AD_____

CONTRACT NO: DAMD17-90-C-0050

TITLE: MOLECULAR STUDIES OF ALPHAVIRUS IMMUNOGENICITY

PRINCIPAL INVESTIGATOR: James H. Strauss, Ph.D.

CONTRACTING ORGANIZATION: California Institute of Technology
1201 E. California Boulevard
Pasadena, California 91125

REPORT DATE: April 23, 1992

TYPE OF REPORT: Annual Report

PREPARED FOR: U.S. ARMY MEDICAL RESEARCH AND DEVELOPMENT COMMAND
Fort Detrick, Frederick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for public release;
distribution unlimited

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION: Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION / AVAILABILITY OF REPORT: Approved for public release; distribution unlimited
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION: California Institute of Technology; 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code): 1201 E. California Boulevard, Pasadena, California 91125	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING / SPONSORING ORGANIZATION: U.S. Army Medical Research & Development Command; 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: Contract No. DAMD17-90-C-0050
8c. ADDRESS (City, State, and ZIP Code): Fort Detrick, Frederick, Maryland 21702-5012	10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.
61102A	3M1-61102BS12	AB	WUDA346101

11. TITLE (Include Security Classification)

MOLECULAR STUDIES OF ALPHAVIRUS IMMUNOGENICITY

12. PERSONAL AUTHOR(S)
James H. Strauss

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Annual Report	FROM 3/30/91 TO 3/29/92	23 April 1992	

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Sindbis Virus; Alphavirus; Ockelbo Disease; RA I; BD; Antigenic Epitope; Immunogenicity
06	02		

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

In the past year, our studies on the immunogenicity of alphaviruses have focused on two areas: 1) Determination of the binding site on viral glycoproteins for neutralizing antibodies. Neutralizing antibodies bind to surface structures of a virus and interfere with the uptake or uncoating of the virus. In one case we localized the binding site for a neutralizing monoclonal antibody by using λgt11 expression libraries, following screening of libraries containing random cDNA inserts from Sindbis virus RNA for reactivity with monoclonal antibodies that neutralized the infectivity of Sindbis virus. When combined with sequencing studies of monoclonal escape variants this has allowed us to define an immunogenic domain of alphavirus E2. 2) Sequence analysis of alphaviruses. We are sequencing a number of different strains of Sindbis virus and of its relatives in order to understand the geographic distribution of these viruses and their disease potential. These viruses include Aura virus from South America and Whataroa virus from New Zealand. We have developed cloning methods that have allowed us to obtain cDNA clones representing all of a viral genome which are suitable for high throughput automated DNA sequencing. This has made it possible to acquire sequence data at a much more rapid rate.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
☐ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code)	22c. OFFICE SYMBOL
Virginia Miller	301-619-7325	SGRD-RMI-S

DD Form 1473, JUN 86 *Previous editions are obsolete.* SECURITY CLASSIFICATION OF THIS PAGE

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the US Army.

_____ Where copyrighted material is quoted, permission has been obtained to use such material.

_____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

_____ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

_____ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

____ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

JHS In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

J H Strauss
PI - Signature

April 23, 1992
DATE

Table of Contents

Introduction

The alphaviruses are a widespread group of human pathogens that are present in many parts of the world (Griffin, 1986; Monath, 1988; Peters and Dalrymple, 1990). They are mosquito-borne and are particularly prevalent in tropical and subtropical areas of the world, but alphaviruses pathogenic for man are also present in temperate areas. Many alphaviruses are capable of causing fever, rash and arthralgia in man that in some cases may be disabling for extended periods of time. Many of the New World alphaviruses can cause encephalitis in man. Our program attempts to understand the molecular basis of alphavirus immunogenicity and to determine the relationships of alphaviruses and strains of alphaviruses to one another.

We reported last year that strains of Sindbis virus from Northern Europe referred to as Ockelbo virus and Karielian fever virus, which cause an illness characterized by polyarthritis whose symptoms can persist for months or years, were very closely related to pathogenic strains of Sindbis virus isolated from South Africa (Shirako et al., 1991). We concluded that a South African strain of Sindbis virus was introduced into Northern Europe, probably in the 1960's, either by the activities of man or by migratory birds, and this led to epidemics of Ockelbo disease in Sweden. The virus then spread to Finland and the Karelian region of the Soviet Union, probably in the 1980's, causing epidemics of disease called Pogosta disease and Karelian fever, respectively. We also found that repeated sequence elements found in the 3' nontranslated region of Sindbis viruses are much more highly conserved than sequences outside these elements, and concluded that these repeated elements must play an important role in RNA replication.

In the past year we have continued our sequencing efforts on alphaviruses in order to determine the relations of these viruses to one another. These have included the nsP3-nsP4 regions of a South African strain of Sindbis virus in in order to examine the relationships within these domains between South African Sindbis viruses and the Northern European Ockelbo viruses

begun last year. We have also examined Whataroa virus, a virus related to Sindbis virus isolated from New Zealand, an Indian isolate of Sindbis virus, and an Australian isolate of Sindbis virus in order to determine the relationships of these viruses to one another. We have begun sequencing of Aura virus in order to search for emergent viruses. We had previously found that Western equine encephalitis virus, found throughout North and South America, arose by recombination between Eastern equine encephalitis virus and a New World alphavirus related to Sindbis virus. Aura virus has been isolated in Brazil and Northern Argentina and is known from serological studies to be related to Sindbis virus. We wished to determine if Aura virus was the second parent of Western equine encephalitis virus.

We are also interested in the localization of neutralizing antibody-binding sites in alphaviruses. The knowledge of immunogenetic domains is important in developing vaccines. Neutralizing antibodies bind to the glycoproteins of alphaviruses and prevent them from attaching to susceptible cells or prevent them from penetrating cells. The exact mechanisms by which neutralizing antibodies inactivate a virus are somewhat controversial and differ from case to case, but at least in some cases the antibody neutralizes by binding to the structure on the surface of the virus that interacts with a receptor on the cell surface, thus directly blocking the virus from interacting with its receptor. In these cases anti-idiotypic antibodies made against such antibodies may function as anti-receptor antibodies. In studies of antibody escape variants we have identified domains of glycoprotein E2 which appear to be important for virus neutralization (Strauss et al., 1991). Here we report that we have been able to use λgt11 expression libraries to directly demonstrate interaction between a neutralizing antibody and a specific domain of glycoprotein E2. Such a result is significant because cases have been described in which resistance to a monoclonal antibody (mAb) arose from single amino acid substitutions away from the actual antibody binding site (Diamond et al., 1985; Parry et al., 1990). Thus it is possible to induce changes in conformation of the antibody-binding regions with amino acid substitutions outside the epitope, and direct demonstration of antibody binding to a defined region is important. Because

this neutralizing monoclonal antibody used here elicits production of anti-idiotypic antibodies which act as anti-receptor antibodies in chicken cells (Wang et al., 1991), this domain is also implicated in attachment to the surface of a susceptible cell. A complete description of these results has appeared in the *Journal of Virology* **65**, 7037-7040 (1991). A preprint of this paper entitled "Use of a λgt11 expression library to localize a neutralizing antibody-binding site in glycoprotein E2 of Sindbis virus," by K. S. Wang and J. H. Strauss, was submitted to the U.S. Army Medical Research and Development Command at the time of submission to the journal.

Methods Used

Virus Strains. South African strains of Sindbis virus, Whataroa virus, Indian and Australian isolates of Sindbis virus, and Aura virus were obtained from Dr. J. M. Dalrymple of USAMRIID. Viruses were grown and purified as previously described (Shirako et al., 1991).

cDNA clones for most of the viruses were produced using standard methods (Sambrook et al., 1989). First strand cDNA was made using oligo(dT) as a primer and second strand synthesis was by the method of Gubler and Hoffman (Gubler and Hoffman, 1983). In some cases *Hin*dIII fragments of the cDNA were cloned into vector pGem3Z. In other cases *Eco*RI linkers were added to double-stranded cDNA and the cDNA cloned into the *Eco*RI site of pGem3Z. DNA sequencing and RNA sequencing used standard technology that is in common use in our laboratory (Hahn et al., 1989; Rice et al., 1985; Rice and Strauss, 1981; Shirako et al., 1991; Strauss et al., 1984).

Construction of a Random cDNA Library of Virus RNA. We have also developed methods suitable for high throughput automated DNA sequencing in order to speed up the acquisition of sequence data. For this we used Whataroa virus, strain M78, isolated in 1962 at Westland, New Zealand, from *Culex pervigilans*, as a test virus. The virus was propagated once in primary chicken fibroblast cells and purified by sucrose gradient centrifugation. The RNA was

extracted by an SDS/phenol method, precipitated in ethanol, and suspended in water. First strand cDNA was synthesized with 2 μg of virus RNA using 200 pmol of $pd(N)_6$ and 2 pmol of dT_{17} by AMV reverse transcriptase. The second strand DNA was synthesized by the Gubler and Hoffman (Gubler and Hoffman, 1983) method. The double-stranded cDNA was blunt ended with T4 DNA polymerase in the presence of RNase A, extracted with phenol/chloroform, and precipitated with ethanol. After methylating internal *Eco*RI sites with *Eco*RI methylase, the DNA was electrophoresed in an LMP agarose gel and a 2-4 kb fraction was isolated by a CTAB method as described elsewhere (Shirako and Strauss, 1992). The isolated DNA was kinased with T4 DNA polynucleotide kinase and ligated to kinased *Eco*RI linkers. The ligation products were digested with *Eco*RI, extracted with phenol/chloroform, precipitated with ethanol, and electrophoresed in an LMP agarose gel. The 2-4 kb fraction was isolated by a CTAB method and ligated to an *Eco*RI-digested, CIAP-treated pGEM3Z vector. The ligated DNA was transformed into *E. coli* JM109. One hundred clones that appeared to contain inserts were selected randomly and characterized by restriction analysis of the DNA prepared from 0.5 ml of bacterial cultures. Ninety-six clones were were found to contain inserts larger than 1.0 kb. Fifty clones containing larger inserts were further selected and the DNA was prepared from 10 ml of bacterial cultures by a modified boiling method.

Construction and Screening of the Bacteriophage Library. Sindbis virus strain AR339, from A. Schmaljohn of USAMRIID, was grown in monolayers of primary chicken embryo fibroblasts (Pierce et al., 1974). Virus was purified as described (Bell et al., 1979), disrupted with 0.5% SDS, and 49S genomic RNA extracted with phenol/chloroform (Hsu et al., 1973). After two ethanol precipitations, RNA was suspended in distilled water and stored at -70°C until use as a template for cDNA synthesis.

A λgt11 library containing short inserts of Sindbis cDNA was constructed by a modification of the procedure of Young and Davis (Young and Davis, 1983). cDNA synthesis

was randomly primed with sonicated salmon testis DNA; [^{32}P]dCTP was included during cDNA synthesis in order to monitor the product. After flush-ending with the Klenow fragment of DNA polymerase I, methylation with *Eco*RI methyltransferase, and addition of *Eco*RI linkers (Collaborative Research), the modified cDNA was digested with an excess of *Eco*RI restriction enzyme. The digested cDNA was then fractionated on a Sephadex CL-6B column, and Sindbis cDNA fragments 100-300 base pairs in size were pooled and ligated to dephosphorylated λgt11 arms (Promega). After *in vitro* packaging into phage heads (Stratagene), the percentage of phage containing Sindbis virus cDNA inserts was found to be 90% by plating phage on *E. coli* Y1090 in the presence of 5-bromo-4-chloro-3-indolyl β-D-galactoside. Plaques were screened for reactivity with the various mAbs. Phage plaques were grown for 6 hrs at 42°C, nitrocellulose disks (Schleicher & Schuell) soaked in 10 mM isopropyl thio-β-D-galactopyranoside were then placed on the top of the agar layer, and the plates were transferred to 37°C for 15 hrs. The filters then were lifted and washed successively in 10 mM Tris-Cl pH 7.5 and 150 mM NaCl containing 5% nonfat milk. The filters were incubated overnight at 4°C with monoclonal antibody (10 μg/ml in PBS containing 5% nonfat milk), washed, ^{125}I-conjugated protein G (0.5 μCi/ml in 5% nonfat milk) added, and the filters were incubated for at least 2 hr at room temperature. After washing and drying, the filters were exposed overnight at -80°C to Kodak-X-OMAT film. Immunoreactive phage were picked and rescreened until a uniformly reactive population was obtained.

Sequence of the nsP3 and nsP4 Region of Alphaviruses

We have obtained the complete sequence of the nsP3-nsP4 region, approximately 3.5 kb, for four new alphaviruses. These are a South African strain of Sindbis virus isolated from a human case of Sindbis disease, Whataroa virus from New Zealand, an Indian isolate of Sindbis virus, and an Australian isolate of Sindbis virus. Information on these strains and on a number of other strains with which we are currently working, is given in Fig. 1. Shown is the name of the

Subgroup	Strain	Source	Year	Location	Cloning/sequencing status
I	AR339	*Culex univittatus*	1952	Egypt	Completed (HRsp)
	MP684	*Mansonia fuscopennata*	1958	Uganda	
	R33	*Acrocephalus scripaceous*	1971	Czechoslovakia	
	AR86	*Culex sp.*	1954	South Africa	50 clones in pGEM
	Girdwood	Human	1963	South Africa	nsP3 nsP4 only
	1038	Turtle dove	1964	Israel	
	Edsbyn82-5 (Ockelbo)	*Culiseta sp.*	1982	Sweden	Completed
II	A1036	*Bdellonyssus bursa*	1953	India	nsP3-nsP4 only
	MM2215	*Culex tritaeniorhynchus*	1955	Indonesia	50 clones in pGEM
	MRM18520	Mosquito pool	1975	Australia	nsP3-nsP4 only
III	M78 (Whataroa)	Mosquito pool	1962	New Zealand	nsP3-nsP4 completed

Figure 1 Strains of Sindbis virus used in this study

strain, the source from which the virus was isolated, the year and place of isolation, and the status of our work with the virus.

The four new sequences obtained are presented in Figs. 2 to 5. These nucleotide sequences and the amino acid sequences deduced from them illustrate the close relationships among these alphaviruses and confirm that South African strains of Sindbis virus are very closely related to Ockelbo virus and its allies. nsP4 in particular is very highly conserved. The C-terminal domain of nsP3, which is not highly conserved among alphaviruses, shows more variability, but in each case there is an opal termination codon between nsP3 and the beginning of nsP4 which must be read through in order to produce nsP4.

The relationships among these viruses are illustrated in numerical fashion in Fig. 6. South African Girdwood and Ockelbo exhibit only 1.3% sequence divergence in nsP4 and only 1.8% divergence in the conserved region of nsP3. The Indian and Australian isolates have diverged by 7-10% from these strains in nsP3 and nsP4. Whataroa virus is clearly related to these Sindbis viruses but differs by 12-16% in amino acid sequence in these regions from the Sindbis virus strains.

High Throughput Automated DNA Sequencing

Several companies, including Applied Biosystems, now make automated DNA sequencers which can greatly speed up the rate of acquisition of sequence data. In order to use such a system, random cDNA clones must constructed which represent the entire viral genome and DNA must be prepared from such clones that is highly purified and suitable for automated sequencing. We have shown that it is feasible to use the Applied Biosystems sequenator to sequence alphaviruses by using Whateroa virus as a test virus. Random cDNA clones were constructed in a plasmid vector and plasmid DNA was subjected to high throughput automated DNA sequencing. Preparation of plasmid cDNA libraries containing a representative sampling of the Whateroa genome required

Figure 2. nsP3/nsP4 of A1036 (1953, India, *Bdellonyssus bursa*)

```
   1  GCUCCGGCCUAUCGCUCGAAACGUGAGAACAUCGCCGAGUGCCUCGAAGAGGCCGUAGUU   60
      A  P  A  Y  R  S  K  R  E  N  I  A  E  C  L  E  E  A  V  V

  61  AAUGCCGCGAAUGCACUCGGACGGCCGGGCGAAGGGGUAUGCAAAGCCAUAUAUAAAAAA  120
      N  A  A  N  A  L  G  R  P  G  E  G  V  C  K  A  I  Y  K  K

 121  UGGCCUAAUAGUUUCGUCGAUUCCGCGACAGAGACUGGAACGGCUAAGCUAGUGUGCUGU  180
      W  P  N  S  F  V  D  S  A  T  E  T  G  T  A  K  L  V  C  C

 181  CAAGGAAAGAAAAUUAUCCACGCCGUCGGACCCGACUUCCGCAAACACUCCGAGGCAGAA  240
      Q  G  K  K  I  I  H  A  V  G  P  D  F  R  K  H  S  E  A  E

 241  GCACUGAAGAUUCUCCAGAACACAUACCACGCCAUAGCAGAUUUGGUUAACAAACAUGGA  300
      A  L  K  I  L  Q  N  T  Y  H  A  I  A  D  L  V  N  K  H  G

 301  AUCAAGACUGUAGCGAUCCCGCUACUAUCCACCGGGAUUUACGCAGCGGGAAAAGACAGA  360
      I  K  T  V  A  I  P  L  L  S  T  G  I  Y  A  A  G  K  D  R

 361  CUCGAGGUCUCCUUAAACUGUCUUACCACCGCCCUGGACAGAACAGACGCAGACGUCACA  420
      L  E  V  S  L  N  C  L  T  T  A  L  D  R  T  D  A  D  V  T

 421  AUCUACUGUCUAGACAAAAAAUGGAAAGAAAGGAUCGAUGCGGUUAUACAAUUGAAGGAG  480
      I  Y  C  L  D  K  K  W  K  E  R  I  D  A  V  I  Q  L  K  E

 481  UCGGUGACGGAACUGAAGGAUGAGGAUAUGGAGAUCGACGAUGAGUUAGUAUGGAUCCAC  540
      S  V  T  E  L  K  D  E  D  M  E  I  D  D  E  L  V  W  I  H

 541  CCGGAUAGUUGUCUCAAGGGCAGGAAAGGGUAUAGCACAACAAAAGGUAAACUUUAUUCG  600
      P  D  S  C  L  K  G  R  K  G  Y  S  T  T  K  G  K  L  Y  S

 601  UACUUUGAGGGGACUAAGUUUCAUCAGGCAGCAAAAGACAUGGCGGAGAUUAAAGUACUU  660
      Y  F  E  G  T  K  F  H  Q  A  A  K  D  M  A  E  I  K  V  L

 661  UUUCCCGAUGAGCAAGAGUGCAACGAGCAGUUGUGUGCAUACAUCCUUGGUGAAACCAUG  720
      F  P  D  E  Q  E  C  N  E  Q  L  C  A  Y  I  L  G  E  T  M

 721  GAAGCCAUCAGGGAAAAAUGUCCAGUGGACUUUAAUCCGUCGUCCAGUCCGCCGAAGACA  780
      E  A  I  R  E  K  C  P  V  D  F  N  P  S  S  S  P  P  K  T

 781  CUCCCCUGUUUGUGCAUGUAUGCCAUGACGCCUGAGAGAGUGCACCGUCUGCGUAGCAAC  840
      L  P  C  L  C  M  Y  A  M  T  P  E  R  V  H  R  L  R  S  N

 841  AACGUCAAGUCCAUCACAGUGUGUUCGUCUACCCCACUUCCGAAGCACAAGAUCAAGAAC  900
      N  V  K  S  I  T  V  C  S  S  T  P  L  P  K  H  K  I  K  N

 901  GUUCAGAAAGUACAGUGCACGAAAGUGGUCUUGUUCAAUCCACAGACCCCUGAAUUUGUC  960
      V  Q  K  V  Q  C  T  K  V  V  L  F  N  P  Q  T  P  E  F  V

 961  CCUGCCCGUAAGUACAUAGAAGCACAACCAAAAGACGUAAGCCAAGAUGCAGAAGAAAGC 1020
      P  A  R  K  Y  I  E  A  Q  P  K  D  V  S  Q  D  A  E  E  S

1021  CCUGCCGCAGCCGCCCGAGAUAACACCUCACGGGACGUAACAGACAUAUCCCUGGAUGUG 1080
      P  A  A  A  A  R  D  N  T  S  R  D  V  T  D  I  S  L  D  V

1081  GAAGAAAGUCAAGCCGCAGCCGGCCAACCAGAGGAGCGCUCGGGGGACAACACUUCCCGG 1140
      E  E  S  Q  A  A  A  G  Q  P  E  E  R  S  G  D  N  T  S  R

1141  GAUGUAACAGAUAUAUCCCUAGAUCACGACAGCGAUAGUGAGGUGGGCUCCAUCUUCUCU 1200
      D  V  T  D  I  S  L  D  H  D  S  D  S  E  V  G  S  I  F  S

1201  AACCUUAGCUGCUCCAGUCAAUCCAUCACUAGUAUGGACAGCUGGUCCUCCGGACCGGGA 1260
      N  L  S  C  S  S  Q  S  I  T  S  M  D  S  W  S  S  G  P  G
```

Figure 2. nsP3/nsP4 of A1036 (1953, India, *Bdellonyssus bursa*)

```
1261 UCGAUCACGAUAAACGAGAACCGCACCAUUCAGGUCACGGCGGAGAUACACAAUGCUCCU 1320
     S  I  T  I  N  E  N  R  T  I  Q  V  T  A  E  I  H  N  A  P

1321 GCCGCGUUGCCUGUUCCACCACCACGCCUUAAGAAACUGGCACGCUUAGCAGCCCAGAAG 1380
     A  A  L  P  V  P  P  P  R  L  K  K  L  A  R  L  A  A  Q  K

1381 CCCAAUCCGCCAUCCGACCCGCCUUCGACGGUCGAGGACGUGUCGAUGCGCUUGUCCUUC 1440
     P  N  P  P  S  D  P  P  S  T  V  E  D  V  S  M  R  L  S  F

1441 CCUGCCACGGUGUCGUUCGGAUCAUUCUCCGACGGAGAAGUCGACGACCUUAGCCGCGAU 1500
     P  A  T  V  S  F  G  S  F  S  D  G  E  V  D  D  L  S  R  D

1501 AAAGCAGUGUCAGAACCGGUGGUCUUUGGUGCUUUCGAGCCUGGAGAGGUAACCUCUAUC 1560
     K  A  V  S  E  P  V  V  F  G  A  F  E  P  G  E  V  T  S  I

1561 AUCGAAUCAAGGUCUGUCGUGUCAUUCCCCGUGCAUAAACGCCGGCGCAGAAGACGGGGC 1620
     I  E  S  R  S  V  V  S  F  P  V  H  K  R  R  R  R  R  R  G

1621 AAAAGAACCGAAUAUUGACUAACCGGGGUAGGUGGGUACAUCUUCUCAACUGACACGGGA 1680
     K  R  T  E  Y  *  L  T  G  V  G  G  Y  I  F  S  T  D  T  G

1681 CCGGGCCACCUCCAGAAGAAGUCAGUUCUGCAAAACCAGCUUACUGAACCGACCCUCGAG 1740
     P  G  H  L  Q  K  K  S  V  L  Q  N  Q  L  T  E  P  T  L  E

1741 CGCAAUCAAUUAGAACGAAUGUAUGCGCCCAGUCUCGAUGUCAAGAAAGAGGAACUUCUG 1800
     R  N  Q  L  E  R  M  Y  A  P  S  L  D  V  K  K  E  E  L  L

1801 AAACUUAAGUACCAAAUGAUGCCCACCGAAGCCAAUAAAAGUAGGUACCAGUCUAGAAAG 1860
     K  L  K  Y  Q  M  M  P  T  E  A  N  K  S  R  Y  Q  S  R  K

1861 GUUGAAAAUCAAAAAGCGGUAACCACCGAGAGGUUACUGUCGGGACUGAAGAUGUACAUC 1920
     V  E  N  Q  K  A  V  T  T  E  R  L  L  S  G  L  K  M  Y  I

1921 CACUCAGAGAACCAACCUGAGUGUUAUAAGGUCACUUAUCCGAAACCGUCGUACUCCAGC 1980
     H  S  E  N  Q  P  E  C  Y  K  V  T  Y  P  K  P  S  Y  S  S

1981 AGUGUCCCUCUUAGUUACCAGAACCCUGAAUUCGCCGUAGCUGUUUGCAAUAACUACCUG 2040
     S  V  P  L  S  Y  Q  N  P  E  F  A  V  A  V  C  N  N  Y  L

2041 CAUGAGAACUACCCGACGGUUGCCUCCUAUCAGAUUACGGACGAAUAUGAUGCCUACCUC 2100
     H  E  N  Y  P  T  V  A  S  Y  Q  I  T  D  E  Y  D  A  Y  L

2101 GACAUGGUGGACGGCACUGUUGCGUGUCUCGACACUGCAACAUUCUGCCCUGCGAAAUUA 2160
     D  M  V  D  G  T  V  A  C  L  D  T  A  T  F  C  P  A  K  L

2161 CGUAGCUUUCCGAAGAAACAUGAGUACCGCGCACCUAACAUCAGGAGUGCCGUGCCGUCU 2220
     R  S  F  P  K  K  H  E  Y  R  A  P  N  I  R  S  A  V  P  S

2221 GCUAUGCAGAACACUCUACAGAACGUCCUGAAUGCAGCAACAAAGAGGAAUUGCAACGUU 2280
     A  M  Q  N  T  L  Q  N  V  L  N  A  A  T  K  R  N  C  N  V

2281 ACUCAGAUGAGAGAACUACCGACCCUAGACUCCGCGACCUUUAACGUGGAAUGCUUCCGA 2340
     T  Q  M  R  E  L  P  T  L  D  S  A  T  F  N  V  E  C  F  R

2341 AAGUACGCGUGCAAUGACGAGUAUUG[illegible]GCUGAAUUCUCCGAAAAACCAAUCAGGAUCACC 2400
     K  Y  A  C  N  D  E  Y  W  A  E  F  S  E  K  P  I  R  I  T

2401 ACGGAGUUUGUUACGGCGUACGUGGCGAGAUUGAAGGGACCAAAGGCUGCUGCUCUGUUU 2460
     T  E  F  V  T  A  Y  V  A  R  L  K  G  P  K  A  A  A  L  F

2461 GCAAAACGCAUAACCUAGUCCCAUUGCAAGAAGUACCUAUGGACAGGUUUGUGAUGGAC 2520
     A  K  T  H  N  L  V  P  L  Q  E  V  P  M  D  R  F  V  M  D
```

Figure 2. nsP3/nsP4 of A1036 (1953, India, *Bdellonyssus bursa*)

```
2521  AUGAAGCGAGAUGUCAAGGUGACUCCGGGCACAAAACACACCGAAGAAAGGCCUAAGGUG  2580
      M  K  R  D  V  K  V  T  P  G  T  K  H  T  E  E  R  P  K  V

2581  CACGUAAUCCAAGCGGCUGAGCCUUUUGCUACAGCCUACCUUUGUGGCAUCCACCGAGAG  2640
      Q  V  I  Q  A  A  E  P  F  A  T  A  Y  L  C  G  I  H  R  E

2641  CUGGUACGCCGGCUUACCGCGGUUCUACUCCCGAACGUACACACCCUGUUUGACAUGUCU  2700
      L  V  R  R  L  T  A  V  L  L  P  N  V  H  T  L  F  D  M  S

2701  GCGGAGGAUUUCGACGCGAUUAUUGCCGAGCAUUUCCGACAAGGUGACGCCGUGCUCGAG  2760
      A  E  D  F  D  A  I  I  A  E  H  F  R  Q  G  D  A  V  L  E

2761  ACAGACAUCGCGUCAUUCGAUAAGAGUCAGGACGAUGCGAUGGCCCUGACUGGGCUGAUG  2820
      T  D  I  A  S  F  D  K  S  Q  D  D  A  M  A  L  T  G  L  M

2821  AUCCUGGAGGACCUCGGCGUCGAUCAACCGCUGCUGGACCUCAUCGAGUGUGCCUUCGGA  2880
      I  L  E  D  L  G  V  D  Q  P  L  L  D  L  I  E  C  A  F  G

2881  GAAAUAUCAUCUACGCAUCUGCCUACUGGGACACGGUUUAAGUUCGGCUCAAUGAUGAAA  2940
      E  I  S  S  T  H  L  P  T  G  T  R  F  K  F  G  S  M  M  K

2941  UCCGGAAUGUUUCUUACGCUCUUCGUGAACACCAUCUUGAAUGUCGUGAUCGCUAGUCGC  3000
      S  G  M  F  L  T  L  F  V  N  T  I  L  N  V  V  I  A  S  R

3001  GUGCUUGAGCACAGGUUAACAGGAUCACGAUGUGCCGCAUUCAUUGGAGACGAUAACAUC  3060
      V  L  E  H  R  L  T  G  S  R  C  A  A  F  I  G  D  D  N  I

3061  AUCCACGGCGUGGUAUCAGACAAGGAAAUGGCCGAAAGGUGCGCCACUUGGCUGAAUAUG  3120
      I  H  G  V  V  S  D  K  E  M  A  E  R  C  A  T  W  L  N  M

3121  GAGGUAAAAAUCAUUGACGCGGUGAUCGGCGAGCGUCCUCCGUAUUUCUGUGGUGGCUUU  3180
      E  V  K  I  I  D  A  V  I  G  E  R  P  P  Y  F  C  G  G  F

3181  AUACUACAGGACUCUGUCACCCAAACAGCCUGUCGAGUGGCUGACCCCCUAAAAAGACUG  3240
      I  L  Q  D  S  V  T  Q  T  A  C  R  V  A  D  P  L  K  R  L

3241  UUCAAGCUAGGAAAACCUUUGCCCGCAGAUGAUGACCAAGAUGAAGACAGAAGAAGGGCU  3300
      F  K  L  G  K  P  L  P  A  D  D  D  Q  D  E  D  R  R  R  A

3301  UUGCUGGAUGAGACUAAGGCGUGGUUUAGAGUGGGCAUAACCGAAACAUUGGCUACUGCG  3360
      L  L  D  E  T  K  A  W  F  R  V  G  I  T  E  T  L  A  T  A

3361  GUAGCAACGCGGUACGAAGUUGAUAACAUCACGCCUGUCCUGCUGGCACUGAGGACCCUU  3420
      V  A  T  R  Y  E  V  D  N  I  T  P  V  L  L  A  L  R  T  L

3421  GCGCAAAGCAAGAGAUCCUUUCAGUCCAUAAGAGGGGAAAUGAAGCAUCUCUACGGUGGU  3480
      A  Q  S  K  R  S  F  Q  S  I  R  G  E  M  K  H  L  Y  G  G

3481  CCUAAAUAG  3489
      P  K  *
```

Figure 2. Translated sequence of the nsP3-nsP4 region of Sindbis strain A1036, using the single letter amino acid code. nsP3 and nsP4 are translated as part of a polyprotein encoded by nucleotides (nts) 4100 to 7600 in the type virus genome. In this and the following 3 figures, nts are numbered from the amino terminus of nsP3. The star at nt 1636 indicates the opal codon separating nsP3 and nsP4; the star at nt 3489 is the termination codon of nsP4. The amino terminal residue of processed nsP4 is Tyr (Y) encoded by nts 1657-1659.

Figure 3. nsP3/nsP4 of MRM18520 (1975, Australia, mosquito pool)

```
   1  GCUCCGGCCUACCGCUCGAAACGUGAGAAUAUCGCCGAAUGCCUUGAAGAGGCCGUAGUU   60
      A  P  A  Y  R  S  K  R  E  N  I  A  E  C  L  E  E  A  V  V

  61  AACGCCGCGAACCCACUCGGACGUCCGGGCGAAGGGGUGUGUAAAGCCAUAUAUAAAAAA  120
      N  A  A  N  P  L  G  R  P  G  E  G  V  C  K  A  I  Y  K  K

 121  UGGCCCAAUAGUUUUGUCGAUUCUGCGACAGAGACUGGAACAGCUAAGCUAGUGUGCUGU  180
      W  P  N  S  F  V  D  S  A  T  E  T  G  T  A  K  L  V  C  C

 181  CAAGGAAAAAAGAUUAUCCAUGCCGUCGGACCUGACUUCCGUAAACACCCCGAGGCAGAA  240
      Q  G  K  K  I  I  H  A  V  G  P  D  F  R  K  H  P  E  A  E

 241  GCGCUGAAGAUUCUCCAGAACACAUACCACGCCAUCGCAGAUUUGGUUAACAAACAUGGA  300
      A  L  K  I  L  Q  N  T  Y  H  A  I  A  D  L  V  N  K  H  G

 301  AUCAAGACCGUAGCGAUCCCGCUUCUAUCCACCGGGAUUUACGCAGCGGGAAAAGACAGA  360
      I  K  T  V  A  I  P  L  L  S  T  G  I  Y  A  A  G  K  D  R

 361  CUUGAGGUCUCUUUAAACUGCCUCACUACCGCCCUGGACAGAACUGACGCAGACGUCACA  420
      L  E  V  S  L  N  C  L  T  T  A  L  D  R  T  D  A  D  V  T

 421  AUCUACUGCCUUGACAAAAAAUGGAAAGAACGGAUUGAUGCGUUUAUACAGUUGAAGGAG  480
      I  Y  C  L  D  K  K  W  K  E  R  I  D  A  F  I  Q  L  K  E

 481  UCGGUGACGGAACUGAAGGAUGAUGACAUGGAGAUCGACGACGAAUUAGUAUGGAUCCAC  540
      S  V  T  E  L  K  D  D  D  M  E  I  D  D  E  L  V  W  I  H

 541  CCGGAUAGUUGCCUCAAGGGUAGGAAAGGGUUUAGUACGACGAAGGGCAAGCUCUACUCG  600
      P  D  S  C  L  K  G  R  K  G  F  S  T  T  K  G  K  L  Y  S

 601  UACUUUGAGGGGACUAAAUUUCAUCAAGCAGCAAAAGACAUGGCUGAGAUCAAGGUACUU  660
      Y  F  E  G  T  K  F  H  Q  A  A  K  D  M  A  E  I  K  V  L

 661  UUUCCCGAUGAGCAAGAGUGCAACGAGCAACUGUGUGCAUACAUUCUAGGCGAAACCAUG  720
      F  P  D  E  Q  E  C  N  E  Q  L  C  A  Y  I  L  G  E  T  M

 721  GAAGCCAUCAGGGAAAAAUGUCCAGUGGACUUUAAUCCGUCGUCCAGUCCGCCGAAGACG  780
      E  A  I  R  E  K  C  P  V  D  F  N  P  S  S  S  P  P  K  T

 781  CUUCCCUGUUUGUGUAUGUACGCCAUGACGCCCGAGAGAGUGCACCGCUUGCGUAGCAAU  840
      L  P  C  L  C  M  Y  A  M  T  P  E  R  V  H  R  L  R  S  N

 841  AACGUCAAAUCCAUCACAGUAUGCUCGUCAACCCCGCUUCCGAAGCACAAAAUUAAGAAC  900
      N  V  K  S  I  T  V  C  S  S  T  P  L  P  K  H  K  I  K  N

 901  GUUCAGAAAGUACAGUGCACGAAAGUAGUCCUAUUCAACCCACAAACGCCUGAAUUUGUC  960
      V  Q  K  V  Q  C  T  K  V  V  L  F  N  P  Q  T  P  E  F  V

 961  CCUGCCCGCAAGUACAUAGAAACACAACCGAAGGACGACAGUCAAGAGGCGGAAGAAAAC 1020
      P  A  R  K  Y  I  E  T  Q  P  K  D  D  S  Q  E  A  E  E  N

1021  CCUGCCGCAGCCGAUAACACUUCACGGGAUGUAACAGACGUAUCUCUAGAUGUGGAAGGA 1080
      P  A  A  A  D  N  T  S  R  D  V  T  D  V  S  L  D  V  E  G

1081  GAUCGCGUUGCGGCCAACCGAUCAGAGGUGCACUCAGAGGACAACACCUCCCGAGAUGUA 1140
      D  R  V  A  A  N  R  S  E  V  H  S  E  D  N  T  S  R  D  V

1141  ACAGACAUAUCUCUAGACCACAACAGUGAUAGCGAGGUGGGCUCCAUUUUCUCUGACCUC 1200
      T  D  I  S  L  D  H  N  S  D  S  E  V  G  S  I  F  S  D  L

1201  AGCUGCUCCAGUCAUUCCAUCACCAGCAUGGACAGCUGGUCCUCCGGACCGAGCUCGAUC 1260
      S  C  S  S  H  S  I  T  S  M  D  S  W  S  S  G  P  S  S  I
```

Figure 3. nsP3/nsP4 of MRM18520 (1975, Australia, mosquito pool)

```
1261  AUGCUAAACGGGAAUCACACCAUCCAGGUCACGGCAGAGAUACACAACGCUCCUGCUGCA  1320
      M  L  N  G  N  H  T  I  Q  V  T  A  E  I  H  N  A  P  A  A

1321  CCGCCCGUACCACCACCACGCCUCAAGAAACUGGCGCGCUUGGCAGCUCAGAAGUCCGAU  1380
      P  P  V  P  P  P  R  L  K  K  L  A  R  L  A  A  Q  K  S  D

1381  CCGCCAUCCAGCCCGCCCUCAACGGUUGAGGACGUGUCGAUGCGCCUGUCAUUCCCUGCC  1440
      P  P  S  S  P  P  S  T  V  E  D  V  S  M  R  L  S  F  P  A

1441  ACGGUGUCAUUCGGAUCUUUUUCUGACGGCGAAGUCGACGAUCUUAGUCGCGAAAAAGCA  1500
      T  V  S  F  G  S  F  S  D  G  E  V  D  D  L  S  R  E  K  A

1501  GUGUCAGAACCAGUGGUCUUUGGUGCUUUCGAGCCAGGAGAGGUAACAUCUAUCAUUGAA  1560
      V  S  E  P  V  V  F  G  A  F  E  P  G  E  V  T  S  I  I  E

1561  GCAAGGUCUGUCGUGUCAUUCCCCGUGAAUAAACGCCGGCGCAGGAGACGGGGCCAAAAG  1620
      A  R  S  V  V  S  F  P  V  N  K  R  R  R  R  R  R  G  Q  K

1621  AAAACCGAAUAUUGACUAACCGGGGUAGGUGGGUAUAUCUUCUCGACUGACACGGGACCG  1680
      K  T  E  Y  *  L  T  G  V  G  G  Y  I  F  S  T  D  T  G  P

1681  GGUCACCUCCAGAAAAAAUCGGUUCUACAAAACCAGCUUACGGAACCGACCCUCGAGCGU  1740
      G  H  L  Q  K  K  S  V  L  Q  N  Q  L  T  E  P  T  L  E  R

1741  AAUCAAUUAGAACGAGUGUAUGCACCCAGUCUUGAUGCCAAGAAAGAGGAACUCUUGAAA  1800
      N  Q  L  E  R  V  Y  A  P  S  L  D  A  K  K  E  E  L  L  K

1801  CUCAAGUACCAAAUGAUGCCCACCGAAGCCAAUAAAAGUAGGUACCAGUCUAGAAAGGUA  1860
      L  K  Y  Q  M  M  P  T  E  A  N  K  S  R  Y  Q  S  R  K  V

1861  GAAAACCAAAAAGCCGUAACCACCGAGAGGUUACUGUCGGGAUUGAAGAUGUACAUUCAC  1920
      E  N  Q  K  A  V  T  T  E  R  L  L  S  G  L  K  M  Y  I  H

1921  UCAGAGAACCAACCCGAGUGUUACAAGGUCACCUAUCCGAAACCGUCGUACUCUAGCAGU  1980
      S  E  N  Q  P  E  C  Y  K  V  T  Y  P  K  P  S  Y  S  S  S

1981  GUUCCCCUUAGUUACCAGAGCCCCGAAUUCGCCGUAGCCGUCUGCAAUAACUACCUGCAU  2040
      V  P  L  S  Y  Q  S  P  E  F  A  V  A  V  C  N  N  Y  L  H

2041  GAGAAUUAUCCAACGGUUGCCUCCUAUCAGAUUACGGAUGAAUAUGACGCCUACCUUGAC  2100
      E  N  Y  P  T  V  A  S  Y  Q  I  T  D  E  Y  D  A  Y  L  D

2101  AUGGUGGACGGCACCGUAGCGUGUCUCGACACCGCUACAUUUUGCCCCGCGAAAUUACGC  2160
      M  V  D  G  T  V  A  C  L  D  T  A  T  F  C  P  A  K  L  R

2161  AGCUUCCCGAAGAAACACGAGUACCGAGAACCUAACAUCAGGAGCGCCGUACCGUCCGCU  2220
      S  F  P  K  K  H  E  Y  R  E  P  N  I  R  S  A  V  P  S  A

2221  AUGCAGAACACUCUACAGAACGUCCUGAACGCAGCAACAAAGAGGAAUUGCAAUGUUACU  2280
      M  Q  N  T  L  Q  N  V  L  N  A  A  T  K  R  N  C  N  V  T

2281  CAGAUGAGAGAACUACCGACUUUAGACUCCGCAACCUUUAAUGUGGAAUGCUUUCGAAAG  2340
      Q  M  R  E  L  P  T  L  D  S  A  T  F  N  V  E  C  F  R  K

2341  UACGCGUGCAACGACGAGUAUUGGGCUGAAUUCUCCGAAAAACCAAUUAGGAUCACCACA  2400
      Y  A  C  N  D  E  Y  W  A  E  F  S  E  K  P  I  R  I  T  T

2401  GAGUUUGUCACGGCGUACGUGGCGAGAUUGAAGGGACCAAAGGCUGCUGCACUGUUUGCU  2460
      E  F  V  T  A  Y  V  A  R  L  K  G  P  K  A  A  A  L  F  A

2461  AAAACGCAUAACCUAGUCCCACUGCAAGAAGUACCUAUGGACAGGUUUGUGAUGGACAUG  2520
      K  T  H  N  L  V  P  L  Q  E  V  P  M  D  R  F  V  M  D  M
```

Figure 3. nsP3/nsP4 of MRM18520 (1975, Australia, mosquito pool)

```
2521  AAGCGAGACGUUAAGGUGACUCCGGGCACGAAGCACACCGAAGAAAGACCCAAAGUGCAG  2580
      K  R  D  V  K  V  T  P  G  T  K  H  T  E  E  R  P  K  V  Q

2581  GUAAUCCAAGCGGCAGAGCCUCUAGCUACAGCCUAUUUAUGCGGCAUCCACCGUGAGCUG  2640
      V  I  Q  A  A  E  P  L  A  T  A  Y  L  C  G  I  H  R  E  L

2641  GUACGCAGGCUUACCGCAGUCCUGCUUCCGAACGUACACACCCUUUUUGAUAUGUCUGCG  2700
      V  R  R  L  T  A  V  L  L  P  N  V  H  T  L  F  D  M  S  A

2701  GAAGAUUUCGAUGCUAUCAUUGCCGAGCAUUUUCACCAGGGUGACGCUGUGCUCGAGACA  2760
      E  D  F  D  A  I  I  A  E  H  F  H  Q  G  D  A  V  L  E  T

2761  GACAUCGCGUCGUUCGAUAAGAGCCAAGACGAUGCGAUGGCCCUGACGGGGCUGAUGAUC  2820
      D  I  A  S  F  D  K  S  Q  D  D  A  M  A  L  T  G  L  M  I

2821  CUGGAGGACCUCGGAGUCGACCAGCCAUUGCUGGACCUCAUCGAGUGCGCCUUCGGGGAA  2880
      L  E  D  L  G  V  D  Q  P  L  L  D  L  I  E  C  A  F  G  E

2881  AUAUCAUCUACGCACCUGCCGACCGGGACACGGUUUAAGUUCGGCUCAAUGAUGAAAUCC  2940
      I  S  S  T  H  L  P  T  G  T  R  F  K  F  G  S  M  M  K  S

2941  GGAAUGUUCCUCACGCUCUUUGUGAACACCAUCUUGAAUGUCGUGAUAGCUAGUCGCGUG  3000
      G  M  F  L  T  L  F  V  N  T  I  L  N  V  V  I  A  S  R  V

3001  CUCGAGCACAGGUUAGCAGAAUCACGAUGCGCCGCAUUCAUCGGAGACGACAAUAUUAUU  3060
      L  E  H  R  L  A  E  S  R  C  A  A  F  I  G  D  D  N  I  I

3061  CACGGCGUGGUAUCCGACAAAGAAAUGGCUGAAAGGUGCGCCACUUGGCUGAAUAUGGAG  3120
      H  G  V  V  S  D  K  E  M  A  E  R  C  A  T  W  L  N  M  E

3121  GUAAAAAUUAUUGACGCAGUAAUUGGCGAACGUCCUCCGUACUUCUGUGGCGGCUUUAUA  3180
      V  K  I  I  D  A  V  I  G  E  R  P  P  Y  F  C  G  G  F  I

3181  CUGCAGGACUCAGUCACCCAAACAGCCUGCCGAGUGGCGGACCCCCUAAAAAGAUUGUUC  3240
      L  Q  D  S  V  T  Q  T  A  C  R  V  A  D  P  L  K  R  L  F

3241  AAAUUAGGAAAACCAUUACCUGCAGAUGAUGACCAAGAUGAAGACAGAAGAAGGGCUCUG  3300
      K  L  G  K  P  L  P  A  D  D  D  Q  D  E  D  R  R  R  A  L

3301  CUGGAUGAGACCAAGGCGUGGUUUAGAGUGGGCAUAACUGAGACACUGGCUACUGCGGUA  3360
      L  D  E  T  K  A  W  F  R  V  G  I  T  E  T  L  A  T  A  V

3361  GCAACGCGGUAUGAAGUUGAUAACAUCACACCGGUCCUGCUGGCACUGAGGACCCUUGCG  3420
      A  T  R  Y  E  V  D  N  I  T  P  V  L  L  A  L  R  T  L  A

3421  CAAAGCAAGAGAUCUUUUCAGGCCAUAAGGGGGAAAAUGAAGCAUCUCUACGGUGGUCCU  3480
      Q  S  K  R  S  F  Q  A  I  R  G  K  M  K  H  L  Y  G  G  P

3481  AAAUAG  3486
      K  *
```

Figure 3. Translated sequence of the nsP3-nsP4 region of the MRM18520 strain of Sindbis from Australia. Conventions are the same as in Figure 2.

Figure 4. nsP3/nsP4 of Girdwood (1963, South Africa, human)

```
   1  GCACCGUCAUACCGCACUAAAAGGGAGAACAUUGCUGAUUGUCAAGAGGAAGCAGUUGUC   60
      A  P  S  Y  R  T  K  R  E  N  I  A  D  C  Q  E  E  A  V  V

  61  AAUGCAGCCAAUCCGCUGGGCAGACCAGGCGAAGGAGUCUGCCGUGCCAUCUAUAAACGU  120
      N  A  A  N  P  L  G  R  P  G  E  G  V  C  R  A  I  Y  K  R

 121  UGGCCGAACAGUUUCACCGAUUCAGCCACAGAGACCGGCACCGCAAAACUGACUGUGUGC  180
      W  P  N  S  F  T  D  S  A  T  E  T  G  T  A  K  L  T  V  C

 181  CAAGGAAAGAAAGUGAUCCACGCGGUUGGCCCUGAUUUCCGGAAACACCCAGAGGCAGAA  240
      Q  G  K  K  V  I  H  A  V  G  P  D  F  R  K  H  P  E  A  E

 241  GCCCUGAAAUUGCUGCAAAACGCCUACCAUGCAGUGGCAGACUUAGUAAAUGAACAUAAU  300
      A  L  K  L  L  Q  N  A  Y  H  A  V  A  D  L  V  N  E  H  N

 301  AUCAAGUCUGUCGCCAUCCCACUGCUAUCUACAGGCAUUUACGCAGCCGGAAAAGACCGC  360
      I  K  S  V  A  I  P  L  L  S  T  G  I  Y  A  A  G  K  D  R

 361  CUUGAAGUAUCACUUAACUGCUUGACAACCGCGCUAGAUAGAACUGAUGCGGACGUAACC  420
      L  E  V  S  L  N  C  L  T  T  A  L  D  R  T  D  A  D  V  T

 421  AUCUACUGCCUGGAUAAGAAGUGGAAGGAAAGAAUCGACGCGGUGCUCCAACUUAAGGAG  480
      I  Y  C  L  D  K  K  W  K  E  R  I  D  A  V  L  Q  L  K  E

 481  UCUGUAACAGAGCUGAAGGAUGAGGAUAUGGAGAUCGACGACGAGUUAGUAUGGAUCCAU  540
      S  V  T  E  L  K  D  E  D  M  E  I  D  D  E  L  V  W  I  H

 541  CCGGACAGUUGCCUGAAGGGAAGAAAGGGAUUCAGUACUACAAAAGGAAAGUUGUAUUCG  600
      P  D  S  C  L  K  G  R  K  G  F  S  T  T  K  G  K  L  Y  S

 601  UACUUUGAAGGCACCAAAUUCCAUCAAGCAGCAAAAGAUAUGGCGGAGAUAAAGGUCCUG  660
      Y  F  E  G  T  K  F  H  Q  A  A  K  D  M  A  E  I  K  V  L

 661  UUCCCAAAUGACCAGGAAAGCAACGAGCAACUGUGUGCCUACAUAUUGGGGGAGACCAUG  720
      F  P  N  D  Q  E  S  N  E  Q  L  C  A  Y  I  L  G  E  T  M

 721  GAAGCAAUCCGCGAAAAAUGCCCGGUCGACCACAACCCGUCGUCUAGCCCGCCAAAAACG  780
      E  A  I  R  E  K  C  P  V  D  H  N  P  S  S  S  P  P  K  T

 781  CUGCCGUGCCUCUGCAUGUAUGCCAUGACGCCAGAAAGGGUCCACAGACUCAGAAGCAAC  840
      L  P  C  L  C  M  Y  A  M  T  P  E  R  V  H  R  L  R  S  N

 841  AACGUCAAAGAAGUUACAGUAUGCUCCUCCACCCCCCUUCCAAAGUACAAAAUCAAGAAC  900
      N  V  K  E  V  T  V  C  S  S  T  P  L  P  K  Y  K  I  K  N

 901  GUUCAGAAGGUUCAGUGCACAAAAGUAGUCCUGUUUAACCCGCAUACCCCUGCAUUCGUU  960
      V  Q  K  V  Q  C  T  K  V  V  L  F  N  P  H  T  P  A  F  V

 961  CCCGCCCGUAAGUACAUAGAAGCGCCAGAACAGCCUGCAGCUCCGCCUGCACAGGCCGAG 1020
      P  A  R  K  Y  I  E  A  P  E  Q  P  A  A  P  P  A  Q  A  E

1021  GAGGCCCCCGAAGUUGCAGCAACACCAACACCACCUGCAGCUGAUAACACCUCGCUUGAU 1080
      E  A  P  E  V  A  A  T  P  T  P  P  A  A  D  N  T  S  L  D

1081  GUCACGGACAUCUCACUGGACAUGGAAGACAGUAGCGAAGGCUCACUCUUUUCGAGCUUU 1140
      V  T  D  I  S  L  D  M  E  D  S  S  E  G  S  L  F  S  S  F

1141  AGCGGAUCGGACAACUCUAUUACUAGUAUGGACAGUUGGUCGUCAGGACCUAGUUCACUA 1200
      S  G  S  D  N  S  I  T  S  M  D  S  W  S  S  G  P  S  S  L

1201  GAGAUAGUAGACCGAAGGCAGGUGGUGGUGGCUGACGUCCAUGCCGUCCAAGAGCCUGCC 1260
```

Figure 4. nsP3/nsP4 of Girdwood (1963, South Africa, human)

```
      E  I  V  D  R  R  Q  V  V  V  A  D  V  H  A  V  Q  E  P  A
1261  CCUGUUCCACCGCCAAGGCUAAAGAAGAUGGCCCGCCUGGCAGCGGCAAGAAUGCAGGAA  1320
      P  V  P  P  P  R  L  K  K  M  A  R  L  A  A  A  R  M  Q  E
1321  GAGCCAACUCCACCGGCAAGCACCAGCUCUGCGGACGAGUCCCUUCACCUUUCUUUUGGU  1380
      E  P  T  P  P  A  S  T  S  S  A  D  E  S  L  H  L  S  F  G
1381  GGGGUAUCCAUGUCCUUCGGAUCCCUUUUCGACGGAGAGAUGGCCCGCUUGGCAGCGGCA  1440
      G  V  S  M  S  F  G  S  L  F  D  G  E  M  A  R  L  A  A  A
1441  CAACCCCCGGCAAGUACAUGCCCUACGGAUGUGCCUAUGUCUUUCGGAUCGUUUUCCGAC  1500
      Q  P  P  A  S  T  C  P  T  D  V  P  M  S  F  G  S  F  S  D
1501  GGAGAGAUUGAGGAGCUGAGCCGCAGAGUAACCGAGUCUGAGCCCGUCCUGUUUGGGUCA  1560
      G  E  I  E  E  L  S  R  R  V  T  E  S  E  P  V  L  F  G  S
1561  UUUGAACCGGGCGAAGUGAACUCAAUUAUAUCGUCCCGAUCAGCCGUAUCUUUUCCACCA  1620
      F  E  P  G  E  V  N  S  I  I  S  S  R  S  A  V  S  F  P  P
1621  CGCAAGCAGAGACGUAGACGCAGGAGCAGGAGGACCGAAUACUGACUAACCGGGGUAGGU  1680
      R  K  Q  R  R  R  R  R  S  R  R  T  E  Y  *  L  T  G  V  G
1681  GGGUACAUAUUUUCGACGGACACAGGCCCUGGGCACUUGCAAAAGAAGUCCGUUCUGCAG  1740
      G  Y  I  F  S  T  D  T  G  P  G  H  L  Q  K  K  S  V  L  Q
1741  AACCAGCUUACAGAACCGACCUUGGAGCGCAAUGUUCUGGAAAGAAUCUACGCCCCGGUG  1800
      N  Q  L  T  E  P  T  L  E  R  N  V  L  E  R  I  Y  A  P  V
1801  CUCGACACGUCGAAAGAGGAACAGCUCAAACUCAGGUACCAGAUGAUGCCCACCGAAGCC  1860
      L  D  T  S  K  E  E  Q  L  K  L  R  Y  Q  M  M  P  T  E  A
1861  AACAAAAGCAGGUACCAGUCUAGAAAAGUAGAAAAUCAGAAAGCCAUAACCACUGAGCGA  1920
      N  K  S  R  Y  Q  S  R  K  V  E  N  Q  K  A  I  T  T  E  R
1921  CUGCUUUCAGGGCUACGACUGUAUAACUCUGCCACAGAUCAGCCAGAAUGCUAUAAGAUC  1980
      L  L  S  G  L  R  L  Y  N  S  A  T  D  Q  P  E  C  Y  K  I
1981  ACCUACCCGAAACCAUCGUAUUCCAGCAGUGUACCGGCGAACUACUCUGACCCAAAGUUU  2040
      T  Y  P  K  P  S  Y  S  S  S  V  P  A  N  Y  S  D  P  K  F
2041  GCUGUAGCUGUUUGCAACAACUAUCUGCAUGAGAAUUACCCGACGGUAGCAUCUUAUCAG  2100
      A  V  A  V  C  N  N  Y  L  H  E  N  Y  P  T  V  A  S  Y  Q
2101  AUCACCGACGAGUACGAUGCUUACUUGGAUAUGGUAGACGGGACAGUCGCUUGUCUAGAU  2160
      I  T  D  E  Y  D  A  Y  L  D  M  V  D  G  T  V  A  C  L  D
2161  ACUGCAACUUUUUGCCCCGCCAAGCUUAGAAGUUACCCGAAAAGACACGAGUAUAGAGCC  2220
      T  A  T  F  C  P  A  K  L  R  S  Y  P  K  R  H  E  Y  R  A
2221  CCAAACAUCCGCAGUGCGGUUCCAUCAGCGAUGCAGAACACGUUGCAAAACGUGCUCAUU  2280
      P  N  I  R  S  A  V  P  S  A  M  Q  N  T  L  Q  N  V  L  I
2281  GCCGCGACUAAAAGAAACUGCAACGUCACACAAAUGCGUGAAUUGCCAACACUGGACUCA  2340
      A  A  T  K  R  N  C  N  V  T  Q  M  R  E  L  P  T  L  D  S
2341  GCGACAUUCAACGUUGAAUGCUUUCGAAAAUAUGCAUGUAAUGACGAGUAUUGGGAGGAG  2400
      A  T  F  N  V  E  C  F  R  K  Y  A  C  N  D  E  Y  W  E  E
2401  UUUGCCCGAAAGCCAAUUAGGAUCACUACUGAGUUCGUUACCGCAUACGUGGCCAGACUG  2460
      F  A  R  K  P  I  R  I  T  T  E  F  V  T  A  Y  V  A  R  L
2461  AAAGGCCCUAAGGCCGCCGCACUGUUCGCAAAGACGCAUAAUUUGGUCCCAUUGCAAGAA  2520
```

Figure 4. nsP3/nsP4 of Girdwood (1963, South Africa, human)

```
      K  G  P  K  A  A  A  L  F  A  K  T  H  N  L  V  P  L  Q  E
2521  GUGCCUAUGGAUAGGUUCGUCAUGGACAUGAAAAGAGACGUGAAAGUUACACCUGGCACG  2580
      V  P  M  D  R  F  V  M  D  M  K  R  D  V  K  V  T  P  G  T
2581  AAACACACAGAAGAAAGACCGAAAGUACAAGUGAUACAAGCCGCAGAACCCCUGGCGACC  2640
      K  H  T  E  E  R  P  K  V  Q  V  I  Q  A  A  E  P  L  A  T
2641  GCUUACCUGUGCGGGAUCCACCGGGAGUUAGUGCGCAGGCUUACAGCCGUCUUGCUACCC  2700
      A  Y  L  C  G  I  H  R  E  L  V  R  R  L  T  A  V  L  L  P
2701  AACAUUCACACGCUUUUUGACAUGUCGGCGGAGGACUUUGAUGCAAUCAUAGCAGAACAC  2760
      N  I  H  T  L  F  D  M  S  A  E  D  F  D  A  I  I  A  E  H
2761  UUCAAGCAAGGUGACCCGGUACUGGAGACGGAUAUCGCCUCGUUCGACAAAAGCCAAGAC  2820
      F  K  Q  G  D  P  V  L  E  T  D  I  A  S  F  D  K  S  Q  D
2821  GACGCUAUGGCGUUAACUGGCCUGAUGAUCUUGGAAGACCUGGGUGUGGACCAACCACUA  2880
      D  A  M  A  L  T  G  L  M  I  L  E  D  L  G  V  D  Q  P  L
2881  CUCGACUUGAUCGAGUGCGCCUUUGGAGAAAUAUCAUCCACCCAUCUGCCCACGGGUACC  2940
      L  D  L  I  E  C  A  F  G  E  I  S  S  T  H  L  P  T  G  T
2941  CGUUUCAAAUUCGGGGCGAUGAUGAAAUCCGGAAUGUUCCUCACGCUCUUUGUCAACACA  3000
      R  F  K  F  G  A  M  M  K  S  G  M  F  L  T  L  F  V  N  T
3001  GUUCUGAAUGUCGUUAUCGCCAGCAGAGUAUUGGAGGAGCGGCUUAAAACGUCCAAAUGU  3060
      V  L  N  V  V  I  A  S  R  V  L  E  E  R  L  K  T  S  K  C
3061  GCAGCAUUUAUCGGCGACGACAACAUCAUACACGGAGUAGUAUCUGACAAAGAAAUGGCU  3120
      A  A  F  I  G  D  D  N  I  I  H  G  V  V  S  D  K  E  M  A
3121  GAGAGGUGUGCCACCUGGCUCAACAUGGAGGUUAAGAUCAUUGACGCAGUCAUCGGCGAG  3180
      E  R  C  A  T  W  L  N  M  E  V  K  I  I  D  A  V  I  G  E
3181  AGACCGCCUUACUUCUGCGGUGGAUUCAUCUUGCAAGAUUCGGUUACCUCCACAGCGUGU  3240
      R  P  P  Y  F  C  G  G  F  I  L  Q  D  S  V  T  S  T  A  C
3241  CGCGUGGCGGACCCCUUGAAAAGGCUGUUUAAGUUGGGUAAACCGCUCCCAGCCGACGAC  3300
      R  V  A  D  P  L  K  R  L  F  K  L  G  K  P  L  P  A  D  D
3301  GAGCAAGACGAAGACAGAAGACGCGCUCUGCUAGAUGAAACAAAGGCGUGGUUUAGAGUA  3360
      E  Q  D  E  D  R  R  R  A  L  L  D  E  T  K  A  W  F  R  V
3361  GGUAUAACAGACACCUUAGCAGUGGCCGUGGCAACUCGGUAUGAGGUAGACAACAUCACA  3420
      G  I  T  D  T  L  A  V  A  V  A  T  R  Y  E  V  D  N  I  T
3421  CCUGUCCUGCUGGCAUUGAGAACUUUUGCCCAGAGCAAAAGAGCAUUUCAAGCCAUCAGA  3480
      P  V  L  L  A  L  R  T  F  A  Q  S  K  R  A  F  Q  A  I  R
3481  GGGGAAAUAAAGCAUCUCUACGGUGGUCCUAAAUAG  3516
      G  E  I  K  H  L  Y  G  G  P  K  *
```

Figure 4. Translated sequence of the nsP3-nsP4 region of the Girdwood strain of Sindbis isolated in South Africa. Conventions are the same as in Figure 2.

Figure 5. nsP3/nsP4 of Whataroa M78 (1962, New Zealand, mosquito pool)

```
   1  GCGCCAUCGUACAAAUCAAGGAGAGGAAACAUCAUCGAAUGCACCGAAGAAGCCGUCGUG   60
      A  P  S  Y  K  S  R  R  G  N  I  I  E  C  T  E  E  A  V  V

  61  AACGCUGCCAACGCACUAGGACGCCCCGGAGAAGGGGUCUGCAAGGCGAUUUACAAGAAG  120
      N  A  A  N  A  L  G  R  P  G  E  G  V  C  K  A  I  Y  K  K

 121  UGGCCGAACAGCUUCACCGGUUCCGCAACAGAAGUAGGGACUGCAAAAAUGACCACAAGC  180
      W  P  N  S  F  T  G  S  A  T  E  V  G  T  A  K  M  T  T  S

 181  CUAGGCAAGAAAGUCAUACAUGCCGUCGGACCGGAUUUUAAGAAGCACUCUGAAGAAGAA  240
      L  G  K  K  V  I  H  A  V  G  P  D  F  K  K  H  S  E  E  E

 241  GCCCUUAAACUGCUGCAGAAUGCCUACCACGCCAUCGCAGAUAUUAUUAAUGAGAACAAC  300
      A  L  K  L  L  Q  N  A  Y  H  A  I  A  D  I  I  N  E  N  N

 301  AUCAAAUCAGUGGCCAUUCCAUUGCUAUCAACUGGUAUAUACGCUGCAGGGAAGGACAGA  360
      I  K  S  V  A  I  P  L  L  S  T  G  I  Y  A  A  G  K  D  R

 361  CUAGAGACUUCUUUGCACUGUUUGACCACAGCGAUGGACAGGACGGACGCCGACGUAACG  420
      L  E  T  S  L  H  C  L  T  T  A  M  D  R  T  D  A  D  V  T

 421  GUAUACUGCCUUGACAAGAAAUGGCAGCAGCGAAUUGACGCAGUCCUUAGAUUGAAAGAA  480
      V  Y  C  L  D  K  K  W  Q  Q  R  I  D  A  V  L  R  L  K  E

 481  GAGGUAACGGAGCUAAAAGACGACGACAUGGAAAUUGAUGAGGAGCUGGUUUGGAUCCAC  540
      E  V  T  E  L  K  D  D  D  M  E  I  D  E  E  L  V  W  I  H

 541  CCUGACAGCUGUUUGAAAGGACGUAAAGGCUUUAGCACCACCAAAGGCAAACUGUAUUCA  600
      P  D  S  C  L  K  G  R  K  G  F  S  T  T  K  G  K  L  Y  S

 601  UACUUCGAAGGAACUAAAAUUUCACCAGGCAGCGAAAGACAUGGCAGAAAUCAAUGUAUUG  660
      Y  F  E  G  T  K  F  H  Q  A  A  K  D  M  A  E  I  N  V  L

 661  UUUCCAGACACCAUUGAGGCUAACGAGCAAAUCUGUAUGUAUAUCCUUGGAGAAAGCAUG  720
      F  P  D  T  I  E  A  N  E  Q  I  C  M  Y  I  L  G  E  S  M

 721  GAAGCUAUCCGCGAAAAAUGCCCCGUCGACUACAACCCUUCGUCAAGUCCGCCCAAAACC  780
      E  A  I  R  E  K  C  P  V  D  Y  N  P  S  S  S  P  P  K  T

 781  UUACCCUGCCUGUGCAUGUAUGCUAUGACACCUGAGAGGGUGCAUAGACUCAGAAGCAAC  840
      L  P  C  L  C  M  Y  A  M  T  P  E  R  V  H  R  L  R  S  N

 841  AAUGUCAAAGAAAUUACGGUAUGCUCCUCGACUCCACUUCCAAAACAUAAAAUCAAGAAC  900
      N  V  K  E  I  T  V  C  S  S  T  P  L  P  K  H  K  I  K  N

 901  GUACAACGAAUCCAGUGUUCAAAAAUCGUCUUGUUUAAUCCCCAGACUCCAGCUUUUGUA  960
      V  Q  R  I  Q  C  S  K  I  V  L  F  N  P  Q  T  P  A  F  V

 961  CCUGCACGUAAGUUCAUAGAAACCGAACCCAAAGAAACAGAAGACGAUGCGGCUCAGCCG 1020
      P  A  R  K  F  I  E  T  E  P  K  E  T  E  D  D  A  A  Q  P

1021  GACCCGACACCGGUAGUGCAGGCGAGUGUUUCGACCCCGGUCCCACAACGUCAGCAAGAC 1080
      D  P  T  P  V  V  Q  A  S  V  S  T  P  V  P  Q  R  Q  Q  D

1081  CCGUUAGAGUUGAUAAUAUCCGCAGACUCUUUAACCGAAGUAAACGACACCUCUGACGAC 1140
      P  L  E  L  I  I  S  A  D  S  L  T  E  V  N  D  T  S  D  D

1141  AUUUCCGACAUACCCUUUGACACAUCUGUAUAUGCUAGUACUUCCUCACUGAGCUCGGUU 1200
      I  S  D  I  P  F  D  T  S  V  Y  A  S  T  S  S  L  S  S  V

1201  UUGGACUGCCACAAUGUAGUCGAGGUCGAGGCGGAAAUUCACGUCGUCCCGCAGACUCCG 1260
      L  D  C  H  N  V  V  E  V  E  A  E  I  H  V  V  P  Q  T  P
```

Figure 5. nsP3/nsP4 of Whataroa M78 (1962, New Zealand, mosquito pool)

```
1261  GUGGCACCGCCGAGAAAGAAGAAGUUAGCACGUUUAGCGGCGCUAUCAAGAGCAUCUAGC  1320
      V  A  P  P  R  K  K  K  L  A  R  L  A  A  L  S  R  A  S  S

1321  AUUUCCUCCAUCGAAUCCAACCCACCAAUCACUUUUGGAUCAUUUGAGGAUGGAGAAAUA  1380
      I  S  S  I  E  S  N  P  P  I  T  F  G  S  F  E  D  G  E  I

1381  GACAACUUGCAGAAGAAGUGCACUUCAGAACCAUUUAUGUUCGGCUCGUUCGAACCAGGC  1440
      D  N  L  Q  K  K  C  T  S  E  P  F  M  F  G  S  F  E  P  G

1441  GAAGUCAACAGCCUGAUAGAAACCAGGUCGGAGCCACCACGUAGGGGGCGCAGACGUCGC  1500
      E  V  N  S  L  I  E  T  R  S  E  P  P  R  R  G  R  R  R  R

1501  AACAAGAACCGACAGGAGUAUUGACUAACCGGGGUAGGUGGGUACAUCUUCUCGACGGAC  1560
      N  K  N  R  Q  E  Y  *  L  T  G  V  G  G  Y  I  F  S  T  D

1561  ACUAAUGAAGGACACCUCCAGAAGAAAUCGGUACUCCAAAAUGAUCUGGCAGUCACCAUU  1620
      T  N  E  G  H  L  Q  K  K  S  V  L  Q  N  D  L  A  V  T  I

1621  UUAGAACGGAACAUAUUGGAAAAAGUCCAUGCACCCGUGUACAACGCUGAAAAAGAGGAG  1680
      L  E  R  N  I  L  E  K  V  H  A  P  V  Y  N  A  E  K  E  E

1681  AUACUGAAAAUGAAGUACCAGAUGAUGCCCACCGAAACCAACAAGAGUCGGUACCAAUCG  1740
      I  L  K  M  K  Y  Q  M  M  P  T  E  T  N  K  S  R  Y  Q  S

1741  AGAAAAGUAGAAAAUCAAAAAGCAGUAACUACCCAACGUCUAUUAUCAGGACUGAAACUU  1800
      R  K  V  E  N  Q  K  A  V  T  T  Q  R  L  L  S  G  L  K  L

1801  UAUACAUAUGAGCCUAACCAACCGGAGUGCUACAAAACCACAUAUCCGAGACCAUUGUAU  1860
      Y  T  Y  E  P  N  Q  P  E  C  Y  K  T  T  Y  P  R  P  L  Y

1861  UCUAGUAGCAUACCAGUUAGUUACGAUAGCGCACAAGUGGCGGUCGCAGUGUGCAAUAAC  1920
      S  S  S  I  P  V  S  Y  D  S  A  Q  V  A  V  A  V  C  N  N

1921  UACCUGCAUGAAAACUAUCCGACUGUCGCAUCUUACCAGAUUACCGACGAGUACGACGCU  1980
      Y  L  H  E  N  Y  P  T  V  A  S  Y  Q  I  T  D  E  Y  D  A

1981  UACCUAGACAUGGUGGAUGGCGCUGUCGCUUGUCUGGACACAGCUACAUUUUGUCCAGCU  2040
      Y  L  D  M  V  D  G  A  V  A  C  L  D  T  A  T  F  C  P  A

2041  AAGCUCAGGAGCUUCCCGAAGAAGCAUGAAUAUAAGACUCCCGAAAUUCGCAGCGCUGUC  2100
      K  L  R  S  F  P  K  K  H  E  Y  K  T  P  E  I  R  S  A  V

2101  CCCUCCGCCAUGCAGAACACACUACAGAAUGUACUCAUUGCCGCGACGAAACGAAACUGC  2160
      P  S  A  M  Q  N  T  L  Q  N  V  L  I  A  A  T  K  R  N  C

2161  AACGUUACUCAGAUGCGAGAAUUACCAACAUUGGAUUCAGCGACUUUUAACGUGGAAUGC  2220
      N  V  T  Q  M  R  E  L  P  T  L  D  S  A  T  F  N  V  E  C

2221  UUCAAAAAAUUUGCGUGUAAUGACGAGUACUGGAGCGAAUUUCGUGACAAACCCAUCAGA  2280
      F  K  K  F  A  C  N  D  E  Y  W  S  E  F  R  D  K  P  I  R

2281  AUAACAACCGAAUUCGUUACCUCGUACGUAGCGCGACUAAAAGGACCAAAGGCAGCGGCG  2340
      I  T  T  E  F  V  T  S  Y  V  A  R  L  K  G  P  K  A  A  A

2341  UUGUUCGCAAAAACUCAUAACCUAGUUCCCUUGCAAGAAGUUCCUAUGGAUAGGUUUGUC  2400
      L  F  A  K  T  H  N  L  V  P  L  Q  E  V  P  M  D  R  F  V

2401  AUGGACAUGAAGAGGGACGUUAAAGUCACACCCGGAACAAAACACACAGAAGAGAGACCA  2460
      M  D  M  K  R  D  V  K  V  T  P  G  T  K  H  T  E  E  R  P

2461  AAAGUCCAAGUCAUCCAGGCCGCUGAGCCGCUAGCUACCGCAUACUUAUGCGGAAUCCAC  2520
      K  V  Q  V  I  Q  A  A  E  P  L  A  T  A  Y  L  C  G  I  H
```

Figure 5. nsP3/nsP4 of Whataroa M78 (1962, New Zealand, mosquito pool)

```
2521  CGAGAACUGGUUAGGAGGCUGACUGCUGUACUACUUCCGAACAUUCACACCCUGUUCGAU  2580
      R  E  L  V  R  R  L  T  A  V  L  L  P  N  I  H  T  L  F  D

2581  AUGUCGGCCGAAGAUUUUGACGCUAUCAUAGCUGAACAUUUCAACUAUGGGGACCCUGUC  2640
      M  S  A  E  D  F  D  A  I  I  A  E  H  F  N  Y  G  D  P  V

2641  UUAGAAACCGACAUCGCGUCGUUCGACAAAAGUCAGGACGACGCCAUGGCCCUGACCGGC  2700
      L  E  T  D  I  A  S  F  D  K  S  Q  D  D  A  M  A  L  T  G

2701  CUGAUGAUCCUUGAAGACUUGGGUGUCGACCAGCCCCUUUUAGACCUUAUUGAAUGUGCG  2760
      L  M  I  L  E  D  L  G  V  D  Q  P  L  L  D  L  I  E  C  A

2761  UUCGGCGAAAUCUCCUCGACGCAUCUCCCGACAGGUACGAGAUUCAAAUUUGGAUCGAUG  2820
      F  G  E  I  S  S  T  H  L  P  T  G  T  R  F  K  F  G  S  M

2821  AUGAAAUCUGGAAUGUUCCUCACCCUGUUUGUCAACACUGUGCUGAAUGUUGUAAUCGCC  2880
      M  K  S  G  M  F  L  T  L  F  V  N  T  V  L  N  V  V  I  A

2881  AGCAGGGUCCUAGAGCAUAGACUGAAAGAGUCGCGAUGCGCCGCAUUCAUCGGUGAUGAC  2940
      S  R  V  L  E  H  R  L  K  E  S  R  C  A  A  F  I  G  D  D

2941  AACAUAAUACACGGCGUAGUGUCUGACAAGGAAAUGGCAGAAAGAUGCGCUACCUGGCUU  3000
      N  I  I  H  G  V  V  S  D  K  E  M  A  E  R  C  A  T  W  L

3001  AACAUGGAAGUGAAGAUCAUCGACGCCGUCAUAGGCAUCAGACCUCCAUAUUUUUGUGGU  3060
      N  M  E  V  K  I  I  D  A  V  I  G  I  R  P  P  Y  F  C  G

3061  GGAUUCAUCCUUCAAGAUGAGACGACAUUAACCACAUGUCGCGUCGCCGAUCCGCUUAAG  3120
      G  F  I  L  Q  D  E  T  T  L  T  T  C  R  V  A  D  P  L  K

3121  AGGCUCUUUAAACUAGGUAAACCACUACCCGCGGAGGACACGCAAGAUGAAGACAGAAGA  3180
      R  L  F  K  L  G  K  P  L  P  A  E  D  T  Q  D  E  D  R  R

3181  CGUGCCCUUAUGGACGAAACCAAAGCAUGGUUCCGGGUAGGAAUUAGGAACACUCUCGCA  3240
      R  A  L  M  D  E  T  K  A  W  F  R  V  G  I  R  N  T  L  A

3241  GUUGCCGUAUCGACCAGGUACGAGGUAGAAGAUAUUACACCCGUUCUAUACGCGCUUAGA  3300
      V  A  V  S  T  R  Y  E  V  E  D  I  T  P  V  L  Y  A  L  R

3301  ACAUUCGCUCAAAGCAAAAAGGCCUUCCAGACUAUACGAGGAGAAAUAAGACAGCUCUAC  3360
      T  F  A  Q  S  K  K  A  F  Q  T  I  R  G  E  I  R  Q  L  Y

3361  GGCGGUCCUAAAUAG  3375
      G  G  P  K  *
```

Figure 5. Translated sequence of the nsP3-nsP4 regions of Whataroa virus, isolated in New Zealand. It is clear from this sequence that Whataroa virus is closely related to Sindbis virus.

Amino acid differences in N-terminal half of nsP3 (%)

	Girdwood	Ockelbo	A1036	MRM18520	Whataroa
AR339	1.8	1.8	10.2	9.5	15.4
Girdwood		0.3	8.9	8.6	15.4
Ockelbo			9.2	8.9	15.4
A1036				1.5	16.0
MRM18520					16.3

Amino acid differences in nsP4 (%)

	Girdwood	Ockelbo	A1036	MRM18520	Whataroa
AR339	1.3	1.6	7.7	7.7	11.5
Girdwood		0.3	7.4	7.4	12.1
Ockelbo			7.7	7.7	12.4
A1036				1.6	12.4
MRM18520					11.8

Figure 6. Percent amino acid differences between the different isolates of Sindbis virus in two regions of the nonstructural proteins.

development of techniques for construction of such a library. The details of the methods developed are presented in the Methods section and required a careful attention to detail in order to obtain a random library. With the automated DNA sequencer, 24 DNA samples can be analyzed at one time and each sample can be read automatically to more than 400 nucleotides. Thus about 10 kb of sequence is obtained from a single run. To obtain the complete sequence of a viral RNA requires over-sequencing of the genome because of compression artifacts and occasional misreading by the machines and a slightly nonrandom distribution of the sequences obtained. The procedure developed as the most efficient is to over-sequence about three-fold, that is to obtain about 30 kb of sequence for the 12 kb RNA, align this sequence using computer programs and using the homology between different alphaviruses, and then to fill in any gaps that might still exist by designing PCR primers that can be used to obtain double-stranded cDNA for the missing segments and sequence this DNA manually. Fig. 7 illustrates the distribution of cDNA clones obtained using the technology developed for Whateroa virus and shows the random nature of the clones obtained. Fig. 8 illustrates sequence output from the Applied Biosystems sequenator for one clone (the original output is in four colors, a different color for each of the four nucleotides, which aids in interpreting the data). This sequence is automatically recorded in a computer file. Fig. 9 shows the DNA sequence obtained from this clone of Whateroa virus and compares the sequence to that of Sindbis virus. The technology is highly suitable to obtaining large amounts of sequence from alphaviruses and makes it conceptually feasible to examine a large number of different alphaviruses or of strains of one alphavirus isolated from different locations of the world in order to examine the relationships of the viruses to one another. The sequence of the nsP3-nsP4 region of Whataroa virus obtained by this method was shown in Fig. 5.

Mapping of a Neutralizing Antibody-Binding Site in Glycoprotein E2

A λgt11 library containing randomly generated 100-300 base pair Sindbis cDNA inserts in the lacZ gene was tested for reactivity with 6 mAbs, using ^{125}I-protein G to detect the presence of mAb (all were IgGs) bound to immunoreactive phage clones on nitrocellulose filters. Four

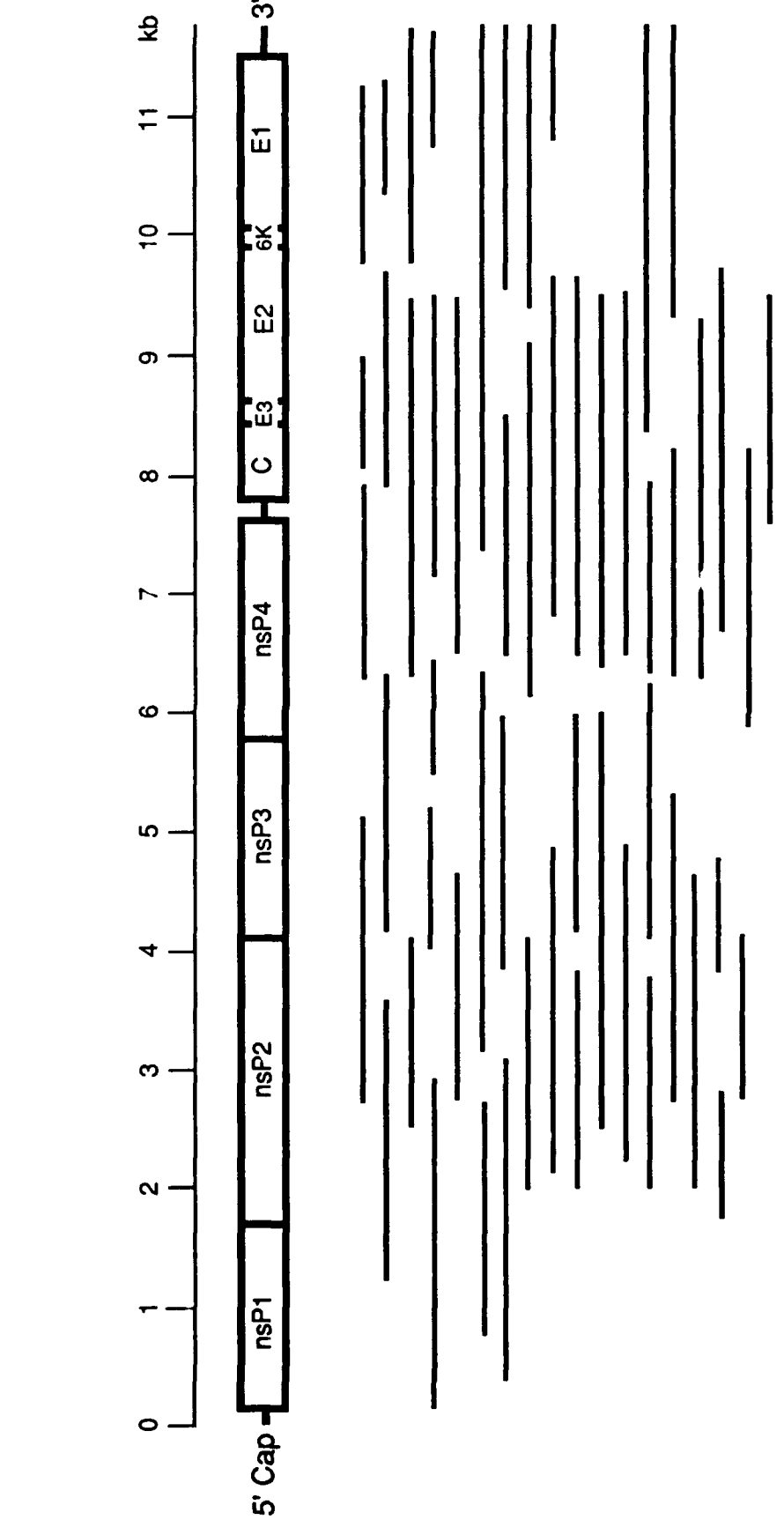

Figure 7. Map of the random cDNA clones of Whataroa virus which have been characterized. The top line is a scale in kilobases, and below that is shown a simple map of the alphavirus genome. Each line below represents a cloned insert which has been mapped by aligning its sequence with that of Sindbis virus.

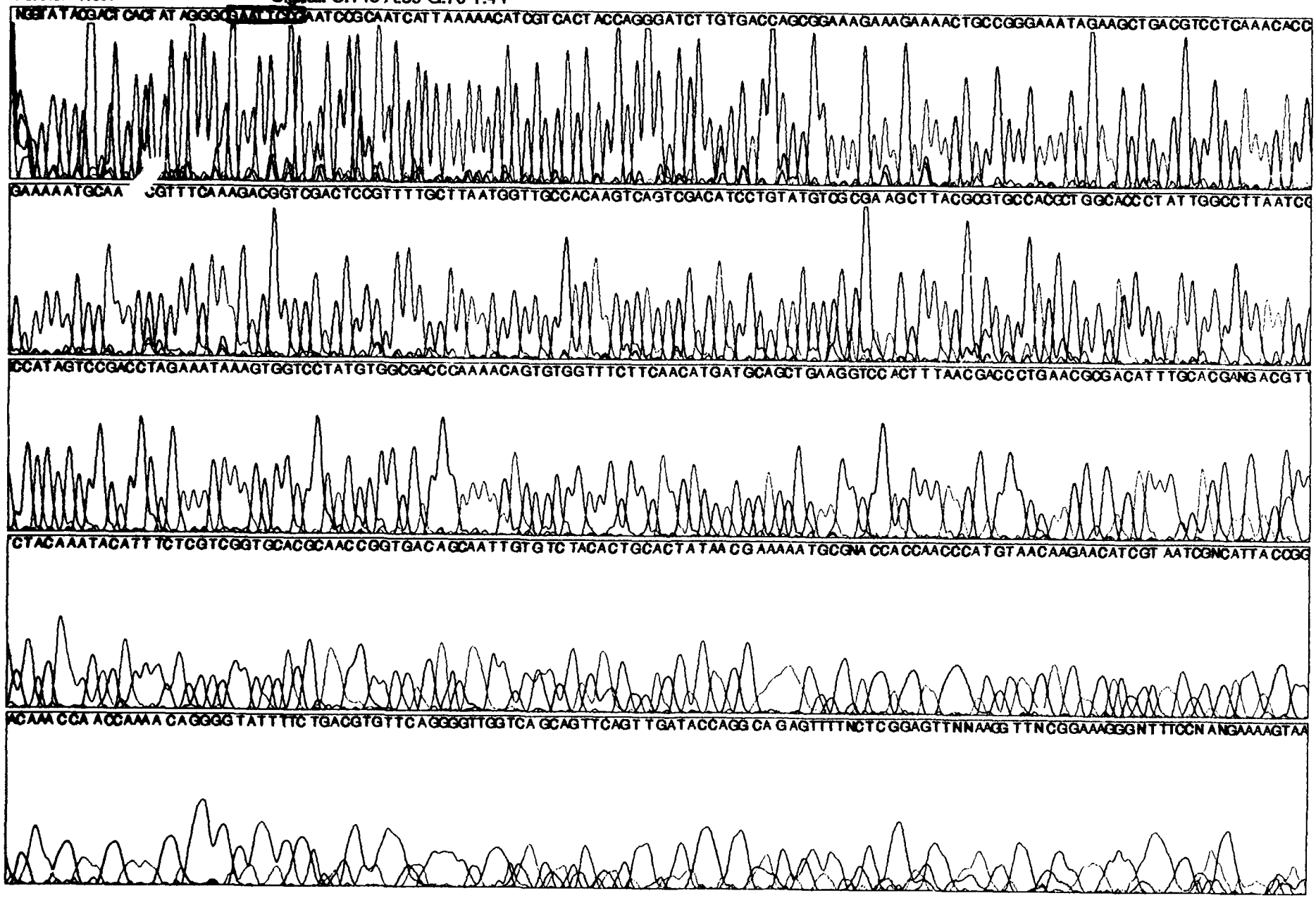

Figure 8. Automated sequence analysis of clone NZ430 of Whataroa virus.

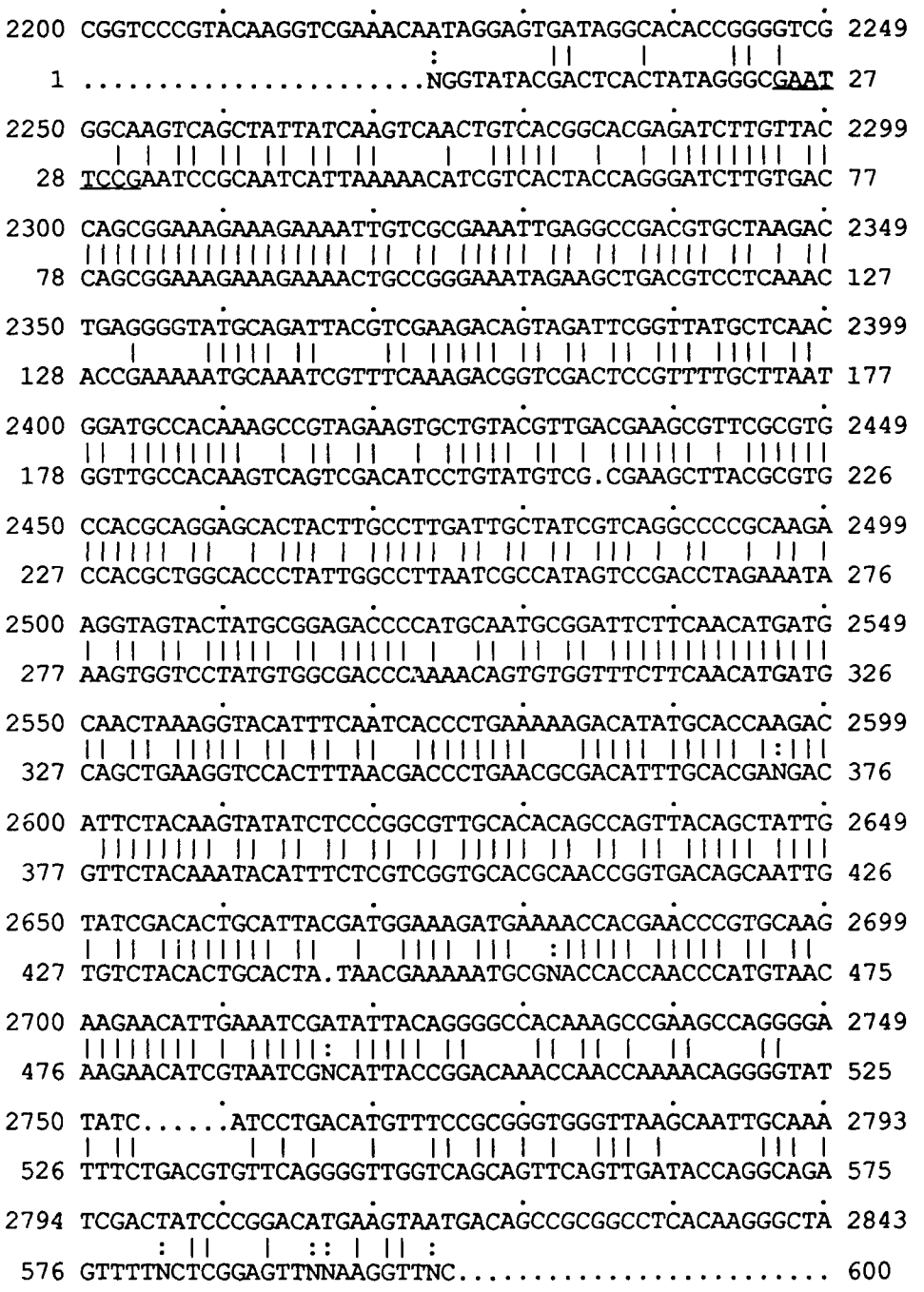

Figure 9. Comparison of the sequence of clone NZ430 of Whataroa virus (from data in Figure 8) with the sequence of Sindbis virus. Vertical lines between nucleotides highlight identity;, vertical dots show ambiguities. Underlined sequence is that of the *Eco*RI linker used to construct the clone; sequence upstream of this is vector.

positive phage clones were identified when mAb 23 was used to screen 10^6 plaques, designated λ23a, λ23b, λ23c and λ23d. Results with two of these clones are illustrated in Fig. 10; also shown is a control in which a nonreactive region of E2 is present as an insert in the λgt11 clone. These four phages were plaque purified and DNA prepared from each (Young and Davis, 1983). The inserts were removed with *Eco*RI, subcloned into vector M13mp18, and sequenced by the dideoxy chain termination procedure (Sanger et al., 1977). The four inserts contained overlapping sequences from the central region of glycoprotein E2 (Fig. 11). The insert in λ23a comprised E2 residues 155-258, that in λ23b comprised residues 173-251, that in λ23c 145-223, and that in λ23d 169-220. Thus the domain from residues 173 to 220 is present in all four inserts, and the neutralizing epitope recognized by mAb 23 must lie within this region. It is of note that this overlap region is 2-3 fold larger that the 15-22 amino acid residues found to contact antibody in epitopes defined by X-ray diffraction analysis (Laver et al., 1990), and it is conceivable that the epitope could be formed by a folded structure with contributions from residues throughout this region.

We also attempted to identify fusion proteins immunoreactive with four other E2-specific neutralizing mAbs, namely mAbs 18, 50, 51, and 49, as well as fusion proteins immunoreactive with mAb 33, specific for glycoprotein E1. In each case 10^6 plaques were screened. No positive plaques could be identified with any of these antibodies. We concluded that these antibodies probably react with conformational epitopes not present in the λgt11 library, either because these epitopes are discontinuous or consist of conformations not assumed by the fusion proteins.

Conclusions

Rapid Sequencing of Virus RNAs. High throughput automated DNA sequencing is ideally suited to obtaining large amounts of sequence data for strains of alphaviruses or for other viruses. The methods that we have developed can be used for any RNA virus and are suitable to

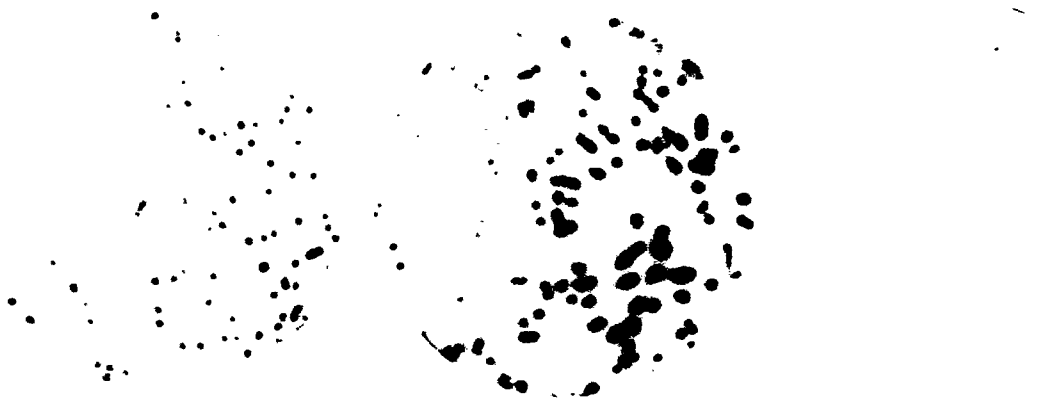

Figure 10. Reactivities of phage clones λ23a and λ23b with MAb 23. Immunoreactive phage plaques were picked and rescreened until a uniformly reactive population was obtained. Illustrated are the final populations for two reactive clones and a nonreactive clone expressing amino acids 129 - 192 of E2 (control). Phage stocks were plated on *E. coli* Y1090 in the presence of the inducer and the plaques transferred to nitrocellulose. Filters were incubated with MAb 23 followed by ^{125}I conjugated protein G and autoradiographed. Comparison of the autoradiogram with the pattern of plaques on the petri plate showed that all of the λ23a and λ23b plaques reacted with the antibody.

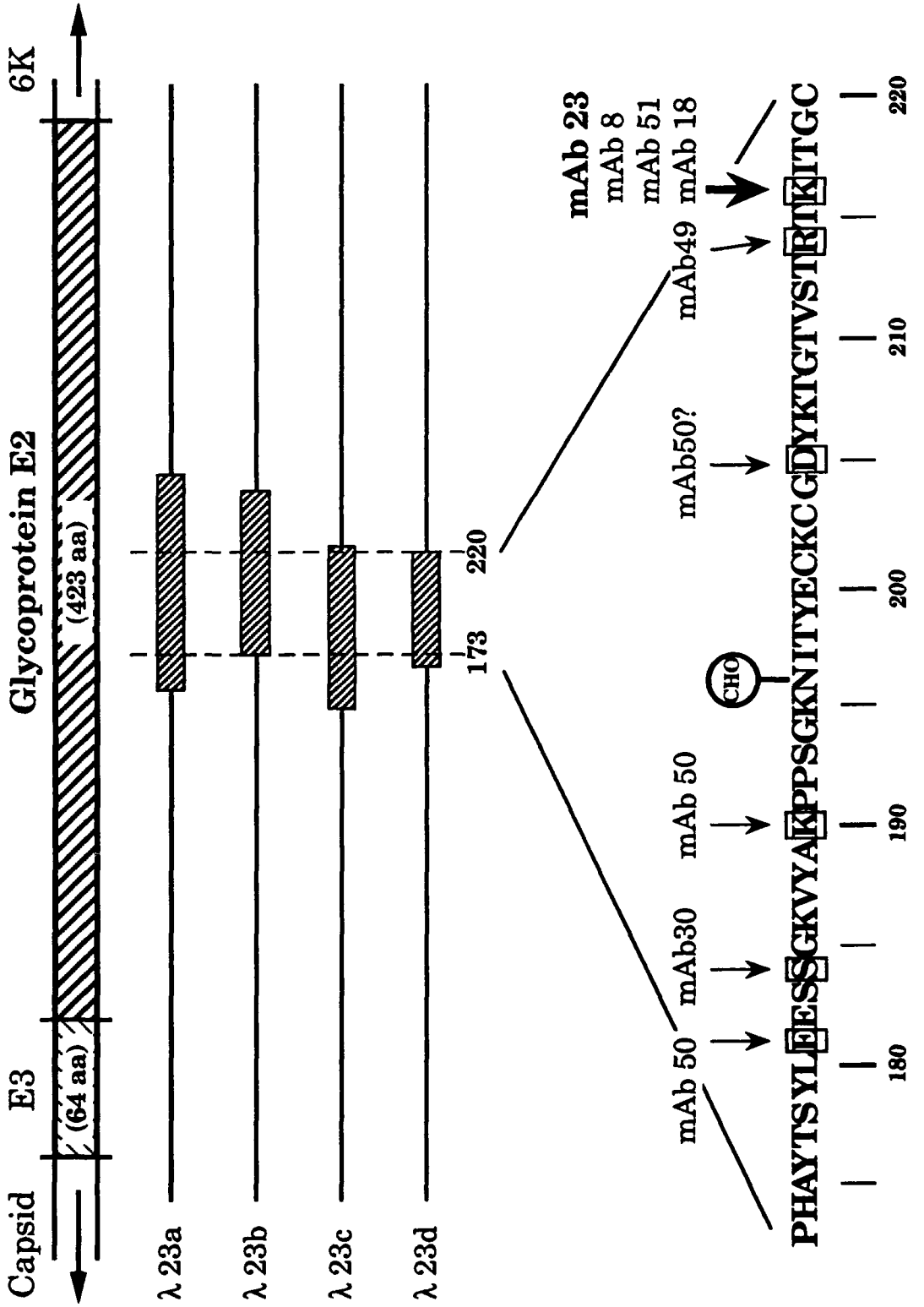

Figure 11. Schematic representation of an antigenically important domain of Sindbis virus glycoprotein E2. The relative locations of the inserts in four λgt11 clones reactive with MAb 23 are mapped. The overlap region in these four clones between residues 173 and 220 of E2 is expanded below, with a number of key features indicated. Residues altered in variants resistant to MAbs are boxed and a carbohydrate attachment site is indicated with a stalked symbol (CHO).

obtaining sequence information for different viruses belonging to a virus group or to obtaining sequence for strains of the same virus isolated from different geographic regions. This makes it feasible to examine a large number of isolates and therefore to determine the relationships among a group of viruses or to search for emergent viruses that differ in certain fundamental ways from other members of the group. The sequences presented in this report are an example of what it is possible to do. These sequences examine the relationships among a number of different geographic isolates of Sindbis-like viruses.

An Antibody Binding Domain in E2. The λgt11 system provides a rapid, specific, and sensitive strategy for the physical mapping on large viral genomes of the genes encoding proteins for which antibody reagents are available. We used small Sindbis virus genomic inserts in an attempt to define the immunoreactive domain of the protein more precisely. The limitation of the λgt11 system is the fact that these protein domains are expressed as part of a fusion protein and thus may not fold in the same way as the native protein, and only antibodies that interact with contiguous linear domains of the proteins of interest may be reactive with phage plaques.

From the sequence of the inserts in the four clones immunoreactive with mAb 23, it is clear that this antibody can react with a single continuous region of the Sindbis glycoprotein E2, and that the neutralization epitope must lie within the 48 residues between amino acids 173 and 220. This result is consistent with the results from mapping of antibody escape variants resistant to mAb 23 (Fig. 10). Sequencing of 3 independent variants resistant to mAb 23 and of 2 independent revertants selected to be sensitive again to mAb 23, as well as of other variants, has shown that residue 216 is important for reactivity with mAb 23 (Strauss et al., 1991). Virus with Lys-216 were fully sensitive to mAb 23, virus with Val-216 or Ile-216 demonstrated a reduced sensitivity to mAb 23, and virus with Glu-216 were resistant to mAb 23. From the results obtained here, it appears likely therefore that residue 216 interacts directly with mAb 23.

Although the remaining antibodies tested failed to react with the λgt11 library, presumably because they react with conformational epitopes not present in the library, it seems likely that the

E2-specific mAbs 50, 51, 49, and 18 also bind to epitopes at least partially encompassed within this same domain. Variants selected to be resistant to these mAbs were all found to have amino acid changes responsible for the escape from neutralization within the domain from residues 181 to 216 (Fig. 1) (Strauss et al., 1991). Furthermore, mAb 23 and these mAbs all react with closely spaced or overlapping epitopes as defined by competition assays or by the pattern of cross reactivity of different variants resistant to the various antibodies (Davis et al., 1987; Schmaljohn et al., 1983; Strauss et al., 1991). The results are all consistent with the hypothesis that the E2 domain between 173 and 220 forms a major antibody binding region important for neutralization of virus infectivity. This domain is illustrated in Fig. 11 with the locations of antibody escape variants shown and the region selected by mAb 23 indicated. This domain is hydrophilic, containing 25% charged residues, and has a glycosylation site at Asn-196, and thus is almost certainly exposed on the surface of the glycoprotein spike (Strauss and Strauss, 1986).

We have previously found that an antiidiotypic antibody to mAb 23, as well as antiidiotypic antibodies to mAbs 49 and 50, function as antireceptor antibodies in chicken cells (Wang et al., 1991). This suggests that the E2 domain defined by the fusion protein reactive with mAb 23 and by the antibody escape variants might form part of the antireceptor on the virus spike that binds to the cellular receptor. This hypothesis is supported by the observation that two strains of Sindbis virus that differ only in having Gly or Arg at residue 172 of E2 differ in their ability to bind to neuroblastoma cells in culture (Tucker and Griffin, 1991) and differ in their neurovirulence for mice (Lustig et al., 1988).

These results make clear that the region of E2 between residues 170 and 220 contains a number of dominant epitopes, and that this region is a key region for the development of vaccines.

References

Bell, J. R., Strauss, E. G., and Strauss, J. H. (1979). Purification and amino acid compositions of the structural proteins of Sindbis virus. *Virology* **97,** 287-294.

Davis, N. L., Pence, D. F., Meyer, W. J., Schmaljohn, A. L., and Johnston, R. E. (1987). Alternative forms of a strain-specific neutralizing antigenic site on the Sindbis virus E2 glycoprotein. *Virology* **161,** 101-108.

Diamond, D. C., Jameson, B. A., Bonin, J., Kohara, M., Abe, S., Itoh, H., Komatsu, T., Arita, M., Kuge, S., Nomoto, A., Osterhaus, A. D. M. E., Crainic, R., and Wimmer, E. (1985). Antigenic variation and resistance to neutralization in poliovirus type 1. *Science* **229,** 1090-1093.

Griffin, D. E. (1986). Alphavirus pathogenesis and immunity. *In* "The Togaviridae and Flaviviridae" (S. Schlesinger and M. J. Schlesinger, Ed.), pp. 209-250. Plenum Publishing Corp., New York.

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

Hahn, C. S., Strauss, E. G., and Strauss, J. H. (1989). Dideoxy sequencing of RNA using reverse transcriptase. *Methods in Enzymology* **180,** 121-130.

Hsu, M. T., Kung, H. J., and Davidson, N. (1973). An electron microscope study of Sindbis virus RNA. *Cold Spring Harbor Symp. Quant. Biol.* **38,** 943-950.

Laver, W. G., Air, G. M., Webster, R. G., and Smith-Gill, S. J. (1990). Epitopes on protein antigens: Misconceptions and realities. *Cell* **61,** 553-556.

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

Parry, N., Fox, G., Rowlands, D., Brown, F., Fry, E., Acharya, R., Logan, D., and Stuart, D. (1990). Structural and serological evidence for a novel mechanism of antigenic variation in foot-and-mouth disease virus. *Nature* **347,** 569-572.

Peters, C. J., and Dalrymple, J. M. (1990). Alphaviruses. *In* "Virology" (B. N. Fields and D. M. Knipe, Ed.), pp. 713-761. Raven Press, New York.

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

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

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual, Second, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74,** 5463-5467.

Schmaljohn, A. L., Kokubun, K. M., and Cole, G. A. (1983). Protective monoclonal antibodies define maturational and pH-dependent antigenic changes in Sindbis virus E1 glycoprotein. *Virology* **130,** 144-154.

Shirako, Y., Niklasson, B., Dalrymple, J. M., Strauss, E. G., and Strauss, J. H. (1991). Structure of the Ockelbo virus genome and its relationship to other Sindbis viruses. *Virology* **182,** 753-764.

Shirako, Y., and Strauss, J. H. (1992). A rapid and simple CTAB method for isolating DNA fragments from low melting point agarose gels. *submitted for publication*

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., Stec, D. S., Schmaljohn, A. L., and Strauss, J. H. (1991). Identification of antigenically important domains in the glycoproteins of Sindbis virus by analysis of antibody escape variants. *J. Virol.* **in press,**

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

Tucker, P. C., and Griffin, D. E. (1991). Mechanism of altered Sindbis virus neurovirulence associated with a single-amino-acid change in the E2 glycoprotein. *J. Virol.* **65,** 1551-1557.

Wang, K.-S., Schmaljohn, A. L., Kuhn, R. J., and Strauss, J. H. (1991). Antiidiotypic antibodies as probes for the Sindbis virus receptor. *Virology* **181,** 694-702.

Young, R. A., and Davis, R. W. (1983). Efficient isolation of genes by using antibody probes. *Proc. Natl. Acad. Sci. USA,* **80,** 1194-1198.