Functional Analysis of Viral Nonstructural and Structural Proteins

Thesis by

Young Shin Lim Hahn

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1989

(Submitted February 8, 1989)

Acknowledgement

First of all, I would like to thank my advisor, Dr. James Strauss for his support and for providing such an excellent lab in which to learn a nice work. I also would like to thank my thesis committee members, Drs. Giuseppe Attardi, Scott Emr, Ellen Rothenberg, and Paul Sternberg for their enthusiasm to discuss my thesis work.

I am also grateful to Dr. Ellen Strauss for her collaborations as well as her considerable help in writing a paper and to Dr. Charles Rice for teaching me the basic techniques including basketball which was one of his favorite exercises.

I would like to thank Ms. Edith Lenches for her excellent technical assistance as well as good friendship and Reef Hardy, Frank Preugschat, Kangsheng Wang, Drs. Richard Kuhn, Bert Niesters, and Yukio Shirako for many helpful discussions. Especially, Richard and Bert led me to finish my thesis just on time with their encouragement. And also thanks to Dr. Lee Hood and Tim Hunkapiller for letting me use their computer facility and teaching me how to use it.

I would like to thank my parents for their support and cheer and my little son for waking me up during the night to help me keep my late schedule. Finally, to my husband, Chang Soo, for all those discussions, encouragements, and arguments which he always brought from a different point of view. Without his companionship, this work would not have been possible.

Abstract

The genera Alphavirus and Flavivirus contain enveloped RNA viruses which are similar in size, morphology, and RNA content, and were once considered to belong to the same family, the Togaviridae. Recently the flaviviruses were reclassified as a separate family, the Flaviviridae, because they differ markedly from alphaviruses in replication strategy and mode of assembly as well as genome organization. In this thesis, representatives of both groups have been studied in order to understand the functions of the various virus-encoded proteins in replication and pathogenesis. Part I describes the mapping of temperature-sensitive (*ts*) RNA^- mutants of the alphavirus Sindbis virus, to elucidate the function of each nonstructural protein during RNA replication. Part II includes the determination of the complete nucleotide sequence of the flavivirus dengue 2, comparative analysis of conserved elements in the 3' untranslated region of various flaviviruses, and expression of dengue 2 structural proteins in a recombinant vaccinia virus.

During alphavirus replication, the parental 49S plus-strand RNA is transcribed into a complementary minus-strand RNA which serves as the template for the synthesis of both plus-strand 49S genomic RNA and 26S subgenomic RNA. The nonstructural proteins, which are involved in viral RNA replication, are translated from the genomic 49S RNA as two polyprotein precursors that are processed by cotranslational or posttranslational cleavage into four final polypeptide products.

ts RNA^- mutants of Sindbis virus have been isolated previously and grouped by complementation into four groups (A, B, F, G); these mutants fail to make RNA after infection at a nonpermissive temperature and were presumed to contain *ts*

lesions in the nonstructural proteins. Over a number of years, work in our own and other laboratories has established the details of the RNA$^-$ phenotypes of these mutants, which include defects in RNA chain elongation, in initiation of genomic and 26S RNA synthesis, in regulation of minus-strand template synthesis, and in processing of polyprotein precursors. However, it is only with the mapping described here that it has been possible to unambiguously assign these functions to particular nonstructural polypeptides. The mutations responsible for the *ts* phenotype of Sindbis complementation group F mutants *ts*6, *ts*110, and *ts*118, have been determined. *ts*6 and *ts*110 have a single base substitution in nsP4 resulting in a replacement of Gly by Glu at position 153 or position 324, respectively. It is of interest that nsP4 contains the Gly-Asp-Asp motif characteristic of a number of viral replicases, and this together with the fact that all RNA synthesis in *ts*6-infected cells is shut off upon shift-up from the permissive to the nonpermissive temperature suggests that nsP4 is the viral RNA polymerase or elongation enzyme. *ts*118 is a double mutant where one mutation is in nsP2 (the Val at residue 425 is changed to Ala). This mutation alone causes the formation of minute plaques at the nonpermissive condition without a reduction in the plaque number. The second change (Gln-93 of nsP4 changed to Arg) has little apparent phenotype on its own, but in combination with the change in nsP2 it leads to a temperature-sensitive phenotype. This suggests that nsP2 and nsP4 interact with one another in a complex.

We also have mapped representatives of RNA$^-$ complementation groups A, B, and G. Mutants belonging to groups A and G have been found to map in nsP2, suggesting that this protein is required for initiation of 26S RNA synthesis, proteolytic processing of the nonstructural precursor, and shut off of minus-strand

synthesis. *ts*11, the only member of group B, has a mutation in nsP1, indicating that this protein is responsible for initiation of minus-strand synthesis.

To understand the role of viral structural proteins in the pathogenesis of flaviviruses, we studied one of the dengue viruses, which constitute a worldwide health problem of increasing dimensions, and determined the complete nucleotide sequence of the PR159-S1 strain of dengue 2 virus except for 15 nucleotides at the 5' end. There is one long open reading frame which is translated to give the structural proteins, capsid (C), membrane-like protein (M), and envelope protein (E), followed by nonstructural proteins, NS1, NS2, NS3, NS4, and NS5. The individual proteins appear to be produced by posttranslational cleavage of a precursor polyprotein. There are nucleotide sequences in the 5' terminal region (in the coding region for the capsid protein) and in the 3' terminal region (in the 3' untranslated sequence) that are invariant among flaviviruses examined to date and that may be involved in cyclization of the RNA. These sequences are presumed to be important for viral replication. In addition, the 3' terminal 79 nucleotides are capable of forming a hairpin structure.

We have expressed the structural proteins of dengue 2 using a recombinant vaccinia virus to study the role of these proteins in the immunological response. The vaccinia recombinant containing a cDNA copy of the 5' region of the dengue genome virus expressed dengue structural proteins which are correctly cleaved and modified. This suggests that the sequences encoding the structural proteins specify all the necessary catalytic activities or recognition signals required to ensure the proper synthesis and maturation of the polypeptides. And also, this recombinant can generate dengue-specific antibodies in mice which neutralize viral infectivity.

Table of Contents

Introduction

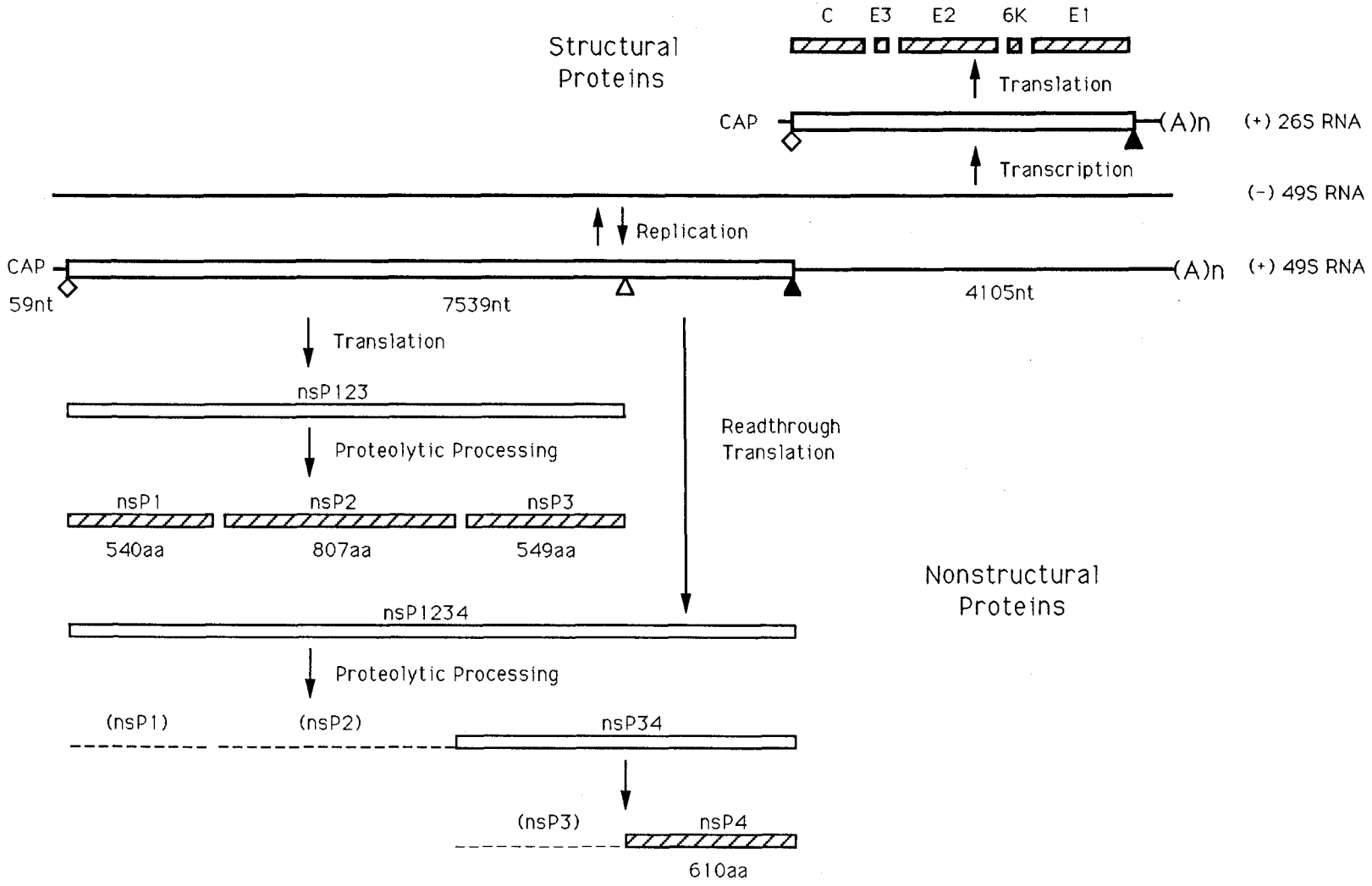
Introduction

Recent developments in recombinant DNA technology have greatly expanded our knowledge of the molecular biology of viruses. Nucleotide sequencing has established the structure of viral genomes, the nature of the encoded proteins and the translation strategies used by the viruses, as well as made possible various expression experiments for vaccine development. 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. Such clones make it possible to explore the functions of viral nonstructural and structural proteins.

The alphavirus genus of the family Togaviridae contains about 25 members, many of which are important human or veterinary pathogens. They are transmitted in nature by blood-sucking arthropods, usually mosquitoes, and thus replicate alternately in an arthropod host and in a vertebrate host. Sindbis virus is one of the most extensively studied alphaviruses since it is avirulent for man (Griffin, 1986) and can be conveniently worked with in the laboratory. Its genome is a single molecule of plus-strand RNA 11,703 nucleotides in length, that is capped at the 5' end and polyadenylated at the 3' end (Strauss et al., 1984). In the virus this RNA is surrounded by a single species of basic protein, the capsid protein, to form a nucleocapsid with cubic symmetry. As the final event in maturation this capsid buds through the plasma membrane of the infected cells and acquires an envelope containing lipids of host cell origin and two virus-encoded transmembrane glycoproteins, E2 and E1 (reviewed in Strauss and Strauss, 1986).

The replication strategy of Sindbis virus, the type alphavirus, is shown in Figure 1. Sindbis virus produces two messenger RNAs after infection. The first, 49S RNA, appears to be identical to the genomic RNA found in virions. The nonstructural proteins required to replicate the virus RNA and transcribe a subgenomic 26S RNA are translated from this RNA (Schlesinger and Kaariainen, 1980). Translation of the incoming RNA is a very early event in the virus replication cycle, and the naked RNA is itself infectious when transfected into a susceptible cell. After translation of the virus replicase, replication of the virus RNA can occur, as can transcription of the subgenomic 26S RNA. Use of a subgenomic mRNA to produce the structural proteins allows for the production of these proteins in great molar excess over the nonstructural proteins. The four nonstructural proteins, nsP1, nsP2, nsP3 and nsP4, are translated from the genomic 49S RNA as two polyprotein precursors that are subsequently processed by posttranslational cleavage (Figure 1). In Sindbis virus, as in many but not all alphaviruses, translation of nsP4 requires readthrough of an opal termination codon and nsP4 is underproduced relative to nsP1, nsP2, and nsP3 (Hardy and Strauss, 1988; Lopez et al., 1985; Strauss et al., 1983). It has been postulated that the processing of the nonstructural polypeptides is catalyzed by a virus-encoded protease located within one of the nonstructural polypeptides (reviewed in Strauss et al., 1987). The structural proteins are also translated as a single polypeptide from the subgenomic 26S mRNA and the capsid protein is believed to possess an autoproteolytic activity that cleaves the capsid protein from the nascent polyprotein during translation (Hahn et al., 1985). The remaining portion of the nascent polypeptide is then inserted into the rough endoplasmic reticulum of the cell where posttranslational modifications, including glycosylation and acylation as well as further processing and cleavage, occur.

Figure 1. Replication strategy of Sindbis virus, an alphavirus. RNAs of plus polarity are shown 5' to 3' from left to right. Untranslated regions of the RNAs are shown as single lines; translated open reading frames are shown as open boxes. Precursors of the nonstructural proteins are shown as open boxes and final structural and nonstructural polypeptides are shown as shaded boxes. (◇) indicate the initiating methionine codons; (▲) indicate termination codons; (Δ) is the opal codon readthrough to produce nsP4.


C
E3
E2
6K
E1
Structural Proteins
Translation
CAP
(A)n
(+) 26S RNA
Transcription
(-) 49S RNA
Replication
CAP
(A)n
(+) 49S RNA
59nt
7539nt
4105nt
Translation
nsP123
Proteolytic Processing
Readthrough Translation
nsP1
nsP2
nsP3
540aa
807aa
549aa
Nonstructural Proteins
nsP1234
Proteolytic Processing
(nsP1)
(nsP2)
nsP34
(nsP3)
nsP4
610aa

The role of the nonstructural proteins during alphavirus replication

During alphavirus replication the parental 49S plus-strand RNA is transcribed into a complementary minus-strand RNA which serves as a template for the synthesis of both genome 49S RNA and 26S subgenomic mRNA (reviewed in Strauss and Strauss, 1986). The synthesis of both plus- and minus-strand RNA increases during the first 3hr postinfection, but at 3 to 3.5hr after infection the synthesis of minus-strand ceases, whereas the synthesis of both 49S and 26S plus-strand RNA continues (Bruton and Kennedy, 1975; Sawicki and Sawicki, 1980). In this early phase, plus-strand RNA is produced in about 5-fold molar excess over minus-strand RNA.

A large number of temperature-sensitive (*ts*) mutants of the HR strain of Sindbis virus have been isolated and characterized (Burge and Pfefferkorn, 1966a; Strauss et al., 1976; Strauss and Strauss, 1980). Mutants may be defective in RNA replication (RNA^- mutants) or in the production of the structural proteins (RNA^+ mutants), and have been grouped by complementation into four RNA^- groups (A, B, F, G) and three RNA^+ groups (C, D, E) (Burge and Pfefferkorn, 1966b; Strauss et al., 1976). Representative mutant-revertant pairs from RNA^+ groups have been analyzed by sequence analysis, and there is an excellent correlation between specific nucleotide changes and *ts* phenotypes (Arias et al., 1983; Hahn et al., 1985; Lindqvist et al., 1986). A summary of the *ts* phenotypes of a number of these mutants and the localization of the mutations within the genome is shown in Figure 2.

RNA^- mutants presumably contain *ts* lesions in the viral nonstructural proteins which function to replicate viral RNA, and some of these RNA^- mutants have been characterized in depth (Keranen and Kaariainen, 1979; D. L. Sawicki

Figure 2. Functional defects of the temperature-sensitive mutants of Sindbis virus. Data are from this thesis and from Strauss and Strauss, 1980; Arias et al., 1983; Hahn et al., 1985; Lindqvist et al., 1986; Sawicki, D. L., et al., 1981; Sawicki and Sawicki, 1985. All *ts* mutants by definition are unable to produce progeny virions at the nonpermissive temperature. In the column labelled "RNA synthesis" (+/+) indicates that RNA is made at both 30° and 40°C; (-/+) indicates that RNA is not made at 40°C but once initiated at the permissive temperature, RNA synthesis continues after a shift to the nonpermissive temperature; (-/-) indicates no RNA synthesis at 40° or after a shift to 40°C. In the columns headed "Temperature-sensitive Phenotypes", functions which are *ts* are noted, and for those marked "shift", *ts* means that the function in question does not occur after a shift to the nonpermissive temperature, usually 40°C. nd = not determined.

COMPLEMENTATION GROUP: Mutant #		RNA SYNTHESIS (40°C/Shift)	PROTEIN MUTATED	Temperature Sensitive Phenotypes					
				Regulation of Minus Strand (Shift)	26sRNA Synthesis	p230 Cleavage	Induction Interferon (Shift)	Minus Strand Synthesis (Shift)	RNA Pol.
GROUP A: I	*ts* 15	-/+	nd	-	ts	ts	ts	-	-
	ts 17	-/+	nsP2	ts	ts	ts	nd	-	-
	ts 24	-/+	nsP2/nsP3	ts	ts	ts	nd	-	-
	ts 133	-/+	nd	ts	ts	ts	nd	-	-
	ts 21	-/+	nsP2	-	ts	-	ts	-	-
A:II	*ts* 4	-/+	nd	-	*	-	nd	ts	nd
	ts 14	-/+	nd	-	-	-	nd	-	nd
	ts 16	-/+	nd	-	-	-	nd	-	nd
	ts 19	-/+	nd	-	*	-	nd	-	nd
	ts 138	-/+	nd	-	-	-	nd	-	nd
GROUP G	*ts* 7	-/+	nsP2/nsP3	-	-	-	ts	-	-
	ts 18	-/+	nsP2	-	ts	ts	nd	-	-
GROUP B	*ts* 11	-/+	nsP1	-	-	-	(ts)	ts	-
GROUP F	*ts* 6	-/-	nsP4	-	ts	-	-	ts	ts
	ts 110	-/-	nsP4	-	nd	-	nd	nd	nd
	ts 118	-/-	nsP2/nsp4	-	nd	-	nd	nd	nd
GROUP C	*ts* 2, *ts* 5, *ts*13, *ts*128	+/+ +/+	Capsid	ts autoprotease, no capsids made, p140 precursor accumulates					
GROUP D	*ts*10, *ts* 23	+/+	E1	ts hemagglutinin, no virion assembly					
GROUP E	*ts* 20	+/+	E2	ts cleavage PE2→E2, no virion assembly					

*26S RNA synthesis is also defective at 30°C.

et al., 1981; S. G. Sawicki et al., 1981; Sawicki and Sawicki, 1985; Sawicki and Sawicki, 1986). Group A mutants were divided by Sawicki and Sawicki (1985) into two phenotypic subgroups, where subgroup I mutants (*ts*15, *ts*17, *ts*21, *ts*24 and *ts*133) were temperature-sensitive in synthesis of 26S mRNA synthesis, whereas subgroup II mutants (*ts*4, *ts*14, *ts*16, *ts*19 and *ts*138) were not. In addition, *ts*17, *ts*24 and *ts*133 failed to cleave the polyprotein precursor nsP123 and failed to shut off minus-strand synthesis upon shift to 40°C. One member of group G, *ts*18, was also found to be temperature-sensitive in synthesis of 26S mRNA and as well in the cleavage of nsP123, whereas a second mutant of group G, *ts*7, demonstrated neither of these defects (Figure 2). These studies suggest that groups A and G define functions involved in both initiation of 26S RNA synthesis and proteolytic cleavage of polyprotein precursors, and that the group A mutants might be also involved in the regulation of minus-strand synthesis.

The single mutant in group B, *ts*11, has a defect in minus-strand RNA synthesis, since upon shift from the permissive to a nonpermissive temperature, synthesis of plus-strand RNA continued, whereas minus-strand RNA synthesis ceased (D. L. Sawicki et al., 1981). This led to the hypothesis that group B mutants are involved in the initiation of minus-strand synthesis. Finally, mutant *ts*6 in complementation group F was found to cease all RNA synthesis upon shift to a nonpermissive temperature, leading to the hypothesis that group F mutants produce a temperature-sensitive polymerase, the elongation component of the viral replicase (D. L. Sawicki et al., 1981; Barton et al., 1988).

A full-length cDNA clone of Sindbis virus has been constructed and can be transcribed *in vitro* by SP6 RNA polymerase to produce infectious full-length transcripts (Rice et al., 1987). Viruses produced from *in vitro* transcripts are

identical to Sindbis virus and show strain-specific phenotypes reflecting the source of RNA used for cDNA synthesis (Lustig et al., 1988; Rice et al., 1987). The full-length clone can thus be used to map interesting phenotypes of Sindbis virus, and we have utilized this technology to map the location of mutations responsible for the *ts* phenotype of mutants belonging to all four RNA^- complementation groups. Mutations were approximately mapped by replacing restriction fragments from cDNA clones of each mutant into the full-length clone and the phenotype of the recombinant viruses assayed; the fragment responsible for the *ts* phenotype could then be sequenced to determine the nucleotide change. The same technology was used to reexamine the *ts* phenotypes by putting individual mutations into the same genetic background (Toto1101) to determine whether the phenotypes of the original mutants had been affected by unknown alterations in other portions of the genome. We found that three group F mutants have mutations in nsP4, that the single group B mutant had an alteration in nsP1, and that three group A and two group G mutants mapped to nsP2. Most of the mutations were the result of a single amino acid substitution in one protein, but in at least two mutants, changes in two different proteins contributed to the *ts* phenotype. From the mapping results and the mutant characterizations described above, some of the functions of the nonstructural proteins can be deduced. nsP1 appears to be required for the initiation of minus-strand synthesis, nsP4 appears to be the viral RNA polymerase (elongation enzyme), and nsP2 has several functions. It is required for initiation of 26S subgenomic RNA synthesis, it functions as a protease to process the polyprotein precursors of the nonstructural proteins, and it may be involved in the temporal regulation of minus-strand synthesis. The results are reported in detail in Chapters 1 and 2.

Structure of the flavivirus genome

The flaviviruses were formerly classified as a genus in the family Togaviridae, but have now been elevated to family status, family Flaviviridae, in part because of differences in replication and assembly (Westaway, 1980) and in part because their genome structure differs significantly from that of the alphaviruses (Rice et al., 1985). The family contains approximately 70 members, many of which are important human or veterinary pathogens. Most flaviviruses are transmitted by blood-sucking arthropods, mosquitoes or ticks, although some members of this group apparently lack an anthropod vector (Shope, 1980). The mosquito-borne flaviviruses have been divided into three subgroups on the basis of serological cross-reactions, the yellow fever (YF), dengue and Japanese encephalitis (JE) virus groups (Porterfield, 1980).

There are four serotypes of dengue viruses, which are important human disease agents causing millions of cases of dengue fever annually. These viruses have a very wide geographical distribution and are present in both the Old World (India, Southeast Asia, Oceania, Africa) and the New World (Central America, the Caribbean, and South America). Uncomplicated dengue fever is usually an acute illness characterized by joint pain, fever, and rash, but in a small percentage of cases complications develop leading to dengue hemorrhagic fever or dengue shock syndrome which have a significant mortality. The causes of dengue hemorrhagic fever or dengue shock syndrome are not yet understood. It has been hypothesized that this syndrome results from immune enhancement and usually develops when a person is infected with a second dengue virus within 6 to 12 months after infection by one of the other dengue serotypes (Halstead, 1981). However, at least some cases of primary infection by dengue virus also appear to result in dengue

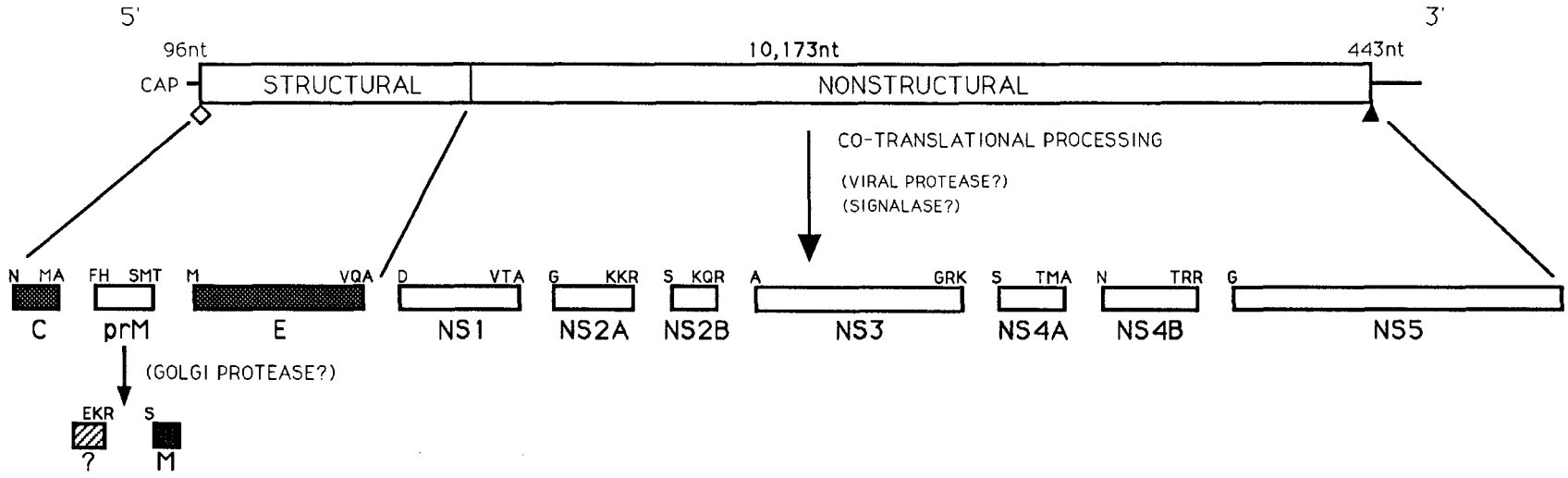
shock syndrome and it has been suggested that strains of increased virulence may play a role in the epidemiology of the disease (reviewed in Monath, 1986; Rosen, 1986).

To understand the molecular biology of the replication of these viruses and make possible various expression experiments that may lead to the development of vaccines against dengue viruses, the complete nucleotide sequence of the RNA of dengue 2 virus (S1 candidate vaccine strain derived from the PR-159 isolate), with the exception of about 15 nucleotides at the 5' end, has been determined (Hahn et al., 1988; Chapter 3). The genome organization of dengue virus, which is shown schematically in Figure 3, is the same as that previously found for YF virus (Rice et al., 1985) and for West Nile (WN) virus (Castle et al., 1985, 1986; Castle and Wengler, 1987; Wengler et al., 1985; Wengler and Castle, 1986). There are 96 nucleotides of 5' untranslated sequence and 443 of 3' untranslated flanking one long open reading frame (of 10,173 nucleotides) which is translated to give the structural proteins capsid (C), a protein precursor (prM) which contains the sequences of the membrane-like protein (M), and envelope protein (E), followed by nonstructural proteins NS1, NS2, NS3, NS4, and NS5. As is the case for other flaviviruses, the individual proteins appear to be produced by posttranslational cleavage of a precursor polyprotein. Comparison of the dengue 2 sequence and that of the encoded proteins with those of viruses from two other serological subgroups of mosquito-borne flaviviruses, YF and WN, and with the structural proteins of the Jamaican dengue 2 (Deubel et al., 1986) and of dengue 4 (Zhao et al., 1986) is instructive for what it tells us of the relationships of the flaviviruses to one another (see Chapter 3).

Figure 3. Replication strategy of dengue 2 virus, a flavivirus. Conventions are similar to those in Figure 1. The single long open reading frame in the genome is shown as an open box. The protein products are shown below. Open boxes are the nonstructural proteins, filled boxes are polypeptides found in the virion. The shaded moiety has not been detected directly. The amino acids at the cleavage sites are indicated. Data for the beginnings of the various proteins are from direct protein sequencing for E protein (Bell et al., 1985) and NS1, NS3, and NS5 (Biedrzycka et al., 1987) and from homology with the known cleavage sites in other flaviviruses for the remainder (Rice et al., 1985; Speight et al., 1988).

DENGUE 2 GENOME

(10,712nt)



RNA sequence elements

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; 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, for example, there are three conserved RNA sequence elements, 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, 1982b, 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). These elements are postulated to play important roles during RNA replication. As another example, many plant viruses have a tRNA-like structure at the 3' end of the genome which appears to serve an important function in replication (Hall, 1979). Or as a final example, segmented viruses, such as influenza virus and the bunyaviruses, contain conserved sequences at the ends of each RNA segment. In addition, these negative-strand viruses possess self-complementary sequences of 12-15 nucleotides at the ends of their genomes, such that the sequences at the 3' ends of both the plus- and minus-strand are the same. These sequences have been postulated to be binding sites for RNA replicases, and thus that for these viruses the same binding site is used for both plus- and minus-strand synthesis (reviewed in Strauss and Strauss, 1983).

In case of flaviviruses, the 3' terminal 79 nucleotides of dengue 2 RNA are capable of forming a hairpin structure (Hahn et al., 1987; Chapter 4), as has been observed for other flaviviruses (Rice et al., 1985; Brinton et al., 1986; Wengler and Castle, 1986; Zhao et al., 1986; Takegami et al., 1986). 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., 1987). These sequences are presumed to be important for virus replication. These elements and their significance will be further discussed in Chapter 4.

The role of structural proteins in flavivirus pathogenesis

Virions of dengue virus are relatively simple and contain only three virus coded proteins, designated capsid (C) protein, membrane-like (M) protein, and envelope (E) glycoprotein (Stollar, 1969). Intracellular virus lacks M but contains another glycoprotein, prM, which can be proteolytically cleaved to produce M (Shapiro et al., 1972; Rice et al., 1986). The E glycoprotein contains the major epitopes responsible for neutralization of infectivity by specific antibodies (Smith et al., 1970; Stevens et al., 1965). The envelope glycoprotein also exhibits hemagglutinating activity and is responsible for adsorption to the cell surface (Sweet and Sabin, 1954). Although immunity to flavivirus infection has generally been related to the presence of neutralizing antibodies against the virion envelope proteins, immunization with flavivirus nonstructural protein NS1 protects mice or monkey against lethal flavivirus infection (Schlesinger et al., 1985, 1986, 1987; Cane et al., 1988; Gould et al., 1988).

At the current time there is no effective vaccine for any of the dengue viruses; attempts to produce one have met with only limited success. In addition, the complications of dengue shock syndrome have led to caution being exercised in the development of a vaccine. The problem is exacerbated by the lack of an animal model system for the disease. Inactivated virus vaccines have been prepared from virus passed intracerebrally in mice (Sabin and Schlesinger, 1945; Schlesinger et al., 1956). Recently, tissue culture derived attenuated dengue viruses have been considered as candidate vaccines (Bancroft et al., 1984; Eckels et al., 1984). However, the development of recombinant DNA methods may provide novel alternatives for the development of a satisfactory dengue vaccine. In this thesis, the production of dengue 2 structural proteins via a recombinant vaccinia virus and the immunologic response of the recombinant will be discussed in Chapter 5.

References

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

Argos, P., Kamer, G., Nicklin, M. J. H., and Wimmer, E. (1984). Similarity in gene organization and homology between proteins of animal picornaviruses and a plant comovirus suggest common ancestry of these virus families. Nuc. 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.

Bancroft, W. H., Scott, R. M., Eckels, K. H., Hoke, C. H., Simms, T. E., Jesrani, K. D. T., Summers, P. L., Dubois, D. R., Tsoulos, D., and Russell, P. K. (1984). Dengue virus type 2 vaccine: Reactogenicity and immunogenicity in soldiers. J. Inf Dis. 149: 1005-1010.

Barton, D. J., Sawicki, S. G., and Sawicki, D. L. (1988). Demonstration *in vitro* of temperature-sensitive elongation of RNA in Sindbis virus mutant *ts6*. J. Virol. 62: 3597-3602.

Bell, J. R., Kinney, R. W., Trent, D. W., Lenches, E. M., Dalgarno, L., and Strauss, J. H., (1985). Amino-terminal amino acid sequences of structural proteins of three flaviviruses. Virology 143: 224-229.

Biedrzycka, A., Cauchi, M. R., Bartholomeusz, A., Gorman, J. J., and Wright, P. J., (1987). Characterization of the protease cleavage sites involved in the formation of envelope glycoprotein and three nonstructural proteins of dengue virus type 2, New Guinea C strain. J. Gen. Virol. 68: 1317-1326.

Brinton, M. A., Fernandez, A. V., and Dispoto, J. H. (1986). Sequence analysis of the 3' terminus of West Nile virus, strain E101, genome RNA. Virology 153: 113-121.

Bruton, C. J., and Kennedy, S. I. T. (1975). Semliki Forest virus intracellular RNA: Properties of the multistranded RNA species and kinetics of positive- and minus-strand synthesis. J. Gen. Virol. 32: 413-430.

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.

Cane, P. A., and Gould, E. A. (1988). Reduction of yellow fever virus mouse neurovirulence by immunization with a bacterially synthesized nonstructural protein (NS1) fragment. J. Gen. Virol. 69: 1241-1246.

Castle, E. and Wengler, G. (1987). Nucleotide sequence of the 5' terminal untranslated part of the genome of the flavivirus West Nile virus. Arch. Virol. 92: 309-313.

Castle, E., Nowak, T., Leidner, U., Wengler, G. and Wengler, G. (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: 227-236.

Castle, E., Nowak, T., Leidner, U., Wengler, G. and Wengler, G. (1986). Primary structure of the West Nile flavivirus genome region coding for all nonstructural proteins. Virology 149: 10-26.

Deubel, V., Kinney, R. M., and Trent, D. W. (1986). Nucleotide sequence and deduced amino acid sequence of the structural proteins of dengue type 2 virus, Jamaica genotype. Virology 155: 365-377.

Eckels, K. H., Scott, R. McN., Bancroft, W. H., Brown, J., Dubois, D. R., Summers, P. L., and Russell, P. K., and Halstead, S. B. (1984). Selection of attenuated dengue 4 viruses by serial passage in primary kidney cells. V. Human response to immunization with a candidate vaccine prepared in fetal rhesus lung cells. Am. J. Trop. Med. Hyg. 33: 684-689.

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.

Gould, E. A., Buckley, A., Barrett, A. D. T., and Cammack, N. (1986). Neutralizing (54K) and non-neutralizing (54K and 48K) monoclonal antibodies against structural and nonstructural yellow fever virus proteins confer immunity in mice. J. Gen. Virol. 67: 591-595.

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

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. U.S.A. 82: 4648-4652.

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

Hahn, Y. S., Galler, R., Hunkapiller, T., Dalrymple, J. M., Strauss, J. H. and Strauss, E. G. (1988). Nucleotide sequence of dengue 2 RNA and comparison of the encoded proteins with those of other flaviviruses. Virology 162: 167-180.

Hahn, Y. S., Grakoui, A., Rice, C. M., Strauss, E. G., and Strauss, J. H. (1989). Mapping of RNA$^-$ temperature-sensitive mutants of Sindbis virus: Complementation group F mutants have lesions in nsP4. J. Virol. In press.

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.

Halstead, S. (1981). The pathogenesis of dengue: Molecular epidemiology in infectious disease. Amer. J. Trop. Med. Hyg. 114: 632-648.

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

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

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

Keranen, S., and Kaariainen, L. (1979). Functional defects of RNA$^-$ *ts* mutants of Sindbis and Semliki Forest viruses. J. Virol. 32: 19-29.

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.

Lopez, S., Bell, J. R., Strauss, E. G., and Strauss, J. H. (1985). The nonstructural proteins of Sindbis virus as studied with an antibody specific for the C terminus of the nonstructural readthrough polyprotein. Virology 141: 235-247.

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. 62: 2329-2336.

Monath, T. P. (1986). Pathology of flaviviruses. In: *The Togaviridae and Flaviviridae*, S. Schlesinger and M. Schlesinger, eds., Plenum Publishing Corp., New York, pp. 375-440.

Ou, J. H., Rice, C. M., Dalgano, I., Strauss, E. G., and Strauss, J. H. (1982a). Sequence studies of several alphavirus genomic RNAs on the region encoding the start of the subgenomic RNA. Proc. Natl. Acad. Sci. U.S.A. 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.

Pelham, H. R. B. (1978). Leaky UAG termination codon in tobacco virus RNA. Nature (London) 272: 469-471.

Porterfield, J. S., Casals, J., Chumakov, M. P., Gaidamovich, S. Ya, Hannoun, C., Holmes, I. H., Horzinek, M. C., Mussgay, M., Oker-Blom, N., Russell, P. K., and Trent, D. W. (1978). Togaviruses. Intervirology 9: 129-148.

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: Implication for flavivirus gene expression and evolution. Science 229: 726-733.

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

Rice, C. M., Levis, R., Strauss, J. H., 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.

Rosen, L. (1986). The pathogenesis of dengue hemorrhagic fever. Supplement to South Amer. J. Med. 11: 40-42.

Sabin, A. B., and Schlesinger J. J. (1945). Production of immunity to dengue with virus modified by propagation in mice. Science 101: 640-642.

Sawicki, D. L., Sawicki, S. G., Keranen, S., and Kaarianen, L. (1981). Specific Sindbis virus coded function for minus-strand RNA synthesis. J. Virol. 39: 348-358.

Sawicki, D. L. and Sawicki, S. G. (1980). Short lived minus-strand polymerase for Semliki Forest virus. J. Virol. 34: 108-118.

Sawicki, D. L., and Sawicki, S. G. (1985). Functional analysis of the A complementation group mutants of Sindbis HR virus. Virology 144: 20-34.

Sawicki, S. G., and Sawicki, D. L. (1986). The effect of overproduction of nonstructural proteins on alphavirus plus-strand and minus-strand synthesis. Virology 152: 507-512.

Sawicki, S. G., Sawicki, D. L., Kaariainen, L., and Keranen, S. (1981). A Sindbis virus mutant temperature-sensitive in the regulation of minus-strand synthesis. Virology 115: 161-172.

Schlesinger, M. J., and Kaariainen, I. (1980). Translation and processing of alphavirus proteins. In *The Togaviruses*, R. W. Schlesinger, ed., Academic Press, New York, pp. 371-392.

Schlesinger, R. W., Gordon, I., Frankel, J. W., Winter, J. W., Patterson, P. R., and Dorrance, W. R. (1956). Clinical and serologic response of man to immunization with attenuated dengue and yellow fever viruses. J. Immunol. 77: 352-364.

Schlesinger, J. J., Brandriss, M. W., and Walsh, E. E. (1985). Production against 17D yellow fever encephalitis in mice by passive transfer of monoclonal antibodies to the nonstructural glycoprotein gp48 and by active immunization with gp48. J. Immunol. 135: 2805-2809.

Schlesinger, J. J., Brandriss, M. W., Cropp, C. B., and Monath, T. P. (1986). Production against yellow fever in monkeys by immunization with yellow fever virus nonstructural protein NS1. J. Virol. 60: 1153-1155.

Schlesinger, J. J., Brandriss, M. W., and Walsh, E. E. (1987). Protection of mice against dengue 2 encephalitis by immunization with dengue 2 virus nonstructural glycoprotein NS1. J. Gen. Virol. 68: 853-857.

Shapiro, D., Brant, W. E., and Russell, P. K. (1972). Change involving a viral membrane glycoprotein during morphogenesis of group B arboviruses. Virology 50: 906-911.

Shope, R. E. (1980). Medical significance of Togaviruses: An overview of diseases caused by Togaviruses in man and in domestic and wild vertebrate animals. In *The Togaviruses*, R. W. Schlesinger, ed., Academic Press, New York, pp. 47-82.

Smith, T. J., Brandt, W. E., Swanson, J. L., McCown, J. M., and Buesher, E. L. (1970). Physical and biological properties of dengue 2 virus and associated antigens. J. Virol. 5: 524-532.

Speight, G., Coia, G., Parker, M. D. and Westaway, E. G. (1988). Gene mapping and positive identification of the nonstructural proteins NS2A, NS2B, NS3, NS4B and NS5 of the flavivirus Kunjin and their cleavage sites. J. Gen. Virol. 69: 23-24.

Stevens, T. M., Schlesinger, R. W., and Schulze, I. T. (1965). Studies on the nature of dengue viruses. I. Correlation of particle density, infectivity and RNA content of type 2 virus. Virology 27: 103-112.

Stollar, V. (1969). Studies on the nature of dengue viruses. IV. The structural proteins of type 2 dengue virus. Virology 39: 426-438.

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., Rice, C. M., and Strauss, J. H. (1983). Sequence coding for the alphavirus nonstructural proteins is interrupted by an opal termination codon. Proc. Natl. Acad. Sci. U.S.A. 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*, R. W. Schlesinger, ed., 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. Current Topics in Microbiology and Immunology 105: 1-98.

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

Strauss, J. H., Strauss, E. G., Hahn, C. S., Hahn, Y. S., Galler, R., Hardy, W. R., and Rice, C. M. (1987). Replication of alphaviruses and flavivirus: Proteolytic processing of polyproteins. In *UCLA Symposium on Molecular and Cellular Biology*, Vol. 54, M. A. Brinton and R. Rueckert, eds., Alan R. Lisds, Inc., New York, pp. 209-225.

Sweet, B. H., and Sabin, A. B. (1954). Properties and antigenic relationships of hemmagglutinins associated with the dengue viruses. J. Immunol. 73: 363-373.

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

Wengler, G., Castle, E., Leidner, U., Nowak, T., and Wengler, G. (1985). Sequence analysis of the membrane protein V3 of the flavivirus West Nile virus and of its gene. Virology 147: 264-274.

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. (1980). Replication of flaviviruses. In *The Togaviruses*, R. W. Schlesinger, ed., Academic Press, New York, pp. 531-581.

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

PART I

The Role of the Nonstructural Proteins during Alphavirus Replication

Chapter 1

Mapping of RNA$^-$ Temperature-Sensitive Mutants of Sindbis Virus: Complementation Group F Mutants Have Lesions in nsP4

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

Mapping of RNA$^-$ Temperature-Sensitive Mutants of Sindbis Virus: Complementation Group F Mutants Have Lesions in nsP4

YOUNG S. HAHN[1], ARASH GRAKOUI[2], CHARLES M. RICE[2], ELLEN G. STRAUSS[1], and JAMES H. STRAUSS[1]

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

[2]*Department of Microbiology and Immunology, Washington University, St. Louis, MO 63110*

Running title: Sindbis group F *ts* mutants

ABSTRACT

Temperature-sensitive mutants of Sindbis virus belonging to complementation group F, *ts*6, *ts*110, and *ts*118, are defective in RNA synthesis at the nonpermissive temperature. cDNA clones of these group F mutants, as well as of ts^+ revertants, have been constructed. To assign the *ts* phenotype to a specific region in the viral genome, restriction fragments from the mutant cDNA clones were used to replace the corresponding regions of the full-length clone Toto1101 of Sindbis virus. These hybrid plasmids were transcribed *in vitro* by SP6 RNA polymerase to produce infectious transcripts, and the virus recovered was tested for temperature sensitivity. After mapping the *ts* lesion of each mutant to a specific region of 400-800 nucleotides using this approach, this region of the cDNA clones of both the *ts* mutant and ts^+ revertants was sequenced in order to determine the precise nucleotide change and amino acid substitution responsible for each mutation. Rescued mutants, which have a uniform background except for one or two defined changes, were examined for viral RNA synthesis and complementation to show that the phenotypes observed were the result of the mutations mapped. *ts*6 and *ts*110 have a single base substitution in nsP4 resulting in a replacement of Gly by Glu at position 153 or position 324, respectively. It is of interest that nsP4 contains the Gly-Asp-Asp motif characteristic of a number of viral replicases, and together with the fact that all RNA synthesis in *ts*6-infected cells and, to a lesser extent, in *ts*110-infected cells, shuts off when cells are shifted from a permissive to a nonpermissive temperature, suggests that nsP4 is the virus polymerase. *ts*118 is a double mutant. It contains a single base substitution in nsP2 resulting in a replacement of Val by Ala at position 425 that results in the formation of minute plaques, but not in a reduction in the plaque number at the nonpermissive condition. The second change, a substitution of Gln by Arg in *ts*118 at residue 93 in nsP4, has little apparent phenotype on its own, but

in combination with the change in nsP2 leads to a temperature sensitive phenotype. Thus in each case the mutation responsible for temperature sensitivity of the three known complementation group F mutants lies in nsP4. In addition, the result with *ts*118 suggests that nsP2 and nsP4 may interact with one another in a complex.

INTRODUCTION

Sindbis virus is a well studied member of the alphavirus family. Its genome is a single molecule of plus stranded RNA 11,703 nucleotides in length, that is capped at the 5' end and polyadenylated at the 3' end (35). During replication, the parental 49S plus-strand RNA is transcribed into a complementary minus strand which serves as a template for the synthesis of both 49S plus strand genomic RNA and for a 26S subgenomic RNA. Nonstructural polypeptides are translated from the genomic 49S RNA as two polyprotein precursors that are processed by cotranslational or post translational cleavage into four nonstructural proteins, called nsP1,2,3,4, which are required for RNA replication (11). Three structural polypeptides are produced by processing of a polyprotein precursor translated from the subgenomic 26S mRNA.

Large numbers of temperature-sensitive (*ts*) mutants of the HR strain of Sindbis virus have been isolated and characterized (3, 30,34). Mutants may be defective in RNA replication (RNA^- mutants) or in the production of the structural proteins (RNA^+ mutants), and have been grouped by complementation into four RNA^- groups (A,B,F,G) and three RNA^+ groups (C,D,E) (4, 30). Representative mutant-revertant pairs from RNA^+ groups have been analyzed by sequence analysis, and there is an excellent correlation between specific nucleotide changes and phenotypes (1, 10, 15). None of the RNA^- mutant defects has been rigorously assigned to specific nonstructural proteins or RNA

sequences. These mutants presumably contain *ts* lesions in the viral nonstructural proteins which function to replicate viral RNA.

A full length cDNA clone of Sindbis virus has been constructed that can be transcribed *in vitro* by SP6 RNA polymerase to produce infectious full-length transcripts (24). Viruses produced from *in vitro* transcripts are identical to Sindbis virus and show strain-specific phenotypes reflecting the source of RNA used for cDNA synthesis (17, 24). This full length clone can be used to map interesting phenotypes of Sindbis virus.

We have used this approach to define precisely the mutations responsible for the *ts* phenotypes of Sindbis complementation group F mutants. Mutants *ts*6, *ts*110, and *ts*118 of complementation group F are defective in RNA synthesis at the nonpermissive temperature. The best characterized member, *ts*6, ceases all viral RNA synthesis upon shift from permissive to nonpermissive condition, and it has been postulated that *ts*6 has a defect in the elongation activity of the replicase (2, 14, 26). cDNA clones of these mutants, as well as of ts^+ revertants, have been constructed, and restriction fragments from the mutant cDNA clones were used to replace the corresponding regions of a full-length clone of Sindbis. These plasmids were transcribed *in vitro* by SP6 RNA polymerase to produce infectious transcripts which were then tested for *ts* phenotype. The viruses recovered from these transcripts have been characterized. Together with sequence analysis of the cDNA clones, these experiments have defined the mutations responsible for the group F mutants.

MATERIALS AND METHODS

Virus stocks, growth and purification. Mutant *ts*6, obtained originally from Dr. B. Burge, was isolated from the HR strain of Sindbis virus following mutagenesis with nitrosoguanidine (3). Mutants *ts*110 and *ts*118 were isolated

from a small plaque strain of HR Sindbis virus following mutagenesis with nitrous acid (30). Revertants were isolated by plaquing mutant stocks at 30°C and 40°C. A single virus plaque of a ts^+ revertant was picked from the 40°C plate, and the virus was eluted into 1 ml of Eagle medium containing 5% fetal calf serum. This revertant plaque was used to infect primary chicken cells at 40°C, and the resulting stocks, following plaque assay at 30°C and 40°C, were used as infecting stocks for RNA preparation. Viruses were grown in primary or secondary chicken embryo fibroblasts and harvested 10-20 h after infection, depending on the mutant. All mutants and revertants were grown at 30°C or 40°C respectively. Viral RNA was isolated as described (23).

cDNA cloning. cDNA synthesis for *ts*6, *ts*110, and *ts*118, as well as for their revertants, followed the procedure of Okayama and Berg (20). A primer complementary to a sequence near the start codon of capsid protein (nt7642 - nt7661 of the genomic RNA) was made; this primer also contained the recognition site for *Xba*I restriction endonuclease at its 5' end. This primer was used for the first strand synthesis, and second strand synthesis used *E. coli* DNA Pol I, *E. coli* RNase H, and *E. coli* DNA ligase. Phosphorylated *Eco*RI linkers were ligated to the double stranded cDNA to facilitate later digestion with *Xba*I (see below). The double stranded cDNA was divided into two portions for cloning. For the 5' library, the *Acc*I (nt83)-*Sac*II (nt2771) fragment of the cDNA, which encodes nsP1 and the N-terminal half of nsP2, was cloned into Kahn5. Kahn 5 is a plasmid containing a cDNA copy of the 5' terminus of the Sindbis genome in Proteus 1, a vector consisting of the replicon and β-lactamase genes of pBR322 and an SP6 RNA polymerase promoter (H. V. Huang and C. M. R., unpublished; 24). The 3' library was constructed by cloning the *Bgl*II (nt2268)-*Xba*I (nt7662) fragment of the cDNA, encoding the C-terminal half of nsP2 and all of nsP3 and nsP4 into plasmid pMT21, an ampicillin resistant cloning vector derived from pBR322 (the

*Xba*I site is not present in this viral RNA but was introduced by the primer, as noted above).

Construction of hybrid genomes. Hybrid genomes were produced by replacing restriction fragments in Sindbis clone Toto1101 (24) with the corresponding regions from cDNA clones derived from the mutants or their revertants (17). Details of restriction sites used are included in the figure legends. Full-length hybrid plasmids that contained one of three non-overlapping intervals (A, B and C) from the mutants substituted into Toto1101 were first constructed for gross mapping. Plasmids with interval A contained the sequence of the mutant from the *Ssp*I (nt504) to the *Cla*I (nt2713) site in Toto1101. (Toto1101 contains approximately 13638 nucleotides; numbering begins from the first nucleotide of the Sindbis genome.) Since the *Ssp*I site is not unique, a shuttle vector, πnsP12 (S. A. Chervitz and C. M. R., unpublished), containing the *Sac*I (nt 13552, a site in the vector upstream of the SP6 promoter) to *Eco*RV (nt2750) region of Toto1101 cloned in πAN7 (Lutz *et al.*, 1987), was digested with *Ssp*I (nt504) and *Cla*I (nt2713) and ligated with the corresponding fragment of 5' cDNA of the *ts* mutant. The *Sac*I-*Cla*I fragment of the resulting clone was then cloned into Toto1101 which had been cut with *Sac*I and *Cla*I and treated with calf intestinal alkaline phosphatase. Interval B plasmids contained the sequence of the mutant from *Cla*I (nt2713) to *Spe*I (nt5262) in Toto1101, and was constructed by replacing this fragment in Toto1101 with the corresponding fragment from the 3' cDNA library of the *ts* mutant. Interval region C plasmids contained the *Spe*I (nt5262) to *Aat*II (nt7999) region of the *ts* mutant in Toto1101, and was constructed by digesting the 3' cDNA library of the *ts* mutant with *Spe*I (nt5262) and *Bam*HI (nt7334) and cloning into shuttle vector πnsP34, which is a πAN7 derivative containing the *Pvu*II (nt5160) to *Nco*I (nt8038) fragment of Toto1101

(obtained from H. V. Huang). The *Spe*I to *Aat*II fragment of the resulting clones was then used to replace the corresponding fragment of Toto1101.

For fine mapping of the B region, plasmids were constructed that contained three overlapping subregions referred to as B1, B2, B3. Subregion B1, covering the region *Cla*I (nt2713) - *Avr*II (nt4280), and B3, covering the *Avr*II (nt4280)-*Spe*I (nt5262) region, were directly cloned into Toto1101. Subregion B2, covering the *Ava*I (nt3546)-*Bam*HI (nt4633) region, was obtained from the shuttle vector Kahn5B, consisting of the *Cla*I (nt2713) to *Eco*RI (nt5869) fragment of Sindbis subcloned into Kahn5. Three clones containing overlapping subregions C1, C2, and C3 were constructed for the fine mapping of C region, using πnsP34. Fragments *Spe*I (nt5262)-*Hin*dIII (nt6267), *Pst*I (nt5824)-*Hpa*I (nt6919), or *Nsi*I (nt6461)-*Bam*HI (nt7334) of the *ts* mutant were cloned into πnsP34, and the *Spe*I (nt5262)-*Aat*II (nt7999) fragment was used to replace the corresponding fragment in Toto1101.

***In vitro* transcription and transfection.** RNA transcripts were synthesized *in vitro* with SP6 RNA polymerase, using supercoiled plasmid template or plasmid DNA digested with the appropriate restriction endonuclease for the production of run off transcripts, as described previously (24). The resulting transcripts were transfected into confluent monolayers of secondary chick cells in 35 mm multi well tissue culture plates, and the phenotype of the recovered virus was tested. Plaques were quantitated by overlaying the monolayers with 2 ml of 1% agarose in Eagle medium containing 2% fetal calf serum, followed by incubation at 30°C and 40°C. Plaques were visualized by staining with neutral red or crystal violet after incubation for 36 to 40 hr at 40°C or for 60 to 72 h at 30°C.

Analysis of viral RNA synthesis. Chick embryo fibroblast monolayers (60 mm plate) were infected with Sindbis HR or *ts* mutants or recombinant viruses recovered from the hybrid cDNA clones at a multiplicity of 50 pfu/cell in PBS (6) containing 1% fetal calf serum and 1 μg/ml ActD and incubated at 30°C or 40°C

for 1 h. At the end of the adsorption period, the inocula were removed, the cells were washed with prewarmed medium, and incubated at 30°C or 40°C, respectively, in Eagle medium containing 3% fetal calf serum and 1 μg/ml ActD. For shift from 30°C to 40°C, at 3.5h postinfection (p.i.) one set of 30°C plates was washed once with prewarmed medium lacking ActD, prewarmed medium containing 1 μg/ml ActD was then added and the plates were shifted to 40°C. At 10 h p.i. at 30°C or 6 h p.i. at 40°C, or at 8 h p.i. following shift to 40°C, cells were harvested.

The amount of viral RNA present was quantitated by the cytoplasmic dot hybridization method of White and Bancroft (37). Equal numbers of cells were washed with cold PBS, then lysed with 1% Nonidet P-40 in TE buffer (10 mM Tris-Cl, pH 7.5, 0.1 mM EDTA). Nuclei were pelleted and the supernatant was treated with 14.8% formaldehyde in 1 X SSC (150 mM sodium chloride, 15 mM sodium citrate) at 60°C for 15 min. 10-20 μl of RNA samples were blotted onto the nitrocellulose membranes and probed with ^{32}P-labeled minus strand RNA from the region of the genome (the "26S" region) encoding the structural proteins. This probe was transcribed with SP6 RNA polymerase from a cDNA clone of Sindbis virus that contained the structure protein region only inserted in an inverted sense downstream from an SP6 promoter. Relative amounts of RNA were determined by a beta-scanning counter. All results were corrected for the amount of incorporation into mock infected plates, which was between 0.5% and 1% of the incorporation into cells infected with the parental strain of Sindbis virus.

Alternatively, RNA synthesis following shift was assayed by examining the incorporation of [^{3}H]uridine into infected cells. Following infection at 30°C, cells were shifted at 3 hr p.i. to 40°C, and labeled with [^{3}H]uridine (20 μCi/ml) in the presence of ActD from 3.5 to 8 hr p.i. Monolayers were then washed with PBS,

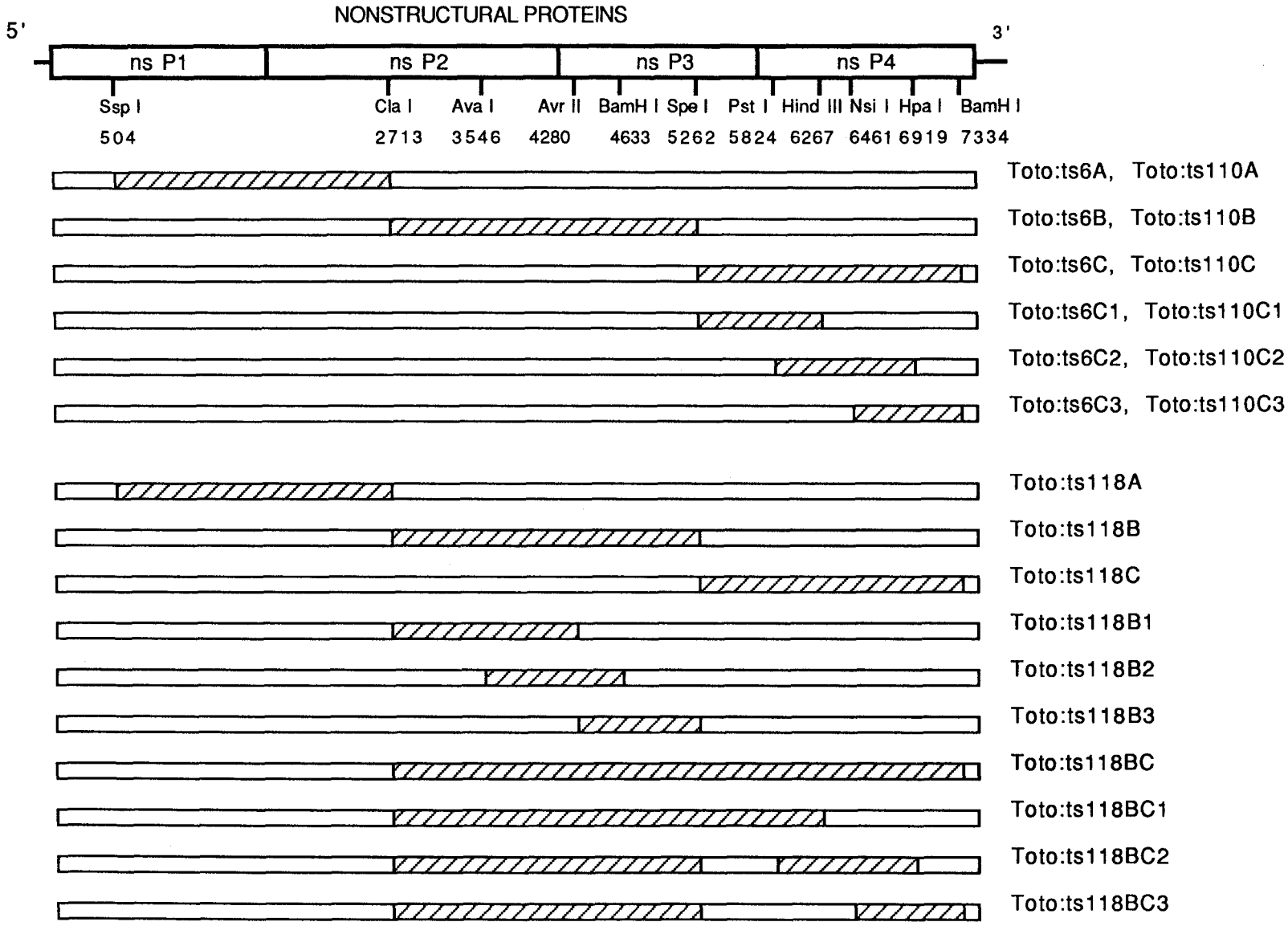
lysed with 0.5 ml of 2% SDS, and 50 µl was precipitated with TCA and the incorporated radioactivity quantitated by liquid scintillation counting.

Complementation analysis. Complementation tests were performed as described by Strauss *et al.* (30) but employed 35 mm multiwell plates and an moi of 20 pfu/cell for each mutant. A complementation index was calculated as the yield from the mixed infection divided by the sum of the yields following infection by each parent alone. A complementation index was calculated separately for each mutant in tests in which the two mutants differed markedly in plaque size. The absolute magnitude of the complementation index is dependent upon the yield of parental viruses (i.e., the extent of leakage of the parents), as complementation is always inefficient (4, 5, 30), not exceeding 1-10% of the wild type yield, and in the case of *ts*118 only one way complementation could be demonstrated because of relatively high yields of *ts*118 at 40°C.

RESULTS

Construction of recombinant plasmids. In order to localize the *ts* mutations of *ts*6, *ts*110, and *ts*118, we have constructed and tested a number of recombinant plasmids. The constructs are illustrated schematically in Fig. 1. In each case small (873-2584 nucleotides) restriction fragments in the Sindbis virus cDNA clone Toto1101, from which infectious RNA can be transcribed *in vitro* with SP6 RNA polymerase (24), were replaced with cDNA from a mutant or its revertant. In this figure the restriction sites used to construct the hybrid genomes and their numbering from the 5' end of the RNA (33) are also shown. The genomic region encoding the nonstructural proteins was first divided into three large nonoverlapping regions, A, B, and C, for gross mapping, as shown. For fine mapping, regions B and C were each subdivided into 3 overlapping regions (B1, B2, B3 or C1, C2, C3, respectively). The 5' 444 nucleotides and the 3' 265 nucleotides

Fig. 1. Construction of hybrid genomes. A schematic of the nonstructural protein-coding region of Sindbis virus cDNA clone Toto1101 (24) is shown together with a number of restriction sites and their position in the Sindbis genome numbered from the 5' end according to Strauss *et al.* (33). Translated regions are shown as the open box in which the names of the various proteins are indicated. Nontranslated regions are shown as a single line. The dashed boxes show the location of restriction fragments in clone Toto 1101 that were replaced with the corresponding restriction fragments from the temperature-sensitive mutants or their revertants. The names used to refer to these hybrid clones are shown at the right.


NONSTRUCTURAL PROTEINS
5'
3'
ns P1
ns P2
ns P3
ns P4
Ssp I
Cla I
Ava I
Avr II
BamH I
Spe I
Pst I
Hind III
Nsi I
Hpa I
BamH I
504
2713
3546
4280
4633
5262
5824
6267
6461
6919
7334
Toto:ts6A, Toto:ts110A
Toto:ts6B, Toto:ts110B
Toto:ts6C, Toto:ts110C
Toto:ts6C1, Toto:ts110C1
Toto:ts6C2, Toto:ts110C2
Toto:ts6C3, Toto:ts110C3
Toto:ts118A
Toto:ts118B
Toto:ts118C
Toto:ts118B1
Toto:ts118B2
Toto:ts118B3
Toto:ts118BC
Toto:ts118BC1
Toto:ts118BC2
Toto:ts118BC3

of the coding region for the nonstructural proteins, which were not covered by hybrid genome constructions, were sequenced in each case to insure that no changes had occurred within these regions.

RNA was transcribed *in vitro* from the recombinant plasmids with SP6 RNA polymerase and transfected onto monolayers of chicken cells. Monolayers were incubated under agarose at 30°C or 40°C to determine whether or not the virus recovered in each case was temperature sensitive.

Localization of the mutations in *ts*6, *ts*110, and *ts*118. The results obtained with the constructs tested are summarized in Table 1. For *ts*6, of the three large interval replacement clones (Toto:*ts*6A, Toto:*ts*6B, Toto:*ts*6C) tested, plasmids Toto:*ts*6A and Toto:*ts*6B gave rise to virus that exhibited wild type growth at the nonpermissive temperature, while *ts* virus was obtained from plasmid Toto:*ts*6C. This localized the *ts* mutation to the interval 5262-7334 of the genome. Plasmids Toto:*ts*6C1, Toto:*ts*6C2, and Toto:*ts*6C3, containing three smaller intervals in the C region, were then constructed and tested. Plasmids Toto:*ts*6C1 and Toto:*ts*6C2 gave rise to temperature sensitive virus whereas Toto:*ts*6C3 did not (Table 1). Thus, *ts*6 has one or more mutations in the region between *Spe*I (nt5262) and *Hpa*I (nt6919); if only a single mutation is involved it must lie in the region of overlap of C1 and C2, between *Pst*I (nt5824) and *Hin*dIII (nt6267), which is located near the N-terminus of nsP4.

For *ts*110, virus recovered from plasmid Toto:*ts*110C was *ts*, whereas virus from Toto:*ts*110A and Toto:*ts*110B was not. Plasmids Toto:*ts*110C1, Toto:*ts*110C2, and Toto:*ts*110C3 were then constructed and tested for fine mapping of the interval region C of *ts*110. *ts* viruses were obtained from recombinant plasmids Toto:*ts*110C2 and Toto:*ts*110C3, whereas plasmid Toto:*ts*110C1 gave rise to wild type virus. From this we conclude that *ts*110 has one or more mutations in the region between nucleotides 5824 to 7334 of the genome, and that if a single mutation is involved, it must lie between nucleotides 6461 and 6919 of the genome. This region is also in nsP4.

Table 1

Summary of the constructs tested for ts phenotype at the nonpermissive temperature

Recombinant clone[a]	Replaced fragment	Phenotype[b]	Mutant localization
Toto:ts6A	nt 504-2713	wt	
Toto:ts6B	nt 2713-5262	wt	
Toto:ts6C	nt 5262-7334	ts	nt 5824-6267
Toto:ts6C1	nt 5262-6267	ts	
Toto:ts6C2	nt 5824-6919	ts	
Toto:ts6C3	nt 6461-7334	wt	
Toto:ts110A	nt 504-2713	wt	
Toto:ts110B	nt 2713-5262	wt	
Toto:ts110C	nt 5262-7334	ts	nt 6461-6919
Toto:ts110C1	nt 5262-6267	wt	
Toto:ts110C2	nt 5824-6919	ts	
Toto:ts110C3	nt 6461-7334	ts	
Toto:ts118A	nt 504-2713	wt	
Toto:ts118B	nt 2713-5262	(ts)[c]	
Toto:ts118B1	nt 2713-4280	(ts)[c]	
Toto:ts118B2	nt 3546-4633	wt	nt 2713-3546
Toto:ts118B3	nt 4280-5262	wt	nt 5824-6267
Toto:ts118C	nt 5262-7334	wt	
Toto:ts118BC	nt 2713-7334	ts	
Toto:ts118BC1	nt 2713-6267	ts	
Toto:ts118BC2	nt 2713-5262,5824-6919	ts	
Toto:ts118BC3	nt 2713-5262,6461-7334	(ts)[c]	

[a] The clones are described in Fig. 1.

[b] RNA transcripts were transfected onto cells at 30 or 40 C and the plaque titer determined as described in Materials and Methods.

[c] Partially ts, in that plaque size but not plaque number, is reduced at 40 C

Mapping of *ts*118 suggested that it was a double mutant in which one mutation is located in the nsP4 region as was the case for *ts*6 and *ts*110 and the second mutation was in a different region. Viruses from plasmids Toto:*ts*118A and Toto:*ts*118C were apparently wild type whereas that from plasmid Toto:*ts*118B was partially *ts* (Tables 1 and 2). This partial temperature-sensitivity manifested itself as a change in plaque size from small size plaques at 30°C to minute plaques at 40°C, although the number of plaques at 30°C and 40°C were the same; and RNA synthesis at 40°C was reduced relative to Toto1101 virus (Table 2). When a construct Toto:*ts*118BC was tested, the virus once again formed small plaques at 30°C, whereas at 40°C minute plaques were formed with the plaque number being reduced by 4 orders of magnitude (Table 2), as is the case for the parental *ts*118. Furthermore, RNA synthesis at 40°C is reduced to very low levels (Table 2). Thus, we conclude that *ts*118 is a double mutant in which a mutation in the B region results in formation of minute plaques at 40°C but no change in plaque number, and that a second mutation in the C region, although having little apparent phenotype on its own, when combined with the change in the B region, results in a reduction of plaque number (and thus scoring as *ts* in plaque assays).

Fine mapping of the two *ts*118 mutations was done by constructing Toto:*ts*118B1, B2, and B3 (Fig. 1) and the change responsible for formation of minute plaques at 40°C mapped to region B1 (Table 1). Similarly, the change in the C region that when combined with the B region change, results in reduction in plaque count was mapped to the region of overlap in C1 and C2 between nt 5824 and nt 6267.

Sequence analysis of *ts*6, *ts*110, *ts*118 and their revertants. In order to define the *ts* lesions of *ts*6, *ts*110, *ts*118, regions shown by the mapping experiments to contain the *ts* lesions were sequenced by the chemical method of Maxam and Gilbert (19) as modified by Smith and Calvo (29), using cDNA clones of

Table 2

Plaque morphology and RNA synthesis by ts118 at 40°C and 30° C

Virus	Plaque morphology		E.O.P.[a]	RNA synthesis[b]		
	40°C	30°C	(40 °C /30°C)	40°C	30°C	shift to 40°C
Toto1101	Large plaque	Large plaque	2.6×10^{-1}	1.00	1.00	1.00
ts118	Minute plaque	Small plaque	5.0×10^{-5}	0.07	0.70	0.32
Toto:ts118B1	Minute plaque	Small plaque	4.0×10^{-1}	0.28	0.99	0.58
Toto:ts118C	Large plaque	Large plaque	5.0×10^{-1}	0.71	0.93	0.84
Toto:ts118BC	Minute plaque	Small plaque	5.0×10^{-5}	0.08	0.73	0.48
Toto:ts118BC*[c]	Minute plaque	Small plaque	2.6×10^{-5}	0.08	0.76	0.44
ts118R	Large plaque	Large plaque	5.4×10^{-1}	0.70	1.14	0.81

[a] Plaque titer at 40° C divided by that at 30°C

[b] Viral RNA synthesis was assayed by dot hybridization following infection at 40°C, 30°C, or after shift to 40°C following infection at 30°C, as described in the Materials and Methods.

[c] Toto:ts118BC* contains the B fragment from ts118 and the C fragment from ts118R.

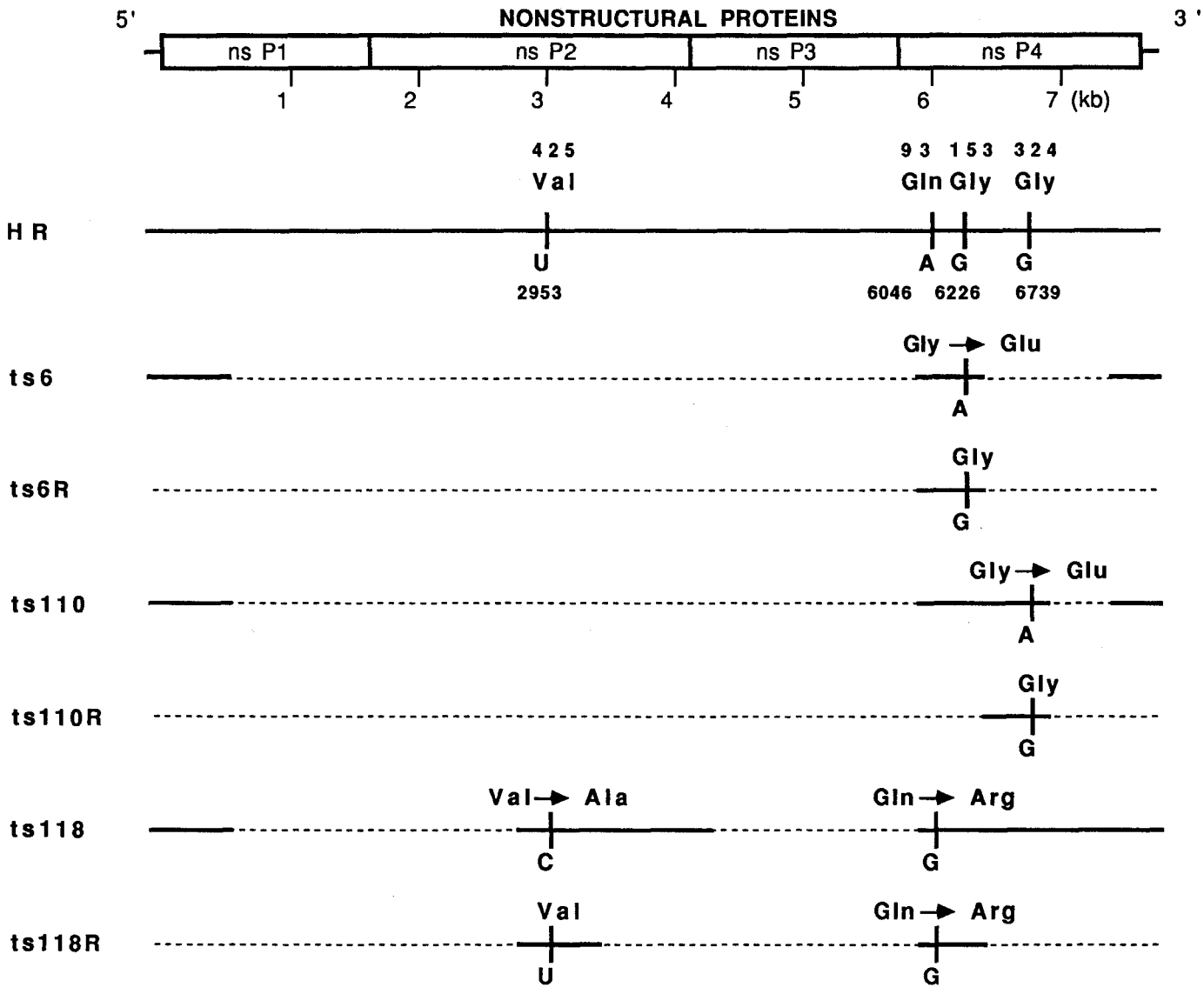
mutants as well as of their revertants. The sequences obtained are shown in Fig. 2.

*ts*6 has a single base substitution in the region sequenced. Comparing the *ts*6 sequence with that of its revertant, and from the results in Table 1, we found the mutation responsible for temperature sensitivity in *ts*6 is a change of G to A at nucleotide 6226, which leads to the replacement of Gly (GGG) at position 153 of nsP4 by Glu (GAG). In the revertant the changed nucleotide reverts to the original nucleotide, restoring the parental amino acid. *ts*110 also has only one change throughout the sequenced region. The change is G to A at nucleotide 6739, resulting in the change of Gly (GGG) to Glu (GAG) at position 324 of nsP4. In the *ts*110 revertant, this nucleotide reverts to the original nucleotide.

*ts*118 has a single base substitution in the region between nt2713 and nt3546. The change is U to C at nucleotide 2953, resulting in the change of Val (GUG) to Ala (GCG) at position 425 of nsP2. As discussed earlier, this change results in reduction in plaque size but not number at 40°C. In the *ts*118 revertant, this nucleotide reverts to the original nucleotide.

The second change in *ts*118 was found to be a change of A to G at position 6046 resulting in a Gln to Arg substitution at position 93 of nsP4. This change combined with the change in nsP2 renders the virus temperature sensitive in that plaque number as well as plaque size are reduced at 40°C. The revertant selected, *ts*118R, retains the change in nsP4 which, as noted earlier, has little apparent phenotype on its own. An additional construct was made and tested in order to show that this change is in fact responsible in part for temperature sensitivity. The *ts*118R C region was combined with Toto:*ts*118B to produce a virus Toto:*ts*118BC* which upon assay was temperature sensitive, forming tiny plaques at 40°C in reduced numbers (Table 2).

Fig. 2. Localization of temperature-sensitive mutations. A schematic of the nonstructural coding region of Sindbis virus is shown. Below are shown sequencing schematics for HR Sindbis virus (33), the parental strain from which the *ts* mutants were isolated (34), and for mutants *ts*6, *ts*110, *ts*118, and their revertants. Sequenced regions are shown as solid lines. Any change from the HR sequence on the first line is indicated. Where no changes are shown the sequence is identical to that of HR. Nucleotides are numbered from the 5' end of the RNA, amino acids from the N-terminus of each protein.


5'
NONSTRUCTURAL PROTEINS
3'
ns P1
ns P2
ns P3
ns P4
1
2
3
4
5
6
7 (kb)
425
Val
93 153 324
Gln Gly Gly
HR
U
2953
A G G
6046 6226 6739
Gly → Glu
ts6
A
Gly
ts6R
G
Gly → Glu
ts110
A
Gly
ts110R
G
Val → Ala
Gln → Arg
ts118
C
G
Val
Gln → Arg
ts118R
U
G

Characterization of the *ts* mutations rescued from *ts*6, *ts*110, and *ts*118. As a control to establish that the mutations mapped and defined here are the ones responsible for the phenotypes described previously for these mutations, and in order to establish the phenotype of these mutations in a uniform background, recombinant viruses containing a defined region from each of the *ts* mutants in a Toto1101 background were studied. Monolayers were transfected with dilutions of RNA transcribed from recombinant plasmids Toto:*ts*6C2, Toto:*ts*110C2, Toto:*ts*118B1, Toto:*ts*118C, Toto:*ts*118BC, and Toto:*ts*118BC*, and incubated at 30°C and 40°C, with results similar to those shown in Table 1. A single plaque of each virus was isolated from the 30°C plate, a stock grown at 30°C and these plaque purified virus stocks were further characterized. To start, these recombinant viruses were titered at 30°C and 40°C to determine the relative efficiency of plaque formation, and the results are shown in Tables 2 and 3. The virus stocks derived from these infectious transcripts clearly showed the temperature sensitivity of the parental mutant in the case of *ts*6 and *ts*110, and the efficiency of plaque formation at 40°C relative to 30°C was low (Table 3) (the relatively high apparent reversion rate of *ts*6 is due to use of a stock that had not been recently plaque purified; during passage revertants are amplified in most *ts* mutant stocks). However, as noted earlier, the virus recovered from Toto:*ts*118B1 formed minute plaques at 40°C and the number of plaque forming units was only slightly reduced (Table 2). The amino acid change of Val to Ala in nsP2 is thus responsible for the minute plaque phenotype at the nonpermissive temperature but not for the reduced number of plaques. The mutation in nsP4 (construct Toto:*ts*118C) had little effect on its own, but in combination with the nsP2 change (constructs Toto:*ts*118BC and Toto:*ts*118BC*) led to a pronounced decrease in the efficiency of plaque formation at 40°C (Table 2).

Table 3

Efficiency of plaque formation and RNA synthesis by ts6 and ts110 at 40°C and 30°C

Virus	Titer (PFU/ml)		E.O.P.[a]	RNA synthesis		
	40°C	30°C	(40°C /30°C)	40°C	30°C	shift$_o$ to 40°C
Toto1101	5.2×10^{8}	2.0×10^{9}	2.6×10^{-1}	1.00	1.00	1.00
ts 6	2.0×10^{6}	1.4×10^{9}	1.4×10^{-3}	0.04	0.38	0.02
Toto:ts6C2	2.7×10^{5}	8.0×10^{9}	3.4×10^{-5}	0.02	0.46	0.05
ts110	4.0×10^{4}	3.0×10^{9}	1.3×10^{-5}	0.05	0.72	0.16
Toto:ts110C2	1.2×10^{4}	2.1×10^{9}	5.7×10^{-6}	0.05	0.64	0.12

[a] Definitions are the same as in Table 2

To further examine these viruses, RNA synthesis was analyzed after infection at 30°C, after infection at 40°C, and at 40°C following a shift from 30°C. The parental mutants *ts*6, *ts*110, and *ts*118 were included, as was virus recovered from clone Toto1101. Total viral RNA synthesis was analyzed by the cytoplasmic hybridization method of White and Bancroft (1982), using ^{32}P-labeled minus strand RNA transcribed from the structural protein region as a probe. The values determined, relative to Toto1101 virus, are shown in Tables 2 and 3. The synthesis of minus strand RNA is significantly less than that of plus strand RNA and therefore the amount of plus strand RNA detected by hybridization was assumed to be the total viral RNA. Following infection at 40°C, RNA synthesis by the viruses recovered from Toto:*ts*6C2 and Toto:*ts*110C2 was reduced to a level similar to that seen after infection by their respective parents. However, the virus recovered from Toto:*ts*118B1 showed a higher level of RNA synthesis than that of the parental *ts*118 at the nonpermissive temperature (Table 2). This is consistent with the observation that the mutation in nsP2 gives rise to minute plaques at 40°C but does not reduce plaquing efficiency. RNA synthesis by virus from Toto:*ts*118C, containing only the change in nsP4, was only slightly reduced at 40°C relative to Toto1101 virus, and is the same as that by *ts*118R. However, RNA synthesis by the double mutant Toto:*ts*118BC or Toto:*ts*118BC* at 40°C was low and identical to that of the parental *ts*118, and thus the nsP4 mutation at position 93 markedly reduces the RNA synthesis at the nonpermissive temperature when it is combined with the nsP2 change.

The mutants and the constructs were also tested for RNA synthesis after establishing infection at 30°C and shifting to 40°C, to compare with the results of Keranen and Kaariainen (14) and Sawicki *et al.* (26) for *ts*6 (see also ref. 2). *ts*6 makes very little RNA after shift to 40°C, as found by Sawicki *et al.* (26), and the virus recovered from Toto:*ts*6C2 exhibits the same phenotype (Table 3). *ts*110

also makes little RNA after shift to 40°C, as does Toto:*ts*110C2 (Table 3). From a more detailed study of the kinetics of cesation of RNA synthesis in *ts*6-infected cells upon shift up (14, 26), and from studies of *ts*6 replication complexes *in vitro* (2), it was concluded that the elongation of RNA chains, as opposed to initiation, is temperature sensitive in the case of *ts*6, and it was postulated that the F group function, here shown to lie in nsP4, defines the RNA polymerase. The *ts*118 constructs such as Toto:*ts*118BC make significant amounts of RNA after a shift, much more RNA than when infection and incubation are continuously at 40°C (Table 2). Thus the replicase complexes of *ts*118, once formed at 30°C, are active upon shift to 40°C, in contrast to those specified by *ts*110 or *ts*6.

It is worth noting that the assay used for these shift experiments examines total virus plus strand RNA in the infected cells at 8 hr post-infection, after shifting at 3.5 hr. The results make clear that the RNA present at 3.5 hr of infection at 30°C does not contribute significantly to the RNA pool at 8 hr, although the replicase enzymes needed for essentially a full yield of virus RNA are present. As a control for these results, we have repeated the shift experiments and examined labeled RNA made between 3.5 hr and 8 hr in the presence of ActD and [^{3}H]uridine. The results were similar to those presented in Tables 2 and 3.

Complementation analysis of rescued mutants. We also examined the ability of these viruses to complement representative *ts* mutants from the three other complementation groups of RNA$^-$ mutants. The complementation indices shown in Table 4 demonstrate that the viruses derived from Toto:*ts*6C2 and Toto:*ts*110C2 complement the other three complementation groups of RNA$^-$ mutants, as do the parental viruses, but do not complement their parental viruses or each other, in agreement with previous complementation results (30).

Interpretation of results with virus derived from Toto:*ts*118B1 and Toto:*ts*118BC is complicated because of the high titer of virus produced at 40°C (the virus from Toto:*ts*118B1 is only marginally temperature sensitive as noted

Table 4

Complementation analysis of virus derived from Toto:ts6C2, Toto:ts110C2, Toto:ts118B1 and Toto:ts118BC [a]

Virus	A	B	G	F		
	ts 24	ts 11	ts 18	ts 6	ts 110	ts 118
Toto:ts6C2	53	43	55	0.4	–	–
Toto:ts110C2	67	42	42	3	0.3	–
Toto:ts118B1	12	7	7	4	–	0.2
Toto:ts118BC	16	104	4	1	–	0.2

[a] Complementation indexes shown are the yield from mixedly infected cells divided by the sum of the yields from singly infected cells. For ts118 constructs, the complementation indexes are "one way", as described in the Materials and Methods.

earlier, and virus from Toto:*ts*118BC also leaks at 40°C compared to the other mutants studied). In these cases the complementation indices shown are one way indices. Because viruses from Toto:*ts*118B1 and Toto:*ts*118BC form minute plaques, it is possible to distinguish these plaques from those formed by the other *ts* mutants used. The complementation indices shown are the yield of large plaques in mixed infection divided by the yield of the large plaque parents during single infection. Even so, complementation by Toto:*ts*118B1 was marginal, and we cannot assign it to a complementation group, although it did seem to complement all of the other mutants tried. Analysis of the double mutant Toto:*ts*118BC showed that it did not complement its parental virus or *ts*6 and complemented only poorly with *ts*18 in group G, but complemented better with the A mutant *ts*24 and quite well with B mutant *ts*11. Thus except for the marginal complementation of Toto:*ts*118BC with *ts*18, the results with the rescued mutations are in good agreement with previous complementation results (30).

DISCUSSION

In this paper we have localized the mutations responsible for the *ts* phenotype of complementation group F mutants. *ts*6 and *ts*110 have been mapped to nsP4. Each of them has a single base substitution resulting in a replacement of Gly by Glu at positions 153 and 324 of nsP4, respectively. Analysis of these mutations in a Toto1101 background in a variety of ways, including RNA synthesis at 40°C and ability to complement other RNA^- mutants, demonstrates that these mutations are in fact those responsible for the *ts* phenotype and the mutant phenotypes previously described in the literature. *ts*118 turned out to be a double mutation. It has a defect in nsP2 (a Val to Ala change at position 425) which only partially disables it at 40°C, resulting in production of small plaques and reduced RNA synthesis although the plaque number is unchanged. A second mutation in

nsP4 combined with this change in nsP2 results in true temperature-sensitivity in that the plaque titer as well as plaque size is reduced at 40°C, and RNA synthesis after infection at 40°C is reduced to the level characteristic of RNA^- mutants. It is of interest that this change in nsP4 alone has little apparent phenotype, but because it is the one responsible for temperature-sensitivity leading to reduced plaque titer in combination with the nsP2 change, the double mutant *ts*118 complements as a group F mutant, with *ts*6 and *ts*110. Thus mutations in nsP4 lead to complementation as group F.

Experiments of Fuller and Marcus (8) in which the complementation groups were ordered by the relative rate of UV inactivation of their ability to complement gave a gene order for the complementation groups as NH_2-G-A-B/F-COOH. Mutants of B and F could not be precisely localized, but both occurred downstream of group G and group A. Thus, the UV mapping data are consistent with our results that group F mutants contain defects in nsP4. However, mapping of the other complementation groups of Sindbis *ts* RNA^- mutants, currently being carried out, indicates that the Fuller and Marcus (8) order is not correct (Y. S. Hahn *et al.*, in preparation).

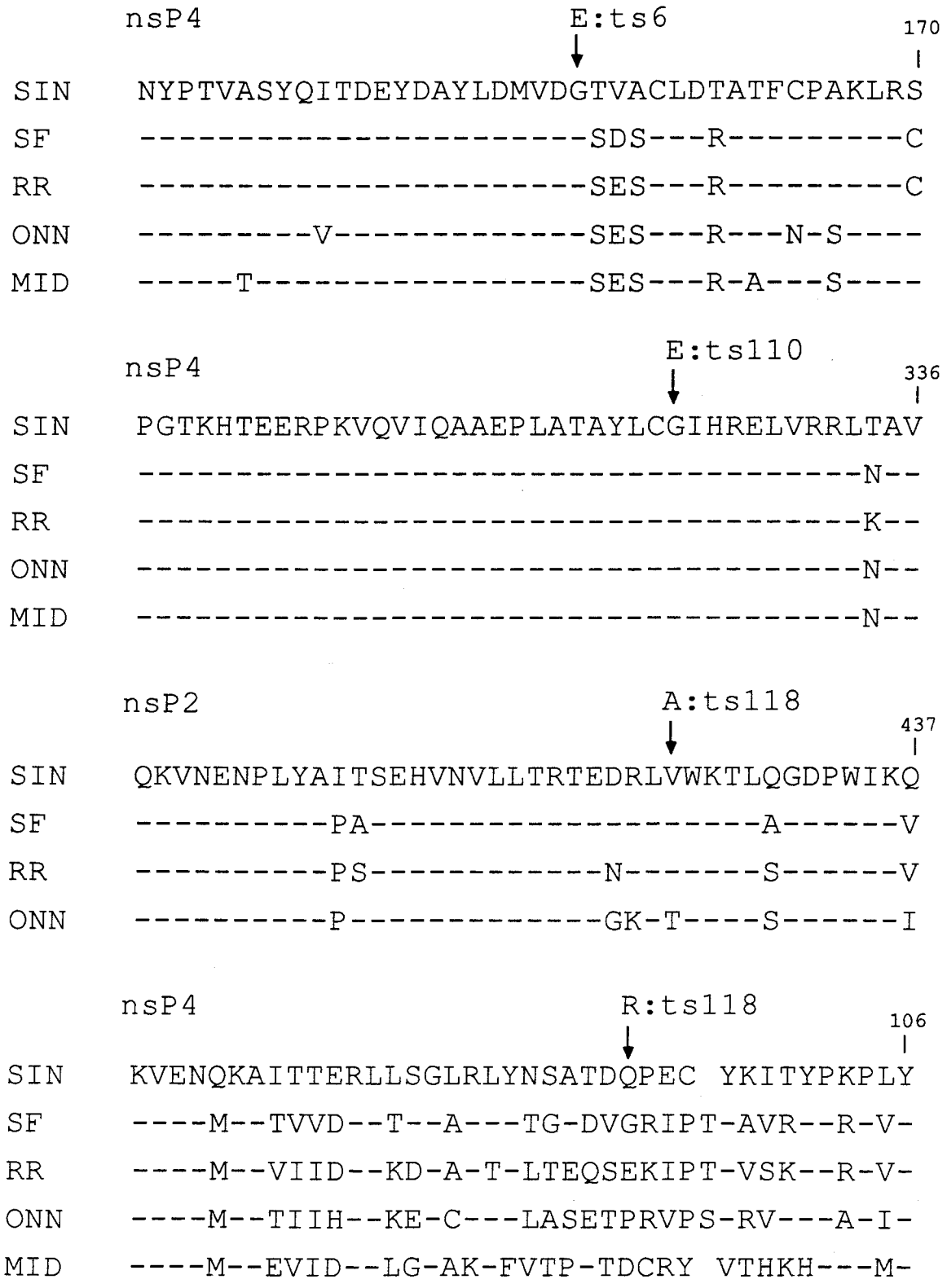
Keranen and Kaariainen (14) and Sawicki *et al.* (26) had demonstrated that *ts*6 ceased genomic, subgenomic, and minus strand RNA synthesis upon shift from permissive to nonpermissive temperature, and postulated that there was a temperature sensitive lesion in the elongation component of the replicase. Recently, Barton *et al.* (2) have extended such studies to *in vitro* studies of replication complexes isolated from *ts*6-infected cells, and concluded that such transcription complexes were temperature sensitive in elongation. In contrast other mutants examined retain activity at 40°C if replication complexes are formed first at 30°C (14). These results, together with the results presented here, suggest that nsP4 is an RNA polymerase or the elongation component of the

alphavirus replicase. In this regard it is of considerable interest that nsP4 contains the Gly-Asp-Asp sequence and surrounding hydrophobic amino acids that have been found to be present in the replicase proteins of several other RNA viruses (13, 22).

The mutations in *ts*6, *ts*110, and *ts*118 are shown in Fig. 3 in which sequences of up to five alphaviruses are compared in the regions affected. It is readily evident that the glycines affected in nsP4 of *ts*6 and *ts*110 are conserved in all alphaviruses sequenced to date, and that each is found within a long stretch of highly conserved amino acids. In view of this conservation in sequence, it is evident that changes in the sequence such as in *ts*6 and *ts*110 might affect function. The nsP4 polypeptide is composed of 610 amino acids and is on average 71 to 74% conserved among alphaviruses which have been examined (31). However, this conservation is not uniform throughout the protein; amino acids 1 to 125 from the N-terminus (including the location of the change in *ts*118) are less highly conserved, as are amino acids 550 to 604 near the C terminus. The canonical sequence Gly-Asp-Asp, flanked by hydrophobic amino acids, is found at residues 464 to 466. This sequence, which has been found in a number of RNA-dependent RNA replicases (13; see also discussion in ref. 22) is located well separated (on the linear sequence) from either the *ts*6 or *ts*110 lesion.

The change of Val→Ala in nsP2 of *ts*118 is found in a domain that is well conserved among alphaviruses, although conservation is not absolute. In particular, the Val affected in *ts*118 is not totally conserved, being replaced by Thr in O'Nyong-nyong virus. The Gln→Arg change in nsP4 of *ts*118 affects an amino acid in a domain that is not conserved (Fig. 3). The Gln is replaced by Gly, Glu, Pro, or Asp in different alphaviruses. Thus the failure of this change to have much effect by itself is perhaps not surprising. The pronounced effect of this substitution in combination with the change in nsP2 suggests that nsP2 and nsP4

Fig. 3. Amino acid sequences from five alphaviruses in the region near the Group F mutations. Sequence data are from the following sources: SIN=Sindbis virus, from Strauss *et al.* (33); SF=Semliki Forest virus, from Takkinen, (36); RR=Ross River virus, from Faragher *et al.* (7) and Strauss *et al.*, (31); ONN=O'Nyong-nyong virus from Strauss *et al.* (31) and unpublished data; MID=Middelburg virus, from Strauss *et al.* (32).


nsP4
E:ts6
170
SIN NYPTVASYQITDEYDAYLDMVDGTVACLDTATFCPAKLRS
SF ----------SDS---R---------C
RR ----------SES---R---------C
ONN ---------V----------SES---R---N-S----
MID -----T----------SES---R-A---S----
nsP4
E:ts110
336
SIN PGTKHTEERPKVQVIQAAEPLATAYLCGIHRELVRRLTAV
SF ----------N--
RR ----------K--
ONN ----------N--
MID ----------N--
nsP2
A:ts118
437
SIN QKVNENPLYAITSEHVNVLLTRTEDRLVWKTLQGDPWIKQ
SF ----------PA----------A------V
RR ----------PS----------N-------S------V
ONN ----------P----------GK-T----S------I
nsP4
R:ts118
106
SIN KVENQKAITTERLLSGLRLYNSATDQPEC YKITYPKPLY
SF ----M--TVVD--T--A---TG-DVGRIPT-AVR--R-V-
RR ----M--VIID--KD-A-T-LTEQSEKIPT-VSK--R-V-
ONN ----M--TIIH--KE-C---LASETPRVPS-RV---A-I-
MID ----M--EVID--LG-AK-FVTP-TDCRY VTHKH---M-

may interact to form a functional complex, although other explanations for such a synergistic effect are possible.

In Sindbis virus, translation of nsP4 requires readthrough of an opal termination codon so that nsP4 is underproduced relative to nsP1, nsP2, or nsP3 (11, 16, 32). The active form of nsP4 may be the polypeptide nsP34, which accumulates during Sindbis virus infection whereas little or no nsP4 is detected. The finding that nsP4 may be the viral RNA polymerase, based upon the results with complementation group F mutants and the presence of motifs within this protein that are shared with other RNA polymerases of animal viruses (13), is then reminiscent of the control of virus replicases in other systems in which readthrough of a termination codon is required to produce the polymerase. In tobacco mosaic viruses readthrough of an amber codon is required (9, 21), and the readthrough portion of the protein is homologous to the Sindbis protein (12). Similarly, in the case of the retroviruses, translation of the reverse transcriptase requires readthrough of an amber codon or frame shifting to eliminate an amber codon (25, 27, 28). We presume that regulation of the amount of RNA polymerase produced is important during a virus life cycle. As has been pointed out, however, Semliki Forest virus and O'Nyong-nyong virus lack this termination codon (31, 36), and regulation of the activity of the polymerase seems to be different for these alphaviruses.

Acknowledgments

The expert technical assistance of E. M. Lenches is gratefully acknowledged. We thank Dr. H. V. Huang and S. A. Chervitz for plasmid DNAs. This work was supported by Grants AI 10793, AI 20612, and AI 24134 from N.I.H. C. M. R. is a Pew scholar in the biomedical sciences.

REFERENCES

1. Arias, C., J. R. Bell, E. M. Lenches, E. G. Strauss, and J. H. Strauss. 1983. Sequence analysis of two mutants of Sindbis virus defective in the intracellular transport of their glycoproteins. J. Mol. Biol. **168:** 87-102.
2. Barton, D. J., S. G. Sawicki, and D. L. Sawicki. 1988. Demonstration *in vitro* of temperature-sensitive elongation of RNA in Sindbis virus mutant *ts*6. J. Virol. **62:**3597-3602.
3. Burge, B. W., and E. R. Pfefferkorn. 1966a. Isolation and characterization of conditional-lethal mutants of Sindbis virus. Virology **30:** 204-213.
4. Burge, B. W., and E. R. Pfefferkorn. 1966b. Complementation between temperature-sensitive mutants of Sindbis virus. Virology **30:** 214-223.
5. Burge, B. W. and E. R. Pfefferkorn. 1967. Temperature-sensitive mutants of Sindbis virus: Biochemical correlates of complementation. J. Virol. **1:**956-962.
6. Dulbecco, R. and M. Vogt. 1954. One step growth curve of Western equine encephalomyelitis virus on chick embryo cells grown in vitro and analysis of virus yields from single cells. J. Exptl. Med. **99:**183-189.
7. Faragher, S. G., A. D. J. Meek, C. M. Rice, and L. Dalgarno. 1988. Genome sequences of a mouse-avirulent and a mouse-virulent strain of Ross River virus. Virology **163:** 509-526.
8. Fuller, F. J., and P. I. Marcus. 1980. Sindbis virus. I. Gene order of translation in *in vivo*. Virology **107:** 441-451.
9. Goelet, P., G. P. Lomonssoff, P. J. G. Butler, M. E. Akam, M. J. Gait, and J. Karn. 1982. Nucleotide sequence of tobacco mosaic virus RNA. Proc. Natl. Acad. Sci. USA **79:** 5818-5822.

10. Hahn, C. S., E. G. Strauss, and J. H. Strauss. 1985. Sequence analysis of three Sindbis virus mutants temperature-sensitive in the capsid autoprotease. Proc. Natl. Acad. Sci. USA **82:** 4648-3652.

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

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

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

14. Keranen, S., and L. Kaariainen. 1979. Functional defects of RNA(-) *ts* mutants of Sindbis and Semliki Forest viruses. J. Virol. **32:** 19-29.

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

16. Lopez, S., J. R. Bell, E. G. Strauss, and J. H. Strauss. 1985. The nonstructural proteins of Sindbis virus as studied with an antibody specific for the C terminus of the nonstructural readthrough polyprotein. Virology **141:** 235-247.

17. Lustig, S., A. C. Jackson, C. S. Hahn, D. E. Griffin, E. G. Strauss, and J. H. Strauss. 1988. Molecular basis of Sindbis virus neurovirulence in mice. J. Virol. **62:** 2329-2336.

18. Lutz, C. T., W. M. Hollifield, B. Seed, J. M. Davie, and H. V. Huang. 1987. Syoinx 2A: An improved λ phage vector designed for screening DNA libraries by recombination *in vivo*. Proc. Natl. Acad. Sci. USA **84**, 4379-4383.

19. Maxam, A. M., W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods in Enzymology **65:** 499-560.

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

21. Pelham, H. R. B. 1978. Leaky UAG termination codon in tobacco virus RNA. Nature (London) **272:** 469-471.

22. Rice, C. M., E. G. Strauss, and J. H. Strauss. 1986. Structure of the flavivirus genome, pp. 35-90. *In* S. Schlesinger and M. Schlesinger (eds.), The Togaviridae and Flaviviridae. Chapter 10, Plenum Publishing Corp., New York.

23. Rice, C. M., and J. H. Strauss. 1981. Synthesis, cleavage and sequence analysis of DNA complementary to the 26S messenger RNA of Sindbis virus. J. Mol. Biol. **150:** 313-340.

24. Rice, C. M., R. Levis, J. H. Strauss, H. V. Huang. 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.

25. Rice, N. R., R. M. Stephens, A. Burny, and R. V. Gilden. 1985. The *gag* and *pol* genes of bovine leukemia virus: Nucleotide sequence and analysis. Virology **142:** 357-377.

26. Sawicki, D. L., S. G. Sawicki, S. Keranen, L. Kaariainen. 1981. Specific Sindbis virus coded function for minus strand RNA synthesis. J. Virol. **39:** 348-358.

27. Schwartz, D. E., R. Tizard, and W. Gilbert. 1983. Nucleotide sequence of Rous sarcoma virus. Cell **32:** 853-869.

28. Shinnick, T. M., R. A. Lerner, and J. G. Sutcliff. 1981. Nucleotide sequence of Moloney murine leukemia virus. Nature (London) **293:** 543-548.

29. Smith, D. R., and J. M. Calvo. 1980. Nucleotide sequence of the *E. coli* gene coding for dihydrofolate reductase. Nucl. Acids Res. **8:** 2255-2274.

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

31. Strauss, E. G., R. Levinson, C. M. Rice, J. Dalrymple, and J. H. Strauss. 1988. Nonstructural proteins nsP3 and nsP4 of Ross River and O'Nyong-nyong viruses: Sequence and comparison with those of other alphaviruses. Virology **164:** 265-274.

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

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

34. Strauss, E. G., and J. H. Strauss. 1980. Mutants of Alphaviruses: Genetics and Physiology, pp. 393-426. *In* R. W. Schlesinger (ed.), The Togaviruses. Academic Press, New York.

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

36. Takkinen, K. 1986. Complete nucleotide sequence of the nonstructural protein genes of Semliki Forest virus. Nucl. Acids Res. **14:** 5667-5682.

37. White, B. A., and F. C. Bancroft. 1982. Cytoplasmic dot hybridization. Simple analysis of relative mRNA levels in multiple small cell or tissue samples. J. Biol. Chem. **257:** 8569-8572.

Chapter 2

Mapping of RNA$^-$ Temperature-Sensitive Mutants of Sindbis Virus: Assignment of Complementation Groups A, B, and G to Nonstructural Proteins

This chapter was prepared for submission.

Mapping of RNA$^-$ Temperature-Sensitive Mutants of Sindbis Virus: Assignment of Complementation Groups A, B, and G to Nonstructural Proteins

Young S. Hahn, Ellen G. Strauss and James H. Strauss

Division of Biology, California Institute of Technology

Pasadena, CA 91125

Running title: Mapping of RNA$^-$ *ts* mutants of Sindbis

ABSTRACT

Four complementation groups of temperature-sensitive mutants of Sindbis virus are known that fail to make RNA at the nonpermissive temperature, and we have previously shown that group F mutants have defects in nsP4. Here we map representatives of groups A, B and G. Restriction fragments from a full-length clone of Sindbis virus, Toto1101, were replaced with the corresponding fragments from the various mutants. These hybrid plasmids were transcribed *in vitro* by SP6 RNA polymerase to produce infectious RNA transcripts and the virus recovered was tested for temperature sensitivity. After mapping each lesion to a specific region, cDNA clones of both mutants and revertants were sequenced in order to determine the precise nucleotide change responsible for each mutation. Synthesis of viral RNA and complementation by rescued mutants were also examined in order to study the phenotype of each mutation in a uniform genetic background. The single mutant of group B, *ts*11, has a defect in nsP1, Ala-348 to Thr. All of the group A and group G mutants examined had lesions in nsP2, Ala-517 to Thr in *ts*17, Cys-304 to Tyr in *ts*21, and Gly-736 to Ser in *ts*24 for three group A mutants, and for two group G mutants Phe-509 to Leu in *ts*18 and Asp-522 to Asn in *ts*7. In addition *ts*7 has a change in nsP3, Phe-312 to Ser, which also renders the virus temperature- sensitive and RNA^-. Thus changes in any of the four nonstructural proteins can lead to failure to synthesize RNA at a nonpermissive temperature, indicating that all four are involved in RNA synthesis. From the results presented here and from previous results several of the activities of the nonstructural proteins can be deduced. It appears that nsP1 may be involved in initation of minus-strand RNA synthesis. nsP2 appears to be involved in the initiation of 26S RNA synthesis, and in addition to be a protease that cleaves the nonstructural

polyprotein precursors. It may also be involved in shut off of minus-strand RNA synthesis. nsP4 appears to function as the viral polymerase or elongation factor. The functions of nsP3 are as yet unresolved.

INTRODUCTION

Sindbis virus is an alphavirus whose genome is a single-stranded RNA of 11,703 nucleotides of plus polarity (32). During replication the parental 49S plus-strand RNA is transcribed into a complementary minus-strand RNA which serves as a template for the synthesis of both genome-length RNA and 26S subgenomic mRNA (reviewed in 34). The synthesis of both plus- and minus-strand RNA increases during the first 3 h post-infection, but at 3 to 3.5 h after infection the synthesis of minus-strand ceases, whereas the synthesis of both 49S and 26S plus-strands continues (4, 22). In this early phase plus-strand RNA is produced in about 5-fold molar excess over minus-strand RNA. Replication of viral RNA is presumed to require the activities of four nonstructural proteins called nsP1, nsP2, nsP3 and nsP4, which are translated from the genomic 49S RNA as two polyprotein precursors that are subsequently processed by posttranslational cleavage (14). It has been postulated that the processing of the nonstructural polypeptides is catalyzed by a virus-encoded protease located within one of the nonstructural polypeptides (reviewed in 35).

Temperature-sensitive mutants of Sindbis virus that fail to synthesize RNA at a nonpermissive temperature (called RNA^-), some of which are defective in processing of the nonstructural polypeptides, have been isolated and grouped by complementation into four RNA^- groups, A, B, F and G (5, 6, 29, 33). These mutants have been presumed to contain *ts* lesions in the viral nonstructural proteins active in viral RNA synthesis. Some of these RNA^- mutants have been characterized in depth (17, 23, 24, 26). Group A mutants were divided by Sawicki

and Sawicki (24) into two phenotypic subgroups, where subgroup 1 mutants (*ts*15, *ts*17, *ts*21, *ts*24 and *ts*133) were temperature-sensitive for synthesis of 26S mRNA, whereas subgroup 2 mutants (*ts*4, *ts*14, *ts*16, *ts*19 and *ts*138) were not. In addition, *ts*17, *ts*24 and *ts*133 failed to cleave the polyprotein precursor nsP123 and failed to shut off minus-strand synthesis upon shift to 40°C. One member of group G, *ts*18, was also found to be temperature-sensitive in synthesis of 26S mRNA and as well in the cleavage of nsP123, whereas a second mutant of group G, *ts*7, demonstrated neither of these defects. These studies suggested that groups A and G define functions involved in both initiation of 26S RNA synthesis and in cleavage of polyprotein precursors, and that the A group mutants might also define a function involved in the shut off of minus-strand synthesis.

The single mutant in group B, *ts*11, has a defect in minus-strand RNA synthesis. Upon shift from a permissive to a nonpermissive temperature, synthesis of plus-stranded RNA continued, whereas minus-strand RNA synthesis ceased (23). This led to the hypothesis that group B mutants are involved in the initiation of minus-strand synthesis. Finally, mutant *ts*6 (complementation group F) was found to cease all RNA synthesis upon shift to a nonpermissive temperature, leading to the hypothesis that group F mutants are in the elongation component of the viral replicase (3, 23).

We have previously shown that three group F mutants contain lesions in nsP4, suggesting that it is the replicase (13). Here we map representatives of RNA^- groups A, B and G using the same approach. Groups A and G have been found to map in nsP2, suggesting that this protein is required for initiation of 26S RNA synthesis, proteolysis of the nonstructural precursor, and shut off of minus-strand synthesis. Together with complementation results presented here, it appears that complementation among A and G mutants is intracistronic, and group G should be considered a subgroup of the A group. *ts*11 of group B has been found

to have a lesion in nsP1, which may therefore be required for initiation of minus-strand synthesis.

MATERIALS AND METHODS

Virus growth and purification. Mutants used were obtained originally from Dr. B. Burge and had been isolated from the HR strain of Sindbis virus. *ts*17, *ts*21, *ts*24, *ts*11 and *ts*7 were obtained from a stock mutagenized with nitrosoguanidine, whereas *ts*18 was isolated after mutagenesis with nitrous acid (5). Viruses were grown in primary or secondary chick embryo cells and harvested 10-20 h after infection at 30°C. Revertants of each mutant were isolated by plaquing mutant stocks at 40°C. A single virus plaque was picked from the plate and the virus eluted into 1 ml of Eagle medium containing 5% fetal calf serum. This plaque pick was used to infect primary chick embryo 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. Viral RNA was isolated as previously described (20).

Construction of hybrid genomes. cDNA clones of the nonstructural coding region of mutants and revertants were obtained as previously described (13). Hybrid genomes were produced by replacing restriction fragments in a full-length clone of Sindbis virus, Toto1101 (21), with the corresponding regions from the mutant or revertant cDNA clones (13). Details of the restriction sites used are included in the figure legends. Plasmids containing the large intervals A, B and C from the mutants, as well as subintervals B1, B2, and B3, were constructed as previously described (13). To obtain recombinant plasmids containing subregions A1 and A2 from the mutants, the restriction fragment SspI (nt504) to PstI (nt1507) or fragment PstI (nt1507) to ClaI (nt2713), respectively, of the *ts* mutant was cloned into the shuttle vector πnsP12, which contains the SacI (nt13552) to EcoRV

(nt2750) region of Toto1101 cloned into πAN7 (13). The SacI (nt13552) to ClaI (nt2713) fragment of the resulting plasmid was then used to replace the corresponding fragment in Toto1101.

Subintervals B4 and B5 were constructed for *ts7* to separate two changes found in subinterval B3. Toto:*ts7*B4 contained the AvaI (nt3546)-SalI (nt4845) fragment of *ts7* in Toto1101, and Toto:*ts7*B5 contained the SalI (nt4845)-SpeI (nt5262) fragment of *ts7* in Toto1101; construction of these two plasmids utilized the shuttle vector Kahn 5B (see 13).

***In vitro* transcription and transfection.** RNA transcripts were synthesized by transcribing plasmids *in vitro* with SP6 RNA polymerase (21). The transcribed RNA was assayed for the production of *ts* virus by transfecting confluent monolayers of secondary chicken cells in 35 mm multi-well tissue culture plates (13).

Viral RNA synthesis. RNA synthesis following infection by the various mutants or revertants was assayed at 30°C, at 40°C, or at 40°C following a shift from 30°C by using the cytoplasmic dot hybridization method of White and Bancroft (38) as previously described (13).

RESULTS

Construction of recombinant plasmids. In order to localize the *ts* lesions of three group A mutants (*ts17*, *ts21* and *ts24*), of the sole group B mutant (*ts11*), and of two group G mutants (*ts7* and *ts18*), we have constructed a number of recombinant plasmids. These recombinant plasmids are derivatives of Toto1101, a full-length cDNA clone of Sindbis virus, from which infectious RNA can be transcribed *in vitro* with SP6 RNA polymerase (21). Restriction fragments in Toto1101 were replaced with the corresponding fragments from cDNA clones of

the mutants and the resulting constructs are illustrated schematically in Fig. 1. In this figure the restriction sites used to construct the hybrid genomes are indicated together with their numbering from the 5' end of the RNA (32). For gross mapping the region of the genome encoding the nonstructural proteins was first divided into three large intervals, A, B and C, as shown. For finer mapping the interval A was subdivided into two subregions, A1 and A2, and region B was subdivided into three overlapping subregions, B1, B2 and B3, as shown. For *ts*7, two additional subregions, B4 and B5, were constructed and tested as described below.

RNA was transcribed from the recombinant plasmids *in vitro* with SP6 RNA polymerase and transfected onto monolayers of chicken cells. Monolayers were incubated under agarose at 30°C or 40°C to determine whether or not the virus recovered was *ts*. The names used to refer to the recombinant plasmids or to the virus derived from them are also indicated in Fig. 1.

Localization of the lesions in group A mutants. The results obtained for the group A recombinant viruses tested are summarized in Table 1. For *ts*17, among the three large interval replacement clones (Toto:*ts*17A, Toto:*ts*17B, and Toto:*ts*17C), plasmids Toto:*ts*17A and Toto:*ts*17C gave rise to virus that exhibited wild-type growth at the nonpermissive temperature, while *ts* virus was obtained from plasmid Toto:*ts*17B. This localized the *ts* mutation to the interval 2713-5262 of the genome. Plasmids Toto:*ts*17B1, Toto:*ts*17B2, and Toto:*ts*17B3 containing three smaller intervals in the B region were then constructed and tested. Plasmid Toto:*ts*17B1 gave rise to temperature-sensitive virus, whereas the other two plasmids did not. Thus *ts*17 has one or more mutations in the region encoding nsP2, between ClaI (nt2713) and AvaI (nt3546).

Using a similar approach the mutation in *ts*21 was localized to the region between nucleotides 1507 and 2713 of the genome represented in plasmid Toto:*ts*21A2. This region includes the C-terminus of nsP1 and the N-terminal half of nsP2.

Fig. 1. Construction of hybrid genomes. The top line is a schematic map of the nonstructural protein coding region of Sindbis virus cDNA clone Toto1101. The translated region of the genome is shown as an open box and nontranslated regions as single lines. Within the box, the boundaries of the nonstructural proteins are indicated, as are the locations of a number of restriction sites used to construct the hybrid genomes, numbered from the 5' end (32). In the lines below, the dashed boxes show the location of restriction fragments in clone Toto1101 that were replaced with the corresponding restriction fragments from the *ts* mutants or their revertants. The names used for these clones are shown at the right, e.g., Toto:*ts*11A2 is a construct in which the PstI (1507) to ClaI (2713) fragment comes from a cDNA clone of mutant *ts*11.

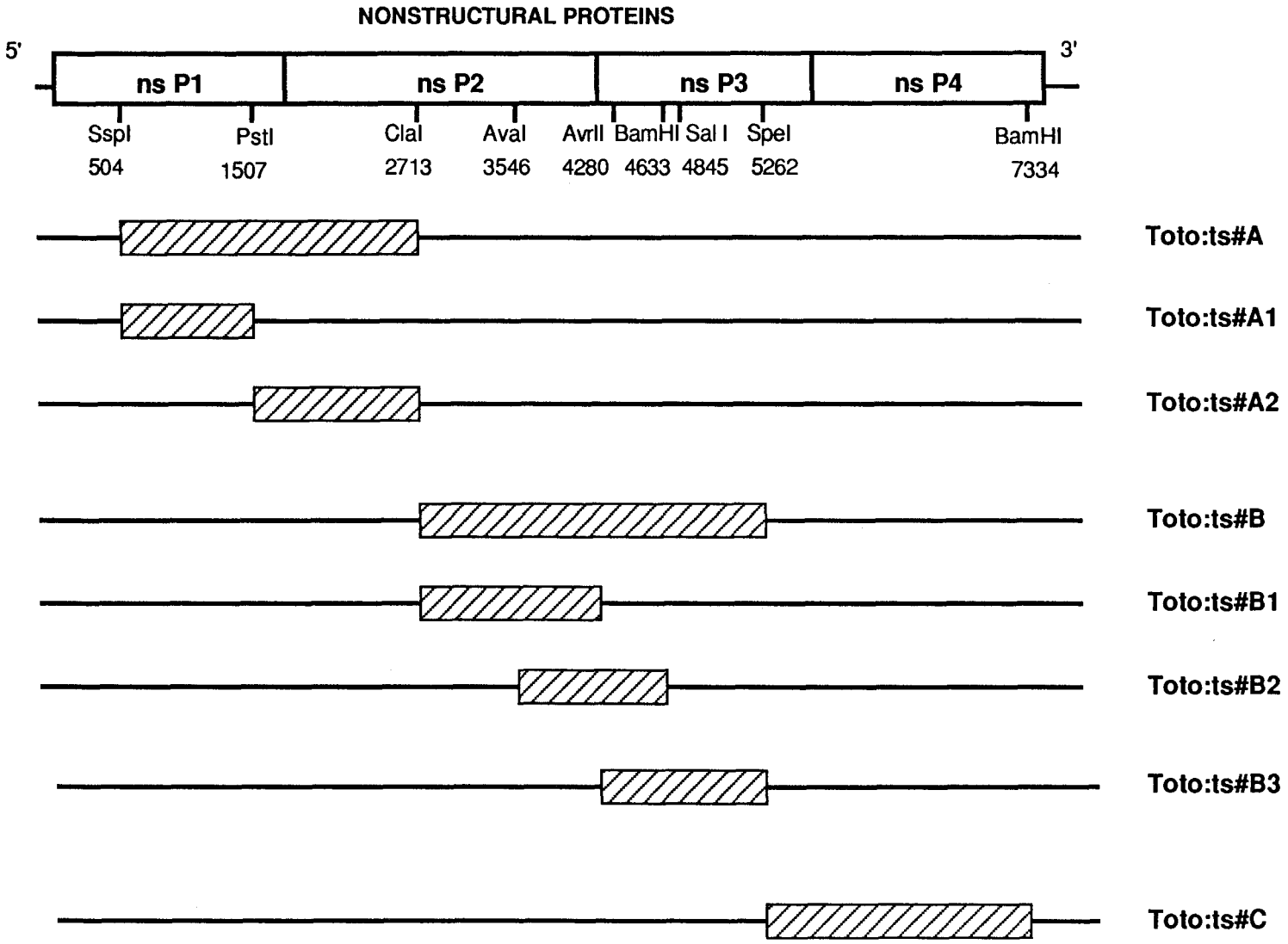

NONSTRUCTURAL PROTEINS
5'
3'
ns P1
ns P2
ns P3
ns P4
SspI
504
PstI
1507
ClaI
2713
AvaI
3546
AvrII
4280
BamHI
4633
Sal I
4845
SpeI
5262
BamHI
7334
Toto:ts#A
Toto:ts#A1
Toto:ts#A2
Toto:ts#B
Toto:ts#B1
Toto:ts#B2
Toto:ts#B3
Toto:ts#C

Table 1. Summary of the constructs of group A mutant tested for ts phenotype at the nonpermissive temperature

Recombinant clone	Replaced fragment	Phenotype	Mutant localization
Toto:ts17A	nt 504-2713	wt	nt 2713-3546
Toto:ts17B	nt 2713-5262	ts	
Toto:ts17B1	nt 2713-4280	ts	
Toto:ts17B2	nt 3546-4633	wt	
Toto:ts17B3	nt 4280-5262	wt	
Toto:ts17C	nt 5262-7334	wt	
Toto:ts21A	nt 504-2713	ts	nt 1507-2713
Toto:ts21A1	nt 504-1507	wt	
Toto:ts21A2	nt 1507-2713	ts	
Toto:ts21B	nt 2713-5262	wt	
Toto:ts21C	nt 5262-7334	wt	
Toto:ts24A	nt 504-2713	wt	nt 3546-4280
Toto:ts24B	nt 2713-5262	ts	
Toto:ts24B1	nt 2713-4280	ts	
Toto:ts24B2	nt 3546-4633	ts	
Toto:ts24B3	nt 4280-5262	wt	
Toto:ts24C	nt 5262-7334	wt	

The lesion in *ts24* was mapped to the B region. Testing of the subregions then showed that virus from Toto:*ts*24B1 and Toto:*ts*24B2 was temperature-sensitive. Assuming that only a single nucleotide change is involved, the mutation in *ts24* lies in the overlap between these two subregions, that is, between nucleotides 3546 and 4280 of the genome. Once again the region includes mostly nsP2 sequences, with the N-terminus of nsP3 also present.

Localization of the mutation in *ts*11. The results obtained with the *ts*11 constructs tested are summarized in Table 2. By the same procedure as before the *ts*11 lesion was localized to the region between nucleotides 504 and 1507 of the RNA genome, and is therefore found in nsP1.

Localization of the lesions in group G mutants. The results for *ts7* and *ts*18 are also shown in Table 2. For *ts*18 the mapping was straightforward and the *ts* lesion was localized to the region between 2713 and 3546 of the genome (and is therefore in nsP2).

The results with *ts7* were more complex, as *ts7* is a multiple mutant with two lesions contributing to temperature sensitivity. Of the large interval plasmids tested only Toto:*ts*7B gave rise to *ts* virus. However, upon testing the subregions, both plasmids Toto:*ts*7B1 and Toto:*ts*7B3 gave rise to *ts* virus, whereas plasmid Toto:*ts*7B2 gave rise to wild-type virus. Thus we conclude that *ts7* has a *ts* lesion between nucleotides 2713 and 3546 of the genome and a second lesion between nucleotides 4633 and 5262. The first region lies within nsP2, whereas the second lesion lies within nsP3.

Sequence analysis of *ts* mutants and of their revertants. In order to precisely define the *ts* lesions of the six mutants being examined, regions of cDNA clones from each mutant that had been shown by the mapping experiments to contain the *ts* lesions were sequenced by the chemical method (19, 27). In each

Table 2. Summary of the constructs of group B and G mutants tested for ts phenotype at the nonpermissive temperatures

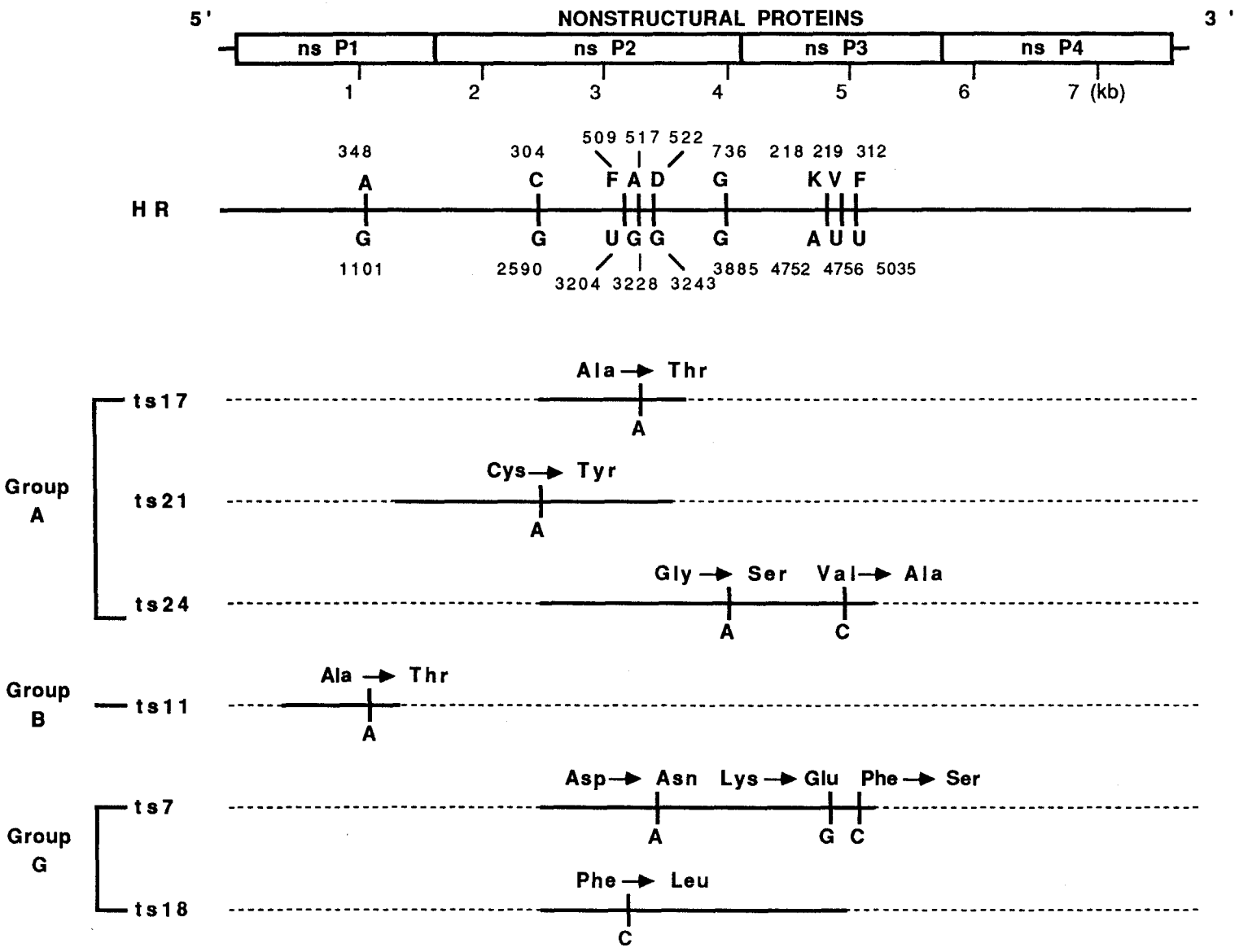
Recombinant clone	Replaced fragment	Phenotype	Mutant localization
Toto:ts11A	nt 504-2713	ts	nt 504-1507
Toto:ts11A1	nt 504-1507	ts	
Toto:ts11A2	nt 1507-2713	wt	
Toto:ts11B	nt 2713-5262	wt	
Toto:ts11C	nt 5262-7334	wt	
Toto:ts7A	nt 504-2713	wt	nt 2713-3546 nt 4633-5262
Toto:ts7B	nt 2713-5262	ts	
Toto:ts7B1	nt 2713-4280	ts	
Toto:ts7B2	nt 3546-4633	wt	
Toto:ts7B3	nt 4280-5262	ts	
Toto:ts7C	nt 5262-7334	wt	
Toto:ts18A	nt 504-2713	wt	nt 2713-3546
Toto:ts18B	nt 2713-5262	ts	
Toto:ts18B1	nt 2713-4280	ts	
Toto:ts18B2	nt 3546-4633	wt	
Toto:ts18B3	nt 4280-5262	wt	
Toto:ts18C	nt 5262-7334	wt	

case the corresponding region of a revertant clone was also sequenced. The results obtained are shown in Fig. 2.

Mutant *ts*17 has a single base substitution in the region sequenced. Comparing the *ts*17 sequence with that of HR and with that of a revertant we found the mutation responsible for temperature sensitivity to be a change of G to A at nucleotide 3228 which results in the replacement of Ala (GCC) at position 517 of nsP2 by Thr (ACC). In the revertant the altered nucleotide reverts to the parental nucleotide, restoring the parental amino acid. *ts*21 also has a single change in the sequenced region, a G to A substitution at nucleotide 2590 of the genome which results in replacement of Cys (UGC) by Tyr (UAC) at position 304 of nsP2. In the *ts*21 revertant sequenced this nucleotide reverts to the original nucleotide.

The *ts* lesion in *ts*24 was found to lie between nucleotides 3546 and 4280 (Table 1). In this region only a single nucleotide substitution was found, a G to A at 3885 which results in the substitution of Gly (GGC) by Ser (AGC) at position 736 of nsP2. In a revertant of *ts*24 sequenced this nucleotide reverted to the parental nucleotide. In part because of the unusual reversion frequency of *ts*24 (see below), the entire region encompassed by interval B was sequenced to rule out the possibility of a second mutation that might result in temperature sensitivity. During this sequencing we found a second nucleotide substitution, a change of U to C at nucleotide 4756 which results in a change of Val (GUC) by Ala (GCC) at position 219 of nsP3 (Fig. 2). This change, which would be present in construct Toto:*ts*24B3, does not lead to temperature sensitivity, however (Table 1). Because the *ts*24 revertant sequenced had the parental nucleotide at this position, we suspect that this substitution resulted from cloning a minor variant in the *ts*24 population. In any event, because it does not give rise to temperature sensitivity, it has not been further studied.

Fig. 2. Localization of the *ts* lesions of six mutants. A schematic of the nonstructural coding region of Sindbis virus is shown. Below are shown sequencing schematics for HR Sindbis virus (32), the parental strain from which the *ts* mutants were isolated (5), and for mutants *ts*17, *ts*21, *ts*24, *ts*11, *ts*7, and *ts*18. Any change from the HR sequence is indicated. Nucleotides, shown below the lines, are numbered from the 5' end of the RNA, and amino acids, shown above the lines, from the N-terminus of each protein (amino acids are indicated in the single letter code for the HR strain in order to reduce crowding in the figure). For each mutant only the region shown by the solid line was sequenced.


5'
NONSTRUCTURAL PROTEINS
3'
ns P1
ns P2
ns P3
ns P4
1
2
3
4
5
6
7 (kb)
HR
348
304
509 517 522
736
218 219 312
A
C
F A D
G
K V F
G
G
U G G
G
A U U
1101
2590
3204 3228 3243
3885 4752 4756 5035
Group A
ts17
Ala → Thr
A
ts21
Cys → Tyr
A
ts24
Gly → Ser
Val → Ala
A
C
Group B
ts11
Ala → Thr
A
Group G
ts7
Asp → Asn
Lys → Glu
Phe → Ser
A
G C
ts18
Phe → Leu
C

The mutation responsible for temperature sensitivity in *ts*11 was found to be a change of G to A at nucleotide 1101 which leads to the replacement of Ala (GCC) by Thr (ACC) at position 348 of nsP1. In a revertant sequenced, the changed nucleotide had reverted to the original nucleotide, restoring the parental amino acid.

*ts*7 is a double mutant as described earlier and the situation with this mutant is complex. Three base substitutions were found in the two regions previously shown to be responsible for temperature sensitivity (Table 2). In the nsP2 region there is a change of G to A at nucleotide 3243 resulting in a change of Asp (GAC) to Asn (AAC) at position 522. In the nsP3 region two changes were found, a change of A to G at position 4752 resulting in a Lys (AAG) to Glu (GAG) substitution at position 218, and a U to C change at nucleotide 5035 resulting in a Phe (UUU) to Ser (UCU) substitution at amino acid 312. As will be described in more detail below, only the Asp to Asn change in nsP2 and the Phe to Ser change at position 312 of nsP3 result in temperature sensitivity. A partial revertant of *ts*7 was isolated and sequenced and found to retain all three amino acid changes, and thus appeared to be a second site pseudorevertant (see below).

The change in *ts*18 responsible for temperature sensitivity is a U to C substitution at nucleotide 3204 resulting in the change of Phe (UUU) to Leu (CUU) at position 509 of nsP2 (Figure 2). A revertant of *ts*18 was sequenced and found to revert to the original nucleotide, restoring the parental amino acid.

Characterization of rescued mutations. In order to establish that the mutations mapped are the ones responsible for the phenotypes previously described for these mutants and in order to characterize the phenotypes of these mutations in a uniform background, recombinant viruses containing a defined

region from each of the *ts* mutants in a Toto1101 background were studied. Monolayers were transfected with RNA transcribed from the recombinant plasmids described earlier and a single plaque of each virus was isolated from the 30°C plate and used to obtain a stock of the rescued mutant.

These stocks were titered at 30°C and at 40°C and the results are shown in Table 3. The virus stocks derived from the infectious transcripts of all but *ts*7 showed approximately the same apparent reversion frequency as did the parental viruses. (In the case of *ts*21 and *ts*18 the apparent reversion frequency of the parental stock is somewhat higher than that of the stocks derived from the infectious transcripts because the parental stocks had not been as recently plaque purified; revertants are amplified in most stocks of *ts* mutants during passage because they have a selective advantage even at the permissive temperature). For *ts*17, *ts*21, *ts*11 and *ts*18 this apparent reversion frequency was between 10^{-3} and 10^{-5}, consistent with the temperature sensitivity arising from a single nucleotide change which reverts with a frequency on the order of 10^{-4} (16, 28, 36). The results with *ts*24, however, demonstrate a much lower reversion frequency. It is clear that the virus derived from the construct Toto:*ts*24B1 differs from Toto1101 by a single nucleotide change (Fig. 2) and yet the reversion frequency is very low, less than 10^{-8} as is the case for the parental *ts*24. Thus it seems clear that in some cases a single nucleotide change can revert to the parental nucleotide very infrequently and that although in general the rate of nucleotide substitution in RNA virus genomes is quite high, there are nucleotide substitutions that in certain contexts revert very infrequently. Another example of a specific mutation in the Sindbis virus genome with very low reversion frequencies has been described by Durbin and Stollar (9). The results in Table 3 also demonstrate that the virus derived from construct Toto:*ts*24B3 which

Table 3. Efficiency of plaque formation and viral RNA synthesis at 40°C and 30°C

Virus	Titer (PFU/ml)		E.O.P.	RNA synthesis		
	40°C	30°C	(40°C /30°C)	40°C	30°C	shift to 40°C
Toto1101	5.2×10^{8}	2.0×10^{9}	2.6×10^{-1}	1.00	1.00	1.00
ts17	4.0×10^{4}	3.0×10^{9}	1.3×10^{-5}	0.04	1.14	0.19
Toto:ts17B1	2.0×10^{5}	2.0×10^{9}	8.0×10^{-4}	0.03	1.25	0.27
ts21	4.8×10^{6}	4.9×10^{9}	1.0×10^{-3}	0.03	0.84	0.32
Toto:ts21A2	1.2×10^{5}	1.2×10^{9}	1.0×10^{-4}	0.03	0.87	0.40
ts24	<10	1.2×10^{9}	<10 $^{-8}$	0.03	0.99	0.38
Toto:ts24B	<10	8.0×10^{8}	<10 $^{-8}$	0.07	0.99	0.79
Toto:ts24B1	<10	1.5×10^{9}	<10 $^{-8}$	0.12	1.18	0.89
Toto:ts24B3	2.4×10^{9}	2.7×10^{9}	8.2×10^{-1}	1.16	1.16	1.14
ts24R	7.0×10^{8}	2.6×10^{9}	2.7×10^{-1}	1.11	1.14	1.18
ts11	4.0×10^{5}	9.0×10^{8}	4.4×10^{-4}	0.03	0.73	0.47
Toto:ts11A1	4.4×10^{5}	1.5×10^{9}	2.9×10^{-4}	0.04	0.94	0.55
ts7	<10	2.0×10^{9}	<10 $^{-8}$	0.02	1.16	0.53
Toto:ts7B	<10	5.8×10^{9}	<10 $^{-8}$	0.02	1.12	0.70
Toto:ts7B1	8.2×10^{6}	1.0×10^{9}	8.2×10^{-3}	0.38	1.10	0.80
Toto:ts7B3	2.0×10^{6}	1.7×10^{9}	1.2×10^{-3}	0.05	1.16	0.71
ts7R	3.0×10^{6}	7.2×10^{8}	4.2×10^{-3}	0.50	1.25	1.20
ts18	4.0×10^{5}	1.6×10^{9}	1.6×10^{-4}	0.02	1.30	0.37
Toto:ts18B1	2.4×10^{4}	2.7×10^{9}	2.7×10^{-5}	0.05	1.52	0.56

contains the Val to Ala substitution in nsP3 is temperature insensitive with respect to plaquing efficiency at 40°C or RNA synthesis at 40°C (see below).

Because of the very low reversion frequency of *ts24* it was difficult to obtain a revertant, but one was obtained by passing the virus stock at 40°C in liquid culture for three passages, followed by plaque assay at 40°C. Because of the tendency of revertants, once they arise, to accumulate in the virus population, this method led to the isolation of a revertant which was temperature insensitive for plaque assay or for RNA synthesis.

The results with *ts7* are consistent with it being a double mutant. The parental virus, *ts7*, and the virus derived from construct Toto:*ts7*B demonstrated a very low reversion frequency, less than 10^{-8}. Virus derived from the individual constructs Toto:*ts7*B1 or Toto:*ts7*B3 which contain a single change in nsP2 or two changes in nsP3, respectively, showed a much higher apparent reversion frequency, on the order of 10^{-2} to 10^{-3}, consistent with the temperature sensitivity of these viruses being due to a single nucleotide alteration (although this high apparent reversion frequency suggests that there may be some leakage). Note that both of these virus constructs are in fact temperature-sensitive for plaque formation at 40°C. As shown in Table 3 and as will be described in more detail below, the change(s) in nsP3 (i.e., Toto:*ts7*B3) leads to temperature sensitivity in RNA synthesis, that is, an RNA^- phenotype, whereas the change in nsP2 results in an $RNA^{\pm}$ phenotype.

In order to separate the two changes in nsP3 of *ts7*, we took advantage of a SalI site at nt4845 that is located between these two changes. Construct Toto:*ts7*B4 containing the Lys to Glu change at position 218 gave rise to virus with wild type properties, plaquing efficiently and making wild type levels of RNA at 40°C (data not shown). Construct Toto:*ts7*B5 containing the Phe to Ser change at

residue 312 gave rise to *ts* virus that proved to be RNA^-, and thus this substitution alone in nsP3 is responsible for the *ts* phenotype.

A partial revertant of *ts*7, *ts*7R, was isolated that plaqued with an efficiency of 4×10^{-3} at 40°C. This virus was $RNA^{\pm}$ and made RNA at wild type levels after shift to 40°C (Table 3; see also below). As described above, *ts*7R was a pseudorevertant, containing all three amino acid changes found in *ts*7. Because of the unusual nature of this result, the sequence data were confirmed by preparing a new library of *ts*7R cDNA clones for sequence analysis, and the same three amino acid substitutions were found in this library. The position of the suppressing change has not been mapped.

RNA synthesis by the rescued mutants. To examine whether the rescued mutants were as defective in RNA synthesis as the parental viruses, RNA synthesis was analyzed after infection at 40°C, after infection at 30°C, or at 40°C following a shift from 30°C. Virus from clone Toto1101, from the parental mutants, and from the rescued mutants were compared. Total viral RNA synthesis was analyzed by a cytoplasmic dot hybridization method as described (13, 38) using a ^{32}P-labeled probe consisting of minus-strand RNA transcribed from the structural region of the virus. The values determined relative to RNA synthesis by virus recovered from Toto1101 are shown in Table 3.

In the case of *ts*17, *ts*21, *ts*11 and *ts*18, the virus recovered from the constructs is RNA^- as was the parental virus. In the case of *ts*24 the virus recovered from Toto:*ts*24B1 is also RNA^- but does show a slightly elevated level of RNA synthesis at 40°C relative to the *ts*24 parent, and there may be other as yet unmapped mutations in the *ts*24 genome that lead to a decreased level of RNA synthesis at 40°C (see also below).

In the case of virus derived from Toto:*ts*7B3, which contains the defect in nsP3, the virus is RNA^- as is the parental *ts*7. Virus derived from Toto:*ts*7B1

containing the defect in nsP2 is RNA$^{\pm}$, showing substantial RNA synthesis at 40°C.

These different virus strains were also tested for RNA synthesis after establishing infection at 30°C and shifting to 40°C. Under these conditions all of the mutants studied make substantial amounts of RNA following shift, from about 20% as much as Toto1101 virus in the case of *ts*17 to equivalent amounts in the case of some of the other constructs. In the case of *ts*17, *ts*21, *ts*11 and *ts*18, the recovered virus demonstrated the same level of RNA synthesis following shift as did the parental virus. In the case of *ts*24, however, virus derived from Toto:*ts*24B or Toto:*ts*24B1 exhibited virtually full RNA synthesis following shift, whereas that derived from the parental virus was about half as much, suggesting once again that the parental *ts*24 contains one or more unmapped mutations that depress RNA synthesis. Note that RNA synthesis following shift with Toto:*ts*24B or Toto:*ts*24B1 are about equivalent, suggesting that the change in the nsP3 region has little effect upon RNA synthesis.

In the case of *ts*7, the viruses recovered from the various constructs all show a substantial level of RNA synthesis following shift.

Complementation analysis of the rescued mutants. We examined the ability of the viruses recovered from these constructs to complement representative *ts* mutants from the other complementation groups of RNA^{-} mutants (Table 4). Complementation values obtained are defined as the yield following mixed infection at the nonpermissive temperature divided by the sum of the yields following infection with each mutant alone under the same conditions. The absolute magnitude of the complementation index thus depends upon the extent of leakage of the parental viruses, as complementation is inefficient, never exceeding 1-10% of a wild-type yield (6, 7, 29). Virus from the three A group mutants demonstrated ready complementation with representatives of the B group (*ts*11) and the F group (*ts*6) but showed only poor or marginal complementation

Table 4. Complementation between Sindbis ts RNA⁻ Mutants

Virus		A	B	F	G	Parental viruses
		ts 24	ts 11	ts 6	ts 18	
A	Toto:ts17B1	ND[a]	191	195	4	1
	Toto:ts21A2	ND	20	11	2	1
	Toto:ts24B1	<1	7	10	2	<1
B	Toto:ts11A1	105	1	120	118	1
C	Toto:ts7B	<1	19	35	ND	<1
	Toto:ts7B1	<1	3	2	ND	1
	Toto:ts7B3	<1	4	18	ND	1
	Toto:ts18B1	2	55	125	<1	<1

[a] ND : not determined

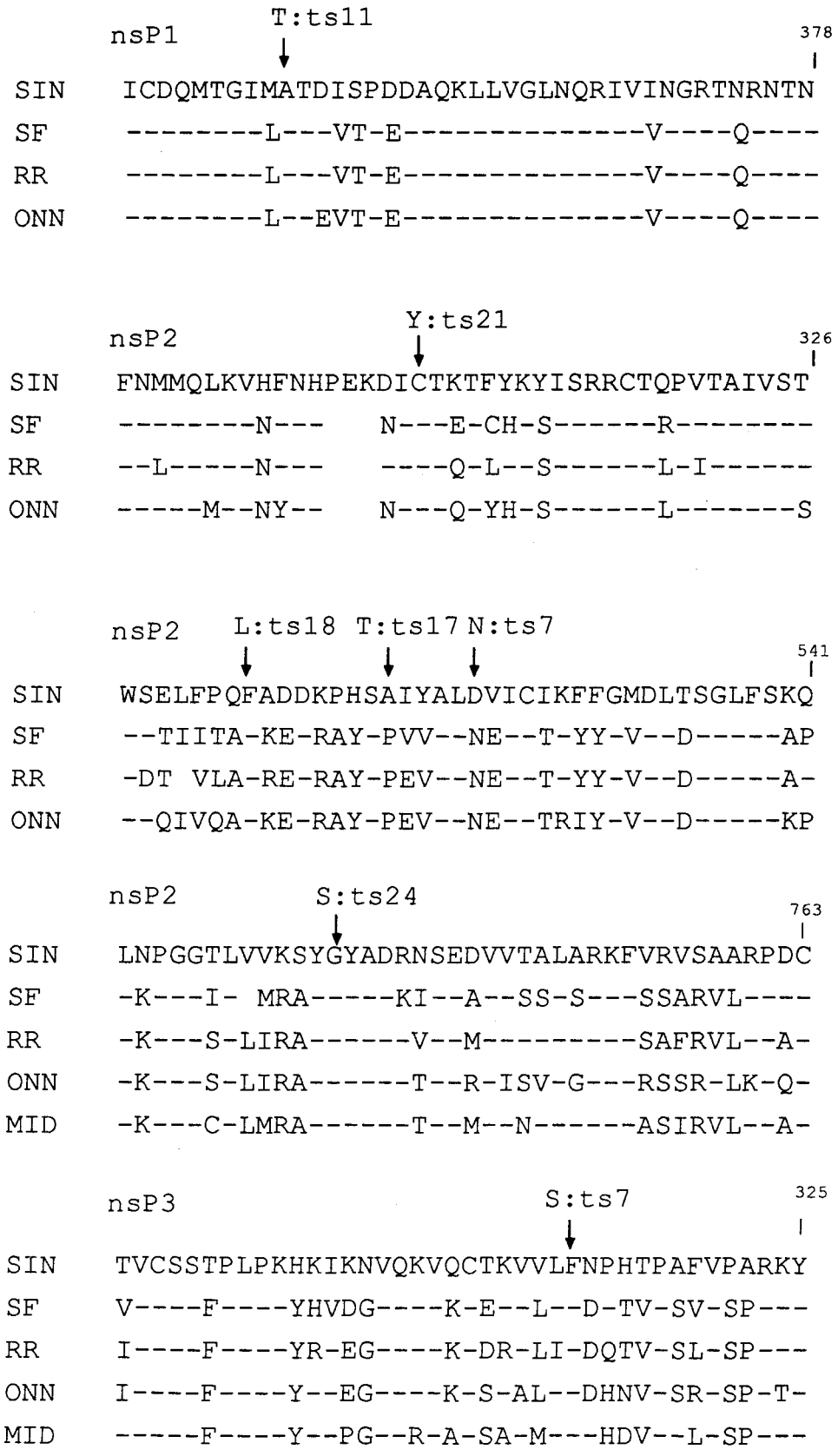
with G group mutant *ts*18. Virus derived from the B group mutant *ts*11 demonstrated ready complementation with members of groups A, F and G. Finally the two members of the G group examined complemented with representatives of group B and F, but not with members of group A. Complementation by the separated changes in *ts*7, represented by Toto:*ts*7B1 and Toto:*ts*7B3, was difficult to demonstrate because of the high reversion frequency or high degree of leakage of these mutants which leads to a high background during the complementation assays, although complementation appears to occur with B (*ts*11) and F (*ts*6) group mutants but not with the A group mutant *ts*24.

From these complementation results and from the finding that mutants in group A and group G both arise from changes in nsP2, it appears that the previous complementation demonstrated between *ts*7 and *ts*18 (G group mutants) with representatives of group A (29) represents intragenic complementation, and that the G group mutants should be considered as a subgroup of group A. We had hoped that the mutation in nsP3 of *ts*7 might define another complementation group, but in part because of the high background in complementation assays with this mutant we were unable to do so (complementation with Group A was not seen).

DISCUSSION

The nature of the *ts* lesions. We have localized the mutations responsible for the temperature-sensitive phenotype of mutants belonging to complementation groups A, B and G. The mutations found are shown in Fig. 3 in which sequences of four or five alphaviruses are compared in the regions affected. The mutations in *ts*11, *ts*21, *ts*18 and *ts*24, as well as the change in nsP3 of *ts*7, affect conserved amino acids. In the case of *ts*11 and *ts*21, as well as *ts*24, these conserved amino acids are found in domains that exhibit a high degree of conservation, whereas the

Fig. 3. Comparison of amino acid sequences from four or five alphaviruses in the regions near *ts* mutations. Arrows indicate the change in amino acid sequence for each mutant. Sequence data are from the following sources: SIN = Sindbis virus, from Strauss *et al.* (32); SF = Semliki Forest virus, from Takkinen (37); RR = Ross River virus, from Faragher *et al.* (10) and Strauss *et al.* (30); ONN = O'Nyong-nyong virus from Strauss *et al.* (30) and unpublished data; MID = Middelburg virus, from Strauss *et al.* (31) and unpublished data.


nsP1
T:ts11
378
SIN ICDQMTGIMATDISPDDAQKLLVGLNQRIVINGRTNRNTN
SF --------L---VT-E-------------V----Q----
RR --------L---VT-E-------------V----Q----
ONN --------L--EVT-E-------------V----Q----
nsP2
Y:ts21
326
SIN FNMMQLKVHFNHPEKDICTKTFYKYISRRCTQPVTAIVST
SF --------N--- N---E-CH-S------R--------
RR --L-----N--- ----Q-L--S------L-I------
ONN -----M--NY-- N---Q-YH-S------L-------S
nsP2
L:ts18 T:ts17 N:ts7
541
SIN WSELFPQFADDKPHSAIYALDVICIKFFGMDLTSGLFSKQ
SF --TIITA-KE-RAY-PVV--NE--T-YY-V--D-----AP
RR -DT VLA-RE-RAY-PEV--NE--T-YY-V--D-----A-
ONN --QIVQA-KE-RAY-PEV--NE--TRIY-V--D-----KP
nsP2
S:ts24
763
SIN LNPGGTLVVKSYGYADRNSEDVVTALARKFVRVSAARPDC
SF -K---I- MRA-----KI--A--SS-S---SSARVL----
RR -K---S-LIRA------V--M---------SAFRVL--A-
ONN -K---S-LIRA------T--R-ISV-G---RSSR-LK-Q-
MID -K---C-LMRA------T--M--N------ASIRVL--A-
nsP3
S:ts7
325
SIN TVCSSTPLPKHKIKNVQKVQCTKVVLFNPHTPAFVPARKY
SF V----F----YHVDG----K-E--L--D-TV-SV-SP---
RR I----F----YR-EG----K-DR-LI-DQTV-SL-SP---
ONN I----F----Y--EG----K-S-AL--DHNV-SR-SP-T-
MID -----F----Y--PG--R-A-SA-M---HDV--L-SP---

change in *ts*18 and the *ts*7 change, although affecting conserved amino acids, are found within domains that are not otherwise highly conserved. The change in nsP2 of *ts*7 is intriguing. This Asp→Asn substitution in nsP2 changes the Sindbis amino acid to that found in the other alphaviruses sequenced to date, and this change of the Sindbis amino acid to the "consensus" amino acid results in temperature sensitivity.

It is noteworthy that *ts*18, *ts*17 and *ts*7 affect amino acids in nsP2 over a region of only 14 amino acids. This region is not particularly well conserved among alphaviruses, but its precise sequence appears to be important because of the fact that multiple changes in this region lead to temperature sensitivity. In the course of this study we also found two changes in nsP3, affecting consecutive amino acids, that have no apparent phenotype, Lys-218→Glu in *ts*7 and Val-219→Ala in *ts*24. Neither of these amino acids is conserved in other alphaviruses and the surrounding domain is not well conserved, suggesting that this region of nsP3 can accommodate a number of amino acid substitutions.

Mutants *ts*7, *ts*11, *ts*17, *ts*21, and *ts*24 were produced by mutagenesis with nitrosoguanidine (5). Of the six mutations in these five mutants that resulted in temperature sensitivity, five are G→A transitions and one is A→G. Transition of G→A is the most common change to be expected from the action of nitrosoguanidine (8), although our past results on mapping of *ts* lesions in Sindbis virus have more often found other changes produced (2, 12, 13, 18). Mutant *ts*18 was produced by mutagenesis by nitrous acid. The U→C transition responsible for temperature sensitivity is not one of the changes expected to arise from HNO_2 treatment (8), but once again our previous mapping studies have often found other changes produced, including U→C transitions (2, 12, 13).

From our mapping studies, the gene order for the complementation groups is NH_2-B-A/G-(G)-F-COOH. These results differ from the results of Fuller and

Marcus (11) in which the complementation groups were ordered as NH_2-G-A-B/F-COOH, determined from the relative rate of UV inactivation of the ability of the mutants to complement. The major discrepancy between the two orders is the location of complementation group B and it is unclear why the UV inactivation data placed this complementation group in the wrong position.

Functions of the nonstructural proteins. All four RNA^- complementation groups have now been assigned to nonstructural proteins, and these results are summarized in Fig. 4. From these data hypotheses can be developed for the functions of the nonstructural proteins during RNA replication. Firstly it is clear that mutations in any of the four nonstructural proteins can lead to an RNA^- phenotype, implying that all four proteins are important for RNA replication (note that the change in nsP3 of *ts7* is responsible for the RNA^- phenotype of this mutant). Secondly, from the phenotypes of the RNA^- mutants that have been characterized, at least some of the functions of nsP1, nsP2 and nsP4 can be deduced. The studies of the F group mutants *ts6* and *ts110* are consistent with the hypothesis that nsP4 is the polymerase or major elongation component of the viral replicase (13). Previous studies of the single group B mutant, *ts11*, had shown that minus-strand RNA synthesis ceased when infected cultures were shifted to 40°C from 30°C, whereas synthesis of both 49S and 26S plus-strand RNA continued unabated (23). This suggests that nsP1 may function as an initiation factor for synthesis of minus-strand RNA.

The possible functions of nsP2 deduced from previous studies of the group A and group G mutants are more varied and it is probably the case that each of the virus nonstructural proteins has more than a single function during RNA replication. *ts17*, *ts21*, and *ts24* of group A (as well as *ts133* which has not yet been mapped), and *ts18* of group G, have been found to be temperature-sensitive in the synthesis of 26S RNA upon shift from a permissive to a nonpermissive condition (17, 24), suggesting that nsP2 may be required for the initiation of 26S

Fig. 4. Summary of mutations of Sindbis RNA^- *ts* mutants. The change in amino acid sequence from the Sindbis HR parental sequence is shown for each *ts* mutant mapped to date. Amino acids are numbered from the N-terminus of each protein. $^*P$ in *ts*17, *ts*24, and *ts*18 indicates these mutants have a defect in the cleavage of nonstructural proteins at the nonpermissive temperature. The changes found in nsP3 that do not result in temperature sensitivity, Val-219→Ala in *ts*24 and Lys-218→Glu in *ts*7, are not shown. Note that the nsP3 change shown for *ts*7 Phe-312→Ser, results in an RNA^- phenotype, whereas the nsP2 change, Asp-522→Asn, results in an $RNA^\pm$ phenotype. The Group F changes are from Hahn *et al.* (13).

Summary : Mapping of Sindbis ts RNA $^-$ Mutants

Group A	ts17*P	ns P2	Ala - 517 → Thr
	ts21	ns P2	Cys - 304 → Tyr
	ts24*P	ns P2	Gly - 736 → Ser
Group B	ts11	ns P1	Ala - 348 → Thr
Group G	ts 7	ns P2	Asp - 522 → Asn
		ns P3	Phe - 312 → Ser
	ts18*P	ns P2	Phe - 509 → Leu
Group F	ts 6	ns P4	Gly -153 → Glu
	ts110	ns P4	Gly - 324 → Glu
	ts118	ns P2	Val - 425 → Ala
		ns P4	Gln - 93 → Arg

RNA synthesis. In addition, three of the mutants with changes in nsP2, *ts*17 and *ts*24 of group A and *ts*18 in group G, as indicated in Fig. 4, fail to process the polypeptide precursor nsP123 (also known as ns230) upon shift to the nonpermissive temperature (17, 24; W. R. Hardy and Y. S. Hahn, unpublished data). This suggests that nsP2 is the protease that processes the nonstructural polyprotein precursors. Two of the mutations in nsP2 that result in failure to cleave the polyprotein precursor, those in *ts*17 and in *ts*18, are found close to one another (Fig. 3), suggesting that this region might form part of the protease, although the mutation in *ts*7 is also found within the same region and does not affect proteolytic processing. The mutation in *ts*24 which also leads to failure of processing is found some 200 amino acids downstream, suggesting that this region might also form part of the protease, and that nsP2 is folded so as to bring these regions together to form an active site. We presume that the enzyme is a cysteine protease, and propose that one of the conserved cysteines in the domains defined by the *ts* lesions is the active residue. It is also possible, however, that these various mutations affect protease function by altering the overall folding of the protein rather than because they are situated near the active site.

*ts*17, *ts*133, and *ts*24 of group A were also found to have a third phenotypic defect, namely, that upon shift from permissive to nonpermissive temperature, the normal shut-down of minus-strand synthesis failed to occur (24, 25, 26). In these mutants, not only does this shut-down fail to occur at a nonpermissive temperature, but shifting infected cells from permissive to nonpermissive conditions after shut-down of minus-strand synthesis results in resumption of minus-strand synthesis. ts^+ revertants of *ts*17 and *ts*133 demonstrated normal shut off, whereas ts^+ revertants of *ts*24 did not. Thus in the case of *ts*24 the defect in shut down appears to arise from an unmapped mutation, consistent with the results in Table 3 implying that *ts*24 has unmapped defects related to RNA

synthesis. Further studies on these mutants and their revertants will be required to ascertain if the regulation of minus-strand synthesis is a function of nsP2, although the results with *ts*17 suggest that it is, albeit that another protein (defined by the unmapped change in *ts*24) may be involved as well.

Thus it appears that alterations in nsP2 can result in *ts* synthesis of 26S RNA, *ts* proteolysis of nonstructural proteins, and *ts* regulation of minus-strand synthesis, implying that this protein is involved in all of these functions. However, the *ts* synthesis of 26S RNA does not in itself lead necessarily to the *ts* phenotype, defined as a failure to form plaques at the nonpermissive temperature, as a ts^+ revertant of *ts*133 examined remained *ts* in 26S RNA synthesis (25), and as discussed above for *ts*24 the *ts* regulation of minus-strand synthesis does not necessarily result in temperature sensitivity of plaque formation. All of the ts^+ revertants of proteolysis mutants examined, however, did show normal processing, so that in the case of these mutants the temperature sensitivity might result from the failure to process the nonstructural proteins at elevated temperatures. From the results with *ts*24 and *ts*133, and because other RNA^- mutants in Group A show none of these phenotypes, it appears clear that nsP2 must possess other functions that are also required for RNA replication.

It has been shown that nonstructural proteins nsP1, nsP2 and nsP4 of Sindbis virus share amino acid sequence homology with nonstructural proteins from several RNA plant viruses including alfalfa mosaic virus, bromegrass mosaic virus, and tobacco mosaic virus (1, 15). The plant viruses share similarities in replication strategies with the alphaviruses, including production of a subgenomic messenger RNA for the translation of structural proteins, as well as the obvious need to produce genomic length plus- and minus-strands. The amino acid sequence similarities suggest that the plant virus proteins perform the same functions

during replication of the virus RNAs as do the corresponding proteins of Sindbis virus, and it will be of interest to compare the functions of these proteins as more information becomes known.

ACKNOWLEDGEMENTS

The expert technical assistance of E. M. Lenches is gratefully acknowledged. This work was supported by grants AI10793 and AI20612 from NIH and by grant BMD8617372 from NSF.

LITERATURE CITED

1. **Ahlquist, P., E. G. Strauss, C. M. Rice, J. H. Strauss, J. Haseloff, and D. Zimmern.** 1985. Sindbis virus proteins nsP1 and nsP2 contain homology to nonstructural proteins from several RNA plant viruses. J. Virol. **53**:536-542.
2. **Arias, C., J. R. Bell, E. M. Lenches, E. G. Strauss, and J. H. Strauss.** 1983. Sequence analysis of two mutants of Sindbis virus defective in the intracellular transport of their glycoproteins. J. Mol. Biol. **168**:87-102.
3. **Barton, D. J., S. G. Sawicki, and D. L. Sawicki.** 1988. Demonstration *in vitro* of temperature-sensitive elongation of RNA in Sindbis virus mutant *ts*6. J. Virol. **62**:3597-3602.
4. **Bruton, C. J. and S. I. T. Kennedy.** 1975. Semliki Forest virus intracellular RNA: Properties of the multistranded RNA species and kinetics of positive and minus-strand synthesis. J. Gen. Virol. **32**:413-430.
5. **Burge, B. W. and E. R. Pfefferkorn.** 1966. Isolation and characterization of conditional-lethal mutants of Sindbis virus. Virology **30**:204-213.
6. **Burge, B. W. and E. R. Pfefferkorn.** 1966. Complementation between temperature-sensitive mutants of Sindbis virus. Virology **30**:214-223.
7. **Burge, B. W. and E. R. Pfefferkorn.** 1967. Temperature-sensitive mutants of Sindbis virus: Biochemical correlates of complementation. J. Virol. **1**:956-962.
8. **Drake, J. W.** 1970. The Molecular Basis of Mutation, Chap. 13, pp. 146-159. Holden-Day Inc., San Francisco.
9. **Durbin, R. K. and V. Stollar.** 1986. Sequence analysis of the E2 gene of a hyperglycosylated, host-restricted mutant of Sindbis virus and estimation of mutation rate from frequency of revertants. Virology **154**:135-143.

10. **Faragher, S. G., A. D. J. Meek, C. M. Rice, and L. Dalgarno.** 1988. Genome sequences of a mouse-avirulent and a mouse-virulent strain of Ross River virus. Virology **163:**509-526.

11. **Fuller, F. J., and P. I. Marcus.** 1980. Sindbis virus. I. Gene order of translation in *in vivo*. Virology **107:**441-451.

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

13. **Hahn, Y. S., A. Grakoui, C. M. Rice, E. G. Strauss, and J. H. Strauss.** 1989. Mapping of RNA^- temperature-sensitive mutants of Sindbis virus: Complementation group F mutants have lesions in nsP4. J. Virol., in press.

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

15. **Haseloff, J., P. Goelet, D. Zimmern, P. Ahlquist, R. Dasgupta, and P. Kaesberg.** (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.

16. **Holland, J., K. Spindler, F. Horodyski, E. Grabau, S. Nichol, and S. VandePol.** 1982. Rapid evolution of RNA genomes. *Science* **215,** 1577-1585.

17. **Keranen, S., and L. Kaariainen.** 1979. Functional defects of RNA-negative temperature-sensitive mutants of Sindbis and Semliki Forest viruses. J. Virol. **32:**19-29.

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

19. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base specific chemical cleavages. Methods in Enzymol. **65:**499-560.

20. **Rice, C. M. and J. H. Strauss.** 1981. Synthesis, cleavage, and sequence analysis of DNA complementary to the 26S messenger RNA of Sindbis virus. J. Mol. Biol. **150:**313-340.

21. **Rice, C. M., R. Levis, J. H. Strauss, and H. V. Huang.** 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.

22. **Sawicki, D. L. and S. G. Sawicki.** 1980. Short-lived minus-strand polymerase for Semliki Forest virus. J. Virol. **34:**108-118.

23. **Sawicki, D. L., S. G. Sawicki, S. Keranen, and L. Kaarianen.** 1981. Specific Sindbis virus coded function for minus-strand RNA synthesis. J. Virol. **39:**348-358.

24. **Sawicki, D. L., and S. G. Sawicki.** 1985. Functional analysis of the A complementation group mutants of Sindbis HR virus. Virology **144:**20-34.

25. **Sawicki, S. G., and D. L. Sawicki.** 1986. The effect of loss of regulation of minus-strand RNA synthesis on Sindbis virus replication. Virology **151:**339-349.

26. **Sawicki, S. G., D. L. Sawicki, L. Kaariainen, and S. Keranen.** 1981. A Sindbis virus mutant temperature-sensitive in the regulation of minus-strand synthesis. Virology **115:**161-172.

27. **Smith, D. R., and J. M. Calvo.** 1980. Nucleotide sequence of the *E. coli* gene coding for dihydrofolate reductase. Nucl. Acids Res. **8:**2244-2274.

28. **Steinhauer, D. A., and J. J. Holland.** 1987. Rapid evolution of RNA viruses. Ann. Rev. Microbiol. **41:**409-433.

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

30. **Strauss, E. G., R. Levinson, C. M. Rice, J. Dalrymple, and J. H. Strauss.** 1988. Nonstructural proteins nsP3 and nsP4 of Ross River and O'Nyong-nyong viruses: Sequence and comparison with those of other alphaviruses. Virology **164:**265-274.

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

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

33. **Strauss, E. G., and J. H. Strauss.** 1980. Mutants of alphaviruses: genetics and physiology, pp. 393-426. *In* R. W. Schlesinger (ed.), The Togaviruses. Academic Press, New York.

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

35. **Strauss, J. H., E. G. Strauss, C. S. Hahn, Y. S. Hahn, R. Galler, R., W. R. Hardy, and C. M. Rice.** 1987. Replication of alphaviruses and flaviviruses: Proteolytic processing of polyproteins, pp. 209-225. *In* M. A. Brinton and R. Rueckert (eds.) UCLA Symposium on Molecular and Cellular Biology, Vol. 54, Alan R. Liss, Inc., New York.

36. **Strauss, J. H. and E. G. Strauss.** 1988. Evolution of RNA viruses. Ann. Revs. Micro. **42:**657-683.

37. **Takkinen, K.** 1986. Complete nucleotide sequence of the nonstructural protein genes of Semliki Forest virus. Nucl. Acids Res. **14:**5667-5682.

38. **White, B. A., and F. C. Bancroft.** 1982. Cytoplasmic dot hybridization. Simple analysis of relative mRNA levels in multiple small cell or tissue samples. J. Biol. Chem. **257:**8569-8572.

PART II

Structure of the Flavivirus Genome

Chapter 3

Nucleotide Sequence of Dengue 2 RNA and Comparison of the Encoded Proteins with Those of Other Flaviviruses

This chapter was published in *Virology* and is reproduced with the permission of the publisher.

VIROLOGY 162, 167–180 (1988)

Nucleotide Sequence of Dengue 2 RNA and Comparison of the Encoded Proteins with Those of Other Flaviviruses

YOUNG S. HAHN,* RICARDO GALLER,*,[1] TIM HUNKAPILLER,* JOEL M. DALRYMPLE,† JAMES H. STRAUSS,* AND ELLEN G. STRAUSS*,[2]

*Division of Biology, California Institute of Technology, Pasadena, California 91125, and †Virology Division, U. S. Army Medical Research Institute of Infectious Diseases, Ft. Detrick, Frederick, Maryland 21701

Received July 13, 1987; accepted September 2, 1987

We have determined the complete sequence of the RNA of dengue 2 virus (S1 candidate vaccine strain derived from the PR-159 isolate) with the exception of about 15 nucleotides at the 5′ end. The genome organization is the same as that deduced earlier for other flaviviruses and the amino acid sequences of the encoded dengue 2 proteins show striking homology to those of other flaviviruses. The overall amino acid sequence similarity between dengue 2 and yellow fever virus is 44.7%, whereas that between dengue 2 and West Nile virus is 50.7%. These viruses represent three different serological subgroups of mosquito-borne flaviviruses. Comparison of the amino acid sequences shows that amino acid sequence homology is not uniformly distributed among the proteins; highest homology is found in some domains of nonstructural protein NS5 and lowest homology in the hydrophobic polypeptides ns2a and 2b. In general the structural proteins are less well conserved than the nonstructural proteins. Hydrophobicity profiles, however, are remarkably similar throughout the translated region. Comparison of the dengue 2 PR-159 sequence to partial sequence data from dengue 4 and another strain of dengue 2 virus reveals amino acid sequence homologies of about 64 and 96%, respectively, in the structural protein region. Thus as a general rule for flaviviruses examined to date, members of different serological subgroups demonstrate 50% or less amino acid sequence homology, members of the same subgroup average 65–75% homology, and strains of the same virus demonstrate greater than 95% amino acid sequence similarity.

INTRODUCTION

The four serotypes of dengue virus form a serological group within the *Flaviviridae*, a family comprised of approximately 70 enveloped viruses, primarily transmitted to vertebrates by mosquitoes and ticks, which contain a single molecule of plus-sense RNA of about 11 kb as their genome. The dengue viruses are important human disease agents causing millions of cases of dengue fever annually. Uncomplicated dengue fever is usually an acute illness characterized by joint pain, fever, and rash, but in a small percentage of cases complications develop leading to dengue hemorrhagic fever or dengue shock syndrome which have a significant mortality. The causes of dengue hemorrhagic fever or dengue shock syndrome are not yet understood. It has been hypothesized that this syndrome results from immune enhancement and usually develops when a person is infected with dengue virus within 6 to 12 months after infection by one of the other dengue serotypes (Halstead, 1981), but at least some cases of primary infection by dengue virus also appear to result in dengue shock syndrome and it has been suggested that strains of increased virulence may play a role in the epidemiology of the disease (reviewed in Monath, 1986; Rosen, 1986).

The dengue viruses have a very wide geographical distribution and are present in both the Old World (India, Southeast Asia, Oceania, Africa) and the New World (Central America, the Caribbean, and South America). They have caused explosive epidemics in which the majority of the population of an area is infected (see, for example, Brès, 1979; Carey, 1971). An epidemic of dengue 1 currently occurring in Brazil has led to an estimated 500,000 cases of dengue fever through 1986 (Schatzmayr *et al.*, 1986). The increasing frequency of epidemics in the Caribbean and the recent spread into Brazil has led to fears that the geographic range of the virus will continue to expand to include the United States, where epidemics of dengue have occurred in the past, the most recent being in 1941 (Ehrenkranz *et al.*, 1971). It is unclear whether the recent epidemic spread of the virus is simply a cyclical phenomenon involving introduction of the viruses into susceptible populations or is related to abandonment of efforts to control the mosquito vector *Aedes aegypti* or may even be due partially to the control of yellow fever that began in 1902 (since there may be interference at the population level between

[1] Present address: Fundacao Oswaldo Cruz, Dept. Bioquimica e Biologic Molecular, Rio de Janiero, Brazil, CEP 21040.

[2] To whom requests for reprints should be addressed.

0042-6822/88 $3.00

these viruses; see, for example, Brès, 1979; Groot *et al.*, 1979; Frederiksen, 1955; Theiler and Anderson, 1975).

At the current time there is no effective vaccine for any of the dengue viruses; attempts to produce one have met with only limited success. In addition the complications of dengue shock syndrome possibly caused by sequential infection with dengue viruses have led to caution being exercised in the development of a vaccine. The problem is exacerbated by the lack of an animal model system for the disease. Dengue 1 virus (Hawaii strain) and dengue 2 virus (New Guinea strain) were first isolated in 1944–1945, and candidate vaccines for these viruses were developed by mouse brain passage and tested in a small number of volunteers (reviewed in Sabin, 1952); extensive field trials were never conducted, however, and mouse brain passages would appear undesirable for current vaccines. More recently, a candidate live virus vaccine for dengue 2 has been produced by 19 passages in cell culture involving several plaque isolations (Eckels *et al.*, 1976, 1980). This S1 strain is a small plaque, temperature-sensitive derivative of a dengue 2 virus isolated in 1969 from a patient with dengue fever in Puerto Rico, referred to as the PR-159 strain. The S1 strain exhibits reduced virulence for suckling mice and is markedly attenuated in rhesus monkeys. It has been tested in more than 100 human volunteers and found to be effective if the recipient had previously been immunized against yellow fever; in the absence of prior immunity to yellow fever the outcome is less predictable. However, the frequency and severity of side reactions to the vaccine make it unsuitable for general use (Bancroft *et al.*, 1981; Scott *et al.*, 1983; Bancroft *et al.*, 1984).

We report here virtually the complete nucleotide sequence of this S1 candidate vaccine strain derived from dengue 2 (DEN 2) and compare this sequence and that of the encoded proteins with those of viruses from two other serological subgroups of mosquito-borne flaviviruses, yellow fever virus (YF), and West Nile virus (WN), and with the structural proteins of the Jamaican DEN 2 and of dengue 4 (DEN 4). With the complete sequence of representatives of each of three serological subgroups of mosquito-borne flaviviruses now known, the relationships of the flaviviruses to one another are becoming clearer.

MATERIALS AND METHODS

Preparation of dengue 2 virus RNA

The original source of DEN 2 virus used to prepare virion RNA was the Dengue Vaccine (Type 2) Live Attenuated Virus Strain PR-159 (S1) taken directly from Lot. No. 1 of the vaccine prepared by the Department of Biologics Research, WRAIR, Washington, D.C. 20012, and dated January 1976 (Eckels *et al.*, 1976, 1980). A single passage of this virus in the *Aedes albopictus* C6/36 cell line served as infectious inoculum for all virion propagation experiments. Confluent monolayers of C6/36 cells in 150-cm^2 flasks were infected at low multiplicity (<0.01 PFU/cell) and culture fluids were harvested following 5 days incubation at 28°. Selected cultures were radiolabeled by adding [^{3}H]uridine (20 μCi/ml Amersham) at 2 days postinfection. Infected culture fluids were clarified by centrifugation at 3000 *g* for 30 min and virus was concentrated by precipitation using polyethylene glycol and additional NaCl. Virions were purified by potassium tartrate–glycerol density gradient centrifugation followed by rate-zonal centrifugation on linear sucrose gradients (Repik *et al.*, 1983). RNA was obtained from pelleted virions that were disrupted with SDS and extracted with a mixture of phenol/cresol/8-hydroxyquinoline/chloroform followed by ethanol precipitation. Sucrose gradient analysis of [^{3}H]uridine-labeled DEN 2 RNA revealed a single peak of radioactivity sedimenting at approximately 42 S.

Cloning and sequencing of DEN 2 RNA

Details of the methods used to obtain clones of DEN 2 cDNA have been described (Rice *et al.*, 1987). Briefly, first-strand cDNA synthesis was performed using random primers and avian myeloblastosis virus reverse transcriptase in the presence of actinomycin D to inhibit fold back. Second-strand synthesis used the methods of Okayama and Berg (1982). After methylation of *Eco*RI sites with *Eco*RI methylase, treatment with T4 polymerase, and attachment of *Eco*RI linkers by blunt-end ligation, the double-strand cDNA copies were size selected on agarose gels and inserted into the *Eco*RI site of the vector pGem1 (Promega Biotech). Ampicillin-resistant colonies were screened for the size of the insert, and colonies containing large inserts were screened by sequencing the ends of the insert from labeling sites within the polylinker. Computer programs were used to align this sequence with that for yellow fever virus, using the significant similarities that exist between the amino acid sequences of dengue virus and yellow fever virus. Characterized clones were used to screen the library by colony hybridization. Clones representing the entire nucleotide sequence less ~15 nucleotides from the 5′ end (by comparison with other flaviviruses) and ~330 nucleotides from the 3′ end were characterized and sequenced by the chemical methods of Maxam and Gilbert (1980) as modified by Smith and Calvo (1980).

To obtain the 3′ end sequence of DEN 2 RNA, which was not present in the random library, the genomic RNA was tailed with poly(A) as previously described (Rice *et al.*, 1985), and cDNA was synthesized with oligo(dT) primers. The resultant 3′ clones were selected by colony hybridization with a probe from the randomly primed clone closest to the 3′ end (clone 8). Twelve clones were obtained and found to have the same restriction map. Three of these were sequenced and were found to have an identical sequence preceding a poly(A) tract, and to be homologous to the 3′ terminal sequences reported for other flaviviruses (Hahn *et al.*, 1987b). Because of this homology and because two of the sequenced clones (which may not be independent isolates) had poly(A) tracts of 22 nucleotides and the third had a tract of 35 nucleotides, which were longer than the oligo(dT) used as primer, it seems unlikely that these clones resulted from internal priming at an oligo(A) tract, but rather arose from the 3′ terminus.

RESULTS

Sequence of dengue 2 RNA

The sequencing strategy used to obtain the sequence of DEN 2 RNA is shown in Fig. 1. Clones containing inserts from 1.1 to 5.2 kb were ordered along the genome by sequencing of the termini of these inserts and comparing the deduced amino acid sequence to that of YF virus (Rice *et al.*, 1987). Probes derived from clones thus characterized were used to rescreen the library, in particular to identify those rare clones representing the ends of the genome. Ultimately it was concluded that the randomly primed library included the entire DEN 2 sequence with the exception of 15 nucleotides at the 5′ end and about 330 nucleotides at the 3′ end. Clones representing the 3′ end were obtained by tailing the genomic RNA with poly(A) and priming cDNA synthesis wih oligo(dT), as described under Materials and Methods (Fig. 1).

Once these clones were aligned, they were systematically sequenced by the chemical method of Maxam and Gilbert (1980) in order to obtain the complete sequence. Virtually all of the sequence was obtained on at least two independent clones so that cloning artifacts could be excluded and clonal heterogeneity would be detected; moreover, more than 75% of the sequence was obtained by sequencing both strands, often on two independent clones (Fig. 1). A number of clonal differences were observed and these are reported below. During the propagation of virus as a source of RNA for cloning experiments, no attempt was made to select for the small plaque phenotype

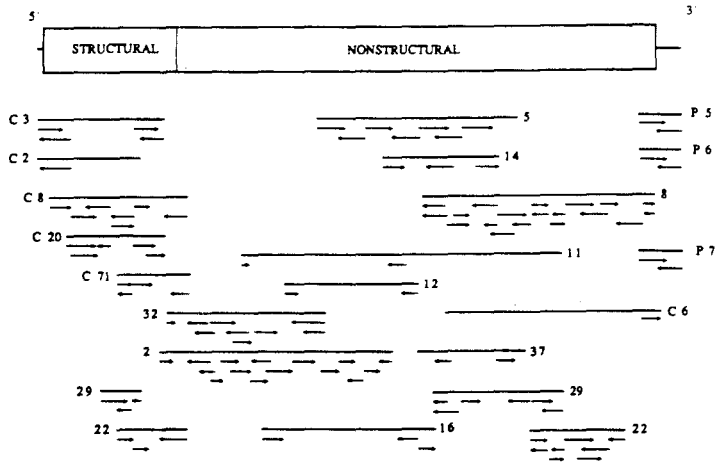


FIG. 1. Sequencing strategy for the DEN 2 genome. The various clones used to determine the sequence are mapped together with a schematic representation of the RNA genome. Clones isolated in the original screening are indicated by numbers without prefixes. Clones indicated with the prefix C were isolated after colony hybridization using probes derived from clones representing the 5′ or 3′ end of the previously established sequence. The 5′ end probe was the *Eco*RI (958) to *Nco*I (1221) fragment, and the 3′ probe was the *Eco*RI (8753) to *Nco*I (10025) fragment, in either case labeled by filling the single-strand protrusions with the Klenow fragment and isolated by gel electrophoresis in low-melting-temperature agarose or in acrylamide. Clones indicated by the prefix P were obtained after poly(A) tailing of the RNA and use of oligo(dT) to prime first-strand cDNA synthesis; selection was with the 3′ probe described above. Arrows pointing to the right indicate regions sequenced 3′ to 5′ on the minus-sense strand; arrows to the left, 3′ to 5′ on the plus-sense strand. Note that clones 22 and 29 have two inserts from noncontiguous regions of the genome.

and some large plaques were evident in the plaque assay of the suspensions used for purification and RNA extraction. Reversion to large plaque upon passage of the dengue 2 (S1) virus has been reported (Eckels *et al.*, 1976) and the possibility that initial cDNA clones represented each of the responsible RNA species cannot be excluded. However, virus with small plaque morphology represented the predominant phenotype in all preparations.

We also found one clone (clone 8) that possessed an extra nucleotide in a stretch of adenine residues such that the sequence GAAAAAAC in DEN 2 RNA at positions 6835–6842, confirmed by sequence on four other clones (Fig. 1), was replaced in clone 8 by the sequence GAAAAAAAC, which would result in a frame shift in the deduced amino acid sequence. This phenomenon is related to that reported by Lindqvist *et al.* (1986) in which, during cloning of alphavirus RNAs, two examples were found in which stretches of adenine residues had single residue deletions. It appears that some enzyme used during the cloning stages stutters when copying poly(A) tracks and can either add or delete nucleotides during this process. This enzyme is presumably reverse transcriptase although other enzymes cannot be excluded (e.g., *Pol* I, or even the viral replicase if it produced a defective RNA that was subsequently cloned).

The nucleotide sequence found was 10,703 nucleotides (including 15 unidentified bases at the 5′ end) in length, with a base composition of 33% A, 25.4% G, 21% U, and 20.7% C. The dinucleotide frequencies show the usual deficit of the CG doublet which has been reported in eukaryotic genomes (Bird, 1980) and in flaviviruses (Rice *et al.*, 1985).

Genome organization of dengue 2

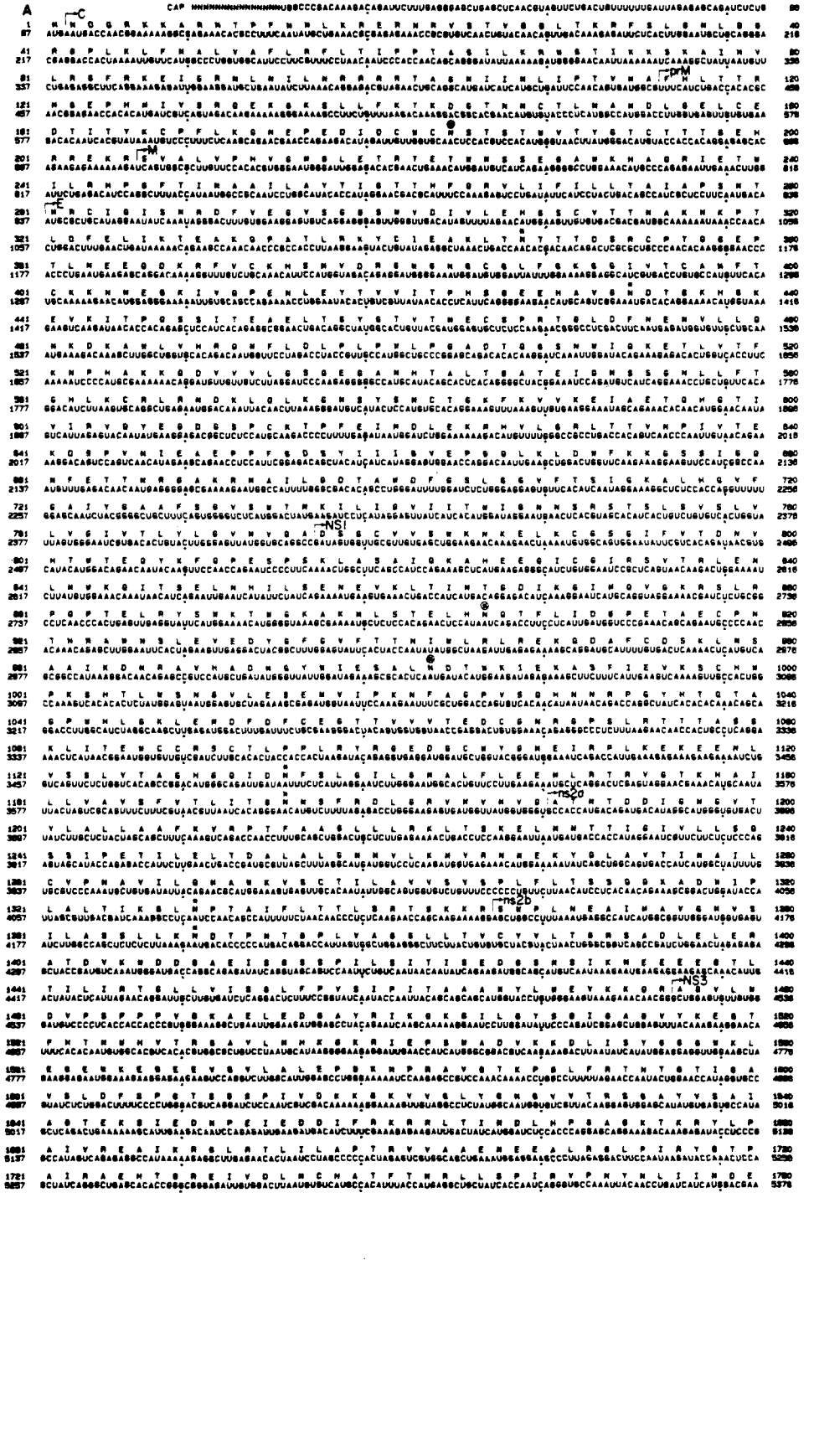
The translated nucleotide sequence of DEN 2 RNA is shown in Fig. 2. The genome organization of the RNA is the same as that previously found for YF RNA (Rice *et al.*, 1985) and for WN RNA (Castle *et al.*, 1985, 1986; Wengler *et al.*, 1985). There are 96 nucleotides of 5′ untranslated sequence and 443 of 3′ untranslated flanking one, long open reading frame (of 10,164 nucleotides) which is translated to give the structural proteins capsid (C), membrane protein precursor (prM) which contains the sequences for the membrane-like protein M, and envelope protein (E), followed by nonstructural proteins NS1, ns2, NS3, ns4, NS5. The start point of protein E has been assigned by direct amino acid sequencing of the protein isolated from purified DEN 2 (Bell *et al.*, 1985) and the start points of NS1, NS3, and NS5 have been also confirmed by direct amino acid sequencing (Biedrzycka *et al.*, 1987). The start points of the remaining proteins (see below) have been assigned by homology with the start points of yellow fever proteins or proteins of other flaviviruses (Rice *et al.*, 1985, 1986a; Castle *et al.*, 1985, 1986; Wengler *et al.*, 1985).

Cleavages in the dengue 2 polyprotein

As is the case for other flaviviruses (reviewed in Rice *et al.*, 1986b), the individual proteins of DEN 2 appear to be produced by post-translational cleavage of a precursor polyprotein. Four of the DEN 2 protein start points are known by direct amino acid sequencing (E, NS1, NS3, NS5) and the rest have been assigned by homology. As proposed for other flaviviruses, the structural proteins and NS1 appear to be processed by the cellular enzyme signalase. Each of these proteins, C, prM, E, and NS1, has a hydrophobic domain at or near the C-terminus that may form a membrane-spanning anchor. Cleavage in each case satisfies the -1, -3 rule for signalase (von Heijne, 1986; Perlman and Halvorson, 1983). However, the cleavage site after NS1 has not been positioned by direct amino acid sequencing for any flavivirus, and for the site after C the assignment is from a comparison with known sites in other flaviviruses; possible sites for signalase cleavage are usually not unique (other nearby sites often would satsify the -1, -3 rule) and the exact positioning of such sites awaits further data.

FIG. 2. Translated nucleotide sequence of DEN 2 RNA (PR159/S1 strain). The sequence obtained begins approximately 15 nucleotides from the 5′ end of the genome [by comparison with the DEN 2 (Jamaica) sequence of Deubel *et al.*, 1986]. The predicted start points of the various proteins are shown. For proteins E (Bell *et al.*, 1975) and NS1, NS3, NS5 (Biedrzycka *et al.*, 1987), the start points have been established by direct amino terminal sequencing of the respective proteins. For the remaining proteins, start points have been assigned from homologies with the yellow fever sequence (Rice *et al.*, 1985) and the West Nile sequence (Castle *et al.*, 1985, 1986; Wengler *et al.*, 1985). Asterisks indicate potential sites of N-linked glycosylation; circled asterisks are sites of glycosylation conserved among flaviviruses. The single letter amino acid code is used: A, Ala, C, Cys, D, Asp, E, Glu, F, Phe, G, Gly, H, His, I, Ile, K, Lys, L, Leu, M, Met, N, Asn, P, Pro, Q, Gln, R, Arg, S, Ser, T, Thr, V, Val, W, Trp, Y, Tyr. The sequence reported is derived from sequencing a number of clones and several clonal differences were found. Nucleotide 959 is A in clone C8 and G in clone C20 (leading to a change in the amino acid assignment from N → S); 1314 is G in C8 and A in 29 (silent); 4965 is C in 5 and T in 2 (silent); 5815 is G in 5 and 2 and C in 14 (V → L); the stretch of six A's beginning at 6836 is replaced by seven A's in clone 8, leading to a frame shift, as discussed in the text; 10021 is A in 8 and P7 and C in P5 and P6 (which may not represent independent isolates, as noted under Materials and Methods; leading to a N → H substitution).

SEQUENCE OF DENGUE 2 RNA



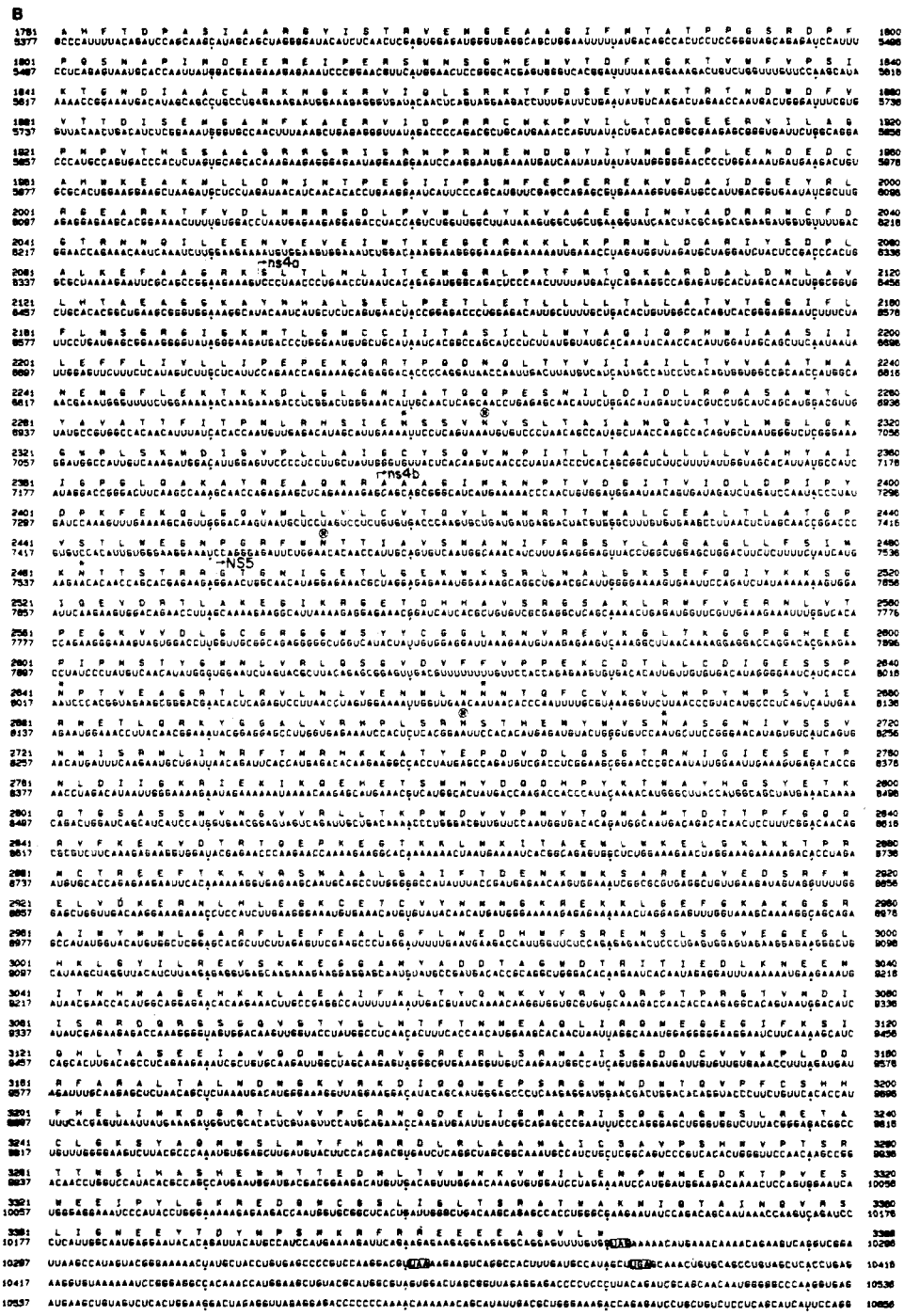


FIG. 2—*Continued.*

The nonstructural proteins appear to be cleaved after two basic amino acid residues in succession, either Lys–Arg or Arg–Arg with one exception (that to produce the N-terminus of NS3), which are flanked on either or both sides by amino acids with short side chains (Fig. 2) (Rice *et al.*, 1985, 1986a). The starts of NS3 and NS5 have been confirmed (Biedrzycka *et al.*, 1987). The sites in the ns2 and ns4 regions have not been confirmed by direct amino acid sequence in any flavivirus and confirmation of these sites is necessary.

The cleavage of prM to produce M occurs after translation, probably during virus maturation, and follows the canonical sequence Arg–X–Arg/Lys–Arg. This type of cleavage event occurs in many enveloped virus glycoprotein precursors and has been hypothesized to be catalyzed by an enzyme of the Golgi complex (reviewed in Strauss and Strauss, 1985).

Glycosylation sites

The nature and location of potential sites of glycosylation in flaviviruses show relatively little conservation among the various members of the family. The glycosylated precursor prM of the membrane-associated protein M has a single site of glycosylation in DEN 2 at Asn-69 (both the PR-159 S1 strain sequenced here and the Jamaican strain sequenced by Deubel *et al.*, 1986), which is also present in DEN 4 (Zhao *et al.*, 1986), but which is not present in the prM of other flavivirus subgroups [the YF subgroup or the WN/Murray Valley encephalitis (MVE)/St. Louis encephalitis (SLE) subgroup]. YF has a site of glycosylation at Asn-13 (Rice *et al.*, 1985) and the WN/MVE/SLE group at Asn-15 (Castle *et al.*, 1985; Dalgarno *et al.*, 1986; Trent *et al.*, 1987). In addition YF has two additional sites for potential glycosylation in prM.

Similarly, glycosylation sites in the E protein are not conserved. Both DEN 2 strains examined to date and the DEN 4 strain sequenced have potential glycosylation sites at Asn-67 and Asn-153. MVE also has a (single) site at Asn-154 and SLE has sites at Asn-154 (shared with MVE) and Asn-314, whereas for WN (Wengler *et al.*, 1985) and Kunjin (Wright, 1982) the E protein has no sites for N-linked glycosylation. YF virus has a single potential site at Asn-309.

Only in the case of NS1 do there appear to be glycosylation sites conserved among all the flaviviruses. The four flaviviruses sequenced to date through the NS1 region (DEN 2, this paper; YF, Rice *et al.*, 1985; MVE, Dalgarno *et al.*, 1986; WN, Castle *et al.*, 1986) have potential sites of NS1 glycosylation at Asn-130 and Asn-207 (Asn-208 in YF). DEN 2 has two additional sites at Asn-359 and Asn-399 which are not shared with the others; and MVE and WN have a potential site at Asn-175. Conservation of at least some sites suggests that glycosylation plays an important role in NS1 function.

The remaining nonstructural proteins are not known to be glycosylated. There is a potential site in each of the proteins NS5, ns4a, and ns4b that are found in other flaviviruses (YF and WN). NS5 in many flaviviruses has been examined on acrylamide gels and found not to be glycosylated, but ns4a has not been conclusively identified on protein gels (it is for this reason that we describe it with lower case letters).

Conservation of cysteines

As has been noted previously (Rice *et al.*, 1986b; Dalgarno *et al.*, 1986), cysteine residues appear to be strictly conserved in the structural polypeptides of flaviviruses and dengue is no exception to this. All cysteines in prM (M) and E, as well as in NS1, are conserved in all flaviviruses examined to date. Conversely, only one cysteine of the four in DEN 2 NS3 is conserved and there is little or no conservation of cysteine in ns2a, 2b, 4a, 4b. In NS5 the majority of the cysteines are once again conserved in the four flaviviruses examined in this region.

Comparison of dengue 2 S1 strain with other dengue viruses

The sequence of the structural protein region of a DEN 2 strain (an isolate from Jamaica) has recently been obtained by Deubel *et al.* (1986) and the sequence of the structural protein domain of a strain of DEN 4 RNA has been reported by Zhao *et al.* (1986). In addition, partial sequence of the prototype New Guinea C strain of DEN 2 in the nonstructural region has been published by Yaegashi *et al.* (1986). Within the structural protein region there are 29 amino acid changes between the two DEN 2 strains (96.3% homology; Table 1). There are 211 nucleotide differences between the two strains, mostly third-codon position silent changes, so that the nucleotide sequence conservation is 91%. Thus, the two strains of DEN 2 are quite closely related as would be expected, but nucleotide sequence divergence is more pronounced than amino acid sequence divergence.

In the nonstructural protein region, DEN 2 (S1) and DEN 2 (New Guinea) demonstrate 97.6% amino acid sequence homology in the regions sequenced to date, but only 91% nucleotide sequence homology. Although the structural and nonstructural domains diverge at somewhat different rates, the close correspondence in nucleotide sequence divergence as well as amino acid sequence divergence among the three DEN 2 strains suggests that they separated from a common ancestor at approximately the same time in the past.

TABLE 1

HOMOLOGIES AMONG DENGUE VIRUSES

Structural		DEN 2 (PR) vs DEN 2 (JAM)				DEN 2 (PR) vs DEN 4			
		Amino acid divergence		Nucleotide divergence		Amino acid divergence		Nucleotide divergence	
Protein	Amino acid no.[a]	Δ[b]	%	Δ[c]	%	Δ[b]	%	Δ[c]	%
C	114	4	3.5	30	8.8	37	32.4	116	33.9
pr[d]	91	7	7.7	30	11.0	26	28.6	78	28.6
M	75	3	4.0	19	8.4	27	36.0	81	36.0
E	495	15	3.0	132	8.9	187	37.8	417	28.0
Total	775	29	3.7	211	9.1	277	35.7	692	29.8

Nonstructural		DEN 2 (PR) vs DEN 2 (NG)			
		Amino acid divergence		Nucleotide divergence	
Protein	Amino acid no.	Δ[b]	%	Δ[c]	%
NS1[e]	(..145)	3	2.1	38	8.7
ns2a	157	3	1.9	42	8.9
ns2b	130	3	2.3	43	11.0
NS3	(453..)	13	2.9	116	8.5
ns4a	(..261)	6	2.3	68	8.7
ns4b	112	3	2.7	28	8.3
NS5	(269..)	6	2.2	65	8.0
Total	1527	37	2.4	400	8.7

[a] The number of amino acids in the protein in question.
[b] The number of changed amino acids.
[c] The number of changed nucleotides.
[d] pr is that portion of prM which is not found in mature M protein (i.e., amino acids 115 to 205 of the open reading frame.)
[e] The complete sequence of the nonstructural proteins of dengue 2 (New Guinea C) is not known. (..145) indicates that the carboxy terminal 145 amino acids of NS1 have been compared to the dengue 2 (PR) sequence. (453..) indicates that the amino terminal 453 amino acids of NS3 are being compared.

DEN 2 is more closely related to DEN 4 than it is to either YF or WN virus. DEN 2 and DEN 4 belong to the same serological subgroup and exhibit 64% homology in the structural region (Table 1). The envelope proteins of MVE, SLE, and WN viruses, three members of a different serological subgroup, share 72–78% amino acid sequence homology (Trent *et al.*, 1987) in contrast to the 62% homology between the envelope proteins of DEN 2 and DEN 4 (Table 1). A comparison of the amino acid sequences of the various dengue strains throughout the structural protein region is shown in Fig. 3.

Comparison of dengue 2 with flaviviruses in other subgroups

With the complete sequence of DEN 2 RNA, there now exists complete nucleotide sequences for viruses representing three of the major serological subgroups of the mosquito-borne flaviviruses. In Fig. 4 the amino acid sequence of the DEN 2 polyprotein is compared with that of YF and of WN viruses, representing two other serological subgroups. In the top three panels are shown pairwise comparisons between the viruses in which the amino acid sequence homology is plotted as a moving average using a window of 20 amino acids. The plotted values vary from almost 100%, i.e., complete conservation of a 20 amino acid stretch, to virtually 0%. To construct this figure the amino acid sequences have been aligned using the minimum number of gaps necessary to keep the sequences in register.

The level of amino acid sequence conservation varies among the different proteins. The nonstructural proteins ns2a and ns2b show little sequence conservation, ns4a and ns4b show intermediate sequence conservation, and NS3 and NS5 show a high degree of

TABLE 2

Structural proteins

PR[a]	JAM[b]	No.[c]	PR	JAM	No.	PR	JAM	No.
G →	A	42	G →	C	2	C →	G	2
A →	G	39	G →	U	4	C →	A	4
C →	U	59	A →	C	1	U →	G	1
U →	C	48	A →	U	7	U →	A	2
			Pu →	Py	14	Py →	Pu	9
Transitions		188	Transversions			23		

Nonstructural proteins

PR[a]	NG[d]	No.[c]	PR	NG	No.	PR	NG	No.
G →	A	70	G →	C	4	C →	G	1
A →	G	66	G →	U	3	C →	A	8
C →	U	119	A →	C	6	U →	G	3
U →	C	103	A →	U	6	U →	A	10
			Pu →	Py	19	Py →	Pu	22
Transitions		358	Transversions			41		

[a] PR is the dengue 2 (PR159-S1) strain.

[b] JAM is the Jamaica strain (Deubel *et al.*, 1987). The two have been compared throughout the structural protein region.

[c] No indicates the number of each type of change. Transitions are purine → purine or pyrimidine → pyrimidine changes. Transversions are purine (Pu) to pyrimidine (Py) changes and vice versa.

[d] NG is the New Guinea C strain of dengue 2. Regions compared in the nonstructural proteins are the same as in Table 1.

sequence conservation. Within the structural proteins, the capsid protein shows limited sequence similarity, prM exhibits somewhat higher sequence homology, and in E the overall homology is still greater. Within any particular protein the amino acid sequence homology varies widely. Within E there are regions of high sequence homology that presumably represent conserved domains required for flavivirus envelope function and may correspond to cross-reactive group-specific epitopes in the viral envelope protein. The domains of low sequence homology may correspond to type-specific or subgroup-specific antigens.

Below the homology plots are shown hydrophobicity profiles for each of the three viruses, calculated by the method of Kyte and Doolittle (1982). It is remarkable how similar the hydrophobicity profiles are among the various viruses. Whether regions of high sequence homology are being compared as in domains of NS5 or regions which show only limited amino acid sequence homology such as ns2, the hydrophobicity profiles are highly conserved, and moreover, every rise and fall in the hydrophobicity index appears to be mirrored in the other viruses such that these hydrophobicity profiles are virtually superimposable. Thus, in some domains of the various proteins the exact amino acid sequence appears to be essential for virus function, whereas in others considerable amino acid sequence disparity can be accommodated as long as the hydrophobicity profile is conserved. These profiles also illustrate the hydrophobic nature of the ns2 and ns4 polypeptides, and the hydrophobic domains at the C-terminal ends of C, prM, E and NS1. These C-terminal hydrophobic domains have been postulated to act as membrane-spanning anchors and as internal signal sequences that lead to the insertion of the following protein (Rice *et al.*, 1985, 1986b).

The overall homologies between YF, WN, and DEN 2 are summarized in Table 3.

Nucleotide sequence divergence versus amino acid sequence divergence

It has been noted before that codon usage for conserved amino acids has essentially been randomized between flaviviruses belonging to different serological subgroups (Rice *et al.*, 1986b). Comparison of DEN 2 with YF or WN viruses confirms this general rule. An example of such divergence is shown in Fig. 5 in which a conserved amino acid sequence in the envelope proteins is shown together with the nucleotide se-

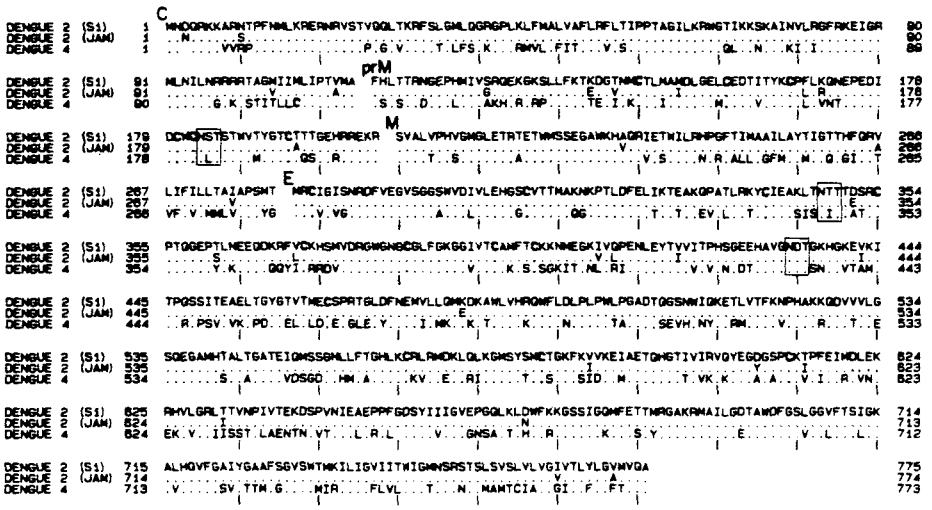


FIG. 3. Aligned deduced amino acid sequences for DEN 2 (PR159/S1), DEN 2 (Jamaica) (Deubel *et al.*, 1986), and DEN 4 (Zhao *et al.*, 1986) in the region encoding the structural proteins. Cysteine residues are highlighted by a dotted overlay and conserved glycosylation sites are boxed.

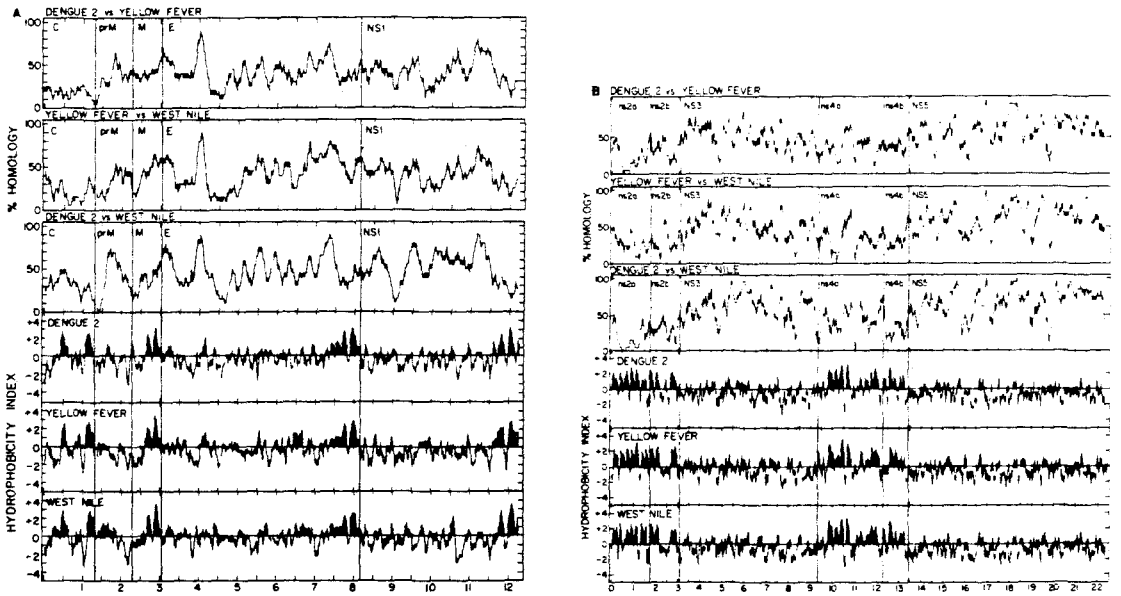


FIG. 4. Homology between DEN 2, YF, and WN viruses. The top three panels show the homology, in pairwise combinations, between the deduced amino acid sequences of DEN 2 (this paper) and YF (Rice *et al.*, 1985) and WN (Castle *et al.*, 1985, 1986; Wengler *et al.*, 1985) plotted as a moving average with a window of 20 amino acids. Below are shown individual hydrophobicity plots for the three viruses using the algorithm of Kyte and Doolittle (1982) and a string length of nine. Hydrophobic domains (filled outlines) are above the midline. The abscissa gives the amino acid position in the aligned files in hundreds, including gaps introduced for alignment; because of these gaps, the numbering does not correspond exactly to that in Fig. 2 and is included only for scale. (A) The structural proteins and NS1. (B) The nonstructural proteins.

TABLE 3

HOMOLOGY BETWEEN DENGUE 2 AND OTHER FLAVIVIRUSES

Protein		Yellow fever			West Nile		
		Identical aa	Positions[a]	% Homology[b]	Identical aa	Positions[a]	% Homology[b]
C	(114aa)	24	135	17.8	39	125	31.2
pr[c]	(91aa)	30	95	31.5	38	94	40.4
M	(75aa)	28	75	37.3	25	75	33.3
E	(495aa)	223	515	43.3	243	511	47.5
NS1	(413aa)	172	421	40.8	213	414	51.4
ns2a	(157aa)	34	170	20.0	29	171	16.9
ns2b	(130aa)	44	136	32.3	36	131	27.5
NS3	(615aa)	321	624	51.4	379	618	61.3
ns4a	(286aa)	104	291	35.7	117	293	39.9
ns4b	(112aa)	41	114	36.0	36	115	31.3
NS5	(900aa)	536	907	59.1	597	908	65.7
Total	(3388aa)	1557	3483	44.7	1752	3455	50.7

[a] Number of positions is number of amino acids plus gaps for alignment.
[b] Percentage homology is number of identical residues/total number of positions.
[c] pr is that portion of prM which is not present in mature M protein.

quence that encodes it. As this figure illustrates, codon usage is (for conserved amino acids) essentially randomized not only among viruses belonging to different subgroups, but also among different viruses belonging to the same subgroup (compare the MVE, SLE, WN sequences). In contrast, as noted above, strains of the same virus, represented by the different DEN 2 strains compared, have not yet randomized codon usage although the nucleotide sequence divergence exceeds the amino acid sequence divergence (Table 1). As a specific example, comparison of DEN 2 (PR159/S1) and DEN 2 (Jamaica) in the structural region reveals 29 coding changes and 184 silent changes in the codons for 775 amino acids. Of the 746 conserved amino acids, 25% are encoded by different codons, establishing a clear trend toward randomization but still far from random (~65% of conserved amino acids encoded by different codons).

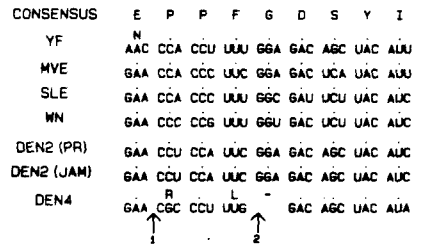


FIG. 5. Codon divergence in flaviviruses. A stretch of translated nucleotide sequence encoding a conserved domain of the E protein is shown for several flaviviruses. A dot over the codon indicates that the amino acid encoded is that shown in the consensus sequence in the first line, if the amino acid is not that of the consensus, it is shown above the codon encoding it. Note the use of different codons to encode the same amino acid. Arrows 1 and 2 indicate positions in the DEN 4 sequence and are discussed in the text. Data for MVE are from Dalgarno *et al.* (1986), for SLE from Trent *et al.* (1987), and for the other viruses as in Figs. 3 and 4.

Figure 5 also illustrates an interesting gap in the DEN 4 sequence. The two DEN 2 structural protein sequences align perfectly with no gaps required to keep the sequences in register (775 total amino acids). However, two gaps are required to align the DEN 4 sequence with the DEN 2 sequence. One gap occurs at amino acid 3 of the polyprotein (second amino acid of the mature capsid protein). The second gap occurs within the sequence illustrated in Fig. 5. In this domain there are 50 nucleotides perfectly conserved between the two DEN 2 strains, and as indicated the amino acid sequence exhibits a high degree of conservation among all flaviviruses. Insertion of a C at the first arrow in the figure and of GN (N being any nucleotide) at the second arrow would make the DEN 4 amino acid sequence identical to the flavivirus consensus sequence, with no gaps. Because only one clone was sequenced in the case of DEN 4 strain, it is unclear whether the gaps are representative of the population average (they could conceivably have arisen from cloning artifacts), but if so it suggests that nucleotide deletions leading to shifts in reading frame may contribute to strain divergence (see also Vrati *et al.*, 1986).

Since codon usage divergence among flaviviruses is so pronounced, the presence of certain conserved nucleotide sequence elements is particularly striking.

There are nucleotide sequences in the 5′ terminal region (in the coding region for the capsid protein) and in the 3′ terminal region (in the 3′ untranslated sequence) that are invariant among flaviviruses examined to date and that may be involved in cyclization of the RNA (Hahn *et al.*, 1987b). These sequences are presumed to be important for viral replication. In addition, the 3′ terminal 79 nucleotides of DEN 2 RNA are capable of forming a hairpin structure (Hahn *et al.*, 1987b), as has been observed for other flaviviruses (Rice *et al.*, 1985; Brinton *et al.*, 1986; Wengler and Castle, 1986; Zhao *et al.*, 1986; Takegami *et al.*, 1986).

DISCUSSION

The dengue viruses represent a world-wide health problem of increasing dimensions. The sequence presented here represents a step forward in understanding the molecular biology of the replication of these viruses and makes possible various expression experiments that may lead to the development of vaccines against dengue. Because of the difficulty in obtaining attenuated strains suitable to use as vaccines, the lack of an animal model to study the disease, and in part because of the complications of dengue shock syndrome, vaccine development in the dengue viruses has been slow and we need to understand in greater detail the genome organization, translation strategy, and molecular biology of replication of these viruses. In addition, these and related studies on other flaviviruses or other strains of dengue highlight the sequence variability from virus to virus, which has direct implications for vaccine development.

The geographic origin of new epidemic strains and the amount of genetic and antigenic drift that these strains undergo has been examined in several ways, including T1 oligonucleotide mapping (Trent *et al.*, 1983), antigen signature analysis (Monath *et al.*, 1986), and use of hybridization probes consisting of synthetic oligonucleotides of defined sequences (Kerschner *et al.*, 1986). From these studies it has been shown that at least six groups or topotypes of DEN 2 exist, which differ from one another by roughly 10% nucleotide sequence divergence. These analyses also suggest that new strains of DEN 2 are introduced at times into an area; for example, Trent *et al.* (1983) hypothesize that the DEN 2 Jamaican strain may have been introduced into the Caribbean from West Africa. By these criteria, the three isolates of DEN 2 for which nucleotide sequence data exist, viz., the S1 strain derived from the PR 159 (isolated in 1959 in Puerto Rico), the New Guinea C prototype strain (isolated in 1944 in New Guinea), or the Jamaica strain (isolated in 1983 in Jamaica), belong to different topotypes. Comparison of the nucleotide and deduced amino acid sequences (Table 1) showed about 9% nucleotide sequence divergence among the DEN 2 strains, but amino acid divergence of only 3.7% among the structural proteins and 2.4% among the nonstructural proteins. This is two to four times the amino acid sequence divergence, but 20 times the nucleotide sequence divergence, found between Asibi YF and 17D YF after 240 passages *in vitro* (Hahn *et al.*, 1987a). Thus the various dengue strains are separated by the equivalent of 5000 *in vitro* passages. What this corresponds to in terms of passages in nature cannot be calculated accurately from our current knowledge, but crude estimates suggest that it exceeds 1000 man–mosquito–man cycles, which would require more than 35 years of continuous passage from the time of divergence of any two strains. It is also of interest that in the natural environment the ratio of silent nucleotide changes to amino acid changes is much higher than following *in vitro* passage, suggesting that some of the amino acid substitutions found after *in vitro* passage were positively selected.

In Table 2 the distribution of transversions (purine to pyrimidine changes and vice versa) and transitions is shown. Transitions outnumber transversions by more than 8:1. The predominance of transitions is presumably due to the fact that most of the changes occur during replication by misincorporation of the wrong purine or the wrong pyrimidine; in addition, transitions in the third codon position are less likely to lead to a change in coding assignment than are transversions and thus are more likely to survive selection pressures.

The sequences obtained here make clear that all the flaviviruses are closely related and have evolved from a common ancestor. There are conserved domains within the protein sequence that are clearly essential for virus replication. In other domains, the linear amino acid sequence is less important than the hydrophobicity profile of the protein. It is to be hoped that variable and conserved regions of the proteins can ultimately be correlated with specific protein epitopes important for interaction with the immune system or with other biological properties, which would aid in designing expression experiments and could help in developing second-generation dengue vaccines.

ACKNOWLEDGMENTS

We are grateful to E. M. Lenches for excellent technical assistance, to C. S. Hahn for helpful discussions, and to P. Wright and D. Trent for furnishing us with results prior to publication. This work was supported by Grants AI-20612 and AI-10793 from NIH, Grant DMB 86-17372 from NSF, by a contract from the U. S. Army Medical Development Command, and by a grant from the World Health Organization. R.G. was supported by a fellowship from the National Research Council of Brazil (CNPq).

REFERENCES

BANCROFT, W. H., SCOTT, R. M., ECKELS, K. H., HOKE, C. H., SIMMS, T. E., JESRANI, K. D. T., SUMMERS, P. L., DUBOIS, D. R., TSOULOS, D., and RUSSELL, P. K. (1984). Dengue virus type 2 vaccine: Reactogenicity and immunogenicity in soldiers. *J. Infect. Dis.* **149,** 1005–1010.

BANCROFT, W. H., TOP, F. H., ECKELS, K. H., ANDERSON, J. H., MCCOWN, J. M., and RUSSELL, P. K. (1981). Dengue-2 vaccine: Virological, immunological, and clinical responses of six yellow fever-immune recipients. *Infect. Immun.* **31,** 698–703.

BELL, J. R., KINNEY, R. M., TRENT, D. W., LENCHES, E. M., DALGARNO, L., and STRAUSS, J. H. (1985). Amino-terminal amino acid sequences of structural proteins of three flaviviruses. *Virology* **143,** 224–229.

BIEDRZYCKA, A., CAUCHI, M. R., BARTHOLOMEUSZ, A., GORMAN, J. J., and WRIGHT, P. J. (1987). Characterization of protease cleavage sites involved in the formation of envelope glycoprotein and three non-structural proteins of dengue virus type 2, New Guinea C strain. *J. Gen. Virol.* **68,** 1317–1326.

BIRD, A. P. (1980). DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res.* **8,** 1499–1504.

BRÈS, P. (1979). Historical review of dengue-1: Implications of its introduction in the Western Hemisphere in 1977. *In* "Dengue in the Carribean, 1977," pp. 4–18. Pan American Health Organization Scientific Pub. No. 375, Washington, DC.

BRINTON, M. A., FERNANDEZ, A. V., and DISPOTO, J. H. (1986). Sequence analysis of the 3′ terminus of West Nile virus, strain E101, genome RNA. *Virology* **153,** 113–121.

CAREY, D. E. (1971). Chikungunya and dengue: A case of mistaken identity? *J. Hist. Med. Allied Sci.* **26,** 243–262.

CASTLE, E., LEIDNER, U., NOWAK, T., WENGLER, G., and WENGLER, G. (1986). Primary structure of the West Nile flavivirus genome region coding for all nonstructural proteins. *Virology* **149,** 10–26.

CASTLE, E., NOWAK, T., LEIDNER, U., WENGLER, G., and WENGLER, G. (1985). Sequence analysis of the viral core protein and the membrane-associated proteins VI and NV2 of the flavivirus West Nile virus and of the genome sequence for these proteins. *Virology* **145,** 227–236.

DALGARNO, L., TRENT, D. W., STRAUSS, J. H., and RICE, C. M. (1986). Partial nucleotide sequence of Murray Valley encephalitis virus: Comparison of the encoded polypeptides with yellow fever virus structural and nonstructural proteins. *J. Mol. Biol.* **187,** 309–323.

DEUBEL, V., KINNEY, R. M., and TRENT, D. W. (1986). Nucleotide sequence and deduced amino acid sequence of the structural proteins of dengue type 2 virus, Jamaica genotype. *Virology* **155,** 365–377.

ECKELS, K. H., BRANDT, W. E., HARRISON, V. R., MCCOWN, J. M., and RUSSELL, P. K. (1976). Isolation of a temperature-sensitive dengue-2 virus under conditions suitable for vaccine development. *Infect. Immun.* **14,** 1221–1227.

ECKELS, K. H., HARRISON, V. R., SUMMERS, P. L., and RUSSELL, P. K. (1980). Dengue-2 vaccine: Preparation from a small-plaque virus clone. *Infect. Immun.* **27,** 175–180.

EHRENKRANZ, N., VENTURA, A., CUADRADO, R., POND, W., and PORTER, J. (1971). Pandemic dengue in Caribbean countries and the southern United States—Past, present, and potential problems. *N. Engl. J. Med.* **285,** 1460–1469.

FREDERIKSEN, H. (1955). Historical evidence for interference between dengue and yellow fever. *Amer. J. Trop. Med. Hyg.* **4,** 483–491.

GROOT, H., MORALES, A., ROMERO, M., VIDALES, H., LESMES-DONALDSON, C., MARQUEZ, G., ESCOBAR DE CALVACHE, D., and SAENZ, M. (1979). Recent outbreaks of dengue in Colombia. *In* "Dengue in the Caribbean, 1977," pp. 4–18. Pan American Health Organization Scientific Pub. No. 375.

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. Natl. 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.*, in press.

HALSTEAD, S. (1981). The pathogenesis of dengue: Molecular epidemiology in infectious disease. *Amer. J. Trop. Med. Hyg.* **114,** 632–648.

KERSCHNER, J. H., VORNDAM, A. V., MONATH, T. P., and TRENT, D. W. (1986). Genetic and epidemiological studies of dengue type 2 viruses by hybridization using synthetic deoxyoligonucleotides as probes. *J. Gen. Virol.* **67,** 2645–2661.

KYTE, J., and DOOLITTLE, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157,** 105–132.

LINDQVIST, B. H., DI SALVO, 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.

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, pp. 499–560. Academic Press, New York.

MONATH, T. P. (1986). Pathology of flaviviruses. *In* "The Togaviridae and Flaviviridae" (S. Schlesinger and M. Schlesinger, Eds.), Chap. 12, pp. 375–440. Plenum, New York.

MONATH, T. P., WANDS, J. R., HILL, L. J., BROWN, N. V., MARCINIAK, R. A., WONG, M. A., GENTRY, M. K., BURKE, D. S., GRANT, J. A., and TRENT, D. W. (1986). Geographic classification of dengue-2 strains by antigen signature analysis. *Virology* **154,** 313–324.

OKAYAMA, H., and BERG, P. (1982). High efficiency cloning of full length cDNA. *Mol. Cell. Biol.* **2,** 161–170.

PERLMAN, D., and HALVORSON, H. O. (1983). A putative signal peptidase recognition site and sequence in eukaryotic and prokaryotic signal peptides. *J. Mol. Biol.* **167,** 391–409.

REPIK, P. M., DALRYMPLE, J. M., BRANDT, W. E., MCCOWN, J. M., and RUSSELL, P. K. (1983). RNA fingerprinting as a method for distinguishing dengue-1 virus strains. *Amer. J. Trop. Med. Hyg.* **32,** 577–589.

RICE, C. M., AEBERSOLD, R., TEPLOW, D. B., PATA, J., BELL, J. R., VORNDAM, A. V., TRENT, D. W., BRANDEIS, M. W., SCHLESINGER, J. J., and STRAUSS, J. H. (1986a). Partial N-terminal amino acid sequences of three nonstructural proteins of two flaviviruses. *Virology* **151,** 1–9.

RICE, C. M., DALGARNO, L., GALLER, R., HAHN, Y. S., STRAUSS, E. G., and STRAUSS, J. H. (1987). Molecular cloning of flavivirus genomes for comparative analysis and expression. *In* "Modern Trends in Virology" (H. Bauer, H. D. Klenk, and C. Scholtissek, Eds.), Proceedings of the International Symposium, June 1986, Giessen, Springer-Verlag, Berlin, in press.

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., STRAUSS, E. G., and STRAUSS, J. H. (1986b). Structure of the flavivirus genome. *In* "Togaviridae and the Flaviviradae" (M. Schlesinger and S. Schlesinger, Eds.), Chap. 10, pp. 279–326. Plenum, New York.

ROSEN, L. (1986). The pathogenesis of dengue haemorrhagic fever, Supplement to *South Amer. J. Med.* **11,** 40–42.

SABIN, A. B. (1952). Research on dengue during World War II. *Amer. J. Trop. Med. Hyg.* **1,** 30–50.

SCHATZMAYER, H. G., NOGUEIRA, R. M. R., and TRAVASSOS DA ROSA, A. P. A. (1986). An outbreak of dengue virus at Rio de Janeiro (1986). *Mem. Inst. Oswaldo Cruz* **81**, 245–246.

SCOTT, R. M., ECKELS, K. H., BANCROFT, W. H., SUMMERS, P. L., MCCOWN, J. M., ANDERSON, J. H., and RUSSELL, P. K. (1983). Dengue 2 vaccine: Dose response in volunteers in relation to yellow fever status. *J. Infect. Dis.* **148**, 1055–1060.

SMITH, V. R., and CALVO, J. M. (1980). Nucleotide sequence of the *E. coli* gene for coding for dihydrofolate reductase. *Nucleic Acids Res.* **8**, 2255–2274.

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

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., and ANDERSON, C. R. (1975). The relative resistance of dengue-immune monkeys to yellow fever virus. *Amer. J. Trop. Med. Hyg.* **24**, 115–117.

TRENT, D. W., GRANT, J. A., ROSEN, L., and MONATH, T. P. (1983). Genetic variation among dengue-2 viruses of different geographic origin. *Virology* **128**, 271–285.

TRENT, D. W., KINNEY, R. M., JOHNSON, B. J., VORNDAM, A. V., GRANT, J. A., DEUBEL, V., RICE, C. M., and HAHN, C. (1987). Partial nucleotide sequence of St. Louis encephalitis virus RNA: Structural proteins, NS1, ns2a, and ns2b. *Virology* **156**, 293–304.

VON HEIJNE, G. (1986). A new method for predicting signal sequence cleavage sites. *Nucleic Acids Res.* **14**, 4683–4690.

VRATI, S., FARAGHER, S. G., WEIR, R. C., and DALGARNO, L. (1986). Ross River virus mutant with a deletion in the E2 gene: Properties of the virion, virus-specific macromolecule synthesis, and attenuation of virulence for mice. *Virology* **151**, 222–232.

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.

WENGLER, G., CASTLE, E., LEIDNER, U., NOWAK, T., and WENGLER, G. (1985). Sequence analysis of the membrane protein V3 of the flavivirus West Nile virus and of its gene. *Virology* **147**, 264–274.

WRIGHT, P. J. (1982). Envelope protein of the flavivirus Kunjin is apparently not glycosylated. *J. Gen. Virol.* **59**, 29–38.

YAEGASHI, T., VAKHARIA, V. N., PAGE, K., SASAGURI, Y., FEIGHNY, R., and PADMANABHAN, R. (1986). Partial sequence analysis of cloned dengue virus type 2 genome. *Gene* **46**, 257–267.

ZHAO, B., MACKOW, E., BUCKLER-WHITE, A., MARKOFF, L., CHANNOCK, 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.

Chapter 4

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, Australia 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.

0022-2836/87/210033-09 $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

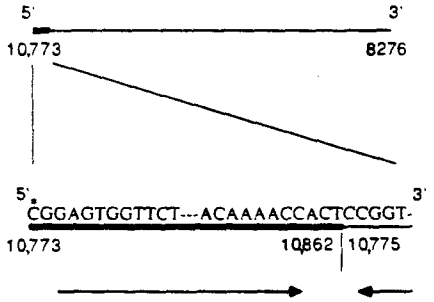


Figure 1. Structure of clone 3'-1. A schematic diagram of the structure of this clone in which the 3'-terminal sequences as plus strand (thick line) are attached to minus strand sequences (thin line) is shown. Nucleotide sequences around the joint between plus and minus strand sequence and at the end 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

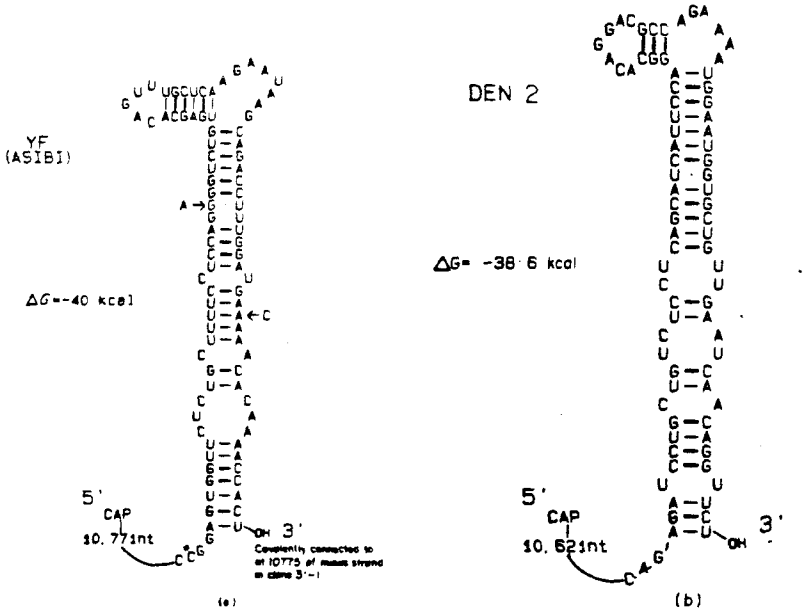


Figure 2. Putative secondary structure found at the 3' end of Asibi YF RNA and of DEN 2 RNA. (a) A possible secondary structure at the 3' terminus of Asibi YF RNA is shown. This structure is virtually identical to that proposed by Rice *et al.* (1985) for the 3' terminus of 17D YF RNA with 2 nucleotide 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.

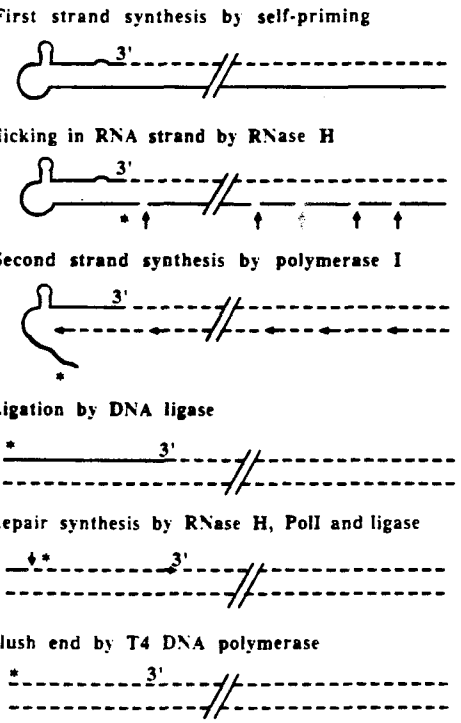


Figure 3. Model for the origin of clone 3'-1. The first step describes first strand cDNA (broken line) synthesis in which self-priming occurs onto the 3' terminus of the genomic RNA (continuous line). The next 3 steps describe second strand cDNA synthesis following nicking of the RNA strand with *E. 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.

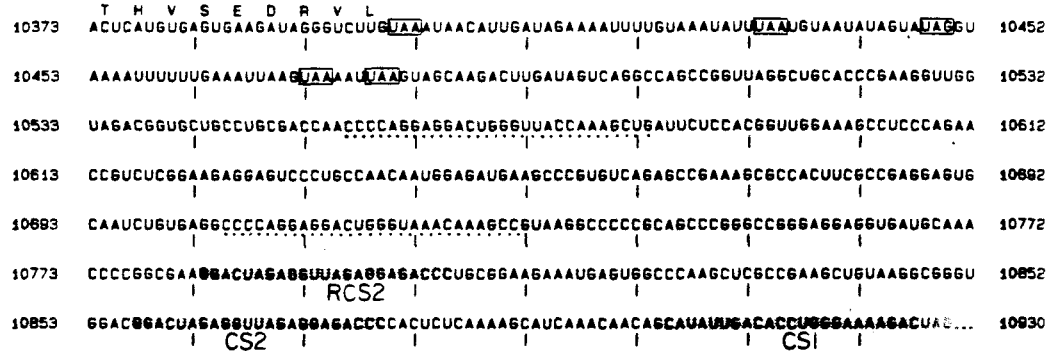


Figure 4. Nucleotide sequence of the 3′ untranslated region of MVE RNA. The last few nucleotides of the long open reading frame are shown as translated sequence. The entire untranslated region is shown except for the putative 3′-terminal secondary structure which has yet to be determined. The first few 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.

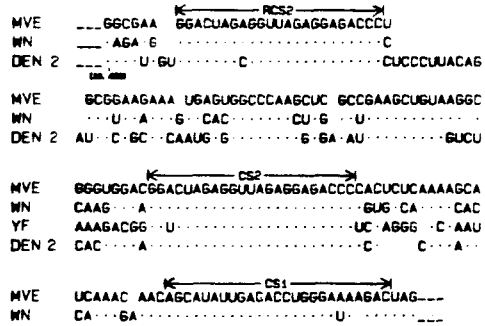


Figure 5. Comparison of conserved and repeated sequences in the 3′ untranslated region of flavivirus RNAs. Conserved sequences found in DEN 2 RNA, in MVE RNA (Fig. 4), in YF RNA (Rice *et al.*, 1985), and in WN RNA (Wengler & Castle, 1986) are compared. There are 2 such conserved nucleotide sequences (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

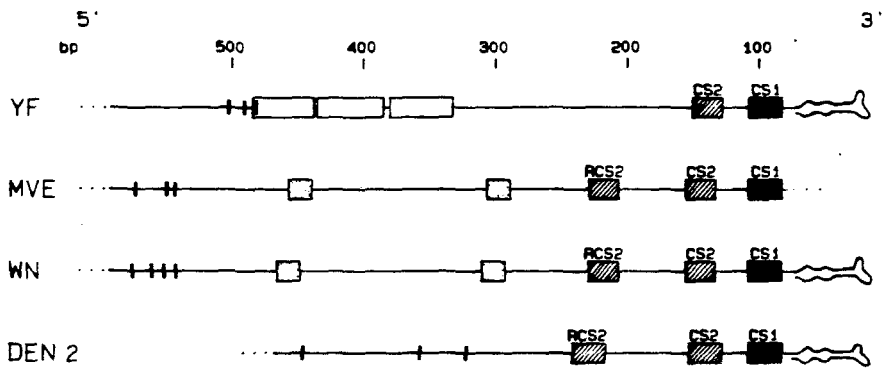


Figure 6. Schematic diagram of the organization of the 3′ untranslated region of flavivirus RNAs. The location of repeated nucleotide sequences, conserved nucleotide sequences and the 3′-terminal secondary structure are indicated for YF RNA, MVE RNA, WN RNA and DEN 2 RNA. In addition to CS1 (filled boxes) 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

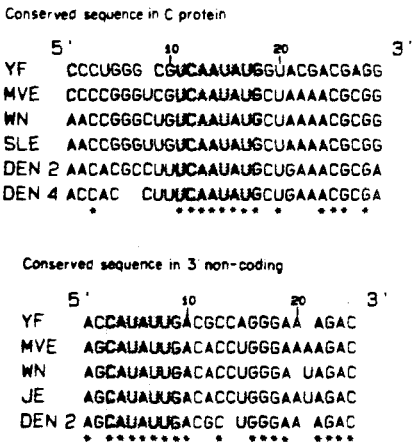


Figure 7. Conserved nucleotide sequence elements in flavivirus RNAs. In the top half of the Figure are shown nucleotide sequences from 6 flaviviruses found in the coding region for the capsid protein. This sequence begins at nucleotide 147 of YF RNA (Rice *et al.*, 1985), at 128 of MVE RNA (Dalgarno *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).

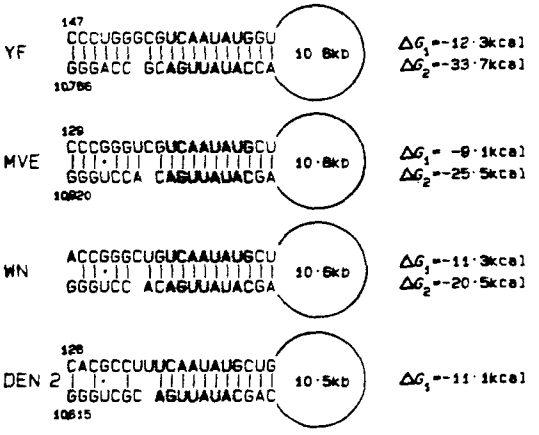


Figure 8. Possible cyclization figures for flavivirus RNAs. The complementary sequences shown in Fig. 7 are aligned for 4 flaviviruses as circular figures. The calculated thermal stability of these circles is indicated (Tinoco *et al.*, 1973). The 8-nucleotide core that is perfectly conserved is indicated 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 ΔG values are shown. ΔG_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 $-13{\cdot}5$ kcal at 25°C (Frey *et al.*, 1979). The ΔG_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

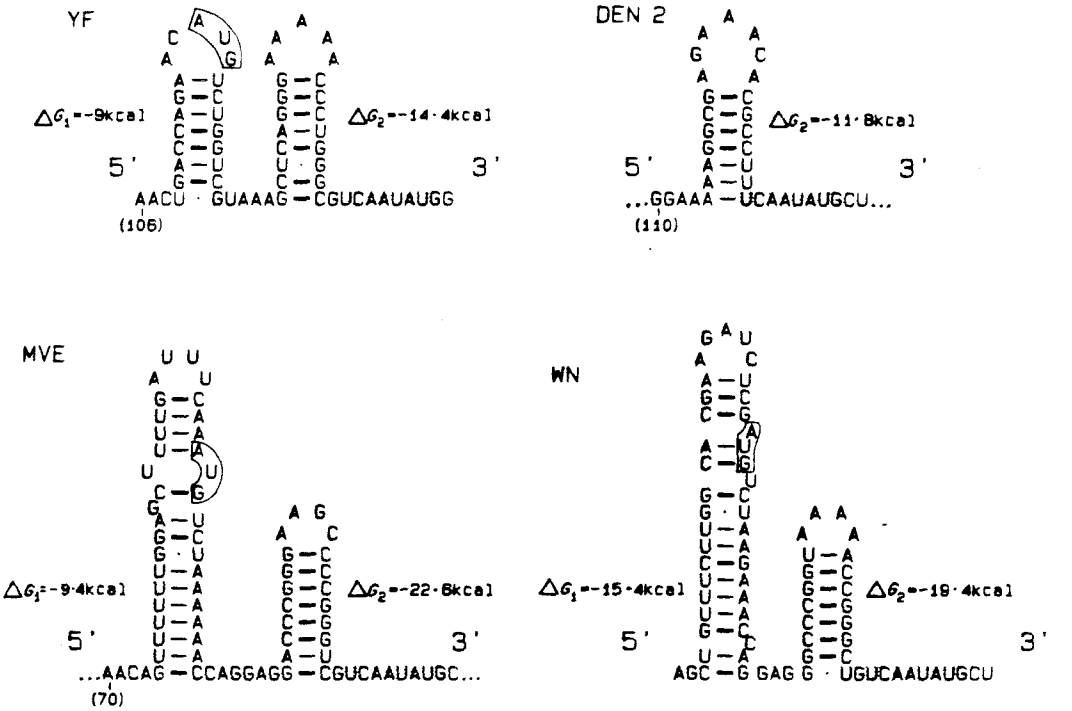


Figure 9. Possible hairpin structures in the 5′ regions of flavivirus RNAs. The AUG codon that initiates the very long open reading frame is boxed and the core conserved sequence that might be involved in cyclization is shaded. Free energies were calculated as for Fig. 2.

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 (ΔG_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, C. (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

Chapter 5

Expression of Dengue 2 Virus Structural Proteins by a Recombinant Vaccinia Virus

Expression of Dengue 2 Virus Structural Proteins by a Recombinant Vaccinia Virus

Young S. Hahn[1], Joel Dalrymple[2], Edith Lenches[1], and James H. Strauss[1]

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

[2] Virology Division, U.S. Army Medical Research Institute of Infectious Diseases, Ft. Detrick, Frederick, Maryland 21701

Abstract

We have obtained a vaccinia virus recombinant which contains cDNA sequences coding for the three structural proteins of dengue 2 virus (PR159/S1 strain) within the thymidine kinase gene of the vaccinia virus genome. The dengue 2 sequences are expressed under the control of the vaccinia 7.5K early/late promoter, and cells infected with this recombinant produced the structural proteins of dengue 2 virus. E protein and prM expressed from the recombinant appeared to be identical to those found in dengue 2 infected cells, indicating that these proteins were apparently cleaved and glycosylated in a manner similar to authentic dengue virus proteins. Mice immunized with the recombinant vaccinia virus showed a dengue specific immune response that included low levels of neutralizing antibodies.

Introduction

The four serotypes of dengue virus are members of the family Flaviviridae and are transmitted to vertebrates by *Aedes* mosquitoes (Gubler, 1988). The dengue viruses are important human disease agents causing millions of cases of dengue fever annually and have a very wide geographical distribution, including Southeast Asia, the South Pacific, the Caribbean, and Central and South America. Uncomplicated dengue fever is usually an acute illness characterized by joint pain, fever, and rash, but in a small percentage of cases complications develop leading to dengue hemorrhagic fever or dengue shock syndrome which is hypothesized to result from immune enhancement (Halstead, 1988; Monath, 1986). Since the epidemiology of dengue shock syndrome suggests that it can be caused by sequential infection with dengue viruses of different serotypes, extreme caution has been used in the development of a vaccine, for fear that vaccinees might be rendered even more susceptible. Moreover, no animal model for dengue

infection and disease exists, making it difficult to test prospective candidate vaccines in other than man. For these reasons, at the current time there is no effective vaccine against any dengue serotype.

Virions of dengue virus are relatively simple and contain only three virus coded proteins, designated capsid (C) protein, membrane-like (M) protein, and envelope (E) glycoprotein (Stollar, 1969). Intracellular virus lacks M but contains another glycoprotein prM which is cleaved to produce M (Shapiro et al., 1972; Rice et al., 1986). The E glycoprotein is the major site responsible for neutralization of infectivity by specific antibodies (Smith et al., 1970; Stevens et al., 1965). The envelope glycoprotein also exhibits hemagglutinating activity and is responsible for adsorption to the cell surface (Sweet, 1954).

Recently cDNA clones of the live attenuated vaccine strain PR159 (S1) of dengue 2 virus have been obtained and the complete sequence of the genome determined (Hahn et al., 1988). The dengue 2 virus genome consists of a single molecule of positive-strand RNA, 10712 nucleotides in length having a 5' cap but lacking a poly(A) tail at the 3' end. It codes for a polyprotein which is cleaved to generate the individual viral proteins (Cleaves, 1985; Rice et al., 1986). The three structural proteins, C-prM(M)-E, are encoded at 5' end of genome, whereas nonstructural proteins NS1-NS2a,2b-NS3-NS4a,4b-NS5 are encoded at 3' end of genome. Here we report the production of dengue 2 structural proteins via recombinant vaccinia virus and the immunological response to this vaccinia recombinant.

Materials and Methods

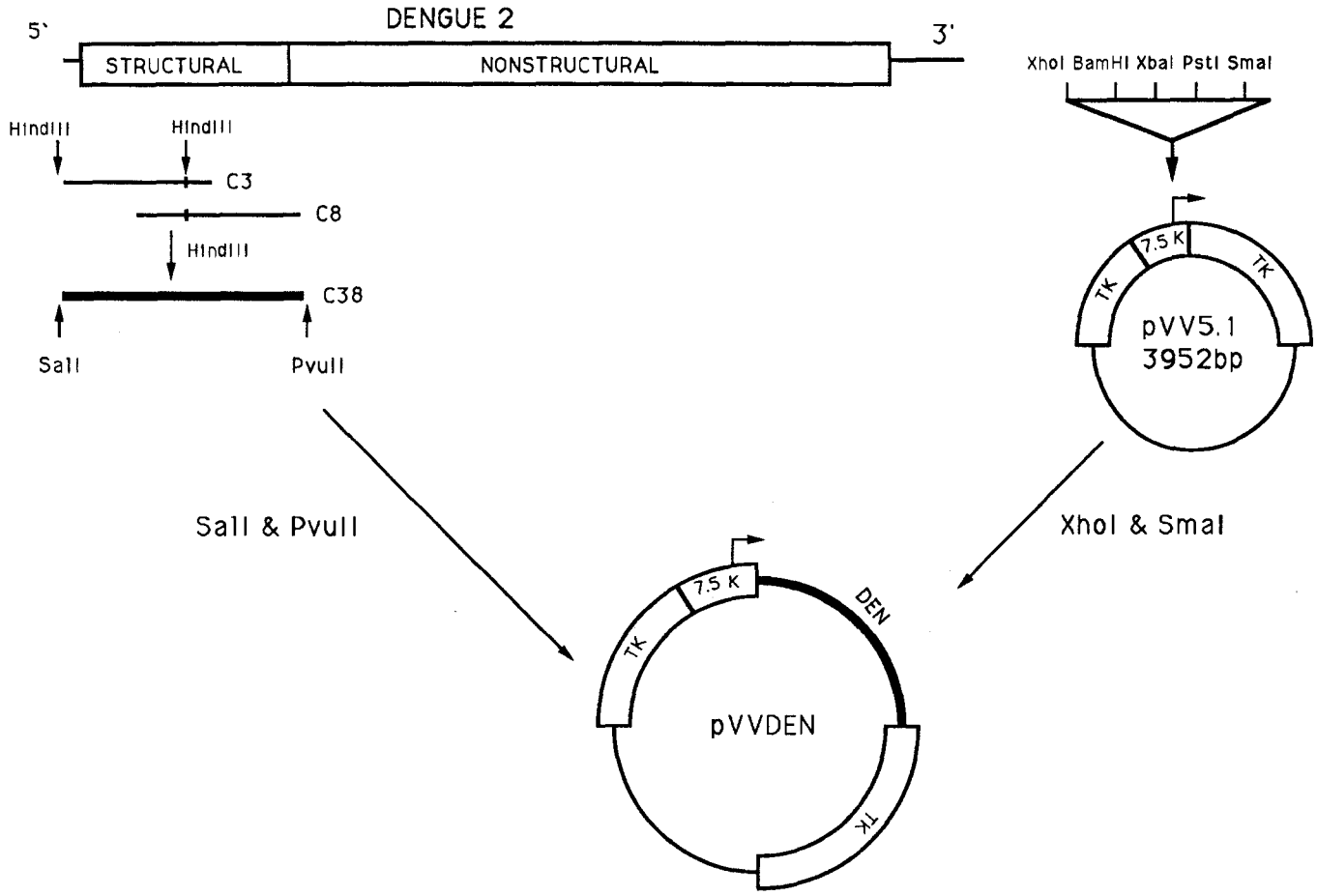
Cells and virus

Vaccinia virus (WR strain) was propagated and titered as previously described (Hruby et al., 1979). Dengue 2 virus (S1 strain, PR159 isolate) was grown and titered as previously described (Eckels et al., 1976, 1980; Hahn et al., 1988). BSC40 cells, LTK^- cells, MDBK cells, CHO cells, and BHK-21 cells (American Type Culture collection) were maintained at 37°C under 5% CO_2 in Eagle minimum essential medium supplemented with 10% heat-inactivated fetal calf serum, 2mM L-glutamine, and 10 μg of gentamicin sulfate per ml. Virus infections were carried out as previously described (Hruby et al., 1979).

Construction of chimeric plasmids

Plasmids were constructed, propagated and purified using standard techniques (Maniatis et al., 1982). Dengue 2 cDNA clones C3 and C8 (Hahn et al., 1988) were joined to obtain a plasmid encoding the structural proteins of dengue 2 virus. C3 was digested with HindIII (one site exists in the pGEM1 vector and the other site is in dengue genome) and ligated with HindIII digested plasmid C8 to produce the plasmid designated as C38. Then the SalI-PvuII fragment (both sites reside in pGEM1 vector) was excised from plasmid C38 and was inserted into plasmid pVV5.1 which had been digested with XhoI and SmaI, to yield pVVDEN (Figure 1). The pVV5.1 plasmid was a derivative of pVV2, a plasmid derived from pBR322 which contained the vaccinia thymidine kinase (TK) gene interrupted with the 7.5Kd vaccinia early promoter (Rice et al., 1985). In pVV5.1 the promoter has been truncated to improve expression efficiency and a polylinker inserted at the deletion site. The construction of pVV5.1 is described briefly below and in more detail in the Appendix. pVV2 was linearized at the BamHI site and Bal31 nuclease

Figure 1. Construction of the vaccinia-dengue 2 recombinant plasmid. The SalI-PvuII fragment containing the genome region encoding the structural proteins C, prM, and E protein of dengue 2 virus was inserted into plasmid pVV5.1 immediately downstream of the vaccinia 7.5Kd promoter. In the recombinant plasmid, the transcription unit is flanked by thymidine kinase sequences. The 3' end sequence of the fusion protein is also shown at the bottom of the figure, using the single letter amino acid code (X = stop codon).



Amino acid sequences of 3' end of fusion protein

E NS1 pGEM1 TK-COOH

LGVMVQA..DSGCVVSWKNKELKCGSGIFVTDNVHTWTGIPVSLXXVVLISISQGNSVSVWQTKEKXLXX

digestion was used to trim back the 3' end of the promoter by 36 nucleotides from the transcription start site and the plasmid reclosed with XhoI linkers. The PstI-XhoI fragments containing the truncated promoter was excised from this plasmid and it and a XhoI-BglII polylinker was ligated to pVV4 which had been digested with PstI-BamHI. (pVV4 in turn was a second pVV2 derivative in which the XhoI and XbaI sites had been destroyed, and from which the HindIII-XbaI fragment had been removed.)

Isolation of a recombinant vaccinia virus

Recombinants were constructed using the marker rescue technique as described previously (Rice et al., 1985). Plasmid pVVDEN DNA was coprecipitated with wild type vaccinia DNA and calf thymus DNA with calcium phosphate (Graham and Van der Eb, 1973). The DNA precipitates were added to LTK^- cells (without bromodeoxyuridine) which had been infected for 3h with wild type vaccinia (strain WR) at a multiplicity of 0.05 PFU/cell. Viruses were harvested and titered following incubation of transfected cells at 37°C for 72h. TK^- virus was amplified by low multiplicity passage through LTK^- cells in the presence of 25 µg/ml bromodeoxyuridine and then screened for recombinants containing the dengue-specific insert by using *in situ* plaque hybridization procedures (Villareal and Berg, 1977). A dengue specific probe of ^{32}P-labeled RNA was obtained by transcription *in vitro* with SP6 polymerase of clone C38, in which the dengue sequence encoding the structural proteins is downstream of the SP6 promoter. Recombinants containing the dengue specific insert were named VVDEN and plaque purified through at least two rounds until 100% of the plaques scored as positive with the dengue specific probe.

Protein synthesis

BSC40 cells in 60mm plates were infected with recombinant virus VVDEN or wild type vaccinia virus at a multiplicity of 10 PFU/cell and labeled with 5 to 25 μCi/ml of [^{35}S]methionine (>600 Ci/mmol; Amersham Corp.) for 15h after infection. Monolayers were washed three times with ice-cold PBS, and lysed with 400 μl of lysis solution containing 0.5% Triton X-100, 25mM Tris-Cl, pH 7.6, 0.5mM EDTA, 20 μg/ml phenylmethylsulfonylfluoride. After 20min on ice, the lysates were centrifuged in a microfuge to remove nuclei and vaccinia cores and stored at -70°C. 50 to 100 μl of cell lysates were used for immunoprecipitation (Anderson and Blobel, 1983; Hardy and Strauss, 1988). Immunoprecipitated proteins were separated on 8 to 15% linear gradient SDS-polyacrylamide gels (Laemmli, 1970). The gels were fixed for 30min, washed for 30min in distilled water and soaked in Fluoro-Hance (Research Products International) for 30min before being dried. Gels were exposed on Kodak XAR-5 film at -70°C.

Radioimmunoprecipitation

Radioimmune assays were performed as previously described (Summers et al., 1984). Vaccinia or dengue-2 infected cells and uninfected cells were lysed, diluted and dried onto polyvinylchloride 96 well plates. Aliquots of serially diluted mouse sera were added and the plates incubated overnight. After washing the plates with PBS containing 10% bovine serum, ^{125}I-labeled goat anti-mouse immunoglobulin was added for 4h and plates were washed, cut and counted in a gamma counter.

Virus neutralization assays

Equal volumes of test sera diluted in growth medium containing 2% guinea pig complement and ~100 PFU of virus diluted in growth medium were mixed and

incubated overnight at 4°C. The mixture was then added to LLC-MK2 cells in 6-well culture dishes, the virus allowed to absorb for 1h and the cells overlayed with 0.8% agarose in Eagle medium. After incubation for 7 days at 37°C, the plaques were visualized by staining the cells overnight with neutral red.

Results

Construction of a vaccinia-dengue 2 recombinant plasmid

Plasmid pVV5.1, which is derived from pGS20, a vaccinia virus insertion vector that has been used to construct a number of vaccinia recombinants (Mackett et al., 1984; Smith et al., 1983), was used for construction of a vaccinia-dengue recombinant plasmid. The SalI-PvuII fragment of clone C38, containing the 5' end of dengue 2 cDNA encoding the structural proteins C, prM, and E, was excised and inserted into plasmid VV5.1 at the XhoI-SmaI sites, immediately downstream of the vaccinia 7.5Kd promoter (Fig. 1). In the hybrid recombinant plasmid pVVDEN, the transcription unit is flanked by vaccinia virus DNA sequences containing the 5' and 3' ends of the vaccinia TK gene. The 3' end sequence of the fusion protein is shown in Fig. 1. The first stop codon is preceded by 29 amino acids of protein NS1 and 6 amino acids from the pGEM1 vector.

Isolation of a vaccinia recombinant containing dengue 2 cDNA

The hybrid plasmid, pVVDEN, was coprecipitated with wild type vaccinia DNA and calf thymus DNA by the calcium phosphate method and adsorbed to vaccinia infected LTK^- cells at 3h postinfection. Recombinant viruses which possessed a TK^- phenotype, caused by the insertional inactivation of the viral thymidine kinase gene, were amplified by a low multiplicity passage through LTK^- cells in the presence of bromodeoxyuridine. Plaque hybridization to isolate recombinants was done using a dengue specific ^{32}P-labeled RNA probe. Individual

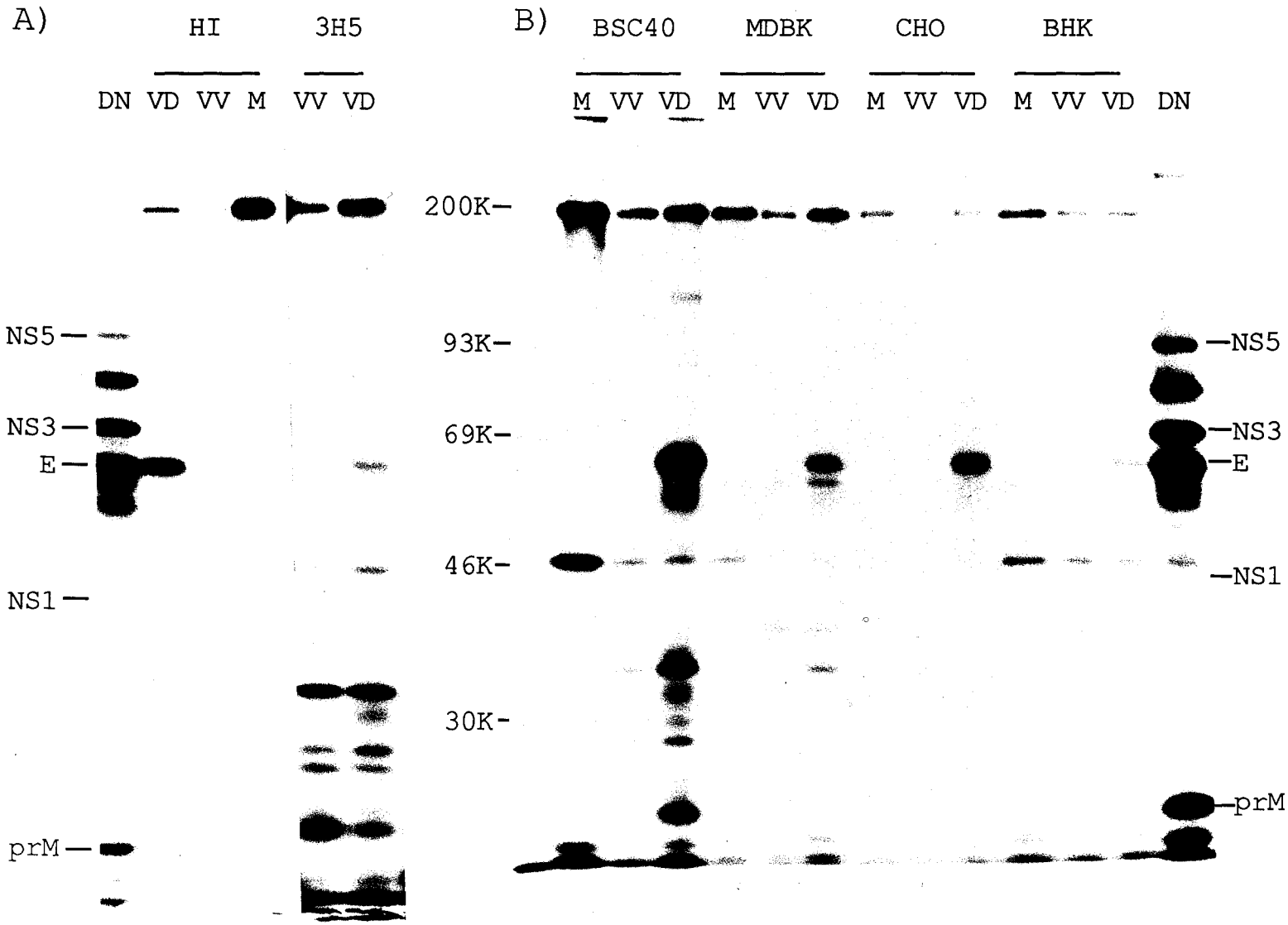
positive plaques were picked and purified through at least two additional plaque isolations until all plaques were positive in a plaque hybridization assay. One recombinant virus was selected and called VVDEN and used for all further experiments.

Expression of dengue 2 proteins

To determine what protein products were translated from the dengue specific transcripts, BSC40 cells infected with wild type vaccinia virus or vaccinia recombinant VVDEN were labeled with [^{35}S]methionine. The radioactively labeled proteins were immunoprecipitated with mouse hyperimmune antisera or with monoclonal antibodies (3H5) directed against the dengue E protein, and the precipitated proteins were separated on 8 to 15% linear gradient SDS-polyacrylamide gels (Fig. 2).

Fig. 2a shows that the polyvalent hyperimmune mouse antiserum precipitated the E glycoprotein and protein prM from cells infected with pVVDEN, which migrated with the same mobility as those found in dengue-infected cells. The relative intensity of the prM band is variable, depending on the cell line used and in the best case is still less than 20% that of E, and it is unclear if this results from more rapid turnover of prM or from incomplete precipitation of prM. A minor band of larger molecular size was precipitated nonspecifically and many minor bands of small size were also present. When an E glycoprotein-specific monoclonal antibody (3H5) was used, the E glycoprotein precipitated again migrated with the same mobility as E protein formed in dengue-infected cells. In any event, in recombinant virus-infected cells, the two glycoproteins encoded in the dengue cDNA appeared to be proteolytically cleaved and glycosylated in a manner similar to that observed during dengue virus infection. This suggests that the dengue virus structural proteins can be properly processed in the absence of

Figure 2. Expression of dengue proteins. Infected or mock infected cells were labeled with [^{35}S]methionine for 15h in medium containing 1/10 the normal concentration of methionine. The labeled proteins were analyzed by immunoprecipitation followed by SDS-polyacrylamide gel electrophoresis. Abbreviations: VV = cells infected with wild type vaccinia virus; VD = cells infected with recombinant vaccinia; DN = cells infected with dengue 2 virus; M = mock-infected cells. a) ^{35}S-labeled BSC40 cell lysates were immunoprecipitated with hyperimmune anti-dengue mouse ascites fluid (HI) or with monoclonal antisera (3H5) directed against dengue E protein as indicated. b) Dengue protein expression in various cell types. BSC40, MDBK, CHO, and BHK cells were infected as above and lysates were immunoprecipitated with anti-dengue hyperimmune mouse ascites fluid.


A)
HI
3H5
DN VD VV M VV VD
NS5
NS3
E
NS1
prM
B)
BSC40
MDBK
CHO
BHK
M VV VD M VV VD M VV VD M VV VD DN
200K
93K
69K
46K
30K
NS5
NS3
E
NS1
prM

the nonstructural proteins. Each of the dengue viral glycoproteins, prM and E is preceded by a stretch of hydrophobic amino acids which could serve as a signal sequence, and hence cellular signalase is most likely responsible for these proteolytic cleavages. Presumably, the cleavage events that generated the two glycoproteins also yielded the capsid protein, which is located amino terminal to the prM glycoprotein, but in these experiments capsid protein was not seen.

We also examined several other cell lines, both permissive and nonpermissive for vaccinia replication, for the production of dengue proteins after infection by VVDEN, and the results are shown in Fig. 2b. CHO, MDBK, and BHK cells infected with VVDEN all produced the same series of dengue proteins as infected BSC40 cells. Thus, even in cell lines nonpermissive for vaccinia growth (CHO and MDBK cells), the expression of the dengue specific transcripts and protein products is efficient, and proteolytic processing of the polyprotein precursor occurs. VVDEN-infected BSC40 cells expressed the highest levels of dengue proteins of the four cell lines used (Fig. 2b).

Immunologic response of CD-1 and C57BL6 mice inoculated with recombinant VVDEN virus

Two different mouse strains, CD-1 and C57BL6, were immunized with recombinant VVDEN virus, and their immunologic responses were evaluated. Mice were inoculated with VVDEN by tail scarification or foot pad inoculation. Control mice received the same dose of wild type vaccinia virus (strain WR). Animals were bled 3 weeks after immunization to determine if dengue specific antibodies were present. Antibody production was monitored by radioimmune assays and by neutralization of virus infectivity, and the results are shown in Tables 1, 2 and 3, and Fig. 3.

Figure 3. Radioimmunoassay results. Cells were infected with vaccinia virus (solid diamonds), with dengue-2 virus (open squares), or were not infected (solid triangles), lysed, and dilution dried onto 96 well plastic plates. Sera from individual mice were diluted 100 fold, and serial 2-fold dilutions from the diluted sera tested by radioimmunoassay as described in the Materials and Methods. Panels a and b are controls, and panels c-f contain the results with the serum from a single mouse, as follows: a) serum from a vaccinia infected mouse; b) anti-dengue hyperimmune mouse ascites fluid; c) CD-1 mouse #8 infected with VVDEN by tail scarification; d) CD-1 mouse #10 infected with VVDEN by footpad inoculation; e) C57BL6 mouse #24 infected with VVDEN by tail scarification; f) C57BL6 mouse #30 infected with VVDEN by footpad inoculation.

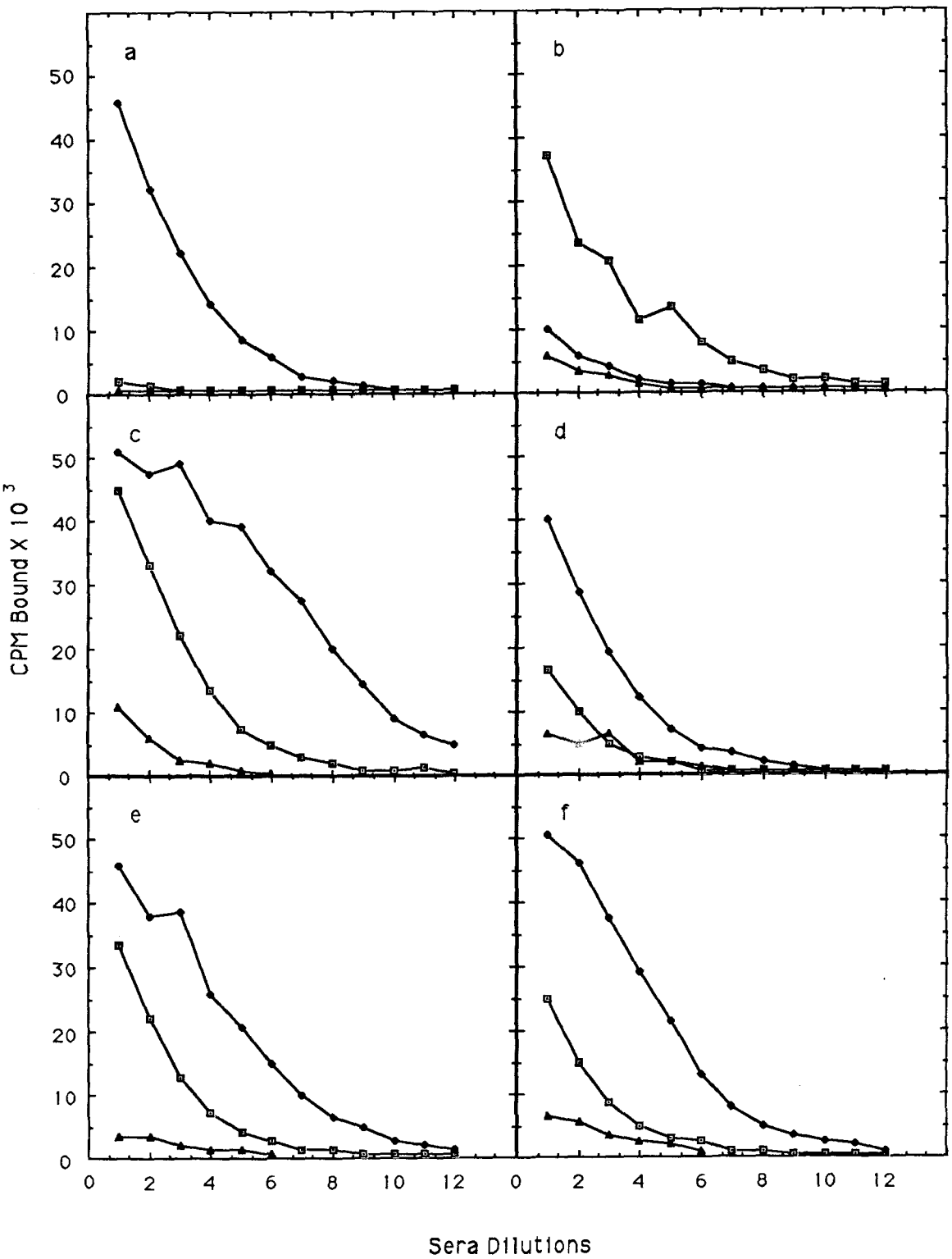

a
b
c
d
e
f
CPM Bound X 10 3
0
10
20
30
40
50
0
2
4
6
8
10
12
Sera Dilutions

Table 1. Antibody production on CD-1 and C57BL6 mice

Viruses [a]	Route of virus inoculation	Mouse strain	Antibody Production [b]	
			Dengue 2	Vaccinia
Vaccinia	Tail Scarification	CD-1	0/4	4/4
	Tail Scarification	C57BL6	0/5	5/5
VVDEN	Tail Scarification	CD-1	8/8	8/8
	Foot Inoculation	CD-1	6/8	8/8
	Tail Scarification	C57BL6	8/8	8/8
	Foot Inoculation	C57BL6	4/8	8/8

[a] Mice were inoculated with a virus dose of 4 $X10^6$ PFU.
[b] Mice were bled 21 days after virus inoculation and antibody production was measured by RIA titer.

Table 2. Dengue 2 neutralization by antisera from CD-1 mice

Virus	Mouse #[a]	Route of virus inoculation	Neutralization titers[b]	
			80%	50%
Vaccinia	CI-1	Tail Scarification	<20	80
	CI-2		<20	160
	CI-3		<20	40
	CI-4		ND [c]	ND
	CI-5		<20	160
VVDEN	1	Tail Scarification	<20	160
	2		320	640
	3		<20	160
	4		<20	<20
	5		40	320
	6		ND	ND
	7		<20	320
	8		80	320
	9	Footpad Inoculation	40	160
	10		<20	160
	11		<20	160
	12		40	320
	13		<20	160
	14		<20	160
	15		<20	160
	16		<20	80
Dengue 2 hyperimmune mouse ascitic fluid			>640	>640

[a] Mice were inoculated with a virus dose of 4×10^6 PFU of VVDEN or the same dose of vaccinia as controls.

[b] The neutralization test was done by incubating virus and serum together overnight at $4^{\circ}C$ in the presence of 2% guinea pig complement. The titers are the reciprocals of the serum dilutions which reduced the plaque count by 50% or 80% when mixed with a constant amount of DEN2 virus (approximately 100 PFU).

[c] N D = not determined.

Table 3. Dengue 2 neutralization by antisera from C57BL6/J mice

Viruses	Mouse # [a]	Route of virus inoculation	Neutralization titers [b]	
			80%	50%
Vaccinia	CC-1	Tail Scarification	<20	80
	CC-2	↓	<20	40
	CC-3		<20	40
	CC-4		<20	40
	CC-5		<20	40
VVDEN	17	Tail Scarification	40	320
	18	↓	40	160
	19		20	320
	20		40	160
	21		40	320
	22		80	320
	23		<20	320
	24	↓	<20	320
	25	Footpad Inoculation	<20	320
	26	↓	<20	160
	27		<20	160
	28		<20	160
	29		<20	160
	30		<20	80
	31		<20	160
	32		<20	160
Dengue 2 hyperimmune mouse ascitic fluid			>640	>640
Normal human serum			<20	<20
Normal mouse serum			0	0

[a] Mice were inoculated with a virus dose of 4×10^6 PFU of VVDEN or the same dose of vaccinia as controls.

[b] The neutralization test was done by incubating virus and serum together overnight at 4 C^{o} in the presence of 2% guinea pig complement. The titers are the reciprocals of the serum dilutions which reduced the plaque count by 50% or 80% when mixed with a constant amount of DEN2 virus (approximately 100PFU).

CD-1 mice immunized with recombinant VVDEN by tail scarification or foot pad inoculation showed anti-dengue antibody responses by RIA in 8 out of 8 mice or 6 out of 8 mice, respectively. In C57BL6 mice which were immunized with VVDEN recombinant virus, 8 out of 8 mice by tail scarification and 4 out of 8 mice by foot pad inoculation produced anti-dengue antibody as detected by RIA. Mice immunized via tail scarification thus demonstrated a more reliable antibody response than the foot pad route, and the antibody titers were higher (Fig. 3). RIA curves for some of the CD-1 mice or C57BL6 mice which were immunized are shown in Fig. 3. The first dilution used was 1:100 and twofold dilutions were used thereafter. Mice inoculated with vaccinia or dengue virus were used as negative or positive controls for an anti-dengue antibody response (Fig. 3a, 3b). CD-1 and C57BL6 mice which were immunized with VVDEN by tail scarification produced high titers of anti-dengue antibody as well as of vaccinia specific antibody, similar to that produced in dengue- or vaccinia-inoculated mice (Fig. 3a, 3b). Dengue specific antibody was lower in mice immunized with VVDEN by foot pad inoculation (Fig. 3d, 3f). No anti-dengue antibody was produced in control mice.

To determine whether the antibodies produced could neutralize virus infectivity, these sera obtained 21 days postinoculation were tested by a neutralization assay. The prototype PR159 strain of dengue 2 was incubated with different sera overnight at 4°C in the presence of 2% guinea pig complement, and the remaining infectivity titrated. The reciprocal dilutions of antisera that when mixed with a constant virus plaque dose of approximately 100 PFU reduced the plaque count by 80% or 50% are shown in Tables 2 and 3. Mice immunized with VVDEN developed neutralization antibody responses that were more pronounced when the route of inoculation was tail scarification than when foot pad inoculation was used. Thus three of seven CD-1 mice tested gave positive results using an 80% endpoint and six of eight C57BL6/5 mice were positive. Mouse #2

demonstrated the best response, a titer of 1:320 at 80% neutralization. The results using a 50% endpoint were more difficult to interpret because of the neutralization obtained with sera from control vaccinia-infected mice, an effect not seen with normal mouse sera. Using footpad inoculation, two of eight CD-1 mice and zero of eight C57BL6 mice were positive using an 80% endpoint.

Discussion

We have constructed a vaccinia virus recombinant which contains a cDNA copy of the structural protein encoding region of the dengue 2 virus (S1) inserted into the vaccinia TK gene. This insert is expressed and translated into dengue C, prM, and E proteins which are correctly cleaved and modified, a result similar to that seen previously with a flavivirus recombinant (Zhao et al., 1987). This demonstrates that the sequences encoding the structural proteins specify all the necessary catalytic activities or recognition signals required to ensure the proper synthesis and maturation of the polypeptides. Each of the viral glycoproteins prM and E is preceded by a stretch of hydrophobic amino acids which could potentially serve as a signal sequence and hence cellular signalase is most likely responsible for the proteolytic cleavages resulting in the mature proteins. We have also shown that the vaccinia recombinant induces the production of dengue specific antibodies in mice which can neutralize viral infectivity, a result which has not been previously obtained with vaccinia-flavivirus recombinants (Zhao et al., 1987). Studies to test the ability of VVDEN to protect mice from dengue virus infection are underway.

Among the mosquito-borne flaviviruses, the dengue viruses are notable for their global distribution and the frequency of large-scale epidemics. Attempts to produce a satisfactory vaccine by conventional methods have met with only limited success to date. Killed virus vaccines have been prepared by intracerebral

passage in mice and tested in a small number of volunteers (Sabin and Schlesinger 1945; Schlesinger et al., 1956). Recently, a tissue culture derived attenuated dengue 2 virus has been considered as candidate vaccines (Bancroft et al., 1984, Eckels et al., 1984). The development of recombinant DNA methods may provide reasonable alternatives for the development of a satisfactory dengue vaccine.

Recombinant vaccinia viruses encoding surface glycoproteins of human respiratory syncytial virus (RSV) (Elango et al., 1986; Wertz et al., 1988) and influenza A virus (Smith et al., 1988) protected rodents from intranasal challenge with RSV and influenza viruses, respectively. Furthermore, recombinant vaccinia virus expressing Venezuelan equine encephalitis virus capsid protein and glycoproteins E1 and E2 protects mice from peripheral VEE virus challenge (Kinney et al., 1988). However, dengue 4 virus structural proteins and nonstructural protein NS1 expressed by a recombinant vaccinia virus were not effective in inducing an antibody response (Zhao et al., 1987). Infection of cotton rats with the recombinant virus induced anti-NS1 antibodies in only 1 of 11 animals, whereas an immune response to the prM or E glycoproteins was not detected (Zhao et al., 1987). In a related study, it has been found that dengue 4 structural proteins and nonstructural protein NS1 expressed by a baculovirus recombinant in insect cells were able to induce an antibody response to M, E, and NS1 rabbits immunized with these dengue virus protein products, although the titers were low, especially to prM and E. These dengue virus antigens produced by the recombinant baculovirus were also found to induce resistance in mice to fatal dengue encephalitis (Zhang et al., 1988). Gould et al. (1986) showed that monoclonal antibody against both E glycoprotein and NS1 nonstructural protein of yellow fever virus can protect mice against infection. Overall, there is no clear cut relationship between the ability of a monoclonal antibody to neutralize virus infectivity *in vitro* and to protect mice *in vivo*. It will be of interest to see

whether the recombinant VVDEN can protect mice from dengue 2 virus infection.

Although immunity to flavivirus infection has generally been correlated with the presence of neutralizing antibodies against the virions envelope glycoproteins, immunization with flavivirus nonstructural protein NS1 protects mice or monkey against lethal flavivirus infection (Schlesinger et al., 1985, 1986, 1987; Cane et al., 1988). This protein is highly conserved among flaviviruses and is the soluble complement-fixing (CF) antigen which is expressed on the surface of infected cells. Monoclonal antibodies to NS1 of the 17D vaccine strain of yellow fever virus confer passive protection in mice against lethal intracerebral challenge with 17D YFV. The protective capacity of monoclonal antibodies to NS1 correlated with their CF activity and with their ability to lyse virus infected cells. Active immunization with purified NS1 or NS1 expressed in bacteria can protect animals in the absence of virus-neutralizing antibodies. The mechanism by which passive or active immunization to NS1 protein protect against flavivirus is not obvious, although the simplest hypothesis is that cytotoxic T-cell reactions are involved.

While sequencing cDNA clones and constructs for this work, we discovered that nine nucleotides were omitted from the sequence reported by Hahn et al. (1988), caused by the close juxtaposition of two EcoRI sites used for sequencing. These nine nucleotides should be inserted between nucleotides 6351 and 6352 of the originally reported sequence, and the correct sequence should read:

K	E	F	K	E	F	A
AAA	GAA	UUC	<u>AAG</u>	<u>GAA</u>	<u>UUC</u>	GCA

where the nine nucleotide insertion is underlined. This sequence arises from a direct repeat of three amino acids, and results in two EcoRI sites spaced by three

nucleotides. Thus the total length of the dengue 2 PR159-S1 RNA is 10712 nucleotides and the (translated) open reading frame is 10173 nucleotides in length.

References

Anderson, D. J., and Blobel, G. (1983). Immunoprecipitation of proteins from cell-free translations. Methods Enzymol. 96: 111-120.

Bancroft, W. H., Scott, R. M., Eckels, K. H., Hoke, C. H., Simms, T. E., Jesrani, K. D. T., Summers, P. L., Dubois, D. R., Tsoulos, D., and Russell, P. K. (1984). Dengue virus type 2 vaccine: reactogenicity and immunogenicity in soldiers. J. Inf. Dis. 149: 1005-1010.

Cane, P. A., and Gould, E. A. (1988). Reduction of yellow fever virus mouse neurovirulence by immunization with a bacterially synthesized nonstructural protein (NS1) fragment. J. Gen. Virol. 69: 1241-1246.

Cleaves, G. R. (1985). Identification of dengue type 2 virus specific high molecular weight proteins in virus-infected BHK cells. J. Gen. Virol. 66: 2767-2771.

Eckels, K. H., Brant, W. E., Harrison, V. R., McCown, J. H., and Russel, P. K. (1976). Isolation of a temperature-sensitive dengue-2 virus under conditions suitable for vaccine development. Infect. Immun. 14: 1221-1227.

Eckels, K. H., Harrison, V. R., Summers, P. L., and Russell, P. K. (1980). Dengue-2 vaccine: Preparation from a small-plaque virus clone. Infect. Immun. 27: 175-180.

Eckels, K. H., Scott, R. McN., Bancroft, W. H., Brown, J., Dubois, D. R., Summers, P. L., and Russell, P. K., and Halsted, S. B. (1984). Selection of attenuated dengue 4 viruses by serial passage in primary kidney cells. V. Human response to immunization with a candidate vaccine prepared in fetal rhesus lung cells. American Journal of Tropical Medicine and Hygiene 33: 684-689.

Elango, N., Prince, G. A., Murphy, B. R., Venkatesan, S., Chanock, R. M., and Moss, B. (1986). Resistance to human respiratory syncytial virus (RSV) infection induced by immunization of cotton rats with recombinant vaccinia virus expressing the RSV G glycoprotein. Proc. Natl. Acad. Sci. USA 83: 1906-1910.

Gould, E. A., Buckley, A., Barrett, A. D. T., and Cammack, N. (1986). Neutralizing (54K) and nonneutralizing (54K and 48K) monoclonal antibodies against structural and nonstructural yellow fever virus proteins confer immunity in mice. J. Gen. Virol. 67: 591-595.

Graham, F. L., and Van der Eb, A. J. (1973). A new technique for the assay of infectivity of human adenovirus 5 DNA. Virology 52: 456-467.

Gubler, D. J. (1988). Dengue, In: *The Arboviruses: Epidemiology and Ecology*, (T. P. Monath, ed.) Vol. II, Chapter 23, pp. 223-260, C. R. C. Press, Boca Raton, FL.

Hahn, Y. S., Galler, R., Hunkapiller, T., Dalrymple, J. M., Strauss, J. H. and Strauss, E. G. (1988). Nucleotide sequence of dengue 2 RNA and comparison of the encoded proteins with those of other flaviviruses. Virology 162: 167-180.

Halstead, S. B. (1988). Pathogenesis of dengue: challenge to molecular biology. Science 239: 476-481.

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

Hruby, D. E., Guarino, L. A., and Kates, J. R. (1979). Vaccinia virus replication. J. Virol. 29: 705-715.

Kinney, R., Esposito, J. J., Mathews, J. H., Johnson, B. J. B, Roehrig, J. T., Barrett, A. D. T., and Trent, D. W. (1988). Recombinant vaccinia virus/Venezuelan Equine Encephalitis (VEE) virus protects mice from peripheral VEE challenge. J. Virol. 62: 4697-4702.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriphage T4. Nature (London) 227: 680-685.

Mackett, J., Smith, G. L., and Moss, B. (1984). General method for production and selection of infectious vaccinia recombinants expressing foreign genes. J. Virol. 49: 857-864.

Mackow, E., Makino, Y., Zhao, B., Zhang, Y. M., Markoff, L., Buckler-White, A., Guiller, M., Chanock, R., and Lai, C. J. (1987). The nucleotide sequence of dengue type 4 virus: analysis of genes coding for structural proteins. Virology 159: 217-228.

Maniatis, T., Fritch, E. F., and Sambrook, J. (1982). Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Monath, T. P. (1986). Pathology of flaviviruses. In *The Togaviridae and Flaviviridae* (S. Schlesinger and M. Schlesinger, eds.), Vol. 12, pp. 375-440. Plenum, New York.

Rice, C. M., Franke, C. A., Strauss, J. H., and Hruby, D. (1985). Expression of Sindbis virus structural proteins via recombinant vaccinia virus: synthesis, processing and incorporation into mature Sindbis virions. J. Virol. 56: 227-239.

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

Sabin, A. B., and Schlesinger, J. J. (1945). Production of immunity to dengue with virus modified by propagation in mice. Science 101: 640-642.

Schlesinger, R. W., Gordon, I., Frankel, J. W., Winter, J. W., Patterson, P. R., and Dorrance, W. R. (1956). Clincal and serologic response of man to immunization with attenuated dengue and yellow fever viruses. J. Immunol. 77: 352-364.

Schlesinger, J. J., Brandriss, M. W., and Walsh, E. E. (1985). Protection against 17D yellow fever encephalitis in mice by passive transfer of monoclonal antibodies to the nonstructural glycoprotein gp48 and by active immunization with gp48. J. Immunol. 135: 2805-2809.

Schlesinger, J. J., Brandriss, M. W., Cropp, C. B., and Monath, T. P. (1986). Protection against yellow fever in monkeys by immunization with yellow fever virus nonstructural protein NS1. J. Virol. 60: 1153-1155.

Schlesinger, J. J., Brandriss, M. W., and Walsh, E. E. (1987). Protection of mice against dengue 2 encephalitis by immunization with dengue 2 virus nonstructural glycoprotein NS1. J. Gen. Virol. 68: 853-857.

Shapiro, D., Brant, W. E., and Russell, P. K. (1972). Change involving a viral membrane glycoprotein during morphogenesis of group B arboviruses. Virology 50: 906-911.

Smith, G. L., Murphy, B. R., and Moss, B. (1983). Construction and characterization of an infectious vaccinia virus recombinant that expresses the influenza hemagglutinin gene and induces resistance to influenza virus infection in hamsters. Proc. Natl. Acad. Sci. USA 80: 7155-7159.

Smith, T. J., Brandt, W. E., Swanson, J. L., McCown, J. M., and Buesher, E. L. (1970). Physical and biological properties of dengue 2 virus and associated antigens. J. Virol. 5: 524-532.

Stevens, T. M., Schlesinger, R. W., and Schulze, I. T. (1965). Studies on the nature of dengue viruses. I. Correlation of particle density, infectivity and RNA content of type 2 virus. Virology 27: 103-112.

Stolar, V. (1969). Studies on the nature of dengue viruses. IV. The structural proteins of type 2 dengue virus. Virology 39: 426-438.

Summers, P. L., Eckels, K. H., Dalrymple, J. M., Scott, R. M., and Boyd, V. A. (1984). Antibody response to dengue-2 vaccine measured by two different radioimmunoassay methods. J. Clin. Micro. 195: 651-659.

Sweet, B. H., and Sabin, A. B. (1954). Properties and antigenic relationships of hemagglutinins associated with the dengue viruses. J. Immunol. 73: 363-373.

Villarreal, L. P., and Berg, P. (1977). Hybridization *in situ* of SV40 plaques: detection of recombinant SV40 virus carrying specific sequences of nonviral DNA. Science 196: 183-186.

Wertz, G. W., Stott, E. J., Young, K. K. Y., Anderson, K., and Ball, L. A. (1987). Expression of the fusion protein of human respiratory syncytial virus from recombinant vaccinia virus vectors and protection of vaccinated mice. J. Virol. 61: 293-301.

Zhang, Y., Hayes, E. P., McCarty, T. C., Dubois, D. R., Summers, P. L., Eckels, K. H., Chanock, R. M., and Lai, C. (1988). Immunization of mice with dengue structural proteins and nonstructural proteins NS1 expressed by baculovirus recombinant induces resistance to dengue virus encephalitis. J. Virol. 62: 3027-3031.

Zhao, B., Prince, G., Horswood, R., Eckels, K., Summers, P., Chanock, R., and Lai, C. (1987). Expression of dengue virus structural proteins and nonstructural protein NS1 by a recombinant vaccinia virus. J. Virol. 61: 4019-4022.

Acknowledgements

This work has been supported by a grant from the World Health Organization.

APPENDIX

3'-5' Deletion Analysis of the Vaccinia Virus 7.5k Early Promoter Region: Viral Transcription Can Initiate within Foreign Sequences

3'-5' Deletion Analysis of the Vaccinia Virus 7.5k Early Promoter Region: Viral Transcription Can Initiate within Foreign Sequences

Charles M. Rice[1,3], Young S. L. Hahn[1], James H. Strauss[1], Walter M. Hodges[2], Chris A. Franke[2], and Dennis E. Hruby[2]

[1]Division of Biology, California Institute of Technology, Pasadena, California 91125. [2]Center for Gene Research and Department of Microbiology, Oregon State University, Corvallis, Oregon 97331.

[3]Present address: Department of Microbiology and Immunology, Washington University School of Medicine, Box 8093, 660 South Euclid Avenue, St. Louis, Missouri 63110-1093.

ABSTRACT

A series of 3' to 5' deletions extending up to and through the transcription initiation site of the early promoter for the vaccinia virus 7.5 kilodalton gene (7.5 kd) were constructed and fused to the bacterial chloramphenicol acetyl transferase (CAT) gene. The effects of these deletions on early and late viral transcription were studied by assaying CAT activity and by 5' mapping of CAT-specific transcripts in cells infected with vaccinia recombinants containing the chimeric genes. A deletion to +12 nucleotides (relative to the major start for early transcription) had little effect on either the level of CAT activity or CAT-specific transcripts. Deletions to +2 and -7 drastically reduced the level of both the early transcripts and CAT activity, and a deletion to -30 abolished early transcription. The 5' termini of these transcripts were mapped using both S1 nuclease and primer extension and the results indicate that in the -7 deletion, proper initiation of early transcription can occur, although inefficiently, within this foreign insert.

INTRODUCTION

Vaccinia virus, a poxvirus (reviewed in ref. 1), has been used successfully as an expression vector for a number of foreign genes (2-12). As an expression system, its advantages include a broad host range (13), accommodation of large inserts of foreign DNA (14), cytoplasmic transcription with no evidence of splicing (15-17), and the potential for using engineered vaccinia recombinants as live vaccine strains (6,7,12,18). It is thus useful for expression of heterologous genes lacking introns, such as cDNA copies of RNA viruses which replicate in the cytoplasm.

Distinct sets of early and late vaccinia virus genes are transcribed by virus encoded enzymes which recognize unique viral regulatory regions (17,19-23).

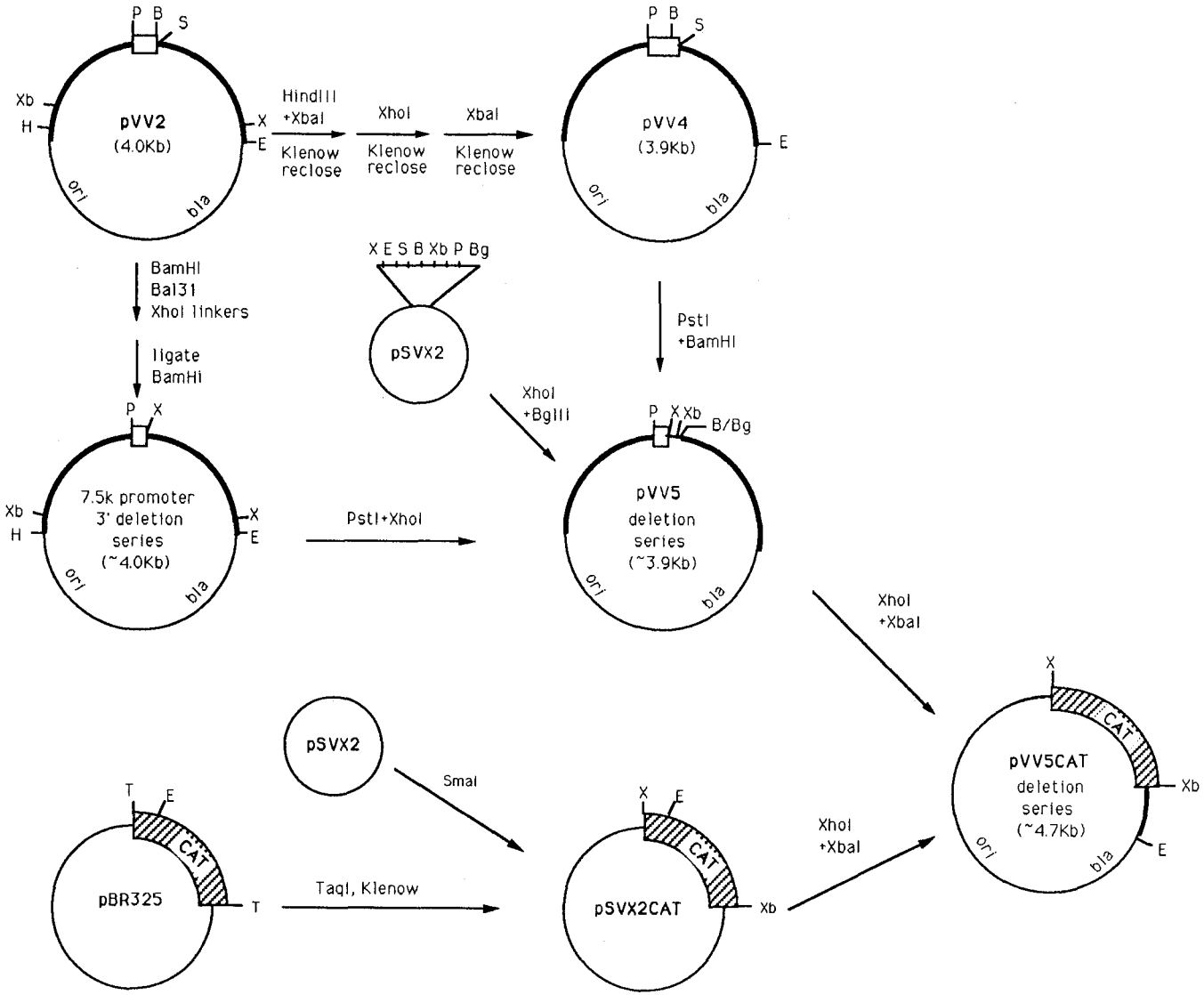
Previous studies have shown that the upstream sequences of the vaccinia 7.5 kd cistron (17) can be used to efficiently drive the expression of heterologous sequences throughout the infection cycle (11,24,25) as well as in cell lines which are nonpermissive for vaccinia replication (11). This region contains two major promoters; one which is active early in infection and a second, about 55 nucleotides upstream, which functions during expression of the late viral genes (11,25). Analysis of 5' to 3' deletion mutants in this region has demonstrated that cis-regulatory sequences necessary for proper transcription initiation lie, in both cases, within 31 bases upstream of their transcription start sites (25). In this study, we have constructed a series of 3' to 5' deletion mutants of the 7.5 kd early promoter region to define the 3' boundary of the promoter necessary for efficient transcription of foreign genes, in this case the bacterial chloramphenicol acetyl transferase (CAT) gene.

MATERIALS AND METHODS

Plasmid constructions

Plasmids were constructed, propagated, and purified using essentially standard techniques (26). pVV2 (11) was linearized at the BamHI site and deletions were produced using Bal 31 nuclease and reclosed with XhoI Linkers. The PstI-XhoI fragments from these plasmids were ligated to a PstI-BamHI cut derivative of pVV2, pVV4 (which had the HindIII-XbaI fragment removed, and both the XhoI and XbaI sites destroyed), and a XhoI-BglII polylinker fragment from pSVX2 (C.M.R., unpublished). This series is designated pVV5. The TaqI fragment of pBR325 containing the CAT gene was filled in and subcloned into the SmaI site of pSVX2. The CAT gene was removed from these plasmids with XhoI and XbaI (yielding the CAT gene in either orientation with respect to the XhoI site) and cloned directionally into the pVV5 series, and designated pVV5CAT. A schematic of these constructions is shown in Fig. 1. The deletion breakpoints were determined by DNA sequencing procedures using the chemical method (27).

Figure 1. Construction of the VV5 deletion series. The open boxes indicate the vaccinia virus 7.5 kd promoter region. Vaccinia virus tk DNA (bold lines), the polylinker region (wavy lines), and the CAT gene (hatched boxes) are shown. B = BamHI, Bg = BglII, E = EcoRI, H = HindIII. P = PstI, S = SmaI, T = TaqI, X = XhoI, Xb = XbaI.


P B S
Xb
H
pVV2
(4.0Kb)
X
E
ori
bla
HindIII
+XbaI
Klenow
reclose
XhoI
Klenow
reclose
XbaI
Klenow
reclose
pVV4
(3.9Kb)
BamHI
Bal31
XhoI linkers
ligate
BamHI
X E S B Xb P Bg
pSVX2
PstI
+BamHI
XhoI
+BglII
P X
7.5k promoter
3' deletion
series
(~4.0Kb)
PstI+XhoI
P X Xb B/Bg
pVV5
deletion
series
(~3.9Kb)
XhoI
+XbaI
pSVX2
SmaI
T E
CAT
pBR325
T
TaqI, Klenow
X E
pSVX2CAT
Xb
XhoI
+XbaI
X
pVV5CAT
deletion
series
(~4.7Kb)
Xb
E

Construction of vaccinia recombinants

Recombinants were constructed using parental vaccinia (strain WR) and the marker rescue technique as described previously (11), and purified by several successive plaque purifications. Small scale preparations of DNA from recombinant viruses (11) were digested with appropriate restriction enzymes, separated by electrophoresis on agarose gels (28), transferred to nitrocellulose and analyzed by hybridization with appropriate nick-translated probes (29).

CAT assay

Monolayers of BSC40 cells were infected with wild-type or vaccinia recombinants at a multiplicity of 5 pfu/cell, and extracts were prepared (30) at either 2 h or 8 h post-infection for early and late expression, respectively. Assays were performed essentially as previously described (30), and the products separated by thin layer chromatography. The corresponding areas of the thin layer plate were excised, dissolved in a toluene-based fluor, and counted by liquid scintillation. CAT activity was quantitated relative to a CAT standard (Sigma) included with each set of reactions.

RNA analyses

Cytoplasmic RNAs were harvested from monolayers of infected BSC40 cells at either 2 h or 8 h post-infection (31). Early RNAs were also prepared at 6 h post-infection in the presence of 100 μg/ml cycloheximide. 10 μg of total RNA were used for each analysis. For S1 nuclease analysis, 5' ^{32}P end-labeled probes were prepared from each of the pVV5CAT plasmids by digestion with MspI, followed by alkaline phosphatase treatment and kinasing with γ-^{32}P ATP and T4 polynucleotide kinase (27). After digestion with Sau3AI the fragments were isolated by preparative gel electrophoresis, elution and precipitation with ethanol (32). RNA samples and the corresponding end-labeled probes were mixed with 25 μg carrier tRNA and co-precipitated with ethanol. After resuspension in 10 μl annealing buffer (40 mM Pipes, pH 6.4, 400 mM NaCl, 1 mM EDTA, 80% deionized

formamide) the samples were heated to 85°C for 10 min and incubated at 40°C at least 12 h. The annealing reactions were terminated by the addition of 150 µl cold S1 buffer (280 mM NaCl, 30 mM NaOAc, pH 4.4, 4.5 mM $ZnCl_2$) containing 20 µg/ml denatured calf thymus DNA, 10 µg/ml duplex calf thymus DNA, and 200 U/ml S1 nuclease. Unless otherwise indicated, incubation was for 30 min at room temperature, followed by isopropanol precipitation (33), and heated to 90°C immediately prior to electrophoresis on 5% polyacrylamide, 8.3 M urea sequencing gels (27). Sequence ladders for size markers were generated by the chemical method (27) from probes used for S1 mapping or from fragments whose sequence had previously been determined.

For primer extensions, the primer 5'-TTAGCTCCTGAAAATCTC-3' was synthesized (complementary to the sequence starting 28 nucleotides 5' to the AUG codon of the CAT gene), end-labeled with γ-^{32}P ATP and T4 polynucleotide kinase, and isolated by preparative gel electrophoresis, elution, and ethanol precipitation (32). Approximately 10 µg total RNA and 0.05 pmole 5'-end-labeled primer were co-precipitated and resuspended in 5 µl reverse transcriptase buffer (32). After heating to 65°C for 10 min the samples were cooled to 30°C and dithiothreitol was added to 10 mM followed by the addition of 12.5 U RNasin. After 3 h annealed primers were extended with reverse transcriptase under conditions previously described (34) in the presence of 40 µg/ml actinomycin D at 37°C for 45 min. The reactions were stopped by the addition of EDTA and SDS to 10 mM and 0.5%, respectively, extracted with a mixture of phenol and chloroform (1:1 v/v), and precipitated with ethanol. The samples were resuspended in sequencing sample buffer, heated to 90°C, and analyzed as described above.

RESULTS

Plasmid and recombinant constructions

The construction of the vaccinia insertion plasmids is outlined in Fig. 1. The thymidine kinase (tk) gene and flanking sequences with the 7.5 kd gene promoter region inserted into the tk coding region were originally derived from pGS20 (6,24). A set of 3' to 5' deletion mutants were constructed using Bal 31 digestion, beginning at the BamHI site 36 nucleotides 3' to the major early transcription start site (17), and then fused to the CAT gene with the AUG initiation codon 44 nucleotides downstream from the deletion end-points. The position of each deletion was determined by chemical sequence analysis of the plasmid DNA as indicated in Fig. 2. As controls, two plasmids (with deletion end-points at +2 and +36 nucleotides) were constructed which contained the CAT gene in the opposite orientation with respect to the 7.5 kd gene promoter.

Although a transient expression protocol could theoretically be used as a rapid assay for vaccinia promoter activity, this method seems to favor expression driven by late promoters (35). Thus it was necessary to produce infectious recombinants containing each of these chimeric constructions by using the marker rescue technique (36,37). CsCl-banded plasmid DNAs were co-precipitated with wild-type vaccinia DNA using the $CaPO_4$ method and transferred to monolayers of tk^- L cells which had been previously infected with wild-type vaccinia. Recombinants arising by insertional inactivation of the tk gene were amplified in the presence of bromodeoxyuridine, identified by *in situ* hybridization with a CAT-specific probe, and purified by several rounds of plaque purification. The structure of the CAT gene relative to the 7.5 kd gene promoter was verified for each recombinant by Southern blot analysis (29) and is shown in Fig. 3. Recombinant DNAs digested with PstI and PvuII (which cut at -270 nucleotides in the 7.5 kd promoter and at 158 within the CAT insert, respectively) were transferred to nitrocellulose and hybridized to a CAT-specific probe. In each

Figure 2. Sequence of the 7.5 kd promoter region and deletion breakpoints. The sequence of the +36 7.5 kd/CAT construct is shown. The position of the 3' deletion breakpoints are indicated and numbered relative to the 7.5 kd major early transcription start (↓, 17). The juxtaposed polylinker and CAT sequences (43) are 3' to the position marked by (↑). An oligonucleotide complementary to the underlined sequence was used for primer extension analyses, and the MspI site indicated was end-labeled for S1 analyses. The major late start region (25) is indicated by an open box, and the early start region in the -7 deletion mutant is shown by a solid overline. The start codon for translation of the CAT gene is boxed.

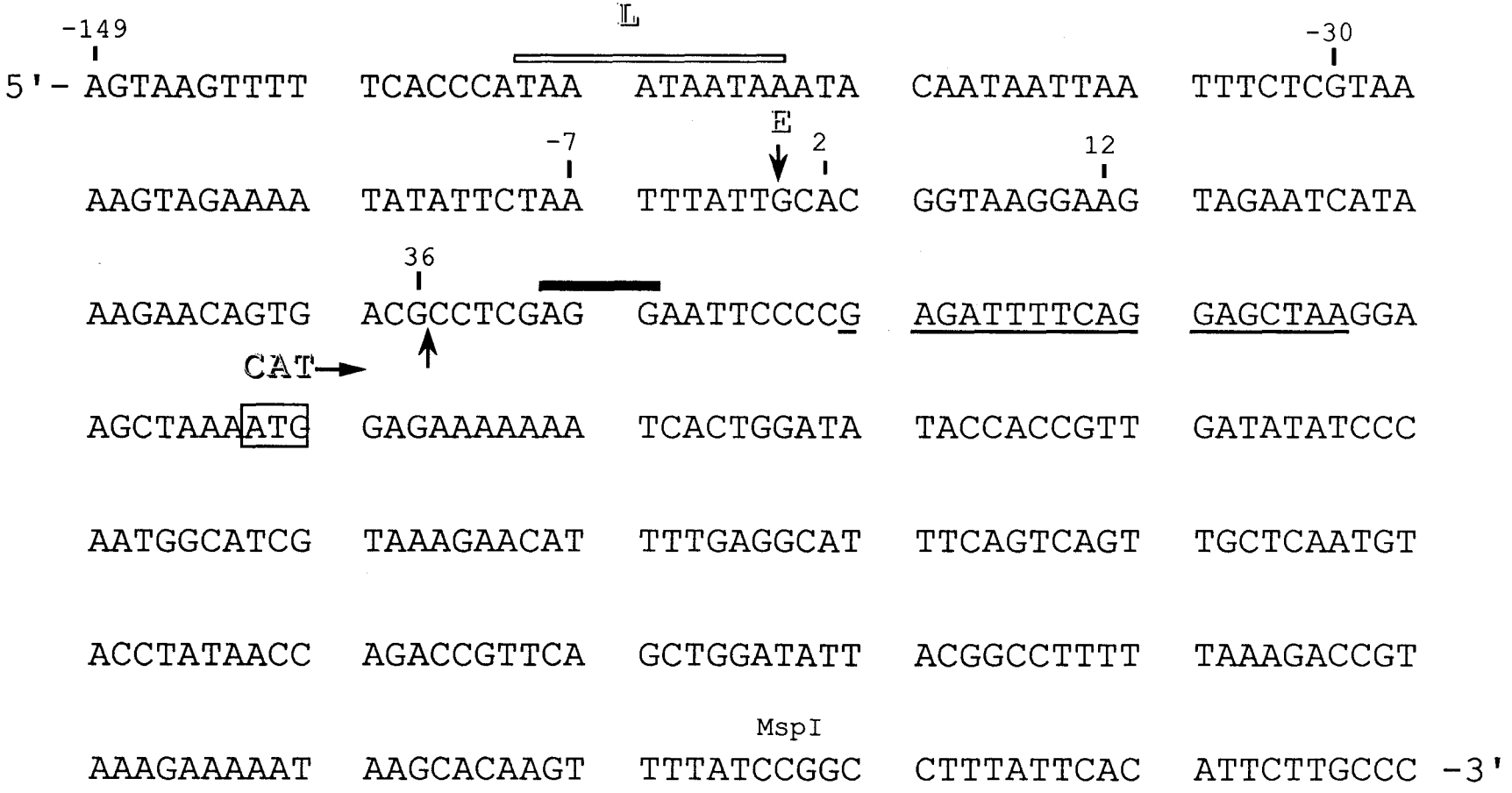
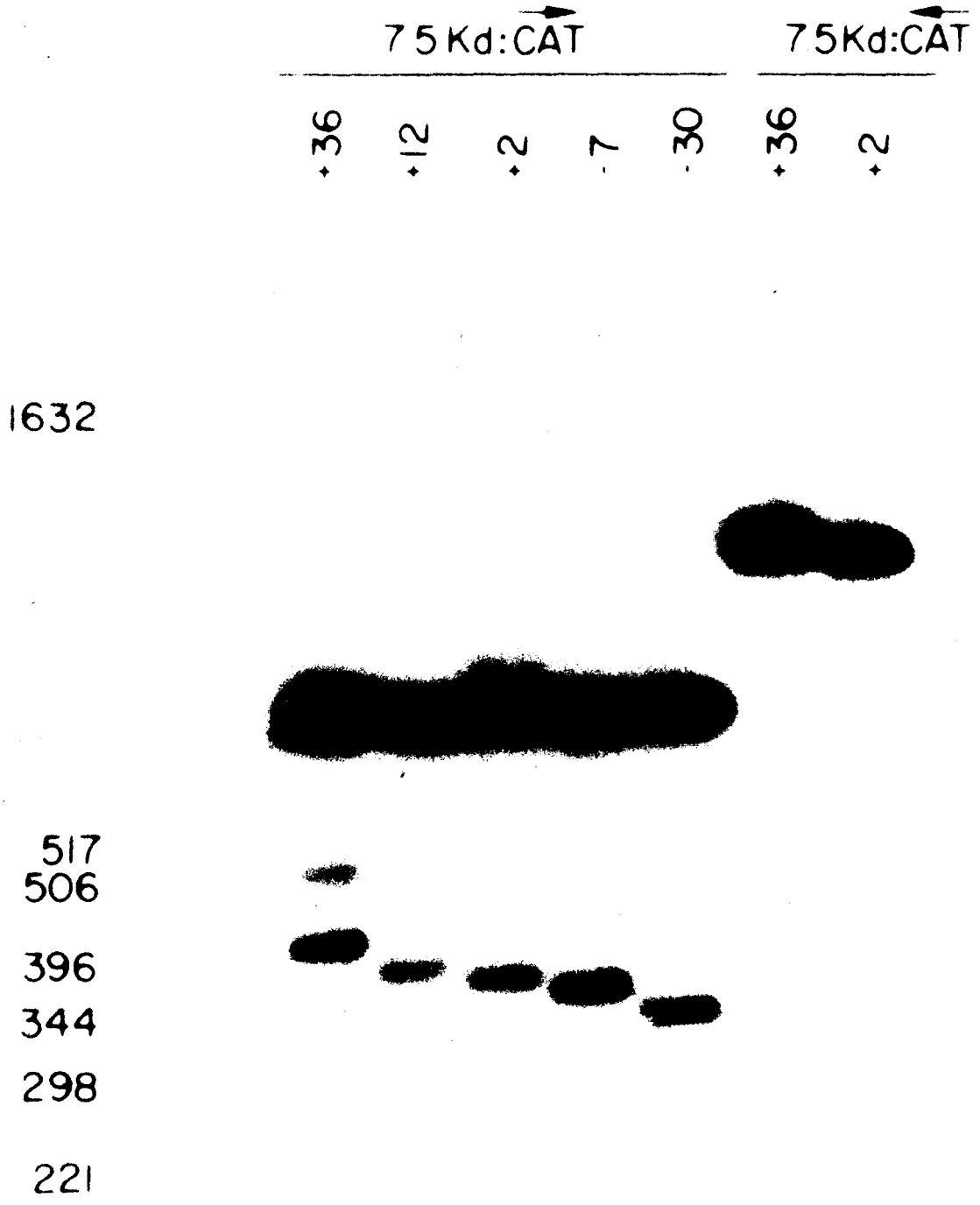

-149
L
-30
5'- AGTAAGTTTT TCACCCATAA ATAATAAATA CAATAATTAA TTTCTCGTAA
E
-7
2
12
AAGTAGAAAA TATATTCTAA TTTATTGCAC GGTAAGGAAG TAGAATCATA
36
AAGAACAGTG ACGCCTCGAG GAATTCCCCG AGATTTTCAG GAGCTAAGGA
CAT
AGCTAAAATG GAGAAAAAAA TCACTGGATA TACCACCGTT GATATATCCC
AATGGCATCG TAAAGAACAT TTTGAGGCAT TTCAGTCAGT TGCTCAATGT
ACCTATAACC AGACCGTTCA GCTGGATATT ACGGCCTTTT TAAAGACCGT
MspI
AAAGAAAAAT AAGCACAAGT TTTATCCGGC CTTTATTCAC ATTCTTGCCC -3'

Figure 3. Structure of vaccinia/CAT recombinants. DNA from each of the 7.5 kd:CAT deletion mutant was digested with PstI and PvuII, separated on a 1.2% agarose gel, blotted to nitrocellulose and hybridized to a CAT-specific probe. Numbers in the left-hand margin indicate the sizes in base pairs of fragments of a HindIII digest of pBR322 included as markers.


75Kd:CAT
75Kd:CAT
+36
+12
+2
-7
-30
+36
+2
1632
517
506
396
344
298
221

case, hybridization to a fragment of the expected size was detected, verifying the proper structure of the 7.5 kd-CAT chimeric gene in each of the vaccinia recombinants. Similar analyses using HindIII digests with vaccinia or CAT-specific probes demonstrated that insertion had occurred in the vaccinia tk locus presumably via homologous recombinantion with the tk flanking regions (data not shown).

CAT expression in recombinant-infected cells

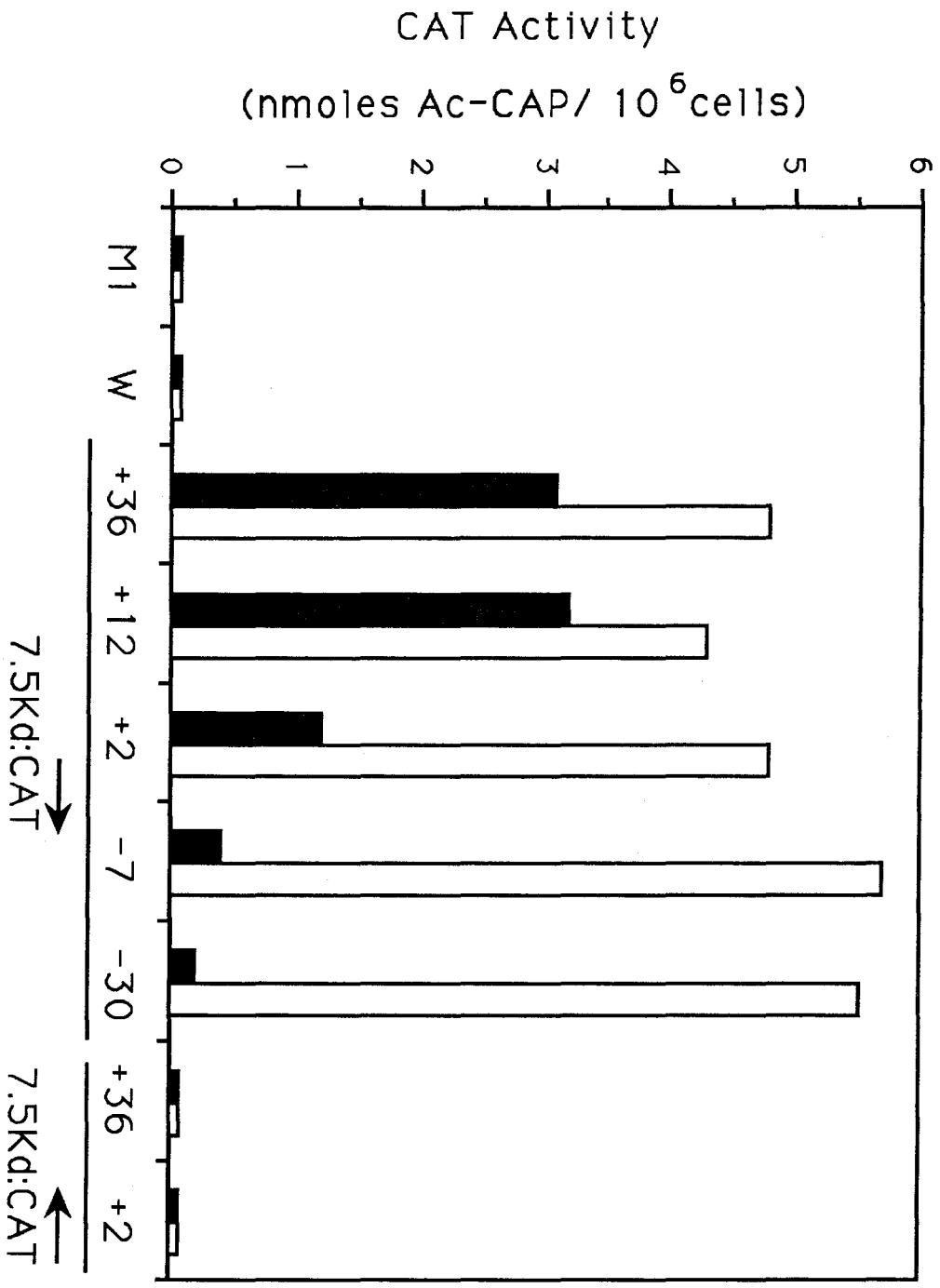
Extracts were prepared from monlayers at 2 h or 8 h after infection with either the recombinants or wild-type vaccinia at equivalent multiplicities and assayed for CAT activity (Fig. 4). No activity was detected in extracts from mock-infected or wild-type vaccinia infected cells. In addition, cell extracts from recombinants containing the CAT in the incorrect orientation with respect to the 7.5 kd promoter had little or no CAT activity. Relative to the +36 construction, the +12 deletion reduced early expression of CAT activity by only 9%. Deletions to +2 and -7, however, reduced the level of early CAT activity by 63% and 86%, respectively. A deletion to -30 reduced the level by more than 95%. After 8 h of infection all of the mutants produced a similar level of CAT activity. These results suggested that sequences between -7 and +12 are important for the efficiency of the vaccinia 7.5 kd early promoter, although a direct effect mediated by the non-vaccinia sequences inserted into this region cannot be ruled out.

Analysis of CAT-specific transcripts

To map the 5'-termini of CAT-specific RNAs and to see if alterations in CAT activity were reflected by different levels of CAT-specific transcripts, total cytoplasmic RNA was prepared at either 2 h or 8 h post-infection from cells infected with wild-type vaccinia or with each of the recombinants. In the first experiment an equivalent proportion of each RNA preparation was annealed with an excess of a 5'-end-labeled CAT-specific probe, digested with S1 nuclease,

Figure 4. CAT activity in recombinant-infected cells Extracts were prepared at either 2 h (solid bars) or 8 h (open bars) post infection and assayed for CAT activity as described in the text. Mock-infected (MI), wild-type vaccinia-infected (WT), and recombinants containing the CAT gene in the opposite orientation with respect to the 7.5 kd promoter

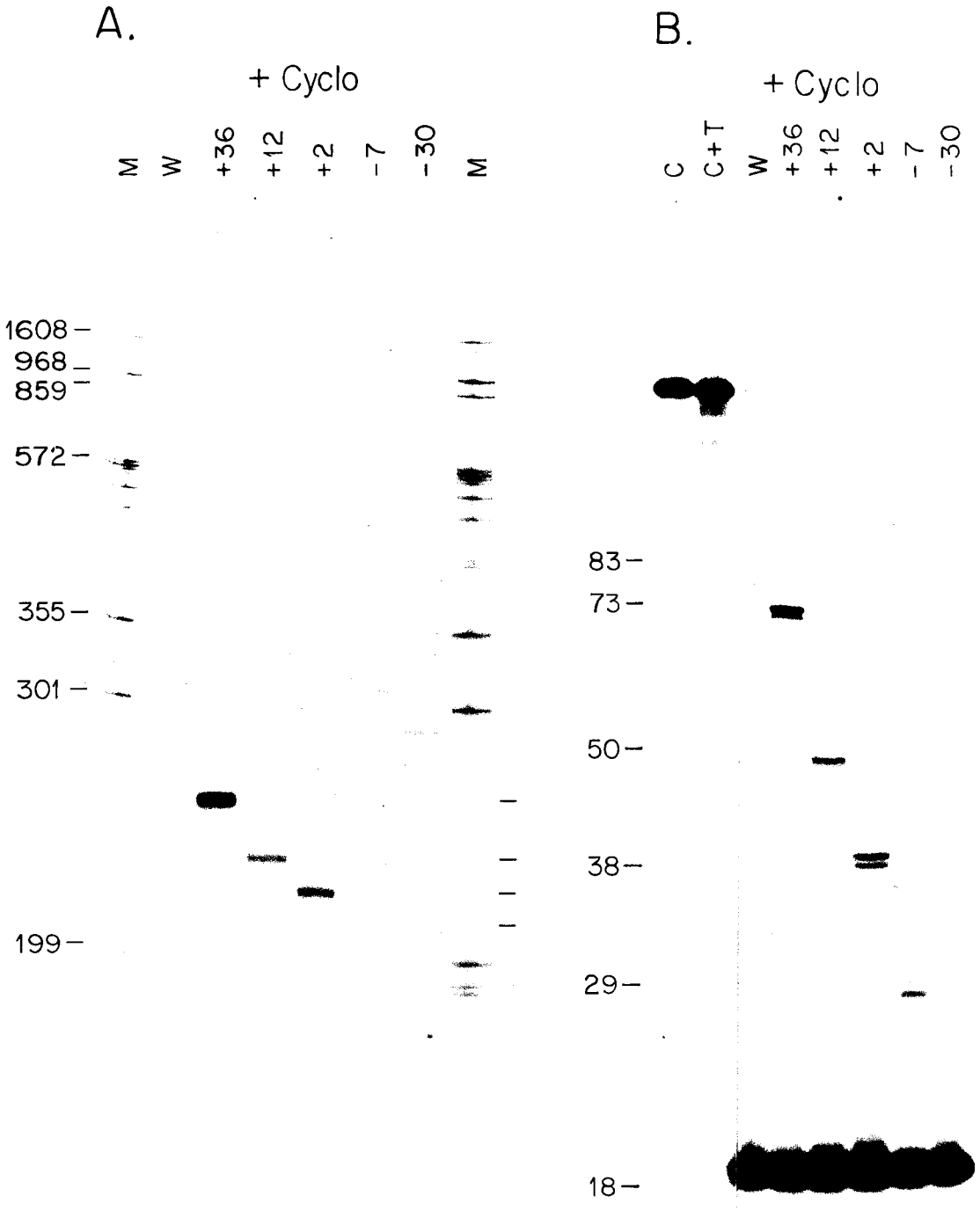
region (7.5 kd:CAT) are shown as controls.


CAT Activity
(nmoles Ac-CAP/ 10^6cells)
0
1
2
3
4
5
6
M1
W
+36
+12
+2
-7
-30
+36
+2
7.5Kd:CAT
7.5Kd:CAT

denatured and run on a sequencing gel (Fig. 5A). The probes were labeled at the MspI site 168 nucleotides downstream from the CAT AUG codon and extended to the Sau3AI site approximately 55 nucleotides upstream of the major 7.5 kd late promoter start site, and would therefore protect transcripts originating from both the early and late 7.5 kd gene promoters (Fig. 2) as well as transcripts from the tk or other promoters further upstream (these protected fragments would comigrate with undigested probe). The sizes of the major protected fragments in both early and late RNA (not shown) preparations for the +36, +12, and +2 constructions (~250, 226, and 216 nucleotides, respectively) corresponded to the 7.5 kd transcription initiation sites which have been previously mapped (25), and again verifies that these deletion mutants have the proper structures. However, the pattern of protected fragments from the late RNA preparations showed considerably more heterogeneity than has been previously reported (21,25), suggesting the presence of a number of weak late transcription initiation sites in the 7.5 kd upstream region in these constructions (not shown). Altering the S1 digestion conditions (temperature or enzyme concentration) did not markedly affect these results (not shown). In the case of the -7 recombinant discrete early transcripts were clearly present, although at a reduced level, and originated in the foreign sequences at precisely the position expected for proper early initiation. No discrete early transcripts originating in the 7.5 kd region were detected when the deletion extended to -30 nucleotides.

The results of these S1 analyses were verified using reverse transcriptase to extend a 5'-end-labeled synthetic oligonucleotide which hybridizes from nucleotides 11 to 28 5' to the CAT AUG codon (see Fig. 2). This method is more sensitive than S1 analysis using end-labeled probes, requires only a single oligonucleotide to analyze transcripts from a series of 3'→5' deletion mutants, and in most cases allows a more precise estimate of the 5' termini of the RNAs since the heterogeneity due to S1 digestion is eliminated. Artifactual bands can be

Figure 5. S1 and primer extension analyses of CAT-specific early transcripts. S1 analyses of RNA at 2 h post-infection. Sau3AI-MspI restriction fragments 5' end-labeled at the MspI site (indicated in Figure 2) were used as probes. Markers were prepared from end-labeled restriction fragments (M; sizes indicated at the left in bases). A mixture of of the probes used for S1 protection (P) is also shown. Control RNA was from wild-type vaccinia-infected cells (W). The positions of the protected fragments are indicated in the right margin. Panel B: Primer extension analyses of early RNAs (not necessarily equivalent amounts) prepared at 6 h post-infection from cells infected in the presence of cycloheximide. A 5' end-labeled synthetic oligonucleotide complementary to the underlined sequence in Figure 2 was used to prime cDNA synthesis on CAT-specific transcripts. The sizes of markers in bases (G + A) are indicated in the left margin.


A.
+ Cyclo
M
W
+36
+12
+2
-7
-30
M
1608
968
859
572
355
301
199
B.
+ Cyclo
C
C+T
W
+36
+12
+2
-7
-30
83
73
50
38
29
18

produced by strong stops for reverse transcriptase, however, and thus a combinantion of both methods is preferable. The results of these experiments are shown in Fig. 5B, and verify that the starts of the transcripts in the -7 deletion mutant are in fact in the foreign sequences. Minor late transcription initiation sites are also seen when the primer extension assay is used (not shown). It should be noted that extended products are often doublets apparently due to partial copying of the RNA cap by reverse transcriptase (38). The probable position for the major initiation site(s) for early transcription in the -7 recombinant is indicated in Fig. 2.

DISCUSSION

The results of these studies suggest that sequences immediately downstream from an early vaccinia virus transcription initiation site can markedly effect the efficiency of transcription. Deletion of the major early initiation site in the 7.5 kd gene promoter region and an additional 6 nucleotides upstream still allows early transcription to occur with initiation at the same relative position downstream in the foreign sequences. However, in the -7 and +2 deletion mutants both the levels of CAT activity and early CAT-specific transcripts were drastically reduced relative to the +12 and +36 constructions. A deletion to -30 completely abolished detectable early transcription. Thus since previous studies have shown that 5'→3' deletions from -31 to -24 abolish the early promoter activity (25), the sequences necessary for proper early initiation lie between positions -31 and -7 relative to the major early start. Whether these observations regarding spatial constraints will apply to other early vaccinia promoters will require additional studies.

Thus far no clear consensus sequence or structure for early vaccinia promoters has emerged although two AT-rich regions are commonly found in the upstream 45 nucleotides (17,19,20), spaced approximately 20-24 nucleotides apart

(39). Since the necessary signals for transcription lie between -31 and -7 in the case of the 7.5 kd early promoter, the AAAATATA sequence located from positions -17 to -10 would seem to be a likely candidate for future mutagenesis studies.

When sequences between -7 and +12 were deleted, the reduction in the level of CAT activity and CAT-specific early transcripts is most simply interpreted as a reduction in the efficiency of the early promoter. This suggests that signals important for regulating vaccinia early transcription may actually extend into transcribed region, and further distinguishes this promoter region from those typically recognized by eucaryotic RNA polymerase II (23). The economy of this promoter fits with the observation that the transcription units in the vaccinia genome which have thus far been examined tend to be extremely closely packed. mRNAs often have short 5' untranslated regions (19,20) and transcription termination signals which overlap with downstream transcription units (15,40-42). The functional significance of this gene organization is unknown and perhaps somewhat surprising given that large inserts of foreign DNA are stable in vaccinia recombinants, at least in tissue culture.

The effects of these deletions around the early start site on late transcription in the 7.5 kd promoter region are also complex. In contrast to other studies (21,25), in these experiments we found a number of weak late transcription initiation sites in this region, in addition to the major site which had been previously reported (25). A possible difference is the tandem orientation of the 7.5 kd-CAT insert with respect to tk promoter in our constructs. In any case, deletions past the early transcription start progressively lead to the loss of transcripts originating from these minor late initiation sites (21).

If vaccinia promoters (or sequences which affect them) generally extend beyond their transcription initiation sites it may be difficult to use this system to produce transcripts with precise 5' termini which originate in the foreign insert at

predictable transcription levels. This study has shown, however, that initiation of early transcription can occur within foreign DNA at the expected position. When vaccinia virus early transcription termination signals are defined, this property may allow the use of this expression system for production of heterologous unspliced transcripts (such as those from cDNA copies of RNA viruses) which require precise 5' and 3' termini for their biological activity.

ACKNOWLEDGEMENTS

We thank E. Lenches for expert technical assistance, Dr. B. Moss for the plasmid pGS20, and Dr. L. Hood and T. Hunkapiller for the use of their computing facility. The work was supported by grants PCM-8316390 (to D.E.H.) and DMB-8316856 (to J.H.S.) from the National Science Foundation, by Public Health Service grants AI-20563 (to D.E.H.) and AI-10793 and AI-20612 (to J.H.S.) from the National Institute of Health, and by Biomedical Research Support grant 557 RR 07003.

REFERENCES

1. Dales, S. and Pogo, B. G. T. (1981) Virology Monographs 18, 1-109.
2. Mackett, M., Smith, G. L. and Moss, B. (1982) Proc. Nat. Acad. Sci. USA 79, 7415-7419.
3. Panicali, D. and Paoletti, E. (1982) Proc. Nat. Acad. Sci. USA 79, 4927-4931.
4. Panicali, D., Davis, S. W., Weinberg, R. L. and Paoletti, E. (1983) Proc. Nat. Acad. Sci. USA 80, 5364-5368.
5. Smith, G. L., Mackett, M. and Moss, B. (1983) Nature (London) 302, 490-495.
6. Smith, G. L., Murphy, B. R. and Moss, B. (1983) Proc. Nat. Acad. Sci. USA 80, 7155-7159.
7. Paoletti, E., Lipinskas, B. R., Samsonoff, C., Mercer, S. and Panacali, D. (1984) Proc. Nat. Acad. Sci. USA 81, 193-197.
8. Smith, G. L., Godson, G. N., Nussenzweig, V., Nussenzweig, R. W., Barnwell, J. and Moss, B. (1984) Science 224, 297-399
9. Kieny, M. P., Lathe, R., Drillien, R., Spehner, D., Skory, S., Schmitt, D., Wiktor, T., Koprowski, H. and Lecocq, J. P. (1984) Nature (London) 312, 163-166.
10. Mackett, M., Yilma, T., Rose, J. K. and Moss, B. (1985). Science 227, 433-435.
11. Rice, C. M., Franke, C. A., Strauss, J. H and Hruby, D. E. (1985) J. Virol 56, 227-239.
12. Perkus, N. E., Piccini, A., Lipinskas, B. R. and Paoletti, E. (1985) Science 229, 981-984.
13. Hruby, D. E., Lynn, D. L., Condit, R. C. and Kates, J. R. (1980) J. Gen. Virol. 47, 485-488.

14. Smith, G. L. and Moss, B. (1983) Gene 25, 21-28.

15. Wittek, R., Cooper, J. A., Barbarosa, E. and B. Moss (1980) Cell 21, 487-493.

16. Wittek, R. and Moss, B. (1982) J. Virol. 42, 447-455.

17. Venkatesen, S., Baroudy, B. M. and Moss, B. (1981) Cell 25, 805-813.

18. Moss, B., Smith, G. L., Gerin, J. L. and Purcell, R. H. (1984) Nature (London) 311, 67-69.

19. Venkatesen, S., Gershowitz, A. and Moss, B. (1982) J. Virol. 44, 637-646.

20. Weir, J. P. and Moss, B. (1983) J. Virol. 46, 530-537.

21. Weir, J. P. and Moss, B. (1984) J. Virol. 51, 662-669.

22. Hruby, D. E., Maki, R. A., Miller, D. B. and Ball, L. A. (1983) Proc. Nat. Acad. Sci. USA 80, 3411-3415.

23. Puckett, C. and Moss, B. (1983) Cell 35, 441-448.

24. Mackett, M., Smith, G. L. and Moss, B. (1984) J. Virol. 49, 857-864.

25. Cochran, M. A., Puckett, C. and Moss, B. (1985) J. Virol. 54, 30-37.

26. Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) Molecular cloning: A laboratory manual (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).

27. Maxam, A. and Gilbert, W. (1980) Meth. Enzymol. 65, 499-560.

28. Studier, F. W. (1973) J. Mol. Biol. 79, 237-248.

29. Southern, E. M. (1975) J. Mol. Biol. 98, 503-517.

30. Gorman, C. M., Moffat, L. F. and Howard, B. H. (1982) Mol. Cell Biol. 2, 1044-1051.

31. Ou, J.-H., Strauss, E. G. and Strauss, J. H. (1981) Virology 109, 281-289.

32. Rice, C. M. and Strauss, J. H. (1981) J. Mol. Biol. 150, 315-340.

33. Favaloro, J. M., Treisman, R. H. and Kamen, R. (1980) Meth. Enzymol. 65, 718-749.

34. Strauss, E. G., Rice, C. M. and Strauss, J. H. (1984) Virology 133, 92-110.

35. Cochran, M. A., Mackett, M. and Moss, B. (1985) Proc. Nat. Acad. Sci. USA 82, 19-23.

36. Nakano, E., Panicali, D. and Paoletti, E. (1982) Proc. Nat. Acad. Sci. USA 79, 1593-1596.

37. Weir, J. P., Bajszár, G. and Moss, B. (1982) Proc. Nat. Acad. Sci. USA 79, 1210,1214.

38. Gupta, K. C. and Kingsbury, D. W. (1984) Nucl. Acids Res. 12, 3829-3841.

39. Pluciennicak, A., Schroeder, E., Zettlmeissl, G. and Streeck, R. E. (1985) Nucl. Acids Res. 13, 985-998.

40. Mahr, A. and Roberts, B. E. (1984) J. Virol. 49, 497-509.

41. Golini, F. and J. R. Kates (1984) J. Virol. 49, 459-470.

42. Bajszár, G., Wittek, R., Weir, J. P. and Moss, B. (1983) J. Virol. 45, 62-72.

43. Alton, N. K. and Vapnek, D. (1979) Nature 282, 864-869.