EXPLORING AND EXPLOITING THE RNA GENOME OF

TOMATO SPOTTED WILT VIRUS

Peter de Haan

Promotor: **Dr R.W. Goldbach**

hoogleraar in de Virologie

Co-promotor: **Dr Ir D. Peters**

universitair hoofddocent bij de Vakgroep

Virologie

P. T. DE HAAN

EXPLORING AND EXPLOITING THE RNA GENOME

OF TOMATO SPOTTED WILT VIRUS

Proefschrift

ter verkrijging van de graad van doctor

in de landbouw- en milieuwetenschappen

op gezag van de rector magnificus,

Dr H.C. van der Plas,

in het openbaar te verdedigen

op woensdag 23 oktober 1991

des namiddags te vier uur in de Aula

van de Landbouwuniversiteit te Wageningen

BIBLIO THEEK
LANDBOUWUNIVERSITEIT
WAGENINGEN

STELLINGEN

1. Het Impatiens necrotic spot virus isolaat beschreven door Law & Moyer (1990) is een morfologisch defectieve mutant zoals omschreven door Resende et al. (1991) en als zodanig niet representatief voor dit virus species.

    Law, M.D. & Moyer, J.W. (1990). A tomato spotted wilt-like virus with a serologically distinct N protein. J. Gen. Virol. 71: 933-938.

    Resende, R. de O., De Haan, P., Avila, A.C. de., Kormelink, R., Goldbach, R. & Peters, D. (1991). Generation of envelope and defective interfering RNA mutants of tomato spotted wilt virus by mechanical passage. J. Gen. Virol. 72 (10) in press.

2. De uitspraak van Pfannenstiel et al. (1987) en Muthukumar & Nickerson (1987) dat de Bacillus thuringiensis kristaleiwitten glycoproteinen zijn, gevormd via N-glycosylering, suggereert ten onrechte dat deze bacterie beschikt over een endoplasmatisch reticulum en een Golgi-apparaat.

    Pfannenstiel, M.A., Muthutkumar, G., Couche, G.A. & Nickerson, K.W. (1987). Amino sugars in the glycoprotein toxin from Bacillus thuringiensis subsp. israelensis. J. Bacteriol. 169: 796-801.

    Muthutkumar, G. & Nickerson, K.W. (1987). The glycoprotein toxin of Bacillus thuringiensis subsp. israelensis indicates a lectinlike receptor in the larval mosquito gut. Appl. Envir. Microbiol. 53: 2650-2655.

3. De experimenten beschreven door Kawchuk et al. (1990, 1991), tonen niet onomstotelijk aan dat de waargenomen resistentie tegen het aardappelbladrolvirus een direct gevolg is van de introductie van het virale manteleiwitgen in het genoom van de onderzochte planten.

    Kawchuk, L.M., Martin, R.R. & McPherson, J. (1990). Resistance in transgenic potato expressing the potato leafroll virus coat protein gene. Mol. Plant-Microbe Inter. 3:301-307.

    Kawchuk, L.M., Martin, R.R. & McPherson, J. (1991). Sense and antisense RNA-mediated resistance to potato leafroll virus in Russet burbank potato plants. Mol. Plant-Microbe Inter. 3:301-307.

4. Bij het streven naar een taxonomie van RNA-virussen, gebaseerd op phylogenetische verwantschappen, verdient het aanbeveling geen taxa hoger dan familie te introduceren.

    Gibbs, A. (1987). Molecular evolution of viruses: 'Trees','clocks' and 'modules'. J. Cell. Sci. (suppl.) 7: 319-337.

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

    Zimmern, D. (1987). Evolution of RNA viruses. In: RNA Genetics, pp. 211-240. Edited by J. Holland, E. Domingo & P. Ahlquist. Boca Raton: CRC Press

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

5. Het ligt voor de hand dat tenuivirussen, net als alle andere negatief-strengs RNA-virussen, van nature beschikken over een lipide membraan.

    Gingery, R. (1988). The rice stripe virus group, In: The plant viruses, vol. 4, pp. 297-329. Edited by R.G. Milne. New York: Plenum Press.

    Ishikawa, K., Omura, T. & Hibino, H. (1990). Morphological characteristics of rice stripe virus. J. Gen. Virol. 70: 3465-3468.

6. De suggestie van Turpen (1989) om sequentie homologie in de 3'-uiteinden van de genomische RNAs van potyvirussen te gebruiken als taxonomisch criterium, wordt door de auteur zelf niet overgenomen.

    Turpen, T. (1989). Molecular cloning of a potato virus Y genome: nucleotide sequence homology in non-coding regions of potyviruses. J. Gen. Virol. 70: 1951-1960.

7. De oproep van De Zoeten (1991) aan wetenschappers die gebruik maken van genetisch gemodificeerde planten om vooral risico-analyses te maken, geeft blijk van zijn volslagen onbekendheid met het onderzoek op dit gebied.

    De Zoeten, G.A. (1991). Risk assessment: do we let history repeat itself? Phytopathology 81: 585-586.

8. De vorming van cylindrische insluitsels in protoplasten geïnfecteerd met potyvirussen is niet in tegenspraak met de veronderstelling dat deze betrokken zijn bij het cel-naar-cel transport van deze virussen.

    Langenberg, W.G. (1986). Virus protein association with cylindrical inclusions of two viruses that infect wheat. J. Gen. Virol. 67: 1161-1168.

    Murphy, J.F., Järlfors, U. & Shaw, J.G. (1991). Development of cylindrical inclusions in potyvirus-infected protoplasts. Phytopathology 81: 371-374.

9. Het gebruik van de benaming 'onecht' kind voor een buiten het huwelijk geboren kind, getuigt van minachting voor mensen die kiezen voor alternatieve samenlevingsvormen.

Stellingen behorend bij het proefschrift:

EXPLORING AND EXPLOITING THE RNA GENOME OF TOMATO SPOTTED WILT VIRUS

Randwijk, 27 september 1991   Peter de Haan

CONTENTS

voor Joanneke en Luuk

VOORWOORD

Ik zou graag de volgende personen willen bedanken voor hun bijdrage aan de totstandkoming van dit proefschrift: Mijn begeleiders Dick Peters en Rob Goldbach; de mensen van de TSWV werkgroep Lia Wagemakers, Claire Huguenot, Richard Kormelink, Renato de Oliveira-Resende, Antonio de Avila, Ineke Wijkamp, Frank van Poelwijk, Wies Smeets en Cor Meurs; de (ex) studenten Marja Korsman, Gerard Jongedijk, Jan Mosterd, Titia Sijen, Lonneke van der Geest, Anne Adema, Paul Brassé, Jaap Uijthof en Maurice van der Heyden; de 'sequencers' Marc Krueger en Yoshitaka Sano; de collega's van de Zaadunie Ad Kool, André Schramm, Mart van Grinsven en Jan Gielen; mijn kamergenoten Annelien Roenhorst en René van der Vlugt; de overige medewerkers van de Vakgroep Virologie en als laatste en meest belangrijke Joanneke Krabman.

CHAPTER 1   INTRODUCTION

1.1 Tomato spotted wilt virus: history and economical impact

The disease in tomato described as 'spotted wilt' was first observed by Brittlebank in Australia in 1915 (Brittlebank, 1919). Later studies revealed that the causal agent was a virus, ever since denoted tomato spotted wilt virus (TSWV) (Samuel et al., 1930). TSWV mainly occurs in (sub)tropical climate zones and is known under various different names (Table 1.1 and Best, 1968; Ie, 1970; Smith, 1972; Sakimura, 1962).

Kat river disease virus	Vira cabeça virus
Kromnek virus	Lycopersicon virus 3
Pineapple side rot virus	Lycopersicon virus 7
Pineapple yellow spot virus	Peanut bud necrosis virus
Tomato bronze leaf virus	Makhorka tip chlorosis virus
Tomato bronzing virus	Tip blight virus
Carcova virus	

Table 1.1 : List of synonyms of 'tomato spotted wilt virus'.

The economical impact of TSWV is enormous, due to its wide geographical distribution, its extremely broad host range and its devastating effects on infected plants. Up to 400 species from 50 different families, both monocotyledons and dicotyledons, can be infected (Smith, 1972; Francki et al., 1985; McRitchie, 1986; Peters, unpublished results). Infection with TSWV causes a wide variety of different symptoms such as necrosis, chlorosis, enation, stunting and local lesions, depending on the host plant species and virus isolate. Both genetical and environmental factors seem to affect the susceptibility of the host and the severity of the induced symptoms (Francki et al., 1985).

Since indications have been reported for the presence of different 'strains' of TSWV (Best & Gallus, 1955; Best, 1968), serological differentiation of isolates is currently a

1

topic of intensive research (Cho et al., 1988; Kameyi-Iwaki et al., 1988; Wang & Gonsalves, 1990; Avila et al., 1990). These studies have revealed that TSWV isolates can be separated into a number of serogroups and serotypes (Law & Moyer, 1990; Avila et al., 1990).

The most important plant species, for which considerable yield losses have been reported, include tomato, potato, tobacco, groundnut, (sweet) pepper, lettuce and papaya. Although TSWV mainly occurs in (sub)-tropical regions or in areas with relatively warm summers, severe outbreaks have recently also been reported in many countries within the temperate climate zones. In the Netherlands, for example, TSWV has become an actual threat in the cultivation of tomato, sweet pepper, egg-plant and a growing number of ornamental plants, like chrysanthemum, Ageratum and Impatiens, grown in greenhouses.

1.2 Transmission and disease management

TSWV is transmitted by thrips (Thysanoptera). Sofar eight species have been reported as possible vectors, from which the first four listed in Table 1.2, seem to be the most important ones (Sakimura, 1962; Kobatake et al., 1984; Paliwal, 1974). The recently observed expansion of TSWV into Western Europe and North America is most likely due to the introduction of the Western Flower thrips, Frankliniella occidentalis into these areas.

There is only little information on the virus transmission mechanism. As far as known, the mode of transmission seems analogous to that of the group of persistently transmitted aphid-borne plant viruses. Acquisition of virus by its vector takes only place during the first and second larval stages with an acquisition threshold of 5 to 45 min. The latency period is about 5 to 10 days and virus can be retained in the vector for a long period (probably for its whole lifespan) (Sakimura, 1962; Paliwal, 1974; Cho et al., 1988). Recently, virus particles have been detected in the vector. It is however not known whether TSWV multiplies in the vector (Ullman et al., 1989).

2

Thrips tabaci   Lindeman
Frankliniella schultzei   Trybom
Frankliniella occidentalis   Pergande
Frankliniella fusca   Hinds
Frankliniella tenuicornis   Uzel
Thrips setosus   Moulton
Thrips palmi   Karny
Scirtothrips dorsalis   Hood

Table 1.2 : Thrips species reported as vectors of tomato spotted wilt virus.

In the laboratory TSWV is easily transmittable from plant to plant by sap inocula-
tion, but in the field mechanical transmission, as well as transmission by seed or pollen,
does not play a significant role in the virus spread (Ie, 1970; Peters et al., 1990).

From studies in the USA and Japan it is known that a large number of naturally
occurring weed plants can serve as virus reservoirs, thereby playing a crucial role in the
survival and distribution of the virus (Cho et al., 1986; Bond et al., 1983; Kobatake et al.,
1984). The epidemiology of vector and virus is poorly understood and it is clear that
more detailed studies on this topic are urgently needed to develop feasible management
procedures.

To limit the incidence of TSWV infections, several sanitary measures can be taken,
such as removal of weed host plants in the vicinity of threatened crops, early destruction
of infected plants and biological or chemical control of thrips. In addition, plant breeders
continuously try to obtain crops with increased TSWV resistance or tolerance levels.
Plant resistance to TSWV has intensively been studied during the last decades. Studies
on tomato, tobacco and lettuce have revealed however, that a simple 'Mendelian'
inheritance of natural resistance against TSWV does almost never occur in these crops.
It seems therefore likely that natural TSWV-resistance is predominantly polygenic, based
on complex interactions between virus, vector and plant (Smith & Gardner, 1951; Finlay,
1953; Borchers, 1956; Holmes, 1958; Best, 1968; Moldovan & Chokan, 1972; Stoyanova
& Konotop, 1975; Vinogradov et al., 1982; Gajos, 1983; Cupertino et al., 1986; O'Malley
& Hartmann, 1989; Paterson et al., 1989).

1.3 Virus structure

In the past thirty years extensive electron microscopical studies have been performed on TSWV in infected plant cells. These analyses have shown that TSWV particles are spherically shaped (80-110 nm in diameter) and consist of a granular core of nucleocapsids, bounded by a lipid membrane, which is covered with surface projections (Best & Katekar, 1964; Best & Palk, 1964; Ie, 1964; Martin, 1964; Kitajima, 1965). Virus particles are found clustered within dilated cysternae of the rough endoplasmatic reticulum (Best & Palk, 1964; Ie, 1964; Kitajima, 1965; Martin, 1964; Francki & Grivell, 1970) and most likely mature by budding of nucleocapsids through the ER membrane (Milne, 1970; Kitajima et al., 1991). Sofar, there is no evidence that the Golgi complex is involved in maturation or transport of the virus (Best & Palk, 1964; Ie, 1964; Milne, 1970). Besides enveloped virus particles, other structures, associated with TSWV-infection, are observed by electron microscopy. The cytoplasm of infected cells also contains clusters of electron-dense masses (also described as diffuse masses or viroplasm) and fibers, fibrous
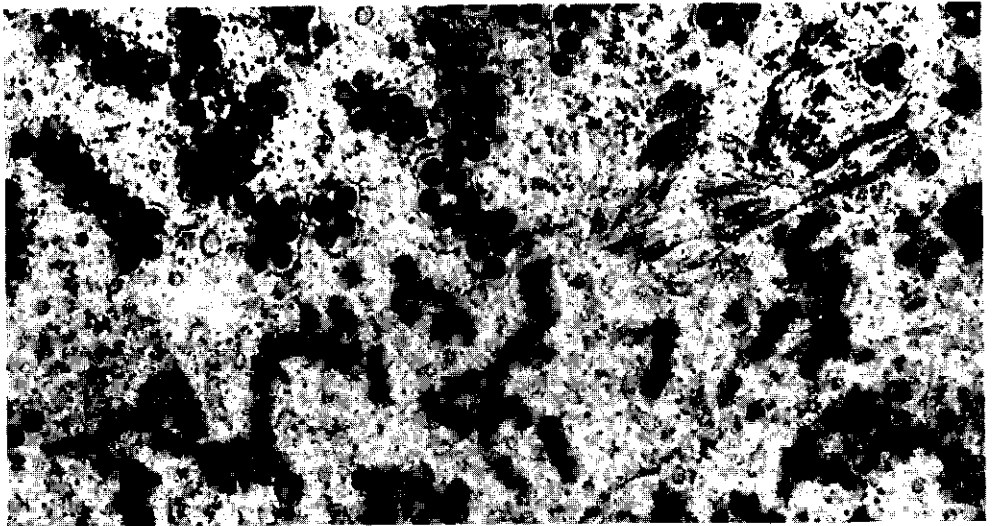

Fig.1.1: Electron micrograph of a TSWV infected Nicotiana rustica cell. Virus particles (V), diffuse electron-dense masses (DM) and fibrous structures (F) can be observed.

4

structures or tubuli (Francki & Grivell, 1970; Milne, 1970; Law & Moyer, 1990; Kitajima et al., 1991). The electron-dense masses represent aggregates of non-enveloped nucleocapsids (Ie, 1964; Kitajima et al., 1991) (Fig.1.1).

In vitro, TSWV particles are highly instable, as can be concluded from the short half-life in plant sap (30-60 min) and the thermal inactivation point of 45 °C. (Ie, 1970). A few methods have been developed to purify TSWV particles. They are based on differential centrifugation, followed by banding in sucrose gradients, in the presence of a mild reducing agent, such as sodiumsulfite or cystein. (Black et al., 1963; Best & Palk, 1964; Martin, 1964; Van Kammen et al., 1966; Best, 1968; Gumpf & Weathers, 1972; Tsakiridis & Gooding, 1972; Joubert et al., 1974; Paliwal, 1974;). The use of antisera against uninfected plant material, to remove host contaminants during purification, is a valuable tool to obtain highly purified virus preparations (Tas et al., 1977b).

Purified virus particles contain four to five structural proteins, with molecular masses of respectively 29,000 (29K), 52K, 58K, 78K and approximately 200K (Mohamed et al., 1973; Tas et al., 1977a). Treatment of virus particles with a non-ionic detergent like Nonidet-P40 and subsequent sucrose gradient centrifugation shows that the 29K protein is tightly associated with the genomic RNA and forms stable circular nucleocapsid struc-tures. This protein is therefore denominated the nucleocapsid (N) protein. Virus preparations contain three distinct nucleocapsids, each containing a copy of genomic RNA (Mohamed, 1981). The nucleocapsids, together with few copies of the 200K protein (denoted L protein) form the interior of the virus particle. The function of the L protein is thusfar unknown. The 78K and 58K proteins are glycosylated and are denominated glycoproteins, G1 and G2, respectively. The 52K protein most likely is a partial degra-dation product of the 58K protein, since it is not present in all virus preparations (Mohamed et al., 1973).
Iodination experiments and protease treatment of virus particles have revealed that the G proteins are located at the surface. It can therefore be assumed that one or both glycoproteins form the observed surface projections (spikes). The structure of these spikes however has remained unknown (Mohamed et al., 1973; Tas et al., 1977a). The

5
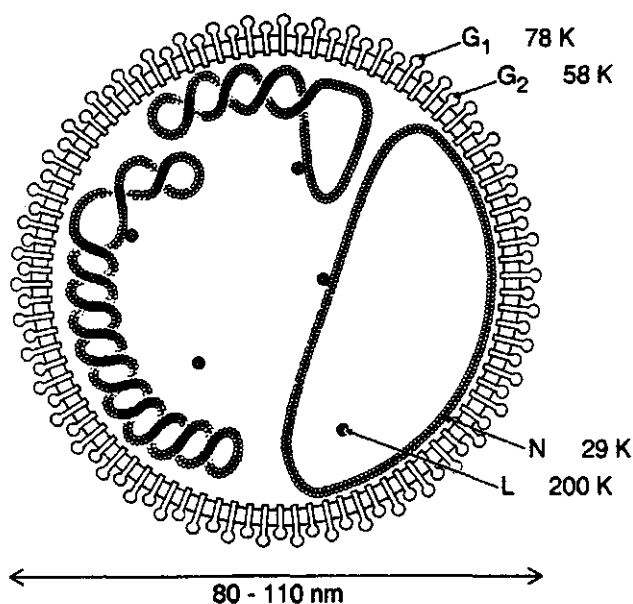
Fig.1.2: Schematical representation of a TSWV particle. The three linear, single strand RNA segments are tightly encapsidated with nucleocapsid (N) proteins and form circles, which may be coiled. The lipid envelop contains two types of glycoproteins denoted G1 and G2. A large (L) protein, present in minor amounts, is associated with the internal nucleocapsids.

putative morphology of a TSWV particle is drawn schematically in Fig.1.2.

TSWV has a genome consisting of three linear single stranded RNA molecules, denominated (small) S RNA, (middle) M RNA and (large) L RNA. Polyacrylamide and agarose gel electrophoresis have shown that the S RNA is approximately 3000 nucleotides long, with a molecular mass of $1,1 \times 10^6$; the M RNA is approx. 5000 nucleotides long ($1,7 \times 10^6$) and the L RNA has a length of approximately 8000 nucleotides ($2,7 \times 10^6$) (Van den Hurk et al., 1977; Verkleij et al., 1982). Chromatography of genomic RNA on oligo-dT cellulose columns has demonstrated that the genomic RNA molecules are not poly-adenylated at their 3' termini (Verkleij et al., 1982). At the start of the studies described in this thesis, the structure and coding capacity of the RNA segments was completely unknown.

6

1.4   Taxonomy

Polyclonal antisera have been raised against purified virus preparations, or against TSWV nucleocapsids, in a number of laboratories (Joubert et al., 1974; Paliwal, 1974; Tas et al., 1977b; Gonsalves & Trujillo, 1986; Law & Moyer, 1990; Resende et al., 1991). With these antisera sensitive, serological detection assays have been developed, for diagnostic purposes and for supporting fundamental research on TSWV. Recently, also monoclonal antibodies have been obtained to the nucleocapsid protein and to the membrane glycoproteins to analyse different TSWV isolates in more detail (Huguenot et al.,1989; Sherwood et al.,1989; Avila et al.,1990).

Sofar plant viruses have been classified on the basis of serology, particle morphology, genome structure and biological properties, such as transmission and host range. A criterium of increasing importance for the taxonomy of RNA viruses has become the structure and genetical organization of the genome (Goldbach, 1986; Strauss & Strauss, 1988). Hence, the polarity of the genomic RNA is an important mainstay for virus classification. At the beginning of the molecular studies presented in this thesis, the polarity of the genome of TSWV was not unequivocally determined. Purified TSWV RNA is not infectious, in contrast to purified nucleocapsids, indicating a negative polarity (Van den Hurk et al., 1977; Mohamed, 1981). On the contrary it has been reported that TSWV RNA encodes virus specific proteins in cell-free translation assays. In addition, no in vitro transcriptase activity could thusfar be detected under conditions described for RNA polymerases of other negative-strand viruses. These results would suggest a positive polarity (Verkleij et al., 1982).

If TSWV is indeed an enveloped, positive-strand RNA virus, such as the Corona-viridae, Flaviviridae, Togaviridae and Toroviridae, then it would be the only enveloped positive-strand RNA virus with a tripartite genome. When alternatively, TSWV has a negative-stranded genome, it has many properties in common to members of the Bunyaviridae, a large family of arthropod-borne viruses, sofar only found in the animal kingdom.

The similarities of TSWV to members of the Bunyaviridae are so striking that it has

7

been proposed to consider TSWV as a possible member of this virus family (Milne & Francki, 1984; De Haan et al., 1989) (Table 1.3).

The current lack of detailed knowledge on the structure and genetic properties of the genome and on the biology of TSWV makes a reasonable comparison and definitive

Property	TSWV	Bunyaviridae
Morphology		
Shape	spherical	spherical
Diameter (nm)	80-110	90-120
Envelop	+	+
Surface projections	+	+
Circular nucleocapsids	+	+
$S^o_{20,w}$	520-530 S	350-470 S
Buoyant density in		
CsCl, g/cm^3	1.21	1.20
Morphogenesis		
Maturation	budding into RER	budding into RER
Localization	cysternae of ER	Golgi-complex
Structural proteins		
N (Mr x 10^{-3})	29	19-54
G1	78	55-120
G2	58 (52)	29-70
L	200	145-250
Genome		
Type	ssRNA	ssRNA
Number of segments	3	3
Polarity	?	negative/ambisense
Length of the segments(Kb)		
and coding properties		
S RNA	3.0 (?)	0.8-2.0 (N, NSs)
M RNA	5.0 (G1 and G2,?)	3.2-4.6 (G1,G2,NSm)
L RNA	8.0 (?)	6.5-15.0 (L)
Transmission		
Vector	Thrips	Ticks, mosquitos, sandflies and other arthropods
Vertical transmission	?	+
Replication in the vector	?	+

Table 1.3: Some properties of TSWV compared with those of the Bunyaviridae (Milne & Francki, 1984)

8

classification thusfar impossible. Among the plant viruses, TSWV is certainly unique and for this reason it has sofar been classified as the single representative of a distinct group of plant viruses: the tomato spotted wilt virus group (Ie, 1970; Matthews, 1982).

1.5 Scope of the investigation

As outlined in the previous paragraphs, among the plant viruses TSWV takes a unique position. Moreover, this virus has become a limiting factor in the cultivation of many crops. Despite the considerable economical importance of TSWV, our knowledge on the molecular structure of this virus has thusfar remained very limited and fragmentary. In order to classify this virus properly and to design effective management strategies, it is of great importance to gain insight into the coding capacity and genetic organization of the genomic RNAs and into the functions of the viral proteins in replication, virulence and transmission by its vector. For these reasons research on the molecular properties of TSWV has been initiated. The obtained nucleotide sequence data will enable development of faithful detection techniques and moreover, transformation strategies to gain virus-resistance can be employed. In addition, these data will confirm or deny the phylogenetical relationships between TSWV and the Bunyaviridae as proposed for several years.

In view of this possible relationship, Chapter 2 of this thesis starts with a description of the molecular properties of the Bunyaviridae in relation to other negative-strand virus families.

The next three Chapters (3, 4 and 5) will describe the molecular cloning and sequence determination of the S and L RNA segments of the TSWV genome.

Chapter 6 subsequently describes the application of the cDNA clones obtained for diagnosis and sensitive detection of TSWV. Last (but not least), Chapter 7 deals with the utilization of the cloned nucleocapsid protein gene of TSWV, for creating resistance of host plants to this virus. This genetically engineered resistance is a first step on the way to control TSWV-induced diseases.

REFERENCES

Avila A.C. de, Huguenot, C., Resende, R. de O., Kitajima, E.W., Goldbach, R. & Peters, D. (1990). Serological differentiation of 20 isolates of tomato spotted wilt virus. J. Gen. Virol. 71: 2801-2807.

Avila A.C. de, De Haan, P., Kitajima, E.W., Kormelink, R., Resende, R. de O., Goldbach, R. & Peters, D. (1991). Characterization of a distinct tomato spotted wilt virus (TSWV) isolate from Impatiens sp. in the Netherlands. submitted for publication.

Best, R.J. & Gallus, H.P.C. (1955). Strains of tomato spotted wilt virus. Austr. J. Sci. 15: 212-214.

Best, R.J. & Katekar, G.F. (1964). Lipid in a purified preparation of tomato spotted wilt virus. Nature 203: 671-672.

Best, R.J. & Palk, B.A. (1964). Electron microscopy of strain E of tomato spotted wilt virus and comments on its probable biosynthesis. Virology 23: 445-460.

Best, R.J. (1968). Tomato spotted wilt virus. Adv. Virus Res. 13: 65-146.

Black, L.M., Brakke, M.K. & Vatter, A.E. (1963). Purification and electron microscopy of tomato spotted wilt virus. Virology 20: 120-130.

Bond, W.P., Whitman, H.K. & Black, L.L. (1983). Indigenous weeds as reservoirs of tomato spotted wilt virus in Louisiana. Phytopathology 73: 499.

Borchers, E.A. (1956). A study of the nature of resistance of the Mignonette variety of lettuce to the tomato spotted wilt virus. PhD. Thesis, University of California. Davis, California.

Brittlebank, C.C. (1919). Tomato diseases. J. Agr. Victoria 17: 213-235.

Cho, J.J., Mau, R.F.L., Gonsalves, D. & Mitchell, W.C. (1986). Reservoir weed hosts of tomato spotted wilt virus. Plant Disease 70: 1014-1017.

Cho, J.J., Mau, R.F.L., Hamasaki, R.T. & Gonsalves, D. (1988). Detection of Tomato spotted wilt virus in individual Thrips by Enzyme-linked Immunosorbent Assay. Phytopathology 78: 1348-1352.

Cupertino, F.P., Avila de, A.C., Araujo, M.T. & Maluf, W.R. (1986). Sources of resistance to tomato spotted wilt virus in Lycopersicon. Fitopatologia Brasileira 11: 330-331.

De Haan, P., Wagemakers, C.A.M., Goldbach, R. & Peters, D. (1989). Tomato spotted wilt virus, a new member of the Bunyaviridae? In: Genetics and pathogenicity of Negative strand viruses, pp. 287-291. Edited by D. Kolakofski & B.W.J. Mahy. Amsterdam: Elsevier Science Publishers B.V.

Finlay, K.W. (1953). Inheritance of spotted wilt resistance in the tomato. Austr. J. Biol. Sci. 6: 153-163.

Francki, R.I.B. & Grivell, C.J. (1970). An electron microscope study of the distribution of tomato spotted wilt virus in systemically infected Datura stramonium leaves. Virology 42: 969-978.

Francki, R.I.B., Milne, R.G. & Hatta, T. (1985). Atlas of plant viruses. vol 1. 222 pp. Boca Raton CRC Press.

Gajos, Z. (1983). Further progress in the breeding Tobacco resistant to tomato spotted wilt virus (Lycopersicon virus 3). Wyadom. Tyton 27: 12-17.

Goldbach, R. (1986). Molecular evolution of plant RNA viruses. Ann. Rev. Phyopathol. 24: 289-310.

Gonsalves, D. & Trujillo, E.E. (1986). Tomato spotted wilt virus in papaya and detection of virus by ELISA. Plant Dis. 70: 501-506.

Gumpf, D.J. & Weathers, L.G. (1972). Identification and purification of tomato spotted wilt virus isolated from Ageratum. Plant Dis. Rep. 56: 869-672.

Holmes, F.O. (1958). A sigle-gene resistance test for viral relationship as applied to strains of spotted wilt virus. Virology 5: 382-390.

Huguenot, C., Van den Dobbelsteen, G., De Haan, P., Wagemakers, C.A.M., Drost, G.A., Osterhaus, A.D.M.E. & Peters, D. (1989). Detection of tomato spotted wilt virus using monoclonal antibodies and riboprobes. Arch. Virol. 110: 47-62.

Ie, T.S. (1964). An electron microscope study of tomato spotted wilt virus in the plant cell. Neth. J. Plant Pathol. 70: 114-115.

Ie, T.S. (1970). Tomato spotted wilt virus. CMI/AAB Descr. of Plant Viruses 39.

Ie, T.S. (1982). A sap transmissable, defective form of tomato spotted wilt virus. J. Gen. Virol. 59: 387-391.

10

Joubert, J.J., Hahn J.S., Von Wechmar, M.B. & Van Regenmortel, M.H.V. (1974). Purification and properties of tomato spotted wilt virus. Virology 57: 11-19.

Kameyi-Iwaki, M., Hanada, K., Honda, Y. & Tochihara, H. (1988). A watermelon strain of tomato spotted wilt virus (TSWV-W) and some properties of its nucleocapsid (abstract). Fifth International Congress of Plant Pathology. Japan.

Kitajima, E.W. (1965). Electron microscopy of vira-cabeca virus (Brazilian tomato spotted wilt virus) within the host cell. Virology 26: 89-99.

Kitajima, E.W., Avila, A.C. de, Resende, R. de O., Goldbach, R. & Peters, D. (1991). Comparative cytological and immunogold-labelling studies on different isolates of tomato spotted wilt virus. Submitted for publication.

Kobatake, H., Osaki, T. & Inouye, T. (1984). The vector and reservoirs of tomato spotted wilt virus in Nara prefecturev (Japan). Ann. Phytopathol. Soc. Japan 50: 541-544.

Law, M.D. & Moyer, J.W. (1990). A tomato spotted wilt-like virus with a serologically distinct N protein. J. Gen. Virol. 71: 933-938.

Martin, M.M. (1964). Purification and electron microscope studies of tomato spotted wilt virus (TSWV) from tomato roots. Virology 22: 645-649.

Matthews, R.E.F. (1982). Classification and nomenclature of viruses. Fourth report of the International Committee on Taxonomy of viruses. Intervirology 17: 1-200.

McRitchie, J.J. (1986). Tomato spotted wilt. Plant Pathol. Circ. 287.

Milne, R.G. (1970). An electrone microscope study of tomato spotted wilt virus in sections of infected cells and in negative stain preparations. J. Gen. Virol. 6: 267-276.

Milne, R.G. & Francki, R.I.B. (1984). Should tomato spotted wilt virus be considered as a possible member of the family Bunyaviridae? Intervirology 22: 72-76.

Mohamed, N.A., Randles, J.W. & Francki, R.I.B. (1973). Protein composition of tomato spotted wilt virus. Virology 56: 12-21.

Mohamed, N.A. (1981). Isolation and characterization of subviral structures from tomato spotted wilt virus. J. Gen. Virol. 53: 197-206.

Moldovan, M.YA. & Chokan, N.G. (1972). Resistance of wild species of Nicotiana to spotted wilt (Lycopersicon virus 3 (Smith)). Izv. Akad. Nauk Mold. SSR. Ser. Biol. Khim. Nauk 1972: 38-44.

O'Malley, P.J. & Hartmann, R.W. (1989). Resistance to tomato spotted wilt virus in lettuce. Hort. Science 24: 360-362.

Paliwal, Y.C. (1974). Some properties and thrips transmission of tomato spotted wilt virus in Canada. Can. J. Bot. 52: 1177-1182.

Paterson, R.G., Scott, S.J. & Gergerich, R.C. (1989). Resistance in two Lycopersicon species to an Arkansas isolate of tomato spotted wilt virus. Euphytica 43: 173-178.

Peters, D., Avila, A.C. de, Kitajima, E.W., Resende, R. de O., De Haan, P. & Goldbach, R. (1991). An overview of tomato spotted wilt virus. In: Virus-Thrips-Plant Interactions of TSWV, Proc. USDA workshop, Belltsville U.S.A., pp.1-14. Edited by H-.T. Hsu & R.H. Lawson. Nat. Techn. Inf. Serv. Springfield (VA).

Resende, R. de O., Avila, A.C. de, Goldbach, R. & Peters, D. (1991). Detection of tomato spotted wilt virus using polyclonal antisera in double antibody sandwich (DAS) ELISA and cocktail ELISA. J. Phytopathol. 132: 46-56.

Sakimura, K. (1962). The present status of thrips-borne viruses. In: Biological transmission of disease agents, pp. 33-40. Edited by K. Maramorosch. New York: Academic Press.

Samuel, G., Bald, J.G. & Pitman, H.A. (1930). Spotted wilt of tomatoes. Council Sci. Ind. Res. (Australia) Bull. 44.

Sherwood, J.L., Sanborn, M.R., Keyser, G.C. & Myers, L.D. (1989). Use of monoclonal antibodies in detection of tomato spotted wilt virus. Phytopathology 79: 61-64.

Smith, K.M. (1972). Tomato spotted wilt virus. In: A textbook of plant virus diseases, pp. 245-249. Edited by K.M. Smith. 3rd. ed. London: Longman Group Ltd.

11

Smith, P.G. & Gardner, M.W. (1951). Resistance in tomato to the spotted wilt virus. Phytopathology 41: 257-260.

Stoyanova, M. & Konotop, A. (1975). A study on the resistance to Lycopersicon virus 3 (Smith) in progenies of Nicotiana tabacum (L.) X N. glauca (Grah.) interspecies hybridization. Compt. Rend. Acad. Agr. George Dimitrov 8: 43-44.

Strauss, J.H. & Strauss, E.G. (1988). Evolution of RNA viruses. Ann. Rev. Microb. 42: 657-683.

Tas, P.W.L., Boerjan M.L. & Peters, D. (1977a). The structural proteins of tomato spotted wilt virus. J. Gen. Virol. 36: 267-279.

Tas, P.W.L., Boerjan, M.L. & Peters, D. (1977b). Purification and serological analysis of tomato spotted wilt virus. Neth. J. Plant Pathol. 83: 61-72.

Tsakiridis, J.P. & Gooding, G.V. (1972). Tomato spotted wilt virus in Greece. Phytopathol. Mediterr. 11: 42-47.

Ullman, D.E., Westcot, D.M., Hunter, W.B. & Mau, R.F.L. (1989). Internal anatomy and morphology of Frankliniella occidentalis (Pergande) (Thysanoptera: Thripidae) with special reference to interaction between thrips and tomato spotted wilt virus. Int. J. Insect Morphol. & Embryol. 18: 289-310.

Van den Hurk, J., Tas, P.W.L. & Peters, D. (1977). The ribonucleic acid of tomato spotted wilt virus. J. Gen. Virol. 36: 81-91.

Van Kammen, A., Henstra, S. & Ie, T.S. (1966). Morphology of tomato spotted wilt virus. Virology 30: 574-577.

Verkleij, F.N., De Vries, P. & Peters, D. (1982). Evidence that tomato spotted wilt virus RNA is a positive strand. J. Gen. Virol. 58: 329-338.

Vinogradov, V.A., Trofimova, L.I. & Sarychev, Y.F. (1982). Field resistance of varieties, hybrids and chemical mutants of tobacco to tobacco thrips and tomato spotted wilt virus. Tabak 1982: 43-44, no. 3.

Wang, M. & Gonsalves, D. (1990). ELISA detection of various tomato spotted wilt virus isolates using specific antisera to structural proteins of the virus. Plant Dis. 74: 154-158.

12

CHAPTER 2   REPLICATION OF NEGATIVE-STRAND RNA VIRUSES

2.1 General introduction

Viruses consist of a relatively small amount of genetic information, protected by protein, which is in some cases also surrounded by a lipid envelop. Despite the obligate requirement for host cell structural and metabolic compounds, the viral genomes encode and regulate essential functions for their own replication, maturation, transport and hence survival of the virus. From this point of view, these intracellular parasitic entities can be regarded as 'independently operating genetic matter' and the intriguing question can be asked, whether there exist evolutionary relationships between the many different viruses, or in other terms, what is the origin of viral matter?

Especially, RNA viruses are interesting subjects to unravel their evolution. They are very diverse and widespread in nature and possess the capacity to adapt rapidly to changing environmental circumstances, due to the high rate at which mutation and recombination of their genomes occurs (Holland et al., 1982; Reanney, 1982; Goldbach, 1986; Strauss & Strauss, 1988; Smith & Palese, 1989). Besides, it is now commonly accepted that RNA is an older biological agent than DNA or protein (Reanney, 1979), which means that the study on viral RNA replication may reveal insight into one of the earliest molecular biological processes on earth (Eigen & Schuster, 1982; Zimmern, 1982).

The classical taxonomy of viruses is based on parameters like serological relationships, particle morphology, genome structure, host range, transmission, disease syndrome etc. (Matthews, 1982). Since no comparative virus material from the past is available, such as fossiles as found for other living organisms, evolution of viruses can only be investigated at the molecular level by analysis of their genomes.

Modern cloning and sequence determination techniques have led to the elucidation of complete nucleotide sequences of the genomes from a large number of viruses. This

13

has enabled the possibility to compare viruses and their genes in terms of nucleotide and amino acid sequences, and a considerable number of such comparative studies on the genomes of RNA viruses have demonstrated that, although sequence homologies are virtually absent, some of the presently recognized virus families may be clustered into 'supergroups' (Goldbach, 1986; Strauss & Strauss, 1988; Goldbach et al., 1991). This grouping is based on similarities in genome structure and expression, and on the preservation of conserved amino acid sequence motifs in viral proteins among different members. It implicates that viruses with RNA genomes are more or less related, and all evolved from a small number of common ancestor viruses (Strauss & Strauss, 1988).

At least seven supergroups of eukaryotic RNA viruses are presently recognized: the carmo-like, the corona-like, the flavi-like, the picorna-like, the sobemo-like, the negative-strand, and the double-strand virus supergroup. The supergrouping concept fits remarkably well with the other properties, previously used for virus taxonomy (Goldbach, 1986; Goldbach & Wellink, 1989; Strauss & Strauss, 1988; Gorbalenya et al., 1989; Candresse et al., 1990).

The most homogeneous supergroup is undoubtedly formed by the negative-strand viruses, which include the Rhabdoviridae, Filoviridae, Paramyxoviridae, Orthomyxoviridae, Bunyaviridae and Arenaviridae. All negative-strand viruses have single stranded RNA genomes, which are tigthly encapsidated with protein to nucleocapsids. The RNA is of negative polarity and after entrance in a cell, genomic RNA is transcribed into mRNAs by a viral RNA polymerase, present in the virus particles. The particles are enveloped and covered with surface projections. The majority of negative-strand viruses is transmitted by arthropod vectors, although other transmission strategies are also used. Besides these common general properties, all members of the separate families are characterized by their own distinct features. The main molecular properties of the different families of this supergroup, will be described in the next paragraphs.

14

2.2 The Mononegavirales: Rhabdoviridae, Filoviridae and Paramyxoviridae

2.2.1 Taxonomy

Rhabdo-, filo- and paramyxoviruses are characterized by unsegmented RNA genomes with great similarities in genetical organization and with considerable sequence homologies. Since they share many molecular biological properties, these virus families have recently been placed in a higher taxon by the ICTV, the order of the Mononegavirales.

The family of Rhabdoviridae includes members that can infect either animals or plants. More than a hundred viruses have been classified as members of this family on the basis of their characteristic baciliform or bullet-shape structure. Some of the animal rhabdoviruses can be divided into two genera, the vesiculoviruses (with vesicular stomatitis virus, VSV, as the prototype virus) and the lyssaviruses (with rabies virus, RV, as prototype). Others await further classification. The plant rhabdoviruses are poorly studied, which makes it almost impossible to classify them properly. There are however some arguments to divide the plant rhabdoviruses into the subgroup A (lettuce necrotic yellows virus, LNYV) and the subgroup B (potato yellow dwarf virus, PYDV), in analogy to their animal counterparts (Peters, 1981) (Table 2.1).

Members of the Paramyxoviridae are characterized by particles that are polymorphic in size and shape, ranging from spherical to filamentous. The present family includes the parainfluenza- (Newcastle disease virus, NDV), morbilli- (measles virus, MV) and pneumovirus (respiratory syncytial virus, RSV) genera (Kingsbury, 1985) (Table 1).

The Filoviridae, including Marburg and Ebola virus are characterized by having pleomorphic, U-shaped (often branched) virus particles. Sofar, both viruses are poorly studied, but they have many properties in common with the rhabdo- and paramyxoviruses.

The infection cycle of rhabdoviruses usually involves replication in an arthropod

15

Rhabdoviridae

    vesiculovirus: vesicular stomatitis virus (VSV), indiana, New Jersey, Chandipura etc.
                                          serotypes

    Lyssavirus: rabies virus (RV), Duvenhage, Lagos, Mokola etc. serotypes

    subgroup A: lettuce necrotic yellows virus (LNYV),
                                            broccoli necrotic yellows virus,
                                          sonchus virus,
                                          wheat striate mosaic virus.

    subgroup B: potato yellow dwarf virus (PYDV),
                                          sonchus yellow net virus (SYNV),
                                          eggplant mottled dwarf virus,
                                          sowthistle yellow vein virus.

           plus many unassigned members

Paramyxoviridae

    parainfluenzavirus: parainfluenza virus (PIV) type 1 to 6 (Sendai virus (SV) is the murine
                                          parainfluenzavirus type 1 and Newcastle disease virus (NDV) is the
                                          avian parainfluenzavirus type 1),
                                          mumps virus (MuV).

    morbillivirus: measles virus (MV),
                                              canine distemper virus (CDV),
                                          Rinderpest virus (RV),
                                          peste-des-petits-ruminants.

    pneumovirus: Respiratory syncytial virus (RSV),
                                          murine pneumonia virus (MPV).

Table 2.1: Classification of the Rhabdoviridae and Paramyxoviridae

vector (with the exception of lyssaviruses). Virus particles enter their host cells by receptor-mediated endocytosis and nucleocapsids are released in the cytoplasm after fusion of the virus-containing endosomes with lysosomes. The paramyxoviruses are all air-borne and viral nucleocapsids enter their host cells by direct fusion of viral envelopes with cellular membranes. Transcription and replication takes place in the cytoplasm of infected cells.

16

The best studied members from the Mononegavirales thusfar are, VSV and RV of the Rhabdoviridae, and NDV, Sendai virus (SV) and MV of the Paramyxoviridae (Banerjee, 1987; Vainionpaa et al., 1989).

2.2.2 Virus structure and genome organization

Virus particles of the mononegavirales are bounded by a lipid envelop, which is acquired by budding of the nucleocapsids through the cellular membrane or, in case of a number of plant rhabdoviruses, through the inner nuclear membrane. The viral envelopes are studded with glycoprotein spikes. Rhabdoviruses contain only one membrane glycoprotein, denoted G, which is involved in receptor-mediated cell attachment and membrane fusion. Parainfluenza viruses have two types of glycoproteins, a membrane-fusion protein (F) and an HN protein with both haemagglutinin and neuraminedase activity. Both glycoproteins are involved in receptor-binding and virus-release. The membrane protein of morbilliviruses, corresponding to HN, is denoted H, which has only haemagglutinin activity, whereas the second membrane protein of the pneumoviruses, denoted G, has neither of the two activities. In case of paramyxoviruses, cleavage of the surface glycoproteins (F, HN, H or G) by host cell proteases is necessary to yield activated surface proteins, essential for infectivity (Sato et al., 1988).

The viral nucleocapsids consist of one linear single stranded RNA molecule of 11-16 kilobases (kb) long, tightly encapsidated with nucleocapsid protein (N or NP). In addition, minor amounts of a large (L) protein, representing the functional polymerase subunit, and a highly phosphorylated protein, denoted NS (vesiculoviruses), M1 (lyssaviruses) or P (paramyxoviruses), are associated with the internal nucleocapsids. Together, both L and NS (M1 or P) form the active transcriptase-complex (Emerson & Yu, 1975; Banerjee, 1987).

The final structural component is the hydrophobic matrix protein (M or M2 for lyssaviruses), which bridges the internal nucleocapsid with the surface glycoprotein(s). It plays a crucial role in viral transcription, assembly of virus particles (Galinski et al., 1987; Vainionpaa et al., 1989) and cytopathogenesis (Blondel et al., 1990).

17

The genetical organization of the genomes have been determined by mapping the genes in transcription/ translation experiments, using UV-treated virus particles and by nucleotide sequence analysis. The genes in the genome of VSV are arranged in the order 3' l-N-NS-M-G-L 5', where l stands for the short non-coding leader sequence, which is the first and most abundant transcript. It is involved in the inhibition of the cellular RNA synthesis (the 'host shut off') (Eisemann Crone & Keene, 1989). Compared to the VSV genome, the RV genome contains one extra 'remnant' gene (rg) between the G and L· genes (Tordo et al., 1986). The only plant rhabdovirus from which sequence data are available (SYNV), contains one extra gene (sc4) between the NS and M genes, while the fish rhabdoviruses also have one additional gene (NV) between the G and L genes (Heaton et al., 1989). The function of the proteins encoded by these extra genes has remained unresolved sofar.

Paramyxoviruses have a similar genetic organization, but normally possess more genes. Pneumoviruses, for example, contain ten genes in their genomes. Strikingly, the membrane glycoprotein genes (G and F) of RSV are located on the genome in reversed order, compared to the other paramyxo- and rhabdoviruses. Moreover, pneumovirus RSV contains a second matrix protein of 22K and a third membrane associated protein (1a) of 7.5K. Its function is unknown, just as that of other 'non-structural' proteins (1b, 1c).

The genes on the genomic RNA of filo-, paramyxo- and rhabdoviruses are separated by gene junction sequences. A gene junction region consists of a polyadenylation signal, some non-transcribed nucleotides and a transcription start signal (Table 2.2).

Paramyxo- and rhabdoviral genes usually encode a single protein, with only one exception. The NS gene of VSV additionally encodes a small C protein from an overlapping reading frame. Another small 7K protein is translated from an internal start codon, located at the carboxy-terminal end of the NS ORF (Herman, 1986; Hudson et al., 1986). The corresponding P genes of paramyxoviruses also express multiple proteins. Usually, the P protein is encoded by normal templated mRNAs, whereas the mRNAs encoding other proteins contain non-templated G insertions in the middle of the gene (Fig.2.2). This mechanism, known as RNA editing (Thomas et al., 1988; Vidal et al.,

18
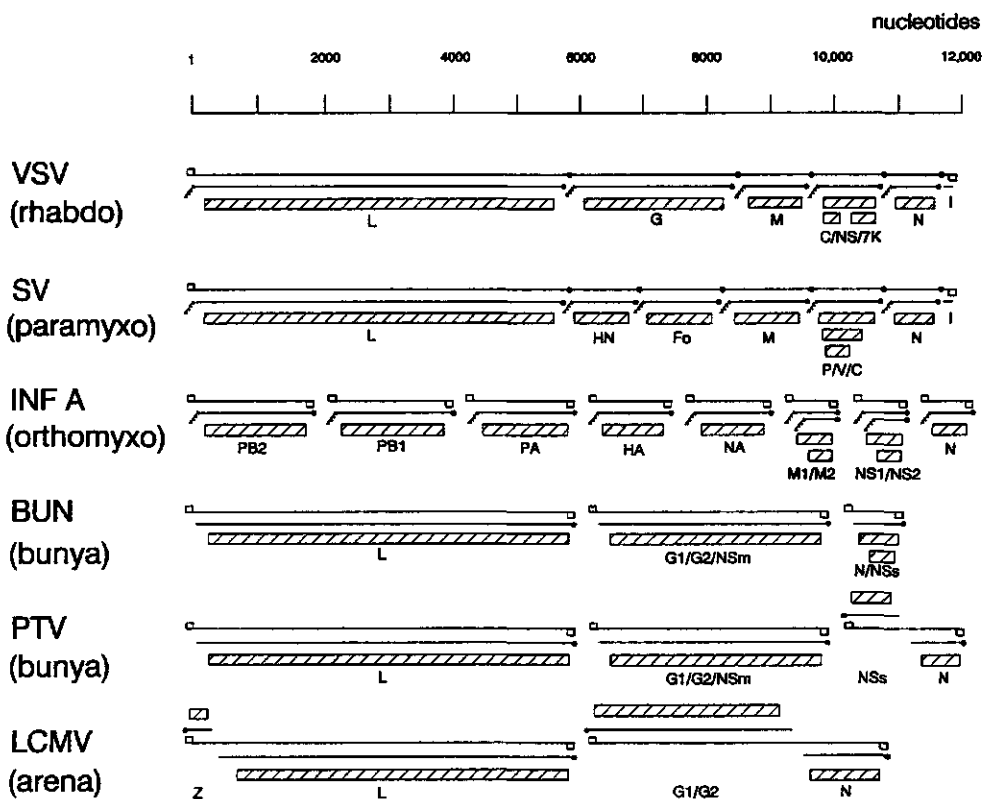
Fig.2.1: Schematical representation of the structure and expression of the genomes of representative negative-strand viruses. The lines represent RNA, the boxes represent protein. The gene junction regions in the unsegmented RNAs are indicated by circles. The cap structures at the 5' ends of the mRNAs are indicated by black circles. The terminal complementary sequences are represented by small boxes. Viral sense mRNAs are drawn above the genomic RNAs, the viral complementary sense mRNAs below. The genome segments are approximately drawn to scale

1990; Ohgimoto et al., 1990), most likely occurs by a stuttering mechanism of the viral polymerase (Vidal et al., 1990; Pelet et al., 1991). In addition, smaller C proteins are translated from overlapping open reading frames in case of MV, SV, PIV1, PIV3 and CDV (Bellini et al., 1985; Spriggs & Collins, 1986) (Fig. 2.2).

virus	termination signals	intergenic region	initiation signals
VSV	....UAUG(A_7)	CU	AACAG...UC.
RV	......UG(A_{7-8})	C(N_{1-422})	AACA...CU..
SNYV	....UAAG(A_5)	CC	AACA.......
SV	..A.UAAG(A_5)	CUU	AGGGU.AAAG.
PIV3	.AA.UA..(A_5)	CUU	AGGA..AAAG.
MuV	....$^{AA}_{UU}$AG(A_{6-7})	N_{1-7}	A^G_AG$^{CC}_{AA}$.....
MV	....UUAU(A_6)	CUU	AGGA.C.....
RSV	........(A_4)	(N_{0-51})U	GGGGAAAU...

Table 2.2: The intergenic regions and transcription signals of rhabdo- and paramyxoviruses.

2.2.3 Transcription and replication

After entrance of a paramyxo- or rhabdovirus in a host cell, the viral polymerase most likely enters its encapsidated RNA template at the extreme 3' end and sequentially synthesizes the short leader (l) RNA (which remains unencapsidated) followed by the mRNAs, by terminating and restarting at each gene junction region (Vidal & Kolakofski, 1989). At the first gene junction transcription terminates and since this junction sequence does not contain a polyadenylation signal, the l RNA remains unmodified. All mRNAs are immediately capped at the 5' end by the viral polymerase, while poly-A tails are added at the 3' ends by slippage of the polymerase at the U-rich polyadenylation signals (Horikami & Moyer, 1982; Sanchez et al., 1985). As a consequence of this sequential and polar mode of transcription, decreasing amounts of 3'-distal mRNAs are found in infected cells, which leads to relatively large amounts of N mRNA and low amounts of L mRNA (Fig.2.3). The amount of each mRNA is also regulated by the nucleotide sequences around the gene junctions. The RSV M2 and L genes are overlapping for 68 nucleotides. Due to this overlap, the transcription termination site of the M2 gene is located downstream of the initiation site of the L gene. This causes premature

20
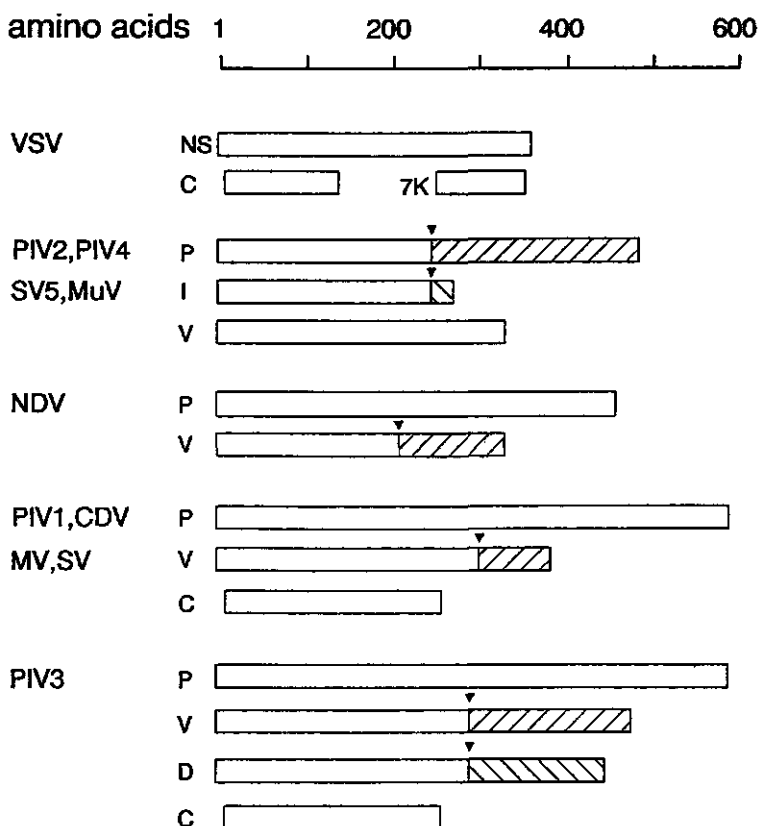
Fig.2.2: The NS and P gene products of rhabdo- and paramyxoviruses. The arrows correspond with the RNA editing sites. The arced regions are only expressed after introduction of one or more G nucleotides in the corresponding mRNA.

termination of most L transcripts, hence a very low level expression of viral polymerase is observed in infected cells (Collins et al., 1986). The unusually long 5' non-translated regions of the mRNAs encoding the F protein of morbilliviruses (CDV and RV) are GC-rich, contain multiple AUG translational start codons and are capable of folding into extensive secondary structures. It has been demonstrated that they play a strong regulatory role in the synthesis of F protein, an important determinant of viral pathogenicity (Evans et al., 1990).

At later stages of infection, when viral proteins are accumulating in infected cells,
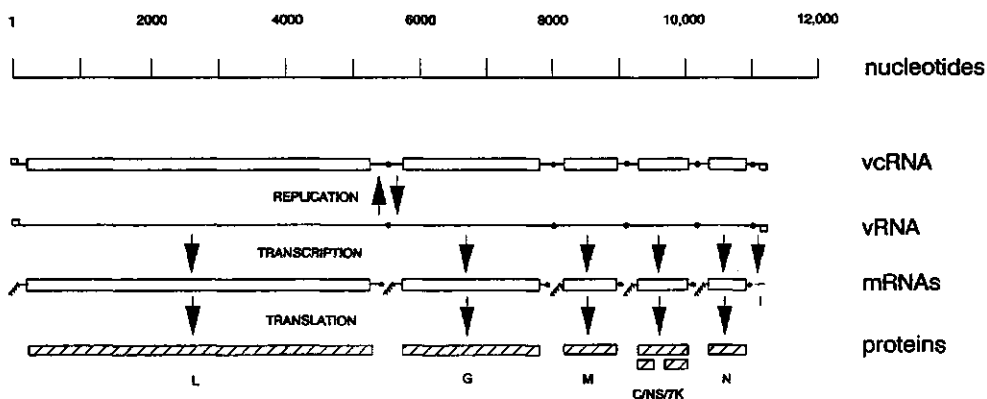
Fig. 2.3: Transcription and replication of rhabdo- and paramyxoviruses, examplified with vesicular stomatitis virus (VSV). The open bars represent ORFs, the arced bars correspond with proteins. vRNA, viral sense RNA; vcRNA, viral complementary sense RNA.

the polymerase switches to the replicative mode and unmodified, genome-length antigenomes are synthesized, which serve as intermediates in RNA replication. Both viral RNA and viral complementary RNA are immediately encapsidated with nucleocapsid protein, in contrast to the subgenomic mRNAs. The origin of encapsidation is located on the leader at the extreme 5' end of the RNA. The transcription to replication switch is most likely affected by the amount of intracellular, unassembled nucleocapsid protein (Vidal & Kolakofski, 1989). Encapsidation of the nascent RNA strand with N protein makes that the polymerase overrides the termination signal of the first gene junction and consequently, replicative intermediates are synthesized. However, polycistronic transcripts, spanning contiguous genes, have been described several times, demonstrating that there are considerable amounts of readthrough mRNAs present in infected cells (Moscona & Galinski, 1990). In some cases, the 'minimal' model for transcription and replication needs further extension to completely explain all observed events. In MV-infected cells for example, free leader sequences are almost completely absent (Chan et

22

al., 1989). Leader-containing subgenomic RNAs (l-N, l-N-P and even l-N-P-M) are present, but always encapsidated with nucleocapsid protein (Castaneda & Wong, 1990). This demonstrates that at least MV uses alternative regulatory mechanisms for transcription and replication.

The NS (M1) or P proteins play a dual regulatory role in viral transcription and replication. Their interaction with L proteins and nucleocapsids enables binding of the polymerase to the encapsidated template RNA, while its specific binding to free N (NP) protein prohibits its precipitation or interaction with non-specific RNAs. The nature of the association between L and NS (M1) or P has remained unknown sofar (Paul et al., 1988; Ryan & Portner, 1990).

2.3 Orthomyxoviridae

2.3.1 Taxonomy

The family Orthomyxoviridae consists of the single genus influenzavirus, which is divided into three types of viruses (A, B and C), based on serologically distinct nucleocapsid (NP) and matrix (M) proteins. The viruses are further classified into a considerable number of antigenic subtypes, based on the immunological properties of their specific surface glycoproteins. Influenzaviruses are responsible for severe epidemic outbreaks in man, mammals and birds, due to the extraordinary high rate of mutation (antigenic drift) and recombination (antigenic shift) in their genomes.

Besides these air-borne influenzaviruses, a group of tick-transmitted viruses (Thogoto and Dhori viruses) should also be included into this virus family (Clerx et al., 1983).

2.3.2 Virus structure and genome organization

The morphology of the influenzaviruses resembles that of the paramyxoviruses. Besides, the surface glycoproteins exhibit similar biological properties. Influenzaviruses

23

A and B contain two different types of surface proteins, haemagglutinins (HA) and neuraminedases (ND), which are responsible for receptor binding, membrane fusion and release of the virions from the cellular membrane. This indicates that ortho- and paramyxoviruses are evolutionary related. These are the major reasons why these viruses were in the past classified as one, undivided virus family, the Myxoviridae.

In addition to having segmented genomes, the Orthomyxoviridae differ from the Paramyxoviridae in a number of other properties. Influenzaviruses enter their host cells by receptor-mediated endocytosis and viral nucleocapsids are released in the cytoplasm by membrane-fusion, after acidification of the virus-containing endosomes. Cell-attachment and membrane-fusion are both mediated by the HA protein, while the receptor-destroying, neuraminedase activity of the NA is responsible for the release of progeny virions from the cellular membrane. In contrast, paramyxoviruses have both haemagglutinin and neuraminedase activity joined in one HN glycoprotein. Paramyxo-viruses enter their host cells by fusion with cellular membranes, mediated by their fusion (F) proteins. Although the F proteins are structurally related to influenzavirus HA proteins, they are not involved in adsorption of virions to the host membranes.

The most striking difference, however, is the localization in infected cells. When released in the cytoplasm, influenzavirus nucleocapsids directly migrate to the nucleus, where transcription, replication and assembly of progeny nucleocapsids take place.

The genome of the Orthomyxoviridae consists of six (Dhori and Thogoto), seven (influenza C) or eight (influenza A and B) linear single-stranded RNA molecules. In virions these genomic RNAs occur in a circular conformation, due to base-pairing of the termini that form panhandles of approximately 16 nucleotides long (Hsu et al., 1987). The morphology, genome structure and coding capacity of influenzaviruses have recently been reviewed extensively by Lamb (1989) and Krug et al. (1989) (Fig.2.1).

The three largest RNA segments denoted one, two and three, encode the three polymerase subunits PB2, PB1 and PA, respectively. These P proteins are present as complexes in infected cells (Detjen et al., 1987) and represent the equivalents of the L proteins of the other negative-strand viruses.

RNA segment four encodes the haemagglutinin (HA), located on the viral envelope

24

as trimers. After translation and translocation on the endoplasmatic reticulum membrane, the HA precursor molecules are cleaved by a cellular trypsin-like protease to yield activated HA1- HA2 dimers. Influenza C viruses have only one single type of membrane glycoprotein, with both haemagglutinin and neuraminedase activity (Vlasak et al., 1987), denoted HEF (standing for Haemagglutinin Esterase Fusion protein) (Herrler et al., 1988).

RNA segment five encodes the nucleocapsid protein (NP of 56K) and since this sequence contains a nuclear accumulation signal, NP accumulates in the nucleus, where viral nucleocapsids are assembled (Davey et al., 1985).

RNA segment six of the influenza A and B viruses encodes the viral neuraminedase (NA), which is not further processed and occurs as mushroom-shaped tetramers on the viral envelope. Influenza B viruses encode another protein (NB) from a second reading frame on the same mRNA. This small membrane-bound glycoprotein is found in large amounts on plasma membranes, but not on the viral envelopes. The role of this third surface protein in influenza B virus infections still awaits to be determined (Shaw et al., 1983).

RNA segment seven (A and B) or six (C) encodes two proteins. In influenza A infected cells, the viral matrix protein (M1) and a non-glycosylated M2 protein have been detected. M1 is expressed from a colinear mRNA molecule, whereas M2 is translated from a second mRNA, derived from the colinear transcript by splicing. Splicing occurs by the host nuclear splicing machinery. M2 is abundantly present on the cellular membrane, but only in very small amounts in virions. This protein might be involved in virion maturation (Zebedee & Lamb, 1988) (Fig.2.4). Strikingly, in case of influenza C viruses, the spliced mRNA encodes the structural M protein (Yamashita et al., 1988). In addition, influenza B viruses use an alternative strategy to express a second overlapping ORF. A coupled translational stop and start mechanism leads to the synthesis of M2 with a size of 12-15K (Fig.2.4). Influenza B M2 is a cytoplasmatic protein, with an unknown function (Horvath et al., 1990).

The smallest RNA segment of influenza A, B (eight) and C (seven) encodes two non-structural proteins NS1 (26K) and NS2 (14K). The latter protein is also expressed

SEGMENT 7   SEGMENT 8

vRNA

mRNA

PROTEINS

M1 (28K)   NS1 (26K)
M2 (11K)   NS2 (14K)
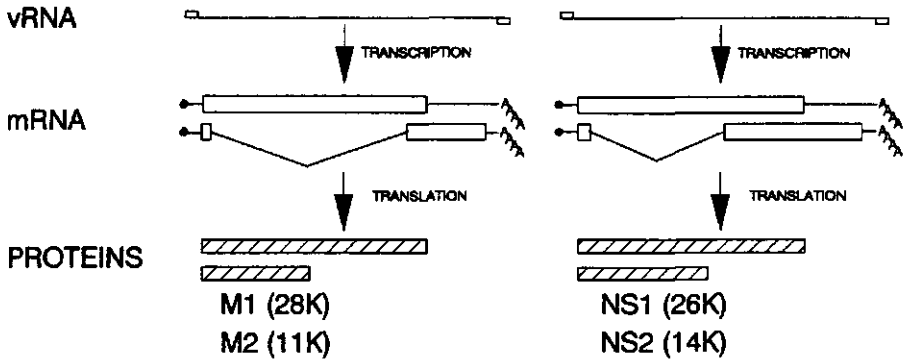
Fig.2.4: Organization of RNA segments 7 and 8 of influenza A and B viruses, which code for the matrix proteins and non-structural proteins from unspliced (M1, NS1) and spliced (M2, NS2) mRNAs. For explanation of symbols see Fig. 2.1 and Fig. 2.3.

from a mRNA molecule derived by splicing (Fig.4). Both proteins accumulate in the nucleus of infected cells and are essential for virus multiplication but their functions have thusfar remained unknown.

2.3.3 Transcription and replication

After entrance of viral nucleocapsids in the cytoplasm and subsequent transport to the nucleus, primary transcription takes place by the polymerase subunits PA, PB1 and PB2 (Ishihama & Nagata, 1988). In contrast to the polymerases of the Mononegavirales, those of the segmented negative-strand viruses do not possess capping activity. In order to obtain 5'-capped mRNAs, transcription is initiated by capped primers, derived from nascent host cell mRNAs (a mechanism known as 'cap snatching'). Capped primers are generated by cleavage of a host mRNA behind an A residue 12-14 nucleotides downstream of the 5' cap, most likely by the PB2 subunit. Transcription is initiated by

26

addition of a G residue to the primer, complementary to the second C residue at the 3' end of the viral RNA strand (Krug et al., 1989). Chain elongation is performed by the PB1 subunit and proceeds up to 17-22 nucleotides from the 5' end of the viral RNA, where a short stretch of U residues serve as polyadenylation signal (Braam et al., 1983). Poly-A tails are added to the 3' ends of the mRNAs by slippage of the polymerase. Using purified influenza polymerase and in vitro RNA transcripts as templates (Parvin et al., 1989), it has been demonstrated that approximately 25 3'-terminal nucleotides are essential for template recognition by the viral polymerase. The precise mechanism by which transcription is performed by the three different subunits is not completely understood thusfar. It is clear however, that splicing and cap-snatching makes the replication of influenzaviruses dependent of the host nuclear transcription machinery.

The synthesized mRNAs are transported into the cytoplasm where translation takes place. Influenzavirus gene expression is regulated mainly at the transcriptonal level, but the mechanisms by which this occurs, are not completely elucidated yet (Shapiro et al., 1987). As with the non-segmented negative-strand viruses, the amount of free NP protein in the nucleus discriminates between transcription and replication (Beaton & Krug, 1986; Shapiro & Krug, 1988). Later during infection, when NP starts to accumulate in the nucleus, the viral polymerase-complex is switched to the replicative mode. Viral complementary RNA molecules are synthesized, which are exact copies of the viral RNAs. These antigenomic RNA molecules are immediately encapsidated with NP protein, which indicates that the origins of encapsidation are most likely located at the 5' ends of the RNA molecules. Encapsidation by NP is essential for elongation of anti-genomic RNA synthesis but not for synthesis of cap endonuclease-primed mRNAs (Krug et al., 1989). The capped primers at the 5' ends of the mRNAs most likely interfere with encapsidation of the nascent RNA strands by NP.

The antigenomic RNAs remain in the nucleus and serve as intermediates in replication, while the nascent encapsidated viral RNA molecules are transported out of the nucleus as nucleocapsid structures to yield progeny virions (Shapiro et al., 1987).

27

2.4 Arenaviridae

2.4.1 Taxonomy

The Arenaviridae represent a relatively small negative-strand virus family and are characterized by a bipartite genome. On the basis of serology, approximately 11 different viruses have been placed into 2 groups within this family; the 'new world' viruses (with Pichinde virus as the prototype virus) and the 'old world' species (with lymphocytic choriomeningitis virus as prototype). Arenaviruses usually have rodent hosts and are transmitted by aerosolized excretions to a variety of other mammals, including humans. The only exception is Tacaribe virus, a new world species, which has been isolated from fruit-eating bats (Johnson, 1985).

2.4.2 Virus structure and genome organization

Arenavirus particles are extremely pleomorphic, measuring 50-300 nm in diameter. The lipid envelopes are studded with surface projections, which consist of one or two different glycoproteins, denoted GP or respectively G1, G2. The internal components include two different nucleocapsid structures with minor amounts of L protein (250K). In addition, a small zinc-binding protein, denoted Z or P_{11} of 10K has been detected in virus particles (Vezza et al., 1978; Salvato & Shimomaye, 1989; Iapalucci et al., 1989b). L and Z (P_{11}) most likely represent the viral polymerase components (Fig.2.1). The characteristic morphology may be explained by the fact that a number of host ribosomes, RNAs and enzymes are also enclosed in virus particles (Pedersen, 1979).

Arenaviruses enter their host cells by direct fusion with the plasma membrane, while subsequent transcription and replication take place in the cytoplasm of infected cells. Progeny virus particles are released by budding from the plasma membrane (Johnson, 1985).

The viral genome consists of two species of linear single stranded RNA, denoted S (small) RNA (3.4 kb) and L (large) RNA (7.2 kb) (Fig.2.1 and 2.5). The genomic RNA

28

segments contain complementary ends, which most likely serve as recognition signals for the viral polymerase and which are involved in the formation of circular nucleocapsids (Palmer et al., 1977). Both RNAs have an ambisense gene arrangement. The S RNA encodes the major structural proteins, the membrane glycoprotein(s) from a viral sense gene and the nucleocapsid protein (NP of 62-72K) from a viral complementary sense gene (Auperin et al., 1984; Romanowski et al., 1985) (Fig.2.5). For those members with two membrane glycoproteins, a GP precursor is glycosylated, transported to the dictyosomes and subsequently cleaved to yield G1 and G2. The cleavage mechanism is sofar unknown (Buchmeier et al., 1978; Wright et al., 1990).

The L RNA encodes the minor structural proteins, the Z or P_{11} proteins from a viral sense gene and the L protein from a viral complementary sense gene (Salvato et al., 1989; Iapalucci et al., 1989a; Salvato & Shimomaye, 1989) (Fig.2.5).

All arenavirus genes are expressed by subgenomic mRNA species. The non-encapsidated mRNAs possess cap structures and 1-5 additional nontemplated nucleotides at their 5' ends, that have been obtained by cap-snatching as found for influenzaviruses (Raju et al., 1990). The intergenic regions are rich in G and C nucleotides. Mapping of the 3' ends of the mRNAs has revealed that 3'-terminal secondary structures of the transcripts most likely signal transcription termination, rather than particular sequences at the intercistronic regions. The mRNAs do not contain poly-A tails at their 3' ends. It is tempting to speculate that the viral N protein, which tightly binds viral and viral complementary RNA, functions as antiterminator by preventing the folding of the RNA (Auperin et al., 1984; Iapalucci et al., 1991).

2.4.3 Transcription and replication

The intracellular molecular events during arenavirus infection, are thusfar poorly studied. Time course experiments in the presence or absence of inhibitors of protein synthesis have demonstrated that, ongoing viral protein synthesis is required for synthesis of the viral sense mRNAs (Franze-Fernandez et al., 1987; Fuller-Pace & Southern, 1988). This indicates that, in analogy to the orthomyxo-, paramyxo- and rhabdoviruses, NP is
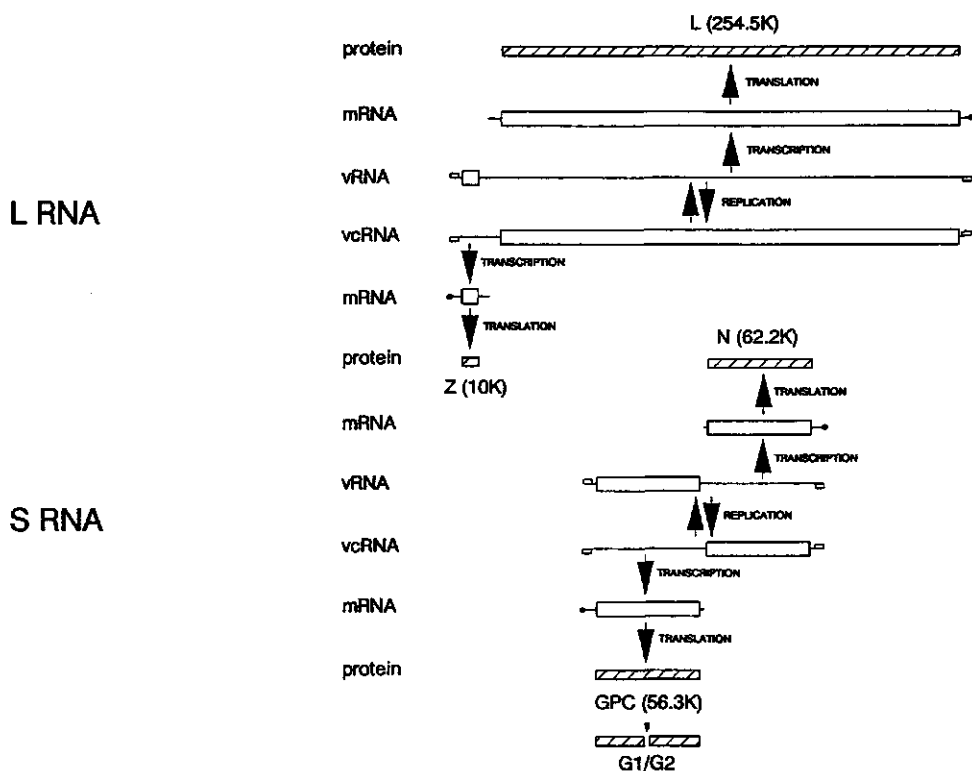
Fig.2.5: Ambisense gene arrangement of arenavirus (LCMV) S and L RNA. The mature membrane glycoproteins are cleaved from a precursor protein (GPC). For explanation of symbols see Fig. 2.1 and Fig. 2.3.

involved in the switch from transcription to replication and (during replication) in anti-termination at the intergenic regions by the viral polymerase. An ambisense coding strategy implies that all genes are expressed under control of distinct promoters, and furthermore, enables a time regulated gene expression. The NP and L proteins are the first viral proteins that accumulate in infected cells, while the 'late' viral sense mRNAs, encoding the membrane glycoprotein(s) and the Z or P_{11} proteins are synthesized after RNA replication (Franze-Fernandez et al., 1987). It is unclear, why both components of the presumed active polymerase are encoded in an ambisense gene arrangement (Fig.2.5).

2.5 Bunyaviridae

2.5.1 Taxonomy

The family Bunyaviridae is a very large and heterogeneous virus family. On the basis of serology and some limited biochemical analyses, more than 200 different species belonging to this virus family, have been separated into five genera, the bunyaviruses, phleboviruses, uukuviruses, hantaviruses and nairoviruses. The genera bunyavirus and phlebovirus consist of numerous serogroups, based on the antigenic properties of the more conserved internal nucleocapsid proteins (Bishop & Shope, 1979; Travasso da Rossa et al., 1983). Each serogroup contains serotypes or species, which are distinguished by the antigenic properties of the more variable surface glycoproteins (Bishop et al., 1980). Since the uukuviruses and phleboviruses exhibit considerable sequence homologies, it has been proposed to classify both genera within one subfamily, the Phlebovirinae (Bishop, 1985).

Except for hantaviruses, the Bunyaviridae are arthropod-borne and many warm-blooded vertebrates have been shown to be hosts. Some members can infect the central nervous system, others cause haemorrhagic fevers in susceptable hosts (Karabatsos, 1985). Hantaviruses are not transmitted by arthropod vectors. They cause symptomless persistent infections in rodents, from which they are spread by aerosols and urine (Schmaljohn et al., 1986).

2.5.2 Virus structure and genome organization

Bunyaviruses are spherically shaped and approximately 90-120 nm in diameter. The viral envelopes are covered with surface projections, which consist of two glycoproteins, denoted G1 and G2. Virus particles contain nucleocapsids, which consist of nucleocapsid protein (N), a few copies of polymerase (L) and genomic RNA (Bishop et al., 1980). Conform influenzaviruses and arenaviruses, bunyavirus nucleocapsids have a circular structure, due to base-pairing of the terminal RNA sequences (Petterson & Von

31

Bonsdorff, 1975; Raju & Kolakofski, 1989). Each nucleocapsid contains one copy of genomic RNA, denoted S (small), M (medium) or L (large) RNA. Considerable nucleotide sequence information is available of at least the type members of four genera. The data obtained demonstrate that members of the Bunyaviridae use rather different coding strategies (Fig.2.1).

Members of the genus bunyavirus have a simple negative-stranded coding arrangement for all three genomic RNA species. The S RNA segment codes for the viral nucleocapsid protein (N, 26-27K) and a small non-structural protein (NSs, 10.4-11K) by a single viral complementary mRNA species. Both proteins are translated from overlapping reading frames (Bishop et al., 1982; Akashi & Bishop, 1983; Cabradilla et al., 1983; Akashi et al., 1984; Gerbaud et al., 1987; Elliott & McGregor, 1989; Elliott, 1989) (Fig.2.6).
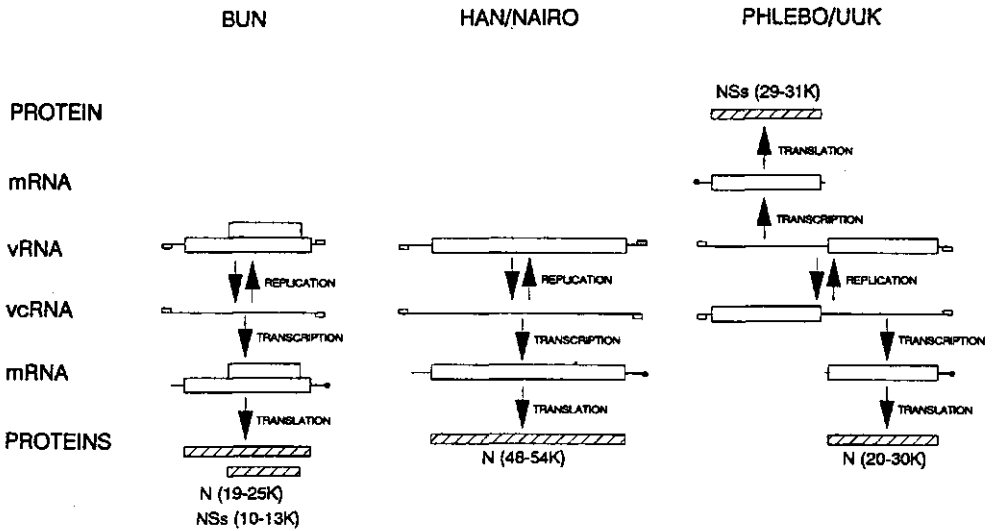
Fig.2.6: Structure and expression of the S RNA segments of members of the Bunyaviridae. For explanation of symbols see Fig. 2.1 and Fig. 2.3.

32

The M RNA segment encodes a large primary translation product (162.6K) from a viral complementary mRNA, which is approximately 100 nucleotides shorter than genome-length (Eshita & Bishop, 1984; Lees et al., 1986; Grady et al., 1987; Pardigon et al., 1988). The primary translation product is co-translationally processed on endoplasmatic reticulum membranes to yield the mature viral glycoproteins (G1 of 108-120K, G2 of 29-41K) and a non-structural protein, NSm of 15-18K. The order of the proteins translated from the large open reading frame is 5' G2-NSm-G1 3' (Fazakerley et al., 1988) (Fig.2.7).

The L RNA encodes the viral transcriptase (L, 259K) which is translated from a viral complementary mRNA species (Elliott, 1989).

Hantaviruses have an identical, but even more simple coding arrangement. The S RNA only encodes the relatively large nucleocapsid protein (N, 49K) (Schmaljohn et al., 1986; Stohwasser et al., 1990; Parrington & Yong Kang, 1990), while the M RNA codes for both membrane glycoproteins, in the order 5' G1-G2 3' (Schmaljohn et al., 1987; Yoo & Kang, 1987; Giebel et al., 1989) (Fig.2.6 and 2.7). The L RNA codes for the viral polymerase (L, 246K) (Schmaljohn, 1990).

The nairovirus S RNA segment is of negative polarity and only encodes the nucleocapsid protein of 48-54K (Ward et al., 1990). Sofar no sequence information is available for the other genomic RNAs of nairoviruses. The length of the M RNA segment (4.6 Kb), compared to the molecular weight of both membrane glycoproteins (G1, 71-84K; G2, 30-40K), predicts that nairovirus M RNA is simply of negative polarity (Clerx et al., 1981). Strikingly, the L RNA is extremely long, approximately 13 kb (Clerx & Bishop, 1981) and for this reason it would be very interesting to unravel the coding properties of this genomic RNA segment.

Phleboviruses have an ambisense S RNA segment, an arrangement similar to that found in both RNAs of the Arenaviridae. The nucleocapsid protein (N, 25-27K) is encoded in viral complementary sense and a non-structural protein (NSs, 29-30K) in viral sense. Both proteins are expressed by subgenomic mRNA species, which terminate in a secondary structure located at the intergenic region of the RNA (Ihara et al., 1984; Marriott et al., 1989) (Fig.2.6).

33

The M RNA segment encodes the precursor protein for the two membrane glycoproteins and a non-structural protein (NSm, 14-30K) in viral complementary sense. The gene product order in the glycoprotein precursor is 5' NSm-G2-G1 3', which is slightly different from that of members of the genus bunyavirus (Ihara et al., 1985; Collett et al., 1985) (Fig.2.7). Rift Valley fever virus M RNA encodes an additional non-structural protein of 78K (Kakach et al., 1988). The biogenesis of these four M RNA-encoded proteins is remarkably complex and remains to be completely elucidated (Suzich et al., 1990).

The structure and genetic organization of the S RNA segment of uukuvirus is similar to that of the S RNA of phleboviruses, including the ambisense gene arrangement (Simons et al., 1990). Uukuvirus M RNA however does not encode a non-structural protein, but only the precursor protein for the envelop glycoproteins in order 5' G1-G2 3' (Ronnholm & Petterson, 1987) (Fig.2.6 and 2.7).

2.5.3 Transcription and replication

Viral nucleocapsids are released in the cytoplasm of host cells after adsorption of virions and entrance by endocytosis (Vezza et al., 1979). Transcription and replication occur in the cytoplasm of infected cells. Like influenza- and arenaviruses, transcription is primed by host derived, capped RNA sequences of 10-18 nucleotides in length (cap-snatching) (Bishop et al., 1983; Patterson & Kolakofski, 1984; Eshita et al., 1985; Collett, 1986; Bouloy et al., 1990). The synthesized mRNAs are approximately 60-120 nucleotides shorter than the corresponding viral RNAs. Like arenaviruses, bunyavirus mRNAs are not polyadenylated at their 3' ends. It is not known what signals transcription termination. The secondary structures located in the intergenic regions of the ambisense S RNA segments of phlebo- and uukuviruses may act as such (Emery & Bishop, 1987). Although relatively poorly studied, ongoing viral protein synthesis is essential for RNA replication (Ihara et al., 1985). This implies that, like for other negative-strand viruses, the amount of intracellular free nucleocapsid protein regulates the switch from transcription to replication. It has been reported for La Crosse virus (genus bunyavirus),

34

BUN

G2 (32K) NSm (11K)   G1 (109K)

HAN

G1 (64K)   G2 (54K)

UUK

G1 (55K)   G2 (55K)

PHLEBO

NSm (30K)   G1 (61K)   G2 (55K)
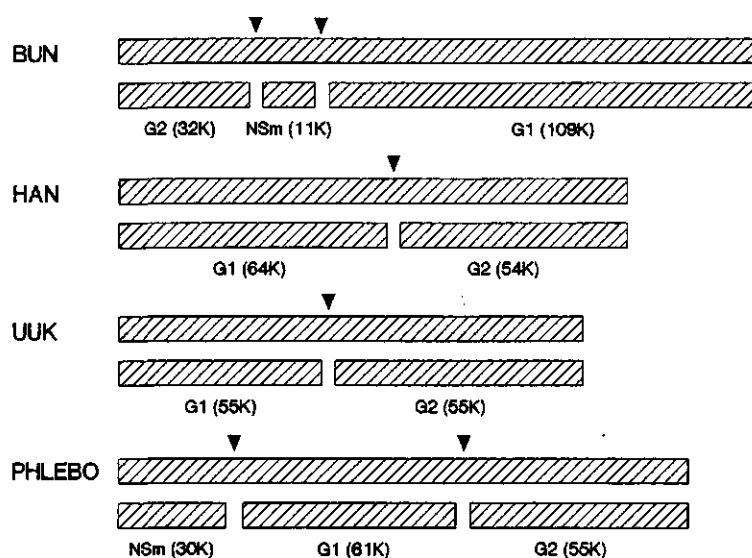
Fig.2.7: Translation products of the M RNA segments of members of the Bunyaviridae. The processing sites within the primary translation products are indicated by arrowheads.

that ongoing protein synthesis is also required to prevent premature termination of transcription. Loading of nascent mRNAs by host ribosomes seems to be necessary, due to the presence of certain, unknown cellular factors (Bellocq et al., 1987). This translational requirement however, is 'cell-type' dependent and, for example, does not occur during infection of mosquito cells (Raju et al., 1989).

The S and L RNA specific mRNAs are most likely translated in the cytoplasm, whereas the M RNA specific mRNAs are translated on membranes of the endoplasmatic reticulum. The M RNA-encoded primary gene products are co-translationally processed and translocated, to yield mature membrane glycoproteins (Ulmanen et al., 1981; Kakach et al., 1988). The mature glycoproteins are subsequently transported to dictyosomes and

35

virus particles are formed by budding of nucleocapsids into Golgi vesicles. Progeny virus particles accumulate in the dictyosomes, from where they are transported out of the cell (Murphy et al., 1968; Bishop & Shope, 1979).

2.6 Negative-strand viruses: variations on a few themes

The previous paragraphs have demonstrated that negative-strand viruses share many features. Molecular studies on their genome structure and coding properties have revealed that they form a relatively homogeneous supergroup of virus families. They share the following properties:

Morphology and protein composition:

Virus particles contain lipid envelopes covered with surface projections, which consist of clusters of one or two membrane proteins. These projections play an important role in cell attachment, maturation of the virus particles, their release from host cells and they are the main determinants of virus transmission and disease development (Smith & Palese, 1989; Shope et al., 1981). The genomes consist of single stranded RNA, complexed with protein to form stable nucleocapsids, which have either an amorphous or distinctly helical structure, when visualized by electron microscopy. Matrix proteins are linking the internal nucleocapsids to the envelopes of the virus particles. Although Bunyaviridae and Arenaviridae lack distinct matrix proteins, one of the membrane glyco-proteins may have taken over this function.

Assembly and transmission:

Virus particles enter susceptible host cells by membrane-fusion, endocytosis or by mechanical damage (plant viruses). Nascent virus particles maturate by budding of nucleocapsids from the inner nuclear membrane, the endoplasmatic reticulum or directly

36

from the plasma membrane. Released particles are transmitted by either arthropod vectors, vertebrates or even by air, to new hosts. Plant viruses are not released from the host cells, but accumulate intracellularly and are acquired by their arthropod vectors during ingestion of food from plants.

Genome structure:

Negative-strand viruses can have either a monopartite or a segmented single stranded RNA genome, which is tigthly wrapped with nucleocapsid protein to stable nucleocapsids. In these structures the sugar-phosphate backbone of the RNA is protected against ribonuclease activity, while the base-pairing capacity remains unaffected. As far as analyzed, the genomic RNAs possess pyrophospate groups at their 5' termini and free hydroxyl groups at their 3' ends. The sequences at the 3' and 5' ends are complementary and serve as recognition sites for the viral RNA polymerase. The segmented negative-strand viruses have circular nucleocapsids, due to base-pairing of the terminal nucleotides. The arrangement of the genes in the genomes among representative members shows to a lesser or greater extend resemblance to that of VSV, the prototype genome of the supergroup (Fig.2.1).

Transcription and replication:

The genomic RNA is transcribed into mRNAs by a viral polymerase or polymerase-complex, brought along with the virus particle. The nascent mRNAs subsequently encode the proteins involved in virus multiplication. RNA replication is a process, clearly distinct from transcription and takes place at later stages in the course of infection. Both processes are regulated by the intracellular free N (NP) concentration, which is the central regulatory event in the multiplication cycle of negative-strand viruses.

The common properties summarized above suggest that the negative-strand viruses are all evolutionary interrelated. Based on molecular and epidemiological data, it seems

37

likely that the ancestral virus was an arthropod virus with an unsegmented RNA genome. Indeed, a major part of the negative-strand viruses is able to replicate in their arthropod vectors. Since virus spread obviously follows the host feeding behaviour, virus evolution is narrowly associated with that of its host. This resulted in adaptation of viruses to all kinds of animals and plants. Some viruses even became independent of their vectors for their transmission (arena-, influenza- and paramyxoviruses, rabies virus and hantaan virus).

The two driving forces for molecular evolution are mutation and recombination. The absence of proofreading activity of RNA-dependent RNA polymerases causes point mutation frequencies of 10^{-3} /10^{-5}. This will lead to a rapid divergence of sequence information, when selective pressures alter (Holland et al., 1982; Reanny, 1982) and may be an explanation why the amino acid homologies between related viral proteins are often virtually absent, or only preserved in functional domains (Kamer & Argos, 1984; Argos, 1988; Hodgman, 1988). The other mechanism to obtain variability, is recombination of genetic information. Several models have been developed to explain this phenomenon. A 'polymerase jumping' model, as outlined by several authors (Perrault, 1981; Lazzarini et al., 1981; Cascone et al., 1990) is an elegant model to explain the mechanism of rearrangement of RNA sequences. This model, based on copy choice, where the viral polymerase changes from template during elongation at a certain frequency, may also explain the formation of defective interfering (DI) particles, segmented genomes or even ambisense RNAs. In addition, extension of RNA genomes by duplication and subsequent mutation or by introduction of 'host' RNA sequences, may also be explained by this model (For example, the spliced genes in influenzavirus genomes or the additional genes in some rhabdo- or paramyxovirus genomes, compared to VSV). Reassortment of RNA segments, as observed for influenza- and bunyaviruses (antigenic shift), hereby, is an extra mechanism leading to viral variation.

Our knowledge on the molecular biology of negative-strand viruses has rapidly expanded over the last few years. Research on this group of viruses is rather intense, since many members cause severe diseases in animals, man and plants. In the near future many questions concerning their replication strategies will undoubtly be addressed. The

recent development of in vitro systems that yield infectious negative-strand virus particles from cloned cDNA, opens the way to detailed analysis of gene functions after site-directed mutagenesis, tools previously applicable only to positive strand viruses and retroviruses (Mirakhur & Peluso, 1988; Luytjes et al., 1989; Ballart et al., 1990).

REFERENCES

Akashi, H. & Bishop, D.H.L. (1983). Comparison of the sequences and coding of La Crosse and snowshoe hare bunyavirus S RNA species. J. Virol. 45: 1155-1158.

Akashi, H., Gay, M., Ihara, T. & Bishop, D.H.L. (1984). Localized conserved regions of the S RNA gene products of bunyaviruses are revealed by sequence analyses of the Simbu serogroup Aino virus. Virus Res. 1: 51-63.

Auperin, D.D., Romanowski, V., Galinski, M. & Bishop, D.H.L. (1984). Sequencing studies of Pichinde arenavirus S RNA indicate a novel coding strategy, an ambisense viral S RNA. J. Virol. 52: 897-904.

Ballart, I., Eschle, D., Cattaneo, R., Schmid, A., Metzler, M., Chan, J., Pifko-Hirst, S., Udem, S.A. & Billeter, M.A. (1990). Infectious measles virus from cloned cDNA. EMBO J. 9: 379-384.

Banerjee, A.K. (1987). Transcription and replication of rhabdoviruses. Microbiol. Rev. 51: 66-87.

Beaton, A.R. & Krug, R.M. (1986). Transcription antitermination during influenza virus template RNA synthesis requires the nucleocapsid protein and the absence of a 5' capped end. Proc. Nat. Acad. Sci. U.S.A. 83: 6282-6286.

Bellini, W.J., Englund, G., Rozenblatt, S., Arnheiter, H. & Richardson, C.D. (1985). Measles virus P gene codes for two proteins. J. Virol. 53: 908-919.

Bellocq, C. & Kolakofski, D. (1987). Translational requirement for La Crosse virus S-mRNA synthesis: a possible mechanism. J. Virol. 61: 3960-3967.

Bishop, D.H.L. & Shope, R.E. (1979). Bunyaviridae. In: Comprehensive Virology, vol. 14. pp. 1-156. Edited by H. Fraenkel-Conrat & R.R. Wagner. New York: Plenum Press.

Bishop, D.H.L., Calisher, C., Casals, J., Chumakov, M.P., Gaidamovich, S.Y.A., Hannoun, C., Lvov, D.K., Marshall, I.D., Oker-Blom, N., Petterson, R.F., Porterfield, J.S., Russell, P.K., Shope, R.E. & Westaway, E.G. (1980). Bunyaviridae. Intervirology 14: 125-143.

Bishop, D.H.L., Gould, K.G., Akashi, H. & Clerx-Van Haaster, C.M. (1982). The complete sequence and coding content of snowshoe hare bunyavirus small (S) viral RNA species. Nucl. Ac. Res. 10: 3703-3713.

Bishop, D.H.L., Gay, M.E. & Matsuoka, Y. (1983). Nonviral heterogeneous sequences are present at the 5' end of one species of snowshoe hare bunyavirus S complementary RNA. Nucl. Ac. Res. 11: 6409-6418.

Bishop, D.H.L. (1985). Replication of arenaviruses and bunyaviruses. In: Virology, pp. 1083-1110. Edited by B.N. Fields, R.M. Chanock, D.M. Knipe, J.L. Melnick, B. Roizman & R.E. Shope. New York: Raven Press.

Blondel, D., Harmison, G.G. & Schubert, M. (1990). Role of matrix protein in cytopathogenesis of vesicular stomatitis virus. J. Virol. 64: 1716-1725.

Bouloy, M., Pardigon, N., Vialat, P., Gerbaud, S. & Girard, M. (1990). Characterization of the 5'- and 3'-ends of viral mRNAs isolated from BHK21 cells infected with Germiston virus (Bunyavirus). Virology 175: 50-58.

Braam, J., Ulmanen, I. & Krug, R.M. (1983). Molecular model of a eukaryotic transcription complex: Functions and movements of influenza P proteins during capped RNA-primed transcription. Cell 34: 609-618.

Buchmeier, M.J., Elder, J.H. & Oldstone, M.B.A. (1978). Protein structure of lymphocytic choriomeningitis virus: identification of the virus structural and cell-associated polypeptides. Virology 89: 133-145.

Cabradilla, C.D., Holloway, B.P. & Obeijeski, J.F. (1983). Molecular cloning and sequencing of the La Crosse virus S RNA. Virology 128: 463-468.

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

Cascone, P.J., Carpenter, C.D., Li, X.H. & Simon, A.E. (1990). Recombination between satellite RNAs of turnip crinkle virus. EMBO J. 9: 1709-1715.

Castaneda, S.J. & Wong, T.C. (1990). Leader sequence distinguishes between translatable and encapsidated measles virus RNAs. J. Virol. 64: 222-230.

Chan, J. Pifko-Hirst, S. & Udem, S.A. (1989). Where is the measles virus (+) strand leader RNA transcript? In: Genetics and Pathenogenicity of Negative-strand viruses. pp. 222-231. Edited by D. Kolakofski & B.W.J. Mahy. Amsterdam: Elsevier Science Publishers B.V.

Clerx, J.P.M. & Bishop, D.H.L. (1981). Qalyub virus, a member of the newly proposed nairovirus genus (Bunyaviridae). Virology 108: 361-372.

Clerx, J.P.M., Casals, J. & Bishop, D.H.L. (1981). Structural characteristics of nairoviruses (genus Nairovirus, Bunyaviridae). J. Gen. Virol. 55: 165-178.

Clerx, J.P.M., Fuller, F. & Bishop, D.H.L. (1983). Tick-borne viruses structurally similar to orthomyxoviruses. Virology 127: 205-219.

Collett, M.S., Purchio, A.F., Keegan, K.,Frazier, S., Hays, W., Anderson, D.K., Parker, M.D., Schmaljohn, C.M., Schmidt, J. & Dalrymple, J.M. (1985). Complete nucleotide sequence of the M RNA segment of Rift Valley fever virus. Virology 144: 228-245.

Collett, M.S. (1986). Messenger RNA of the M RNA segment RNA of Rift Valley Fever virus. Virology 151: 151-156.

Collins, P.L., Dickens, L.E., Buckler-White, A., Olmsted, R.A., Spriggs, M.K., Camargo, E. & Goelingh, K.V.W. (1986). Nucleotide sequences for the gene junctions of human respiratory syncytial virus reveal distinctive features of intergenic structure and gene order. Proc. Nat. Acad. Sci. U.S.A. 83: 4594-4598.

Davey, J., Dimmock, N.J. & Colman, A. (1985). Identification of the sequence responsible for the nuclear accumulation of the influenza virus nucleoprotein in Xenopus oocytes. Cell 40: 667-675.

Detjen, B.M., St. Angelo, C., Katze, M.G. & Krug, R.M. (1987). The three influenza virus polymerase (P) proteins not associated with viral nucleocapsids in the infected cell are in the form of a complex. J. Virol. 61: 16-22.

Eigen, M. & Schuster, P. (1982). Stages of emerging life: principles of early organization. J. Mol. Evol. 19: 47.

Eisemann Crone, D. & Keene J.D. (1989). Viral transcription is necessary and sufficient for vesicular stomatitis virus to inhibit maturation of small nuclear ribonucleoproteins. J. Virol. 63: 4172-4180.

Elliott, R.M. (1989). Nucleotide sequence analysis of the small (S) RNA segment of Bunyamwera virus, the prototype of the family Bunyaviridae. J. Gen. Virol. 70: 1281-1285.

Elliott, R.M. & McGregor, A. (1989). Nucleotide sequence and expression of the small (S) RNA segment of Maguari virus. Virology 171: 516-524.

Elliott, R.M. (1989). Nucleotide sequence analysis of the large (L) genomic RNA segment of Bunyamwera virus, the prototype of the family Bunyaviridae. Virology 173: 426-436.

Emerson S.U. & Yu, Y.H. (1975). Both NS and L protein are required for in vitro RNA synthesis by vesicular stomatitis virus. J. Virol. 15: 1348-1356.

Emery, V.C. & Bishop, D.H.L. (1987). Characterization of Punta Toro S mRNA species and identification of an inverted complementary sequence in the intergenic region of Punta Toro phlebovirus ambisense S RNA that is involved in mRNA transcription termination. Virology 156: 1-11.

40

Eshita, Y. & Bishop, D.H.L. (1984). The complete sequences of the M RNA of snowshoe hare bunyavirus reveals the presence of internal hydrophobic domains in the viral glycoproteins, Virology 137: 227-240.

Eshita, Y., Ericson, B., Romanowski, V. & Bishop, D.H.L. (1985). Analyses of the mRNA transcription processes of snowshoe hare bunyavirus S and M RNA species. J. Virol. 55: 681-689.

Evans, R.A. (1988). The steroid and thyroid hormone receptor superfamily. Science 240: 889-895.

Evans, S.A., Belsham, G.J. & Barrett, T. (1990). The role of the 5'-nontranslated regions of canine distemper virus and rinderpest virus. Virology 177: 317-323.

Fazakerley, J.K., Gonzalez-Scarano, F., Strickler, J., Dietzschold, B., Karush, F. & Nathanson, N.,(1988). Organization of the middle RNA segment of snowshoe hare bunyavirus. Virology 167: 422-432.

Franze-Fernandez, M. T., Zetina, C., Iapalucci, S., Lucero, M.A., Bouissou, C., Lopez, R., Rey, O., Daheli, M., Cohen, G.N. & Zakin, M.M. (1987). Molecular structure and early events in the replication of Tacaribe arenavirus S RNA. Virus Res. 7: 309-324.

Fuller-Pace, F.V. & Southern, P.J. (1988). Temporal analysis of transcription and replication during accute infection with lyphocytic choriomeningitis virus. Virology 162: 260-263.

Galinski, M., Mink. M., Lambert, D.M., Wechsler, S., Pons, M.W. (1987). Molecular cloning and sequence analysis of the human parainfluenza 3 virus RNA encoding the matrix protein, Virology 157: 24-30.

Gerbaud, S., Vialat, P., Pardigon, N., Wychowski, C., Girard, M. & Bouloy, M. (1987). The S segment of Germiston virus RNA genome can code for three proteins. Virus Res. 8: 1-13.

Giebel, L.B., Stohwasser, R., Zoller, L., Bautz, E.K.F. & Darai, G. (1989). Determination of the coding capacity of nephropathia epidemica virus strain hallnas B1 by molecular cloning and nucleotide sequence analysis. Virology 172: 498-505.

Goldbach, R. (1986). Molecular evolution of plant RNA viruses. Ann. Rev. Phytopathol. 24: 289-310.

Goldbach, R. & Wellink, J. (1988). Evolution of plus-strand RNA viruses. Intervirology 29: 260-267.

Goldbach, R., Le Gall, O. & Wellink, J. (1991). Alpha-like viruses in plants. Seminars in Virology 2: 19-25.

Gorbalenya, A.E., Koonin, E.V., Donchenko, A.P. & Blinov, V.M. (1989). A novel superfamily of nucleoside triphosphate-binding motif-containing proteins which are probably involved in duplex unwinding in DNA and RNA replication and recombination. FEBS Letters 235: 16-24.

Grady, L.J., Sanders, M.L. & Campbell, W.P. (1987). The sequence of the M RNA of an isolate of La Crosse virus. J. Gen. Virol. 68: 3057-3071.

Heaton, L.A., Hillman, B.I., Hunter, B.G., Zuidema, D. & Jackson, A.O. (1989). Physical map of the genome of sonchus yellow net virus, a rhabdovirus with six genes and conserved gene junction sequences. Proc. Nat. Acad. Sci. U.S.A. 86: 8665-8668.

Herman, R.C. (1986). Internal initiation of translation on the vesicular stomatitis virus phosphoprotein mRNA yields a second protein. J. Virol. 58: 797-804.

Herrler, G., Durkop, I., Becht, H. & Klenk, H.D. (1988). The glycoprotein of influenza C virusis the haemagglutinin, esterase and fusion factor. J. Gen. Virol. 69: 839-846.

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

Holland, J., Spindler, K., Horodyski, F., Grabau, E., Nichol, S. & Van de Pol S. (1982). Rapid evolution of RNA genomes. Science 215: 1577-1585.

Horikami, S.M. & Moyer, S.A. (1982). Host range mutants of vesicular stomatitis virus defective in in vitro RNA methylation. Proc. Nat. Acad. Sci. U.S.A. 79: 7694-7698.

Horvath, C.M., Williams, M.A. & Lamb, R.A. (1990). Eukaryotic translation of tandem cistrons: identification of the influenza B virus BM2 polypeptide. EMBO J. 9: 2639-2647.

Hsu, M.T., Parvin, J.D., Gupta, S., Krystal, M. & Palese, P. (1987). Genomic RNAs of influenza viruses are held in a circular conformation in virions and in infected cells by a terminal panhandle. Proc. Nat. Acad. Sci. U.S.A. 84: 8140-8144.

Hudson, L.D., Condra, C. & Lazzarini, R.A. (1986). Cloning and expression of a viral phosphoprotein: structure suggests vesicular stomatitis virus NS may function by mimicking an RNA template. J. Gen. Virol. 67: 1571-1579.

Iapalucci, S., Lopez, R., Rey, O., Lopez, N., Franze-Fernandez, M.T., Cohen, G.N., Lucero, M., Ochoa, A. & Zakin, M.M. (1989a). Tacaribe virus L gene encodes a protein of 2210 amino acid residues. Virology 170: 40-47.

Iapalucci, S., Lopez, N., Rey, O., Zakin, M.M., Cohen, G.N. & Franze-Fernandez, M.T. (1989b). The 5' region of Tacaribe virus L RNA encodes a protein with a potential metal binding domain. Virology 173: 357-361.

Iapalucci, S., Lopez, N. & Franze-Fernandez, M.T. (1991). The 3' end of the Tacaribe arenavirus subgenomic RNAs. Virology 182: 269-278.

Ihara, T., Akashi, H. & Bishop, D.H.L. (1984). Novel coding strategy (ambisense genomic RNA) revealed by sequence analysis of Punta Toro phlebovirus S RNA. Virology 136: 293-306.

Ihara, T., Matsuura, Y. & Bishop, D.H.L. (1985). Analyses of the mRNA transcription processes of Punta Toro phlebovirus (Bunyaviridae). Virology 147: 317-325.

Ihara, T., Smith, J., Dalrymple, J.M. & Bishop, D.H.L. (1985). Complete sequences of the glycoproteins and mRNA of Punta Toro phlebovirus compared to those of Rift Valley fever virus. Virology 144: 246-259.

Ishihama, I. & Nagata, K. (1988). Viral RNA polymerases. Crit. Rev. Biochem. 23: 27-76.

Johnson, K.M. (1985). Arenaviruses. In: Virology, pp. 1033-1053. Edited by B.N. Fields, R.M. Chanock, D.M. Knipe, J.L. Melnick, B. Roizman & R.E. Shope. New York: Raven Press.

Kakach, L.T., Wasmoen, T.L. & Collett, M.S. (1988). Rift Valley fever virus M segment: use of recombinant vaccinia viruses to study phlebovirus gene expression. J. Virol. 62: 826-833.

Kakach, L.T., Suzich, J.A. & Collett, M.S. (1989). Rift Valley fever virus M segment: Phlebovirus expression strategy and protein glycosylation. Virology 170: 505-510.

Kamer, G. & Argos, P. (1984). Primary structural comparison of RNA-dependant polymerases from plant, animal and bacterial viruses. Nucl. Ac. Res. 12: 7269-7282.

Karabatsos, N. (1985). International Catalogue of Arboviruses. American Society of Tropical Medicine and Hygiene.

Kingsbury, D.W. (1985). Orthomyxo- and paramyxoviruses and their replication. In: Virology, pp. 1157-1178. Edited by B.N. Fields, R.M. Chanock, D.M. Knipe, J.L. Melnick, B. Roizman & R.E. Shope. New York: Raven Press.

Kingsbury, D.W., Jones, I.M. & Murti, K.G. (1987). Assembly of influenza ribonucleocapsids. J. Virol. 4: 219-225.

Kirkegaard, K. & Baltimore, D. (1986). The mechanism of RNA recombination in poliovirus. Cell 47: 433-443.

Krug, R.M., Alonso-Caplen, F.V., Julkunen, I. & Katze, M.G. (1989). Expression and replication of the influenza virus genome. In: The influenza viruses, pp. 89-152. Edited by R.M. Krug. New York: Plenum Press.

Lamb, R.A. (1989). Genes and proteins of the influenza viruses. In: The influenza viruses, pp. 1-87. Edited by R.M. Krug. New York: Plenum Press.

Lazzarini, R.A., Keene, J.D. & Schubert, M. (1981). The origins of defective interfering particles of the negative strand RNA viruses. Cell 26: 145-154.

Lees, J.F., Pringle, C.R. & Elliott, R.M. (1986). Nucleotide sequence of the Bunyamwera virus M RNA segment conservation of structural features in the in the bunyavirus glycoprotein gene product. Virology 148: 1-14.

Luytjes, W., Krystal, M., Enami, M., Parvin, J.D. & Palese, P. (1989). Amplification, expression and packaging of a foreign gene by influenza virus. Cell 59: 1107-1113.

Marriott, A.C., Ward, V.K. & Nuttall, P.A. (1989). The S RNA segment of sandfly fever Sicilian virus: Evidence for an ambisense genome. Virology 169: 341-345.

Matthews, R.E.F. (1982). Classification and nomenclature of viruses. Fourth Report of the International Committee on Taxonomy of viruses, Intervirology 17: 1-200.

Mirakhur, B. & Peluso, R.W. 91988). In vitro assembly of a functional nucleocapsid from the negative-stranded genome RNA of a defective interfering particle of vesicular stomatitis virus. Proc. Nat. Acad. Sci. U.S.A. 85: 7511-7515.

42

Moscona, A. & Galinski, M.S. (1990). Characterization of human parainfluenza virus type 3 persistent infection in cell culture. J. Virol. 64: 3213-3218.

Murphy, F.A., Whitfield, S.G., Coleman, P.H., Calisher, C.H., Rabin, E.R., Jenson, A.B., Melnick, J.L.,Edwards, M.R. & Whiney, E. (1968). California group arboviruses: Electron microscopic studies. Exp. Mol. Pathol. 9: 44-56.

Ohgimoto, S., Bando, H., Kawano, M., Okamoto, K., Kondo, K., Tsurudome, M., Nishio, M. & Ito, Y. (1990). Sequence analysis of P gene of human parainfluenza type 2 virus: P and cysteine-rich proteins are translated by two mRNAs that differ by two nontemplated G residues. Virology 177: 116-123.

Palmer, E.L., Obijeski, J.F., Webb, P.A. & Johnson, K.M. (1977). The circular segmented nucleocapsid of an arenavirus-tacaribe virus. J. Gen. Virol. 36: 541-545.

Pardigon, N., Vialat, P., Gerbaud, S., Girard, M. & Bouloy, M. (1988). Nucleotide sequence of the M segment of Germiston virus: comparison of the M gene product of several bunyaviruses. Virus Res. 11: 73-85.

Parrington, M.A. & Yong Kang, C. (1990). Nucleotide sequence analysis of the S genomic segment of Prospect Hill virus: comparison with the prototype hantavirus. Virology 175: 167-175.

Parvin, J.D., Palese, P., Honda, A., Ishihama, A. & Krystal, M. (1989). Promoter analysis of influenza virus RNA polymerase. J. Virol. 63: 5142-5152.

Patterson, J.L., Holloway, B. & Kolakofski, D. (1984). La Crosse virions contain a primer-stimulated RNA polymerase and a methylated cap-dependant endonuclease. J. Virol. 52: 215-222.

Paul, P.R., Chattopadhyay, D. & Bannerjee, A.K. (1988). The functional domains of the phosphoprotein (NS) of vesicular stomatitis virus (Indiana strain). Virology 166: 350-357.

Pedersen, I.R. (1979). Structural components and replication of arenaviruses. Adv. Virus Res. 24: 277-330.

Pelet, T., Curran, J. & Kolakofski, D. (1991). The P gene of bovine parainfluenza virus 3 expresses all three reading frames from a single mRNA editing site. EMBO J. 10: 443-448.

Perrault, J. (1981). Origin and replication of defective interfering particles. Current Topics Microbiol. Immunol. 93: 151-207.

Peters, D. (1981). Plant rhabdovirus group. CMI/AAB Descr. of Plant Viruses 244.

Petterson, R. & Von Bonsdorff, C.H. (1975). Ribonucleoproteins of Uukuniemi virus are circular. J. Virol. 15: 386-392.

Raju, R. & Kolakofski, D. (1989). The ends of La Crosse virus genome and antigenome RNAs within nucleocapsids are base paired. J. Virol. 63: 122-128.

Raju, R., Raju, L. & Kolakofski, D. (1989). The translational requirement for complete La Crosse virus mRNA synthesis is cell-type dependent. J. Virol. 63: 5159-5165.

Raju, R., Raju, L., Hacker, D., Garcin, D., Compans, R. & Kolakofski, D. (1990). Non-templated bases at the 5' ends of Tacaribe virus mRNAs. Virology 174: 53-59.

Reanney, D.C. (1982). The evolution of RNA viruses. Ann. Rev. Microb. 36: 47-73.

Romanowski, V., Matsuura, Y. & Bishop, D.H.L. (1985). Complete sequence of the S RNA of lymphocytic choriomeningitis virus (WE strain) compared to that of Pichinde arenavirus. Virus Res. 3: 101-114.

Ronnholm, R. & Petterson, R.F. (1987). Complete nucleotide sequence of the M RNA segment of Uukuniemi virus encoding the membrane glycoproteins G1 and G2. Virology 160: 191-202.

Ryan, K.W. & Portner, A. (1990). Separate domains of Sendai virus P proteins are required for binding to viral nucleocapsids. Virology 174: 515-521.

Salvato, M., Shimomaye, E.M. & Oldstone, M.B.A. (1989). The primary structure of the lymphocytic choriomeningitis virus L gene encodes a putative RNA polymerase. Virology 169: 377-384.

Salvato, M. & Shimomaye, E.M. (1989). The completed sequence of lymphocytic choriomeningitis virus reveals a unique RNA structure and a gene for a zinc-finger protein. Virology 173: 1-10.

Sanchez, A., De, D.P. & Banerjee,A.K. (1985). In vitro phosphorylation of the NS protein by the L protein of vesicular stomatitis virus. J. Gen. Virol. 66: 1025-1036.

Sato, T.A., Kohama,T. & Sugiura, A. (1988). Intracellular processing of measles virus fusion protein. Arch. Virol. 98: 39-50.

Schmaljohn, C.S., Jennings, G.B., Hay, J. & Dalrymple, J.M. (1986). Coding strategy of the S genome segment of Hantaan virus. Virology 155: 633-643.

Schmaljohn, C.S., Schmaljohn, A.L. & Dalrymple, J.M. (1987). Hantaan virus M RNA: Coding strategy, nucleotide sequence and gene order. Virology 157: 31-39.

Schmaljohn, C.S. (1990). Nucleotide sequence of the L genome segment of Hantaan virus. Nucl. Ac. Res. 18: 6728.

Shapiro, G.I., Gurney, T. & Krug, R.M. (1987). Influenza virus gene expression: Control mechanisms at early and late times of infection and nuclear cytoplasmatic transport of virus specific RNAs. J. Virol. 61: 764-773.

Shapiro, G.I. & Krug, R.M. (1988). Influenza virus RNA replication in vitro: Synthesis of viral template RNAs and virion RNAs in the absence of an added primer. J. Virol. 62: 2285-2290.

Shaw, M.W., Choppin, P.W. & Lamb, R.A. (1983). A previously unrecognized influenza B virus glycoprotein from a bicistronic mRNA that also encodes the viral neuraminedase. Proc. Nat. Acad. Sci. U.S.A. 80: 4879-4883.

Shope, R.E., Rozhon, E.J. & Bishop, D.H.L. (1981). Role of the middle-sized bunyavirus RNA segment in mouse virulence. Virology 114: 273-276.

Simons, J.F., Hellman, U. & Petterson, R.F. (1990). Uukuniemi virus S RNA segment: Ambisense coding strategy, packaging of complementary strands into virions and homology to members of the genus phlebovirus. J. Virol. 64: 247-255.

Smith, F.I. & Palese, P. (1989). Variation in influenza virus genes. in: The influenza viruses, pp. 319-359. Edited by R.M. Krug. New York: Plenum Press.

Spaan, W.J.M., Den Boon, J.A., Bredenbeek, P.J., et al. (1990). Comparative and evolutionairy aspects of coronaviral, arteriviral and togaviral genome structure and expression. In: Positive stranded RNA viruses, pp. - . Edited by M.A. Brinton & F.X. Heinz, Washington D.C.: American Soc. for Microbiol.

Spriggs, M.K. & Collins, P.L. (1986). Sequence analysis of the P and C protein genes of human parainfluenza virus type 3: Patterns of amino acid sequence homology among paramyxovirus proteins. J. Gen. Virol. 67: 2705-2719.

Stohwasser, R., Giebel, L.B., Zoller, L., Bautz, E.K.F. & Darai, G. (1990). Molecular characterization of the RNA S segment of nephropathia epidemica virus strain hallnas B1. Virology 174: 79-86.

Strauss, J.H. & Strauss, E.G. (1988). Evolution of RNA viruses. Ann. Rev. Microbiol. 42: 657-683.

Suzich, J.A., Kakach, L.T. & Collett, M.S. (1990). Expression strategy of a phlebovirus: biogenesis of proteins from the Rift Valley fever virus M segment. J. Virol. 64: 1549-1555.

Thomas, S.M., Lamb, R.A. & Paterson, R.G. (1988). Two mRNAs that differ by two nontemplated nucleotides encode the amino coterminal proteins P and V of the paramyxovirus SV5. Cell 54: 891-902.

Tordo, N., Poch, O., Ermine, A., Keith, G. & Rougeon, F. (1986). Walking along the rabies genome: is the large G-L intergenic region a remnant gene ? Proc. Nat. Acad. Sci. U.S.A. 83: 3914-3918.

Travassos Da Rossa, A.P.A., Tesh, R.B., Pinheiro, F.P., Travassos Da Rossa, J.F.S. & Peterson, N.E. (1983). Characterization of eight new phlebotomus fever serogroup arboviruses (Bunyaviridae: phlebovirus) from the Amazon region of Brazil. Amer. J. Trop. Med. Hygiene 32: 1164-1171.

Ulmanen, I., Seppala, P. & Petterson, R.F. (1981). In vitro translation of Uukuniemi virus-specific RNAs: Identification of a nonstructural protein and a precursor to the membrane glycoproteins. J. Virol. 37: 72-79.

Vainionpaa, R., Marusyk, R. & Salmi, A. (1989). The Paramyxoviridae: Aspects of molecular structure, pathogenesis and immunity. Adv. Virus Res. 137: 211-242.

Vezza, A.C., Clewley, J.P., Gard, G.P., Abraham, N.Z., Compans, R.W. & Bishop, D.H.L. (1978). Virion RNA species of the arenavirus Pichinde, Tacaribe and Tamiami. J. Virol. 26: 485-497.

Vezza, A.C., Repik, P.M., Cash, P. & Bishop, D.H.L. (1979). In vivo transcription and protein synthesis capabilities of bunyaviruses: Wild-type snowshoe hare virus and its temperature-sensitive group I, group II and group I/II mutants. J. Virol. 31: 426-436.

Vidal, S. & Kolakofski, D. (1989). Modified model for the switch from Sendai virus transcription to replication. J. Virol. 63: 1951-1958.

Vidal, S., Curran, J. & Kolakofski, D. (1990). A stuttering model for paramyxovirus P mRNA editing. EMBO J. 9: 2017-2022.

44

Vlasak, R., Krystel, M., Nacht, M. & Palese, P. (1987). The influenza C virus glycoprotein (HE) exhibits receptor-binding (haemagglutinin) and receptor-destroying (esterase) activities. Virology 160: 419-425.

Ward, V.K., Marriott, A.C., El-Ghorr, A.A. & Nuttall, P.A. (1990). Coding strategy of the S RNA segment of Dugbe virus (nairovirus: Bunyaviridae). Virology 175: 518-524.

Yamashita, M., Krystel, M. & Palese, P. (1988). Evidence that the matrix protein of influenza C virus is coded for by a spliced mRNA. J. Virol. 62: 3348-3355.

Wright, K.E., Spro, R.C., Burns, J.W. & Buchmeier, M.J. (1990). Post-translational processing of the glycoproteins of lymphocytic choriomeningitis virus. Virology 177: 175-183.

Yoo, D. & Kang, C.Y. (1987). Genomic comparisons among members of hantavirus group. In: The biology of negative strand viruses, pp. 424-431. Edited by D. Kolakofski & B.W.J. Mahy. New York: Elsevier/North Holland.

Zebedee, S.L. & Lamb, R.A. (1988). Influenza A virus M2 protein: monoclonal antibody restriction of virus growth and detection of M2 in virions. J. Virol. 62: 2762-2772.

Zimmern, D. (1982). Do viroids and RNA viruses derive from a system that exchanges genetic information between eukaryotic cells? Trends in Biochem. Sci. 7: 205-207.

CHAPTER 3 TERMINAL SEQUENCE ANALYSIS AND MOLECULAR CLONING OF THE GENOMIC RNA SEGMENTS OF TOMATO SPOTTED WILT VIRUS

Peter de Haan, Lia Wagemakers, Dick Peters and Rob Goldbach

SUMMARY

Complementary DNA (cDNA) to the genomic RNA of tomato spotted wilt virus (TSWV) was synthesized and cloned in either pUC19 or lambda gt10. Restriction endonuclease maps were constructed from cDNA clones specific for the S and the L RNA segment, extending 3.0 and 8.9 kbp respectively. The nucleotide sequences of the 3' and 5' termini of both RNA molecules have been determined. The S and L RNA contain inverted repeats at their termini, probably involved in RNA replication and in the formation of circular nucleocapsids in virions. The terminal structures of the TSWV genome resemble in these aspects those of the Bunyaviridae.

47

3.1   Introduction

Tomato spotted wilt virus (TSWV) causes important plant diseases in tropical, subtropical and temperate regions. At least 400 species in 50 plant families, both mono- and dicotyledons are susceptable to TSWV (Matthews, 1982; D. Peters, personal communication). The virus is exclusively transmitted by thrips (Thysanoptera) (Sakimura,1962).

The virion of TSWV is a spherical membrane-bound particle 80-110 nm in diameter, covered with surface projections (Van Kammen et al., 1966). The core consists of three different ribonucleocapsid structures each containing a single species of genomic RNA. Virus particles contain three distinct structural proteins: an internal nucleocapsid protein (N) of M, 27,000 (27K) and two membrane glycoproteins of 78K (G1) and 58K (G2). In addition minor amounts of a large protein (L) of approximately 200K have been detected in virus particles (Mohamed et al., 1973; Mohamed, 1981; Tas et al., 1977).

The genome consists of three linear single stranded RNA molecules of approximately 3000 nucleotides (S RNA), 5000 nucleotides (M RNA) and 8000 nucleotides (L RNA). Analysis of transmission-defective strains has provided evidence that the M RNA encodes the G1 membrane glycoprotein (Verkleij & Peters, 1983). The coding properties of the other RNA molecules and the polarity of the genomic RNA remain unknown, although preliminary in vitro translation studies provided indications that the genome of TSWV is of positive polarity (Verkleij et al., 1982).

Since TSWV differs in morphology and genome structure from any other known plant virus, it has been classified as the single representative of a distinct group (TSWV group) (Ie, 1970; Matthews, 1982). However, TSWV has a number of properties in common with the Bunyaviridae, a large family of arthropod-associated viruses (Bishop et al., 1980; Milne & Francki, 1984). Viruses from this family have enveloped particles and tripartite single-stranded RNA genomes of negative polarity. Sofar, insufficient molecular information is available to determine whether TSWV should be regarded as a member of the Bunyaviridae.

To gain more insight into the polarity, structure and coding capacity of the different

48

RNA segments of TSWV and to study its possible relationship to the Bunyaviridae, nucleotide sequence data are required. As a first step in the unravelling of the molecular biological properties of TSWV, we here describe the determination of the terminal nucleotide sequences of the genomic RNA molecules. In addition, the cloning and physical mapping of complementary DNA (cDNA) to the majority of the sequences from the S and L RNA of TSWV is described.

3.2   Methods

3.2.1   Virus and plants

TSWV CNPH1, a Brasilian isolate from tomato, was maintained in tomato by grafting or in Nicotiana rustica var. America by mechanical passage. Virus was isolated from mechanically inoculated N. rustica leaves as described by Tas et al. (1977). Nucleocapsids were purified from infected leaves according to Verkleij & Peters (1983).

3.2.2   RNA purification and terminal sequence determination

RNA was recovered from purified virions, or from purified nucleocapsids by adding sodium dodecyl sulfate (SDS) to 1 % (w/v) and phenol extraction followed by ethanol precipitation (Verkleij & Peters, 1983).

In order to determine the nucleotide sequence at the 3' ends, the TSWV RNAs were labelled using RNA ligase and $(5'-^{32}P)pCp$ (England & Uhlenbeck, 1978). The 5' ends were labelled using $(Y-^{32}P)ATP$ and T4 polynucleotide kinase after treatment with tobacco acid pyrophosphatase (Promega) and/or calf intestinal alkaline phosphatase (Pharmacia), to remove possible caps or (pyro) phosphate groups at the 5' end.   In order to detect potential genome-linked proteins, extracted RNA was also iodinated with ^{125}I according to Markwell (1982).

End-labelled RNA molecules were resolved by electrophoresis in 'Low gelling

3.3.2 Cloning and physical mapping of TSWV cDNA

RNA was isolated from virus particles, purified from infected N. rustica leaves according to Tas et al. (1977). Generally, 100 g of leaf material yielded approximately 0.5 mg of virus, from which 1 - 5 μg of RNA could be extracted. The intactness of the isolated RNA molecules was tested by agarose gel electrophoresis (Fig.3.2). From a large number of gel patterns apparent sizes of 3000 nucleotides (S RNA), 5000 nucleotides (M RNA) and 8000 nucleotides (L RNA) could be estimated for the three genomic TSWV RNA molecules (data not shown). These values are in rather good agreement with previous calculations (Van den Hurk et al.,1977; Mohamed, 1981; Verkleij & Peters, 1983). Two different approaches have been followed for the synthesis and cloning of cDNA. A first series of cDNA clones has been obtained by random priming using fragmented single stranded calf thymus DNA, followed by blunt-end cloning in the SmaI site of plasmid pUC19.

A second series has been created by priming with random primers and synthetic oligonucleotides complementary to the 3' ends of the genomic RNAs and subsequent cloning in phage lambda gt10, using EcoR1 linkers. To select cDNA derived from S, M or L RNA sequences, clones from both plasmid and phage lambda cDNA libraries were further characterized by Northern blot analysis (Fig.3.2). In order to select sets of over-lapping cDNA clones representing the complete S and L RNA sequences, plasmid and phage 'walking' was performed by Southern blot analysis, using ^{32}P-labelled cDNA inserts as probes. In this way, a restriction map of approximately 3.0 kilobasepairs (kbp.) could be constructed for TSWV S RNA (Fig.3.3). To avoid any misinterpretations, all regions in the physical map were covered by at least 2 independent cDNA clones.

Orthomyxoviridae

influenza B   ........CCUGCUUUUGCU$_{OH}$
influenza A   ........UUCGCUUCUGCU$_{OH}$
influenza C   ........CCUGCUUUUGCU$_{OH}$
Dhori (segm. 2)   .....CUGCUUGUUUUUGCU$_{OH}$
Thogoto (segm. 3)   .....UGCUUU<u>GAUUU</u>U<u>CUCU</u>$_{OH}$

Bunyaviridae

tospo (TSW)
    L RNA   .....UUACCU<u>GAUUGCUCU</u>$_{OH}$
    M RNA   .....UGUAUU<u>GAUUGCUCU</u>$_{OH}$
    S RNA   .....UGACAC<u>GAUUGCUCU</u>$_{OH}$

bunya (Bunyamwera)
    L RNA   .....UAGGAGUACACUACU$_{OH}$
    M RNA   .....CGGUAGUACACUACU$_{OH}$
    S RNA   .....GUGGAGUACACUACU$_{OH}$

hanta (Hantaan)
    L RNA   .....AGGGAGUCUACUACUA$_{OH}$
    M RNA   .....GCGGAGUCUACUACUA$_{OH}$
    S RNA   .....AGGGAGUCUACUACUA$_{OH}$

uuku (Uukuniemi)
    M RNA   .....UAGCCGUCUUUGUGU$_{OH}$
    S RNA   .....UGGACGUCUUUGUGU$_{OH}$

phlebo (Punta Toro)
    M RNA   .....GUGCCGUCUUUGUGU$_{OH}$
    S RNA   .....CAGGGAGCUAUGUGU$_{OH}$

nairo (Qualyub)
    L RNA   .....UUAAUUUCUUAGAGA$_{OH}$
    M RNA   .....AGUAUUUCUUAGAGA$_{OH}$
    S RNA   .....CGUCCGUCUUAGAGA$_{OH}$

Arenaviridae

old world (LCM)
    L RNA   .....AGGAUCUUCGGUGCG$_{OH}$
    S RNA   .....AGGAUCCACUGUGCG$_{OH}$

new world (TAC)
    L RNA   .....AGGAUCCUCGGUGCG$_{OH}$
    S RNA   .....AGGAUCCACUGUGCG$_{OH}$

Fig.3.1: Comparison of the 3'-terminal sequences of the genomic RNA molecules of TSWV to those of members of the Arenaviridae, Bunyaviridae and Orthomyxoviridae. Nucleotides conserved between TSWV and Thogoto virus are underlined.

53
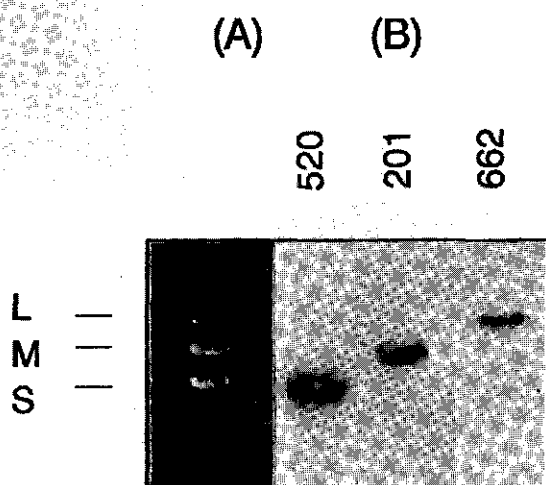
Fig.3.2: Electrophoretic analysis of TSWV genomic RNA (A) and selection of cDNA clones. After electrophoresis under denaturating conditions (Bailey & Davidson, 1976) TSWV RNA was transferred to nitrocellulose and hybridized to ^{32}P-labelled inserts of cDNA clones: 662, 201 and 520 as denoted by the numbers (B).

The Northern blot experiments furthermore revealed that the previously reported restriction map of TSWV M RNA (De Haan et al., 1989), actually represents that of a defective L RNA molecule of 4.7 kb in length. This defective RNA molecule is abundantly present in the TSWV line used in this study and almost completely masks the authentic M RNA segment (5.0 kb). This TSWV line has been maintained by mechanical passage of the virus for many years (see Chapter 5). To obtain cDNA clones to the full-length L RNA sequence, the original cDNA libraries were screened again, and additional cDNA clones to TSWV L RNA could be aligned, yielding a restriction

Fig.3.3: Restriction endonuclease maps of aligned cDNA clones to the L RNA ((A) and the S RNA (B) of TSWV. The arrows correspond with synthetic oligonucleotides used for primer extension sequencing. The numbers refer to the cDNA clones. Restriction enzymes are abbreviated as follows: Bc, BclI; Bg, BglII; E, EcoRI; H, HindIII; Hp, HpaI; K, KpnI; P, PstI; S, SphI; Ss, SstI; X, XbaI.

map covering approximately 8,900 nucleotides (Fig.3.3). Since clones 280, 803, 806 and 808 hybridized only to the full-length L RNA and not to the defective L RNA molecule (results not shown), it can be assumed that the latter molecule is the result of an internal deletion in TSWV L RNA (Resende et al., 1991).

The lengths of the two physical maps are in rather good agreement with the estimated sizes of the corresponding RNA molecules, which indicates that the physical maps almost completely cover both genomic RNA sequences. The polarity of both restriction maps in relation to the viral RNAs was determined by sequence analysis of the termini of both S and L RNA as described in the next paragraph.

55

3.3.3 Determination of the 5'-terminal sequence of the S and L RNA

To determine whether the the genomic RNA molecules possess special structures at their 5' ends, the RNAs were subjected to various labelling protocols. No genome-linked protein could be detected after iodination with ^{125}I, The 5' termini could only be labelled with ^{32}P by T4 polynucleotide kinase after treatment of the RNAs with alkaline phosphatase (15000 c.p.m./μg RNA), or even better after successive treatment with acid pyrophosphatase and alkaline phosphatase (35000 c.p.m./μg RNA). These results suggest the presence of pyrophosphate groups: (p)ppN, at the 5' termini. Direct RNA sequence determination experiments revealed that all three RNA segments start with 5' AGAGC.. 3' (results not shown).

In order to further analyse the nucleotide sequences at the 5' ends of both S and L RNA and to determine the polarity of the restriction maps shown in Fig.3.3, synthetic oligonucleotides corresponding to the ultimate 3' ends of the cloned regions (S3 and L2) were synthesized and used for direct sequencing on the RNA. In this way 'run off' transcripts were obtained at the 5' termini, indicating that the oligonucleotides used, were of viral complementary sense. The sequence data of the 5'-terminal regions of the genomic S and L RNA molecules are included in Fig.3.4. Both S and L RNA have complementary ends, S RNA over a length of 65-70 nucleotides and L RNA over a length of 62-66 nucleotides.

3.4 Discussion

Two sets of cDNA clones have been aligned, covering the S and L RNA sequence for more than 95 % (Fig.3.3). The terminal sequences of both RNA molecules were determined using direct 'enzymatic' RNA sequencing techniques and primer extension on RNA with reverse transcriptase, or by sequence analysis of a cDNA clone containing the complete genetic information of the 3' region of the corresponding RNA segment. The polarity of the S and L RNA specific cDNA clones could be verified by primer extension on RNA.

Fig.3.4: The complementary sequences at the termini of the S (left) and L (right) RNA of TSWV. The numbers represent the position to the 5' end of the RNA, derived from the complete sequences as described in the Chapters 4 and 5.

From these preliminary sequence data it can be concluded that both TSWV RNA molecules have complementary ends of 65-70 (S RNA) and 62-66 (L RNA) nucleotides in length. These terminal sequences may contain important regulatory signals, such as the recognition sites for the viral polymerase (Strauss & Strauss, 1983, 1988). Since nucleocapsids of TSWV are circular (Peters et al., 1991), whereas the purified viral RNA molecules are linear, the complementary ends may be involved in maintaining a 'pseudo circular' state of the nucleocapsids, as has been reported for bunyaviruses (Petterson & Von Bonsdorf, 1975; Obijeski et al., 1976; Pardigon et al., 1982; Raju & Kolakofski, 1989). Both the tripartition of the genome and the

57

structure of the termini, support the hypothesis that TSWV is related to the Bunyavirid-ae. An alignment of the 3'-terminal sequences of the RNAs from segmented negative-strand viruses is shown in Fig.3.1. On the basis of terminal nucleotide sequence homology, the animal bunyaviruses can be clustered into three groups, the nairoviruses (Clerx-Van Haaster et al., 1982), the uuku-/phleboviruses (Ihara et al., 1984; Ihara et al., 1985; Ronnholm & Petterson, 1987; Simons et al., 1990) and the hanta-/bunyaviruses (Schmaljohn et al., 1986; Schmaljohn et al., 1987; Clerx-Van Haaster et al., 1982). Members of the Arenaviridae and Orthomyxoviridae have their own distinct terminal sequences (Fig.3.3. and Desselberger et al., 1980; Auperin et al., 1982). The terminal sequences of TSWV RNA differ from those of all members of the Bunyaviridae, which might indicate that TSWV represents a distinct genus within this virus family. Strikingly, the termini of the TSWV RNAs show considerable sequence homology to that of RNA segment 3 of Thogoto virus and segment 2 of Dhori virus, members of the Orthomyxo-viridae, which might reflect ancestral relationships between these viruses. Further sequence analysis should answer the question of the polarity and coding capacity of the TSWV genome and will allow proper classification of this virus. In addition, the cDNA clones characterized in this paper may be utilized for sensitive and unequivocal detection of TSWV (Chapter 6).

ACKNOWLEDGEMENTS

58

REFERENCES

Auperin, D.D., Compans, R.W. & Bishop, D.H.L. (1982). Nucleotide sequence conservation at the 3' termini of the virion RNA species of New world and Old world arenaviruses. Virology 121: 200-203.

Bailey, J.M. & Davidson, N. (1976). Methylmercury as a reversible denaturating agent for agarose gel electrophoresis. Anal. Biochem. 70: 75-85

Bishop, D.H.L., Calisher, C., Casals, J., Chumakov, M.P., Gaidamovich, S.Ya., Hannoun, C., Lvov, D.K., Marshall, I.D., Okerblom, N., Petterson, R.F., Porterfield, J.S., Russel, P.K., Shope, R.E. & Westaway, E.G. (1980). Bunyaviridae. Intervirology 14: 125-143.

Clerx-Van Haaster, C.M. & Bishop, D.H.L. (1980). Analysis of the 3'-terminal sequences of Snowshoe Hare and La Crosse bunyaviruses. Virology 105: 564-574.

Clerx-Van Haaster, C.M., Clerx, J.P.M., Ushijima, H., Akashi, H., Fuller, F. & Bishop, D.H.L. (1982). The 3' terminal RNA sequences of Bunyaviruses and Nairoviruses (Bunyaviridae): Evidence of end sequence generic differences within the virus family. J. Gen. Virol. 61: 289-292.

Clerx, J.P.M., Fuller, F. & Bishop, D.H.L. (1983). Tick-borne viruses structurally similar to Orthomyxoviruses. Virology 127: 205-219.

De Haan, P., Wagemakers, L., Peters, D. & Goldbach, R. (1989). Molecular cloning and terminal sequence determination of the S and M RNA of tomato spotted wilt virus. J. Gen. Virol. 70: 3469-3473.

Desselberger, U., Racaniello, V.R., Zazra, J.J. & Palese, P. (1980). The 3' and 5'-terminal sequences of influenza A, B and C virus RNA segments are highly conserved and show partial inverted complementarity. Gene 8: 315-328.

England, T.E. & Uhlenbeck, O.C. (1978). 3'-Terminal labelling of RNA with T4 RNA ligase. Nature 275: 560-561.

Grunstein, J.M. & Hogness, D.S. (1975). Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. Proc. Nat. Acad. Sci. U.S.A. 72: 3961-3965.

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

Hewlett, M.J., Petterson, R.F. & Baltimore, D. (1977). Circular forms of Uukuniemi virion RNA: an electron microscopic study. J. Virol. 21: 1085-1093.

Huynh, T.H., Young, R.A. & Davis, R.W. (1985). Construction and screening cDNA libraries in lambda gt10 and lambda gt11. In: DNA cloning techniques: a practical approach, pp. 49-78. Edited by D. Glover. Oxford: IRL press.

Ie, T.S., (1970). Tomato spotted wilt virus. CMI/AAB Descr. of Plant Viruses. 39.

Ihara, T., Akashi, H. & Bishop, D.H.L. (1984). Novel coding strategy (ambisense genomic RNA) revealed by sequence analysis of Punta Toro phlebovirus S RNA. Virology 136: 293-306.

Ihara, T., Smith, J., Dalrymple, J.M. & Bishop, D.H.L. (1985). Complete sequences of the glycoproteins and M RNA of Punta Toro phlebovirus compared to those of Rift Valley fever virus. Virology 144: 246-259.

Lin, D.A., Roychoudhury, S., Palese, P., Clay, W.C. & Fuller, F.J. (1991). Evolutionary relatedness of the predicted gene product of RNA segment 2 of the tick-borne Dhori virus and the PB1 polymerase gene of influenza viruses. Virology 182: 1-7.

Maniatis, T., Fritsch, E.F. & Sambrook, J. (Eds.)(1982). Molecular cloning: a laboratory manual, New York: Cold Spring Harbor Laboratory.

Markwell, M.A. (1982). A new solid-state reagent to iodinate proteins: 1. Conditions for the efficient labeling of antiserum. Anal. Biochem. 125: 427-432.

Matthews, R.E.F. (1982). Classification and nomenclature of viruses: Fourth report of the international committee on Taxonomy of Viruses. Intervirology 17: 1-200.

59

Milne, R.G. & Francki, R.I.B. (1984). Should tomato spotted wilt virus be considered as a possible member of the family Bunyaviridae. Intervirology 22: 72-76.

Mohamed, N.A., Randles, J.W. & Francki, R.I.B. (1973). Protein composition of tomato spotted wilt virus. Virology 56: 12-21.

Mohamed, N.A. (1981). Isolation and characterization of subviral structures from tomato spotted wilt virus. J. Gen. Virol. 53: 197-206.

Parker, M.D. & Hewlett, M.J. (1981). The 3'-terminal sequences of Uukuniemi and Inkoo virus RNA genome segments. In: Replication of Negative Strand Viruses, pp. 125-145. Edited by D.H.L. Bishop & R.W. Compans. New York: Elsevier/North-Holland.

Peters, D., Avila, A.C. de, Kitajima, E.W., Resende, R. de O., De Haan, P. & Goldbach, R. (1991). An overview of tomato spotted wilt virus. In: Virus-Thrips-Plant Interactions of TSWV, Proc. USDA workshop, Belltsville U.S.A., pp.1-14. Edited by H-.T. Hsu & R.H. Lawson. Nat. Techn. Inf. Serv. Springfield (VA).

Raju, R. & Kolakofski, D. (1989). The ends of La Crosse virus genome and antigenome RNAs within nucl eocapsids are base paired. J. Virol. 63: 122-128.

Resende, R. de O., De Haan, P., Avila, A.C. de, Kormelink, R., Goldbach, R. & Peters, D. (1991). Generation of envelope and defective interfering RNA mutants of tomato spotted wilt virus by mechanical passage. J. Gen. Virol. in press.

Ronnholm, R. & Petterson, R.F. (1987). Complete nucleotide sequence of the M RNA segment of Uukuniemi virus encoding the membrane glycoproteins G1 and G2. Virology 160: 191-202.

Sakimura, K. (1962). The present status of thrips-borne viruses. In: Biological transmission of disease agents, pp. 33-40. Edited by K. Maramorosch. New York: Academic press.

Sanger, F., Nicklen, S. & Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Nat. Acad. Sci. U.S.A. 74: 5463-5467.

Schmaljohn, C.S., Jennings, G.B., Hay, J. & Dalrymple, J.M. (1986). Coding strategy of the S genome segment of Hantaan virus. Virology 155: 633-643.

Schmaljohn, C.S., Schmaljohn, A.L. & Dalrymple, J.M. (1987). Hantaan virus M RNA: coding strategy, nucleotide sequence and gene order. Virology 157: 31-39.

Schmaljohn, C.S. (1990). Nucleotide sequence of the L genome segment of Hantaan virus. Nucl. Ac. Res. 18: 6728.

Simons, J.F., Hellman, U. & Petterson, R.F. (1990). Uukuniemi virus S RNA segment: Ambisense coding strategy, packaging of complementary strands into virions and homology to members of the genus phlebovirus. J. Virol. 64: 247-255.

Staunton, D., Nuttall, P.A. & Bishop, D.H.L. (1989). Sequence analysis of Thogoto viral RNA segment 3: Evidence for a distant relationship between an arbovirus and members of the Orthomyxoviridae. J. Gen. Virol. 70: 2811-2817.

Strauss, E.G. & Strauss, J.H. (1983). Replication strategies of the single stranded RNA viruses of eukaryotes. Current Topics Microb. Immun. 105: 1-98.

Strauss, J.H. & Strauss, E.G. (1988). Evolution of RNA viruses. Ann. Rev. Microbiol. 42: 657-683.

Tas, P.W.L., Boerjan M.L. & Peters, D. (1977). The structural proteins of tomato spotted wilt virus. J. Gen. Virol. 36: 267-279.

Van den Hurk, J., Tas, P.W.L. & Peters, D. (1977). The ribonucleic acid of tomato spotted wilt virus. J. Gen. Virol. 36: 81-91.

Van Kammen, A., Henstra, S. & Ie, T.S. (1966). Morphology of tomato spotted wilt virus. Virology 30: 574-577.

Verkleij, F.N., De Vries, P. & Peters, D. (1982). Evidence that tomato spotted wilt virus RNA is a positive strand. J. Gen. Virol. 58: 329-338.

Verkleij, F.N. & Peters, D. (1983) Characterization of a defective form of tomato spotted wilt virus. J. Gen. Virol. 64: 677-686.

60

Wieslander, R. (1979). A simple method to recover intact high molecular weight RNA and DNA after elec
trophoretic separation in low gelling temperature agarose gels. Anal. Biochem. 98: 305-309.
Yanisch-Perron, C., Vieira, J. & Messing, J. (1985). Improved M13 phage cloning vectors and host strains:
nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33: 103-119.

CHAPTER 4   THE S RNA SEGMENT OF TOMATO SPOTTED WILT VIRUS HAS AN AMBISENSE CHARACTER

Peter de Haan, Lia Wagemakers, Dick Peters and Rob Goldbach

SUMMARY

The complete nucleotide sequence of the S RNA of tomato spotted wilt virus (TSWV) has been determined. The RNA is 2916 nucleotides long and has an ambisense coding strategy. The sequence contains two open reading frames (ORFs), one in the viral sense which encodes a protein with a predicted M_r of 52,400 (52.4K) and one in the viral complementary sense which encodes the viral nucleocapsid protein of 28.8K. Both proteins are expressed by translation of two subgenomic RNA species that possibly terminate at a long stable hairpin structure, located at the intergenic region. The structure of this RNA segment resembles that of the arthropod-borne phlebo- and uukuviruses (family Bunyaviridae). The absence of significant sequence homology between TSWV and bunyaviruses infecting animals suggests that TSWV should be considered as a representative of a new genus within the Bunyaviridae.

4.1   Introduction

Among the plant viruses, tomato spotted wilt virus (TSWV) is unique in its particle morphology and genome structure. It has therefore been classified as the only representative of a distinct group, the tomato spotted wilt virus group (Ie, 1970; Matthews, 1982). The virus has a broad host range and causes large yield losses in many economically important crops. It is transmitted exclusively by thrips species in a persistent manner (Sakimura, 1962; Paliwal, 1974).

TSWV virions are spherical enveloped particles, about 80-110 nm in diameter, that contain four different proteins: an internal nucleocapsid protein (N) of M_r 27,000 (27K), two membrane glycoproteins (G1 and G2) of 78K and 58K respectively, and a large protein (L) of approximately 200K (Mohamed et al., 1973; Tas et al., 1977). The genome consists of three linear ssRNA molecules approximately 3000 (S RNA), 5000 (M RNA) and 8000 (L RNA) nucleotides long, each complexed with N proteins to form circular nucleocapsids (Chapters 1, 3 and Van den Hurk et al., 1977; Mohamed, 1981). Based on this and other properties it has been suggested that TSWV should be considered as a member of the arthropod-borne Bunyaviridae (Milne & Francki, 1984; De Haan et al., 1989). Recently, the S and L RNA segments of TSWV have been cloned and their terminal sequences determined (Chapter 3). This analysis has revealed that, like RNAs of the Bunyaviridae and other negative-stranded RNA viruses, TSWV RNAs have complementary ends.

To gain more insight into the molecular properties of TSWV and to study its possible relationship to the Bunyaviridae in more detail, information on the nucleotide sequence of the TSWV genome is required. In this paper we report the complete nucleotide sequence of the TSWV S RNA. The data show that this RNA molecule has an ambisense gene arrangement with a putative 'non-structural' protein gene located on the viral RNA strand and the nucleocapsid protein gene on the viral complementary RNA strand. Evidence will be provided that both genes are expressed by the synthesis of two subgenomic mRNAs.

64

4.2 Methods

4.2.1 Cloning of cDNA and sequence determination

Complementary DNA to the S RNA segment of TSWV isolate CPNH1 was synthesized and cloned as described in the previous Chapter. DNA sequencing was performed by the dideoxy chain termination method (Sanger et al., 1977), after subcloning of restriction fragments in M13 mp18 or mp19 vectors (Yanisch-Perron et al., 1985). The RNA sequences of the 3' and 5' ends of the TSWV S RNA were determined as described before (Chapter 3). Sequence data were stored, edited and analysed using the Staden programs (Staden, 1982)

4.2.2 RNA isolation and Northern blot analysis

Total RNA from healthy and TSWV-infected Nicotiana rustica var. America plants was isolated according to De Vries et al. (1982). RNAs in samples of five μg were separated by electrophoresis in 1 % agarose gels, under denaturating conditions (Bailey & Davidson, 1976), transferred to nitrocellulose and hybridized to ^{32}P-labelled DNA probes by standard methods (Maniatis et al.,1982).

4.2.3 In vitro transcription and cell free translation

For in vitro expression purposes, viral cDNA was cloned in the transcription vector Bluescript SK+ (Stratagene). Two μg of plasmid DNA was linearized with HindIII, downstream of the cDNA insert and messenger sense 'run-off' transcripts were synthesized using T7 RNA polymerase according to the manufacturer's conditions. After incubation for 2 hr at 40 °C, the RNA was selectively precipitated in 2 M-lithium chloride and 1 μg of the transcript was incubated in 25 μl of rabbit reticulocyte lysate (Boehringer, Mannheim) containing 50 μCi (^{35}S)methionine according to protocols provided by the manufacturer. Translation products were either analysed directly by

65

```
                                              M S S S V V E S I I Q T R A S V W
AGAGCAAUUGUGUCAGAAUUUUGUUCAUAAUCAAACCUCACUUAGAAAAUCACAAUACUGUAAUAAGAACACAGUACCAAUAACCAUAAUGUCUUCAAGUGUUUAUGAGUCGAUCAUUCAGACAAGAGCUUCAGUCUGGG
                             70                                          140

G S T A S G K A V V D S Y W I K E L G T G S Q L V Q T Q L Y S D S R S K V V L W L Y C K V G I
GAUCAACUGCAGCUGGGUAAAGCUGUAUGUAGAUUCUUACUGGAUUCAUGAACUUGGUACUGGUCUCAACUAAGUUCAGACCCAGCUGUAUUCUGAUUCAAGAAGCAAAGUAGUCCUUUGGCUAUACUGCAAAGUAGGGAUC
                            210                                          280

F P V K K K R F L S Q H V Y I P I F D D I D F S I K I Q N S V L A L S V C S N T V N A H G V K
UUCCCUGUGAAGAAGAAGAAUUUCVUUCUCAGCAUGUGUAUAUCCCUAUUUUUGAUGAUAUUGAUUGUUGAGCAUCAAUAUUGAUAACUCUGUUCUGGCACUAUCUGUUUGCUCAAAUACAGUCAAUGCUAACGGAGUGAA
                            350                                          420

M Q G H L K V L S P A Q L M S I E S I M R R S D I T D R F Q L Q E K Q I I P N D K Y I E A A
ACAUCAAGGUCAUUUGAAGGUUUUGUCUCCUGCCCAGCUCCACUCUCAUUUGAAUCUAUCAACACAUCUGGAUAUUCAGACCGAUUCCAGCUCCAAGAAAAAGACAUAAUUCCCAAUGACAAAUACAUUGAAGCUGCAA
                            490                                          560

N K G S L S C V K E H T Y K I E N C Y R Q A L G K V N V L S P N R N V H E W L Y S F K P N F N
ACAAAGGCUCUUUGUCUUGUGUCAAAGCAUACCUAUAAGAUCGAUGAGAUGUGCUAUAAUCAGGCUUUUAGGCAAAGUGAAGUUUCUAUCUCCUAACAGAAAUGUCCAAGAAUGGCUGUACAGUUUCAAGCCAAAUUUCAAU
                            630                                          700

Q V E S H N R T V N S L A V K S L L M S A E N N I H P N S Q A S T D S H F K L S L W L R V P K
CAAGUUGAAAGCAACAACAGAACUGUAAAUUCUUGCAGUGAAAUCUCUGCCUCAUGUCAGCAGAAAACAACAUCAUGCCUAACUCUCAAGCUUCCACUGAUUCUCAUUUCAAGCUGAGCCUCUGGCUAAGGGUUCCAAA
                            770                                          840

V L K Q V S I Q K L F K V A G D E T N K T F Y L S I A C I P K H N S V E T A L N I T V I C K
GGAUUUGAAGCAGGUUUCCAUUCAGAAAUUGUUCAAGGUUGCAGGAGAUGAAACAAACAAAACAUUUUAUUUAUCUAUUGCCUGCAUUCCAAACCAUAACAGUGUUGAGACAGCUUUAAACAUUACUGUUAUUUGCAAGC
                            910                                          980

H Q L P I R K C K A P F E L S M H F S D L K E P Y N I V H D P S Y P K G S Y P M L M L E T H Y
AUCAGCUCCCCAAUUCGCAAAUGCAAAGCUCCUUUUGAAUUAUCAAUGAUGUUUUCUGAUUUAAAGGAGCCUUAACAACAUUGUUCAUGACCCUUCUAUACCCAAAGGAUCGGUUCCAAUGCUCUGGCUCUGAAACUCACACA
                            1050                                         1120

S L H K F F A T N L Q E D V I I Y T L N N L E L T P G K L D L G E R T L N Y S E D A Y K R K Y
UCVUUGCACAAGUUCAAUUAUCUUCUGCAUGUCAUUACUACACUUUGAACAACCUUGAGCUAACUCCUGGAAGAGUUAGAUUAGGUGAAGAAGAACCUUGAAUUUACAGUCAAGAUGCCUACAAAAGGAAAUA
                            1190                                         1260

F L S K T L E C L P S N T Q T M S Y L D S I Q I P S W K I Q F A R G E I K I S P Q S I S V A
UUUCCUUUCAAAAACACUUGAAUGCUUCCAUCUAACACACAAGAACAGCAUCCAAAUCCCUUCAUGGAAGAUAGACUUUGCCAGAGGAGAAAAUUAAAAUUCUCCACAAUCUAUUUCAGUUGCAA
                            1330                                         1400

K S L L K L Q L S G I K K K E S K V K E A Y A S G S K *
AAUCUUUGUUUAAAGCUUGAUUUUAAGCGGGAUCAAAAAGAAAGAAUCUAAGGUUAAGGAAGCGUAUGCUUCAGGAUCAAAAUAAUCUUGCUUUGUUUUCUAAUUAUGUUAUGUUUAUUUUUCUUUCUUUACUUUAU
                            1470                                         1540

AAUUAUUUUCUCUGUUUGUUCAUCUCUUUCAAAUUCCUCCCUGUCUAGUAGAAACCAUAAAAACAAAAAUAAAAAUUAAAAUAAAAUAAAAUCAAAAAUGAAAUAAAAACAACAAAAAUUAAAAAACGA
                            1610                                         1680

AAAACCAAAAAGACCCGAAAAGGGACCAAUUUUGGCCAAAUUUGGGUUUGUUUGUUUGUUUGUUUGUUUUUUGUAUUUUUAUUUUUUUAUAUUUAUUUUUAAUUUUAUUUUUAUUUUUUUGAUUUUUUUUUGUCUU
                            1750                                         1820

UCGUUUGUUUUUGUUAUUUUUAUUUAUUUAUUUAUUAAGCACAACACACAGAAAGCAAACUUUUAAUUAAAACACACUUAUUUAAAAAUUUAACACACUAAAGCAAGCACAAGCAAUAAAAGAUAAAGAAAGCUUUAUAUAUUUAUAGGCUU
                            1890                                         1960

UUUUAUUAAUUUUAACUUUACAGCUGCUUUCAAGCAAGUUCUGCGAGUUUUUGCCUGCUUUUUAACCCCGAACAUUUCAUAGAACUUUGUUAAGAGUUUCACUGUAAUGUUCCAUAGCAACACUCCCUUUAGCAUUAGGAUUGCU
                            2030                                         2100
                                         * A L E A L K A Q K K V G F M E Y F K N L T E S Y H E M A V S G K A N P N S
GGAGCUAAGUAUAGCAGCAUCACUCUUUCCCCUCUUCACCUGAUCUUUCAUUUCAAAUGCUUUCAGCACAGUGCAAACUUUUCCUUAAGGCUUCCUUGGUGUCAUACUUUCUUUGGGUCGAUACUUCUUUGGGUCGAUACCCGAGGUCCU
                            2170                                         2240
S S L I A A Y E K G K K V Q D E N M E F A K S K L V T C V K G L A E K T D Y K K P D I G L D K
UGUAUAUUUGCAUCCUGAUAUAUAGCCAAGACAACACUGAUCAUCUCAAAGCUAUCAACUGAAGCAAUAAGAGGUAAGCUACCUCCCAGCAUUAUGGCAAGUCUCACAGACUUUGCAUCAUCGAGAGGUAAAUCCAUAGGCU
                            2310                                         2380
Y K A Q Q Y I A L V V S I M E F S Q V S A I L P L S G G L M I A L R V S K A D D L P L G Y A
UGAAUCAAAGGAUGGGAAGCAACUUUGAGAUUUGAUAGUAUUGAGAUUCUCAGAAUUCCCAGUUGUCUUACAAACCUGACCCUGAUCAAGCUAUCAAGCCUUCUGAAGGUCAUGUCAGUGCCUCCAAUCCUUGUCCUUGAUUGAAGU
                            2450                                         2520
Q I L P N S A I K S K I T N L N E S N G T E E V L R V R I L S D L R R F T M D T G G I R D S T
UUUCUUUGAUGGUUAAUUUUACCAAAAGUAAAAUCGCUUUUUGCUUUGCUAAUAACCUUGUUUUUAUUAUGCUCUGACAUAUGCUCUGAUCAUAUAUGCUCUGACAUAUGCUCUGACAUAUAUGUCUCAGACAAAAAGUU
                            2590                                         2660
K K I T I K G F T F D S Q K I V K M I S Q R N K L F T L C S I V S M K K I Q D I N E L C F T K
UGAAAGUUGAAUGCUACCAGAUUCUGAUCGUCCUCAAACUGAACAAGGUCUUUGUCCCUUAGUGAGCUUUAACCUUUAGACAUGAUCGUAAAAGUUGGUUAUAGUCUUUGACCGUAUG
                            2730                                         2800
F N F A V L N Q D E E F E L D K G Q T L L A V I S E K T L K V K S M
UAACUCAAGGUGCGAAAGUGCAACUCUGUAUCCCGCAGUCGUUUCUUAGGUUCUUAAUGUGAUGACUGAGUGGUUAAGGUAUGAACACAAAAUUGACACGAUUGCUCU
                            2870
```

Fig.4.1: *The complete nucleotide sequence of TSWV S RNA (numbered from the 5' end) and the predicted gene products. The deduced amino acid sequence of the protein encoded by the viral RNA is written above the RNA sequence. The sequence of the protein encoded by the viral complementary strand is written below the RNA sequence. Potential N-glycosylation sites are underlined. The asterisks (*) indicate termination codons.*

66

SDS-PAGE, or immunoprecipitated using polyclonal antibodies against purified TSWV nucleocapsids, prior to electrophoresis, as described by Bernstein & Hruska (1981).

4.3   Results

4.3.1   Cloning and sequencing of the TSWV S RNA

The nucleotide sequences of 30 nucleotides at the 3' end of the S RNA and five nucleotides at the 5' end, were determined after end-labelling of the RNA, followed by partial degradation with base-specific ribonucleases. From the deduced sequence, a synthetic oligonucleotide was synthesized to be complementary to the 20 3'-terminal nucleotides (S1). This oligonucleotide was used for synthesis and cloning of cDNA, and for determination of the nucleotide sequence of approximately 200 nucleotides at the 3' end of the RNA, by primer extension (Chapter 3). Based on the restriction map of S RNA-specific cDNAs, the clones 514, 520, 608 and 614 were selected for further sequence analysis (Fig.3.3). The sequence of each DNA restriction fragment was determined from both strands. Finally, a second synthetic oligonucleotide (S3) was synthesized, corresponding to the most 5'-proximal sequence of cDNA clone 520 and the 5'-terminal sequence of the S RNA (5' AGAGCAA.. 3') was verified by primer extension (Chapter 3).

The complete sequence of the TSWV S RNA is 2916 nucleotides long (Fig.4.1). This is approximately 100 to 500 nucleotides shorter than previously reported values, which were based on electrophoretic mobilities (Verkleij & Peters, 1983). Its base composition is 31.6% A, 32.9% U, 19.3% C and 16% G. The 3'- and 5'-terminal sequences are complementary over a stretch of 65 to 70 nucleotides and can be folded into a stable panhandle structure, with a free energy of $\Delta G = -254.1$ kJ/mol (Fig.3.4). An internal inverted complementary sequence of U-rich stretches followed by A-rich stretches, is located between position 1582 and 1834 (numbered from the 5'-end of the viral RNA). This sequence can be folded into a stable hairpin structure with a $\Delta G = -452.7$ kJ/mol (Fig.4.3).
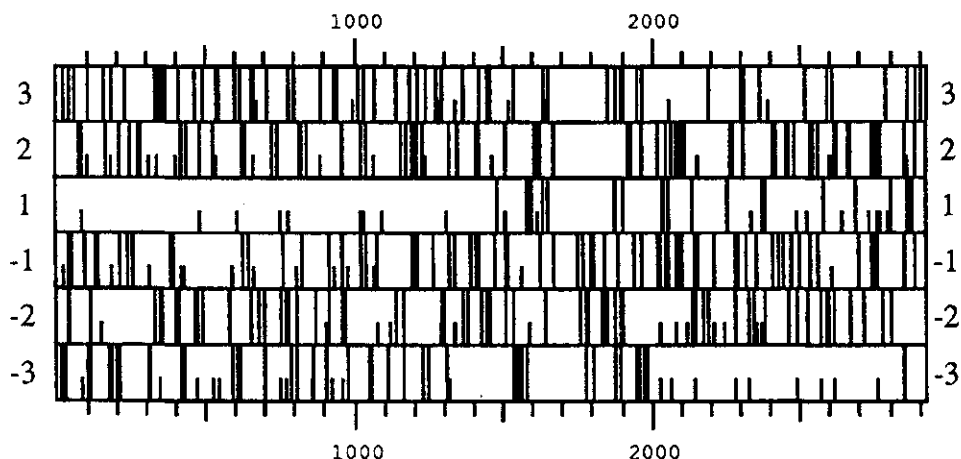
Fig.4.2: Distribution of translation initiation (short vertical bars) and termination (long vertical bars) codons in the three possible reading frames of the viral (1, 2 and 3) and viral complementary (-1, -2 and -3) S RNA strands.

4.3.2 Predicted gene products of TSWV S RNA

Analysis of the six different reading frames of the viral and viral complementary RNA strand revealed two long open reading frames (ORFs), one on each strand. No additional ORFs of significant length (i.e. ORFs encoding proteins larger than 5K) could be detected in any of the other reading frames (Fig.4.2). The ORF on the viral RNA strand starts with an AUG codon at position 88 and terminates at an UAA stopcodon at position 1481 (numbered from the 5' end of the viral strand), which corresponds to a protein of 465 amino acids and an M_r of 52.4K. The amino acid sequence of this putative protein does not contain hydrophobic regions that might function as signal peptides or transmembrane domains, according to the hydropathy algorithms of Hopp & Woods (1981) and Kyte & Doolittle (1982) (data not shown). Although the sequence contains five potential N-glycosylation sites, it is not known whether these sites are indeed glycosylated in vivo. No significant homology could be detected to any other protein in the EMBL protein and nucleotide sequence data base.

68
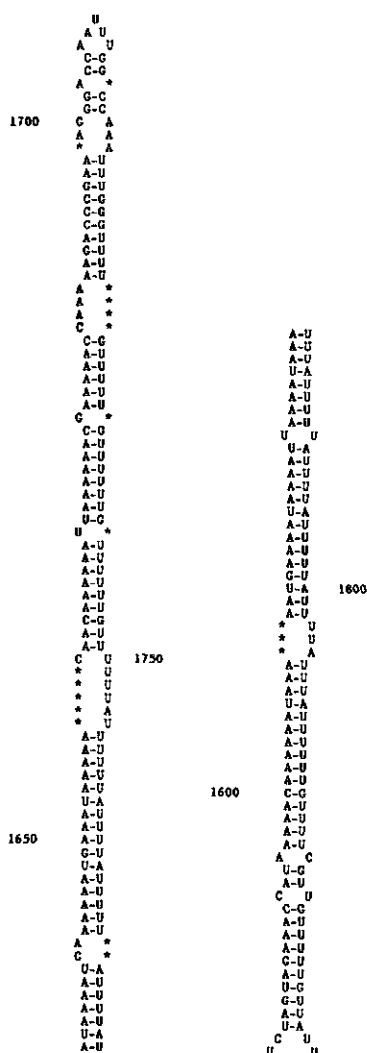
The ORF on the viral complementary RNA strand starts with an AUG codon at position 2763 and terminates at an UGA stopcodon at position 1989 (numbered from the 5' end of the viral strand); it encodes a protein of 259 amino acids with an M_r of 28.8K. The amino acid sequence also does not contain any strongly hydrophobic regions and there are no possible N-glycosylation sites present. The length of this ORF suggests that it encodes the nucleocapsid protein. In order to verify this, a cDNA fragment containing this ORF was expressed in vitro. To this end a 387 bp EcoRI cDNA fragment of clone 614 was fused to the 494 bp HindIII/EcoRI cDNA fragment of clone 520 and subsequently cloned in Bluescript SK+ (pTSWV-vcORF). Run-off transcripts were synthesized and translated in a rabbit reticulocyte lysate. (Fig.4.4). One discrete translation product was obtained (Fig.4.4, lane 4), which comigrated with purified viral nucleocapsid protein (Fig.4.4, lane 8), and moreover, which reacted with an antiserum raised against purified nucleocapsids (Fig.4.4, lane 7). Translation products of unrelated plant virus RNAs (tobacco mosaic virus (TMV) and cowpea chlorotic mottle virus (CCMV)) were not precipitated by this antiserum (Fig.4.4, lanes 5 and 6). These results indicate that the ORF located on the

*Fig.4.3: The secondary structure at the intergenic region of TSWV S RNA. The nucleotide positions are numbered from the 5' end. Asterisks (*) represent gaps corresponding to unpaired nucleotides in the sequence.*

69

viral complementary strand of the TSWV S RNA represents the nucleocapsidgene. When searching the EMBL protein and nucleotide sequence data base, no significant homology could be detected to other nucleocapsid proteins of positive- or negative-strand viruses, or to any other protein.

4.3.3   Detection of subgenomic RNA species in TSWV-infected plant cells

The ambisense gene arrangement of the S RNA predicts that the two ORFs are expressed by the formation of two subgenomic mRNA species, as found for other ambisense RNA molecules of members of the Arenaviridae and Phlebovirinae (Auperin
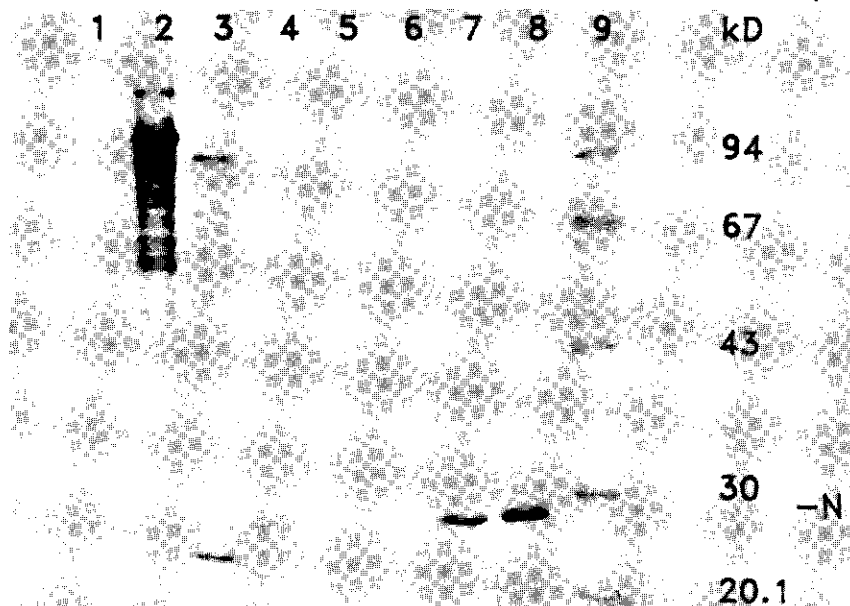
Fig.4.4: Identification of the TSWV nucleocapsid gene by cell-free translation of the in vitro transcripts of plasmid pTSWV-vcORF in a rabbit reticulocyte lysate. Samples were: lanes 1 and 5, control i.e. the endogenous translation products in the absence of RNA transcripts; lane 2, translation products of TMV RNA (M_r 126,000 (126K) and 183K); lane 3, translation products of CCMV RNA (23K, 35K, 100K and 105K); lanes 4 and 7, translation products directed by the pTSWV-vcORF transcripts; Lane 6, translation products directed by TMV RNA plus CCMV RNA. Translation products were analysed directly (lanes 1 to 4) or immunoprecipitated using polyclonal antibodies against purified nucleocapsids prior to electrophoresis (lanes 5 to 7). Lanes 8 and 9 contain Coomassie-blue stained purified nucleocapsid proteins (N) and marker proteins respectively. M_r values are indicated on the right.

et al.,1984; Salvato, 1989; Giorgi et al., 1991). To identify possible mRNA species corresponding to the nucleocapsid protein and the putative non-structural protein, Fig.4.5 shows Northern blots of total RNA isolated from healthy and infected plants, hybridized to ^{32}P-labelled probes. The probe corresponding to the ORF on the viral strand (probe NSs) hybridized to the genomic S RNA and to a subgenomic RNA species of approximately 1.7 kb. With the probe corresponding to the nucleocapsid gene (probe N), a second smaller subgenomic mRNA (1.2 kb) was detected, in addition to full-length S RNA. These results indicate that the expression strategy of the TSWV S RNA is similar to that of the ambisense RNAs of arenaviruses and phlebo-/ uukuviruses.
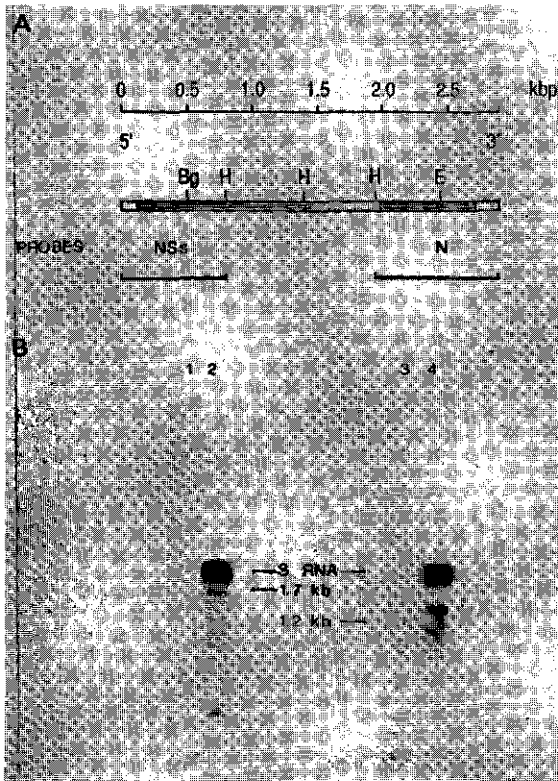
Fig.4.5: Identification of subgenomic RNA species in TSWV-infected tobacco cells.
(a) Position of cDNA probes used, on the restriction map of TSWV S RNA. The shaded regions in the open bar, correspond to the two ORFs shown in Fig.2. The horizontal lines correspond to the cDNA fragments used as probes (denoted NSs and N). (b) Total RNA from healthy (lanes 1 and 3) and TSWV-infected (lanes 2 and 4) tobacco leaf tissue was resolved in agarose gels, blotted to nitrocellulose and hybridized to probes NSs (left panel) and N (right panel), as described in Methods.

4.4 Discussion

The sequence data presented here, show that the TSWV S RNA is 2916 nucleotides long. From the sequence it can be deduced that the TSWV S RNA has an ambisense gene arrangement. The Northern blot hybridization experiments demonstrate that the two genes on this RNA are expressed by two subgenomic RNA species. This genome strategy is also found for the S RNA segments of viruses of the genera phlebovirus and uukuvirus (Giorgi et al., 1991) and for the S and L RNA segments of viruses of the family Arenaviridae (Auperin et al., 1984; Romanowski et al., 1985; Clegg & Oram, 1985; Salvato, 1989), but has not yet been reported for any plant virus.

Together with morphological data (Chapter 1), the structure and genetic organisation of the S RNA provide strong evidence that TSWV should be regarded as a member of the Bunyaviridae. Preliminary nucleotide sequence data demonstrate that TSWV M and L RNA, like the corresponding RNAs of the animal Bunyaviridae, are entirely of negative polarity (Chapter 5).

Compared to the ambisense RNA molecules of the animal phlebo-, uuku- and arenaviruses, TSWV S RNA contains relatively long terminal untranslated regions. This may be related to the formation of a long stable panhandle structure at the termini of the genomic RNA molecule and may explain the circular appearance of isolated TSWV nucleocapsids (Peters et al., 1991). These non-coding sequences probably contain important signals for replication and transcription (Strauss & Strauss, 1988; Raju & Kolakofski, 1989). Another putative regulatory element is located in the intergenic region of the RNA, where the sequence can be folded into a long hairpin structure. A similar hairpin structure has been found in the S RNA of Punta Toro virus (Emery & Bishop, 1987). For this virus it is known that the peak of the hairpin forms the transcription termination point for the two subgenomic RNA molecules produced from this RNA. The estimated sizes of the subgenomic RNA species of TSWV S RNA (1.2 and 1.7 kb) and the location of the hairpin structure are consistent with TSWV using the same mechanism for transcription termination. Investigations are currently in progress to study the regulation of the expression of both genes of TSWV S RNA and to map precisely

72

the transcription initiation and termination sites.

TSWV S RNA appears to encode a non-structural protein which we propose to designate NSs, according to the nomenclature used for the Bunyaviridae. It does not show significant amino acid homology to any of the reported phlebo- or uukuvirus NSs proteins (Giorgi et al., 1991). The protein has not yet been detected in TSWV-infected plant cells and its role in the viral infection process is sofar unknown.

In conclusion, the data presented in this paper indicate that TSWV should be considered as a member of the Bunyaviridae. However, in view of the lack of homology in both the nucleotide sequence and the derived amino acid sequence to any of the animal Bunyaviridae analysed sofar, we propose to place TSWV into a new genus. In view of its proposed taxonomic position the question needs to be answered as to whether TSWV replicates also in its insect vectors. Further studies on its genome organization and replication may reveal how this virus departed evolutionarily from other bunyaviruses as an insect-animal virus to become an insect-plant virus.

73

REFERENCES

Auperin, D.D., Romanowski, V., Galinski, M. & Bishop, D.H.L. (1984). Sequencing studies of Pichinde arenavirus S RNA indicate a novel coding strategy, ambisense viral S RNA. J. Virol. 52: 897-904.

Bailey, J.M. & Davidson, N. (1976). Methylmercury as a reversible denaturating agent for agarose gel elec trophoresis. Anal. Biochem. 70: 75-85.

Bernstein, J.M. & Hruska, J.F. (1981). Respiratory syncitial virus proteins: identification by immunoprecipi tation. J. Virol. 38: 278-285.

Clegg, J.C.S. & Oram, J.D. (1985). Molecular cloning of Lassa virus RNA: nucleotide sequence and expression of the nucleocapsid protein gene. Virology 144: 363-372.

Clerx-van Haaster, C.M. & Bishop, D.H.L. (1980). Analysis of the 3'-terminal sequences of Snowshoe Hare and La Crosse bunyaviruses. Virology 105: 564-574.

De Haan, P., Wagemakers, L., Peters, D. & Goldbach, R. (1989). Molecular cloning and terminal sequence determination of the S and M RNA of tomato spotted wilt virus. J. Gen. Virol. 70: 3469-3473.

De Vries, S.C., Springer, J. & Wessels, J.G.H. (1982). Diversity of abundant mRNA sequences and patterns of protein synthesis in etiolated and greened pea seedlings. Planta 156: 129-135.

Emery, V.C. & Bishop, D.H.L. (1987). Characterization of Punta Toro S mRNA species and identification of an inverted complementary sequence in the intergenic region of Punta Toro phlebovirus S RNA that is involved in mRNA transcription termination. Virology 156: 1-11.

England, T.E. & Uhlenbeck, O.C. (1978). 3'-Terminal labelling of RNA with T4 RNA ligase. Nature 275: 560-561.

Giorgi, C., Accardi, L., Nicoletti, L., Gro, M.C., Takehara, K., Hilditch, C., Morikawa, S. & Bishop, D.H.L. (1991). Sequences and coding strategies of the S RNAs of Toscana and Rift valley fever and Uukuniemi viruses. Virology 180: 738-753.

Hopp, T.P. & Woods, K.R. (1981). Prediction of protein antigenic determinants from amino acid sequences. Proc. Nat. Acad. Sci. U.S.A. 78: 3824-3828.

Ie, T.S., (1970). Tomato spotted wilt virus. CMI/AAB Descr. of Plant Viruses 39.

Kyte, J. & Doolittle, R.F. (1982). A simple method for displaying the hydrophobic character of a protein. J. Mol. Biol. 157: 105-132.

Maniatis, T., Fritsch, E.F. & Sambrook, J. (Eds.)(1982). Molecular cloning: a laboratory manual. New York: Cold Spring Harbor Laboratory.

Matthews, R.E.F. (1982). Classification and nomenclature of viruses: Fourth report of the international committee on Taxonomy of Viruses. Intervirology 17: 1-200.

Milne, R.G. & Francki, R.I.B. (1984). Should tomato spotted wilt virus be considered as a possible member of the family Bunyaviridae? Intervirology 22: 72-76.

Mohamed, N.A. (1981). Isolation and characterization of subviral structures from tomato spotted wilt virus. J. Gen. Virol. 53: 197-206.

Mohamed, N.A., Randles, J.W. & Francki, R.I.B. (1973). Protein composition of tomato spotted wilt virus. Virology 56: 12-21.

Paliwal, Y.C. (1974). Some properties and thrip transmission of tomato spotted wilt virus in Canada. Can. J. Bot. 52: 1177-1182.

Peters, D., Avila, A.C. de, Kitajima, E.W., Resende, R. de O., De Haan, P. & Goldbach, R. (1991). An overview of tomato spotted wilt virus. In: Virus-Thrips-Plant Interactions of TSWV, Proc. USDA workshop, Belltsville U.S.A., pp.1-14. Edited by H-.T. Hsu & R.H. Lawson. Nat. Techn. Inf. Serv. Springfield (VA).

Raju, R. & Kolakofski, D. (1989). The ends of La Crosse virus genome and antigenome RNAs within nucleocapsids are base-paired. J. Virol. 63: 122-128.

74

Romanowski, V., Atsura, Y. & Bishop, D.H.L. (1985). Complete sequence of the S RNA of lymphocytic choriomeningitis virus (WE strain) compared to that of Pichinde arenavirus. Virus Res. 3: 101-114.

Sakimura, K. (1962). The present status of thrips-borne viruses. In: Biological transmission of disease agents, pp. 33-40. Edited by K. Maramorosch. New York: Academic press.

Salvato, M. (1989). Ambisense nature of the L genomic segment of LCMV. In: Genetics and Pathogenicity of Negative-strand viruses, pp. 168-173. Edited by D. Kolakovski & B.W.J. Mahy. Amsterdam: Elsevier. Science Publishers B.V.

Sanger, F., Nicklen, S. & Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Nat. Acad. Sci. U.S.A. 74: 5463-5467.

Staden, R. (1982). An interactive graphics program for comparing and aligning nucleic acid and amino acid sequences. Nucl. Ac. Res. 10: 2951-2961.

Strauss, J.H. & Strauss, E.G. (1988). Evolution of RNA viruses. Ann. Rev. Microbiol. 42: 657-683.

Tas, P.W.L., Boerjan M.L. & Peters, D. (1977). The structural proteins of tomato spotted wilt virus. J. Gen. Virol. 36: 267-279.

Van den Hurk, J., Tas, P.W.L. & Peters, D. (1977). The ribonucleic acid of tomato spotted wilt virus. J. Gen. Virol. 36: 81-91.

Verkleij, F.N. & Peters, D. (1983). Characterization of a defective form of tomato spotted wilt virus. J. Gen. Virol. 64: 677-686.

Wieslander, R. (1979). A simple method to recover intact high molecular weight RNA and DNA after electrophoretic separation in low gelling temperature agarose gels. Anal. Biochem. 98: 305-309

Yanisch-Perron, C., Vieira, J. & Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33: 103-119.

75

CHAPTER 5   TOMATO SPOTTED WILT VIRUS L RNA ENCODES A PUTATIVE RNA POLYMERASE

Peter de Haan, Richard Kormelink, Renato de Oliveira Resende, Frank van Poelwijk, Dick Peters and Rob Goldbach

SUMMARY

The complete nucleotide sequence of the large (L) genome segment of tomato spotted wilt virus (TSWV) has been determined. The RNA is 8897 nucleotides long and contains complementary 3' and 5' ends, comprising 62 nucleotides at the 5' end and 66 nucleotides at the 3' end. The RNA is of negative polarity, with one large open reading frame (ORF) located on the viral complementary strand. This ORF corresponds to a primary translation product of 2875 amino acids in length, with a predicted M_r of 331,500. Comparison with the polymerase proteins of other negative-strand viruses indicates that this protein most likely represents the viral polymerase. The genetic organization of TSWV L RNA is similar to that of the L RNA segments of Bunyamwera and Hantaan virus, animal-infecting representatives of the Bunyaviridae.

5.1 Introduction

Based on its unique properties among other plant viruses, tomato spotted wilt virus (TSWV) has previously been classified as the single representative of a distinct virus group (Ie, 1970; Matthews, 1982). Recently obtained molecular data however, have provided evidence that TSWV should be considered as a member of the arthropod-borne Bunyaviridae, unique in its property to infect plants (Chapter 3 and 4).

Like the established members of the Bunyaviridae (Elliott, 1990), TSWV is characterized by spherical enveloped particles of approximately 80-110 nm in diameter. Two virus-encoded glycoproteins, denoted G1 (M_r 78,000 (78K)) and G2 (M_r 58K) are associated with the virus envelope (Tas et al., 1977). The internal pseudo-circular nucleocapsids consist of three species of single stranded RNA, denoted S RNA (2916 nucleotides), M RNA (approximately 5000 nucleotides) or L RNA (approximately 8000 nucleotides), which are each seperately encapsidated with nucleocapsid (N) protein (M_r 28.8K) (Chapter 1). In addition few copies of a large (L) protein (approximately 200K) are present in the virus particle, which may represent the viral polymerase (Mohamed et al., 1973; Mohamed, 1981; Tas et al., 1977).

As described in Chapters 3 and 4, the genomic RNA segments have been cloned and the complete nucleotide sequence of the S RNA has been determined from a set of overlapping cDNA clones. TSWV S RNA encodes two proteins, the nucleocapsid (N) protein and a non-structural (NSs) protein in an ambisense gene arrangement. The nucleocapsid protein is expressed from a subgenomic mRNA species of approximately 1.2 kb, transcribed from the viral RNA strand, while the NSs protein (M_r 52.4K) is expressed from a mRNA of approximately 1.7 kb, transcribed from the viral complementary RNA strand. The structure of TSWV S RNA conforms that of the phleboviruses and uukuviruses, two genera of the family Bunyaviridae (Giorgi et al., 1991).

Here we report the complete nucleotide sequence of TSWV L RNA. It contains a single large open reading frame (ORF) in viral complementary sense, which most likely corresponds with the viral polymerase gene. The genetic organization of the TSWV L

78

RNA segment further strengthens our previous conclusion, that this virus represents a plant-infecting member of the Bunyaviridae.

5.2   Methods

5.2.1   Virus and plants

TSWV CNPH1, a Brasilian isolate from tomato, was maintained in tomato by grafting and infected leaf tissue was stored in liquid nitrogen. Nicotiana rustica var. America plants were either mechanically inoculated from this original virus stock, or from previously inoculated, systemically infected N. rustica. Virus was purified from infected N. rustica leaves according to Tas et al. (1977) and RNA extracted as described in Chapter 3.

5.2.2   Synthesis, cloning and sequence determination of cDNA

Complementary DNA (cDNA) to TSWV RNA was synthesized and cloned as previously described (Chapter 3). To obtain cDNA clones containing the 3' end of the L RNA, a portion of 5 µg of genomic RNA was polyadenylated at the 3' end, using one unit of poly(A)-polymerase (BRL), according to Devos et al. (1976). First strand cDNA synthesis was primed with oligo(dT), followed by second strand synthesis according to Gubler & Hoffmann (1983). Double stranded cDNA was made blunt-ended using T_4 DNA polymerase and subsequently cloned into the SmaI site of plasmid pUC19.

DNA sequencing was performed by the dideoxy chain termination method (Sanger et al., 1977) on double stranded DNA templates (Zhang et al., 1988), or after subcloning of restriction fragments in M13 mp18 or mp19 vectors (Yanisch-Perron et al., 1985). Nucleotide and amino acid sequences were compiled and analyzed using programs developed by the University of Wisconsin Genetics Computer Group (UWGCG).

79

5.3   Results

5.3.1 Cloning and sequence determination of the TSWV L RNA

Northern blot analysis of genomic RNA, purified from the original CNPH1 virus stock has revealed that the previously reported restriction map of TSWV M RNA (De Haan et al., 1989), actually represents that of a defective L RNA molecule of 4.7 kb in length. This defective RNA molecule is abundantly present in the TSWV CNPH1 line used in this study and masks the authentic M RNA segment (5.0 kb). This TSWV line has been maintained by mechanical passage of the virus for several years.

In order to obtain cDNA clones corresponding to the full-length genomic RNA sequence, the original cDNA library was screened again and additional cDNA clones to TSWV L RNA could be aligned, yielding a restriction map covering approximately 8,900 nucleotides (Fig.5.1 and Chapter 3). The cDNA clones denoted 70, 266, 280, 299, 329, 420, 662, 669, 803, 806, 808 and 810 were selected for sequence analysis.

As described under 3.3.1, clone 669 contained the 3'-terminal sequences of the L RNA segment. Direct dideoxy sequencing, using L RNA as a template and four different synthetic oligonucleotides as primers, was used to obtain the 5'-terminal sequence (Chapter 3) and to verify internal sequences (Fig.5.1).

5.3.2 Characteristics of the TSWV L RNA

The complete nucleotide sequence of the TSWV L RNA is shown in Fig.5.3. The RNA is 8897 nucleotides long, with a base composition of 28.7% A, 37.8% U, 19.0% C and 14.5% G. The length is in rather good agreement with the previously estimated sizes, deduced from electrophoretic mobility (Chapter 1 and Van den Hurk et al., 1977). The L RNA exhibits complementarity between its 3' and 5' ends for 62 nucleotides at the 5' end and to 66 nucleotides at the 3' end, similar in range to the complementary termini of the S RNA (Chapter 4). The resulting 'panhandle' structure has a free energy of $\triangle G = -217.1$ kJ/mol (Fig.3.4).

80
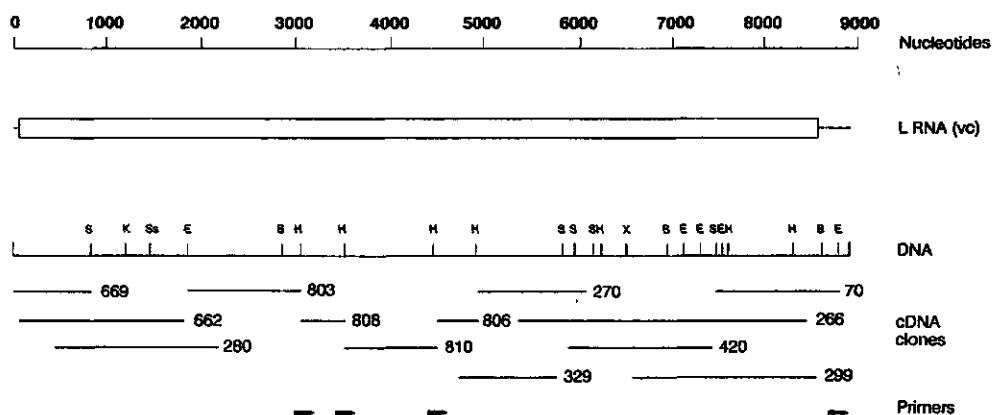
Fig.5.1: Cloning strategy for the TSWV L RNA segment. The viral complementary RNA strand is repre-sented. The box corresponds with the large ORF. The arrows represent the synthetic oligonucleotides used for primer extension sequencing on the L RNA as a template. The numbers correspond with the cDNA clones used. Restriction enzymes are abbreviated as follows: Bg, BglII; E, EcoRI; H, HindIII; K, KpnI; S, SphI; Ss, SstI; X, XbaI.

5.3.3 Predicted gene product encoded by TSWV L RNA

Analysis of the six reading frames of the viral and viral complementary RNA strand revealed only one large ORF, located on the viral complementary RNA strand (Fig.5.2). This ORF starts with an AUG codon at position 34 and extends to an UAA stopcodon at position 8659, hence the non-coding regions of the plus sense RNA are 33 bases long at the 5' end and 235 bases at the 3' end.

81
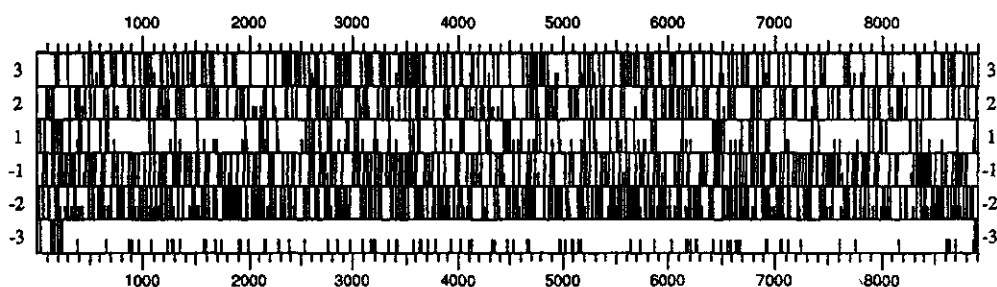
Fig.5.2: Distribution of translation initiation (short vertical bars) and termination (long vertical bars) codons in the three possible reading frames of the viral (1, 2 and 3) and viral complementary (-1, -2 and -3) L RNA strands.

The amino acid sequence derived from this ORF is shown in Fig.5.3. The sequence of the predicted gene product is 2875 amino acids long with an M_r of 331.5K. Analysis of the amino acid sequence of the predicted protein reveals several short hydrophobic regions (Kyte & Doolittle, 1982) and a very acidic carboxy-terminus, as can be seen by the large number of aspartic acid (D) and glutamic acid (E) residues (Fig.5.3). A search in the EMBL protein and nucleotide sequence database showed that the predicted protein encoded by TSWV L RNA is homologous to the L proteins of the animal-infecting Bunyaviridae. Hence, it can be deduced that the L RNA segment of TSWV encodes the L protein.

82

```
                                    M  N  I  Q  K  I  Q  K  L  I  E  N  G  T  T  L  L  L  S  I  E  D
1    AGAGCAAUCA GGUAACAACG AUUUUAAGCA AACAUGAACA UCCAGAAAAU ACAAAAAUUA AUAGAAAUG GAACCACUUU ACUGUGUCU AUUGAGGAUU

      C  V  G  S    N  H  D  L  A  L    D  L  R  K  R  N  S  D  E  I  P  E  D  V  I  I  N  N  N  A  K  N  Y  E
101  GUGUAGGUUC UAACCACGAU CUAGCUUUGG AUUUACAUAA GAGAAAUAGU GAUGCAAUCC CAGAACGAUG DAUUAUAAAU AAUAAUGCAA AAAAUUAUGA

      T  M  R  E  L  L  V  K  I  T  A  D  G  E  G  L  N  K  G  H  A  T  V  D  V  K  K  L  S  E  H  V  S
201  GACAAAUGAGA GAGUUAAUUG UCAAAAUCAC UGCUGAUGGU GAAGGACUAA ACAAAGGGAU GGCAACUGUG CAUGUCAAAA AGCUAAGUGA GAUGGUCUCU

      L  F  E  Q  K  Y  L  E  T  E  L  A  R  H  D  I  P  G  E  L  I  S  R  B  L  R  I  K  P  K  Q  R  N
301  CUGUUUGAGC AAAAAUACCU AGAAACAGAG UUAGCAAGGC AUGACAUUUU UGGAGAGCUG AUCUCCAGGC ACCUGAGAAU AAAGCCCAAA CAAAGAAAUG

      E  V  R  I  E  B  A  L  R  E  Y  L  D  E  L  N  K  K  S  C  I  N  K  L  S  D  D  K  F  E  R  I  N  K
401  AAGUGGAGAU AGAGCAUGCA CUAAGAGAAU AUCUGGAUGA ACUCAACAAA AAGUCCUGCA UUAACAAGCU CUCGAUGAU GAGUUUGAGA GAAUAAAUAA

      E  Y  V  A  T  N  A  T  P  D  N  Y  V  I  Y  K  E  S  K  N  S  E  L  C  L  I  I  Y  D  N  K  I  S
501  AGAAUAUGUA GCAACUAAUG CCACCCCUGA UAACUAUGUG AUAUAUAAAG AAUCAAAAA CAGUGAGCUU UGUUUAAUCA UUUAUGAUUG GAAAAUAAUCU

      V  D  A  R  T  E  T  K  Q  W  R  N  T  Y  K  N  I  W  K  S  F  K  D  I  K  V  N  C  K  P  F  L  E
601  GUCCAUGCCA GGACUGAAAC CAAACAAUGG ACAAAAUACCU ACAACAAUAU UUCGAAAUCU UUCAAACAUA UAAAAGUCAA UGGAAAGCCA UUCCUCGAAG

      E  H  P  V  F  V  S  I  V  I  L  K  P  I  A  G  M  P  I  T  V  T  S  S  R  V  L  E  K  F  E  D  S  P
701  AGCAUCCCGU UUUCGUUUCU AUAGUUAUAU UGAAACCUAU UGCUGGGAUG CCAAUCACUG UUACUAGUAG CAGGGUUUUG GAGAAAUUCG AAGAUUCUCCC

      S  A  L  H  G  E  R  I  K  H  A  K  N  A  K  L  L  N  I  S  Y  V  G  Q  I  V  G  T  T  P  T  V  V
801  AUCAGCAUUG CACGGAGAAA GAAUAAAGCA UGCUAAAAAU GCCAAAUUGC UAAAUAUUUC UUAUGUUGGG CAAAUAGUUG GAACCACACC CACAGUGCUG

      R  N  Y  T  A  N  T  Q  R  I  K  S  E  V  R  G  I  L  C  D  D  F  C  S  K  D  V  F  F  S  H  W  T
901  AGAAACUAUU AUGCAAACAC UCAAAGAAUC AAAUCUGAAG UCAGAGGAAU CUUAGGUGAU GAUUUUUGAU CUAAAGAUGU GUUUUUCAGU CACUGGACCA

      S  K  Y  K  E  R  N  P  T  E  I  A  T  S  E  D  I  E  R  I  I  D  S  L  V  T  D  E  I  P  R  E  K  I
1001 GCAAAUACAA AGAAAGAAAU CCUACUAGAA UAGCCUAUUC CGAAGAAAUU UCGAUGGUAU UGUAACCGAU UGUUACGAGC UAAAAUCCCUA GAGACGGAAU

      I  H  F  L  F  G  N  F  C  F  H  I  E  T  M  N  D  Q  H  I  A  D  K  F  K  G  Y  Q  N  S  C  I  N
1101 AAUACAUUUG UUGUUUUGGAA AUUUCUGUUU CCACAUUGAA ACAAUCAAGG ACCACCCAUAU AGCUGACAAA UUUAAAGGGU ACCAAAACUC UUGUAUCAAU

      L  K  I  E  P  K  A  D  L  A  D  L  K  D  H  L  I  Q  K  Q  Q  I  W  E  S  L  Y  G  K  H  L  F  K
1201 UUAAAAAUAG ACCCAAAAGC UGAUUUAGCU GAUUUGAAAG ACCACUUAUA UCCAAAACAG CAAAUAUGGG AAUCUCUGUA UGGAAAAACAC UUGCAAGAAG

      I  M  L  R  I  R  E  K  K  R  K  E  K  E  I  P  D  I  T  T  A  F  N  Q  N  A  A  E  Y  K  E  R  Y  P
1301 UCAUCGCUUAG AAUUAGAGAA AAAAAGAGAA AAGAAAAAGA GAUAUCCAGAC UAUAACCACAG CUUUUAACCA GAAUGCUGCU GAAUAAUAG AAAAGGUAUCC

      N  C  F  N  D  L  S  K  L  K  L  T  F  H  D  L  V  P  S  L  K  I  E  L  S  S  E  V  D  Y  N  N  A
1401 UAACUGUUGC AAUCAUCUCU CUCAACUAAA ACUAACUUUC CAUGACGUUG UCCCCAGUUU GAAGAUAGAA UUGAGCUCAG AGGUAGAUUA CAACAACGCA

      I  I  N  K  F  R  E  S  F  R  S  S  S  R  V  I  Y  N  S  P  Y  S  S  I  N  N  Q  T  N  K  A  R  D
1501 AUUAUUAACA ACUUUCCGGCA GACCUUCAAA AGUUCUUCAA GCGUUAUUUA UAAUAGCCCA UAUAGUAGCA UAAAUAACCA AACAAAUAAA GCAAGAGAUA

      I  T  N  L  V  R  L  C  L  A  E  L  S  C  D  T  Y  K  M  E  K  Q  E  L  E  D  E  I  D  I  N  T  G  S
1601 UAACAAACUU AGUUAGACUG UGUUUUACCAG AGCUAAGUUG UGAUACAACG AAAAUGGAAA AGCAGGAAUC UGAAGAUGAA AUAGAUAUAA ACACCGGAG

      I  K  V  E  R  T  K  K  S  K  E  W  N  K  Q  G  S  C  L  T  R  N  K  N  E  F  C  M  K  D  T  G  R
1701 UAUCAAAGUU GAGAGAACAA AAAACUCUAA AGAAUGGAAU AAGCAAGGUU CCUGUUUUAC CAGAAACAAA AAUGAAUUUU GCAUGAAAGA UACAGGCAGC

      E  N  K  T  T  Y  F  R  G  L  A  V  H  N  I  G  M  S  K  K  R  I  L  K  R  E  E  I  K  E  R  I
1801 GACAAACAAAA CUACCUAUUU UAAAGGCUUA GCAGUAAUGA AUAUAGGAAU GAGUUCUAAG AAAAGAAUUC UAAAAAAGA AGAAAUAAAA GAAAGGAUCU

      S  K  G  L  E  Y  D  T  S  E  R  Q  A  D  P  N  D  D  Y  S  S  I  D  M  S  S  L  T  H  R  K  L  I
1901 CUAAAGGCUU GGAAUAUGAU ACCUCUGAAA GGCAGGCUGA CCCAAAUGAU GAUUAUUCA GUAAUCGAUA UGUCUUCUCUG ACUCAUAGAA AAAAACUGAU

      R  H  D  N  D  D  S  L  S  G  K  R  F  K  G  S  F  F  L  L  H  N  F  N  I  I  E  D  G  K  I  T  S
2001 AAGGCAUGAC AAUGCAUGAA GCUUAAAGUGG UAAAAGAUGU UUUUUCUUCU UUUUUUAUCA UCAAUAAUUU AAUAUAAUAG AGGAACGGUAA GAUCACAUCU

      V  F  N  N  Y  A  K  N  P  E  C  L  Y  I  Q  D  S  V  L  K  T  E  L  E  T  C  K  K  I  N  K  L  C
2101 GUUUUCAAUA AUUAUGCUAA AAAUCCUGAA UGCUUGUACA UUCAAGAAUUC AGUACUGAAG ACUCAAUUAG AGACUUGCAA AAAGAUAAAA AAAUUAUGCA

      N  D  L  A  I  Y  H  Y  S  E  D  N  M  Q  F  S  K  G  L  H  V  A  Q  R  Y  H  T  K  E  S  F  K  I  L
2201 AUGAUCUAGC CAUUUAUCAU UACUCUGAAG ACAUGAUGCA AUUCUCCAAA GCUUUAAUGG UGGCUCAGAG GUACAGAGACU AAAGAAAGUU UCAAGAUAUU

      T  T  A  N  T  S  M  M  L  L  A  F  K  G  D  G  H  N  T  G  G  S  G  V  P  Y  I  A  L  H  I  V  D
2301 AACCACAGCA AAUACUAGCA UGAUGCUAUU AGCAUUCAAA GGGAAGGAUG GAACACGG AGGAUCGGGAGU GUUCCUUACA UAGCAUUGCA UAUAGUGGAU

      R  D  M  S  D  H  F  N  I  C  Y  T  K  E  I  Y  S  Y  F  R  S  G  S  N  Y  I  Y  I  H  R  P  Q  R
2401 GAAGACAUGU CAGAUCAUUU UAACAUAUGU UAUACAAAAG AAAUUUAUAG CUAUUUCCGA AGUGGUAGUA AUUACAUUUA UAUACACAGG CCGCAGAGAC

      L  H  Q  V  R  L  L  R  L  F  K  T  P  S  K  V  P  V  C  F  P  Q  F  S  E  R  A  N  I  G  K  S  L
2501 UAAACCAGGU GAGGCUGCUG AGCCUUUUCA AAACGCCUAG UAAAGUUCCU GUAUGUUUUC CACAAUUUUC AAAGAAAGCA AAUGAAAUCG GAAAAUCGCU

      K  N  K  D  I  E  K  V  N  L  F  S  H  T  M  T  V  K  Q  I  L  I  N  I  V  F  S  S  V  M  I  G  T
2601 AAAAAAUAAA GAUAUAGAAA AAGUAAAUCU CUUUUCUCAUG ACAAUGACUG UAAAACAGAU AUUAAUAAAU AUUGUUUUAU CAUCUGUCAU GAUAGGAACU

      V  T  K  L  S  R  H  G  I  F  D  F  M  R  Y  A  G  F  L  R  L  S  D  Y  S  N  I  K  E  Y  I  R  D
2701 GUGACAAAGC UCAGUCGACAAU GGGAAUUUUU GAUUUCAUGC GGUAUGCAGG UUUUUUGCGA CUAUCCGAUU AUCUAAACAU AAAAGAAUAC AUUAGAGACA

      K  S  D  P  D  I  T  N  C  C  R  Y  L  F  R  N  G  I  K  K  L  L  F  R  M  E  D  L  N  L  S  T  N  A
2801 AAUCUGAUCC UGAUAUAACU AACUGUGGCA GAUAUCUAUU UCGUAAUGGA AUCAAAAAC UAUUGUUCAG AAUGGAAGAU CUCAAUUUAA GCACAAAUGC

      K  P  V  V  V  D  H  E  N  D  I  I  G  G  I  T  N  L  N  I  K  C  P  I  T  G  S  T  L  L  T  L  E
2901 CAAGCCUGUU GUUGUGGACC ACGAAAAUGA UAUUAUAGGA GGAAUCACAA ACUUGAACAUA AAAAUGUCCU AUAACAGGAU CAACUCUAC UGACAUUGAC

      D  L  Y  N  N  V  Y  L  A  I  Y  H  H  P  K  S  L  H  N  H  V  H  N  L  T  S  L  L  N  L  P  A  E
3001 GACCUGUACA AUAAUGUUUA UUUGGCUAUU UAUAUGAUGC CUAAAUCACU GCACAAUCAU GUUCACAAUC UAACAAGCUU AUUAAAUCUC CCUGCUGAGU

      V  E  L  K  P  R  K  E  L  G  F  N  I  F  E  D  I  Y  P  K  K  A  M  F  D  D  K  D  L  F  S  I  N  G
3101 GGGAGCCUAA GUUCAGAAAA GAAUUAGGUU UCAACAUAUU UGAAGACAUA UACCCUAAGA AAGCAAUGUU UGAUGACAAA GACCUAUUCU CCAUUAAUCG
```

```
          A  L  N  V  K  A  L  S  D  Y   Y  L  G  N  I  E   N  V  G  L   N  R  S  X  I  E   N  K  E  D    F  L  S
3201  AGCUUUGAAC CUGAAAGCAU UAUCUGAUUA CUAUCUACGA AAUAUACAAA AUCUUGGUUU AAUGAGAUCA GAAAUAGAAA AUAAAGAACA UUUCCUAAGC

          P  C  Y   X  I  S  T  L  X  S   S  X  X  C  S  Q  S   N  I  I   S  T  D   E  I  I  E   C  L  Q   D  A  K
3301  CCUUGUUAUA AAAUAUCUAC UUUAAAAUCU UCAAAAAAU CCUCACAAAU AGUACUGAUC AGAUAAUGAC UCUCUUCGAU GAUGCAAAGA

          I  Q  D  I   E  N  W  X  G  N   N  L  A  I   I  K  G   L  I  R   T  Y  N  E   E  K  N   R  L  V   E  F  F  E
3401  UUCAAGAUAU AGAAAUUUGG AAAGGAAAUA ACCUAGCUAU UAUAAAAGGG CUUAUAAGA CCUACAUGA GGAGAAAAAU ACAUUGGUG AAUUUUUUGA

          D  N  C   V  N  S   L  Y  L  I   E  K  L  X   E  I  I  S  S  G   S  I  T   V  G  X   S  V  T  S   X  F  I
3501  AGAUAAUUGU GUCAAUUCAU UAUAUCUUAU AGAAAAGCUU AAAGAGAUAA UUAGUAGUGG AUCAAUAACU GUAGGGAAAU CUGUAACAUC UAAAUUCAUA

          R  N  W   N  P  L  T   V  E  T   Y  L  K   T  K  L  Y   Y  R  N   N  V  T   V  L  K  S   K  K  V   S  E  E
3601  AGAAACAAUC AUCCUUUAAC AGUAGAAACA UAUCUCAAAA CAAAACUAUA UUAUAGAAAU AAUGUAACAG UUUUAAAGUC UAAAAAAGUG UCAGAGGAGC

          L  Y  D  L   V  E  Q   F  B  D   M  M  E  I   O  L  D   S  V  N   N  L  G  K   G  T  E   G  K  K   L  T  F  L
3701  UCUAUGACCU UGUAAAACAG UUCCAUGACA UGAUGGAAAU AGACCUAGAU UCUGUUAUGA ACCUCGGGAA AGGUACAGAA GGAAAAAAC UCACAUUCUU

          Q  N  L   E  F  V   N  S  X  A   K  N  V   T  G  S   V  D  F  L   V  S  V   F  E  K   M  Q  R  Y   K  T  D
3801  GCAGAAUGCUU GAAUUUGUCA UGUCCAAGGC UAAAAAUGUC ACCGGGUCUG UAGAUUUCU UUGUUUCGUU UUUGAAAAAA UGCAGAGAAC CAAAACAGAC

          R  E  I   Y  L  M  S   M  K  V   K  N  M   L  Y  F  I   E  H  T   F  K  H   V  A  Q  S   D  P  S   K  A  I
3901  AGAGAAAUAU ACUUGAUGAG CAUGAAAGUG AAAAUGAUGC UUUAUUUUAU AGAGCACACU UUCAAACAUG UAGCGCAGAG UGAUCCAAGA UCUUAUGCAU

          S  I  S  G   D  N  K   I  R  A  L   S  T  L   S  L  D   T  I  T   S  Y  N  D   I  L  N   K  N  S   K  K  S  R
4001  CUAUAAGUGG AGACAAUAAA AUAAGAGCAC UUUCUACCUU AUCUUUGGAC ACAAUCACCU CUUACAAUGA UAUUUUAAAC AAAAAUUCAA AGAAGUCAAG

          L  A  F  L  S  A   D  Q  S  K   W  S  A  S   G  L  T  T  Y  X   Y  V  L   A  I  I   L  N  P  I   L  T  T
4101  AUUUGCUUUC CUAUCUGCAG AUCAGUCGAA AUGGUCGGCA UCAGGCCUUA CCACCUAUAA AUAUGUUUUA GCAUCAUAU UAAAUCCAAU UUUAACUACU

          G  E  A   S  L  H  I   E  C  I   L  M  Y   V  K  L  K   K  V  C   I  P  T   D  I  F  L   N  L  R   K  A  Q
4201  GGUGAAGCUA GCUUAAUGAU AGAAUGCAUC UUAAUGUAC GUAAAAUUGA AGAAGGUUUGU AUACCAACGG AUAUUUUUUU GAAUCUAAGA AAAGCUCAA

          Q  T  F  G   K  N  E   T  A  I   G  L  L  T   K  G  L   T  T  N   T  Y  P  V   S  M  N   W  L  Q   G  N  L  N
4301  AAACUUUUGG GGAAAAUGAA ACUGCCAUAG GACUUUUGAC CAAAGGCUUG ACGACAAACA CAUACCCUGU UAGCAUGAAU UGGUUGCAAG GCAAUUUAAA

          Y  L  S   S  V  T   H  S  C  A   M  K  A  Y   H  N  T   L  E  C   Y  R  N   C  D  F   Q  T  R  W   I  V  H
4401  UUAUCUGUCU UCUGUUUAUC ACUCUUGUGC AAUGAAAGCU UAUCACAACA CUUUGGAAUG UUACAAAAAC UGUGAUUUCC AACUACAAUG GAUUGUCCAC

          S  D  D   H  A  T  S   L  I  A   S  G  X   V  D  K  H   L  T  D   P  S  S   S  S  L  P   E  M  L   F  R  S
4501  UCUGAACAUA AUGCAACAUC AUUAAUAGCC AGUGGAGAGG UUGAUAAAAU GCUGACAGAC UUUUCAAGCU CAUCUCUGCC AGAAAUGUUG AGAUCUCAA

          I  K  A  H   F  K  S   F  C  I  T   L  N  F   K  K  S   Y  A  S   S  S  E  V   E  F  I   S  E  R   I  S  R  W
4601  UUGAAGCUCA UUUCAAAACU UUUUGCAUAA CUUUCAACCC AAAAAACAGU UAUCCUUCUU CAUCAGAAGU AGAGUUUAUA UCUGAAAGAA UUAGUAAAUG

          S  D  T   S  S  L   L  Q  A  F   S  K  L   L  H  R   I  F  A  Y   K  L  F   D  D  L   M  S  L  S   I  H  V
4701  CAGCGGAUAU UCCUCUCUAU UGCAGGCAUU UAGCAAACUU UUGCACAGAA UCUUUGCCAUA AAGUUAUUU GAUGAUCUAA UGUCACUCAG UAUACAUGUU

          T  M  L   L  R  K  C   C  P  N   E  V  I   F  F  A  Y   G  A  V   Q  V  Q   A  L  S  I   Y  S  M   L  P  G
4801  ACAAUGCUUC UGAGAAAAGG CUGUCCUAAU GAAGUAAUCU UUUUUGCAUA CGGUGCUGUU CAAGUACAAG CGUUAAGCAU CUAUUCAAUG CUUCCUGGGU

          E  V  N  D   S  I  R   I  F  N   K  L  G  V   S  L  K   S  N  E   I  P  T  N   M  G  G   W  L  T   S  P  I  E
4901  AAGUGAAUGA UAGUAUUAGA AUUUUUAACA AGCUUGGAGU AAGUUUAAAA UCAAAUGAAA UUCCCACAAA CAUGGGGGGC UGGUUGACCU CUCCUAUAGA

          P  L  S   I  L  G   P  S  S  N   D  Q  I  I   Y  Y  N  V  I  R   D  F  L   N  K  K   S  L  E  K   Y  V  K  D
4501  GCCGUUGUCU AUAUUAGGUC CAUCAUCAAA UGAUCAAAUA AUCUAUUACA AUGUGAUAAG AGAUUUUUUG AACAAAAAA GUUUAGAAGA ACUAAAAGAU

          S  V  S   S  S  S  Y  L  Q  N   R  F  R   E  L  K  E   K  Y  E   R  G  T   L  E  E  K   D  K  K   H  I  F
5101  AGUGUCUCUU CUUCCAGUUA UCUACAGAAU CGAUUCAGAG AACUCAAAGA AAAGUAUGAA AGAGGAACUC UGGAAGAAAA AGAUAAAAAG CAUAUUUUC

          L  I  N  L   F  E  K   A  S  V   S  E  D  S   D  V  L   T  I  G   M  K  F  Q   T  N  L   T  Q  I   I  K  L  P
5201  UUAUCAAUCU GUUUGAGAAA GCAUCACUGU CUGAACAUUC AGACCUUCUA ACAAUUGGGA UGAAAUUUCA AACUAUCGUUA ACUCAGAUUA UAAAAUUACC

          N  F  I   N  E  N   A  L  N  K   M  S  S   Y  X  D   F  S  K  L   Y  P  N   L  K  K   M  E  D  L   Y  K  S
5301  UAAUUUUAUA AAUGAGAAUG CUUUAAACAA GAUGUCAAGU UAUAAAGAUU UUUCAAAACU UUAAAAAAGA AUGAAGAUUU AUAUAAAAGC

          T  K  N   L  K  I  D   E  D  A   V  L  E   E  D  E  L   Y  K  K   I  A  E   S  L  E  N   E  S  V   H  D  I
5401  ACUAAGAACU UAAAGCAUUAG A CCAGGAUGCU GUUUUAGAAG AAGACGAGUU AUAUAAAAAG AUUGCCAUCUA GUCUGGAAAAU GGAAUCUGUA CAUGACAUAA

          M  I  K  N   P  E  T   I  L  I   A  P  L  N   D  R  D   F  L  L   S  Q  L  F   M  Y  T   S  P  S   K  R  N  Q
5501  UGAUAAAAAA UCCUGAAACA AUUCUCAUAG CACCAUUGAA UGAUAGAGAU UUUUUACUUA GUCAGCUGUU CAUGUACACA AGCCCUUCUA AAAGAAACCA

          L  S  N   Q  S  T   E  K  L  A   L  D  R   V  L  R   S  K  A  R   T  F  V   N  I  S   S  T  V  K   H  T  Y
5601  GUUAUCGAAC CAAUCUACAG AGAAACUUGG UUUAGAUAGA GUGUUAAGGU CAAAAGCUAG AACAUUUGUA AACAUUUCUU CCACUGUUAA GAUGACGUAU

          E  K  N   H  E  K  X   I  L  E   M  L  K   F  D  L  D   S  Y  C   S  F  K   T  C  V  N   L  V  I   K  D  V
5701  GAAGAAAAUC AUGAAAAAAA AAUCCUUGAAA AUGCUAAAAU UUGAUUUGGA UUCAUAUUGU UCAUUUAAAA CAUGUGUAAA UCUUGUUAUC AAAGAUGUUA

          N  F  S  N   L  I  F   I  L  D   S  A  F  P   C  E  S   R  K  R   D  N  Y  N   F  R  W   F  Q  T   E  R  W  I
5801  AUUUCAGCAU GCUGAUUCCA AUAUUAGAUU CUGCAUACCC UUGUGAAUCU AGGAAAAGC AUAACUACAA UUUCAGGUGG UUUCAGACUG AGAUGGAU

          P  V  V   E  G  S   P  G  L  V   V  H  H   A  V  Y   G  S  N  Y   I  K  N   L  G  L   K  N  I  P   L  T  D
3901  ACCGUUGUU GAAGGCUCUC CGGGACUAGU AGUAGAUCAU GCUGUUUAUG GAUCAAAUUA UAUAGAGAAU UUAGGUUUAA AAAACAUCCC UCUAACAGAC

          D  S  I   N  V  L  T   S  T  F   G  T  G   L  I  N  E   D  V  K   S  L  V   N  G  K  D   S  F  E   T  E  A
6001  GAUAGCAUUA AUGUUUUAAC AAGCACGUUU GGAACAGGUU UAAUUAAUGA AGACGUUAAG UCCUUAGUUA AUGGCAAAGA CAGCUUCGAA ACAGAGGCUU

          F  S  N   N  E  C   Q  R  L   V  K  A  C   N  Y  M   I  A  A   Q  N  R  L   L  A  I   N  T  C   F  T  R  R
6101  UUAGCAAUUC UAAUGAAUGU CAAAGAUUGG UCAAGCAUG CAAUUAUAUG AUAGCACCAC AAAACAGCCU UUUAGCAAUU AACACACGCU UUACUAGGAA

          S  F  P   F  Y  S   K  F  N  L   G  R  G   F  I  S   N  T  L  A   L  L  S   T  I  Y   S  K  E  E   S  Y  H
6201  AAGCUUUCCC UUCUAUUCUA AGUUCAAUCU AGGGACGAGG UUUAUCCAA ACACAUUAGC UCUCCUAUCC ACCAUCUACA GUAAAGAAGA AUCCAUCAU

          F  V  S   T  A  S  Y   K  L  D   K  T  I   R  T  V  V   S  A  Q   Q  D  N   N  L  E  K   I  L  D   T  A  V
6301  UUUGUUUCUA CAGCUAGUUA UAAAUUGGAC AAAACUAUCA GAACUGUGGU AAGUGCUCAG CAAGAUAUGA ACUUAGAGAA AAUACUGGAC ACUGCUGUAU
```

```
        Y  I  S  D    K  L  Q    S  L  F    P  T  I  T    R  E  D    I  V  L    I  L  Q  N    V  C  L    D  S  K    P  I  W  Q
6401  ACAUAUCAGA UAAAUUGCAG UCACUUUUCC CAACAAUUAC AAGAGAGGAU AUAGUUUUAA UAUUGCAAAA UGUUUGCCUU GACAGUAAAC CUAUAUGGCA

        S  L  E    D  K  H    K  K  I  N    N  S  T  A  S  G    F  T  V  S    N  V  I    L  S  B    N  S  E  L    N  T  I
6501  GAGUCUAGAA GACAAAAUGA AAAACAUUAA CAAUUCAACA GCAAGUCGCU UCACAGUGUC AAAUGUGAUU CUAUACAUA ACAGUGAAUU GAACACAAUC

        Q  K  Q    I  V  W  M    W  N  M    G  L  C    S  B  R  T    L  D  F    V  I  R    Y  I  R  R    R  D  V    R  Y  V
6601  CAGAAACAAA UUGUCUGGAU GUGGAACAUC GGUUUGUGUU CUCACAGAAC AUUAGCAUUU GUUAUCAGGU AUAGUAGAAG AAGCGGAUCGA AGAUAUCUAA

        K  T  E  E    Q  D  E    S  G  M    Y  V  S  G    T  M  Y    K  I  G    I  M  T  R    S  C  Y    V  E  L    I  A  S  D
6701  AAACUGAAGA ACAAGAUGAA UCAGGAAAUU AUGUCUCUGG AACUAUGUAC AAAAUAGGGA UCAUGACAAG AAGCUGCUAU GUGGAAUUGA UAGCAUCUGA

        Q  D  V    A  V  S    L  R  T  P    F  E  I    L  N  K    R  E  Y  L    F  D  T    Y  R  E    S  I  E  E    L  L  A
6801  UCAAGAUGUA GCAGUUUCUU UGAGAACACC AUUUGAGAUA UUGAAUGAAA GAGAGUAUCU UUUUGACACA UACAGAGAAA GUAUAGAGAA AUUAUUGGCA

        E  I  H    F  D  K  V    N  I  I    N  Q  T    T  T  D  C    F  L  R    T  R  R    S  C  I  R    N  T  T    D  N  K
6901  GAAAUUAUGU UUGAUAAAGU GAACAUAAUA AAUCAAACAA CCACAGAUUG UUUUCUUAGA ACCAGGAGAU CUUGCAUCAG AAUGACCACA GACAACAAAA

        H  I  V  K    V  N  A    T  S  R    Q  I  R  L    E  N  V    K  L  V    V  K  I  K    Y  E  N    V  N  S    D  V  W  D
7001  UCAUUGUAAA GGUUAAUGCU ACAUCAAGAC AAAUAAGACU AGAGAAUGUA AAAUUAGUUG UAAAGAUAAA AUAUGAAAAU GUGAAUUCCG AUGUAUGGGA

        I  I  E    S  Q  E    S  L  V  L    R  L  P    E  V  G    E  F  F  S    D  M  Y    K  T  A    D  S  E  T    K  T  I
7101  UAAUUAUAGA AGCCAAAAAU CUCUAGUCUU AAGGCUCCCU GAACUAGGGG AAUUUUUCUC UCAUAUGUAU AAAACUGCAG ACUCUGAAAC UGAAACAAUC

        K  T  I    K  N  R  L    H  T  S    L  T  F    I  E  A  F    C  N  L    S  Q  Q    I  K  E  I    V  D  D    D  I  R
7201  AAAACCAUAA AAAACAGGCU UAUGCAUUCU UUAACUUUCA UAGAAGCCUU UGGAAACUUA UCACACCAGA UCAAAGAGAU UGUAGCAUCAU GAUAUCAGAC

        E  T  M  D    E  F  L    M  N  I    R  D  T  C    L  E  G    L  E  N    C  K  S  V    E  E  Y    D  S  Y    L  D  E  N
7301  AAACGAUGGA UGAAUUCUUA AUGAACAUCC GCGAUACCUG CUUAGAAGGU UUGGAAAACU GCAAAAGUGU GGAAGAAUAU GAUUCAUAUU UUGAUGAAAA

        G  F  N    D  T  V    E  L  F  E    N  L  L    R  T  H    D  N  F  R    N  R  Y    S  P  L    F  S  R  I    V  D  K
7401  UGGAUUUAAU GACACAGUAG AACUAUUCCA AAACUUGCUA AGAACACAUC ACAACUUUGA AAAUGAGUAU AGUCCUCUUU UUUCAGAGAU UGUUGACAAA

        A  K  Q    Y  T  R  D    L  E  G    F  K  E    I  L  L  H    L  R  K  Y    S  L  I    H  D  A  S    G  F  K    S  Y  R
7501  GCAAAACAGU AUACUAGACA UUUACAAGGU UUCAAAGAAA UACUCCUCAU UCUCUCAAUA AUCAUGCAUC AGGAUUUAAA AGCUAUAGAG

        A  T  G  M    H  A  V    E  L  H    A  K  K  E    I  E  I    G  E  F    N  L  L  G    M  I  Q    L  I  K    A  C  E  T
7601  CCACUGGAAU GCAUGCUGUU GAGCUAAUGG CAAAAAAGCA CAUAGCAAUG GGGCAAUUCA ACUUGUUAGG AAUGAUCCAA UUGAUUAAAG CUUGCGAAAC

        C  H  N    N  D  S    I  L  N  L    A  S  L    R  N  V    L  S  R  T    Y  A  T    F  G  R    R  I  R  L    D  H  D
7701  AUGCCACAAC AAUGACUCUA UAUUAAACUU AGCAAGUUUA AGGAAUGUUC UUAGCAGGAC AUAUGCCACA UUUUGGGAGGA GAAUAAGAUU GGAUCAUGAU

        L  D  L    Q  N  N  L    M  E  K    S  Y  D    F  K  T  L    V  L  P    E  I  K    L  S  E  L    S  R  E    I  L  K
7801  CUGGACUUGC AAAACAACUU AAUGGAAAAA AGUUAUGAUU UCAAGACGCU GGUUUUUACCA GAAAUAAAAU UAUCAGAACU AUCUAGGGAA AUACUGAAAG

        E  N  G  F    V  I  S    G  E  N    L  K  H  D    R  S  D    E  E  F    V  G  L  A    S  F  N    V  L  R    L  D  E  E
7901  AAAAUGGGUU UGUUAUAUCU GGAGAGAAUC UAAAAAUGGA UAGGGUCUGAU GAAGAAUUUG UAGGUCUAGC CAGUUUUAAU GUGUUGAGGU UAGAUGAAGAA

        E  M  Y    E  G  L    I  K  E  H    E  I  K    R  K  K    K  G  F  L    F  P  A    N  T  L    L  L  S  E    L  I  E
8001  AGAAAUGUAU GAAGGUUUGA UCAAACAAAU GAAAAUUAAA AGGAAAAAGA AAGGGUUUUU AUUUCCAGCA AACACACUUC UACUAAGUGA GUUGAUAAAG

        F  L  I    G  G  I  E    G  T  S    F  D  I    E  T  L  L    R  N  S    F  R  P    D  I  F  S    T  D  R    L  G  R
8101  UUCUUGAUUG GAGGAAUAGAA GGGAACCAGC UUUGAUAUAG AGACAUUGUU ACGGAACCAG ACAUAUUUUC AACUGACAGA UUGGGAAGAU

        L  S  S    V  P  A    L  R  V    Y  A  T  V    Y  M  E    Y  K  N    V  N  C  P    L  N  E    I  A  D    S  L  E  G
8201  UAAGUUCCAG UGUACCUGCA CUCAAACUUU AUGCAACUGU UUAUAUGGAA UAUAAGAAUG UCAAUUGUCC UUUAAAUGAG AUAGCUGACA GCUUAGAAGG

        Y  L  K    L  T  K    S  R  S  K    E  B  P    L  S  C    R  V  K  K    A  L  I    Q  L  R    D  E  Q  S    R  T  K
8301  UUAUCUAAAA CUGACAAAAA GCAGGUCCAA GGAACAUUUC UUGUCUGGAA GAGUUAAAAA AGCUUUGAUA CAAUUAAGAG AUGAACAAUC GCGAACUAAA

        K  L  E    V  Y  K  D    I  A  N    F  L  A    R  H  P  L    C  L  S    E  K  T    L  Y  G  R    Y  Y  Y    S  D  I
8401  AAACUAGAGG UCUAUAAGGA UAUCCCAAAU UUCCUUGCUA GGCACCCACU AUGUUUAUCA GAAAAAACAU UGUAUGGAAG AUAUACCUAC UCUGCAUAUCA

        N  D  Y  I    M  Q  T    R  E  I    I  L  S  K    I  S  E    L  D  E    V  V  R  T    D  E  D    N  F  L    L  S  Y  L
8501  AUGAUUAUAU CAUCCAAACA AGAGAGAUUA UUUUGCAGUAA AAUAAGUGAG UUGGACGAGG UUGUUCAAAC AGAUGAAGAC AAUUUCUUCC UUACUUAUCU

        R  G  E    E  D  A    F  D  E  D    E  L  D    E  E  E    D  T  D  *
8601  AAGAGGGGAA GAAGAUGCCCU UUCAUGAAGA UGAGCUUGAU GAAGAAGAAC ACACAGAUUA AAAUUGAAAGU AAUGACUAAC AAUCCAUGAA UAACAGAUUA

8701  CAUAUAACUU ACAAUAUAAA UUUAUUGCUA UUUUAGAAUU AGAUUAGAAC UACUUAGCCU AAAACAAUUU CGUGAACCAA AUCUAUACGU UAUAUAAAUG

8801  UAGAGUCCCG GUAUACUUUC ACUGGAGGGA AUUCUUUAUGU AAUUUGUAAA GUCUGGCUGU GGAGAGGUUA UAUGUUUUAG UUGUACCUGA UUGCCUCU
```

Fig.5.3: The complete nucleotide sequence of TSWV L RNA (numbered from the 5' end of the viral complementary RNA strand) and its predicted gene product. The deduced amino acid sequence of the protein encoded by the viral complementary RNA is written above the RNA sequence. The asterisk (*) indicates the UAA termination codon.

85

The discrepancy between the size reported here (331.5K) and the previously estimated size (approximately 200K) may be due to the gel systems used in those experiments (Mohamed et al., 1973; Tas et al., 1977).

Computer-assisted alignment of the predicted L protein of TSWV to that of Bunyamwera virus (Elliott, 1989) revealed one internal region (approximately 1000 amino acids long) with significant (27% identity) amino acid sequence homology (Fig.5.4). Homology between TSWV and Hantaan virus L proteins, and between those of Bunyamwera and Hantaan virus however, is lower and restricted to a shorter internal stretch of approximately 200-250 residues long (Fig.5.4). For the animal-infecting Bunyaviridae it has been proposed that the L proteins represent the viral RNA polymerases. Proteins involved in transcription and replication of RNA viruses contain conserved signature sequences, such as putative polymerase, helicase or methyltransferase motifs (Kamer & Argos, 1984; Goldbach, 1987; Hodgman, 1988; Gorbalenya et al., 1989). The presence or absence of these motifs, together with other molecular characteristics such as genome structure and expression, are important determinants for evolutionary relationships between viruses and virus families (Goldbach, 1986; Strauss & Strauss, 1988; Poch et al., 1989; Candresse et al., 1990). A search for such conserved sequences in the (putative) polymerase proteins of members of the Bunyaviridae and Orthomyxoviridae reveals five types of short consensus sequences: $GDX_{1-3}K$, GXXNXXS, SDD, $FX_{10-17}KK$ and EFXSXF (Fig.5.5). These amino acid motifs are present in the region where the predicted L protein of TSWV shows sequence homology to Bunyamwera and Hantaan L protein and to influenza A protein PB1, the core polymerase of this virus (Braam et al., 1983; Krug et al., 1989). Hence, it is anticipated that the major ORF in TSWV L RNA represents the polymerase gene.

5.4 Discussion

Determination of the complete nucleotide sequence of the TSWV L RNA demonstrates that TSWV is a negative-strand RNA virus. The presented nucleotide sequence data confirm the previous conclusion, derived from the S RNA sequence, that TSWV should be considered as a member of the Bunyaviridae. Indeed, on the ICTV meeting during the 8th International Congress of Virology in Berlin (1990), TSWV has been accepted as the first member of a newly created genus, tospovirus, within the Bunyaviridae.

The TSWV L RNA segment is 8897 nucleotides long, which is significantly longer than the L RNAs of Bunyamwera (6875 nucleotides) and Hantaan virus (6530 nucleotides) (Elliott, 1989; Schmaljohn, 1990). Additional domains may be present in the gene product of TSWV L RNA, which may reflect adaptation of this bunyavirus to plants.

TSWV L RNA contains complementary ends of 62 to 66 nucleotides in length. Hence, the RNA can be folded into a stable panhandle structure (Fig.3.4), which may be involved in the appearance of circular nucleocapsids in virus particles (Peters et al., 1991), as also found for the Bunyaviridae (Raju & Kolakofski, 1989). Moreover, these terminal sequences will play an important role in genome transcription and replication (Krug et al., 1989; Parvin et al., 1989).

TSWV L RNA contains a single ORF in viral complementary sense, corresponding with a protein of predicted M_r of 331.5K. Analysis of viral RNA species in infected plant cells, indicated that this ORF is expressed by the formation of a mRNA of approximately genome length. No subgenomic RNA species derived from the L RNA could be detected (unpublished results). Remarkably, in several TSWV isolates, defective L RNA species appeared, when maintained under laboratory conditions. In line CNPH1, which has been used for sequence determination of the L RNA, a deleted form of this RNA segment accumulated, which had approximately the size of the M RNA. The genesis and implication of these defective RNA molecules for virus multiplication is currently under investigation (Resende et al., 1991).
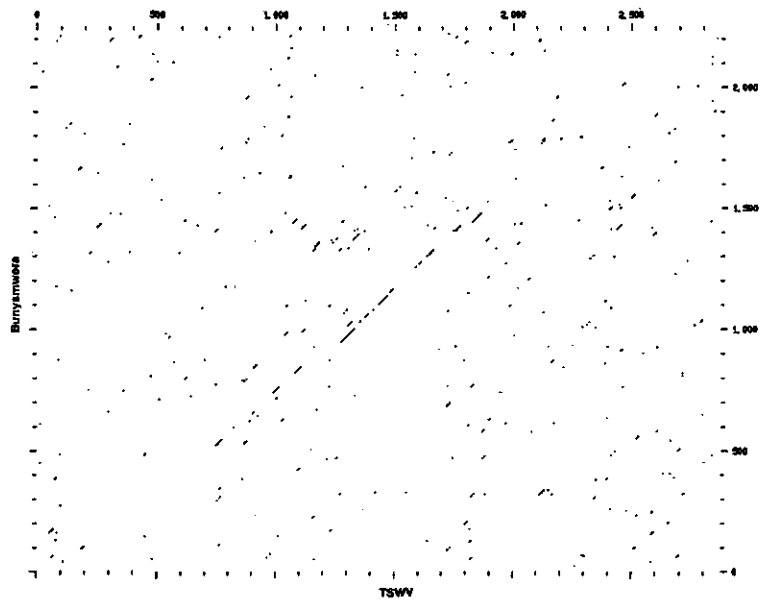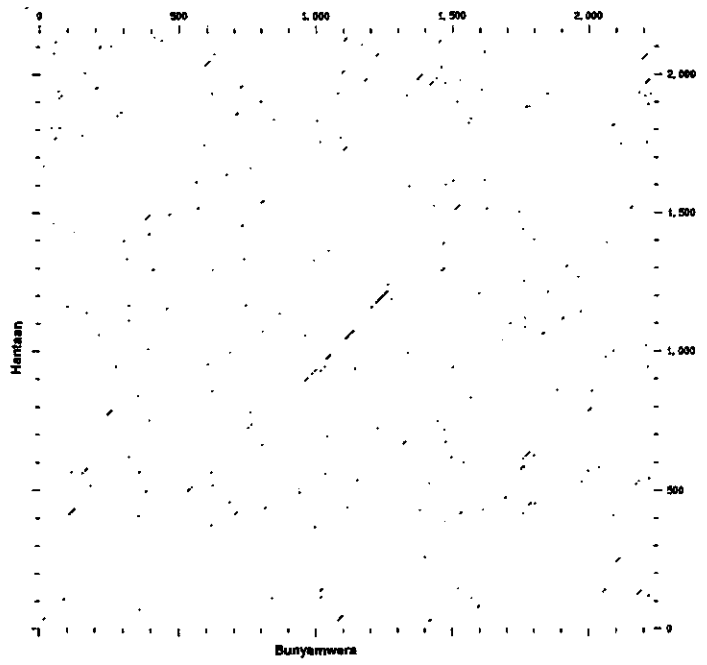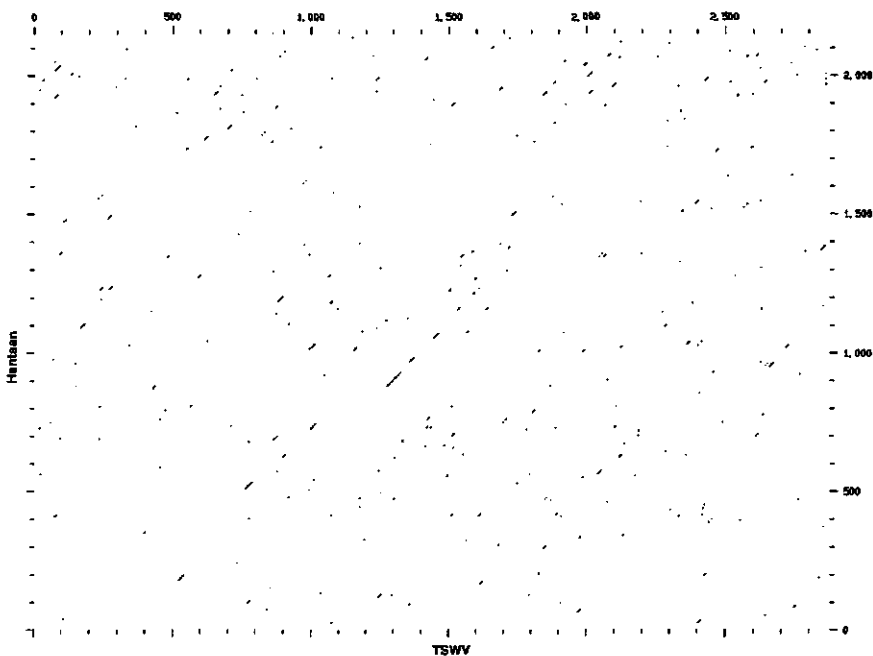
Fig.5.4: Dot plot comparison of the predicted L proteins of TSWV, Bunyamwera and Hantaan virus, using the 'compare' and 'dotplot' programs (window= 30, stringency= 16) of the GCG-package. Sequence data were obtained from: Elliott, 1989 (Bunyamwera virus) and Schmaljohn, 1990 (Hantaan virus).

The predicted 331.5K protein encoded by TSWV L RNA most likely corresponds to the viral polymerase. Comparison of (putative) RNA polymerases from TSW, Bunyamwera, Hantaan and influenza A viruses revealed the presence of amino acid sequence motifs, which are

present in all polymerases showing RNA template specificity and most likely form the active sites for RNA synthesis (Poch et al., 1989). The region in the predicted TSWV L protein, surrounding these 'polymerase' motifs, shows considerable sequence homology (approximately 27% identity) to the putative polymerase of Bunyamwera virus, but to a much lesser extend to that of Hantaan virus. All three L proteins in their turn share

89

```
InfA PB1  (308) FTITGDNT.K WNEN......  .......... ..........  ..QNPRMFLA
TSWVL     (1321) ISISGDN..K IRAL......  .STLSLDTIT SYNDILNKNS  KKSRLAFLSA
BunL      (1020) SIIKGDPS.K ALKL......  .......... ..........  ......EINA
HanL       (936) ISYGGE..KK ILAIQGALEK ALRWASGESF IELSNHKFIR  MKRKLMYVSA
                 GD...K

InfA PB1         MITYITKNQP EW.FRNVL.S IAPIMFSNKM ARLGKGYMFE  SKSMKLRTQI
TSWVL            DQSKWSASGL TTYKYVLAII LNPILTTGEA SLMIECILMY  VKLKKVCIPT
BunL             DMSKWSAQDV F.YKYFWLIA MDPILYPAEK T.RILYFMCN  YMQKLLILPD
HanL             DATKWSPGDN SAKFRRFTSM NNKLKNCVID ALKQVYKTDF  FMSRKLRNYI


InfA PB1         PAEMLASIDL KYFNESTRKK IEKIRPLLID GTASLSPGMM  MGMFNMLSTV
TSWVL            DIFLNLRKAQ QTF.GENATA IGLLTKGLTT NTYPVSMNWL  QGNLNYLSSV
BunL             DLIANILDQK RPY...NDDL ILEMTNGLNY NYVQIKRNWL  QGNFNYISSY
HanL             DSMESLDPHI KQF....... ...LDFFPDG HHGEVKGNWL  QGNLNKCSSL
                                                             G..N..S

InfA PB1         LGVSILNLGQ ......KRYT KTTYWWDGLQ SSDDFALIVN  APNHE.....
TSWVL            YHSCAMKAYH NTL.ECYK.. NCDFQTRWIV HSDDNATSLI  ASGEVDKMLT
BunL             VHSCAMLVYK DILKECMKLL DGDCLINSMV HSDDNQTSLA  IIQNKVSDQI
HanL             .RGVAMSLLF KQVWTNLFP. ELDCFFEFAH HSDDALFIYG  YLEPVDDGTD
                                                    SDD

InfA PB1         .......GIQ AG........ ....VDRFYR TCKLVGINM.  ..SKKKSYIN
TSWVL            DFSSSSLP.. .......... ....EMLFRS IEAHFKSFCI  TLNPKKSYAS
BunL             .......VIQ YA........ ....ANTFES VCLTFGCQA.  ..NMKKTYIT
HanL             WFLFVSQQIQ AGHLHWFSVN TEMWKSMFNL HEHILLLGSI  KISPKKTTVS
                                              F.. .......... ....KK


InfA PB1         RTGTFEFTSF FYRYGFVANFS  (514)
TSWVL            .SSEVEFISE RISKWSDYSSL (1562)
BunL             .HTCKEFVSL FNLHGEPLSVF (1229)
HanL             .PTNAEFLST FFEGCAVSIPF (1197)
                     EF.S. F
```

Fig.5.6: Amino acid sequence homology between the RNA polymerases of members of the Bunyaviridae and protein PB1, the core polymerase of the influenzaviruses. Residues conserved in at least three sequences are indicated in bold. Sequence data were obtained from: Yamashita et al., 1989 (Influenza A); Elliott, 1989 (Bunyamwera virus) and Schmaljohn, 1990 (Hantaan virus).

conserved amino acid motifs, in a stretch of 200-250 residues, with the PB1 polymerase subunits of influenza viruses (Fig.5.5). These findings further underline the importance of these common signature sequences and, moreover, justify the assumption that TSWV L RNA indeed encodes the viral polymerase. Strikingly, on the basis of amino acid homology, TSWV is more closely related to Bunyamwera virus than Hantaan virus is to this prototype bunyavirus. It may be anticipated that the amino acid homology between TSWV L protein and those of phlebo- and uukuviruses is even higher, since these viruses are even more closely related to TSWV, sharing similarly organized ambisense S RNA

90

segments (De Haan et al., 1990; Giorgi et al., 1991).

The presented data furthermore imply that, based on molecular properties, such as terminal sequences, and based on the exclusive host range and mode of transmission, TSWV indeed represents a virus of a new distinct genus (tospovirus) within the Bunyaviridae.

ACKNOWLEDGEMENTS

REFERENCES

Braam, J., Ulmanen, I. & Krug, R.M. (1983). Molecular model of a eukaryotic transcription complex: Func tions and movements of influenza P proteins during capped RNA-primed transcription. Cell 34: 609-618.

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

De Haan, P., Wagemakers, L., Peters, D. & Goldbach, R. (1989b). Molecular cloning and terminal sequence determination of the S and M RNA of tomato spotted wilt virus. J. Gen. Virol. 70: 3469-3473.

Devos, R.G.E., Gillis, E. & Fiers, W. (1976). The enzymatic addition of poly(A) to the 3' end of RNA using bacteriophage MS2 RNA as a model system. Eur. J. Biochem. 62: 401-410.

Elliott, R.M. (1989). Nucleotide sequence analysis of the large genomic RNA segment of Bunyamwera virus, the prototype of the family Bunyaviridae. Virology 173: 426-436.

Elliott, R.M. (1990). Molecular biology of the Bunyaviridae. J. Gen. Virol. 71: 501-522.

Giorgi, C., Accardi, L., Nicoletti, L., Gro, M.C., Takehara, K., Hilditch, C., Morikawa, S. & Bishop, D.H.L. (1991). Sequences and coding strategies of the S RNAs of Toscana and Rift valley fever viruses compared to those of Punta toro, Sicilian sandfly fever, and Uukuniemi viruses. Virology 180: 738-753.

Goldbach, R. (1986). Molecular evolution of plant RNA viruses. Ann. Rev. Phytopathol. 24: 289-310.

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

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

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

Hodgman, T.C. (1988). A new superfamily of replicative proteins. Nature 335: 22-23.

Ie, T.S. (1970). Tomato spotted wilt virus. CMI/AAB Descr. of Plant Viruses 39.

Kamer, G. & Argos, P. (1984). Primary structural comparison of RNA-dependant polymerases from plant, animal and bacterial viruses. Nucl. Ac. Res. 12: 7269-7282.

Krug, R.M., Alonso-Caplen, F.V., Julkunen, I. & Katze, M.G. (1989). In: The influenza viruses, pp. 89-152. Edited by R.M. Krug. New York: Plenum Press.

Kyte, J. & Doolittle, R.F. (1982). A simple method for displaying the hydrophobic character of a protein. J. Mol. Biol. 157: 105-132.

Maniatis, T., Fritsch, E.F. & Sambrook, J. (Eds.)(1982). Molecular cloning: a laboratory manual. New York: Cold Spring Harbor Laboratory.

Matthews, R.E.F. (1982). Classification and nomenclature of viruses: Fourth report of the international committee on Taxonomy of Viruses. Intervirology 17: 1-200.

Mohamed, N.A., Randles, J.W. & Francki, R.I.B. (1973). Protein composition of tomato spotted wilt virus. Virology 56: 12-21.

Mohamed, N.A. (1981). Isolation and characterization of subviral structures from tomato spotted wilt virus. J. Gen. Virol. 53: 197-206.

Parvin, J.D., Palese, P., Honda, A., Ishihama, A. & Krystal, M. (1989). Promoter analysis of influenza virus RNA polymerase. J. Virol. 63: 5142-5152.

Peters, D., Avila, A.C. de, Kitajima, E.W., Resende, R. de O., De Haan, P. & Goldbach, R. (1991). An overview of tomato spotted wilt virus. In: Virus-Thrips-Plant Interactions of TSWV, Proc. USDA workshop, Belltsville U.S.A., pp.1-14. Edited by H-.T. Hsu & R.H. Lawson. Nat. Techn. Inf. Serv. Springfield (VA).

Poch, O., Sauvaget, I., Delarue, M. & Tordo, N. (1989). Identification of four conserved motifs among the RNA-dependant polymerase encoding elements. EMBO J. 8: 3867-3874.

Raju, R. & Kolakofski, D. (1989). The ends of La Crosse virus genome and antigenome RNAs within nucleocapsids are base-paired. J. Virol. 63: 122-128.

Resende, R. de O., De Haan, P., Avila, A.C. de, Kormelink, R., Goldbach, R. & Peters, D. (1991). Generation of envelope and defective interfering RNA mutants of tomato spotted wilt virus by mechanical passage. J. Gen. Virol. in press.

Sanger, F., Nicklen, S. & Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Nat. Acad. Sci. U.S.A. 74: 5463-5467.

Schmaljohn, C.S. (1990). Nucleotide sequence of the L genome segment of Hantaan virus. Nucl. Ac. Res. 18: 6728.

Strauss, J.H. & Strauss, E.G. (1988). Evolution of RNA viruses. Ann. Rev. Microbiol. 42: 657-683.

Tas, P.W.L., Boerjan M.L. & Peters, D. (1977). The structural proteins of tomato spotted wilt virus. J. Gen. Virol. 36: 267-279.

Van den Hurk, J., Tas, P.W.L. & Peters, D. (1977). The ribonucleic acid of tomato spotted wilt virus. J. Gen. Virol. 36: 81-91.

Yamashita, M., Krystal, M. & Palese, P. (1989). Comparison of the three large polymerase proteins of inluenza A, B and C viruses. Virology 171: 458-466.

Yanisch-Perron, C., Vieira, J. & Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33: 103-119.

Zhang, H., Scholl, R., Browse, J. & Somerville, C. (1988). Double stranded DNA sequencing as a choice for DNA sequencing. Nucl. Ac. Res. 16: 1220.

92

CHAPTER 6 DETECTION OF TOMATO SPOTTED WILT VIRUS IN INFECTED PLANTS BY MOLECULAR HYBRIDIZATION AND PCR

Peter de Haan[1], Jan Gielen[2], Mart van Grinsven[2], Dick Peters[1] & Rob Goldbach[1].

SUMMARY

In addition to serological techniques, of which ELISA is the most important one, two molecular techniques have been developed for sensitive detection of tomato spotted wilt virus (TSWV). Using a dot blot hybridization assay, approximately 1 pg of viral RNA could be detected in leaf extracts from TSWV-infected plants. Using the polymerase chain reaction (PCR) technique the sensitivity is further increased to approximately 0.1 pg of viral RNA.

The usefulness of the two molecular techniques in sensitive and early detection of TSWV will be discussed.

[1]Dept. of Virology, Agricultural University Wageningen, Binnenhaven 11, P.O. Box 8045, 6709 PD Wageningen, The Netherlands.
[2]Zaadunie Research, Westeinde 62, P.O. Box 26, 1600 AA Enkhuizen, The Netherlands.

6.1   Introduction

Tomato spotted wilt virus (TSWV) ranks among the most damaging plant viruses, causing worldwide yield losses in many crops and ornamental plants (Peters et al., 1990). TSWV is transmitted by thrips species, of which Frankliniella occidentalis (Perg.) and Thrips tabaci (Lind.) are the most important (Sakimura, 1962).

The virus particles of TSWV are enveloped, spherically shaped (80-110 nm. in diameter) and studded with membrane spikes. The genome consists of three linear single stranded RNA segments, denoted S RNA (2.9 kb), M RNA (5.0 kb) and L RNA (8.9 kb), which are tightly associated with nucleocapsid proteins to form stable circular nucleocapsids (Chapters 1 and 3).

In general, plant virus diagnosis is based on symptom development on test plants, on detection of viral protein using serological techniques, such as ELISA, or based on detection of viral nucleic acids using molecular hybridization or the polymerase chain reaction (PCR) technique.

TSWV diagnosis by means of serology is hampered for a number of reasons. For instance, since TSWV is characterized by its fairly labile virus particles, it is difficult to obtain highly purified antigen preparations required for the production of specific antisera. Furthermore, the occasional generation of defective isolates of this virus may cause serious misinterpretations of results (Avila et al., 1990).

Another complication is the considerable natural variability among the different TSWV isolates. Recently performed serological studies on isolates, obtained from many different sources, revealed that they can be separated into a number of serogroups and serotypes (Wang & Gonsalves, 1990; Avila et al., 1990). For example, recently a new 'TSWV-like' virus has been characterized (Law & Moyer, 1989; Avila et al., 1991), which is serologically distinct from TSWV and escapes detection when using antisera raised against other isolates.

The availability of both cDNA clones and nucleotide sequence information on the genomic RNAs of TSWV, enables identification and detection of this virus using

94

molecular techniques. Although the use of cDNA clones in TSWV diagnosis has been reported previously (Ronco et al., 1990; German et al., 1990), the lack of sensitivity makes it thusfar less valuable for practical application.

We here describe two methods for detection of TSWV, one involving Northern dot blot hybridization, using synthetic RNA transcripts (riboprobes), and another involving the 'polymerase chain reaction' (PCR) technique. Both methods enable sensitive and unequivocal detection of TSWV.

6.2 Methods

6.2.1 Virus and plants

TSWV CNPH1, a Brasilian isolate from tomato, has been used in all experiments. TSWV-infected plants have been obtained from professional growers in the Netherlands. Virus was purified from infected Nicotiana rustica var. America according to Tas et al. (1977) and RNA was extracted from purified virions using SDS-phenol and subsequent ethanol precipitation.

6.2.2 Probe preparation

Complementary DNA (cDNA) to TSWV RNA was synthesized and cloned as described in Chapter 3. DNA fragments of several clones were inserted in bluescribe vectors pSK^+ (Stratagene), flanked by bacteriophage T3 and T7 RNA polymerase promoters (Fig.6.1). After linearization of the templates, ^{32}P-labelled, run-off transcripts were synthesized using T3 or T7 RNA polymerase (Gibco-BRL), according to conditions recommended by the manufacturer. Routinely a specific incorporation of 10^9 c.p.m. was obtained per µg linearized plasmid DNA. The strand-specific RNA probes were tested by Northern blot analysis of RNA, extracted from purified virus, or from TSWV-infected and healthy N. rustica leaves.

6.2.3 Northern dot blot hybridization

Deproteinized RNA samples were prepared by grinding 1 g leaf material in 5 ml extraction buffer (10 mM Tris-HCl (pH = 7.5), 10mM NaCl, 5 mM EDTA, 0.1 % Triton X-100 and 0.1 % 2-mercaptoethanol), followed by phenol extraction. Purified virus and RNA preparations were diluted and spotted on membranes in quantities of 3 µl per spot. Several membrane types (nitrocellulose, nylon) obtained from different companies were tested. Nylon membranes, such as Hybond N$^+$ (Amersham) or Genescreen-plus (NEN) gave the most satisfactory results. The RNA was UV-cross-linked, prehybridized for 4 h at 60 °C in a buffer consisting of 5 X SSC, 5 X Denhardts, 50 mM NaH$_2$PO$_4$, 50 mM EDTA, 100 µg/ml denatured herring sperm DNA and subsequently hybridized in the same buffer, for 18 h at 60 °C, after adding 10^5 c.p.m./ml RNA probe. The filters were washed twice with 2 X SSC, 0.1 % SDS for 5 min at room temperature, washed once with 0.1 % SSC, 0.1 % SDS for 5 min at 50 °C and exposed to X-ray films.

6.2.4 PCR reactions

Crude RNA samples were prepared by grinding 1 g leaf material in 5 ml of sterile water, immediately followed by phenol extraction. One µl of RNA sample was reverse transcribed and subsequently amplified in a final volume of 100 µl, containing 100 units (u) MuMLV-reverse transcriptase (Gibco-BRL) and 2.5 u Taq polymerase (Promega), using 4 µg/ml primer S1 and S2 or L1 and L3 (see Fig.6.1) for 30 min at 37 °C followed by 30 amplification cycles (1 min denaturation at 93 °C, 1.5 min annealing at 55 °C and 2 min extension at 72 °C). The amplified DNA fragments were visualised on 1% agarose gels.

6.3   Results

6.3.1   RNA probe synthesis and dot blot hybridization

A number of methods can be used for the preparation of labelled probes to detect viral RNA by molecular hybridization. Hybridization techniques using labelled synthetic RNA transcripts have been shown to be particularly sensitive (Maule et al., 1983; Baulcombe et al., 1984; Roy et al., 1988; Varveri et al., 1988; Koenig et al., 1988). To obtain templates for synthesis of strand specific RNA probes to be used for detection of TSWV, several cDNA fragments derived from the TSWV S, M and L RNA segment (70, 201, 270 and 514) were subcloned in plasmid pSK$^+$. For each template, RNA transcripts complementary to the viral RNA were synthesized using T7 RNA polymerase (Fig.6.1). The experiments described in this Chapter have all been performed with a L RNA-specific probe, i.e. clone 70, which contained a viral insert of 1175 nucleotides.

As a next step in the development of a sensitive detection assay, experiments were performed to optimize the hybridization procedure. The resulting favourable membrane type, hybridization and washing conditions have been described under methods. To prepare samples from infected plants, a number of different grinding buffers were tested. Buffers giving the best results contained 0.1 % Triton X-100 and 0.1 % 2-mercaptoethanol. An extraction step with phenol was used to remove proteins that cause non-specific hybridization signals. Polyphenols present in some test samples such as tomato, were removed by addition of 1 % L-polyclar (BDH) during extraction. Polysaccharides present in many ornamental plant leaves were removed by addition of 1 % cetyl-trimethyl-ammoniumbromide (CTAB) to the grinding buffer. Samples of purified viruses were spotted directly on membranes without any pre-treatment. Fig.6.2 shows that TSWV can be detected in tobacco leaf extracts when diluted up to 1/10,000, which corresponds to approximately one picogram of viral RNA.
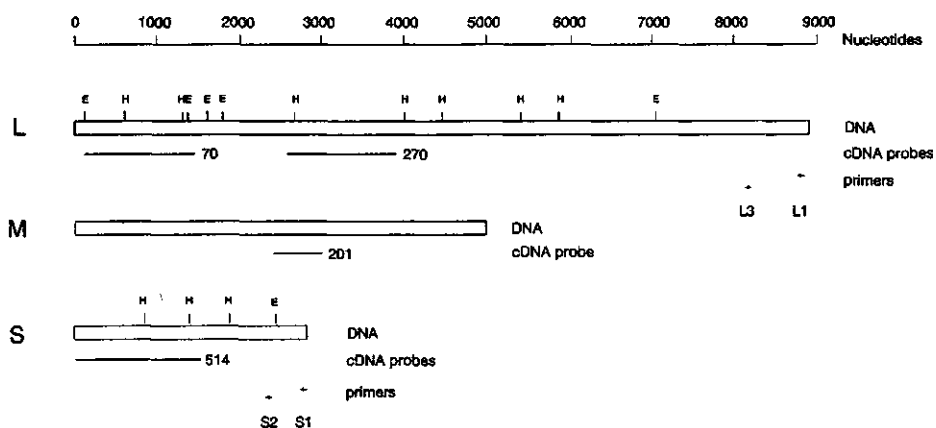
Fig.6.1: Restriction endonuclease map of aligned cDNA clones to TSWV S and L RNA, together with the location of the primers used for PCR (S1, S3, L1, L3) and the cDNA clones (70, 201, 270 and 514) used as templates for synthesis of ^{32}P-labelled RNA probes.

6.3.2 PCR detection

The most recent and a most promising technique for virus detection, is the polymerase chain reaction (PCR), where viral nucleic acids are enzymatically amplified using reverse transcriptase and the thermo-stable Taq-polymerase (Ehrlich, 1989; Innis et al., 1990). Using primers derived from the 3' region of the S RNA sequence (see Fig. 6.1), DNA fragments of expected size (500 bp) were generated from crude RNA samples of TSWV-infected plants. In a same assay primers derived from the 3' region of TSWV L RNA yielded 800 bp DNA fragments.

Under these standard amplification conditions approximately 100 fg viral RNA could be detected (Fig. 6.3). Virus could be detected in plant sap when diluted up to 50,000 times. Based on the band intensities on the agarose gels, it can be concluded that the standard PCR conditions used, were apparently sub-optimal for detection of extremely low amounts of RNA. The detection levels may be further increased by using more reverse transcriptase and/or Taq-polymerse or by using longer extension times. In

98

PURIFIED VIRUS   TOBACCO LEAVES
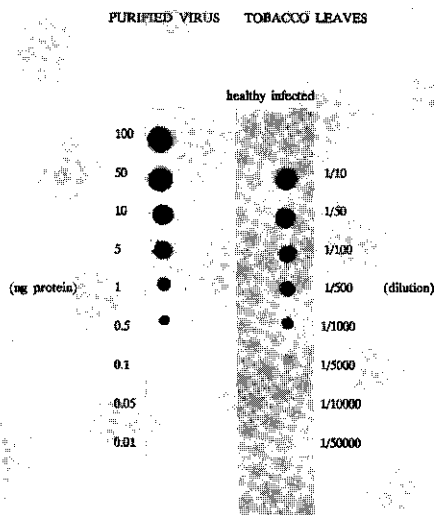
Fig.6.2: Detection of TSWV in dilutions of a purified virus preparation and in extracts of infected plants, using the dot blot hybridization assay.

contrast to the hybridization method, no additives or additional steps in the sample preparation were necessary for proper PCR reactions.

6.3.3 Indexing of plants for TSWV infection

In order to determine the value of both molecular detection methods a large number of uninfected and TSWV-infected plants, obtained from individual growers, were analyzed.

Routinely, leaf extracts were diluted tenfold prior to analysis. Examples are shown in Fig.6.4, where pepper plants have been tested. These experiments revealed that although both methods are sensitive and reliable, the critical step in both assays is the choice, which leaf to select as sample to be analysed. In our hands, the use of the upper (young) leaves showing systemic, non-necrotic symptoms gave the best results.

99
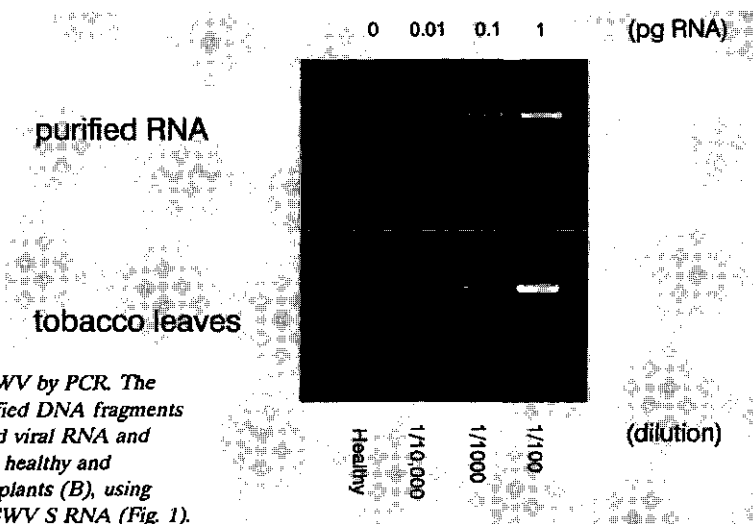
Fig.6.3: Detection of TSWV by PCR. The agarose gels show amplified DNA fragments from dilutions of purified viral RNA and sap extracts (A) or from healthy and TSWV-infected tobacco plants (B), using primers derived from TSWV S RNA (Fig. 1).

6.4 Discussion

In standard double sandwich ELISA procedures routinely up to 1 ng of TSWV protein can be detected, corresponding with approximately 10 pg of viral RNA (Resende et al., 1991). With the described dot blot hybridization assay up to 1 pg of viral RNA can be detected. This method, which is approximately ten times more sensitive than ELISA, may therefore be useful for detection of TSWV at early stages of infection. However, hybridization techniques are relatively expensive and time consuming, compared to serological techniques. Working with radio-isotopes requires facilities, usually not available in plant pathology laboratories that carry out routine diagnosis. In the future, these problems might be circumvented by the use of non-radio-isotope labelling techniques, or by the use of the PCR technique. The described PCR assay is indeed more sensitive (0.1 pg of viral RNA can easily be detected) and less laborious than hybridization. At this moment the costs of PCR are significantly higher than those of serology. When PCR becomes a more widely applied technique, the prices of used equipment and biochemicals will certainly decrease significantly.

100
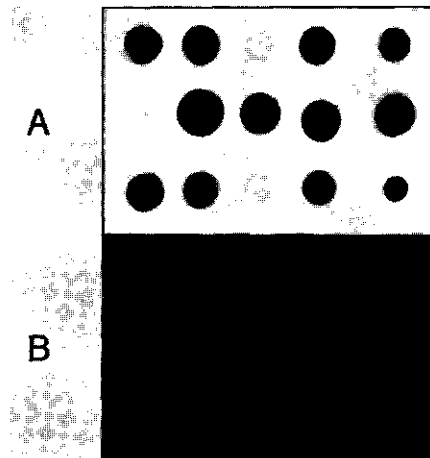
*Fig.6.4: Dot blot hybridization (A) and PCR (B)
to index pepper plants for infection with TSWV*

Although both molecular methods are highly valuable for virus identification under laboratory conditions (Fig.6.4), additional refinements and simplifications are necessary to make them suitable for diagnostic purposes in practise.

As outlined in the introduction, several serogroups of TSWV can be recognized. Preliminary hybridization experiments revealed that nucleotide sequence homology between representatives of different serogroups is low (Avila et al., 1991). This implies that other TSWV isolates may not be identified using the RNA probes and primers described here. Nucleotide sequence data on other TSWV isolates are therefore urgently required to select for conserved sequences, which can be used for the production of more universal riboprobes or PCR-primers, as reported recently for luteo- and potyviruses (Robertson et al., 1991; Langeveld et al., 1991).

REFERENCES

Avila, A.C. de, Huguenot, C., Resende, R. de O., Kitajima, E.W., Goldbach, R. & Peters, D. (1990). Serological differentiation of 20 isolates of tomato spotted wilt virus. J. Gen. Virol. 71: 2801-2807.

Avila, A.C. de, De Haan, P., Kitajima, E.W., Kormelink, R., O. Resende, R. de O., Goldbach, R. & Peters, D. (1991). Characterization of a distinct tomato spotted wilt virus (TSWV) isolate from Impatiens sp. in the Netherlands. Submitted for publication.

Baulcombe, D., Flavell, R.B., Boulton, R.E. & Jellis, G.J. (1984). The sensitivity and specificity of a rapid nucleic acid hybridization method for the detection of potato virus X in crude sap samples. Plant Pathol. 33: 361-370.

Ehrlich, H.A. (Ed.)(1989). (Ed.). PCR technology. Priciples and applications for DNA amplification. New York: M. Stockton Press.

German, T.L. & Hu, Y. (1990). Detection of tomato spotted wilt virus RNA in Thrips using strand specific probes. Phytopathology 80: 996.

Huguenot, C., Van den Dobbelsteen, G., De Haan, P., Wagemakers, C.A.M., Drost, G.A., Osterhaus, A.D. M.E. & Peters, D. (1990). Detection of tomato spotted wilt virus using monoclonal antibodies and riboprobes. Arch. Virol. 110: 47-62.

Innis, M.A., Gilfand, M.A., Sninski, J.J. & White, T.J. (Eds.)(1990). A guide to methods and applications. PCR Protocols. New York: Academic Press.

Koenig, R., An, D. & Burgermeister, W. (1988). The use of filter hybridization techniques for the identification, differentiation and classification of plant viruses. J. Virol. Methods 19: 57-68.

Langeveld, S.A., Dore, J.-M., Memelink, J., Derks, A.F.L.M., Van der Vlught, C.I.M., Asjes, C.J. & Bol, J.F. (1991). Identification of potyviruses by polymerase chain reactions with degenerate primers. J. Gen. Virol. 72: 1531-1541.

Law, M.D. & Moyer, J.W. (1990). A tomato spotted wilt-like virus with a serologically distinct N protein. J. Gen. Virol. 71: 933-938.

Maule, A.J., Hull, R. & Donson, J. (1983). The application of spot hybridization to the detection of DNA and RNA viruses in plant tissues. J. Virol. Methods 6: 215-224.

Peters, D., Avila, A.C. de, Kitajima, E.W., Resende, R. de O., De Haan, P. & Goldbach, R. (1991). An overview of tomato spotted wilt virus. In: Virus-Thrips-Plant Interactions of TSWV, Proc. USDA workshop, Belltsville U.S.A., pp.1-14. Edited by H.-T. Hsu & R.H. Lawson. Nat. Techn. Inf. Serv. Springfield (VA).

Resende, R. de O., Avila, de A.C., Goldbach, R. & Peters, D. (1991). Detection of tomato spotted wilt virus using polyclonal antisera in double antibody sandwich (DAS) ELISA and cocktail ELISA. J. Phytopathol. 132: 46-56.

Robertson, N.L., French, R. & Gray, S.M. (1991). Use of group-specific primers and the polymerase chain reaction for the detection and identification of luteoviruses. J. Gen. Virol. 72: 1473-1477.

Ronco, A.E., Dal Bo, E., Ghiringhelli, P.D., Medrano, C., Romanowski, V., Sarachu, A.N. & Grau, O. (1989). Cloned cDNA probes for the detection of tomato spotted wilt virus. Phytopathology 79: 61-64.

Roy, B.P., AbouHaidar, M.G., Sit, T.L. & Alexander, A. (1988). Construction and use of cloned cDNA biotin and ^{32}P-labeled probes for detection of papaya mosaic potexvirus RNA in plants. Phytopathology 78: 1425-1429.

Sakimura, S. (1962). The present status of thrips-borne viruses. In: Biological transmission of disease agents, pp. 33-40. Edited by K. Maramorosch. New York: Academic Press.

Tas, P.W.L., Boerjan, M.L. & Peters, D. (1977). The structural proteins of tomato spotted wilt virus. J. Gen. Virol. 36: 267-279.

Varveri, C., Ravelonarandro, M. & Dunez, J. (1987). Construction and use of a cloned cDNA probe for the detection of plum pox virus in plants. Phytopathology 77: 1221-1224.

Wang, M. & Gonsalves, D. (1990). ELISA detection of various tomato spotted wilt virus isolates using specific antisera to structural proteins of the virus. Plant Dis. 74: 154-158.

CHAPTER 7 ENGINEERED RESISTANCE TO TOMATO SPOTTED WILT VIRUS, A NEGATIVE-STRAND RNA VIRUS

Peter de Haan[1], Jan Gielen[2], Ad Kool[2], Dick Peters[1], Mart van Grinsven[2] & Rob Goldbach[1]

ABSTRACT

For a growing number of positive-strand RNA viruses, it has been demonstrated that transformation of host plants with the viral coat protein gene confers resistance to the corresponding virus. Sofar, successful transformation strategies to gain resistance to negative-strand RNA viruses have not been reported. Here we show that genetically engineered resistance can be obtained to tomato spotted wilt virus, an enveloped virus with a negative-strand RNA genome, by transforming tobacco with the gene encoding the viral nucleocapsid (N) protein, an internal RNA-binding protein.

[1]Dept. of Virology, Agricultural University Wageningen, Binnenhaven 11, P.O. Box 8045, 6709 PD Wageningen, The Netherlands
[2]Zaadunie Research, Westeinde 62, P.O. Box 26, 1600 AA Enkhuizen, The Netherlands.

7.1 Introduction

Among plant viruses tomato spotted wilt virus (TSWV) is unique in its particle morphology and genome structure, and moreover, it is the only plant virus transmitted by thrips species (Sakimura, 1962; Paliwal, 1974). TSWV particles are enveloped, spherically shaped (80-110 nm in diameter) and are studded with surface projections (Best & Palk, 1964; Ie, 1964; Kitajima, 1965). The genome consists of three species of linear single stranded RNA denoted S, M and L. The genomic RNAs are tightly associated with nucleocapsid (N) proteins to form pseudo-circular nucleocapsids (Van den Hurk et al., 1977; Mohamed, 1981; Peters et al., 1991).

The TSWV S RNA segment is 2916 nucleotides long and has an ambisense gene arrangement, exactly as found for bunyaviruses belonging to the genera uukuvirus and phlebovirus (Giorgi et al., 1991). This RNA contains two genes, the nucleocapsid (N of 28.8K) protein gene in viral complementary sense and a gene encoding a non-structural (NSs of 52.4K) protein in viral sense (Chapter 4).

TSWV L RNA is 8897 nucleotides long and completely of negative polarity. It encodes a primary translation product of 331.5K, which most likely corresponds with the viral polymerase, which is present in the virus particles (Chapter 5). Preliminary sequence data on the M RNA (approximately 5000 nucleotides long) indicate that it is also of negative polarity. It most likely encodes the membrane glycoproteins G1 of 78K and G2 of 58K.

The properties summarized here indicate that TSWV represents a member of the arthropod-borne Bunyaviridae, a large family of negative-strand RNA viruses (Elliott, 1989). Indeed at the latest meeting of the International Committee on Taxonomy of viruses in Berlin 1990, TSWV has been classified as the prototype of the genus tospovirus within the Bunyaviridae, being unique in its property to infect plants.

The worldwide distribution of TSWV, together with the current dramatic expansion of one of its major vectors, the Western Flower thrips (Frankliniella occidentalis), makes this virus one of the most harmful plant viruses. To date more than 400 plant species, both mono- and dicotyledons, are known to act as susceptible hosts for TSWV and

106

considerable yield losses have been reported in the cultivation of many important crops, such as tomato, tobacco, lettuce, groundnut, pepper and ornamentals such as Impatiens, Ageratum and Chrysanthemum (Iwaki et al., 1984; Barnes & Halliwell, 1985; Reddick et al., 1987; Allen & Matteoni, 1988; Brown, 1988; Mantel & Van de Vrie, 1988; Gebré-Selassie et al., 1989; Cho et al., 1989). Mainly due to the resistance of thrips species to insecticides and routine sanitary measures are no longer adequate to limit the incidence of TSWV infections. Therefore, natural occurring resistance or tolerance to TSWV has been topic of intensive breeding research. However, studies on tobacco, tomato and lettuce have shown that natural resistance or tolerance to TSWV is predominantly polygenic and based on complex interactions between virus, vector and host plant (Smith & Gardner, 1951; Finlay, 1953; Borchers, 1956; Holmes, 1958; Best, 1968; Moldovan & Chokan, 1972; Vinogradov et al., 1982; Cupertino et al., 1986; O'Malley & Hartmann, 1989; Paterson et al., 1989).

We therefore addressed the question whether engineered resistance to TSWV infections can be achieved by expression of the viral N protein in tobacco. The rational behind this molecular approach is the dual function of this protein. It has been demonstrated for other negative-strand viruses (i.e. arena-, bunya- and influenzaviruses) that the N protein is involved in wrapping the viral RNAs to yield nucleocapsids. Moreover, the amount of intracellular, free, unassembled N protein discriminates whether the viral polymerase is active in transcription or replication of the RNA genome (Ihara et al., 1985; Beaton & Krug, 1986; Franze-Fernandez et al., 1987; Vidal & Kolakofski, 1989). It is envisaged that accumulation of the TSWV N protein in a susceptible host plant could lead to to blocking of the transcriptional activity of the viral polymerase, causing abortive replication of incoming viral RNAs.

7.2   Methods

7.2.1 Virus and plants

TSWV isolate CNPH1, originating from Brasil, was maintained in tomato by grafting. Virus was purified from mechanically inoculated tobacco (Nicotiana rustica var. America) as described by Tas et al. (1977). For the transformation experiments, in vitro grown N. tabacum var. SR1 was used as recipient. Transgenic tobacco plants were grown under greenhouse conditions (PKII), according to the legislation imposed by the Dutch authorities (voorlopige Commissie Genetisch Modificatie: vCOGEM).

7.2.2 Construction of the plant expression vectors

All manipulations involving DNA or RNA were performed according to standard procedures (Ausubel et al., 1990). A cDNA fragment containing the TSWV N gene and 124 nucleotides of its 5' untranslated leader sequence and 6 nucleotides of its 3' trailer sequence (Chapter 4) was provided with Pst1 linkers and subsequently cloned into the expression vector pZU-A, yielding pTSWVN-A (Fig.7.1a). The vector pZU-A contains the cauliflower mosaic virus (CaMV) 35S promoter sequences and the transcription-terminator sequences from the nopalin synthase (nos) gene of the Agrobacterium tumefaciens Ti plasmid. A second gene cassette was constructed by removal of the 5' un-translated leader sequence from the TSWV N gene, using site directed mutagenesis (an EcoRV restriction site was created just in front of the AUG startcodon). The resulting DNA fragment was cloned in expression vector pZU-B, yielding pTSWVN-B (Fig.7.1b). The vector pTSWVN-B is identical to pTSWVN-A, except that the 5' untranslated leader sequence of the TSWV N gene has been replaced by that of tobacco mosaic virus (TMV). Both chimaeric gene cassettes were cloned as XbaI DNA fragments in binary transformation vector pBIN19 (Bevan, 1984).
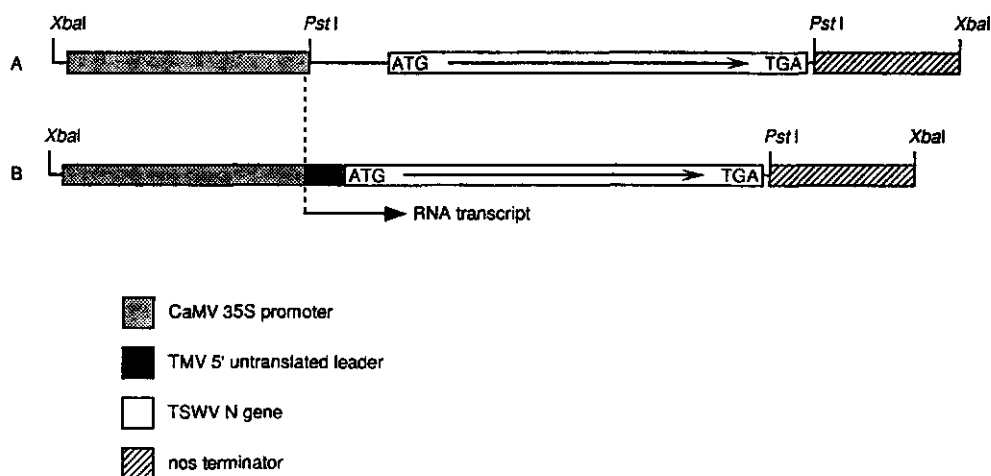
Fig. 7.1: Schematical representation of the chimaeric gene cassettes containing the TSWV N gene. A, pTSWVN-A; B, pTSWVN-B.

7.2.3 Transformation of tobacco

The pBIN19 derived tranformation vectors were mated into the non-oncogenic A. tumefaciens strain LBA4404 (Hoekema et al., 1983) by a triparental plasmid transfer using pRK2013 as a helper plasmid (Ditta et al., 1980). The resulting recombinant A. tumefaciens strains were checked for the integrity of the TSWV N gene by Southern blot analysis. Transformation and regeneration of in vitro grown N. tabacum var. SR1 was performed by the leaf disk method essentially according to Horsch et al.(1984). Tobacco transformants were selected for resistance to kanamycin (100 μg/ml), rooted, potted in soil and subsequently transferred to the greenhouse.

7.2.4 Southern and Northern blot analysis

Total DNA was extracted according to Ausubel et al. (1987). Portions of 10 μg DNA were digested with XbaI, fractionated by electrophoresis in 0.8 % agarose gels and

transferred to Hybond-C membranes by capillary blotting. The resulting Southern and Northern blots were subsequently hybridized to a ^{32}P-labelled cDNA fragment, containing the TSWV N gene (See 7.2.1).

Total RNA was extracted from transgenic tobacco plantlets according to Logemann et al. (1987). RNA samples of 20 μg were fractionated on 1.2% agarose gels under denaturating conditions and tranferred to Hybond-C membranes (Amersham) by capillary blotting.

7.2.5 Protein analysis in transgenic plants

The amounts of N protein in young tobacco leaves were quantified by DAS-ELISA using a polyclonal rabbit antiserum to purified TSWV nucleocapsids (Resende et al., 1991). Purified nucleocapsid protein was included as standard. The soluble protein content of the leaf extracts was determined using the Bio-Rad protein assay. The steady-state levels were calculated and presented as percentage of total soluble protein.

The integrity of the TSWV N protein accumulating in the transgenic tobacco plants was verified by Western blot analysis. Leaf tissue was grinded in PBS-T (phosphate-buffered saline supplemented with 0.1% Tween-20) and portions of 50 μg of soluble protein were fractionated by electrophoresis in 12.5 % SDS-polyacrylamide gels (Laemmli, 1970). The proteins were blotted to Immobilon-P membranes (Millipore) and incubated with antiserum, conjugated with alkaline phosphatase (Avila et al., 1990).

7.2.6 Analysis of protection against TSWV infection

Progeny plants of self-pollinated initial tobacco transformants were analyzed for the segregation of the introduced N gene cassette by DAS-ELISA (See 7.2.5) and subsequently inoculated with TSWV, approximately 6 weeks after sowing (two-leaves stage). The plants which did not receive a copy of the N gene after self-pollination were

110

used as controls. The largest leaf was dusted with carborundum and inoculated with 25 µl inoculum, containing approximately 5-10 µg virus. The inocula were prepared by grinding 1 g of systemically infected N. rustica leaves in 5 ml 0.1 M sodiumphosphate (pH = 7.0) supplemented with 0.01 M Na_2SO_3. Since TSWV is one of the most unstable plant viruses known (Ie, 1970; Francki et al., 1985), all inocula were prepared freshly and kept on ice. The transgenic plants were inoculated first, followed by the control plants. After inoculation, the leaves were rinsed with water and the plants were monitored daily for the development of local and systemic symptoms. Upon mechanical inoculation of tobacco var. SR1 with TSWV, usually, necrotic primary infection spots appear on the inoculated leaves within 4-6 days. In addition, chlorosis and/or mosaic symptoms can be observed on the systemically infected leaves within 6-10 days. The plants will die a few
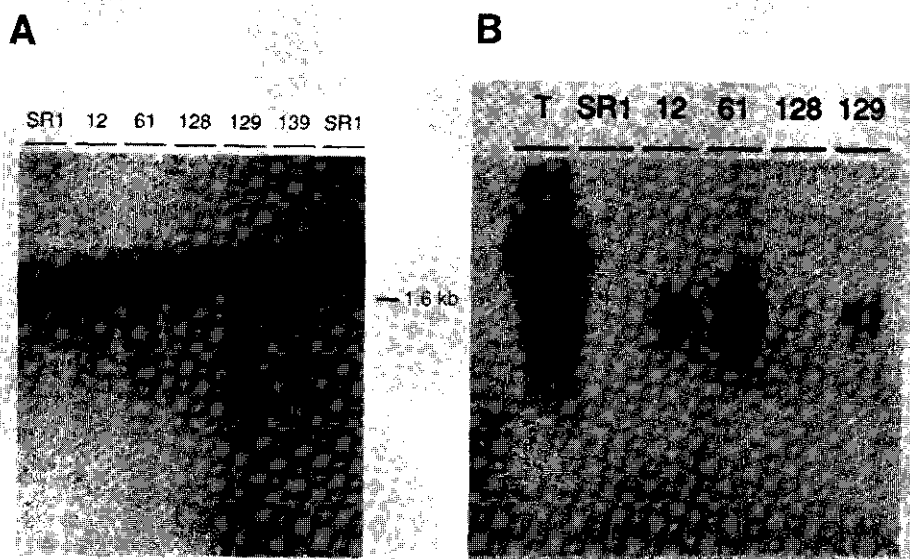
Fig. 7.2: Panel A: Southern blot analysis of the primary transformants. Genomic DNA was digested with Xba1 to reveal the complete TSWV N gene cassette, fractionated on a 0.8 % agarose gel, blotted to a Hybond-C membrane and hybridized to a ^{32}P-labelled DNA fragment, containing the TSWV N gene. Panel B: Northern blot analysis of the N gene transcripts in transgenic tobacco. Total plant RNA was fractionated on a 1.2 % agarose gel, transferred to a Hybond-C membrane and hybridized to the same probe as panel A. The numbers above the lanes correspond to the transgenic tobacco lines. SR1, untransformed tobacco SR1; C, control plasmid DNA; I, RNA from TSWV-infected tobacco leaves; H, RNA from healthy leaves

days later. In the inoculation experiments, plants were scored as being susceptible, when any leaf younger than the inoculated leaf showed typical systemic symptoms. Leaf samples were collected from visually healthy and from infected plants to check for the presence of virus by DAS-ELISA using monoclonal antibodies directed to the membrane glycoproteins of the virus (Avila et al., 1990).

7.3   Results

7.3.1 Transformation of tobacco with the TSWV N gene

A chimaeric gene cassette (pTSWVN-A) was constructed which contained the TSWV N gene with 124 nucleotides of the 5'-untranslated leader sequence, downstream of the CaMV 35S promoter (Fig.7.1a). To enhance the expression levels of the N gene, a second construct (pTSWVN-B) was made, in which the original TSWV-specific 5'-leader was replaced by the leader from TMV, which was known to function as a translational enhancer (Gallie et al., 1987)(Fig.7.1b). The polyadenylation signal was derived from the 3'flanking region of the nopalin synthase (nos) gene. Both gene cassettes were cloned into the binary transformation vector pBIN19 and subsequently transferred to N. tabacum var. SR1 by means of A. tumefaciens mediated leaf disc transformation. The integrity of the introduced N gene cassettes was verified by Southern blot analysis (Fig.7.2a). Northern blot analysis showed that mRNAs of expected size were transcribed from these chimaeric genes (Fig.7.2b). In total, 65 pTSWVN-A transformants (numbered 1-65) and 55 pTSWVN-B transformants (numbered 101-155) have been obtained. Except for one transformant, all transgenic tobacco plants exhibited normal phenotypic appearances and set seeds after self-pollination.
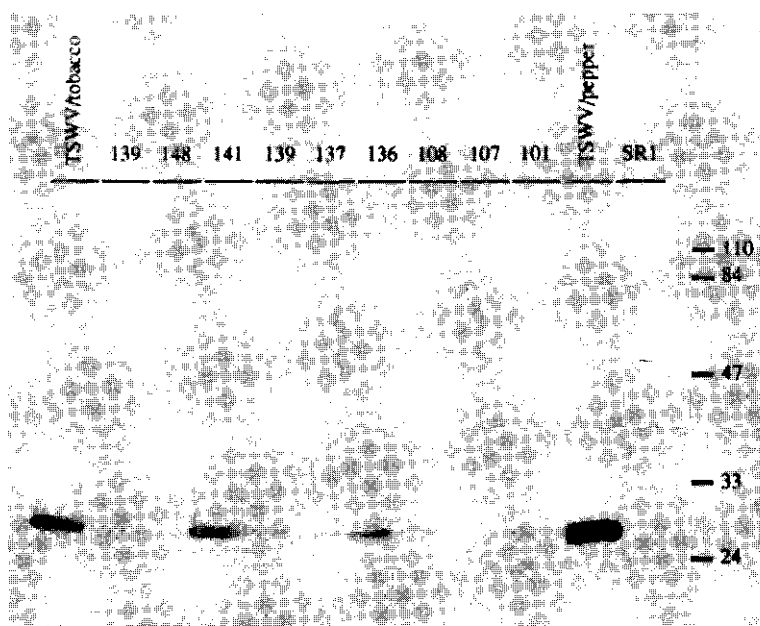
Fig. 7.3: Western blot analysis of nucleocapsid proteins accumulating in transgenic tobacco. Protein extracted from leaves, were separated on a 12.5% SDS-polyacrylamide gel, blotted to an Immobilon-P membrane and N protein was detected using an antiserum to purified TSWV nucleocapsids. Lane 1, marker proteins with the M_r values indicated on the left; Lane 2, TSWV-infected leaf extract; Lane 3, healthy leaf extract; Lanes 4 to 10, leaf extracts of transformants 10, 12, 61, 129, 139, 141 and 142 respectively.

7.3.2 Expression of the TSWV N gene in transgenic tobacco

The presence of N protein in the transgenic plants was determined by Western blot analysis (Fig.7.3). These experiments showed that the N protein produced in these plants, comigrated with that extracted from tobacco plants infected with TSWV. The expression levels were quantified by DAS-ELISA (Fig.7.4). The amounts of N protein differed considerably between individual transformants, ranging from below detection level to 1.5 % of the soluble leaf protein fraction. The differences in the amounts of N protein in the transgenic plants are most likely due to positional effects of the local genomic environment on the introduced gene cassettes. The average amount of N

protein accumulating in the pTSWVN-B transformants was approximately twice as high as that of the pTSWVN-A transformants.
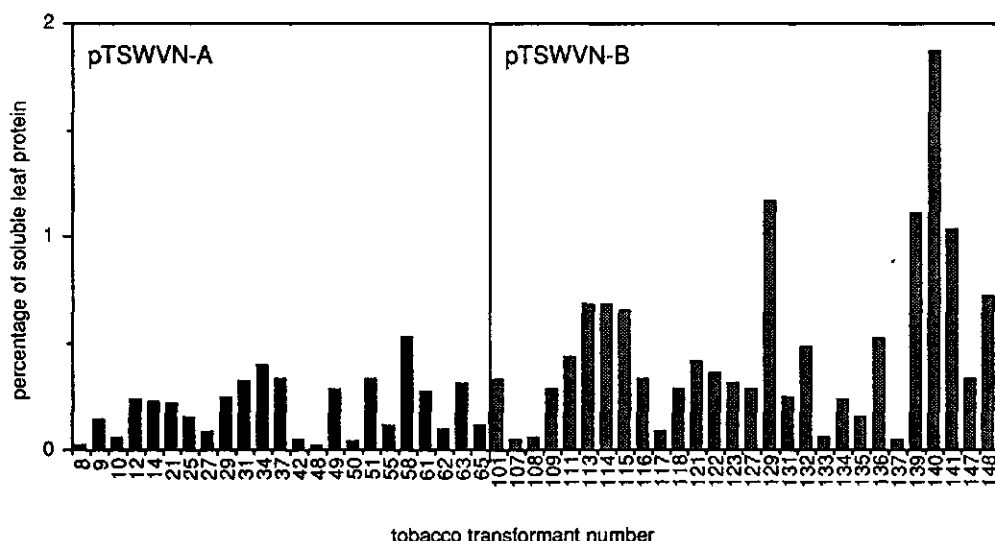
Fig.7.4: *Accumulation of TSWV N protein in transgenic tobacco quantified by DAS-ELISA. The accumulation levels are presented as percentages of total soluble leaf protein content. The numbers refer to the primary tobacco transformants.*

DAS-ELISA analysis of S1 progeny plants, obtained by self-pollination of the initial transformants, revealed the segregation ratios of the introduced gene cassettes. In most cases a segregation ratio of 3:1 (expressor : non-expressor) was obtained, indicating that the N gene cassette behaved as a single dominant gene. Southern blot experiments confirmed that most transgenic tobacco plants contained an unrearranged copy of the TSWV N gene. However, transformant nr. 139 (Fig.7.4) contained two or more active copies of the N gene. The expression levels of the TSWV N transgene in the S1 progeny plants were similar to that of the initial transformants. A correlation between the level of expression and the zygosity (homo- or heterozygote) of the plant could not be observed.

114

7.3.3 Virus susceptibility of the transgenic plants

To determine the susceptibility of the transgenic tobacco plants to TSWV-infections, progeny S1 plants were challenged by mechanical inoculation with TSWV. Over three separate experiments 80 progeny plants of each line were inoculated with 5-10 μg virus (isolate CNPH1), approximately 6 weeks after sowing. Control plants consisted of S1 individuals which did not inherit the N transgene after self-pollination of the original transformants. After inoculation, plants were monitored daily for the appearance of systemic symptoms. On average, four out of ten transgenic N protein-expressing (N^+) tobacco lines showed reduced susceptibility to TSWV-infection, compared to the control plants, the non-expressing (N^-) segregants. Figure 7.5a shows a typical experiment, using pTSWVN-A transformed lines 12 and 61, which both accumulated the N protein to approximately 0.25 % of total soluble protein (Fig.7.4). The N^- plants of both lines all showed severe systemic symptoms, 6 days after inoculation. The progeny N^+ plants of line 12 exhibited a delay in the symptom-development and moreover, approximately 70 % of the plants escaped from TSWV-infection. The progeny N^+ plants of line 61 all became systemically infected, albeit one day later than the control plants. Similar results have been obtained with the pTSWVN-B transformed tobacco plants (Fig.7.5b). Again, approximately four out of ten lines showed significant levels of protection to TSWV-infection. The results obtained with three lines, in which the N protein accumulated to approximately 1% of total soluble protein are shown in Fig.7.5b. Approximately 90 % of the N^+ progeny plants of line 139 were protected, whereas no protection could be observed for line 141. Line 129, which accumulates N protein to a level comparable to 139, displayed an intermediate level of resistance, since only approximately 25% of the N^+ progeny plants were protected. In addition, compared to non-expressing control plants, N^+ progeny plants of line 129 and 139 exhibited a delay of three to four days in symptom-development upon mechanical inoculation. In tobacco lines that showed reduced susceptibility to TSWV the numbers of primary infection spots were significantly lower (results not shown). It is evident from our study that no correlation could be found between the amounts of N protein accumulating in the transgenic plants and the levels
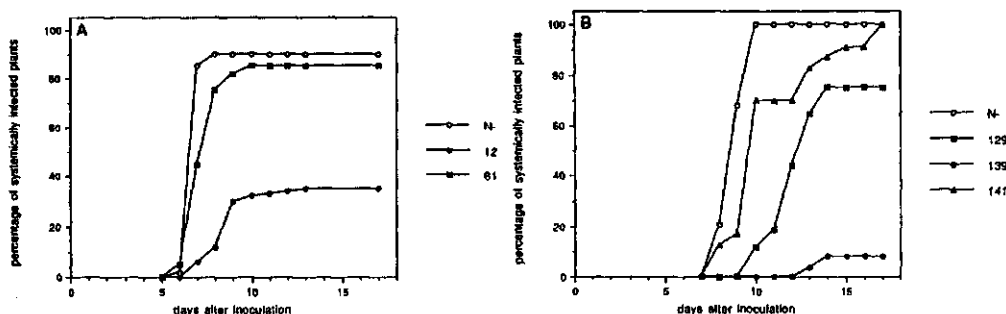
Fig.7.5: Development of systemic disease symptoms in transgenic S1 progeny plants, after mechanical inoculation with TSWV. Control plants consisted of plants that did not inherit a copy of the N gene after self-pollination of the initial transformant. Plants were inoculated six weeks after sowing with approximately 5-10 μg of virus. The development of systemic symptoms was monitored daily. Panel A: two representative lines nr. 12 and 61, of pTSWVN-A transformed tobacco plants. Panel B: Three pTSWVN-B transformed tobacco lines, nr 129, 139 and 141.

of protection to TSWV infection.

DAS-ELISA experiments using monoclonal antibodies to the membrane glycoproteins of the virus demonstrated that TSWV could only be detected in plants showing symptoms. Virus could not be detected in the protected transgenic plants, except for the few primary infection spots that appeared on the inoculated parts of the leaf.

7.4   Discussion

In this report we demonstrated that transformation of tobacco with the TSWV N gene confers resistance to this virus. Hence, this is the first example of genetically engineered resistance to a negative-strand virus. Transgenic tobacco plants have been obtained that express the introduced gene to high levels (up to 1.5 % of the total amount

116

of soluble protein). The TSWV N gene is stably inherited by inbreeding, through the third generation produced to date. In general, the transgene does not have any deleterious effects on phenotype and fertility of the initial transformants or their progeny.

The amounts of TSWV N protein in transgenic plants is not affected by gene dosage effects (either zygosity or gene copy number). Thus, a correlation between amount of N protein and number of loci expressing the gene could not be observed. This might be due to the overall high levels of expression. Generally, the steady-state levels of N protein in the pTSWVN-B transformed plants were twice as high as those of the pTSWVN-A transformants. This may indicate that in these plants the TMV 5'-untranslated leader sequence indeed served as a translational enhancer (Gallie et al., 19-87).

Upon mechanical inoculation, a number of transgenic tobacco lines expressing the TSWV N gene, show reduced susceptibility to TSWV. In comparison to control plants, transgenic plants escaped from infection and moreover, the few plants that became infected showed a significant delay in symptom development. DAS-ELISA experiments revealed that virus could only be detected in those parts of the leaves that showed systemic symptoms, or in the primary infection spots. Since, both the number of primary infections is decreased and systemic transport of virus is inhibited (the infection remains localized), these results suggest that this genetically engineered resistance operates at several levels.

It has been shown that the CaMV 35S promoter consists of multiple regulatory elements, which can each be differently affected according to its position in the chromosome. As a consequence, even when the overall amounts of N protein in different plants are equal, the 35S promoter in these plants may exhibit divergent levels of transcriptional activity among different tissues and cell types (Benfey et al., 1990a, 1990b). As we measured only overall N protein accumulation and not specifically at the site of infection, this could explain why there is no correlation between the amount of N protein and the observed level of resistance.

It should be noticed that the transcripts of the N gene cassettes are complementary

117

to the viral S RNA molecules. Hence, these transcripts could exhibit 'antisense' activity. To address this question, tobacco has to be transformed, with a TSWN N gene cassette, in which the open reading frame is distorted by frame shift mutation. Challenging of these transformants with TSWV should reveal whether the observed protection to viral infection is caused by accumulation of the N protein or by antisense inhibition. If the resistance indeed resides at the protein level, we could assume that the presence of N protein in the transgenic plants will block transcription by the viral polymerase, which is present in the infecting virus particles. This then implies that the mechanism on which the protection is based, would differ from coat protein-mediated protection as reported for plus-strand RNA viruses (Beachy et al., 1990; Hemenway et al., 1990).

This successful approach to obtain genetically engineered resistance to TSWV in tobacco will be of great value to combat this devastating plant pathogen. Therefore, this strategy will be extended to other economical important crops. Moreover, this approach may work for other negative-strand viruses, infecting plants.

ACKNOWLEDGEMENTS

118

REFERENCES

Allen, W.R. & Matteoni, J.A. (1988). Cyclamen ringspot: Epidemics in Ontario greenhouses caused by the tomato spotted wilt virus. Can. J. Plant Pathol. 10: 41-46.

Ausubel, F.A., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. & Struhl, K. (Eds.)(1990). Current protocols in Molecular Biology. New York: Green Publishing and Wiley-Interscience.

Avila, A.C. de, Huguenot, C., Resende, R. de O., Kitajima, E.W., Goldbach, R. & Peters, D. (1990). Serological differentiation of 20 isolates of tomato spotted wilt virus. J. Gen. Virol. 71:2801-2807.

Beachy, R.N., Loesch-Fries, S. & Tumer, N.E. (1990). Coat protein-mediated resistance against virus infection. Ann. Rev. Phytopathol. 28:451-474.

Beaton, A.R. & Krug, R.M. (1986). Transcription antitermination during influenza virus template RNA synthesis requires the nucleocapsid protein and the absence of a 5' capped end. Proc. Nat. Acad. Sci. U.S.A. 83: 6282-6286.

Benfey, P.N., Ren, L. & Chua, N.-H. (1990a). Tissue-specific expression from CaMV 35S enhancer subdomains in early stages of plant development. EMBO J. 9: 1677-1684.

Benfey, P.N., Ren, L. & Chua, N-.H. (1989b). Combinatorial and synergistic properties of CaMV 35S subdomains. EMBO J. 9: 1685-1696.

Best, R.J. & Palk, B.A. (1964). Electron microscopy of strain E of tomato spotted wilt virus and comments on its probable biosynthesis. Virology 23: 445-460.

Bevan, M. (1984). Binary Agrobacterium vectors for plant transformation. Nucl. Ac. Res. 12:8711-8721.

Borchers, E.A. (1956). A study of the nature of resistance of the Mignonette variety of lettuce to the tomato spotted wilt virus. PhD. Thesis, University of California. Davis, California.

Brown, L.G. (1988). Tomato spotted wilt virus in ornamentals. Plant Pathol. circ. 313.

Cho, J.J., Mau, R.F.L., Gonsalves, D. & Mitchell, W.C. (1986). Reservoir weed hosts of tomato spotted wilt virus. Plant Dis. 70: 1014-1017.

Cupertino, F.P., Avila de, A.C., Araujo, M.T. & Maluf, W.R. (1986). Sources of resistance to tomato spotted wilt virus in Lycopersicon. Fitopatologia Brasileira 11: 330-331.

Ditta, G., Stanfield, S., Corbin, D. & Helinski, D.R. (1980). Broad host range DNA cloning system for Gram-negative bacteria: construction of a gene bank of Rhizobium meliloti. Proc. Nat. Acad. Sci. U.S.A. 80:7347-7351.

Elliott, R.M. (1990). Molecular biology of the Bunyaviridae. J. Gen. Virol. 71: 501-522.

Finlay, K.W. (1953). Inheritance of spotted wilt resistance in the tomato. Austr. J. Biol. Sci. 6: 153-163.

Francki, R.I.B., Milne, R.G., Hatta, T.,(1985). Atlas of plant viruses. vol. 1 Boca Raton CRC Press, 222 p.

Franze-Fernandez, M. T., Zetina, C., Iapalucci, S., Lucero, M.A., Bouissou, C., Lopez, R., Rey, O., Daheli, M., Cohen, G.N. & Zakin, M.M. (1987). Molecular structure and early events in the replication of Tacaribe arenavirus S RNA. Virus Res. 7: 309-324.

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

Gebré-Selassie, K., Chabriere, C. & Marchoux, G. (1989). Un virus qui se réveille. Phytoma 410: 30-35.

Giorgi, C., Accardi, L., Nicoletti, L., Gro, M.C., Takehara, K., Hilditch, C., Morikawa, S. & Bishop, D.H.L. (1991). Sequences and coding strategies of the S RNAs of Toscana and Rift Valley fever viruses compared to those of Punta Toro, Sicilian sandfly fever, and Uukuniemi viruses. Virology 180: 738-753.

Hemenway, C., Haley, L., Kaniewski, W.K., Lawson, E.C., O'Connell, K.M., Sanders, P.R., Thomas, P.E. & Tumer, N.E. (1990). Genetically engineered resistance: transgenic plants, in: Plant viruses. vol.2: Pathology. pp. 347-362. Edited by C.L. Mandahar. Boca Raton: CRC Press.

Hoekema, A., Hirsch, P.R., Hooykaas, P.J.J. & Schilperoort, R.A. (1983). A binary vector system based on separation of vir- and T-region of the Agrobacterium tumefaciens Ti-plasmid. Nature 303: 179-180.

Holmes, F.O. (1958). A single-gene resistance test for viral relationship as applied to strains of spotted wilt virus. Virology 5: 382-390.

Horsch, R.B., Fry, J.E., Hoffman, N.L., Eichholtz, D., Rogers, S.G. & Fraley, R.T. (1984). A simple and general method for transferring genes into plants. Science 227: 1229-1231.

Ie, T.S. (1964). An electron microscope study of tomato spotted wilt virus in the plant cell. Neth. J. Plant Pathol. 70: 114-115.

Ie, T.S. (1970). Tomato spotted wilt virus. CMI/AAB Descr. of Plant Viruses 39.

Ihara, T., Matsuura, Y. & Bishop, D.H.L. (1985). Analyses of the mRNA transcription processes of Punta Toro phlebovirus (Bunyaviridae). Virology 147: 317-325.

Iwaki, M., Honda, Y., Hanada, K., Tochihara, H., Yonaha, T., Hokama, K. & Yokohama, T. (1984). Silver mottle disease of watermelon caused by tomato spotted wilt virus. Plant Dis. 68: 1006-1008.

Kitajima, E.W. (1965). Electron microscopy of vira-cabeca virus (Brazilian tomato spotted wilt virus) within the host cell. Virology 26: 89-99.

Laemmli, U.K. 91970). 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. Anal. Biochem. 163: 16-20.

Mantel, W.P. & Van de Vrie, M. (1988). The Western flower thrips, Frankliniella occidentalis, a new thrips species causing damage in protected cultures in the Netherlands. Ent. Ber., Amst. 48: 140-144.

Mohamed, N.A. (1981). Isolation and characterization of subviral structures from tomato spotted wilt virus. J. Gen. Virol. 53: 197-206.

O'Malley, P.J. & Hartmann, R.W. (1989). Resistance to tomato spotted wilt virus in lettuce. Hort. Sci. 24: 360-362.

Paliwal, Y.C. (1974). Some properties and thrips transmission of tomato spotted wilt virus in Canada. Can. J. Bot. 52: 1177-1182.

Paterson, R.G., Scott, S.J. & Gergerich, R.C. (1989). Resistance in two Lycopersicon species to an Arkansas isolate of tomato spotted wilt virus. Euphytica 43: 173-178.

Peters, D., Avila, A.C. de, Kitajima, E.W., Resende, R. de O., De Haan, P. & Goldbach, R. (1991). An over-view of tomato spotted wilt virus. In: Virus-Thrips-Plant Interactions of TSWV, Proc. USDA workshop, Belltsville U.S.A., pp.1-14. Edited by H-.T. Hsu & R.H. Lawson. Nat. Techn. Inf. Serv. Springfield (VA).

Resende, R. de O., Avila, A.C. de, Goldbach, R. & Peters, D. (1991). Detection of tomato spotted wilt virus using polyclonal antisera in double antibody sandwich (DAS) ELISA and cocktail ELISA. J. Phytopathol. 132: 46-56.

Sakimura, K. (1962). The present status of thrips-borne viruses. In: Biological transmission of disease agents, pp. 33-40. Edited by K. Maramorosch. New York: Academic Press.

Shapiro, G.I. & Krug, R.M. (1988). Influenza virus RNA replication in vitro: Synthesis of viral template RNAs and virion RNAs in the absence of an added primer. J. Virol. 62: 2285-2290.

Smith, P.G. & Gardner, M.W. (1951). Resistance in tomato to the spotted wilt virus. Phytopathology 41: 257-260. Van den Hurk, J., Tas, P.W.L. & Peters, D. (1977). The ribonucleic acid of tomato spotted wilt virus. J. Gen. Virol. 36: 81-91.

Vidal, S. & Kolakofski, D. (1989). Modified model for the switch from Sendai virus transcription to repli-cation. J. Virol. 63: 1951-1958.

Vinogradov, V.A., Trofimova, L.I. & Sarychev, Y.F. (1982). Field resistance of varieties, hybrids and chemical mutants of tobacco to tobacco thrips and tomato spotted wilt virus. Tabak 1982 (no.3): 43-44.

120

CHAPTER 8   CONCLUDING REMARKS

Tomato spotted wilt virus has become a limiting factor in the cultivation of many crops and ornamental plant species. The rapid spread of its most important vector species (Frankliniella occidentalis), is the most plausible explanation of its recent revival in the temperate climate zones of both the Old and New World. Besides the economic importance, also scientific reasons may explain the renewed interest to this plant virus.

TSWV is characterized by its typical enveloped, spherically-shaped particles, which are covered in a layer of spikes. The genome consists of three species of linear single stranded RNA, which are tightly wrapped with nucleocapsid protein to form pseudo-circular nucleocapsids. These unique properties among plant viruses formerly led to a classification into a monotypic plant virus group; the tomato spotted wilt virus group (Ie, 1970). The analysis of the structure and coding properties of two of the three genomic RNA segments (Chapters 3, 4 and 5), has revealed that TSWV actually belongs to the Bunyaviridae, a large family of viruses with arthropod and mammalian hosts (Bishop et al., 1980; Elliott, 1990).

TSWV S RNA is 2916 nucleotides long and encodes the nucleocapsid protein (N of 28.8K) in a viral complementary sense subgenomic mRNA species and a putative non-structural protein (NSs of 52.4K) in a viral sense subgenomic mRNA. No sequence relationships can be identified between TSWV N and NSs to any other protein (Fig.8.1a). This typical coding arrangement, termed ambisense, is also found for the S RNA segments of uuku- and phleboviruses, two genera within the Bunyaviridae (Giorgi et al., 1991) and for the S and L RNA segments of arenaviruses (Auperin et al., 1984; Salvato & Shimomaye, 1989). Recently, ambisense coding strategies have also been reported for the RNA segments 3 and 4 of tenuiviruses, i.e. rice stripe virus (Gingery, 1988; Kakutani et al., 1990, 1991; Zhu et al., 1991) and maize stripe virus (Huiet et al., 1991). Tenuiviruses represent a group of segmented negative-strand viruses, which seem to lack a lipid envelope. Rice stripe virus is the prototype of this small family of plant viruses. The intergenic region of TSWV S RNA consists of an A-rich stretch of

nucleotides followed by an U-rich stretch and can be folded into a stable 'hairpin' structure. Among the ambisense RNA segments, two types of intergenic regions can be distiguished. The S RNA segments of TSWV, Punta Toro virus (PTV), Uukuniemi virus and the segments 3 and 4 of tenuiviruses have A-U rich intergenic regions. The S RNAs of Rift Valley fever virus, Toscana virus, Sandfly Sicilian fever virus and the S and L RNAs of arenaviruses have G-C rich intercistronic regions. The 3' ends of the subgenomic mRNAs of only two viruses (PTV and Tacaribe virus) have been characterized. Transcription termination of both the N and the NSs mRNAs of PTV occurs in the vicinity of the loop of the proposed hairpin structure. On the basis of the lengths of the subgenomic mRNAs of TSWV S RNA (1.2 kb and 1.7 kb), it seems likely that both transcripts also terminate in the top of the intergenic hairpin. For Tacaribe virus, no special sequences that could function as termination signals were identified in the S RNA. In this case the structure at the 3' ends of the mRNAs, rather than particular sequences in the template RNA may be involved in transcription termination. The sequences which signal termination of transcription have not been elucidated in the other viruses with ambisense RNA segments.

TSWV L RNA is 8897 nucleotides long and is completely of negative polarity. One large ORF is located on the viral complementary strand, which corresponds with a predicted protein of 331.5K In analogy to other bunyaviruses it can be assumed that TSWV L RNA encodes the L protein, the viral polymerase (Fig.8.1). Comparison of the predicted L protein of TSWV to the putative polymerases of two other bunyaviruses, Bunyamwera and Hantaan virus and to PB1 of Influenza A virus, revealed the presence of five conserved amino acid sequence motifs, which are specific for all polymerases using RNA as a template (Poch et al.,1989). These signature sequences most likely represent the catalytic domains of the polymerase molecules. The comparative studies furthermore demonstrated that obviously, TSWV is more closely related to Bunyamwera virus than Hantaan virus is to the prototype of the Bunyaviridae (Fig.8.2). Bunya- and arenaviral L proteins are much larger than PB1, of the influenzaviruses. The active RNA polymerase of influenzaviruses consists of three subunits (PA, PB1 and PB2), which are encoded by the three largest RNA segments (Chapter 2). Since PB1 is the
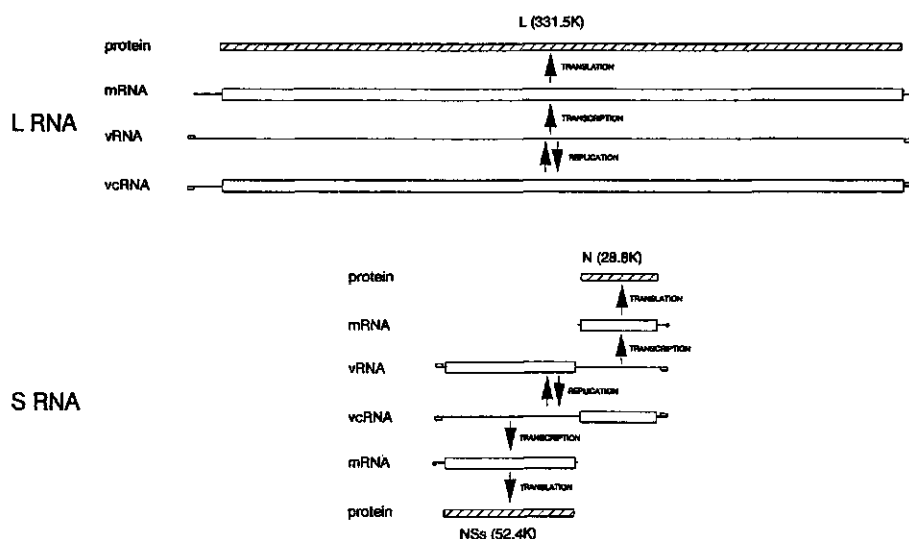
122

Fig.8.1: Schematic representation of the ambisense gene arrangement of TSWV S RNA and the negative-stranded L RNA segment.

core polymerase subunit, it can be anticipated that the catalytic domains responsible for 'cap-snatching' and polyadenylation of the mRNAs are most likely present in the PA and PB2 subunits.

As outlined in Chapters 3 and 5, a defective L RNA molecule was present in the TSWV line used in these experiments. Truncated genomic L RNAs are abundantly present in some TSWV isolates, when maintained by mechanical transmission (Resende et al., 1991).

Both TSWV S and L RNA have complementary terminal sequences, which is a property typical for negative-strand viruses. The termini can be folded in 'panhandle' structures, which are involved in the formation of circular nucleocapsids in the virus particle (Chapter 3). Moreover, they certainly will contain recognition signals for the viral polymerase and hence, play an important role in transcription and replication of the genomic RNAs (Chapter 2). The terminal sequences of members of the Bunyaviridae are more or less genus-specific. On the basis of these sequences, the genera can be

clustered into three groups, the nairoviruses, the phlebo/uukuviruses and the bunya/hantaviruses. Remarkably the termini of the TSWV RNAs show remarkable sequence homology to those of RNA segment 3 of Thogoto virus, a tick-borne member of the Orthomyxoviridae (chapter 5). This suggests that these viruses share some genetical interrelationship.

Taken all morphological, serological and molecular data into account, it is clear that TSWV should be classified as the representative of a distinct genus within the Bunyaviridae. Recently the genus name tospovirus has been accepted by the ICTV (International Committee on Taxonomy of Viruses). During the coarse of the investigations described in this thesis, other TSWV-like viruses have been described such
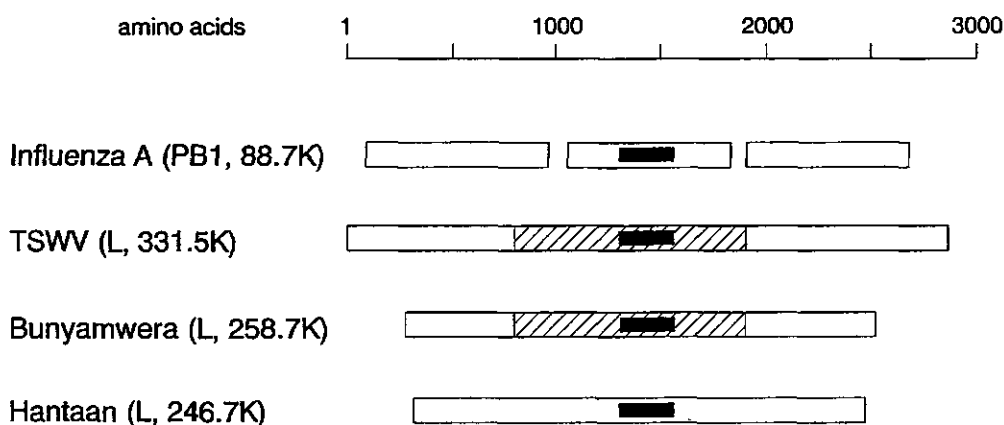
Fig.8.2: Schematical representation of the L proteins of TSWV, Bunyamwera and Hantaan virus and comparison with the polymerase P proteins of influenza A virus. The arced boxes represent regions within the TSWV and Bunyamwera L proteins with significant amino acid sequence homology (27% identity). The black boxes represent the regions harbouring the putative polymerase motifs.

124

as isolates from Impatiens, watermelon and peanut (Kameyi-Iwaki et al., 1988; Law & Moyer, 1990; Avila et al., 1991; Reddi et al., 1991). These viruses may represent different virus species within this newly created genus.

Despite the detailed knowledge on the molecular properties of many members of the negative-strand virus supergroup, the evolutionary relationships between the different families and genera remain unclear. For example, as outlined above, on the basis of amino acid sequence homology in the polymerase proteins, TSWV is more closely related to Bunyamwera virus than Hantaan virus is. Nucleotide sequence homology at the RNA termini however, suggests the opposit. The identical termini of the phlebo/uukuviral RNAs and those of tenuiviruses suggests a close ancestral relationship, although tenuiviruses do not have enveloped virus particles and contain 4 genomic RNA species. In summary, no phylogenetic relationships can be deduced from the complex patterns of similarities between these viruses. This is a general problem in the definition of higher taxa for viruses.

Now nucleotide sequence data and cDNA clones have become available, new molecular detection methods can be developed, which may serve as alternatives of the currently applied ELISA assays. Chapter 6 describes the use of two molecular techniques for sensitive detection of TSWV. The first assay is based on dot blot hybridization, using synthetic radioactively labelled RNA probes. It enables detection of approximately 1 pg of viral RNA in TSWV-infected leaf material. The second method (PCR) is based on amplification of DNA fragments, using specific oligonucleotides, reverse transcriptase and taq-polymerase. In this assay, quantities as low as 0.1 pg of viral RNA can be detected. In principle, after further optimizations, both methods can be applied in TSWV diagnosis.

Another result of the research described in this thesis is that cloned TSWV genes have become available. This has opened the way to investigate whether they can be utilized for creating genetically engineered resistance to this harmful pathogen. Sofar, no plant resistance genes to TSWV have been characterized, which can be of potential use in breeding programs. Chapter 7 deals with the introduction and expression of the TSWV nucleocapsid protein (N) gene in tobacco. For several other negative-strand

viruses it has been reported that the N protein plays an important regulatory role in the viral infection cycle. The amount of free unassembled N protein in the cytoplasm discriminates whether transcription or replication takes place by the viral RNA polymerase (Chapter 2). It may be anticipated that high levels of N protein in transgenic plants lead to immediate and abortive replication of incoming viral RNA. The experiments described in Chapter 7 demonstrate that high levels of TSWV N protein accumulated in transgenic tobacco. Moreover, a number of lines show reduced susceptibility to TSWV-infection. Upon mechanical inoculation, the numbers of primary infections that appear on the leaves of protected plants are drastically decreased. The few plants that still become systemically infected show a delay in symptom development. Obviously the resistance works at two levels, i.e. primary infection and cell-to-cell or systemic movement. In addition to coat protein-mediated resistance for positive-strand RNA viruses (Beachy et al., 1990; Hemenway et al., 1990), the experiments described in chapter 7 clearly show successfull genetically engineered resistance to a negative-strand RNA virus. Although the mechanisms by which this resistance works has remained unclear, this approach may be generally applied for creating resistance to minus-strand RNA viruses of plants (tospo-, tenui- and rhabdoviruses).

REFERENCES

Auperin, D.D., Romanowski, V., Galinski, M. & Bishop, D.H.L. (1984). Sequencing studies of Pichinde arenavirus S RNA indicate a novel coding strategy, ambisense viral S RNA. J. Virol. 52: 897-904.

Avila, A.C. de, De Haan, P., Kitajima, E.W., Kormelink, R., Resende, R. de O., Goldbach, R. & Peters, D. (1991). Characterization of a distinct tomato spotted wilt virus (TSWV) isolate from Impatiens sp. in the Netherlands. submitted for publication.

Beachy, R.N., Loesch-Fries, S. & Tumer, N.E. (1990). Coat protein-mediated resistance against virus infection. Ann. Rev. Phytopathol. 28:451-474.

Bishop, D.H.L., Calisher, C., Casals, J., Chumakov, M.P., Gaidamovich, S.Ya., Hannoun, C., Lvov, D.K., Marshall, I.D., Okerblom, N., Petterson, R.F., Porterfield, J.S., Russel, P.K., Shope, R.E. & Westaway, E.G. (1980). Bunyaviridae. Intervirology 14: 125-143.

Elliott, R.M. (1990). Molecular biology of the Bunyaviridae. J. Gen. Virol. 71: 501-522.

Gingery, R.E. (1988). The rice stripe virus. In: The plant viruses, vol. 4, The filamentous plant viruses, pp. 297-329. Edited by R.G. Milne. New York, London: Plenum Press.

Giorgi, C., Accardi, L., Nicoletti, L., Gro, M.C., Takehara, K., Hilditch, C., Morikawa, S. & Bishop, D.H.L. (1991). Sequences and coding strategies of the S RNAs of Toscana and Rift valley fever viruses compared to those of Punta toro, Sicilian sandfly fever, and Uukuniemi viruses. Virology 180: 738-753.

Hemenway, C., Haley, L., Kaniewski, W.K., Lawson, E.C., O'Connell, K.M., Sanders, P.R., Thomas, P.E. & Tumer, N.E. (1990). Genetically engineered resistance: transgenic plants. In: Plant viruses, vol. 2: Pathology, pp. 347-362. Edited by C.L. Mandahar. Boca Raton: CRC Press.

Huiet, L., Klaasen, V., Tsai, J.H. & Falk, B.W. (1991). Nucleotide sequence and RNA hybridization analyses reveal an ambisense coding strategy for maize stripe virus RNA3. Virology 182: 47-53.

Iapalucci, S., Lopez, N. & Franze-Fernandez, M.T. (1991). The 3' end of the Tacaribe arenavirus subgenomic RNAs. Virology 182: 269-278.

Kakutani, T., Hayano, Y., Hayashi, T. & Minobe, Y. (1990). Ambisense segment 4 of rice stripe virus: possible evolutionary relationship with phleboviruses and uukuviruses (Bunyaviridae). J. Gen. Virol. 71: 1427-1432.

Kakutani, T., Hayano, Y., Hayashi, T. & Minobe, Y. (1991). Ambisense segment 3 of rice stripe virus: the first instance of a virus containing two ambisense segments. J. Gen. Virol. 72: 465-468.

Kameyi-Iwaki, M., Hanada, K., Honda, Y. & Tochihara, H. (1988). A watermelon strain of tomato spotted wilt virus (TSWV-W) and some properties of its nucleocapsid (abstract). Fifth International Congress of Plant Pathology. Japan.

Law, M.D. & Moyer, J.W. (1990). A tomato spotted wilt-like virus with a serologically distinct N protein. J. Gen. Virol. 71: 933-938.

Poch, O., Sauvaget, I., Delarue, M. & Tordo, N. (1989). Identification of four conserved motifs among the RNA-dependant polymerase encoding elements. EMBO J. 8: 3867-3874.

Reddi, D.V.R., Sudarshana, M.R., Ratna, A.S., Reddi, A.S., Amin, P.W., Kumar, I.K. & Murthy, A.K. (1991). The occurrence of yellow spot virus, a member of tomato spotted wilt virus group, on peanut (Arachis hypogaea L.) in India. In: Virus-Thrips-Plant Interactions of TSWV, Proc. USDA workshop, Belltsville U.S.A., pp.77-88. Edited by H-.T. Hsu & R.H. Lawson. Nat. Techn. Inf. Serv. Springfield (VA).

Resende, R. de O., De Haan, P., Avila, A.C. de., Kormelink, R., Goldbach, R. & Peters, D. (1991). Generation of envelope and defective interfering RNA mutants of tomato spotted wilt virus by mechanical passage. J. Gen. Virol. in press.

Salvato, M. & Shimomaye, E.M. (1989). The completed sequence of lymphocytic choriomeningitis virus reveals a unique RNA structure and a gene for a zinc-finger protein. Virology 173: 1-10.

Zhu, Y., Hayakawa, T., Toriyama, S. & Takehashi, M. (1991). Complete nucleotide sequence of RNA 3 of rice stripe virus: an ambisense coding strategy. J. Gen. Virol. 72: 763-767.

SAMENVATTING

Het bronsvlekkenvirus virus van de tomaat, tomato spotted wilt virus (TSWV) was voor lange tijd de grote onbekende onder de plantevirussen. Ondanks het feit dat het een economisch erg belangrijk plantepathogeen is, kwam het onderzoek aan dit virus pas relatief laat op gang. Dit wordt in de eerste plaats veroorzaakt door het feit dat het een labiel en moeilijk hanteerbaar virus is. In de tweede plaats is TSWV pas echt onderwerp van intensieve studie geworden nadat één van de belangrijkste vectoren voor verspreiding, de Californische thrips (Frankliniella occidentalis) zich vanuit de Noord-Amerikaanse westkust is gaan verspreiden over het noordelijk halfrond en wellicht de gehele wereld. Daarnaast is TSWV een wetenschappelijk interessant virus. De virusdeeltjes (met een diameter van 80-110 nm) bestaan uit nucleocapsiden, omgeven door een lipidemembraan, die bedekt is met 'spikes'. Het genoom bestaat uit drie enkelstrengs lineaire RNA segmenten, die S (small), M (medium) en L (large) RNA worden genoemd. De complexe morfologie van de virusdeeltjes samen met het feit dat er nagenoeg geen moleculaire gegevens zijn, heeft lange tijd een goede classificatie in de weg gestaan.

Dit proefschrift beschrijft een moleculair genetisch onderzoek aan TSWV, met als doel enerzijds de moleculaire eigenschappen en de taxonomische positie te bepalen en anderzijds een bijdrage te leveren aan de beheersing van dit schadelijke virus.

De hoofdstukken drie, vier en vijf beschrijven de moleculaire clonering van het virale genoom en de vaststelling van de nucleotidenvolgorde van respectievelijk het kleinste (S) en het grootste (L) genomische RNA segment. De nucleotidenvolgorde van het middelgrote (M) RNA segment bleef bij dit onderzoek buiten beschouwing.

Het TSWV S RNA is 2916 nucleotiden lang en bevat twee genen in een zogenaamde 'ambisense' rangschikking. Op de virale RNA streng ligt een non-structureel eiwit (NSs, 52,4K) gecodeerd, terwijl het nucleocapside eiwit (N, 28,8K) op de viraal complementaire RNA streng gecodeerd ligt. Voor beide eiwitten kan geen significante aminozuur sequentie homologie gevonden worden met enig ander eiwit. De twee genen komen tot

129

expressie middels de synthese van subgenomische mRNA moleculen. De regio tussen beide genen, rijk aan A en U residuen, kan gevouwen worden in een haarspeld structuur en bevat waarschijnlijk belangrijke signalen voor transcriptie door het virale polymerase.

Het TSWV L RNA is 8897 nucleotiden lang en heeft een negatieve polariteit, met één groot open leesraam op de viraal complementaire streng. Het voorspelde genproduct bevat een vijftal geconserveerde aminozuur sequentie motieven, die typerend zijn voor alle polymerasen met RNA als matrijs. Dit maakt het erg waarschijnlijk dat het L RNA codeert voor het virale RNA polymerase (L, 331,5K). Zoals beschreven in de hoofdstukken 3 en 5, werd een verkort L RNA molecuul aangetroffen in het gebruikte TSWV isolaat. Verkorte L RNA segmenten zijn in diverse isolaten aangetroffen, die middels mechanische inoculatie in stand worden gehouden.

Zowel het S als het L RNA bezitten complementaire uiteinden. Deze vormen waarschijnlijk de herkenningsplaatsen voor het virale RNA polymerase. Vergelijking met andere RNA virussen laat zien dat er homologie bestaat tussen de terminale sequenties van TSWV en die van Thogoto virus, een door teken overgebracht orthomyxovirus. Dit kan mogelijk duiden op evolutionaire verwantschap tussen TSWV en de Orthomyxoviridae. De verkregen moleculaire gegevens samen met de morfologische eigenschappen laten zien dat TSWV behoort tot de familie der Bunyaviridae, een grote groep door arthropoden verspreide virussen. De structuur en genetische organisatie van het TSWV genoom lijkt sterk op dat van phlebo- en uukuvirussen, die twee genera vormen binnen deze grote virusfamilie. Omdat er geen serologische verwantschap bestaat tussen TSWV en de andere bunyavirussen en omdat er ook geen homologie in nucleotiden- of aminozuurvolgorde bestaat (behalve de geconserveerde 'polymerase motieven' in de L eiwitten natuurlijk), moet TSWV beschouwd worden als de vertegenwoordiger van een nieuw genus binnen de Bunyaviridae. Recentelijk is de genus naam tospovirus geaccepteerd door de ICTV (Internationaal Comité voor Taxonomie van Virussen).

Met behulp van de verkregen cDNA clonen en de nucleotidenvolgorde informatie is het thans mogelijk geworden om gevoelige detectiesystemen te onwikkelen, die een aanvulling kunnen zijn op de reeds bestaande serologische detectiemethoden. Hoofdstuk

130

zes gaat over de onwikkeling van een detectiemethode gebaseerd op 'dot blot' hybridisatie met behulp van radioactief gemerkte synthetische RNA transcripten en van een methode gebaseerd op de 'polymerase chain reaction' (PCR). Respektievelijk 1 en 0,1 picogram viraal nucleinezuur kon met deze methoden worden gedecteerd in TSWV-geïnfecteerd plantemateriaal. In principe zijn beide methoden bruikbaar voor vroegtijdige TSWV diagnose.

Door de beschikbaarheid van gecloneerde virale genen lag het vervolgens voor de hand te zoeken naar mogelijkheden voor resistentie tegen TSWV, door introductie van virale genen in waardplanten. Dit omdat tot op heden geen bruikbare natuurlijke resistentie genen tegen dit virus gevonden zijn. Gekozen is voor het tot expressie brengen van het TSWV N gen in tabak, omdat voor andere negatief-strengs RNA virussen vastgesteld was dat de hoeveelheid intracellulair N eiwit een sterk regulerende rol speelt in de virale infectiecyclus. De hoeveelheid vrij, cytoplasmatisch N eiwit bepaalt immers of er transcriptie danwel replicatie plaatsvindt door het virale polymerase. Wellicht zal een grote hoeveelheid N eiwit in transgene planten leiden tot voortijdige en abortieve replicatie van binnenkomend viraal RNA. De experimenten beschreven in hoofdstuk zeven laten zien dat een aantal transgene tabakslijnen verkregen werden die inderdaad verminderd vatbaar geworden zijn voor TSWV. Hoewel nog erg veel onderzoek nodig is, vormen deze resultaten een eerste en belangrijke stap op weg naar de beheersing van TSWV.

CURRICULUM VITAE

Peter T. de Haan werd op 8 april 1961 geboren te Heteren. In 1979 behaalde hij het VWO-B diploma aan het Heldring College te Zetten. In hetzelfde jaar begon hij aan zijn studie Biologie, met specialisatie 'cel', aan de Landbouw Universiteit te Wageningen. Het doctoraal examen werd behaald in 1986 met als hoofdvakken Moleculaire Biologie en Erfelijkheidsleer.

Van februari 1986 tot februari 1989 was hij als Wetenschappelijk Assistent werkzaam op de vakgroep Virologie van de Landbouw Universiteit, waar hij moleculair genetisch onderzoek verrichtte aan het bronsvlekken virus van de tomaat, zoals beschreven in dit proefschrift. Sedert juni 1989 is hij als Wetenschappelijk Medewerker verbonden aan de vakgroep Virologie, op een samenwerkingsproject met de Zaadunie B.V. te Enkhuizen. Dit onderzoeksproject richt zich op vroegtijdige diagnose van en resistentie tegen het bronsvlekken virus van de tomaat.

133

ACCOUNT

De Haan. P., Wagemakers, L., Goldbach, R. & Peters, D. (1989). Tomato spotted wilt virus, a new member of the Bunyaviridae? In: Genetics and Pathogenicity of Negative-strand viruses, pp. 287-290. Edited by D. Kolakofski & B.W.J. Mahy. Amsterdam: Elsevier Science Publishers B.V.

De Haan, P., Wagemakers, L., Peters, D. & Goldbach, R. (1989). Molecular cloning and terminal sequence determination of the S and M RNA of tomato spotted wilt virus. J. Gen. Virol. 70: 3469-3473.

De Haan, P., Wagemakers, L., Peters, D. & Goldbach, R. (1990). The S RNA segment of tomato spotted wilt virus has an ambisense character. J. Gen. Virol. 71: 1001-1007.

Huguenot, C., Van den Dobbelsteen, G., De Haan, P., Wagemakers, C.A.M., Drost, G.A., Osterhaus, A.D. M.E. & Peters, D. (1990). Detection of tomato spotted wilt virus using monoclonal antibodies and riboprobes. Arch. Virol. 110: 47-62.

Kormelink, R., Kitajima, E.W., De Haan, P., Zuidema, D., Peters, D. & Goldbach, R. (1991). The non-struc tural protein (NSs) encoded by the ambisense S RNA segment of tomato spotted wilt virus is associated with fibrous structures in infected plant cells. Virology 181: 459-468.

Peters, D., Avila, A.C. de, Kitajima, E.W., Resende, R. de O., De Haan, P. & Goldbach, R. (1991). An over- view of tomato spotted wilt virus. In: Virus-Thrips-Plant Interactions of TSWV, Proc. USDA workshop, Belltsville U.S.A., pp.1-14. Edited by H-.T. Hsu & R.H. Lawson. Nat. Techn. Inf. Serv. Springfield (VA).

De Haan, P., Kormelink, R., Peters, D. & Goldbach, R. (1991). Genetic organization and expression of the tomato spotted wilt virus genome. In: Virus-Thrips-Plant Interactions of TSWV, Proc. USDA workshop, Belltsville U.S.A., pp. 60-66. Edited by H-.T. Hsu & R.H. Lawson. Nat. Techn. Inf. Serv. Springfield (VA).

De Haan, P., Kormelink, R., Resende, R. de O., Van Poelwijk, F., Peters, D. & Goldbach, R. (1991). Tomato spotted wilt virus L RNA encodes a putative RNA polymerase. J. Gen. Virol. in press.

Resende, R. de O., De Haan, P., Avila, A.C. de., Kormelink, R., Goldbach, R. & Peters, D. (1991). Generation of envelope and defective interfering RNA mutants of tomato spotted wilt virus by mechanical passage. J. Gen. Virol. in press.

Avila, A.C. de, De Haan, P., Kitajima, E.W., Kormelink, R., Resende, R. de O., Goldbach, R. & Peters, D. (1991). Characterization of a distinct tomato spotted wilt virus (TSWV) isolate from Impatiens sp. in the Netherlands. Submitted for publication.

Gielen, J.J.L., De Haan, P., Kool, A.J., Peters, D., Van Grinsven, M.Q.J.M. & Goldbach, R. (1991). Engi- neered resistance to tomato spotted wilt virus, a negative-strand RNA virus. Submitted for publication.