# Structure-Function Relationships in the Structural Proteins and in the RNAs of Alphaviruses and Flaviviruses

Thesis by

Chang Soo Hahn

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1988

(Submitted May 12, 1988)

**Acknowledgments**

I would like to express my deepest appreciation to my advisor Jim Strauss, for his support and guidance as well as spending hard time to translate my Korean-English to real English throughout my staying at Caltech. I also would like to thank my thesis committee members, Drs. Norm Davidson, Elliot Meyerowitz, Ellen Rothenberg, and Mel Simon for sharing their time to discuss my thesis work.

I am also grateful to Ellen Strauss for her enthusiasm and many collaborations and discussions as well as my English teacher as a part-time job, even though she failed to make me a birdwatcher. And thanks to Charlie Rice who influenced and taught me various techniques during my first three years. We spent countless nights at the lab together doing experiments and having discussions until dawn.

Also, I would like to thank all my collaborators throughout my thesis work, Drs. Lynn Dalgarno, Joel Dalrymple, Diane Griffin, Alan Jackson, Shlomo Lustig, and Charlie Rice. And to Edith Lenches for excellent technical assistance and collaboration, to Reef Hardy, Frank Preugschat, Kangsheng Wang, Drs. Richard Kuhn, Bert Niesters, and Yukio Shirako for many helpful discussions and friendships, and Dr. Lee Hood and Tim Hunkapiller for letting me use their computer facility and teaching me how to use it.

Thank you my parents for your support and cheer which make me keep awake all night throughout those times. (There are 17 hours difference between here and Korea.)

Many thanks to all adminstrative staffs, especially Lody Kempees for managing all funds, Bill Lease for helping me purchase and service equipment, all the others in the stockroom, to the division cryptographers, especially Katie Patterson for typing all the manuscripts, and especially to Frank Ostrander, John Klemic and Tony Solyom in the shop for doing marvelous jobs for making and

fixing many instruments and letting me use all their machines.

Finally to my wife, Young, the best collaborator I ever had and will have, for all those discussions, encouragements, and both physical and mental helps. Without her companionship, none of this would have been possible.

## Abstract

The RNA virus families *Togaviridae* and *Flaviviridae* were considered one family as recently as 1983. These two families contain more than 100 members, many of which are important pathogens for humans and domestic animals. Studies on members of both families which were undertaken to increase our understanding of the functions of the virus structural proteins and of RNA sequence elements that interact with virus proteins, and of the evolution of these two families of RNA viruses, are presented in this thesis. These investigations include two on the nature and function of a virus-encoded self-protease that functions in the processing of the structural proteins, several studies on the role of the virus structural glycoproteins in assembly of progeny virions and viral virulence, and studies on the evolution of these viruses, including the demonstration that recombination has occurred in the *Togaviridae* to produce an important new pathogen, and that RNA sequence elements have been conserved during the evolution of the *Flaviviridae*.

Alphavirus structural proteins are translated from a subgenomic messenger RNA as a polyprotein, which is cleaved to the final products by proteolytic processing. This processing was studied by comparative sequence analysis of three temperature sensitive mutants of Sindbis virus (the type alphavirus) which have a defect in processing of the polyprotein at the nonpermissive temperature. These mutations were localized in the C-terminal region of the capsid protein. From the position of these mutations and from sequence similarities between the alphavirus capsid proteins and animal serine proteses, we hypothesized that the capsid protein was a serine autoprotease whose active site is formed by His-141, Asp-147 and Ser-215. To study this capsid protein protease activity in more detail, we have altered the proposed catalytic triad of the protease by site-directed mutagenesis. We have assayed the protease activity in the mutagenized capsid

proteins by *in vitro* transcription and translation, and attempted to rescue virus from mutagenized full-length "infectious" clones. The results supported our hypothesis.

Sindbis virus matures when preformed nucleocapsids acquire their envelopes by budding through virus-modified areas of the cell surface membrane. *ts*103 is a mutant of Sindbis virus which has a defect in this late maturation step such that it generates multicored particles, and it has provided a good system for studying structure-function relationships during viral assembly and maturation. Hybrid genomes were constructed that were formed from a full-length cDNA clone of wild type Sindbis in which restriction fragments were replaced with cDNA from *ts*103. Virus rescued from these constructs were used to determine the protein responsible for the multicored phenotype and to map the mutation. *ts*103 was found to have a single amino acid substitution in glycoprotein E2. The implications of this mutation for our understanding of virus assembly are discussed.

Virus surface glycoproteins are believed to be important determinants of virulence and tissue tropism. Neurovirulence of Sindbis virus for mice has been used as an animal model system in which to explore the effects of each individual protein or of a particular domain of a given protein, or even of a single amino acid residue, on neurovirulence. By constructing hybrid genomes among various strains of Sindbis virus at the cDNA level and rescuing virus *in vitro* using *in vitro* transcription and transfection, it was possible to evaluate the effect of each protein on neurovirulence. From these studies, we concluded that Sindbis virus glycoproteins are important determinants of neurovirulence, but not the sole determinants. The virulence phenotypes of various recombinant viruses in both weanling mice and suckling mice are discussed, with reference to the role of particular residues in producing the neurovirulent phenotype.

We have also undertaken a study of virulence in flaviviruses. The 17D vaccine strain of yellow fever virus, the type flavivirus, is one of the most reliable and stable live virus vaccines ever developed. By comparison of the nucleotide sequences of the 17D vaccine strain and its parental virulent Asibi virus we have located all of the changes which occurred during the attentuation of yellow fever to produce the 17D vaccine. This comparison led us to the conclusion that changes in the viral envelope protein play an important role in attenuation.

The 25 members of the genus *Alphavirus* have for the most part diverged by linear descent from a common ancestor. We have now found that Western equine encephalitis virus is an exception to this. Western equine encephalitis virus is a close relative of Sindbis virus as determined by immunological cross-reaction, but it is a New World virus that causes encephalitis in humans and horses, whereas Sindbis virus is an Old World virus not normally associated with encephalitis. The nucleotide sequence and deduced amino acid sequence of the structural proteins of Western equine encephalitis virus reveal that it arose by recombination between Eastern equine encephalitis virus (or a recent ancestor of it) and a virus closely related to Sindbis virus. The importance of recombination in the evolution of RNA viruses and in the generation of new potentially pathogenic virus strains, as well as the implications of the amino acid changes which have occurred in Western equine encephalitis virus (subsequent to the initial recombination event) for our understanding of the interaction between the structural proteins of alphaviruses, are discussed.

To study evolution in flaviviruses, sequences at the 5' and 3' ends of several flaviviruses have been compared. Conserved structures or sequence elements have been identified, one pair of which could result in cyclization of flavivirus RNA. The significance of these sequences is discussed.

## Table of Contents

# Introduction

In recent years, there have been rapid developments in the methods used for the determination of primary nucleotide sequences and deduced protein sequences (Sanger et al., 1977; Zimmern and Kaesberg, 1978; Maxam and Gilbert, 1980). In addition, for many plus stranded RNA viruses, including Sindbis virus (Rice et al., 1987), it is now possible to resurrect infectious virus from cDNA clones. These advances have made it possible to explore at the molecular level the structure-function relationships of proteins and/or nucleic acids.

The family Togaviridae and family Flaviviridae (which was considered a genus within the family Togaviridae as late as 1983) together contain some 100 species of RNA viruses, many of which are pathogenic for man and domestic animals (Strauss and Strauss, 1977; Chamberlain, 1980; Matthews, 1982; Westaway et al., 1985a,b; Brown, 1986; Griffin, 1986). There are more than 25 members in the genus Alphavirus of the family Togaviridae. Sindbis virus, the type alphavirus, has been studied extensively, in part because it is one of the least virulent members; most Sindbis isolates are asymptomatic in man (Griffin, 1986). Sindbis virus has a single strand RNA genome of 11,703 nucleotides, excluding the 3' poly(A) tail (Strauss et al., 1984; Strauss and Strauss, 1986). The viral RNA itself can act as an mRNA upon infection to generate large polypeptide precursors which are processed to four nonstructural proteins: nsP1, nsP2, nsP3 and nsP4. Since there is an opal termination codon at the junction between nsP3 and nsP4, translation of nsP4 can occur only by read-through (Strauss et al., 1983; Strauss et al., 1984; Hardy and Strauss, 1988; Strauss et al., 1988).

**Synthesis and processing of alphavirus structural proteins**

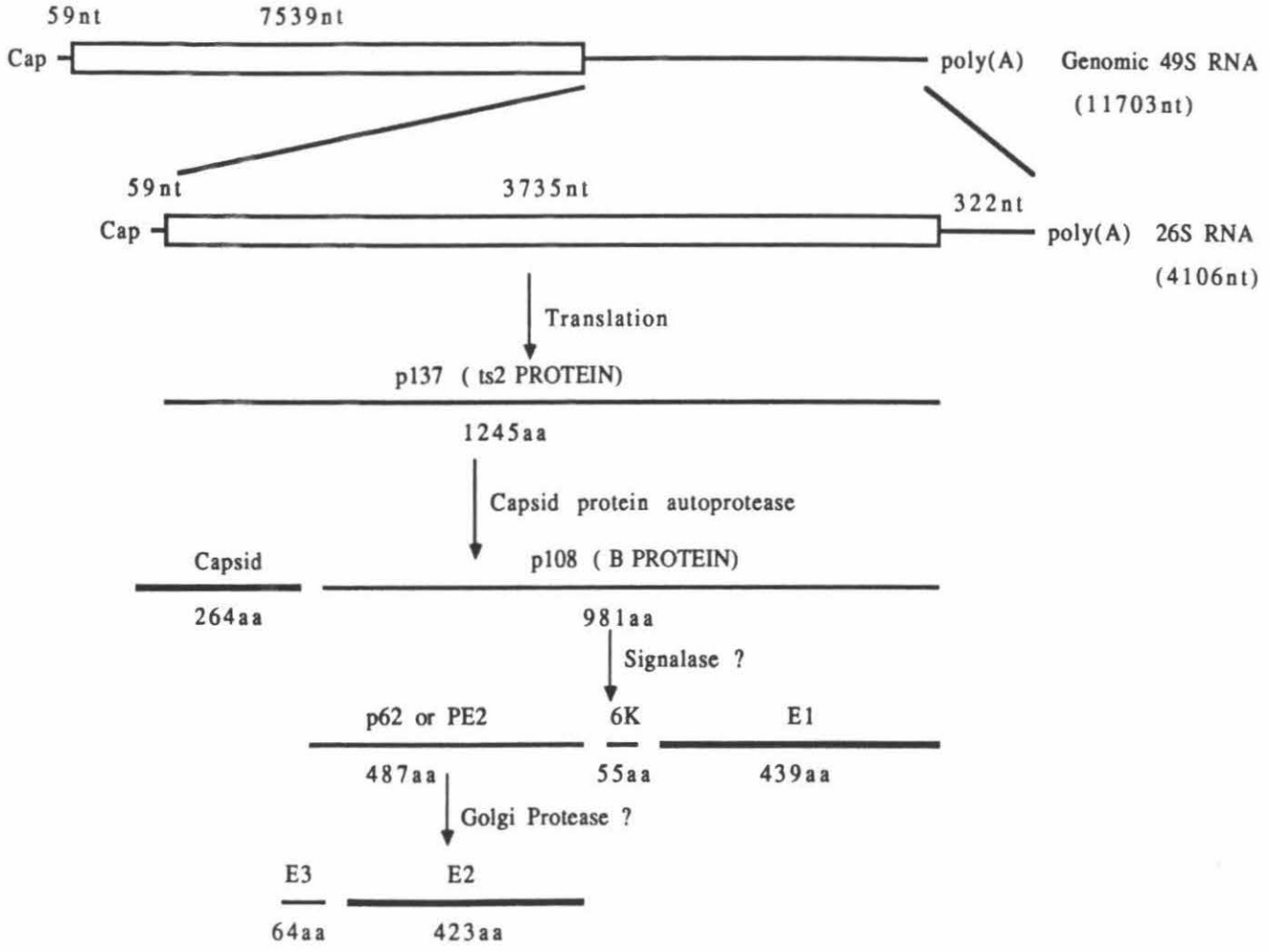
Alphavirus structural proteins are translated as a large polypeptide precursor from a subgenomic messenger, 26S RNA, which is the 3' terminal one-third of the genomic 49S RNA (Simmons and Strauss, 1972). The capsid protein is located at the amino terminus of the precursor and is followed by an envelope

protein precursor (PE2) and the second envelope protein (E1) (Garoff et al., 1980a,b; Rice and Strauss, 1981a,b). A number of proteolytic cleavages occur during processing of this polyprotein (Schlesinger and Kaariainen, 1980; Strauss and Strauss 1987a,b). Several lines of evidence suggest that the first cleavage, which releases the capsid protein from the polyprotein, occurs by autoproteolysis (Simmons and Strauss, 1974; Scupham et al., 1977; Aliperti and Schlesinger, 1978; Hahn et al., 1985; Melancon and Garoff, 1987). The remaining portion of the nascent polypeptide is then inserted into the rough endoplasmic reticulum of the cell. During insertion or shortly thereafter, the polyprotein is glycosylated with mannose-rich oligosaccharides and cleaved to form PE2 and E1 (Sefton, 1977). Late in maturation the final cleavage occurs, which converts PE2 to glycoproteins E2 and E3 (Fig. 1).

There are three complementation groups of temperature sensitive mutants that affect the structural proteins of the virus (Burge and Pfefferkorn, 1966a,b, 1967, 1968; Strauss and Strauss, 1980). Complementation groups D and E have defects in glycoprotein E1 (Arias et al. 1983) and E2 (Lindqvist et al. 1986) respectively. Sequence studies of complementation group C mutants have shown that the mutations are located in the capsid protein (see Part 1, Chapter 1; Hahn et al. 1985). Furthermore there is sequence similarity between the active sites of animal and insect serine proteases and the C-terminal half of the alphavirus capsid protein (Boege et al., 1981; Hahn et al., 1985). Site-directed mutagenesis of the proposed catalytic triad of the capsid protein autoprotease will be discussed in Chapter 2.

**Fig. 1** Synthesis and processing of alphavirus structural proteins. Untranslated regions of both genomic and subgenomic RNAs are shown as single lines and open reading frames for the nonstructural proteins and structural proteins are shown as open boxes (in genomic 49S RNA and subgenomic 26S RNA respectively). Translation products from the subgenomic 26S RNA and processing intermediates are indicated, and the final structural protein products are shown as bold lines.

Figure 1. Synthesis and processing of alphavirus structural proteins



**The role of alphavirus glycoproteins in assembly and virulence**

It has been known for a number of years that glycoproteins E1 and E2 are found closely associated as a heterodimer soon after their synthesis (Bracha and Schlesinger, 1976; Rice and Strauss, 1982). The two proteins are transported to the cell surface as a dimer and incorporated together into the budding virion. Although care must be taken in assigning individual functions to E1 and E2, since each protein in the dimer affects the conformation and availability of sites on the other, in general glycoprotein E2 plays a primary role in the initial interactions with the host cell and glycoprotein E1 contains the hemagglutinin and a fusion activity (E. G. Strauss and J. H. Strauss, 1985; J. H. Strauss and E. G. Strauss, 1985; Dalrymple et al. 1976; Chanas et al., 1982). In addition, interactions between the glycoproteins and the nucleocapsid play a crucial role in the assembly of virions (Fuller, 1987; Rice et al. 1982).

Sindbis virus mutant *ts*103, which was isolated by nitrous acid mutagenesis a decade ago (Strauss et al. 1976), is a minute plaque former which grows very slowly at any temperature and produces, under optimal conditions, 3-10% of the virus yield of the parental HR strain. It has been shown that *ts*103 has a defect in a late maturation stage and forms multi-cored particles. (Strauss et al. 1977). Unlike wildtype alphaviruses, whose virion particles have one nucleocapsid in a close fitting envelope that consists of a lipid bilayer in which are embedded trimeric spikes of the two viral glycoproteins E1 and E2 in an icosahedral array, *ts*103 does not have a fixed form and from one to several nucleocapsids are engulfed in a single envelope during budding. This is presumably due to a weakened interaction between the nucleocapsid and the cytoplasmic domains of the envelope proteins, caused by a defect in E1-E2 dimer formation and/or a conformational change in the trimeric spike which is the basic unit of the icosahedral structure (Fuller, 1987). The nature and location of the *ts*103 mutation will be discussed in Chapter 3.

Another approach to studying structure-function relationships in the alphavirus structural proteins has been to characterize viruses which have been selected for altered neurovirulence in animal systems. Changes in either replicase proteins or in structural proteins might affect tissue tropism and virulence. Alterations in replication enzymes leading to a slower rate of replication could attenuate the virus, as a more rapidly replicating virus is often more pathogenic since infection can become established before the host immune defenses are activated. In addition, alterations in the replication enzymes which change their interaction with host factor(s) (which factors may differ depending upon the type of tissue infected) could cause differential tissue tropism, since the efficiency of replication might be different in different tissues. Changes in the surface membrane protein(s), on the other hand, can change their affinity for certain tissue-specific receptors, which could lead to differential tissue tropism and alteration of the virulence phenotype. Changes in structural proteins can also alter virulence by affecting the kinetics of assembly of progeny virus.

For viruses with segmented genomes, elegant studies have localized the protein(s) responsible for temperature sensitivity and tissue tropism by genome reassortment. Such experiments have been performed, for example, for reoviruses (Ahmed and Fields, 1981) and arenaviruses (Ahmed and Oldstone, 1988), among others. However, for alphaviruses or flaviviruses, which have one long RNA molecule as their genome, such an approach is impossible. With the recent development of an "infectious" cDNA clone of Sindbis virus, from which infectious RNA can be transcribed in vitro, mapping of such genetic markers becomes possible (Rice et al.,1987). By making hybrid genomes between two strains which differ in either tissue tropism or virulence, one can determine which protein is important for a given phenotype. Furthermore it is also possible by site directed mutagenesis to test the effect of changes in a particular protein domain or even at a single amino acid residue.

Sindbis virus infection in mice has been studied as an experimental model of acute viral encephalitis (Johnson, 1965; Griffin, 1976; Olmsted et al., 1984, 1986). Wild-type Sindbis virus (AR339 strain) causes a fatal encephalitis in suckling mice and a nonfatal encephalitis in four-week-old weanling mice (Johnson et al., 1972). A strain of Sindbis virus that is highly lethal for weanling and adult mice was isolated after six intracerebral (IC) passages of wild-type Sindbis strain AR339(SV) alternating between suckling and weanling mice (Griffin and Johnson, 1977). This neuroadapted strain of Sindbis (NSV) has been genetically stable for many passages in cell culture. NSV replicates 5 to 10-fold more efficiently in the mouse brain than SV and its IC 50% lethal dose is 2-20 pfu in weanling mice. After IC inoculation with NSV weanling mice become ruffled, develop kyphoscoliosis and hind limb paralysis and have a high mortality (Griffin, 1986). There is a poliomyelitis, particularly involving the ventral horns, in the thoracic and lumbar spinal cord (Jackson et al., 1987, 1988; Griffin et al., 1988). Stanley et al. (1985, 1986) demonstrated that some anti-Sindbis virus E1 and E2 monoclonal antibodies can discriminate between NSV and SV, suggesting that changes in the surface glycoproteins may be associated with changes in virulence. In another approach, Olmsted et al. (1984,1986) have selected Sindbis virus variants in vitro that are attenuated in suckling mice (Polo et al., 1988). Thus, strains of Sindbis virus can be essentially avirulent for mice of all ages, or virulent for suckling mice but avirulent for weanling mice, or virulent for mice of all ages. Moreover, two different laboratory strains of Sindbis virus (our laboratory isolate HRSP and Toto1101, a recombinant virus between our HRSP and Dr. Schlesinger's HR strain) were also avirulent for mice of all ages. Recombinants among these four strains (NSV, SV, HRSP, and Toto1101) showed a gradient of virulence for mice of all ages (Lustig et al., 1988). These constructs and their biological properties will be discussed in Chapter 4.

**Studies on Flavivirus virulence**

Yellow fever virus, the type flavivirus, is a good system in which to study protein changes associated with altered virulence and tissue tropism. Yellow fever is an arthropod-borne virus, transmitted by mosquitos of the genera *Aedes* and *Haemagogus*. Its natural vertebrate host range is limited to primates in which it is both viscerotropic and neurotropic. In man, the virus causes a serious, often fatal, illness marked by liver and kidney involvement and hemorrhage. For several hundred years the virus caused epidemics in the Americas, Europe, and Africa that led to widespread human suffering. With the control of its urban vector, *Aedes aegypti*, beginning in the early 1900s, epidemic urban yellow fever disappeared. However the virus remains present in an enzootic cycle in the forests of South America and Africa and causes periodic outbreaks in neighboring human populations (Strode, 1951).

Reed (1901) first proved that yellow fever is transmitted by mosquitos and shortly thereafter, that the disease agent was filterable. The Asibi strain of yellow fever virus was isolated from a young Ghanian of that name in 1927 by the Rockefeller Foundation's West Africa Yellow Fever Commission (Stokes et al., 1928). This virus, which was maintained by monkey-monkey passage, causes an invariably fatal disease when inoculated into rhesus monkeys. From Asibi yellow fever virus Theiler and colleagues (Theiler, 1930; Theiler and Smith, 1937a,b) developed a live attenuated vaccine strain referred to as 17D. Starting with the Asibi strain, which had been passed 53 times in monkeys interspersed with passages in *A. aegypti*, the virus was propagated serially in cultures of embryonic mouse tissue (18 passages), minced whole chicken embryos (50 passages), and finally minced chicken embryos without brain and central nerve system tissue (152 passages). Between the 89th and 114th in vitro passage (from the start of the experiment, that is including the passages in embryonic mouse tissue and whole

chicken embryos), a marked change in virulence of the virus occurred. The reason for the change is not clear and attempts to repeat the experiment to develop additional avirulent strains by virus passage have failed. The 17D strain has been widely used as a human vaccine, being safe and highly effective. It causes a mild, generalized infection in humans (or other primates) with involvement of lymphoid tissue and minimal quantities of virus circulation in the blood, and the viscerotropism and neurotropism of the strain are less than that of wild type strains of yellow fever. By comparing the sequence of the entire genome of the parental Asibi virus and the vaccine strain 17D-204 which was derived from it, it was possible to determine the total number of nucleotide and amino acid substitutions that occurred during the derivation of the 17D strain, and to develop hypotheses as to which amino acid substitutions are primarily responsible for the altered tissue tropism and reduction in virulence of 17D (Rice et al.,1985; Hahn et al., 1987a,c; see Chapter 5).

**Evolution of alphaviruses and flaviviruses**

The alphavirus genus contains about 25 members, many of which have geographic variants (Porterfield, 1980). The vertebrate host range of this group of viruses is quite wide, and alphaviruses have been isolated from numerous species of birds and mammals as well as some reptiles and amphibians (Chamberlain, 1980). Alphaviruses have been divided into three to six serological subgroups or complexes, which are clearly separable on the basis of cross-neutralization tests (Porterfield, 1961; Karabatsos, 1975; Chanas et al., 1976). A classification of alphaviruses based on cross-neutralization but modified to take into account sequencing data and geographic range (discussed in more detail below) is given in Table 1.

Eastern equine encephalitis (EEE) virus is a New World virus capable, as its name implies, of causing encephalitis in man and horses. It forms a distinct

subgroup with at least two grographic variants (North American and South American). The Western Equine encephalitis (WEE) virus complex is also a New World complex containing several viruses (Table 1). WEE is also capable of causing encephalitis in man. The Old World viruses grouped with Sindbis virus in Table 1 have also been included in the WEE subgroup to the present, on the basis of neutralization cross reaction. Our recent sequence data (Hahn et al., 1988 and Chapter 6), however, demonstrate the WEE is a recombinant virus with part of its genome derived from an EEE-like virus and the structural proteins (which possess the neutralization epitopes) derived from a Sindbis-like virus. Thus it seems best to classify WEE in a subgroup distinct from the Sindbis group. Within the New World viruses, Highlands J and WEE are very similar antigenically (Calisher et al., 1980; Hayes and Wallis, 1977) and limited sequence data (Ou et al., 1982b, 1983) also suggests that Highlands J, which replaces WEE in the eastern United States, is very similar to WEE (i.e., was also derived from the ancestral recombinant virus that gave rise to WEE).

The Venezuelan equine encephalitis (VEE) virus subgroup consists of at least 6 New World viruses. The Semliki Forest (SF) virus subgroup, which has both New World and Old World members, was formerly considered as part of the VEE complex, but more recent sequencing data suggest that the SF group should be treated as a distinct subgroup. The SF group is unusual in having representatives in both the New World and the Old World. Although sequence data have been obtained for 4 of the Old World representatives, no sequence data for the New World members exist to determine exactly how closely related they are to the Old World viruses.

Table 1. Classification of Alphaviruses

| Geographic range | Subgroup | Virus | Disease symptoms |
|---|---|---|---|
| New World | EEE | EEE | Encephalitis |
| | WEE | WEE | Fever, Encephalitis |
| | | Aura | |
| | | Highlands J | |
| | | Fort Morgan | |
| | VEE | VEE | Fever, Encephalitis |
| | | Bijou Bridge | |
| | | Cabassou | |
| | | Everglades | Encephalitis |
| | | Mucambo | Fever |
| | | Pixuna | |
| | SF | Mayaro | Fever, Arthritis, Rash |
| | | Una | |
| Old World | SIN(WEE) | Sindbis | Fever |
| | | Kyzylagach | |
| | | Whataroa | |
| | MID | Middelburg | |
| | Ndumu | Ndumu | |
| | SF | Semliki Forest | Fever, Encephalitis |
| | | Bebaru | |
| | | Chikungunya | Fever, Arthritis, Rash |
| | | Getah | |
| | | Ross River | Rash, Polyarthritis |
| | | O'Nyong-nyong | Fever, Arthritis, Rash |
| | | Sagiyama | |
| | ? | Barmah Forest | |

Finally, three alphaviruses, Middelburg (MID), Ndumu, and Barmah Forest (BF) are Old World viruses that are best treated as separate subgroups, although MID and Ndumu have sometimes been classified in the WEE/SIN subgroup. Sequence data for MID and BF support their classification as distinct subgroups, and, in fact, MID is more closely related to SF on the basis of sequence than to Sindbis.

It is of interest that a number of New World alphaviruses are capable of causing encephalitis, whereas the Old World viruses typically cause a disease characterized by fever, rash, and, often, arthritis. The reasons for this are not clear at present.

Comparison of the complete and partial nucleotide sequences and deduced amino acid sequences available for alphaviruses demonstrated that they have all diverged from a common ancestor during evolution (Bell et al., 1984; Strauss and Strauss 1986). As indicated above, the relationships among the alphaviruses derived from the sequencing studies are compatible, for the most part, with those from serological cross-reactivity, which examine only the viral structural proteins. One exception has been the case of WEE. It is a New World virus with a wide geographic distribution, being found from western Canada to Mexico and, discontinuously, to Argentina, that is capable of causing encephalitis, although serologically it is most closely related to SIN, an Old World alphavirus which is not normally associated with encephalitis. SIN is widely distributed, being found in Europe, India, southeast Asia, Australia, and Africa. SIN has been implicated in only mild febrile illnesses in man, although close relatives of SIN such as Ockelbo virus in Europe (Niklasson et al., 1984) and Babanki virus in Africa cause a typical Old World alphavirus disease in man characterized by fever, rash, and arthritis. We have recently discovered the basis for the close serological relationship between these otherwise disparate alphaviruses. We have cloned and sequenced

the 3' terminal 4288 nucleotides of the BFS1703 strain of WEE. Close examination of the nucleotide sequence and the deduced amino acid sequence encoded therein clearly shows that WEE is a recombinant virus between EEE and a Sindbis-like progenitor (See Chapter 6). In this recombinant, only the surface glycoproteins appear to be derived from the Sindbis-like parent.

There has been a great deal of speculation about the importance of recombination for RNA virus evolution (Strauss and Strauss, 1988; Hodgman and Zimmern, 1988). In segmented RNA viruses, reassortment of individual genome segments during mixed infection, a form of recombination equivalent to the shuffling of chromosomes in diploid creatures, is readily demonstrable in cell culture. Also reassortment has been well documented as a major mechanism for generating new pandemic strains of influenza virus (Desselberger et al., 1978; Webster et al., 1982) Among the nonsegmented RNA viruses recombination has been in general more difficult to demonstrate, but it has been shown to occur in the picornaviruses (Cooper, 1977; Emini et al. 1984; Kew and Nottay 1984; Kirkegaard amd Baltimore, 1986), the coronaviruses (Makino et al., 1986), and bromoviruses (Bujarski and Kaesberg, 1986). Although well established in principle, evidence for the importance of such recombination in nature as a mechanism for generating successful new strains of virus is limited. WEE, a virus with a wide geographic range, provides a clear example of a successful recombinant that arose naturally, and lends support to the hypothesis that RNA recombination is an important force in the evolution of RNA viruses in general (Hahn et al., 1988; Chapter 6).

During divergent evolution, the rate of change is decidedly dependent on the functions of the proteins or RNA sequences. For example, replicase molecules are more highly conserved than structural proteins (Ahlquist et al. 1985; Argos et al., 1984; Franssen et al. 1984; Goldbach, 1987; Haseloff et al. 1984; Kamer and

Argos, 1984). Similarly, RNA sequence elements that serve as binding sites for virus proteins are more highly conserved. In the case of such elements, either the primary sequence or a secondary structure may be conserved. In alphaviruses, there are three conserved RNA sequences, one near the 5' end of the genome, the second in the junction region near the start of the subgenomic 26S RNA, and the third at the 3' end just before the poly(A) tail (Ou et al., 1982a,b, 1983). In addition, the 5' end of the 49S RNA has a conserved secondary structure in which the primary nucleotide sequence is not conserved (Ou et al., 1983). One of the elements is also conserved in rubella virus, which is in a different genus of the *Togaviridae* (Frey and Marr, 1988). Many plant viruses have a tRNA-like structure at the 3' end of the genome which may serve an important function in replication (Hall, 1979). Conserved sequences 12 to 15 nucleotides in length are found at the ends of the genome segments of influenza virus and of bunyaviruses (Strauss and Strauss, 1983).

In the case of flaviviruses, Rice et al. (1985) proposed that the 3'-terminal 87 nucleotides of yellow fever (YF) virus RNA form a stable secondary structure. Subsequently, West Nile (WN) virus (Brinton et al. 1986; Wengler and Castle, 1986), Japanese encephalitis (JE) virus (Takegami et al., 1986), and Dengue 4 (DEN 4) virus (Zhao et al. 1986) were proposed to have secondary structures very similar to that of YF RNA. Conservation of this structure supports the hypothesis that it is important for viral RNA replication (Rice et al., 1986), but direct evidence for its existence has been limited. Brinton et al. (1986) found that nucleotides within the putative hydrogen-bonded regions of the stem were partially resistant to ribonuclease suggesting that the structure was present in WN RNA in solution. In case of YF, no direct evidence of this conformation had been presented, and in fact Grange et al. (1985) proposed an alternative structure involving the 3'-terminal 120 nucleotides. In this thesis we present evidence for the existence

in solution of the 87 nucleotide structure in YF RNA (Chapter 7). In addition to the 3'-terminal conserved secondary structure there are two other conserved RNA sequences in the 3' noncoding region of flaviviruses and a third conserved sequence at 5' end of the genome near the start of translation. One of the 3' conserved sequences, which is adjacent to the secondary structure, is complementary to the 5' conserved element so that it can form a panhandle structure (Hahn et al., 1987b). These elements and their significance will be further discussed in Chapter 7.

# REFERENCES

Ahlquist, P., Strauss, E.G., Rice, C.M., Strauss, J.H., Haseloff, J., Zimmern, D., 1985. Sindbis virus proteins nsP1 and nsP2 contain homology to nonstructural proteins from several RNA plant viruses. J. Virol. 53: 536-542.

Ahmed, R., and Fields, B.N. 1981. Reassortment of genome segments between reovirus defective interfering particles and infectious virus: Construction of temperature-sensitive and attenuated viruses by rescue of mutations from DI particles. Virology 111: 351-363

Ahmed, R., and Oldstone, M.B.A. 1988. Organ specific selection of viral variants during chronic infection. J. Exp. Med. in press.

Aliperti, G. and Schlesinger, M. 1978. Evidence for an autoprotease activity of Sindbis virus capsid protein. Virology 90: 366-369.

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

Arias, C., Bell, J.R., Lenches, E.M., Strauss, E.G., and Strauss, J.H. 1983. Sequence analysis of two mutants of Sindbis virus defective in the intracellular transport of their glycoproteins. J. Mol. Biol. 168: 87-102.

Bell, J.R., Kinney, R.M., Trent, D.W., Strauss, E.G., and Strauss, J.H. 1984. An evolutionary tree relating eight alphaviruses based on amino terminal sequences of their glycoproteins. Proc. Nat. Acad. Sci. USA 81: 4702-4706.

Boege, U., Wengler, G., Wengler, G., and Wittmann-Liebold, B. 1981. Primary structures of the core proteins of the alphaviruses, Semliki Forest virus and Sindbis virus. Virology 113: 293-303.

Bracha, M., and Schlesinger, M.J. 1976. Defects in $RNA^+$ temperature-sensitive mutants of Sindbis virus and evidence for a complex of PE2-E1 viral glycoproteins. Virology 74: 441-449.

Brinton, M.A., Fernandez, A.V., and Dispoto, J.H. 1986. The 3'-nucleotides of flavivirus genomic RNA form a conserved secondary structure. Virology 153: 113-121.

Brown, F. 1986. The classification and nomenclature of viruses: Summary of results of meetings of the international committee on taxonomy of viruses in Sendai, September 1984. Intervirology 25: 141-143.

Bujarski, J.J., and Kaesberg, P. 1986. Genetic recombination between RNA components of a multipartite plant virus. Nature 321: 528-531.

Burge, B.W., and Pfefferkorn, E.R. 1966a. Isolation and characterization of conditional-lethal mutants of Sindbis virus. Virology 30: 204-213.

Burge, B.W., and Pfefferkorn, E.R. 1966b. Complementation between temperature-sensitive mutants of Sindbis virus. Virology 30: 214-223.

Burge, B.W., and Pfefferkorn, E.R. 1967. Temperature-sensitive mutants of Sindbis virus: Biochemical correlates of complementation. J. Virol. 1: 956-962.

Burge, B.W., and Pfefferkorn, E.R. 1968. Functional defects of temperature-sensitive mutants of Sindbis virus. J. Mol. Biol. 35: 193-205.

Calisher, C.H., Shope, R.E., Brandt, W., Casals, J., Karabatsos, N., Murphy, F.A., Tesh, R.B., and Wiebe, M.E. 1980. Proposed antigenic classification of registered arboviruses. I. Togaviridae, Alphavirus. Intervirology 14: 229-232.

Chamberlain, R.W. 1980. Epidemiology of arthropod-borne togaviruses: The role of arthropods as hosts and vectors and of vertebrate hosts in natural transmission cycles. In The Togaviruses, ed. Schlesinger R. W. (Ed.), Chapter 6, Academic Press, New York, pp. 175-227.

Chanas, A.C., Johnson, B.K., and Simpson, D.I.H. 1976. Antigenic relationships of alphaviruses by a simple microculture cross-neutralization method. J. Gen. Virol. 32: 295-300.

Chanas, A.C., Gould, E.A., Clegg, J.C.S., and Varma, M.G.R. 1982. Monoclonal antibodies to Sindbis virus glycoprotein E1 can neutralize, enhance infectivity and independently inhibit haemagglutination or haemolysis. J. Gen. Virol. 58: 37-46.

Cooper, P.D., 1977. Genetics of Picornaviruses. In: Comprehensive Virology, Vol. 9, Regulation and Genetics, Fraenkel-Conrat, H., and Wagner, R. R., (Eds.) Plenum Press, New York, pp.133-207.

Dalrymple, J., Schlesinger, S., and Russell, P.K. 1976. Antigenic characterization of two Sindbis envelope glycoproteins separated by isoelectric focusing. Virology 69: 93-103.

Desselberger, U., Nakajima, K., Alfino, P., Pedersen, F.S., Haseltine, W.A., Hannoun, C., and Palese, P., 1978. Biochemical evidence that "new" influenza virus strains in nature may arise by recombination (reassortment). Proc. Nat. Acad. Sci. U. S. A. 75: 3341-3345.

Emini, E.A., Leibowitz, J., Diamond, D.C., Bonin, J., and Wimmer, E., 1984. Recombinants of Mahoney and Sabin strain poliovirus type 1: Analysis of in vitro phenotypic markers and evidence that resistance to guanidine maps in the nonstructural proteins. Virology 137: 74-85.

Franssen, H., Leunissen, J., Goldbach, R., Lomonossoff, G., and Zimmern, D. 1984. Homologous sequences in nonstructural proteins from cowpea mosaic virus and picornaviruses. The EMBO Journal 3: 855-861.

Frey, T.K., and Marr, L.D. 1988. Sequence of the region coding for virion proteins C and E2 and the carboxy-terminus of the nonstructural proteins of rubella virus. Gene in press.

Fuller, S.D. 1987. The T=4 envelope of Sindbis virus is organized by interactions with a complementary T=3 capsid. Cell 48: 923-934.

Garoff, H., Frischauf, A.M., Simons, K., Lehrach, H., and Delius, H. 1980a. The capsid protein of Semliki Forest virus has clusters of basic amino acids and prolines in its amino-terminal region. Proc. Nat. Acad. Sci. USA 77: 6376-6380.

Garoff, H., Frischauf, A.M., Simons, K., Lehrach, H., and Delius, H. 1980b. Nucleotide sequence of cDNA coding for Semliki Forest virus membrane glycoproteins. Nature (London) 288: 236-241.

Goldbach, R., 1987. Genome similarities between plant and animal RNA viruses. Microbiol. Sci. 4:197-201.

Grange, T., Bouloy, M., and Girard, M. 1985. Stable secondary structures at the 3'-end of the genome of yellow fever virus (17*D* vaccine strain) FEBS Lett. 188:159-163.

Griffin, D.E. 1976. Role of the immune response in age-dependent resistance of mice to encephalitis due to Sindbis virus. J. Infect. Dis. 133: 456-464.

Griffin, D.E. 1986. Alphavirus pathogenesis and immunity. In The Togaviridae and Flaviviridae, S. Schlesinger and M.J. Schlesinger (Eds.), Plenum Publishing Corp., New York, Chapter 8, pp. 209-249.

Griffin, D.E., Hahn, C.S., Jackson, A.C., Lustig, S., Strauss, E.G., and Strauss, J.H. 1988. The basis of Sindbis virus neurovirulence. In: Cell Biology of Virus Entry, Replication, and Pathogenesis Compans, R., Helenius, A., and Oldstone, M.(Eds.) in press

Griffin, D.E. and Johnson, R. 1977. Role of the immune response in recovery from Sindbis virus encephalitis in mice. J. Immunol. 118: 1070-1075.

Hahn, C.S., Dalrymple, J.M., Strauss, J.H., and Rice, C.M. 1987a. Comparison of the virulent Asibi strain of yellow fever virus with the 17D vaccine strain derived from it. Proc. Nat. Acad. Sci. USA 84: 2019-2023.

Hahn, C.S., Hahn, Y.S., Rice, C.M.,Lee, E., Dalgarno, L., Strauss, E.G., and Strauss, J.H. 1987b. Conserved elements in the 3' untranslated region of flavivirus RNAs and potential cyclization sequences. J. Mol. Biol. 198: 33-41.

Hahn, C.S., Lustig, S., Strauss, E.G., and Strauss, J.H. 1988. Western equine encephalitis virus is a recombinant virus. Proc. Natl. Acad. Sci. USA, in press.

Hahn, C.S., Rice, C.M., Strauss, J.H., and Darlymple, J.M. 1987c. Comparison of the Asibi and 17D strains of yellow fever virus. In: Vaccines 87, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, pp. 316-321.

Hahn, C.S., Strauss, E.G. and Strauss, J.H. 1985. Sequence analysis of three Sindbis virus mutants temperature-sensitive in the capsid protein autoprotease. Proc. Natl. Acad. Sci. USA 82: 4648-4652

Hall, T.C. 1979. Transfer RNA-like structures in viral genomes. In: International Review of Cytology, Vol 60, Bourne, G.H., and Danielli, J.F. (Eds.) Academic press, New York, pp. 1-26.

Hardy, W.R., and Strauss, J.H. 1988. Processing of the nonstructural polyproteins of Sindbis virus: Study of the kinetics in vivo using monospecific antibodies. J. Virol. 62: 998-1007.

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

Hayes, C.G., and Wallis, R.C. 1977. Ecology of western equine encephalomyelitis in the eastern United States. Adv. Virus Res. 21: 37-83

Hodgman,T.C., and Zimmern, D. 1987. Evolution of RNA viruses. In: RNA Genetics, Holland, J.J., Domingo, E., and Ahlquist, P., (Eds.) CRC Press, Boca Raton, in press.

Jackson, A.C., Moench, T.R., Griffin, D.E. and Johnson, R.T. 1987. The pathogenesis of spinal cord involvement in the encephalomyelitis of mice caused by neuroadapted Sindbis virus infection. Lab. Invest. 56: 418-423.

Jackson, A.C., Moench, T.R., Trapp, B.D., and Griffin, D.E. 1988. The basis of neurovirulence in Sindbis virus encephalomyelitis of mice. Lab. Invest., in press

Johnson, R.T. 1965. Virus invasion of the central nervous system: A study of the Sindbis virus infection in the mouse using fluorescent antibody. Am. J. Pathol. 46: 929-943

Johnson, R.T., McFarland, H.F., and Levy, S.E. 1972 Age-dependent resistance to viral encephalitis: Studies of infections due to Sindbis virus in mice. J. Infec. Dis. 125: 257-262

Kamer, G., and Argos, P., 1984. Primary structural comparison of RNA-dependent polymerases from plant, animal, and bacterial viruses. Nucleic Acids Res. 12: 7269-7282.

Karabatsos, N. 1975. Antigenic relationships of Group A arboviruses by plaque reduction neutralization testing. Am. J. Trop. Med. Hyg. 24: 527-532.

Kew, O.M., and Nottay B.K., 1984. Evolution of the oral polio vaccine strains in humans occurs by both mutation and intramolecular recombination. In: Modern Approaches to Vaccines: Molecular and Chemical Basis of Virus Virulence, Chanock, R.M., (Ed.) Cold Spring Harbor Laboratories, N. Y. pp.357-362.

Kirkegaard, K., and Baltimore, D. 1986. The Mechanism of RNA recombination in poliovirus. Cell 47: 433-443.

Lindqvist, B.H., DiSalvo, J., Rice, C.M., Strauss, J.H., and Strauss, E.G. 1986. Sindbis virus mutant ts20 of complementation group E contains a lesion in glycoprotein E2. Virology 151: 10-20.

Lustig, S., Jackson, A.C., Hahn, C.S., Griffin, D.E., Strauss, E.G., and Strauss, J. H. 1988. Molecular basis of Sindbis virus neurovirulence in mice. J. Virol. in press

Makino, S., Keck, J.G., Stohlman, S.T., and Lai, M.M.C. 1986. High-frequency RNA recombination of murine coronaviruses. J. Virol. 57: 729-737.

Matthews, R.E.F., 1982. Classification and nomenclature of viruses. Intervirology 17: 1-199.

Maxam, A.M., and Gilbert, W. 1980. Sequencing end-labeled DNA with base specific chemical cleavage. In Methods in Enzymology. (Grossman, L. and Moldave, K. (Eds.) Vol. 65, Academic Press, Orlando, Fla. pp. 499-560.

Melancon, P., and Garoff, H. 1987. Processing of the Semliki Forest virus structural proteins: Role of the capsid protease. J. Virol. 61: 1301-1309.

Niklasson, B., Espmark, A., LeDuc, J.W., Gargan, T.P., Ennis, W.A., Tesh, R.B., and Main Jr., A.J. 1984. Association of Sindbis-like virus with Ockelbo disease in Sweden. Am. J. Trop. Med. Hyg. 33: 1212-1217.

Olmsted, R.A., Baric, R.S., Sawyer, B.A., and Johnston, R.E. 1984. Sindbis virus mutants selected for rapid growth in cell culture display attenuated virulence in animals. Science 225: 424-427.

Olmsted, R.A., Meyer, W.J. and Johnston, R.E. 1986. Characterization of Sindbis virus epitopes important for penetration in cell culture and pathogenesis in animals. Virology 148: 245-254.

Ou, J.-H., Rice, C.M., Dalgarno, L., Strauss, E.G., and Strauss, J.H. 1982a. Sequence studies of several alphavirus genomic RNAs in the region encoding the start of the subgenomic RNA. Proc. Nat. Acad. Sci. USA 79: 5235-5239.

Ou, J.H., Trent, D.W., and Strauss, J.H. 1982b. The 3' noncoding regions of alphavirus RNAs contain repeating sequences. J. Mol. Biol. 156: 719-730.

Ou, J.-H., Strauss, E.G., and Strauss, J.H. 1983. The 5'-terminal sequences of the genomic RNAs of several alphaviruses. J. Mol. Biol. 168: 1-15.

Polo, J.M., Davis, N.L., Rice, C.M., Huang, H.V., and Johnston, R.E. 1988. Molecular analysis of Sindbis virus pathogenesis in neonatal mice using virus recombinants constructed in vitro J. Virol., in press.

Porterfield, J.S. 1980. Antigenic characteristics and classification of togaviridae. In: The Togaviruses, Schlesinger, R.W. (Ed.), Academic Press, New York, Chapter 2, pp. 13-46.

Porterfield, J.S. 1961. Cross-neutralization studies with Group A arthropod-borne viruses. Bull. W.H.O. 24: 735-741.

Reed, W. 1901. Propagation of yellow fever; observations based on recent researches. Med. Rec. 60: 201-209.

Rice, C.M., Bell, J.R., Hunkapiller, M.W., Strauss, E.G., and Strauss, J.H. 1982. Isolation and characterization of the hydrophobic COOH-terminal domains of the Sindbis virus glycoproteins. J. Mol. Biol. 154: 355-378.

Rice, C.M., Lenches, E.M., Eddy, S.R., Shin, S.J., Sheets, R.L., and Strauss, J.H., 1985. Nucleotide sequence of yellow fever virus: Implications for flavivirus gene expression and evolution. Science **229:** 726-733.

Rice. C.M., Levis, R., Strauss, J.H., and Huang, H.V. 1987. Production of infectious RNA transcripts from Sindbis virus cDNA clones: Mapping of lethal mutations, rescue of a temperature-sensitive marker, and in vitro mutagenesis to generate defined mutants. J. Virol. 61: 3809-3819.

Rice, C.M., Strauss, E.G., and Strauss, J.H. 1986. Structure of the flavivirus genome. In: The Togaviridae and Flaviviridae, Schlesinger. S. and Schlesinger, M.J. (Eds.), Plenum Publishing Corp., New York, Chapter 10, pp. 279-326.

Rice, C.M., and Strauss, J.H. 1981a. Nucleotide sequence of the 26S mRNA of Sindbis virus and deduced sequence of the encoded virus structural proteins. Proc. Nat. Acad. Sci. USA. 78: 2062-2066.

Rice, C.M., and Strauss, J.H. 1981b. Synthesis, cleavage, and sequence analysis of DNA complementary to the 26S messenger RNA of Sindbis virus J. Mol. Biol. 150: 315-340.

Rice, C.M., and Strauss, J.H. 1982. Association of Sindbis virion glycoproteins and their precursors. J. Mol. Biol. 154: 325-348.

Sanger F., Nicklen S., and Coulson A. R. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA, 74: 5463-5468

Schlesinger, M.J., and Kaariainen, L. 1980. Translation and processing of alphavirus proteins. In: The Togaviruses, Schlesinger, W. (Ed.), Chapter 13, Academic Press, New York, pp371-392.

Scupham, R.K., Jones, K.J., Sagik, J.B. and Bose, H.R. 1977. Virus-directed posttranslational cleavage in Sindbis virus-infected cells. J. Virol. 22: 568-571.

Sefton, B.M. 1977. Immediate glycosylation of Sindbis virus membrane proteins. Cell 10: 659-668.

Simmons, D.T., and Strauss, J.H., 1972. Replication of Sindbis virus. I. Relative size and genetic content of 26S and 49S RNA. J. Mol. Biol. 71: 599-613.

Simmons, D.T. and Strauss, J.H. 1974. Translation of Sindbis virus 26S RNA and 49S RNA in lysates of rabbit reticulocytes. J. Mol. Biol. 86: 397-409.

Stanley, J., Cooper, S.J., and Griffin, D.E. 1985. Alphavirus neurovirulence: Monoclonal antibodies discriminating wild-type from neuroadapted Sindbis virus. J. Virol. 56: 110-119.

Stanley, J., Cooper, S.J., and Griffin, D.E. 1986. Monoclonal antibody cure and prophylaxis of lethal Sindbis virus encephalitis in mice J. Virol. 58: 107-115.

Stokes, A., Bauer, J.H., and Hudson, N.P. 1928. Experimental transmission of yellow fever to laboratory animals. Am. J. Trop. Med. 8: 103.

Strode, G.K. 1951. Yellow Fever. McGraw-Hill, New York.

Strauss, E.G., Birdwell, C.R., Lenches, E.M., Staples, S.E., and Strauss, J.H. 1977. Mutants of Sindbis virus II. Characterization of a maturation-defective mutant, ts103. Virology 82: 122-149.

Strauss, E.G., Lenches, E.M., and Strauss, J.H. 1976. Mutants of Sindbis virus I: Isolation and partial characterization of 89 new temperature-sensitive mutants. Virology 74: 154-168.

Strauss, E.G., Levinson, R., Rice, C.M., Dalrymple, J.M., and Strauss, J.H., 1988. Nonstructural proteins nsP3 and nsP4 of Ross River and O'Nyong-nyong viruses: Sequence and comparison with those of other alphaviruses. Virology. in press.

Strauss, E.G., Rice, C.M., and Strauss, J.H., 1983. Sequence coding for the alphavirus nonstructural proteins is interrupted by an opal termination codon. Proc. Nat. Acad. Sci. USA 80: 5271-5275.

Strauss, E.G., Rice, C.M., and Strauss, J.H. 1984. Complete nucleotide sequence of the genomic RNA of Sindbis virus. Virology 133: 92-110

Strauss, E.G., and Strauss, J.H., 1980. Mutants of alphaviruses: Genetics and physiology. In The Togaviruses, Schlesinger, W. (Ed.), Chapter 13, Academic Press, New York, pp. 393-426.

Strauss, E.G. and Strauss, J.H., 1983. Replication strategies of the single-stranded RNA viruses of eukaryotes. Curr. Topics. Micro. Immun. 105: 1-98.

Strauss, E.G. and Strauss, J.H. 1985. Assembly of enveloped animal viruses. In: Virus Structure and Assembly, Casjens, S. (Ed.), Chapter 6, Jones and Bartlett, Boston, pp. 205-234.

Strauss, E.G., and Strauss, J.H. 1986. Structure and replication of the alphavirus genome. In: The Togaviridae and Flaviviridae, Schlesinger, S. and Schlesinger, M.J. (Eds.), Plenum Publishing Corp., New York, Chapter 3, pp. 35-90.

Strauss, J.H. and Strauss, E.G. 1985. Antigenic structure of togaviruses. In: The Immunochemistry of Viruses: the Basis for Serodiagnosis and Vaccines. van Regenmortel, M.H.V., and Neurath, A.R. (Eds.), Chapter 22, Elsevier Amsterdam, pp. 407-424.

Strauss, J.H. and Strauss, E.G. 1988. Evolution of RNA viruses. Ann. Rev. Microbiol. 42, in press.

Strauss, J.H., and Strauss, E.G. 1977. Togaviruses. In: The Molecular Biology of Animal Viruses, Vol. I, Nayak, D.P. (Ed.), Marcel Dekker, Inc., New York, pp. 111-166.

Strauss, J.H., Strauss, E.G., Hahn, C.S., Hahn, Y.S., Galler, R., Hardy, W.R., and Rice, C.M. 1987a. Replication of Alphaviruses and Flaviviruses: Proteolytic processing of polyproteins. In: Positive Strand RNA Viruses. Brinton, M. A. and Reuckert, R. (Eds.) UCLA Symposia in Molecular and Cellular Biology, Vol. 54, Alan R. Liss, Inc., New York, pp. 209-225.

Strauss, J.H., Strauss, E.G., Hahn, C.S., and Rice, C.M. 1987b. The genome of alphaviruses and flaviviruses: Organization and translation. In: The Molecular Biology of the Positive Strand RNA Viruses, Rowlands, D.J., Mayo, M.A., and Mahy, B.W.J. (Eds.) Academic Press, London, pp. 75-104.

Takegami, T., Washizu, M., and Yasui, K. 1986. Nucleotide sequence at the 3' end of Japanese encephalitis virus genomic RNA. Virology 152: 483-486.

Theiler, M. 1930. Susceptability of white mice to virus of yellow fever. Science 71: 367.

Theiler, M. and Smith, H.H. 1937a. Effect of prolonged cultivation in vitro upon pathogenicity of yellow fever virus. J. Exp. Med. 65: 767-786.

Theiler, M. and Smith, H.H. 1937b. Use of yellow fever virus modified by in vitro cultivation for human immunization. J. Exp. Med. 65: 787-800.

Webster, R.G., Laver, W.G., Air, G.M., and Schild, G.C. 1982. Molecular mechanisms of variation in influenza viruses. Nature 296: 115-121.

Wengler, G., and Castle, E., 1986. Analysis of structural properties which possibly are characteristic for the 3'-terminal sequence of the genome RNA of flaviviruses. J. Gen. Virol. **67:** 1183-1188.

Westaway, E.G., Brinton, M.A., Gaidamovich, S.Ya., Horzinek, M.C., Igarashi, A., Kaariainen, L., Lvov, D.K., Poterfield, J.S., Russell, P.K., and Trent, D.W. 1985a. Togaviridae. Intervirology 24: 125-139.

Westaway, E.G., Brinton, M.A., Gaidamovich, S.Ya., Horzinek, M.C., Igarashi, A., Kaariainen, L., Lvov, D.K., Poterfield, J.S., Russell, P.K., and Trent, D.W. 1985b. Flaviviridae. Intervirology 24: 183-192.

Zhao, R., Mackow, E., Buckler-White, A., Markoff, L., Chanock, R. M., Lai, C.-J., and Makino, Y. 1986. Cloning full-length dengue type 4 viral DNA sequences: Analysis of genes coding for structural proteins. Virology 155: 77-88.

Zimmern, D., and Kaesberg, P. 1978. 3' terminal sequence of encephalomyocarditis virus RNA determined by reverse transcriptase and chain-terminating inhibitors. Proc. Nat. Acad Sci. USA 75: 4257-4261.

# Part I

## The Role of the Alphavirus Capsid Autoprotease in the Processing of Alphavirus Structural Proteins

# Chapter 1

## Sequence Analysis of Three Sindbis Virus Mutants Temerature-sensitive in the Capsid Protein Autoprotease

This chapter was published in *Proc. Natl. Acad. Sci. USA*

*Proc. Natl. Acad. Sci. USA*
Vol. 82, pp. 4648–4652, July 1985
Biochemistry

# Sequence analysis of three Sindbis virus mutants temperature-sensitive in the capsid protein autoprotease

(alphavirus/serine protease/RNA virus evolution)

CHANG S. HAHN, ELLEN G. STRAUSS, AND JAMES H. STRAUSS

Division of Biology, California Institute of Technology, Pasadena, CA 91125

*Communicated by Giuseppe Attardi, March 25, 1985*

**ABSTRACT** **We have cloned and sequenced the cDNA made to the region of RNA encoding the structural proteins of three complementation group C mutants of Sindbis virus, *ts*2, *ts*5, and *ts*13, and of their revertants. These mutants possess defects in the posttranslational processing of their structural proteins at the nonpermissive temperature. Comparison of the deduced amino acid sequences of the mutants with those of the revertants and with the parental HR strain of virus showed all three mutants to have single amino acid substitutions in the highly conserved COOH-terminal half of the capsid protein that give rise to temperature sensitivity. *ts*2 and *ts*5 were found to have the same lesion and thus represent independent isolations of the same mutant, whereas *ts*13 possessed a different change. Reversion to temperature insensitivity in all three mutants occurred by reversion of the mutated nucleotide to the parental nucleotide, restoring the original amino acid. It has been previously postulated that the capsid protein possesses an autoproteolytic activity that cleaves the capsid protein from the nascent polyprotein during translation. Comparison of the amino acid sequence of the capsid protein with that of serine proteases leads us to hypothesize that histidine-141, aspartate-147, and serine-215 of the Sindbis capsid protein form the catalytic triad of a serine protease. This hypothesis is supported by the finding that all three temperature-sensitive lesions mapped occur near these residues: *ts*2 and *ts*5 change proline-218 to serine and in *ts*13 lysine-138 has been replaced by isoleucine.**

Sindbis virus is a small enveloped RNA virus that belongs to the genus Alphavirus of the family Togaviridae. It consists of a nucleocapsid containing the single-stranded virus RNA complexed with a basic capsid protein, surrounded by a lipid bilayer containing the virus-encoded integral membrane glycoproteins E1 and E2 (1).

The viral structural proteins are translated as a single large precursor polypeptide from a subgenomic messenger, 26S RNA, that is the 3′ terminal one-third of the genomic 49S RNA (2–6). In this precursor the $NH_2$-terminal capsid protein is followed by the envelope protein precursor PE2 and the second envelope protein, E1. Proteolytic cleavages occur several times during processing of this polyprotein precursor. The first cleavage releases the capsid protein from the polyprotein, and the remaining portion of the nascent polypeptide is then inserted into the rough endoplasmic reticulum of the cell. During their insertion, or shortly after, the glycoproteins are glycosylated with mannose-rich oligosaccharides and cleaved into the proteins PE2 and E1 (7, 8). A final proteolytic cleavage converts the PE2 glycoprotein to proteins E2 and E3. A failure in any of these cleavage and/or insertional steps inhibits virion formation and abnormal polypeptide precursors accumulate in the infected cell.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

There are three complementation groups of temperature-sensitive (*ts*) mutants of Sindbis virus that affect the structural proteins of the virus (9). Mutants in groups D and E are located in glycoproteins E1 (10) and E2 (unpublished data), respectively. Mutants in group C are thought to be in the capsid protein since at the nonpermissive temperature the polypeptide precursor is not cleaved and accumulates in the cytoplasm (11). No capsid protein is formed and little envelope protein is produced. It is thought that failure to remove the capsid protein causes the signal sequences for insertion of the glycoproteins to fail to function; the result is that the glycoprotein portions are not inserted or glycosylated or further processed, and the complete polyprotein results.

Several investigators have suggested that the cleavage of the capsid protein from the polyprotein precursor could occur by autoproteolytic activity of the capsid protein itself (3, 12, 13). In this model autoproteolysis ordinarily results from the nascent polypeptide acting upon itself, rather than from a diffusible protease, because proteolysis is rapid and complete even during *in vitro* translation where very small amounts of products are made (3).

The complete nucleotide sequences of Sindbis virus genomic RNA and of 26S messenger RNA have been published (14, 15). This has made possible the analysis of *ts* mutants of Sindbis virus (10). Here we report the cloning of cDNA from three group C mutants, *ts*2, *ts*5, and *ts*13, and of their revertants and the determination of the nucleotide sequences of the regions of 26S RNA that encode the capsid proteins and parts of the glycoproteins. Mapping of these *ts* mutations together with comparison of the amino acid sequences of alphavirus capsid proteins with those of serine proteases leads us to postulate that the alphavirus capsid protein is a serine protease whose catalytic triad is formed by histidine-141, aspartate-147, and serine-215 in Sindbis virus.

## MATERIALS AND METHODS

**Virus Strains.** The mutants *ts*2, *ts*5, and *ts*13 isolated from the heat-resistant HR strain were kindly provided by B. W. Burge. We used the oldest stocks in our possession, which had not been passed since 1971, as seed stocks in this work. All strains were plaque purified at 30°C immediately before use.

**Isolation of Revertants.** Mutant stocks were assayed for revertants by plaque titration at 40°C and 30°C, and the reversion frequency is expressed as the ratio of plaques at 40°C to plaques at 30°C. Single virus clones of $ts^+$ revertants were picked from the 40°C plate and the virus was eluted into 1 ml of Eagle's medium/5% fetal calf serum (16). These revertant plaque picks were used to infect a Petri plate of primary chicken embryo fibroblast cells at 40°C and the resulting stocks were plaque assayed at 30°C and 40°C and used as the infecting stocks for RNA preparation.

**Virus Purification and RNA Isolation.** Viruses were grown in primary or secondary chicken embryo fibroblast cells and harvested 10–20 hr after infection, depending on the mutant.

All mutants and revertants were incubated at 30°C or 40°C, respectively. Viruses were purified and RNA was isolated as described (17).

**Cloning and Sequencing of DNA.** The methods used for obtaining clones of cDNA made to the entire structural protein region will be described elsewhere (unpublished data). Briefly, a plasmid vector originally derived from pBR322 was cut with *Sma* I, thymidine-tailed with terminal nucleotidyltransferase, and then cut with *Sal* I to remove one of the tails. The resulting tailed vector was used as a primer with 49S RNA from purified virions to synthesize double-stranded cDNA. Conditions for first-strand synthesis were essentially as described by Strauss *et al.* (14) and those for second-strand synthesis were as described by Okayama and Berg (18). The resulting product was cut with *Hin*dIII, ligated with *Escherichia coli* ligase, and cloned. Most of the plasmids that contain inserts have a 5.5-kilobase insert containing cDNA from the 3′ terminal poly(A) tract of 49S RNA to the *Hin*dIII site at nucleotides 6266–6271 in the genomic RNA and thus contain the entire structural protein region. Plasmid DNA was isolated from these clones and sequenced by the chemical sequencing methods of Maxam and Gilbert (19), as modified by Smith and Calvo (20), as described (14).

## RESULTS

**Sequencing of Complementation Group C Mutants.** Clones containing the entire structural protein region of three group C mutants, *ts*2, *ts*5, and *ts*13, and of revertants of these mutants were obtained by using a poly(dT)-tailed vector as a primer for cDNA synthesis with purified virion RNA. The cloned cDNA extended from the 3′ terminal poly(A) tract to the *Hin*dIII site at position 6267 in the genomic RNA sequence. DNA from these clones was sequenced, using the methods of chemical sequencing of Maxam and Gilbert (19), from the first nucleotide of 26S RNA (nucleotide 7598 of the genomic RNA) to nucleotide 1237 of 26S RNA (nucleotide 8835 of the genome). This region encompasses the 5′ untranslated region of 26S RNA, the entire capsid protein region, all of the E3 protein, and the first 68 amino acids of the E2 protein. In each case at least three independent clones were sequenced to ensure that any changes found were characteristic of the RNA population and did not represent minor variants in the population or reverse transcriptase errors. This proved to be necessary in some cases because revertants in the population or other minor variants were sequenced. The results presented in Fig. 1 in all cases represent at least three clones that agree as to sequence.

**Sequence Analysis of *ts*2 and Its Revertant.** *ts*2 was produced by nitrous acid mutagenesis of the HR strain of Sindbis virus (21, 22). The sequences obtained for this mutant and its revertant are illustrated in Fig. 1. Comparing the *ts*2 sequence with that of the parental HR strain (14), we found three nucleotide substitutions in *ts*2 within the region sequenced. Two of these changes do not revert in the revertant and presumably have nothing to do with temperature sensitivity, probably arising during the original mutagenesis even though the changes observed are not those expected for nitrous acid. The first of these nonreverting changes is a U→C substitution at nucleotide 67, which is a silent change in the codon for phenylalanine-6 (UUU→UUC). The second nonreverting change observed is a G→U change at nucleotide 1043, which leads to a change of aspartate-4 to tyrosine (GAC→UAC) in glycoprotein E2.

We conclude that the mutation responsible for temperature sensitivity in *ts*2 is a change of C→U at nucleotide 701, which leads to the replacement of proline-218 by serine (CCG→UCG). In the revertant the changed nucleotide reverts to the original nucleotide, restoring the parental amino acid. It is of interest that this is the only change found in the sequenced region that has the expected characteristics of nitrous acid mutagenesis.

**Sequence Analysis of *ts*5 and Its Revertant.** *ts*5 was produced by ethyl methanesulfonate mutagenesis of the HR strain (21, 22). We found that this mutant had only one change throughout the sequenced region, a C→U change at nucle-

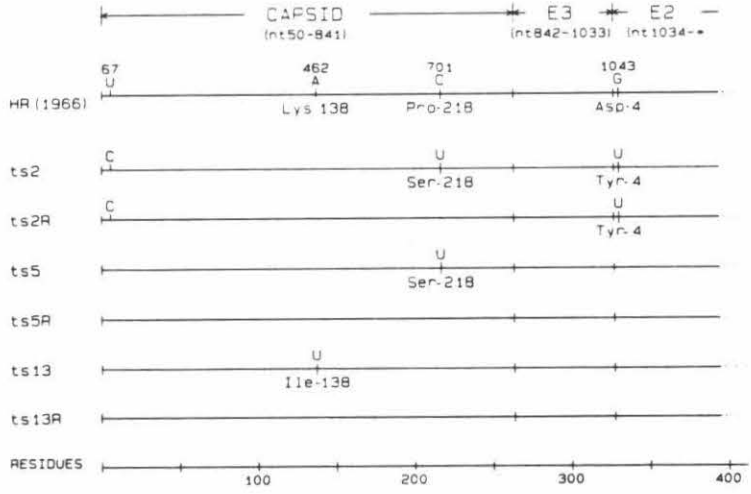



FIG. 1. Schematic representation of the nucleotide sequence of the first one-third of 26S RNA, which encodes the capsid protein, the E3 protein, and part of the E2 glycoprotein, for several strains of Sindbis virus. The top line represents the sequence of the ancestral HR strain (10). The remaining lines represent *ts*2 (derived from HR by mutagenesis with $HNO_2$) and its revertant *ts*2R; *ts*5 (derived from HR by mutagenesis with ethyl methanesulfonate) and its revertant *ts*5R; and *ts*13 (derived from HR by nitrosoguanidine mutagenesis) and its revertant *ts*13R. Nucleotides are numbered from the 5′ end of 26S RNA (which begins at nucleotide 7598 in the 49S RNA). Amino acids are numbered from the $NH_2$ terminus of each protein. All nucleotides or amino acids changed from the ancestral sequence in any of the strains shown are indicated in the representation of the ancestral sequence. In the other strains shown, any change from the ancestral sequence is indicated; if no change is shown, then the sequence is the same as the ancestral sequence. nt, Nucleotide.

otide 701 of 26S RNA that leads to replacement of proline-218 of the capsid protein by serine (CCG→UCG) (Fig. 1). In the *ts5* revertant this change reverts to the original nucleotide (Fig. 1). Thus despite the fact that different mutagens were used to produce them, *ts2* and *ts5* represent independent isolations of the same mutation. Their independent history is confirmed by the fact that *ts5* lacks the two extraneous changes found in *ts2*.

**Sequence Analysis of *ts13* and Its Revertant.** *ts13* was produced by nitrosoguanidine mutagenesis of the HR strain (21, 22). *ts13* also has a single base substitution in the region sequenced. The change is A→U at nucleotide 462, resulting in the change of lysine-138 to isoleucine (AAA→AUA) in the capsid protein. In the *ts13* revertant, this nucleotide reverts to the original nucleotide (Fig. 1).

## DISCUSSION

**Nature of the Mutational Events.** The three temperature-sensitive mutants of the C complementation group isolated by Burge and Pfefferkorn (21, 22) have been widely studied. All mutants in this group studied to date accumulate the large polypeptide precursor called the *ts2* protein or NVP130 at the nonpermissive temperature (9, 11, 23). cDNA sequence analysis of the structural protein region of these three mutants shows that amino acid substitutions in the COOH-terminal half of the capsid protein cause this *ts2* protein accumulation. It has long been thought that group C mutants are defective in the capsid protein, but no firm evidence for this has existed.

In each case, reversion to temperature insensitivity occurred by same-site reversion, which is consistent with our work on *ts* mutants in proteins E1 (10) and E2 (unpublished data). This has greatly simplified the task of identifying the changes responsible for temperature sensitivity. It is not completely clear why second-site pseudorevertants have not been seen. Either they do not occur or they do not accumulate in the virus population. In most cases the *ts* mutant grows more slowly than the wild type even at the permissive temperature, the wild type being highly evolved for rapid growth. Once true revertants arise they are quickly amplified in the population, and it is possible that second-site pseudorevertants, even if they arose, might not possess such a selective advantage and might not be amplified in the virus population.

All three mutants were reported by Burge and Pfefferkorn (21) to have a reversion frequency of $0.5–1 \times 10^{-5}$. We found the same reversion frequency for *ts2* and *ts5* but a considerably higher reversion frequency, $4 \times 10^{-3}$, for *ts13*, which suggests that the *ts13* revertant may have arisen early in this experiment. In any event the reversion frequency of all three mutants is consistent with their being single mutants as determined from nucleotide sequencing.

As we had previously observed with Sindbis *ts* mutants (10), the changes found in this study were often not those most commonly produced by the mutagens used (24). In the case of *ts5*, which was derived by ethyl methanesulfonate mutagenesis, the C→U transition found is one of the changes expected from the action of the alkylating agent. For *ts2*, which was derived by $HNO_2$ mutagenesis, the major changes expected are C→U or A→G transitions produced by deamination. The mutational event giving rise to temperature sensitivity was a C→U transition, but a U→C transition and a G→U transversion were also found, although in the last two cases it cannot be shown that the changes occurred during $HNO_2$ mutagenesis. Finally, for *ts13*, which was derived by nitrosoguanidine mutagenesis, the A→U change found is not one predicted for this agent, but this agent has previously been found to cause many different changes (10).

**Autoproteolysis by the Capsid Protein.** Several investigators have proposed that the formation of capsid protein could be accomplished by autoproteolysis. This was first suggested by the finding that cleavage occurred normally during *in vitro* translation in lysates of rabbit reticulocytes (3). It has also been found that group C mutants can be complemented by other mutants (22) and that during complementation the *ts2* polyprotein is apparently cleaved by a diffusible factor (12). Finally, during *in vitro* translation with amino acid analogues the cleavage of capsid protein from the polyprotein is inhibited (13). The *ts* lesions at lysine-138 and proline-218 of the capsid protein, which result in failure of the cleavage to occur at the nonpermissive temperature, provide further evidence for autoproteolysis by the capsid protein and presumably implicate these regions in proteolysis. *In vitro* translation experiments (25) have shown that cleavage occurs when fewer than 100 amino acids of the PE2 protein have been translated, indicating that translation of the E2 protein is not required for proteolysis. The fact that aspartate-4 of E2 can be replaced by another amino acid without apparent effect on the proteolytic activity (in the *ts* mutant studied here) and aspartate-5 of E2 can similarly be replaced (our large-plaque strain of HR has aspartate-5 replaced by glycine) (10) is consistent with this conclusion. Whether sequences within the E3 region are involved in any way in the proteolysis step is still open.

Cleavage of the capsid protein occurs after a tryptophan residue, tryptophan-264 of the polyprotein precursor, and is therefore a chymotryptic-like cleavage. Furthermore, Pfefferkorn and Boyle (26) have reported that the chymotrypsin inhibitor 1-tosylamido-2-phenylethyl chloromethyl ketone inhibits this cleavage in infected cells. Because this cleavage specificity is characteristic of serine proteases, we searched for possible homology between the highly conserved COOH-terminal half of the alphavirus capsid protein and active sites of mammalian and insect serine proteases. The results are given in Fig. 2 and show that a suggestive homology exists between animal serine proteases and alphavirus capsid proteins. The spacings between the components of serine proteases that form the active site and the corresponding homologous regions of alphavirus capsid proteins are different, but the three amino acids that constitute the catalytic triad (histidine-57, aspartate-102, and serine-195) and their surrounding amino acids show some homology with regions in the COOH-terminal half of the capsid proteins. In particular, serine-215 in Sindbis capsid protein is found in the sequence Gly-Asp-Ser-Gly, which is conserved in all the serine proteases, as has been previously noted (33). Although the sequence homology is limited, we suggest that histidine-141, aspartate-147, and serine-215 of the Sindbis capsid protein perform the same function during autoproteolysis as histidine-57, aspartate-102, and serine-195, respectively, of animal serine proteases. Furthermore, all three mutants have amino acid substitutions within the homologous regions between serine proteases and alphavirus capsid proteins around these three key residues. *ts13* has a change of lysine-138 to isoleucine, which is near the possible charge transmitter, histidine-141. *ts2* and *ts5* have a change of proline-218 to serine, which is just next to the possible catalytic serine-215. The sites of these mutations support the hypothesis that the capsid protein possesses proteolytic activity and lend support to the hypothesis that histidine-141, aspartate-147, and serine-215 are active in this proteolysis. The hypothesis that these three residues form the catalytic triad of a serine protease should be testable in a number of ways; in particular the importance of these residues for proteolysis can be directly tested by site-specific mutagenesis.

Fig. 3 shows the percentage homology as a moving average between the capsid proteins and the E3 proteins of Sindbis

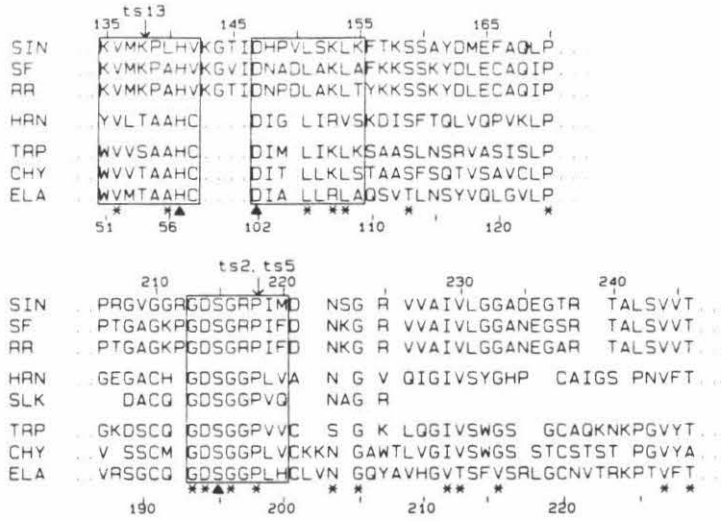



FIG. 2. Homology between the highly conserved COOH-terminal half of the alphavirus capsid protein and mammalian and insect serine protease active sites. Upper three lines, partial amino acid sequences of Sindbis virus (SIN), Semliki Forest virus (SF), and Ross River virus (RR) capsid proteins [sequences are from Rice and Strauss (15), Garoff *et al.* (27), and Dalgarno *et al.* (28), respectively; numbering of amino acids is for the SIN capsid protein]; middle lines, sequences of hornet chymotrypsin (HRN) (29) and silkworm cocoonase (SLK) (30); bottom three lines, active-site domains of three mammalian serine proteases: bovine trypsin (TRP), bovine chymotrypsin (CHY), and porcine elastase (ELA), respectively [the lower numbers are the standard numbering of amino acids for chymotrypsinogen (31, 32)]. Solid triangles indicate the three active amino acids that form the catalytic triad (histidine-57, aspartate-102, and serine-195) and asterisks indicate amino acids that are highly conserved between the capsid proteins and the animal serine proteases; regions of highest homology are boxed. Solid arrows indicate locations of mutations in *ts*2, *ts*5, and *ts*13. The single-letter amino acid code is used.

virus and Ross River virus, which are widely separated alphaviruses (34). These two viruses share up to 95% sequence homology in 20 amino acid strings in the COOH-terminal half of the capsid proteins, and overall the COOH-terminal half of the capsid proteins share 74% amino acid sequence homology. The shaded regions indicate the regions of homology with serine proteases and the solid arrows indicate the sites of mutation in *ts*13 and in *ts*2 and *ts*5, respectively. The very high homology in amino acid sequence together with the fact that two of the capsid protein mutants represent independent isolates of the same mutation suggests that the target window for mutation to temperature sensitivity in the capsid protein might be quite narrow.

Virus-encoded proteases have been described in a number of systems. The best studied example is in picornaviruses (35), in which a protease acts both as an autoprotease, cleaving itself from a nascent polyprotein, and as a diffusible protease, cleaving the precursor of the capsid proteins. Proteases have also been described in retroviruses (36) and in adenoviruses (37), and evidence has been presented that protein VP1 of polyoma virus is a serine protease (38). It is probable that proteases are produced by other virus groups as well and that some viruses produce more than one protease. Organelle-bound cellular proteases such as signalase (which appears to be active only within the lumen of the endoplasmic reticulum) or a Golgi-associated protease that cleaves after double basic amino acids (which appears to be active within vesicles transporting proteins to the cell surface) appear to be involved in processing many virus proteins associated with membranes, such as the Sindbis glycoproteins or other virus glycoproteins (reviewed in ref. 39). However, we have postulated that in general cleavage of viral protein precursors *in the cytosol* requires virus-encoded proteases (15), since no compelling evidence for cellular

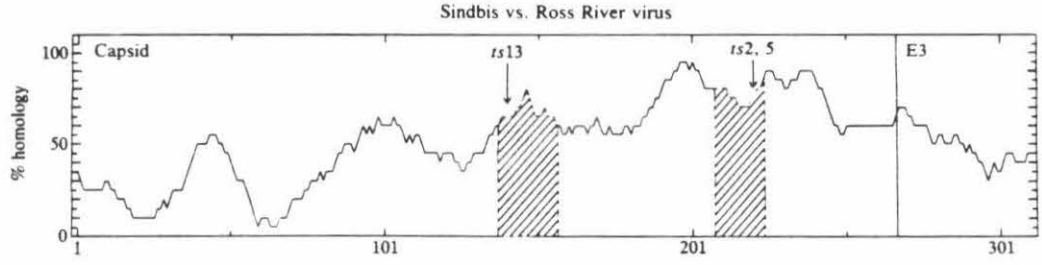



FIG. 3. Homology between the capsid proteins and the E3 proteins of Sindbis virus and Ross River virus plotted as a moving average with a string length of 20. The first shaded region is the possible serine protease active site between lysine-135 and lysine-155, which contains the histidine-141 and aspartate-147 residues (Fig. 2). The second shaded region is the part of the possible serine protease active site between proline-206 and asparagine-222, which contains the $Gly^{213}$-Asp-Ser-$Gly^{216}$ conserved sequence. Arrows indicate the locations of the mutations in *ts*2, *ts*5, and *ts*13.

proteases that might perform such a function has been reported, whereas several virus-encoded proteases active in the cytosol have been described, as noted above. It seems intuitively unlikely that so many viruses would have evolved their own protease activities if cellular proteases were readily available to perform such functions, although more evidence is clearly needed to resolve this issue. Finally, the question arises as to the evolutionary origin of these viral enzymatic activities. The similarities in the capsid protease and the animal serine proteases reported here suggest that these proteases may share a common ancestral origin, although convergent evolution cannot be ruled out in view of the limited extent of homology found. It has previously been found that alphaviruses and certain plant viruses appear to be evolutionarily related (40, 41). Since these plant viruses appear to lack proteases, one possibility is that a proto-alphavirus acquired the protease activity after diverging from the plant virus line and that the enzyme could have been captured from the host cell.

We thank E. Lenches for expert technical assistance in preparation of cells and of virus stocks and are grateful to Dr. Charles Rice for helpful discussions. This work was supported by Grants AI 10793 and AI 20612 from the National Institutes of Health and by Grant DMB 8316856 from the National Science Foundation. C.S.H. was supported by a grant from the California Foundation for Biochemical Research.

1. Strauss, J. H. & Strauss, E. G. (1977) in *The Molecular Biology of Animal Viruses*, ed. Nayak, D. P. (Marcel Dekker, New York), Vol. 1, pp. 111–166.
2. Simmons, D. T. & Strauss, J. H. (1972) *J. Mol. Biol.* **71**, 599–613.
3. Simmons, D. T. & Strauss, J. H. (1974) *J. Mol. Biol.* **86**, 397–409.
4. Clegg, J. C. S. (1975) *Nature (London)* **254**, 454–455.
5. Kennedy, S. I. T. (1976) *J. Mol. Biol.* **108**, 491–511.
6. Wengler, G. & Wengler, A. (1976) *Virology* **73**, 190–199.
7. Katz, F. N., Rothman, J. E., Lingappa, V. R., Blobel, G. & Lodish, H. F. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 3278–3282.
8. Sefton, B. M. (1977) *Cell* **10**, 659–668.
9. Strauss, E. G. & Strauss, J. H. (1980) in *The Togaviruses*, ed. Schlesinger, R. W. (Academic, New York), pp. 393–426.
10. Arias, C., Bell, J. R., Lenches, E. M., Strauss, E. G. & Strauss, J. H. (1983) *J. Mol. Biol.* **168**, 87–102.
11. Scheele, C. M. & Pfefferkorn, E. R. (1970) *J. Virol.* **5**, 329–337.
12. Scupham, R. K., Jones, K. J., Sagik, J. B. & Bose, H. R. (1977) *J. Virol.* **22**, 568–571.
13. Aliperti, G. & Schlesinger, M. (1978) *Virology* **90**, 366–369.
14. Strauss, E. G., Rice, C. M. & Strauss, J. H. (1984) *Virology* **133**, 92–110.
15. Rice, C. M. & Strauss, J. H. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 2062–2066.
16. Strauss, E. G., Lenches, E. M. & Strauss, J. H. (1976) *Virology* **74**, 154–168.
17. Ou, J.-H. & Strauss, J. H. (1981) *Virology* **109**, 281–289.
18. Okayama, H. & Berg, P. (1982) *Mol. Cell. Biol.* **2**, 161–170.
19. Maxam, A. M. & Gilbert, W. (1977) *Methods Enzymol.* **65**, 499–560.
20. Smith, D. R. & Calvo, J. M. (1980) *Nucleic Acids Res.* **8**, 2255–2274.
21. Burge, B. W. & Pfefferkorn, E. R. (1966) *Virology* **30**, 204–213.
22. Burge, B. W. & Pfefferkorn, E. R. (1966) *Virology* **30**, 214–223.
23. Strauss, J. H., Burge, B. W. & Darnell, J. E. (1969) *Virology* **37**, 367–376.
24. Drake, G. R. (1970) *The Molecular Basis of Mutation* (Holden-Day, San Francisco), pp. 146–159.
25. Garoff, H., Simons, K. & Dobberstein, B. (1978) *J. Mol. Biol.* **124**, 587–600.
26. Pfefferkorn, E. R. & Boyle, M. K. (1972) *J. Virol.* **9**, 187–188.
27. Garoff, H., Frischauf, A. M., Simons, K., Lehrach, H. & Delius, H. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 6376–6380.
28. Dalgarno, L., Rice, C. M. & Strauss, J. H. (1983) *Virology* **129**, 170–187.
29. Jany, K. D., Bekelar, K., Pfleiderer, G. & Ishay, J. (1983) *Biochem. Biophys. Res. Commun.* **110**, 1–7.
30. Kramer, K. J., Felsted, R. L. & Law, J. H. (1973) *J. Biol. Chem.* **248**, 3021–3028.
31. Greer, J. (1981) *J. Mol. Biol.* **153**, 1027–1042.
32. Dayhoff, M. O., ed. (1978) *Atlas of Protein Sequence and Structure* (National Biomedical Research Foundation, Washington, DC), Vol. 5, Suppl. 3, pp. 79–83.
33. Boege, U., Wengler, G., Wengler, G. & Wittmann-Liebold, B. (1981) *Virology* **113**, 293–303.
34. Bell, J. R., Kinney, R. M., Trent, D. W., Strauss, E. G. & Strauss, J. H. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 4702–4706.
35. Palmenberg, A. C., Pallansch, M. A. & Rueckert, R. R. (1979) *J. Virol.* **32**, 770–778.
36. Von Der Helm, K. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 911–915.
37. Bhatti, A. R. & Weber, J. (1979) *Virology* **96**, 478–485.
38. Bowen, J. H., Chlumecky, V., D'Obrenan, P. & Colter, J. S. (1984) *Virology* **135**, 551–554.
39. Strauss, E. G. & Strauss, J. H. (1983) *Curr. Top. Microbiol. Immunol.* **105**, 1–98.
40. Haseloff, J., Goelet, P., Zimmern, D., Ahlquist, P., Dasgupta, R. & Kaesberg, P. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 4358–4362.
41. Ahlquist, P., Strauss, E. G., Rice, C. M., Strauss, J. H., Haseloff, J. & Zimmern, D. (1985) *J. Virol.* **53**, 536–542.

# Chapter 2

## Site-Directed Mutagenesis of the Proposed Catalytic Amino Acids of the Sindbis Virus Capsid Protein Autoprotease

Chang S. Hahn and James H. Strauss

This chapter was prepared for submission.

Abstract

The structural proteins of Sindbis virus, consisting of a capsid protein and two envelope glycoproteins, are translated as a polyprotein precursor that is cleaved posttranslationally. The first cleavage event releases the capsid protein from the N-terminus of this polyprotein in a process believed to be autoproteolytic, and we have previously postulated that His-141, Asp-147 and Ser-215 of the capsid protein form a catalytic triad similar to that found in serine proteases. To test this hypothesis each of these amino acids has been replaced with other amino acids using oligonucleotide directed site specific mutagensis. Ser-215 has been changed to Thr, Ile or Cys; His-141 has been changed to Pro; and Asp-147 has been changed to His or Tyr. The effects of these changes were tested by placing each of the mutations into a plasmid containing a cDNA copy of the Sindbis mRNA encoding the structural proteins. RNA was transcribed *in vitro* using T7 RNA polymerase, the RNA translated *in vitro* in a rabbit reticulocyte translation system, and the extent of cleavage of capsid protein was monitored. The change of His-141 to Pro led to complete loss of proteolytic activity. The two changes at Asp-147 had no detectable effect on proteolytic activity in this system, with 100% of the capsid protein being released. The changes at Ser-215 had effects that differed with the mutation induced. The change to Cys resulted in a loss of only 20% of the protease activity in this system (that is 20% of the polyprotein molecules were not cleaved to produce capsid protein). The change to Thr resulted in the loss of 90% of the activity, whereas the change to Ile resulted in total loss of proteolytic activity. In a second test for the effect of these mutations, they were placed in a full-length cDNA copy of Sindbis virus RNA from which infectious RNA can be transcribed *in vitro*. The resulting RNA was found to be nonviable, that is no virus could be recovered after RNA transcription, indicating that all of the changes are lethal. This indicates that either production

of infectious virus is a more sensitive measure of proteolytic activity than the *in vitro* cleavage assay, or that these residues have necessary functions in addition to their role in proteolysis.

## INTRODUCTION

Sindbis virus is a small enveloped RNA virus belonging to the genus Alphavirus of the family Togaviridae. It consists of a nucleocapsid containing the single-stranded viral RNA of positive polarity complexed with a basic capsid protein C, surrounded by a lipid bilayer containing the virus-encoded membrane glycoproteins E1 and E2 (Strauss and Strauss, 1976, 1986). The virion structural proteins are translated as a single, large precursor polyprotein from a subgenomic mRNA, 26S RNA, which forms the 3' terminal one-third of the genomic 49S RNA (Simmons and Strauss, 1972; Clegg, 1975). In this precursor polyprotein the capsid protein is amino terminal and is followed by glycoprotein precursor PE2, a small hydrophobic peptide referred to as 6K, and the second glycoprotein E1 (Garoff *et al.*, 1980a,b; Rice and Strauss, 1981; Strauss *et al.*, 1984). Proteolytic processing occurs several times during the processing of this polyprotein precursor (Schlesinger and Kaariainen, 1980; Strauss *et al.*, 1987a,b). The first cleavage releases the capsid protein from the nascent polyprotein. The remaining cleavages occur during insertion of the protein into the endoplasmic reticulum, subsequent processing, and transport to the cell membrane. These latter events are postulated to be effected by cellular proteases active in subcellular organelles. Under conditions in which the glycoprotein precursor is not inserted into the endoplasmic reticulum, such as situations in which capsid protein cleavage does not occur, or those in which the N-terminal signal sequence fails to function, the glycoprotein portion is not further processed. In these cases either a complete structural polyprotein (p138) or the complete envelope protein precursor

(p107), respectively, accumulates in the cytoplasm (Strauss and Strauss, 1980).

The cleavage of the capsid protein from the precursor has long been postulated to be an autoproteolytic event (Simmons and and Strauss, 1974; Scupham *et al.*, 1977; Aliperti and Schlesinger, 1978; Boege *et al.*, 1981; Hahn *et al.*, 1985; Melancon and Garoff, 1987). We have previously postulated that the capsid autoprotease is a serine autoprotease whose catalytic triad is formed by His-141, Asp-147 and Ser-215 (Hahn *et al.*, 1985), based in part on the location of mutations that cause the capsid protein autoprotease to be nonfunctional at the nonpermissive temperature, and in part upon amino acid sequence similarities between the sequences around these three residues and those surrounding the amino acids of the catalytic triads of mammalian and insect serine proteases. To test this hypothesis, we have developed an *in vitro* system in which to assay the effects of site specific mutations introduced into this catalytic triad upon protease activity. The results are consistent with our original hypothesis that these three residues in the capsid protein do in fact form the catalytic triad of a serine autoprotease.

## MATERIALS AND METHODS

***General recombinant DNA techniques.*** Restriction endonucleases and DNA-modifying enzymes were purchased from commercial sources and used essentially as recommended by the manufacturers. Plasmids and phages were grown, purified and analyzed using standard methods with minor modifications (Maniatis *et al.*, 1982).

***Construction of M13 recombinant phage.*** A recombinant phage M13 containing an *Nco*I fragment of Sindbis virus cDNA (genomic coordinates 8040 to 8867, covering the C-terminal 133 amino acids of the capsid protein and the N-terminal 144 amino acids of PE2) was constructed for use as a site-specific

mutagenesis vehicle. A plasmid containing Sindbis cDNA from the *Hin*dIII site at nt 6267 to the 3' poly (A) termed SIN3' and derived from strain HR*sp* by the method described in Lindqvist *et al.* (1986) and Hahn *et al.* (1985), was digested with *Nco*I, the resulting fragments were flush-ended by reaction with the Klenow fragment of *E. coli* polymerase 1 and dNTPs, and ligated into *Sma*I digested, alkaline phosphatase-treated M13 mp19 RF. This results in the regeneration of the *Nco*I site at both ends of the insert so that this same enzyme can be used to remove the insert from the recombinant phage M13.

***Site-specific mutagenesis.*** Oligonucleotide-directed site specific mutagensis was performed using recombinant phage M13 single-stranded DNA as a template and synthetic oligonucleotides as a primer essentially as described (Shortle *et al.*, 1981; Dalbadie-McFarland *et al.* 1982; Zoller and Smith, 1984). The His-141 to Pro change was effected by using as primer TCCTTTCACGGGCAGAGGTT, the Asp-147 to His or Tyr was effected using as primer ACAGGGTGGT(G,A)GATGGGTTCC, the changes of Ser-215 to Thr or Ile effected with GGACGACCG(G,T)TGTCTCCTCT, and the Ser-215 to Cys change with GGACGACCGCAGTCTCCTCT. Following mutagenesis the resulting virus plaques were screened by plaque lift hybridization with sequentially higher stringency washes, using the same primers as were used in the mutagenesis reactions as a radiolabelled probe. Mutant plaques were sequenced by the dideoxy method using synthetic oligonucleotides as primers (Sanger *et al.*, 1977; Strauss *et al.* 1986).

***Construction of an in vitro* transcription/translation vector.** Plasmid SIN3' was partially digested with *Ava*II and the recessed ends filled in with the Klenow fragment of *E. coli* DNA polymerase I and all four dNTPs. The linearized plasmid was isolated following electrophoresis in a low-melting temperature agarose gel, treated with restriction enzyme *Eco*RI to cut just following the poly(A) tract in

the cDNA insert, and the fragment of 4811 nucleotides isolated by electrophoresis in a low melting temperature agarose gel. This fragment was ligated into vector pGEM4 (Promega Biotech), which had been prepared by digestion with restriction enzyme ***Sal***I, the recessed ends filled in with the Klenow fragment, followed by digestion with *Eco*RI and dephosphorylation with calf intestinal alkaline phosphatase. The resulting plasmid was called T7SVSP and was determined to be correct by both restriction enzyme analysis and by sequencing the regions where the vector and the Sindbis cDNA insert were joined.

***Construction of mutagenized plasmids.*** The 847 nucleotide *Nco*I fragment of T7SVSP was replaced with the corresponding fragments from mutagenized recombinant phage M13 mp19 RF. The orientation of the *Nco*I insert in the substituted T7SVSP was analyzed by restriction enzyme analysis.

***In vitro transcription and translation.*** Large-scale preparations of plasmid T7SVSP and its mutagenized derivatives were purified by isopycnic centrifugation twice in CsCl. The purified plasmids were linearized by digestion with *Eco*RI, which cuts just past the poly(A) tail of the Sindbis cDNA insert, and the linearized plasmids used as templates for run-off transcription with T7 RNA polymerase. The reaction mixture contained 40 mM Tris-Cl (pH 8.0), 8 mM MgCl, 25 mM NaCl, 2 mM spermidine-HCl, 5 mM dithiothreitol, 1 mM each of the four NTPs, and 10-50 μg/ml of linearized DNA template. To increase the efficiency of the translation, a cap analog was present at 1 mM during transcription. The RNA was isolated from the transcription mixture by phenol-chloroform extraction followed by ethanol precipitation, and translated *in vitro* in a rabbit reticulocyte translation system (Promega Biotech) for 1 hr at 30°C as described by the manufacturer, in the presence of [$^{35}$S]methionine. The resulting translated polypeptides were analyzed by electrophoresis in 15% SDS-containing polyacrylamide gels.

***Test of the effects of the mutations on RNA infectivity.*** The 827 nucleotide *Nco*I fragment from mutagenized recombinant M13 mp19 RF was excised by *Nco*I digestion and inserted into *Nco*I-digested SIN3'. The orientation of the insert was checked by restriction anlaysis and plasmids containing an insert in the correct orientation were cleaved with restriction endonucleases *Hpa*I and *Bss*HII, which cleave unique sites in the Sindbis virus genome. The corresponding region of a full-length cDNA clone of Sindbis, Toto1101 (Rice *et al.*, 1987), from which infectious RNA can be transcribed *in vitro* with SP6 RNA polymerase, was replaced with this mutagenized *Hpa*I-*Bss*HII fragment. The resulting clones were checked by dideoxy sequencing for the presence of the predicted mutation. RNA was transcribed *in vitro* and used in transfection assays as previously described (Rice *et al.*, 1987; Lustig *et al.*, 1988).

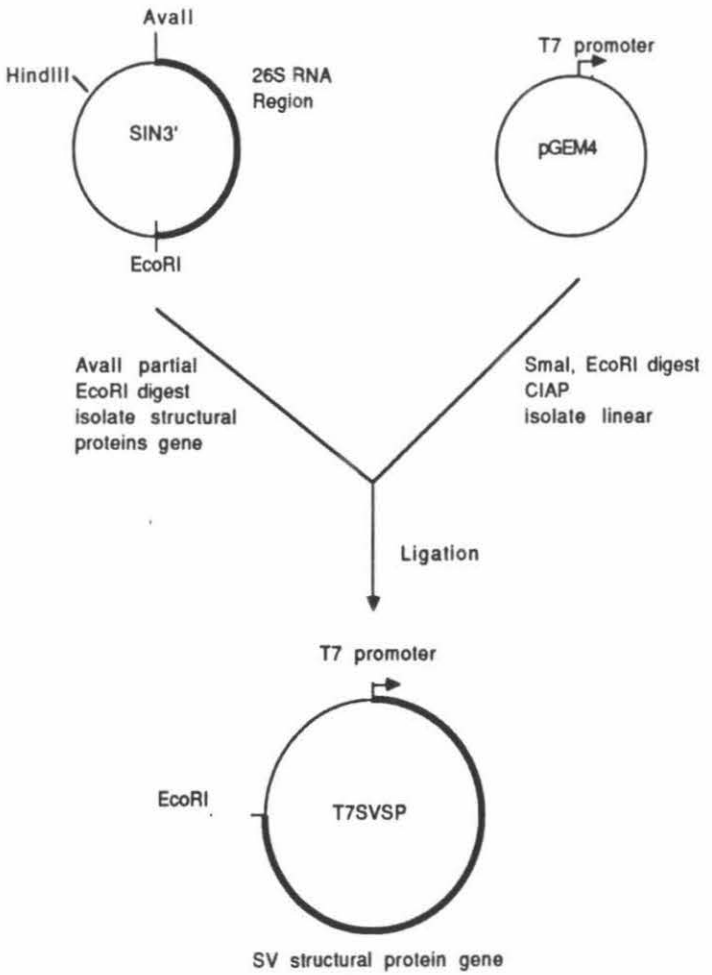
## RESULTS

***Processing of the structural proteins in vitro.*** The processing of the Sindbis structural proteins *in vitro* was studied using *in vitro* transcription and translation systems. The Sindbis structural proteins are translated *in vivo* from a 26S subgenomic mRNA which is 4106 nucleotides in length exclusive of the 3' terminal poly(A) tract. Translation of this mRNA *in vitro* has been shown to result in production of capsid protein and of other structural proteins (Simmons and Strauss, 1974; Garoff *et al.*, 1978). We constructed a plasmid expression vector which contained the 26S RNA region of Sindbis virus inserted next to a T7 RNA polymerase promoter (Fig. 1a). Transcription *in vitro* of the resulting plasmid, named T7SVSP, produces an RNA transcript which has 40 nucleotides extra at the 5' end (derived from the plasmid vector) and 3 nucleotides extra following the poly(A) tract, but which is otherwise identical to 26S mRNA (Fig. 1b). Translation of this *in vitro* transcribed RNA in a rabbit reticulocyte system gives a pattern of

**Fig. 1.** Constructs and procedures for transcription and translation *in vitro*. A) Construction of the vector for *in vitro* transcription with T7 RNA polymerase of the structural protein genes (T7SVSP). B) Outline of the procedure for *in vitro* transcription and translation. Abbreviations: CIAP, calf intestinal alkaline phosphatase.

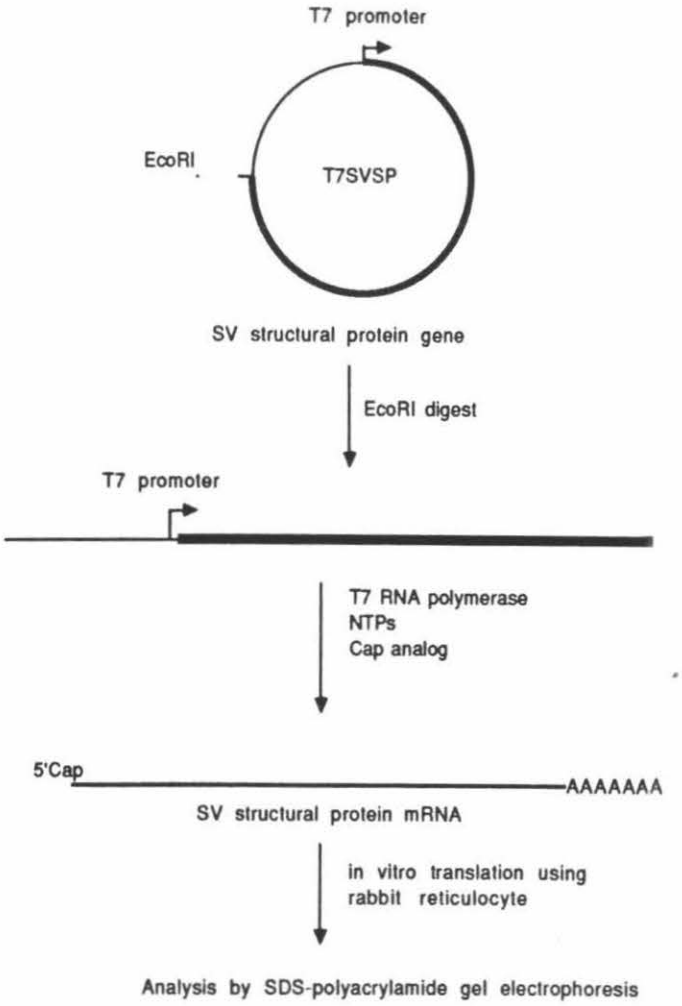
A

**Construction of in vitro transcription vector for Sindbis virus structural proteins**




B

**IN VITRO TRANSCRIPTION / TRANSLATION**



proteins synthesized which is indistinguishable from that translated from 26S mRNA isolated from infected cells (data not shown). Proteins produced *in vitro* include the capsid protein (32,000 daltons) and polypeptide p108 which contains the sequences of the envelope protein region (PE2, 6K, and E1) (data not shown, but see below and Simmons and Strauss, 1974). This *in vitro* system does not contain membranes and thus the p108 polypeptide cannot be inserted or processed and accumulates after translation. Besides these two proteins there are a few minor bands present which are probably premature termination products, perhaps due to termination in the hydrophobic domains in the C-terminus of E2 and the internal signal sequence in the 6K protein region.

***In vitro translation of mutagenized RNA.*** We have previously identified three amino acid changes in the Sindbis virus capsid protein that result in a nonfunctional capsid autoprotease at 40°C (Hahn *et al.*, 1985; Strauss *et al.*, 1987a). From the position of these mutations and sequence similarities between alphavirus capsid proteins and serine proteases from mammals and insects, we proposed that the capsid protein autoprotease possessed a catalytic triad similar to serine proteases (Hahn *et al.*, 1985). The three amino acids of the catalytic triad are diagrammed schematically in Fig. 2 together with changes that we effected in the sequence by oligonucleotide directed site specific mutagenesis. These changes were inserted into the expression vector T7SVSP, and mutant mRNA produced by *in vitro* transcription (Fig. 1b). This RNA was then translated *in vitro*, using the same amount of RNA in each reaction, in a rabbit reticulocyte translation system, and the resulting polypeptides were analyzed by SDS-polyacrylamide gel electrophoresis. The results are shown in Fig. 3. The translation products from wild-type RNA show complete cleavage of the polyprotein precursor to capsid protein and p108, as do those from the two mutations introduced at Asp-147. The change introduced at His-141 abolishes all

**Fig. 2.** Summary of alterations made by site-directed mutagenesis. Both amino acid and nucleotide changes are shown on a map of the capsid protein drawn to scale. The locations of His-141, Asp-147, and Ser-215 are indicated by vertical bars. Domains of the capsid protein which share amino acid sequence similarity with cellular serine proteases are shaded. The cleavage site between the capsid protein and glycoprotein PE2 is also shown. The location of four *ts* mutants which fail to cleave the capsid protein at the non-permissive temperature are also indicated (Hahn *et al.*, 1985; Strauss *et al.*, 1987a).

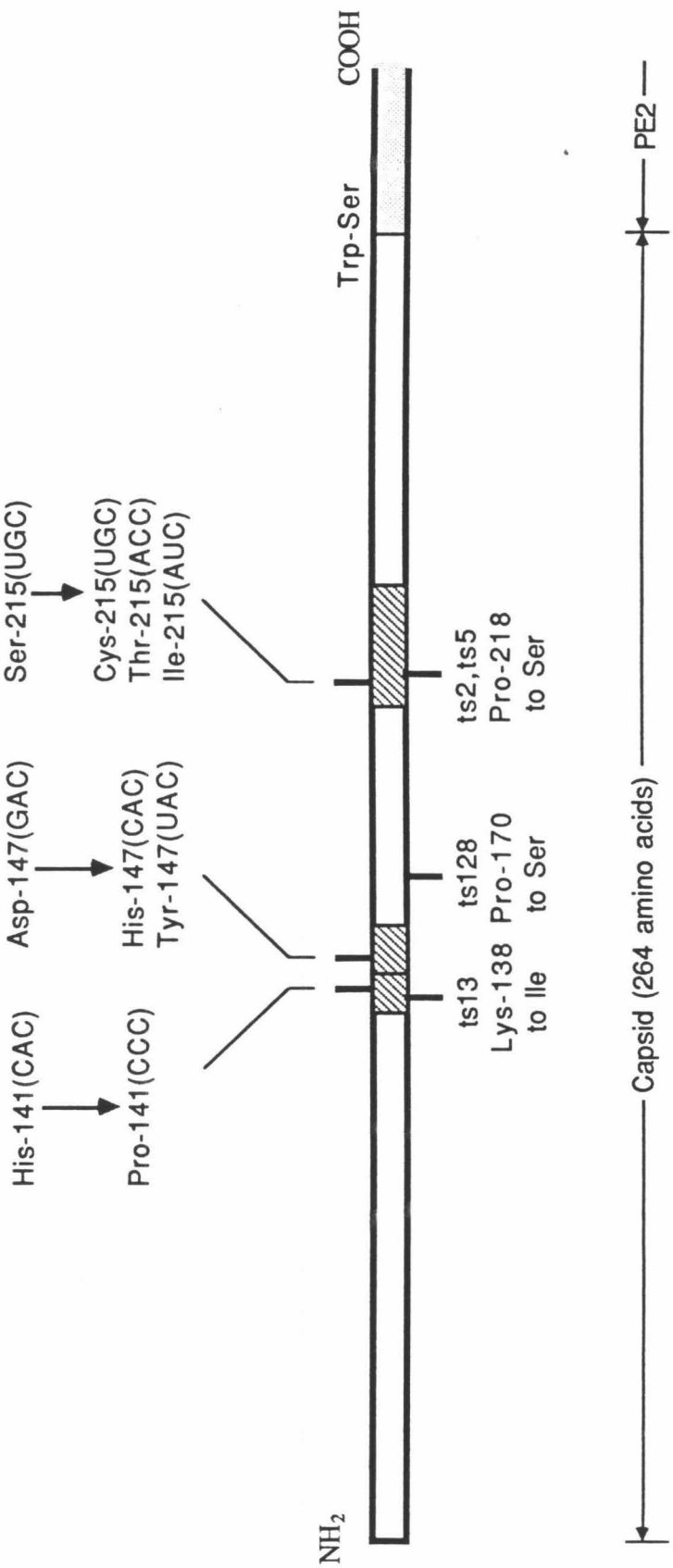
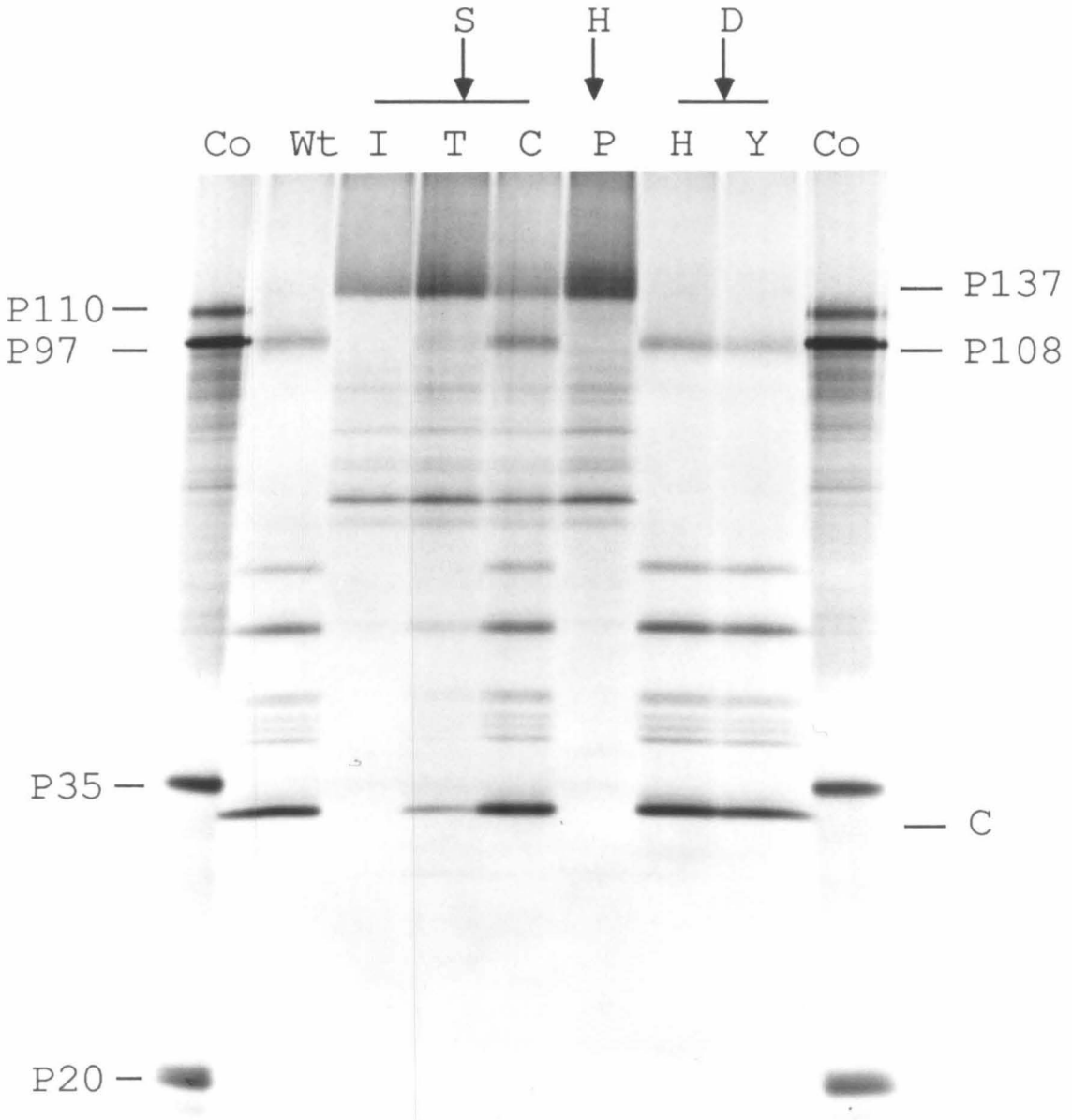

His-141(CAC)
Pro-141(CCC)
Asp-147(GAC)
His-147(CAC)
Tyr-147(UAC)
Ser-215(UGC)
Cys-215(UGC)
Thr-215(ACC)
Ile-215(AUC)
NH2
COOH
Trp-Ser
ts13
Lys-138
to Ile
ts128
Pro-170
to Ser
ts2,ts5
Pro-218
to Ser
Capsid (264 amino acids)
PE2

**Fig. 3.** SDS polyacrylamide gel analysis of *in vitro* translation products from *in vitro* transcripts of T7SVSP and its mutants. Outside lanes (Co) contain the products of *in vitro* translation of total brome mosaic virus RNA, and their molecular weights are indicated at the left. Lane 1 contains transcripts from T7SVSP (wild type HRSP). In lanes 2, 3, and 4, Ser-215 (S) has been changed to Ile (I), Thr (T), or Cys (C), respectively, in the plasmids used for transcription and translation. Similarly, in lane 5 His-141 (H) has been replaced with Pro (P), and in lanes 6 and 7 Asp-147 (D) is replaced with His (H) and Tyr (Y). The amount of RNA used was normalized before *in vitro* translation. The positions of capsid (C), p108 (the polypeptide containing the sequences of E3, E2, 6K, and E1), and p137 (the complete translation product containing C and p108 sequences) are indicated at the right. The extra bands located between capsid and p108 are probably due to premature termination occurring at the stretch of hydrophobic amino acids located at the C-terminus of glycoprotein E2 and within the 6K protein.


S
H
D
Co
Wt
I
T
C
P
H
Y
Co
P110
P97
P35
P20
P137
P108
C

protease activity in this reaction. The changes introduced at Ser-215 lead from 0 to 80% cleavage. These results are summarized in Fig. 4. The effects of changing His-141 and Ser-215 are clearly consistent with the hypothesis that these residues are part of the active site of the protease. The results at Asp-147 are ambiguous. Either this amino acid is not important for proteolysis, or if it is, the residual protease activity is sufficient to lead to complete cleavage under the assay conditions used (see also below). This latter possibility seems quite possible in view of the fact that substitution of Ser-215 by Cys results in reduced but detectable proteolytic activity in this system (Fig. 4).

***Effect of the mutations on virus infectivity.*** The four amino acid substitutions for which at least some proteolytic activity was detectable *in vitro*, namely Asp-147 to His or Tyr, and Ser-215 to Cys or Thr, were tested for their effect upon viability of virus. A restriction fragment containing the changed nucleotide was used to replace the corresponding restriction fragment in a full-length cDNA clone of Sindbis virus, Toto1101, from which infectious RNA can be transcribed *in vitro* (Rice *et al.*, 1987). RNA was transcribed from the plasmids and tested for its infectivity in secondary chicken embryo fibroblast cells at both 30°C and 40°C. No infectious virus could be recovered after transfection with RNA from any of these mutants (Fig. 4), although controls in the same experiment using RNA from Toto1101 were positive. Thus these mutations, although retaining all or some proteolytic activity as assayed *in vitro*, are lethal for the virus.

## DISCUSSION

We have tested the hypothesis that His-141, Asp-147 and Ser-215 of the Sindbis virus capsid protein form the catalytic triad of a serine autoprotease by substituting other amino acids for these residues. The results strongly support the hypothesis that His-141 and Ser-215 are important for proteolysis. In the case of

**Fig. 4.** Summary of protease activity and viability of capsid mutants. Percent cleavage is defined as the amount of capsid protein cleavage that occurs during *in vitro* translation. Quantitation was performed by densitometry of autoradiographs such as in Figure 3, and the amount of radioactivity in capsid protein for the various mutants was compared with that for wild type RNA (taken as 100% cleavage). Each mutant with demonstrable capsid cleavage was inserted into a cDNA clone of Sindbis virus, Toto1101, from which infectious RNA can be transcribed *in vitro*, RNA was transcribed, and an attempt was made to rescue virus by RNA transfection. All of these attempts were negative (-). NT (not tested) indicates that virus rescue was not attempted from mutants in which no cleavage could be detected.

| Wild-type sequence | Mutant sequence | Percent cleavage | Infectious virus |
|---|---|---|---|
| His-141 | Pro | 0 % | NT |
| Asp-147 | His | 100% | - |
| | Tyr | 100% | - |
| Ser-215 | Cys | 80 % | - |
| | Thr | 10 % | - |
| | Ile | 0 % | NT |

these two amino acids any substitution led to decreased proteolytic activity *in vitro* and were lethal for the virus. We have previously shown that a change of Lys-138 to Ile, three residues removed from His-141, or a change of Pro-218 to Ser, three residues removed from Ser-215, results in a temperature sensitive protease (Fig. 2 and Hahn *et al.*, 1985), and Melancon and Garoff (1987) have shown that insertion of Arg adjacent to Ser-215 results in an inactive protease. There is also sequence similarity between animal serine proteases and alphavirus capsid proteins around Ser-215 (Boege *et al.*, 1981) as well as around His-141 (Hahn *et al.*, 1985). Thus there is evidence from many sources that His-141 and Ser-215 form in fact part of the active site of the capsid protein.

The results at Asp-147 were less clear cut. Changes at this residue were lethal to the virus, but proteolytic activity assayed *in vitro* appeared unchanged. However, the cleavage event in this polyprotein precursor is different from other systems that have been studied in the past, in that the cleavage event is self-cleavage. The nascent protein takes up a conformation as it is being synthesized that presumably brings together the active site and the bond to be cleaved, and cleavage is thought to be virtually instantaneous. Thus even if the cleavage rate were slowed by several orders of magnitude proteolysis might still occur in the *in vitro* system at some time during translation. On the other hand, a reduced catalytic rate could lead to abortive processing of the structural polyprotein precursor. If the capsid moiety were not removed rapidly enough, the signal sequence at the N-terminus of PE2 could fail to function to insert the glycoprotein precursor into the endoplasmic reticulum, and if so the glycoproteins would not be modified, cleaved or transported to the site of virus budding, with the result that no virus would be produced. It is, of course, also possible that these mutations are lethal for reasons other than their effect on autoproteolysis. The capsid protein does have other essential functions in virus replication and

assembly which may be adversely affected. Further studies on the effects of these mutations *in vivo* will be useful to decide among these alternative explanations for the lethal phenotype.

There are two families of serine proteases known that have a catalytic triad consisting of an aspartic acid, a histidine and a serine; the first is represented by mammalian and insect serine proteases and the second by bacterial subtilisin. Studies of these enzymes using chemical (Neet and Koshland, 1966) or site-specific mutagenesis (Craik *et al.*, 1987; Carter and Wells, 1987, 1988) has shown that changes at any of these three amino acids have dramatic effects upon proteolytic activity. Craik *et al.* (1987) found that when the active site Asp-102 of rat trypsin was replaced by Asn, the catalytic activity decreased to $10^{-4}$ to $6x10^{-2}$ of the wild type rate, depending upon the pH. More extensive studies of the amino acids in the catalytic triad of subtilisin were reported by Carter and Wells (1988). A change of Asp-32 to Ala led to a decrease in the activity to $4x10^{-5}$ of the wild type rate, whereas changing either Ser-221 to Ala or His-64 to Ala resulted in a decrease in activity to $6x10^{-7}$. Even in the latter cases, however, the catalytic rate was some three orders of magnitude faster than proteolysis in buffer alone, suggesting that the mutant protein lacking a functional catalytic triad had activity analogous to that of catalytic antibodies (Tramontano *et al.*, 1986; Pollack *et al.*, 1986) in which the binding of the protein to the substrate enhances proteolysis in aqueous solution by stabilizing transition state complexes.

These results suggest that a change at the Asp in the catalytic triad produces a less severe inhibition of activity than do changes at either the Ser or His. In the case of the Sindbis capsid autoprotease, if the model is correct, the changes induced at Asp-147 could lead to inhibition of catalytic proteolysis by several orders of magnitude, but self-cleavage by the molecule could still go to

completion during the *in vitro* translation reaction. In contrast, changes in the Ser residue or His residue, which lead to a more extensive inhibition of activity, would result in incomplete proteolysis *in vitro*. The results with Ser-215 are of particular interest. Substitution of Cys for Ser resulted in 80% cleavage, and it seems reasonable to assume that the sulfhydryl group of the Cys residue is functioning in proteolysis in the same way that the hydroxyl of the Ser functions. Substitution of Ser by Cys in other serine proteases has given rise to an enzyme which retains some esterase activity but in which the protease activity has been lost (Neet and Koshland, 1966). In these assay systems, however, a proteolytic rate of less than $10^{-3}$ or $10^{-4}$ of the activity of the wild type protease would not have been detected. Similarly the residual activity from substitution of Ser by Thr also suggests that the hydroxyl group of Thr might serve in the nucleophilic attack during proteolysis as does the hydroxyl group of Ser, but at a greatly reduced rate because of the extra methyl group on the carbon containing the hydroxyl group.

## ACKNOWLEDGEMENTS

We thank Dr. J. Richards and Y. Hahn for many helpful discussions. This work was supported by NIH Grants AI10793 and AI20612.

# REFERENCES

ALIPERTI, G. and SCHLESINGER, M. (1978). Evidence for an autoprotease activity of Sindbis virus capsid protein. *Virology* **90,** 366-369.

BOEGE, U., WENGLER, G., WENGLER, G. and WITTMANN-LIEBOLD, B. (1981). Primary structure of the core proteins of the alphaviruses, Semliki Forest virus and Sindbis virus. *Virology* **113,** 293-303.

CARTER, P. and WELLS, J. A. (1987). Engineering enzyme specificity by "substrate-assisted catalysis." *Science* **237,** 394-399.

CARTER, P. and WELLS, J. A. (1988). Dissecting the catalytic triad of a serine protease. *Nature* **332,** 564-568.

CLEGG, J. C. S. (1975). Sequential translation of capsid and membrane protein genes of alphaviruses. *Nature* **254,** 454-455.

CRAIK, C. S., ROCZNIAK, S., LARGMAN, C. and RUTTER, W. J. (1987). The catalytic role of the active site aspartic acid in serine proteases. *Science* **237,** 909-913.

DALBADIE-McFARLAND, G., COHEN, L. W., RIGGS, A. D., MORIN, C., ITAKURA, K. and RICHARDS, J. H. (1982). Oligonucleotide-directed mutagenesis as a general and powerful method for studies of protein function. *Proc. Natl. Acad. Sci. USA* **79,** 6409-6413.

GAROFF, H., FRISCHAUF, A. M., SIMONS, K., LEHRACH, H. and DELIUS, H. (1980a). The capsid protein of Semliki Forest virus has clusters of basic amino acids and prolines in its amino-terminal region. *Proc. Natl. Acad. Sci. USA* **77,** 6376-6380.

GAROFF, H., FRISCHAUF, A. M., SIMONS, K., LEHRACH, H. and DELIUS, H. (1980b). Nucleotide sequence of cDNA coding for Semliki Forest virus membrane glycoproteins. *Nature (London)* **288,** 236-241.

GAROFF, H., SIMONS, K., and DOBBERSTEIN, B. (1978). Assembly of the Semliki Forest virus membrane glycoproteins in the membrane of the endoplasmic reticulum *in vitro*. *J. Mol. Biol.* **124**, 587-600.

HAHN C. S., STRAUSS, E. G. and STRAUSS, J. H. (1985). Sequence analysis of three Sindbis virus mutants temperature-sensitive in the capsid autoprotease. *Proc. Natl. Acad. Sci. USA* **82**, 4648-4652.

LINDQVIST, B. H., DiSALVO, J., RICE, C. M., STRAUSS, J. H. and STRAUSS, E. G. (1986). Sindbis virus mutant *ts* 20 of complementation group E contains a lesion in glycoprotein E2. *Virology* **151**, 10-20.

LUSTIG, S., JACKSON, A. C., HAHN, C. S., GRIFFIN, D. E., STRAUSS, E. G. and STRAUSS, J. H. (1988). Molecular basis of Sindbis virus neurovirulence in mice. *J. Virology*, in press.

MANIATIS, T., FRITSCH, E. F. and SAMBROOK, J. (1982). Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

MELANCON, P. and GAROFF, H. (1987). Processing of the Semliki Forest virus structural proteins: Role of the capsid protease. *J. Virol.* **61**, 1301-1309.

NEET, K. E. and KOSHLAND, D. E. (1966). The conversion of serine at the active site of subtilisin to cysteine: A "chemical mutation." *Proc. Natl. Acad. Sci. USA* **56**, 1606-1611.

POLLACK, S. J., JACOBS, J. W. and SCHULTZ, P. G. (1986). Selective chemical catalysis by an antibody. *Science* **234**, 1570-1573.

RICE, C. M., LEVIS, R., STRAUSS, J. H. and HUANG, H. V. (1987). Production of infectious RNA transcripts from Sindbis virus cDNA clones: Mapping of lethal mutations, rescue of a temperature-sensitive marker, and *in vitro* mutagenesis to generate defined mutants. *J. Virol.* **61**, 3809-3819.

RICE, C. M. and STRAUSS, J. H. (1981). Nucleotide sequence of the 26S mRNA of Sindbis virus and deduced sequence of the encoded virus structural proteins. *Proc. Natl. Acad. Sci. USA* **78,** 2062-2066.

SANGER, F., NICKLEN, S. and COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74,** 5463-5468.

SCHEELE, C. M. and PFEFFERKORN, E. R. (1970). Virus-specific protein synthesis in cells infected with $RNA^+$ temperature-sensitive mutants of Sindbis virus. *J. Virol.* **5,** 329-337.

SCHLESINGER, M. J. and KAARIAINEN, L. (1980). Translation and processing of alphavirus proteins. *In* "The Togaviruses (ed. Schlesinger, W.), Chapter 13, Academic Press, New York, pp. 371-392.

SCUPHAM, R. K., JONES, K. J., SAGIK, J. B. and BOSE, H. R. (1977). Virus-directed posttranslational clevage in Sindbis virus-infected cells. *J. Virol.* **22,** 568-571.

SEFTON, B. M. (1977). Immediate glycosylation of Sindbis virus membrane proteins. *Cell* **10,** 659-668.

SHORTLE, D., DiMAIO, D. and NATHANS, D. (1981). Directed mutagenesis. *Ann. Rev. Genet.* **15,** 265-294.

SIMMONS, D. T. and STRAUSS, J. H. (1972). Replication of Sindbis virus. I. Relative size of and genetic content of 26S and 49S RNA. *J. Mol. Biol.* **71,** 599-613.

SIMMONS, D. T. and STRAUSS, J. H. (1974). Translation of Sindbis virus 26S RNA and 49S RNA in lysates of rabbit reticulocytes. *J. Mol. Biol.* **86,** 397-409.

STRAUSS, E. C., KOBORI, J. A., SIU, J. and HOOD, L. E. (1986). Specific-primer-directed DNA sequencing. *Anal. Biochem.* **154,** 353-360.

STRAUSS, E. G., RICE, C. M. and STRAUSS, J. H. (1984). Complete nucleotide sequence of the genomic RNA of Sindbis virus. *Virology* **133,** 92-110.

STRAUSS, E. G. and STRAUSS, J. H. (1980). Mutants of alphaviruses: Genetics and physiology. *In* "The Togaviruses: Biology, Structure, Replication" R. Schlesinger (ed.) Chapter 14, Academic Press, New York, pp. 393-426.

STRAUSS, E. G. and STRAUSS, J. H. (1986). Structure and replication of the alphavirus genome. *In* "The Togaviridae and Flaviviridae" S. Schlesinger and M. Schlesinger (eds.) Chapter 3, Plenum Publishing Corp., New York, pp. 35-90.

STRAUSS, J. H. and STRAUSS, E. G. (1976). Togaviruses. *In* "The Molecular Biology of Animal Viruses" Vol. 1, D. P. Nayak (ed.) Chapter 4, Marcel Dekker, Inc., New York, pp. 111-166.

STRAUSS, J. H., STRAUSS, E. G., HAHN, C. S., HAHN, Y. S., GALLER, R., HARDY, W. R. and RICE, C. M. (1987a). Replication of alphaviruses and flaviviruses: Proteolytic processing of polyproteins. *In* "Positive Strand RNA Viruses" M. A. Brinton and R. Rueckert (eds.) UCLA Symposia on Molecular and Cellular Biology, Vol. 54, Alan R. Liss, Inc., New York, pp. 209-225.

STRAUSS, J. H., STRAUSS, E. G., HAHN, C. S. and RICE, C. M. (1987b). The genomes of alphaviruses and flaviviruses: Organization and translation. *In* "The Molecular Biology of the Positive Strand Viruses" D. J. Rowlands, M. Mayo and B. W. J. Mahy (eds.) Chapter 6, Academic Press, London, pp. 75-104.

TRAMONTANO, A., JANDA, K. D., and LERNER, R. A. (1986). Catalytic antibodies. ***Science* 234**, 1566-1570.

ZOLLER, M. J. and SMITH, M. (1984). Laboratory Methods: Oligonucleotide-directed mutagensis: A simple method using two oligonucleotide primers and a single-strand DNA template. ***DNA* 3**, 477-488.

# Part II

## The Role of Glycoproteins in Assembly and Virulence

# Chapter 3

## Sindbis Virus *ts*103 Has a Mutation in Glycoprotein E2 That Leads to Defective Assembly of Virions

This chapter was prepared for submission.

# Sindbis Virus *ts*103 Has a Mutation in Glycoprotein E2 That Leads to Defective Assembly of Virions

CHANG S. HAHN, CHARLES M. RICE*, ELLEN G. STRAUSS, EDITH M. LENCHES, and JAMES H. STRAUSS

*Division of Biology, California Institute of Technology, Pasadena, California 91125*

Running title: Assembly Mutant of Sindbis Virus

*Present Address: Department of Microbiology and Immunology, Washington University School of Medicine, St. Louis, Missouri 63110-1093

We have determined that a mutation in glycoprotein E2 of Sindbis virus *ts*103 is the change that leads to defects in the late stages of maturation and to formation of multi-cored particles. The mutant phenotype was mapped to glycoprotein E2 by constructing recombinant viruses between the parental strain of Sindbis virus and the mutant. To accomplish this, we replaced restriction fragments in a full-length cDNA clone of Sindbis virus, from which infectious RNA can be transcribed *in vitro*, with the corresponding fragments from cDNA clones of *ts*103. Viruses were rescued from these recombinant clones and only recombinant viruses containing glycoprotein E2 of *ts*103 were phenotypically *ts*103, forming minute plaques at 30° or 40°, growing slowly to low yield, and forming multi-cored particles. Sequence analysis of *ts*103 and of its parent (HR) showed that there was only one difference in glycoprotein E2 between the two strains: a change of alanine to valine at position 344. A partial revertant of *ts*103, called *ts*103R, was also mapped and sequenced. These studies clearly showed that *ts*103R is a second site revertant in which a change in glycoprotein E1 from lysine to methionine at position 227 partially suppresses the phenotypic effects of the change at E2 position 344. The assembly defect in *ts*103 appears to result from a weakened interaction between the virus membrane glycoproteins and the nucleocapsid during budding. Both the E2 mutation leading to the defect in virus assembly and the suppressor mutation in glycoprotein E1 are in the domains external to the lipid bilayer, and thus in domains that cannot interact directly with the nucleocapsid. This suggests that in *ts*103, either the E1-E2 heterodimers or the trimeric spikes (consisting of three E1-E2 hetero-dimers) have an aberrant conformation, and thus cannot interact properly with the nucleocapsid.

## INTRODUCTION

Sindbis virus is a small enveloped RNA virus that belongs to the genus Alphavirus of the family *Togaviridae*. The nucleocapsid of the virus is an icosahedral structure (T=3) that contains the single stranded RNA of 11.7 kb complexed with 180 copies of a 30,000 dalton capsid protein. During the final stages of virus assembly, the nucleocapsid, assembled in the cytoplasm, buds through the plasma membrane to acquire an envelope consisting of a lipid bilayer containing two virus encoded integral membrane glycoproteins, E2 and E1. In the final virion structure, 240 copies of E1 and E2 are present in the membrane in an icosahedral array (T=4). Each external spike of the virus consists of a trimer of E1-E2 heterodimers, and these spikes and the nucleocapsid fit together in a precise fashion (Fuller, 1987). The precision of virus assembly and rigid exclusion of nonviral proteins from the virus (Strauss, 1978) has led us and others to hypothesize that there is a specific interaction between the glycoproteins and the nucleocapsid that furnishes the free energy required to drive virus budding.

Mutant *ts*103 was isolated more than a decade ago following mutagenesis with nitrous acid (Strauss *et al.*, 1976). It is a minute plaque former which grows slowly at any temperature and produces, under optimal conditions, virus yields of 3-10% of those of the parental HRSP strain of Sindbis virus (Strauss *et al.*, 1977). Following *ts*103 infection, RNA synthesis and protein synthesis as well as nucleocapsid formation are virtually indistinguishable from those following infection by the parental strain of virus, and the very slow rate of *ts*103 virus production appears to be due to a defect in the final stages of virus maturation, the budding of nucleocapsids through the capsid membrane to produce the infectious virus. Electron microscopy of *ts*103 infected cells reveals the presence of large numbers of nucleocapsids apparently in the process of budding. Yet the release of mature virus is delayed and the final yield of virus is reduced.

Examination of the released virions by sedimentation velocity centrifugation or by electron microscopy showed that many virus particles contain from 2 to 5 or more nucleocapsids in a single envelope (Strauss *et al.*, 1977). Thus the interactions of the nucleocapsids with the glycoproteins appear to be weak, leading to a slow rate of virus assembly and to the formation of aberrant virions.

Here we report the sequence of the entire structural protein region of *ts*103 and of a partial revertant *ts*103R, and a comparison of these sequences with that of the parental HRSP strain. In addition we have mapped the location of the mutation resulting in the *ts*103 phenotype by constructing recombinant viruses between *ts*103 and HRSP, using a cDNA clone of the HR strain of Sindbis virus from which infectious RNA can be transcribed *in vitro* (Rice *et al.*, 1987), and in this way shown that the *ts*103 phenotype maps to a change in glycoprotein E2.

## MATERIALS AND METHODS

### Cells and viruses

Culturing of chicken embryo fibroblast cells, infection with Sindbis virus, and plaque assay were as previously described (Pierce *et al.*, 1974; Strauss *et al.*, 1976), with the exception that in all experiments, 1 to 1.2% agarose was used for overlay during plaque formtion (Strauss *et al.*, 1977; Rice *et al.*, 1987). Isolation of mutant *ts*103 and of the revertant *ts*103R have been previously described (Strauss *et al.*, 1976, 1977).

### Virus purification and isolation of RNA

*ts*103 and *ts*103R were grown in primary chick embryo fibroblast cells in hypotonic saline for 10 to 20 hours and virus was harvested in hypertonic saline as described (Pierce *et al.*, 1974; Strauss *et al.*, 1977). Virus was purified and RNA extracted as previously described (Rice and Strauss, 1981).

**Isolation of cDNA clones**

Clones containing the 3' terminal 5438 nucleotides [plus a variable length of poly(A)] of the genomes of *ts*103 and of *ts*103R were obtained by using a T-tailed plasmid vector for first strand synthesis as previously described (Lindqvist *et al.*, 1986; Hahn *et al.*, 1985). A library of cDNA clones representing the rest of the genome was obtained by using calf thymus random primers for first strand synthesis and *Eco*RI linkers to clone the double-stranded cDNA into a plasmid vector essentially as described (Okayama and Berg, 1982; Rice *et al.*, 1988). To identify clones containing inserts representing the 5' end of the genomes, colonies were probed with a radiolabeled RNA consisting of the 5' terminal 500 nucleotides of Sindbis virus HRSP, derived by transcribing RNA *in vitro* from clone Toto1101 (Rice *et al.*, 1987).

**Sequence analysis**

Sequence of *ts*103 and of *ts*103R was obtained by the chemical sequencing methods of Maxam and Gilbert (1980) as modified by Smith and Calvo (1980). Most of the sequence of the structural region of *ts*103 and *ts*103R was obtained from sequencing of single stranded cDNA restriction fragments produced by *Hae*III digestion (Rice and Strauss, 1981; Arias *et al.*, 1983). This method has the advantage that a consensus sequence is immediately obtained. These results were confirmed and extended by sequencing of cDNA derived from plasmid clones as previously described (Strauss *et al.*, 1984). To confirm the nature of recombinant viruses produced from construction of hybrid genomes, certain regions of the recombinant virus genomes were sequenced by direct RNA sequencing, using reverse transcriptase, synthetic primers, and RNA templates as described (Zimmern and Kaesberg, 1978).

**General recombinant DNA techniques**

Restriction endonucleases and DNA modifying enzymes were purchased from commercial sources and used essentially as recommended by the manufacturers. Plasmids were grown, purified, and analyzed using standard methods with minor modifications (Maniatis *et al.*, 1982).

**Construction of recombinant viruses**

Hybrid genomes were constructed by replacing restriction fragments in Sindbis cDNA clone Toto1101 or Toto50 (Rice *et al.*, 1987) with the corresponding regions from cDNA clones derived from either *ts*103 or *ts*103R. Details of the restriction sites used are given in the figure legends. Virus was rescued from these recombinant clones essentially as described (Lustig *et al.*, 1988) and the virus tested for its biological properties.

**Analysis of virus stocks for multi-cored particles**

Viral RNA was labeled with [$^3$H] uridine during replication, and 1 to 4 ml of the high salt harvest from a 100 mm tissue culture plate was layered on to a 10 ml sucrose gradient (15 to 30% w/v sucrose in 200 mM sodium chloride, 50 mM tris pH 7.4, 1 mM EDTA). Centrifugation was at 5°C for 80 to 90 min at 27,000 RPM or 50 to 60 min at 36,000 RPM in a Beckman SW40 or SW41 rotor. The gradients were fractionated and assayed for radioactivity in a liquid scintillation counter.

## RESULTS

**Sequence analysis of the structural proteins of *ts*103 and *ts*103R.**

Since *ts*103 has a defect in the assembly of virions, it seemed probable that the mutation would lie in one of the structural proteins. To start, the entire structural protein region of the genomes of *ts*103 and of *ts*103R were sequenced and compared to the sequence of the parental HRSP strain. Most of the sequence

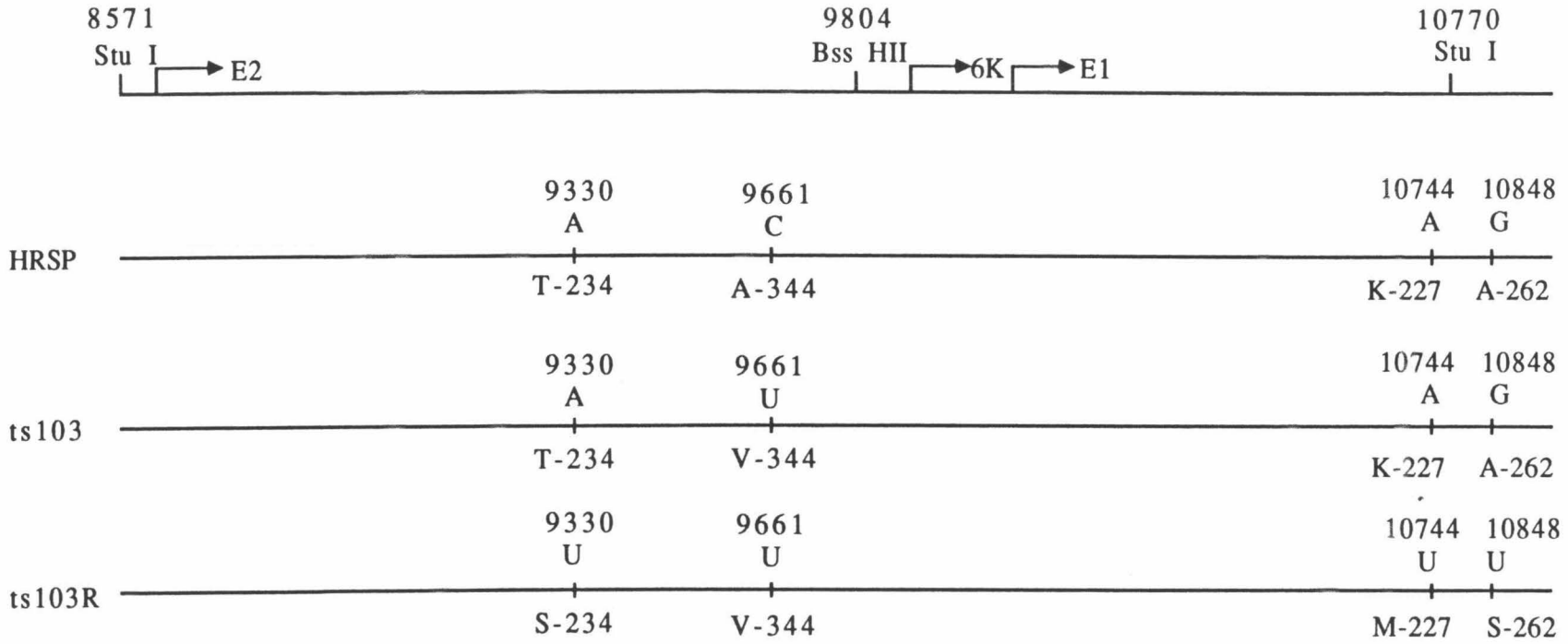
was obtained by direct sequencing of first strand cDNA after digestion with *Hae*III (Rice and Strauss, 1981) which gives directly the consensus sequence in the RNA population. These sequencing results were confirmed and extended by sequencing of cDNA clones of *ts*103 and *ts*103R. The results are shown in Fig. 1.

There were no changes in the 5' or 3' nontranslated regions flanking the structural protein region, and there were no coding differences in the capsid proteins of the viruses (although there was a silent change in the second codon of the capsid protein, AAU → AAC, in *ts*103 and *ts*103R). In the remainder of the structural region, there was only a single change between HRSP and *ts*103, a C to U change at position 9661 leading to the substitution of Ala-344 in the E2 glycoprotein in HRSP by Val in *ts*103 (GCC → GUC). *ts*103R was found to contain Val-344, as did *ts*103, but contained one additional change in E2 and two differences in the E1 region from both *ts*103 and HRSP. These were changes of nt9330 from A to U (leading to the substitution of Thr-234 in E2 by Ser), of nt10744 from A to U (leading to the substitution of Lys-227 in E1 by Met) and of nt10848 from G to U (leading to the substitution of Ala-262 in E1 by Ser). These data indicated that if, in fact, the assembly defect of *ts*103 was due to a change within the structural protein region, that change must be the Ala to Val change at position 344 of E2, and that the partial revertant isolated was a second site revertant, presumably involving one or more of the changes in either glycoprotein E2 or E1.

**Localization of the mutation responsible for the *ts*103 phenotype**

In order to map the location of the change responsible for the *ts*103 phenotype, recombinant viruses were constructed. Sindbis virus cDNA clones Toto1101 and Toto50 contain complete cDNA copies of Sindbis virus inserted downstream from an SP6 RNA polymerase promoter, and infectious RNA can be transcribed from these clones *in vitro* (Rice *et al.*, 1987). Hybrid cDNA clones

**Fig. 1.** Sequence differences between HRSP, *ts*103, and *ts*103R in the glycoprotein region. A schematic diagram of the glycoprotein region of the Sindbis genome is shown with the location of certain restriction enzymes sites. Nucleotide coordinates are numbered from the 5' end of the RNA according to Strauss *et al.* (1984). Below are shown sequencing schematics of HRSP (Strauss *et al.*, 1984), *ts*103, and *ts*103R. Where differences in any of the strains occur, nucleotides are shown above the line together with the nucleotide coordinate, and the encoded amino acids are shown below the line, using the single letter code, together with amino acid position numbered from the amino terminus of glycoprotein E2 or of glycoprotein E1.


8571
Stu I
E2
9804
Bss HII
6K
E1
10770
Stu I
9330
A
9661
C
10744
A
10848
G
HRSP
T-234
A-344
K-227
A-262
9330
A
9661
U
10744
A
10848
G
ts103
T-234
V-344
K-227
A-262
9330
U
9661
U
10744
U
10848
U
ts103R
S-234
V-344
M-227
S-262

were constructed by replacing restriction fragments within one of these clones with the corresponding sequences from cDNA clones of *ts*103 or *ts*103R. RNA was transcribed from the recombinant clones and transfected onto secondary chick embryo fibroblast cells. Monolayers were incubated at 30°C and virus was harvested when the cells showed a full cytopathic effect. The parental HR strain or phenotypically wild-type recombinants required about 72 hr for this, whereas *ts*103 took up to 6 days. Monolayers were also incubated at 30°C or 40°C under 1 to 1.2% agarose for plaque assay. The constructs tested are diagrammed in Fig. 2, which also includes the coordinates of the restriction sites used to make the hybrid genomes. Note that in each case most of the viral sequences are derived from clone Toto1101 with small, contiguous elements derived from *ts*103. For simplicity these various constructs and the viruses rescued from them will be referred to as 103A through 103F as indicated in Fig. 2.

The phenotypes of each of the six recombinants illustrated in Fig. 2 are shown in Fig. 3. The plaque size of 103A at 30°C and of 103B and 103C at both 30° and 40° were indistinguishable from that of Toto1101 virus. Recombinants 103D, 103E, and 103F showed minute plaques at both 30°C and 40°C, as does *ts*103. Furthermore, the size of virus plaques correlates with the kinetics of virus release and with the formation of multi-cored particles. Virus from 103A, 103B and 103C have the same growth rate as that from Toto1101, whereas *ts*103 and recombinants 103D, 103E, and 103F have a very slow kinetics of virus release. Similarly, tests for the production of multi-cored particles by sucrose gradient sedimentation showed that virus from Toto1101 or from 103A, 103B, and 103C gave a single sharp virus peak sedimenting at 280 S. In contrast, *ts*103 and recombinant 103D showed multiple peaks of virus sedimenting between 280 S and 700 S (recombinants 103E and 103F were not tested). Since recombinants 103D, 103E, and 103F share only the E2 region of *ts*103, we conclude that glycoprotein

**Fig. 2.** Construction of Sindbis virus genomes recombinant between HR virus and *ts*103. Restriction fragments in Sindbis virus cDNA clone Toto1101 (Rice *et al.*, 1987) were replaced with the corresponding fragments from cDNA clones of *ts*103 (indicated by the diagonal hatching). The restriction sites used for these constructs and their coordinates in the viral genome are indicated. These constructs were named 103A, 103B, etc. in the approximate order, from 5' to 3' in the genome, in which sequence was derived from *ts*103, as indicated.

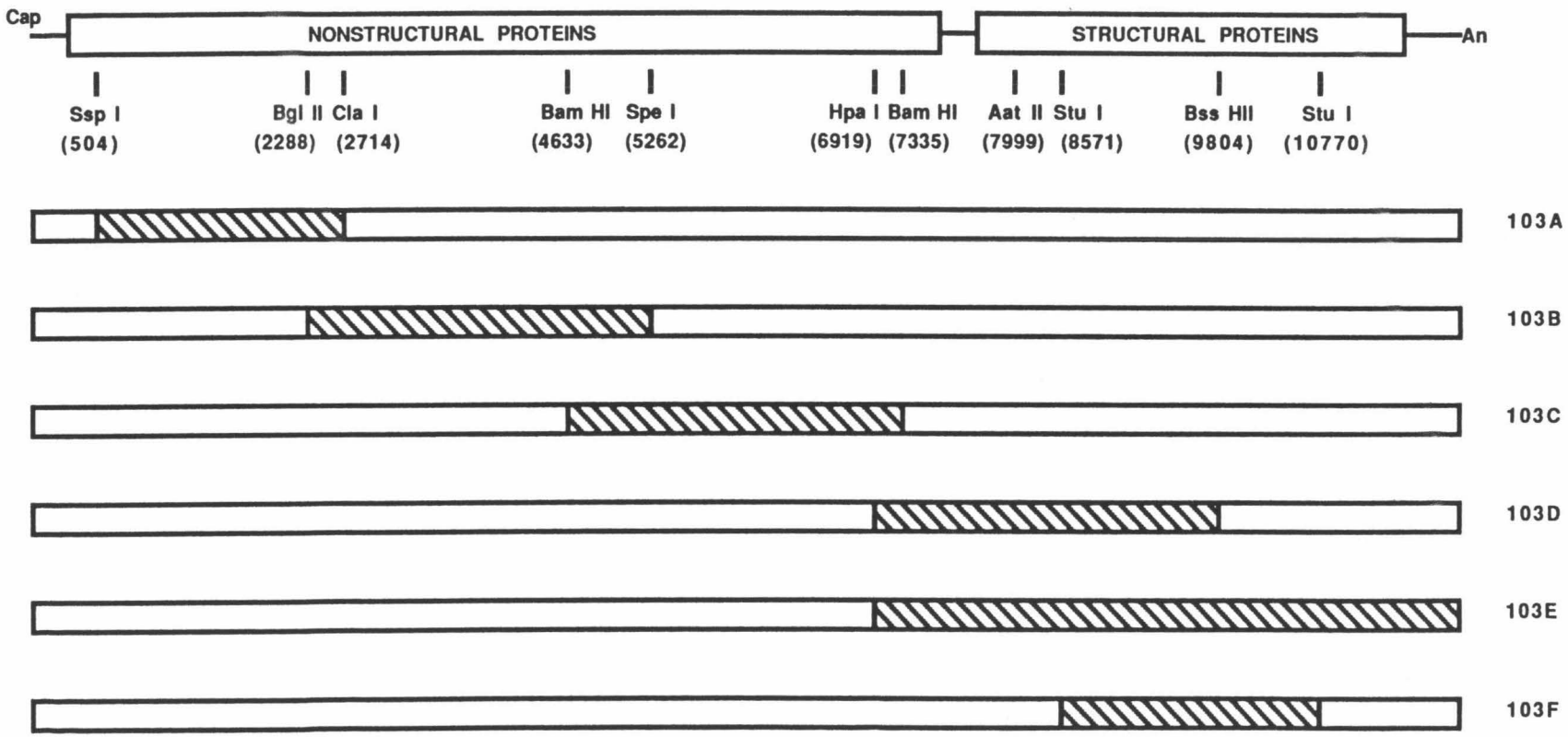

Cap
NONSTRUCTURAL PROTEINS
STRUCTURAL PROTEINS
An
Ssp I (504)
Bgl II (2288)
Cla I (2714)
Bam HI (4633)
Spe I (5262)
Hpa I (6919)
Bam HI (7335)
Aat II (7999)
Stu I (8571)
Bss HII (9804)
Stu I (10770)
103A
103B
103C
103D
103E
103F

**Fig. 3.** Functional assay of recombinant viruses. The recombinant viruses rescued from the constructs shown in Fig. 2, together with virus rescued from clone Toto1101 and virus *ts*103, were tested for plaque size at 40°C and at 30°C, for the growth kinetics of the virus, and for the formation of multi-cored particles by sucrose gradient velocity sedimentation (NT = not tested).

| | Plaque Size 40 °C/30° C | Growth Pattern | Multicore Particle |
|---|---|---|---|
| Toto1101 | Normal/Normal | Normal | No |
| ts103 | Minute/Minute | Very Slow | Yes |
| 103A | Small/Normal | Normal | No |
| 103B | Normal/Normal | Normal | No |
| 103C | Normal/Normal | Normal | No |
| 103D | Minute/Minute | Very Slow | Yes |
| 103E | Minute/Minute | Very Slow | N/T |
| 103F | Minute/Minute | Very Slow | N/T |

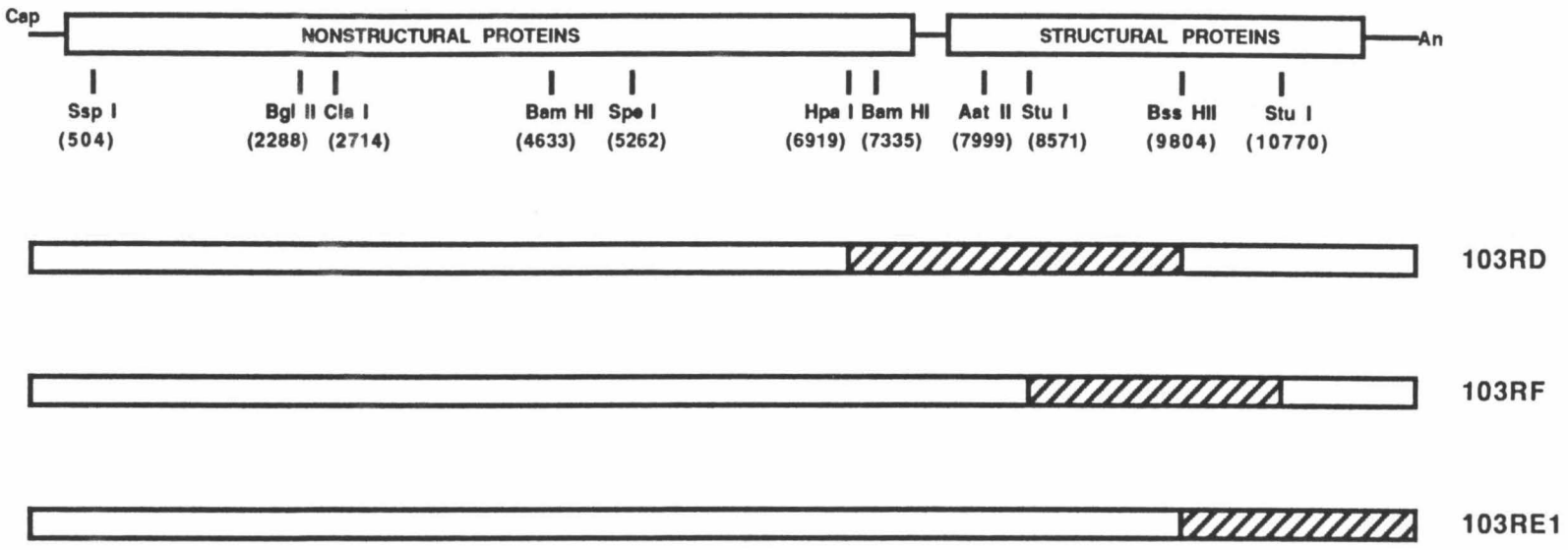
E2 from *ts*103 is necessary and sufficient to obtain the *ts*103 phenotype, and that the plaque-size, growth pattern, and production of multi-cored particles are all due to the same mutation, namely the substitution of Ala-344 in E2 by Val (Fig. 1).

**Second site reversion in *ts*103R**

The partial revertant *ts*103R was found to be intermediate in phenotype between *ts*103 and HR, in terms of virus release, production of multi-cored particles and plaque morphology. Since *ts*103R possesses Val-344 in E2, as does *ts*103, the altered phenotype must be due to partial suppression of this mutation by a second site mutation. To confirm this, a recombinant virus between Toto1101 and *ts*103R was made corresponding to construct 103D of Fig. 2. This recombinant virus was tested for its biological properties and found to possesses the *ts*103 phenotype, showing that the original *ts*103 mutation was in fact still present and that the suppressing mutation must lie outside the region defined by construct 103D, that is, outside the coordinates 6919 to 9804 in the viral RNA (Fig. 4).

The location of the suppressor mutation was defined by a second recombinant virus tested, recombinant 103RF, illustrated in Fig. 4. Recombinant 103RF had the same phenotype as *ts*103R, namely it made small plaques (as opposed to minute plaques made by *ts*103) at 40°C and 30°C, released virus more rapidly than *ts*103 (but still more slowly than HR), and and made a smaller proportion of multi-cored particles. This result shows that the suppressor mutation must lie between coordinates 9804 and 10770 (Fig. 4), and thus that the suppressor mutation must be the Lys-227 to Met change in glycoprotein E1 (Fig. 1). Note, however, that 103RF contains Ser-234 of E2 and, although this change alone will not suppress the *ts*103 mutation (construct 103D), we cannot rule out an effect of this change upon suppression, acting in concert with Met-227 in E1.

**Fig. 4.** Construction of Sindbis virus genomes recombinant between *ts*103R and Toto1101. Restriction fragments in clone Toto1101 (Rice *et al.*, 1987) were replaced by the corresponding fragments from cDNA clones of *ts*103R (indicated by the diagonal cross hatching). The restriction sites used and their coordinates in the viral genome are indicated. Constructs were named as indicated to the right of the figure. Note that in construct 103RD, the same *Hpa*I to *Bss*HII restriction fragment of Toto 1101 is substituted from *ts*103R cDNA as was used to form construct 103D from clones of *ts*103, and that the similar case holds for clone 103RF, which uses a *Stu*I to *Stu*I restriction fragment.


Cap
NONSTRUCTURAL PROTEINS
STRUCTURAL PROTEINS
An
Ssp I (504)
Bgl II (2288)
Cla I (2714)
Bam HI (4633)
Spe I (5262)
Hpa I (6919)
Bam HI (7335)
Aat II (7999)
Stu I (8571)
Bss HII (9804)
Stu I (10770)
103RD
103RF
103RE1

A third recombinant virus, 103RE1, was also constructed and tested. This virus contains the E1 region from *ts*103R with the rest of the genome derived from Toto1101. This recombinant virus formed large plaques (larger than the plaques formed by Toto1101 virus), had a normal growth pattern, and did not lead to the formation of multi-cored particles. Thus, the suppressor mutation, when separated from the *ts*103 mutation, leads to the formation of large plaques, and may be a generalized suppressor that increases the efficiency of virus assembly and release in some way.

**Revertant *ts*103RR**

At the time of the isolation of *ts*103R, since it seemed to be only a partial revertant, we attempted to isolate a complete revertant from the *ts*103R population. One such isolate made large (rather than intermediate) plaques and produced no multi-cored particles; this was called *ts*103RR. Limited sequence analysis of *ts*103RR has showed that it has Ala-344 at E2 (like HR) so that it is a true (same site) revertant. The analysis of *ts*103RR is not yet complete and it is unknown whether it retains the E1 change of *ts*103R as well.

**Reversion frequency of *ts*103**

During the isolation of revertants from *ts*103, the reversion frequency was found to be less than $10^{-7}$ (Strauss *et al.*, 1977). This reversion frequency is quite low in view of the fact that the *ts*103 phenotype is obtained with a single nucleotide change in the E2 coding region and partial suppression of this mutant can be accomplished by a single nucleotide change in the E1 region of *ts*103R. Because *ts*103 was originally isolated following nitrous acid mutagenesis of HRSP, it seemed possible that other unmapped changes in the genome might contribute to the difficulties in obtaining revertants. We have reexamined the reversion frequency of *ts*103 by using recombinant 103D which differs from Toto1101 by only 1 coding difference, the change in position 344 of glycoprotein E2. Construct

**Fig. 5.** Functional assay of viruses recombinant between Toto1101 and *ts*103R. Assays included plaque size at 30°C and 40°C, growth kinetics, and the formation of multi-cored particles as assayed by sucrose gradient velocity sedimentation. (NT = not tested.)

| | Plaque Size 40 °C/30° C | Growth Pattern | Multicore Particle |
|---|---|---|---|
| Toto1101 | Normal/Normal | Normal | No |
| ts103R | Small/Small | Slow | Yes |
| 103RD | Minute/Minute | Very Slow | N/T |
| 103RF | Small/Small | Slow | Yes |
| 103RE1 | Large/Large | Normal | No |

103D was plaque purified once and passed at low multiplicity in chick cells. This stock was plaqued at 40°C. More than 40,000 plaques were screened and three large plaques were obtained, a reversion frequency of $7 \times 10^{-5}$. This reversion frequency is consistent with a single nucleotide change being responsible for reversion. These three new revertant viruses, however, all differed in plaque size and titer produced from virus from Toto 1101, suggesting that they may all be second site revertants and that in each case the suppressor mutation is different. From this and from the original reversion frequency found for *ts*103, it appears that the *ts*103 mutation is very stable and that reversion to the wild-type nucleotide occurs at very low frequency.

## DISCUSSION

We have shown here that the *ts*103 phenotype is due to a C → U change at position 9661 of Sindbis RNA, leading to the substitution of alanine by valine at position 344 of glycoprotein E2. A C → U transition is consistent with the action of nitrous acid, in which C in the viral RNA is deaminated to produce U. A substitution of alanine by valine is normally considered to be conservative but in this case it leads to dramatic effects upon the function of glycoprotein E2 during virus assembly.

We had originally postulated (Strauss *et al.*, 1977) that the defect in *ts*103 might lie in the nucleocapsid protein because it was found that many nucleocapsids isolated from *ts*103-infected cells or from multi-cored virions sedimented more slowly that did those isolated from wild-type virions or from wild-type infected cells. More recent results from another laboratory, however, have shown that there is an immature form of the nucleocapsid that sediments more slowly, which upon maturation becomes a more rapidly sedimenting form (Coombs *et al.*, 1984). Thus, the finding here that the nucleocapsid protein of

*ts*103 is identical to that of wild-type virus suggests that the nucleocapsids in *ts*103 infected cells, or many of the nucleocapsids in multi-cored particles, are, in fact, immature forms that are not triggered to assume the mature form because of deficiencies in virus assembly.

It is believed that during virus assembly the cytoplasmic domains of the glycoprotein interact with the nucleocapsid to drive virus budding and to produce the precisely assembled virion. The fact that the *ts*103 mutation is in a domain of glycoprotein E2 external to the lipid bilayer means that the weakened interactions between the glycoproteins and the nucleocapsid proteins are not the results of changes in the domains of these two proteins that interact directly. Instead, it suggests that the E1-E2 heterodimer forms improperly in *ts*103, due either to a change in the conformation of E2 itself induced by the amino acid substitution at position 344 or due to steric hindrance, if this altered amino acid is a contact residue essential for heterodimerization. In either case, the changed conformation could affect assembly of the spikes (each made up of 3 dimers) rather than the assembly of the E1-E2 dimer itself. In any event, the mutant spike unit has a suboptimal geometry and fails to interact properly with the nucleocapsid, leading to defective assembly of virions.

We have also shown here that the mutation in *ts*103 can be partially suppressed by a change of Lys-227 to Met in glycoprotein E1. In some way this change in E1 can compensate in part for the change in E2. One possibility is that Ala-344 of E2 and Lys-227 of E1 are contact residues in E1-E2 dimer formation, and that the two changes partially compensate for one another. A second possibility is that the altered conformation of E2 induced by Val-344 is partially compensated by an altered conformation of E1 induced by the Met-227. A third possibility is that the change in E1 increases the stability of glycoprotein E1-E2 interactions or the affinity of the spike for the nucleocapsid during budding. We

favor this third possibility because of the finding that the change in E1, when separated from the *ts*103 mutation, leads to an increase in plaque size of the virus, suggesting that the change in E1 has a generalized effect rather than being specific for the *ts*103 mutation. Moreover, the *ts*103R mutation only moderates the *ts*103 phenotype rather than specifically suppressing it, suggesting a general effect.

## ACKNOWLEDGMENTS

This work was supported by NIH grants AI10793 and AI20612 and by NSF grant DMB8617372. C. M. R. is a Pew Scholar in the Biomedical Sciences.

## REFERENCES

Arias, C., Bell, J. R. Lenches, E. M., Strauss, E. G., and Strauss, J. H. (1983). Sequence analysis of two mutants of Sindbis virus defective in the intracellular transport of their glycoproteins. *J. Mol. Biol.* **168,** 87-102.

Coombs, K., Brown, B., and Brown, D. T. (1984). Evidence for a change in capsid morphology during Sindbis virus envelopment. *Virus Res.* **1,** 297-302.

Fuller, S. D. (1987). The T=4 envelope of Sindbis virus is organized by interactions with a complementary T=3 capsid. *Cell* **48,** 923-934.

Hahn, C. S., Strauss, E. G. and Strauss, J. H. (1985). Sequence analysis of three Sindbis virus mutants temperature-sensitive in the capsid protein autoprotease. *Proc. Natl. Acad. Sci. USA* **82,** 4648-4652.

Lindqvist, B. H., DiSalvo, J., Rice, C. M., Strauss, J. H., and Strauss, E. G. (1986). Sindbis virus mutant *ts*20 of complementation group E contains a lesion in glycoprotein E2. *Virology* **151,** 10-20.

Lustig, S., Jackson, A., Hahn, C. S., Griffin, D. E., Strauss, E. G., and Strauss, J.H. (1988). Molecular basis of Sindbis virus neurovirulence in mice. *J. Virol.*, in press.

Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982). Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y.

Maxam, A. M., and Gilbert, W. (1980). Sequencing end-labeled DNA with base specific chemical cleavage. *In* "Methods in Enzymology". (L. Grossman and K. Moldave, eds.) Vol. 65, 499-560, Academic Press, Orlando, Florida.

Okayama, H., and Berg, P. (1982). High-efficiency cloning of full-length cDNA. *Mol. Cell. Biol.* **2,** 161-170.

Pierce, J. S., Strauss, E. G., and Strauss, J. H. (1974). Effect of ionic strength on the binding of Sindbis virus to chick cells. *J. Virol.* **13,** 1030-1036.

Rice, C. M., Bell, J. R., Hunkapiller, M. W., Strauss, E. G., and Strauss, J. H. (1982). Isolation and characterization of the hydrophobic COOH-terminal domains of the Sindbis virus glycoproteins. *J. Mol. Biol.* **154**, 355-378.

Rice, C. M., Levis, R., Strauss, J. H., and Huang, H. V. (1987). Production of infectious RNA transcripts from Sindbis virus cDNA clones: Mapping of lethal mutations, rescue of a temperature-sensitive marker, and in vitro mutagenesis to generate defined mutants. *J. Virol.* **61**, 3809-3819.

Rice, C. M., Dalgarno, L., Galler, R., Hahn, Y. S., Strauss, E. G., and Strauss, J. H. (1988). Molecular cloning of flavivirus genomes for comparative analysis and expression. In: Modern Trends in Virology, Bauer, H., Klenk, H.-D. and Scholtissek, C. (Eds.), Proceedings of the International Symposium, June 1986, Giessen, Springer-Verlag, Berlin, in press.

Rice, C. M., and Strauss, J. H. (1981). Synthesis, cleavage, and sequence analysis of DNA complementary to the 26S messenger RNA of Sindbis virus J. Mol. Biol. **150**, 315-340.

Smith, D. R., and Calvo, J. M. (1980). Nucleotide sequence of the E. coli gene coding for dihydrofolate reductase. *Nucl. Acids Res.* **13**, 1717-1731.

Strauss, E. G., Lenches, E. M., and Strauss, J. H. (1976). Mutants of Sindbis virus I: Isolation and partial characterization of 89 new temperature-sensitive mutants. *Virology* **74**, 154-168.

Strauss, E. G., Birdwell, C. R., Lenches, E. M., Staples, S. E., and Strauss, J. H. (1977). Mutants of Sindbis virus. II. Characterization of a maturation-defective mutant, *ts*103. *Virology* **82**, 122-149.

Strauss, E. G. (1978). Mutants of Sindbis virus III. Host polypeptides present in purified HR and *ts*103 virus particles. *J. Virol.* **28**.

Strauss, E. G., Rice, C. M., and Strauss, J. H. (1984). Complete nucleotide sequence of the genomic RNA of Sindbis virus. *Virology* **133**, 92-110.

Zimmern D., and Kaesberg, P. (1978). 3' terminal sequence of encephalomyocarditis virus RNA determined by reverse transcriptase and chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **75,** 4257-4261.

# Chapter 4

Molecular Basis of Sindbis Virus Neurovirulence in Mice

This chapter is in press in the *Journal of Virology*.

# Molecular Basis of Sindbis Virus Neurovirulence in Mice

SHLOMO LUSTIG[1,3], ALAN C. JACKSON[2,4], CHANG S. HAHN[1], DIANE E. GRIFFIN[2], ELLEN G. STRAUSS[1], and JAMES H. STRAUSS[1]

[1]*Division of Biology, California Institute of Technology, Pasadena, California 91125*

[2]*Departments of Medicine and Neurology, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205*

[3]*Present address: Israel Institute for Biological Research, Ness-Ziona, 70450, P.O. Box 19, Israel*

[4]*Present address: Departments of Medicine and Microbiology and Immunology, Queen's University, Kingston, Ontario, Canada, K7L 3N6*

Running title: Sindbis virus neurovirulence

**Abstract**

We have examined a variety of strains of Sindbis virus for the genetic changes responsible for differences in neurovirulence for mice. A low passage of the AR339 strain of Sindbis virus (SV), a neuroadapted Sindbis virus (NSV), and two laboratory strains of Sindbis virus (HRSP and Toto1101) were examined. NSV causes a severe encephalomyelitis with hind limb paralysis and high mortality after intracerebral inoculation in weanling mice. In contrast, SV causes only mild, nonfatal, disease in weanling mice; however, in suckling mice SV causes a fatal encephalomyelitis after either intracerebral or subcutaneous inoculation. The two laboratory strains used have greatly reduced neurovirulence for suckling mice and are avirulent for weanling mice. The nucleotide sequences and encoded amino acid sequences of the structural glycoproteins of these four strains were compared. Hybrid genomes were constructed by replacing restriction fragments in a full-length cDNA clone of Sindbis virus, from which infectious RNA can be transcribed in vitro, with fragments from cDNA clones of the various strains. These recombinant viruses allowed us to test the importance of each amino acid difference between the various strains for neurovirulence in weanling mice and suckling mice. Glycoproteins E2 and E1 were of paramount importance for neurovirulence in adult mice. Recombinant viruses containing the nonstructural protein region and the capsid protein region from an avirulent strain and the E1 and E2 glycoprotein regions from NSV were virulent, although less virulent than NSV. Furthermore, changes in either E2 (His 55 in NSV to Gln in SV) or in E1 (Ala 72 in NSV to Val in SV and/or Asp 313 in NSV to Gly in SV) would reduce virulence. For virulence in suckling mice we have found that a number of changes in E2 and E1 can lead to decreased virulence and, in fact, a gradient of virulence exists.

**Introduction**

Sindbis virus is the type species of the genus Alphavirus in the family Togaviridae. It is among the least pathogenic of the alphaviruses, which include such important pathogens as Eastern, Western, and Venezuelan equine encephalitis viruses (10). In nature, Sindbis virus is transmitted by mosquitoes and its vertebrate hosts are usually birds or mammals (41). The viral genome is a plus stranded RNA of 11,703 nucleotides. In the virion it is complexed with a capsid protein C to form an icosahedral nucleocapsid that is surrounded by a lipid bilayer in which two integral membrane glycoproteins, E2 and E1, are embedded (reviewed in 40). Both E1 and E2 are anchored in the membrane by a conventional C-terminal hydrophobic anchor, and are closely associated with each another as a heterodimer that probably forms during synthesis of the glycoproteins (28,31). This heterodimer is believed to be the functional subunit, and three heterodimers are associated to form a trimeric spike in the virus (8). The spike is the structure that binds to susceptible cells to initiate infection, and also possesses neutralizing epitopes.

Sindbis virus infection of mice has been studied as an experimental model of acute viral encephalitis (9,14). Wild-type Sindbis virus (AR339 strain) causes a fatal encephalitis in suckling mice and a nonfatal encephalitis in 4 week old weanling mice (15). A strain of Sindbis virus that is highly lethal for weanling and adult mice was isolated after six intracerebral (IC) passages of wild-type Sindbis strain AR339 (SV) alternating between suckling and weanling mice (11). This neuroadapted strain of Sindbis (NSV) has been genetically stable after many passages in cell culture. NSV replicates 10-50 fold more efficiently in the brain of mice than SV and its IC 50% lethal dose is 2-20 pfu in weanling mice. After IC inoculation with NSV weanling mice become ruffled, develop kyphoscoliosis and hind limb paralysis and have a high mortality (11,12). There is a poliomyelitis,

particularly involving the ventral horns, in the thoracic and lumbar spinal cord (12). Stanley et al. (36) demonstrated that some anti-Sindbis virus E1 and E2 monoclonal antibodies discriminate between NSV and SV, suggesting that changes in the surface glycoproteins may be associated with changes in virulence.

Olmsted et al. (21,22) have selected Sindbis virus variants *in vitro* that are attenuated in suckling mice. Thus, strains of Sindbis virus can be essentially avirulent for mice of all ages, or virulent for suckling mice but avirulent for weanling mice, or virulent for mice of all ages.

Recently a complete cDNA clone of Sindbis virus has been constructed from which infectious RNA can be transcribed *in vitro* with SP6 RNA polymerase (29). This has made it possible to map changes between the various strains of Sindbis virus which lead to differences in virulence in mice, an approach similar to that used by Kohara et al. (16) to map virulence determinants in poliovirus.

## MATERIALS AND METHODS

**Virus stocks, preparation of virus and viral RNA.** Sindbis virus wild-type AR339 strain (a low passage stock obtained from the American Type Culture Collection and referred to herein as SV) and neuroadapted Sindbis virus (NSV) were grown and titred on monolayers of primary and secondary chick embryo fibroblasts (CEF) as previously described (37). Virus was precipitated from the culture fluid with polyethylene glycol and purified by successive velocity sedimentation and equilbrium density centrifugation (24). The isolation of viral RNA was as described (30).

**cDNA cloning.** cDNA synthesis followed essentially the procedure of Okayama and Berg (20) as modified by Lindqvist et al. (17). A plasmid vector referred to as proteus 1 was tailed with dT (40-60 residues) and used as primer for first strand cDNA synthesis. Second strand synthesis then used *E. coli* DNA

polymerase I, *E. coli* RNase H and *E. coli* DNA ligase. After second strand synthesis the double stranded DNA was cut with HindIII and ligated intramolecularly. The resulting clones have an insert of 5438 nucleotides [plus a variable length of poly(A)] and contain the entire structural protein coding region of Sindbis virus.

**General recombinant DNA techniques.** Restriction endonucleases and DNA modifying enzymes were purchased from commercial sources and used essentially as recommended by the manufacturer. Plasmids were grown, purified and analyzed using standard methods with minor modifications (18).

**Construction of hybrid genomes.** Hybrid genomes were produced by replacing restriction fragments in Sindbis clone Toto1101 or Toto50 (29) with the corresponding regions from cDNA clones derived from NSV, with cDNA from strain SV, or with cDNA from a clone of Sindbis virus HRSP strain. Details of the restriction sites used are included in the figure legends. Virus was rescued from these recombinant clones and tested for its biological properties.

**In vitro transcription and RNA transfection.** RNA transcripts were synthesized as described (29). Briefly, either supercoiled plasmid templates or plasmid DNAs which had been digested with an appropriate restriction endonuclease to produce a run-off transcript were transcribed *in vitro* by SP6 RNA polymerase in 40 mM Tris-Cl, pH 7.6, 6 mM MgCl, 2 mM spermidine, 1 mM ATP, CTP, GTP, UTP, 100 μg/ml nuclease-free bovine serum albumin, 5 mM DTT, 1 mM $m^7G(5')ppp(5')G$ cap analogue, 500 U/ml human placental RNase inhibitor, 400 U/ml SP6 polymerase, and 10-100 μg/ml template DNA. Quantitation of the RNA transcript was effected by including a trace amount of [$^{32}$P] CTP in the reaction and counting an aliquot of the product after adsorption to Whatman DE81 paper. Confluent monolayers of secondary CEF in 35 mm tissue culture plates (about $10^6$ cells) were transfected with the resulting RNA. After washing once

with Eagle's minimal essential medium (7) containing Earles salts but without serum, the cells were incubated with 1.5 ml Eagle's medium containing 50 mM Tris-Cl, pH 7.3 at 25°C, and 200 μg/ml DEAE dextran (500,000 average molecular weight from Sigma) for 60 min at 37°C. This medium was removed and 200 μl of *in vitro*-transcribed RNA diluted in phosphate buffered saline was added to the cells and incubated at room temperature for 60 min with occasional rocking. Plaque-forming units were quantitated by overlaying the monolayer with 3 ml of 1.2% agarose (SeaKem ME from FMC) in Eagle's medium containing 2% fetal calf serum, followed by incubation at 30°C or 40°C for 3 days or 2 days, respectively. Plaques were visualized by staining with neutral red. Virus stocks were produced by removing the transfection mix and incubating the cells in 3 ml of Eagle's medium containing 3% fetal calf serum for 48 hr or longer at 30°C.

**Sequence determination of SV and NSV cDNA.** DNA sequencing was carried out on cloned cDNA using the methods of Maxam and Gilbert (19) as modified by Smith and Calvo (35), using restriction fragments 3' end labeled with the Klenow fragment of *E. coli* DNA polymerase (Bethesda Research Laboratories). NSV clones NSV5-37 and NSV5-44 and SV clones SV1A-3 and SV1A-21 were analyzed. Chain termination sequencing of RNA with reverse transcriptase as described by Ou et al. (23) and Zimmern and Kaesberg (42), using intracellular viral RNA as template and synthetic oligonucleotides as primers, was also used to obtain part of the sequence. To test that virus derived from hybrid cDNAs had the predicted sequence, certain regions of the RNA genome sequence were confirmed by chain termination sequencing in the same way. All of the sequencing was obtained on two clones, or checked by the dideoxy sequencing of an RNA template (which gives the majority nucleotide at any position), in order to rule out cloning artifacts or sequencing of minor variants.

**Animal tests.** Viruses for animal inoculation were grown and assayed by plaque formation on BHK-21 cells. Three- to four-week old BALB/cJ weanling mice of either sex (Jackson Laboratories, Bar Harbor, ME) were inoculated intracerebrally (IC) with 1000 pfu of virus in 0.03 ml Hanks balanced salt solution containing 1% fetal calf serum. One- to three-day old CD-1 (ICR) suckling mice (Charles River Breeding Laboratories, Wilmington, MA) were inoculated subcutaneously with 500 pfu of virus in each hindlimb (footpad). Mice were observed for 25 days. The percent mortality was determined for weanling mice and the percent mortality and mean day of death determined for suckling mice. At least 10 weanling mice and two litters of suckling mice were inoculated with each virus strain. For recombinant strains, viruses grown from 2-4 different plaques were tested separately. These data have been pooled for presentation. For viruses which killed weanling mice at 1000 pfu, the intracerebral dose for 50% mortality ($ICLD_{50}$) was determined by the method of Reed and Muench (26) using groups of 10 mice at serial 10-fold dilutions.

## RESULTS

**Sindbis virus strains.** Four strains of Sindbis virus were used in these studies as well as recombinants between various strains (Table 1). The passage history of these strains significantly affects the interpretation of the results. The AR339 strain of Sindbis, here designated as SV, is a low passage stock (7-9 passages in suckling mouse brain) received from the American Type Culture Collection and subsequently passaged five times in BHK cells (including three sequential plaque purifications) and twice in CEF. The AR339 strain that served as the parent for NSV had a more extensive passage history. It was obtained originally from Dr. H. Hineberg (Cleveland Metropolitan General Hospital) after an unknown passage history, and subsequently passed ten times in mouse brain,

TABLE 1 *Sindbis virus strains*[a]

| Virus Strain | Passage History[b] | Virulence | |
|---|---|---|---|
| | | Suckling Mice | Weanling Mice |
| SV | 9 SMB | + | - |
| | 5 BHK | | |
| | 2 CEF | | |
| NSV | Unknown passages in | | |
| | cell culture | + | + |
| | 10 MB | | |
| | 1 BHK | | |
| | 10 CEF | | |
| | 3 + 3 SMB + MB | | |
| | 2 CEF | | |
| Toto1101 | Extensive passage in | | |
| | cell culture | ± | - |
| HRSP | Extensive passage in | | |
| | cell culture | ± | - |

[a]All strains were derived by passage of the AR339 strain, isolated originally from a pool of Culex univittatus in August, 1952 in Sindbis, Egypt by inoculation into suckling mouse brain (41).

[b]SMB = suckling mouse brain; MB = mouse brain; BHK = BHK21 cells; CEF = Chick embryo fibroblasts.

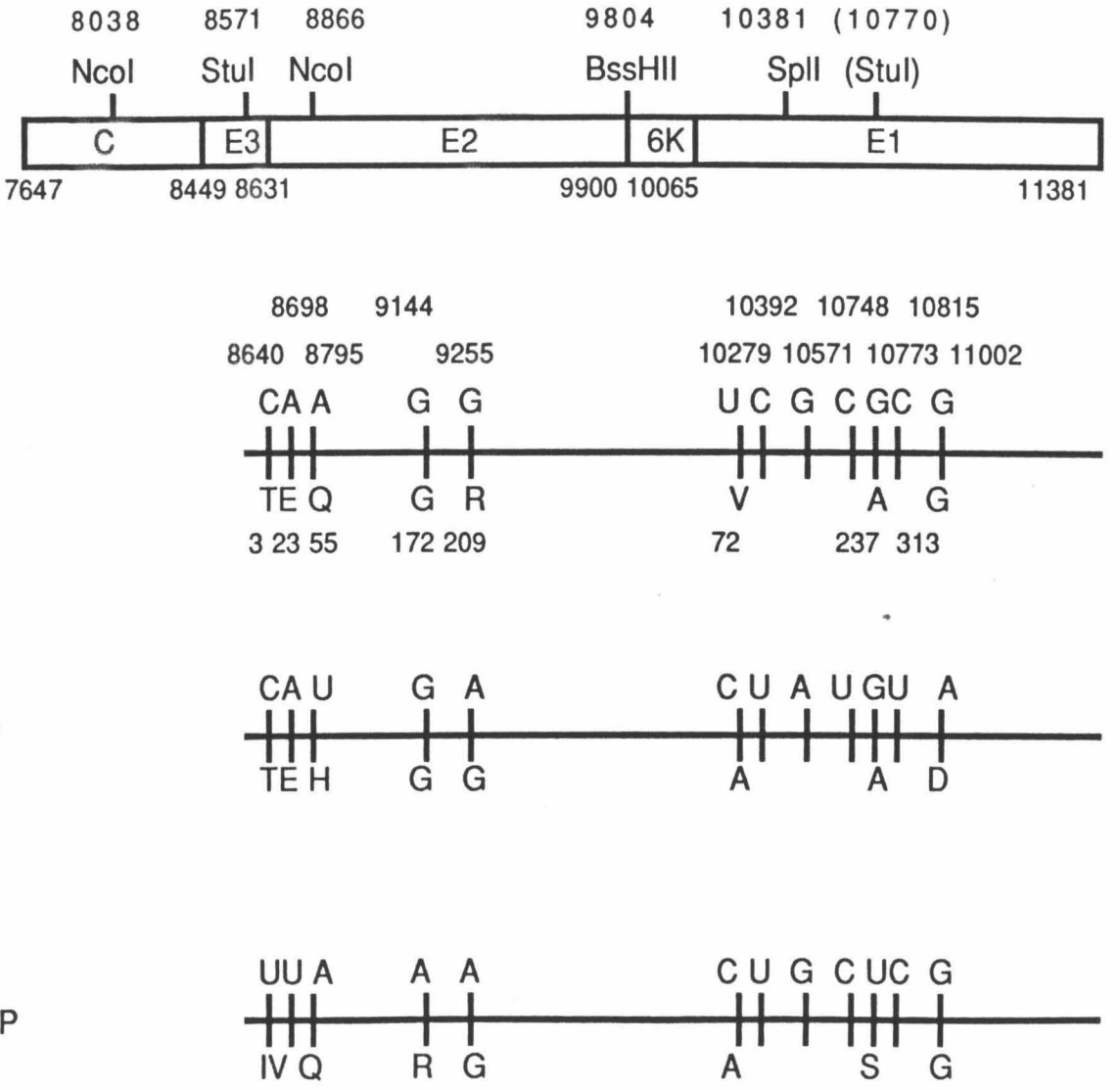
once in BHK cells, and more than ten times in CEF prior to neuroadaption (15). NSV was obtained from this AR339 strain by serial transmission six times by IC inoculation alternately in suckling and weanling mice (11). This virus was subsequently passed twice in CEF in order to isolate viral RNA for sequence analysis.

In addition, two laboratory strains of Sindbis virus derived from infectious RNA rescued from cDNA clones (29) were used; both were derived from the AR339 strain. The first of these laboratory variants was rescued from cDNA clone Toto1101 and contains nucleotides 1-2713 and 9805-11703 from the HRSP strain described below, and nucleotides 2714-9804 from a strain of uncertain passage history prior to cloning as Toto1101. The second strain, from clone Toto50, was derived from cDNA made to the HRSP strain of Sindbis virus (heat resistant small plaque strain). The HR strain was obtained by Burge and Pfefferkorn (5) from AR339 by selecting variants able to survive heating to 56°C. It had been subsequently passaged multiple times in both CEF and BHK cells before isolation of the small plaque variant (39). HRSP was then passaged several times in CEF prior to cloning as Toto50. The sequence of the HRSP strain has been published (29,38).

**Sequence analysis of the glycoproteins of NSV and of SV.** Because monoclonal antibodies discriminate between NSV and SV (36) and attenuating mutations have been found in E2 (6,22), it seemed likely that the glycoproteins would be important for neurovirulence in Sindbis virus. We therefore obtained the nucleotide sequence of the region encoding the glycoproteins, and from this the deduced amino acid sequence of glycoproteins E3, E2 and E1, for both SV and NSV, and compared this sequence with that previously published for HRSP (Fig. 1).

In the glycoprotein E3 region sequenced there were no nucleotide differences among the various strains, SV, NSV, or HRSP. In glycoprotein E2,

**Fig. 1.** Sequence differences between SV, NSV, and HRSP. Above is shown a schematic of the structural region of the Sindbis virus genome. The positions of a number of restriction enzyme sites are shown, as well as the boundaries between the different proteins encoded, together with their coordinates numbered from the 5' end of the RNA according to Strauss et al. (38). Below are shown sequencing schematics of three Sindbis strains. Where differences between any of the strains occur, nucleotides are shown above the line (numbered from the 5' end as before) and encoded amino acids (if a change in coding is involved) are shown below the line, using the single letter amino acid code. The amino acid numbers refer to the position within the glycoprotein; thus Q55 is residue 55 of glycoprotein E2, etc. The sequence from nucleotide 8571 in E3 to the end of the RNA was determined for SV and for NSV, as described in the Materials and Methods. The sequence for HRSP is taken from Strauss et al. (38).


8038
8571
8866
9804
10381 (10770)
NcoI
StuI
NcoI
BssHII
SpII (StuI)
C
E3
E2
6K
E1
7647
8449 8631
9900 10065
11381
8698
9144
10392 10748 10815
8640 8795
9255
10279 10571 10773 11002
SV
CA A
G G
U C G C GC G
TE Q
G R
V
A G
3 23 55
172 209
72
237 313
NSV
CA U
G A
C U A U GU A
TE H
G G
A
A D
HRSP
UU A
A A
C U G C UC G
IV Q
R G
A
S G

however, there were two nucleotide changes between SV and NSV, both of which led to amino acid substitutions. Nucleotide 8795 is A in SV and U in NSV, resulting in a Gln to His change at position 55 of E2. Nucleotide 9255 is G in SV and A in NSV, leading to the replacement of Arg in SV by Gly in NSV at position 209 of E2. Amino acid 55 is Gln in the HRSP strain (as is found in SV), while amino acid 209 is Gly in HRSP (as is the case for NSV).

NSV and SV differ from HRSP in three additional amino acids of E2. Position 3 of NSV and SV is Thr, whereas HRSP has Ile at this position. Position 23 of SV and NSV is Glu and is Val in HRSP [Glu 23 was previously shown to be the ancestral amino acid in HR (1)]. Finally, position 172 of E2 is Gly in SV and NSV and is Arg in HRSP. The changes at positions 3 and 172 may have arisen during selection of the HR strain.

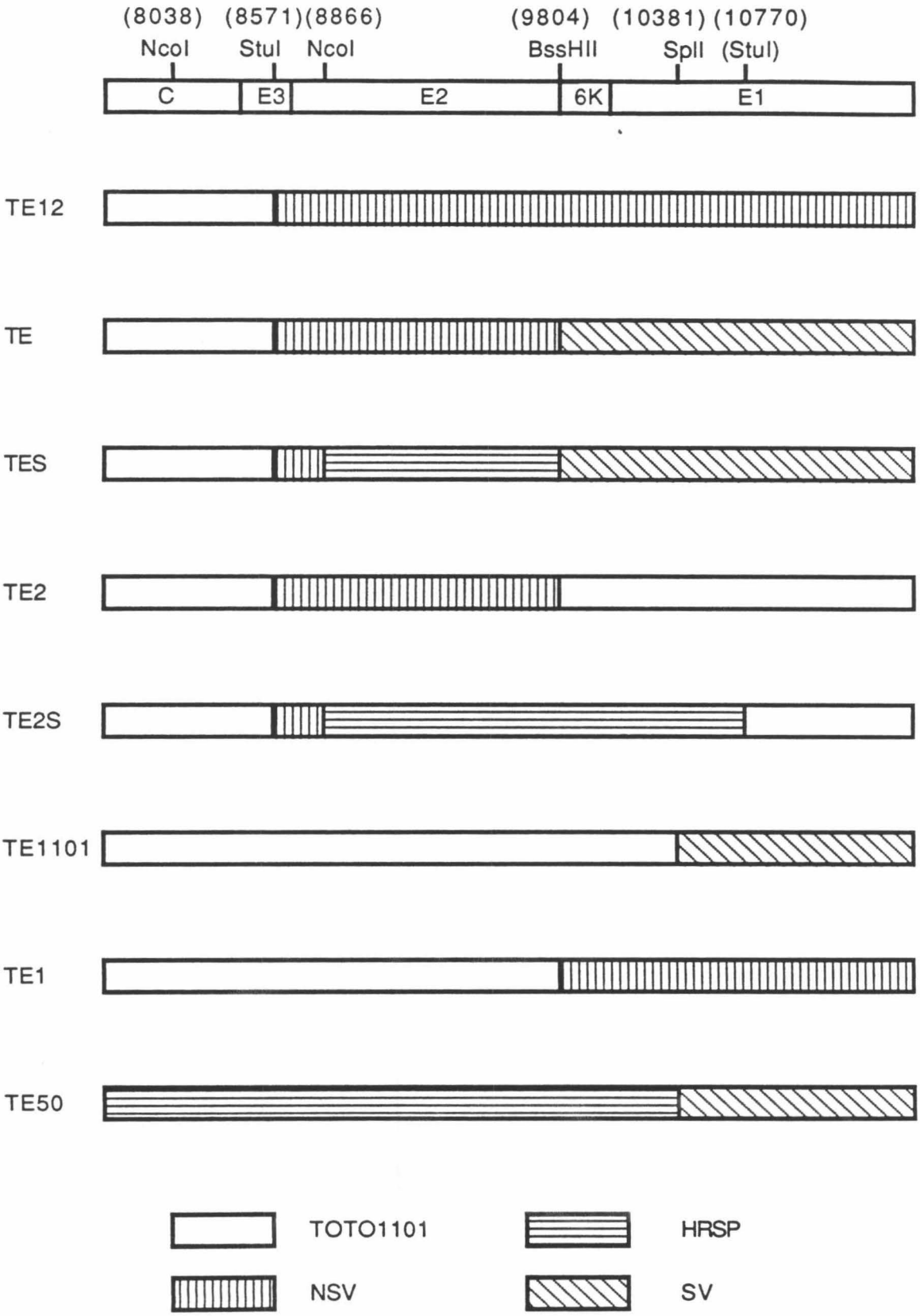
In glycoprotein E1 there were six nucleotide differences between SV and NSV leading to two amino acid changes. Silent changes were found at positions 10392 (C in SV → U in NSV), 10571 (A → G), 10748 (C → U), and 10815 (C → U). The coding changes were at position 10279 (U in SV → C in NSV) leading to substitution of Val in SV by Ala in NSV at position 72 of E1, and nucleotide 11002 (G → A) leading to the substitution of Gly by Asp at position 313. Position 72 is Ala in HRSP (as in NSV), whereas amino acid 313 is Gly in HRSP (as in SV). In addition to these changes SV and NSV also differ at position 237 (Ala) from HRSP (Ser); this change may also have arisen during selection of the HR strain.

Of the eight nucleotide differences between SV and NSV seven were transitions (three C → U, three G → A, and one U → C). The transversion was the A → U change leading to the substitution of Gln by His.

There were no changes between SV and NSV in the 3' noncoding region.

**Construction of recombinant viruses.** In order to examine the effect of the different amino acid substitutions on neurovirulence of Sindbis virus, a number of

**Fig. 2.** Construction of Sindbis virus genomes recombinant in the glycoprotein region. Restriction fragments in clone Toto1101 (29) were replaced by the corresponding fragments from three other strains of virus, NSV, SV, or HRSP (derived from Toto50), as indicated in the diagram. The restriction sites used and their coordinates in the viral genome are indicated.


(8038)
NcoI
(8571)
StuI
(8866)
NcoI
(9804)
BssHII
(10381)
SpII
(10770)
(StuI)
C
E3
E2
6K
E1
TE12
TE
TES
TE2
TE2S
TE1101
TE1
TE50
TOTO1101
HRSP
NSV
SV

recombinant viruses were constructed. To accomplish this, restriction fragments from clones of SV, NSV, or HRSP cDNA were used to replace the corresponding restriction fragments in clone Toto1101 or, in one case, Toto50. These clones contain a complete cDNA copy of Sindbis virus inserted downstream from an SP6 RNA polymerase promoter from which infectious RNA can be transcribed *in vitro* (29). RNA was transcribed from the recombinant clones and transfected onto CEF cells. Monolayers were incubated at 30°C under liquid medium in order to rescue infectious virus or under agarose at 30°C and 40°C for plaque assays. None of the recombinant viruses were temperature sensitive. The titres of the rescued viruses were between $10^8$ and $10^9$ pfu/ml. They were passaged once in BHK cells and then examined for their biological properties.

The constructs tested are illustrated schematically in Fig. 2. The restriction sites used to construct the hybrid genomes and their coordinates numbered from the 5' end of the RNA (38) are also shown. The StuI site at 10770 is not present in all of the virus strains. Those with Ser at position 237 of E1 possess this site, whereas those that have Ala at this site do not. In all cases the recombinants possess the nonstructural protein region and the capsid protein region from Toto1101 or from Toto50 and only the glycoprotein region or portions of the glycoprotein region are derived from other strains.

**Neurovirulence of recombinant viruses for weanling mice.** The neurovirulence of each of the recombinant viruses, the parental viruses NSV, SV, HRSP, and viruses rescued from Toto50 and Toto1101, were tested in 3-4 week old weanling mice, using 1000 pfu of each virus by IC inoculation. In each case two or more independent virus stocks were tested. The results were pooled and the relevant data are summarized in Table 2.

Only NSV and the recombinant TE12 were virulent for weanling mice. TE12 has the nucleotide sequence from 8571 to 11552 of NSV and nucleotides 1 to

**TABLE 2** *Neurovirulence of Sindbis strains in weanling mice*[a]

| Virus Strain | Dead/Total | % Mortality |
|---|---|---|
| NSV | 21/21 | 100 |
| TE12 | 15/35 | 44 |
| TE | 2/55 | 4 |
| TES | 0/20 | 0 |
| TE2 | 0/20 | 0 |
| TE2S | 0/30 | 0 |
| TE1 | 0/21 | 0 |
| SV | 0/20 | 0 |
| Toto1101 | 0/10 | 0 |
| Toto50 | 0/20 | 0 |

[a]Challenged with 1000 pfu by IC inoculation.

8570 and 11553 to 11703 from Toto1101, that is, glycoproteins E2 and E1 from NSV and the remainder of the genome from Toto1101 (Fig. 2). Toto1101 is avirulent for weanling mice whereas TE12 is virulent, demonstrating that the envelope proteins alone are sufficient to confer virulence to an otherwise avirulent strain. However TE12 has a higher LD50 (50 versus 3, calculated by the method of Reed and Muench [26]), than NSV and the overall mortality is reduced (Table 2), suggesting that nucleotide sequences outside the envelope protein region are also important for neurovirulence. These sequences are presumably in the region encoding the replicase proteins, thereby affecting the efficiency of virus replication, although changes in the capsid protein or in noncoding regulatory sequences cannot be ruled out.

All of the other strains are avirulent in weanling mice. Thus strains containing only glycoprotein E1 or only E2 from NSV are avirulent, demonstrating that changes in both E1 and E2 were necessary for the transition in neurovirulence from SV to NSV. The significance of these findings in terms of amino acid substitutions will be discussed below.

**Neurovirulence of recombinant viruses for suckling mice.** Our major interest in these studies was to define the changes that led to attenuation in weanling mice. Sindbis virus has also been used as a model system for the study of neurovirulence in suckling mice, however (6,21,22,25), and it seemed of interest to compare the virulence of these constructs in suckling mice as well (Table 3). Strains NSV and TE12 were essentially indistinguishable in their neurovirulence (although the mean day of death may be slightly extended in TE12). Thus although Toto1101 is much less virulent for suckling mice than NSV, possession of the glycoproteins from NSV is sufficient to render Toto1101 virus fully virulent for suckling mice. Strains that contain E2 from NSV and E1 from NSV, SV, or Toto1101 were also neurovirulent in suckling mice, causing 100% mortality after

**TABLE 3** *Neurovirulence of Sindbis strains in suckling mice*[a]

| Virus Strain[b] | Dead/Total | % Mortality | MDOD[c] |
|---|---|---|---|
| NSV | 31/31 | 100 | 3.3 |
| TE | 77/77 | 100 | 3.6 |
| TE12 | 20/20 | 100 | 3.9 |
| SV | 20/20 | 100 | 4.5 |
| TES | 47/47 | 100 | 4.8 |
| TE2 | 20/20 | 100 | 5.2 |
| TE2S | 31/36 | 86 | 8.3 |
| TE1101[d] | 31/44 | 70 | 9.4 |
| TE1 | 27/40 | 68 | 9.9 |
| Toto1101 | 33/66 | 50 | 8.1 |
| Toto50(HRSP) | 21/49 | 41 | 10.8 |
| TE50[d] | 4/21 | 19 | 12.8 |

[a]Challenged with 1000 pfu by SC inoculation at 1-3 days of age.

[b]Described in Fig. 2. At least two independent stocks were tested and the results averaged.

[c]MDOD is mean day of death after subcutaneous injection of 1000 pfu.

[d]The neurovirulence of these constructs is dependent on the exact age of the mice and drops from 100% mortality in 1-2 day old mice to 20% in 3-4 day old mice (see text).

**Fig. 3.** Amino acid differences among Sindbis strains. Amino acid differences in the E2 and E1 glycoprotein region among the parental strains and various recombinant strains of Sindbis virus are indicated using the one letter amino acid code. The amino acid position in E2 or in E1, respectively, is shown. Strains are listed in order of decreasing virulence. The sources of the sequences are described in the text.

| STRAIN | E2 | | | | | | E1 | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 3 | 23 | 55 | 172 | 209 | 251 | 72 | 75 | 237 | 313 |
| NSV | T | E | H | G | G | A | A | D | A | D |
| TE12 | T | E | H | G | G | A | A | D | A | D |
| TE | T | E | H | G | G | A | V | D | A | G |
| SV | T | E | Q | G | R | A | V | D | A | G |
| TES | T | E | H | R | G | A | V | D | A | G |
| TE2 | T | E | H | G | G | A | A | G | S | G |
| TE2S | T | E | H | R | G | A | A | D | S | G |
| TE1101 | I | E | Q | G | G | V | A | G | A | G |
| TE1 | I | E | Q | G | G | V | A | D | A | D |
| TOTO1101 | I | E | Q | G | G | V | A | G | S | G |
| TOTO50 | I | V | Q | R | G | A | A | D | S | G |
| TE50 | I | V | Q | R | G | A | A | D | A | G |

subcutaneous inoculation, as did a construct containing only the N-terminal domain of E2 from NSV with the remainder of E2 from HRSP and E1 from SV (construct TES). The survival time of the animals varied from 3 to 5 days depending upon the construct tested. When E2 was derived from Toto1101 (construct TE1 or virus from Toto1101) or Toto50 (virus HRSP) the resultant virus was less virulent. With these strains 30 to 60% of the inoculated mice survived, and those that died had an extended survival time. The construct TE50 will be discussed below. Thus there is a gradient of neurovirulence extending from a mortality of 100% in suckling mice with a survival time of 3 days to a mortality of 20-40% with a 12 day survival time. The significance of these findings with regard to individual amino acid substitutions is described in the following section.

**Correlation of amino acid changes with neurovirulence.** The amino acid changes among the various strains of Sindbis virus in the E2 and E1 glycoproteins that have been assayed for neurovirulence have been summarized in Fig. 3, with the strains being listed in descending order of neurovirulence. Sequence data for NSV and SV or for constructs containing sequence from these viruses are from Fig. 1, the Toto50 sequence is the HRSP sequence of Strauss et al. (38), while the data for Toto1101 are from Polo et al. (25) and R. E. Johnston (personal communication).

For weanling mice, as noted above, changes in both E2 and E1 contribute to the differences seen between SV and NSV. In E2 the change at position 55 from His (in NSV) to Gln (in SV) appears to be primarily responsible for attenuation (compare NSV and construct TE12 with SV and construct TE1) although the changes from Thr to Ile at position 3 and Gly to Arg at position 209 cannot be independently evaluated from these constructs. In glycoprotein E1 either the change Ala to Val at position 72 or the change Asp to Gly at 313, or both, appear to attenuate the virus (compare NSV and construct TE12 with SV and construct TE).

The situation for neurovirulence in suckling mice is more complex. For one, as noted above, the different strains do not exhibit an all or none effect. For another, it is difficult to exclude cooperative interactions between the different changes found. With these provisos we can draw a number of conclusions, however. Firstly, as pointed out earlier, only changes in glycoproteins E1 and E2 lead to attenuation in the constructs tested. Secondly, in glycoprotein E2 Gln 55 appears to be required for attenuation, but is not sufficient. Attenuated E2's also have Ile 3, Arg 172, and/or Gly 209. The Gly to Arg change at 172 does lead to a slight increase in survival time when combined with His 55 (compare TE with TES and TE2 with TE2S) and could be involved in attenuation. The Thr to Ile at position 3 could also be attenuating from the data presented here, although Polo et al. (1988) have argued that it is not involved in attenuation. The change at position 209 (as well as the changes at positions 23 and 251) do not appear to be involved in attenuation. Thirdly, in glycoprotein E1 two changes appear to be important in this group of constructs, the Ala to Val change at position 72 and the Ala to Ser at 237. Comparing TE to TE2 or TES to TE2S, it is clear that the two changes together lead to decreased mortality and extended survival.

In order to separate the two changes in E1, the constructs TE1101 and TE50 were made. In initial experiments TE1101 demonstrated an intermediate virulence, with 70% of the suckling mice dying and mean day of death 9.4 days (Table 3). However, the results were variable from litter to litter, in contrast to the results obtained with other constructs, and it appeared that this virus might be more age dependent. Further experiments to test this revealed that results with TE1101 were, in fact, strongly age-dependent. In 1-2 day old mice the mortality was 100% (two litters tested, mean day of death 5.7 days). In slightly older mice, 3-4 days old, mortality was 20% (2 of 10 mice in two litters died with mean day of death 14.5 days). Similarly, results with construct TE50 were age dependent in

the same way. In 1-2 day old mice 10/10 mice died (two litters) with mean day of death 10.2 days, in 3-4 da old mice 2/10 died (mean day of death 10 days). Similar experiments with Toto50 and Toto1101 exhibited no age dependence between 1-2 day old mice and 3-4 day old mice. These results taken together suggest that the Ala to Ser change at 237 is somewhat attenuating in suckling mice, but that the Ala to Val change at 72 may also contribute. The Asp to Gly at 75 does not appear to make a significant contribution to attenuation in this system and the effect of the Asp to Gly at 75 is unclear.

## DISCUSSION

The AR339 strain of Sindbis virus demonstrates age-dependent virulence in mice. Mice up to 8 days of age develop an acute encephalitis that is fatal; older mice also develop encephalitis, but the infection is not fatal (15,27). The virus grows to higher titer in the brain and other tissues of suckling mice than in older mice. The AR339 strain was originally isolated by inoculation of suckling mice and passaged several times in this host (41), which may have been partially responsible for the neurovirulence characteristics of the strain. Starting from AR339, more virulent strains have been isolated which are neurovirulent in weanling mice by alternate passage IC in suckling and weanling mice (11); presumably the passage in weanling mice selected for virus able to replicate better in the central nervous system of older mice. More attenuated strains of Sindbis virus have also been obtained from the AR339 strain. Olmsted et al. (21,22) have used several selection procedures, including selection for rapid penetration and growth *in vitro*, to isolate variants that were attenuated in suckling mice (see also 25). Barrett and Atkins (2) have shown that temperature-sensitive mutants are often attenuated. In this paper and in Polo et al. (25) it is also shown that laboratory strains represented by HRSP (and Toto50) and Toto1101

are attenuated for suckling mice. These strains have been maintained by extensive passage in cell culture, usually chick cells or BHK cells (and the HR strains had been selected for the ability to tolerate high temperatures). Passage of virus in cell culture is a classical way to select attenuated virus for use as vaccines, and appears in this case also to have selected variants which are reduced in virulence for mice. We should note, however, that the virulence properties of NSV and of SV have been maintained upon passage in cell culture, although an extensive series of passages has not been tested for its effects.

By constructing recombinant viruses among strains that differ in virulence we have mapped a number of virulence determinants. The situation in weanling mice seems clearcut. There are determinants in both glycoproteins E2 and E1 that lead to attenuation, and determinants in the non-glycoprotein regions as well. In glycoprotein E2 change of His to Gln at position 55 is attenuating, and in glycoprotein E1 change of Ala 72 to Val and/or change of Asp 313 to Gly lead to attenuation. The results with suckling mice were more confusing. As noted, it appears that a spectrum of neurovirulence exists with different virus strains, and in at least some cases the virulence is sharply age dependent. In addition, changes at multiple sites appear to have unpredictable effects, and thus the genetic background of the virus is important. Finally, in comparing our results to those of Johnston and colleagues (6,21,22,25), it also appears that there may be differences that depend upon the strain of mouse. Although we have tested a large number of constructs, it would require testing many more, comparing the changes in different combinations, to completely resolve the ambiguities. However, our results implicate His 55 and Gly 172 (and possibly Thr 3) in E2 and Ala 72 and Ala 237 (and possibly Asp 313) in E1 as involved in neurovirulence in suckling mice (the situation with E1 is peculiar; one interpretation of the results is that forms of E1 virulent for weanling mice are less virulent for suckling mice and vice versa).

Davis et al. (6) have previously shown the change of Ser to Arg at position 114 of E2 is attenuating. The fact that changes in the glycoproteins can attenuate the virus is consistent with studies of several other virus systems in which structural proteins have been implicated as determinants of virulence (3,4,32,33).

The concept of neurovirulence is complex. The efficiency of replication of the virus in peripheral tissues, the efficiency of crossing the blood-brain barrier, and the efficiency of replication in the central nervous system once the virus invades, are all of importance. There appear to be many ways to change a virulent virus so that it becomes less virulent. Changes in the replicase so that the virus replicates less rapidly may do so. Changes in the glycoproteins may affect speed of virus penetration or maturation, thus affecting the growth rate, or may affect tissue tropism and thus the ability of the virus to invade certain tissues such as the central nervous system.

One simple hypothesis to explain the results here is that the changes in the surface glycoproteins that lead to attenuation result in changes in the affinity of the virus for receptors on uninfected cells, thus altering its cell tropism. The effect need not be all or none. If altered virus bound to receptors in certain cells or tissues with reduced affinity, or if the alteration produced virus that bound to a different class of receptors with variable expression in different tissues, the alteration in tissue tropism or cell type preferences could be a relative one affecting the kinetics of virus replication in different organs, allowing host defenses to clear the infection and prevent death.

It is not known whether the changes in the Sindbis glycoproteins leading to differences in neurovirulence for mice lead to differences in attachment to cellular receptors. Smith and Tignor (34) compared the binding of two different field isolates of Sindbis virus differing in virulence and found that neuroblastoma cells (N-18) as well as non-neural cells (CER) had increased numbers of receptors

for the virulent AR86 strain compared to the avirulent AR339 strain. There were postulated to be distinct receptors for the two strains based on sensitivity to enzymatic degradation. Whether the same would be true for the more closely related strains used in this study (both derived from AR339) is not known. Comparative studies of the replication of NSV and SV in the brains and spinal cords of intracerebrally inoculated weanling mice have shown that both viruses infect the central nervous system and that the target cells for replication are the same, primarily neurons and ependymal cells (13). However, more virus is produced and neurons show greater injury after infection with NSV than SV. Whether this is due to a decrease in the effective concentration of receptors for SV, leading to slower spread of virus, or whether both SV and NSV recognize the same receptors on the same cells but subsequent steps in NSV replication are more efficient, requires further study. The surface glycoproteins of alphaviruses are also important in other steps of virus replication such as penetration, fusion with lysosome or phagosome membranes to release the RNA, and for interaction with the capsid protein during virion maturation. Many of these steps involve conformational changes in the E1-E2 heterodimer which might be affected by changes in the amino acid structure of these glycoproteins.

The fact that there are a variety of changes in E1 and E2 that affect neurovirulence, and that the effects of each change are different, has obvious implications for vaccine development. Classically one of the problems encountered during virus passage to develop attenuated strains has been overattenuation such that the vaccine strain is no longer efficacious. On the other hand multiple attenuating mutations are desirable so that the frequency of reversion to virulence is negligible. The ability to test the effect of individual attenuating changes and to mix them at will in recombinant strains in the approach used here could be of great value if applied to other viruses for which a vaccine is desired.

Acknowledgements

We thank Y. S. Hahn for participating in the rescue of some of the virus constructs, E. M. Lenches for expert technical assistance, and Dr. R. Johnston and colleagues for sharing their results prior to publication. This work was supported by NIH grants AI 10793 and AI 20612 (to JHS) and by NIH grant NS 18596 (to DEG). ACJ is the recipient of a postdoctoral fellowship from the Multiple Sclerosis Society of Canada.

## LITERATURE CITED


1. Arias, C., Bell, J. R., Lenches, E. M., Strauss, E. G. and Strauss, J. H. 1983. Sequence analysis of two mutants of Sindbis virus defective in the intracellular transport of their glycoproteins. J. Mol. Biol. 168, 87-102.
2. Barrett, P. N. and Atkins, G. J. 1979. Virulence of temperature-sensitive mutants of Sindbis virus in neonatal mice. Infec. Immunity 26, 848-852.
3. Bassel-Duby, R., Spriggs, D. R., Tyler, K. and Fields, B. N. 1986. Identification of attenuating mutations on the reovirus type 3 S1 double-stranded RNA segment with a rapid sequencing technique. J. Virol. 60, 64-67.
4. Beaty, B. J., Holterman, M., Tabachnick, W., Shope, R. E., Rozhon, E. J. and Bishop, D. H. L. 1981. Molecular basis of bunyavirus transmission by mosquitoes. The role of the M RNA segment. Science 211, 1433-1435.
5. Burge, B. W. and Pfefferkorn, E. R. 1966. Isolation and characterization of conditional-lethal mutants of Sindbis virus. Virology 30, 204-213.
6. Davis, N. L., Fuller, F. J., Dougherty, W. G., Olmsted, R. A. and Johnston, R. E. 1986. A single nucleotide change in the E2 glycoprotein gene of Sindbis virus affects penetration rate in cell culture and virulence in neonatal mice. Proc. Natl. Acad. Sci. USA 83, 6771-6775.
7. Eagle, H. (1959). Amino acid metabolism in mammalian cell cultures. Science 130, 432-437.
8. Fuller, S. D. 1987. The T=4 envelope of Sindbis virus is organized by interactions with a complementary T=3 capsid. Cell 48, 923-934.
9. Griffin, D. E. 1976. Role of the immune response in age-dependent resistance of mice to encephalitis due to Sindbis virus. J. Infect. Dis. 133, 456-464.

10. Griffin, D. E. 1986. Alphavirus pathogenesis and immunity, pp 209-249. In S. Schlesinger and M. J. Schlesinger (eds.), The Togaviridae and Flaviviridae. Plenum Press, New York.

11. Griffin, D. E. and Johnson, R. T. 1977. Role of the immune response in recovery from Sindbis virus encephalitis in mice. J. Immunol. 118, 1070-1075.

12. Jackson, A. C., Moench, T. R., Griffin, D. E. and Johnson, R. T. 1987. The pathogenesis of spinal cord involvement in the encephalomyelitis of mice caused by neuroadapted Sindbis virus infection. Lab. Invest. 56, 418-423.

13. Jackson, A. C., Moench, T. R., Trapp, B. D., and Griffin, D. E. 1988. The basis of neurovirulence in Sindbis virus encephalomyelitis of mice. Lab. Invest., in press.

14. Johnson, R. T. 1965. Virus invasion of the central nervous system: A study of the Sindbis virus infection in the mouse using fluorescent antibody. Am. J. Pathol. 46, 929-943.

15. Johnson, R. T., McFarland, H. F. and Levy, S. E. 1972. Age-dependence to viral encephalitis: Studies of infections due to Sindbis virus in mice. J. Infec. Dis. 125, 257-262.

16. Kohara, M., Omata, T., Kameda, A., Semler, B., Itoh, H., Wimmer, E., and Nomoto, A. 1985. In vitro phenotypic markers of a poliovirus recombinant constructed from infectious cDNA clones of the neurovirulent Mahoney strain and the attenuated Sabin 1 strain. J. Virol. 53, 786-792.

17. Lindqvist, B. H., DiSalvo, J., Rice, C. M., Strauss, J. H. and Strauss, E. G. 1986. Sindbis virus mutant ts20 of complementation group E contains a lesion in glycoprotein E2. Virology 151, 10-20.

18. Maniatis, T., Fritsch, E. F. and Sambrook, J. 1982. Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

19. Maxam, A. M. and Gilbert, W. 1980. Sequencing end-labeled DNA with base specific chemical cleavage. In L. Grossman and K. Moldave (eds.) Methods in Enzymology, Vol. 65, pp. 499-560. Academic Press, Orlando, Fl.

20. Okayama, H. and Berg, P. 1982. High-efficiency cloning of full-length cDNA. Mol. Cell. Biol. 2, 161-170.

21. Olmsted, R. A., Baric, R. S., Sawyer, B. A. and Johnston, R. E. 1984. Sindbis virus mutants selected for rapid growth in cell culture display attenuated virulence in animals. Science 225, 424-427.

22. Olmsted, R. A., Meyer, W. J. and Johnston, R. E. 1986. Characterization of Sindbis virus epitopes important for penetration in cell culture and pathogenesis in animals. Virology 148, 245-254.

23. Ou, J. H., Trent, D. W. and Strauss, J. H. 1982. The 3' noncoding regions of alphavirus RNAs contain repeating sequences. J. Mol. Biol. 156, 719-730.

24. Pierce, J. S., Strauss, E. G. and Strauss, J. H. 1974. Effect of ionic strength on the binding of Sindbis virus to chick cells. J. Virol. 13, 1030-1036.

25. Polo, J. M., Davis, N. L., Rice, C. M., Huang, H. V. and Johnston, R. E. 1988. Molecular analysis of Sindbis virus pathogenesis in neonatal mice using virus recombinants constructed in vitro. J. Virol. in press.

26. Reed, L. J. and Muench, H. 1938. A simple method of estimating fifty percent end points. Amer. J. Hyg. 27, 493-497.

27. Reinarz, A. B. G., Broome, M. G. and Sagik, B. P. 1971. Age-dependent resistance of mice to Sindbis virus infection: Viral replication as a function of host age. Infec. Immunity 3, 268-273.

28. Rice, C. M., Bell, J. R., Hunkapiller, M. W., Strauss, E. G. and Strauss, J. H. 1982. Isolation and characterization of the hydrophobic COOH-terminal domains of the Sindbis virus glycoproteins. J. Mol. Biol. 154, 355-378.

29. Rice, C. M., Levis, R., Strauss, J. H. and Huang, H. V. 1987. Production of infectious RNA transcripts from Sindbis virus cDNA clones: Mapping of lethal mutations, rescue of a temperature-sensitive marker, and in vitro mutagenesis to generate defined mutants. J. Virol. 61, 3809-3819.

30. Rice, C. M. and Strauss, J. H. 1981. Synthesis, cleavage, and sequence analysis of DNA complementary to the 26S messenger RNA of Sindbis virus. J. Mol. Biol. 150, 315-340.

31. Rice, C. M. and Strauss, J. H. 1982. Association of Sindbis virion glycoproteins and their precursors. J. Mol. Biol. 154, 325-348.

32. Seif, I., Coulon, P., Rollin, P. E. and Flamand, A. 1985. Rabies virulence: Effect on pathogenicity and sequence characterization of rabies virus mutations affecting antigenic site III of the glycoprotein. J. Virol. 53, 926-934.

33. Shope, R. E., Tignor, E. J., Rozhon, E. J., Bishop, D. H. L. 1981. The association of the bunyavirus middle sized RNA segment with mouse pathogenicity. In D. H. L. Bishop and R. W. Compans (eds.). The Replication of Negative Strand Viruses, pp. 147-152, Elsevier, New York.

34. Smith, A. L. and Tignor, G. H. 1980. Host cell receptors for two strains of Sindbis virus. Arch Virol. 66, 11-26.

35. Smith, D. R. and Calvo, J. M. 1980. Nucleotide sequence of the E. coli gene coding for dihydrofolate reductase. Nucl. Acids. Res. 13, 1717-1731.

36. Stanley, J., Cooper, S. J. and Griffin, D. E. 1985. Alphavirus neurovirulence: Monoclonal antibodies discriminating wild-type from neuroadapted Sindbis virus. J. Virol. 56, 110-119.

37. Strauss, E. G., Lenches, E. M. and Strauss, J. H. 1976. Mutants of Sindbis virus. I. Isolation and partial characterization of 89 new temperature-sensitive mutants. Virology 74, 154-168.

38. Strauss, E. G., Rice, C. M. and Strauss, J. H. 1984. Complete nucleotide sequence of the genomic RNA of Sindbis virus. Virology 133, 92-110.

39. Strauss, E. G. and Strauss, J. H. 1980. Mutants of alphaviruses: Genetics and physiology. In R. W. Schlesinger (ed.) The Togaviruses, Chapter 14, pp. 371-426. Academic Press, New York.

40. Strauss, E. G. and Strauss, J. H. 1986. Structure and replication of the alphavirus genome. In S. Schlesinger and M. J. Schlesinger (eds.). The Togaviridae and Flaviviridae, pp. 35-90. Plenum Press, New York.

41. Taylor, R. M., Hurlbut, H. S., Work, T. H., Kingston, J. R. and Frothingham, T. E. 1955. Sindbis virus: A newly recognized arthropod-transmitted virus. Am. J. Trop. Med. Hyg. 4, 844-862.

42. Zimmern, D. and Kaesberg, P. 1978. 3' terminal sequence of encephalomyocarditis virus RNA determined by reverse transcriptase and chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 75, 4257-4261.

## Chapter 5

Comparison of the Virulent Asibi Strain of Yellow Fever Virus with the 17D Vaccine Strain Derived from It.

This chapter was published in *Proc. Natl. Acad. Sci. USA*

*Proc. Natl. Acad. Sci. USA*
Vol. 84, pp. 2019–2023, April 1987
Medical Sciences

# Comparison of the virulent Asibi strain of yellow fever virus with the 17D vaccine strain derived from it

CHANG S. HAHN*, JOEL M. DALRYMPLE†, JAMES H. STRAUSS*, AND CHARLES M. RICE*‡

*Division of Biology, California Institute of Technology, Pasadena, CA 91125; and †Virology Division, U.S. Army Medical Research Institute of Infectious Diseases, Fort Detrick, Frederick, MD 21701-5011

*Communicated by Ray D. Owen, November 17, 1986 (received for review October 30, 1986)*

**ABSTRACT** **We have sequenced the virulent Asibi strain of yellow fever virus and compared this sequence to that of the 17D vaccine strain, which was derived from it. These two strains of viruses differ by more than 240 passages. We found that the two RNAs, 10,862 nucleotides long, differ at 68 nucleotide positions; these changes result in 32 amino acid differences. Overall, this corresponds to 0.63% nucleotide sequence divergence, and the changes are scattered throughout the genome. The overall divergence at the level of amino acid substitution is 0.94%, but these changes are not randomly distributed among the virus proteins. The capsid protein is unchanged, while proteins NS1, NS3, and NS5 contain 0.5% amino acid substitutions, and proteins ns4a and ns4b average 0.8% substitutions. In contrast, proteins ns2a and ns2b have 3.0 and 2.3% amino acid divergence, respectively. The envelope protein also has a relatively high rate of amino acid change of 2.4% (a total of 12 amino acid substitutions). The large number of changes in ns2a and ns2b, which are largely conservative in nature, may result from lowered selective pressure against alteration in this region; among flaviviruses, these polypeptides are much less highly conserved than NS1, NS3, and NS5. However, many of the amino acid substitutions in the E protein are not conservative. It seems likely that at least some of the difference in virulence between the two strains of yellow fever virus results from changes in the envelope protein that affect virus binding to host receptors. Such differences in receptor binding could result in the reduced neurotropism and vicerotropism exhibited by the vaccine strain.**

Yellow fever virus belongs to the *Flavivirus* genus of the family Flaviviridae, a group of some 70 closely related viruses many of which cause serious human illness (1). Yellow fever is arthropod-borne, transmitted by mosquitos of the genera *Aedes* and *Haemagogus*. Its natural vertebrate host range is limited to primates in which it is viscerotropic and neurotropic. In man, the virus causes a serious, often fatal, illness marked by liver and kidney involvement and hemorrhage. For several hundred years the virus caused epidemics in the Americas, Europe, and Africa that led to widespread human suffering. With the control of the urban vector of yellow fever *Aedes aegypti* beginning in the early 1900s, epidemic urban yellow fever disappeared. However, the virus remains present in an enzootic cycle in the forests of South America and Africa and causes periodic outbreaks in neighboring human populations.

Reed (2) first proved that yellow fever is transmitted by mosquitos and, shortly thereafter, that the disease agent was filterable. Because there was no recognized, susceptible laboratory host, many years elapsed before the virus responsible was isolated by the Rockefeller Foundation's West Africa yellow fever commission. In 1927, these workers succeeded in isolating a virus from the blood of a young Ghanian named Asibi by monkey/monkey passage (3). This Asibi strain of yellow fever causes an invariably fatal disease when inoculated into rhesus monkeys.

Theiler (4) developed a live, attenuated vaccine strain, which he referred to as 17D, from the Asibi strain. Starting with the Asibi strain that had been passaged 53 times in monkeys, with intermittent passages in *A. aegypti*, the virus was propagated serially in cultures of embryonic mouse tissue (18 passages), minced whole chicken embryo (50 passages), and finally minced chicken embryos without nervous tissue (152 passages). Between the 89th and 114th *in vitro* passage (from the start of the experiment, that is including the passages in embryonic mouse tissue and whole chicken embryos) a marked change in virulence of the virus occurred. The reason for the change in virulence is unknown and attempts to repeat these experiments by virus passage have failed to develop additional avirulent strains. The 17D strain has been widely used as a human vaccine, being safe and highly effective. It causes a mild, generalized infection in humans (or other primates) with involvement of lymphoid tissue and minimal quantities of virus circulating in the blood, and both the viscerotropism and neurotropism of the parental Asibi virus are markedly reduced.

We have sequenced the genome of a plaque-purified virus derived from the 17D-204 vaccine strain of yellow fever (5). The 17D-204 strain, supplied by the American Type Culture Collection, had been passed an additional 14 times in chicken embryo tissue culture, for a total of 234 *in vitro* passages. To obtain RNA for cloning and sequencing, the virus was passed twice in chicken embryo fibroblasts in our laboratory, plaque-purified in Vero cells, and passed twice in BHK cells; virus for RNA preparation was then grown in SW-13 cells (5). Thus, the 17D strain sequenced by us had been passed 240 times *in vitro*.

We now report the sequence of virtually the entire genome of the Asibi strain of yellow fever and compare the sequences of the 17D and Asibi strains at both the nucleotide level and the amino acid level. In addition to defining the changes that have occurred during 240 serial passages, this work represents a necessary prelude for studying the biological significance of these changes as related to the different virulence of the two strains.

## MATERIALS AND METHODS

**Asibi Strain Propagation and RNA Isolation.** The Asibi strain of yellow fever virus was obtained from the Yale Arbovirus Research Unit reference collection as viremic monkey serum (supplied by R. Shope). This virus was originally isolated in rhesus monkeys (3) and had undergone ≈45 serial monkey passages consistently producing a fatal illness (6); it is unknown how many of these 45 passages

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

‡Present address: Department of Microbiology and Immunology, Washington University School of Medicine, Box 8093, 660 South Euclid Avenue, St. Louis, MO 63110.

correspond to the monkey passages of the laboratory strain of Asibi used as the starting virus for *in vitro* passages that led to 17D. The virulence of this Asibi strain was demonstrated by inoculation of additional rhesus monkeys, all of which suffered a lethal infection with yellow fever pathology. Virus was isolated in LLC-$MK_2$ cell culture by direct inoculation of viremic plasma from one of these monkeys, and the infectious cell culture fluids were used to infect cultures of Vero cells. This Vero cell-passaged Asibi strain virus was titered at $7.8 \times 10^7$ plaque-forming units/ml and was used as inoculum for all subsequent virus production.

Confluent monolayers of Vero cells in 150-$cm^2$ flasks were infected at a multiplicity of ≈0.1, and the cultures were incubated at 36°C in Eagle's medium supplemented with 5% (vol/vol) heated fetal bovine serum and antibiotics. Virus-containing culture medium was harvested 4–5 days after infection (depending upon the first evidence of virus induced cytopathology) and concentrated by polyethylene glycol precipitation. Virus was purified by density-gradient centrifugation on potassium tartrate/glycerol gradients followed by rate-zonal centrifugation on linear sucrose gradients, and the RNA was extracted from $NaDodSO_4$-disrupted virions using a phenol/cresol/8-hydroxyquinoline/chloroform mixture (7).

**Cloning of Asibi cDNA.** A cDNA library from Asibi genomic RNA was constructed as described (5, 8). Ampicillin-resistant colonies from this library were screened by colony hybridization using nick-translated restriction fragments derived from the library of 17D yellow fever clones as probes, as described (9). Colonies with larger inserts (obtained using cDNA size class of 2 kilobases or larger) were screened with a 5′ probe of 2280 base pairs that extends from the 5′ end of 17D yellow fever to the first *Eco*RI site and with a 3′ probe of 2580 base pairs, which extends from nucleotide 8280 in 17D to the 3′ end, or with probes derived from other regions of the 17D genome. To obtain clones containing the extreme 3′ end of the RNA ≈4000 clones from the smaller insert class (derived from Asibi double-stranded cDNA of 0.8–2 kilobases long) were screened with a fragment derived from the extreme 3′ end of 17D. This fragment, 150 nucleotides long, extends from the *Xba* I site at nucleotide 10,708 to the 3′ end of 17D yellow fever. We found three positive clones with this probe, two of which were identical. One of these clones had an unusual structure and probably arose by self-priming; it was used to obtain the Asibi sequence through to the 3′-terminal nucleotide.

**Sequence Analysis of Asibi cDNA.** Plasmid DNA from selected Asibi clones was sequenced using the chemical method (10, 11) as described (12).

## RESULTS

**Sequence of Asibi Yellow Fever.** The sequencing strategy used to obtain the nucleotide sequence of Asibi yellow fever is diagramed in Fig. 1. The sequence obtained was completely overlapped and was determined for at least two independent clones throughout virtually the entire genome. This allowed the detection of heterogeneity in the cloned cDNA population (either due to heterogeneity in the RNA genomes cloned or due to errors introduced during reverse transcription and subsequent cloning). Since the RNA templates that were used for cloning were not derived from plaque-purified virus, such clonal differences might be expected, and we found six nucleotides that differed between two clones. The sequence obtained totaled 10,848 nucleotides and represents the entire Asibi genomic RNA sequence with the exception of the 5′-terminal 14 nucleotides.

**Comparison of Nucleotide Sequences of 17D and Asibi Yellow Fever.** All of the nucleotide differences found between Asibi and 17D yellow fever are shown in Fig. 2. Clonal

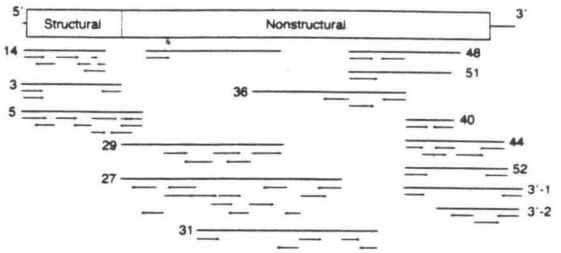



FIG. 1. Sequencing strategy used to obtain Asibi yellow fever sequence from cDNA clones. A representation of the yellow fever genome is shown with the coding region indicated as a box. Below are shown various cDNA clones drawn to scale. The regions of each clone sequenced are indicated. All sequencing was 3′ to 5′ on the cDNA, and the direction of the arrow indicates whether the plus strand (←) or the complementary minus strand (→) was sequenced. Clones 3 and 5 start with the 15th and 17th nucleotide, respectively, from the 5′ end of the yellow fever genome. The 3′-terminal 111 nucleotides were present in only one clone, 3′-1, whose structure is complex and which probably arose by self-priming. In all other regions at least two independent clones were sequenced.

differences found among the Asibi clones are also presented. A total of 68 nucleotide differences were found between the two strains that had been fixed and an additional 6 nucleotides that differed in two different clones of Asibi yellow fever. In each clonal difference detected, one of the nucleotides found at a given position was the same as that found in 17D yellow fever. The sequence of 17D yellow fever that has been reported is a consensus sequence derived from sequencing more than one clone throughout the entire region (5). With a total of 68 changes (0.63% of the genome) there was an average of about 0.27 change fixed per passage. Assuming an effective multiplication of $2^{10}$ at each passage, this would represent $2 \times 10^{-6}$ change fixed per nucleotide per generation. Changes presumably occur more frequently than this but do not survive selection pressure during continued passage.

The nucleotide changes are summarized as to transitions and transversions in Fig. 3. Transitions are five times more common than transversions, as expected if most of the changes occur by mispairing during RNA replication.

Nucleotide changes are scattered throughout the genome but occurred more frequently in the envelope protein region (1.0% nucleotide sequence difference), in the ns2a region (1.6% difference), in the ns2b region (1.0% difference), and in the 3′-noncoding region (1.2% difference) (Fig. 4). Changes in the untranslated region may be more frequent because of a relative lack of selective pressure against changes that occur in these regions, although it is of interest that no changes were found in the 5′-untranslated region. Within the coding region, selection against deleterious changes in protein sequence would be expected, and it is notable that approximately half of the nucleotide changes (36 of 68 changes) do not result in a change in a coding assignment. Outside the envelope protein region and the ns2a regions, in which a disproportionate number of the nucleotide changes lead to coding changes, 32 of 47 nucleotide changes do not result in an amino acid substitution.

**Changes in Amino Acid Sequence.** Amino acid differences between 17D and Asibi yellow fever are given in Fig. 2 and summarized in Fig. 4. Overall, there are 32 amino acid changes (0.94% difference) between the two strains and an additional three clonal differences in the Asibi strain. Thus, there has been an average of 0.12 change fixed per passage, or $4 \times 10^{-6}$ change per amino acid per generation. As is clear from Fig. 4, these changes are not randomly distributed.

| | No | NT | AA |
|---|---|---|---|
| Capsid | 304 | A G | |
| | 370 | C U | |
| M | 854 | U C | F L |
| | 883 | G A | |
| | 1127 | A G | R G |
| | 1140 | U C | V A |
| | 1482 | U C | V A |
| | 1491 | U C | I T |
| | 1572 | C A | T K |
| | 1750 | U C | |
| | 1819 | U C | |
| Envelope | 1870 | A G | I M |
| | 1887 | U C | F S |
| | 1946 | U C | S P |
| | 1965 | G A | R K |
| | 2112 | G C | R T |
| | 2142 | A C | H P |
| | 2219 | A G | T A |
| | 2356 | U C | |
| | 2687 | U C | F L |
| | 2704 | G A | |
| NS1 | 3274 | A G | |
| | 3371 | G A | V I |
| | 3613 | A G | |
| | 3817 | G G,A | |
| | 3860 | G A | V M |
| | 3915 | U U,A | |
| ns2a | 4007 | G A | A T |
| | 4013 | U C | F L |
| | 4022 | G A | A T |
| | 4054 | U C | |
| | 4056 | U C | F S |
| | 4289 | C A | L I |
| ns2b | 4387 | G A | |
| | 4505 | C A | L I |
| | 4507 | C U | |
| | 4612 | C U | |
| | 4864 | G G,A | |
| | 4873 | G U | |
| | 5131 | G G,U | M M,I |
| | 5153 | G A | V I |
| NS3 | 5194 | C U | |
| | 5431 | U C | |
| | 5473 | U C | |
| | 5641 | A G | |
| | 6013 | U C | |
| | 6023 | A G | N D |
| | 6448 | U G | |
| | 6529 | C U | |
| ns4a | 6758 | G A | V I |
| | 6829 | C U | |
| | 6876 | C U | A V |
| | 7171 | G A | M I |
| ns4b | 7571 | A C | |
| | 7580 | C U | H Y |
| | 7642 | C U | |
| | 7701 | G A | R Q |
| | 7945 | U C | |
| | 8008 | C U | |
| | 8629 | U C | |
| NS5 | 9605 | G A | D N |
| | 10075 | G G,U | M M,I |
| | 10142 | A G | K E |
| | 10243 | A G | |
| | 10285 | C U | |
| | 10312 | G A | |
| | 10316 | U U,C | S S,P |
| | 10338 | U C | L P |
| | 10367 | C U | |
| 3' Non- | 10418 | C U | |
| coding | 10454 | G A | |
| | 10550 | C U | |
| | 10800 | A G | |
| | 10847 | C A | |

FIG. 2. Summary of differences between Asibi yellow fever and 17D yellow fever. The changes are grouped by nucleotide number into the various regions of the genome. Nucleotide number (No) is from the 5' terminus of 17D. To the right is shown the nucleotide (NT) in the 17D genome followed by the nucleotide in the Asibi genome. Where clonal differences were found in Asibi yellow fever, both nucleotides as well as both amino acids if appropriate are shown. If the nucleotide change results in an amino acid substitution, the amino acid (AA) in 17D is shown, followed by the amino acid in Asibi. Note that in every case of clonal differences in the Asibi strain, one of the nucleotides is the same as that found in 17D strain.

Proteins E, ns2a, and ns2b exhibit a disproportionately high rate of change, 2.4% for E, 3% for ns2a, and 2.3% for ns2b. Only 13 changes, or less than half of the total changes found between Asibi and 17D, are found in other regions, an amino acid sequence divergence of only 0.5%.

The function of the various nonstructural proteins in virus replication is unknown. However, NS3 and NS5 probably form components of the viral replicase responsible for replicating RNA. If these proteins do possess enzymatic functions, then amino acid changes might be expected to be deleterious and selected against during continued passage. It is of note that most of the changes found in NS5 occur near the ends of the molecule, with only one change occurring within the central 75% of the protein. It is unknown whether NS1 has enzymatic activity, but it has been postulated that since it is a glycoprotein, it may be involved in virus assembly. The fact that amino acid substitutions in NS1 are rare, as was the case for NS3 and NS5, suggests that whatever role it plays in virus replication, it requires a precise amino acid sequence for that function. NS1, NS3, and NS5 demonstrate a high degree of conservation among different flaviviruses (13).

The small polypeptides ns2a, ns2b, ns4a, and ns4b are hydrophobic in nature and are not highly conserved among flaviviruses (13). Their hydrophobicity profiles are remarkably conserved, however, suggesting that as long as the hydrophobicity profile is unchanged, a large number of amino acid substitutions can be accommodated without affecting the normal function of these proteins. The high frequency of change observed in the ns2 region may simply be a reflection of this; many of the changes that arise may not affect function and are, therefore, not selected against. The amino acid changes found in the ns2 region (Fig. 2) would have only marginal effects upon the hydrophobicity profile.

| ASIBI | 17D | No | ASIBI | 17D | No | ASIBI | 17D | No |
|---|---|---|---|---|---|---|---|---|
| G | A | 11 | G | C | 0 | C | G | 1 |
| A | G | 17 | G | U | 1 | C | A | 2 |
| C | U | 20 | A | C | 4 | U | G | 3 |
| U | C | 14 | A | U | 1 | U | A | 0 |
| Transition | | 62 | Pu | Py | 6 | Py | Pu | 6 |
| | | | Transversion | | 12 | | | |

FIG. 3. Transitions and transversions that have occurred during passage of 17D strain.

**Amino Acid Changes Within the Structural Protein Region.** No changes were found in the capsid protein, and only one change was found in prM (this change occurs within M). There are, however, a large number of changes in the envelope protein (Figs. 2 and 4). Furthermore, 15 nucleotide changes in the envelope gene have led to 12 amino acid substitutions (in contrast to the rest of the coding region where 53 nucleotide changes lead to 20 amino acid substitutions), suggesting that in the E protein, some of the amino acid substitutions may have been positively selected for during passage, rather than simply being neutral in effect. Although it is impossible to predict the effect of any particular amino acid alteration without detailed knowledge of the three-dimensional structure and function of the protein, five nonconservative amino acid substitutions are likely candidates to significantly alter envelope protein structure and function. These are Gly-52 to Arg, Thr-173 to Ile, Lys-200 to Thr, Pro-320 to Ser, Thr-380 to Arg, and Pro-390 to His. It is of note that the last two of these changes occur within a conserved domain of the E protein. The E protein sequences from flaviviruses representing the three serological sub-

| REGION | TOTAL NT/AA | CHANGE NT/AA | % CHANGE NT/AA |
|---|---|---|---|
| 5' Noncoding | 118 / - | 0 / - | 0 / - |
| Capsid | 363 / 121 | 2 / 0 | 0.55% / 0 |
| prM (-M) | 267 / 89 | 0 / 0 | 0 / 0 |
| M | 225 / 75 | 2 / 1 | 0.89% / 1.89% |
| Envelope | 1479 / 493 | 15 / 12 | 1.01% / 2.43% |
| NS1 | 1227 / 409 | 5 / 2 | 0.41% / 0.49% |
| ns2a | 501 / 167 | 6 / 5 | 1.20% / 2.99% |
| ns2b | 390 / 130 | 4 / 2 | 1.03% / 2.31% |
| NS3 | 1869 / 623 | 9 / 2 | 0.48% / 0.32% |
| ns4a | 861 / 287 | 6 / 3 | 0.70% / 0.78% |
| ns4b | 336 / 112 | 2 / 1 | 0.60% / 0.89% |
| NS5 | 2715 / 905 | 11 / 4 | 0.40% / 0.44% |
| 3' Noncoding | 511 / - | 6 / - | 1.17% / - |
| TOTAL | 10862 / 3411 | 68 / 32 | 0.63% / 0.97% |

FIG. 4. Summary of the differences between Asibi and 17D strains at the nucleotide and the amino acid levels. The number of changes and the percent change at both nucleotide and amino acid levels are shown for various regions of the genome.

groups of the mosquito-borne flaviviruses are aligned in Fig. 5; the domain from Pro-369 to Gly-448 shows a high degree of conservation among flaviviruses. In this region there are the following three amino acid changes between Asibi and 17D: Thr-380 to Arg, Pro-390 to His, and Ala-416 to Thr. It is remarkable that in all three cases, 17D changes resulted in the same amino acid as in the Murray Valley encephalitis/West Nile/St. Louis encephalitis subgroup of flaviviruses. In one case the amino acid is also present in dengue-2 virus (Thr-416), and in a second case (His-390) there is also an aromatic amino acid (phenylalanine) at the equivalent position in dengue-2 virus. It seems unlikely, therefore, that these alterations in the yellow fever E protein sequence are due to random events, and it is tempting to speculate that these substitutions alter receptor affinities. In the Murray Valley encephalitis subgroup, birds are an important natural host, and 17D was selected for efficient multiplication in chicken cells. It should also be kept in mind that only a limited number of amino acid substitutions might be tolerated.

**The 3′-Untranslated Region.** Of the six nucleotide changes found in the 3′-untranslated region, two occur within the predicted secondary structure at the 3′ end of the RNA. One of these would result in an extra A·U base pair in the 3′-terminal 88 nucleotides, while the second results in the loss of an A·U base pair; thus these two changes should have only a minor effect on the stability of the proposed secondary structure.

There are three repeated-sequence elements in the 3′-untranslated region that are 42 nucleotides long, each of which differs from the others by 4 or 5 nucleotides. These repeats are shown for Asibi RNA in Fig. 6. Note that within this repeat, there are 16 contiguous nucleotides that are repeated identically. In 17D RNA there has been an adenosine to guanosine transition at nucleotide 10,454 that results in a mismatch within one copy of this perfect repeat. The significance of this change is unknown.

## DISCUSSION

The 17D strain of yellow fever virus is one of the safest and most effective live virus vaccines ever developed, especially after stabilization of seed lots (for a study of the different seed lots see ref. 17). In this report we detail the differences in nucleotide and amino acid sequence between this vaccine strain and the parental Asibi isolate from which it was derived. The change or changes responsible for the altered virulence or attenuation of 17D cannot be determined at the current time. Indeed, the situation is complicated by the fact that the 17D vaccine contains a mixture of variants that differ in a number of biological properties (17, 18). It is unclear which of these variants was sequenced to obtain the 17D sequence, since the vaccine strain was plaque-purified before amplification and sequencing. It is unknown if this mixture of variants is important for the vaccine properties of the 17D strain; however, reversion to virulence has never been documented. It is known that a change occurred in the virulence of yellow fever between the 89th and 114th passage; that this change, once it occurred, was stable; and that comparable changes were not obtained when the passage history of the 17D strain was repeated. The Asibi strain sequenced was not plaque purified and a number of clonal differences were found.

Attenuation of a virus by propagation in tissue culture presumably results from selection for variants better adapted for replication in tissue culture and, conversely, less-well adapted for replication in their natural hosts, although accumulation and fixation of nonselected mutations may also be important. In particular, variants might be selected that bind more readily to receptors on cultured chicken cells, leading to more rapid attachment and penetration. Such an alteration in receptor binding could, as a consequence, lead to less-efficient binding to receptors found in hepatocytes or neurons in primates, and thus to an alteration in tissue tropism and reduction in virulence. The large number of changes found in the envelope protein is consistent with such an hypothesis. Changes elsewhere in the genome may also be important, however, particularly if they affect the efficiency of virus replication. Because of the stability of the vaccine strain, it is likely that more than one change is important for its avirulence. If it becomes possible to rescue infectious virus from a cloned cDNA copy of yellow fever, as has been done for several RNA viruses (19–21), it will be possible to test the effect of the individual changes found on the virulence of the virus. Such an approach has been used with the Sabin and

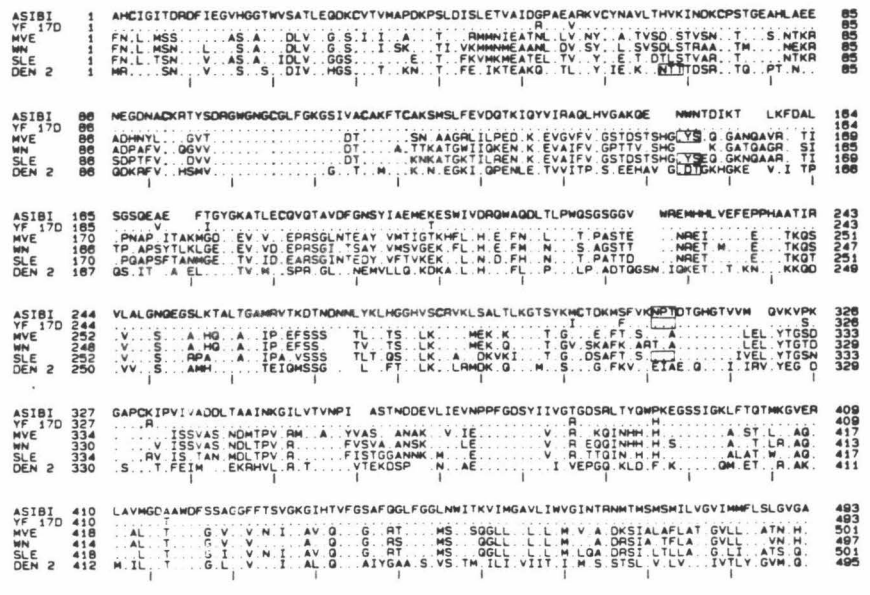


FIG. 5. Envelope protein sequences of five flaviviruses. The aligned amino acid sequence of five flavivirus E proteins are shown in the single-letter amino acid code. Dots indicate that the amino acid is the same as in Asibi yellow fever strain. Murray Valley encephalitis (MVE) (14), West Nile (WN) (15), and St. Louis encephalitis (SLE) (16) viruses belong to a separate serological subgroup of mosquito-borne flaviviruses. Dengue 2 (DEN 2) virus is a member of a third subgroup (Y. S. Hahn, R. Galler, J.M.D., J.H.S., and E. G. Strauss, unpublished data). Cysteine residues are shaded, and potential carbohydrate addition sites are boxed.

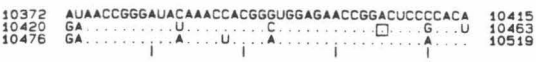



FIG. 6. Repeated sequences in the 3'-untranslated region. The sequences of three repeats in the 3'-untranslated region of Asibi RNA are aligned. Nucleotide 10,454 is a guanosine in the 17D strain, which introduces a mismatch in a 16-nucleotide perfect repeat.

Mahoney strains of poliovirus type I, with the finding that several changes within the genome are involved in the change in virulence between the two strains (22, 23).

Studies by Schlesinger *et al.* (24) have shown that monoclonal antibodies against the envelope protein of 17D yellow fever are often able to discriminate between 17D and Asibi yellow fever. Interestingly, some of these antibodies neutralize Asibi virus but not 17D virus. This suggests that changes in the E proteins are important for determining the structure of E and/or are found in regions important for antibody binding. These authors also found that E proteins of 17D were present in both glycosylated and nonglycosylated forms, whereas Asibi E protein was present in only one form (presumably nonglycosylated from its mobility relative to that of 17D E protein, since we show here that the two proteins are identical in size). No change in the single potential glycosylation site was found between 17D and Asibi, although the change of Phe-305 to Ser is found just upstream of this site. Presumably this change, or other changes upstream that affect the folding of the proteins, leads to changes in accessibility of the site to the glycosylation enzymes.

This study is also of interest because it represents a detailed study of differences that have arisen during 240 passages in which the passage history is known with some clarity. RNA replicases lack proofreading activity and the error frequency of these enzymes is estimated to be on the order of about $10^{-4}$ per nucleotide per generation, based upon theoretical considerations of the free energy of an A·U or G·C base pair, or upon measured mutation frequencies in a number of virus systems (25, 26). However, measurements in other systems have yielded lower estimates of mutation frequency. Parvin *et al.* (27) found a substitution frequency in the NS gene of influenza of $4 \times 10^{-6}$ substitution per nucleotide per generation and $<5 \times 10^{-7}$ substitution per generation in the poliovirus VP1 gene. We report here that yellow fever underwent $2 \times 10^{-6}$ substitution per nucleotide per generation during passage of the 17D strain. The difficulty in relating the various mutation rates observed comes in estimating the fraction of nucleotide substitutions that are capable of survival under the conditions used, which will differ with the protein domain under consideration (some domains or even entire proteins tolerate changes more than others) and with the method of selection [simple ability to persist in a population and form a plaque, as was used by Parvin *et al.* (27), or ability to compete head to head with other viruses in the population, as was the case with yellow fever 17D]. Domingo *et al.* (28) have shown with bacteriophage Qβ that in direct competition, the RNA population consists of an average nucleotide sequence that is maintained during passage, but in which variants arise at high frequency that are subsequently selected against during continued passage. Since many nucleotide changes studied were almost certainly silent changes, the RNA secondary structure of Qβ appears to be important for rapid growth and even silent changes may be selected against, a situation that may well prevail with other RNA viruses and especially with plus-strand RNA viruses. Thus it is unclear whether even the incidence of silent change can be used to assay the inherent mistake frequencies of the viral RNA polymerases. In the case of 17D yellow fever, those alterations that have been fixed presumably represent mutations that are truly neutral (or possibly fixed during plaque purification) as well as positive changes that lead to more efficient growth in tissue culture.

We thank Ms. E. M. Lenches for expert technical assistance. This work was supported by a grant from The World Health Organization and in part by Grants AI 20612 and AI 10793 from the National Institutes of Health and Grant PCM 83 16856 from the National Science Foundation.

1. Monath, T. P. (1986) in *The Togaviridae and Flaviviridae*, eds. Schlesinger, S. & Schlesinger, M. J. (Plenum, New York), pp. 375–440.
2. Reed, W. (1901) *Med. Rec.* **60**, 201–209.
3. Stokes, A., Bauer, J. H. & Hudson, N. P. (1928) *Am. J. Trop. Med.* **8**, 103–164.
4. Theiler, M. (1951) in *Yellow Fever*, ed. Strode, K. (McGraw-Hill, New York), pp. 39–136.
5. Rice, C. M., Lenches, E. M., Eddy, S. R., Shin, S. J., Sheets, R. L. & Strauss, J. H. (1985) *Science* **229**, 726–733.
6. Theiler, M. & Anderson, C. R. (1975) *Am. J. Trop. Med. Hyg.* **24**, 115–117.
7. Repik, P. M., Dalrymple, J. M., Brandt, W. E., McCown, J. M. & Russell, P. K. (1983) *Am. J. Trop. Med. Hyg.* **32**, 577–589.
8. Rice, C. M., Dalgarno, L., Galler, R., Hahn, Y. S., Strauss, E. G. & Strauss, J. H. (1987) in *Molecular Trends in Virology*, Proceedings of the International Symposium, June 1986, eds. Bauer, H., Klenk, H. D. & Scholtissek, C. (Springer, Giessen, F.R.G.), in press.
9. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
10. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
11. Smith, D. R. & Calvo, J. M. (1980) *Nucleic Acids Res.* **8**, 2255–2274.
12. Strauss, E. G., Rice, C. M. & Strauss, J. H. (1984) *Virology* **133**, 92–110.
13. Strauss, J. H. & Strauss, E. G. (1987) in *RNA Genetics*, eds. Domingo, E., Holland, J. & Ahlquist, P. (CRC, Boca Raton, FL), in press.
14. Dalgarno, L., Trent, D. W., Strauss, J. H. & Rice, C. M. (1986) *J. Mol. Biol.* **187**, 309–323.
15. Wengler, G., Castle, E., Leidner, U., Nowak, T. & Wengler, G. (1985) *Virology* **147**, 264–274.
16. Trent, D. W., Kinney, R. M., Johnson, B. J. B., Vorndam, A. V., Grant, J. A., Deubel, V., Rice, C. M. & Hahn, C. (1987) *Virology*, in press.
17. Monath, T. P., Kinney, R. M., Schlesinger, J. J., Brandriss, N. W. & Bres, B. (1983) *J. Gen. Virol.* **64**, 627–637.
18. Liprandi, F. (1981) *J. Gen. Virol.* **56**, 363–370.
19. Taniguchi, T., Palmieri, M. & Weissmann, C. (1978) *Nature (London)* **274**, 223–228.
20. Racaniello, B. R. & Baltimore, D. (1981) *Science* **214**, 916–919.
21. Ahlquist, P., French, R., Janda, M. & Loesch-Fries, L. S. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 7066–7070.
22. Omata, P., Kohara, M., Kuge, S., Komatsu, T., Abe, S., Semler, B. L., Kameda, A., Itoh, H., Arita, M., Wimmer, E. & Nomoto, A. (1986) *J. Virol.* **58**, 348–358.
23. LaMonica, N., Meriam, C. & Racaniello, V. R. (1986) *J. Virol.* **57**, 515–525.
24. Schlesinger, J. J., Brandriss, N. W. & Monath, T. P. (1983) *Virology* **125**, 8–17.
25. Holland, J., Spindler, K., Horodyski, F., Grabau, E., Nichol, S. & VandePol, S. (1982) *Science* **215**, 1577–1585.
26. Steinhauer, D. A. & Holland, J. J. (1986) *J. Virol.* **57**, 219–228.
27. Parvin, J. D., Moscona, A., Pan, W. T., Leider, J. M. & Palese, T. (1986) *J. Virol.* **59**, 377–383.
28. Domingo, E., Sabo, D., Taniguchi, T. & Weissmann, C. (1978) *Cell* **13**, 735–744.

# Appendix

## Comparison of the Asibi and 17D Strains of Yellow Fever Virus

This was published in *Vaccines 87*,
Cold Spring Harbor Press, Cold Spring Harbor, N.Y.
and is reproduced with the permission of the publisher.

# Comparison of the Asibi and 17D Strains of Yellow Fever Virus

**Chang S. Hahn, Charles M. Rice,* and James H. Strauss**
*Division of Biology, California Institute of Technology, Pasadena, California 91125*

**Joel M. Dalrymple**
*Virology Division, United States Army Medical Research Institute of Infectious Diseases Fort Detrick, Federick, Maryland 21701*

The 17D strain of yellow fever virus is an excellent live-virus vaccine, being highly efficacious and having only limited side effects. It is the only effective live-virus vaccine obtained to date for any of the flaviviruses, a group of about 70 closely related viruses classified as family Flaviviridae, many of which cause serious human illness. We were interested in comparing the sequences of the 17D and Asibi strains of yellow fever virus in order to define the changes that have occurred during 240 serial passages of a virus in culture, as well as to provide a framework for studying the biological significance of the changes that have occurred between the two strains.

## 17D Strain of Yellow Fever Virus

The Asibi strain of yellow fever virus was isolated from a young Ghanian of that name in 1927 by the Rockefeller Foundation's West Africa Yellow Fever Commission (Stokes et al. 1928). This virus, which was maintained by monkey-monkey passage, is neurotropic and viscerotropic in humans and causes an invariably fatal disease when inoculated into rhesus monkeys. Beginning with this strain, M. Theiler and colleagues (for review, see Strode 1951) developed a live attenuated vaccine strain referred to as 17D. The passage history of the 17D strain is reviewed in Table 1. Somewhere between the 89th and 114th passage in mouse embryo tissue culture and chicken embryo tissue culture, a marked change in virulence occurred.

The vaccine strain known as 17D-204 (provided by the American Type Culture Collection) had been through a total of 234 in vitro passages. Subsequently, to obtain enough RNA for molecular cloning, we subjected this strain in our laboratory to additional passages in chick cells, BHK cells, and SW13 cells, as well as to plaque purification in VERO cells, as outlined in Table 1. A cDNA library representing the entire nucleotide sequence of yellow fever was obtained from virion RNA after virus purification by precipitation with polyethylene glycol, followed by velocity and isopycnic sedimentation in gradients of glycerol and tartrate. The entire sequence of 17D yellow fever was obtained from this library (Rice et al. 1985, 1986a) and was used to deduce the amino acid sequences of all the yellow-fever-encoded proteins. We found that the yellow fever genome was translated as a polyprotein and subsequently

*Present address: Department of Microbiology and Immunology, Washington University of St. Louis, St. Louis, Missouri 63110.


*Vaccines 87*. © 1987 Cold Spring Harbor Laboratory. 0-87969-302-9/87 $1.00 + .00

**Table 1**
Passage History of 17D Yellow Fever Virus

| |
|---|
| In vivo passages |
| 53 passages in monkey (with intermittent periods of life in *Aedes aegypti)* |
| In vitro passages to produce 17D yellow fever[a] |
| 18 passages in minced mouse embryo tissue culture |
| 50 passages in minced whole chicken embryo tissue culture |
| 152 passages in minced chicken embryo without CNS[b] |
| 14 more passages in CE (with or without CNS) = 17D-204 strain, supplied by ATCC |
| In vitro passages to produce RNA for cDNA cloning |
| 2 passages in CEF |
| plaque purification in VERO cells |
| 2 passages in BHK cells |
| 1 passage in SW13 cells |

The passage history of yellow fever to produce the 17D strain is indicated. See text for details and authors. (CNS) Central nervous system; (CE) chick embryo; (ATCC) American type-culture collection; (CEF) chick embryo fibroblast; (BHK) baby hamster kidney; (SW13) a human cell line derived from an adenocarcinoma.

[a]Total in vitro passages is 234.

[b]A marked change in virulence occurred somewhere between the 89th and 114th in vitro passage and was not able to be repeated.

processed to produce the various structural and nonstructural proteins of the virus. The genome organization and the location of the flavivirus-encoded proteins have been confirmed by direct amino-terminal sequencing of yellow fever proteins (Bell et al. 1985; Rice et al. 1986b), as well as by sequencing of proteins of other flaviviruses (Bell et al. 1985; Castle et al. 1985, 1986; Wengler et al. 1985; Rice et al. 1986b). Because we are ultimately interested in comparing strains of yellow fever virus, we obtained the entire sequence of the 17D strain on at least two independent cDNA clones so that strain variation and/or errors occurring during reverse transcription and cloning would not affect the sequence obtained.

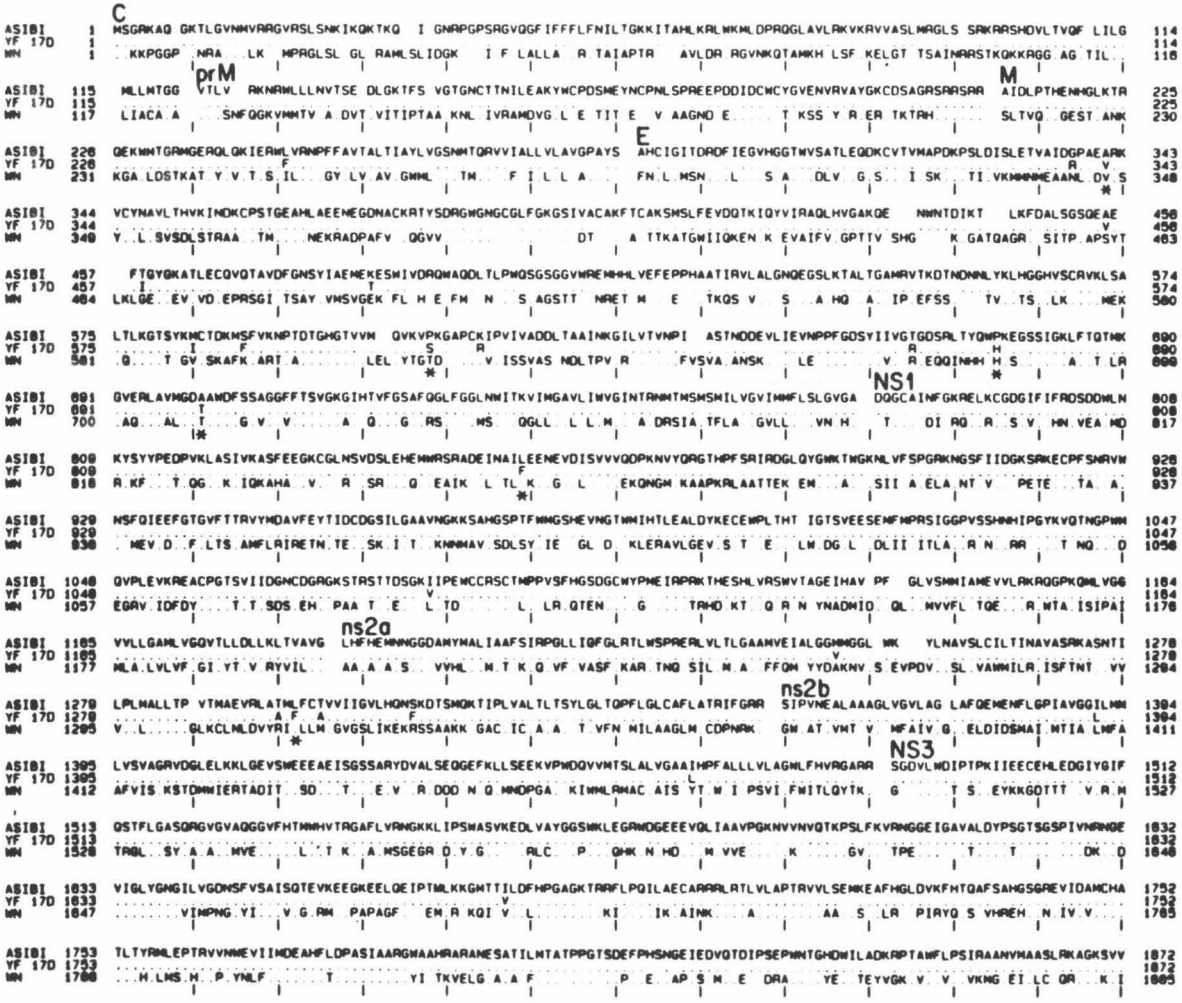
### Asibi Strain of Yellow Fever Virus

The Asibi strain used had been passaged 45 times in rhesus monkeys by blood-to-blood transfer (Table 2). It is not known how many of these monkey-monkey passages are the same as the passages in the monkey of the 17D strain reported in Table 1, as the records have been lost in antiquity. At the 46th passage, the virus was still

**Table 2**
Passage History of Asibi Yellow Fever

| |
|---|
| In vivo passages |
| 45 passages in rhesus monkey = YARU[a] reference virus |
| 1 additional passage in monkey |
| In vitro passage to produce RNA for cDNA cloning |
| 1 passage in LLC cells |
| 2 passages in VERO cells |

[a]Yale Arbovirus Research Unit. Virus supplied by R. Shope.


C
prM
M
E
NS1
ns2a
ns2b
NS3

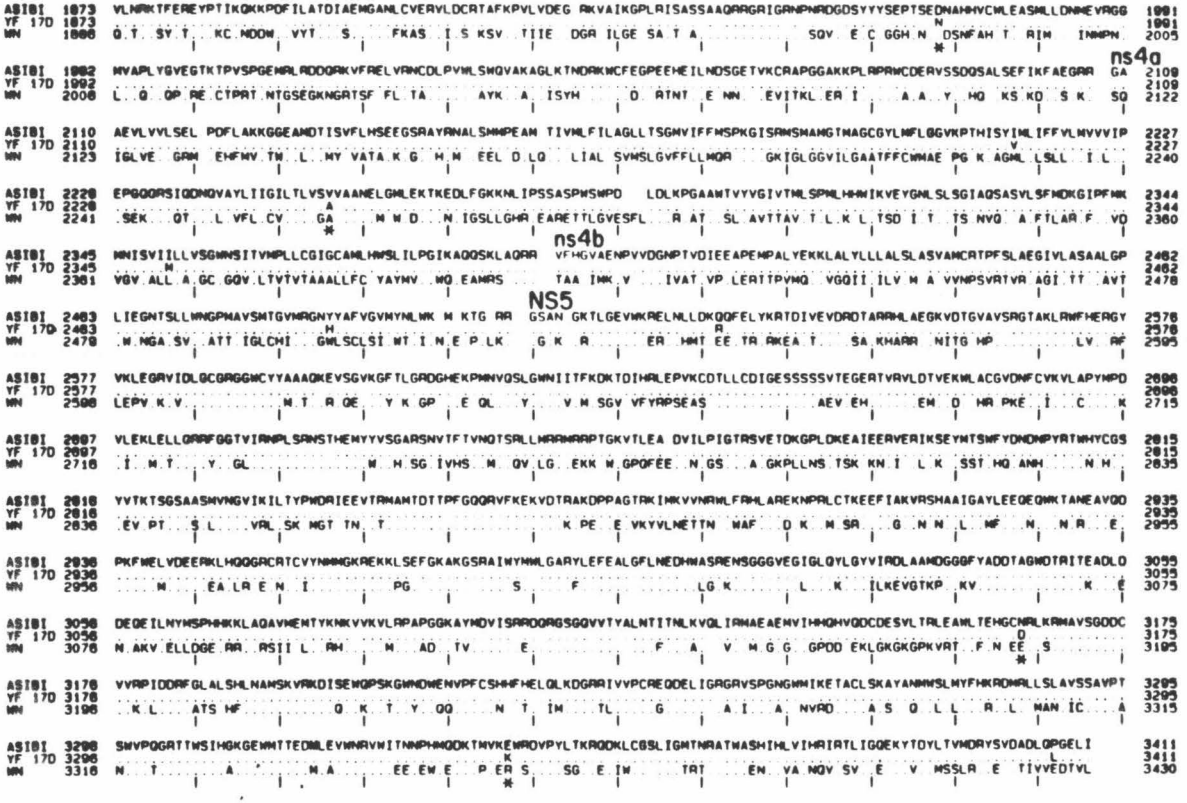



**Figure 1**

Amino acid sequences of the polyproteins encoded by two strains of yellow fever, Asibi and 17D, are shown together with the amino acid sequence of the polyprotein encoded by West Nile virus. Spaces have been introduced for alignment. The sequence for 17D yellow fever is from Rice et al. (1985), that for Asibi yellow fever is from Hahn et al. (1987), and that for West Nile virus is from Wengler et al. (1985) and also Castle et al. (1985, 1986). Asterisks indicate substitutions in which either Asibi or the 17D yellow fever has the amino acid (or homologous amino acid) of West Nile rather than that of the other strain.

virulent and killed the monkey. Blood from this infected monkey was passed once in LLC cells and twice in VERO cells to produce virus for purification and RNA extraction. Cloning of cDNA followed the same procedures used for the 17D strain, and clones representing different regions of the genome were identified by colony hybridization using restriction fragments from the 17D library. The entire genome of the Asibi strain has now been sequenced, once again using at least two clones to obtain each nucleotide position, and the amino acid sequences of the proteins encoded by the Asibi strain of yellow fever virus have been deduced.

The amino acid sequences of the polyproteins specified by Asibi and 17D yellow fever are compared in Figure 1. For purposes of comparison, we also included the amino acid sequence of the polyprotein encoded by West Nile virus. Ignoring clonal differences, the Asibi and 17D polyproteins differ at 32 amino acid positions, a 0.94% divergence in the amino acid sequence. These amino acid substitutions are not uniformly distributed along the polyprotein. In particular, the envelope protein contains 40% of the total amino acid substitutions (12 of 32), a sequence divergence of 2.4%. Many of these changes are not conservative in nature, such as Arg-52 to glycine, Ile-173 to threonine, Thr-200 to lysine, Arg-380 to threonine, and His-390 to proline. One domain of E is particularly intriguing. The domain from Pro-369 to Gly-448 is highly conserved among flaviviruses, as can be seen by comparing the yellow fever and West Nile sequences in Figure 1. In this domain, there are three amino acid substitutions between Asibi and 17D. Thr-380 in Asibi is arginine in 17D, Pro-390 is histidine in 17D, and Ala-416 is threonine in 17D. None of these changes are conservative in nature, and it is of interest that in each case during selection in tissue culture, the sequence of the virulent Asibi yellow fever has changed in 17D to the sequence of West Nile virus. It is tempting to speculate that these changes alter receptor affinities; birds serve as an important reservoir in nature for the West Nile subgroup of mosquito-borne flaviviruses, and as noted, 17D yellow fever arose by selection for rapid growth in chick embryo tissue culture. In any event, it seems unlikely that these three changes represent random changes in the yellow fever E protein during passage in culture.

## SUMMARY AND CONCLUSIONS

We have no way of knowing at present which changes are important for the difference in virulence between the two strains of yellow fever virus. The changes could arise in part from differences in tissue tropism caused by altered affinities of the virus for cellular receptors. Viral surface proteins have been shown to be important in other systems in determining virulence of strains. However, it is also possible that changes in nonstructural proteins could be important determinants of avirulence in this case, and it seems likely from the stability of the 17D strain that more than one change was involved in the attenuation of Asibi. Now that the amino acid differences between the two strains are known, it should be possible to design experimental systems to test the effect of the specific changes on virulence.

## ACKNOWLEDGMENTS

This work was supported by grants from the World Health Organization, by grants AI-20612 and AI-10793 from the National Institutes of Health, and by grant PCM-8316856 from the National Science Foundation.

## REFERENCES

Bell, J.R., R.W. Kinney, D.W. Trent, E.M. Lenches, L. Dalgarno, and J.H. Strauss. 1985. Amino-terminal amino acid sequences of structural proteins of three flaviviruses. *Virology* **143:** 224.

Castle, E., U. Leidner, T. Nowak, G. Wengler, and G. Wengler. 1986. Primary structure of the West Nile flavivirus genome regions coding for all nonstructural proteins. *Virology* **149:** 264.

Castle, E., T. Nowak, U. Leidner, G. Wengler, and G. Wengler. 1985. Sequence analysis of the viral core protein and the membrane associated proteins V1 and NV2 of the flavivirus West Nile virus and of the genome sequence for these proteins. *Virology* **145:** 227.

Hahn, C.S., J.M. Dalrymple, J.H. Strauss, and C.M. Rice. 1987. Comparison of the virulent Asibi strain of yellow fever virus with the 17D vaccine strain derived from it. *Proc. Natl. Acad. Sci.* **84:** (in press).

Rice, C.M., E.M. Lenches, S.R. Eddy, S.J. Shin, R.L. Sheets, and J.H. Strauss. 1985. Nucleotide sequence of yellow fever virus: Implications for flavivirus gene expression and evolution. *Science* **229:** 726.

Rice, C.M., L. Dalgarno, R. Galler, Y.S. Hahn, E.G. Strauss, and J.H. Strauss. 1986a. Molecular cloning of flavivirus genomes for comparative analysis and expression. In *Proceedings of the International Symposium,* June 1986, Giessen (ed. H. Bauer et al.). Springer-Verlag, Berlin. (In press.)

Rice, C.M., R. Aebersold, D.B. Teplow, J. Pata, J.R. Bell, A.V. Vorndam, D.W. Trent, M.W. Brandriss, J.J. Schlesinger, and J.H. Strauss. 1986b. Partial N-terminal amino acid sequences of three nonstructural proteins of two flaviviruses. *Virology* **151:** 1.

Stokes, A., J.H. Bauer, and N.P. Hudson. 1928. Experimental transmission of yellow fever to laboratory animals. *Am. J. Trop. Med.* **8:** 103.

Strode, G.K. 1951. *Yellow fever.* McGraw-Hill, New York.

Wengler, G., E. Castle, U. Leidner, T. Nowak, and G. Wengler. 1985. Sequence analysis of the membrane protein V3 of the flavivirus West Nile virus and of its gene. *Virology* **147:** 264.

# Part III

## Evolution of Alphaviruses and Flaviviruses

# Chapter 6

## Western Equine Encephalitis Virus Is a Recombinant Virus

This chapter is in press in *Proc. Natl. Acad. Sci. USA*

Classification: EVOLUTION

# Western Equine Encephalitis Virus is a Recombinant Virus

Key words: (RNA recombination/Alphavirus/evolution of RNA viruses)

Chang S. Hahn, Shlomo Lustig*, Ellen G. Strauss, and James H. Strauss

Division of Biology - 156-29, California Institute of Technology,
Pasadena, California 91125, USA

Communicated by: James Bonner

*Present address: Israel Institute for Biological Research, Ness-Ziona, 70450, P. O. Box 19, Israel

Address correspondence to: James H. Strauss, Division of Biology 156-29, California Institute of Technology, Pasadena, CA 91125, Phone (818) 356-4903

ABSTRACT The alphaviruses are a group of 26 mosquito-borne viruses that cause a variety of human diseases. Many of the new world alphaviruses cause encephalitis whereas the old world viruses more typically cause fever, rash, and arthralgia. The genome is a single-stranded non-segmented RNA molecule of plus polarity; it is about 11,700 nucleotides in length. Several alphavirus genomes have been sequenced in whole or in part, and these sequences demonstrate that alphaviruses have descended from a common ancestor by divergent evolution. We have now obtained the sequence of the 3' terminal 4288 nucleotides of the RNA of the new world alphavirus Western equine encephalitis virus (WEE). Comparisons of the nucleotide and amino acid sequences of WEE with those of other alphaviruses clearly show that WEE is recombinant. The sequences of the capsid protein and of the (untranslated) 3' terminal 80 nucleotides of WEE are closely related to the corresponding sequences of the new world alphavirus Eastern equine encephalitis (EEE) virus, whereas the sequences of glycoproteins E2 and E1 of WEE are more closely related to those of an old world virus, Sindbis (SIN) virus. Thus, WEE appears to have arisen by recombination between an EEE-like virus and a SIN-like virus to give rise to a new virus with the encephalogenic properties of EEE but the antigenic specificity of SIN. There has been speculation that recombination might play an important role in the evolution of RNA viruses. The current finding that a widespread and successful RNA virus is recombinant provides support for such an hypothesis.

## INTRODUCTION

The 26 members of the alphavirus genus of the family *Togaviridae* are mosquito-borne viruses that form an important group of disease agents (1-3). The new world alphaviruses include Western equine encephalitis (WEE) and Eastern equine encephalitis (EEE) viruses, both of which are capable, as their names imply, of causing encephalitis in man, as well as of causing severe disease in horses. WEE has a wide geographic distribution, being found from western Canada to Mexico and, discontinuously, to Argentina. WEE is transmitted in the western United States by the mosquito *Culex tarsalis*; birds serve as an important vertebrate reservoir. In the eastern United States WEE is replaced by Highlands J (HJ) virus, whose primary vector is *Culiseta melanura*. From serological studies (3,4) and from limited sequencing studies (5,6), WEE and HJ are known to be very closely related, and HJ can be considered to be a strain of WEE (2). In the eastern United States the range of HJ overlaps that of EEE, whose primary vector is also *Cs. melanura*. Other new world alphaviruses include Venezuelan equine encephalitis virus (VEE), found in Central and South America; Fort Morgan virus, found in Colorado; and Aura virus, found in South America.

The old world alphaviruses include Sindbis virus (SIN), the prototype alphavirus; Semliki Forest virus (SF); Chikungunya virus (CHIK); O'Nyong-nyong virus (ONN); and Ross River virus (RR). SIN and SF have been intensively studied as models for alphavirus replication (7). SIN is widely distributed, being found in Europe, India, southeast Asia, Australia, and Africa. Close relatives of this virus, such as Ockelbo virus in Europe (8) and Babanki virus in Africa, cause disease in man characterized by fever, rash, and arthritis. CHIK and ONN have caused large epidemics in Africa of a dengue-like disease also characterized by fever, rash, and arthralgia. RR is the causative agent of epidemic polyarthritis in Australia and the South Pacific.

The single-stranded RNA genome of the alphaviruses is nonsegmented and about 11,700 nucleotides in length. Complete or partial RNA sequences have been obtained for SIN (9), SF (10-12), RR (13), EEE (14), and VEE (15). Comparison of these nucleotide sequences, and more importantly, their encoded amino acid sequences, have demonstrated that the alphaviruses are related by linear descent from a common ancestor (7). The relationships among the various alphaviruses derived from the sequencing studies are compatible, for the most part, with those derived from studies of serological cross-reactivity, which depends only upon antigenic epitopes in the structural proteins. In serological studies, however, WEE has always been something of a puzzle. It is a new world virus that often causes encephalitis, but serologically it is most closely related to SIN, an old world alphavirus not normally associated with encephalitis. In order to explore the relationship of WEE to other alphaviruses, we have now obtained the nucleotide sequence of the 3' terminal 4,288 nucleotides of the WEE genome. Comparison of this sequence and of the encoded amino acid sequences with those from other alphaviruses suggests that WEE arose by a recombination event between an EEE-like virus and a SIN-like virus.

## MATERIALS AND METHODS

**Virus RNA Preparation.** WEE RNA (strain BFS1703) was obtained from Drs. Mark Stanley and James Hardy of the University of California, Berkeley. The BFS1703 strain of WEE was isolated from *Culex tarsalis* mosquitos in July, 1953 in Kern county, California (16). The virus had been passed twice by IC inoculation of suckling mice and four times (including three plaque isolations) in Vero cells. For RNA preparation, virus was grown in Vero cells and purified by pelleting onto a 30% sucrose cushion followed by isopycnic banding in Nicodenz. Purified virus was pelleted, dissociated in SDS, and the RNA extracted by phenol-chloroform

treatment. Following ethanol precipitation, the RNA was sedimented in a discontinuous sucrose gradient, the RNA band recovered, and the RNA was concentrated by ethanol precipitation.

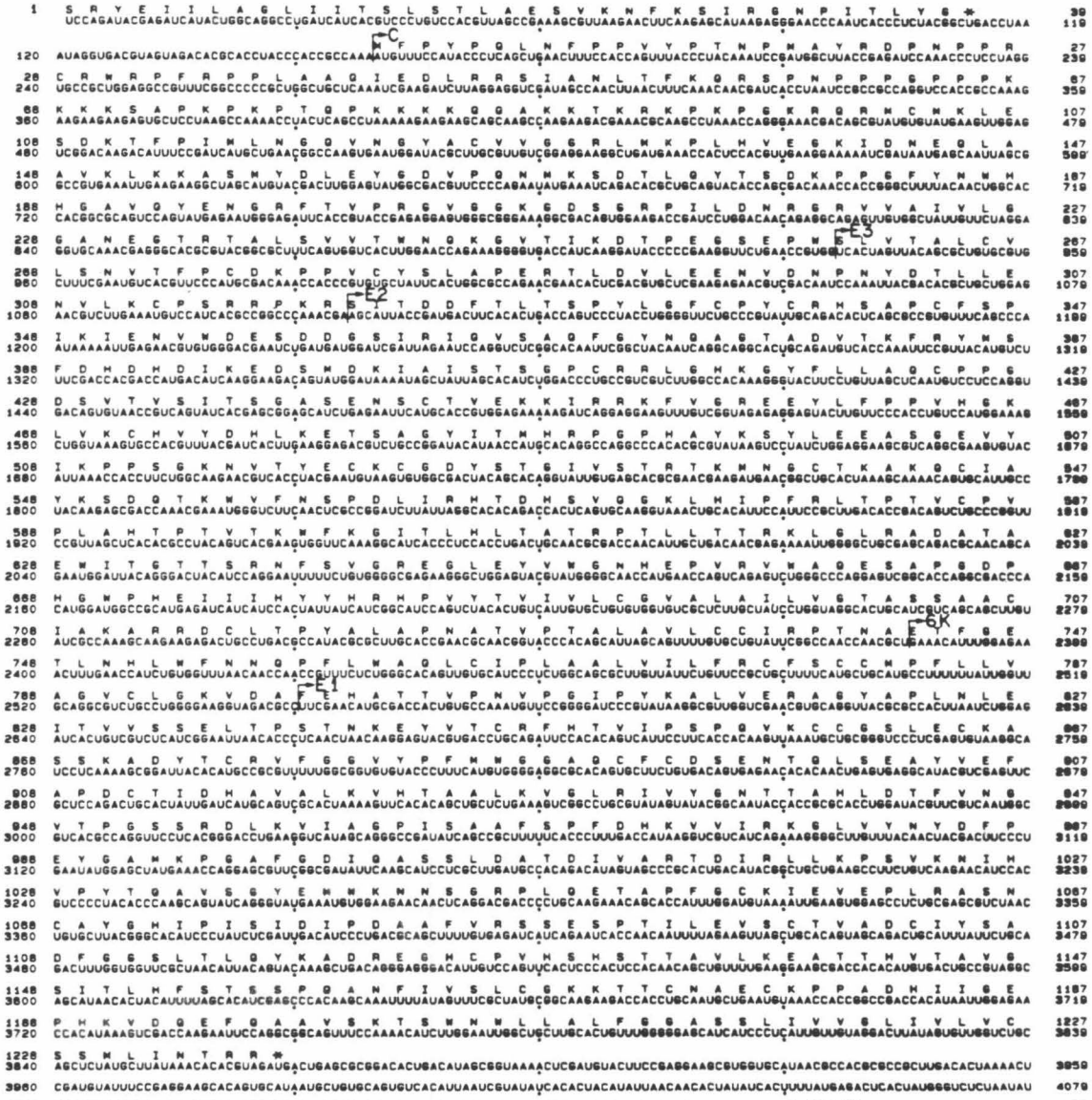
**Cloning and Sequencing.** Clones containing the 3' terminal 4288 nucleotides of WEE RNA [plus a variable length of poly(A)] were obtained using an oligo(dT)-tailed vector as a primer as described (17). Clones were sequenced using the chemical sequencing method (18,19). More than 99% of the nucleotide sequence was obtained on two independent clones in order to rule out cloning artifacts (17) and to detect possible clonal variation.

## RESULTS

**Partial Sequence of WEE RNA.** The translated sequence of the 3' terminal 4,170 nucleotides of the WEE genome is shown in Fig. 1. This sequence begins in the region encoding the carboxy terminus of nonstructural protein 4, continues through the junction region between the nonstructural and structural proteins [this region contains the start of the subgenomic mRNA that is translated to give the structural proteins and is also believed to contain nucleotide sequence elements required for transcription of this subgenomic RNA (20)], progresses through the coding sequence of the three structural proteins of the virus (a nucleocapsid protein and two envelope glycoproteins E2 and E1), and finally through the 3' terminal untranslated sequence which ends in a poly(A) tract.

We have previously sequenced the N termini of the three structural proteins of the McMillan strain of WEE (isolated in 1941 in Canada from the brain of a fatal human case) and thus established the start points of the structural proteins (21). Comparison of the amino acid sequence of the McMillan strain with that deduced here for the BFS1703 strain (isolated from mosquitos in 1953 in California) reveals four amino acid differences in 142 amino acids for which

**Fig. 1** Sequence of the 3' terminal 4170 nucleotides of WEE RNA (strain BFS1703). The start points of the structural proteins are indicated. Astericks indicate the termination codons for the nonstructural and structural polyproteins. Two independent clones were sequenced and only one clonal difference was found: the GAC encoding Asp 72 of E2 was replaced by UAC encoding Tyr in the second clone.

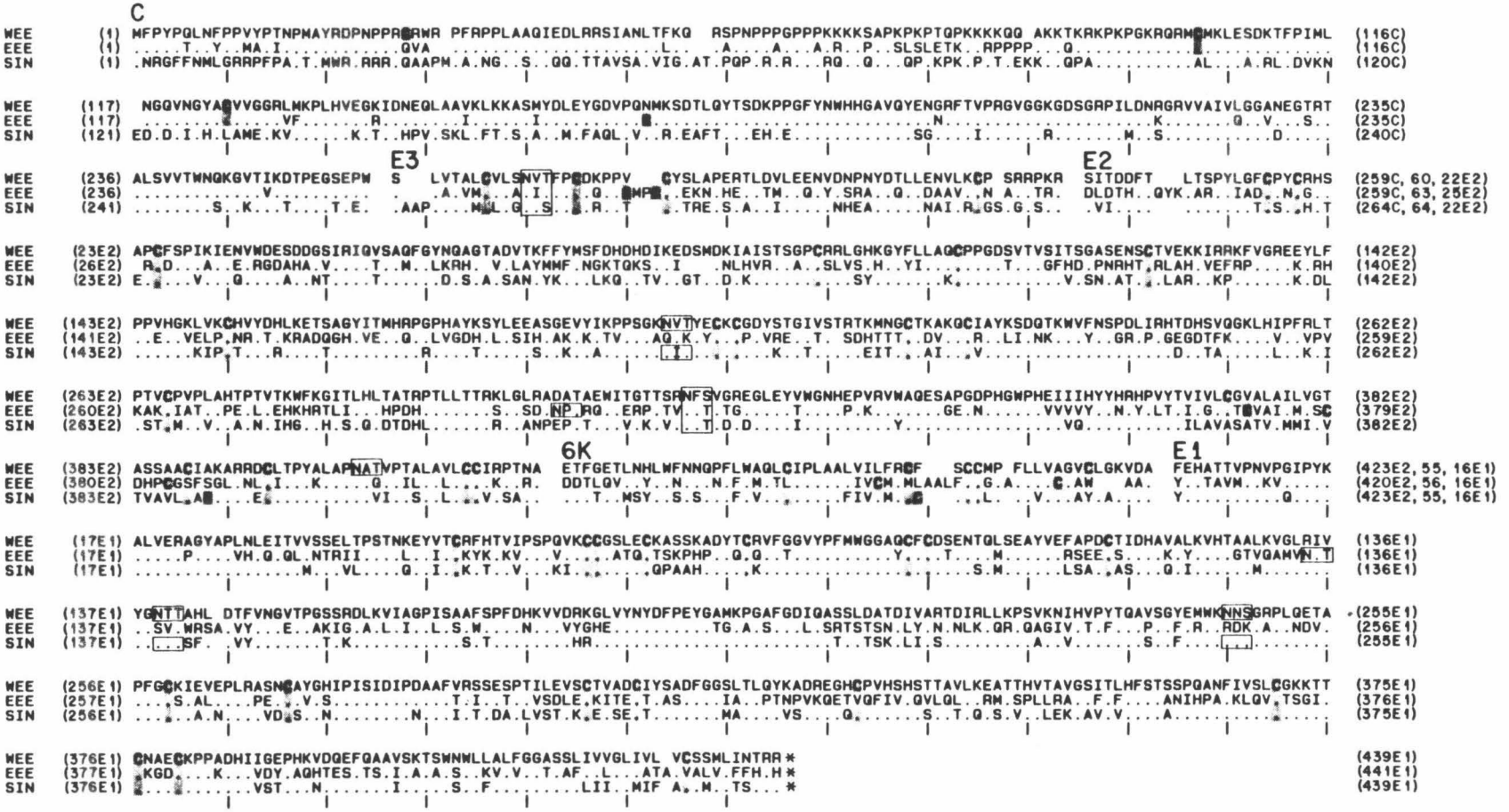

C
E1
E2
E3
6K

comparison is possible (1 difference in C, 1 in E2, and 2 in E1). However, reevaluation of the original data for the McMillan strain suggests that the apparent difference in the capsid proteins may result from a misscall in the McMillan sequence, and that there are no differences in the capsid proteins of these two strains in the region for which comparison is possible. The amino acid sequence divergence between the two strains is 2.8% (or 2.1% if the apparent difference in the capsid proteins is ignored). We have also reported the sequence of the 3' terminal 351 nucleotides of McMillan RNA (6). Comparison of this sequence with that for BFS1703 shows three nucleotide substitutions and one nucleotide deletion (in McMillan) between these two strains, a nucleotide sequence divergence of 1.1%. These comparisons establish that the widely studied McMillan strain (the prototype WEE virus), and the BFS1703 strain are the same virus. Since these two strains were isolated 12 years apart in different geographic areas, the calculated rate of divergence of WEE in nature is, at most, 0.1-0.2% per year, which is low in comparison to rates of divergence that have been established for several RNA viruses (22,23).

**WEE is a Recombinant.** The amino acid sequences of the WEE structural proteins are compared to those of EEE and of SIN in Fig. 2. Inspection of this figure clearly reveals that the WEE capsid protein C is most closely related to that of EEE whereas the glycoproteins E2 and E1 are more closely related to the corresponding proteins of SIN.

The relationships among the proteins of these viruses are summarized in Table 1. The N terminal and C terminal domains of the capsid protein are considered separately because of the fact that the C termini of all alphavirus capsid proteins are closely related. The N terminal 132 amino acids of the WEE and EEE capsid proteins share 78% amino acid sequence identity whereas the corresponding figures for WEE and SIN and for EEE and SIN are 39% and 36%,

**Fig. 2** Comparison of the amino acid sequences of the structural proteins of WEE, EEE, and SIN. A dot in the EEE or SIN sequence means that the amino acid is the same as that of WEE on the first line. Gaps have been introduced for alignment. The sequence for WEE is from Fig. 1, that for EEE is from (14), and that for SIN is from (9) Potential glycosylation sites are boxed and cysteines are highlighted with dotted overlay.


C
E3
E2
6K
E1
WEE (1) MFPYPQLNFPPVYPTNPMAYRDPNPPRGRWR PFRPPLAAQIEDLRRSIANLTFKQ RSPNPPPGPPPKKKKSAPKPKPTQPKKKKQQ AKKTKRKPKPGKRQRMGMKLESDKTFPIML (116C)
EEE (1) .....T..Y..MA.I.....QVA.....L...A...A..A.R..P..SLSLETK..RPPPP...Q..... (116C)
SIN (1) .NRGFFNMLGRRPFPA.T.MWR.RRR.QAAPM.A.NG..S..QQ.TTAVSA.VIG.AT.PQP.R.R...RQ..Q...QP.KPK.P.T.EKK..QPA.....AL...A.RL.DVKN (120C)
WEE (117) NGQVNGYAGVVGGRLMKPLHVEGKIDNEQLAAVKLKKASMYDLEYGDVPQNMKSDTLQYTSDKPPGFYNWHHGAVQYENGRFTVPRGVGGKGDSGRPILDNRGRVVAIVLGGANEGTRT (235C)
EEE (117) .....V.....R.....I.....I.....N.....K.....Q..V...S.. (235C)
SIN (121) ED.D.I.H.LAME.KV.....K.T..HPV.SKL.FT.S.A..M.FAQL.V..R.EAFT...EH.E.....SG....I.....R.....M..S.....D..... (240C)
WEE (236) ALSVVTWNQKGVTIKDTPEGSEPN S LVTALGCVLSNVTFPCDKPPV CYSLAPERTLDVLEENVDNPNYDTLLENVLKCP SRRPKR SITDDFT LTSPYLGFCPYCRHS (259C, 60, 22E2)
EEE (236) .....V..... .A.VM..A.I...Q..MPR..EKN.HE..TM..Q.Y.SRA..Q..DAAV..N A..TR. DLDTH..QYK.AR..IAD..N.G.. (259C, 63, 25E2)
SIN (241) .....S..K...T....T.E. .AAP....MVL.G..S...R..T ..TRE.S.A..I.....NHEA.....NAI.R.GS.G.S.. .VI.... .....T.S..H.T (264C, 64, 22E2)
WEE (23E2) APCFSPIKIENVWDESDDGSIRIQVSAQFGYNQAGTADVTKFFYMSFDHDHDIKEDSMDKIAISTSGPCRRLGHKGYFLLAQCPPGDSVTVSITSGASENSCTVEKKIRRKFVGREEYLF (142E2)
EEE (26E2) R..D...A..E.RGDAHA.V....T..M..LKRH. V.LAYMMF.NGKTQKS..I .NLHVR..A..SLVS.H..YI.....T...GFHD.PNRHT..RLAH.VEFRP....K.RH (140E2)
SIN (23E2) E.....V...Q....A..NT....T.....D.S.A.SAN.YK...LKQ..TV..GT..D.K.....SY.....K.....V.SN.AT..LAR..KP.....K.DL (142E2)
WEE (143E2) PPVHGKLVKCHVYDHLKETSAGYITMHRPGPHAYKSYLEEASGEVYIKPPSGKNVTYECKCGDYSTGIVSTRTKMNGCTKAKGCIAYKSDQTKWVFNSPDLIRHTDHSVQGKLHIPFRLT (262E2)
EEE (141E2) ..E..VELP.NR.T.KRADQGH.VE..Q..LVGDH.L.SIH.AK.K.TV...AQ.K.Y...P.VRE..T. SDHTTT..DV...R..LI.NK...Y..GR.P.GEGDTFK....V..VPV (259E2)
SIN (143E2) ......KIP.T...R....T.....R....T.....S..K..A.....I.....K..T....EIT.....AI...V.....D..TA.....L..K.I (262E2)
WEE (263E2) PTVCPVPLAHTPTVTKWFKGITLHLTATRPTLLTTRKLGLRADATAEWITGTTSRNFSVGREGLEYVWGNHEPVRVWAQESAPGDPHGWPHEIIIHYYHRHPVYTVIVLCGVALAILVGT (382E2)
EEE (260E2) KAK.IAT..PE.L.EHKHRTLI...HPDH.....S..SD.NP.RQ..ERP.TV..T.TG.....T....P.K.....GE.N.....VVVVY..N.Y.LT.I.G..TGVAI.M.SC (379E2)
SIN (263E2) .ST.M..V..A.N.IHG..H.S.Q.DTDHL.....R..ANPEP.T...V.K.V..T.D.D....I.....Y.....VQ.....ILAVASATV.MMI.V (382E2)
WEE (383E2) ASSAACIAKARRDCLTPYALAPNATVPTALAVLCCIRPTNA ETFGETLNHLWFNNGPFLWAGLCIPLAALVILFRCF SCCMP FLLVAGVCLGKVDA FEHATTVPNVPGIPYK (423E2, 55, 16E1)
EEE (380E2) DHPCGGSFSGL.NL.I...K.....Q..IL..L...K..R. DDTLQV..Y..N...N.F.M.TL.....IVCM.MLAALF..G.A....C.AW AA. Y..TAVM..KV..... (420E2, 56, 16E1)
SIN (383E2) TVAVL.AG....E.....VI..S..L...V.SA.. ...T..MSY..S.S...F.V.....FIV.M..G ...L. ..V...AY.A.... Y.....Q.... (423E2, 55, 16E1)
WEE (17E1) ALVERAGYAPLNLEITVVSSELTPSTNKEYVTCRFHTVIPSPQVKCCGSLECKASSKADYTCRVFGGVYPFMWGGAQCFCDSENTQLSEAYVEFAPDCTIDHAVALKVHTAALKVGLRIV (136E1)
EEE (17E1) .....P....VH.Q.QL.NTRII....L..I..KYK.KV...V.....ATQ.TSKPHP..Q.Q..T.....Y....T....M.....RSEE.S....K.Y....GTVQAMVN.T (136E1)
SIN (17E1) .....M...VL....Q..I..K.T..V...KI.....QPAAH....K.....S.M.....LSA..AS...Q.I.....M..... (136E1)
WEE (137E1) YGNTTAHL DTFVNGVTPGSSRDLKVIAGPISAAFSPFDHKVVDRKGLVYNYDFPEYGAMKPGAFGDIQASSLDATDIVARTDIRLLKPSVKNIHVPYTQAVSGYEMWKNNSGRPLQETA (255E1)
EEE (137E1) ..SV.WRSA.VY...E..AKIG.A.L.I..L.S.M....N...VYGHE.....TG.A.S...L.SRTSTSN.LY.N.NLK.QR.QAGIV.T.F...P..F.R..RDK.A..NDV. (256E1)
SIN (137E1) ..[...]SF. .VY.....T.K.....S.T.....HR.....T..TSK.LI.S.....A..V.....S..F....[...]..... (255E1)
WEE (256E1) PFGCKIEVEPLRASNCAYGHIPISIDIPDAAFVRSSESPTILEVSCTVADCIYSADFGGSLTLQYKADREGHCPVHSHSTTAVLKEATTHVTAVGSITLHFSTSSPQANFIVSLCGKKTT (375E1)
EEE (257E1) ...S.AL....PE..V.S.....T.I..T..VSDLE.KITE.T.AS....IA..PTNPVKQETVQFIV.QVLQL..RM.SPLLRA..F.F....ANIHPA.KLQV.TSGI. (376E1)
SIN (256E1) ...A..A.N....VD.S..N.....N...I.T.DA.LVST.K.E.SE.T.....MA....VS....Q.....S..T.Q.S.V..LEK.AV.V....A.....A..... (375E1)
WEE (376E1) CNAECKPPADHIIGEPHKVDQEFQAAVSKTSWNWLLALFGGASSLIVVGLIVL VCSSMLINTRR* (439E1)
EEE (377E1) .KGD....K...VDY.AQHTES.TS.I.A.A.S..KV.V..T.AF..L...ATA.VALV.FFH.H* (441E1)
SIN (376E1) .....VST...N.....I.....S..F.....LII..MIF A..M..TS...* (439E1)

respectively. The C terminal domain of the capsid protein of WEE is also more closely related to EEE (91% sequence identity) than it is to SIN (69%). The relationships are reversed in the case of the envelope proteins. Overall, WEE and SIN share 71% amino acid sequence similarity in the envelope glycoproteins compared to only 47% between WEE and EEE or 46% between EEE and SIN. Figures for the C terminal domain of nsP4 are also included. Although in general this protein is highly conserved among alphaviruses, its C terminal domain is more variable and WEE and EEE are much more closely related in this region than are WEE and SIN or EEE and SIN.

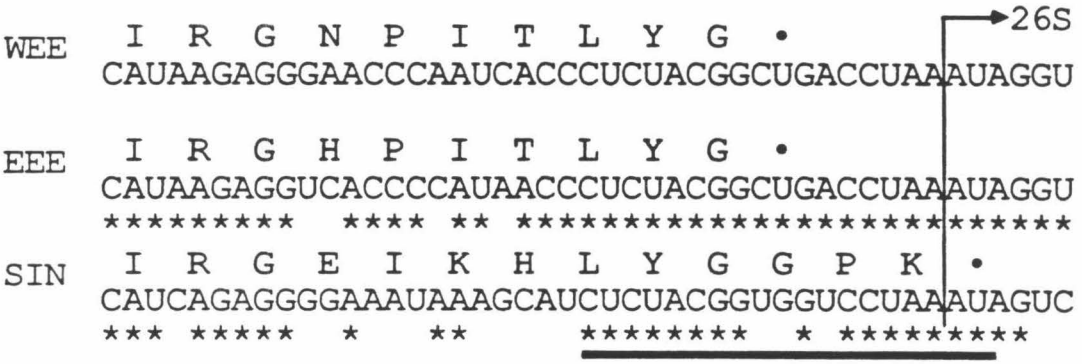
Also included in this table are comparisons with another alphavirus, VEE, to illustrate that alphaviruses in general differ from one another in a uniform and consistent way. Sequence data for SF or RR lead to similar results (not shown). WEE is exceptional in that it is closely related to SIN in the region of the genome encoding E1 and E2, but to EEE in other regions.

Nucleotide sequences in the carboxyterminal region of nsP4 and in the junction region between structural and nonstructural proteins, which are believed to encode important signals for transcription of a subgenomic RNA (20), are compared for the three viruses in Fig. 3a. Note that EEE and WEE nucleotide sequences are very similar to one another and that, in particular, the sequences flanking the start of the subgenomic 26S RNA are identical. The sequence of SIN in this region is similar but not identical. Notice also that the nsP4 proteins of EEE and WEE terminate at the same residue whereas the SIN protein terminates downstream.

The sequences at the 3' termini of WEE, EEE, and SIN are shown in Fig. 3b. The 3' terminal 19 nucleotides have been proposed to form an important element in alphavirus RNA replication because they are highly conserved among members of this genus (6), and this sequence element (underlined in Fig. 3b) is

**Fig. 3** Comparison of the nucleotide sequences in the junction regions of EEE, WEE, and SIN (panel a) or at the 3' end of the RNAs (panel b). Asterisks denote conserved nucleotides. The heavy underlines denote conserved nucleotide sequences in the alphaviruses, 21 nucleotides around the start of the subgenomic 26S RNA (20) and 19 nucleotides at the 3' end of the RNA (6) which are believed to form important regulatory elements for RNA transcription (7). The termination codons that end the nonstructural open reading frames are marked with black circles.

a. JUNCTION CONSERVED SEQUENCE




b. 3' END CONSERVED SEQUENCE

**Table 1** Percent Sequence Identity Among WEE, EEE, SIN and VEE Proteins

| | WEE/EEE | WEE/SIN | EEE/SIN | EEE/VEE | SIN/VEE | WEE/VEE |
|---|---|---|---|---|---|---|
| nsP4 (C-Terminus) | **70%** | 35% | 40% | | | |
| Capsid | | | | | | |
| N-Terminus[1] | **78** | 39 | 36 | 42 | 27 | 49 |
| C-Terminus[1] | **91** | 69 | 64 | 76 | 61 | 77 |
| Overall | **85** | 53 | 50 | 59 | 44 | 63 |
| Envelope | | | | | | |
| E3 | 50 | **58** | 42 | 59 | 49 | 56 |
| E2 | 44 | **68** | 42 | 46 | 40 | 41 |
| 6K | 44 | **67** | 45 | 54 | 40 | 40 |
| E1 | 49 | **76** | 51 | 58 | 51 | 50 |
| Overall | 47 | **71** | 46 | 53 | 46 | 46 |

1 N-terminus refers to amino acids 1 to 132 of the Sindbis capsid protein or the corresponding positions in the aligned files in Figure 2. C-terminus includes the remaining amino acids in the capsid proteins in the aligned files.

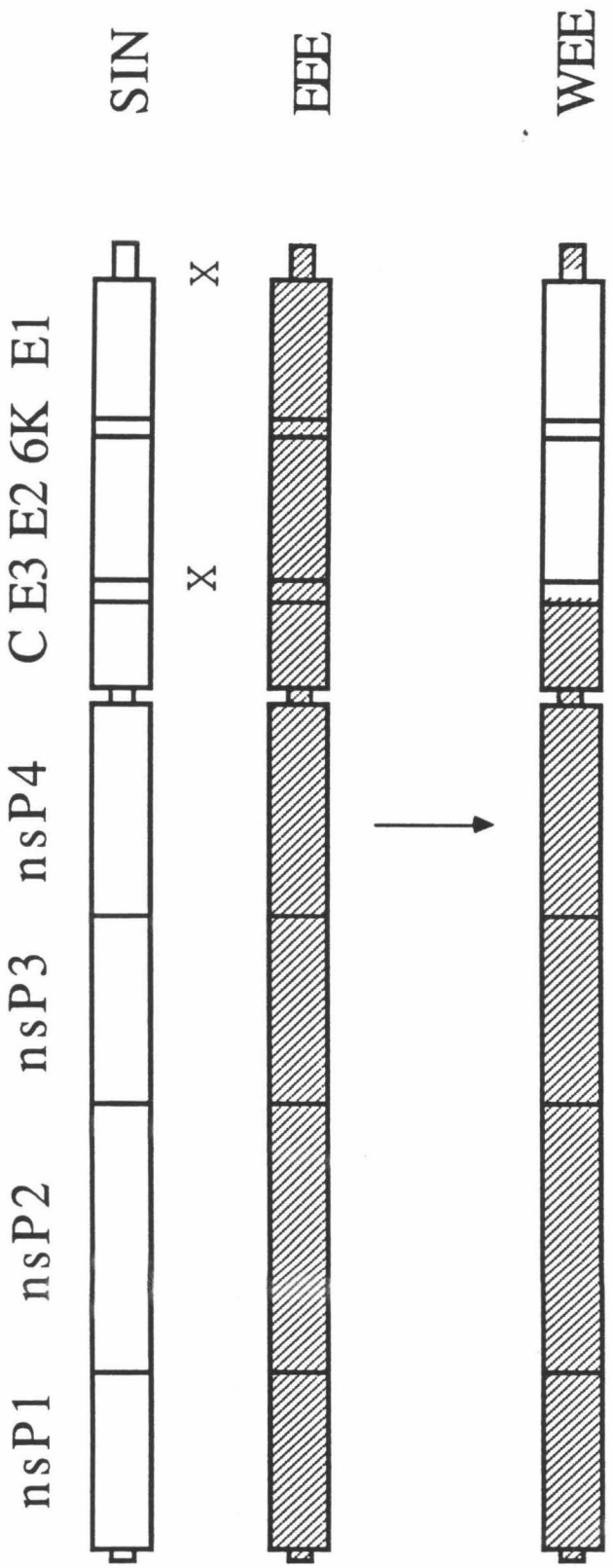
invariant among these three viruses with the exception of the sixth nucleotide from the end. The nucleotides upstream of this are AU rich and not particularly conserved among alphaviruses, but in this domain the sequences of WEE and EEE are almost identical whereas that of SIN is more variable.

These results show that within the region examined the WEE nucleotide sequence is recombinant, with both the 5' and 3' ends derived from an EEE-like virus and the intervening glycoprotein genes derived from a SIN-like virus. We presume that the 5' terminal two-thirds of the genome, which has not yet been sequenced, is also derived from the EEE-like virus. Partial support for this comes from our previous finding that the 5' terminal sequence of HJ is similar to that of EEE (5).

**The Recombination Events.** Our interpretation of the sequence information is shown schematically in Fig. 4. This figure is included in part to illustrate the structure of the alphavirus genome, as well as to describe the most likely scenario for the origin of WEE.

In this model, close inspection of the aligned sequences in Fig. 2 suggests that the 5' crossover occurred in E3. Gaps must be introduced into the amino acid sequences to align them and the two gaps of three amino acids each in E3 are of particular interest. The first gap, following residue 1, is shared by WEE and EEE, and upstream of this WEE and EEE are in almost perfect alignment (only one gap of one amino acid must be introduced into each sequence to maintain alignment), whereas several gaps must be introduced to keep the SIN sequence aligned. Conversely, the gap following residue 21 of WEE E3 is shared by SIN and WEE, and downstream of this the SIN and WEE sequences are in almost perfect register (only one gap of one amino acid in E3 is required to maintain alignment), whereas numerous gaps

**Fig. 4** Schematic representation of the recombination event that produced WEE. The genome structures of SIN, EEE, and WEE are diagrammed, and the crossover points to produce WEE are indicated.


nsP1
nsP2
nsP3
nsP4
C E3 E2 6K E1
X
X
SIN
EEE
WEE

are required to keep the EEE sequence in register. This suggests that the recombination event occurred between these two gaps of three amino acids each in E3, which is compatible with the sequence similarities exhibited by the capsid proteins and the glycoproteins in Table 1.

The 3' crossover appears to have occurred in the 3' untranslated region. The 60 nucleotides of WEE RNA following the structural protein stop codon are very similar to the SIN sequence, whereas the last 80 nucleotides of the RNA are similar to EEE, with no sequence similarity detectable in between. Although a double crossover seems inherently less likely than a single crossover, the presence of important replication signals at the 3' end may require such an event to produce viable (or at least efficiently replicating) virus (6,7).

There is a formal possibility that WEE is one of the parental viruses in a cross that resulted in the reciprocal recombinants SIN and EEE. Because RNA recombination is believed to occur by a copy choice mechanism, however, in which reciprocal recombinants are not produced (24), and because of the apparent rarity of viable recombinant viruses, this possibility appears remote.

**Interaction of the Nucleocapsid and Glycoproteins During Virus Budding.** Alphaviruses mature when preassembled nucleocapsids, which are icosahedral structures consisting of 180 copies of the nucleocapsid protein and one molecule of the virus RNA, acquire an envelope by budding through the plasma membrane (25,26). The envelope consists of a lipid bilayer derived from the host cell in which are embedded two virus-encoded glycoproteins, E2 and E1. The nucleocapsid and the glycoproteins are thought to interact specifically with one another, so as to exclude nonvirus proteins from the structure; the free energy for driving virus budding is

derived from these specific interactions. During evolution, certain domains of the glycoproteins of a particular virus must have been selected for maximal specific interaction with the capsid of that virus. In a recombinant virus which contains the capsid protein from one virus and the glycoproteins from another, the interactions during budding might not be optimal. During passage of such a recombinant virus, selection pressure would favor variants in which the nucleocapsid and glycoprotein interactions were improved. It is thus of considerable interest that there are only seven amino acid differences between WEE and EEE in the C terminal 104 amino acids of the capsid protein, and for 6 of these WEE has the SIN amino acid (Fig. 2). This suggests that this domain of the capsid protein interacts with the glycoproteins during virus assembly, and that following the recombination event selection pressure has led to some of the EEE capsid amino acids being replaced with SIN amino acids to allow more efficient interaction with the SIN glycoproteins. Conversely, in the C terminal 16 amino acids of E2 there are 6 amino acid differences between WEE and SIN, and for 4 of these WEE has the EEE amino acid, suggesting by the same logic that this domain of E2 interacts with the capsid during budding. Other examples can be found in other regions of the structural proteins. The hypothesis that these are in fact involved in capsid-glycoprotein interactions can now be tested by site specific mutagenesis, using a cDNA clone of Sindbis virus from which infectious RNA can be transcribed *in vitro* (27). We are in the process of testing this model by constructing site-specific mutants in the domains identified by this analysis.

## DISCUSSION

**The Origin of WEE Virus.** The two parents of WEE and the time of the recombination event cannot be determined at the current time. As described earlier, the McMillan strain of WEE isolated in 1941 in Canada and the BFS1703 strain isolated in 1953 in California are clearly strains of the same virus. They have nearly identical capsid proteins, glycoproteins E2 and E1, and 3' terminal sequences. Thus the recombination event could not have occurred during passage of the virus in culture, as this would have required the identical recombination event to have occurred twice, in different laboratories (see, for example, Fig. 4). By the same logic, the recombination event must have predated the isolation of the McMillan strain of WEE in 1941. Furthermore, as noted earlier, all of the sequence information obtained so far is compatible with the hypothesis that the recombinant virus arose before the separation of WEE and HJ viruses. On the other hand, the N terminal portions of the capsid proteins of WEE and EEE are very similar, a domain not well conserved among alphaviruses (28). This domain is lysine- and arginine-rich, has a high proportion of proline, and appears to interact electrostatically with the virus RNA to stabilize the capsid structure. Thus the similarity in the WEE and EEE sequences, together with the fact that RNA viruses diverge rapidly (21), suggests that the recombination event must be relatively recent. We propose that one of the parents was EEE itself. The sequence similarities with SIN in the envelope protein regions are not as pronounced and suggest that the second parent was not SIN itself, but a relative of it. Because WEE and EEE are new world viruses, we suggest that the recombination event occurred in the new world between EEE, or an immediate ancestor of it, and a SIN-like virus that has yet to be identified. It seems most likely that the recombination event would take place in the

mosquito vector, in which the virus sets up a persistent life-long infection. EEE and HJ overlap in geographic ranges and mosquito vector. Thus HJ might represent the ancestral recombinant virus which radiated to produce WEE.

**Recombination in RNA Virus Evolution.** There has been a great deal of speculation about the importance of recombination in the evolution of RNA viruses (for recent reviews see 29,30). In segmented RNA viruses, reassortment of individual genome segments during mixed infection, a form of recombination equivalent to the shuffling of chromosomes in diploid creatures, is readily demonstrated in cell culture. Reassortment has been well documented as a major mechanism for generating new pandemic strains of influenza virus (31,32), and it can be argued that the ability to undergo ready recombination conveys significant selective advantage on RNA viruses with segmented genomes. Among the nonsegmented RNA viruses, recombination has been in general more difficult to demonstrate, but it has been shown to occur in the picornaviruses (33,34), the coronaviruses (35,36), and the bromoviruses (37), although not before now in the alphaviruses. In poliovirus, recombination occurs by a copy choice mechanism during RNA replication (24), and it is assumed that all RNA recombination (as opposed to reassortment) occurs by this mechanism. Although well established in principle, evidence for the importance of recombination in nature as a mechanism that leads to successful new strains is limited. In the case of poliovirus, recombination has been shown to occur in vaccinees who have simultaneously received high doses of three attenuated viruses (34), but this is a somewhat artificial system. The finding that WEE, a virus with a wide geographic range, is a naturally occurring recombinant lends support to the hypothesis that RNA recombination is an important force in the evolution of

RNA viruses. In this particular case it has given rise to a new virus that combines the disease-causing potential of EEE with new antigenic properties from a SIN-like virus.

**Acknowledgments.** We thank Drs. M. Stanley and J. Hardy for the WEE RNA used in this project. This work was supported by Grants AI20612 and AI10793 from NIH and Grant DMB86-17372 from NSF.

## References


1. Griffin, D. E. (1986) in *The Togaviridae and Flaviviridae* (Plenum Press, New York), pp. 209-249.
2. Chamberlain, R. W. (1980) in *The Togaviruses* (Academic Press, New York), pp. 175-227.
3. Calisher, C. H., Shope, R. E., Brandt, W., Casals, J., Karabatsos, N., Murphey, F. A., Tesh, R. B. & Wiebe, M. E. (1980) *Intervirology* **14**, 229-232.
4. Hayes, C. G., & Wallis, R. C. (1977) *Adv. Virus Res.* **21**, 37-83.
5. Ou, J.-H., Strauss, E. G. & Strauss, J. H. (1983) *J. Mol. Biol.* **168**, 1-15.
6. Ou, J.-H., Trent, D. W. & Strauss, J. H. (1982) *J. Mol. Biol.* **156**, 719-730.
7. Strauss, E. G. & Strauss, J. H. (1986) in *The Togaviridae and Flaviviridae* (Plenum Press, New York), pp. 35-90.
8. Niklasson, B., Espmark, A., LeDuc, J. W., Gargan, T. P., Ennis, W. A., Tesh, R. B., & Main, A. J. (1984) *Am. J. Trop. Med. Hyg.* **33**, 1212-1217.
9. Strauss, E. G., Rice, C. M. & Strauss, J. H. (1984) *Virology* **133**, 92-110.
10. Garoff, H., Frischauf, A.-M., Simons, K., Lehrach, H., & Delius, H. (1980) *Nature* **288**, 236-241.
11. Garoff, H., Frischauf, A.-M., Simons, K., Lehrach, H. & Delius, H. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 6376-6380.
12. Takkinen, K. (1986) *Nuc. Acids Res.* **14**, 5667-5682.
13. Dalgarno, L., Rice, C. M., & Strauss, J. H. (1983) *Virology* **129**, 170-187.
14. Chang, G.-J. J. & Trent, D. W. (1987) *J. Gen. Virol.* **68**, 2129-2142.

15. Kinney, R. M., Johnson, R. J. B., Brown, V. C., & Trent, D. W. (1986) *Virology* **152**, 400-413.

16. Hardy, J. L., Reeves, W. C., Rush, W. A., & Nir, Y. D. (1974) *Infection and Immunity* **10**, 553-564.

17. Lindqvist, B. H., DiSalvo, J., Rice, C. M., Strauss, J. H., & Strauss, E. G. (1986) *Virology* **151**, 10-20.

18. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499-560.

19. Smith, D. R. & Calvo, J. M. (1980) *Nuc. Acids Res.* **8**, 2225-2274.

20. Ou, J.-H., Rice, C. M., Dalgarno, L., Strauss, E. G., & Strauss, J. H. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 5235-5239.

21. Bell, J. R., Bond, M. W., Hunkapiller, M. W., Strauss, E. G., Strauss, J. H., Yamamoto, K. & Simizu, B. (1983) *J. Virology* **45**, 708-714.

22. Steinhauer, D. A. & Holland, J. J. (1987) *Ann. Rev. Microbiol.* **41**, 409-433.

23. Smith, D. B. & Inglis, S. C. (1987) *J. Gen. Virol.* **68**, 2729-2740.

24. Kirkegaard, K., & Baltimore, D. (1986) *Cell* **47**, 433-443.

25. Strauss, E. G. & Strauss, J. H. (1985) in *Virus Structure and Assembly* (ed. Casjens, S.) (Jones & Bartlett, Boston), pp. 205-234.

26. Fuller, S. D. (1987) *Cell* **48**, 923-934.

27. Rice, C. M., Levis, R., Strauss, J. H., & Huang, H. V. (1987) *J. Virol.* **61**, 3809-3819.

28. Rice, C. M. & Strauss, J. H. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 2062-2066.

29. Strauss, J. H. & Strauss, E. G. (1988) *Ann. Rev. Microbiol.* **42**, in press.

30. Hodgman, T. C. & Zimmern, D. (1988) in *RNA Genetics* (ed. Holland, J. J., Domingo, E. & Ahlquist, P.) CRC Press, in press.

31. Desselberger, U., Nakajima, K., Alfino, P., Pederson, F. S., Haseltine, W. A., Hannoun, C., & Palese, P. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 3341-3345.

32. Webster, R. G., Laver, W. G., Air, G. M., & Schild, G. C. (1982) *Nature* **296**, 115-121.

33. Cooper, P. D. (1977) in *Comprehensive Virology* (ed. Fraenkel-Conrat, H. and Wagner, R. R) (Plenum Press, New York), Vol. 9, pp. 133-207.

34. Kew, O. M. & Nottay, B. K. (1984) in *Modern Approaches to Vaccines: Molecular and Chemical Basis of Virus Virulence* (ed. Channock, R. M.), (Cold Spring Harbor Laboratory, New York), pp. 357-362.

35. Lai, M. M. C., Baric, R. S., Makino, S., Keck, J. G., Egbert, J., Leibowitz, J. L., & Stohlman, S. A. (1985) *J. Virol.* **56**, 449-456.

36. Makino, S., Keck, J. G., Stohlman, S. A., & Lai, M. M. C. (1986) *J. Virol.* **57**, 729-737.

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

# Chapter 7

## Conserved Elements in the 3' Untranslated Region of Flavivirus RNAs and Potential Cyclization Sequences

This chapter was published in the *Journal of Molecular Biology* and is reproduced with the permission of the publisher.

*J. Mol. Biol.* (1987) **198**, 33–41

# Conserved Elements in the 3′ Untranslated Region of Flavivirus RNAs and Potential Cyclization Sequences

**Chang S. Hahn[1], Young S. Hahn[1], Charles M. Rice[1]†, Eva Lee[2] Lynn Dalgarno[2], Ellen G. Strauss[1] and James H. Strauss[1]**

[1] *Division of Biology*
*California Institute of Technology*
*Pasadena, CA 91125, U.S.A.*

[2] *Biochemistry Department*
*The Faculties, Australian National University*
*P.O. Box 4, Canberra, Austrilia 2601*

*(Received 6 May 1987, and in revised form 17 July 1987)*

We have isolated a cDNA clone after reverse transcription of the genomic RNA of Asibi yellow fever virus whose structure suggests it was formed by self-priming from a 3′-terminal hairpin of 87 nucleotides in the genomic RNA. We have also isolated a clone from cDNA made to Murray Valley encephalitis virus RNA that also appears to have arisen by self-priming from a 3′-terminal structure very similar or identical to that of yellow fever. In addition, 3′-terminal sequencing of the S1 strain of dengue 2 RNA shows that this RNA is also capable of forming a 3′-terminal hairpin of 79 nucleotides. Furthermore, we have identified two 20-nucleotide sequence elements which are present in the 3′ untranslated region of all three viruses; one of these sequence elements is repeated in Murray Valley encephalitis and dengue 2 RNA but not in yellow fever RNA. In all three viruses, which represent the three major serological subgroups of the mosquito-borne flaviviruses, the 3′-proximal conserved sequence element, which is found immediately adjacent to the potential 3′-terminal hairpin, is complementary to another conserved domain near the 5′ end of the viral RNAs, suggesting that flavivirus RNAs can cyclize (calculated $\Delta G < -11$ kcal; 1 kcal = 4·184 kJ).

## 1. Introduction

Short nucleotide sequences located near or at the 3′ termini of numerous viral RNAs have been postulated to play important roles in virus RNA replication (for a review, see Strauss & Strauss, 1983). These elements may be either elaborate secondary structures or linear nucleotide sequences, and are believed to form specific binding sites recognized by the viral encoded replicase, analogous to promoters in DNA sequences. The best-studied examples of structures with specific roles in replication are found in plant virus RNAs (e.g. see Hall, 1979; Pleij *et al.*, 1985; Smith & Jaspers, 1980). A number of linear sequences located at the 3′ termini or animal viruses, which appear to be essential for RNA replication, have also been well-studied. In the case of alphaviruses there is a conserved 19-nucleotide sequence at the 3′ terminus (Ou *et al.*, 1982) which has been shown by truncation experiments using defective interfering RNAs to be required for replication and/or packaging of the RNAs (Levis *et al.*, 1986). Conserved sequences 12 to 15 nucleotides in length are found at the 3′ ends of the genome segments of influenza virus and of Bunyaviruses. Bunyaviruses with identical 3′-terminal sequence elements can exchange genome segments, giving rise to recombinant viruses and, in fact, these conserved terminal sequences have been used to define genera within the Bunyaviridae (for a review, see Strauss & Strauss, 1983).

In the case of flaviviruses, Rice *et al.* (1985) proposed that the 3′-terminal 87 nucleotides of yellow fever (YF)‡ RNA form a stable secondary structure. Subsequently, West Nile (WN) virus (Brinton *et al.*, 1986; Wengler & Castle, 1986).

† Present address: Department of Microbiology and Immunology, Box 8093, Washington University School of Medicine, 660 So. Euclid Avenue, St. Louis, MO 63110, U.S.A.

‡ Abbreviations used: YF, yellow fever; WN, West Nile; JE, Japanese encephalitis; DEN 4, dengue 4; DEN 2, dengue 2; MVE, Murray Valley encephalitis; kb, $10^3$ bases or base-pairs.



0022-2836/210033-09 $03.00/0

Japanese encephalitis (JE) virus (Takegami *et al.*, 1986), and dengue 4 (DEN 4) virus (Zhao *et al.*, 1986) were proposed to have secondary structures very similar to that of YF RNA. The conservation of this structure supports the hypothesis that it is important for viral RNA replication, but so far direct evidence for its existence has been limited. Brinton *et al.* (1986) found that nucleotides within the putative hydrogen-bonded regions of the stems were partially resistant to ribonuclease, suggesting that the structure was present in WN RNA in solution. In the case of YF, no direct evidence of this conformation has been presented, and in fact Grange *et al.* (1985) proposed an alternative structure involving the 3′-terminal 120 nucleotides. We now report the isolation of a clone for YF cDNA which could easily have arisen by self-priming of cDNA synthesis from the 87-nucleotide 3′-terminal structure previously proposed, but which is otherwise difficult to explain. We also report the isolation of a clone from Murray Valley encephalitis (MVE) RNA which apparently begins at the same corresponding nucleotide as this YF clone, and which thus may also have arisen by self priming. In addition, we have obtained the sequence of the 3′ untranslated region of MVE RNA, excluding the putative 3′-terminal structure, as well as the entire 3′ untranslated sequence of dengue 2 (DEN 2) RNA, and found conserved and repeated sequences within this RNA domain: one conserved sequence could be involved in cyclization of flavivirus RNA. Models of these structures as well as comparisons of these various sequence elements with the corresponding sequences in WN virus (Brinton *et al.*, 1986; Wengler & Castle, 1986), DEN 4 virus (Zhao *et al.*, 1986), and JE virus (Takegami *et al.*, 1986) are presented.

## 2. Materials and Methods

### (a) *Virus strains and cloning of flavivirus cDNA*

The preparation of Asibi YF RNA (Hahn *et al.*, 1987) and MVE RNA (Dalgarno *et al.*, 1986) have been described and the preparation of DEN 2 RNA will be described elsewhere. The methods used for obtaining a cDNA library from these genomes have also been described (Rice *et al.*, 1987). Briefly, first strand cDNA was synthesized using AMV reverse transcriptase and degraded calf thymus DNA as random primers in the presence of human placental RNase inhibitor and actinomycin D. Second strand synthesis was done according to the conditions described by Okayama & Berg (1982). In this method *Escherichia coli* RNase H (Bethesda Research Laboratories) is used to introduce nicks into the RNA strand of the RNA–cDNA hybrid duplex; primer extension occurs at these nicks and ultimately these RNA-primed fragments are repaired and ligated to form the second DNA strand. The double-stranded cDNA resulting was methylated at the *Eco*RI sites. After treatment with phage T4 DNA polymerase, *Eco*RI linkers were attached and the cDNA was fractionated in agarose gels. Selected size fractions were then inserted into the *Eco*RI site of vector pGEM1 (Promega Biotech) (Asibi YF and DEN 2) or pMT21 (MVE). Plasmids for ampicillin-resistant colonies were screened for insert size and restriction digestion pattern. The clone containing the extreme 3′ end of Asibi YF RNA was identified by colony hybridization (Grunstein & Hogness, 1975) using a fragment 150 nucleotides in length derived from the extreme 3′ end of a clone of 17D YF RNA (Rice *et al.*, 1985, 1987) as a probe, with the library obtained from the 0·8 to 2 kb double-stranded cDNA size class. The probe fragment extends from the *Xba*I site at nucleotide 10,708 to the 3′ end of 17D YF. Of approximately 4000 clones screened with this probe, 3 colonies were found to be positive. Two of these clones started at the 113th nucleotide from the 3′ terminus. The 3rd clone had sequences from both plus and minus strand RNA joined together and its structure is reported in detail in the main text. The MVE clone containing sequences in the 3′ untranslated region was identified during characterization of the random MVE library (Rice *et al.*, 1987). To obtain clones containing the 3′-terminal sequence of DEN 2, the RNA was polyadenylated with poly(A) polymerase (Rice *et al.*, 1985) and oligo(dT) was used to prime first strand cDNA synthesis. After second strand synthesis and insertion into plasmid pMT21, ampicillin-resistant colonies were screened by colony hybridization, using as a probe a restriction fragment derived from a clone of dengue 2 cDNA containing sequences in the NS5 region of the genome (Rice *et al.*, 1987; Y. S. Hahn *et al.*, unpublished results). Twelve positive colonies were obtained which all had the same restriction pattern; 3 of these clones were sequenced in the 3′ untranslated region and all had the identical sequence terminating in a poly(A) tract. For this reason, and because of homologies in structure and sequence with other flavivirus RNAs, we are confident that the 3′-terminal sequence has been obtained.

### (b) *Sequence analysis of cDNA*

Plasmid DNAs from cDNA clones were sequenced using the methods of Maxam & Gilbert (1980) as modified by Smith & Calvo (1980), using restriction fragments 3′ end-labeled with the Klenow fragment of *E. coli* DNA polymerase (Bethesda Research Laboratories).

## 3. Results

### (a) *Asibi yellow fever clone 3′-1*

A library of cDNA made to Asibi YF RNA was constructed using random priming. A probe which represented the 3′-terminal 150 nucleotides of 17D YF RNA, derived from a restriction fragment obtained from the 17D YF cDNA library previously constructed and characterized (Rice *et al.*, 1985, 1987), was used to screen the Asibi library. A clone designated 3′-1 was obtained which had an unusual structure, illustrated schematically in Figure 1. This clone has the 3′-terminal nucleotides from 10,773 to 10,862 from one strand connected to the sequence from the complementary strand beginning at nucleotide 10,775 and proceeding upstream, that is towards the 5′ end of the genome (nucleotide numbers refer to the virion plus-strand RNA). The simplest interpretation of how this clone might have arisen is self-priming of first strand cDNA synthesis by the 3′ terminus of YF genomic RNA. A 3′-terminal secondary structure for Asibi YF RNA, virtually identical to

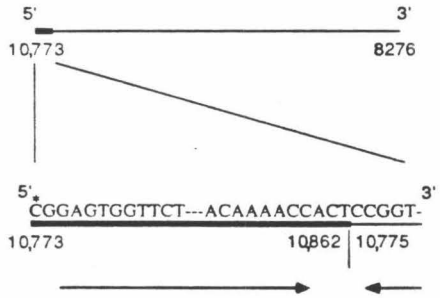



**Figure 1.** Structure of clone 3′-1. A schematic diagram of the structure of this clone in which the 3′-terminal sequences as plus strand (thick line) are attached to minus strand sequences (thin line) is shown. Nucleotide sequences around the joint between plus and minus strand sequence and at the end of the clone are indicated. Arrows indicate polarity of 5′ to 3′ expressed as the plus sense genomic RNA. The asterisk denotes the 5′-terminal nucleotide in the insert as indicated, and marks the same nucleotide as in Figs 2 and 3.

that proposed by Rice *et al.* (1985) for YF 17D RNA. is illustrated in Figure 2(a) and shows the 3′-terminal uridylic acid of YF RNA (nucleotide 10,862) hydrogen-bonded into the structure. Self-priming in which the 3′-terminal hydroxyl is extended by reverse transcriptase would lead to the 3′-terminal RNA nucleotide (10,862) being attached to nucleotide 10,775 of minus-strand cDNA. Second strand cDNA synthesis and repair illustrated in Figure 3 could then result in the observed clone. For second strand cDNA synthesis *E. coli* RNase H was used to cleave the RNA strand in RNA–DNA duplexes forming new primers that can be extended to produce the second strand. This makes it possible, given the proper conditions of *E. coli* RNase H concentration and activity of *E. coli* DNA polymerase I, to copy back part of the RNA strand during second strand cDNA synthesis (step 3 of Fig. 3), an activity that polymerase I is known to be capable of (for a review, see Kornberg, 1980). (Although we here assume that the RNase H and *E. coli* polymerase I added for second strand synthesis are responsible for the copy back, it is difficult to rule out the possibility that the RNase H activity in AMV reverse transcriptase might lead to copy back by reverse transcriptase itself during first strand synthesis). Repair of RNA–DNA duplex in this region could lead to double-stranded cDNA. Presumptively, some RNA sequence at the 5′ end of the RNA–DNA duplex could be lost because of the inability of this RNA to be replaced with DNA. However, the resulting clone would have the structure shown in Figure 3 and the key point is the

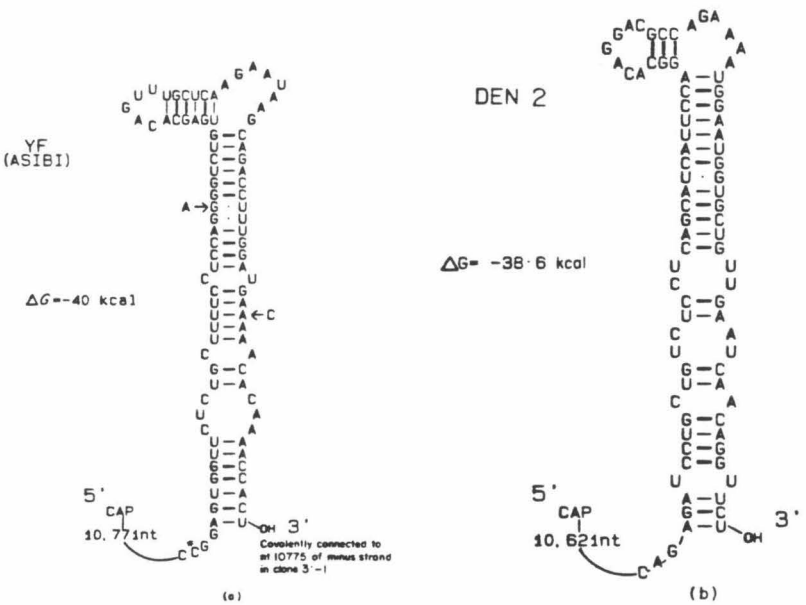



**Figure 2.** Putative secondary structure found at the 3′ end of Asibi YF RNA and of DEN 2 RNA. (a) A possible secondary structure at the 3′ terminus of Asibi YF RNA is shown. This structure is virtually identical to that proposed by Rice *et al.* (1985) for the 3′ terminus of 17D YF RNA with 2 nucleotide changes as shown. Note that the use of the 3′-terminal hydroxyl group for self-priming for reverse transcription would result in the 3′-terminal uridylic acid of the RNA genome being covalently attached to the complement of nucleotide 10,775 (in the minus strand cDNA). The asterisk indicates the 5′ end of the clone and has the same meaning as in Figs 1 and 3. Free energies (at 25 °C) in this and other Figures are calculated by the method of Tinoco *et al.* (1973). (b) A possible secondary structure at the 3′ terminus of DEN 2 RNA is shown. nt, nucleotide.

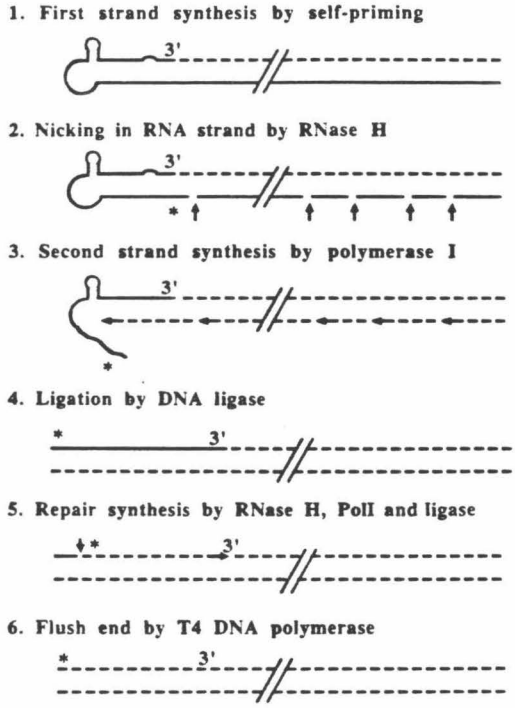



**Figure 3.** Model for the origin of clone 3′-1. The first step describes first strand cDNA (broken line) synthesis in which self-priming occurs onto the 3′ terminus of the genomic RNA (continuous line). The next 3 steps describe second strand cDNA synthesis following nicking of the RNA strand with *E. coli* RNase H (step 2), priming of second strand cDNA synthesis by the RNA fragments so produced such that foldback and copying of template RNA occurs during second strand synthesis (step 3), and finally (step 4) completion of the second strand. The RNase H activity would be expected to lead to replacement of the RNA strand in step 4 with the DNA strand as in step 5, which would cover all but an undetermined number of nucleotides at the left-hand end of this double-stranded cDNA. The asterisk indicates the 5′ terminus of the final double-stranded cDNA as found in the insert and has the same meaning as in Figs 1 and 2.

attachment of nucleotide 10,862 of one sense RNA to nucleotide 10,775 of the opposite sense RNA. It is difficult to explain the existence of this clone by mechanisms other than self-priming by the RNA template and this clone provides strong support for the existence of a 3′-terminal structure in which the terminal U can be paired with nucleotide 10,776 as proposed by Rice *et al.* (1985).

Clones containing cDNA inserts with both RNA senses in one strand have been observed and have been ascribed to self-priming and copy back of the first strand cDNA during second strand synthesis (Fields & Winter, 1981; Volckaert *et al.*, 1981). The formal mechanism is thus analogous to that described in Figure 3, but in these previous findings the template RNA was removed prior to second strand synthesis, and first strand cDNA was required to self prime for second strand synthesis. These conditions are quite different from the Okayama & Berg (1982) conditions for second strand cDNA synthesis used here.

Although clone 3′-1 provides strong support for the occurrence of self priming during cDNA synthesis, such self-priming appears to be a rare event. We have attempted, without success, to demonstrate self priming directly by incorporation of radiolabeled nucleotides into first strand cDNA in the absence of added primer. However, the complexity of the library screened was approximately 3000 (that is, following transformation we obtained approximately 3000 independent colonies containing inserts), and only one clone of this type was found. If self priming occurred at only 1/3000 the frequency of priming by added primers it would have been difficult to detect incorporation under our conditions. The analysis is complicated by the fact that the double-stranded cDNA used for the construction of the library screened was size selected (see Materials and Methods), and thus the proportion of self-primed cDNAs might not be representative. In addition, depending upon the relative concentrations of reverse transcriptase, primer and template, clones in a given size class arising by self priming might be depressed in the presence of added primer.

In the screening process, we also found two clones with identical inserts which were unremarkable in structure, beginning at the 113th nucleotide from the 3′ terminus and proceeding upstream to nucleotide 8275 of virion RNA. The two inserts presumably arose from the same transformation event, since the colonies screened with the 3′-specific probe were obtained by plating a portion of the original library and were, therefore, not necessarily independent isolates.

### (b) *MVE clone 2/2/38*

During characterization of a library of clones from double-stranded cDNA made to MVE RNA a clone was obtained which began with a nucleotide corresponding to nucleotide 10,775 of YF, termed clone 2/2/38 (Rice *et al.*, 1987). If a 3′-terminal structure exists in MVE that is equivalent to that found in YF, then this clone could have arisen by self priming in the same way as Asibi YF clone 3′-1. In this case, however, second strand synthesis and repair would not have led to copy back of the RNA primer. The existence of this clone is quite suggestive that in fact it did arise by self priming and that the 3′ terminus of MVE RNA has a similar structure to that found for YF.

The existence of this clone as well as other clones originating in the 3′ untranslated region upstream from this has allowed us to obtain the sequence of the 3′ untranslated region of MVE RNA (excluding the sequence in the putative 3′-terminal secondary structure) and this sequence is shown in Figure 4.

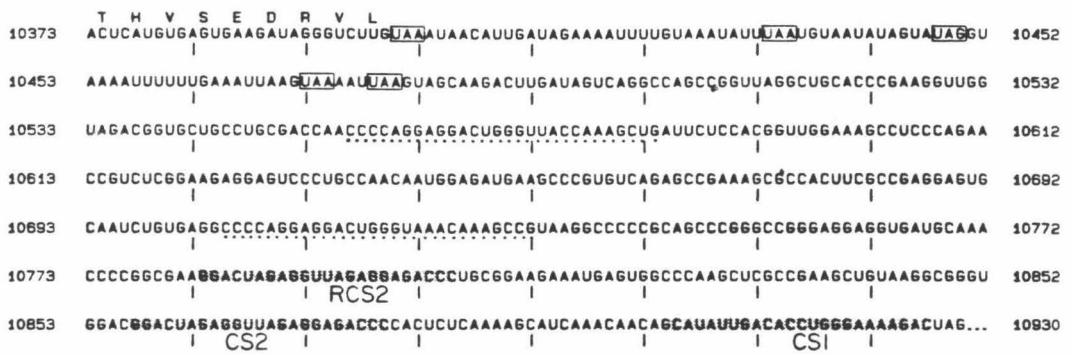



**Figure 4.** Nucleotide sequence of the 3′ untranslated region of MVE RNA. The last few nucleotides of the long open reading frame are shown as translated sequence. The entire untranslated region is shown except for the putative 3′-terminal secondary structure which has yet to be determined. The first few in-phase termination codons are boxed, a repeated sequence of 28 nucleotides is indicated by dotted underlining, and sequences conserved between MVE and YF (and other flaviviruses as well) are shaded (see also Figs 5 and 6).

Firstly, there are two 20-nucleotide sequences which are highly conserved between MVE and YF. A similar observation has been reported by Wengler & Castle (1986) for WN virus compared to YF. These conserved nucleotide sequences are shown in Figure 5. It is of considerable interest that the first of these, which we refer to as CS1, is found immediately upstream from the 3′-terminal structure hypothesized to exist in YF RNA. Secondly, one of these conserved sequences (CS2) is repeated identically in MVE approximately 75 nucleotides upstream. This conserved sequence is present and also repeated in the 3′ untranslated region of WN RNA, but is not repeated in YF RNA. On the other hand, we have reported repeated sequences further upstream in the 3′ untranslated region of YF (Rice *et al.*, 1985) which are not shared with either MVE or WN RNA, and there are short repeated sequences found in MVE and WN that are not found in YF. A schematic diagram of these conserved and/or repeated sequences is shown in Figure 6.

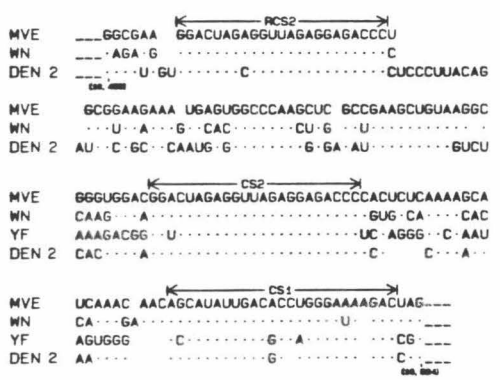



**Figure 5.** Comparison of conserved and repeated sequences in the 3′ untranslated region of flavivirus RNAs. Conserved sequences found in DEN 2 RNA, in MVE RNA (Fig. 4), in YF RNA (Rice *et al.*, 1985), and in WN RNA (Wengler & Castle, 1986) are compared. There are 2 such conserved nucleotide sequences (CS1 and CS2). Conserved sequence 2 is repeated in DEN 2, WN and MVE RNA (RCS2) but not in YF RNA.

### (c) *The 3′ untranslated region of DEN 2 RNA*

The mosquito-borne flaviviruses can be grouped into three major serological subgroups. The subgroups differ in their vertebrate host range and in tissue tropism within the vertebrate host. YF virus, representing one subgroup, is viscerotropic and neurotropic and its vertebrate host range is limited to primates. The MVE/JE/WN subgroup is also neurotropic, but has a wider vertebrate host range with birds as a major reservoir in nature. The four dengue viruses constitute the third subgroup; these viruses replicate primarily within cells of lymphoid origin in man, their only natural host. To compare the features of the 3′ untranslated region of dengue RNA with those of the RNAs of the two other subgroups, clones constaining the 3′ untranslated sequence of DEN 2 RNA (S1 candidate vaccine strain derived from the PR159 strain) were obtained from polyadenylated RNA, using oligo(dT) to prime first strand synthesis, and sequenced. The 3′-terminal 79 nucleotides can be folded into a structure that is very similar to that postulated for YF RNA (Fig. 2(b)). This structure has a calculated thermal stability of −38 kcal/mol (1 kcal = 4·184 kJ) and would be expected to exist in solution. The fact that all flaviviruses examined to date, including representatives from three subgroups, can potentially form a very similar structure argues that this structure plays a role in virus replication.

The 3′-terminal dinucleotide in DEN 2 is CU as has been found in all flaviviruses sequenced to date, and the 3′ terminus may be hydrogen-bonded into a secondary structure similar to that postulated for other flaviviruses (Fig. 2(b)). However, the

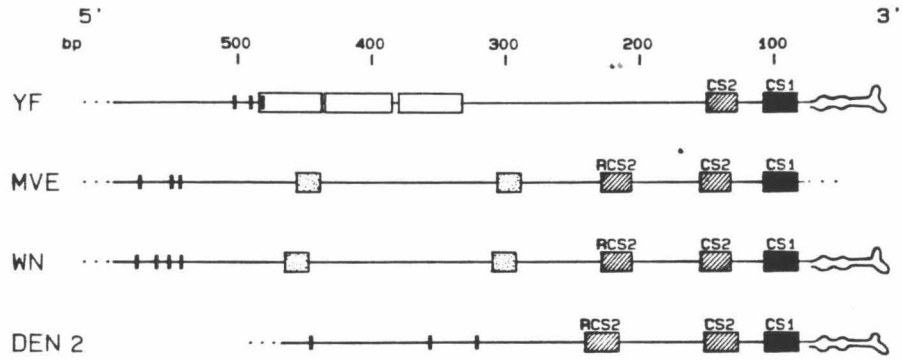



**Figure 6.** Schematic diagram of the organization of the 3′ untranslated region of flavivirus RNAs. The location of repeated nucleotide sequences, conserved nucleotide sequences and the 3′-terminal secondary structure are indicated for YF RNA, MVE RNA, WN RNA and DEN 2 RNA. In addition to CS1 (filled boxes) and to CS2 and RCS2 (hatched boxes) shown in Fig. 5, these include a sequence element repeated 3 times in YF RNA (open boxes) not found in WN or MVE, and a shorter sequence element (stippled boxes) present as 2 copies in both MVE and WN but not found in YF or DEN 2. The first few in-phase stop codons that terminate the long open reading frame are shown as vertical bars. bp, base-pairs.

sequence ACACA found just upstream from the 3′ terminus of YF, WN and JE, and which is present near the 3′ terminus of the minus strand as well as of the plus strand, is not found in DEN 2 (Fig. 2(b)), nor in DEN 4 (Zhao *et al.*, 1986). There is a related sequence CAACA, but the complement of this sequence is not found near the 5′ end of DEN 2 RNA (Deubel *et al.*, 1986).

The sequence elements CS1 and CS2 are also present in DEN 2 RNA (Figs 5 and 6). As was the case for MVE, CS2 is present in two copies, with a single nucleotide substitution in the second copy (Fig. 5). The conservation of these sequence elements is striking and suggests that they are important in replication.

The 3′ untranslated region of DEN 2 is somewhat shorter than that of other flaviviruses examined to date, 443 nucleotides, and the cluster of in-phase termination codons that end the long open reading frames of YF, MVE and WN are not found in DEN 2 (Fig. 6).

### (d) *Possible involvement of conserved sequences in flavivirus RNA cyclization*

Conserved nucleotide sequences located near the 5′ end of the flavivirus genome (within the coding region of the capsid protein) are shown in Figure 7 for six different flaviviruses. This sequence, which begins at nucleotide 147 of YF, demonstrates a high degree of conservation. In particular, eight contiguous nucleotides (indicated by the shaded overlay) are perfectly conserved among these six different flaviviruses representing three subgroups. In the lower panel the conserved sequence within the 3′ untranslated region labeled CS1 in Figures 5 and 6 is shown for five flaviviruses, again representing three serological subgroups. There are once again eight contiguous nucleotides (indicated

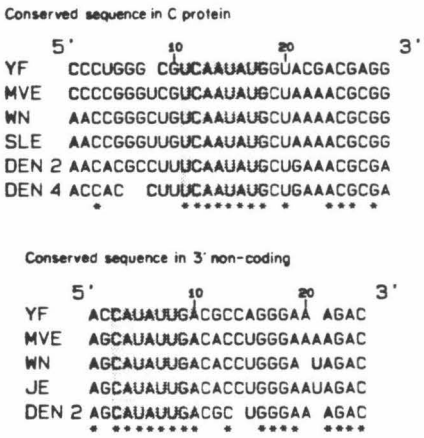



**Figure 7.** Conserved nucleotide sequence elements in flavivirus RNAs. In the top half of the Figure are shown nucleotide sequences from 6 flaviviruses found in the coding region for the capsid protein. This sequence begins at nucleotide 147 of YF RNA (Rice *et al.*, 1985), at 128 of MVE RNA (Dalgarno *et al.*, 1986), at 128 of DEN 4 RNA (Zhao *et al.*, 1986), at 129 of St Louis encephalitis virus (SLE) RNA (Trent *et al.*, 1987), at an unknown nucleotide of WN RNA (since the 5′-terminal sequence of WN is incomplete) (Castle *et al.*, 1985), and at 124 of DEN 2 RNA (Y. S. Hahn, unpublished results). Asterisks denote nucleotides that are invariant among the 6 viruses; note that there is also a high degree of similarity among nucleotides in this region even when they are not invariant. The lower panel compares CS1 for 5 flaviviruses (see also Figs 5 and 6). Note that CS1 is complementary to the nucleotide sequence in the upper panel beginning with uridine 20, and that the 8 contiguous nucleotides perfectly conserved in the 5′ region are complementary to 8 contiguous nucleotides perfectly conserved beginning with nucleotide 3 of CS1 (indicated by the shaded overlays). The JE sequence is from Takegami *et al.* (1986).

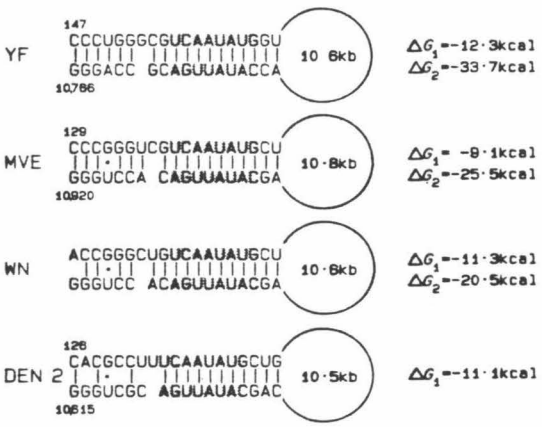



**Figure 8.** Possible cyclization figures for flavivirus RNAs. The complementary sequences shown in Fig. 7 are aligned for 4 flaviviruses as circular figures. The calculated thermal stability of these circles is indicated (Tinoco *et al.*, 1973). The 8-nucleotide core that is perfectly conserved is indicated by the shaded overlay.

by the shaded overlay) perfectly conserved among the five flaviviruses shown and these are complementary to the eight-nucleotide conserved domain in the 5′ region. Thus, it is possible that these sequences could be used to cyclize the RNA to form a panhandle structure. Possible panhandle structures are shown in Figure 8 for four different flaviviruses. In each case there are 11 to 12 contiguous nucleotides that are perfectly base-paired, and these include the eight-nucleotide conserved core (indicated by the overlay). Four to six additional hydrogen bonds can be formed from adjacent sequences which could contribute to the stability of cyclization for all but DEN 2. Two $\Delta G$ values are shown. $\Delta G_1$ is based only upon the 11 to 12 contiguous base-pairs and provides sufficient free energy (−9 to −11 kcal at 25°C) to cyclize the RNA. Alphavirus RNAs are known to cyclize under physiological conditions (Hsu *et al.*, 1973) and the free energy of cyclization measured thermodynamically is −13·5 kcal at 25°C (Frey *et al.*, 1979). The $\Delta G_2$ value includes the possible contribution of the additional four to six base-pairs upstream, which lead to a much more stable structure ($\Delta G = -20$ to $-33$ kcal), but which are not present in DEN 2 RNA.

The RNA sequence upstream from the 5′ conserved sequence may also form a single or double hairpin structure as illustrated in Figure 9 for four viruses. The 5′ hairpin illustrated cannot be formed in DEN 2 RNA and is variable in structure in the other viruses; lack of conservation suggests it may not play an important role in replication but it is interesting that it includes the initiating AUG

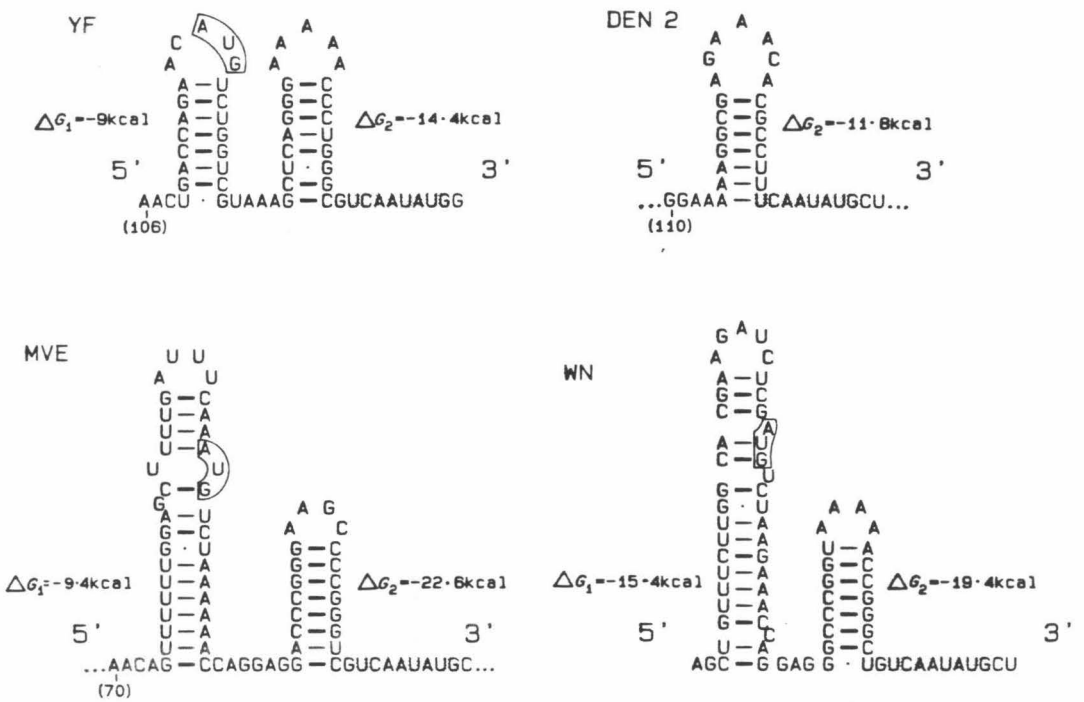



**Figure 9.** Possible hairpin structures in the 5′ regions of flavivirus RNAs. The AUG codon that initiates the very long open reading frame is boxed and the core conserved sequence that might be involved in cyclization is shaded. Free energies were calculated as for Fig. 2.

and could thus have an influence upon translation. The 3′ hairpin, on the other hand, has a high calculated thermal stability ($\Delta G = -12$ to $-23$ kcal), is conserved in all four viruses and is found immediately upstream from the core conserved sequence of Figures 7 and 8 (indicated by the shaded overlay). Thus, it could compete with the upstream portion of the panhandle structure of Figure 8 and destabilize cyclization but, as discussed above, the 11 to 12 contiguous nucleotide pairs alone could lead to cyclization ($\Delta G_1$ of Fig. 8).

## 4. Discussion

The 3′-terminal sequences or structures on the end of viral RNAs are thought to play an important role in viral RNA replication. Sequence analysis of several flavivirus RNAs including YF RNA (Rice *et al.*, 1985), DEN 2 RNA (this paper), WN RNA (Brinton *et al.*, 1986; Wengler & Castle, 1986), JE RNA (Takegami *et al.*, 1986), and DEN 4 RNA (Zhao *et al.*, 1986) have shown that these all can form a 3′-terminal hairpin, suggesting that this hairpin is essential for flavivirus replication. However, to date the only direct evidence for this structure in solution was obtained by Brinton *et al.* (1986) by RNase digestion of WN RNA. The structure of clone 3′-1 from Asibi YF RNA presented here also provides strong support for the existence of the 3′-terminal structure in solution. This YF structure, which involves the 3′-terminal 87 nucleotides, has a calculated thermal stability of $-42$ to $-46$ kcal/mol and was first proposed by Rice *et al.* (1985) on the basis of the sequence data alone. It is of interest that an alternative structure for YF RNA was proposed by Grange *et al.* (1985), based upon primary sequence data, in which the 3′-terminal 120 nucleotides were involved; that is, the predicted structure was one in which the 3′-terminal U was paired with A-10743 rather than with A-10776. The Grange *et al.* (1985) structure has a greater calculated stability, $-52$ kcal/mol, and was identified by computer analysis as being the most favored structure. However, the data presented here make it seem more likely that the Rice *et al.* (1985) structure is the one that acutally forms in solution.

We have also identified nucleotide sequence elements approximately 20 nucleotides in length present in the 3′ untranslated region that are highly conserved among flaviviruses, suggesting that these are also important in RNA replication or packaging (see also Wengler & Castle, 1986). One of these conserved sequence elements, found immediately upstream from the 3′-terminal secondary structure, could be used to cyclize flavivirus RNA. The calculated free energy of cyclization is significant, suggesting that flavivirus RNAs do in fact cyclize. No other evidence for cyclization of flavivirus RNAs exists, but it should be noted that alphavirus RNAs have long been known to cyclize (Hsu *et al.*, 1973) as do the RNAs of the Bunyaviruses (Hewlett *et al.*, 1977). The function of cyclization in virus replication is unknown, but could be used to help ensure that virus RNA molecules that are replicated are full-length RNA, if a viral RNA replicase were required to bind to both 5′ and 3′ regions simultaneously in order to initiate RNA replication.

It should be noted that the eight-nucleotide core sequence near the 3′ end of the plus-strand genomic RNA will also be present near the 3′ end of the minus-strand template (as the complement of the 5′ core sequence) and, conversely, the complementary sequence will be present near the 5′ ends of both plus and minus-strand RNA. Thus, an alternative role to cyclization of such self-complementary sequence elements is to serve as signals in replication, perhaps as replicase recognition sites, in which case the same signal may be utilized for transcription of both plus and minus strands. Electron microscopy or physical studies of the RNA in solution will be required to establish whether the RNA physically cyclizes.

This work was supported by grant DMB-86-17372 from the NSF, by grants AI-20612 and AI-10793 from the NIH, by a grant from the World Health Organization, and by a grant from the National Health and Medical Research Council of Australia. C.M.R. is supported by a grant from the Pew Memorial Trust.

## References

Brinton, M., Fernandez, A. V. & Dispoto, J. H. (1986). *Virology*, **153**, 113–121.

Castle, E., Nowak, T., Leidner, U., Wengler, G. & Wengler, G. (1985). *Virology*, **145**, 227–236.

Dalgarno, L., Trent, D. W., Strauss, J. H. & Rice, C. M. (1986). *J. Mol. Biol.* **187**, 309–323.

Deubel, V., Kinney, R. M. & Trent, D. W. (1986). *Virology*, **155**, 365–377.

Fields, S. & Winter, G. (1981). *Gene*, **15**, 207–214.

Frey, T. K., Gard, D. L. & Strauss, J. H. (1979). *J. Mol. Biol.* **132**, 1–18.

Grange, T., Bouloy, M. & Girard, M. (1985). *FEBS Letters*, **188**, 159–163.

Grunstein, M. & Hogness, D. (1975). *Proc. Nat. Acad. Sci., U.S.A.* **72**, 3961–3965.

Hahn, C. S., Dalrymple, J. M., Strauss, J. H. & Rice, C. M. (1987). *Proc. Nat. Acad. Sci., U.S.A.* **84**, 2019–2023.

Hall, T. C. (1979). In *International Review of Cytology* (Bourne, G. H. & Danielli, J. F., eds), vol. 60, pp. 1–26. Academic Press, New York.

Hewlett, M. J., Pettersson, R. F. & Baltimore, D. (1977). *J. Virol.* **21**, 1085–1093.

Hsu, M. T., Kung, H. J. & Davidson, N. (1973). *Cold Spring Harbor Symp. Quant. Biol.* **38**, 943–950.

Kornberg, A. (1980). *DNA Replication*, 2nd edit., W. H. Freeman and Co., San Francisco.

Levis, R., Weiss, B. G., Tsiang, M., Huang, H. & Schlesinger, S. (1986). *Cell*, **44**, 137–145.

Maxam, A. M. & Gilbert, W. (1980). *Methods Enzymol* **65**, 449–560.

Okayama, H. & Berg, P. (1982). *Mol. Cell. Biol.* **2**, 161–170.

Ou, J. H., Trent, D. W. & Strauss, J. H. (1982). *J. Mol. Biol.* **156**, 719–730.

Pleij, C. W. A., Rietveld, K. & Bosch, L. (1985). *Nucl. Acids Res.* **13**, 1717–1731.

Rice, C. M., Lenches, E. M., Eddy, S. R., Shin, S. J., Sheets, R. L., & Strauss, J. H. (1985). *Science*, **229**, 726–733.

Rice, C. M., Dalgarno, L., Galler, R., Hahn, Y. S., Strauss, E. G. & Strauss, J. H. (1987). In *Modern Trends in Virology. Proceedings of the International Symposium in Giessen, June 1986*, Springer-Verlag, Heidelberg, in the press.

Smit, C. H. & Jaspers, E. M. J. (1980). *Virology*, **104**, 454–461.

Smith, D. R. & Calvo, J. M. (1980). *Nucl. Acids Res.* **8**, 2255–2274.

Strauss, E. G. & Strauss, J. H. (1983). *Current Topics Microbiol. Immunol.* **105**, 1–98.

Takegami, T., Washizu, M. & Yasui, K. (1986). *Virology*, **152**, 483–486.

Tinoco, I., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M. & Gralla, J. (1973). *Nature New Biol.* **246**, 40–41.

Trent, D. W., Kinney, R. M., Johnson, B. J. B., Vorndam, A. V., Grant, J. A., Deubel, V., Rice, C. M. & Hahn, Ç. (1987). *Virology*, **156**, 293–304.

Volckaert, G., Tavernier, J., Derynck, R., Devos, R. & Fiers, W. (1981). *Gene*, **15**, 215–223.

Wengler, G. & Castle, E. (1986). *J. Gen. Virol.* **67**, 1183–1188.

Zhao, B., Mackow, E., Buckler-White, A., Markoff, L., Chanock, R. M., Lai, C.-J. & Makino, Y. (1986). *Virology*, **155**, 77–88.

*Edited by S. Brenner*