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 East California Blvd.
Pasadena, California 91125

REPORT DATE: December 3, 1992

TYPE OF REPORT: Final 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.
# Molecular Studies of Alphavirus Immunogenicity

**James H. Strauss, Ph.D.**

The alphaviruses consist of a group of 26 closely related viruses. Many of these viruses can cause disease in man, characterized by encephalitis, polyarthritis, fever or rash, depending upon the virus. In the 2.5 years of research supported under this contract we have mapped antigenic epitopes in the structural glycoproteins of alphaviruses that lead to neutralization of virus infectivity upon reaction with antibody, and have determined the sequence relationships of a number of Sindbis-like alphaviruses to one another and to other alphaviruses. We found that a domain of glycoprotein E2 of alphaviruses, between residues 170 and 220, was an important region for binding of monoclonal antibodies that neutralize virus infectivity, making it critical importance for the immune response required for protection from infection by the virus. In the determination of the relationships of alphaviruses to one another, we have determined complete or partial sequences of 8 different alphavirus RNAs. These include Ockelbo virus, a virus causing epidemic polyarthritis in northern Europe, strains of Sindbis virus from Africa, India, Australia and New Zealand and...
Aura virus from South America. We found that the Sindbis-like viruses possess certain key features in common and are all closely related to one another. Aura virus in the first true representative of the Sindbis viruses found in the Americas and demonstrates that these viruses are global in their distribution. We have also developed improved methods for rapidly sequencing large viral RNA genomes.
Contract No. DAMD17-90-0050

Final Report

List of personnel receiving pay, effort percentage and graduate degrees resulting from the contract support.

James H. Strauss, Professor, Principal Investigator 10% Effort 3/30/90-9/30/92

Richard Kuhn, Senior Research 50% Effort 4/1/90-12/31/91

Kangsheng Wang, Graduate Student 50% 4/1/90-6/30/91
Kangsheng Wang, Research Fellow 50% Effort 7/1/91-5/31/92
completed Ph.D requirements 5/91

Hong Zhang Research Assist. 75% Effort 4/23/90-8/23/92
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.

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

12-3-92
PI - Signature DATE
# TABLE OF CONTENTS

Front Cover .................................................................................. 1  
Report Documentation Page ........................................................... 2  
Foreword ..................................................................................... 3  
Table of Contents ........................................................................... 4  
Introduction .................................................................................. 5  
Methods Used ............................................................................... 5  
Mapping of Neutralizing Antigenic Epitopes of Alphaviruses .......... 7  
Alphaviruses Examined for Sequence Relationships ...................... 10  
Sequence Analysis of Ockelbo Virus ............................................. 10  
Analysis of 3' Terminal Nontranslated Sequence ............................ 15  
Sequence Studies of Aura RNA .................................................... 18  
Sequence Analysis of Whataroa Virus ........................................... 26  
Sequence Analysis of Other Alphaviruses ...................................... 26  
Conclusions ................................................................................... 39  
References ................................................................................... 41
INTRODUCTION

The alphaviruses are a widespread group of human pathogens that are present virtually everywhere in the world (Griffin, 1986; Monath, 1988; Peters and Dalrymple, 1990). They are mosquito-borne viruses and thus are particularly prevalent in tropical areas where mosquitoes abound and problems of overwintering by the virus do not arise, but are also present in temperate areas of the world including the United States. They have the capacity to replicate in the mosquito vector as well as in human hosts or in various species of birds and mammals. Old World alphaviruses are, in general, capable of causing a painful and disabling disease in man characterized by fever, rash and arthralgia. In the cases of the Ockelbo strain of Sindbis virus and of Ross River virus, this arthralgia manifests as a polyarthritis that may in some cases last for months or years. Many of the New World alphaviruses can cause fatal encephalitis in man. Our program attempts to understand the molecular basis of alphavirus immunogenicity and to determine the relationships of alphaviruses to one another, and has developed in collaboration with Drs. Alan Schmaljohn and Joel Dalrymple of USAMRIID.

METHODS USED

Virus Strains. Viruses used in this study were from the collection of Dr. J. M. Dalrymple of USAMRIID. Viruses were grown in BHK cells, in secondary chicken embryo fibroblast cells, or in mosquito cells, purified, and RNA prepared as described (Ou et al., 1981; Shirako et al., 1991).
cDNA Clones. cDNA clones were made in one of two ways. The first method used standard procedures in which first strand cDNA was made using oligo(dT) as primer and second strand synthesis was by the method of Gubler and Hoffman (Gubler and Hoffman, 1983; Sambrook et al., 1989). These cloning methods, as well as the methods of DNA sequencing and RNA sequencing, have been described in numerous publications from our laboratory over the years (Hahn et al., 1985; Rice et al., 1985; Rice and Strauss, 1981; Shirako et al., 1991; Strauss et al., 1984).

In a second approach, we developed methods suitable for high throughput automated DNA sequencing, in order to speed up the acquisition of sequence data. Whataroa virus was chosen as a test virus. The methods were described in detail in our annual report of 4/23/92. Briefly, first strand cDNA synthesis used random priming and second strand cDNA was synthesized by the method of Gubler and Hoffman (Gubler and Hoffman, 1983). After blunt-ending the double-stranded cDNA, the internal EcoRI restriction sites were methylated and the DNA was electrophoresed in an agarose gel. EcoRI linkers were attached to the 2-4 kb fraction and the DNA cloned in the EcoRI site of a suitable vector. One hundred clones that resulted from this cloning were characterized by restriction analysis and many of them were sequenced using an Applied Biosystems automated DNA sequencer.

Construction and Screening of the Bacteriophage Lambda Library. Sindbis virus strain AR339 from A. Schmaljohn at USAMRIID was grown in monolayers of primary chicken cells (Pierce et al., 1974). Virus was purified as described (Bell et al., 1978), 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 used as a template for cDNA
synthesis. A λgt11 expression 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. After flush-ending the product with the Klenow fragment of DNA polymerase I, methylation with EcoRI methyltransferase, and addition of EcoRI linkers, the modified cDNA was digested with an excess of EcoRI restriction enzyme. The digested DNA was fractionated on a Sephadex CL-6B column, and 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), phage plaques were grown for 6 h at 42°C. Nitrocellulose disks soaked in 10 mM isopropyl thio-β-D-galactopyranoside were then placed on top of the agar layer, and the plates were transferred to 37°C for 15 h. The filters 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 a monoclonal antibody in PBS solution containing 5% nonfat milk, washed, and the filters were then incubated at least two hours at room temperature in the presence of 125I-conjugated protein G (0.5 μCi/ml in 5% nonfat milk). 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.

**MAPPING OF NEUTRALIZING ANTIGENIC EPITOPES OF ALPHAVIRUSES**

We have localized a site in alphavirus glycoprotein E2 that binds neutralizing antibodies. Characterization of such immunogenic domains is important in developing vaccines, because neutralizing antibodies are thought to be particularly important in protecting a vaccinee from viral infection. We
developed a novel approach in which λgt11 expression libraries were constructed that expressed parts of the Sindbis genome, and these were screened with neutralizing monoclonal antibodies (MAbs). Many neutralizing antibodies react with discontinuous epitopes and thus will not react with a chimeric protein expressed in a λgt11 library. However, we succeeded in identifying one MAb which bound to specific clones within the λgt11 library (Wang and Strauss, 1991). Four λgt11 clones were found that reacted with MAb23, and a schematic of these four clones in relation to the Sindbis virus genome is shown in Fig. 1. The four clones all contain overlapping inserts from the E2 region of the genome, and the sequence of E2 from residues 173 to 220 is present in all. This demonstrates directly that this neutralizing MAb binds to glycoprotein E2 of Sindbis virus between residues 173 and 220.

The result with MAb23 confirmed and extended results in which variants of the virus selected to be resistant to neutralizing MAbs were sequenced in order to identify the regions within the glycoproteins of the virus with which the antibodies react (Strauss et al., 1991). This is illustrated in Fig. 1 in which the sequence of E2 between residues 173 and 220 is shown, and the location of many variants that render the virus resistant to neutralization by several MAbs is indicated. It is clear that the domain between residues 170 and 220 of glycoprotein E2 of alphaviruses is particularly important for the antibody response of a host. We have estimated that 90% of the neutralizing antibodies produced by an infected mouse are directed against this E2 domain (Strauss et al., 1991)

This domain of E2 identified as being important for reactivity with neutralizing antibodies also appears to be important for virus attachment to host cell receptors. First, many neutralizing antibodies are thought to inactivate the virus by binding to the domain that interacts with the cell receptor, thus blocking virus binding to the cell. Second, antiidiotypic antibodies made to MAbs that bind
Figure 1. 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).
to this domain of Sindbis E2 function as antireceptor antibodies (Wang et al., 1991). Third, changes in this region of E2 alter the ability of the virus to bind to neuronal cells (Ubol and Griffin, 1991). The simplest interpretation of these results is that the E2 domain between 170 and 220 binds to a cell receptor to initiate infection.

**ALPHAVIRUSES EXAMINED FOR SEQUENCE RELATIONSHIPS**

We have examined 12 strains of alphaviruses for their relationships to one another. These 12 viruses are shown in Fig. 2 together with the source from which they were isolated, their year of isolation, and the location in which they were isolated. Strains to be examined were chosen in consultation with Dr. Joel Dalrymple of USAMRIID, and were chosen on the basis of geography, year of isolation, potential for human disease, and, in the case of Aura virus, as a possible parent for the emergent virus Western equine encephalitis virus.

**SEQUENCE ANALYSIS OF OCKELBO VIRUS**

We have determined the complete nucleotide sequence of the genome of Ockelbo virus. This virus was chosen for analysis because it causes epidemics of polyarthritis in humans, a disabling disease that can last for months. The sequence of the virus isolated in 1982 in Edsbyn, Sweden, is shown in Fig. 3. The viral genome is 11,708 nucleotides in length excluding the poly(A) tail. The genome is identical in organization to that of the Sindbis virus AR339 strain (Strauss et al., 1984) isolated in Sindbis, Egypt in 1952 (Taylor et al., 1955). There are only 672 nucleotide differences between the two viruses (5.7% divergence) that result in 97 amino acid changes (2.6) divergence. Thus more that 85% of all
<table>
<thead>
<tr>
<th>Name</th>
<th>Strain</th>
<th>Source</th>
<th>Year</th>
<th>Location</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subgroup I</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sindbis</td>
<td>AR339</td>
<td>Mosquito (Culex univittatus)</td>
<td>1952</td>
<td>Egypt</td>
<td>Taylor et al., (1955)</td>
</tr>
<tr>
<td>Sindbis</td>
<td>MP684</td>
<td>Mosquito (Mansonia fuscopennata)</td>
<td>1958</td>
<td>Uganda</td>
<td></td>
</tr>
<tr>
<td>Sindbis</td>
<td>Girdwood</td>
<td>Human</td>
<td>1963</td>
<td>South Africa</td>
<td>Malherbe et al., (1963)</td>
</tr>
<tr>
<td>Sindbis</td>
<td>R33</td>
<td>Reed Warbler (Acrocephalus scirpaceus)</td>
<td>1971</td>
<td>Czechoslovakia</td>
<td></td>
</tr>
<tr>
<td>Sindbis</td>
<td>1038</td>
<td>Turtle Dove (Streptopelia turtur)</td>
<td>1964</td>
<td>Israel</td>
<td></td>
</tr>
<tr>
<td>Ockelbo</td>
<td>Edsbyn 83M107</td>
<td>Mosquito (Culiseta morsitans)</td>
<td>1983</td>
<td>Edsbyn village, Sweden</td>
<td></td>
</tr>
<tr>
<td><strong>Subgroup II</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sindbis</td>
<td>MM2215</td>
<td>Mosquito (Culex tritaeniorhynchus)</td>
<td>1955</td>
<td>Indonesia</td>
<td></td>
</tr>
<tr>
<td>Sindbis</td>
<td>A-1036</td>
<td>Mite (Bdellonyssus bursa)</td>
<td>1953</td>
<td>India</td>
<td>Shah et al., (1960)</td>
</tr>
<tr>
<td>Sindbis</td>
<td>MRM18520</td>
<td>Mosquito (unidentified)</td>
<td>1975</td>
<td>Queensland, Australia</td>
<td></td>
</tr>
<tr>
<td><strong>Subgroup III</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whataroa</td>
<td>M78</td>
<td>Mosquito pool</td>
<td>1962</td>
<td>New Zealand</td>
<td></td>
</tr>
<tr>
<td><strong>Subgroup IV</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aura</td>
<td>AR10315</td>
<td>Mosquito (Culex spp.)</td>
<td>1959</td>
<td>Brazil</td>
<td>Causey et al. 1963</td>
</tr>
</tbody>
</table>

**Figure 2.** Strains of Sindbis virus and related viruses used in this study.
Figure 3a. See legend on last page of this sequence.
Figure 3b. See legend on the last page of this sequence.
Figure 3c. Complete nucleotide sequence of the Ockelbo virus genome. The sequence is shown from 5’ to 3’ and translated using the single letter amino acid code. Nucleotides different from those in HRSP are underlined, and changed amino acids are boxed. Deletions relative to HR are indicated by solid triangles pointing upward and the number of residues deleted. Insertions have both amino acids and nucleotides boxed together, and an open triangle pointing downward. Termination codons are labelled Am (Amber, UAG) or Op (Opal, UGA) as appropriate. Nucleotides are numbered 5’ to 3’; amino acid numbering begins again at the beginning of each final protein product.
nucleotide differences are silent, illustrating the importance of conservation of amino acid sequence.

Glycoprotein E2 is particularly important for antigenicity, as described above, and changes in E2 have been associated with changes in virulence (Lustig et al., 1988; Olmsted et al., 1986; Strauss et al., 1991; Tucker and Griffin, 1991). The differences in glycoprotein E2 between six strains of Sindbis virus are listed in Fig. 4. The residues at positions 172, 209, 212, and 216 are known to be important determinants of the antigenicity of the virus (Strauss et al., 1991), and the changes in these positions are important for the differences in the cross-reactivity of the viruses with antibodies. The residues at 55 and 172 are known to be important determinants of the neurovirulence of the virus in a mouse model (Lustig et al., 1988), and it is possible that the amino acid difference at position 55 may be important for the increased virulence of Ockelbo virus compared to the other strains of Sindbis virus in Fig. 4.

**ANALYSIS OF 3' TERMINAL NONTRANSLATED SEQUENCE**

To study the relationships among a number of Sindbis viruses present in nature, the sequences of the 3' nontranslated regions (NTR) were obtained for a number of strains. These sequences are shown in Fig. 5. The sequence identity throughout this region is greater than 80% for all viruses shown, and the sequence organization is identical except for a few scattered insertions and deletions. In the 3' NTR there are three repeated elements that are highly conserved (boxed in the figure). As an example of the conservation of these elements, there are 49 differences in the 3' NTRs of the Australian and AR339 strains that occur outside the repeated elements (24.1% divergence) but only 7 changes within these elements (5.8% divergence), and the overall divergence is
<table>
<thead>
<tr>
<th>RESIDUE</th>
<th>AR339</th>
<th>S.A. AR86</th>
<th>OCKELBO</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HRSP</td>
<td>DG</td>
<td>AS</td>
</tr>
<tr>
<td>1</td>
<td>S</td>
<td>S</td>
<td>R</td>
</tr>
<tr>
<td>3</td>
<td>T</td>
<td>E</td>
<td>T</td>
</tr>
<tr>
<td>23</td>
<td>V</td>
<td>E</td>
<td>E</td>
</tr>
<tr>
<td>29</td>
<td>V</td>
<td>V</td>
<td>V</td>
</tr>
<tr>
<td>55</td>
<td>Q</td>
<td>Q</td>
<td>Q</td>
</tr>
<tr>
<td>61</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>69</td>
<td>L</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>70</td>
<td>K</td>
<td>K</td>
<td>E</td>
</tr>
<tr>
<td>116</td>
<td>V</td>
<td>V</td>
<td>V</td>
</tr>
<tr>
<td>126</td>
<td>L</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>172</td>
<td>R</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>209</td>
<td>G</td>
<td>R</td>
<td>G</td>
</tr>
<tr>
<td>212</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>216</td>
<td>E</td>
<td>E</td>
<td>K</td>
</tr>
<tr>
<td>243</td>
<td>L</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>247</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>277</td>
<td>I</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>312</td>
<td>V</td>
<td>V</td>
<td>V</td>
</tr>
<tr>
<td>375</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>386</td>
<td>V</td>
<td>V</td>
<td>V</td>
</tr>
</tbody>
</table>

Figure 4. Amino acid differences in the glycoprotein E2 of various Sindbis strains. The sequence of HRSP is from Strauss et al. (1984); The sequence marked DG is the SV1A strain published in Lustig et al. (1988). AS is our unpublished sequence of the strain used by A. Schmaljohn for the isolation of antigenic escape mutants (Stec. et al., 1986); RJ is the sequence from Davis et al. (1986) of a laboratory strain from Robert Johnston. The sequence of AR86 was reported in Russell et al. (1989), and the Ockelbo sequence was presented in Figure 3.
Figure 5. Sequence of the 3' termini of several Sindbis viruses. The sequences of Ockelbo 83M107, Karelian fever, and South African Sindbis (Girdwood) were determined directly from RNA by dideoxy sequencing using reverse transcriptase and a T12GA primer. The Ockelbo 82 sequence is from Fig. 3 and that of AR339 (HRSP) is from Strauss. Three repeated sequence elements of 40 nucleotides are boxed. The translated sequence for AR339 (HRSP) and any amino acid that differs in the other viruses is boxed. This figure is from Shirako et al. (1991).
18.1%. From such analysis, we propose that these repeated and conserved elements play an important role in viral RNA replication, and this role is probably more important in mosquito cells than in vertebrate cells (Kuhn et al., 1990).

The relationships among these viruses is illustrated in Fig. 6. Three points are obvious from this diagram. One is that the Sindbis strains analyzed can be divided into a European-African group and an Asian-Australian group. The second point is that Ockelbo virus and Karelian fever virus are virtually identical. The third point is that Ockelbo virus is more closely related to the South African strain of Sindbis virus isolated in 1963 (and which is also capable of causing human illness) than it is to the Egyptian strain isolated in 1952. We conclude from this last point that Ockelbo virus was probably introduced into Northern Sweden from South Africa in the 1960s, from where it spread into Finland (where it causes the disease called Pogosta) and the Karelian region of Russia.

**SEQUENCE STUDIES OF AURA RNA**

We have obtained the sequence of essentially all of the genome of Aura virus and are currently assembling this sequence. We were particularly interested in this virus because we have previously shown that Western equine encephalitis virus (WEE), previously thought to be closely related to Sindbis virus, is in fact a recombinant virus in which most of the genome was derived from Eastern equine encephalitis virus and only the surface glycoproteins were derived from a Sindbis-like virus (Hahn et al., 1988). Thus the question arose as to whether there is a virus found in the Americas that is closely related to Sindbis and that could have served as the second parent of WEE. The question is of particular interest because WEE emerged from a recombination event.
Figure 6. Evolutionary tree of strains of Sindbis virus. The vertical distances indicate the number of nucleotide differences between any two strains in the 3' terminal 420 nucleotides. The horizontal distances are arbitrary. The tree was constructed from a difference matrix by iteration to give the best possible representation with the minimum number of branch points. Nucleotide differences between any two strains can be calculated by summing the numbers on the vertical branches connecting the two strains to be compared.
The sequence of about 5000 nucleotides of Aura RNA in the nonstructural protein coding region is shown in Fig. 7. This sequence begins in the 5' NTR and continues through nsP1, nsP2, and part of nsP3. From this sequence, it is clear that Aura virus is closely related to Sindbis virus. Comparison of the amino acid sequences of Sindbis virus and of Aura virus in the region represented by the Aura sequence in Fig. 7 shows that the two sequences are 80% identical, illustrating that Aura is in fact a Sindbis-like virus. We also found that the 3' NTR of Aura RNA is Sindbis-like. As described above, Sindbis-like viruses contain three copies of a conserved sequence element that we postulate is important for RNA replication. Although other alphaviruses often contain repeated sequence elements, these elements are completely different in sequence from the Sindbis sequence. Furthermore, WEE lacks the characteristic Sindbis 3' NTR, and contains instead a chimeric 3' NTR. Thus Aura virus represents the first known example of a true Sindbis-like virus in the Americas.

Aura virus is widely distributed in South America, having been isolated in Brazil and in Northern Argentina. Analysis of the data is not yet complete, but it is possible that Aura virus represents the ancestral Sindbis-like virus, and that it was transmitted to the Old World to serve as the founder of the Sindbis viruses in the Old World, as we previously postulated (Levinson et al., 1990). Aura virus may have served as one of the parents of WEE, contributing its glycoproteins to this recombinant virus (Hahn et al., 1988).
Figure 7a. See legend on last page of this sequence
Figure 7b. See legend on last page of this sequence
<table>
<thead>
<tr>
<th>Line</th>
<th>Sequence</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>2221/719</td>
<td>GAA ACA GTG GGG GTC ATT GGA ACA CCG GGG TCA GGT AAG TCG GCC ATC ATA AAA TCT ACG</td>
<td>glu thr leu gly val ile gly thr pro gly ser gly lys ser ala ile ile lys ser thr</td>
</tr>
<tr>
<td>2281/739</td>
<td>GTA ACA CTA AAA GAC CTC GTA ACT AGC GGT AAG AAA GAA AAT TGC AAA GAA ATA GAG AAT</td>
<td>val thr leu lys asp leu val thr ser gly lys glu asn cys lys glu ile gla asn</td>
</tr>
<tr>
<td>2341/759</td>
<td>GAG GTG CAG AAA ATG CGG GTA ATT GTA ACT ATA GCT AGC AGA ACG GTA GAC TCG GTA CTG TCT</td>
<td>asp val gln lys met arg gly met thr ile ala thr arg thr val asp ser val leu leu</td>
</tr>
<tr>
<td>2401/779</td>
<td>AAT GGA TGG AAG AAA GCA GTA GAC GTC CTA TAT GTG GAT GAA GCG TTT GTA GCA GTG CAT GCA</td>
<td>glu thr ileu met ala leu ile ala ile val lys pro arg arg lys gly val leu cys gly</td>
</tr>
<tr>
<td>2461/799</td>
<td>GCC ACC TTA ATG GCA TGG ATT GCC ATG AAA ATA GAG TAA GTA CTG AGT GCC GCG GGT</td>
<td>gly thr leu met ala leu ile ala ile val lys pro arg arg lys gly val leu cys gly</td>
</tr>
<tr>
<td>2521/819</td>
<td>GAC CCG AAG CAG TGG CCC TTC TTT AAT TTA ATG CAA CTG AAG GTA AAG TCG GCC ATG</td>
<td>asp pro lys gln trp pro phe asn leu met gln lys val asn phe asn asn pro</td>
</tr>
<tr>
<td>2581/839</td>
<td>GAG CGA GAC CTG TGT ACT CTC ACC CAT TAT AAA TAT ATC TCT CGC AGG TGC ACC GAA CCT</td>
<td>glu arg asp leu cys thr ser thr his tyr lys tyr ile ser arg arg cys thr gln pro</td>
</tr>
<tr>
<td>2641/859</td>
<td>GCT ACA GCC ATA GTG TCT ACA TTA CAC TAT GAC GGA AAG ATG AGT ACT ACG AAT CCC TGC</td>
<td>val thr ileu ileu val thr leu thr leu his tyr asp gly lys met arg thr thr asn pro cys</td>
</tr>
<tr>
<td>2701/879</td>
<td>AAA AGG GCT GAA ATA GAC GTA AAG TCA TCT TGAG ACT AAG CCC AAG AAA GGA GAC ATA GTG</td>
<td>lys arg ala ile glu ile asp gly asp gly lys thr lys gly asp ile val</td>
</tr>
<tr>
<td>2761/899</td>
<td>TGG ACG GTG TTC CGT GGT TGG GTT AAG CAG GGA AAA ATC GAT TAC CCC GGA CCC GGA GGT</td>
<td>leu thr cys phe arg gly trp val lys gln gly gln ile asp tyr pro gly pro gly gly</td>
</tr>
<tr>
<td>2821/919</td>
<td>CAT GAC CGT GCA GCT TCT CAA GGG CTA ACC AGA AGG GGC GTT TAT GCG GTC AGA CAG AAA</td>
<td>his asp arg ala ala ser gln gly leu thr arg arg gly val tyr ala val leu cys lys</td>
</tr>
<tr>
<td>2881/939</td>
<td>GTA AAT GAA AAC CCA CTA TAT GCA GAG AAG TCA GAA CAC GTC CAA GTG TAA GTA ATC CTC</td>
<td>val asn gly val pro leu tyr ala glu lys ser gly lys asn val leu leu thr arg</td>
</tr>
<tr>
<td>2941/959</td>
<td>ACG GAA GAT CGC ATA GTG TGG AAG ACA CTG CAA GGG GAT CTT TGG ATT AAG TAC TCT ACT</td>
<td>thr glu asp arg ile val trp lys thr leu lys gly glp pro trp ile lys tyr leu thr</td>
</tr>
<tr>
<td>3001/979</td>
<td>AAG CGT CCA AAA GGG AAC TTT ACA GCC ACT TTA GAA TGG CAG GCC GAA CCC GAG CAC</td>
<td>asn val pro lys gly asp phe thr ala thr leu glu trp gln ala glu his asp</td>
</tr>
<tr>
<td>3061/999</td>
<td>ATT ATG AAG GCC ATT AAT TCT ACA TCC ACA GTC TCT GAC CCT TCC GCC AGC AAA GTG AAT</td>
<td>ile met lys ala ala ser gln gly leu thr arg ser asp pro phe ala ser lys val asn</td>
</tr>
<tr>
<td>3121/1019</td>
<td>ACA TGG TGG GTCT AAA GCT ATT ATA CCC ATC CTA AGA AGC GCA GGG GTA AAG CTT ACA TCC</td>
<td>thr cys trp ala lys ala ile ile pro ile leu thr thr thr thr thr thr thr thr thr</td>
</tr>
<tr>
<td>3181/1039</td>
<td>GAG CAG TGG GAA GAT CTA TCC CGG CAA TTT CTT CAG CAA CCT TAC TCC GTG ATG TAT</td>
<td>glu glu trp glu asp leu phe pro gln phe arg asn asp gln pro tyr ser val met tyr</td>
</tr>
<tr>
<td>3241/1059</td>
<td>GCC CTA GAT GTG ATG ACC AAG ATG TCC GGC ATG GAT CTG ACG AGT GGG GAT TCC GTC TCT</td>
<td>ala leu asp ile cys thr lys met phe gly met asp leu ser ser gly ile phe ser</td>
</tr>
<tr>
<td>3301/1079</td>
<td>CGT CCT GAG ATA CCT CTA AGC TTC CAT CCC GCG GAC GTC GGC CCA GTG GAG GCT CAC TGG</td>
<td>arg pro glu ile pro leu thr phe his pro ala asp val gly arg val arg ala his trp</td>
</tr>
</tbody>
</table>

Figure 7c. See legend on last page of this sequence.
Figure 7e. Translated sequence of Aura virus. This sequence starts near the 5'terminus of the genome, although the exact 5' end is not known. The translated sequence shown encompasses nsP1, nsP2, and the N-terminal (conserved) region of nsP3. Nucleotides are numbered from the beginning of the sequence; amino acids are numbered from the beginning of the open reading frame.
SEQUENCE ANALYSIS OF WHATAROA VIRUS.

We have obtained most of the sequence of Whataroa virus RNA, 11.7 kb in length. This sequence is being assembled to give the complete sequence of this virus RNA. We were interested in this virus because it represents a geographically isolated Sindbis-like virus, being found in New Zealand and presumably transferred there by migratory birds.

The sequences of a stretch of the nonstructural protein coding region of the Whataroa genome is shown in Figs 8. The sequence begins near the beginning of the nsP2 gene and continues through to the end of the nsP2 region of the virus genome, a stretch of about 2000 nucleotides. From the analysis of this sequence, Whataroa virus can clearly be considered to be a strain of Sindbis virus that has spread to New Zealand. The amino acid sequence deduced from the nucleotide sequence in Fig. 8 is compared to that of the AR339 strain of Sindbis virus, isolated from Egypt in 1952, in Fig. 9. These amino acid sequences are 84% identical. Furthermore, we found that Whataroa virus RNA has the characteristic 3' NTR of the Sindbis viruses.

SEQUENCE ANALYSIS OF OTHER ALPHAVIRUSES

We have obtained the nucleotide sequence encoding the nsP3 and nsP4 genes of several other alphaviruses, in order to examine the relationships of viruses isolated from Australia, India, and South Africa to other alphaviruses. Sequences of this region for Sindbis virus isolated from India in 1953 is shown in Fig. 10, that for a Sindbis virus isolated in Australia in 1975 is shown in Fig. 11, and that for a Sindbis virus isolated from South Africa in 1963 is shown in Fig. 12. The South African isolated came from a human patient exhibiting symptoms of
Figure 8a. See legend on next page.
Figure 8b. Translated nucleotide sequence of Whataroa virus in the region encoding nonstructural protein nsP2. By homology with Sindbis virus, the sequence shown begins at amino acid 97 of nsP2 and continues to the nsP2/nsP3 cleavage site.
Figure 9. Aligned deduced amino acid sequences of the nonstructural protein regions of Whataroa virus and Sindbis virus, beginning with amino acid 97 of Sindbis virus nsP2. The upper sequence in each case is Whataroa virus, and amino acid identity in the Sindbis sequence is indicated with a dot.
Figure 10a. See legend at the end of this sequence.
Figure 10b. See legend at the end of this sequence.
Figure 10c. Nucleotide sequence of the region of the genome encoding nonstructural proteins nsP3 and nsP4 of Sindbis A1036, isolated in India in 1953. The sequence has been translated using the single letter amino acid code.
Figure 11a. See legend on the last page of this sequence.
Figure 11b. See legend on the last page of this sequence.
Figure 11c. Nucleotide sequence of the region of the genome encoding nonstructural proteins nsP3 and nsP4 of an isolate of Sindbis virus isolated from a mosquito pool from Australia in 1975.
Figure 12a. See legend on last page of this sequence.

36
Figure 12b. See legend on the last page of this sequence.
Figure 12c. Nucleotide sequence of the region of the genome encoding nonstructural proteins nsP3 and nsP4 for the Girdwood South African strain of Sindbis virus isolated in 1963.
viral disease. These viruses are all closely related, exhibiting 90% or greater amino acid sequence identity in the conserved region of nsP3 or in nsP4. Conclusions as to sequence relationships are similar to conclusions drawn from the analysis of the 3' NTR.

CONCLUSIONS

We have identified an important antigenic epitope present in E2 of the alphaviruses. This epitope, located in whole or in part within the domain of E2 between residues 170 and 220, depending upon the antibody, is clearly of major importance for the neutralization of the virus infectivity and thus for vaccine design.

We have established the relationships between many of the Sindbis-like alphaviruses. The Sindbis-like viruses, which are found throughout the Old World from Northern Europe to Africa, India, the Philippines and the Australasian region including New Guinea, are a clearly identifiable group of viruses. They share a minimum of 80% amino acid sequence identity in the nonstructural proteins and possess a characteristic and conserved 3' NTR. Virulent strains exist that can cause significant disease in man, and the relationship of the virulent strains to avirulent strains has been established. It is of considerable interest that viruses belonging to this group coexist in many parts of the world with other alphaviruses that are demonstrably different in their epidemiology, serology, organization of the 3' NTR, and evolutionary history, even though many of these non-Sindbis alphaviruses cause diseases very similar to those caused by the virulent Sindbis-like viruses.

We found that a strain of Sindbis virus from Northern Europe that causes Ockelbo disease in Sweden, Pogosta disease in Finland, or Karelian fever in
Russia, a disease characterized by a polyarthritis whose symptoms can persist for months or years, are very closely related to pathogenic strains of Sindbis virus isolated from South Africa. We concluded that a South African strain of Sindbis was introduced into Northern Europe, probably in the 1960s, where it continues to cause epidemics of a significant human disease (Shirako et al., 1991).

We have shown that Aura virus is a New World representative of the Sindbis viruses. Further analysis is required to determine whether it is one of the parents of Western equine encephalitis virus, but the hypothesis that Western equine encephalitis virus is a virus that emerged from a recombination event has received further support from these studies.

We have also shown that high throughput automated DNA sequencing is ideally suited to the rapid analysis of an RNA virus family such as the alphaviruses. These procedures are rapid and generate large amounts of useful information very quickly. Such procedures would be very useful in defining the origin and spread of an epidemic virus.
REFERENCES


