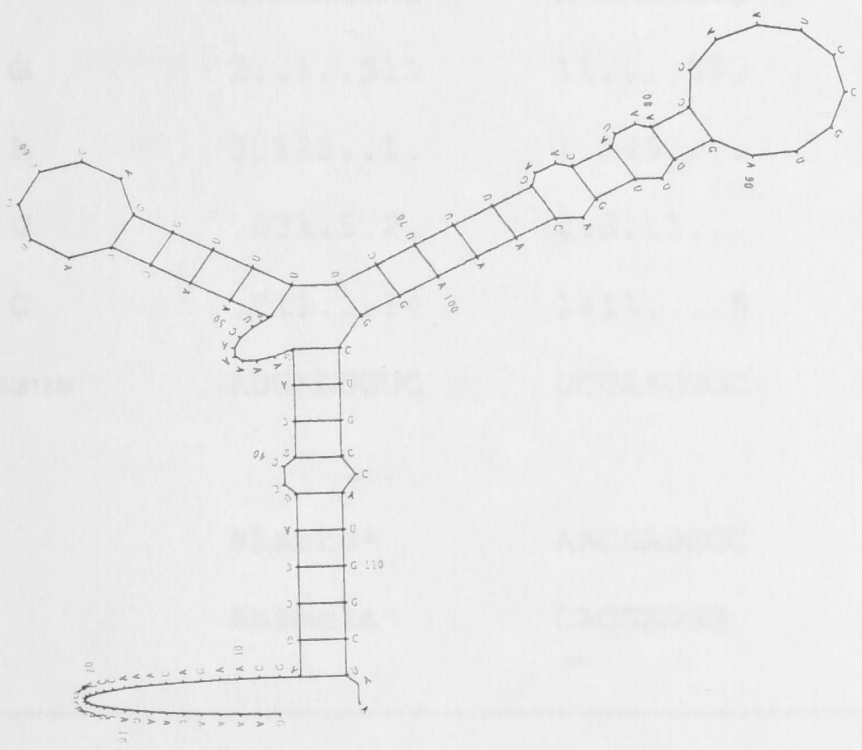
AlMV



BMV



TMV



WClMV

34 shows an alignment of the nucleotide sequences surrounding the three initiation codons of ORFs of five tymoviruses. It can be seen that the initiation tracts of ELV genes differ from those of other tymoviruses, and do not confirm an earlier report that the tracts around the initiation codons of RP and VP genes of three tymoviruses, but not around the OP genes, were more like those of plant genes, which have a consensus of AACAAUGGC (Kozak, 1986b; Lütche *et al.*, 1987), than of animal genes, CACCAUGG (Kozak, 1984; Heidecker and Messing, 1986; Kozak, 1986c, 1987; Lütche *et al.*, 1987). The crucial G residue at position +4 is found in both the OP and RP, but not the VP, whereas the four nucleotides immediately to the 5'-side of the AUG are close to the plant consensus sequence. It is likely that the translational controls are complex, and that more nucleotides, than have been considered so far, are involved in these control regions.

Figure 34. Alignment of sequences surrounding the AUG initiation codons of tymoviral ORFs.

Virus	OP	RP	VP
ELV	AUUU**AUG**GC	GCCA**AUG**GC	UACA**AUG**AC
EMV-Trin	GUCA**AUG**CC	CCUC**AUG**GC	AUCA**AUG**GA
KYMV-JB	ACUG**AUG**UC	UCAA**AUG**GC	CUAG**AUG**GC
OYMV-Tin	AUUC**AUG**UC	UCUA**AUG**GC	AAUC**AUG**GA
TYMV-CL	GCAA**AUG**AG	AGUA**AUG**GC	CGAC**AUG**GA
G	2..1..511	11....55.	.1.1..54.
A	3.125..1.	1.145....	22225..13
U	.331.5.2.	2.3..5...	121..5...
C	.211...14	1411....5	2.22....2
Consensus	AUUA**AUG**UC	UCUA**AUG**GC	AAAA**AUG**GA
			CUCC
	Plants*	AACA**AUG**GC	
	Animals*	CACC**AUG**G	

* From Kozak (1986b) and Lütcke *et al.* (1987).

The termination codons of each protein are also different (Figure 13, 31 and Table 4). There seems to be no consistent pattern of their usage in different proteins. Furthermore the possible readthrough codon, UAG, terminates the RP genes of EMV and TYMV (Morch *et al.*, 1982 b), and although this might be suppressed by an appropriate host tRNA, this seems unlikely as it is not present in ELV, which shares hosts with TYMV.

Comparisons of Possible Subgenomic Promoters in the Genomes of Tymoviruses and Other Viruses

Tymoviral VPs are translated from subgenomic mRNAs that are transcribed from the (-) genomic strand. It is likely that this transcription is controlled by genomic sequences near the start of the VP gene. Ding *et al.* (1990a) reported a conserved sequence in this region of all tymoviral genomes, and named it the tymobox. Figure 35 shows an alignment of this region of 17 tymoviral genomes. It can be seen that the tymobox is 7 or 8 nucleotides from the 'initiation box', (CAA[U/C]), of the subgenomic mRNA, and this is 1 to 16 nucleotides to the 5'-side of the initiation codon of the VP. Furthermore the tymobox, and associated regions, of the ELV genome are closely similar to those of other tymoviruses, despite the fact that the ELV VP, and other parts of the genome, are least like those of other tymoviruses.

There have been many reports of searches for conserved sequence motifs in the sub-genomic promoter regions of 'Sindbis-like' viruses (Zuidema *et al.*, 1989; Forster *et al.* (1988) and Skryabin *et al.*, 1988 *a,b*; Zavriev *et al.*, 1991; Solis and Garcia-Arenal, 1990; French and Ahlquist, 1988). Figure 35 also shows the aligned nucleotide sequences of the same regions of the genomes of several 'Sindbis-like' viruses. It can be seen that although the tymobox resembles the potexvirus promoters in being U-rich, there are no obvious sequence similarities between the promoters of these two groups, nor are there any similarities with the promoter regions of other 'Sindbis-like' viruses. It can also be seen that the promoter regions of different tymoviruses are closely similar (4 differences in 17 tymoboxes, each of 16 nucleotides), and thus as conserved as the promoters of alphaviruses (2 differences in 4 'alphaboxes', each of 21 nucleotides), whereas the others are much more variable.

Figure 35. The possible domains of subgenomic promoters reported for Sindbis-like viruses.

Virus	Conserved region

Tymoviruses (Ding *et al.*, 1990)

APLV-Hu	GAGUCUGAAUUGCUUC	7nt	CAA U	5nt	AUG
BdMV-Eur	GAGUCUGAAUUGCUUC	8nt	CAA U	8nt	AUG
CoYMV*	GAGUCUGAAUUGCUUC	8nt	CAA U	7nt	AUG
CYVV	GAGUCUGAAUUGCUUC	8nt	CAA C	16nt	AUG
DMV	GAGUCUGAAUUGCUUC	8nt	CAA U	6nt	AUG
ELV	GAGUUUGAAUUGCUUC	7nt	CAA U	5nt	AUG
EMV-Trin	GAGUCUGAAUUGCUUC	8nt	CAA U	6nt	AUG
KYMV-BP	GAGUCUGAAUUGCUUC	8nt	CAA U	5nt	AUG
KYMV-JB	GAGUUUGAAUUGCUUC	8nt	CAA^U	5nt	AUG
KYMV-PD	GAGUCUGAAUUGCUUC	8nt	CAA G	6nt	AUG
OYMV-Tin	GAGUCUGAAUUGCUUA	7nt	CAA^U	1nt	AUG
TYMV-BL	GAGUCUGAAUUGCUUC	8nt	CAA U	16nt	AUG
TYMV-Caul*	GAGUCUGAAUUGCUUC	8nt	CAA U	11nt	AUG
TYMV-CL	GAGUCUGAAUUGCUUC	8nt	CAA U	16nt	AUG
TYMV-Roth	GAGUCUGAAUUGCUUC	8nt	CAA U	11nt	AUG
TYMV-type	GAGUCUGAAUUGCUUC	8nt	CAA^U	16nt	AUG
WCuMV	GAGUCUGAAUUGCAUC	7nt	CAA U	6nt	AUG
consensus	**GAGUcUGAAUUGCuUc**		**CAA u**		

Potexviruses (Zuidema *et al.*, 1989)

NMV CP	UGGAACGGGGUUAAGUUUCCUUUAU	54nt	AUG	
PVX CP	UUGAAC--GGUUAAGUUUCCAUUGA	10nt	AUG	
WClMV CP	AACCAC-GGGUUAAGUUUACCAUCU	12nt	AUG	
PAMV CP	UGGAACGGGGUUAAGUUUCCAUCUA	46nt	AUG	
PMV CP	CUACACGGGCUUAGGAACUAGCUUC	25nt	AUG	
NMV 26K	CGCUACCGGGUUAAGUUCAUUGCCU	16nt	AUG	
PVX 25K	UUUAAÇC--GUUAAUUUACUUUAUA	12nt	AUG	
WClMV 26K	GACUAC-GGGUUAAGAGACCUUAAG	5nt	AUG	
consensus	**ACgGGgUUAaguuuccuu**			

Alfalfa mosaic virus (Barker *et al.*, 1983)

AlMV RNA3	**GAUCauUGauCGG**	**UAAU**		

Alphaviruses (Ou *et al.*, 1982)

MBV	ACCUCUACGGCGGUCCUAAAU	48nt	CACC	AUG	AAU
RRV	ACCUCUACGGCGGUCCUAAAU	46nt	AAAC	AUG	AAU
SFV	ACCUCUACGGCGGUCCUAAAU	49nt	CACC	AUG	AAU
Sindbis	AUCUCUACGGUGGUCCUAAAU	47nt	CACC	AUG	AAU
consensus	**AcCUCUACGGcGGUCCUAAAU**		**cAcC**		**AAU**

Bromoviruses (Ahlquist *et al.*, 1981 and Allison *et al.* unpublished result)

BMV RNA3	AGAUCUAUG	UCCUAAUU	4nt	GUA-UUAAUA	AUG
CCMV RNA3	AAAUCUAUG	UU-UAAUU	5nt	GUAAUUUAUC	AUG
consensus	**AUCUAUG**	**UcCUAAUU**		**AaUU-AU-**	

Tobamoviruses (Solis and Garcia-Arenal, 1990)

TMGMV 30K	GGUUC GUUUGCUU	53nt	UUAA	AUG GCU
TMV 30K	GGUUC GUUUGUUU	53nt	AUAG	AUG GCU
ToMV 30K	GGUUC GUUUGUUU	53nt	CUUG	AUG GCU
TMGMV CP	AGUAC GUUUUUAA	1nt	CAAU	AUG CCU
TMV CP	GAUUC GUUUUUAA	-2nt	UAAA	AUG UCU
ToMV CP	GAUUC GUAUUUAA	-2nt	UAAA	AUG UCU
consensus	**ggUuC GUuUNuNN**			**gCU**

* CoYMV (Ding *et al.*, 1990c)

TYMV cauliflower strain (Anne Mackenzie unpublished data)

Abbreviations used for the viruses here and in the text are: Andean potato latent (APLV); belladonna mottle (BdMV); brome mosaic (BMV); cacao yellow mosaic (CoYMV); clitoria yellow vein (CYVV); cowpea chlorotic mottle (CCMV); dulcamara mottle (DMV); erysimum latent (ELV); eggplant mosaic (EMV); kennedya yellow mosaic (KYMV); Middelburg (MBV); narcissus mosaic (NMV); ononis yellow mosaic (OYMV); potato aucuba mosaic (PAMV); papaya mosaic (PMV); potato virus X (PVX); Ross River (RRV); Semliki Forest (SFV); Sindbis; tobacco mosaic (TMV); tomato mosaic (ToMV); tobacco mild green mosaic (TMGMV); white clover mosaic (WClMV) and wild cucumber mosaic (WCuMV).

CHAPTER 6 - CONSTRUCTION OF CLONED DNA ENCODING THE ELV GENOME

INTRODUCTION

Another way to study the genetic organization, replication and expression of the tymoviral RNA genome, once the complete genomic sequence of ELV has been determined, is to make a clone of DNA encoding the ELV genome, and from this to generate infectious transcripts by *in vitro* transcription.

CLONE CONSTRUCTION

Construction and Selection of Full-size cDNA Clones

DNA encoding the entire genome of ELV was synthesized by two methods: the RNase H method (Gubler and Hoffmann, 1983) and the polymerase chain reaction (PCR) as described in Chapter 2. However, methyl mercuric hydroxide-treated RNA (Gall *et al.*, 1988) was not a suitable template for ELV DNA synthesis, as reported already in Chapter 2, perhaps because the primers for first strand synthesis, dT_8dG or PS9, were ineffective for some unknown reason.

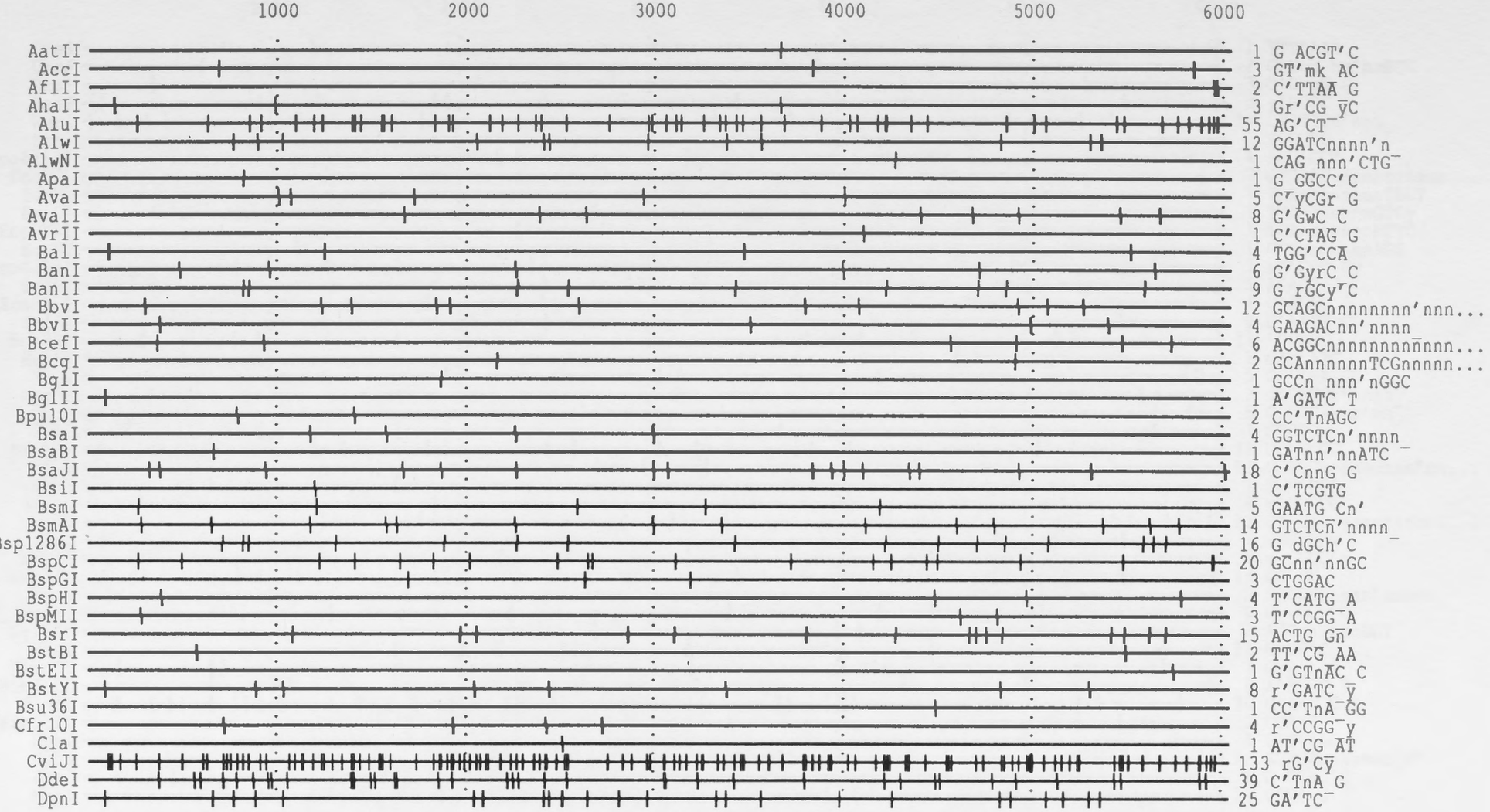
DNA prepared by the RNase H method, hydrolysed by *Sph* I and cloned in vector pTZ18U and pTZ19U, yielded a few hundred clones. At all stages in the preparation of the full-length ELV DNA, the cloned inserts were examined by detailed restriction fragment analysis and sequencing, and compared with the restriction endonuclease map (Figure 36) of the genomic sequence that had been determined before.

Two DNA clones encoding the ELV genome, and named pELV35 and pELV81, were assembled in pBluescript SK(+). First, six smaller clones were ligated together; their names, genomic positions encoded and the restriction endonuclease sites that formed their termini were:

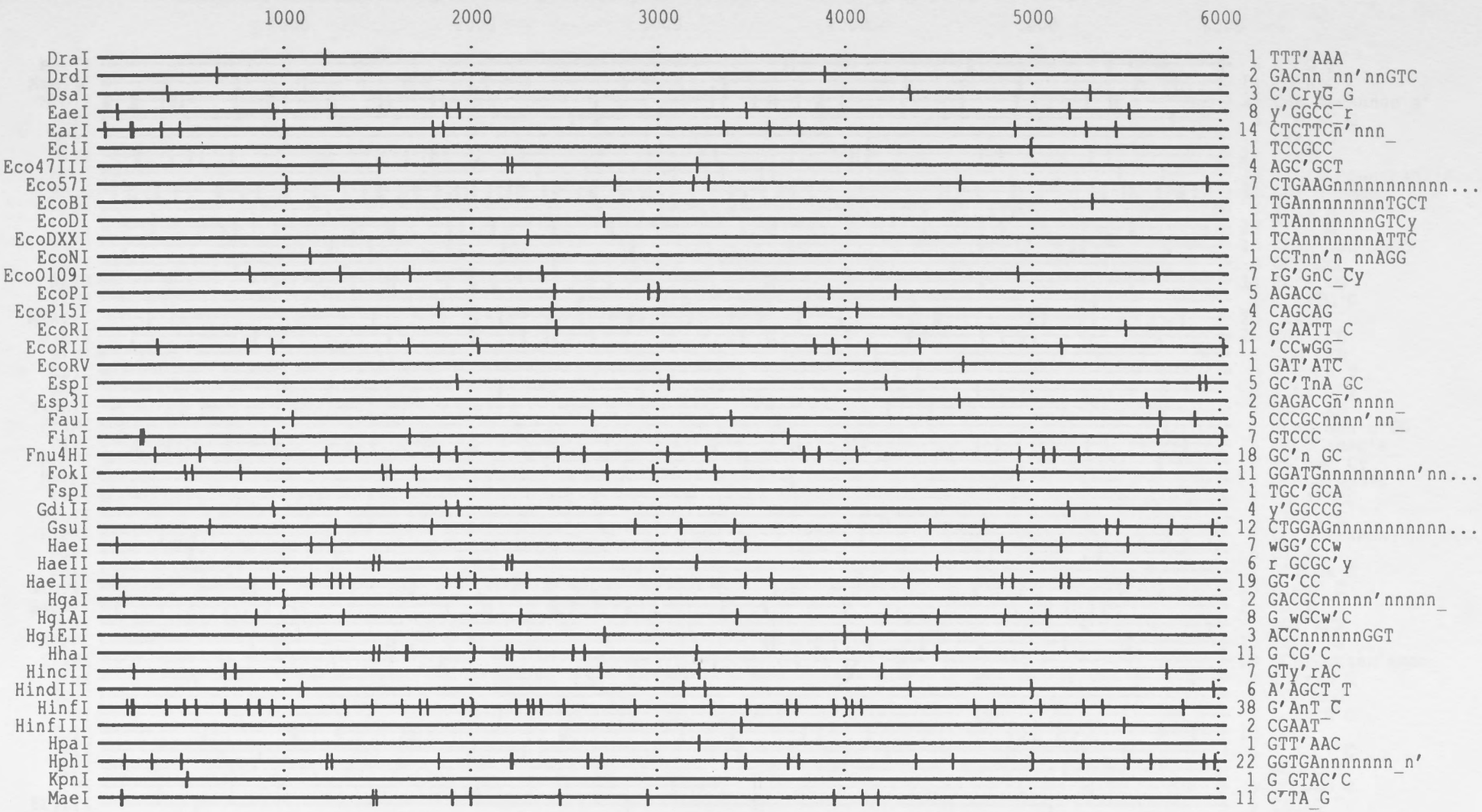
- pELV37, pTZ19U: (nts 1-2142), *EcoR* I/*Sph* I;
- pELV25, pTZ19U: (nts 2142-2461), *Sph* I/*EcoR* I;
- pELV40, pTZ19U: (nts 2461-4630), *EcoR* I/*EcoR* V;

Figure 36. Map of restriction endonuclease sites in DNA encoding the ELV genome. These restriction sites are those recognized by four and six base restriction endonuclease enzymes and were identified by the UWGCG program Version 6.0.

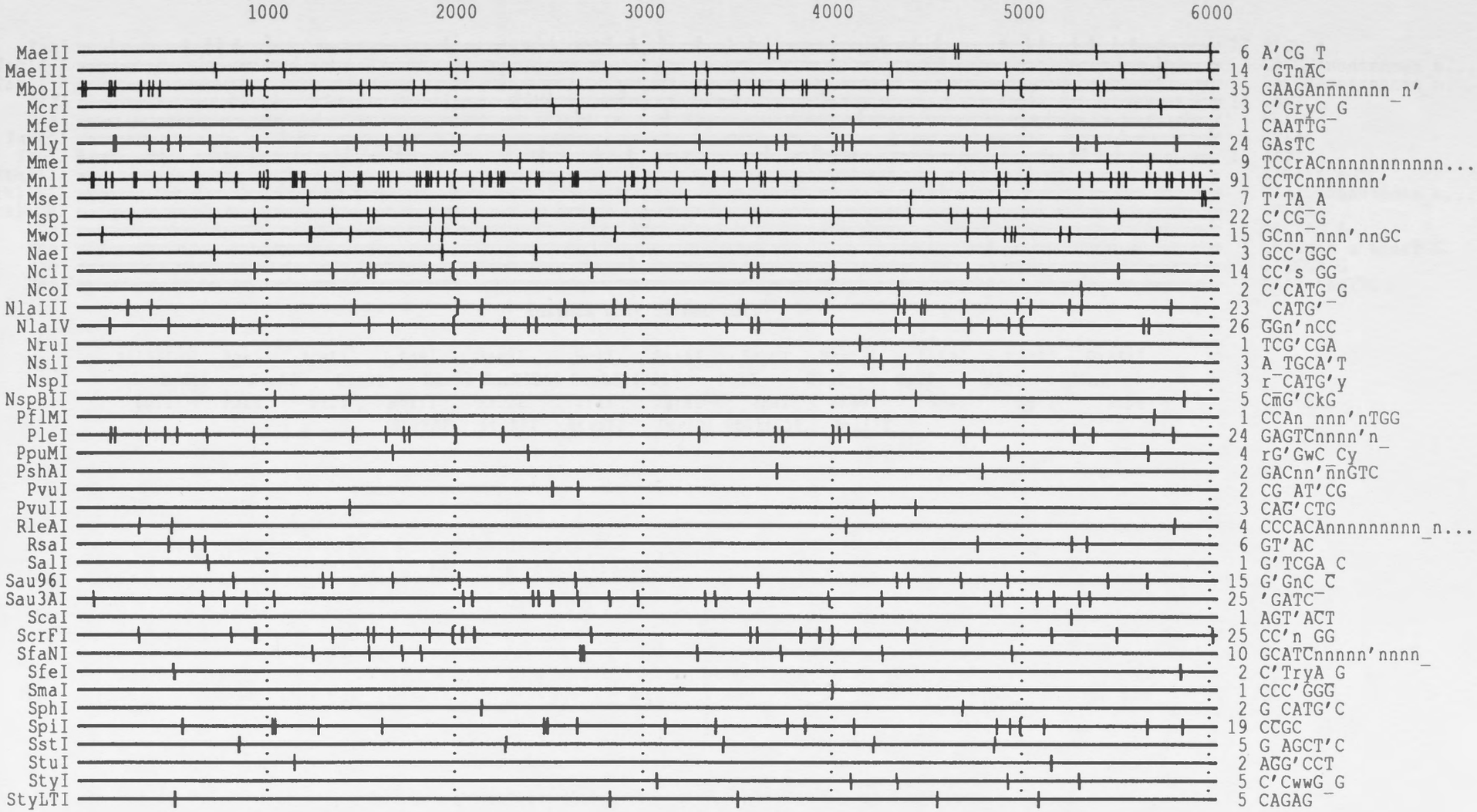
(Linear) MAPPLOT of: Elv.Dna ck: 7837, 1 to: 6034 January 24, 1991 15:23.



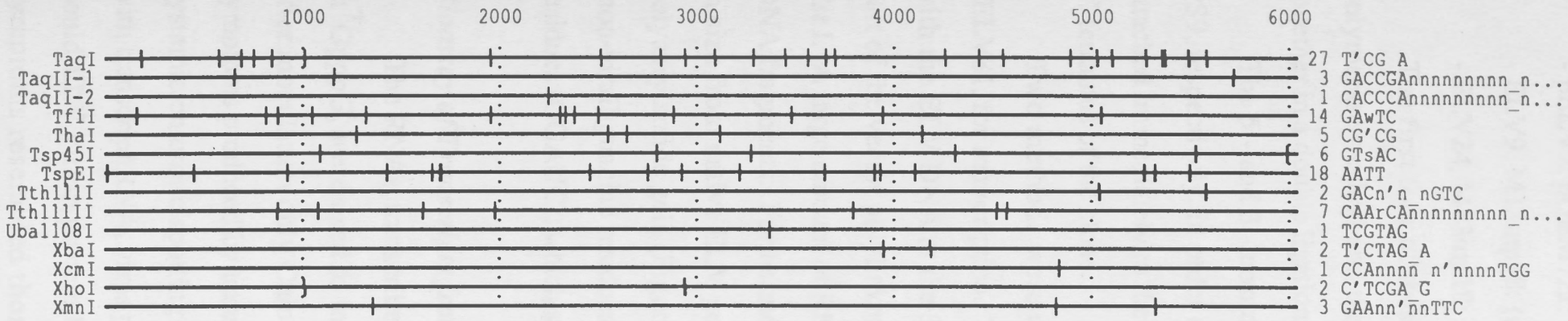
(Linear) MAPPLOT of: Elv.Dna ck: 7837, 1 to: 6034 January 24, 1991 15:24.



(Linear) MAPPLOT of: Elv.Dna ck: 7837, 1 to: 6034 January 24, 1991 15:24.



(Linear) MAPPLOT of: Elv.Dna ck: 7837, 1 to: 6034 January 24, 1991 15:25.



Enzymes that do not cut:

AflIII	AgeI	ApaLI	AseI	BamHI	BclI	BsaAI	BspMI	BssHII	BstXI	CfrAI	DraIII			
DrdII	EcoAI	EcoEI	EcoKI	EcoR124I	EcoR124/3I	FseI	MluI	NarI	NdeI	NheI				
NotI	PmlI	PstI	RsrII	SacII	SfiI	SgrAI	SnaI	SnaBI	SpeI	SplI	SspI			
			StySBI	StySJI	StySPI	StySQI	Uba1105I	XmaIII						

- pELV2, pGEM7Zf: (nts 4630-5494), *EcoR* V/*EcoR* I;

- ELV9, M13mp18: (nts 5494-5745), *EcoR* I/*BstE* II;

- ELV24, M13mp18: (nts 5745-6034), *BstE* II/*BamH* I.

The first four of these clones were obtained by the RNase H method and *Taq* polymerase method and the last two by PCR. Some of these fragments were ligated together in a double ligation reaction as outlined in Figure 37 and Figure 38.

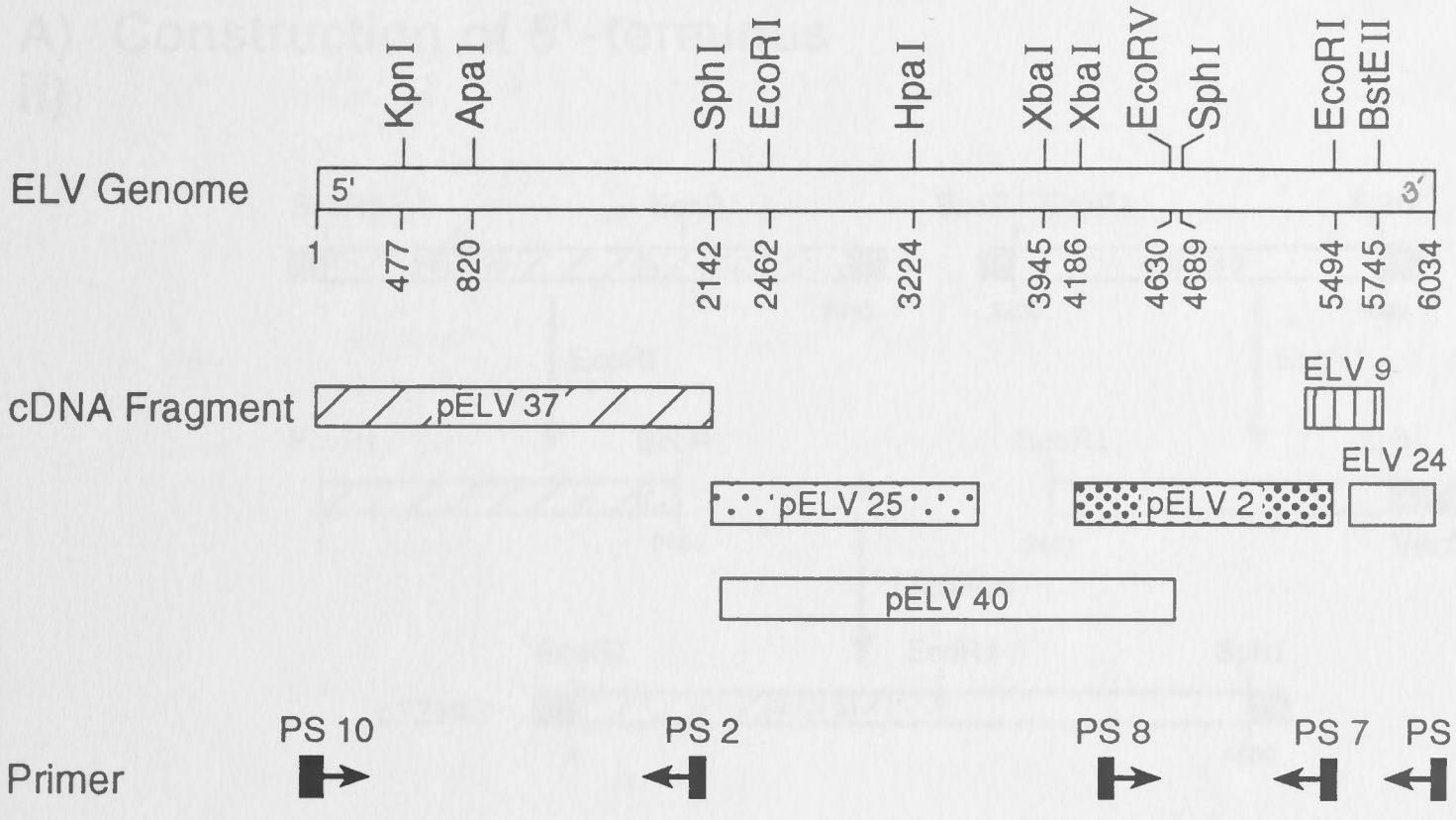
The 5'- and 3'-termini were then replaced sequentially using primers PS10 and PS9, respectively, in order to provide a unique *Pst* I site followed by the T7 promoter attached immediately to the 5'-terminus of the clone and a *BamH* I site adjacent to the 3'-terminus of the clone.

Two methods were used for preparing the template plasmids, pELV35 and pELV81, for transcription. They were hydrolysed with *BamH* I, resulting in molecules with the ELV DNA at the 3'-end of part of the PS9 primer sequence and the 5'-end part of the vector polylinker, or they were hydrolysed using a mixture of *BamH* I and *Pst* I. A large amount of RNAs were obtained by transcription of those two linearised DNA fragments. Furthermore the RNAs were of the same size, and indistinguishable in size from native ELV genomic RNA, as judged by electrophoresis in denaturing polyacrylamide gels (Figure 40). These transcripts were expected to have one extra (non-viral) guanine residues at their 5'-termini and, at their 3'-termini, up to five extra residues (GGAUC), which are part of the *BamH* I site (Figure 38).

Bioassay of Transcripts from the ELV DNA Clones.

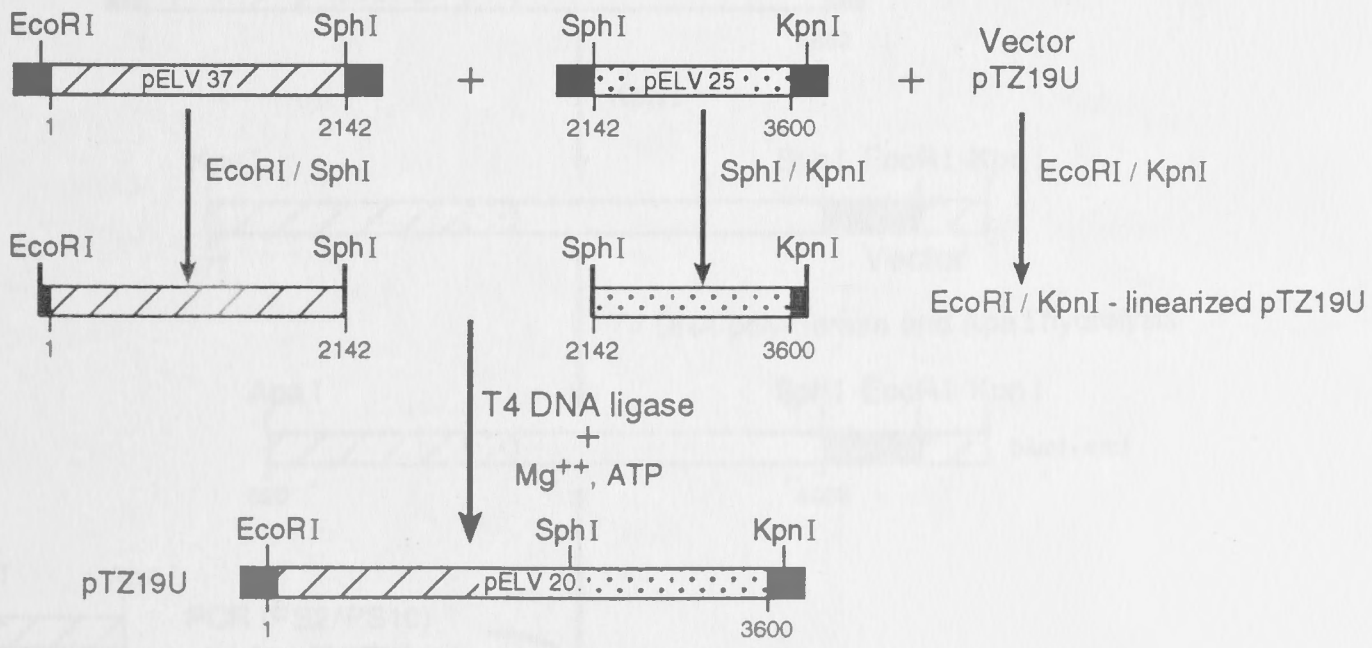
The RNAs, transcribed from DNA clones in the presence of the cap homologue, m^7GpppG, were used to inoculate Chinese cabbage and swede plants. Two weeks after inoculation only Chinese cabbage showed chlorotic mottling (Figure 39). The symptoms produced by transcript RNA and 'parent' ELV were not identical; the severe systemic chlorotic spotting and mottling induced by the parent RNA did not develop with transcript RNA, virus particles could not be detected by electron microscopy, nor could ELV virion protein be detected by immunoprecipitation. However the mild symptoms resembled those found sometimes when virus inoculum was used at terminal dilutions, but when Chinese cabbage plants were inoculated with sap from the transcript-inoculated plants showing mild symptoms, no symptoms developed.

Figure 37. Diagrams showing details of the stages involved in constructing and assembling dsDNA encoding the ELV genome; the location of each of these stages is given the same lettering in the Figure 38 which shows the general plan of DNA synthesis. Locations of important restriction sites are indicated and those DNA portions with different origins are shaded differently. Part of vector genomes are shaded as solid bars. The positions of the primers, and the direction of DNA synthesis they primed, are shown; their sequences are given in the text.

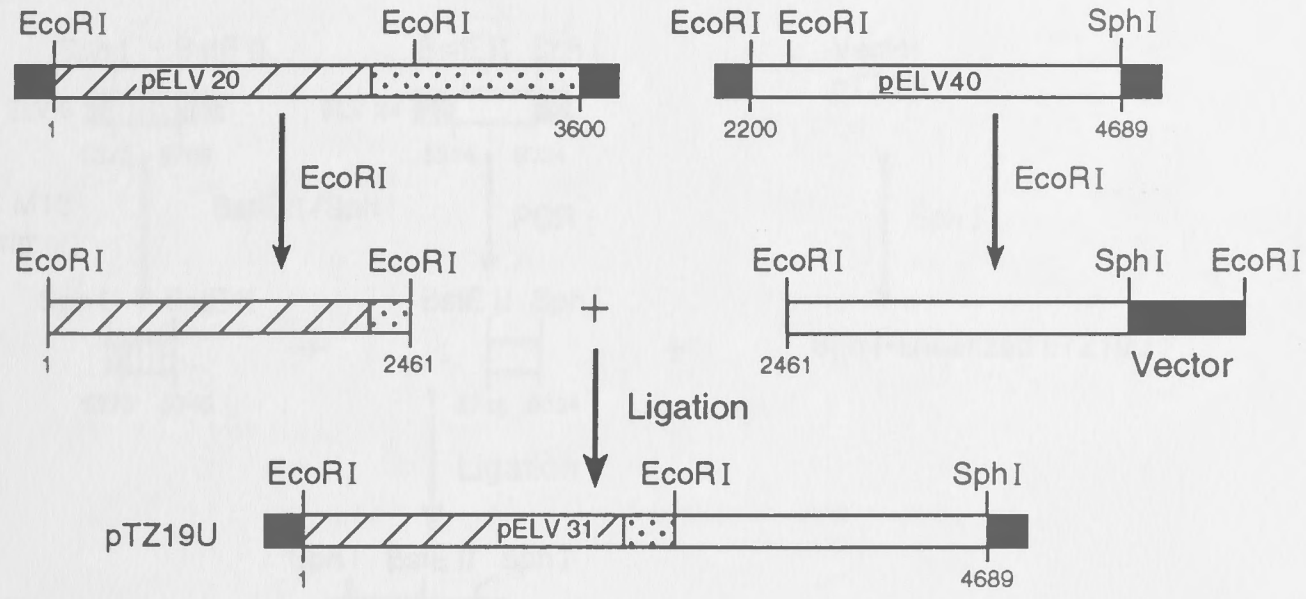


A) Construction of 5'-terminus

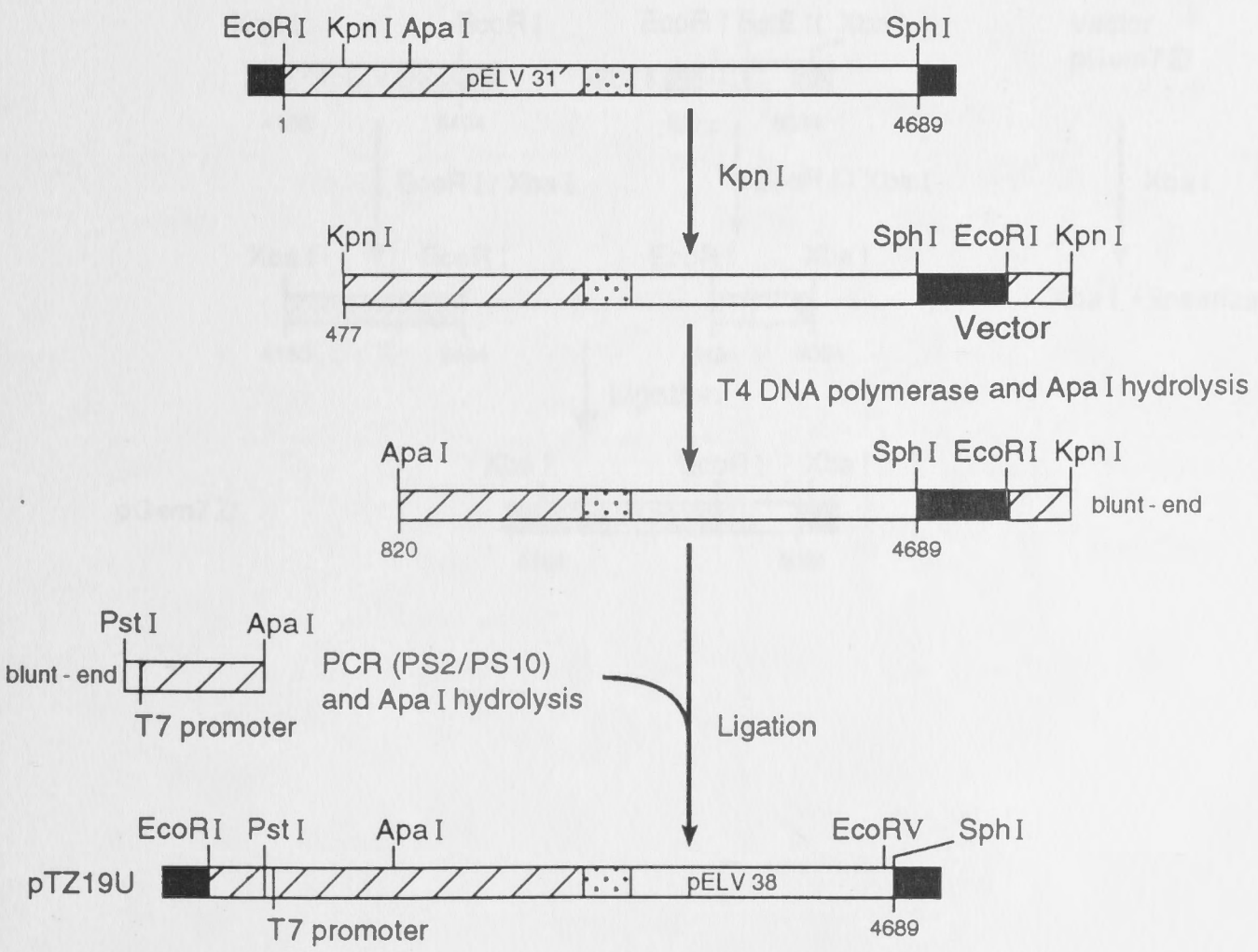
i)



A) Construction of 5'-terminus
ii)

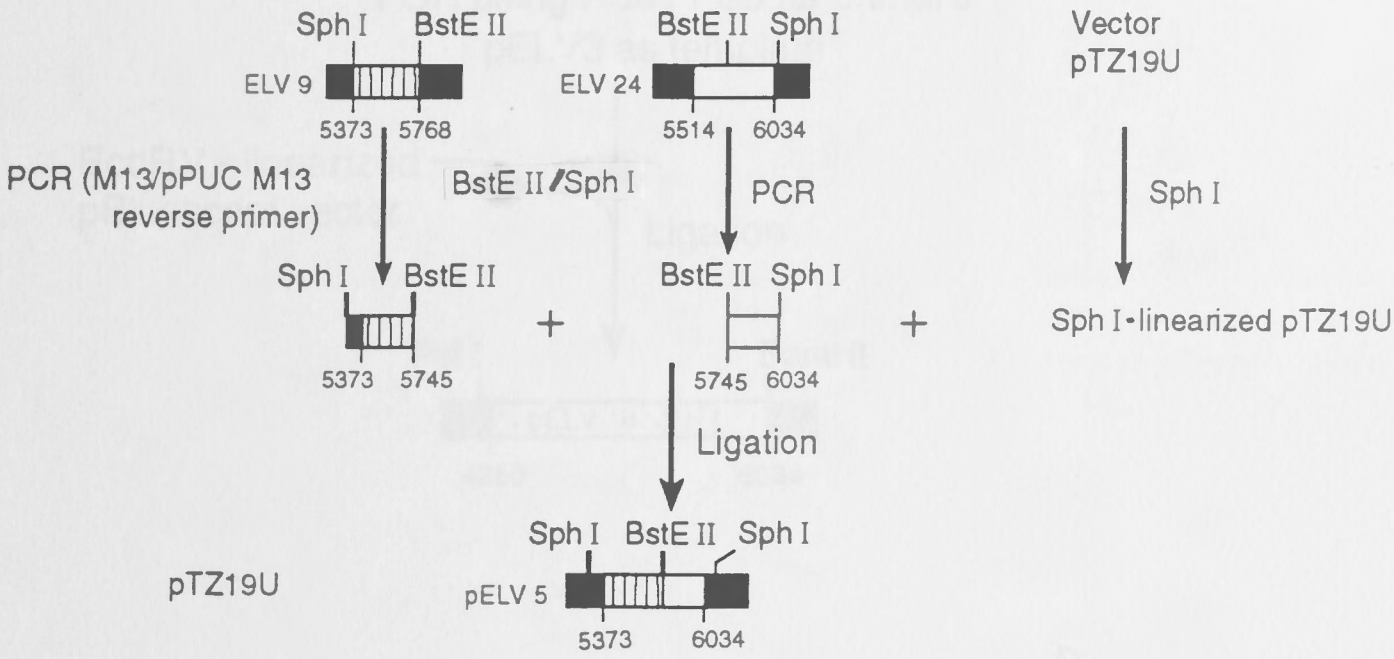


B) Substitution of 5'-terminus (PS10 T7RNA polymerase primer)

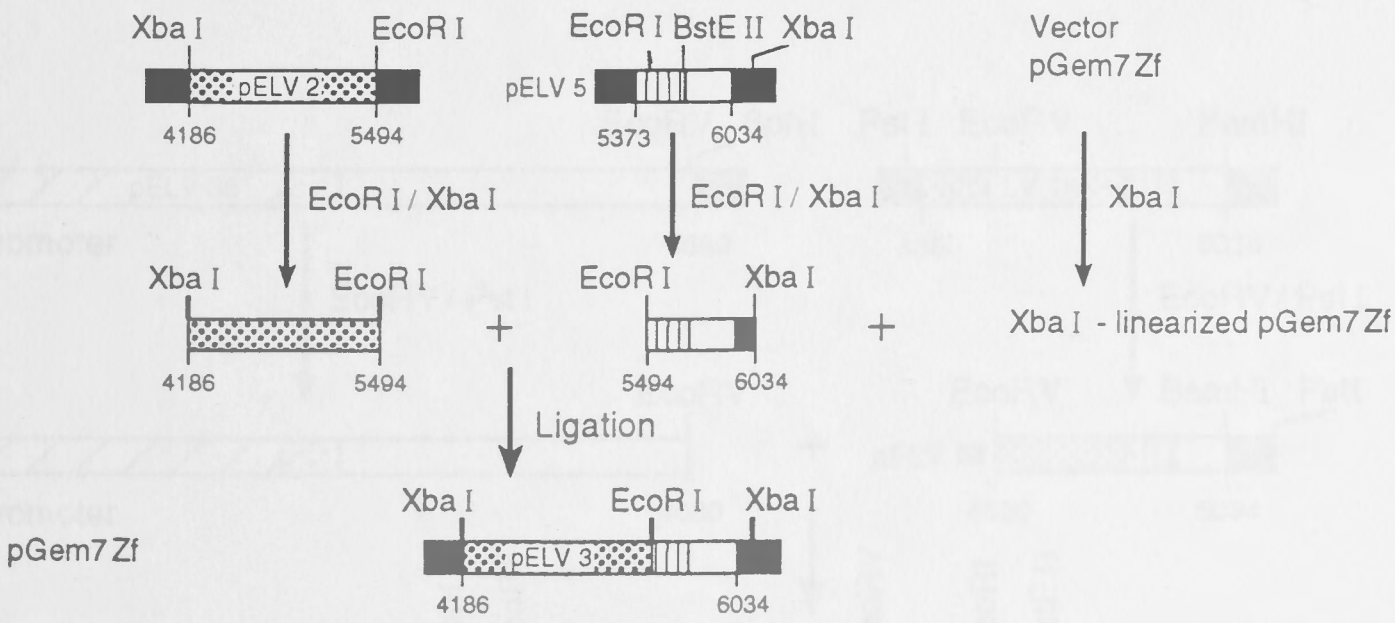


C) Construction of 3'-terminus

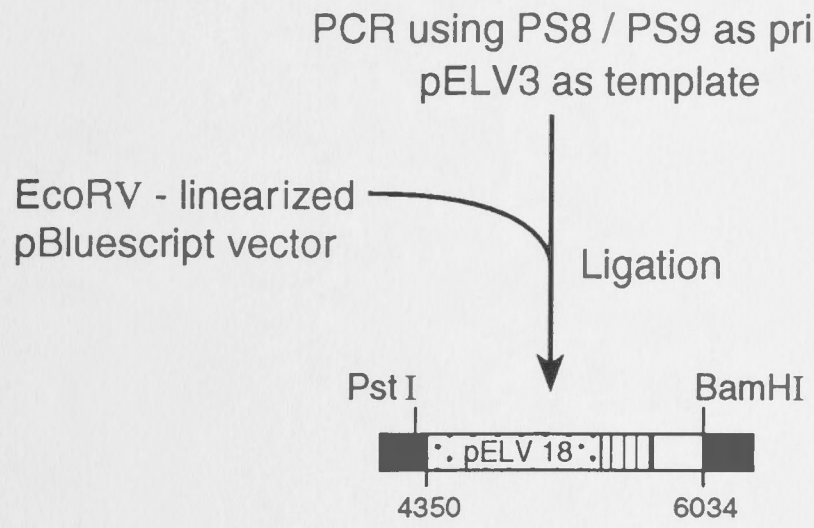
i)



ii)



D) Substitution of 3' - terminus



E) Construction of full-length ELV genome

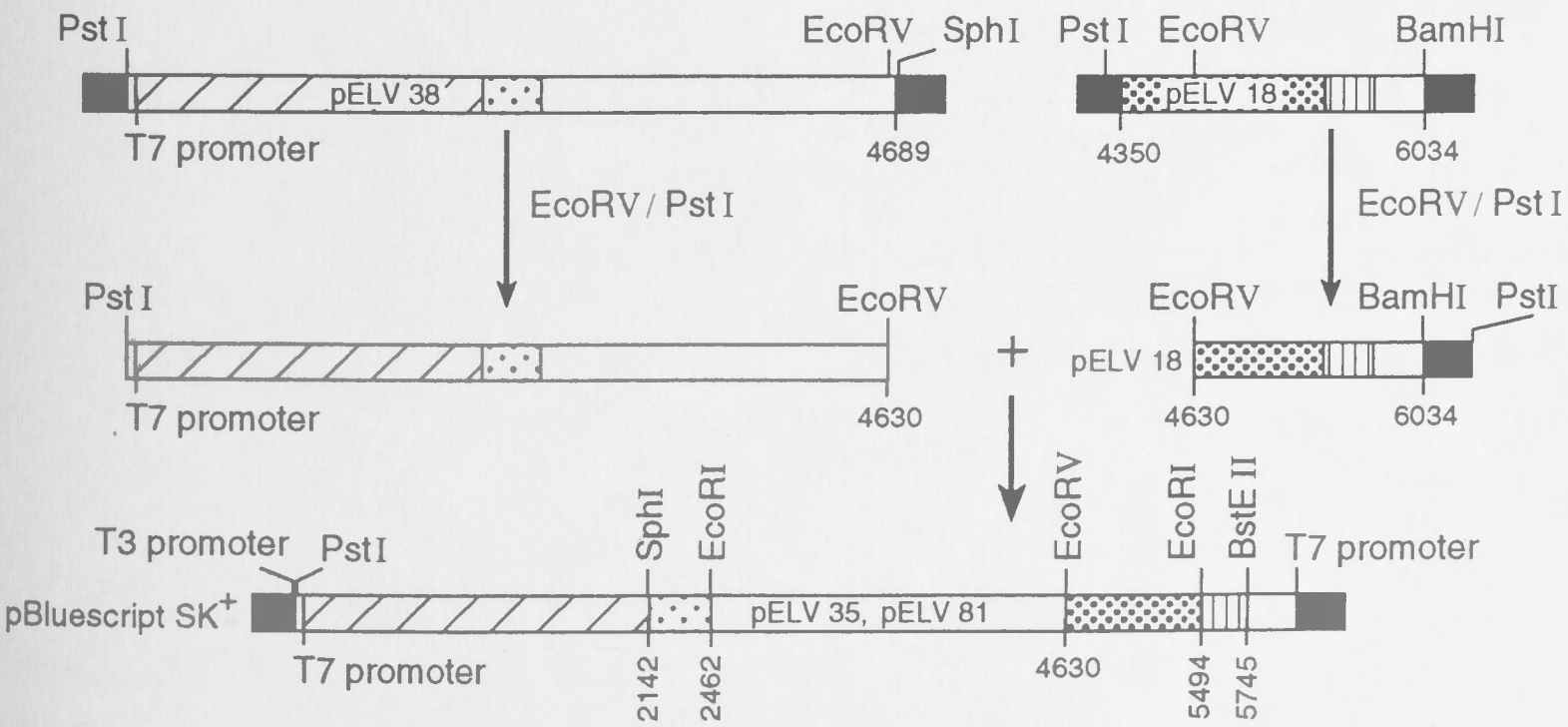
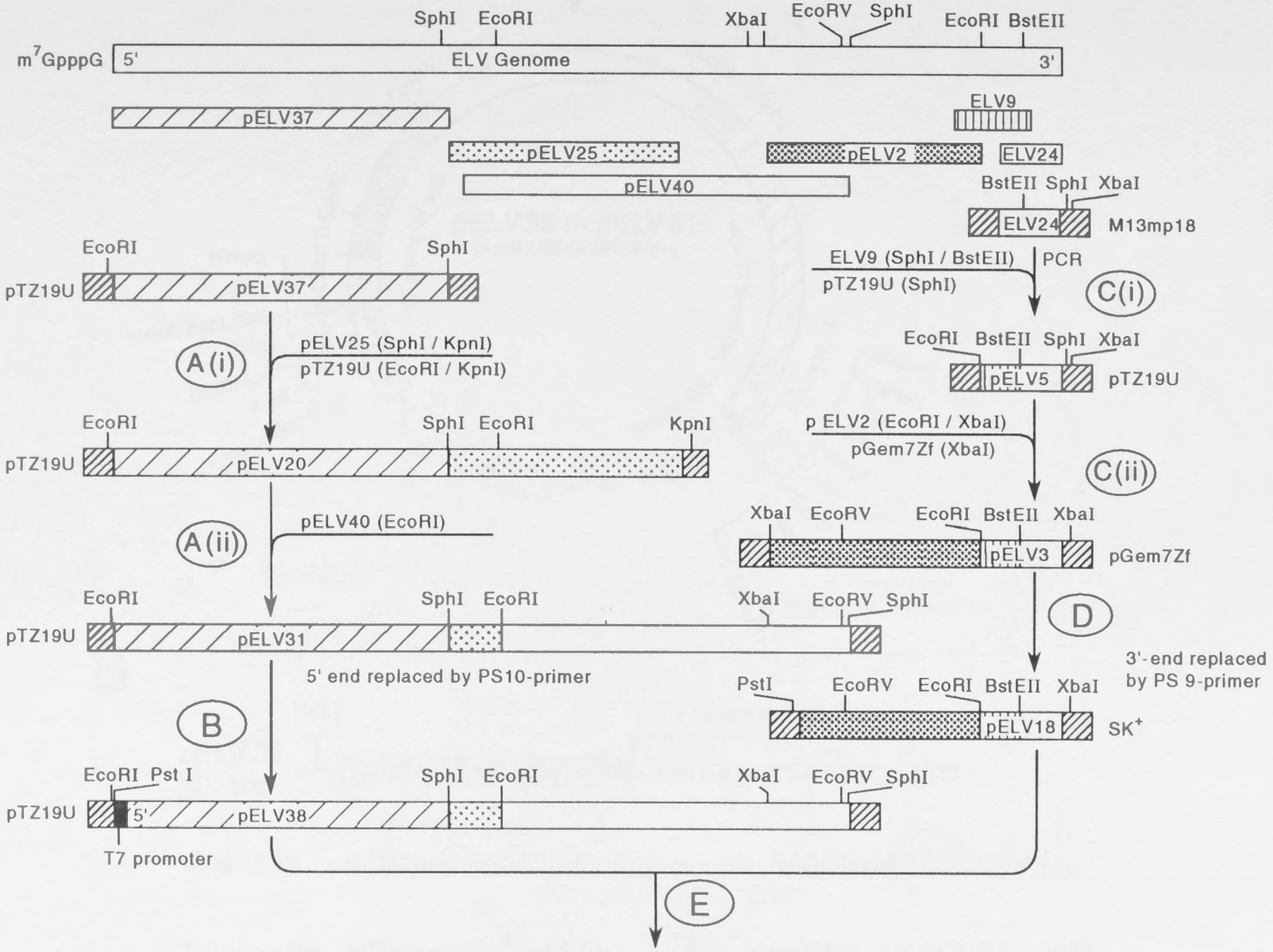
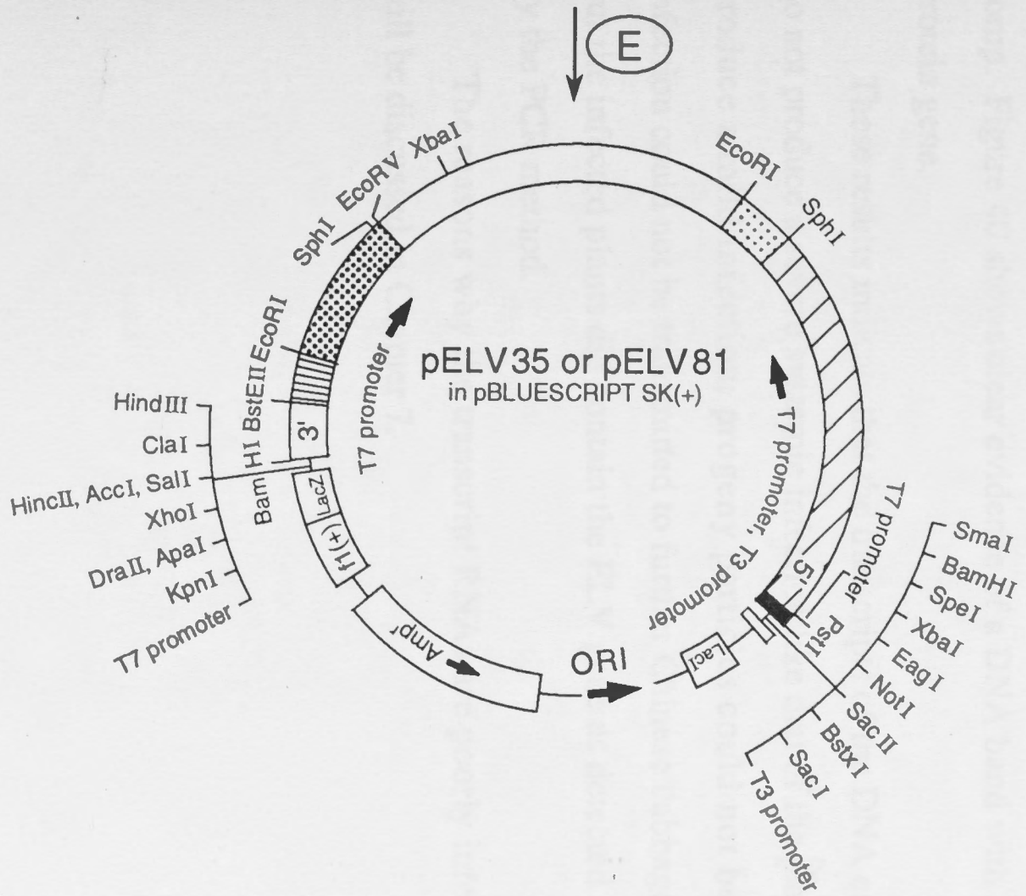


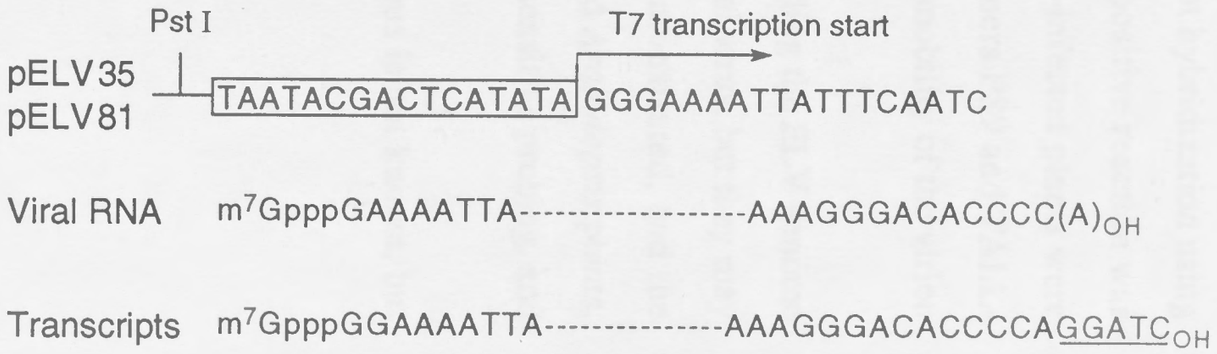
Figure 38. (A) General plan of the stages used to construct and assemble dsDNA encoding the ELV genome. Locations of important restriction sites are indicated, and all the intermediate stages are illustrated in Figure 37 and described in the text. The final plasmids pELV35 and pELV81 contained the entire ELV encoding DNA. (B) The expected terminal sequences of transcript products of *BamH* I-linearised pELV35 and pELV81 are presented.

A


SphI EcoRI
XbaI
EcoRV SphI
EcoRI BstEII
m7GpppG
5'
ELV Genome
3'
pELV37
ELV9
pELV25
pELV2
ELV24
pELV40
BstEII SphI XbaI
ELV24
M13mp18
EcoRI
SphI
pTZ19U
pELV37
ELV9 (SphI / BstEII)
PCR
pTZ19U (SphI)
C(i)
A(i)
pELV25 (SphI / KpnI)
pTZ19U (EcoRI / KpnI)
EcoRI BstEII SphI XbaI
pELV5
pTZ19U
EcoRI
SphI EcoRI
KpnI
pTZ19U
pELV20
p ELV2 (EcoRI / XbaI)
pGem7Zf (XbaI)
C(ii)
XbaI EcoRV
EcoRI BstEII XbaI
A(ii)
pELV40 (EcoRI)
pELV3
pGem7Zf
EcoRI
SphI EcoRI
XbaI EcoRV SphI
D
pTZ19U
pELV31
5' end replaced by PS10-primer
3'-end replaced
by PS 9-primer
PstI
EcoRV
EcoRI BstEII XbaI
B
pELV18
SK+
EcoRI Pst I
SphI EcoRI
XbaI EcoRV SphI
pTZ19U
5'
pELV38
T7 promoter
E


E
pELV35 or pELV81
in pBLUESCRIPT SK(+)
EcoRI
SphI
XbaI
EcoRV
SphI
EcoRI
BstEII
HI
HindIII
ClaI
HincII, AccI, SalI
XhoI
DraII, ApaI
KpnI
T7 promoter
Bam
3'
LacZ
f1(+)
Ampr
ORI
lacI
T7 promoter
T3 promoter
5'
PstI
T7 promoter
SmaI
BamHI
SpeI
XbaI
EagI
NotI
SacII
BstXI
SacI
T3 promoter

B


Pst I
T7 transcription start
pELV35
pELV81
TAATACGACTCATATA GGGAAAATTATTTCAATC
Viral RNA m7GpppGAAAATTA----------AAAGGGACACCCC(A)OH
Transcripts m7GpppGGAAAATTA----------AAAGGGACACCCCAGGATCOH

Further tests were done to establish the nature of the agent causing the symptoms in plants inoculated with RNA transcribed from plasmids pELV35 and pELV81. RNAs extracted from these plants did not yield a band of RNA with the mobility of ELV RNA, when the mixture was analysed by electrophoresis in polyacrylamide gels (Figure 40), although the infected plants yielded RNA with a considerably greater concentration of plant ribosomal RNAs than in plants infected with parental ELV. The RNA extracts were also examined for ELV RNA by dot-blot hybridization using labelled DNA of the virion protein gene as probe, and a faint positive reaction was obtained (Figure 40). The RNAs extracted from the transcript-infected plants were also tested for amplification by the PCR method with two primers PS9 and TALL-comp. Figure 40 shows clear evidence of a DNA band with the mobility of the virion protein gene.

These results indicate that the transcripts of the DNA encoding the ELV genome do not produce a severe systemic infection like that of the parental virus, but they may produce a mild infection; progeny particles could not be demonstrated, and the infection could not be transmitted to further Chinese cabbage and *Arabidopsis* plants, but the infected plants did contain the ELV gene as detected by sensitive probing, and by the PCR method.

The reasons why the transcript RNAs are poorly infectious is not known, but will be discussed in Chapter 7.

Figure 39. ELV-infected plants: (A) and (B) Chinese cabbage plants showing severe necrotic and chlorotic spotting and mottling 10 days after inoculation with ELV and (C) Chinese cabbage plants showing mild chlorotic mottling after inoculation with transcript RNA. (D) *Arabidopsis thaliana* showing very severe symptoms of chlorotic mottling and veinclearing after inoculation with ELV; these plants later died.

A

B

C

Figure 40. Analysis of RNAs from Chinese cabbage plants inoculated with RNA transcript.

(A) By polyacrylamide gel electrophoresis. Tracks (from left to right) were loaded with RNA from: uninoculated leaf; ELV-infected leaf; leaf sham-inoculated with buffer; leaf inoculated with RNA transcript from pELV35 and pELV81; purified ELV genomic RNA; RNA transcribed from pELV35.

(B) Analysis of the RNAs analysed in (A) for presence of ELV VP gene by PCR using labelled precursors and PS9 primer complementary to the 3'-terminal of ELV genome and TALL-comp tymobox primer. Tracks (from left to right) were loaded with DNA products obtained from: uninoculated leaf; buffer-inoculated leaf; ELV-inoculated leaf; Markers (pBR325/ *Hinf* I); leaf inoculated with transcript of pELV81 and pELV35.

(C) Dot-blot hybridization analysis of RNA extracted from (from left to right and anticlockwise): uninoculated leaf; buffer-inoculated leaf; leaf inoculated with transcript from pELV35; purified ELV genomic RNA and also 2 dilutions of pELV35 DNA.

C
Healthy
Buffer
pELV35
ELV
cDNA

A
Healthy
ELV
Buffer
pELV35
pELV81
ELV RNA
ELV Transcript

B
Healthy
Buffer
ELV
Marker
pELV81
pELV35
150
220
290
344
396
512
652

CHAPTER 7 - DISCUSSION and CONCLUSIONS

GENOMIC ELV SEQUENCE

My work has provided the complete genomic nucleotide sequence of erysimum latent tymovirus (ELV). This sequence has been compared, in detail, with the four other known tymovirus genomic sequences. The ELV genome is the shortest of those tymoviruses. Like all known tymoviruses, it has a genome with an unusually large cytosine content and small guanine content (Symons *et al.*, 1963; Gibbs *et al.*, 1966; Keese *et al.*, 1989), and the proteins it encodes are rich in amino acids with cytosine-dominated codons such as leucine, proline, serine, and correspondingly poor in amino acids with purine-dominated codons.

The genome has three major open reading frames (ORFs), which encode the overlapping, replicase and virion proteins (OP, RP and VP , respectively). These are flanked by short 5' and 3' non-coding regions with that at the 5'-end probably capped with m^7GpppG as reported for other tymoviruses (Ahlquist and Janda, 1984; Klein *et al.*, 1976; Pleij *et al.*, 1976); the cap structure is probably not essential for translation (Shatkin, 1976), but may improve its efficiency by helping to unwind loops in the 5'-untranslated sequence (Ray *et al.*, 1985; Lawson *et al.*, 1986).

The terminal untranslated regions

The 5'-untranslated region of ELV, like comparable regions of other tymoviruses, has a smaller content of guanine and cytosine than the remainder of the genome, and in this respect resembles a number of other plant viruses (brome mosaic virus, Ahlquist *et al.*, 1981, 1984; tobacco mosaic virus, Goelet *et al.*, 1982; cowpea mosaic virus, Lomonossoff and Shanks, 1983 and van Wezenbeck *et al.*, 1983; cucumber mosaic virus, Rezaian *et al.*, 1984 and 1985; southern bean mosaic sobemovirus, Wu *et al.*, 1987).

Computer analysis, to assess possible folding of the terminal regions into stem loop structures, has shown that the 5'-end of the ELV genome may fold into two non-identical weak stem loops, but this region of the genome shows little similarity with comparable regions of the genomes of four other tymoviruses or other 'Sindbis-like

viruses' (Barker *et al.*, 1983; Bergh *et al.*, 1985; Hamilton *et al.*, 1987; Gallie *et al.*, 1988; Ding, 1989). Secondary structures in the 5'-end of the genome possibly control translation (Barker *et al.*, 1983) as loop structures in single-stranded RNAs probably control interactions between RNA and protein (Gregory and Zimmerman, 1986; Mougel *et al.*, 1987; Thomas and Nomura, 1987; Christiansen *et al.*, 1987; Romaniuk *et al.*, 1987; Turner *et al.*, 1988), and the 5'-untranslated region of TMV has been shown by mutation analysis to be an efficient enhancer of protein translation (Sleat *et al.*, 1987; Gallie *et al.*, 1988) and viral multiplication (Takamatsu *et al.*, 1991). The bias in base composition of the 5'-terminal region may also be a consequence of the secondary control structures which form in this region.

The 3'-end of the ELV genomic RNA is of great interest because it is unlikely that it is able to fold into the full five-stemmed base-paired tRNA-like cloverleaf structure proposed for the 3'-untranslated region of the genomes of TYMV and most other tymoviruses (Meshi *et al.*, 1981; Florentz *et al.*, 1982; Joshi *et al.*, 1982; Rietveld *et al.*, 1982, van Belkum *et al.*, 1987), and for cytoplasmic tRNAs (Florentz *et al.*, 1982; Joshi *et al.*, 1983a, b; Rietveld *et al.*, 1982; van Belkum *et al.*, 1988). The tRNA-like structures of tymoviruses have been shown to be aminoacylated *in vitro* by valyl-tRNA synthetase from wheat germ (Dresher *et al.*, 1988b) or yeast cells (Giège *et al.*, 1978) and *in vivo* in Chinese cabbage (Joshi *et al.*, 1982). Sequences in the 3'-terminus that are able to form tRNA-like structures have been found in the four complete tymovirus genome sequences, and also in twelve partial sequences, that include the 3'-terminal regions of APLV (Maria Osorio-Keese, unpublished data), BMV (Ding *et al.*, 1990b), CoYMV (Ding *et al.*, 1990c), CYVV (Anne Mackenzie, unpublished data), EMV-Trin (Osorio-Keese *et al.*, 1989), KYMV-BP (Anne Mackenzie, unpublished data), KYMV-JB (Ding *et al.*, 1990d), KYMV-PD (Paul Keese, unpublished data), OYMV-Tin (Ding *et al.*, 1989), TYMV-BL (Drew Meek, unpublished data), TYMV-CL (Keese *et al.*, 1989), TYMV-type (Morch *et al.*, 1988), TYMV-Cauliflower (Anne Mackenzie, unpublished data), TYMV-Roth (Marjo Torronen, unpublished data), WCMV (Anne Mackenzie, unpublished data), and OYMV-Trin (Ding *et al.*, 1989) but not DMV (Jennie Gibbs, unpublished data).

It is uncertain whether the 3'-terminal tRNA-like structures of tymoviruses are essential. On the one hand, it is clear that in some viruses they are involved in viral replication (Ahlquist *et al.*, 1984; Miller *et al.*, 1985; Bujarski *et al.*, 1986; Dreher and

Hall, 1988a, b), but the details are uncertain. Variants of TMV that lack the 3'-terminal 5-10 nucleotides are not infectious (Salomon *et al.*, 1976), and certain point mutations within the tRNA-like structure of brome mosaic are lethal (Dreher and Hall, 1988b), are not amino-acylated (Hall, 1979) and do not function as template for synthesis of the genomic minus-strand. By contrast, the fact that the 3'-terminal region of ELV probably forms only part of the tRNA-like structure found in other tymoviruses, and that it seems to be have been replaced by a poly-A tail in dulcamara mottle tymovirus, seems to indicate that a tRNA-like structure is not essential for tymovirus survival, and that other parts of the ELV or DMV genomes can replace its function.

Nonetheless it is clear that further experiments involving mutagenesis and recombination are required to establish the functions of the untranslated regions of viral genomes, and the variety of 3'-termini found among tymoviruses offers many advantages for this work.

Internal Control Regions

Comparisons reported in this thesis do not support earlier suggestions that the consensus of tymoviral translation initiation regions of RP and VP ORFs, but not the OP ORF, are resemble more closely the plant gene initiation consensus at the crucial G (position +4).

It is, however, likely that stem loop structures, that can be detected by computer algorithms, adjacent to the initiation codons of the OP and RP genes (together with the possible poor context of the OP AUG) may control translation of the ORFs (Osorio-Keese *et al.*, 1989). Kozak (1986c) suggested that very stable stem loop structures (i.e with a Gibbs energy of formation of around -50 kcal/mol) decrease translation of associated ORFs. However less stable structures, such as those found in several flaviviruses having free energies of about -20 kcal/mole, seem to affect translation (Brinton and Dispoto, 1988), and thus the possible stem loop found adjacent to the OP initiation codon of several tymoviruses, including ELV, may regulate differential translation of OPs and RPs (Osorio-Keese *et al.*, 1989).

Recently, Ding *et al.* (1990a) reported a conserved sequence, -GAGU[C/U]UGA AUUGC[U/A]UC-, just to the 5'-side of the initiation region of the VP ORF of tymoviruses, and named it the tymobox. It presumably functions as a promoter for sub-

genomic mRNA trascription, like the sequence -ACCUCUACGGCGGUCCUAAAU- found in alphavirus genomes (Ou *et al.*, 1982). Similar conserved sequences have been found in other viruses (bromoviruses, Ahlquist *et al.*, 1981; potexviruses, Zuidema *et al.*, 1989; and tobamoviruses, Solis and Garcia-Arenal, 1990), however the promoters seem to be very specific and there is little obvious homology between the promoters of viruses from different groups.

Encoded Proteins

The genome of ELV shows a strong similarity to those of other tymoviruses in its organization and its sequence, and in the sequences of the encoded proteins (Ding *et al.*, 1989 and 1990a; Keese *et al.*, 1989 and Osorio-Keese *et al.*, 1989). The tymoviruses differ from one another in the lengths of the various components of the genome; the ORFs and the non-coding regions.

The first ORF encodes a 48K OP protein, which, like other tymovirus OPs, has a high isoelectric point because of large amount of basic amino acids (his, lys and arg), reflecting the fact that it has a large number of codons with cytosine in the first position. The tymovirus OP proteins differ more in size and sequence than the RPs and VPs; the greatest sequence differences are in the C-terminal halves of the OPs. By contrast the RPs are less variable and contain the conserved ancestral polymerase domain that links all viruses (Franssen *et al.*, 1984; Haseloff *et al.*, 1984; Kamer and Argos, 1984; Hodgman, 1988; Argos, 1988).

The function of the OP is at present unknown. It is probably functional, as it is found in all tymoviruses, and Weiland and Dresher (1989) provided mutational evidence of its required expression; TYMV mutants with changes in the OP initiation codon replicated poorly in protoplasts, whereas changes in the RP initiation codon destroyed the infectivity of the virus.

It is possible that the OPs of tymoviruses control host ranges and enable them to adapt during evolution to infect new hosts, but there is no evidence for this in the comparative analyses of OP sequences reported in this thesis. However, the hydrophilic N-terminal portion of the OYMV and KYMV OPs proteins led Ding (1989) to propose that the OPs may be membrane-binding proteins, that could participate in forming the vesicles they induce in the peripheries of chloroplasts (Lesemann, 1977). Ding (1989) also suggested that the OP may be involved in

symptom production like the PrP27-30 prion protein of animals (Prusiner, 1988), the protein product of gene VI of cauliflower mosaic virus (CaMV) (Baughman *et al.*, 1988; Schoelz and Shepherd, 1988) and the cell-to-cell movement 30K protein of TMV (Deom *et al.*, 1987). However all the OPs have high isoelectric points, and thus another possibility is that this protein has a histone-like function enabling replicating tymoviral nucleic acids to pack into the vesicles in the peripherals of the chloroplast.

The RP protein (194K) of ELV is clearly homologous to the RPs of other tymoviruses, indeed the RPs are the most conserved of the tymoviral proteins. The RPs probably form the viral contribution to the viral RNA replicase (Keese *et al.*, 1989). The RPs all contain two well-known consensus sequences; the GDD domain of RNA polymerases, which are conserved in all plant and animal single-stranded plus-sense RNA-containing viruses (Argos, 1988; Gorbalenya *et al.*, 1989a, b; Poch *et al.*, 1989) and also the NTP-binding motif (Hodgman, 1988; Gorbalenya *et al.*, 1988b; Gorbalenya *et al.*, 1989). These two sequences clearly identify this protein as being involved in RNA synthesis.

The third ELV ORF, closest to the 3' terminus, encodes the virion protein or VP (21K). There is no evidence that the VP gene in the intact ELV genome is translated directly to produce VP, however the presence of the conserved sequence, named the tymobox (Ding *et al.*, 1990a), suggests that the VP is translated from a subgenomic mRNA rather than from the genomic RNA. The VP that can be deduced from the VP gene sequence has an amino acid composition (Srifah *et al.*, 1990) similar to that obtained by chemical analysis (Shukla *et al.*, 1980). ELV VP has a unique extra 11 amino acids near its N-terminus which is not represented in the sequences of the other sixteen known tymovirus VPs. Tymovirus VPs probably have an eight-stranded anti-parallel β-barrel structure (J.Varghese, personal communication) like that of the VPs found in most other simple isometric virions. The N-terminal portions of such VPs form a network on the inner surface of virion shell, and this network interacts with the encapsidated genomic RNA and seems to direct and stabilize virion assembly (Carrington *et al.*, 1987 and Hogle *et al.*, 1990). This amino terminal region is more variable in other structurally related viruses such as luteoviruses (Vincent *et al.*, 1990) and thus may account for the variability displayed by ELV.

The relationships of the VPs of ELV and other tymoviruses, based on their amino acid sequences, was found to be closely related to their reported serological

relationships (Shukla and Schmelzer, 1972; Koenig, 1976; Shukla and Gough, 1980), although the particles of ELV are serologically the most distinct of all tymoviruses, examined so far, and are only very distantly serologically related to those of APLV and OYMV (Shukla *et al.*, 1980).

The relationships of the five fully sequenced tymoviruses, based on those sequences, is closely similar whether the relationship is calculated from the OPs, the RPs or the VPs. They fall into three subgroups; EMV-OYMV, TYMV-KYMV and ELV which has its own long separate branch (Figure 25). The average branch lengths for different proteins are, however, different, reflecting differences in the rates of evolutionary change of the three proteins; the maximum FJD distance of RPs, VPs and OPs are about 34%, 54% and 90% respectively showing that the greatest rate of change is in the OPs.

There was, unfortunately, no evidence of which features of the genomic or viral protein sequences determine their host preferences. I had hoped that some sequence features shared between ELV and TYMV would indicate why both cause such similar diseases of brassicas, but attempts to correlate the amino acids at homologous positions in the five sets of proteins with the host preferences of the viruses showed no obvious regions or motifs of greater correlation. It is, of course, probable that the host range preferences are controlled by features of the sequences, but those features are probably more subtle than those which the SEQCORR program was examining.

RELATIONSHIPS WITH OTHER VIRUSES BASED ON SEQUENCE COMPARISONS

The RNA polymerase motif (Kamer and Argos, 1984) and nucleotide binding fold (Hodgman, 1988; Gorbalenya *et al.*, 1988a, b) found in the RP proteins of tymoviruses (Morch *et al.*, 1988; Keese *et al.*, 1989) are shared with many other viruses. Comparisons of these regions from a range of viruses places the tymoviruses among the Sindbis-like viruses (Goldbach, 1987; Goldbach and Wellink, 1988; Skryabin *et al.*, 1988a, b; Strauss and Strauss, 1988; Candresse *et al.*, 1990; Gorbalenya and Koomin, 1989; Habili and Symons, 1989; Poch *et al.*, 1989; German *et al.*, 1990), and, in particular, shows them to be closely related to the potexviruses (Morozov *et al.*, 1989, 1990; Rozanov *et al.*, 1990), potato virus M carlavirus

(Rupasov *et al.*, 1989; Memelink *et al.*, 1990; Morozov *et al.*, 1990) and apple chlorotic leafspot closterovirus (German *et al.*, 1990). However the genomes of the Sindbis-like viruses differ greatly in most of their features (Figure 31), and they differ more in size, terminal genomic structures and translation strategy than, say, the picornavirus group (Goldbach *et al.*, 1989). Comparisons between the Sindbis-like viruses give clear evidence that their evolution has depended crucially on genetic recombination, otherwise known as modular evolution (Botstein, 1980; Gibbs, 1987), and illustrates the evolutionary flexibility of viral genomes.

INFECTIOUS RNA TRANSCRIPTS FROM CLONES OF ELV

Full-length DNA clones encoding the genome of ELV were constructed, and I have obtained some limited evidence that these may be transcribed to produce RNA transcripts that are poorly infectious and produce atypical symptoms.

It is uncertain why these transcripts differ from the parental virus, but there are many possibilities including:

1). The presence of extra nonviral nucleotides at 5'-end of the transcripts. This has been shown to decrease the infectivity of at least nine different plant viruses, namely brome mosaic virus, BMV (Janda *et al.*, 1987), beet necrotic yellow vein, BNYVV (Bouzoubaa *et al.*, 1987), barley yellow dwarf virus (BYDV ; Young *et al.*, 1991), cowpea mosaic virus (CPMV; Vos *et al.*, 1988), plum pox potyvirus (PPV; Riechmann *et al.*, 1990), turnip crinkle virus (TCV; 14 extra nonviral residues; Heaton *et al.*, 1989), TMV (Dawson *et al.*, 1986), tobacco rattle virus (TRV; Angenent *et al.*, 1989), and tobacco vein mottle virus (TVMV; Domier *et al.*, 1989). It has been thought that these nucleotides may interfere with viral replication and thereby diminish the infectivity of *in vitro* transcripts. Nevertheless, there are several papers reporting that these short extra nonviral nucleotides at the 5'-end are not maintained during replication in plants and that the sequence at the 5'-end of the progeny RNA population is restored to wild-type (Ahlquist *et al.*, 1984; Dasmahapatra *et al.*, 1986; Dawson *et al.*, 1986; Janda *et al.*, 1987; Ziegler-Graff *et al.*, 1988; Vos *et al.*, 1988; Heaton *et al.*, 1989 and Reichman *et al.*, 1990). Thus, if these additional nucleotides or partially capped transcripts caused the reduction in infectivity of the viruses described above, they were apparently corrected in the progeny RNA. Such corrections could then

account for the restored infectivity of the progeny virus that was observed upon passage to new plants. This may be why full-length cDNA copies of TYMV-BL and TYMV-CL (Mary Skotnicki and Anne Mackenzie, personal communication), which have an extra nucleotide at the 5'-terminus, are fully infectious.

2). Inefficient capping of the 5'-terminus. The methylated cap nucleotide at the 5'-end of the genome is thought to increase the stability of the RNA genomes and enhance their translation (Shih *et al.*, 1976; Furuichi *et al.*, 1977; Shimotohmo *et al.*, 1977; Sonenberg, 1988). This cap has been shown to have this effect with several viruses including BMV (Janda *et al.*, 1987), barley stripe mosaic virus (BSMV; Petty *et al.*, 1989), BYDV (Young *et al.*, 1991) and TMV (Dawson *et al.*, 1986). Preliminary results have shown that infectivity could be increased by improving transcript capping. The infectivity of the transcripts decreased in the absence of a 5'-terminal cap homologue in TCV (Heaton *et al.*, 1989); and WClMV (Beck *et al.*, 1990), TMV strain L (Meshi *et al.* , 1986), and BMV (Janda *et al.*, 1987). No infectivity was found in its absence in other experiments involving BMV (Ahlquist *et al.*, 1984), TMV (strain U1)(Dawson *et al.*,1986), BSMV (Petty *et al.*, 1989), TRV (Angenent *et al.*, 1989; Hamilton and Baulcombe, 1989), cowpea chlorotic mottle virus (CCMV; Allison *et al.*, 1989) or TYMV-BL and TYMV-CL (Mary Skotnicki and Anne Mackenzie, personal communication). The functions of the cap is uncertain, as is that of the VPg found in other viruses, and the absence of either type in a few viruses.

3). The extra nucleotides at the 3'-terminus of the transcripts. In contrast, the addition of five to seven extra nonviral nucleotides at the 3'-end of several plant virus transcripts, including BMV (Ahlquist *et al.*, 1984), TMV (Dawson *et al.*, 1986), CPMV (Vos *et al.*, 1988), TCV (Heaton *et al.*, 1989), TYMV (Weiland and Dreher, 1989), WClMV (with 6, 14 and 198 extra residues)(Beck*et al.*, 1990), BNYVV RNA3 (Jubin *et al.*, 1990), PPV (Riechmann *et al.*, 1990) and TYMV-BL and TYMV-CL (Mary Skotnicki and Anne Mackenzie, personal communication) did not appear to affect infectivity of the transcripts. Furthermore progeny of normal length appeared during their replication. This might reflect that the ability of the viral replication machinery to recognize and/or replicate only viral sequence (Dawson *et al.*, 1986). However, when a longer 3' extension of nonviral nucleotides was added to the genome (945 nt in TMV and 2700 nt in BMV), the RNA transcripts were not infectious.

4). Sequence differences between the transcripts and the parental virus. There is always the possibility that the clone encoding the genome does not have the same sequence as the infective genome found in the parental virus isolate. This could be caused by transcriptional errors during cloning, or the production of a clone with one or more atypical 'wild-type' nucleotides, as an isolate is often itself a population of slightly different virus genomes (Domingo *et al.*, 1978). This is thought to be the reason for the variable infectivity of the transcripts obtained from BMV (Ahlquist *et al.*, 1984), CPMV (Vos *et al.*, 1988), PPV (Riechmann *et al.*, 1990) and TMV (Dawson *et al.*, 1986; Meshi *et al.*, 1986). This fact, combined with the ability of transcripts to change rapidly could explain why the progeny derived from transcript-infected plants are often so much more infectious than the *in vitro* synthesized transcripts.

Nonetheless there is great value in obtaining DNA clones encoding RNA viral genomes, as these offer great scope for research. The ELV DNA clones I have obtained may be of direct use to provide defined subgenomic fragments, but to obtain more infectious transcripts it may be necessary to replace progressively parts of the clones with newly cloned fragments. The ELV clone will then take its place alongside a wide range of viral genomes now cloned as DNA and producing fully infectious RNA transcripts. These include the following plant viruses:

Tripartite genome

Alfalfa mosaic virus	Alfalfa mosaic viruses (AlMV)	Dore and Pinck (1988)
Bromovirus	Brome mosaic virus (BMV)	Ahlquist *et al.* (1984)
	Cowpea chlorotic mottle virus (CCMV)	Allison *et al.* (1988)
Cucumovirus	Cucumber mosaic virus (CMV)	Rizzo and Palukaitis (1990)
Hordeivirus	Barley stripe mosaic virus (BSMV)	Petty *et al.* (1989)

Bipartite genome

Comovirus	Cowpea mosaic virus (CPMV)	Vos *et al.* (1988)
Tobravirus	Tobacco rattle virus (TRV)	Hamilton and Baulcombe (1989)

Monopartite genome

Carmovirus	Turnip crinkle virus (TCV)	Heaton *et al.* (1989)
Luteovirus	Barley yellow dwarf virus (BYDV-PAV serotype)	Young *et al.* (1991)
Potexvirus	Potato virus X (PVX)	Hemenway *et al.* (1990)
	White clover mosaic virus (WClMV)	Beck *et al.* (1990)
Potyvirus	Plum pox potyvirus (PPV)	Riechmann *et al.* (1990)
	Tobacco vein mottle virus (TVMV)	Domier *et al.* (1989)
Tobamovirus	Tobacco mosaic virus (TMV)	Dawson *et al.* (1986)
	TMV (tomato strain L)	Meshi *et al.* (1986)
Tombusvirus	Tomato bushy stunt virus (TBSV)	Hearne *et al.* (1990)

Tymovirus	Eggplant mosaic virus (EMV-Trin)	(Mary Skotnicki unpublished data)
	Ononis yellow mosaic virus (OYMV-Tin)	(Shouwei Ding unpublished data)
	Turnip yellow mosaic virus (TYMV)	Weiland and Dreher (1989)
	TYMV-BL	(Mary Skotnicki unpublished data)
	TYMV-CL	(Anne Mackenzie unpublished data)

Also plant viruses with DNA genomes have been cloned including, for instance, the bipartite genome of maize streak (Lazarowitz, 1988), squash leaf curl (Lazarowitz and Lazding, 1991) and tomato golden mosaic (Hamilton *et al.*, 1984) geminiviruses, as well as viruses of other types of hosts, including bacteriophage Qβ (Taniguchi *et al.*,1978), human rhinovirus (Mizutani and Colonno, 1985); poliovirus (Racaniello and Baltimore, 1981; van der Werf *et al.*, 1986), and black beetle nodavirus, Dasmahapatra *et al.*, 1986).

REFERENCES

Ahlquist, P., Dasgupta, R., and Kaesberg, P. (1984). Nucleotide sequence of the brome mosaic virus genome and its implications for viral replication. *J. Mol. Biol.* **172**, 369-383.

Ahlquist, P. and Janda, M. (1984). cDNA cloning and *in vitro* transcription of the complete brome mosaic virus genome. *Molecular and Cellular Biology* **4**, 2876-2882.

Ahlquist, P., Luckow, V., and Kaesberg, P. (1981). Complete nucleotide sequence of brome mosaic virus RNA3. *J. Mol. Biol.* **153**, 23-38.

Ahlquist, P., Strauss, E.G., Rice, C.M., Strauss, J.M., Haseloff, J., and Zimmern, D. (1985). Sindbis virus proteins nsP1 and nsP2 containing homology to nonstructural proteins from several RNA plant viruses. *J. Virology* **53,** 536-542.

Allison, R.F., Janda, M., and Ahlquist, P. (1988). Infectious *in vitro* transcripts from cowpea chlorotic mottle virus cDNA clones and exchange of individual RNA components with brome mosaic virus *J. Virol.* **62,** 3581-3588.

Allison, R.F., Janda, M., and Ahlquist, P. (1989). Sequence of cowpea chlorotic mottle virus RNAs 2 and 3 and evidence of a recombination event during bromovirus evolution. *Virology* **172,** 321-330.

Angenent, G.C., Posthumus, E., and Bol, J.F. (1989). Biological activity of transcripts synthesized *in vitro* from full-length and mutated DNA copies of tobacco rattle virus RNA-2. *Virology* **173,** 68-76

Argos, P. (1981). Secondary structure prediction of plant virus coat proteins. *Virology* **110,** 55-62.

Argos, P. (1988). A sequence motif in many polymerases. *Nucleic Acids Res.* **16**, 9909-9916.

Argos, P., Kamer, G., Nicklin, M.J.H., and Wimmer, E. (1984). Similarity in gene organization and homology between proteins of animal picornaviruses and a plant comovirus suggest common ancestry of these virus families. *Nucleic Acids Res.* **12,** 7251-7267.

Barker, R.F., Jarvis, N.P., Thompson, D.V., Loesch-Fries, L.S. and Hall, T.C. (1983). The complete nucleotide sequence of alfalfa mosaic virus RNA3. *Nucleic Acids Res.* **11**, 2881-2891.

Barradas, M.M. (1983). Identification of tomato white necrosis virus as a tymovirus. PhD Thesis, Institututo de Biociencias da U.S.P. Sao Paulo, France.

Baughman, G.A., Jacobs, J.D., and Howell, S.H. (1988). Cauliflower mosaic virus gene VI produces a symptomatic phenotype in trangenic tobacco plants. *Proc. Natl. Acad. Sci. USA* **85**, 733-737.

Beachy, R.N., Zaitin, M., Bruening, G., and Israel, H.W. (1976). A genetic map for the cowpea strian of TMV. *Virology* **73,** 498-507.

Beck, D.L., Forster, R.L.S., Bevan, M.W., Boxen, K.A., and Lowe, S.C. (1990). Infectious transcripts and nucleotide sequence of cloned cDNA of the potexvirus white clover mosaic virus. *Virology* **177**, 152-158.

Benicourt, C., Pere, J.P., and Haenni, A.L. (1978). Translation of TYMV RNA into high molecular weight proteins. *FEBS Lett.* **86**, 268-272.

Bercks, R. (1973). Scorphularia mottle virus. *CMI/AAB Descriptions of Plant Viruses.* No. 113.

Bergh, S.T., Koziel, M.G., Huang, S.C., Thomas, R.A., Gilley, D., and Siegel, A. (1985). The nucleotide sequence of tobacco rattle virus RNA-2 (CAM). *Nucleic Acids Res.* **13**, 8507-8518.

Blok, J., Gibbs, A., and Mackenzie, A. (1987a). The classification of tymoviruses by cDNA-RNA hybridization and other measures of relatedness. *Arch. Virol.* **96**, 225-240.

Blok, J., Mackenzie, A., Guy, P., and Gibbs, A. (1987b). Nucleotide sequence comparisons of turnip yellow mosaic virus isolates from Australia and Europe. *Arch. Virol.* **97**, 283-295.

Bock, K.R. and Guthrie, E.J. (1977). Clitoria yellow vein virus. *CMI/AAB Descriptions of Plant viruses*. No. 171.

Bonnet, J., Ebel, J.P., Dirheimer, G., and Shershneva, L.P. (1974). The corrected nucleotide sequence of valine tRNA from baker's yeast. *Biochimie* **56,** 1211-1213.

Both, G.W. and Air, G.M. (1979). Nucleotide sequence coding for the N-terminal region of the matrix protein of influenza virus. *Eur. J. Biochem.* **96**, 363-372.

Botstein, D. (1980). A theory of modular evolution for bacteriophages. *Ann. N.Y. Acad. Sci.* **354**, 484-491.

Bouley, J.P., Briand, J.P., Genevaux, M., Pinck, M., and Witz, J. (1976). The structure of eggplant mosaic virus: Evidence for the presence of low molecular weight RNA in top component. *Virology* **69**, 775-781.

Bouzoubaa, S., Quillet, L., Guilley, H., Jonard, G., and Richards, K. (1987). Nucleotide sequence of beet necrotic yellow vein virus RNA-1. *J. Gen. Virol.* **68**, 615-626.

Bovė, C., Macquot, B., and Bovė, J.M. (1972). Turnip yellow mosaic virus RNA synthesis in the plastids: Partial purification of a virus specific DNA independent enzyme template complex. *Symp. Biol. Hung.* **13**, 43-59.

Bovė, J.M., Bovė, C., Rondot, M.J., and Morel, G. (1965). Chloroplasts and virus synthesis. In *Biochemistry of Chloroplasts* (Proc. NATO Adv. Stud. Inst., Alberystwyth, Aug. 1965), vol.2, pp 329-339. Academic Press, New York.

Briand, J.P., Jonard, G., Guilley, H., Richards, K., and Hirth, L. (1977). Nucleotide sequence (n=159) of the amino-acid-accepting 3' OH extremity of turnip yellow mosaic virus RNA and the last portion of its coat protein cistron. *Eur. J. Biochem.* **72**, 453-463.

Briand, J.P., Keith, G., and Guilley, H. (1978). Nucleotide sequence at the 5' extremity of TYMV genome RNA. *Proc. Natl. Acad. Sci. USA* **75**, 3168-3172.

Brinton, M.A. and Dispoto, J.H. (1988). Sequence and secondary structure analysis of the 5'-terminal region of flavivirus genome RNA. *Virology*, **162,** 290-299.

Bruenn, J.A. (1991). Relationships among the positive strand and double-strand RNA viruses as viewed through their RNA-dependent RNA polymerase. *Nucleic Acids Res.* **19,** 217-226.

Brunt, A.A. (1970). Cacao yellow mosaic virus. *CMI/AAB Descriptions of Plant Viruses*. No 11.

Bujarski, J.J., Ahlquist, P., Hall, T.C., Dreher, T.W., and Kaesberg, P. (1986). Modulation of replication, aminocylation, and adenylation *in vitro* and infectivity *in vivo* of BMV RNAs containing deletions within the multifunctional 3'end. *EMBO J.* **5,** 1769-1774.

Candresse, T., Batisti, M., Renaudin, C., Mouchés, C., and Bové, J.M. (1987). Immunodetection of turnip yellow mosaic virus non-structural proteins in infected leaves and protoplasts. *Ann. Inst. Pasteur Virol.* **138,** 217-227.

Candresse, T., Morch, M.D., and Dunez, J. (1990). Multiple alignment and hierarchical clustering of conserved amino acid sequences in the replication-associated proteins of plant RNA viruses. *Res. Virol.* **141,** 315-329.

Candresse, T., Mouches, C., and Bove, J.M. (1986). Characterization of the virus encoded subunit of turnip yellow mosaic virus RNA replicase. *Virology* **152**, 322-330.

Carrington, J.C., Heaton, L.A., Zuidema, D., Hillman, B.I., and Morris, T.J. (1989). The complete genome structure of turnip crinkle virus. *Virology* **170,** 219-226.

Carrington, J.C., Morris, T.J., Stockley, P.G., and Harrison, S.C. (1987). Structure and assembly of turnip crinkle virus. IV. Analysis of the coat protein gene and implications of the subunit primary structure. *J. Mol. Biol.* **194,** 265-276.

Cellier, F., Zaccomer, B., Morch, M.D., Haenni, A.L., and Tepfer, M. (1990). Strategies for interfering *in vivo* with replication of turnip yellow mosaic virus. (Abstract) *VIIIth International Congress of Virology*, August 26-31, 1990. An Interdivisional Meeting of the International Union of Microbiological Societies (IUMS), Berlin.

Christiansen, J., Brown, R.S., Sproat, B.S., and Garrett, R.A. (1987). *Xenopus* transcription factor IIIA binds primarily at junctions between double helical stems and internal loops in oocyte 5S RNA. *EMBL J.* **6**, 453-460.

Cohen, S.S. and Greenberg, M.L. (1981). Spermidine, an intrinsic component of turnip yellow mosaic virus. *Proc. Natl. Acad. Sci. USA* **78,** 5470.

Colman, P.M., Tulloch, P.A., Shukla, D.D., and Gough, K.H. (1980). Particles and crystal symmetry of Erysimum latent virus. *J. Mol. Biol.* **142,** 263-268.

Cornelissen, B.J.C. and Bol, J.F., (1984). Homology between the proteins encoded by tobacco mosaic virus and two tricornaviruses. *Pl. Mol. Biol.* **3,** 379-384.

Cornelissen, B.J.C., Brederode, F.T., Veeneman, G.H., van Boom, J.H., and Bol, J.F. (1983). Complete nucleotide sequence of alfalfa mosaic virus RNA2. *Nucleic Acids Res.* 3019-3025.

Crestani, O.A., Kitajimi, E.W., Lin, M.T., Marinho, V.L.A., and Pimentel, J.P. (1984). A new virus disease in passion-fruit yellow mosaic cause by a tymovirus in Brazil. *Fitopatologia Brasilerira* **9,** 394.

Dale, J., Gardiner, J.E. and Gibbs, A. (1975). Viruses in *Kennedya rubicunda*. *Newsl. Aust. Pl. Path. Soc.* **4**, 13-14.

Dasmahapatra, B., Dasgupta, R., Saunders, K., Selling, B., Gallagher, T., and Kaesberg, P. (1986). Infectious RNA derived by transcription from cloned cDNA copies of the genomic RNA of an insect virus. *Proc. Natl. Acad. Sci. USA* **83,** 63-66.

Dawson, W.O., Beck, D.L., Knorr, D.A., and Grantham, G.L. (1986). cDNA cloning of the complete genome of tobacco mosaic virus and production of infectious transcripts. *Proc. Natl. Acad. USA* **83**, 1832-1836.

Dayhoff, M.O., Schwartz, R.M., Chen, H.R., and Orcutt, B.C. (1978). A model of evolutionary change in proteins. In: *Atlas of Protein Sequence and Structure*. (Dayhoff, M.O. eds), vol.5, suppl.3., National Biomedical Research Foundation, Washington D.C.

Deom, C.M., Oliver, M., and Beachy, R.N. (1987). The 30-kilodalton gene product of tobacco mosaic virus potentiates virus movement. *Science* **237**, 389-394.

Devereux, J., Haeberli, P., and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12,** 387-395.

Ding, S. (1989). Molecular biology and evolution of tymoviruses, especially ononis yellow mosaic and kennedya yellow mosaic viruses. PhD thesis, The Australian National University, Canberra, Australia.

Ding, S., Howe, J., Keese, P., Mackenzie, A., Meek, D., Osorio-Keese, M., Skotnicki, M., Srifah, P., Torronen, M., and Gibbs, A. (1990a). The tymobox, a sequence shared by most tymoviruses: Its used in molecular studies of tymoviruses. *Nucleic Acids Res.* **18,** 1181-1187.

Ding, S., Howe, J., Keese, P., Mackenzie, A., Skotnicki, M., and Gibbs, A. (1990b). Nucleotide sequence of the virion protein gene of belladonna mottle tymovirus. *Nucleic Acids Res.* **18,** 6138.

Ding, S., Mackenzie, A., Torronen, M., and Gibbs, A. (1990c). Nucleotide sequence of the virion protein gene of cacao yellow mosaic tymovirus. *Nucleic Acids Res.* **19,** 5886.

Ding, S., Keese, P., and Gibbs, A. (1990d). The nucleotide sequence of the genomic RNA of kennedya yellow mosaic tymovirus-Jervis Bay isolate: Relationships with potex-and carlaviruses. *J. Gen. Virol.* **71,** 925-931.

Ding, S., Keese, P. and Gibbs, A. (1989). Nucleotide sequence of the genome of ononis yellow mosaic tymovirus. *Virology* **172**, 555-563.

Domier, L.L., Franklin, K.M., Hunt., A.G., Rhoads, R.E., and Shaw, J.G. (1989). Infectious *in vitro* transcripts from cloned cDNA of a potyvirus, tobacco vein mottling virus. *Proc. Natl. Acad. Sci. USA* **86**, 3509-3513.

Domingo, E., Sabo, D., Taniguchi, T., and Weissmann, C. (1978). Nucleotide sequence heterogeneity of an RNA phage population. *Cell* **13,** 735-744.

Dore, J.-M. and Pinck, L. (1988). Plasmid DNA containing copy of RNA3 can substitute for RNA3 in alfalfa mosaic virus RNA inocula. *J. Gen Virol.* **69,** 1331-1338.

Dreher, T.W. and Hall, T.C. (1988a). RNA replication of brome mosaic virus and related viruses. In: *RNA Genetics* (Domingo, E., Holland, J.J., and Ahlquist, P., eds), vol.I, pp91-113, CRC Press, Boca Raton, Florida.

Dreher, T.W., and Hall, T.C. (1988b). Mutational analysis of the tRNA mimicry of brome mosaic virus RNA: Sequence and structural requirements for aminoacylation and 3'adenylation. *J. Mol. Biol.* **201,** 41-55.

Dumas, P., Moras, D., Florentz, C., Giége, R., Verlann, P., van Belkum, A., and Pleij, C.W.A. (1987). 3-D graphics modelling of the tRNA-like 3'-ends of turnip yellow mosaic virus RNA: Structural and functional implications. *J. Biomolecular Structure and Dynamics*. **4,** 707-728.

Fauquet, C.A., Monsarrat, A., and Thouvenel, J.C., (1984). Voandzeia necrotic mosaic virus. *CMI/AAB Descriptions of Plant Viruses*. No. 279.

Feng, D.F. and Doolittle, R.F. (1987). Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *J. Mol. Evol.* **25**, 351-360.

Feng, D.F., Johnson, M.S., and Doolittle, R.F. (1985). Aligning amino acid sequences: Comparison of commonly used methods. *J. Mol. Evol.* **21,** 112-125.

Florentz, C., Briand, J.P., Romby, P., Hirth, L., Ebel, J.P., and Giége, R. (1982). The tRNA-like structure of turnip yellow mosaic virus RNA: Structural organization of the last 159 nucleotides from 3' OH terminus. *EMBO J.* **1**, 269-276.

Forster, R.L.S., Bevan, M.W., Harrison, S.-A., and Gardner, R.C. (1988). The complete nucleotide sequence of the potexvirus white clover mosaic virus. *Nucleic Acids Res.* **16**, 291-303.

Fowlkes, D.M. and Shenk, T. (1980). Transcriptional control regions of the adenovirus VAI RNA gene. *Cell* **22**, 405-413.

Franssen, H., Leunissen, J., Goldbach, R., Lomonossoff, G., and Zimmern, D. (1984). Homologous sequences in non-structural proteins from cowpea mosaic virus and picornaviruses. *EMBO J.* **3**, 855-861.

Fribourg, C.E., Jones, R.A.C., and Koenig, R. (1976). Desmodium yellow mottle virus. *CMI/AAB Descriptions of Plant Viruses*. No. 168.

French, R. and Ahlquist, P. (1988) Characterization and engineering of sequences controlling *in vivo* synthesis of brome mosaic virus subgenomic RNA. *J. Virology* **62**, 2411-2420.

Furuichi, Y., La Fiandra, A., and Shaktin, A.J. (1977). 5'-terminal structure and mRNA stability. *Nature (London)* **266**, 235-239.

Gall, O.L., Candresse, T., Brault, Bretout, C., Hibrand, L., and Dunez, J. (1988). Cloning full-length cDNA of grapevine chrome mosaic nepovirus. *Genes* **73**, 67-75.

Gallie, D.R., Sleat, D.E., Watts, J.W., Turner, P.C., and Wilson, T.M.A. (1988). Mutational analysis of the tobacco mosaic virus 5'-leader for altered ability to enhance translation. *Nucleic Acids Res.* **16**, 883-893.

Gargouri, R., Joshi, R.L., Bol, J.F., Manifacier, S.A., and Haenni, A.L. (1989). Mechanism of synthesis of turnip yellow mosaic virus coat protein subgenomic RNA *in vivo*. *Virology* **171**, 386-393.

Garnett, A.L., (1973). Plantago mottle virus, a new member of the tymovirus group. *Phytopathology* **63,** 1313.

Garnier, M., Mamoun, R., and Bové, J.M. (1980). TYMV RNA replication *in vivo*, replicative intermediate is mainly single stranded. *Virology* **104,** 357.

Geneveaux,M., Pinck, M., and Duranton, H.M. (1976). Amino acid accepting structures in tymovirus RNAs. *Ann. Microbiol. (Inst Pasteur)* **127A**, 47-53.

German, S., Candresse, T., Laenneau, M., Huet, J.C., Pernolet, J.C., and Dunez, J. (1990). Nucleotide sequence and genomic organization of apple chlorotic leaf spot closterovirus. *Virology* **179,** 104-112.

Gibbs, A.J. (1978). Kennedya yellow mosaic virus. *CMI/AAB Descriptions of Plant Viruses*. No. 193.

Gibbs, A.J. (1980). How ancient are the tobamoviruses ? *Intervirology* **14**, 101-108.

Gibbs, A.J. (1987) Molecular evolution of viruses; 'trees', 'clocks' and 'modules'. *J. Cell Sci. Suppl.* **7**, 319-337.

Gibbs, A.J., Hecht-Poinar, E., and Woods, R.D. (1966). Some properties of three related viruses: Andean potato latent, dulcamara mottle, and ononis yellow mosaic. *J. gen. Microbiol.* **44**, 177-193.

Gibbs, A.J., Knowles, G.R. (1977). A method for assessing the size of a protein from its composition: A correction. *J. Gen. Virol.* **38,** 167-168.

Gibbs, A.J., and Harrison, B.D. (1973). Eggplant mosaic virus. *CMI/AAB Descriptions of Plant Viruses*. No. 124.

Giège, R., Briand, J.P., Mengual, R., Ebell, J.P., and Hirth, L. (1978). Valylation of the two RNA component of TYMV and specificity of the tRNA aminoacylation reaction. *Eur. J. Biochem.* **84,** 251-256.

Givord, l., and Koenig, R. (1974). Okra mosaic virus. *CMI/AAB Descriptions of Plant Viruses*. No. 128.

Globet, C., Prost, E., and Whalen, R.G. (1989). One-step amplication of transcripts in total RNA using the polymerase chain reaction. *Nucleic Acids Res.* **17,** 2144.

Goelet, P., Lomonosoff, G.P., Butler, P.J.G., Akam, M.E., Gait, M.J., and Kran, J. (1982). Nucleotide sequence of tobacco mosaic virus RNA. *Proc. Natl. Acad. Sci. USA.* **79**, 5818-5822.

Goldbach, R. (1986). Molecular evolution of plus-strand RNA viruses. *Ann. Rev. Phytopath.* **24,** 289-310.

Goldbach, R. (1987). Genome similarities between plant and animal RNA viruses. *Microbiol. Sci.* **4**, 197-202.

Goldbach, R., Eggen, C., de Jager, van Kammen, A., van Lent, J., Rezelman, G., and Wellink, J. (1989). Genetic organization, evolution and expression of plant viral RNA genomes. In *Recognition and Response in Plant-virus Interactions* (R.S.S. Fraser, eds) pp.147-164. Pringer-Verlag, Berlin.

Goldbach, R., and Wellink, J. (1988). Evolution of plus-strand RNA virus. *Intervirology* **29,** 260-267.

Gorbalenya, A.E., Blinov, V.M., Donchenko, A.P., and Koonin, E.V. (1989a). An NTP-binding motif is the most conserved sequence in a highly diverged monophylogenetic group of proteins involved in a positive strand RNA viral replication. *J. Molec. Evol.* **28,** 256-268.

Gorbalenya, A.E., Blinov, V.M., Donchenko, A.P., and Koonin, E.V. (1989b). Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. *Nucleic Acids Res.* **17,** 4713-4730.

Gorbalenya, A.E. and Koonin, E.V. (1989). Viral proteins containing the purine NTP-binding sequence pattern. *Nucleic Acids Res.* **17,** 8413-8440.

Gorbalenya, A.E., Koonin, E.V., Donchenko, A.P., and Blinov, V.M. (1988a). A novel superfamily of nucleoside triphosphate-binding motif containing proteins which are probably involved in duplex unwinding in DNA and RNA replication and recombination. *FEBS Lett.* **253,** 16-24.

Gorbalenya, A.E., Blinov, V.M., Donchenko, A.P., and Koonin, E.V. (1988b). A conserved NTP motif in putative helicases. *Nature* **333,** 22.

Gough, K.H., Lilley, G.G., Shukla, D.D., and Woods, E.F. (1982). Some physiocochemical properties of erysimum latent virus. *Aust. J. Biol. Sci.* **35,** 5-11.

Gregory, R.J. and Zimmermann, R.A. (1986). Site-directed mutagenesis of the binding site for ribosomal protein S8 within 16S ribosomal RNA from *Escherichia coli*. *Nucleic Acids Res.* **14**, 5761-5776.

Gubler, U. and Hoffman, B.J. (1983). A simple and very efficient method for generating cDNA libraries. *Genes* **25**, 263-269.

Guilley, H. and Briand, J.P. (1978). Nucleotide sequence of turnip yellow mosaic virus coat protein mRNA. *Cell* **15**, 113-122.

Guilley, H., Carrington, J.C., Balazs, E., Jonard, G., Richards, K., and Morris, T.J. (1985). Nucleotide sequence and genome organization of carnation mottle virus RNA. *Nucleic Acids Res.* **13**, 6663-6677.

Gupta, K. and Kingsbury, D. (1984). Complete sequences of the intergenic and mRNA start signals in the Sendai virus genome: Homologies with the genome of vesicular stomatitis virus. *Nucleic Acids Res.* **12**, 3829-3841.

Güssow, D., and Clarkson, T. (1989). Direct clone characterization from colonies by the polymerase chain reaction. *Nucleic Acids Res.* **17,** 4000.

Gustafson, G., Hunter, B., Hanau, R., Armour, S.L., and Jackson, A.O. (1987). Nucleotide sequence and genetic organization of barley stripe mosaic virus RNAγ. *Virology* **158**, 394-406.

Guy, P.L., Dale, L.J., Adena, M.A. and Gibbs, A. (1984). A taxonomic study of the host ranges of tymoviruses. *Plant Pathol.* **33**, 337-346.

Habili, N., and Symons, R.H. (1989). Evolutionary relationship between luteoviruses and other RNA plant viruses based on sequence motifs in their putative RNA polymerases nucleic acid helicase. *Nucleic Acids Res.* **17,** 9543-9555.

Hall, T.C. (1979). Transfer RNA-like structures in viral genomes. *Int. Rev. Cytol.* **60,** 1-26.

Hall, T.C., Shih, D.S., and Kaesberg, P. (1972). Enzyme-mediated binding of tyrosine to brome mosaic virus ribonucleic acid. *Biochem J.* **129,** 969-976.

Hamilton, W.D.O, and Baulcombe, D.C. (1989). Infectious RNA produced by *in vitro* transcription of a full-length tobacco rattle virus RNA-1 cDNA. *J. Gen. Virol.* **70,** 963-968.

Hamilton, W.D.O., Boccara, M., Robinson, D.J., and Baulcombe, D.C. (1987). The complete nucleotide sequence of tobacco rattle virus RNA-1. *J. Gen. Virol.* **68**, 2563-2575.

Hamilton, W.D.O., Stein, V.E., Coutts, R.H.A., and Buck, K.W. (1984). Complete nucleotide sequence of the infectious cloned DNA components of tomato golden mosaic virus: Potential coding regions and regulatory sequences. *EMBO J.* **3,** 2197-2205.

Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* **166**, 557-580.

Harrison, B.D., Finch, J.T., Gibbs, A.J., Hollings, M., Shepherd, R.J., Valenta, V., and Wetter, C. (1971). Sixteen groups of plant viruses. *Virology* **45,** 356-363.

Haseloff, J., Goelet, P., Zimmern, D., Ahlquist, P., Dasgupta, R., and Kaesberg, P. (1984). Striking similarities in amino acid sequence among nonstructural proteins encoded by RNA viruses that have dissimilar genomic organization. *Proc. Natl. Acad. Sci. USA* **81**, 4358-4362.

Hatta, T. and Matthews, R.E.F. (1976). Sites of coat protein accumulation in turnip yellow mosaic virus-infected cells. *Virology* **73**, 1-16.

Heaton, L.A. Carrington, J.C., and Morris, T.J. (1989). Turnip crinkle virus infection from RNA synthesized *in vitro*. *Virology* **170,** 214-218.

Hearne, P.Q., Knorr, D.A., Hillman, B.I., and Morris, T.J. (1990). The complete genome structure and synthesis of infectious RNA from clones of tomato bushy stunt virus. *Virology* **177,** 141-151.

Heidecker, G. and Messing, J. (1986). Structural analysis of plant genes. *Ann. Rev. Pl. Physiol.* **37,** 439-466.

Hemenway, C., Weiss, J., O'Connell, K., and Tumer, N. (1990). Characterization of infectious transcripts from a potato virus X cDNA clone. *Virology* **175,** 365-371.

Hodgman, T.C. (1988). A new superfamily of replicative proteins. *Nature* **333**, 22-23.

Holmes, D.S., and Quigley, M. (1981). A rapid boiling method for the preparation of bacterial plasmids. *Anal. Biochem.* **114,** 193-197.

Hogle, J.M., Syed, R., Yeates, T.O., Jacobson, D., and Filman, D.J., (1990). Structural determinants of serotype specificity, host range and thermostability in poliovirus. In: *Use of X-ray Crystallography in the Design of Antiviral Agents* (W.G. Laver and G.M. Air, eds.). Academic press, New York.

Huisman, M.J., Linthorst, H.J.M., Bol, J.F., and Cornelissen, B.J.C. (1988). The complete nucleotide sequence of potato virus X and its homologies at the amino acid level with various plus-stranded RNA viruses. *J. Gen. Virol.* **69**, 1789-1798.

Ishikawa, M.T., Meshi, T., Motoyoshi, F., Takamatsu, N., and Okada, Y. (1986). *In vitro* mutagenesis of the putative replicase genes of tobacco mosaic virus. *Nucleic Acids Res.* **14,** 8291-8305.

Janda, M., French, R., and Ahlquist, P. (1987). High efficiency T7 polymerase synthesis of infectious RNA from cloned brome mosaic virus cDNA and effects of 5'extensions on transcript infectivity. *Virology* **158,** 259-262.

Joshi, S., Chapeville, F., and Haenni, A.L. (1982). Length requirements for tRNA-specific enzymes and cleavage specificity at the 3'ends of the turnip yellow mosaic virus RNA. *Nucleic Acids Res.* **10,** 1945-1962.

Joshi, R.L., Joshi, S., Chapeville, F., and Haenni, A.L. (1983a). tRNA-like structures of plant viral RNAs: Conformational requirements for adenylation and aminoacylation. *EMBO. J.* **2,** 1123-1127.

Joshi, S., Joshi, R.L., Haenni, A.L., and Chapeville, F. (1983b). tRNA-like structures in genomic RNAs of plants viruses. *TIBS*, 402-404.

Joshi, S., Haenni, A.L., Hubert, E., Huez, G., Marbaix, G. (1978). *In vivo* aminoacylation and processing of turnip yellow mosaic virus RNA in *Xenopus leavis* oocytes. *Nature* **275,** 339.

Joshi, R.L., Ravel, J.M., and Haenni, A.L. (1986). Interaction of turnip yellow mosaic virus Val-RNA with eukaryotic elongation factor EF-1α: Search for a function. *EMBO J.* **5,** 1143-1148.

Jubin, I., Richards, K., Jonard, G., Guilley, H., and Pleij, C.W.A. (1990). Mapping sequences required for productive replication of beet necrotic yellow vein virus RNA 3. *Virology* **178,** 273-280.

Kamer, G. and Argos, P. (1984). Primary structural comparison of RNA-dependent polymerases from plant, animal and bacterial viruses. *Nucleic Acids Res.* **12,** 7269-7281.

Kaper, G. (1975). The chemical basic of virus structure, dissociation and reassembly. In: *Frontiers of Biology* (Neuberger, A. and Tatum, E.L., eds). North-Holland, Amterdam.

Kaper, J.M. and Steere, R.L. (1959). Isolation and preliminary studies of soluble protein and infectious nucleic acid from turnip yellow mosaic virus. *Virology* **8,** 527-530.

Keeling, J., Collins, E.R., and Matthews, R.E.F. (1979). Behavior of turnip yellow mosaic virus nucleoproteins under alkaline conditions. *Virology* **97,** 100-111.

Keese, P., Mackenzie, A., and Gibbs, A. (1989). Nucleotide sequence of an Australian isolate of turnip yellow mosaic tymovirus. *Virology* **172,** 536-546.

Keese, P., Martin, R.P., Kawchuk, L.M., Waterhouse, P.M., and Gerlach, W.L. (1990). Nucleotide sequence of an Australian and a Canadian isolate of potato leafroll

luteovirus and their relationships with two European isolates. *J. Gen. Virol.* **71,** 719-724.

Klein, C., Fritsch, C., Briand, J.P., Richards, K.E., Jonard, G., and Hirth, L. (1976). Physical and functional heterogeneity in TYMV RNA: Evidence for the existence of an independent messenger coding for the coat protein. *Nucleic Acids Res.* **3,** 3043-3061.

Koenig, R. (1976). A loop-structure in the serological classification system of tymoviruses. *Virology* **72,** 1-5.

Koenig, R. (1988). Polyhedral plant viruses with monopartite RNA genomes, introduction and summary of important properties. In: *The Plant Viruses* Vol III (Koenig, R. eds.), p1-12. Plenum Press, New York.

Koenig, R. and Givord, L. (1974). Serological interrelationships in the turnip yellow mosaic virus group. *Virology* **58,** 119-125.

Kohl, R.J. and Hall, T.C. (1974). Aminoacylation of RNA from several viruses: Amino acid specificity and differential activity of plant, yeast, and bacterial synthetase. *J. Gen. Virol.* **25,** 257-261.

Kozak, M. (1984). Point mutations close to the AUG initiation codon affect the efficiency of translation of rat prepoinsulin *in vitro*. *Nature* **308,** 214-247.

Kozak, M. (1986a). Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. *Proc. Natl. Sci. USA* **83,** 2850-2854.

Kozak, M. (1986b). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44,** 283-292.

Kozak, M. (1986c). Regulation of protein synthesis in virus-infected animal cells. *Adv. Virus Res.* **31,** 229-292.

Kozak, M. (1987). An analysis of the 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* **15,** 8125-8148.

Lafleche, D., Bove, C., Dupont, G., Mouches, C., Astier, T., Garnier, M., and Bove, J.M. (1972). Site of viral RNA replication in the cells of higher plants. TYMV (turnip yellow mosaic virus)-RNA synthesis on the chloroplast outer membrane system. *Proc. Fed. Eur. Biochem. Soc.* **72,** 43-71.

Lawson, T.G., Ray, B.K., Dodds, J.T., Grifo, J.A., Abramson, R.D., Merrick, W.C., Betsch, D.F., Weith, H.L., and Thach, R.E. (1986). Influence of 5'-proximal secondary structure on the translational efficiency of eukaryotic mRNAs and on their interaction with initiation factors. *J. Biol. Chem.* **261,** 13979-13989.

Lazarowitz, S.G. (1988). Infectivity and complete nucleotide sequence of the genome of a South African isolate of maize streak virus. *Nucleic Acids Res.* **16,** 229-249.

Lazarowitz, S.G. and Lazdins, I.B. (1991). Infectivity and complete nucleotide sequence of the cloned genomic components of a bipartite squash leaf curl geminivirus with a broad host range phenotype. *Virology* **180,** 58-69.

Lesemann, D.E. (1977). Virus group-specific and virus-specific cytological alterations induced by the members of the tymovirus group. *Phytopathol. Z.* **90,** 315-336.

Lomonossoff, G.P. and Shanks, M. (1983). The nucleotide sequence of cowpea mosaic virus B RNA. *EMBO J.* **2,** 2253-2258.

Lütcke, H.A., Chow, K.C., Michel, F.S., Moss, K.A., Kem, H.F., and Scheele, G.A. (1987). Selection of AUG initiation codons differs in plants and animals. *EMBO J.* **6,** 43-48.

Maizel, J.V., and Lenk, R.P. (1981). Enhanced graphic matrix analysis of nucleic acid and protein sequences. *Proc. Natl. Acad. Sci. USA.* **78,** 7665-7669.

Man, R., van Steeg, M., Pleij, C.W.A., and Bosch, L. (1990). Site-directed mutagenesis of the pseudoknot in the tRNA-like structure of TYMV RNA: Effects on structure and aminoacylation. (abs.). *VIII the International Congress of Virology*, August 26-31, 1990. An interdivisional meeting of the International Union of Microbiological Societies (IUMS), Berlin.

Maniatis, T., Frotsch, E.F., and Sambrook, J. (1982). *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbour Laboratory, Cold Spring Harbor, NY.

Markham, R., Matthews, R.E.F., and Smith, K.M. (1948). Specific crystalline protein and nucleoprotein from a plant virus having insect vectors. *Nature* **162**, 88-90.

Markham, R. and Smith, K.M. (1946). A new crystalline plant virus. *Nature* **157**, 300.

Markham, R. and Smith, K.M. (1949). Studies on the virus of turnip yellow mosaic. *Parasitology* **39**, 330-342.

Marsh, L.E., and Hall, T.C. (1987). Evidence implicating a tRNA heritage for the promoters of (+) strand RNA synthesis in brome mosaic virus and related viruses. *Cold Spring Harbor Symp. Quant. Biol.* **52,** 331-341.

Marsh, L.E., Pogue, G.P., and Hall, T.C. (1989). Similarity among plant virus (+) and (-) RNA termini imply a common ancestry with promoters of eukaryotic tRNAs. *Virology* **172,** 415-427.

Matthews, R.E.F. (1960). Properties of nucleoprotein fractions isolated from turnip yellow mosaic virus preparations. *Virology* **12,** 521.

Matthews, R.E.F. (1970). Turnip yellow mosaic virus. *CMI/AAB Descriptions of Plant Viruses*. No. 2.

Matthews, R.E.F. (1988). Turnip yellow mosaic virus. In : *Portraits of Viruses. A History of Virology*. (Fenner, F., and Gibbs, A. eds), pp 87-110, S.Karger AG, Basel.

Matthews, R.E.F. and Sarkar, S. (1976). A light-induced structural change in chloroplasts of Chinese cabbage cells infected with turnip yellow mosaic virus. *J. Gen. Virol.* **33**, 435-446.

Mellema, J.R., Bénicourt, C., Haenni, A.-L., Noort, A., Pleij, C.W.A., and Bosch, L. (1979). Translational studies with turnip yellow mosaic virus RNAs isolated from major and minor virus particles. *Virology* **96,** 38-46.

Melton, D.A., Krieg, P.A. Rebagliati, M.R., Maniatis, T., Zinn, K., and Green, M.R. (1984). Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Res.* **12**, 7035-7056.

Memelink, J., van der Vlugt, C.I.M., Linthorst, H.J.M., Derks, A.F.L.M., Asjes, C.J., and Bol, J.F. (1990). Homologies between the genomes of a carlavirus (lily symptomless virus) and a potexvirus (lily virus X) from lily plants. *J. Gen. Virol.* **71,** 917-924.

Meshi, T., Ishikawa, M.,Motoyoshi, F., Semba, K., and Okada, Y. (1986). *In vitro* transcription of infectious RNAs from full-length cDNAs of tobacco mosaic virus. *Proc. Natl. Acad. Sci. USA.* **83,** 5043-5047.

Meshi, T., Ohno, T., Iba, H., and Okada, Y. (1981). Nucleotide sequence of a cloned cDNA copy of TMV (cowpea strain) RNA, including the assembly origin, the coat protein cistron, and the 3' non-coding region. *Mol. Gen. Genet.* **184**, 20-25.

Mi, S., Durbin, R., Haung, H.R., Rice, C.M., and Stollar, V. (1989). Association of the sindbis virus RNA methyltransferase activity with the nonstructural protein nsP1. *Virology* **170,** 385-391.

Miller, W.A., Dreher, T.W. and Hall, T.C. (1985). Synthesis of brome mosaic virus subgenomic RNA *in vitro* by internal initiation on (-)-sense genomic RNA. *Nature* **313**, 68-70.

Mizutani, S. and Colonno, R.J. (1985). *In vitro* synthesis of an infectious RNA from cDNA clones of human rhinovirus type 14. *J. Virology* **56,** 628-632.

Morch, M.D. and Benicourt, C. (1980). Post-translational proteolytic cleavage of *in vitro*-synthesized turnip yellow mosaic virus RNA-coded high-molecular-weight proteins. *J. Virol.* **34**, 85-94.

Morch, M.D., Drugeon, G., and Benicourt, C. (1982a). Analysis of the *in vitro* coding properties of the 3' region of turnip yellow mosaic virus genomic RNA. *Virology* **119**, 193-198

Morch, M.D., Drugeon, G., Szafranski, P., and Haenni, A.L. (1989). Proteolytic origin of the 150-kilodalton protein encoded by turnip yellow mosaic virus genome RNA. *J. Virol.* **63,** 5153-5158.

Morch, M.D., Joshi, R.L., Denial, T.M., and Haenni, A.L. (1987). A new 'sense' RNA approach to block viral RNA replication *in vitro*. *Nucleic Acids Res.* **15,** 4123-4130.

Morch, M.D., Zagorski, W., and Haenni, A.L. (1982b). Proteolytic maturation of the turnip yellow mosaic virus polyprotein coded *in vitro* occurs by internal catalysis. *Eur. J. Biochem.* **127**, 259-265.

Morch, M.D., Boyer, J.C. and Haenni, A.L. (1988). Overlapping open reading frames revealed by complete nucleotide sequencing of turnip yellow mosaic virus genomic RNA. *Nucleic Acids Res.* **16**, 6157-6173.

Morozov, S.Y., Dolja, V.V., and Atabekov, J.G. (1989). Probable reassortment of genomic elements among elongated RNA-containing plant viruses. *J. Mol. Evol.* **29,** 52-62.

Morozov, S.Y., Kenyuka, K.V., Levay, K.E., and Zaviriev, S.V. (1990). The putative RNA replicase of potato virus M: Obvious sequence similarity with those of potex- and tymoviruses. *Virology* **179,** 911-914.

Mouches, C., Candresse, T., and Bove, J.M. (1984). Turnip yellow mosaic virus RNA-replicase contains host and virus-encoded subunits. *Virology* **134**, 78-90.

Mougel, M., Eyermann, F., Westhof, E., Romby, P., Expert-Bezancon, A., Ebel, J.-P., Ehresmann, B., and Ehresmann, C. (1987). Binding of *Escherichia coli* ribosomal protein S8 to 16S rRNA: A model for the interaction and the tertiary structure of the RNA binding site. *J. Mol. Biol.* **198**, 91-107.

Needleman, S.B., and Wunsch, C.D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **48,** 443-453.

Nutter, R.C., Scheets, K., Panganiban, L.C., and Lommel, S.A. (1989). The complete nucleotide sequence of the maize chlorotic mottle virus genome. *Nucleic Acids Res.* **17**, 3163-3177.

Oberg, B. and Philipson, L. (1972). Binding of histidine to tobacco mosaic virus RNA. *Biochem. Biophys. Res. Commun.* **48,** 927-932.

Osorio-Keese, M.E., Keese, P., and Gibbs, A. (1989). Nucleotide sequence of eggplant mosaic tymovirus genome. *Virology* **172**, 547-554.

Osterburg, R. and Sommer, R. (1981). Computer support of DNA sequence analysis. *Computer Programs in Biomedicine* **13,** 101-109.

Ou, J.-H., Rice, C.M., Dalgarno, L., Strauss, E.G., and Strauss, J.H. (1982). Sequence studies of several alphavirus genomic RNAs in the region containing the start of the subgenomic RNA. *Proc. Natl. Acad. Sci. USA* **79**, 5235-5239.

Paul, H.L. (1971). Belladonna mottle virus, *CMI/AAB Descriptions of Plant Viruses* No. 52.

Paul, H.L., Gibbs, A.J., and Wittman-Liebold, B. (1980). The relationships of certain tymoviruses assessed from the amino acid composition of their coat proteins. *Intervirology* **13**, 99-109.

Pearson, W.R., and Lipman, D.J. (1988). Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA*. **85,** 2444-2448.

Peter, D., and Derks, A.F.L.M. (1974). Host range and some properties of physalis mosaic virus a new virus of the turnip yellow mosaic virus group. *Neth. J. Pl. Pathol.* **80,** 124.

Petty, I.T.D., Hunter, B.G., Wei, N., and Jackson, A.O. (1989). Infectious barley stripe mosaic virus RNA transcribed *in vitro* from full-length genomic cDNA clones. *Virology* **171,** 342-349.

Pinck, M., Yot, P., Chapeville, F., and Duranton, H.M. (1970). Enzymatic binding of valine to the 3' end of TYMV-RNA. *Nature* **226**, 954-956.

Pleij, C.W.A. (1990). Pseudoknots: A new motif in the RNA genome. *TIBS* **15,** 143-147.

Pleij, C.W.A., Needleman, A., van Vloten-Doting, L., and Bosch, L. (1976). Translation of turnip yellow mosaic virus RNA *in vitro:* A closed and open coat protein cistron. *Proc. Natl. Acad. Sci. USA* **73**, 4437-4441.

Pleij, C.W.A., Rietveld, K., and Bosch, L. (1985). A new principle of RNA folding based on pseudoknotting. *Proc. Natl. Acad. Sci. USA.* **13,** 1717-1731.

Poch, O., Sauvaget, I., Delarue, M., and Tordo, N. (1989). Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J.* **8,** 3867-3874.

Pogue, G.P, Marsh, L.E., and Hall, T.C. (1990). Point mutations in the ICR2 motif of brome mosaic virus RNAs debilitate (+)-strand replication. *Virology* **178,** 152-160.

Proeseler, G., and Schmelzer, K. (1977). Laboratory and field tests about the transmission of erysimum latent virus by *Phyllotreta* species (Chrysomelidae). *Zentralbl. Bakt. Abt. II* **132,** 716-721.

Prusiner, S.B. (1988). Molecular structure, biology, and genetics of prions. *Adv. Virus. Res.* **35**, 83-136.

Racaniello, V.R. and Baltimore, D. (1981). Molecular cloning of poliovirus cDNA and determination of the complete nucleotide sequence of the viral genome. *Proc. Natl. Acad. Sci. USA.* **78,** 4887-4891.

Ralph, R.K. and Clark, M.F. (1966). Intracellular location of double-stranded plant viral ribonucleic acid. *Biochim. Biophys. Acta* **119**, 29-36.

Rezaian, M.A., Williams, R.H.V., Gordon, K.H.J., Gould, A.R., and Symons, R.H. (1984). Nucleotide sequence of cucumber mosaic virus RNA 2 reveals a translocation product significantly homologous to corresponding proteins of other viruses. *Eur. J. Biochem.* **143**, 277-284.

Rezaian, M.A., Williams, R.H.V., and Symons, R.H. (1985). Nucleotide sequence of cucumber mosaic virus RNA 1. *Eur. J. Biochem.* **150**, 331-339.

Ricard, B., Barreau, C., Renaudin, H., Mouchès, C., and Bové, J.M. (1977). Messenger properties of turnip yellow mosaic virus RNA. *Virology* **79,** 231-235.

Riechman, J.L., Laín, S., and García J.A. (1990). Infectious *in vitro* transcripts from a plum pox potyvirus cDNA clone. *Virology* **177,** 710-716.

Rietveld, K., van Poelgeest, R., Pleij, C.W.A., van Bloom, J.H., and Bosch, L. (1982). The tRNA-like structure at the 3' terminus of turnip yellow mosaic virus RNA. Differences and similarities with canonical tRNA. *Nucleic Acids Res.* **10**, 1929-1946.

Rizzo, T. and Palukaitis, P. (1990). Construction of full-length cDNA clones of cucumber mosaic virus RNAs 1, 2 and 3: Generation of infectious RNA transcripts. *Mol. Gen. Genet.* **222,** 249-256.

Robinson, D.J., Hamilton, W.D.O., Harrison, B.D., and Baulcombe, D.C. (1987). Two anomalous tobamo-virus isolates: Evidence for RNA recombination in nature. *J. Gen. Virol.* **68,** 2551-2561.

Romaniuk, P.J., Lowry, P., Wu, H.N., Stormo, G., and Uhlenbeck, O.C. (1987). RNA binding site of R17 coat protein. *Biochemistry* **26**, 1563-1568.

Rozanov, M.N., Morozov, S.Y., and Skryabin, K.G. (1990). Unexpected close relationship between the large non-virion proteins of filamentous potexviruses and spherical tymoviruses. *Genes* **3,** 373-379.

Rupasov, V.V., Morozov, S.Y., Kanyuka, K.V., and Zavriev, S.V. (1989). Partial nucleotide sequence of potato virus M RNA shows similarities to potexviruses in gene arrangement and the encoded amino acid sequences. *J. Gen. Virol.* **70,** 1861-1869.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., Erlich, H.A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487-491.

Saitou, N., and Nei, M. (1987). The neighbour-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4,** 406-425.

Salomon, R., Sela, I., Soreq, H., Giveon, D., and Littauer, U.Z. (1976). Enzymatic acrylation of histidine to tobacco mosaic virus RNA. *Virology* **71,** 74-84.

Sanger, F. Nicklen, S. and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74,** 5463-5467.

Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H., and Roe, B.A. (1980). Cloning in a single-stranded bacteriophage as an aid to rapid DNA sequencing. *J. Mol. Biol.* **143**, 161-178.

Schoelz, J., Shepherd, R.J., and Daubert, S. (1986). Region VI of cauliflower mosaic virus encodes a host range determinant. *Mol. Cel. Biol.* **6**, 2632-2637.

Schroeder, J.L., and Blattner, F.R. (1982). Formal description of a DNA oriented computer language. *Nucleic Acids Res.* **10,** 69-84.

Sellers, P.H. (1974). On the theory and computation of evolutionary distances. *SIAM J. App. Math.* **26,** 787-793.

Shatkin, A.J. (1976). Capping of eucaryotic mRNAs. *Cell* **9,** 645-653.

Shih, D.S., Dasgupta. R., and Kaesberg, P. (1976). 7-Methylguanosine and efficiency of translation. *J. Virol.* **19,** 736-742.

Shimotohno, K., Kodama, Y., Hashimoto, J., and Miura, K.I. (1977). Importance of 5'-terminal blocking structure to stabilize mRNA in eukaryotic protein synthesis. *Proc. Natl. Acad. Sci. USA* **74,** 2734-2738.

Shukla, D.D. and Gough, K.H. (1980). Erysimum latent virus. *CMI/AAB Descriptions of Plant viruses*, No. 222.

Shukla, D.D., Koenig, R., Gough, K.H., Huth, W., and Lesemann, D.-E. (1980). Erysimum latent virus-further characterization as a tymovirus. *Phytopathology* **70,** 382-384.

Shukla, D.D., Proll, E., Schmelzer, K., and Schmidt, H.B. (1973a). Studies on viruses and virus diseases of cruciferous plants. VIII. Purification, properties and serology of erysimum latent virus. *Acta Phytopathologica Academiae Scientiarum Hungaricae* **8,** 91-103.

Shukla, D.D., and Schmelzer, K. (1972). Studies on viruses and virus diseases of cruciferous plants. IV. The previously undescribed erysimum latent virus. *Acta Phytopathologica Academiae Scientiarum Hungaricae* **7,** 157-167.

Shukla, D.D., and Schmelzer, K. (1973b). Studies on viruses and virus diseases of cruciferous plants. IX.*Brassica napus* var. *napobrassica*, a sensitive and reliable local lesion host of erysimum latent virus. *Acta Phytopathologica Academiae Scientiarum Hungaricae* **8,** 105-111.

Shukla, D.D., and Schmelzer, K. (1973c). Studies on viruses and virus diseases of cruciferous plants. X. Effects of temperature on symptom expression and concentration of erysimum latent virus in *Brassica chinensis* and *B. juncea*. *Acta Phytopathologica Academiae Scientiarum Hungaricae* **8,** 113-119.

Shukla, D.D., and Schmelzer, K. (1973d). Studies on viruses and virus diseases of cruciferous plants. XI. Effects of light on symptom expression and concentration of

erysimum latent virus in *Brassica chinensis*. *Acta Phytopathologica Academiae Scientiarum Hungaricae* **8,** 121-126.

Shukla, D.D., and Schmelzer, K. (1975). Studies on viruses and virus diseases of cruciferous plants. XIX. Analysis of the results obtained with ornamental and wild species. *Acta Phytopathologica Academiae Scientiarum Hungaricae* **10,** 217-229.

Sleat, D.E., Gallie, D.R., Jefferson, R.A., Bevan, M.W., Turner, PC., and Wilson, T.M.A. (1987). Characterization of the 5'-leader sequence of tobacco mosaic virus RNA as a general enhancer of translation *in vitro*. *Genes* **60,** 217-225.

Skryabin, K.G., Kraev, A., Morozov, S.Y., Rozanov, M.N., Chernov, B.K., Lukasheva, L.I., and Atabekov, J.G. (1988a). Nucleotide sequence of potato virus X RNA. *Nucleic Acids Res.* **16,** 10929.

Skryabin, K.G., Morozov, S.Y., Kraev, A.S., Rozanov, M.V., Chernov, B.K., Lukasheva, L.I., and Atabekov, J.G. (1988b). Conserved and variable elements in RNAs genomes of potexviruses. *FEBS Lett.* **240,** 33-40.

Sneath, P.H.A. and Sokal, R.R. (1973). Numerical Taxonomy: The Principles and Practice of Numerical Classification. Freeman, W.H., San Francisco.

Solis, G. and Garcia-Arenal, F. (1990). The complete nucleotide sequence of the genomic RNA of the tobamovirus, tobacco mild green mosaic virus. *Virology* **177,** 553-558.

Sonenberg, N. (1988). Cap-binding proteins of eukaryotic messenger RNA: Functions in initiation and control of translation. *Prog. Nucleic Acids Res.* **35,** 173-207.

Srifah, P., Keese, P., Shukla, D.D., and Gibbs, A. (1990). The primary structure of the virion protein gene and encoded protein of erysimum latent virus. *J. Gen. Virol.* **71,** 3085-3088.

Staden, R. (1982). Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. *Nucleic Acids Res.* **10**, 4731-4751.

Strauss, E.G., Rice, C.M., and Strauss, J.H. (1984). Complete nucleotide sequence of the genomic RNA of Sindbis virus. *Virology* **133**, 92-110.

Strauss, J.H. and Strauss, E.G. (1988). Evolution of RNA viruses. *Ann. Rev. Microbiol.* **42**, 657-683.

Studier, J.A. and Keppler, K.L. (1988). A note on neighbour joining algorithm of Saito and Nei. *Mol. Biol. Evol.* **5**, 729-731.

Symons, R.H., Rees, M.W., Short, M.N., and Markham, R. (1963). Relationships between the ribonucleic acid and protein of some plant viruses. *J. Mol. Biol.* **6**, 1-15.

Szybiak, U., Bouley, J.P., and Fritsch, C. (1978). Evidence for the existence of a coat protein messenger RNA with the top component of each of three tymoviruses. *Nucleic Acids Res.* **5**, 1821-1831.

Tabor, S. and Richardson, C.C. (1987). DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci. USA.* **84**, 4767-4771.

Takamatsu, N., Watanabe, Y., Iwasaki, T., Shiba, T., Meshi, T., and Okada, Y. (1991). Deletion analysis of the 5'untranslated leader sequence of tobacco mosaic virus RNA. *J. Virol.* **65**, 1619-1622.

Taniguchi, T., Palmieri, M., and Weissman, C. (1978). Qβ DNA-containing hybrid plasmids give rise to Qβ phage formation in the bacterial host. *Nature (London)* **274**, 223-228.

Thomas, M.S. and Nomura, M. (1987). Translational regulation of the L1 ribosomal protein operon of *Escherichia coli* mutations that define the target site for repression by L1. *Nucleic Acids Res.* **15**, 3085-3096.

Turner, D.R., Joyce, L.E., and Butler, P.J.G. (1988). The tobacco mosaic virus assembly origin RNA: Functional characteristics defined by directed mutagenesis. *J. Mol. Biol.* **203**, 531-547.

Unger, R., Harel, D., and Sussman, J.L. (1986). DNAMAT: An efficient graphic matrix sequence homology algarithm and its application to structural analysis. *Computer Applications in the Biosciences* **2**, 283-289.

van Belkum, A., Jiang, B., Rietveld, K., Pleij, C.W.A., and Borsch, L. (1987). Structural similarities among valine-accepting tRNA-like structures in tymoviral RNAs and elongator tRNAs. *Biochemistry* **26**, 1141-1151.

van Belkum, A., Verlaan, P., Jiang, B., Pleij, C. and Bosch, L. (1988). Temperature dependent chemical and enzymatic probing of the tRNA-like structure of TYMV RNA. *Nucleic Acids Res.* **16**, 1931-1950.

van der Werf, S., Bradley, J., Wimmer, E., Studier, F.W., and Dunn, J.J. (1986). Synthesis of infectious poliovirus RNA by purified T7 RNA polymerase. *Proc. Natl. Acad. Sci. USA* **83**, 2330-2334.

van Regenmortel, M.H.V. (1972). Wild cucumber mosaic virus. *CMI/AAB Descriptions of Plant Viruses*. No. 105.

van Wezenbeck, P., Verver, J., Harmsen, J., Vos, P. and van Kammen, A. (1983). Primary structure and gene organization of the middle-component RNA of cowpea mosaic virus. *EMBO J.* **2**, 941-946.

Veidt, I., Lot, H., Leiser, M., Scheidecker, D., Guilley, H., Richards, K., and Jonard, G. (1988). Nucleotide sequence of beet western yellows virus RNA. *Nucleic Acids Res.* **16,** 9917-9933.

Verwoerd, T.C., Dekker, B.M.M., and Hoekema, A. (1989). A small-scale procedure for the rapid isolation of plant RNAs. *Proc. Natl. Acad. Sci. USA* **17,** 2362.

Vincent, J.R., Ueng, P.P., Lister, R.M., and Larkins, B.A. (1990). Nucleotide sequences of coat protein genes for three isolates of barley yellow dwarf virus and their relationship to other luteoviruses coat protein sequences. *J. Gen. Virol.* **71,** 2791-2799.

Vos, P., Jaegle, M., Wellink, J., Verver, J., Eggen, R., Kammen, A.V., and Goldbach, R. (1988). Infectious RNA transcripts derived from full-length DNA copies of the genomic RNAs of cowpea mosaic virus. *Virology* **165**, 33-41.

Weiland, J.J., and Dreher, T.W. (1989). Infectious TYMV RNA from cloned cDNA: Effects *in vitro* and *in vivo* of point substitutions in the initiation codons of two overlapping ORFs. *Proc. Natl. Acad. Sci. USA* **17,** 4675-4687.

Wu, S.X., Rinehart, C.A., and Kaesberg, P. (1987). Sequence and organization of southern bean mosaic virus genomic RNA. *Virology* **161**, 73-80.

Young, M.J., Kelly, P., Larkin, P.J., Waterhouse, P.M., and Gerlach, W.L. (1991). Infectious *in vitro* from a cloned cDNA of barley yellow dwarf virus. *J. Gen. Virol.* **70,** 2245-2251.

Yot, P., Pinck, M., Haenni, A.L., Duranton, H., and Chapeville, F. (1970). Valine specific tRNA-like structure in turnip yellow mosaic virus RNA. *Proc. Natl. Acad. Sci. USA* **67,** 1345.

Zavriev, S.K., Kanyuka, K.V., and Levay, K.E. (1991). The genome organization of potato virus M RNA. *J. Gen. Virol.* **72,** 9-14.

Ziegler-Graff, V., Bouzoubaa, S., Jubin, I., Guilley, H., Jonard, G., and Richards, K. (1988). Biologically active transcripts of beet necrotic yellow vein virus RNA-3 and RNA-4. *J. Gen. Virol.* **69,** 2347-2357.

Zuidema, D., Linthorst, H.J.M., Huisman, M.J., Asjes, C.J., and Bol, J.F. (1989). Nucleotide sequence of narcissus mosaic virus RNA. *J. Gen. Virol.* **70**, 267-276.

Zuker, M. and Steigler, P. (1981). Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.* **9**, 133-148.

Ahlquist, P., French, R., Janda, M., and Loesch-Fries, L.S. (1984). Multicomponent RNA plant virus infection derived from cloned viral cDNA. *Proc. Natl. Acad. Sci. USA* **81**, 7066-7070.