DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
57

COCKSFOOT MOTTLE VIRUS: THE GENOME ORGANISATION AND TRANSLATIONAL STRATEGIES

TIINA TAMM

TARTU 2000

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
57

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
57

COCKSFOOT MOTTLE VIRUS: THE GENOME ORGANISATION AND TRANSLATIONAL STRATEGIES

TIINA TAMM

TARTU UNIVERSITY
PRESS

Department of Molecular Biology, Institute of Molecular and Cell Biology, University of Tartu, Estonia

Dissertation is accepted for the commencement of the degree of Doctor Philosophy (in Molecular Biology) on November 11, 1999 by the Council of the Institute of Molecular and Cell Biology, University of Tartu

Opponent: Prof. Jari Valkonen (Swedish University of Agricultural Sciences, Uppsala, Sweden)

Commencement: January 21, 2000

The publication of this dissertation is granted by the Faculty of Biology and Geography, University of Tartu

Tartu Ülikooli Kirjastuse trükikoda
Tiigi 78, 50410 Tartu
Tellimus nr. 831

To my parents

CONTENTS

LIST OF ORIGINAL PUBLICATIONS

Current thesis is based on the following original publications which will be referred to by their Roman numerals:

I. **Mäkinen, K., Tamm, T., Næss, V., Truve, E., Puurand, Ü., Munthe, T. and M. Saarma.** 1995. Characterization of cocksfoot mottle sobemovirus genomic RNA and sequence comparison with related viruses. *Journal of General Virology* **76:** 2817–2825.

II. **Mäkinen, K., Næss, V., Tamm, T., Truve, E., Aaspõllu, A. and M. Saarma.** 1995. The putative replicase of the cocksfoot mottle sobemovirus is translated as a part of the polyprotein by –1 ribosomal frameshift. *Virology* **207:** 566–571.

III. **Tamm, T., Mäkinen, K. and E. Truve.** 1999. Identification of genes encoding for the cocksfoot mottle virus proteins. *Archives of Virology,* **144:** 1557–1567.

LIST OF ABBREVIATIONS

aa	amino acid(s)
BMYV	beet mild yellowing virus
BSSV	blueberry shoestring virus
BWYV	beet western yellows virus
BYV	beet yellows virus
CABYV	cucurbit aphid-borne yellows virus
cDNA	complementary DNA
CfMMV	cocksfoot mild mosaic virus
CfMV	cocksfoot mottle virus
CfMV-NO	Norwegian isolate of CfMV
CfMV-RU	Russian isolate of CfMV
CMCT	1-cyclohexyl-3-(2-morpholinoethyl) carbodiimide metho-*p*-toluene sulfonate
CnMoV	*Cynosurus* mottle virus
CP	coat protein
CYDV-RPV	cereal yellow dwarf virus RPV serotype
DMS	dimethyl sulphate
ds	double-stranded
GCFV	ginger chlorotic fleck virus
GUS	β-glucuronidase
HAstV-2	human astrovirus 2
ICTV	International Committee on Taxonomy of Viruses
kb	kilobase
kDa	kilodalton
kethoxal	β-ethoxy-α-ketobutyraldehyde
LTSV	lucerne transient streak virus
MBV	mushroom bacilliform virus
mRNA	messenger RNA
nt	nucleotide(s)
ORF	open reading frame
P1	ORF1 encoded protein
P2a	ORF2a encoded protein
P2b	ORF2b encoded protein
P3	ORF3 encoded protein
PEMV-1	pea enation mosaic virus-1
PLRV	potato leafroll virus
RCNMV	red clover necrotic mosaic virus
R-domain	random domain
RdRp	RNA dependent RNA polymerase

RNPC	ribonucleoprotein complex
RYMV	rice yellow mottle virus
RYMV-CI	Ivory Coast isolate of RYMV
RYMV-NG	Nigerian isolate of RYMV
SBMV	southern bean mosaic virus
SBMV-Ark	Arkansas isolate of SBMV
SBMV-S	resistance-breaking mutant of SBMV-Ark
SCMoV	subterranean clover mottle virus
SCPMV	southern cowpea mosaic virus
S-domain	shell or surface domain
SeMV	*Sesbania* mosaic virus
sgRNA	subgenomic RNA
SHMV	sunn-hemp mosaic virus
SNMoV	*Solanum nodiflorum* mottle virus
SoMV	sowbane mosaic virus
ss	single-stranded
TMV	tobacco mosaic virus
TNV	tobacco necrosis virus
tRNA	transfer RNA
TRoV	turnip rosette virus
UTR	untranslated region
VPg	viral genome linked protein
VTMoV	velvet tobacco mottle virus
WGE	wheat germ extract

INTRODUCTION

Most plant viruses characterised so far cause a disease in infected plants. Therefore, the majority of virus names in common use include terms that describe an important symptom in a major host or the host from which the virus was first described. Some viruses may cause mild symptoms and a moderate yield losses. Others may lead to rapid death of the whole plant and result severe epidemics, such as those causing rice tungro disease. At the same time, many viruses infect host plants symptomlessly, i.e., without producing any obvious signs of disease.

In recent years, the understanding of the plant viruses has increased rapidly in parallel with the development of molecular biology techniques. We have improved our understanding of the virion structure, the genome organisation, the gene expression, the replication, the movement, the symptom determination, the host interactions, the protein-RNA interactions, and the disease resistance.

Viruses are genetically very diverse. For the present, it is known that the viruses infecting plants may contain single-stranded RNA (ssRNA, about 75% of plant viruses), double-stranded RNA (dsRNA, e.g. phytoreo- and alphacryptoviruses), single-stranded DNA (ssDNA, e.g. gemini- and nanoviruses), or double-stranded DNA (dsDNA, e.g. caulimo- and badnaviruses) genome (Murphy *et al.*, 1995). The majority of plant viruses have ssRNA genome of messenger polarity [(+)-strand] and are classified into 49 different genera (Pringle, 1999). Plant ssRNA viruses show also a wide variation in capsid morphology ranging from the rod shaped (e.g. tobamo-, tobra-, hordei-, and furoviruses), and the filamentous (e.g. potex-, carla-, poty-, and closteroviruses) to the icosahedral viruses (e.g. bromo-, como-, polero-, sobemo-, and tymoviruses).

Eukaryotic mRNA is usually monocistronic in nature, i.e., only the first open reading frame (ORF) is translated. The genomes of (+)-strand RNA viruses are, in fact, also mRNAs. They are capable of directing protein synthesis *in vivo* and *in vitro*. Usually these genomes carry a battery of genes necessary for the viral life cycle, i.e., for the virus replication, particle formation, movement of the virus from cell to cell, interaction of the virus with its vector, and maturation of the polyprotein. Therefore (+)-strand RNA viruses express their genes through a large variety of strategies at the level of transcription as well as at the level of translation. The main strategies used by plant viruses to allow protein synthesis in an eukaryotic system from (+)-strand RNA genome containing more than one gene are: production of a subgenomic RNA (sgRNA), proteolytic cleavage of a polyprotein, overlapping genes, readthrough of leaky termination codons, ribosomal frameshifting, and genome segmentation. It is

important to note that most plant RNA viruses combine two or more of the above mechanisms for successful expression of their proteins.

Plant viruses differ from animal viruses in several aspects. Plant viruses do not enter the cell with the help of the corresponding receptors, as do animal viruses. The wounding of the plant cell wall by viral vectors, or by mechanical injuries is needed to enable the plant viruses to enter a cell. In addition, the plant viruses require specific proteins, designated movement proteins, to migrate to other cells and to spread throughout the plant. The RNA segments, that compose the genome of animal viruses, are encapsidated within the same particle. This is not the case of plant RNA viruses whose RNA segments frequently reside in distinct particles, so that a plant cell must be infected by distinct particles for the virus to multiply. Finally, plant RNA viruses are frequently accompanied by satellites, a rather rare phenomenon among animal viruses.

The genus *Sobemovirus* contains plant viruses having a single (+)-strand RNA genome encapsidated in a spherical particle. Like many other RNA viruses, expression of the sobemoviral genes is regulated at the level of translation and by selective transcription of portions of the genome. The literature review of this thesis focuses on the characterisation of sobemoviruses. An overview about the biological properties, genome organisation, protein functions, gene expression regulation, replication, particle structure and movement is given. The experimental part of this work reports the nucleotide sequence, genome organisation and gene expression of cocksfoot mottle sobemovirus.

1. REVIEW OF LITERATURE

1.1. Genus *Sobemovirus*

Sobemoviruses are plant RNA viruses named after their type member, ***so***uthern ***be***an ***mo***saic ***virus*** (SBMV). In 1969, Walters proposed to combine single-component-RNA beetle-transmitted viruses into a "southern bean mosaic virus group". In 1977, Hull made a recommendation to establish this plant virus group on the basis of similarities in protein subunit molecular weight, capsid stabilisation, sedimentation coefficient, and the distribution of the particles within the cell (Hull, 1977a). The group was accepted by the International Committee on Taxonomy of Viruses (ICTV) under the name Sobemovirus (Matthews, 1982). In 1995 ICTV recognised the group as an unassigned genus *Sobemovirus* (Hull, 1995). At the moment the genus contains 11 definitive species (Table 1) (Fauquet & Mayo, 1999). The tentative species of the genus are the cocksfoot mild mosaic virus (CfMMV), the *Cynosurus* mottle virus (CnMoV), and the ginger chlorotic fleck virus (GCFV) (Table 1) (Fauquet & Mayo, 1999). It should be noted that three other viruses recently regarded as tentative species of the *Sobemovirus* genus (maize chlorotic mottle virus, olive latent virus 1, and *Panicum* mosaic virus) have now been assigned to other virus genera. However, several viruses presently not recognised by ICTV have been proposed to be closely related to the sobemoviruses. Such viruses include, for instance, *Rottboellia* yellow mottle virus (Brunt *et al.*, 1996), ryegrass mottle virus (Brunt *et al.*, 1996), and *Sesbania* mosaic virus (SeMV) (Bhuvaneshwari *et al.*, 1995).

The viruses in the *Sobemovirus* genus are characterised by icosahedral particles of about 30 nm in diameter (Hull, 1995). The virions contain a single coat protein, approximately 30 kDa in size. The capsid is constructed according to T=3 symmetry of 180 subunits. The genome of sobemoviruses consists of one single-stranded messenger-sense RNA molecule, approximately 4 to 4.5 kb in size. The 5' terminus of the RNA has a genome-linked protein (VPg) and the 3' end lacks a poly(A) tail.

1.1.1. Biological properties of sobemoviruses

Geographical distribution, host range and transmission. The first sobemovirus to be isolated was SBMV from a garden bean in Maryland, Louisiana and California, USA (Zaumeyer & Harter, 1943). Later it was demonstrated that the sobemoviruses are spread all over the world. The cocksfoot mottle virus (CfMV), for example, has been reported to occur in Great Britain (Serjeant,

1964), Denmark (Engsbro, 1978), France (Hariri & Lapierre, 1978), Germany (Rabenstein & Schmidt, 1979), New Zealand (Mohamed, 1980), Japan (Toriyama, 1982), Norway (Norwegian isolate, CfMV-NO; Munthe, 1988), and Russia (Russian isolate, CfMV-RU; Ryabov *et al.*, 1996).

The natural host range of each species of the virus is relatively narrow (Table 1). However, sobemoviruses in general infect plant species from not less than 15 different families, both from dicot and monocot ones (Brunt *et al.*, 1996).

Four major strains of SBMV have been described, which are differentiated by the legume hosts they infect and by the symptoms they induce (Tremaine & Hamilton, 1983). All strains infect only a few species in the family *Leguminosae*. The bean strain (the type strain) of SBMV infects most bean (*Phaseolus vulgaris*) cultivars systemically and fails to infect cowpea (*Vigna unguiculata*). In some bean cultivars, e.g. cv. Pinto, it causes local lesions. The cowpea strain infects most cowpea cultivars systemically. It is not transmitted to beans except cv. Pinto, in which it is symptomless. The Ghana strain infects many cowpea cultivars and induces local or systemic symptomless infection in some cultivars of *Phaseolus vulgaris*. The severe bean or Mexican strain induces more severe symptoms in *Phaseolus vulgaris* than the bean strain. It gives systemic symptoms in the bean cv. Pinto and also infects the cowpeas. Recently, the bean strain and the cowpea strain of SBMV has been distinguished as SBMV and southern cowpea mosaic virus (SCPMV), respectively (Fauquet & Mayo, 1999). They are accepted as independent sobemovirus species due to the differences in host ranges, antigenicity, and substantial differences in sequences (Fauquet, personal communication).

The sobemoviruses are transmitted by vectors, many also through seeds, and readily mechanically (Table 1). The efficient vectors for CfMV, rice yellow mottle virus (RYMV), *Solanum nodiflorum* mottle virus (SNMoV), SBMV, SCPMV, and turnip rosette virus (TRoV) are the beetles. Blueberry shoestring virus (BSSV) is transmitted by aphids and velvet tobacco mottle virus (VTMoV) by mirids. Sowbane mosaic virus (SoMV) is transmitted by leafminers and leafhoppers.

Table 1. Biological properties of sobemoviruses.

Virus	Natural host	Vector	Seed transmission	Reference
A. Definitive species				
BSSV	*Vaccinium corymbosum, V. angustifolium*	Aphids	No	Ramsdell, 1979
CfMV	*Dactylis glomerata, Triticum aestivum*	Beetles	No	Serjeant, 1967; Mohamed & Mossop, 1981
LTSV	*Medicago sativa*	ND[a]	No	Blackstock, 1978; Forester & Jones, 1979
RYMV	*Oryza sativa, O. longistaminata*	Beetles	No	Bakker, 1974
SBMV	*Phaseolus vulgaris*	Beetles	Yes	Tremaine & Hamilton, 1983
SCMoV	*Trifolium subterraneum*	ND	Yes	Francki *et al.*, 1983
SCPMV	*Vigna unguiculata*	Beetles	Yes	Tremaine & Hamilton, 1983
SNMoV	*Solanum nodiflorum, S. nitidibaccatum, S. nigrum*	Beetles	No	Greber, 1981; Jones & Mayo, 1984
SoMV	*Chenopodium spp., C. murale, Vitis sp., Prunus domestica, Atriplex suberecta*	Leafminers, leafhoppers	Yes	Kado, 1967; 1971
TRoV	*Brassica campestris ssp. napus, Brassica campestris ssp. rapa*	Beetles	ND	Hollings & Stone, 1973
VTMoV	*Nicotiana velutina*	Mirids	No	Randles *et al.*, 1981
B. Tentative species				
CfMMV	*Phleum pratense, Dactylis glomerata, Agrostis stolonifera, Bromus mollis, Festuca pratensis, Poa trivialis*	Aphids, beetles	No	Huth & Paul, 1972
CnMoV	*Cynosurus cristatus, Lolium perenne, Agrostis tenuis, A. stolonifera*	Aphids	ND	Mohamed & Mossop, 1981
GCFV	*Zingiber officinale*	ND	ND	Thomas, 1988

[a] not determined.

Symptomatology and cytopathology. The sobemovirus infections can cause a variety of disease symptoms in plants: mild or severe chlorosis and mottling, stunting, necrotic lesions, vein clearing, and sterility of plants (Hollings & Stone, 1973; Chamberlain & Catherall, 1976; Randles *et al.*, 1981; Francki *et al.*, 1983; Hull, 1988). The symptoms vary from symptomless infections to severe disease and death of plants. For instance, in cocksfoot (*Dactylis glomerata*) and wheat (*Triticum aestivum*) CfMV causes a conspicuous yellow streaking and mottling of the leaves, and heavily infected plants are killed (Serjeant, 1967; Chamberlain & Catherall, 1976). In oat (*Avena sativa*) and barley (*Hordeum vulgare*) the symptoms are much milder but include some necrotic spotting.

The sobemoviruses have been detected most dominantly in mesophyll and vascular tissues, but have been reported also for instance from the epidermal and guard cells and bundle sheath cells (Hartman *et al.*, 1973; Rabenstein & Stannarius, 1984; Brunt *et al.*, 1996; Opalka *et al.*, 1998). In the vascular tissues, the localisation of virus particles in xylem parenchyma cells and xylem vessels seems to be characteristic, as only occasional presence in phloem parenchyma cells and sieve elements has been reported (Hartman *et al.*, 1973; Brunt *et al.*, 1996; Opalka *et al.*, 1998). However, for CfMV, it has been reported that in the vascular tissues it localises namely to phloem (Chamberlain & Catherall, 1976; Rabenstein & Stannarius, 1984).

Subcellularly, particles of sobemoviruses have been detected in the cytoplasm, cell nuclei and vacuoles of infected cells (Hull, 1977a; Francki *et al.*, 1985). These viruses also form crystalline arrays in the cytoplasm (Hull, 1977a). The cells infected with several sobemoviruses contain cytoplasmic fibrils, some of which are enveloped in the endoplasmic reticulum-derived vesicles (Francki *et al.*, 1985). The characteristic tubules, often aggregated into bundles, are found in the cells of plants infected, for example, with RYMV, BSSV, and SNMoV. The nature of these structures is unknown. No particles have been detected in either chloroplasts or mitochondria of cells infected with any of these viruses (Hull, 1988).

Resistance. In susceptible hosts, several sobemoviruses can cause severe diseases with concurrent economic losses.

CfMV first occurred in epidemic proportions at Aberystwyth in the mid-1960s (Catherall, 1985). In the middle of the 1980s, a severe outbreak of CfMV was observed in the cocksfoot fields in Norway (Rognli *et al.*, 1995). Most cultivars of cocksfoot are highly susceptible, only few of them are tolerant or resistant (Catherall, 1985; Rognli *et al.*, 1995). There are at least two types of resistance to the CfMV in cocksfoot cultivars (Caterhall, 1986). Progenies from the crosses between resistant cocksfoot genotypes and between the resistant and susceptible cocksfoot genotypes possessed (i) a greater resistance to becoming infected and (ii) either post-infection resistance or tolerance. Both types of the

resistance are clearly heritable. Although CfMV is experimentally transmitted to cereals, the infection of cereals in fields is not a serious problem (Benigno & A'Brooks, 1972).

Natural resistance against sobemoviruses has been detected also for the subterranean clover mottle virus (SCMoV) in *Trifolium subterraneum* (subterranean clover) (Wroth & Jones, 1992), for CnMoV in *Cynosurus cristatus* (crested dog's tail) (Catherall, 1985), for SBMV in *Phaseolus vulgaris* (bean) (Zaumeyer & Harter, 1943). It has been demonstrated that the resistance against SCPMV in *Vigna unguiculata* (cowpea) is controlled by a single gene according to a classical gene-to-gene interaction model (Hobbs *et al.*, 1987). There is no current knowledge of the molecular features of the resistance genes against the sobemoviruses.

Very few natural resistance sources are available for the RYMV, economically the most important sobemovirus causing a major limiting disease in Africa rice production. Here, recent advances have been reported in constructing transgenic rice plants expressing the putative RNA dependent RNA polymerase (RdRp) sequence of Nigerian isolate of RYMV (RYMV-NG) and displaying resistance to several different RYMV strains (Pinto *et al.*, 1999). The resistance achieved was based on the RNA homology-dependent resistance. It is the only reported case for the transgenic resistance against the sobemoviruses.

1.1.2. Genome organisation

The complete nucleotide sequences of several sobemoviruses have been determined (Table 2). As CfMV was under investigation in this study the genome organisation of CfMV and translational strategies used by this virus are discussed more thoroughly in section 3, "Results and discussion".

Table 2. Viruses in the genus *Sobemovirus* which complete nucleotide sequences have been reported.

Virus	Isolate	Accession number	Reference
CfMV	Norwegian	Z48630	Paper I
CfMV	Russian	L40905	Ryabov *et al.*, 1996
LTSV		U31286	Jeffries *et al.*, 1995
RYMV	Ivory Coast	L20893	Ngon A Yassi *et al.*, 1994
RYMV	Nigerian	U23142	Pinto & Baulcombe, 1995
SBMV		L34672	Othman & Hull, 1995
SBMV	Arkansas	AF055887	Lee & Anderson, 1998
SBMV	Arkansas, S[a]	AF055888	Lee & Anderson, 1998
SCPMV		M23021	Wu *et al.*, 1987

[a] resistance-breaking mutant.

The sobemoviral genome is compact and most of the predicted ORFs overlap each other. The coding region contains regularly four ORFs (Fig. 1). All sequenced sobemoviruses contain a small ORF1 at the 5' end of the genome and the 3' proximal ORF which encodes the viral coat protein (CP). Exceptionally, the nucleotide sequence of the lucerne transient streak virus (LTSV) genome available from GenBank database (Jeffries *et al.*, 1995) indicates the presence of two small ORFs at the beginning of the genome, ORF1a and ORF1b. Later, this was shown to be a sequencing error (Rathjen, personal communication). Thus, LTSV contains a single ORF1 similarly to the other sobemoviruses.

The polyprotein of the SCPMV is encoded by the large continuous ORF2 (Fig. 1) (Wu *et al.*, 1987). The genome of SCPMV contains also an internal coding region ORF3, situated in -1 reading frame within ORF2. Similar genome organisations have been reported for the Arkansas isolate of SBMV (SBMV-Ark), the Ivory Coast isolate of RYMV (RYMV-CI), and the LTSV (Ngon A Yassi *et al.*, 1994; Jeffries *et al.*, 1995; Lee & Anderson, 1998).

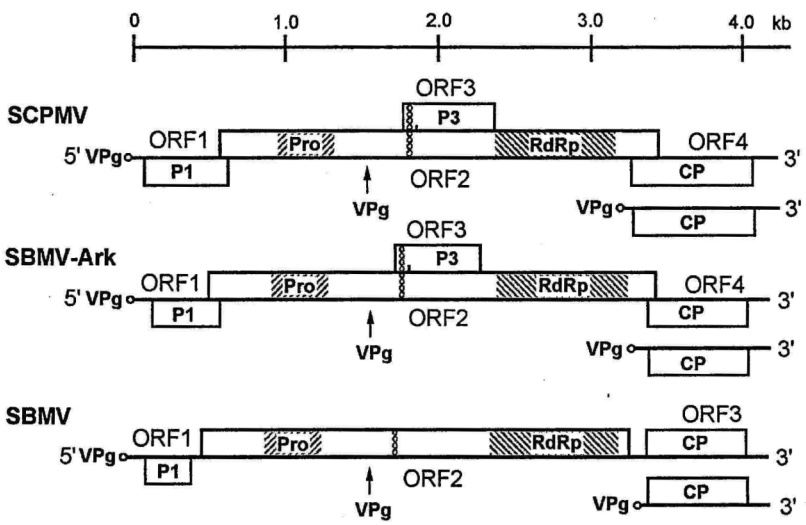


Figure 1. The genome organisation of the SCPMV, SBMV-Ark, and SBMV. A diagram of genomic and subgenomic RNAs is shown. The ORFs are illustrated as boxes. VPg, the peptide covalently attached to the 5' terminus of the RNA genome, is shown as a small circle. The approximate location of the putative protease and putative RdRp domains are marked *Pro* and *RdRp*, respectively. The position of the VPg N-terminus is indicated by an arrow. The sites of –1 ribosomal frameshifting consensus signals are shown by a chain. The position of the first AUG in ORF3 of SCPMV and SBMV-Ark is indicated by the vertical line. P1, ORF1 encoded protein; CP, coat protein.

The genome organisations of the RYMV-CI and the LTSV are similar to that of the SCPMV and the SBMV-Ark.

The genome organisation of SBMV is different from the above-mentioned (Fig. 1) (Othman & Hull, 1995). The SBMV lacks the small ORF3, characteristic for the SCPMV and the SBMV-Ark, and the three remaining ORFs do not overlap each other. Unlike the SBMV, the SBMV-Ark contains four putative overlapping ORFs, making it more similar in genome organisation (but not in amino acid sequences of individual proteins) to the SCPMV (Lee & Anderson, 1998). It has been assumed that the considerable differences in the genome organisation of the SBMV and SBMV-Ark may result from the mutations or the sequencing errors in the SBMV that resulted in mis-identification of ORFs (Lee & Anderson, 1998).

1.1.3. Sobemoviral proteins and their functions

P1. All the characterised sobemoviruses encode a small P1 protein from 5' terminal ORF. However, the nucleotide sequences of ORF1 as well as the primary structures of P1 differ among the members of the *Sobemovirus* genus. They are also unrelated to any other known proteins.

It has been hypothesised that the ORF1 of RYMV-CI could encode two polypeptides due to readthrough of the UGA stop codon at nt 553 (Ngon A Yassi *et al.*, 1994), which would be different from other viruses of this genus. Two proteins with apparent molecular masses of 18 and 19 kDa were translated *in vitro* in wheat germ extract (WGE) from ORF1 sequence of RYMV-CI and immunoreacted with P1 specific antibody (Bonneau *et al.*, 1998). From RYMV-CI infected rice plants and protoplasts, P1 was also detected as a doublet of protein. Surprisingly, *in vitro* translation of both constructs, containing ORF1 coding region with two stop codons (nt 553 and 599) or with one stop codon at nt 553 only, produced two products of 18 and 19 kDa. Therefore, it is unlikely that p19 of RYMV-CI is translated by a readthrough of the stop codon at nt 533. It was proposed that the proteins detected *in vitro* and *in vivo* most likely result from degradation, posttranslational modification(s), or from structural or other characteristics of the RYMV-CI P1 (Bonneau *et al.*, 1998).

Recently, the full-length cDNA clones of the SCPMV and the RYMV-CI were constructed and used to study the functions of P1 in the viral life cycle (Bonneau *et al.*, 1998; Sivakumaran *et al.*, 1998). Analysis of the mutants incapable to produce P1 or producing only truncated version of P1 indicated that RYMV-CI or SCPMV P1 is not needed for virus replication (Bonneau *et al.*, 1998; Sivakumaran *et al.*, 1998). At the same time, the absence of full-length P1 abolished the viral cell-to-cell and systemic movement in the cowpea and the rice plants, respectively. The RYMV-CI mutant that did not express P1 due to a mutation at the initiation codon replicated efficiently in rice protoplasts but at a level lower than wild-type cDNA transcript. Transgenic rice plants, expressing wild-type P1, were able to complement this initiation codon mutant

and produced systemic infection. These data demonstrate that one or more of the functions of P1 act *in trans* and are essential during the infection in plants.

Interesting is the fact that RYMV symptoms appeared more rapidly in P1 overexpressing plants inoculated with RYMV-CI full-length transcript than in nontransformed plants (Bonneau *et al.*, 1998). The observation that expression of P1 *in trans* enhances the infection process *in planta* suggests that P1 can act as an enhancing factor for genome amplification. Recently, it was demonstrated that P1 of RYMV-NG functions as a suppressor of posttranscriptional gene silencing (Pinto, personal communication).

Polyprotein. All the characterised sobemoviruses encode a relatively large protein with its molecular mass around 100 kDa. The genome structure of LTSV, RYMV-CI, SBMV, SBMV-Ark, and SCPMV allows the synthesis of this protein from a continuous single ORF. Based on experimental data as well as on analysis *in silico*, one can distinguish at least the following functional domains in the polyproteins of sobemoviruses — the serine protease-like domain, VPg, and the RdRp-like domain. Whether these are the only functions of the polyprotein or not, is not clear at present.

Gorbalenya *et al.* (1988) identified a putative serine protease motif in the N-terminal part of the polyprotein of SCPMV by sequence comparison with cellular and viral proteases. This putative protease is similar to the $3C^{Pro}$ cystein proteases of picornaviruses and is characteristic to all the members of the *Sobemovirus* genera, members of the *Polerovirus* [formerly known as subgroup II luteoviruses — the beet western yellows virus (BWYV, Veidt *et al.*, 1988), the potato leafroll virus (PLRV, Van der Wilk *et al.*, 1989), the cereal yellow dwarf virus RPV serotype [CYDV-RPV (formerly known as the barley yellow dwarf virus RPV serotype), Vincent *et al.*, 1991], the cucurbit aphid-borne yellows virus (CABYV, Guilley *et al.*, 1994), and the beet mild yellowing virus (BMYV, Guilley *et al.*, 1995)], and the *Enamovirus* [RNA1 of the pea enation mosaic enamovirus (PEMV-1, Demler & de Zoeten, 1991)] genera in the *Luteoviridae* family, the mushroom bacilliform barnavirus (MBV) (Revill *et al.*, 1994), and the human astrovirus 2 (HAstV-2) in *Astroviridae* family (Jiang *et al.*, 1993). The proposed protease sequence is unique among the plant viral proteases in that it resembles a cellular serine protease in possessing a serine residue instead of a cystein in the catalytic triad (Gorbalenya *et al.*, 1988; Dougherty & Semler, 1993). The consensus amino acid sequence of the catalytic triad is $\mathbf{H}(X_{32\text{-}35})\mathbf{[D/E]}(X_{61\text{-}62})\mathbf{TXXGXSG}$ (Koonin & Dolja, 1993). The glycine and histidine residues downstream from the putative catalytic residues are suggested to be a site that is involved in substrate binding. However, biochemical demonstration of protease activity and the role it plays in the sobemovirus life cycle are lacking.

A 12 kDa protein has been reported to be covalently linked to the 5' end of the SBMV genome (Ghosh *et al.*, 1981), and a 10 kDa VPg is attached to the

SCPMV genome (Mang *et al.*, 1982). Recently, the N-terminal sequence of the VPg of SBMV has been determined (Van der Wilk *et al.*, 1998). The obtained amino acid sequence started at position 327 of the ORF2 product of SBMV. Comparison of the obtained N-terminal sequence of the VPg of SBMV revealed 63% identity to residues 326-345 of the ORF2 product of SCPMV (Van der Wilk *et al.*, 1998).

The position of the VPg in the sobemovirus genome differs from the genome arrangement characteristic of the many RNA virus families, including *Picornaviridae* and *Comoviridae*: VPg-protease-RdRp (Koonin & Dolja, 1993). The N-terminal sequence of SBMV VPg places it between the putative serine protease and the putative RdRp motifs in the polyprotein. Recently, it has been shown that VPg of CfMV-RU has analogous position in CfMV-RU polyprotein (Mäkinen *et al.*, unpublished data). Similar polyprotein arrangement — protease-VPg-RdRp — has also been shown for PLRV, MBV, and PEMV-1 (Van der Wilk *et al.*, 1997; Revill *et al.*, 1998; Wobus *et al.*, 1998). As indicated, these viruses share several common features with sobemoviruses.

The C-terminal region of the sobemoviral polyprotein is predicted to encode a putative RdRp based on the presence of the GDD motif and the surrounding conserved motifs characteristic of RdRp-s (Koonin, 1991, Koonin & Dolja, 1993). The putative sobemoviral RdRp-s show extensive similarities to RdRp-s of a number of positive-stranded ssRNA viruses, which include again the poleroviruses (BMYV, BWYV, CABYV, CYDV-RPV, PLRV), the enamoviruses (PEMV-1), the barnaviruses (MBV), and the astroviruses (HAstV-2). Such similarities have been used to evaluate the taxonomic position of SCPMV in relation to other (+)-strand RNA viruses (Koonin, 1991; Koonin & Dolja, 1993). SCPMV as well as BWYV, PLRV, and PEMV-1 have been grouped into the Sobemo lineage of the polymerase supergroup 1 (Koonin & Dolja, 1993), indicating that the RdRp-s of the poleroviruses are more similar to those of the sobemoviruses than they are to those of the luteoviruses (formerly known as subgroup I luteoviruses).

P3. The genomes of LTSV, RYMV-CI, SBMV-Ark, and SCPMV have a small ORF3 nested in the middle of the ORF2 in –1 reading frame. The N-terminal half of the ORF3 encoded proteins (P3) is similar to the N-terminal part of the BWYV and PLRV ORF2 encoded proteins, and the MBV and PEMV-1 ORF3 encoded products (Ryabov *et al.*, 1996). At the moment, the function(s) and the translational mechanism of this ORF are unknown. The genome-length cDNA clone of SCPMV has been used to characterise of the phenotypes of the ORF3 mutants (Sivakumaran *et al.*, 1998). The mutant expressing the truncated P3 was not infectious in the cowpeas, indicating that P3 is needed for the SCPMV infectivity in plants. Interestingly, the mutant having a nonsense codon upstream of a potential ORF3 AUG codon, also was unable to infect the cowpeas.

Experiments in the protoplasts demonstrated that the mutations in ORF3 had no effect on the viral RNA synthesis and the SCPMV assembly.

Coat protein. The coat protein of sobemoviruses is encoded by the 3' proximally located ORF (in case of LTSV, RYMV, SBMV-Ark, and SCPMV ORF4, in case of SBMV ORF3). The amino acid sequence of SCPMV coat protein reported by Hermodson *et al.* (1982) showed that the translation initiation occurs at the second AUG (nt 3271–3273) of the ORF4 and is followed by the post-translational replacement of the N-terminal methionine with an acetyl group. The direct sequencing of the N-terminus of RYMV-CI coat protein showed that it commences at the first AUG codon of the last ORF at nt 3447 (Ngon A Yassi *et al.*, 1994).

The tentative phylogenetic tree generated by aligning the coat protein sequence of SCPMV together with several other viruses consisted of three distinct subdivisions (Dolja & Koonin, 1991). Interestingly, the coat protein of SCPMV grouped together with the tobacco necrosis necrovirus (TNV) coat protein rather than with the coat proteins of poleroviruses. The coat protein of RYMV-CI is also more closely related to that of TNV (Ngon A Yassi *et al.*, 1994), indicating that coat proteins of sobemoviruses and coat proteins of necroviruses are phylogenetically related (Dolja & Koonin, 1991). This is different from the sobemoviral putative proteases, VPg-s, and the putative RdRp-s, which are most closely related to poleroviruses, as indicated above.

Sobemoviral particles are found in the nuclei of infected cells (Hull, 1977a; Francki *et al.*, 1985). Ngon A Yassi *et al.* (1994) noted that the N-terminal part of RYMV-CI coat protein (aa 3–22) contains the sequence which is identical to the bipartite nuclear targeting motif (Dingwall & Laskey, 1991). The similar bipartite nuclear targeting motif can be found at the N-terminus of all the sobemoviral coat proteins (Ngon A Yassi *et al.*, 1994; Paper I; Lee & Anderson, 1998). This finding may explain the observation why have the virus particles been found in the nuclei of the infected cells during the sobemovirus infection. Except for this observation, no molecular determinants have been attributed to the subcellular or tissue-specific localisation of sobemovirus particles.

The full-length cDNA clones of RYMV-CI and SCPMV have been used to study the function(s) of coat protein (Brugidou *et al.*, 1995; Sivakumaran *et al.*, 1998). The tested mutants (C-terminal deletion and frameshift mutants of RYMV-CI; initiation codon and insertion mutations of SCPMV) were not infectious in plants. No virus accumulation was detected in the inoculated or systemic leaves, indicating that the coat protein is essential for the cell-to-cell and the systemic virus movement. At the same time, the coat protein was not required for the RNA synthesis of RYMV-CI and SCPMV in the rice and cowpea protoplasts, respectively. In rice plants, RNA replication for both RYMV-CI mutants was detected in inoculated leaves four weeks after inocula-

tion, indicating the importance of the coat protein dominantly in the virus long distance movement.

Direct evidence of the coat protein determining the host range of the sobemoviruses was provided by characterising the resistance-breaking mutant of SBMV-Ark, SBMV-S (Lee & Anderson, 1998). SBMV-S is able to move systemically in bean cvs. Pinto and Great Northern, although the wild-type SBMV-Ark is restricted to the inoculated leaves in this host. Sequence analysis of the genomes of SBMV-Ark and SBMV-S revealed seven nucleotide differences, but only four deduced amino acid changes. Three amino acid changes were identified in the R-domain of the virus coat protein (see also section 1.1.6). The changes in the R-domain of coat protein may have an effect on specific coat protein-RNA interaction needed for correct capsid assembly/disassembly and through that determine the host range of the virus.

1.1.4. Translational control of gene expression

The first data on proteins encoded by sobemoviruses (besides structural coat protein) were obtained from *in vitro* translation experiments. The RNAs of several sobemoviruses have been translated in the rabbit reticulocyte lysate and WGE systems. Both RNAs of SBMV and SCPMV induce the translation of four major proteins in the cell-free systems: 105, 75, 29, 14 kDa, and 100, 70, 30, 20 kDa, respectively (Saleno-Rife *et al.*, 1980; Mang *et al.*, 1982). Four polypeptides can also be translated from CfMV-NO (Papers II and III), LTSV (Morris-Krsinich & Forster, 1983), SNMoV (Kiberstis & Zimmern, 1984), and TRoV (Morris-Krsinich & Hull, 1981) RNAs, displaying only slight differences in their molecular weights (Table 3).

Table 3. *In vitro* translation products of some sobemoviruses in cell-free systems.

Virus	*In vitro* translation products					Reference
	Poly-protein (kDa)	Unknown (kDa)	Unknown (kDa)	Coat protein (kDa)	Unknown (kDa)	
CfMV	100	71		34	12	Papers II and III
LTSV	105	78		33	18	Morris-Krsinich & Forster, 1983
SBMV	105	75		29	14	Salerno-Rife *et al.*, 1980; Mang *et al.*, 1982
SCPMV	100	70		30	20	Mang *et al.*, 1982
SNMoV	100	67		38	28	Kiberstis & Zimmern, 1984
TRoV	105	67	35	30		Morris-Krsinich & Hull, 1981

Previous studies have demonstrated that the 100 kDa and 70 kDa proteins of SCPMV are related to each other and are translated from the full-length RNA, and the 20 kDa protein is presumably encoded by ORF1 (Ghosh *et al.*, 1981; Mang *et al.*, 1982; Wu *et al.*, 1987). It has also been proposed that the polyprotein, encoded by ORF2 of SCPMV is processed by proteolytic cleavage to give the 70 kDa translation product (Wu *et al.*, 1987). The 30 kDa protein (viral coat protein) was shown to be translated from a smaller, sgRNA (Rutgers *et al.*, 1980; Ghosh *et al.*, 1981; Mang *et al.*, 1982).

1.1.4.1. Subgenomic RNA

The expression of internal genes of the (+)-strand RNA viruses is frequently mediated *via* sgRNAs. In general, sgRNAs are the mRNAs for the 3' proximal genes of the polycistronic viral RNAs, and are identical in sequence to the 3' end of the genomic RNA.

A sgRNA has been detected in the sobemovirus infected tissues as well as in the virus particles. For instance both SBMV and SCPMV encapsidate the subgenomic component into viral particles (Rutgers *et al.*, 1980; Weber & Sehgal, 1982). The sgRNA of 0.3-0.4 X 10^6 of SCPMV and SBMV has the VPg linked to its 5' end as does the genomic RNA (Ghosh *et al.*, 1981; Mang *et al.*, 1982). Sobemoviral RNAs smaller than genomic length have also been reported for CfMV and RYMV-CI (Paper I; Ryabov *et al.*, 1996; Bonneau *et al.*, 1998).

The 5' end of the sgRNA can be determined by direct RNA sequencing, primer extension or RNAse protection experiments. The primer extension experiments are used to identify the precise 5' ends of the sobemoviral sgRNAs. The 5' ends of SCPMV and SBMV sgRNAs map to the bases 3241 and 3163, respectively (Hacker & Sivakumaran, 1997) and based on the sequences of SCPMV and SBMV genomes (Wu *et al.*, 1987; Othman & Hull, 1995), are possessing the 5' end primary sequence ACAAAA. At the same time, the 5' terminal nucleotides of SCPMV genomic RNA were also identified as ACAAAA (Hacker & Sivakumaran, 1997). The 5' ends of the genomic RNAs of the two other sobemoviruses, RYMV-CI and LTSV, have been reported to be ACAA and ACAAA, respectively (Ngon A Yassi *et al.*, 1994; Jeffries *et al.*, 1995). The 5' ends of SBMV and SBMV-Ark genomic RNAs start with the similar motif, CACAAAA (Othman & Hull, 1995; Lee & Anderson, 1998). All sobemoviruses are also characterised to have polypurine track, including the sequence aAGgAAA just in the beginning of the genomic RNA (Hacker & Sivakumaran, 1997).

Using the similarity between the 5' ends of genomic and subgenomic RNAs it is possible to predict the 5' termini of as yet uncharacterised sobemoviral sgRNAs. The sequence ACAAAA (nt 3222-3227) is present upstream of the SBMV-Ark coat protein initiation codon. For RYMV-CI, the sequence

ACAAA (nt 3441–3445) is located 6 nt upstream from the ORF4 AUG codon. The putative transcriptional start site for LTSV sgRNA has also ACAAAA sequence motif (nt 3285–3290).

Similarities between the 5' ends of the sobemoviral genomic and subgenomic RNAs indicate that the translation initiation from these RNAs may occur by a similar mechanism. The addition of an AUG codon to the 5' UTR of the genomic RNA substantially reduced translation of P1 compared to its translation from the wild-type RNA (Hacker & Sivakumaran, 1997). Similarly, the additional AUG codon in the 5' UTR of the sgRNA of SCPMV reduced translation of the coat protein about 80% relative to its translation from the wild-type sgRNA, indicating that the translation of this protein occurs by a 5' end-dependent scanning mechanism rather than by the internal ribosome binding.

1.1.4.2. Initiation of translation: leaky scanning

A scanning model has been proposed to explain the mechanism by which eukaryotic mRNAs initiate translation (Kozak, 1989). According to this model, the 40S ribosomal subunit selects the initiation codon for the protein synthesis as it scans the RNA beginning at the 5' end. Two modifications are introduced to the model to explain the translation of the polycistronic eukaryotic mRNAs. "Leaky scanning" involves a mechanism in which a portion of the ribosomes or initiation complexes bypass the first AUG which is in an unfavourable context for the translation initiation and initiate the translation from downstream AUG codons (Kozak, 1986; 1989). Internal initiation codon may also become accessible if, after the termination of translation of the preceding cistron, some of the ribosomes remain associated with the mRNA and continue scanning until another AUG is encountered ("termination-reinitiation") (Kozak, 1989). Two independently initiated proteins are thus produced from the same mRNA. In either case, the scanning model excludes the third possibility of independent internal ribosome binding on eukaryotic mRNAs at a sequence near the internal start codon ("internal ribosome entry"). The internal initiation mechanism has been most extensively studied for picornaviruses (Belsham & Sonenberg, 1996).

Translation of sobemoviral polyprotein is not initiated from the corresponding sgRNA, as no such RNA has been found. Experimental evidence supports the idea that the genomic RNA of sobemoviruses functions as bicistronic messenger RNA: the translation of the ORFs 1 and 2 occurs from the genomic RNA by initiation at their respective AUG codons. Products of the expected size from both ORFs were observed after translation of the genomic RNA of SCPMV (Sivakumaran & Hacker, 1998) and CfMV-NO (Paper III).

Table 4. Sequence context[a] of start codons of overlapping ORFs on plant viral RNAs.

Virus	First ORF[b]	Second ORF[b]	Reference
A. SOBEMOVIRUSES			
Host dicot			
SBMV	C UG**C** AUG AG	G **ACA** AUG UA	Othamn & Hull, 1995
SBMV-Ark	C UG**C** AUG AG	G **ACA** AUG UA	Lee & Anderson, 1998
SCPMV	U U**UC** AUG AU	G **AGA** AUG UA	Wu *et al.*, 1987
LTSV	C UGU AUG CC	G **AAC** AUG CU	Jeffries *et al.*, 1995
Host monocot			
CfMV	U UAG AUG UG	A **AGA** AUG **GG**	Paper I; Ryabov *et al.*, 1996
RYMV-CI	G UGU AUG **AC**	C **GGG** AUG **GG**	Ngon A Yassi *et al.*, 1995
B. POLEROVIRUSES			
Host dicot			
BMYV	G UCU AUG CA	G **AGA** AUG AA	Guilley *et al.*, 1995
BWYV	G UUG AUG CA	A **AUA** AUG UA	Veidt *et al.*, 1988
CABYV	G GUG AUG CA	A UCA AUG **GC**	Guilley *et al.*, 1994
PLRV	G CAU AUG AU	A **AUC** AUG AA	Van der Wilk *et al.*, 1989
Host monocot			
CYDV-RPV	A CGC AUG UU	C **AAA** AUG AA	Vincent *et al.*, 1991
C. OTHER VIRUSES			
Host dicot			
PEMV-1	G UUU AUG CA	C UAG AUG **GC**	Demler & de Zoeten, 1991
Host fungi			
MBV	C UAU AUG AA	A AUA AUG UC	Revill *et al.*, 1994

[a] Consensus sequences of **[A/G]CN AUG GC** in monocots and **AN[A/C] AUG GC** in dicots were identified by Cavener and Ray (1991);

[b] AUG codons are underlined, the bases matching to the consensus are in bold.

A comparison between the sequence surrounding the initiation codon for the ORF1 of sobemoviruses and the consensus sequences established for the plant mRNAs shows that the sequence surrounding the first AUG codon is in poor context for translation by plant ribosomes (Table 4). The poor context is determined by the presence of a pyrimidine at position –3 and the absence of a G at position +4 (Lütcke *et al.*, 1987; Cavener & Ray, 1991). In contrast, the ORF2 initiation codons are present in a more favourable context for the translation by having a purine at position –3 (Table 4). The 5' terminal half of the sobemovirus genome resembles in its organisation the genome of poleroviruses, enamoviruses and barnaviruses. In all of theses cases, the start codon of the first ORF is flanked by suboptimal bases compared to those flanking the

second ORF start site (Table 4) (Demler & de Zoeten, 1991; Miller *et al.*, 1995; Revill *et al.*, 1998).

Sivakumaran and Hacker (1998) further investigated the mechanism of SCPMV ORF2 translation initiation. *In vitro* and *in vivo* studies showed that the addition of one or two AUG codons upstream of ORF2 AUG codon in a favourable sequence context for translation initiation reduced ORF2 expression. The elimination of the ORF1 initiation codon resulted in an increase in ORF2 expression. *In vivo* studies demonstrated that the addition of 19 AUG codons to the 5' UTR abolished ORF2 expression, and the placement of the ORF1 initiation codon within an optimal sequence context for translation initiation reduced ORF2 expression. These results indicate that the ORF2 of SCPMV is translated by leaky scanning rather than by internal ribosome binding or coupled termination-reinitiation.

The cotranslational disassembly of destabilised SBMV was suggested as a mechanism for uncoating of viral nucleic acid (Wilson, 1985). In contrast to some other isometric viruses, which appear to release their RNA rapidly prior to translation, the particles of the Ghana strain of SBMV disassemble only after their RNA has initiated translation (Shields *et al.*, 1989). According to this hypothesis, the ribosome has to "find" the 5' end of the RNA and the translation of the ORFs starts. Further removal of the coat protein subunits occurs as ribosome translocation proceeds. The proposed model supports the finding that the leaky scanning is the mechanism how the ORF2 of sobemoviruses is translated, as the ribosomes must start from the 5' end of the RNA.

1.1.4.3. Proteolytic polyprotein processing

Proteolytic processing of a precursor polyprotein is a translation strategy employed by many viruses to produce more than one protein from a single mRNA. Since the proteases form a part of the polyprotein, the cleavage of the precursor can occur in *cis* and/or in *trans*. The advantage of this strategy is that several functional proteins can be produced from a minimum of genetic information. The disadvantage of using polyprotein processing for genome expression is that the proteins derived from a given polyprotein will be present in equimolar amounts.

The identification of the proteolytic cleavage sites at the VPg N-terminus of SBMV and CfMV-RU gives the first indication how the sobemoviral polyprotein is processed *in vivo*. It was proposed by Gorbalenya *et al.* (1988) that the sobemoviral serine protease could cleave at E/T or E/S sites. In SBMV polyprotein sequence the VPg N-terminal residue (threonine) is preceded by a glutamic acid residue, indicating that the N-terminal proteolytic processing site consists of the residues E/T (Van der Wilk *et al.*, 1998). In contrast, the

N-terminal sequence of VPg of CfMV-RU indicates that in case of CfMV-RU the cleavage site used is E/N (Mäkinen *et al.*, unpublished data). This finding is not in full accordance with the previously predicted cleavage sites and more experimental data is needed to propose the polyprotein processing model for sobemoviruses.

1.1.5. Replication of sobemoviruses

Little is known about the replication signals needed for the initiation of (+)-strand and (–)-strand synthesis of sobemoviruses. The genomic and subgenomic RNAs of the sobemoviruses start with very similar primary sequences (cACAAaa) which are potentially important for the virus replication. The ACAAAa sequence is also present at the 5' termini of genomic and subgenomic RNAs of poleroviruses (Miller *et al.*, 1995). The same sequence motif is found at the 5' end of the red clover necrotic mosaic virus (RCNMV) RNA1 and at the 5' end of its sgRNA (Xiong & Lommel, 1989; Zavriev *et al.*, 1996). The 5' termini of MBV genomic and subgenomic RNAs also begin with this sequence (Revill *et al.*, 1994; Revill *et al.*, 1999). The conservation of this sequence at the 5' ends of genomic and subgenomic RNAs of several viruses, belonging to different groups suggests that it or its complementary sequence in (–)-strand RNA may function in viral RNA synthesis. It has been proposed that the (–)-strand sequence complementary to the ACAAA domain may act as a promoter or enhancer for viral replicase binding and initiation of RNA synthesis (Miller *et al.*, 1995). So far the role(s) of the above sequences in sgRNA and genomic RNA synthesis have not been tested in sobemo- and poleroviruses.

Even less is known about the replication signals at the 3' end of sobemovirus genomic RNAs needed for the initiation of the (–)-strand synthesis. It has long been known that the RNA genomes of certain plant viruses have tRNA-related properties (Mans *et al.*, 1992). These 3' tRNA-like structures have been shown to be involved in (–)-strand synthesis in the case of the brome mosaic bromovirus, the tobacco mosaic tobamovirus (TMV), and the turnip yellow mosaic tymovirus (Miller *et al.*, 1985; Dawson *et al.*, 1988; Tsai & Dreher, 1991). A potential tRNA-like structure has been attributed to the 3' end of some sobemoviral genomic RNAs (Ngon A Yassi *et al.*, 1994; Ryabov *et al.*, 1996). However, for SCPMV and SBMV it has been impossible to find a 3' end RNA sequence potentially folding to a tRNA resembling secondary structure (Wu *et al.*, 1987; Othman & Hull, 1995). It should be emphasised that these reports are based on computer modelling. No experimental data are available on RNA secondary structures characteristic to 5' or 3' ends of sobemoviral RNAs.

1.1.6. Particle structure and viral assembly

Coat protein is the single protein to build up sobemoviral isometric particles. The three-dimensional structure of SCPMV has been determined to 2.8 Å resolution (Abad-Zapatero *et al.*, 1980), and that of SeMV to 3 Å resolution (Bhuvaneshwari *et al.*, 1995). Each icosahedral unit of a particle comprises of three quasi-equivalent subunits A, B, and C whose individual conformations might be slightly different. The A subunits cluster about the fivefold axes, whereas sets of three B and three C subunits cluster about quasi-sixfold vertices (Fig. 2A).

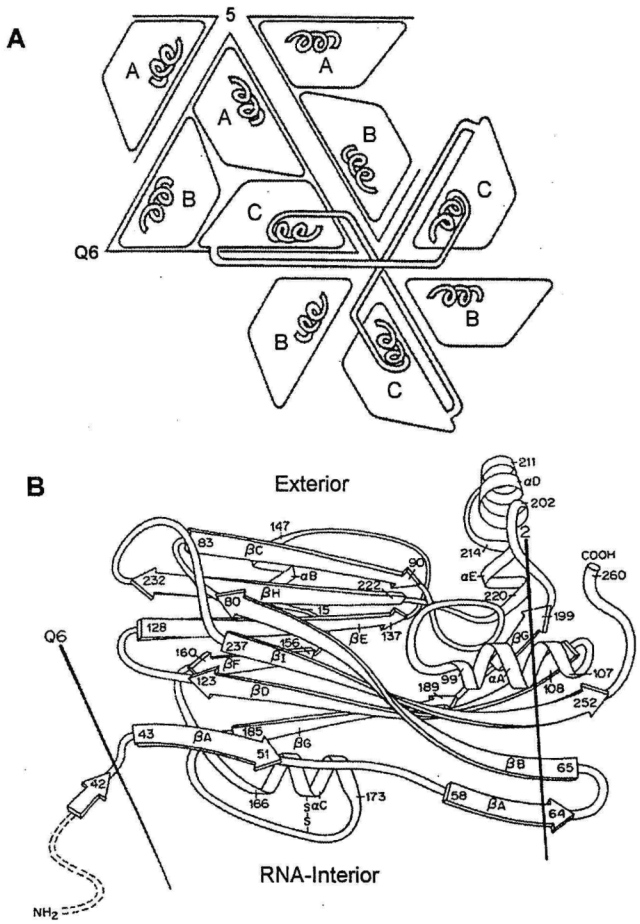


Figure 2. A. Arrangement of subunits A, B, and C within one icosahedral asymmetric unit as viewed from the inside of the virus, looking outward. This view permits a representation of the amino ends of the C subunits (adapted from Abad-Zapatero *et al.*, 1980). **B.** Diagrammatic representation of the SCPMV coat protein fold showing secondary structural elements (adapted from Hermodson *et al.*, 1982).

According to the X-ray structure, the coat protein comprises of two functional domains, the random domain (R-domain) and shell or surface domain (S-domain) connected by an arm (Fig. 2B) (Abad-Zapatero *et al.*, 1980; Hermodson *et al.*, 1982; Rossmann *et al.*, 1983). The S-domain is composed of eight antiparallel β-sheets (termed the β-barrel) and five α-helices and is responsible for subunit-subunit interactions in the virus particle (Hermodson *et al.*, 1982; Rossmann *et al.*, 1983). The R-domain is formed by the N-terminal part of the polypeptide chain and is rich in arginines, lysines, prolines, and glutamines. The basic residues located on the R-domain (together with some similar residues on the inner surface of the S-domain) are responsible for the coat protein contacts with RNA (Hermodson *et al.*, 1982; Rossmann *et al.*, 1983). The pattern of the basic residues on the coat protein surface facing the RNA is able to dock a nine-basepair double-helical A-RNA structure with surprising accuracy. The basic residues are each associated with a different phosphate and the protein can make interactions with five bases in the minor grove. The total number of positive charges associated with the RNA is around 2340, sufficient to cancel about half of the negative charge of the nucleic acid. The interactions between coat protein and viral RNA have been studied using an electrophoretic mobility shift assay (Lee & Hacker, 1999). A recombinant R-domain of the coat protein of SCPMV (aa 1–54) expressed in *E. coli* was shown to have nonspecific RNA binding activity *in vitro*. Alanine substitution mutants showed that the N-terminal arginine-rich motif (aa 22–30) was required for RNA binding.

Hull (1977b) studied the particle stabilisation of TRoV and other sobemoviruses and concluded that the particles of these viruses are stabilised by three types of bonds — pH-dependent interactions between subunits, protein-RNA interactions, and the divalent cation protein-protein bonds. The important cations forming protein-protein bonds in the particle are calcium and magnesium ions. The major Ca^{2+}-binding site is on the quasi-threefold axis between the A, B, and C subunits (Hermodson *et al.*, 1982). The sequence shows that Glu 194 is the ligand associated with this position. Interestingly, in SBMV this residue is lysine (Mang *et al.*, 1982), suggesting a different mode of subunit association. The second Ca^{2+} site is also between the quasi-three-fold-related subunits interacting with Asp138 and Asp141 on one subunit and main chain carbonyls 199 and 259 of the other. The proposed Mg^{2+}-binding sites contain residues His 132, Glu 229, and Glu 77 (Rossmann, 1984). The virus swells on removal of cations at alkaline pH (Hull, 1977b).

Savithri and Erickson (1983) demonstrated the importance of RNA-coat protein interactions in the assembly of SCPMV. The assembly of T=3 particles from SCPMV coat protein and viral RNA requires calcium at neutral or alkaline pH. Low molecular weight RNA component and coat protein formed T=1 particles at acidic and neutral pH. It is possible to convert swollen SBMV into T=1 particles by digestion with trypsin, which removes the basic

N-terminal segment of coat protein (Sehgal *et al.*, 1979). Erickson and Rossmann (1982) have also shown that T=1 particle formation did not require RNA when purified and partially digested coat protein, lacking the basic N-terminal part was used. Native coat protein failed to assemble either T=1 or T=3 mode in the absence of RNA, indicating that an initial RNA-protein interaction is needed and the formation of the T=3 particles requires the interaction between viral RNA and the basic arm of the coat protein. These studies did not demonstrate the requirement for a specific coat protein-RNA interaction in SCPMV assembly, but indicated that the N-terminal R-domain is important for RNA binding.

However, dissociation of SBMV in high salt (0.4 M KCl) at neutral pH yields a ribonucleoprotein complex (RNPC) composed of the viral RNA and about six coat protein subunits (Hsu *et al.*, 1977). Hacker (1995) demonstrated that coat protein subunits present in the RNPC resulting from SCPMV dissociation bind to a specific region of the viral RNA which potentially folds into a hairpin. The specific coat protein binding site is located within the protease coding region in SCPMV ORF2. This site is the most highly conserved region of ORF2 between the sobemoviruses. Unfortunately, these results do not directly demonstrate that coat protein binding to this region serves to nucleate SCPMV assembly.

1.1.7. Cell-to-cell and long distance movement

Most, if not all, plant viruses direct the synthesis of one or more proteins, termed the movement proteins, required for the spread of infection from the initial site to adjacent cells (Hull, 1989; Maule, 1991; McLean *et al.*, 1993; Carrington *et al.*, 1996). For this cell-to-cell movement, the plant viruses seem to circumvent the cell wall by exploiting the intercellular connections termed plasmodesmata. At least two different mechanisms for viral cell-to-cell movement *via* plasmodesmata have been proposed (McLean *et al.*, 1993). The first mechanism of cell-to-cell movement is characteristic for the TMV. TMV-like mechanism involves the movement of the viral genome in a nonvirion form (either the RNPC or free RNA) through plasmodesmata slightly modified by virus encoded movement protein. The second mechanism involves the movement of whole virus particles through tubules in heavily modified plasmodesmata. This cell-to-cell movement mechanism is used by the cowpea mosaic comovirus.

The movement protein functions are not fixed to any sobemoviral gene products. As the functions of P1 have remained relatively indistinct, it is tempting to propose P1 to be the sobemoviral movement protein. However, current data does not allow to state it conclusively.

To achieve the systemic plant infection, the infectious virus particles must enter, circulate within, and leave the vascular tissues. This process has been referred to as long distance movement (Gilbertson & Lucas, 1996; Séron & Haenni, 1996). Mixed infections of the bean with SCPMV, which cannot infect bean systemically, and sunn-hemp mosaic tobamovirus (SHMV) were limited to the short distance movement of SCPMV in the primary inoculated leaves, only SHMV spread systemically (Fuentes & Hamilton, 1991). This SCPMV movement limitation was not due to the inability of the bean cells to support virus replication, as SCPMV replicated efficiency in bean protoplasts. It was proposed that the failure of SCPMV to move systemically in bean was due to the lack of normal SCPMV virion formation (Fuentes & Hamilton, 1993). Examination of thin sections of primary leaves of bean, doubly infected with SHMV and SCPMV, indicated the presence of SCPMV virions having T=1 structure and coat protein clumps in the vacuoles of mesophyll cells. These results show that short-distance (cell-to-cell) and long-distance (vascular) movements of SCPMV are distinct and separate processes in the bean and that the formation of normal T=3 virions is a prerequisite for long distance movement. Supportive evidence for that comes from the work where RYMV-CI virions were found in systemically infected leaves in xylem elements as well as parenhyma cells (Opalka *et al.*, 1998). The predominant location of RYMV-CI virions within xylem implies that the upward flow pattern through xylem may facilitate the systemic spread of infection. Long distance movement through xylem is a relatively rare event for plant viruses. Currently it is not known whether this transport route is characteristic only for RYMV or whether it is the feature common for all sobemoviruses. So far it is not known, if the encapsidation is needed also for short-distance movement of RYMV, although virus-like particles were identified within the plasmodesmata connecting mesophyll cells of RYMV-CI infected leaves (Opalka *et al.*, 1998).

1.2. Regulation of protein synthesis at the level of elongation: –1 ribosomal frameshifting

Ribosomal frameshifting is a strategy frequently employed by various organisms to produce more than one protein from overlapping ORFs. This regulation mechanism leads to the synthesis of two proteins that are identical in their N-terminal region, but differ in their C-terminal part. The advantage of this strategy is that it allows the production of a defined ratio of proteins, i.e., a high ratio of the preframeshift peptide to the frameshift product which is a fusion of two ORFs.

Frameshifting results from movement of ribosomes either in the 5' direction (–1 ribosomal frameshifting) or in the 3' direction (+1 ribosomal frameshifting)

(Brierley, 1995; Farabaugh, 1996). The +1 ribosomal frameshifting occurs in expression of *E. coli* release factor 2, Ty retrotransposons in *S. cerevisiae,* and the mammalian ornithine decarboxylase antizyme, for example. A leftward movement is involved in certain bacterial and bacteriophage genes, bacterial insertion sequences, and *Drosophila* retrotransposons. In addition, ribosomal frameshifting occurs in translation of the genomic RNAs of many retroviruses, coronaviruses, toroviruses, arteriviruses, astroviruses, giardiaviruses, totiviruses, and some plant viruses (Table 5). In most of the systems studied to date, –1 ribosomal frameshifting is involved in the expression of replicases. In retroviruses, it allows the synthesis of the Gag-Pol and Gag-Pro-Pol polyproteins from which reverse transcriptase is derived (Chamorro *et al.*, 1992), and for most other viruses, including plant viruses, frameshifting is required for expression of RdRp-s.

Table 5. Plant viruses that are known or suspected to use ribosomal frameshifting.

Family	Genus	Virus	Reference
+1 ribosomal frameshifting			
Closteroviridae	Closterovirus	BYV; citrus tristeza virus	Agranovsky *et al.*, 1994; Karasev et al., 1995
	Crinivirus	Lettuce infectious yellows virus	Klaassen *et al.*, 1995
–1 ribosomal frameshifting			
Tombusviridae	Dianthovirus	Carnation ringspot virus RNA1; RCNMV	Ryabov *et al.*, 1994; Xiong *et al.*, 1993; Kim & Lommel, 1994
Luteoviridae	Luteovirus	Barley yellow dwarf virus (PAV serotype), soybean dwarf virus	Brault & Miller, 1992; Di *et al.*, 1993; Miller *et al.*, 1995
	Polerovirus	BWYV; PLRV; CYDV-RPV	Prüfer *et al.*, 1992; Garcia *et al.*, 1993; Kujawa *et al.*, 1993; Miller *et al.*, 1995
	Enamovirus	PEMV-1	Demler & de Zoeten, 1991
–	Carlavirus	Potato virus M	Gramstat *et al.*, 1994
–	Sobemovirus	CfMV	Paper II
–	Umbravirus	Groundnut rosette virus; PEMV RNA2	Taliansky *et al.*, 1996; Demler *et al.*, 1993

Two different signals in the mRNA have been identified that facilitate –1 ribosomal frameshifing in eukaryotic systems.

First, the shifty heptanucleotide sequence (also as "slippery sequence") composed of two homopolymeric triplets of the order X XXY YYZ (where X=A, G, or U; Y=A, or U; and Z=A, C, or U; the 0 frame is indicated by spaces) (Jacks *et al.*, 1988; Atkins *et al.*, 1990). The amino acid sequencing of the fusion protein has revealed that this sequence motif is the actual site where the frameshifting occurs (Jacks *et al.*, 1988). The key requirement of the shifty heptanucleotide is that after simultaneous slippage into the –1 frame, partial or full codon-anticodon base-pairing is maintained. Mutations in the slippery sequence reduce or abolish frameshifting (Jacks *et al.*, 1988; Brault & Miller, 1992; Brierly *et al.*, 1992; Kim & Lommel, 1994).

The second signal is a structural element just downstream of the heptanucleotide sequence: the termination codon of the preframeshift ORF (Curran & Yaurus, 1988; Weiss *et al.*, 1988; Brault & Miller, 1992), or a secondary structure (stem-loop structure or RNA pseudoknot structure) immediately downstream (5–8 nt) from the slippery sequence (Le *et al.*, 1991; Prüfer *et al.*, 1992; Xiong *et al.*, 1993; Garcia *et al.*, 1993; Kujawa *et al.*, 1993). It has been suggested that the second element is required for forcing the ribosome complex to pause.

Although the *cis*-acting mRNA signals, which specify ribosomal frameshifting, are reasonably well characterised, the precise mechanism of the process remains unknown. Currently, the most appealing model for –1 ribosomal frameshifting is the pausing model (Jacks *et al.*, 1988). In this model, the RNA structure downstream of the slippery sequence acts as a barrier to translation, pausing ribosomes over the slippery sequence. The pausing increases the likelihood that the tRNAs in the peptidyl and aminoacyl sites on the ribosome simultaneously slip back by one base on the mRNA at the shifty heptanucleotide and the ribosomes continue translating in the –1 frame avoiding the stop codon for the upstream gene.

Ribosomal frameshifting has been studied using two types of assays: *in vitro* transcription and translation of portions of the viral genomic RNA, or by placing the putative frameshift sequence in front of a reporter gene so that frameshifting is required for translation of the gene. This construct is introduced into cells for measurement of reporter gene product activity *in vivo*. The frameshift efficiencies from 1 to over 40%, depending on the specific virus and assay system employed, have been observed (Brierley, 1995).

2. AIMS OF THE PRESENT STUDY

The aims of the present work were:

1. **The determination the complete nucleotide sequence of the cocksfoot mottle sobemovirus (CfMV-NO) genomic RNA.** All sobemoviruses have been characterised to have a small genome and occur at relatively high concentrations in infected plants (Hull, 1988). Therefore, sobemoviruses are ideal subjects for the study of molecular biology of plant RNA virus-host interactions. When this project was initiated, the type member of this genus, SBMV, was extensively studied with respect to its structure and assembly. The complete genomic sequence and three dimensional structure of SCPMV were known (Abad-Zapatero *et al.*, 1980; Hermodson *et al.*, 1982; Wu *et al.*, 1987). Our first aim was to determine the complete sequence of the genomic RNA of CfMV-NO.

2. **The characterisation the genome organisation of CfMV-NO.** The characteristics of genome organisation can be grouped in two sections: the viral genes and the non-coding functions of viral RNA. The complete nucleotide sequence of the virus enabled to identify the genes or ORFs, and deduce amino acid sequences of the viral proteins. In addition to the coding function of the viral genome, viral RNAs contain signals that are necessary for their life cycle. These signals can be important for translation, replication or encapsidation.

3. **The production polyclonal antisera against all proteins encoded by individual ORFs of CfMV-NO.** The availability of antisera specific for viral proteins allows us to study subcellular localisation of the viral proteins in infected plant cells and analyse the role of these proteins in the viral life cycle. Therefore, polyclonal antisera were raised against recombinant CfMV proteins.

4. **The characterisation the translational strategies used by CfMV and other sobemoviruses.** (+)-strand plant RNA viruses have the limited size of genomes which leads to the compactness of the genetic information. They also depend on the host machinery for translation and the regulation of synthesis of the viral products is at the level of RNA transcription and translation. Therefore, plant viruses use a large variety of strategies for expression and regulation of their proteins. The aim of our work was to study the translational strategies used by CfMV-NO and other sobemoviruses.

3. RESULTS AND DISCUSSION

3.1. The complete nucleotide sequence and genome organisation of CfMV-NO (Paper I)

The complete nucleotide sequence of CfMV-NO consists of 4082 nucleotides (Paper I, Fig. 1). It is the shortest sequenced sobemovirus so far. The location of the 5' end of CfMV-NO RNA was confirmed by primer extension experiment (Paper I, Fig. 2). The 3' terminal sequence was obtained from three independent poly(A)-tailed cDNA clones.

For the present, in addition to CfMV-NO the complete nucleotide sequence of genomic RNA of CfMV isolated from Russia (CfMV-RU) has been determined (Ryabov *et al.*, 1996). CfMV-RU shows 95.9% identity to CfMV-NO at the nucleic acid level.

5' and 3' noncoding regions of CfMV-NO RNA. The 5' noncoding region of CfMV-NO genomic RNA is 69 nt long and is identical to those reported for CfMV-RU. This region has a relatively high U and A content (A+U=55%), a common feature among the sobemoviruses and many (+)-strand RNA viruses (Gallie *et al.*, 1987; Wu *et al.*, 1987; Othman & Hull, 1995). A consensus sequence ACAAaa at or near the 5' ends of genomic RNAs of LTSV, RYMV-CI, SBMV, SBMV-Ark, and SCPMV has been found (Hacker & Sivakumaran, 1997). In contrast, this sequence motif is not present at the 5' end of CfMV genome (Paper I; Ryabov *et al.*, 1996). The 5' terminal nucleotides of CfMV-NO and CfMV-RU genomic RNAs were identified to be GAUAAUA and NAUAAUA, respectively. Characteristic to all sobemoviruses is the possession of a polypurine track (aAGgAA) in the beginning of the genomic RNA (Hacker & Sivakumaran, 1997). Similar polypurine track, AAGAAA (nt 13–18) can be found in the 5' noncoding region of CfMV (Paper I; Ryabov *et al.*, 1996).

In RYMV-CI RNA, computer-assisted folding of the 61 most 5' proximal nucleotides, using the STAR 100 computer program, shows a stable stem-loop structure (Ngon A Yassi *et al.*, 1994). A similar type of folding may be predicted for the 30 first 5' residues in CfMV-RU RNA with DNASIS software (Ryabov *et al.*, 1996). These 5' terminal stem-loop structures may function as *cis*-acting signals for (+)-strand synthesis similarly to other plant viruses with (+)-sense RNA genome (Duggal *et al.*, 1994; Rybov *et al.*, 1996).

The 3' noncoding region of CfMV-NO is 225 nt in length and is non-polyadenylated (Paper I). The corresponding region of CfMV-RU is one nucleotide longer and contains 7 nucleotide differences (Ryabov *et al.*, 1996). Secondary structure analysis of the CfMV-NO 3' noncoding region with

RNAFOLD computer program did not reveal the presence of a tRNA-like structure (Paper I). However, using STAR 100 and DNASIS programs, respectively, the last 147 nt of RYMV-CI and 130 nt of CfMV-RU were shown to potentially form stable tRNA resembling secondary structures (Ngon A Yassi *et al.*, 1994; Ryabov *et al.*, 1996). The differences between predictions of the 3' secondary structures of two isolates of CfMV may be due to the different computer programs used for the analysis. It should be mentioned that there are only two nucleotide changes (at position 3918 C or T, at position 4053 C or A) in this 3' region of CfMV-NO and CfMV-RU.

Genome organisation of CfMV-NO. The CfMV-NO genome contains four ORFs. The genome organisation of CfMV-NO is identical to that of CfMV-RU and is presented in Fig. 3.

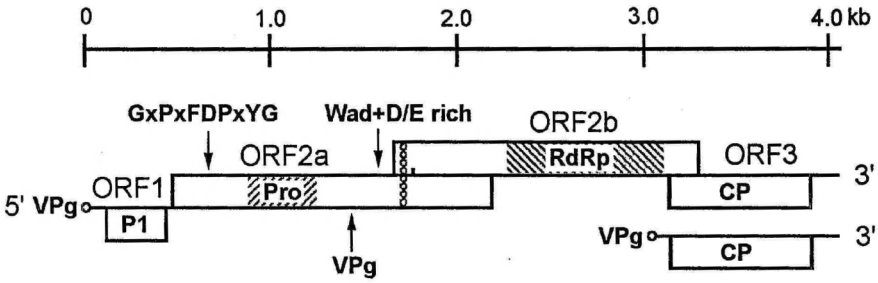


Figure 3. Genome organisation of CfMV-NO. A diagram of genomic and subgenomic RNAs is shown. The ORFs are illustrated as boxes. VPg, the peptide covalently attached to the 5' terminus of the RNA genome, is shown as a small circle. The approximate location of the putative protease and the putative RdRp domains are marked *Pro* and *RdRp*, respectively. The position of the VPg N-terminus is indicated by an arrow. The site of –1 ribosomal frameshift consensus signals is shown by a chain. Different sequence motifs (GxPxFDPxYG and Wad+D/E rich) discussed in the text are indicated. P1, ORF1 encoded protein; CP, coat protein.

Similarly to other sequenced sobemoviruses, CfMV has a small ORF1 at the 5' end of the genome and the 3' terminal ORF which encodes the coat protein. Unlike the others, CfMV lacks the continuous ORF2 and a nested coding region similar to the ORF3 of LTSV, RYMV-CI, SBMV-Ark, and SCPMV (Fig. 1) (Paper I; Ryabov *et al.*, 1996). Instead, CfMV has two overlapping ORFs, ORF2a and ORF2b (Fig. 3). Similar genome organisation is characteristic to the RYMV-NG (Pinto, personal communication). The overlapping arrangement of ORFs 2a and 2b, together with the lacking of an AUG initiation codon near the 5' end of ORF2b, suggests that ORF2b may be expressed by a –1 ribosomal frameshifting mechanism from ORF2a. Indeed, in CfMV, a slip-

pery heptanucleotide sequence followed by a stem-loop structure is present near the beginning of the ORF2a-2b overlapping region (see also section 3.2.4).

The CfMV genome arrangement resembles that reported for MBV (Revill *et al.*, 1994). This similarity extends to both the presence of an ORF1 and 3' proximal coat protein ORF, and two overlapping ORFs for the polyprotein expression. In addition, the arrangement of the CfMV ORFs in the 5' half of the genome is similar to the poleroviruses (BMYV, BWYV, CABYV, CYDV-RPV, PLRV) and the enamoviruses (PEMV-1) (Miller *et al.*, 1995). The 5' gene cluster of these viruses contains a small ORF0, and overlapping ORFs 1 and 2, and the polyprotein is expressed as a translational frameshift fusion of the ORF1 and 2 products.

Proteins encoded by CfMV-NO. The ORF1 encoded proteins (P1) of CfMV-NO and CfMV-RU differ from each other in three amino acids. At the same time, P1 of CfMV is unrelated to the corresponding proteins of other sequenced sobemoviruses. However, CfMV ORF1 sequence (nt 236–312) has 59.2% identity with the beet yellows closterovirus (BYV) 65 kDa protein gene (Paper I). It has been demonstrated that BYV p65, a homologue of HSP70 cell chaperons, is able to complement the cell-to-cell movement of movement protein-deficient potato X potexvirus and barley stripe mosaic hordeivirus (Agranovsky *et al.*, 1998). The function(s) of sobemoviral P1 are currently unknown, but it has been suggested to play a role in virus movement (Othman & Hull, 1995; Sivakumaran *et al.*, 1998), in enhancement of genome amplification (Bonneau *et al.*, 1998) and in posttranscriptional gene silencing (Pinto, personal communication).

Detailed analysis of the amino acid sequences of CfMV ORFs 2a and 2b encoded proteins (P2a and P2b, respectively) reveals the presence of the consensus sequence motifs found in other plant viruses. In particular, P2a contains the putative active site residues for a chymotrypsine-like serine protease while P2b contains the well characterised viral RNA polymerase sequence motifs present in the RdRp-s of all (+)-strand RNA viruses. The CfMV encoded putative serine protease, and those of sobemoviruses (LTSV, RYMV, SBMV, SBMV-Ark, SCPMV), poleroviruses (BMYV, BWYV, CABYV, CYDV-RPV, PLRV), enamoviruses (PEMV-1), barnaviruses (MBV), and astroviruses (HAstV-2), have H, D and S residues in their catalytic sites. The putative serine protease motif is located in the N-terminal third of the polyprotein of all sobemoviruses. It is encoded by ORF2 in case of LTSV, RYMV, SBMV, SBMV-Ark, SCPMV, and by ORF2a in case of CfMV. The putative RdRp domain of CfMV shows again extensive similarity to the putative RdRp-s of sobemoviruses, poleroviruses, enamoviruses, barnaviruses, and astroviruses. According to the classification of Koonin and Dolja (1993), which is based on a tentative phylogeny of (+)-strand RNA virus RdRp-s, all these viruses belong to polymerase supergroup I.

Recently, the VPg of CfMV-RU has been characterised (Mäkinen *et al.*, unpublished results). It has been shown that VPg of CfMV-RU has molecular mass of 12 kDa. The N-terminal sequence of the VPg was determined. The obtained amino acid sequence started at position 320 of the ORF2a product of CfMV-RU and was 100% identical to the corresponding amino acids of CfMV-NO polyprotein. Similarly to the VPg of SBMV, the VPg of CfMV is located downstream of the putative serine protease motif, indicating that the sobemoviruses have the polyprotein arrangement: protease-VPg-RdRp (Van der Wilk *et al.*, 1998; Mäkinen *et al.*, unpublished results). This genome arrangement is characteristic also for several viruses related to sobemoviruses, e.g. polerovirus (PLRV), enamoviruses (PEMV-1), and barnaviruses (MBV) (Van der Wilk *et al.*, 1997; Revill *et al.*, 1998; Wobus *et al.*, 1998).

In all sobemoviruses a conserved WAD/WGD amino acid sequence followed by a D/E-rich region is present downstream of the putative serine protease motif (Paper I, Fig. 5). A similar motif can also be found in PLRV, BWYV, PEMV-1, MBV, and HAstV-2 in front of the –1 frameshift signals. The determined sizes of CfMV, MBV, PLRV, and SBMV VPg-s indicate that the WAD/WGD plus E/D rich region is present in VPg-s of these viruses. This motif is the only conserved sequence element between VPg-s of these viruses. It is possible that this conserved motif is characteristic for the VPg-s of viruses with a sobemovirus/polerovirus-like genome arrangement.

Previously, it has been suggested that VPg-s of the sobemoviruses are encoded by the N-terminal part of the polyprotein by analogy to the location of the VPg-s of picornaviruses (Gorbalenya *et al.*, 1988; Ngon A Yassi *et al.*, 1994; Paper I). Indeed, the N-terminal part of the sobemoviral polyproteins (aa 24–120) contains the conserved 14 and 9 amino acid blocks and the sequence motif GxPxFDPxYG (Paper I, Fig. 4). So far, the significance of these motifs is unknown.

The functions of the N- and C-terminal parts of P2a of CfMV are unknown. It has been suggested that the 60 N-terminal amino acids of P2a may form a transmembrane domain due to the presence of hydrophobic residues (Ryabov *et al.*, 1996). The C-terminal part of P2a contains a strong basic region (aa 539–552) and may determine the RNA binding of P2a (Tamm & Truve, unpublished results).

Direct sequencing of the N-terminus of CfMV coat protein revealed that the coat protein is encoded by 3' terminal ORF and it commences at the first AUG codon of this ORF at nt 3093 (Paper I). CfMV has, in common with other sequenced sobemoviruses, a highly basic amino acid sequence at the R-domain (aa 1–36) of coat protein. This region in the coat protein of sobemoviruses resembles the bipartite signal proposed for nuclear targeting (Dingwall & Laskey, 1991). However, the significance of this signal is presently not known. The R-domain of CfMV coat protein may also be important in protein-RNA interactions needed for the stabilisation of virion structure.

3.2. Translational strategies of CfMV-NO (Papers I, II, III)

In vitro translation of CfMV-NO virion-extracted RNA resulted in the synthesis of four major proteins of 100, 71, 34, and 12 kDa in the WGE translation system (Paper III, Fig. 2). Similar *in vitro* translation product pattern has been reported for several sobemoviruses (Table 3), suggesting that sobemoviruses may use similar translational strategies for gene expression.

The individual genes of CfMV-NO from which the *in vitro* translation products are synthesised were identified, using the immunoprecipitation with specific antibodies (Paper III). The 12, 71 and 100 kDa CfMV proteins are synthesised from the genomic RNA of the virus. The CfMV 12 kDa protein is produced from ORF1, the 71 kDa protein from ORF2a and the 100 kDa protein is a polyprotein encoded by ORFs 2a and 2b by a –1 ribosomal frameshifting (see also sections 3.2.4). CfMV 34 kDa *in vitro* translation product is a coat protein synthesised from the virion-packed sgRNA (see also section 3.2.1). Based on these findings it was assumed that the 70 kDa *in vitro* translation product of SCPMV may represent the ORF2-ORF3 transframe fusion protein and is not the product of the proteolytic cleavage as was hypothesised earlier (see also section 3.2.4). The similarities of CfMV and polerovirus genome arrangement indicate related expression strategies for sobemo- and poleroviruses. It has been demonstrated experimentally that the poleroviruses control the gene expression using several mechanisms, including internal initiation/leaky scanning, posttranslational processing, subgenomic RNA, –1 ribosomal frameshifting, and stop codon readthrough (Miller *et al.*, 1995).

3.2.1. Subgenomic RNA (Papers I, III)

The location of the coat protein encoding ORF (ORF3) towards the 3' end of CfMV-NO RNA suggests it may be expressed from sgRNA. To confirm the presence of such sgRNA in virus infected barley plants, total RNA extracts were analysed by Northern blot analysis (Paper I, Fig. 6). Probe, covering the coat protein C-terminal coding region, revealed two RNA species of 4.1 kb and 1.2 kb, corresponding in sizes to the genomic RNA and to the predicted sgRNA, respectively. The translational initiation site for CfMV-NO coat protein is 989 nt upstream from the 3' end of the genome. Therefore, the sgRNA of 1.2 kb may be the template for coat protein translation.

The 34 kDa protein which is translated *in vitro* from CfMV-NO virion-extracted RNA comigrated with ORF3 encoded protein and immunoprecipitated with antiserum, specifically recognising the CfMV-NO coat protein (Paper III, Fig. 2D). When the translation products of full-length *in vitro* transcribed CfMV-NO RNA were immunoprecipitated with antiserum against coat protein, no proteins were detected. These results indicate that CfMV-NO

34 kDa protein is the viral coat protein and is translated from sgRNA encapsidated in virus particles.

The region upstream of coat protein encoding ORF is expected to contain a promoter sequence for synthesis of the sgRNA. The 5' ends of genomic and subgenomic RNAs of SCPMV start with similar primary sequence (ACAAAA) (Hacker & Sivakumaran, 1997). Based on this similarity it was possible to predict the putative transcriptional start sites for not yet characterised sobemoviral sgRNAs (see also section 1.1.4.1). Unfortunately, it is not possible to predict the 5' end of the sgRNA of CfMV, as this sequence motif is not present upstream of the coat protein coding region. The transcriptional start site of CfMV-RU sgRNA has been determined (Ryabov *et al.*, 1996). Unfortunately, later experiments have indicated that the site described could represent an experimental artefact (Mäkinen, personal communication).

3.2.2. Initiation of translation from bicistronic RNA (Paper III)

The 5' terminal half of CfMV genome contains two ORFs in different reading frames (Fig. 3). The immunoprecipitation studies showed that both 5' terminal ORFs (ORF1 and ORF2a, respectively) of CfMV-NO are translated from the full-length genomic RNA. The immunoprecipitation of *in vitro* translated CfMV-NO proteins with P1 antiserum identified a 12 kDa protein (Paper III, Fig. 2A). The immunoprecipitation of CfMV-NO RNA translation mixture using P2a antiserum detected two proteins of 71 kDa and 100 kDa (Paper III, Fig. 2B). These translational products correspond to ORF2a encoded protein and to polyprotein, respectively. According to these results, CfMV-NO, with its compact genome organisation, utilises a bifunctional mRNA to achieve the expression of both 5' proximal and internally located ORFs.

In poleroviruses, ORF0 at the 5' end of the genomic RNA overlaps with the next ORF1. The start codons of theses two ORFs are the first two AUGs in the genome. In all cases, the start codon for the fist ORF has suboptimal context for translational initiation (Table 4). It has been suggested that ORF 1 of poleroviruses is probably translated by leaky scanning mechanism (Mayo *et al.*, 1989; Veidt *et al.*, 1992). Other examples of plant viruses with overlapping genes at the 5' end of genomic RNA are the enamoviruses (PEMV-1), the tymoviruses (turnip yellow mosaic virus) and the machlomoviruses (maize chlorotic mottle virus) (Weiland & Dreher, 1989; Demler & de Zoeten, 1991; Lommel *et al.*, 1991).

Two lines of evidence indicate that the leaky ribosomal scanning is likely strategy for the initiation of translation of CfMV-NO ORF2a (see also section 1.1.4.2). First, the first $A_{70}UG$ of CfMV is probably recognised inefficiently as an initiation codon due to the unsuitable initiation context (Table 4). Second, for SCPMV, it has been shown that the translational initiation of ORF2 is

occurring by leaky scanning rather than by internal ribosome entry or coupled termination-reinitiation (Sivakumaran & Hacker, 1998). However, the experiments carried out so far do not exclude conclusively the other possible mechanisms for the translational initiation of ORF2 of sobemoviruses.

3.2.3. Polyprotein processing (Paper I)

CfMV-NO encodes a polyprotein with molecular mass of 100 kDa (Paper II, Fig. 2B). The expression of a large polyprotein is characteristic for all the sobemoviruses studied until now (Table 3). Based on the sequence analyses and experimental data it is possible to distinguish at least three domains in sobemoviral polyprotein — putative protease, VPg, and putative RdRp. It is believed that sobemoviral polyprotein is proteolytically processed to give active viral proteins. By the analogy to picornavirus protease cleavage sites, Gorbalenya *et al.* (1988) predicted that sobemoviral polyproteins are cleaved at E/S,T sites. Several putative cleavage sites in CfMV-NO P2a, RYMV, SBMV, and SCPMV polyprotein were suggested (Paper I). The determination of N-terminus of two sobemovirus VPg-s gives the first indication about the cleavage sites used by the sobemoviral serine protease (see also section 1.1.4.3). The finding that the N-terminal cleavage site of CfMV-RU VPg is E/N, does not support the proposition of Gorbalenya *et al.* (1988). In summary, it is not possible at the moment to propose the polyprotein processing sites for sobemoviruses. Unfortunately, the cell-free *in vitro* translation system cannot be utilised for this study, as no proteolytic activity has been described in translation products of CfMV-NO virion extracted RNA (Paper III).

3.2.4. –1 ribosomal frameshifting (Papers II, III)

As noted above, the C-terminal half of the CfMV-NO polyprotein is encoded by separate ORF2b (see also section 3.1) and is expressed as an ORF2a-2b translational fusion protein by –1 ribosomal frameshifting (Paper II). The consensus signals for a –1 ribosomal frameshift event can be found in the beginning of the overlapping region between CfMV-NO ORF2a and ORF2b. These signals are the heptanucleotide (slippery) sequence 5' UUUAAAC (nt 1634–1640) and a stem-loop structure located 7 nucleotides downstream from the heptamer (Paper II, Fig. 1B). The presence of the stem-loop structure has been confirmed by chemical "footprinting" (Fig. 4) (Tamm, unpublished data).

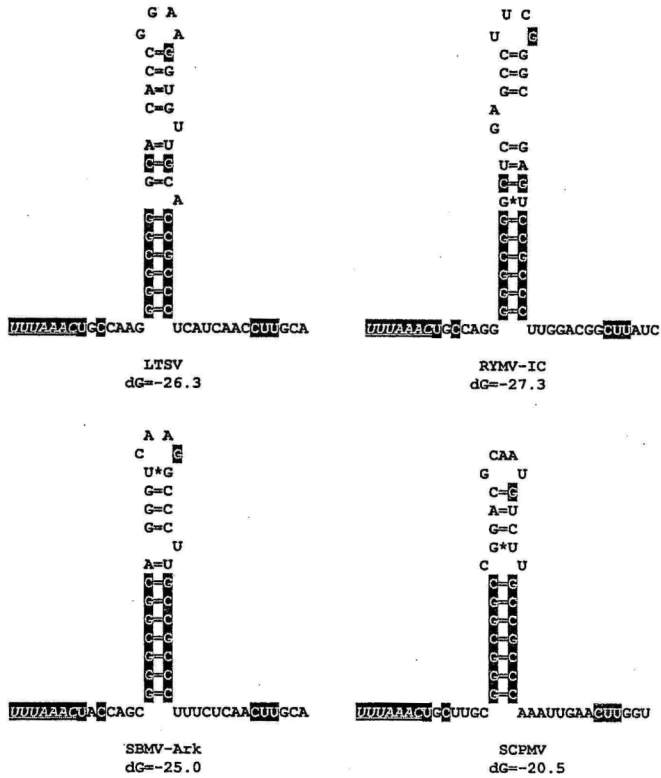
The –1 ribosomal frameshifting directed by CfMV-NO signal sequences was examined *in vitro* by inserting the cDNA fragment representing the nt 1621–2521 of CfMV-NO RNA in the middle of the β-glucuronidase (GUS) sequence

(Paper II, Fig. 2A). This CfMV-NO region contains the overlapping region of ORF2a and ORF2b, including the consensus sequences for frameshifting. When translated in WGE, this sequence directed frameshifting with the efficiency of 26–29%. This experiment showed also that no CfMV-NO specific sequences upstream of the consensus signals are required for an efficient –1 ribosomal frameshift event.

The immunoprecipitation studies showed that ORF2b is not capable of initiating its translation from genomic RNA. The 100 kDa *in vitro* translation product of CfMV-NO was immunoprecipitated with both, P2a and P2b antisera (Paper III, Fig. 2B and C). The 100 kDa polyprotein was the only protein observed after immunoprecipitation with P2b antiserum (Paper III, Fig. 2C), indicating that ORF2b is translated only as a part of the polyprotein by –1 ribosomal frameshifting. When the construct containing the entire ORF2a-ORF2b region was transcribed-translated in WGE, the polyprotein of CfMV-NO was produced with an efficiency of 10.6±1.4% (Paper III). The obtained frameshifting efficiency was lower than described in the reporter gene context.

The consensus signals for –1 ribosomal frameshifting, similar to those characterised for CfMV-NO can be found in the genomes of all sobemoviruses (Fig. 4). The slippery sequence, UUUAAAC followed by the putative stem-loop structure are located upstream of the translational initiation codon of ORF3 of LTSV, RYMV-CI, SBMV-Ark, and SCPMV. There are no stop codons present between the slippery sequence and the initiation codon of ORF3 in both reading frames. It has been proposed that ORF3 is likely translated by the –1 ribosomal frameshifting mechanism similar to that shown for ORF2b in CfMV-NO, but this has not been verified experimentally (Papers II, III). Five facts support this hypothesis: (i) no *in vitro* translation product has been attributed to this ORF; (ii) no sgRNA corresponding to this region has been found; (iii) SCPMV mutant, which had an in-frame stop codon between the predicted frameshift site and potential initiation codon in the ORF3 reading frame, was not able to infect cowpea, suggesting that ORF3 protein was not expressed (Sivakumaran *et al.*, 1998); (iv) the 70 kDa *in vitro* translation product of SCPMV may represent ORF2-ORF3 transframe fusion protein, as the calculated molecular mass for ORF2-ORF3 fusion of SCPMV is 65.8 kDa (Paper III); (v) similarity of P3 to the N-terminal part of the ORF2b encoded protein of CfMV as well as to the N-terminal part of PLRV and BWDV ORF2 encoded proteins, and MBV and PEMV1 ORF3 encoded products starts upstream from the putative ORF3 translation initiation signal (Paper II; Ryabov *et al.*, 1996).

A



B

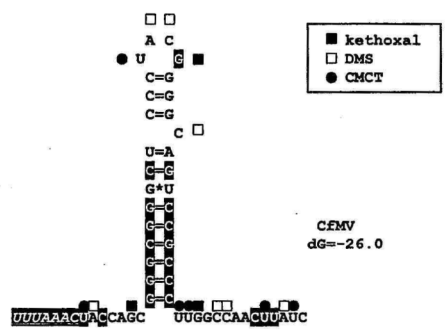


Figure 4. The –1 ribosomal frameshifting consensus signals in sobemoviruses. The shifty heptanucleotide sequences are underlined and italicised. Computer predicted (A) and proven (B) secondary structures downstream from the slippery sequences are presented. Bases conserved among all sobemoviruses are shadowed. The reactivity of nucleotides to the chemical probes dimethyl sulphate (DMS), β-ethoxy-α-ketobutyraldehyde (kethoxal) and 1-cyclohexyl-3-(2-morpholinoethyl) carbodiimide metho-*p*-toluene sulfonate (CMCT) are indicated.

4. CONCLUSIONS AND FURTHER PERSPECTIVES

The genus *Sobemovirus* contains viruses that have similar biological properties (transmission, subcellular localisation, symptomatology), and some similar molecular properties (particle structure, putative protease motif, putative RdRp motif, coat protein sequence). However, the genome organisation of these viruses and therefore the polyprotein expression differs inside the genus.

During the present study the complete nucleotide sequence of CfMV-NO genomic RNA was determined and the genome organisation was characterised. The RNA genome of CfMV-NO, being 4082 nucleotides long, contains four ORFs. In contrast to other sequenced sobemoviruses, CfMV lacks the continuous large ORF that codes for the polyprotein. Instead it has two overlapping ORFs, 2a and 2b. The –1 ribosomal frameshift consensus signals are present in CfMV genome just in the beginning of the overlapping region, suggesting that polyprotein of CfMV is expressed *via* –1 ribosomal frameshifting mechanism. The close relationship between CfMV and the other sobemoviruses is evident from the strong amino acid sequence similarities of putative protease and putative RdRp domains as well as some similarities of their coat proteins.

Both genomic organisation as well as primary structures of several sobemoviral nonstructural proteins (except P1) indicate that these viruses are related to *Polero-* and *Enamovirus* genera in the family *Luteoviridae*. Initially, polero-, enamo- and sobemoviruses were classified into the plant "picorna-like" supergroup (Dolja & Carrington, 1992; Koonin & Dolja, 1993). Viruses belonging to this supergroup have virion RNA with 5' linked VPg and their genomes encode a polyprotein. Their polyprotein arrangement is similar to that of picornaviruses and includes helicase, VPg, chymotrypsine-like protease, and RdRp. Later, "sobemo-like" supergroup of plant (+)-strand RNA viruses have been recognised (Gorbalenya & Koonin, 1993). This supergroup includes enamo-, polero-, and sobemoviruses. The genome of these viruses code for VPg, similarly to "picorna-like" viruses. However, they do not code for the helicase and the polyprotein domains are in order protease-VPg-RdRp, not VPg-protease-RdRp as in the "picorna-like" viruses (Van der Wilk *et al.,* 1997; Revill *et al.,* 1998; Van der Wilk *et al.,* 1998; Wobus *et al.,* 1998; Mäkinen *et al.,* unpublished results).

The translational strategies used by CfMV-NO were under investigation. It has been shown that plant (+)-strand RNA viruses use a large variety of strategies for protein expression. CfMV and other sobemoviruses are not the exceptions. We have demonstrated that the genomic RNA of sobemoviruses functions as bicistronic mRNA, the polyprotein of CfMV is translated by –1 ribosomal frameshifting, and the coat protein of CfMV is translated from sgRNA.

The first indications how the sobemoviral polyprotein is processed are obtained (Van der Wilk *et al.*, 1998; Mäkinen *et al.*, unpublished results). It is important to mention that –1 ribosomal frameshifting consensus signals are present in all sobemovirus genomes and possibly –1 ribosomal frameshifting is used for the expression of ORF3 of LTSV, RYMV, SBMV-Ark, and SCPMV.

Intriguing is the similarity between animal astroviruses, fungal barnaviruses and plant sobemoviruses. This similarity extends to both the genome arrangement and gene function and expression. The genome organisation of MBV resembles that of CfMV. The similarities of expression strategies include (i) translation initiation for ORF2 by the leaky scanning mechanism, (ii) posttranslational processing of the polyprotein, (iii) coat protein translation from sgRNA, (iv) expression of the putative RdRp as a fusion protein with ORF2 encoded protein by a –1 ribosomal frameshifting. Although astroviruses do not have a small ORF in the beginning of the genome, their putative protease and putative RdRp share similarity with corresponding proteins of sobemoviruses (Jiang *et al.*, 1993). Moreover, RdRp of astroviruses is expressed as a fusion with protease *via* –1 ribosomal frameshifting mechanism (Marczinke *et al.*, 1994; Lewis & Matsui, 1996). In summary, genome organisation, primary structure of some nonstructural proteins and translational strategies characteristic to sobemoviruses have been conserved among viruses infecting three large kingdoms of eukaryotes.

During the recent years a great deal of information has been collected on sobemoviruses. Much has been understood about the gene function and expression by (i) comparison with genes of known function, (ii) using the reporter gene assays *in vitro* and *in vivo*, (iii) construction of full-length infectious cDNA constructs and the analysis of mutants, (iv) the N-terminal sequencing of viral VPg-s, (v) mapping of the transcriptional start sites of sgRNAs. However, a better understanding of mechanisms of transmission, replication, and movement would provide new insights into important aspects of the sobemovirus life cycle.

RNA genomes of certain plant viruses present the same characteristics as cellular mRNAs, i.e., a cap structure and a poly(A) tail. It is usually considered that in cellular and viral RNAs, these elements have the same functions during translation. Several plant viruses, including the sobemoviruses, contain a VPg at the 5' end and lack the poly(A) tail. At the same time, these viruses compete effectively with host mRNAs for translational machinery. Therefore, many questions arise, e.g. (i) what is the mechanism of cap-independent translational initiation; (ii) how 5' and 3' ends interact during the translation; (iii) how these structures affect the stability of messenger. At the moment these questions are unanswered.

Genome sizes of plant viruses are mostly within the range of 1 to 10 kb, regardless of the type of virion nucleic acid or expression strategy. The genetic elements that direct the expression of virus genomes are therefore highly efficient, often small, and multifunctional. It is known that sobemoviruses appear at relatively high concentrations in infected plants (Hull, 1988). Unfortunately, studies about the *cis*-acting signals of sobemoviruses, which are required for efficient RNA translation, replication, and encapsidation are almost lacking. The information of structure and function of these genetic elements will make sobemoviruses useful models for the study of minimal forms of expression of genetic information in plant cells, and possible candidates for plant virus vectors.

REFERENCES

Abad-Zapatero, C., S. S. Abdel-Meguid, J. E. Johnson, A. G. W. Leslie, I. Rayment, M. G. Rossmann, D. Suck, and T. Tsukihara. (1980). Structure of southern bean mosaic virus at 2.8 Å resolution. *Nature* **286:** 33–39.

Agranovsky, A. A., E. V. Koonin, V. P. Boyko, E. Maiss, R. Frötschl, N. A. Lunina, and J. G. Atabekov. (1994). Beet yellows closterovirus: complete genome structure and identification of a leader papain-like thiol protease. *Virology* **198:** 311–324.

Agranovsky, A. A., A. S. Folimonov, S. Y. Folimonova, S. Y. Morozov, J. Schiemann, D. Lesemann, and J. G. Atabekov. (1998). Beet yellows closterovirus HSP70-like protein mediates the sell-to-sell movement of a potexvirus transport-deficient mutant and a hordeivirus-based chimeric virus. *J. Gen. Virol.* **79:** 889–895.

Atkins, J. F., R. B. Weiss, and R. F. Gesteland. (1990). Ribosome gymnastics-degree of difficulty 9.5, style 10.0. *Cell* **62:** 413–423.

Bakker, W. (1974). Characterisation and ecological aspects of rice yellow mottle virus in Kenya. *Agric. Res. Rep.* No. 829, Wageningen. 152 pp.

Belsham, G. J., and N. Sonenberg. (1996). RNA-protein interactions in regulation of picornavirus RNA translation. *Microbiol. Rev.* **60:** 499–511.

Benigno, D. A., and J. A'Brook. (1972). Studies on the host range and properties of cocksfoot mottle and phleum mottle viruses. *Plant Pathol.* **21:** 142–144.

Bhuvaneshwari, M., H. S. Subramanya, K. Gopinath, H. S. Savithri, M. V. Nayudu, and M. R. Murthy. (1995). Structure of *Sesbania* mosaic virus at 3 Å resolution. *Structure* **3:** 1021–1030.

Blackstock, J. M. (1978). Lucerne transient streak and lucerne latent, two new viruses of lucerne. *Aust. J. Agric. Res.* **29:** 291–304.

Bonneau, C., C. Brugidou, L. Chen, R. N. Beachy, and C. Fauquet. (1998). Expression of the rice yellow mottle virus P1 protein *in vitro* and *in vivo* and its involvement in virus spread. *Virology* **244:** 79–86.

Brault, V., and W. A. Miller. (1992). Translational frameshifting mediated by a viral sequence in plant cells. *Proc. Natl. Acad. Sci. USA* **89:** 2262–2266.

Brierley, I. (1995). Ribosomal frameshifting on viral RNAs. *J. Gen. Virol.* **76:** 1885–1892.

Brierley, I., A. J. Jenner, and S. C. Inglis. (1992). Mutational analysis of the "slippery sequence" component of a coronavirus ribosomal frameshifting signal. *J. Mol. Biol.* **227:** 463–479.

Brugidou, C., C. Holt, M. N. Ngon A Yassi, S. Zhang, R. Beachy, and C. Fauquet. (1995). Synthesis of an infectious full-length cDNA clone of rice yellow mottle virus and mutagenesis of the coat protein. *Virology* **206:**108–115.

Brunt, A. A., K. Crabtree, M. J. Dallwitz, A. J. Gibbs, L. Watson, and E. J. Zurcher (eds.) Plant Viruses Online: Description and list from the VIDE Database. Version 20th August 1996. [Online] http://biology.anu.edu.au/ Groups/ MES/vide/ [30 November 1999, last date accessed.].

Carrington, J. C., K. D. Kasschau, S. K. Mahajan, and M. C. Schaad. (1996). Cell-to-cell and long-distance transport of viruses in plants. *Plant Cell* **8:** 1669–1681.

Catherall, P. L. (1985). Resistances of grasses to two sobemoviruses, cocksfoot mottle and cynosurus mottle. *Grass Forage Sci.* **40:** 311–316.

Catherall, P. L. (1986). Resistance to cocksfoot mottle virus in *Dactylis glomerata*. *Ann. Appl. Biol./Suppl.* **108:** 148–149.

Cavener, D. R., and S. C. Ray. (1991). Eukaryotic start and stop translation sites. *Nucleic Acids Res.* **19:** 3185–3192.

Chamberlain, J. A., and P. L. Catherall. (1976). Electron microscopy of some grasses and cereals infected with cocksfoot mottle, phleum mottle and cocksfoot mild mosaic viruses. *J. Gen. Virol.* **30:** 41–50.

Chamorro, M. M., N. Parkin, and H. E. Varmus. (1992). An RNA pseudoknot and an optimal heptameric shift site are required for highly efficient ribosomal frameshifting on a retroviral messenger RNA. *Proc. Natl. Acad. Sci. USA* **89:** 713–717.

Curran, J. F., and M. Yarus. (1988). Use of tRNA supressors to probe regulation of *Escherichia coli* release factor 2. *J. Mol. Biol.* **203:** 75–83.

Dawson, W. O., P. Bubrick, and G. L. Crantham. (1988). Modifications to tobacco mosaic virus coat protein gene affecting replication, movement and symptomology. *Phytopathology* **78:** 783–789.

Demler, S. A., and G. A. de Zoeten. (1991). The nucleotide sequence and luteovirus-like nature of RNA 1 of an aphid non-transmissible strain of pea enation mosaic virus. *J. Gen. Virol.* **72:** 1819–1834.

Demler, S. A., D. G. Rucker, and G. A. de Zoeten. (1993). The chimeric nature of the genome of pea enation mosaic virus: the independent replication of RNA 2. *J. Gen. Virol.* **74:** 1–14.

Di, R., S. P. Dinesh-Kumar, and W. A. Miller. (1993). Translational frameshifting by barley yellow dwarf virus RNA (PAV serotype) in *Escherichia coli* and in eukaryotic cell-free extracts. *Mol. Plant-Microbe Interact.* **6:** 444–452.

Dingwall, C., and R. A. Laskey. (1991). Nuclear targeting sequences — a consensus? *Trends Biochem. Sci.* **16:** 478–481.

Dolja, V. V., and E. V. Koonin. (1991). Phylogeny of capsid proteins of small icosahedral RNA plant viruses. *J. Gen. Virol.* **72:** 1481–1486.

Dolja, V. V., and J. C. Carrington. (1992). Evolution of positive-strand RNA viruses. *Semin. Virol.* **3:** 315–326.

Dougherty, W. G., and B. L. Semler. (1993). Expression of virus-encoded proteinases: functional and structural similarities with cellular enzymes. *Microbiol. Rev.* **57:** 781–822.

Duggal, R., F. C. Lahser, and T. C. Hall. (1994). *Cis*-acting sequences in the replication of plant viruses with plus-sense RNA genomes. *Annu. Rev. Phytopathol.* **32:** 287–309.

Engsbro, B. (1978). Investigations on virus diseases of grasses in Denmark. *Nordisk Jordbruksforskning* **61:** 51–52.

Erickson, J. W., and M. G. Rossmann. (1982). Assembly and crystallization of a $T = 1$ icosahedral particle from trypsinized southern bean mosaic virus coat protein. *Virology* **116:** 128–136.

Farabaugh, P. J. (1996). Programmed translational frameshifting. *Microbiol. Rev.* **60:** 103–134.

Fauquet, M. C., and M. A. Mayo. (1999). Abbreviations for plant virus names — 1999. *Arch. Virol.* **144:** 1249–1273.

Forster, R. L. S., and A. T. Jones. (1979). Properties of lucerne transient streak virus, and evidence of its affinity to southern bean mosaic virus. *Ann. Appl. Biol.* **93:** 181–189.

Francki, R. I. B., J. W. Randles, T. Hatta, C. Davies, P. W. G. Chu, and G. D. McLean. (1983). Subterranean clover mottle virus: another virus from Australia with encapsidated viroid-like RNA. *Plant Path.* **32:** 47–59.

Francki, R. I. B., R. G. Milne, and T. Hatta. (1985). Sobemovirus group, p. 153–169. *In* R. I. B. Francki, R. G. Milne and T. Hatta (ed.), *Atlas of Plant Viruses*, vol. 1. CRC Press, Boca Raton.

Fuentes, A. L., and R. I. Hamilton. (1991). Sunn-hemp mosaic virus facilitates cell-to-cell spread of southern bean mosaic virus in a nonpermissive host. *Phytopathology* **81:** 1302–1305.

Fuentes, A. L., and R. I. Hamilton. (1993). Failure of long-distance movement of southern bean mosaic virus in a resistant host is correlated with lack of normal virion formation. *J. Gen. Virol.* **74:** 1903–1910.

Gallie, D. R., D. E. Sleat, J. W. Watts, P. C. Turner, and T. M. A. Wilson. (1987). A comparison of eukaryotic viral 5'-leader sequences as enhancers of mRNA expression *in vivo*. *Nucleic Acid Res.* **15:** 8693–8711.

Garcia, A., J. van Duin, and C. W. A. Pleij. (1993). Differential response to frameshift signals in eukaryotic and prokaryotic translational systems. *Nucleic Acid Res.* **21:** 401–406.

Ghosh, A., T. Rutgers, M. Ke-Qiang, and P. Kaesberg. (1981). Characterization of the coat protein mRNA of southern bean mosaic virus and its relationship to the genomic RNA. *J. Virol.* **39:** 87-92.

Gilbertson, R. L., and W. J. Lucas. (1996). How do viruses traffic on the "vascular highway"? *Trends Plant Sci.* **1:** 260-268.

Gorbalenya, A. E., E. V. Koonin, V. M. Blinov, and A. P. Donchenko. (1988). Sobemovirus genome appears to encode a serine protease related to cysteine proteases of picornaviruses. *FEBS Lett.* **236:** 287–290.

Gorbalenya, A. E., and E. V. Koonin. (1993). Comparative analysis of amino-acid sequences of key enzymes of replication and expression of positive-strand RNA viruses: validity of approach and functional and evolutionary implications. *Sov. Sci. Rev. D. Physicochem. Biol.* **11:** 1–84.

Gramstat, A., D. Prüfer, and W. Rhode. (1994). The nucleic acid-binding zinc finger protein of potato virus M is translated by internal initiation as well as by ribosomal frameshifting involving a shifty stop codon and a novel mechanism of P-site slippage. *Nucleic Acid Res.* **22:** 3911–3917.

Greber, R. S. (1981). Some characteristics of *Solanum nodiflorum* virus — a beetle-transmitted isometric virus from Australia. *Aust. J. Biol. Sci.* **34:** 369–378.

Guilley, H., C. Wipf-Scheibel, K. Richards, H. Lecoq, and G. Jonard. (1994). Nucleotide sequence of cucurbit aphid-borne yellows luteovirus. *Virology* **202:** 1012–1017.

Guilley, H., K. E. Richards, and G. Jonard. (1995). Nucleotide sequence of beet mild yellowing virus RNA. *Arch. Virol.* **140:** 1109–1118.

Hacker, D. L. (1995). Identification of a coat protein binding site on southern bean mosaic virus RNA. *Virology* **207:** 562–565.

Hacker, D. L., and K. Sivakumaran. (1997). Mapping and expression of southern bean mosaic virus genomic and subgenomic RNAs. *Virology* **234:** 317–327.

Hariri, D. and H. Lapierre. (1978). Le virus de nécrose et mosaique du dactyle (*Dactylis glomerata* L.) 281, Centre National de Recherches Agronomique INRA, Versailles, France.

Hartmann, J. X., J. E. Bath, and G. R. Hooper. (1973). Electron microscopy of virus like particles from shoestring-diseased high bush blueberry, *Vaccinium corymbosum* L. *Phytopathology* **63:** 432–436.

Hermodson, M. A., C. Abad-Zapatero, S. S. Abdel-Meguid, S. Pundak, M. G. Rossmann, and J. H. Tremaine. (1982). Amino acid sequence of southern bean mosaic virus coat protein and its relation to the three-dimensional structure of the virus. *Virology* **119:** 133–149.

Hobbs, H. A., C. W. Kuhn, K. E. Papa, and B. B. Brantley. (1987). Inheritance of nonnecrotic resistance to southern bean mosaic virus in cowpea. *Phytopathology* **77:** 1624–1629.

Hollings, M., and O. M. Stone. (1973). Turnip rosette virus. *CMI/AAB Description of Plant Viruses* No. 125.

Hsu, C. H., J. A. White, and O. P. Sehgal. (1977). Assembly of southern bean mosaic virus from its two subviral intermediates. *Virology* **81:** 471–475.

Hull, R. (1977a). The grouping of small spherical plant viruses with single RNA components. *J. Gen. Virol.* **36:** 289–295.

Hull, R. (1977b). The stabilization of the particles of turnip rosette virus and of other members of the southern bean mosaic virus group. *Virology* **79:** 58–66.

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

Hull, R. (1989). The movement of viruses in plants. *Ann. Rev. Phytopath.* **27:** 213–240.

Hull, R. (1995). Sobemovirus, p. 376–378. *In* F. A. Murphy, C. M. Fauquet, D. H. L. Bishop, S. A. Ghabrial, A. W. Jarvis, G. P. Martelli, M. A. Mayo, and M. D. Summers (ed.). Virus Taxonomy. Classification and Nomenclature of Viruses. Sixth Report on the International Committee on Taxonomy of Viruses. Springer Verlag, Wien New York.

Huth, W., and H. L. Paul. (1972). Cocksfoot mosaic virus. *CMI/AAB Description of Plant Viruses* No. 107.

Jacks, T., H. D. Madhani, F. R. Masiarz, and H. E. Varmus. (1988). Signals for ribosomal frameshifting in the Rous sarcoma virus *gag-pol* region. *Cell* **55:** 447–458.

Jeffries, A. C., J. P. Rathjen, and R. H. Symons. (1995). Lucerne transient streak virus complete genome. GenBank accession no. U31286.

Jiang, B., S. S. Monroe, E. V. Koonin, S. E. Stine, and R. I. Glass. (1993). RNA sequence of astrovirus: distinctive genomic organization and a putative retrovirus-like ribosomal frameshifting signal that directs the viral replicase synthesis. *Proc. Natl. Acad. Sci. USA* **90:** 10539–10543.

Jones, A. T., and M. A. Mayo. (1984). Satellite nature of the viroid-like RNA-2 of *Solanum nodiflorum* mottle virus and the ability of other plant viruses to support the replication of viroid-like RNA molecules. *J. Gen. Virol.* **65:** 1713–1721.

Kado, C. I. (1967). Biological and biochemical characterisation of sowbane mosaic virus. *Virology* **31:** 217–229.

Kado, C. I. (1971). Sowbane mosaic virus. *CMI/AAB Description of Plant Viruses* No. 64.

Karasev A. V., V. P. Boyko, S. Gowda, O. V. Nikolaeva, M. E. Hilf, E. V. Koonin, C. L. Niblett, K. Cline, D. J. Gumpf, and R. F. Lee. (1995). Complete sequence of the citrus tristeza virus RNA genome. *Virology* **20:** 511–520.

Kiberstis, P. A., and D. Zimmern. (1984). Translational strategy of *Solanum nodiflorum* mottle virus RNA: synthesis of a coat protein precursor *in vitro* and *in vivo*. *Nucleic Acids Res.* **12:** 933–943.

Kim, K. H., and S. A. Lommel. (1994). Identification and analysis of the site of –1 ribosomal frameshifting in red clover necrotic mosaic virus. *Virology* **200:** 574–582.

Klaassen, V. A., M. L. Boeshore, E. V. Koonin, T. Tian, and B. W. Falk. (1995). Genome structure and phylogenetic analysis of lettuce infectious yellows virus, a whitefly-transmitted, bipartite closterovirus. *Virology* **208:** 99–110.

Koonin, E. V. (1991). The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. *J. Gen. Virol.* **72:** 2197–2206.

Koonin, E. V., and V. V. Dolja. (1993). Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences. *Crit. Rev. Biochem. Mol. Biol.* **28:** 375–430.

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

Kozak, M. (1989). The scanning model for translation: an update. *J. Cell Biol.* **108:** 229–241.

Kujawa, A. B., G. Drugeon, D. Hulanicka, and A.-L. Haenni. (1993). Structural requirements for efficient translational frameshifting in the synthesis of the putative viral RNA-dependent RNA polymerase of potato leafroll virus. *Nucleic Acid Res.* **21:** 2165–2171.

Le, S-Y., B. A. Shapiro, J.-H. Chen, R. Nussinov, and J. V. Maizel. (1991). RNA pseudoknots downstream of the frameshift sites of retroviruses. *Genet. Anal. Tech. Appl.* **8:** 191–205.

Lee, L., and E. J. Anderson. (1998). Nucleotide sequence of a resistance breaking mutant of southern bean mosaic virus. *Arch. Virol.* **143:** 2189–2201.

Lee, S. K., and D. Hacker. (1999). *In vitro* analysis of the RNA binding site of the R domain of the southern bean mosaic virus coat protein, abstr. W23-10, p. 100. *In* Abstracts of the 18th Annual Meeting of the American Society for Virology 1999. American Society for Virology, Amherst, MA.

Lewis, T. L., and S. M. Matsui. (1996). Astrovirus ribosomal frameshifting in an infection-transfection transient expression system. *J. Virol.* **70:** 2869–2875.

Lommel, S. A., T. L. Kendall, N. F. Siu, and R. C. Nutter. (1991). Characterisation of maize chlorotic mottle virus. *Phytopathology* **81:** 819–823.

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

Mäkinen, K., N. Arshava, T. Tamm, E. Truve, S. Zavriev, and M. Saarma. Unpublished data.

Mang, K.-Q., A. Ghosh, and P. Kaesberg. (1982). A comparative study of the cowpea and bean strains of southern bean mosaic virus. *Virology* **116:** 264–274.

Mans, R. M., M. H. Van Steeg, P. W. Verlaan, C. W. Pleij, and L. Bosch. (1992). Mutational analysis of the pseudoknot in the tRNA-like structure of turnip yellow mosaic virus RNA. Aminoacylation efficiency and RNA pseudoknot stability. *J. Mol. Biol.* **223:** 221–232.

Marczinke, B., A. J. Bloys, T. D. K. Brown, M. M. Willcocks, M. J. Carter, and I. Brierley. (1994). The human astrovirus RNA-dependent RNA polymerase coding region is expressed by ribosomal frameshifting. *J. Virol.* **68:** 5588–5595.

Matthews, R. E. F. (1982). Classification and nomenclature of viruses. Fourth Report of the International Committee on Taxonomy of Viruses. *Intervirology* **17:** 1–199.

Maule, A. J. (1991). Virus movement in infected plants. *Crit. Rev. Plant Sci.* **9:** 457–473.

Mayo, M. A., D. J. Robinson, C. A. Jolly, and L. Hyman. (1989). Nucleotide sequence of potato leafroll luteovirus RNA. *J. Gen. Virol.* **70:** 1037–1051.

McLean, B. G., E. Waigmann, V. Citovsky, and P. Zambryski. (1993). Cell-to-cell movement of plant viruses. *Trends Microbiol.* **1:** 105–109.

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

Miller, W. A., S. P. Dinesh-Kumar, and C. P. Paul. (1995). Luteovirus gene expression. *Crit. Rev. Plant Sci.* **14:** 179–211.

Mohamed, N. A. (1980). Cocksfoot mottle virus in New Zealand. *N. Z. J. Agric. Res.* **23:** 273–275.

Mohamed, N. A., and D. W. Mossop. (1981). Cynosurus and cocksfoot mottle viruses: a comparison. *J. Gen. Virol.* **55:** 63–74.

Morris-Krsinich, B. A. M., and R. Hull. (1981). Translation of turnip rosette virus RNA in rabbit reticulocyte lysates. *Virology* **114:** 98–112.

Morris-Krsinich, B. A. M., and R. L. S. Forster. (1983). Lucerne transient streak virus RNA and its translation in rabbit reticulocyte lysate and wheat germ extract. *Virology* **128:** 176–185.

Munthe, T. (1988). Norske undersøkelser vedrørende hundergrasmosaikkvirus. *Växtskyddsrapport* **53:** 90. Sveriges lantbruksuniversitet, Uppsala.

Murphy F. A., C. M. Fauquet, D. H. L. Bishop, S. A. Ghabrial, A. W. Jarvis, G. P. Martelli, M. A. Mayo, and M. D. Summers (ed.). (1995). Virus Taxonomy. Classification and Nomenclature of Viruses. Sixth Report on the International Committee on Taxonomy of Viruses. *Arch. Virol./Suppl.* **10:** 586 pp. Springer Verlag, Wein New York.

Ngon A Yassi, M. N., C. Ritzenthaler, C. Brugidou, C. Fauquet, and R. N. Beachy. (1994). Nucleotide sequence and genome characterization of rice yellow mottle virus RNA. *J. Gen. Virol.* **75:** 249–257.

Opalka, N., C. Brugidou, C. Bonneau, M. Nicole, R. N. Beachy, M. Yeager, and C. Fauquet. (1998). Movement of rice yellow mottle virus between xylem cells through pit membranes. *Proc. Natl. Acad. Sci. USA* **95:** 3323–3328.

Othman, Y., and R. Hull. (1995). Nucleotide sequence of the bean strain of southern bean mosaic virus. *Virology* **206:** 287–297.

Pinto, Y. M., and D. C. Baulcombe. (1995). Rice yellow mottle virus from Nigeria, complete genome. GenBank accession no. U23142.

Pinto, Y. M., R. A. Kok, and D. C. Baulcombe. (1999). Resistance to rice yellow mottle virus (RYMV) in cultivated African rice varieties containing RYMV transgenes. *Nat. Biotec.* **17:** 702–707.

Pringle, C. R. (1999). Virus taxonomy-1999. The Universal System of Virus Taxonomy, updated to include the new proposals ratified by the International Committee on Taxonomy of Viruses during 1998. *Arch. Virol.* **144:** 421–429.

Prüfer, D., E. Tacke, J. Schmitz, B. Kull, A. Kaufmann, and W. Rhode. (1992). Ribosomal frameshifting in plants: a novel signal directs the –1 frameshift in the synthesis of the putative viral replicase of potato leafroll luteovirus. *EMBO J.* **11:** 1111–1117.

Rabenstein, F., and H. B. Schmidt. (1979). Nachweis des Knaulgrasscheckungs-Virus (cocksfoot mottle virus) in der DDR. *Arch. Phytopathol. Pfanzenschutz* **15:** 351–354.

Rabenstein, F., and A. Stanarius. (1984). Untersuchungen zum Knaulgrasscheckungs-Virus (cocksfoot mottle virus). *Arch. Phytopathol. Pflanzenschutz* **20:** 15–31.

Ramsdell, D. C. (1979). Blueberry shoestring virus. *CMI/AAB Description of Plant Viruses* No. 204.

Randles, J. W., C. Davies, T. Hatta, A. R. Gould, and R. I. B. Francki. (1981). Studies on encapsidated viroid-like RNA. I. Characterisation of velvet tobacco mottle virus. *Virology* **108:** 111–122.

Revill, P. A., A. D. Davidson, and P. J. Wright. (1994). The nucleotide sequence and genome organization of mushroom bacilliform virus: a single-stranded RNA virus of *Agaricus bisporus* (Lange) Imbach. *Virology* **202:** 904–911.

Revill, P. A., A. D. Davidson, and P. J. Wright. (1998). Mushroom bacilliform virus RNA: the initiation of translation at the 5' end of the genome and identification of the VPg. *Virology* **249:** 231–237.

Revill, P. A., A. D. Davidson, and P. J. Wright. (1999). Identification of a subgenomic mRNA encoding the capsid protein of mushroom bacilliform virus, a single-stranded RNA mycovirus. *Virology* **260:** 273–276.

Rognli, O. A., K. Aastveit, and T. Munthe. (1995). Genetic variation in cocksfoot (*Dactylis glomerata* L.) populations for mottle virus resistance. *Euphytica* **83:** 109–116.

Rossmann, M. G. (1984). Constraints on the assembly of spherical virus particles. *Virology* **134:** 1–11.

Rossmann, M. G., C. Abad-Zapatero, M. A. Hermodson, and J. W. Erickson. (1983). Subunit interactions in southern bean mosaic virus. *J. Mol. Biol.* **166:** 37–73.

Rutgers, T., T. Salerno-Rife, and P. Kaesberg. (1980). Messenger RNA for the coat protein of southern bean mosaic virus. *Virology* **104:** 506–509.

Ryabov, E. V., E. V. Generozov, T. L. Kendall, S. A. Lommel, and S. K. Zavriev. (1994). Nucleotide sequence of carnation ringspot dianthovirus RNA-1. *J. Gen. Virol.* **75:** 243–247.

Ryabov, E. V., A. A. Krutov, V. K. Novikov, O. V. Zheleznikova, S. Y. Morozov, and S. K. Zavriev. (1996). Nucleotide sequence of RNA from the sobemovirus found in infected cocksfoot shows a luteovirus-like arrangement of the putative replicase and protease genes. *Phytopathology* **86:** 391–397.

Salerno-Rife, T., T. Rutgers, and P. Kaesberg. (1980). Translation of southern bean mosaic virus RNA in wheat embryo and rabbit reticulocyte extracts. *J. Virol.* **34:** 51–58.

Savithri, H. S., and J. W. Erickson. (1983). The self-assembly of the cowpea strain of southern bean mosaic virus: formation of $T = 1$ and $T = 3$ nucleoprotein particles. *Virology* **126:** 328–335.

Sehgal, O. P., C. H. Hsu, J. A. White, and M. Van. (1979). Enzymic sensitivity of conformationally altered virions of southern bean mosaic virus. *Phytopathol. Z.* **95:** 167–177.

Serjeant, E. P. (1964). Cocksfoot mottle virus. *Plant Pathol.* **13:** 23–24.

Serjeant, E. P. (1967). Some properties of cocksfoot mottle virus. *Ann. Appl. Biol.* **59:** 31–38.

Séron, K., and A.-L. Haenni. (1996). Vascular movement of plant viruses. *Mol. Plant-Microbe Interact.* **9:** 435–442.

Shields, S. A., M. J. Brisco, T. M. Wilson, and R. Hull. (1989). Southern bean mosaic virus RNA remains associated with swollen virions during translation in wheat germ cell-free extracts. *Virology* **171:** 602–606.

Sivakumaran, K., B. C. Fowler, and D. L. Hacker. (1998). Identification of viral genes required for cell-to-cell movement of southern bean mosaic virus. *Virology* **252:** 376–386.

Sivakumaran, K., and D. L. Hacker. (1998). The 105-kDa polyprotein of southern bean mosaic virus is translated by scanning ribosomes. *Virology* **246:** 34–44.

Taliansky, M. E., D. J. Robinson, and A. F. Murant. (1996). Complete nucleotide sequence and organisation of the RNA genome of groundnut rosette umbravirus. *J. Gen. Virol.* **77:** 2335–2345.

Thomas, J. E. (1988). Gringer chlorotic fleck virus. *CMI/AAB Description of Plant Viruses* No. 328.

Toriyama, S. (1982). Cocksfoot mottle virus in Japan. *Ann. Phytopath. Soc. Jpn.* **48:** 514–520.

Tremaine, J. G., and R. I. Hamilton. (1983). Southern bean mosaic virus. *CMI/AAB Descriptions of Plant Viruses* No. 274.

Tsai, C.-H., and T. W. Dreher. (1991). Turnip yellow mosaic virus RNAs with anti-codon loop substitutions that result in decreased valylation fail to replicate efficiently. *J. Virol.* **65:** 3060–3067.

Van der Wilk, F., M. J. Huisman, B. J. Cornelissen, H. Huttinga, and R. Goldbach. (1989). Nucleotide sequence and organization of potato leafroll virus genomic RNA. *FEBS Lett.* **245:** 51–56.

Van der Wilk, F., M. Verbeek, A. M. Dullemans, and J. F. J. M. van den Heuvel. (1997). The genome-linked protein of potato leafroll virus is located downstream of the putative protease domain of the ORF1 product. *Virology* **234:** 300–303.

Van der Wilk, F., M. Verbeek, A. Dullemans, and J. F. J. M. van den Heuvel. (1998). The genome-linked protein (VPg) of southern bean mosaic virus is encoded by the ORF2. *Virus Genes* **17:** 21–24.

Veidt, I., H. Lot, M. Leiser, D. Scheidecker, H. Guilley, K. Richards, and G. Jonard. (1988). Nucleotide sequence of beet western yellows virus RNA. *Nucleic Acids Res.* **16:** 9917–9932.

Veidt, I., S. E. Bouzoubaa, R.-M. Leiser, V. Ziegler-Graff, H. Guilley, K. Richards, and G. Jonard. (1992). Synthesis of full-length transcripts of beet western yellows virus RNA: messenger properties and biological activity in protoplasts. *Virology* **186:** 192–200.

Vincent, J. R., R. M. Lister, and B. A. Larkins. (1991). Nucleotide sequence analysis and genomic organization of the NY-RPV isolate of barley yellow dwarf virus. *J. Gen. Virol.* **72:** 2347–2355.

Walters, H. J. (1969). Beetle transmission of plant viruses. *Adv. Virus Res.* **15:** 339–363.

Weber, K. A., and O. P. Sehgal. (1982). Subgenomic RNAs in virions of southern bean mosaic virus. *Phytopathology* **72:** 909–913.

Weiland, J. J., and T. W. Dreher. (1989). Infectious TYMV RNA from cloned cDNA: Effects *in vitro* and *in vivo* of point substitutions in the initiation codons of two extensively overlapping ORFs. *Nucleic Acids Res.* **17:** 4675–4687.

Weiss, R. B., D. M. Dunn, A. E. Dahlberg, J. F. Atkins, and R. F. Gesteland. (1988). Reading frame switch caused by base-pair formation between the 3' end of 16S rRNA and the mRNA during elongation of protein synthesis in *Escherichia coli*. *EMBO J.* **7:** 1503–1507.

Wilson, T. M. (1985). Nucleocapsid disassembly and early gene expression by positive-strand RNA viruses. *J. Gen. Virol.* **66:** 1201–1207.

Wobus, C. E., J. S. Skaf, M. H. Schultz, and G. A. de Zoeten. (1998). Sequencing, genomic localization and initial characterization of the VPg of pea enation mosaic enamovirus. *J. Gen. Virol.* **79:** 2023–2025.

Wroth, J. M., and R. A. C. Jones. (1992). Subterranean clover mottle sobemovirus: its host range, resistance in subterranean clover and transmission through seed and by grazing animals. *Ann. Appl. Biol.* **121:** 329–343.

Wu, S., C. A. Rinehart, and P. Kaesberg. (1987). Sequence and organization of southern bean mosaic virus genomic RNA. *Virology* **161:** 73–80.

Xiong, Z., and S. A. Lommel. (1989). The complete nucleotide sequence and genome organization of red clover necrotic mosaic virus RNA-1. *Virology* **171:** 543–554.

Xiong, Z., K. H. Kim, T. L. Kendall, and S. A. Lommel. (1993). Synthesis of the putative red clover necrotic mosaic virus RNA polymerase by ribosomal frameshifting *in vitro*. *Virology* **193:** 213–221.

Zaumeyer, W. J., and L. L. Harter. (1943). Inheritance of symptom expression of bean mosaic virus 4. *J. Agric. Res.* **67:** 295–300.

Zavriev, S. K., C. M. Hickey, and S. A. Lommel. (1996). Mapping of the red clover necrotic mosaic virus subgenomic RNA. *Virology* **216:** 407–410.

KERAHEINA LAIGUVIIRUS: GENOOMNE ÜLESEHITUS JA TRANSLATSIOONI STRATEEGIAD

Kokkuvõte

Keraheina laiguviirus (*cocksfoot mottle virus*, CfMV) kuulub sobemoviiruste perekonda. Sobemoviiruseid on iseloomustatud kui ikosaheedrilisi taimeviirusi, mille genoomiks on üks *cap*'imata ja polüadenüleerimata (+)-ahelaga RNA molekul ning mis levivad taimelt taimele nii mehaaniliste vigastuste teel kui ka mardikatest vektorite abil. Sobemoviiruste peremeestaimede ring on kitsas. Oluliseks muudavad sobemoviiruste perekonna eelkõige asjaolud, et need viirused paljunevad nakatatud taimedes väga efektiivselt, viiruspartiklid on silmapaistvalt stabiilsed ning mõned perekonna liigid on olulised patogeenid kultuurtaimedele.

Käesoleva doktoritöö esimeseks eesmärgiks oli kindlaks määrata CfMV Norra isolaadi (CfMV-NO) genoomi nukleotiidne järjestus ja struktuur. CfMV-NO genoom, mis on 4082 nukleotiidi pikk, sisaldab nelja avatud lugemisraami (*open reading frame*, ORF). Sarnaselt teiste sobemoviirustega on CfMV genoomi 5' terminaalses osas väike ORF1 ja 3' terminaalses osas kattevalku kodeeriv ORF. Teine ORF kodeerib enamusel sekveneeritud sobemoviirustel suurt polüproteiini, mis protsessitakse vähemalt funktsionaalseks proteaasiks, viiruse RNA-ga seonduvaks valguks (*genome-linked protein*, VPg) ning viiruse RNA-sõltuvaks RNA polümeraasiks (*RNA dependent RNA polymerase*, RdRp). Erandiks on siin CfMV, kus see ORF on asendunud kahe osaliselt kattuva avatud lugemisraamiga ORF2a ning ORF2b. Samuti puudub CfMV-l polüproteiini kodeeriva ORF-i sisse –1 lugemisraamis peidetud väike ORF3, mis on iseloomulik ülejäänud sobemoviirustele. Eksperimentaalselt on tõestatud, et CfMV ORF2a kodeerib VPg molekuli, kusjuures VPg kodeeriv regioon paikneb oletatavast proteaasi domeenist allpool (Mäkinen *et al.*, avaldamata tulemused). Samas on proteaasi ja RdRp domeenide olemasolu P2a ja P2b-s senini teada vaid arvutianalüüside põhjal.

In vitro translatsioonil nisuidude ekstraktis sünteesitakse CfMV RNA-lt nelja põhilist valku molekulmassidega 100, 71, 34 ja 12 kDa. Käesoleva töö raames näidati, millistelt geenidelt need produktid transleeritakse. Samuti uuriti, milliseid translatsiooni strateegiaid kasutab CfMV oma valkude ekspressioonil. Selgus, et CfMV genoomne RNA töötab kui bitsistroonne mRNA. CfMV genoomi 5' terminaalne osa sisaldab kahte ORF-i (ORF1 ja ORF2a), mille translatsioon initseeritakse genoomselt RNA-lt. Kasutades immunosadestamist spetsiifiliste polüklonaalsete antikehadega, näidati, et CfMV 12 kDa ja 71 kDa valgud kodeeritakse vastavalt ORF1 ning ORF2a poolt. Seejuures toimub ORF2a

translatsiooni initsiatsioon suure tõenäosusega lekkiva skaneerimise tõttu läbi ORF1, kuna ORF1 translatsiooni initsiatsiooni signaali ümbrus ei ole taimedele soodsas kontekstis.

CfMV oletatav RdRp (kodeeritakse ORF2b poolt) transleeritakse üksnes liitvalguna polüproteiini koosseisus ribosomaalse raaminihke toimumise tõttu –1 suunas. Immunosadestamise katsed näitasid, et 100 kDa suurune CfMV *in vitro* translatsiooni produkt on ainus valk, mille tunneb ära P2b spetsiifiline antikeha. CfMV –1 ribosomaalse raaminihke signaal koosneb kahest iseloomulikust motiivist: nn. libisevast järjestusest (5' UUUAAAC) ning sellele järgnevast tüvi-aas struktuurist. See signaal indutseerib –1 ribosomaalset raaminihet reportergeeni kontekstis efektiivsusega 26–29% ning ORF2a-2b kontekstis efektiivsusega 10,6±1,4%. CfMV-sarnaseid –1 ribosomaalse raaminihke signaale on võimalik leida ka teiste sekveneeritud sobemoviiruste genoomidest enne ORF3 AUG koodonit. Seni pole üheselt näidatud ORF3 ekspresseerumine *in vivo* ning pole teada ORF3 translatsiooni initsiatsiooni mehhanismid. Võiks arvata, et kui üldse, siis ekspresseerub neil viirustel ORF3 –1 ribosomaalse raaminihke abil.

CfMV kattevalk on ainus viiruse valkudest, mida sünteesitakse subgenoomselt RNA-lt. *Northern blot* analüüs näitas, et CfMV-ga nakatatud odrataimedest on võimalik detekteerida lisaks 4,1 kb suurusele genoomsele RNA-le ka ühte CfMV spetsiifilist RNA-d suurusega 1,2 kb. See genoomsest RNA-st lühem RNA molekul võib olla subgenoomne RNA, kuna kattevalgu translatsiooni initsiatsiooni koodon asub 989 nukleotiidi 3' terminusest ülespoole. Kasutades immunosadestamist kattevalgu spetsiifilise antikehaga, selgus, et 34 kDa suurune CfMV *in vitro* translatsiooni produkt on viiruse kattevalk, mis süntesitakse viiruspartiklisse pakitud subgenoomselt RNA-lt.

ACKNOWLEDGEMENTS

This work has been carried out during the years 1994–1999 at the Laboratory of Molecular Genetics, National Institute of Chemical Physics and Biophysics in Tallinn.

First of all I would like to express my sincere, infinite thanks to *Prof. Mart Saarma*, *Prof. Jaanus Remme*, and *Prof. Erkki Truve*, my supervisors during these years, for their guidance, all possible support, and endless patience. I am especially grateful to Erkki for inviting me to work in Tallinn, for introducing me the world of plant viruses, and for his invaluable advice and friendship.

I would like to express my gratitude to *Prof. Henni Kallak*, my first supervisor for teaching me to plan the experiments and work in the lab.

Particularly I would like to thank *Dr. Vigfrid Næss* from DYNAL Microbiology R&D, Oslo, Norway, and *Dr. Kristiina Mäkinen* from Institute of Biotechnology, University of Helsinki, Finland. Viffi initiated this project when she worked at the Norwegian Crop Research Institute, Ås, Norway. Her enthusiasm has been of great support to me throughout the study. I am grateful to Kristiina for the productive and excellent collaboration for so many years and for valuable discussions.

I wish to express my warmest thanks to all the present and former "cellar" lab people. It was a great pleasure to work with You in the same lab. In particular I would like to thank *Lenne Nigul* and *Marie-Lise Bouscaren* for always having good ideas and helpful hand and for creating so friendly lab atmosphere. Special thanks to *M. Sc. Merike Meier*, *M. Sc. Gabriela Lasn* and *Silva Sütt* who have been not only a good colleagues, but also a patient and enthusiastic students.

I wish to express my gratitude to all people at Laboratory of Molecular Genetics, National Institute of Chemical Physics and Biophysics for creating a pleasant and stimulating working environment. Specially I would like to mention *Dr. Heiti Paves*, *Dr. Lilian Järvekülg*, *Dr. Leena Andrejeva*, *Dr. Andres Merits*, *Dr. Ats Metsis*, *Dr. Merike Kelve*, *Anne Kuusksalu*, *Külli Samuel*, *Monika Drews*, *Anu Aaspõllu* for always sharing the knowledge and experience and for friendly company.

The same thought to people from the Institute of Molecular and Cell Biology, University of Tartu. My most sincere thanks are due to *Dr. Tanel Tenson*, *Dr. Urmas Saarma*, *Dr. Aivar Liiv* and *Tõnu Margus* for their readiness to tolerant my questions and give good experimental advise. I also want to thank *M. Sc. Kadri Tomasson* and *Kai Virumäe* for their support and friendship.

During these years I have been lucky to collaborate with many nice people from different laboratories — *Dr. Teemu Teeri* and *Dr. Nisse Kalkkinen* from Institute of Biotechnology, University of Helsinki, Finland; *Dr. Tor Munthe* and *Dr. Dag-Ragnar Blystad* from the Norwegian Crop Research Institute, Ås, Norway; *Dr. Sergei Zavriev* from Institute of Agricultural Biotechnology, Moscow, Russia; *Dr. Bavid C. Baulcombe*, *Dr. Yvonne Pinto*, and *Dr. Isabelle Malcuit* from The Sainsbury Laboratory, John Innes Centre, Norwich, UK; *Prof. Paul Pumpens* and *Dr. Andris Zeltins* from Biomedical Research and Study Centre, University of Latvia, Riga; *Dr. Milvi Agur* from Institute of Experimental Biology, Estonia. My sincere thanks to all of You.

I also want to thank all past and present members of Saarma's "potato" and "neuro" groups for good attitude, specially *Dr. Ülo Puurand*, *Dr. Urmas Arumäe*, *Dr. Mati Reeben*, *Dr. Dejin Guo*, *B.Sc. Jimmy Lucchesi*, and *Dr. Pia Runeberg-Roos*.

Dr. Lilian Järvekülg is acknowledged for critical reading the manuscript, and *Jüri Valge* , *Kersti Valge* and *Mette Prasse Hartov* for correcting the language of this thesis.

Finally, I would like to thank all my friends for being around during these years. I want to thank my parents, *Silvi* and *Uno Tamm*, my sister *Karin* and brother *Peeter* for their support, understanding and love. Most of all, I am indebted to my husband *Erik* for his trust, patience and love.

This work was financially supported by research grants from Estonian Science Foundation, EC INCO-Copernicus programme and by individual fellowships from FEBS, Centre International Mobility, Nordic Council of Ministers.

PUBLICATIONS

I

Journal of General Virology 1995, **76:** 2817–2825.
Reprinted by courtesy of Society for General Microbiology

Journal of General Virology (1995), **76**, 2817–2825. *Printed in Great Britain*

Characterization of cocksfoot mottle sobemovirus genomic RNA and sequence comparison with related viruses

Kristiina Mäkinen,[1]* Tiina Tamm,[2] Vigfrid Næss,[3] Erkki Truve,[2] Ülo Puurand,[1] Tor Munthe[3] and Mart Saarma[1]

[1] *Institute of Biotechnology, PO Box 56, Viikinkaari 9, FIN-00014, University of Helsinki, Helsinki, Finland,* [2] *Institute of Chemical Physics and Biophysics, Estonian Academy of Sciences, Akadeemia tee 23, EE0026 Tallinn, Estonia and* [3] *The Norwegian Crop Research Institute, Fellesbygget, N-1432 Ås, Norway*

The genome of cocksfoot mottle virus (CfMV) is a positive-sense ssRNA molecule of 4082 nucleotides as revealed by sequencing the entire genome. The 5′-untranslated region of the genome is 69 nucleotides and the 3′-untranslated region is 225 nucleotides in length. The coding region contains four open reading frames (ORFs). The organization of CfMV ORFs differs significantly from that of the previously sequenced sobemoviruses southern bean mosaic virus and rice yellow mottle virus. ORF1 encodes a protein having a calculated molecular mass of 12·3 kDa. The function of this protein is unknown. The next ORF codes for the putative VPg and serine protease. The ORF2a product consists of 568 amino acids, with a calculated molecular mass of 60·9 kDa. The replicase of CfMV is translated as part of a polyprotein by −1 ribosomal frameshifting in ORF2a. The calculated molecular mass of the trans-frame protein is 103·4 kDa. ORF3 encodes the 27·6 kDa coat protein. This has been verified by amino acid sequencing of the CfMV coat protein N terminus. Northern blots of total RNA from CfMV-infected barley leaves reveal the 4·1 kb genomic RNA band and one virus-specific band of 1·2 kb, which may represent a subgenomic RNA for coat protein synthesis.

Introduction

Cocksfoot mottle virus (CfMV) is a member of the sobemovirus group (Rybicki, 1991). The virus particles are isometric and 30 nm in diameter. The CfMV genome is a monopartite ssRNA molecule with an M_r of $1{\cdot}45 \times 10^6$. Its main host is cocksfoot (*Dactylis glomerata*), an important herbage grass in the Nordic countries. Wheat, oat and barley are experimental hosts. The entire nucleotide sequence of the RNA genome has been reported for three members of the sobemovirus group: the cowpea strain of southern bean mosaic virus (SBMV-C; Wu *et al.*, 1987), the bean strain of SBMV (SBMV-B; Othman & Hull, 1995) and rice yellow mottle virus (RYMV; Ngon A Yassi *et al.*, 1994). Recently we have reported that the CfMV genome organization differs substantially from that of SBMV-C and RYMV, as its polyprotein is translated from two different open reading frames (ORFs) via −1 ribosomal frameshifting (Mäkinen *et al.*, 1995). In this paper the entire nucleotide sequence of CfMV is reported and compared to related viruses. The possible origins of the *in vitro* translation products produced from CfMV genomic RNA are discussed.

* Author for correspondence. Fax +358 0 4346046. e-mail Makinen @ operoni.helsinki.fi

The nucleotide sequence data reported in this paper have been deposited with EMBL Database under accession number Z48630.

Methods

Virus purification and nucleic acid extraction. CfMV origin, propagation and the purification of virions were as described earlier (Mäkinen *et al.*, 1995). Viral RNA was purified by treating the virus with proteinase K and SDS (Dougherty & Hiebert, 1980) followed by phenol–chloroform extraction and ethanol precipitation (Puurand *et al.*, 1992).

CfMV cDNA synthesis, cloning, sequencing and analysis. CfMV cDNA synthesis and cloning in the Lambda gt11 Vector (Promega), screening of the resulting library, isolation of CfMV-specific cDNA clones and their sequencing have been described (Mäkinen *et al.*, 1995). Both subcloning using standard methods (Sambrook *et al.*, 1989) and generation of nested unidirectional deletions with exonuclease III (Nested Deletion Kit; Pharmacia) were used to make clones suitable for sequencing. Genetics Computer Group programs (University of Wisconsin; Devereux *et al.*, 1984), DNA Strider (Commissariate a l'Energie Atomique, France), DNAid+ (Ecole Polytechnique, France), Clustal V (Higgins *et al.*, 1992) and PCGENE (University of Geneva, Switzerland) were used for sequence analysis.

N-terminal amino acid sequence of the coat protein. The coat protein of CfMV was separated out by SDS–PAGE (12·5% gel) as described by Laemmli (1970). The gel was aged for 3 days at 4 °C prior to electrophoresis (Staunton *et al.*, 1989). The coat protein was electro-

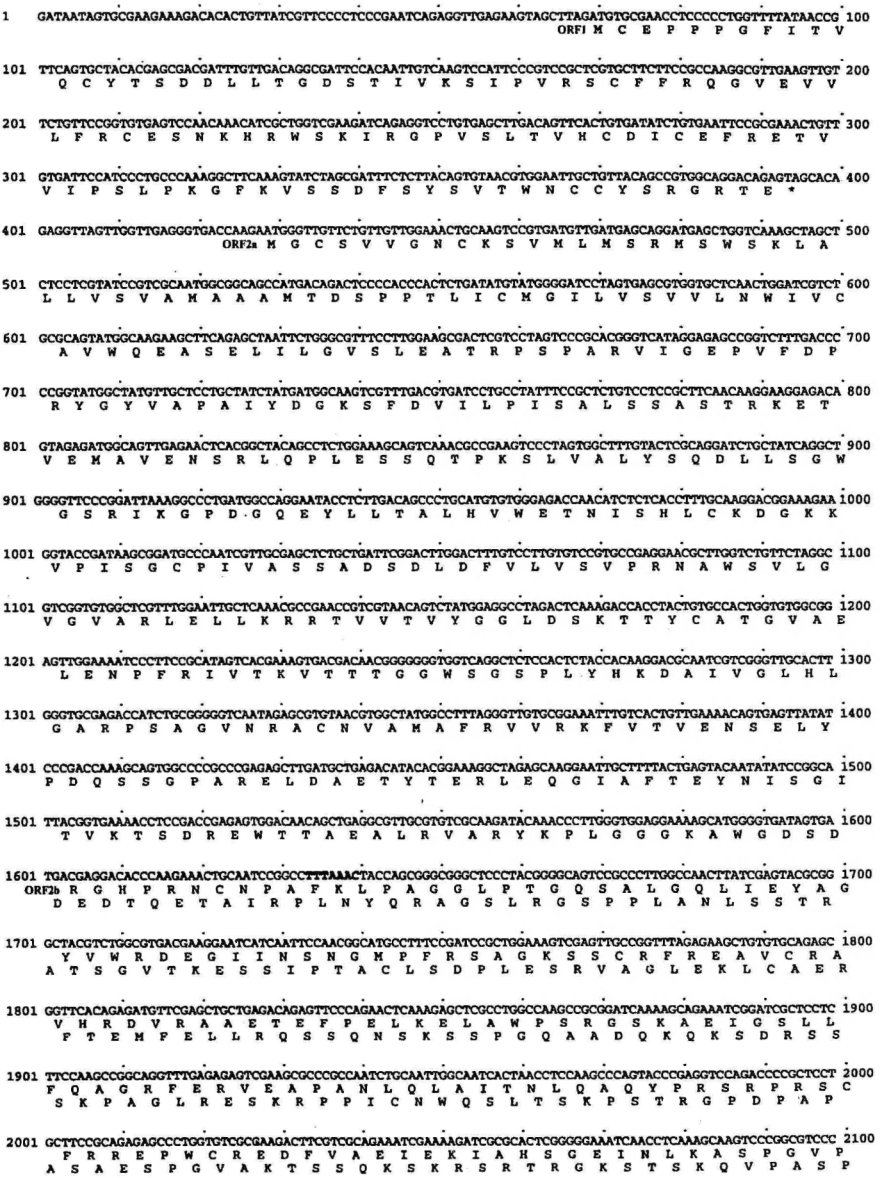


Fig. 1. For legend see opposite.

```
2101 CCTCGCCGAAATCGGGGTCAGCAACCAGCAAGTAATTGATGTAGCTTGGCCCTTGGTGTGTGAGGCTGTGGTGGAAAGGCTCCATGCCTTGGCATCCGTA 2200
      L  A  E  I  G  V  S  N  Q  Q  V  I  D  V  A  W  P  L  V  C  E  A  V  V  E  R  L  H  A  L  A  S  V
      S  P  K  S  G  S  A  T  S  K  *

2201 GACCCGCGCCAGCACGATTGGTCTCCAGAAGAACTCGTGAAAAGAGGATTGTGCGATCCAGTGAGACTCTTTGTCAAACAGGAACCTCATTCTCGACAGA 2300
      D  P  R  Q  H  D  W  S  P  E  E  L  V  K  R  G  L  C  D  P  V  R  L  F  V  K  Q  E  P  H  S  R  Q  K

2301 AGATTGAACAAGGTCGCTTCCGACTTATCTCTTCTGTCTCACTTGTAGATCAACTGGTAGAACGCATGCTTTTCGGACCCCAGAATACCACAGAGATTGC 2400
      I  E  Q  G  R  F  R  L  I  S  S  V  S  L  V  D  Q  L  V  E  R  M  L  F  G  P  Q  N  T  T  E  I  A

2401 TTTGTGGCATTCAAATCCCTCTAAACCCGGGATGGGTCTGTCGAAAGCTAGCCAAGTCGCGCTGTTATGGGAGGATTTGGCGCGCAAGCACCAAACCCAC 2500
      L  W  H  S  N  P  S  K  P  G  M  G  L  S  K  A  S  Q  V  A  L  L  W  E  D  L  A  R  K  H  Q  T  H

2501 CCGGGTGCCATGGCTGATATCTCAGGGTTTGACTGGTCCGTTCAGGATTGGGAACTCTGGGCTGACGTATCTATGAGGATTGAGCTAGGCTCGTTTCCAG 2600
      P  G  A  M  A  D  I  S  G  F  D  W  S  V  Q  D  W  E  L  W  A  D  V  S  M  R  I  E  L  G  S  F  P  A

2601 CGTTGATGGCGAAGGCAGCTATATCGCGGTTTTACTGCTTGATGAACGCTACCTTCCAATTAACGAATGGAGAGCTCTTAGCCCAGGAACTGCCAGGACT 2700
      L  M  A  K  A  A  I  S  R  F  Y  C  L  M  N  A  T  F  Q  L  T  N  G  E  L  L  A  Q  E  L  P  G  L

2701 CATGAAGTCCGGGTCGTACTGTACCTCCAGCTCCAACTCCCGCATTCGCTGCCTTATGGCTGAGCTCATCGGCTCGCCTTGGTGTATCGCAATGGGAGAT 2800
      M  K  S  G  S  Y  C  T  S  S  S  N  S  R  I  R  C  L  M  A  E  L  I  G  S  P  W  C  I  A  M  G  D

2801 GATTCTGTTGAGGGATGGGTCGATGATGCTCCCAGGAAATATTCAGCACTAGGCCATCTCTGTAAAGAGTATGAAGCGTGTCCAGTGCTCCCCAACGGAG 2900
      D  S  V  E  G  W  V  D  D  A  P  R  K  Y  S  A  L  G  H  L  C  K  E  Y  E  A  C  P  V  L  P  N  G  D

2901 ATTTGAAGGAGGTGAGCTTTGCTCCCATCTCATCTCTAAGGGTCGAGCGGAACTGGAGACGTGGCCCAAGTGTCTGTTTCGATATCTCAGTGGACCACA 3000
      L  K  E  V  S  F  C  S  H  L  I  S  K  G  R  A  E  L  E  T  W  P  K  C  L  F  R  Y  L  S  G  P  H

3001 TGATGTGGAAAGCTTGGAGATGGAGCTAAGCTCTTCTCGCCGCTGGGGCCAGATAGTCAGATACCTCCGCCGGATTGGCCGGGTCTCCGGAAATGATGGT 3100
      D  V  E  S  L  E  M  E  L  S  S  S  R  R  W  G  Q  I  V  R  Y  L  R  R  I  G  R  V  S  G  N  D  G
                                                                                    ORF3  M  M  V

3101 GAGGAAAGGAGCAGCAACGAAAGCCCCGCAACAACCAAAACCCAAGGCTCAGCAGCAGCCTGGGGGCCGCCGCAGGAGGCGTGGCCGGTCGATGGAGCCA 3200
      E  E  R  S  S  N  E  S  P  A  T  T  K  T  Q  G  S  A  A  A  W  G  P  P  Q  E  A  W  P  V  D  G  A  S
      R  K  G  A  A  T  K  A  P  Q  Q  P  K  P  K  A  Q  Q  Q  P  G  G  R  R  R  R  R  G  R  S  M  E  P

3201 GTCTCTCGACCTTTGAACCCTCCAGCAGCGGTTGGTTCCACCTTGAAGGCTGGTAGAGGCAGGACCGCTGGTGTGAGTGACTGGTTTGATACCGGCATGA 3300
      L  S  T  F  E  P  S  S  S  G  W  F  H  L  E  G  W  *
      V  S  R  P  L  N  P  P  A  A  V  G  S  T  L  K  A  G  R  G  R  T  A  G  V  S  D  W  F  D  T  G  M  I

3301 TCACCAGTTACCTAGGTGGTTTTCAGCGCACAGCTGGAACTACCGACTCCCAGGTATTCATTGTGTCACCCGCTGCATTGGACCGTGTTGGAACCATAGC 3400
      T  S  Y  L  G  G  F  Q  R  T  A  G  T  T  D  S  Q  V  F  I  V  S  P  A  A  L  D  R  V  G  T  I  A

3401 TAAGGCGTATGCGTTGTGGAGACCTAAGCATTGGGAAATCGTGTATCTCCCCAGATGTTCTACTCAGACAGATGGGTCGATTGAAATGGGATTTCTCTTA 3500
      K  A  Y  A  L  W  R  P  K  H  W  E  I  V  Y  L  P  R  C  S  T  Q  T  D  G  S  I  E  M  G  F  L  L

3501 GATTATGCAGATAGCGTCCCAACCAACACCCGAACTATGGCCTCCTCCACCTCCTTCACAACTTCAAACGTGTGGGGAGGTGGGGATGGCAGCAGTTTGT 3600
      D  Y  A  D  S  V  P  T  N  T  R  T  M  A  S  S  T  S  F  T  T  S  N  V  W  G  G  G  D  G  S  S  L  L

3601 TGCACACGTCCGTGAAGTCCATGGGCAACGCTGTTACGAGCGCTCTTCCTTGTGATGAGTTCTCGAACAAATGGTTTAAGTTATCTTGGAGCACTCCTGA 3700
      H  T  S  V  K  S  M  G  N  A  V  T  S  A  L  P  C  D  E  F  S  N  K  W  F  K  L  S  W  S  T  P  E

3701 GGAATCGGAGAACGCTCATTTAACTGACACCTACGTTCCGGCGCGCTTCGTGGTGAGGTCTGACTTCCCAGTAGTCACGGCTGACCAACCCGGTCATCTC 3800
      E  S  E  N  A  H  L  T  D  T  Y  V  P  A  R  F  V  V  R  S  D  F  P  V  V  T  A  D  Q  P  G  H  L

3801 TGGCTCCGGTCCAGGATCTTGCTCAAGGGATCAGTTTCCCCTTCTACAAATTTGTAGTCAAGTACTGCAGGGATGACAAACCTGTAAAACAGTGTCCGCG 3900
      W  L  R  S  R  I  L  L  K  G  S  V  S  P  S  T  N  L  *

3901 CTTGGTGGAGTCGCGTTCGTCCACTTGGAAATCTAACTCCAAAGCTTTGCTTGAAGATTGACGGATTGGTATCATCTCCGTAAGGATACCGACCACTGGC 4000

4001 TGGGCTCGTCACCCCTCCTTTGTGTCTCACAAACGACGAATCTGCGCTCGGCCAATGACAGGATCATGGCCGGGGCAGGTGC 4082
```

Fig. 1. The complete nucleotide sequence of the CfMV genomic cDNA. The amino acid sequences of the ORFs 1, 2a, 2b and 3 are presented under the nucleotide sequence in single letter code. The slippery heptanucleotide used for the −1 ribosomal frameshift is shown in bold.

transferred onto PVDF membrane and visualized by staining with Coomassie brilliant blue (Staunton *et al.*, 1989). The band corresponding to the coat protein was excised and applied to the reaction cartridge of a gas–pulsed-liquid sequencer equipped with an on-line phenylthiohydantoin amino acid analyser (Kalkkinen & Tilgmann, 1988).

PCR amplification of the cDNA. Some clones were obtained by PCR amplification. Both single- and double-stranded cDNA were used as templates for the PCR amplification using a Perkin Elmer DNA Thermal Cycler or MJ Research Minicycler. The reaction mixture contained 10 mM-Tris–HCl pH 8·8, 1·5 mM-$MgCl_2$, 50 mM-KCl and 0·1 % Triton X-100, 0·1 mM-dNTP, 80 pmol of each primer, 50 ng of template cDNA and 2 units of DynaZyme DNA polymerase (Finnzymes) or *Taq* DNA polymerase (Perkin Elmer, Fermentas). After denaturation of the DNA at 95 °C for 5 min, the reaction mixtures were subjected to 30 cycles of 1·5 min at an annealing temperature calculated for each primer pair used, 2·5 min at 72 °C and 1·5 min at 95 °C. The PCR fragments obtained were cloned into vectors pGEM-T (Promega), pCR1000 (Invitrogen) or pUC57 (Fermentas).

Cloning of the 5′ and 3′ ends. The 5′-terminal sequence of CfMV was determined using the PCR product. The first strand of cDNA was synthesized using an oligodeoxyribonucleotide complementary to nucleotides 357–380 in the CfMV sequence: 5′ CGGCTGTAACAGCAATTCCACGTT (primer #4615). Single-stranded DNA was purified with a Sephaglas PhagePrep Kit (Pharmacia) or Wizard DNA Clean-Up System (Promega) and a poly(A) tail was added by terminal transferase (Promega). The DNA product was amplified by PCR using $(dT)_{25}$dN as the second primer.

To clone the 3′ end, 1 µg of viral RNA was polyadenylated with 5 units of poly(A) polymerase (Pharmacia) according to the manufacturer's instructions. First-strand synthesis of cDNA was primed with 5′ AACTGGAAGAATTCGCGGCCGCAGGAA$(T)_{18}$ oligodeoxynucleotide. The cDNA was then subjected to PCR using the same primer and a specific primer 5′ CAGGATCTTGCTCAAGGGATCAGT (primer #4614; nucleotides 3812–3835 in the CfMV sequence).

Primer extension. Three µg of CfMV RNA was subjected to primer extension using the oligodeoxynucleotide primer 5′ CTAAGCTACTTCTCAACCTC, located 50–70 nucleotides from the 5′ terminus of the obtained 5′-end clones. These clones were also used to provide the size markers for the urea–polyacrylamide gel by using the same primer for chain-termination sequencing reactions (Sambrook *et al.*, 1989).

Isolation of the total RNA and Northern blot analysis of the subgenomic RNA. Total RNA from CfMV-infected barley (cv. Lise) leaves and from control barley leaves was isolated using the method described by Logemann *et al.* (1987). Electrophoresis in formaldehyde-containing agarose gels and hybridizations were done according to the blotting and hybridization protocols for Hybond membranes (Amersham). Blotting of the gel onto Hybond-N nylon filters was done using alkaline transfer of total RNA as described by Löw & Rausch (1994). Eight µg of total RNA was loaded per lane. A probe covering 364 3′-terminal nucleotides of CfMV ORF3 was randomly primed (Feinberg & Vogelstein, 1983) and labelled with [^{32}P]dCTP (NEN).

Results and Discussion

The CfMV cDNA clones picked up from the library or obtained by PCR were sequenced on both strands. The sequence of 98·2 % of the genome was obtained from at least two independent clones. The complete nucleotide sequence of CfMV RNA was 4082 nucleotides long, being the shortest sequenced sobemovirus (Fig. 1). Terminal sequences were verified by polyadenylation of the termini prior to PCR amplification. The nucleotides determined by primer extension matched the sequences obtained with 5′-end clones (Fig. 2). The 3′-end sequence was proven by sequencing three cDNA clones obtained by RT–PCR of three independently isolated and *in vitro* polyadenylated CfMV RNAs. All of the sequences were identical. The 5′-untranslated region was 69 nucleotides and the 3′-untranslated region was 225 nucleotides in length.

Coding capacity of CfMV

The CfMV genome contains four ORFs (Fig. 3), which are compact and overlap each other, except that between ORFs 1 and 2a there is a 32 nucleotide intergenic region. The A of the first AUG is located at nucleotide 70. The context of this AUG is poor when compared to the consensus sequence for translation initiation in plants, AACAAUGGC, described by Lütcke *et al.* (1987). However, as in RYMV, this is the only translation initiation codon in any frame upstream from ORF2a. ORF1 encodes a protein with 109 amino acids and a calculated molecular mass of 12·3 kDa. *In vitro* translation of CfMV genomic RNA revealed four protein products of 100, 71, 34 and 16 kDa (Mäkinen *et al.*, 1995). The smallest protein may represent the ORF1 product. Its function is unknown, but Othman & Hull (1995) have suggested a movement protein function. CfMV ORF1, at nucleotides 236–312, has 59·2 % identity with the beet yellows closterovirus 65 kDa protein gene, which is related to the heat shock protein 70 gene. The 65 kDa protein is believed to provide a virus transport function (Agranovsky *et al.*, 1994).

CfMV, in contrast to other fully sequenced sobemoviruses, lacks the continuous large ORF that codes for the putative VPg, protease and replicase. Instead it has two overlapping ORFs, called 2a and 2b, which code for these proteins (Fig. 3). In the CfMV genome, ORF2a begins 32 nucleotides after the stop codon for ORF1. ORF2a comprises 1710 nucleotides (from 424 to 2133) and encodes a protein of 568 amino acids with a calculated molecular mass of 60·8 kDa. ORF2b is in the −1 reading frame compared to ORF2a. Consensus signals for a −1 ribosomal frameshift event (Jacks *et al.*, 1988) can be found at the beginning of the nucleotides that overlap in ORF2a and ORF2b: the slippery heptanucleotide sequence 5′ UUUAAAC and a stem–loop structure just downstream. We showed previously that the putative replicase of CfMV is translated *in vitro* as a part of a polyprotein by −1 ribosomal frameshifting (Mäkinen *et al.*, 1995). ORF2b has 1650 nucleotides (1603–3253) and encodes a 56·3 kDa protein comprising 504 amino acids from the first AUG. The

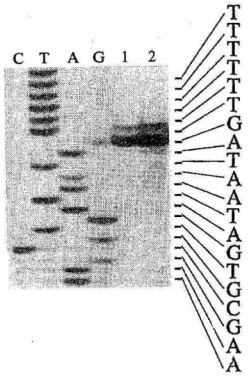


Fig. 2. Primer extension (lanes 1 and 2) and the sequence of the 5′-end cDNA clone of CfMV.

transframe protein is 942 amino acids long and has a calculated molecular mass of 103·4 kDa. The size of the transframe protein and the largest CfMV RNA *in vitro* translation product correlate well (Mäkinen *et al.*, 1995).

The N-terminal part of the ORF2a-encoded protein is presumably the VPg by analogy to the location of the putative VPgs of other sequenced sobemoviruses. In addition to the conserved 14 and 9 amino acid blocks proposed by Ngon A Yassi *et al.* (1994), the sequence motif GxPxFDPxYG can be found in the putative VPg of each of these viruses (Figs 3 and 4). Despite the fact that these three motifs are well conserved among all the sobemoviruses, their location in the putative VPg domain is different (Fig. 4). In general, no common sequence elements in plant viral VPgs have been found (Gorbalenya & Koonin, 1993). However, an amino acid triplet FDP is present in the VPg of the potyviruses tobacco etch virus (Allison *et al.*, 1986; Murphy *et al.*, 1991), potato Y virus (Robaglia *et al.*, 1989) and potato A virus (Puurand *et al.*, 1994). The significance of motif GxPxFDPxYG is unknown.

The next protein coded by ORF2a has features of a typical sobemovirus/luteovirus serine protease. It has the H and D residues common to all serine proteases, and an S at the site where the picornavirus 3C protease has C in its catalytic centre (Gorbalenya *et al.*, 1988). E/S and E/T cleavage sites were proposed for the sobemovirus serine proteases by Gorbalenya *et al.* (1988). There are several putative cleavage sites in the ORF2a-coded protein. One putative protease cleavage site, located after the protease consensus sequence, is conserved among the sobemoviruses SBMV-C, RYMV

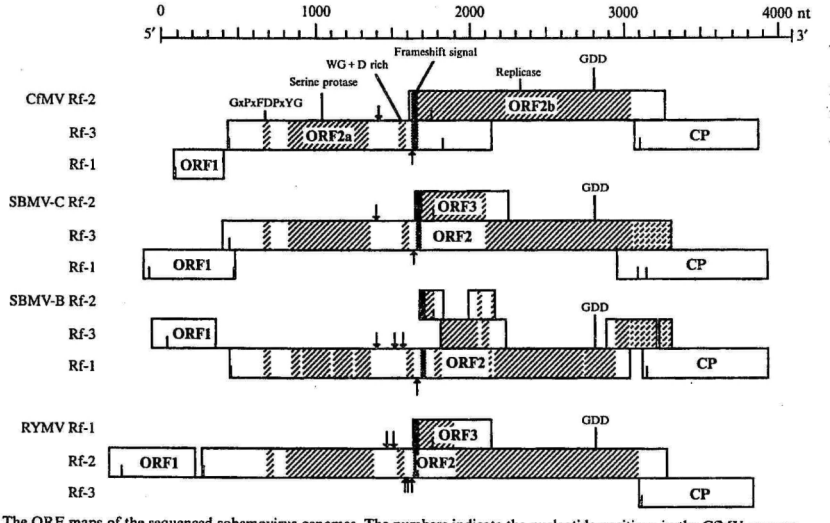


Fig. 3. The ORF maps of the sequenced sobemovirus genomes. The numbers indicate the nucleotide positions in the CfMV genome. Rf, reading frame; CP, coat protein. The first AUG of each ORF is marked by a vertical line. Different sequence motifs discussed in the text are indicated. Putative E/T and E/S protease cleavage sites are indicated by arrows. Amino acid sequences related to the CfMV polyprotein-coding regions are shaded.

```
CfMV     26  VSVAMAAAMTDSPPTLICM-GILVSVVLNWIVCAVWQEASELILGVSL
RYMV     27  AGLIAAAIVSERLPVTASLWAIPSAIIANWIVLSAHESFSRFVEGVEI
SBMV-C   24  AFIYGSYDPSHNIPIVALMTLCATGLWLSTSVVSFGIRYVRVRVSPEK
SBMV-B   25  FDIHRGVTNPAHYPIVAWIPRSLLCL-VERVIRDIPYKYVRTRL---L
                 !  !   !   *!   !       !  !  !  !

CfMV     73  EATRPSP-ARVIGEPVFDP-RY-GYVAPAIYDGKSFDVILPIS----A
RYMV     75  EPMSMLRYGKVQSAPRFDP-SR-GYVVDVSYNGHVIPVILDFTTT-TA
SBMV-C   72  TQNRTI-YVSS-GLPHFDP-VY-GVVKKCEPMGGGPAIELQVNPSWIH
SBMV-B   67  QRRRLPGY-IAQS-AVFVPLVYTDLIARMSETAWTTGTGYQLMSTGIR
                !         !  !  *!*  !  !  !       !     !

CONS             g p FdP  y g
```

Fig. 4. Amino acid alignment of the putative VPg regions of CfMV, RYMV, SBMV-C and SBMV-B. The conserved sequence motif GxPxFDPxYG is indicated. The 14 and 9 amino acid motifs proposed by Ngon A Yassi *et al.* (1994) are underlined. CONS, conserved amino acids; (*), conserved in all sequences shown; (!), partially conserved.

and CfMV (Wu *et al.*, 1987; Ngon A Yassi *et al.*, 1994; Fig. 3) and other viruses; potato leafroll virus (PLRV) and beat western yellows virus (BWYV) (both belonging to luteovirus subgroup II; Mayo *et al.*, 1989; Veidt *et al.*, 1988), pea enation mosaic virus (PEMV; Demler & de Zoeten, 1991), mushroom bacilliform virus (MBV; Revill *et al.*, 1994) and a human astrovirus (H-Ast2; Jiang *et al.*, 1993). The use of this E/S cleavage site in CfMV polyprotein processing would produce proteins with calculated molecular masses of 35·6 and 68·6 kDa. *In vitro* translation of the CfMV genomic RNA revealed a protein of 34 kDa, which correlates well with the joint size of VPg and the protease.

In the ORF2a protein product we have located a highly conserved WAD/WGD amino acid sequence followed by a D/E-rich domain. This amino acid motif or a very similar one can also be found in SBMV, RYMV, PLRV, BWYV, PEMV, MBV and H-Ast2 in front of the putative or proven −1 frameshift signal (Figs 3 and 5). Its function is unknown. The remainder of the ORF2a-encoded protein is very basic.

The amino acid sequence of the putative replicase, encoded by ORF2b, contains the GDD motif typical of RNA-dependent RNA polymerases (Kamer & Argos, 1984) near the C terminus of the protein (Fig. 3). *In vitro* translation of the CfMV RNA results in a 70 kDa product from this ORF. The theoretical size of the replicase produced using the E/T cleavage site in front of the replicase is 60·7 kDa (Fig. 3). If the E/S cleavage site after the protease consensus sequence is exploited instead then the protein product would be 68·6 kDa. This could be the 70 kDa product seen on SDS–PAGE.

The mechanism for the translation initiation of ORF2a is currently unknown. Since the AUG for ORF1 is in a poor context for the initiation of protein synthesis, it could permit the 40S ribosomal subunit to bypass ORF1 by leaky scanning (Kozak, 1989) and initiate translation at the first AUG of ORF2a. This AUG is surrounded by a sequence that satisfies the proposed initiation context in plants (AAGAAUGGG versus AACAAUGGC; proposed by Lütcke *et al.*, 1987).

ORF3 encodes the coat protein. This was verified by comparing the deduced N-terminal amino acid sequence of the coat protein, VRKGAATKAPQQPKPKA, to the obtained cDNA sequence. There was 100% identity between these after the removal of the first two M residues. The calculated molecular mass of the coat protein was 27·6 kDa. The size of the coat protein determined by SDS–PAGE was 32 kDa (data not shown; Rybicki, 1991). This is quite close to the 34 kDa protein produced in the *in vitro* translation of the viral RNA. The coat protein of SBMV is translated from a subgenomic RNA molecule, which is also present within the virus particles (Ghosh *et al.*, 1981). In contrast, it was not possible to detect any subgenomic RNA enclosed in CfMV particles in Northern blot hybridization (Fig. 6). When the *in vitro* translation mixture was subjected to Western blot analysis no coat protein signal was seen, while coat protein isolated from purified virus particles was easily detected (Tamm, 1994). For these reasons it is more likely that the 34 kDa translation product is a cleavage product of the polyprotein.

In Northern blots of total RNA of CfMV-infected leaves, two virus-specific bands of 4·1 kb and 1·2 kb were detected (Fig. 6) using a probe from the C-terminal part of ORF3 coding for the coat protein. A total RNA sample from uninfected barley did not give a CfMV-specific signal. Whether the 1·2 kb CfMV-specific RNA molecule is the subgenomic RNA for coat protein synthesis and the true initiation site for the synthesis of CfMV subgenomic RNA remains to be determined. As the translation initiation site of CfMV coat protein is 989 nucleotides upstream from the 3′ end of the genome, a subgenomic RNA of 1·2 kb seems highly possible.

```
CfMV      343   GGGKAWGDSDDEDTQETAIRPLNYQ
RYMV      384   SGQPSWADRFGDDSGEDVDIETSHP
SBMV-C    370   PKGKAWADMLDDDDLPLPPKMVNAA
SBMV-B    371   ESGKYWADDDDDSLPPPPKVVDGKM
PLRV      435   LTDKNWADDYDSDEDYGLEREAATN
BWYV      434   KTGKYWGDMEDDEDIFFESKEDLSG
PEMV      568   GPIGLWADDTEDDESAPRRSGNGLF
MBV       481   LSGFDWTDDAPMDFDELPVFESTMV
H-Ast2    836   SYDFDWDEDDAKFILPAPHRLTKAD
                   *||
CONS              Wad
```

Fig. 5.

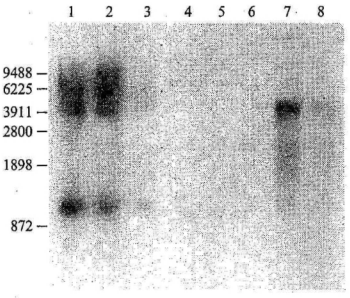


Fig. 6.

Fig. 5. Amino acid alignment of the WAD/WGD plus D/E-rich region upstream from putative or proven frameshift signals in CfMV and other viruses.

Fig. 6. Northern blot analysis of total RNA from CfMV-infected barley plants. Lanes 1–3, 8 µg of total RNA isolated from three different CfMV-infected barley plants; lanes 4–6, 8 µg of total RNA from three different healthy barley plants; lane 7, 100 ng of CfMV RNA isolated from purified virus particles; lane 8, 50 ng of CfMV RNA from virus particles. Positions of RNA markers are shown on the left.

```
CfMV     G--AUAAUAGUGCGAAGAAAGACACA----------CUGUUAUCGUUCC  37
RYMV     A--CAAUUGAAGCUAGGAAAGGAGCAUAUU-----GCGAAAG-CAUCCC  41
SBMV-C   CACAAAAUAUAAGAAGGAAAAGUGCUGAUUUUCCUACCUUUGUGUUUC-  48
SBMV-B   CACAAAAUAUAAGAAGGAAA---GCUGGAUUUCCUACCUUUGUGUUUCC  46
             |*|*|||| |*|****   |*|          * ||||  |* *

CfMV     -CCUCC----CGA--AUCAGAGGUUGAGAAGU--AGCUUAG-----AUG  72
RYMV     UCCUUC----CGACGAACAAUUGUAGCCACAC--UGCAUCGU--GUAUG  82
SBMV-C   ----------------------------------------------AUG  51
SBMV-B   AUUGUCGAAGCAUUGGUCAACGAUUACAAAACGGUGCAUUUUCUGCAUG  95
                                                     ***
```

Fig. 7. Multiple alignment of the 5′-end sequences of CfMV, SBMV-C, SBMV-B and RYMV. No clear consensus sequence was detected.

Sequence similarities to related viruses

CfMV proteins are remarkably similar to their counterparts in SBMV and RYMV, although there are substantial differences in genome organization. The intergenic region between ORF1 and ORF2a is present in CfMV, RYMV and SBMV-B genomes but is absent in SBMV-C. SBMV-C and RYMV have ORF3 nested in the ORF2 sequence but in another reading frame. The CfMV genome lacks the analogous ORF. ORF3 in the CfMV genome codes for the coat protein.

The lengths of CfMV untranslated termini are similar to those of RYMV: CfMV 5′, 69 nucleotides; 3′, 225 nucleotides; and RYMV 5′, 79 nucleotides; 3′, 245 nucleotides. In the SBMV-C genome the untranslated regions have been reported to be much shorter (48 and 138 nucleotides, respectively). The 5′-untranslated region of CfMV shares 62%, 42·5% and 58·3% identical nucleotides with the corresponding regions of SBMV-C, SBMV-B and RYMV, respectively. However, multiple alignment of the 5′-end sequences of CfMV, SBMV-C, SBMV-B and RYMV reveals only 14·1% identity (Fig. 7). We were not able to detect any clear consensus sequence from sobemovirus 5′-untranslated regions. Obvious sequence similarities were also not found in the 3′-untranslated regions of CfMV, SBMV-C, SBMV-B and RYMV. Secondary structure analysis of the CfMV 3′-untranslated region with RNAFOLD (Devereux *et al.*, 1984) did not reveal the presence of a tRNA-like structure, which can be found in the 3′-untranslated region of RYMV (Ngon A Yassi *et al.*, 1994).

ORF1 of CfMV is unrelated to the corresponding

ORFs of other sequenced sobemoviruses. The putative ORF2a-encoded proteins (VPg, protease and the amino acid sequence after the protease) were separately compared to the polyproteins of SBMV-C, SBMV-B and RYMV. The CfMV ORF2a product was divided into separate parts using the putative E/T and E/S cleavage sites before and after the protease consensus sequence. The potential VPg region was 30% identical between RYMV and CfMV, 23% identical between CfMV and SBMV-C and 19% identical between CfMV and SBMV-B while the corresponding values for the protease region were 37%, 38% and 28%, respectively. Several cellular proteases and proteases of picorna-, poty- and comoviruses have H, D and C or S residues in their catalytic sites (Gorbalenya *et al.*, 1988). These important amino acids can be found in the CfMV serine protease, located exactly as in the RYMV amino acid sequence: H is separated from D by 32 amino acids and the distance between D and S is 66 amino acids. Comparison of the ORF2a amino acid sequence after the WAD/WGD plus D/E-rich region is complicated because in this region CfMV does not have the replicase as in SBMV-C and RYMV, but instead has the end of ORF2a which overlaps with ORF2b. This kind of genome organization resembles that of luteovirus subgroup II. The 5′ end of CfMV ORF2b contains signals for ribosomal −1 frameshifting (Mäkinen *et al.*, 1995). Similar frameshifting consensus signals can be found in the beginning of ORF3 of SBMV-C and RYMV. ORF3 is nested in the ORF2 sequence of these two viruses but in another reading frame and its function is unknown. Amino acid comparisons between the putative replicases of CfMV and the other sequenced members of sobemo- and luteovirus subgroup II viruses have been reported by us earlier (Mäkinen *et al.*, 1995).

Although SBMV-B had an overall similarity to SBMV-C and RYMV, it lacked the analogue of ORF3 of these two viruses. Its genome organization resembles that of CfMV. However, SBMV-B polyprotein was reported to be translated from a single ORF without ribosomal frameshifting. Interestingly, the SBMV-B genome also contains a putative −1 ribosomal frameshift signal: slippery sequence UUUAAAC starting at position 1728 followed by a strong stem–loop structure after 7 additional nucleotides.

Comparison of sobemovirus coat proteins revealed 23% identity between CfMV and SBMV-C, 23% identity between CfMV and SBMV-B and 35% identity between CfMV and RYMV. SBMV-C and RYMV have 29% identical amino acids. CfMV has, in common with the other sequenced sobemoviruses, a highly basic amino acid sequence at the N terminus of the coat protein. This region in CfMV resembles the bipartite signal proposed for nuclear targeting (Dingwall & Laskey, 1991). The significance of this finding is presently not known, but CfMV particles have been reported to be present in the cell nucleus (Chamberlain & Catherall, 1976).

Conclusions

CfMV is a member of the sobemovirus group based on the sequence similarities between the polyproteins and coat proteins of CfMV, SBMV and RYMV. CfMV is more closely related to RYMV than SBMV-C and SBMV-B. However, the genome organization of CfMV is significantly different from other characterized sobemoviruses. It is more closely related to SBMV-B, as both lack the ORF3 of SBMV-C and RYMV. The polyprotein-coding region of CfMV is more similar to that of subgroup II luteoviruses, PLRV and BWYV, than to the reported genomic organizations of sobemoviruses. The evolution of subgroup II luteoviruses has been proposed to have occurred by a recombination event between a sobemovirus and a subgroup I luteovirus (Miller *et al.*, 1995). The polyprotein in this recombination evolved from the sobemovirus counterpart. The finding that the replicase of the CfMV is translated, in a similar way to the subgroup II luteoviruses, as part of the polyprotein by −1 ribosomal frameshifting supports this model.

We acknowledge Anu Aaspõllu for participating in some of the experiments, Dr Nisse Kalkkinen for sequencing the N terminus of the coat protein and M.-L. Bouscaren for the critical reading of the manuscript. This research has been supported by the Academy of Finland, Finnish Ministry of Agriculture and Forestry, and by the Estonian Science Foundation. V.N. was supported by the Norwegian Research Council and a grant from NorFa, E.T. from CIMO and T.T. from FEBS.

References

AGRANOVSKY, A. A., KOONIN, E. V., BOYKO, V. P., MAISS, E., FRÖTSCHL, R., LUNINA, N. A. & ATABEKOV, J. G. (1994). Beet yellows closterovirus: complete genome structure and identification of a leader papain-like thiol protease. *Virology* **198**, 311–324.

ALLISON, R., JOHNSTON, R. E. & DOUGHERTY, W. G. (1986). The nucleotide sequence of the coding region of tobacco etch virus genomic RNA: evidence for the synthesis of a single polyprotein. *Virology* **154**, 9–20.

CHAMBERLAIN, J. A. & CATHERALL, P. L. (1976). Electron microscopy of some grasses and cereals infected with cocksfoot mottle, phleum mottle and cocksfoot mild mosaic viruses. *Journal of General Virology* **30**, 41–50.

DEMLER, S. A. & DE ZOETEN, G. A. (1991). The nucleotide sequence and luteovirus-like nature of RNA 1 of an aphid non-transmissible strain of pea enation mosaic virus. *Journal of General Virology* **72**, 1819–1834.

DEVEREUX, J., HAEBERLI, P. & SMITHIES, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Research* **12**, 387–395.

DINGWALL, C. & LASKEY, R. A. (1991). Nuclear targeting sequences – a consensus? *Trends in Biochemical Sciences* **16**, 478–481.

DOUGHERTY, W. G. & HIEBERT, E. (1980). Translation of potyvirus RNA in a rabbit reticulocyte lysate: cell-free translation strategy and a genetic map of the potyviral genome. *Virology* **104**, 183–194.

FEINBERG, A. P. & VOGELSTEIN, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Analytical Biochemistry* **132**, 6–13.

GHOSH, A., RUTGERS, T., KE-QIANG, M. & KAESBERG, P. (1981). Characterization of the coat protein mRNA of southern bean mosaic virus and its relationship to the genomic RNA. *Journal of Virology* **39**, 87–92.

GORBALENYA, A. E. & KOONIN, E. V. (1993). Comparative analysis of amino-acid sequences of key enzymes of replication and expression of positive-strand RNA viruses: validity of approach and functional and evolutionary implications. *Soviet Scientific Reviews, Section D, Biology Reviews* **11**, 1–84.

GORBALENYA, A. E., KOONIN, E. V., BLINOV, V. M. & DONCHENKO, A. P. (1988). Sobemovirus genome appears to encode a serine protease related to cysteine proteases of picornaviruses. *FEBS Letters* **236**, 287–290.

HIGGINS, D. G., BLEASBY, A. J. & FUCHS, R. (1992). Clustal V: improved software for multiple sequence alignment. *Computer Applications in the Biosciences* **8**, 189–191.

JACKS, T., MADHANI, H. D., MASIARZ, F. R. & VARMUS, H. E. (1988). Signals for ribosomal frameshifting in the Rous sarcoma virus *gag–pol* region. *Cell* **55**, 447–458.

JIANG, B., MONROE, S. S., KOONIN, E. V., STINE, S. E. & GLASS, R. I. (1993). RNA sequence of astrovirus: distinctive genomic organization and putative retrovirus-like ribosomal frameshifting signal that directs viral replicase synthesis. *Proceedings of the National Academy of Sciences, USA* **90**, 10539–10543.

KALKKINEN, N. & TILGMANN, C. (1988). A gas/pulsed-liquid sequencer constructed from a Beckman 890D instrument by using Applied Biosystems delivery and cartridge blocks. *Journal of Protein Chemistry* **7**, 242–243.

KAMER, G. & ARGOS, P. (1984). Primary structural comparison of RNA-dependent polymerases from plant, animal and bacterial viruses. *Nucleic Acids Research* **12**, 7269–7282.

KOZAK, M., (1989). A scanning model for translation: an update. *Journal of Cell Biology* **108**, 229–241.

LAEMMLI, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685.

LOGEMANN, J., SCHELL, J. & WILLMITZER, L. (1987). Improved method for the isolation of RNA from plant tissues. *Analytical Biochemistry* **163**, 16–20.

LÖW, R. & RAUSCH, T. (1994). Sensitive, nonradioactive Northern blots using alkaline transfer of total RNA and PCR-amplified biotinylated probes. *BioTechniques* **17**, 1026–1030.

LÜTCKE, H. A., CHOW, K., MICKEL, F. S., MOSS, K. A., KERN, H. F. & SCHEELE, G. A. (1987). Selection of AUG codons differs in plants and animals. *EMBO Journal* **6**, 43–48.

MÄKINEN, K., NÆSS, V., TAMM, T., TRUVE, E., AASPÖLLU, A. & SAARMA, M. (1995). The putative replicase of the cocksfoot mottle sobemovirus is translated as a part of the polyprotein by −1 ribosomal frameshift. *Virology* **207**, 566–571.

MAYO, M. A., ROBINSON, D. J., JOLLY, C. A. & HYMAN, L. (1989). Nucleotide sequence of potato leafroll luteovirus RNA. *Journal of General Virology* **70**, 1037–1051.

MILLER, W. A., DINESH-KUMAR, S. P. & PAUL, C. P. (1995). Luteovirus gene expression. *Critical Reviews in Plant Sciences* **14**, 179–211.

MURPHY, J. F., RYCHLIK, W., RHOADS, R., HUNT, A. G. & SHAW, J. G. (1991). A tyrosine residue in the small nuclear inclusion protein of tobacco vein mottling virus links the VPg to the viral RNA. *Journal of Virology* **65**, 511–513.

NGON A YASSI, M., RITZENTHALER, C., BRUGIDOU, C., FAUQUET, C. & BEACHY, R. N. (1994). Nucleotide sequence and genome characterization of rice yellow mottle virus RNA. *Journal of General Virology* **75**, 249–257.

OTHMAN, Y. & HULL, R. (1995). Nucleotide sequence of the bean strain of southern bean mosaic virus. *Virology* **206**, 287–297.

PUURAND, Ü., MÄKINEN, K., BAUMANN, M. & SAARMA, M. (1992). Nucleotide sequence of the 3′-terminal region of potato virus A RNA. *Virus Research* **23**, 99–105.

PUURAND, Ü., MÄKINEN, K., PAULIN, L. & SAARMA, M. (1994). The nucleotide sequence of potato virus A genomic RNA and its sequence similarities with other potyviruses. *Journal of General Virology* **75**, 457–461.

REVILL, P. A., DAVIDSON, A. D. & WRIGHT, P. J. (1994). The nucleotide sequence and genome organization of mushroom bacilliform virus: a single-stranded RNA virus of *Agaricus bisporus* (Lange) Imbach. *Virology* **202**, 904–911.

ROBAGLIA, C., DURAND-TARDIF, M., TRONCHET, M., BOUDAZIN, G., ASTIER-MANIFACIER, S. & CASSE-DELBART, F. (1989). Nucleotide sequence of potato virus Y (N strain) genomic RNA. *Journal of General Virology* **70**, 935–947.

RYBICKI, E. P. (1991). The southern bean mosaic virus group. In *Classification and Nomenclature of Viruses. Fifth Report of the International Committee on Taxonomy of Viruses.* Edited by R. I. B. Francki, C. M. Fauquet, D. L. Knudson & F. Brown. *Archives of Virology*, Supplement 2, 327–329.

SAMBROOK, J., FRITSCH, E. F. & MANIATIS, T. (1989). *Molecular Cloning: A Laboratory Manual.* New York: Cold Spring Harbor Laboratory.

STAUNTON, D. E., MERLUZZI, V. T., ROTHLEIN, R., BARTON, R., MARLIN, S. D. & SPRINGER, T. A. (1989). A cell adhesion molecule, ICAM-1, is the major surface receptor for rhinoviruses. *Cell* **56**, 849–853.

TAMM, T. (1994). *Characterization of cocksfoot mottle virus.* MSc thesis, Tartu University.

VEIDT, I., LOT, H., LEISER, M., SHEIDECKER, D., GUILLEY, H., RICHARDS, K. & JONARD, G. (1988). Nucleotide sequence of beet western yellows virus. *Nucleic Acids Research* **16**, 9917–9932.

WU, S., RINEHART, C. A. & KAESBERG, P. (1987). Sequence and organization of southern bean mosaic virus genomic RNA. *Virology* **161**, 73–80.

(*Received 29 March 1995; Accepted 17 July 1995*)

II

Virology 1995, **207:** 566–571.
Reprinted by courtesy of Academic Press, Inc.

VIROLOGY 207, 566–571 (1995)

SHORT COMMUNICATION

The Putative Replicase of the Cocksfoot Mottle Sobemovirus Is Translated as a Part of the Polyprotein by −1 Ribosomal Frameshift

KRISTIINA MÄKINEN,*,[1] VIGFRID NÆSS,†,[2] TIINA TAMM,‡ ERKKI TRUVE,‡
ANU AASPÕLLU,‡ and MART SAARMA*

*Institute of Biotechnology, P.O. Box 45, FIN-00014 University of Helsinki, Finland; †Norwegian Plant Protection Institute, Fellesbygget, N-1432 Ås, Norway; and ‡Institute of Chemical Physics and Biophysics, Estonian Academy of Sciences, Akadeemia tee 23, EE0026 Tallinn, Estonia

Received October 12, 1994; accepted December 16, 1994

The polyprotein of cocksfoot mottle sobemovirus (CfMV) is encoded by two overlapping open reading frames (ORF). The ORF 2a codes for the putative VPg and serine protease and the ORF 2b codes for the putative replicase. The consensus signals for a −1 ribosomal frameshifting event are found at the very beginning of the overlapping region of these ORFs. The shifty heptanucleotide in CfMV is UUUAAAC, and the secondary structure after the shifty sequence is predicted to be a stem–loop. *In vitro* translation of the CfMV RNA in wheat germ extract produced proteins of several sizes, including one of 100 kDa. According to the nucleotide sequence data, no single ORF is capable of directing the synthesis of a 100-kDa protein. A chimeric β-glucuronidase–CfMV cDNA containing the entire ORF 2a and 2b overlap region including frameshift signals was constructed. A trans-frame protein of 108 kDa was produced from this construct with an efficiency of 26–29% by *in vitro* translation in wheat germ extract. CfMV is the first sobemovirus in which the putative replicase is reported to be produced as a part of a polyprotein by a −1 frameshift event. The replicases of the sobemoviruses are related to the luteovirus subgroup II replicases, which are known to be produced by −1 ribosomal frameshift. The reported amino acid sequences of the putative replicases of sobemo- and subgroup II luteoviruses were compared to that of the putative replicase of CfMV. This comparison revealed more extensive homology between these groups than previously reported.

Cocksfoot mottle virus (CfMV) is a member of the sobemovirus genus (*1*). The virus particles are isometric and 30 nm in diameter. The CfMV genome is a monopartite, single-stranded RNA of about 4000 nucleotides. CfMV infects only a few species of the family Gramineae (*2*). Its main host is cocksfoot (*Dactylis glomerata*), an important herbage grass species in the Nordic countries, although wheat, oat, and barley are experimental hosts. CfMV causes yellow streaking and severe mottling, and infected plants die without flowering (*3*). Rapid spread of the virus has resulted in substantial yield losses. Chrysomelidae beetles are the natural vectors for CfMV but, in cocksfoot monocultures, the virus is mainly transmitted mechanically. The genomic organization of two members of the sobemovirus group, southern bean mosaic virus (SBMV) (*4*) and rice yellow mottle virus (RYMV) (*5*), has been reported. The RNA polymerase of SBMV is closely related to the polymerases of two subgroup II luteoviruses, potato leafroll virus (PLRV) and beet western yellows virus (BWYV) (*6*). Both sobemo- and luteoviruses have a translational strategy based on the production of a large polyprotein and the production of the coat protein from a subgenomic RNA (*7*). We have now cloned and sequenced the 5′ untranslated region and the coding region for the polyprotein of CfMV. The polyprotein of CfMV is encoded by two overlapping open reading frames (ORFs), ORF 2a and ORF 2b. The polyproteins of SBMV and RYMV are encoded by one ORF. The deduced ORF 2a protein product is related to the putative VPg and serine protease of SBMV and RYMV (Mäkinen *et al.*, unpublished observation). ORF 2b codes for the putative replicase. In this paper we show that the putative replicase is translated *in vitro* as a part of the polyprotein by a eukaryotic −1 ribosomal frameshift mechanism. We also compare the amino acid alignments of the replicase part to those of the other sequenced sobemoviruses and to the closely related luteoviruses PLRV (*8*, *9*) and BWYV (*10*).

Primary isolate 9 from naturally infected cocksfoot plants (Gran, Hadeland, Norway) was used as a source of CfMV particles. Oat plants cv. Blenda were sap-inoculated and grown for 4–5 weeks. CfMV particles were purified (*11*) and viral RNA was isolated as described

The nucleotide sequence data reported in this paper have been deposited with the EMBL Database under Accession No. Z36903.

[1] To whom correspondence and reprint requests should be addressed. Fax: +358-0-4346046; e-mail: Makinen@operoni.helsinki.fi.

[2] The first and the second author have contributed equally to the work.

0042-6822/95 $6.00

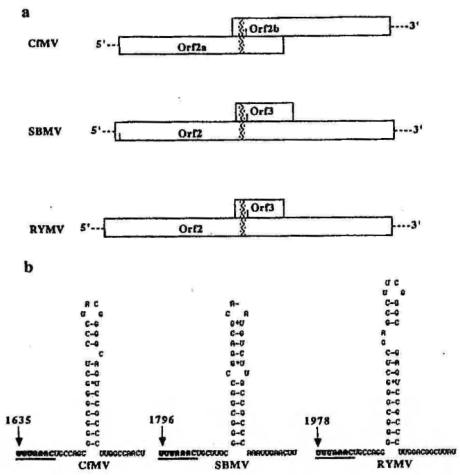


FIG. 1. The frameshift consensus signals in CfMV, SBMV, and RYMV and their locations in the polyprotein region. (a) The ORF maps of the polyprotein regions of CfMV, SBMV, and RYMV. The sites of the frameshift signals are marked by wavy lines on the map and the first ATG in each ORF is marked by a vertical line. (b) The shifty sites of CfMV, SBMV, and RYMV along with the predicted downstream secondary structures. The number pointing to the first nucleotide in a slippery site indicates the actual site in the genome. The shifty heptanucleotides are written in bold and underlined.

previously (*12*). The cDNA synthesis, using 3 μg of CfMV RNA, was randomly primed (Promega Riboclone kit). *Eco*RI adapters were ligated to the blunt-ended cDNA and the products were cloned into the *Eco*RI site of the vector λgt11 (Promega). The λ library was screened initially using CfMV cDNA labeled with digoxigenin (Boehringer-Mannheim), and later using probes closely approximating the missing nucleotide sequences. Some of the clones were obtained by PCR amplification of the cDNA using Dynazyme DNA polymerase (Finnzymes). With these cloning strategies, a total of nine different clones covering the polyprotein coding region were identified and the inserts from λ vectors then subcloned into pBluescript SK$^+$ (Stratagene).

The clones were sequenced from both strands by the dideoxy chain-termination method (*13*) using the Sequenase Version 2.0 sequencing kit (US Biochemicals). Some minor parts of the sequence were obtained using the ALF DNA sequencer (Pharmacia LKB). The PCR fragments were directly sequenced with the CircumVent DNA sequencing kit (New England Biolabs) and, after cloning, with the Sequenase Version 2.0 kit. The polyprotein coding region of CfMV RNA is 2831 nucleotides, revealing two overlapping reading frames called ORF 2a and ORF 2b (Fig. 1). The overlapping region is 541 nucleotides. ORF 2a has 1710 nucleotides from position 424 to 2133 and ORF 2b has 1650 nucleotides from position 1603 to 3253. ORF 2a encodes a protein containing 568 amino acids with a calculated M_r of 60.8 kDa. ORF 2b encodes a protein containing 504 amino acids and a calculated M_r of 56.3 kDa.

In vitro translations of the CfMV RNA were made in wheat germ cell-free extracts (Promega) according to the manufacturer's instructions. One microgram of CfMV RNA was added per reaction after the RNA was denatured at 65° for 10 min. The reaction was carried out at 25° for 60 min. The [^{35}S]Met-labeled (Amersham) translation products were analyzed on 12% SDS–polyacrylamide gels. The *in vitro* translation of the genomic RNA revealed four major products with molecular weights of 100, 71, 34, and 16 kDa, respectively (Fig. 2B). In contrast to the SBMV (*4*) and RYMV (*5*) sequences, the CfMV sequence data (not shown) indicated that no individual ORF is capable of directing the synthesis of a 100-kDa protein. However, it can be expressed as a trans-frame protein by the −1 ribosomal frameshifting event within the overlapping regions between ORF 2a and ORF 2b. This trans-frame protein contains 942 amino acids and has a calculated M_r of 103.4 kDa. In the very beginning of the overlap between ORF 2b and ORF 2a we found the two characteristic structural elements for the −1 frameshift event (*14*): the slippery sequence UUUAAAC and a stem–loop structure just downstream from that (Fig. 1). Similar frameshift signals were also found in the beginning of ORF 3 in SBMV and RYMV. The programs used for folding and drawing the frameshift region were respectively MFOLD (*15–17*) and Loopview (Indiana University). The possibility of pseudoknot formation was also studied, but it

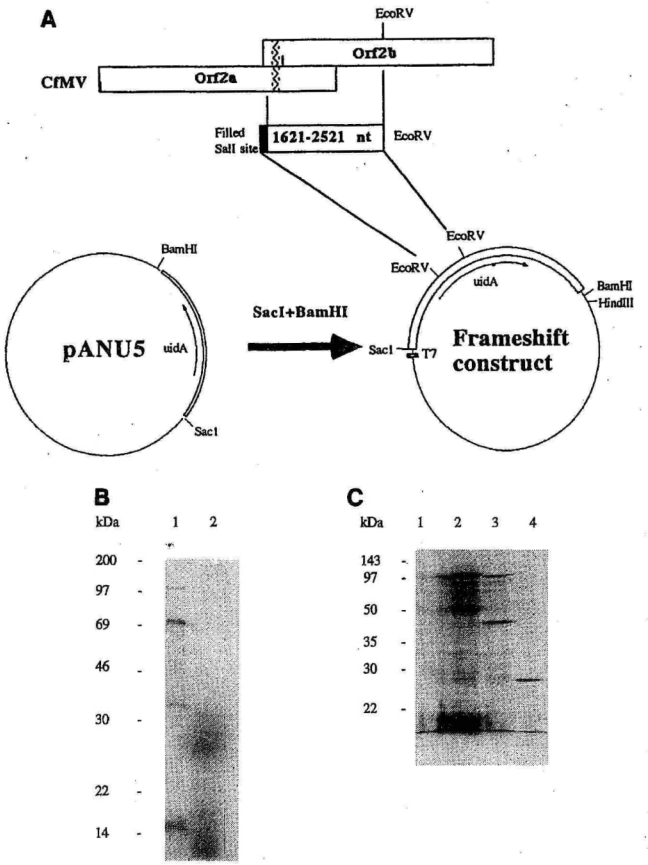


FIG. 2. *In vitro* translations of CfMV RNA and the frameshift construct. (A) The cloning map of the frameshift construct. The GUS gene is labeled *uid*A. The *Eco*RV fragment of the *uid*A gene was replaced with a cDNA fragment corresponding to nucleotides 1621–2521 in CfMV RNA. This fragment contains the frameshift consensus signals from the beginning of ORF 2b, the entire region of overlap between ORF 2a and ORF 2b, and an extra region farther downstream. The part of the *uid*A gene before the *Eco*RV site is called GUS-N and the part after that is called GUS-C in the text. The *Sal*I site at the 5′-terminus of the CfMV cDNA comes from the adaptor of the clone used. (B) The *in vitro* translation of the CfMV: Lane 1 is CfMV RNA translation. The sizes of the proteins produced are 100, 71, 34, and 16 kDa. Lane 2 is a control translation with no added RNA. The molecular weight markers are given on the left. (C) The *in vitro* translation of the GUS-CfMV chimeric frameshift construct: Lane 1 is the control with no DNA added. Lane 2 is the first GUS-CfMV control construct, which should give a single product of 102 kDa. Lane 3 is the GUS-CfMV frameshift construct, which should give a GUS-N-ORF2a protein of 39.5 kDa and the frameshift product GUS-N-ORF2a-ORF2b-GUS-C of 102 kDa. Lane 4 is the second GUS-CfMV control construct, which should give a single product of 21.5 kDa. The apparent sizes of the proteins produced are slightly larger than their calculated sizes: lane 2, 108 kDa; lane 3, 45 kDa and 108 kDa; lane 4, 26 kDa. The molecular weight markers are given on the left.

is not possible to form a pseudoknot 5–9 bases downstream from the shifty site, where it in most cases exists (*18*).

Fusion constructs of the CfMV sequence with the *Escherichia coli* β-glucuronidase gene (GUS, *uid*A in Fig. 2A) for *in vitro* translation experiments on ribosomal frameshifting were constructed as follows: The GUS gene was isolated after *Sac*I/*Bam*HI digestion, which cut at the sites flanking the GUS coding region in pANU5 (generously provided by Kristian Aspegren, Institute of Biotechnology, Helsinki, Finland), and ligated to the corresponding sites in the pGEM3Z(−) (Promega) under the

control of the T7 promoter. The pGEM3Z-GUS construct was digested with *Eco*RV, which removes a 230-bp-long internal GUS fragment. A *SalI*/*Eco*RV cDNA fragment representing nucleotides 1621–2521 of CfMV RNA was ligated to the pGEM3Z-GUS *Eco*RV site (Fig. 2A). This region was selected because it contains the entire region of ORF 2a and 2b overlap, including the putative shifty heptanucleotide and a sequence that can fold into the proposed secondary structure. In this construct, the fusion of the N-proximal half of the GUS gene (GUS-N) to the residual ORF 2a part would yield a chimeric GUS-N-ORF2a protein of 39.5 kDa, while a frameshift into ORF 2b would allow translation to proceed to the GUS carboxy-terminus (GUS-C) and produce a chimeric trans-frame protein GUS-N-ORF2a-ORF2b-GUS-C of 102 kDa. Two chimeric GUS-CfMV control plasmids were also constructed: The first control plasmid is similar to the frameshift construct except that it has a 1-nucleotide deletion before the frameshift signal. This mutation fuses ORF 2a and 2b into one reading frame. The only expected translation product is of the same size as the frameshift product (102 kDa). The second control plasmid is similar to the frameshift construct except that it has a 1-nucleotide insertion before the frameshift signal. This mutation causes the translation to terminate in a stop codon, UAA, formed by the frameshift signal. The expected size of the translation product from the second control plasmid should be 21.5 kDa. Chimeric GUS-CfMV plasmids, linearized with *Hin*dIII, were transcribed and translated in the TNT T7 Coupled Wheat Germ Extract System (Promega). Four different translations were performed: a control with no added DNA, the first control plasmid, the −1 frameshift construct, and the second control construct. The results of the *in vitro* translation of genomic RNA, the −1 frameshift chimeric GUS-CfMV construct, and the controls are shown in Fig. 2C. For quantification of the radioactivity in the polypeptide bands two alternative methods were used: The gel was cut in pieces, swollen in water, and solubilized in NCS (Amersham), and the radioactivity scintillation was counted in OptiPhase "HighSafe" 3 (Wallac LKB). The intensity of the bands in autoradiography film was measured with an Ultroscan XL enhanced laser densitometer. The measured radioactivity of the trans-frame protein GUS-N-CfMV-ORF2a-ORF2b-GUS-C was corrected for background radioactivity and varying methionine content, and frameshift efficiencies were calculated in relation to the 100% control GUS-N-ORF2a. The efficiency of the *in vitro* frameshifting achieved by CfMV cDNA was determined from three different translations by two different methods to be 26–29%.

The programs used for sequence analysis were DNA Strider (Commissariate a l'Energie Atomique, France), DNAid$^+$ (Ecole Polytechnique, France), and Clustal V (EMBL). Amino acid sequence comparisons were made between the CfMV ORF 2b-encoded protein and the protein products of SBMV and RYMV and the two luteoviruses PLRV and BWYV. The ORF 2b-encoded polypeptide contains conserved domains for RNA-dependent RNA polymerases. The N-terminal part of CfMV replicase was found to be similar to the N-terminal part of ORF 3 in SBMV and RYMV. In the SBMV and RYMV sequences, ORF 3 is nested within the polyprotein sequence but in another reading frame. The percentage of similar amino acid sequences of SBMV is 63.7% and of RYMV is 62.0% in this area. The function of ORF 3 in SBMV and RYMV is unknown. The C-terminal part of the CfMV replicase is related to ORF 2 in SBMV and in RYMV. SBMV and RYMV are respectively 70.7 and 74.0% similar in this region. The replicase of SBMV and RYMV is translated from ORF 2. Figure 3 depicts the amino acid sequence alignments of CfMV, SBMV (*4*), RYMV (*5*), PLRV (*8*), and BWYV (*10*).

CfMV is the first sobemovirus reported to translate the replicase by using a −1 frameshift event. In SBMV and RYMV, the replicase is produced as a part of the polyprotein from a single ORF. However, it is possible to find similar consensus sequences for frameshifting in the beginning of the SBMV and RYMV ORF 3 (Fig. 1). That region of the ORF 3 of SBMV and RYMV also contains a high level of sequence similarity to the replicase of CfMV. The frameshifting shifty sequence for CfMV is UUUAAAC. The same shifty sequence is used by infectious bronchitis coronavirus (IBV) (*19*) and human T-cell leukemia virus type I (*20*). The secondary structure after the shifty heptanucleotide sequence in CfMV appears to be a stem–loop. Other viruses, such as human immunodeficiency virus (*21*) and red clover necrotic mosaic virus (*22*), require a simple stem–loop secondary structure, whereas most of the others require a pseudoknot 5 to 9 bases downstream of their shifty sites for efficient frameshifting (*23–25*). The downstream secondary structure is probably needed to slow down the movement of ribosomes along the mRNA or to force the ribosomes to slip (*26, 27*).

The polymerases of sobemoviruses are closely related to subgroup II luteoviral polymerases. The subgroup II luteoviruses and sobemoviruses have been classified as the Sobemovirales (*28*). In luteoviruses, the polymerases are expressed by using a −1 frameshift in the translational reading frame (*25, 29–32*). The overlapping region between the two ORFs coding for the polyprotein is from 475 to 628 nucleotides long in subgroup II luteoviruses. The analogous region in CfMV is 541 nucleotides long, which fits well with the genomic organization of subgroup II luteoviruses. The shifty heptanucleotide for BWYV is GGGAAAC and for PLRV is UUUAAAU. The subgroup II viruses have a weak pseudoknot secondary structure following the shifty sequence (*25, 31*). Another secondary structure for PLRV is a stem–loop (*32*). The nucleotide sequence and the genome organization of mushroom bacilliform virus (MBV) were recently published (*33*). This

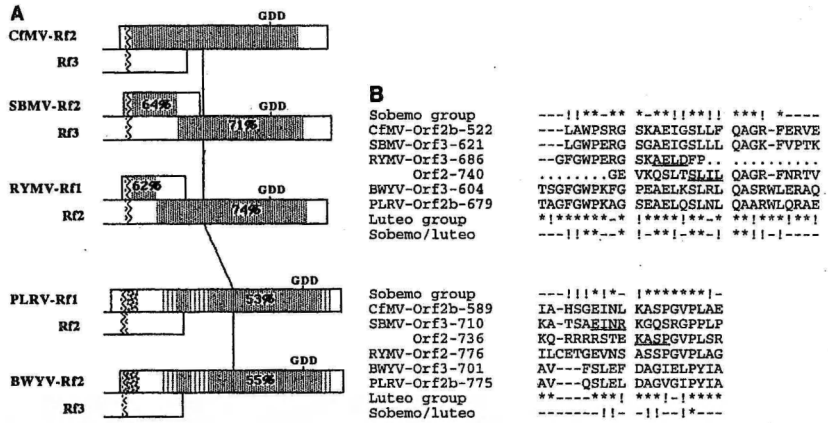


FIG. 3. The amino acid sequence comparison of the putative replicases of SBMV (*4*), RYMV (*5*), PLRV (*8*), BWYV (*10*), and CfMV. (A) A schematic presentation of the amino acid comparison. The solid line from CfMV to BWYV indicates the previously reported extent of the replicase relatedness between sobemo- and luteoviruses (*6*). The position of the main replicase motif GDD is marked. ⊠, frameshift signal; ▩, nucleic acid binding motif; ▥, replicase, <50% homology; ▨, replicase, >50% homology. Rf is an abbreviation for reading frame. (B) The amino acid sequences in the region where the replicase sequences of CfMV switch from ORF 3 to ORF 2 in RYMV and SBMV genomes. The amino acids at the switch point are underlined. The identical amino acids are marked by a star and the similar amino acids are marked by an exclamation mark in separate rows above (sobemo group comparison) and below (luteo group and sobemo/luteo group comparisons) the amino acid sequences.

virus has a replicase which is related to luteo- and sobemoviruses. The polymerase gene of MBV is also predicted to be expressed by a −1 frameshift event. The frameshift signals in MBV genome can be found at the beginning of ORF 3, which codes for the putative replicase. The putative shifty sequence is GUUUUUC, which is followed by a potential stem–loop structure. In CfMV, the frameshifting is highly efficient, with a frequency of about 26–29% *in vitro*. This correlates well with the frameshifting rates reported for other systems, such as 25–30% for IBV *in vitro* (*19*), but is considerably higher than the 1% reported for PLRV (*32*) and BWYV (*25*). The very high efficiency of replication in CfMV-infected cocksfoot could be due to the relatively high frameshifting efficiency in producing the replicase.

A gene exchange between BWYV, BYDV, and SBMV has been proposed (*10*). Recently a recombination model for the origin of luteovirus subgroup II was put forward (*34*). In addition to the polymerase similarity, there is substantial similarity between the first ORFs of sobemo- and subgroup II luteoviruses. The model would require one recombination event in a mixed infection of a subgroup I luteovirus and a sobemovirus, where polymerase switches from the sobemoviral genomic RNA to luteovirus subgroup I subgenomic RNA (*34*). The finding that CfMV polymerase is expressed by −1 ribosomal frameshifting supports this model.

ACKNOWLEDGMENTS

We thank Dr. Thor Mynthe for kindly providing the CfMV-infected leaves and Kristian Aspegren for providing the GUS-gene cDNA. We are indebted to Dr. Alan Schulman for critically reading the manuscript. This work was supported by the Finnish Academy, the Ministry of Agriculture and Forestry, and the Estonian Science Foundation (Grant 609). T.T. was supported by FEBS Summer Fellowship.

REFERENCES

1. Rybicki, E. P., *Arch. Virol. (Suppl.)* **2**, 327–329 (1991).
2. Serjeant, E. P., *Ann. Appl. Biol.* **59**, 31–38 (1967).
3. Serjeant, E. P., *Plant Pathol.* **13**, 23–24 (1964).
4. Wu, S., Rinehart, C. A., and Kaesberg, P., *Virology* **161**, 73–80 (1987).
5. Ngon A Yassi, M., Ritzenthaler, C., Brugidou, C., Fauquet, C., and Beachy, R. N., *J. Gen. Virol.* **75**, 249–257 (1994).
6. Koonin, E. V., *J. Gen. Virol.* **72**, 2197–2206 (1991).
7. Matthews, R. E. F., "Plant Virology," 3rd ed. Academic Press, San Diego (1991).
8. Mayo, M. A., Robinsson, D. J., Jolly, C. A., and Hyman, L. J., *J. Gen. Virol.* **70**, 1037–1051 (1989).
9. van der Wilk, F., Huisman, M. J., Cornelissen, B. J. C., Huttinga, H., and Goldbach, R., *FEBS Lett.* 24551–24556 (1989).
10. Veidt, I., Lot, H., Leiser, M., Scheidecker, D., Guilley, H., Richards, K., and Jonard, G., *Nucleic Acids Res.* **16**, 9917–9932 (1988).
11. Spaer, D., "Pflanzliche Virologie." Akademie-Verlag, Berlin (1980).
12. Puurand, Ü., Mäkinen, K., Baumann, M., and Saarma, M., *Virus Res.* **23**, 99–105 (1992).
13. Sanger, F., Nicklen, S., and Coulsen, A. R., *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467 (1977).
14. Jacks, T., Madhani, H. D., Masiarz, F. R., and Varmus, H. E., *Cell* **55**, 447–458 (1988).
15. Zuker, M., *Science* **244**, 48–52 (1989).
16. Jaeger, J. A., Turner, D. H., and Zuker, M., *Proc. Natl. Acad. Sci. USA* **86**, 7706–7710 (1989).
17. Jaeger, J. A., Turner, D. H., and Zuker, M., *Methods Enzymol.* **183**, 281–306 (1989).

18. ten Dam, E. B., Pleij, C. W. A., and Bosch, L., *Virus Genes* **4**, 121–136 (1990).
19. Brierley, I., Boursnell, M. E. G., Binns, M. M., Bilimoria, B., Blok, V. C., Brown, T. D. K., and Inglis, S. C., *EMBO J.* **6**, 3779–3785 (1987).
20. Inoue, J.-I., Watanabe, T., Sato, M., Oda, A., Toyoshima, K., Yoshida, M., and Seiki, M., *Virology* **150**, 187–195 (1986).
21. Parkin, N. T., Chamorro, M., and Varmus, H. E., *J. Virol.* **66**, 5147–5151 (1992).
22. Kim, K. H., and Lommel, S. A., *Virology* **200**, 574–582 (1993).
23. Brierley, I., Rolley, N. J., Jenner, A. J., and Inglis, S. C., *J. Mol. Biol.* **220**, 889–902 (1991).
24. Chamorro, M., Parkin, N., and Varmus, H. E., *Proc. Natl. Acad. Sci. USA* **89**, 713–717 (1992).
25. Garcia, A., van Duin, J., and Pleij, C. W. A., *Nucleic Acids Res.* **21**, 401–406 (1993).
26. Tu, C., Tzeng, T. W., and Bruenn, J. A., *Proc. Natl. Acad. Sci. USA* **89**, 8636–8640 (1992).
27. Somogyi, P., Jenner, A. J., Brierley, I., and Inglis, S. C., *Mol. Cell. Biol.* **13**, 6931–6940 (1993).
28. Koonin, E. V., and Dolja, V. V., *Crit. Rev. Biochem. Mol. Biol.* **28**, 375–430 (1993).
29. Brault, V., and Miller, W. A., *Proc. Natl. Acad. Sci. USA* **89**, 2262–2266 (1992).
30. Di, R., Dinesh-Kumar, S. P., and Miller, W. A., *Mol. Plant Microbe Interact.* **6**, 444–452 (1993).
31. Kujawa, A., Drugeon, G., Hulanicka, D., and Henni, A.-L., *Nucleic Acids Res.* **21**, 2165–2171 (1993).
32. Prüfer, D., Tacke, E., Schmitz, J., Kull, B., Kaufmann, A., and Rhode, W., *EMBO J.* **11**, 1111–1117 (1992).
33. Revill, P. A., Davidson, D., and Wright, P. J., *Virology* **202**, 904–911 (1994).
34. Miller, W. A., Dinesh-Kumar, S. P., and Paul, C. P., *Crit. Rev. Plant Sci.* in press, (1994).

III

Archives of Virology 1999, **144:** 1557–1567.
Reprinted by courtesy of Springer-Verlag

Arch Virol (1999) 144: 1557–1567

Archives of Virology

Identification of genes encoding for the cocksfoot mottle virus proteins

T. Tamm[1], **K. Mäkinen**[2], and **E. Truve**[1,3]

[1]Institute of Chemical Physics and Biophysics, Tallinn, Estonia
[2]Institute of Biotechnology, Viikki Biocentre, University of Helsinki, Finland
[3]Gene Technology Center, Tallinn Technical University, Tallinn, Estonia

Accepted March 19, 1999

Summary. Cocksfoot mottle sobemovirus (CfMV) has a monopartite single-stranded positive-sense RNA genome. In wheat germ extract, in vitro translation of CfMV virion RNA resulted in the production of four major proteins of 100, 71, 34 and 12 kDa. In this paper we show the individual genes from which these polypeptides are synthesised. Polyclonal antisera against the proteins encoded by all open reading frames (ORFs) of CfMV were produced. Antibodies were used to immunoprecipitate the in vitro translation products of CfMV genomic RNA as well as the coupled in vitro transcription/translation products of individual viral genes. We demonstrate that the 12, 71, and 100 kDa CfMV proteins are synthesised from the bicistronic genomic RNA of the virus. The CfMV 12 kDa protein is produced from ORF1, the 71 kDa protein from ORF2a (and does not result from the proteolytic cleavage of the polyprotein) and the 100 kDa protein is a polyprotein encoded by ORFs 2a and 2b by a −1 ribosomal frameshift. ORF2b alone does not direct any in vitro protein synthesis. CfMV 34 kDa protein is a coat protein synthesised from the virion-packed subgenomic RNA. The translational strategies of different sobemoviruses are also discussed in this paper.

Introduction

Cocksfoot mottle virus (CfMV) is a member of the Sobemovirus genus of plant viruses [10]. The genome of sobemoviruses is a monopartite single-stranded positive-sense RNA molecule linked to a protein VPg at the 5′-terminus and non-polyadenylated at the 3′-terminus [3, 9]. The complete nucleotide sequences of the genomic RNA of two CfMV isolates, Norwegian isolate (CfMV-Nor) and Russian isolate (CfMV-Rus), have been determined [12, 18]. The coding region of CfMV contains four open reading frames (ORFs) (Fig. 1) [12].

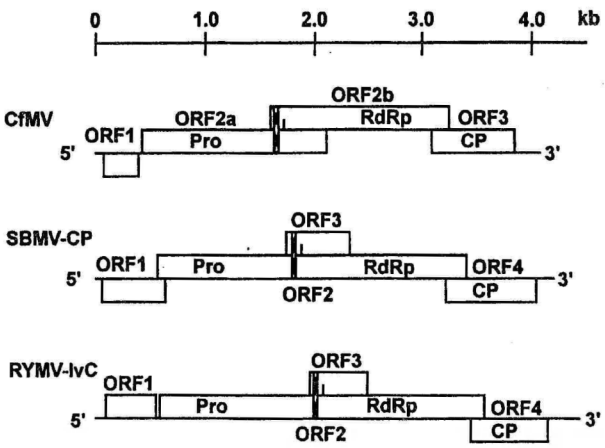


Fig. 1. Genome organization of CfMV, SBMV-CP and RYMV-IvC. The ORFs are illustrated as boxes. The putative serine protease (*Pro*) and RNA-dependent RNA polymerase (*RdRp*) domains are indicated. The ORFs coding for the coat protein are marked as *CP*. The sites of −1 ribosomal frameshift consensus signals are shown as a filled boxes and the first AUG in −1 reading frame after these signals is marked by a vertical line

The organisation of the CfMV genome differs significantly from that of other sequenced sobemoviruses (Fig. 1). First of all, CfMV lacks ORF2 that codes for the polyprotein. Instead CfMV has two overlapping ORFs, ORF2a and ORF2b. The ORF2a encodes a 60.9 kDa protein which contains motifs characteristic for serine proteases and the putative VPg. The putative replicase of CfMV is encoded by ORF2b and is translated as part of a polyprotein by −1 ribosomal frameshift [11]. Although the other sobemoviruses which have been sequenced [i.e. the bean strain of southern bean mosaic virus (SBMV-B), the cowpea strain of southern bean mosaic virus (SBMV-CP), the Ivory Coast isolate of rice yellow mottle virus (RYMV-IvC), the Nigerian isolate of rice yellow mottle virus (RYMV-Nig) and lucerne transient streak virus (LTSV)], do not appear to use the −1 ribosomal frameshifting mechanism [6, 13–15, 22], they do contain the consensus signals for −1 ribosomal frameshift, the heptanucleotide sequence UUUAAAC and the putative stem-loop structure just downstream. CfMV differs also because it lacks a nested coding region similar to the ORF3 of SBMV-CP, RYMV and LTSV (Fig. 1). The function and the translational mechanisms of ORF3 are not known [13, 22].

The 5′-proximal ORF1 of CfMV potentially encodes a 12.3 kDa protein (P1) [12]. Recently it was demonstrated that P1 of RYMV-IvC is required for the infection of plants and is important for virus spread [2]. The function of the CfMV P1 is unknown. The coat protein (CP) of CfMV is encoded by ORF3 (Fig. 1). It was verified by N-terminal amino acid sequencing of the CfMV CP [12]. It has been demonstrated that the CP of SBMV is translated from a subgenomic RNA (sgRNA), which is also present within the virus particles [4, 17]. The genomic

RNA of 4.1 kb and one virus-specific RNA of 1.2 kb were detected in CfMV-Nor infected barley leaf extracts [12]. An RNA molecule of about 1.0 kb in addition to the genomic RNA has been identified also in virus particles of CfMV-Rus [18].

In vitro translation of CfMV-Nor virion-extracted RNA in wheat germ extract (WGE) revealed four major protein products of 100, 71, 34 and 16 kDa [11]. Similarly, RNAs of SBMV-B and SBMV-CP induce the translation of four major proteins in cell-free systems: 105, 75, 29, 14 kDa, and 100, 70, 30, 20 kDa respectively [9, 19]. It has been proposed that the polyprotein of SBMV-CP is translated from ORF2, the 70 kDa protein is derived from polyprotein by proteolytic processing, the 30 kDa protein is viral CP and the 20 kDa protein is presumably encoded by ORF1 [22]. The analysis of the nucleotide sequences of sobemoviruses suggests a number of possible strategies for gene expression. However, thorough studies on the mechanisms of protein translation in CfMV and other sobemoviruses are lacking. Furthermore, it has not been proven which ORFs are encoding the individual translational products of the sobemoviral virion-extracted RNAs. In this study we identified the proteins translated by virion-extracted RNA of CfMV-Nor in WGE system. We report here the production of polyclonal antisera against all proteins encoded by individual ORFs of CfMV-Nor. These antisera allowed us to identify by immunoprecipitation the individual ORFs responsible for coding the described in vitro translation products. As a result of the present study translational strategies used by CfMV and other sobemoviruses are proposed.

Materials and methods

Virus isolate and viral RNA extraction

CfMV-Nor propagation and purification were carried out as described earlier [20]. Total RNA was extracted from purified virus as described by Puurand et al. [16].

Plasmid construction

Specific oligonucleotides corresponding to the sequences located upstream and downstream of individual ORFs of CfMV were synthesised: 5′CCTAGATCTAGCTTAGATGTGCGAACCTC (contains a *Bgl*II site and CfMV nt 62 to 82) and 5′GAGCTGCAGAACAACCCATTCTTGGTCACCCT (contains a *Pst*I site and sequence complementary to CfMV nt 441 to 416) for ORF1, 5′GCATGCAGATCTGTTCTGTTGTTGGAAACTGCAAG (contains a *Bgl*II site and CfMV nt 435 to 458) and 5′TAACTGCAGTCTACGGATGCCAAGGCATGGA (contains a *Pst*I site and sequence complementary to CfMV nt 2203 to 2180) for ORF2a, 5′TAGGATCCGAGGACACCCAAGAAACTGCAATC (contains a *Bam*HI site and CfMV nt 1605 to 1628) and 5′ACTGCAGTCCTGCCTCTACCAGCCTTCA (contains a *Pst*I site and sequence complementary to CfMV nt 3265 to 3239) for ORF2b, 5′ATGGATCCGGTGAGGAAAGGAGCAGCAACGAA (contains a *Bam*HI site and CfMV nt 3098 to 3121) and 5′CATCCCTGCAGTACTTGACTAC (contains a *Pst*I site and sequence complementary to CfMV nt 3875 to 3854) for ORF3. The restriction endonuclease sites are underlined. Base numbering refers to the CfMV genome as in Mäkinen et al. [12].

CfMV cDNA fragments corresponding to the individual ORFs were produced from total viral RNA by RT-PCR using M-MLV-reverse transcriptase (Promega), *Taq* DNA polymerase (Perkin-Elmer), and appropriate primers. The cDNA fragments were cloned into pUC57/T

vector (Fermentas) and sequenced. Expression plasmids pORF1, pORF2b and pORF3 were generated by inserting the coding sequences of P1, P2b and P3, respectively (*Xba*I-*Pst*I fragements) into the *Xba*I and *Nsi*I sites of pGEM- 9Zf(−) (Promega). pORF2b utilises the internal translational initiation site present in ORF2b. The coding sequence of P2a (*Apa*I-*Pst*I fragment) was cloned into the *Apa*I and *Pst*I sites in the pGEM-5Zf(−) to produce pORF2a. To generate pORF2a-2b, the *Sac*II-*Pst*I fragment of pORF2b was cloned into the corresponding sites of pORF2a.

In vitro transcription with T7 RNA polymerase

The plasmid pXL15 was prepared by inserting a DNA fragment prepared by PCR, containing nucleotides 1-4082 of CfMV RNA, into plasmid pUC57. The insert is flanked at the 5′-end by a *Xho*I restriction site followed by a bacteriophage T7 RNA polymerase promoter sequence and at the 3′-end by the *Sal*I restriction site (T. Tamm, E. Truve, unpubl. res.). In vitro transcription reaction was carried out in a volume of 100 µl in 40 mM Tris-HCl (pH 7.5), 6 mM $MgCl_2$, 2 mM spermidine, 10 mM NaCl, 10 mM DTT, 10 µg of BSA, 50 U of RNasin, 1 mM ATP, CTP, and UTP, 0.1 mM GTP, 0.75 mM m^7GpppG, 3 µg of *Sal*I-linerized pXL15, and 100 U of T7 RNA polymerase (Pharmacia). After 2.5 h at 37 °C, 0.9 mM GTP was added and the reaction mixture was incubated for an additional 30 min. The in vitro transcripts were digested with 10 units of RQ1 DNase I (Promega) for 15 min at 37 °C followed by phenol-chloroform extraction and isopropanol precipitation.

Cell-free translation

In vitro translation in WGE was carried out as specified by the manufacturer (Promega) in the presence of [^{35}S]-methionine (Amersham) and using in vitro transcribed capped RNA (1.5 µg of transcript/25 µl of translation mixture) or RNA purified from viral particles (1.0 µg of RNA/25 µl of translation mixture).

In vitro translation was also performed in the coupled transcription-translation WGE system (TNT System; Promega) as described by the supplier. The constructs pORF1, pORF2b and pORF3 were linearized with *Not*I, pORF2a and pORF2a-2b were linearized with *Nde*I before adding to the reaction mixture (0.5 µg of linearized DNA/25 µl of translation mixture). Translation products were separated by SDS-PAGE [7] using either 12.5 % or 15% polyacrylamide gels and detected on dried gels using Bio-Rad Molecular Imager System GS-525. The results were processed using Adobe Photoshop 3.0 program (Adobe Systems). The relative amount of frameshifted product versus P2a was calculated by quantifying radioactive label in specific areas of the gel using Molecular Analyst software (Bio-Rad) and correcting for background and different methionine content of products.

Preparation of antisera

Antisera against the CfMV P1, P2a, P2b and P3 were generated against polypeptides produced in *E. coli* strain M15[pREP4] (T. Tamm, E. Truve, unpubl. res.). The purified recombinant proteins were fractionated on preparative SDS-PAGE, proteins were visualised by copper chloride staining [8] and the appropriate bands were excised. Gel slices were fragmented and emulsified in two volumes of complete Freund's adjuvant. The equivalent of 0.5 mg to 1.0 mg of protein was injected subcutaneously at multiple sites into rabbits. The rabbits were boosted three times at intervals of four weeks using the same amount of protein in incomplete Freund's adjuvant. Serum was collected eight days after each booster injection.

Immunoprecipitation

Immunoprecipitation of in vitro translation products was performed as described by Anderson and Blobel [1] with modifications. Twenty μl of in vitro reaction mixture were mixed with 2 μl of 10% SDS and heated for 4 min at 100 °C. Four volumes of dilution buffer [1.25% Triton X-100,160 mM NaCl, 60 mM Tris-HCl (pH 7.4), 6 mM EDTA] were added to the above sample and incubated overnight at 4 °C with 1.2 μl antiserum. The immunocomplexes were precipitated with protein A-Sepharose CL-4B beads (Pharmacia) for 4 h at room temperature. The protein A-Sepharose beads were pelleted (1 min at 10,000 g) and were washed seven times with wash buffer [0.1% Triton X-100, 0.02% SDS, 150 mM NaCl, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA] and four times with the same buffer without the detergent. Finally, the pellets were resuspended in SDS-PAGE sample buffer and heated for 5 min at 100 °C before analysis by SDS-PAGE on 12.5% or 15% polyacrylamide gels. Gels were dried and immunoprecipitated proteins were visualised using the BioRad Molecular Imager.

Results

Analysis of CfMV RNA in vitro translation products

It has been demonstrated that in WGE CfMV virion-extracted RNA directs the synthesis of four major protein products of 100, 71, 34 and 16 kDa, respectively [11]. In this work we have noted that the smallest major protein product migrates in SDS-PAGE with an apparent Mr of 12 kDa rather than 16 kDa (Fig. 2, lanes 1). The size of 12 kDa correlates well with the calculated Mr of ORF1 encoded protein, which is 12.3 kDa. Therefore we refer this protein as p12. In addition to those, polypeptides of 54, 43 and 23 kDa were translated in relatively low abundance (Fig. 2, lanes 1).

In vitro transcription/translation of ORFs 1, 2a, 2b, 2a-2b and 3

To facilitate investigation of the mechanism of expression of different ORFs of CfMV, cDNAs corresponding to the individual ORFs were cloned downstream of the bacteriophage T7 RNA polymerase promoter. The products of coupled transcription-translation reactions were analysed by SDS-PAGE (Fig. 2A, lane 2; Fig. 2B and C, lanes 2 and 3; Fig. 2D, lanes 3 and 4). The apparent sizes of the products were 12 kDa for ORF1, 71 kDa for ORF2a, 66 kDa for ORF2b and 34 kDa for ORF3. The sizes of these proteins, determined by SDS-PAGE, do not correlate precisely with calculated Mr's, but correlate well with CfMV genomic RNA translation products (except the p66). Expression of plasmid pORF2a-2b containing the entire ORF2a-ORF2b region of CfMV yielded two major protein products (Fig. 2B and C, lanes 2; Fig. 2D, lane 3). The translational product of 71 kDa has an electrophoretic mobility equal to the mobility and size of the ORF2a encoded protein. The upper protein band showed an apparent Mr of 100 kDa as expected for the putative transframe protein. Frameshifting efficiency was determined by counting incorporation of radioactive label into individual proteins and normalising for methionine content. Following the expression of pORF2a-2b, the polyprotein of 100 kDa was produced with an efficiency of 10.6±1.4% (calculated from 18 individual translations). Additional minor translation products were observed when pORF2a, pORF2b, pORF2a-2b or pORF3 constructs

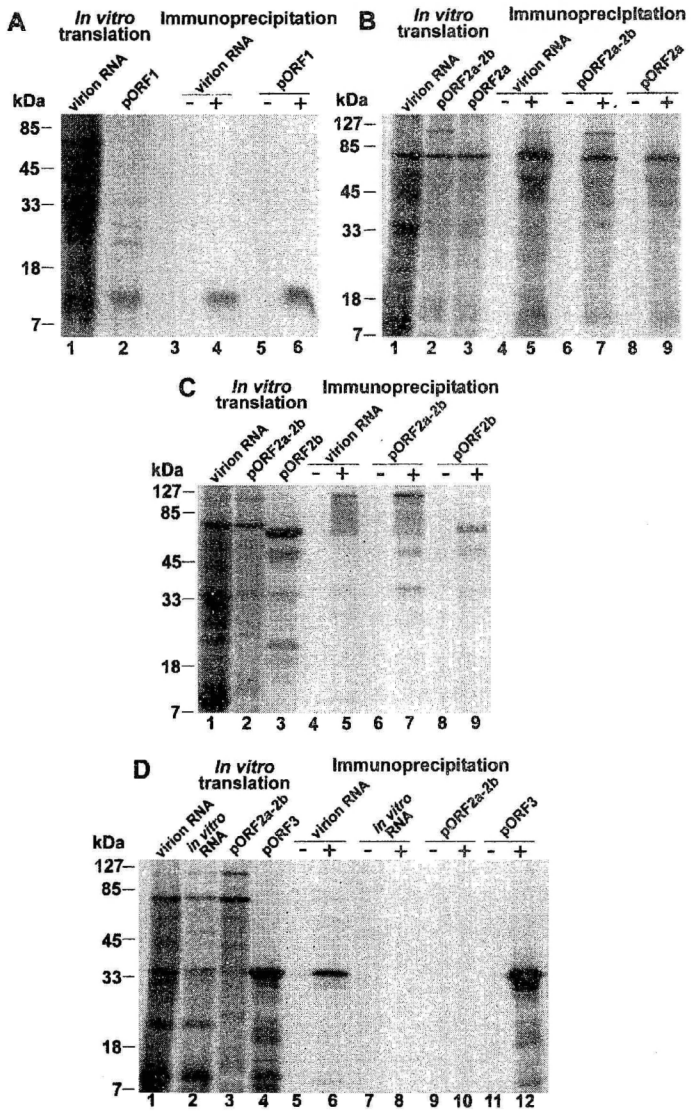


Fig. 2. Immunoprecipitation of CfMV in vitro translated proteins with P1 (**A**), P2a (**B**), P2b (**C**), and P3 antisera (**D**). The RNA templates used for in vitro translation (virion RNA = CfMV RNA purified from virus particles; in vitro RNA = in vitro transcribed CfMV full-length RNA) and the DNA constructs used for coupled transcription-translation (pORF1, pORF2a, pORF2b, pORF2a-2b, pORF3) are indicated above each lane. The immunoprecipitation of the CfMV in vitro translation products were carried out in the presence of specific antiserum (+) or in the presence of preimmune serum (−). The proteins were analysed on 15% (**A**) or 12.5% (**B–D**) SDS-PAGE. The positions of molecular mass markers are shown on the left

were translated. This can be attributed to downstream translational initiation or premature termination.

Immunodetection of CfMV recombinant proteins

The antisera generated against proteins encoded by the individual ORFs of CfMV recognised specifically the recombinant viral proteins expressed in *E. coli* (data not shown). In addition, P3 antiserum as well as polyclonal antibodies against CfMV particles [20] recognized specifically CfMV CP and 34 kDa protein from the CfMV infected leaf material similarly (data not shown).

The antisera were used to immunoprecipitate the products translated in vitro from individual ORFs. P1 antiserum immunoprecipitated the gene product of ORF1 synthesised in vitro in WGE, whereas the preimmune serum did not recognise any product (Fig. 2A, lanes 5 and 6). Furthermore, no cross-reactivity of P1 antiserum with other in vitro translated individual products was observed, proving that the P1 antiserum is specific for P1 (data not shown). Similarly, P2a, P2b and P3 antisera immunoprecipitated specifically the in vitro translation products of ORFs 2a, 2b and 3, respectively (Fig. 2B and C, lanes 9; Fig. 2D, lane 12). The antisera did not precipitate in vitro translation products of brome mosaic virus RNA, which served as a negative control (data not shown). The preimmune sera did not precipitate any of the in vitro translation products (Fig. 2A, lane 5; Fig. 2B and C, lanes 8; Fig. 2D, lane 11).

Identification of CfMV in vitro translation products

The antisera were used to identify the proteins produced from CfMV virion RNA during in vitro translation in WGE. P1 antiserum immunoprecipitated a single protein band which comigrated with ORF1 encoded protein (Fig. 2A, lane 4). This result indicates that CfMV ORF1 is translationally functional in the WGE system and the in vitro translation product of 12 kDa is CfMV P1.

Immunoprecipitation employing a P2a antiserum detected not only a 71 kDa protein, but also significant amount of 100 kDa product (Fig. 2B, lane 5). In a similar analysis of in vitro transcribed-translated pORF2a-2b, 71 and 100 kDa proteins were immunoprecipitated by P2a antiserum as well (Fig. 2B, lane 7). In addition, minor bands with apparent molecular masses of 54 and 43 kDa were also reactive with the P2a antiserum when the translational products of CfMV RNA, pORF2a-2b and pORF2a were studied (Fig. 2B, lanes 5, 7 and 9).

A single protein band was observed after immunoprecipitation with P2b antiserum of CfMV proteins translated in vitro from virion RNA (Fig. 2C, lane 5). This protein of 100 kDa comigrated with the polyprotein in the unprecipitated translation mixture. None of the major low-molecular-mass proteins were precipitated by this antiserum, indicating that ORF2b is not capable of initiation of in vitro translation from genomic RNA and is translated only *via* -1 ribosomal frameshifting mechanism.

When the CfMV in vitro translation products were immunoprecipitated with polyclonal serum against P3, a 34 kDa protein was detected (Fig. 2D, lane 6).

This result suggests that the CfMV 34 kDa protein is the CfMV CP and it is translated from the virion-extracted RNA. To test the possibility that an sgRNA in the viral RNA preparation was responsible for synthesis of p34, the translational products of full-length in vitro transcribed CfMV RNA (Fig. 2D, lane 2) were immunoprecipitated with P3 antiserum. No immunoprecipitated proteins were detected (Fig. 2D, lane 8). Since in vitro transcribed RNA does not contain sgRNA, these results suggest that the p34 is translated from the sgRNA encapsidated in the virus particles.

It should be noted that the 23 kDa minor product, observed in CfMV RNA translation mixtures, was not immunoprecipitated by any antisera used. The source of this minor polypeptide remains unknown.

Discussion

In vitro translation of CfMV virion-extracted RNA resulted in the synthesis of four major proteins of 100, 71, 34 and 12 kDa in the WGE translation system (Fig. 2, lanes 1). As outlined above, the genome organisation of CfMV differs remarkably from other, so far sequenced sobemoviruses. Despite this, the in vitro translation of genomic RNA of all examined sobemoviruses in cell-free systems produces similar patterns of translation products. The data presented here together with the pre-existing information available for sobemoviruses, enables us to propose a model for the gene expression mechanisms used by CfMV and other sobemoviruses.

We have shown by immunoprecipitation that the 5′-terminal ORFs of CfMV are translated from the full-length genomic RNA whereas only the 3′-terminal ORF3 encoding the CP is translated from the sgRNA (Fig. 2A, lane 4; Fig. 2B and C, lanes 5; Fig. 2D, lanes 6 and 8).

The immunoprecipitation of in vitro translated CfMV proteins using the P1 antiserum identified a 12 kDa protein (Fig. 2A, lane 4). A comparison of the sequences surrounding the initiation codons for 5′ ORFs of sobemoviruses with the consensus sequences established for plant mRNAs shows that the sequences surrounding the AUG codon for ORF1 are in poor context for translation by plant ribosomes [12]. So far there was no experimental evidence that the first 5′ terminal ORF of CfMV is translationally active. It was recently demonstrated that two proteins of 18 and 19 kDa are translated in vitro in WGE from ORF1 sequence of RYMV-IvC and both proteins immunoreact with P1 specific antibody [2]. The immunoprecipitation of the CfMV RNA translation mixture with P2a antiserum detected two protein products of 71 and 100 kDa (Fig. 2B, lane 5). According to the immunoprecipitation results, both the 12 and 71 kDa proteins of CfMV are expressed from the same 4.0 kb genomic RNA which thus appears to be functionally bicistronic.

The 5′ terminal half of the CfMV genome resembles in organisation the genome of subgroup II luteoviruses. ORFs 0 and 1 of subgroup II luteoviruses are both translated from genomic RNA, probably by the leaky scanning mechanism [21].

It has been demonstrated that the addition of the AUG codons to the 5′ UTR of SBMV-CP RNA not only decrease the translation of P1 but also reduce the overall translational level of SBMV-CP RNA [5]. By analogy with luteoviruses and considering the results described above it seems reasonable to speculate that ribosomes could bypass the 12 kDa gene because of the suboptimal context of its initiation codon and therefore ORF2a could be translated by a leaky scanning mechanism. However, the experiments carried out so far do not exclude other possible mechanisms for the translational initiation of ORF2a of CfMV.

In general, the polyprotein of the sobemoviruses is encoded by the large ORF2. The only exception of this rule is CfMV, as its polyprotein is translated from two overlapping ORFs, ORF2a and 2b [12]. We have shown by immunoprecipitation that both P2a antiserum and P2b antiserum precipitated the 100 kDa in vitro translation product of CfMV (Fig. 2B and C, lanes 5 and 7). In addition, the 100 kDa protein was the only protein observed after the immunoprecipitation with P2b antiserum (Fig. 2C, lane 5). These results indicate that ORF2b is not capable of initiating its translation from genomic RNA and is translated only by −1 ribosomal frameshift. We also found that when the pORF2a-2b construct was transcribed-translated in WGE, the efficiency of −1 ribosomal frameshift was 10.6±1.4%, which is lower than in the reporter gene context [11].

The genomes of SBMV-CP, RYMV and LTSV have a small ORF3 nested in the middle of ORF2 in −1 reading frame. The functions and the translational mechanisms of this ORF are not known. Interestingly, the ORF3 encoded proteins have homology to the N-terminal part of the ORF2b encoded protein of CfMV [11]. The genomes of all sobemoviruses contain similar consensus signals for −1 ribosomal frameshift as characterised for CfMV. In the case SBMV-CP, RYMV and LTSV these signals are located upstream of the initiation codon of ORF3. It is interesting to note that there are no stop codons present between the slippery sequence and the initiation codon of ORF3 in both reading frames. The most obvious way for expression of this part of the genomes of SBMV-CP, RYMV, LTSV to occur is via −1 ribosomal frameshift since a sgRNA corresponding to this region has not yet been found. The calculated Mr's for ORF2-ORF3 fusions are 65.8 kDa for SBMV-CP, 67.8 kDa for RYMV-IvC, 64.1 kDa for RYMV-Nig and 63.3 kDa for LTSV. It was hypothesized that the 70 kDa in vitro translation product of SBMV-CP RNA is derived from the polyprotein (p100) by proteolytic processing [22]. We propose here that the 70 kDa in vitro translation products of SBMV-CP and RYMV-IvC virion RNA may represent the ORF2-ORF3 trans-frame fusion proteins. The fact that CfMV 71 kDa protein is encoded by ORF2a and is not the product of the proteolytic cleavage support this idea.

Previous studies have indicated that the CP of SBMV is translated from a sgRNA which is encapsidated in the virus particles [4, 17]. Sobemoviral RNAs smaller than unit length have been also reported for RYMV-IvC (1 kb) [2], CfMV-Nor (1.2 kb) [12], CfMV-Rus (about 1 kb) [18]. We have shown that the 34 kDa protein which is translated in vitro from the CfMV virion-extracted RNA comigrated with ORF3 encoded CP and immunoprecipitated with P3 antiserum, specif-

ically recognising the CfMV CP (Fig. 2D, lane 6). When the translation products of full-length in vitro transcribed CfMV RNA were immunoprecipitated with P3 antiserum, no proteins were detected (Fig. 2D, lane 8). These results support the model according to which ORF3 of CfMV is expressed from sgRNA. In earlier report we have been unable to detect the sgRNA from CfMV-Nor particles [12]. However, subsequent experiments have shown that when larger amounts of virion RNA were loaded on the gels or the autoradiographs were exposed for longer times, it was possible to detect the sgRNA also from CfMV-Nor particles. The content of the sgRNA in virions was largely dependent on the stage of the infection when virus particles were isolated from infected material (data not shown).

In conclusion, we suggest that all sobemoviruses have several common features with respect to in vitro messenger function: (i) the genomic RNA appears to be functionally bicistronic and both ORF1 and ORF2 encoded proteins are translated from the genomic RNA; (ii) the CP is translated as a single translational product from a sgRNA, which is also present within the virus particles; (iii) the −1 ribosomal frameshift is probably exploited by all sequenced sobemoviruses. In vitro translation of CfMV genomic RNA defines a set of possible gene products that may appear in vivo during the virus infection. The availability of antisera specific for the proteins encoded by the individual ORFs of CfMV will allow us to determine whether the in vitro synthesised polypeptides can be identified in virus-infected plants.

Acknowledgements

The authors thank Dr. Milvi Agur for help with the immunisation procedure, Prof. Mart Saarma and Dr. Ats Metsis for useful comments throughout the work and Marie-Lise Bouscaren for critical reading of the manuscript. This work was supported by Estonian Science Foundation and EC INCO-Copernicus programme (Contract No. IC15CT960907).

References

1. Anderson DJ, Blobel G (1983) Immunoprecipitation of proteins from cell-free translations. Methods Enzymol 96: 111–120
2. Bonneau C, Brugidou C, Chen L, Beachy RN, Fauquet C (1998) Expression of the rice yellow mottle virus P1 protein in vitro and in vivo and its involvement in virus spread. Virology 244: 79–86
3. Ghosh A, Dasgupta R, Salerno-Rife T, Rutgers T, Kaesberg P (1979) Southern bean mosaic viral RNA has a 5′-linked protein but lacks 3′ terminal poly(A). Nucleic Acids Res 7: 2 137–2 146
4. Ghosh A, Rutgers T, Mang KQ, Kaesberg P (1981) Characterization of the coat protein mRNA of southern bean mosaic virus and its relationship to the genomic RNA. J Virol 39; 87–92
5. Hacker DL, Sivakumaran K (1997) Mapping and expression of southern bean mosaic virus genomic and subgenomic RNAs. Virology 234; 317–327
6. Jeffries AC, Rathjen JP, Symons RH (1995) Lucerne transient streak virus complete genome. GenBank Accession Number U31286

7. Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227: 680–685
8. Lee C, Levin A, Branton D (1987) Copper staining; a five-minute protein stain for sodium dodecyl sulfate-polyacrylamide gels. Anal Biochem 166: 308–312
9. Mang KQ, Ghosh A, Kaesberg P (1982) A comparative study of the cowpea and bean strains of southern bean mosaic virus. Virology 116: 264–274
10. Murphy FA, Fauquet CA, Bishop DHL, Ghabrial SA, Jarvis AW, Martelli GP, Mayo MA, Summers MD (eds) (1995) Virus Taxonomy. Classification and Nomenclature of Viruses. Sixth Report of the International Committe on Taxonomy of Viruses. Springer, Wien New York (Arch Virol [Suppl] 10)
11. Mäkinen K, Næss V, Tamm T, Truve E, Aaspõllu A, Saarma M (1995) The putative replicase of the cocksfoot mottle sobemovirus is translated as a part of the polyprotein by −1 ribosomal frameshift. Virology 207: 566–571
12. Mäkinen K, Tamm T, Næss V, Truve E, Puurand Ü, Munthe T, Saarma M (1995) Characterization of cocksfoot mootle sobemovirus genomic RNA and sequence comparison with related viruses. J Gen Virol 76: 2 817–2 825
13. Ngon A Yassi M, Ritzenthaler C, Fauquet C, Brugidou C, Beachy RN (1994) Nucleotide sequence and genome characterization of rice yellow mottle virus RNA. J Gen Virol 75: 249–257
14. Othman Y, Hull R (1995) Nucleotide sequence of the bean strain of southern bean mosaic virus. Virology 206: 287–297
15. Pinto YM, Baulcombe DC (1995) Rice yellow mottle virus from Nigeria, complete genome. GenBank Accession Number U23142
16. Puurand Ü, Mäkinen K, Paulin L, Saarma M (1992) Nucleotide sequence of the 3′-terminal region of potato virus A RNA. Virus Res 23: 99–105
17. Rutgers T, Salerno-Rife T, Kaesberg P (1980) Messenger RNA for the coat protein of southern bean mosaic virus. Virology 104: 506–509
18. Ryabov EV, Krutov AA, Novikov VK, Zheleznikova OV, Morozov Syu, Zavriev SK (1996) Nucleotide sequence of RNA from the sobemovirus found in infected cocksfoot shows a luteovirus-like arrangement of the putative replicase and protease genes. Phytopathology 86: 391–397
19. Salerno-Rife T, Rutgers T, Kaesberg P (1980) Translation of southern bean mosaic virus RNA in wheat embryo and rabbit reticulocyte extracts. J Virol 34: 51–58
20. Truve E, Næss V, Blystad DR, Järvekülg L, Mäkinen K, Tamm T, Munthe T (1997) Detection of cocksfoot mottle virus particles and RNA in oat plants by immunological, biotechnical and electronmicroscopical techniques. Arch Phytopathol Pflanz 30: 473–485
21. Veidt I, Bouzoubaa SE, Leiser RM, Ziegler-Graff V, Guilley H, Richards K, Jonard G (1992) Synthesis of full-length transcripts of beet western yellows virus RNA: messenger properties and biological activity in protoplasts. Virology 186: 192–200
22. Wu S, Rinehart CA, Kaesberg P (1987) Sequence and organization of southern bean mosaic virus genomic RNA. Virology 161: 73–80

Authors' address: Dr. T. Tamm, Institute of Chemical Physics and Biophysics, Akadeemia tee 23, EE12618 Tallinn, Estonia.

CURRICULUM VITAE

TIINA TAMM

Citizenship: Estonian
Date and place of birth: November 6, 1968, Pärnu, Estonia
Address
Work: Laboratory of Molecular Genetics, National Institute of Chemical Physics and Biophysics, Akadeemia tee 23, 12618 Tallinn, Estonia
Phone: +372 6 398 390, +372 6 398 375; Fax: +372 6 398 382
e-mail: tiina@kbfi.ee
Home: Vilde tee 108–95, 12915 Tallinn, Estonia

Education

1977–1987	Saku Secondary School
1987–1992	University of Tartu, Faculty of Biology and Geography, graduated *cum laude* as cytologist
1992–1994	University of Tartu, Faculty of Biology and Geography, Institute of Molecular and Cell Biology, M. Sc. student
1994	Graduated M. Sc. Title of thesis: "Molecular characterisation of cocksfoot mottle virus"
1994–1999	University of Tartu, Faculty of Biology and Geography, Institute of Molecular and Cell Biology, Ph. D. student

Professional employment

1992–1997	National Institute of Chemical Physics and Biophysics, Laboratory of Molecular Genetics, research assistant
Since 1997	National Institute of Chemical Physics and Biophysics, Laboratory of Molecular Genetics, scientist

Scientific work

Since 1992 I am working with the project which concerns the characterisation of genomic structure and expression of Norwegian isolate of cocksfoot mottle virus (CfMV-NO). We have determined the complete nucleotide sequence of CfMV-NO, characterised viral *in vitro* translation products, identified the individual genes of CfMV-NO from which these products are synthesised and characterised the translational strategies used by this virus. All proteins encoded by individual ORFs of CfMV-NO have been expressed in *E. coli*. Rabbit polyclonal antisera against all CfMV-NO proteins recognising specifically viral proteins have been raised.

ELULOOKIRJELDUS

TIINA TAMM

Kodakondsus: Eesti
Sünniaeg ja -koht: 6. november 1968, Pärnu, Eesti
Aadress
Tööl: Keemilise ja Bioloogilise Füüsika Instituut, molekulaargeneetika laboratoorium, Akadeemia tee 23, 12618 Tallinn, Eesti
Tel.: +372 6 398 390, + 372 6 398 375, fax: +372 6 398 382
e-mail: tiina@kbfi.ee
Kodus: Vilde tee 108-95, 12915 Tallinn, Eesti

Haridus

1977–1987	Saku Keskkool
1987–1992	Tartu Ülikooli bioloogia-geograafiateaduskond, *cum laude* tsütoloogina
1992–1994	Tartu Ülikooli bioloogia-geograafiateaduskond, molekulaar- ja rakubioloogia instituut, magistrant
1994	Teadusmagistri kraad molekulaarbioloogias; väitekiri: "Keraheina laiguviiruse molekulaarne isoloomustamine"
1994–1999	Tartu Ülikooli bioloogia-geograafiateaduskond, molekulaar- ja rakubioloogia instituut, doktorant

Erialane teenistuskäik

1992–1997	Keemilise ja Bioloogilise Füüsika Instituut, molekulaargeneetika laboratoorium, insener
Alates 1997	Keemilise ja Bioloogilise Füüsika Instituut, molekulaargeneetika laboratoorium, teadur

Teadustegevus

Alates 1992. aastast olen tegelenud keraheina laiguviiruse (*Cocksfoot mottle virus*, CfMV) genoomi struktuuri ja bioloogia iseloomustamisega. Oleme kindlaks määranud CfMV Norra isolaadi genoomi nukleotiidse järjestuse, iseloomustanud *in vitro* translatsiooni produkte, näidanud, millistelt geenidelt need produktid transleeritakse ning iseloomustanud viiruse poolt kasutatavaid translatsiooni strateegiaid. Oleme ekpresseerinud kõik viiruse valgud *E. coli*'s ning kasutanud rekombinantseid valke küülikute immuniseerimisel. Käesolevaks ajaks on meil spetsiifilised antiseerumid kõikide selle viiruse valkude vastu.

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS

1. **Toivo Maimets**. Studies of human oncoprotein p53. Tartu, 1991, 96 p.
2. **Enn K. Seppet**. Thyroid state control over energy metabolism, ion transport and contractile functions in rat heart. Tartu, 1991, 135 p.
3. **Kristjan Zobel**. Epifüütsete makrosamblike väärtus õhu saastuse indikaatoritena Hamar-Dobani boreaalsetes mägimetsades. Tartu, 1992, 131 lk.
4. **Andres Mäe**. Conjugal mobilization of catabolic plasmids by transposable elements in helper plasmids. Tartu, 1992, 91 p.
5. **Maia Kivisaar**. Studies on phenol degradation genes of *Pseudomonas* sp. strain EST 1001. Tartu, 1992, 61 p.
6. **Allan Nurk**. Nucleotide sequences of phenol degradative genes from *Pseudomonas sp.* strain EST 1001 and their transcriptional activation in *Pseudomonas putida.* Tartu, 1992, 72 p.
7. **Ülo Tamm**. The genus *Populus* L. in Estonia: variation of the species biology and introduction. Tartu, 1993, 91 p.
8. **Jaanus Remme**. Studies on the peptidyltransferase centre of the *E.coli* ribosome. Tartu, 1993, 68 p.
9. **Ülo Langel**. Galanin and galanin antagonists. Tartu, 1993, 97 p.
10. **Arvo Käärd**. The development of an automatic online dynamic fluorescense-based pH-dependent fiber optic penicillin flowthrought biosensor for the control of the benzylpenicillin hydrolysis. Tartu, 1993, 117 p.
11. **Lilian Järvekülg**. Antigenic analysis and development of sensitive immunoassay for potato viruses. Tartu, 1993, 147 p.
12. **Jaak Palumets**. Analysis of phytomass partition in Norway spruce. Tartu, 1993, 47 p.
13. **Arne Sellin**. Variation in hydraulic architecture of *Picea abies* (L.) Karst. trees grown under different enviromental conditions. Tartu, 1994, 119 p.
13. **Mati Reeben**. Regulation of light neurofilament gene expression. Tartu, 1994, 108 p.
14. **Urmas Tartes**. Respiration rhytms in insects. Tartu, 1995, 109 p.
15. **Ülo Puurand**. The complete nucleotide sequence and infections *in vitro* transcripts from cloned cDNA of a potato A potyvirus. Tartu, 1995, 96 p.
16. **Peeter Hõrak**. Pathways of selection in avian reproduction: a functional framework and its application in the population study of the great tit (*Parus major*). Tartu, 1995, 118 p.
17. **Erkki Truve**. Studies on specific and broad spectrum virus resistance in transgenic plants. Tartu, 1996, 158 p.
18. **Illar Pata**. Cloning and characterization of human and mouse ribosomal protein S6-encoding genes. Tartu, 1996, 60 p.

19. **Ülo Niinemets**. Importance of structural features of leaves and canopy in determining species shade-tolerance in temperature deciduous woody taxa. Tartu, 1996, 150 p.
20. **Ants Kurg**. Bovine leukemia virus: molecular studies on the packaging region and DNA diagnostics in cattle. Tartu, 1996, 104 p.
21. **Ene Ustav**. E2 as the modulator of the BPV1 DNA replication. Tartu, 1996, 100 p.
22. **Aksel Soosaar**. Role of helix-loop-helix and nuclear hormone receptor transcription factors in neurogenesis. Tartu, 1996, 109 p.
23. **Maido Remm**. Human papillomavirus type 18: replication, transformation and gene expression. Tartu, 1997, 117 p.
24. **Tiiu Kull**. Population dynamics in *Cypripedium calceolus* L. Tartu, 1997, 124 p.
25. **Kalle Olli**. Evolutionary life-strategies of autotrophic planktonic microorganisms in the Baltic Sea. Tartu, 1997, 180 p.
26. **Meelis Pärtel**. Species diversity and community dynamics in calcareous grassland communities in Western Estonia. Tartu, 1997, 124 p.
27. **Malle Leht**. The Genus *Potentilla* L. in Estonia, Latvia and Lithuania: distribution, morphology and taxonomy. Tartu, 1997, 186 p.
28. **Tanel Tenson**. Ribosomes, peptides and antibiotic resistance. Tartu, 1997, 80 p.
29. **Arvo Tuvikene**. Assessment of inland water pollution using biomarker responses in fish *in vivo* and *in vitro*. Tartu, 1997, 160 p.
30. **Urmas Saarma**. Tuning ribosomal elongation cycle by mutagenesis of 23S rRNA. Tartu, 1997, 134 p.
31. **Henn Ojaveer**. Composition and dynamics of fish stocks in the gulf of Riga ecosystem. Tartu, 1997, 138 p.
32. **Lembi Lõugas**. Post-glacial development of vertebrate fauna in Estonian water bodies. Tartu, 1997, 138 p.
33. **Margus Pooga**. Cell penetrating peptide, transportan, and its predecessors, galanin-based chimeric peptides. Tartu, 1998, 110 p.
34. **Andres Saag**. Evolutionary relationships in some cetrarioid genera (Lichenized Ascomycota). Tartu, 1998, 196 p.
35. **Aivar Liiv**. Ribosomal large subunit assembly *in vivo*. Tartu, 1998, 158 p.
36. **Tatjana Oja**. Isoenzyme diversity and phylogenetic affinities among the eurasian annual bromes (*Bromus* L., Poaceae). Tartu, 1998, 92 p.
37. **Mari Moora**. The influence of arbuscular mycorrhizal (AM) symbiosis on the competition and coexistence of calcareous crassland plant species. Tartu, 1998, 78 p.
38. **Olavi Kurina**. Fungus gnats in Estonia (*Diptera: Bolitophilidae, Keroplatidae, Macroceridae, Ditomyiidae, Diadocidiidae, Mycetophilidae*). Tartu, 1998, 200 p.
39. **Andrus Tasa**. Biological leaching of shales: black shale and oil shale. Tartu, 1998, 98 p.

40. **Arnold Kristjuhan.** Studies on transcriptional activator properties of tumor suppressor protein p53. Tartu, 1998, 86 p.
41. **Sulev Ingerpuu.** Characterization of some human myeloid cell surface and nuclear differentiation antigens. Tartu, 1998, 163 p.
42. **Veljo Kisand.** Responses of planktonic bacteria to the abiotic and biotic factors in the shallow lake Võrtsjärv. Tartu, 1998, 118 p.
43. **Kadri Põldmaa.** Studies in the systematics of hypomyces and allied genera (Hypocreales, Ascomycota). Tartu, 1998, 178 p.
44. **Markus Vetemaa.** Reproduction parameters of fish as indicators in environmental monitoring. Tartu, 1998, 117 p.
45. **Heli Talvik.** Prepatent periods and species composition of different *Oesophagostomum* spp. populations in Estonia and Denmark. Tartu, 1998, 104 p.
46. **Katrin Heinsoo.** Cuticular and stomatal antechamber conductance to water vapour diffusion in *Picea abies* (L.) karst. Tartu, 1999, 133 p.
47. **Tarmo Annilo.** Studies on mammalian ribosomal protein S7. Tartu, 1998, 77 p.
48. **Indrek Ots.** Health state indicies of reproducing great tits (*Parus major*): sources of variation and connections with life-history traits. Tartu, 1999, 117 p.
49. **Juan Jose Cantero.** Plant community diversity and habitat relationships in central Argentina grasslands. Tartu, 1999, 161 p.
50. **Rein Kalamees.** Seed bank, seed rain and community regeneration in Estonian calcareous grasslands. Tartu, 1999, 107 p.
51. **Sulev Kõks.** Cholecystokinin (CCK) — induced anxiety in rats: influence of environmental stimuli and involvement of endopioid mechanisms and erotonin. Tartu, 1999, 123 p.
52. **Ebe Sild.** Impact of increasing concentrations of O_3 and CO_2 on wheat, clover and pasture. Tartu, 1999, 123 p.
53. **Ljudmilla Timofejeva.** Electron microscopical analysis of the synaptonemal complex formation in cereals. Tartu, 1999, 99 p.
54. **Andres Valkna.** Interactions of galanin receptor with ligands and G-proteins: studies with synthetic peptides. Tartu, 1999, 103 p.
55. **Taavi Virro.** Life cycles of planktonic rotifers in lake Peipsi. Tartu, 1999, 101 p.
56. **Ana Rebane.** Mammalian ribosomal protein S3a genes and intron-encoded small nucleolar RNAs U73 and U82. Tartu, 1999, 85 p.

ISSN 1024–6479
ISBN 9985–56–463–4