Current Topics in Microbiology and Immunology

105

Current Topics in Microbiology and Immunology

105

Editors

M. Cooper, Birmingham/Alabama · P.H. Hofschneider, Martinsried · H. Koprowski, Philadelphia · F. Melchers, Basel
R. Rott, Gießen · H.G. Schweiger, Ladenburg/Heidelberg
P.K. Vogt, Los Angeles · R. Zinkernagel, Zürich

With 35 Figures

Springer-Verlag
Berlin Heidelberg New York Tokyo 1983

Table of Contents

Indexed in Current Contents

ISBN 3-540-12492-6 Springer-Verlag Berlin Heidelberg New York Tokyo
ISBN 0-387-12492-6 Springer-Verlag New York Heidelberg Berlin Tokyo

Library of Congress Catalog Card Number 15-12910
Printed in Germany

Typesetting: Fotosatz Service Weihrauch, Würzburg.
Printing and binding: Universitätsdruckerei H. Stürtz AG, Würzburg.
2121/3321-543210

Replication Strategies of the Single Stranded RNA Viruses of Eukaryotes

ELLEN G. STRAUSS* AND JAMES H. STRAUSS*

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

1 Introduction

Our knowledge of the molecular biology of virus replication has expanded dramatically in the last few years, especially with the advent of rapid techniques for obtaining the nucleotide sequence of viral genomes. Full or partial sequences of virus genomes are appearing monthly, and it seems appropriate at this time to review the subject of the strategies used for replication by RNA animal viruses in the hope of formulating a conceptual framework in which to organize the new sequence information. This chapter will be concerned with the single-stranded RNA viruses which replicate via RNA intermediates and will focus on the animal viruses, but selected plant viruses whose replication strategies are known will also be discussed. The primary topics will be RNA transcription (the production of virus-specific messages), RNA replication (synthesis of viral genomes), and mRNA translation (synthesis and processing of viral proteins).

In terms of replication strategy the RNA viruses can be divided into two groups, the plus stranded viruses and the minus stranded viruses. The plus stranded viruses initiate infection with the translation of the parental genomic RNA to produce the viral replicase/transcriptase enzyme(s). This enzyme complex synthesizes minus strand templates, plus strand genomes, and in many cases plus strand subgenomic messages for virion structural polypeptides. For some viruses, subgenomic messengers are produced for other polypeptides as well. The minus stranded viruses introduce the replicase/transcriptase into the host as a component of the virion and the initial event in virus replication is primary transcription by the parental nucleocapsid to produce messages for all virus-encoded products. Subsequent events include complementary plus strand (antigenome) synthesis, genomic minus strand synthesis, and amplified or secondary transcription. Replication strategies for each group will be discussed in the order of events after infection, starting with translation for plus stranded viruses and with primary transcription for negative stranded viruses.

Replication and transcription of viral RNA involves an initiation event followed by an elongation phase. Temporal and quantitative regulation of transcription versus replication and genome versus antigenome synthesis could reside in the specificity of initiation. We propose that polynucleotide sequences, usually but not always located at the termini of the RNAs, are specifically recognized by the virus-specific replicase/transcriptase enzymes during initiation. Two types of possible recognition signals are described in this review. One is a sequence of 10–20 nucleotides which is highly conserved within a group of related viruses and whose exact RNA sequence may be recognized by the viral enzymes. The second type is a secondary structure composed of up to 200 nucleotides, stabilized by hydrogen bonds in which the structure, and not the sequence, may be recognized. In this case the structure is conserved among related viruses but the primary sequences may diverge. Sequences that may perform these regulatory functions and their significance for replication strategy will be discussed for each virus family.

It has been very difficult to keep the number of literature citations to a manageable number, due both to the scope of this review and to the veritable explosion of published information on viruses in recent years. For this reason we have attempted to make the references selective rather than comprehensive, to cite review articles wherever possible, and to include primarily articles published since 1979.

2 The Plus Stranded Viruses of Animals

All the known plus stranded RNA animal viruses with the exception of the nodaviruses contain a nonsegmented genome. The infecting RNA is translated to produce the viral replicase, which in turn copies the parental RNA to produce the minus strand. The minus strand can then be used to produce plus strands. The production of a full length minus strand from the plus strand, and of a full length plus strand from the minus strand, will be referred to as RNA replication, and the enzyme(s) responsible as the viral replicase(s). If one or more subgenomic RNAs are produced, this event will be referred to as transcription and the enzyme(s) responsible as the viral transcriptase(s).

The most characteristic differences in replication strategies of the plus stranded RNA viruses involve the mechanisms used for production and translation of messenger RNAs. Table 1 summarizes the virus groups according to morphology and lists their best-known members and salient features of their replication. Discussion of the virus groups will be organized according to the replication strategies employed by the viruses.

2.1 The Picornaviruses

The picornavirus family is made up of nonenveloped virions approximately 22–30 nm in diameter with icosahedral symmetry. The genome is a single RNA molecule of molecular weight 2.5×10^6 or 7500 nucleotides (7.5 kb) which has a covalently linked protein at the 5′ terminus and a 3′ terminal poly(A) tail of 80–120 nucleotides. The protein shell is made up of 60 copies of each of four virus-specific polypeptides. VP1, VP2, and VP3 have molecular weights between 23 000 and 37 000 (23–37 K) depending upon the virus; VP4 is variously reported to have a molecular weight of 9–15 K (*Rekosh* 1977; *Matthews* 1982). The most recent classification of the picornaviruses divides them into four genera on the basis of buoyant density, stability to acid pH, and structure of the genome: the enteroviruses, cardioviruses, rhinoviruses, and aphthoviruses (*Matthews* 1982). All of these viruses are mammalian pathogens, and the enterovirus and rhinovirus groups contain a number of significant human pathogens. In addition to these four groups there are a number of unclassified picornaviruses of insects.

2.1.1 Translation Strategy

The RNA sequence of the genome of poliovirus has recently been determined in its entirety (*Kitamura* et al. 1981; *Racaniello* and *Baltimore* 1981), and the translation and processing strategy of the virus is becoming clear. The organization of the genome is illustrated schematically in Fig. 1. The RNA is 7433 nucleotides long; translation of the RNA begins at an AUG codon 741 nucleotides from the 5′ terminus and continues for 6621

Table 1. The plus stranded animal viruses

Virus family[a]	Genera	Representative members	Structure of nucleocapsid or virion	Size of genome (kb)	Terminal modifications of genome	Number of subgenomic mRNAs	Number of polyprotein precursors	Other characteristics
Nonenveloped								
Picornavirus	Enterovirus	Polio, Coxsackie, Echo, SVDV	Icosahedral	7.5	5′ VPg, 3′ poly(A)	0	1	Internal poly (C) in genome RNA
	Cardiovirus	EMC, ME, Mengo						
	Rhinovirus	> 100 Serotypes of human rhinoviruses						
	Aphthovirus	FMDV						Internal poly (C) in genome RNA
Nodavirus	–	Nodamura, black beetle	Icosahedral Two RNAs	3 1.5	no poly(A)	1	1 or 2	Only plus stranded animal virus with segmented genome
Calicivirus	–	VESV, SMSV, feline caliciviruses	Icosahedral	7–8	5′ VPg, 3′ poly(A)	1 or 2	1 or more	
Enveloped								
Togavirus[b]	Alphavirus	Sindbis, SFV, MID, WEE, VEE, EEE, HJ, and others	Icosahedral nucleocapsid	12	5′ cap, 3′ poly(A)	1	2 or more	Replicate in both mammalian host and invertebrate vector
	Flavivirus	Yellow fever, JE, SLE, and others	Icosahedral	12	5′ cap, no poly(A)	0	0?	
Coronavirus[c]	–	IBV, human coronaviruses, MHV, TGEV, bovine coronavirus, and others	Helical nucleocapsid	18	5′ cap, 3′ poly(A)	5 (6)	2 or more	Only plus stranded virus with helical nucleocapsid

Abbreviations: *SVDV,* swine vesicular disease virus; *EMC,* encephalomyocarditis virus; *ME,* Maus-Elberfeld virus; *FMDV,* foot-and-mouth disease virus; *VESV,* vesicular exanthema virus of swine; *SMSV,* San Miguel sea lion virus; *SFV,* Semliki Forest virus; *MID,* Middelburg virus; *WEE,* Western equine encephalitis virus; *VEE,* Venezuelan equine encephalitis virus; *EEE,* Eastern equine encephalitis virus; *HJ,* Highlands J virus; *JE,* Japanese encephalitis virus; *SLE,* St. Louis encephalitis virus; *IBV,* (avian) infectious bronchitis virus; *MHV,* murine hepatitis virus; *TGEV,* transmissable gastroenteritis virus
[a] Matthews (1982); [b] Little is known of the replication of the two other alphavirus genera, the rubiviruses and the pestiviruses; [c] Siddell et al. (1982)

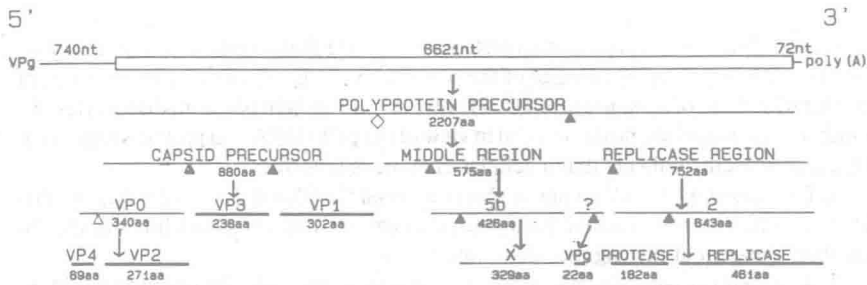

Fig. 1. Translation of poliovirus RNA. The *top line* shows the organization of the genome RNA of the Mahoney strain of poliovirus with a *single line* for untranslated regions at the 5′ and 3′ ends and an *open box* for the translated sequence. Translated products are indicated *below*, with final protein products shown as *heavy lines* and designated according to function (i.e., protease, replicase, etc.). Virion polypeptides are labelled *VP0, VP1, VP2, VP3*, and *VP4*. *VPg* is the genome linked virus polypeptide. *Solid triangles* indicate cleavages between Gln and Gly and are presumably all due to the virus protease encoded in the replicase region. The *open diamond* on the polyprotein precursor is a Tyr-Gly cleavage and the *open triangle* on *VP0* is an Asn-Ser cleavage. Data for this figure are from *Kitamura* et al. (1981), *Semler* et al. (1981a, b), and *Racaniello* and *Baltimore* (1981). *nt*, nucleotide; *aa*, amino acid

nucleotides to a UAG codon which lies 72 nucleotides from the 3′ terminal poly(A) tract. The polyprotein precursor of 2207 amino acids is organized into three regions. The amino terminal region contains the sequences of the four capsid proteins. The central region contains sequences of various polypeptides of unknown function. The carboxy-terminal region contains the sequences of the viral replicase and of the protease responsible for most or all of the processing of the viral polyprotein.

In the processing of the polyprotein precursor at least ten cleavages are known to occur (reviewed in *Rueckert* et al. 1979). Eight of these occur between Gln-Gly pairs (*Semler* et al. 1981a, b; *Larsen* et al. 1982) and are all almost certainly accomplished by a virus encoded protease (*Korant* et al. 1979). This protease is active not only as a free polypeptide of molecular weight 22 K (which appears to be the form active in producing the individual capsid proteins) but also as a part of various precursor polypeptides (which appears to be the form active in processing the replicase/protease precursors) (*Palmenberg* and *Rueckert* 1982). Many of these cleavages occur fairly slowly, with 15–20 min required for processing.

The first two cleavages in the processing pathway, which separate the polyprotein into the three domains referred to above, occur very rapidly, while the polyprotein is still nascent (reviewed in *Korant* 1979 and *Lucas-Lenard* 1979). In fact the entire polyprotein is produced as such in poliovirus only when processing is inhibited, such as when amino acid analogues are incorporated into the precursor or when protease inhibitors are present. The first of these cleavages, which separates the capsid protein precursor from the rest of the polyprotein, occurs between a Tyr-Gly pair (*Semler* et al. 1981b), and the protease responsible thus has a chymotryptic-like activity. Because this cleavage differs in specificity and kinetics from the remainder, the enzyme responsible may be qualitatively different. Three hypotheses can be proposed to explain this difference:

1. A *host cell protease is responsible for this cleavage. Korant* (1972) treated infected cells with inhibitors of trypsin and chymotrypsin and found a differential effect depend-

ing on the host cell, which was interpreted as evidence that a cellular protease was involved in an early stage of cleavage of the polio polyprotein. These data are in apparent conflict with those of *Summers* et al. (1972) however, and in light of recent discoveries of a number of virus-specific proteases, and in view of the polio RNA and protein sequencing data, it now seems unlikely that a cellular enzyme is involved.

2. The *same virus encoded protease that is active on the Gln-Gly bonds also cuts this Tyr-Gly bond.* This also seems unlikely although the enzyme might possess a broader specificity than apparent at first sight as discussed below.

3. A *second virus-encoded protease is responsible.* In this case the situation might be analogous to the cleavage of the nucleocapsid protein of the alphaviruses from its polyprotein precursor. This alphavirus protease activity appears to reside in the highly conserved carboxyterminal region of the nucleocapsid protein itself, possesses chymotryptic-like activity, is active on the nascent polyprotein, and appears to lose much of its activity once the site-specific cleavage it catalyzes has occurred (see Sect. 2.3.1).

The second nascent cleavage, which separates the middle region of the precursor polyprotein from the carboxyterminal region, occurs between a Gln-Gly pair (*Semler* et al. 1981a), and is presumably catalyzed by the viral protease with this specificity. Why this cleavage is so rapid in comparison to other cleavages is not clear.

Seven of the eight subsequent cleavages which are known to occur take place between Gln-Gly pairs as stated above. The eighth cleavage, which is one of the cleavages in the processing of the capsid proteins, occurs between an Asn-Ser pair (*Larsen* et al. 1982). This cleavage is the last to occur, and appears to coincide with the addition of virion RNA to the procapsid to form the mature virion (reviewed in *Rueckert* 1976). It could provide energy for the formation of the virion, or might activate the virion for subsequent disassembly upon infection. An Asn-Ser pair can be viewed as homologous to a Gln-Gly pair, and it is possible that this cleavage is performed by the same viral protease. The delay in processing could be due to a lower affinity of the enzyme for the Asn-Ser pair, perhaps requiring activation by the presence of RNA in the procapsid. Alternatively, another protease, probably also virus encoded, could catalyze this reaction.

It is unknown at present whether other cleavages occur in the processing of the poliovirus polyprotein, and if so what the enzymatic specificities involved might be. One of the major difficulties in working out the processing scheme is the large number of intermediates with varying half lives present in the infected cell. This situation is made more complex by the fact that alternative pathways of processing appear to exist. With the entire nucleotide sequence of the virus RNA now known, however, the complete details of processing should be known shortly.

The 740-nucleotide segment preceding the start codon of the major polyprotein is remarkably long for a 5′ untranslated region and could conceivably encode one or more small polypeptides not yet identified. Alternatively, the length of this region could relate to the fact that polio mRNA; unlike most eukaryotic mRNAs, lacks a cap structure and initiation of translation must recognize other features of the RNA. In this regard it is noteworthy that during in vitro translation of polio RNA, two different initiation sites appear to be used (*Ehrenfeld* 1979). The significance of this observation and its relation to the known sequence is unclear at present.

The other three groups of picornaviruses, the rhinoviruses, cardioviruses, and aphthoviruses, possess processing pathways which are virtually identical to those of the enteroviruses (*Rueckert* et al. 1980; *Sangar* 1979). Thus, the organization of the genome

and the nature of the proteases responsible for processing of protein precursors are probably the same for all of the picornaviruses. However, the amino acids at the cleavage sites in the capsid precursor of aphthoviruses and of mengovirus, a cardiovirus, are different from those in poliovirus shown in Fig. 1. Thus, whereas the three cleavage sites are Asn-Ser, Gln-Gly, and Gln-Gly for poliovirus, as noted above, they have been found to be Ala-Asp, Glu-Gly, and Gln-Thr, respectively, in one strain of foot-and-mouth disease virus (*Boothroyd* et al. 1981), and Ala-Asp, Gln-Ser, and Gln-Gly, respectively, in mengovirus (*Ziola* and *Scraba* 1976). If the virus-encoded protease is responsible for these cleavages, the enzyme may have a specificity less stringent than appears to be the case from a study of the cleavage sites in poliovirus and/or the specificity of the enzyme(s) may vary from virus to virus. We also note that the cardioviruses and the aphthoviruses differ from the other two groups in having a poly(C) tract of 100–500 nucleotides (*Brown* et al. 1974) in the 5′ untranslated region (*Sangar* et al. 1980) whose function is unknown.

Many virus groups inhibit translation of host mRNAs after infection. Because the picornaviruses lack the 5′ cap structure, they could conceivably interfere with translation of host messenger at the level of cap recognition. Such a mechanism has in fact been proposed for poliovirus (*Trachsel* et al. 1980; *Hansen* and *Ehrenfeld* 1981), although encephalomyocarditis virus seems to inhibit host protein synthesis in a different fashion (*Jen* et al. 1980). Of the other RNA viruses whose 5′ terminal structure has been studied, only the caliciviruses and several groups of plant viruses lack a cap and could also use such a mechanism.

Because the entire genome of the picornaviruses is translated as a continuous polypeptide chain, the only possible mechanism for regulation of the relative amounts of capsid proteins versus nonstructural proteins is premature termination of protein synthesis. Such premature termination is known to occur in vitro (*Ehrenfeld* 1979) and results in capsid proteins being produced in proportionately larger amounts. *Rueckert* (1976) has reviewed the evidence that such a regulatory mechanism operates in vivo as well as in vitro.

2.1.2 Replication of Viral RNA

The picornaviruses do not produce subgenomic RNAs and only two enzymatic activities are needed to replicate viral RNA: a minus strand replicase to produce full length minus strands using the plus strand as a template, and a plus strand replicase to produce full length plus strands from a minus strand template (reviewed in *Rekosh* 1977). Because the number of plus strands produced is much greater than the number of minus strands, it is possible that the two activities are not identical.

Although the complete nucleotide sequence has only been reported for poliovirus type 1, a number of authors have determined the 3′ and 5′ terminal sequences of other picornaviruses. *Hewlett* and *Florkiewicz* (1980) examined two strains of poliovirus and one strain of coxsackie virus, and *Nomoto* et al. (1981) examined three strains of poliovirus. The first ten nucleotides from the 5′ terminus are identical in all these viruses and the next ten nucleotides are almost identical in the poliovirus strains, and 50% conserved in coxsackie virus. *Harris* (1980) examined the 5′ sequences of nine aphthoviruses and found that the first 27 nucleotides were highly conserved. As shown in Fig. 2A, this conserved 27 nucleotide stretch of the aphthoviruses is also highly conserved between

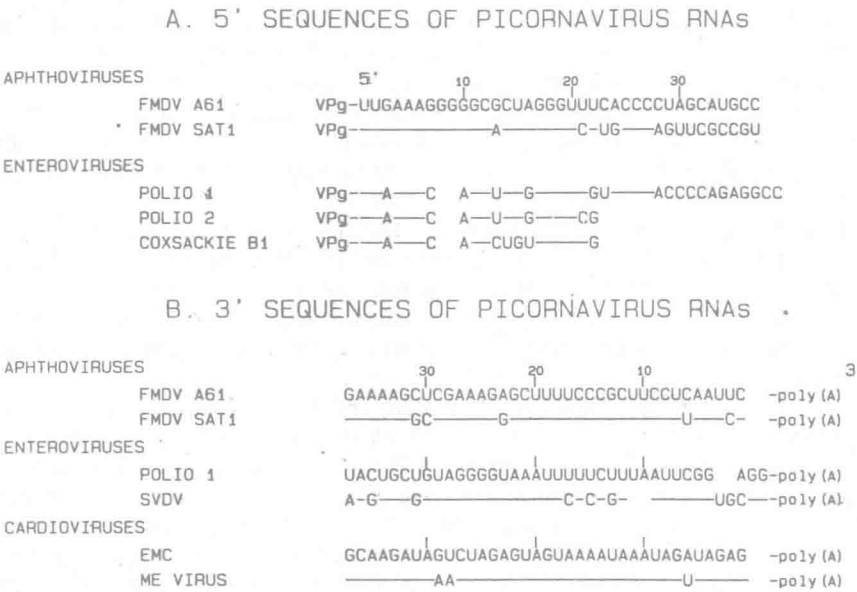

Fig. 2A, B. The 5′ and 3′ terminal sequences of picornaviruses. Sequences are shown from 5′ to 3′ reading left to right. *Horizontal lines* indicate that the nucleotide is identical with the nucleotide in the complete sequence shown above. *Gaps* have been introduced for alignment. Sequencing data are from *Fellner* (1979), *Harris* (1980), *Hewlett* and *Florkiewicz* (1980), *Nomoto* et al. (1981), and *Kitamura* et al. (1981). *FMDV,* foot and mouth disease virus; *SVDV,* swine vesicular disease virus; *EMC,* encephalomyocarditis virus; *ME,* Maus-Elberfeld virus

aphtho and polioviruses: counting deletions as single changes, two-thirds of the nucleotides are conserved.

Hewlett and *Florkiewicz* (1980) suggested the conservation they observed was a recognition site for the host translation system, but we feel it likely that the complement of the conserved sequence in the minus strand forms a recognition site for the viral plus strand replicase. We note also that the first 40 nucleotides or so of poliovirus RNA (*Larsen* et al. 1981) and of aphthovirus RNA (*Harris* 1980) can form a stable hairpin structure which could be involved in replication (see also Sect. 2.3.2) or in translation. The size of the hairpin structure and the nucleotides used to form it differ between the polioviruses and aphthoviruses, however, and this structure and the conserved sequence might serve different functions. Thus the hairpin could be involved in translation and the conserved sequence in replication.

The 3′ terminal sequences of the picornaviruses show strong conservation within a genus but no detectable conservation between genera (*Fellner* 1979). Representative data are shown in Fig. 2B. If production of minus stranded RNA from the plus strand involves a recognition sequence, this sequence has diverged markedly among the picornavirus genera. Note that the 3′ terminus has little or no homology with the complement of the 5′ terminus and if these are the initiation recognition signals then the enzyme complex could differentiate between plus and minus strand synthesis.

Initiation of RNA replication, whether plus stranded or minus stranded, is thought to

involve VPg, the 22 residue (in poliovirus) polypeptide covalently linked to the 5′ terminal U of picornavirus RNAs; linkage is through a phosphodiester bond to a tyrosine residue (*Rothberg* et al. 1978; *Ambrose* and *Baltimore* 1978; *Wimmer* 1979). (Several other virus groups are also known to possess a VPg, see Sect. 2.4 and 3.) All nascent RNA strands, both plus and minus, of poliovirus possess a covalently linked VPg (*Petterson* et al. 1978). Initiation of an RNA strand may involve the 85K polypeptide (or perhaps a shortened version of it), encoded in the 3′ part of the poliovirus genome (and identified as "Replicase Region" in Fig. 1), forming a covalent bond with the initiating U through the VPg component. This is followed by or is concurrent with cleavages which result in VPg being formed and the protease activity being released. The 52K replicase component would also be released in this reaction and would elongate the initiated chain (*Palmenberg* et al. 1979). This model would imply that the picornavirus replicase is not a true catalytic enzyme, but rather that each replicase molecule can produce only a single RNA chain. This could explain why complementation between mutants of poliovirus is both very inefficient and asymmetric. In one case where complementation could be demonstrated between a polymerase mutant and a capsid protein mutant, no polymerase mutant genomes were found in the progeny (*Cooper* 1965; *Cooper* 1969). Furthermore, the observation that defective interfering RNAs of poliovirus must be translated to produce the replicase in order to replicate (see Sect. 5) could be explained by such an RNA replication mechanism. The translation strategy of the virus leads to the production of large amounts of replicase, however, and such a mechanism is not only feasible but may be related to the overall replication strategy of this group of viruses.

Results from in vitro experiments are compatible with this model. A soluble RNA-dependent RNA polymerase, which contains primarily a single virus-specific polypeptide variously identified as p63, p58, or p56, has been isolated from cells infected by poliovirus (*Flanegan* and *Baltimore* 1979; *Etchison* and *Ehrenfeld* 1980) or foot and mouth disease virus (*Lowe* and *Brown* 1981). The most highly purified preparations are template dependent but require an oligo(U) primer to initiate replication of picornaviral RNA. However, a host factor has been partially purified which appears to allow initiation of poliovirus RNA by the replicase (*Dasgupta* et al. 1980), although the RNA product has not been characterized. With either mode of initiation the replicase activity is not picornavirus specific but will replicate other poly(A)-containing RNAs as well. Thus any specificity of the initiation event has been lost by the soluble systems isolated to date. The model presented above predicts that specific initiation requires the precursor labeled "replicase region" in Fig. 1 and further efforts to obtain a specific replicase system are clearly needed.

It is of considerable interest that the aphthoviruses have been found to contain three different VPg's which are equally represented in the virion RNA population. These three VPg's are tandemly arranged in the precursor polyprotein (*J.J. Rowlands*, personal communication). The significance of this observation is unclear at present.

The VPg is removed from the RNA destined to become mRNA, apparently by a cellular enzyme (*Ambros* and *Baltimore* 1980). It is unclear whether removal of the VPg is essential for efficient translation. It is also unclear whether VPg has any function in the virus life cycle other than in the hypothetical initiation of RNA replication. It could conceivably have an encapsidation function (since only VPg-linked RNA is encapsidated) and/or could be used to regulate the amount of RNA to be encapsidated rather than translated.

Replication of virus RNA occurs on membranes in factories called replication complexes (reviewed in *Rekosh* 1977). The function of the membrane association is unclear, although we note that a number of other viruses also replicate in association with membranes (see below). Virus replication apparently occurs completely within the cytoplasm, and transcription of the host DNA is not required after infection.

2.2 The Flaviviruses

The flaviviruses are a group of enveloped viruses which replicate in both their vertebrate hosts and the arthropod vectors, generally ticks or mosquitoes. The virions consist of an icosahedral nucleocapsid surrounded by a lipoprotein envelope. The capsid contains the genomic RNA, 12 kb in length, complexed with a single species of nucleocapsid protein, V2 or C, which has a molecular weight of 13 K. The viral envelope contains a large glycoprotein, V3 or E, of molecular weight 51–59 K, depending on the virus, and a small (7–8 K) membrane-associated protein, V1 or M, which is not glycosylated (*Westaway* et al. 1980). Although the flavivirus group includes a number of important human pathogens, the molecular biology of their replication is not well understood. This reflects the fact that these viruses do not grow well in tissue culture, that the virions are relatively unstable and difficult to purify, and that many of these viruses are severe pathogens. Some of the members of this group are listed in Table 1. The structure of flaviviruses has been recently reviewed by *Russell* et al. (1980) and the replication of these viruses recently reviewed by *Westaway* (1980). We note that these viruses are classified together with the alphaviruses as togaviruses (Table 1), but that the replication strategies of alphaviruses (Sect. 2.3) and flaviviruses differ significantly. Moreover, the recent discovery of six subgenomic polyadenylated RNAs in cells infected with equine arteritis virus, an unclassified nonarthropod borne togavirus, illustrates the diversity of replication strategies within this taxonomic family (*van Berlo* et al. 1982). Not enough is known of the replication of the other two groups of togaviruses, rubiviruses and pestiviruses, to compare them with the former groups.

The flavivirus RNA is capped but lacks poly(A) (*Wengler* and *Wengler* 1981) and is infectious. The infecting RNA must therefore be translated to produce the viral replicase, but the translation strategy of the flavivirus genome has not been definitively established. No evidence for a subgenomic RNA has been reported and it is generally believed that the viral RNA is the only messenger. *Westaway* (1980) has proposed that the structural polypeptides V1, V2, and V3 as well as nonstructural peptides P20, P27, P37, P71, and P100 are separately initiated and terminated during translation, which would make the flavivirus mRNA unique since most other animal mRNAs studied to date have only one or at most two translation initiation sites; see footnote b in Table 3 (Sect. 4). The data supporting this hypothesis (reviewed in *Westaway* 1980) are (a) no evidence for precursor polyproteins has been found in pulse-chase experiments, and peptide mapping has shown that virtually all of the flavivirus polypeptides found in infected cells are distinct, (b) pactamycin mapping indicates that proteins are completed in order of their size, and (c) reinitiation of protein synthesis following a high salt block results in very rapid labeling of all of the virus proteins. On the other hand, *Wengler* et al. (1979) and *Svitkin* et al. (1981) reported that during translation in vitro only a single initiation site appeared to be used and that only structural protein polypeptides were produced. They

Fig. 3. Terminal sequences of flavivirus RNA. The sequences shown are the 5′ (reading from 5′ to 3′) and the 3′ terminal sequences (reading from 3′ to 5′) of West Nile virus RNA (*Wengler* and *Wengler* 1981)

proposed a genetic map based on these results of 5′-V2-V3-(V1, P20, P27, P37, P71, P100)-3′ and it is of note that these results imply that the structural proteins are encoded in the 5′ end of the genome, as is the case for the picornaviruses. The complete sequence of a flavivirus genome, together with sequence information on the proteins, will probably be required to resolve the situation. In this light it has been found recently that the three structural proteins of Saint Louis encephalitis virus are not blocked and two of them do not begin with methionine (*J.R. Bell, R. Kinney, D.W. Trent, J.H. Strauss,* manuscript in preparation). This suggests that these three proteins are produced, at least in their final form, by posttranslational cleavage.

Details of flavivirus replication have yet to be worked out, and nothing is known about the viral replicase. The 5′ and 3′ terminal sequences of flavivirus RNA are different, implying that the recognition sequences for the plus stranded and minus stranded replicase are different (Fig. 3). The flaviviruses can replicate in arthropod cells (mosquito or tick, depending on the virus) and in a wide range of vertebrate cells. This wide host range implies that any functions supplied by the host during replication must be common to a broad phylogenetic range. It is also known that RNA replication is associated with perinuclear membranes (reviewed in *Westaway* 1980).

2.3 The Alphaviruses

Alphaviruses are enveloped viruses, approximately 70 nm in diameter, which replicate in both the arthropod vectors and their mammalian or avian hosts. The virus consists of an icosahedral nucleocapsid surrounded by a lipid bilayer in which are anchored two integral membrane glycoproteins. The three principal virion polypeptides, the capsid protein C, molecular weight 30 K, and the envelope proteins E1 and E2, molecular weights 50–60 K, are present in the virion in equimolar amounts. A third glycoprotein, E3, remains associated with the virion in Semliki Forest virus but is lost into the culture fluid for other alphaviruses. The alphavirus genome is a single stranded RNA of about 12 kb which is capped and polyadenylated, and which is infectious. Most of the molecular biology of these viruses has been determined with either Sindbis virus or Semliki Forest virus, but recent comparative studies with other alphaviruses are giving us better insights into the relationships among the members of this group. An extensive collection of review articles on these viruses has recently appeared (*Schlesinger* 1980).

2.3.1 Translation Strategy

The alphaviruses produce two mRNAs after infection (reviewed in *Strauss* and *Strauss* 1977). One is apparently identical to the virion RNA and is translated into the nonstructural proteins of the virus. The second is a subgenomic RNA identical to the 3′ terminal one-third of the genomic RNA which is translated into the structural proteins of the virus.

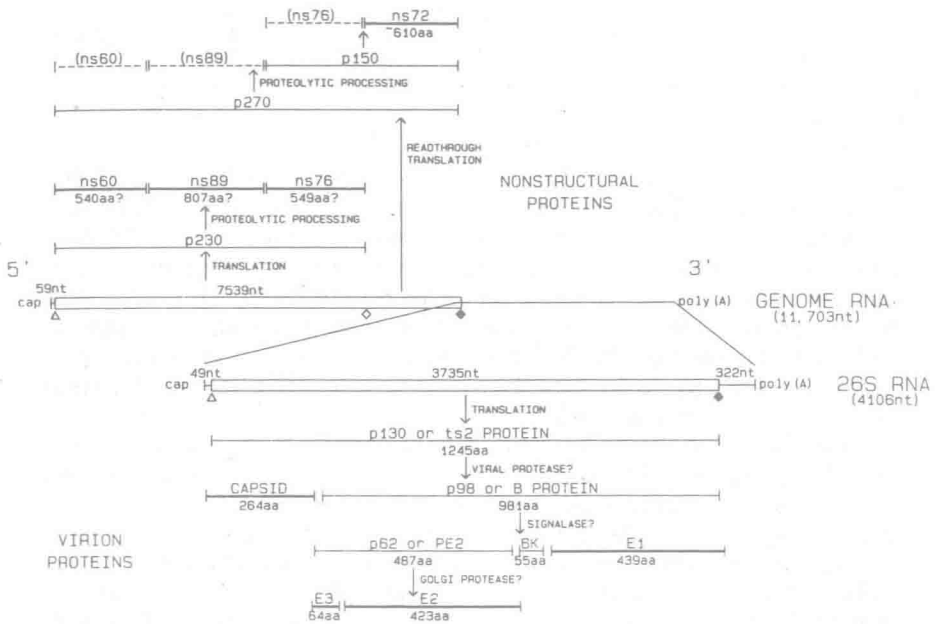

Fig. 4. Replication strategy of Sindbis virus. Untranslated regions of the genomic RNA are shown as *single lines,* and the translated region as an *open box.* The subgenomic RNA region is expanded below using the same convention. Translation products are indicated and the final protein products, both virion and non-structural, are indicated with *heavy lines. Open triangles* are initiation codons, *solid diamonds* are termination codons. The *open diamond* is the UGA codon read through to produce ns72. Data for this figure are from *Ou* et al., 1982a, b, 1983; *Rice* and *Strauss, 1981; Strauss,* et al., 1983a; *E.G. Strauss* unpublished; and *S. Lopez* and *J.R. Bell,* unpublished

Both of these RNAs are capped and polyadenylated. These RNAs and their translation products are illustrated schematically in Fig. 4.

The genomic RNA is a minor message in the infected cell and encodes the non-structural proteins necessary for viral RNA replication. Translation begins at an AUG codon approximately 60–80 nucleotides depending on the alphavirus from the 5′-terminal cap (Ou et al., 1983). As diagrammed in Fig. 4 for Sindbis virus two polyprotein precursors are produced: the major product terminates at an opal codon at nucleotides 5748 to 5750 (open diamond) which interrupts an otherwise open reading frame encoding 2513 amino acids. A minor polyprotein is produced by read through and terminates at multiple in-phase stop codons (solid diamond). The major polyprotein is processed, usually while nascent, to produce the three upstream products, which have been identified following translation in vitro and in extracts of infected cells (reviewed in *Schlesinger* and *Kääriäinen,* 1980; Collins et al., 1982). Genetic analysis had suggested that there were four non-structural polypeptides essential for RNA replication (reviewed in *Strauss* and *Strauss,* 1980). The fourth product, ns72, has recently been identified in Sindbis-infected cell extracts following immunoprecipitation with an antibody directed against a synthetic dodecapeptide with the amino acid sequence of the carboxyl terminus of the 2513 amino acid precursor (*Lopez* and *Bell,* unpublished). The functions of the non-structural polypeptides must include replicase/transcriptase components to replicate the RNA and

transcribe the subgenomic message. In addition, one of the products may be a virus-specific protease, to process these precursors. The read through mechanism allows modulation of the relative amounts of the replicase components; ns72, produced in the smallest amounts, has been suggested to be a regulatory factor for the control of minus strand synthesis (*Strauss* et al., 1983a).

The structural proteins of the virus are translated from a subgenomic messenger. This subgenomic RNA has been completely sequenced in the case of three alphaviruses, Semliki Forest virus (*Garoff* et al. 1980a, b), Sindbis virus (*Rice* and *Strauss* 1981; *Ou* et al. 1982a), and Ross River virus (*L. Dalgarno* et al. 1983 (in press), and corresponds to the 3′ terminal one-third of the genomic RNA. Use of a subgenomic mRNA for the structural proteins allows for amplification of the structural gene products. The subgenomic RNA is produced in about threefold molar excess over the genomic RNA (also see below) and, in addition, much of the genomic RNA is quickly sequestered into nucleocapsids, where it cannot serve as messenger. The result is that 90% of the virus-specific mRNA is the subgenomic species, and only 10% is the genomic RNA; thus a large excess of structural over nonstructural polypeptides is produced (reviewed in *Strauss* and *Strauss* 1977). Because of the use of an infectious genomic RNA and a subgenomic RNA for the structural proteins, the replicase genes are 5′ terminal and the structural protein genes 3′ terminal the inverse order from that of the picornaviruses (see Sect. 2.1.1.).

Translation of the structural proteins from the 4100-nucleotide subgenomic RNA begins at an AUG codon located approximately 50 nucleotides from the 5′ terminal cap (*Ou* et al. 1982a) and proceeds to a termination codon positioned 260–520 nucleotides from the 3′ terminal poly(A) tract (*Rice* and *Strauss* 1981; *Garoff* et al. 1980b; *L. Dalgarno*, et al. 1983 (in press)). Cleavage of the N-terminal capsid protein from the nascent precursor appears to be an autoproteolytic event (reviewed in *Schlesinger* and *Kääriäinen* 1980) and the chymotryptic-like activity, which cuts a tryptophan-serine bond, is thought to reside in the C-terminal region of the capsid protein itself. This cleavage event is not only rapid but quite efficient, and no uncleaved products in found when wild type RNA is translated in vivo or in vitro. Normally the cleavage of the capsid protein from the nascent chain appears to be accomplished by the protease activity in the nascent chain itself, rather than by protease activity in previously released capsid proteins, and much of the proteolytic activity may be lost upon cleavage. At least some proteolytic activity appears to remain, however. Mutants temperature sensitive in the protease activity accumulate large amounts of uncleaved precursor during infection at nonpermissive temperatures. In cells doubly infected with such a mutant, and with mutants defective in the glycoproteins but having a functional protease, the precursor is found in smaller amounts, implying that the mutant polyprotein can be cleaved by a diffusible factor (*Scupham* et al. 1977).

The remaining structural proteins are two integral membrane glycoproteins which traverse the lipid bilayer and are anchored in the bilayer by short hydrophobic stretches found at or near the C-terminus of the proteins (*Garoff* and *Söderlund* 1978; *Rice* et al. 1982). Removal of the capsid from the nascent precursor polyprotein allows a signal sequence of about 19 residues at the N-terminus to function and results in the integration of the first glycoprotein precursor (called PE2, the precursor to glycoprotein E2) into the endoplasmic reticulum, with concomitant core glycosylation (*Garoff* et al. 1979; *Bonatti* et al. 1979; *Bell* et al. 1982). This signal sequence is not cleaved from the precursor at this stage (*Bonatti* and *Blobel* 1979). There is a second, internal signal sequence located

between the two glycoproteins which functions to allow insertion of the second glycoprotein (called E1) into the endoplasmic reticulum (*Hashimoto* et al. 1981), again accompanied by core glycosylation. Removal of this internal signal sequence, which separates the two glycoproteins from one another, requires two proteolytic cleavages, both of which occur after alanine residues. It has been suggested that signalase catalyzes both of these cleavages (*Rice* and *Strauss* 1981).

The glycoproteins, once synthesized and inserted into the endoplasmic reticulum, migrate to the plasma membrane by way of the Golgi apparatus. The cleavage of PE2 to form E2 and E3 has been postulated to occur in the Golgi (*Garoff* et al. 1980b; *Rice* and *Strauss* 1981), catalyzed by the Golgi protease whose specificity is such that it cleaves after clustered basic amino acids, and which cleaves proalbumin, proinsulin, and other precursor proteins (*Dean* and *Judah* 1980). The small glycoprotein produced, E3, is not required for infectivity and may or may not remain associated with the virion. This Golgi protease also appears to cleave glycoproteins of several other enveloped viruses (see below) in addition to the PE2 of alphaviruses.

Thus the cleavage of the alphavirus polyproteins is postulated to require one or more virus proteases active on the nonstructural precursor polyprotein, a virus protease activity present in the capsid protein which acts autoproteolytically, and two cellular proteases, both of which are localized in subcellular organelles.

The alphaviruses inhibit translation of host cell messenger RNAs, apparently by increasing the Na^+ concentration and lowering the K^+ concentration inside the cell (*Garry* et al. 1979a). The virus messengers are efficiently translated under these altered conditions, whereas most host cell mRNAs are not. The interference with translation is at the level of initiation. The virus structural proteins may be implicated in this inhibition (*Atkins* 1976), and it has been suggested that the altered ionic environment inside the cell results from interference with the Na^+/K^+ pump (*Garry* et al. 1979b).

2.3.2 Replication and Transcription of the RNAs

During the course of alphavirus replication three RNA synthesis activities are needed: a minus strand replicase to produce full length minus strands using the plus strand as a template; a plus strand replicase to produce full length plus strands using the minus strand as a template; and a transcriptase to produce the subgenomic messenger RNA for the structural proteins, which uses the minus strand as a template. Each of these activities appears to be independently regulated and different recognition sites for the corresponding enzymes are utilized. We postulate that four activities are involved: an elongation enzyme which synthesizes the RNA chain once properly initiated; and three initiation activities which initiate minus strands, plus strands, and the subgenomic mRNA, respectively. These activities could reside in four different polypeptide chains and compose four different complementation groups, or two or more of these activities could reside in the same polypeptide chain. Similarly, the initiation and elongation functions could be expressed as an enzyme which is a functional complex of several polypeptide chains, or the component parts could function as separate enzymes.

Work with temperature-sensitive mutants has supported this concept of several functions involved in RNA synthesis. In the case of Sindbis virus, four complementation groups are required for normal RNA synthesis after infection (*Strauss* and *Strauss* 1980). One group (F) appears to encode an elongation function because, upon shifting cells