Infectivity Variation and Genetic Diversity among Western Equine Encephalitis Virus Strains

Les P. Nagata[1]*, Wei-Gang Hu[1], Michael Parker[2], Damon Chau[1], George A. Rayner[1], Fay L. Schmaltz[1], and Jonathan P. Wong[1]

1. Chemical and Biological Defence Section, Defence Research and Development Canada – Suffield, Box 4000, Station Main, Medicine Hat, Alberta T1A 8K6, Canada

2. United States Army Medical Research Institute of Infectious Diseases, Frederick, Maryland 21702, USA

*Corresponding author: Chemical and Biological Defence Section, DRDC Suffield; Box 4000, Station Main, Medicine Hat, Alberta T1A 8K6, Canada; Tel: (403) 544-4672; Fax: (403) 544-3388; E-mail: les.nagata@drdc-rddc.gc.ca

Running title: Infectivity variation and genetic diversity among WEEV strains

Words in the main text: 3288

Words in the summary: 233

Number of figures and tables: 7

SUMMARY

Variation in infectivity and genetic diversity in the structural proteins were compared among eight western equine encephalitis virus (WEEV) strains to investigate WEEV virulence at molecular level. A lethal intranasal infectivity model of WEEV was developed in adult BALB/c mice. All eight examined strains were 100% lethal to adult mice in this model, but they varied considerably in the time to death. Based on the time of death, the eight strains could be classified into two pathotypes: a high virulence pathotype consisting of strains California, Fleming, and McMillan; and a low virulence pathotype including strains CBA-87, Mn548, B11, Mn520 and 71V-1658. In order to analyze genetic diversity in the structural genes, 26S RNAs from eight strains were cloned and sequenced. Their deduced amino acid sequences were compared and a cluster diagram was established. More than 96% nucleotide and amino acid identity was found amongst the viruses. The cluster diagram divided the eight WEEV strains into two genotypes, which exactly matched the pathotype grouping, suggesting variation in infectivity may be attributed to genetic diversity in the structural proteins among these eight WEEV strains. Furthermore, a potential amino acid difference in some positions between two groups was found, suggesting the possibility of these amino acid variations contributing to observed differences in virulence. Additional studies need to be undertaken to identify essential amino acids in WEEV infectivity and how these essential amino acids influence virulence.

INTRODUCTION

Western equine encephalitis (WEE), is a severe mosquito-borne infection of the brain of

humans and domestic animals in the Americas. WEE virus (WEEV), was first isolated in 1930, has been responsible for large, periodic, and extensive epizootics and epidemics of encephalitis in equines and humans in United States and Canada [Reisen & Monath, 1988]. It remains an endemic public health concern in North and South Americas, and is a potential biological warfare agent due to its potential aerosol transmissibility.

WEEV is a member of the genus alphavirus, a group of enveloped viruses with a positive sense, single-stranded RNA genome. All alphaviruses share a number of structural, sequence, and functional similarities, including a genome with two polyprotein gene clusters. The nonstructural proteins (nsP) are translated directly from the 5' two-thirds of the genomic RNA. A subgenomic positive-stranded RNA (the 26S RNA) is identical to the 3' one-third of the genome and serves as the translational template for the structural proteins capsid (C), E3, E2, 6K and E1 [reviewed in Schlesinger & Schlesinger, 1996; Strauss & Strauss, 1988; Strauss & Strauss, 1994]. Sequence comparisons of short regions within the nsP4 gene and the E1 protein/3' NTR were determined for many WEEV strains, allowing a preliminary assessment of the nucleic acid phylogenetic relationships within the WEEV antigenic complex [Weaver *et al*., 1997]. These WEEV strains could be grouped into multiple lineages. Further, genetic variation among many WEEV strains from California since 1938 were studied by investigating E2 protein and four major lineages were identified [Kramer & Fallah, 1999].

To better understand alphavirus pathogenesis, studies have been conducted on the virulence of the closely related Sindbis virus (SINV) at molecular level [Davis *et al*., 1986; Lustig *et*

al., 1988; Olmsted *et al*., 1984; Polo *et al*., 1988; Polo & Johnston, 1991]. Certain essential amino acids for SINV virulence have been pinpointed [Davis *et al.*, 1986; Griffin *et al.,* 1989; Tucker *et al.,* 1997]. However, comparable studies on the molecular determinants of virulence of WEEV have not been undertaken, other than a preliminary study of WEEV pathogenicity [Zlotnik, *et al*., 1972] and WEEV virulence differences among different strains [Bianchi *et al*., 1993]. In this study, we compared WEEV viral infectivity in an intranasal route of infection in a mouse model and studied WEEV genetic diversity in a more comprehensive manner, by comparison of the complete 26S structural genome among eight different strains. Relationship between WEEV viral infectivity and phenotypic changes and possible determinants of WEEV virulence at molecular level are discussed.

METHODS

Virus Culture and Purification

Minimal essential media containing 5 % fetal calf serum was used to culture Vero cells (CCL-81) obtained from American Type Culture Collections (ATCC) (Manassas, VA). A summary of eight WEEV strains used was presented in Table 1. A 10 % suckling mouse brain suspension of WEEV 71V-1658 strain was kindly provided by Dr. Nick Karabatsos, Centers for Disease Control, Fort Collins, CO; WEEV Fleming and California strains were purchased from ATCC; WEEV B11 and CBA87 strains were kindly provided by Dr. George Ludwig, United States Army Medical Research Institute of Infectious Disease (Frederick, MD). WEEV strains McMillan, Mn520 and Mn548 were kindly provided by Drs. Mike Drebot and Harvey Artsorb, National Microbiology Laboratory (Winnipeg, MN). Seed

stocks of WEEV strains were made by inoculation of Vero cells with virus suspensions at a multiplicity of infection of less than 0.1. The supernatants were clarified by centrifugation, aliquotted and stored at –70 ^{o}C. All experiments with live virus were carried out in the Defence Research and Development Canada-Suffield (DRDC Suffield) biological level-3 containment facilities, in compliance with Health Canada and Canadian Food Inspection Agency guidelines. Plaque assays were performed as described, in multi 6-well plates, and stained using an agarose neutral red overlay [Greenway *et al*., 1995].

Mouse Infectivity Studies

Female BALB/c mice, 17-25 g (7-20 weeks), were obtained from the pathogen-free mouse breeding colony at DRDC Suffield, with the original breeding pairs purchased from Charles River Canada (St. Constant, QC). The use of these animals was reviewed and approved by Animal Care Committee at DRDC Suffield. Care and handling of the mice followed guidelines set out by the Canadian Council on Animal Care. Virus was administered to the mice by an intranasal route. The volume of inoculum used was 50 µl, containing 1.5×10^3 Plaque Forming Units (PFU) diluted in Hanks Balanced Saline Solution (HBSS). Briefly, mice were anaesthetized with sodium pentobarbital (50 mg/kg body weight) given intraperitoneally. When the animals were unconscious, they were carefully supported by hand with their nose up and the virus suspension in HBSS gently applied with a micropipette into the nostrils. The applied volume was naturally inhaled into the lungs. Infected animals were observed daily, for up to 14 days post-infection. The times to death among groups were compared using the two-tailed t test and one-way analysis of variance, Graph-Pad Prism ver. 4.0 (GraphPad Software, San Diego, CA). Differences were considered statistically

significant at $P < 0.05$.

Extraction of Subgenomic 26S RNA and Reverse Transcriptase-Polymerase Chain Reaction (RT-PCR)

Subgenomic 26S RNA was prepared from cell lysates using the Qiagen RNeasy Mini kit. RNA was precipitated with isopropanol and stored at -70 °C. Prior to use, RNA was washed with 80% (v/v) ethanol, dried and dissolved in nuclease-free water (Promega, Madison, WI). RT-PCR was performed in an Eppendorf Mastercycler Gradient (Mississauga, ON) using an One-Step RT-PCR kit (Qiagen, Mississauga, ON) with 0.1 µg of viral 26S RNA and a pair of primers flanking the Open Reading Frame of the complete structural polyprotein gene encoding C, E3, E2, 6K and E1 proteins. The primers used were: forward primer, 5'-AAGCTTCCGCCAAAATGTTTCCATACCCTCAG-3' and reverse primer, 5'-TCTAGAGTGTATATTAGAGACCCATAGTGAGTC-3'. The reverse transcription reaction was performed in a total volume of 50 µl for 30 min at 45 °C using Omniscript and Sensiscript Reverse Transcriptases (Qiagen). After the reverse transcription step, HotStarTaq polymerase (Qiagen) was activated by an increase in temperature to 95 °C for 15 min followed by 40 cycles of amplification (94 °C for 10 sec, 68 °C x 30 sec, and 68 °C x 4 min) and final extension (72 °C × 10 min). PCR products (3.7 kb) were isolated from 1% agarose gels and purified with QIAquick gel extraction kit (Qiagen).

Cloning and DNA Sequencing

The extracted PCR products were cloned into the pcDNA 3.1 TA vector (Invitrogen, Burlington, ON), following the manufacturer's instructions, and transformed into TOP10

chemically competent *Escherichia coli* cells (Invitrogen). The plasmid DNA was isolated with the QIAprep Miniprep kit (Qiagen). Selected clones were sequenced using the plasmid-specific T7 promoter primer and BGH reverse primer combined with internal, WEEV-specific primers (Table 2) [Netolitzky *et al.,* 2000]. The sequencing reactions were performed using CEQ™ DTCS Quick Start Kit (Beckman, Fullerton, CA). The reaction products were purified by Centri-Sep^{TM} columns (Princeton Separations, Adelphia, NJ), and then run on a CEQ™ 8000 Genetic Analysis System (Beckman). Sequences were assembled and analyzed using Lasergene DNA software (DNASTAR Inc, Madison, WI). The nucleotide sequence data reported in this paper were submitted to GenBank (Fleming, DQ393791; California, DQ393790; McMillan, DQ393792; CBA87, DQ432026; B11, DQ432027; 71V-1658, NC003908; Mn520, DQ393793; and Mn548, DQ393794).

Sequence Analysis

The multiple sequence alignment of complete structural polyproteins from different WEEV strains for creation of a cluster diagram was carried out using the Jotun Hein algorithm in the MegAlign program of the Lasergene Software package (DNASTAR Inc) with the default of gap penalty set to 11, gap length penalty to 3 and PAM250 weight table. Genetic test of selection was conducted by estimating the synonymous and nonsynonymous substitution rates for all eight strains (pairwise analysis) by Nei-Gojobori method using the program of MEGA 3.1 [Kumar *et al*., 2004]. The sequence diversity level among different WEEV strains was designated by values, scoring mismatched amino acids. The values were plotted in windows of 50 amino acids across the complete structural proteins.

RESULTS

Lethal Infection of Mice Using Intranasal Route of Infectivity

When adult BALB/c mice were infected intraperitoneally with WEEV strains Fleming, CBA-87 and 71V-1658 diluted in HBSS to 10^4 PFU in 100 µl, the virus did not induce encephalitis or show overt symptoms of disease. However, if the adult BALB/c mice were inoculated using an intranasal route of infection with eight different WEEV strains, 100 % of adult mice succumbed to a lethal infection using 1.5×10^3 PFU in 50 µl (Fig.1). The different WEEV strains were shown to vary in their range of time to death of the infected mice, and could be divided into two pathotypes ($P<0.05$). Pathotype A included strains California, Fleming, and McMillan, which were more virulent (Fig. 1), with a range of time to death from 5 to 6 days. Pathotype B consisted of strains CBA-87, Mn548, B11, Mn520 and 71V-1658, which were less virulent in mice, with a range of time to death from 8 to 12 days.

Sequence Analysis

Nucleotide sequence comparison of the WEEV structural protein genes showed very high homology with a range of minimum homology (96.8 %) between CBA-87 and Mn548 to a maximum homology (99.4 %) between 71V-1658 and Mn520 (Table 3). Since nucleotide identity was slightly higher than amino acid identity among some pairs of WEEV strains, the synonymous and nonsynonymous substitution rates were estimated for all eight strains (pairwise analysis) by Nei-Gojobori method using the program of MEGA 3.1. The number of nonsynonymous substitutions per nonsynonymous site, the nonsynonymous distance (dN) was greater than the number of synonymous substitutions per synonymous site, the

nonsynonymous distance (dS) for most strains (P<0.05), indicating that the amino acids encoded by the WEEV gene for most strains have evolved by positive selection. A multiple alignment of deduced amino acid sequences of the complete structural proteins of eight strains was listed in figure 2. A comparison of amino acid sequences revealed these differed in less than 4 % of amino acids (Table 3). Small regions of variability were found in C, E1 and E2 proteins, whereas sequences of E3 and 6K proteins were relatively more conserved (Fig.3). The specific sequences of E3 and 6K are required for directing E1 and E2 to endoplasmic reticulum [Liljestrom, 1991; Schlesinger & Schlesinger, 1972]. The relatively conserved region was also noted in the sequences of C-terminal of E1 or E2 proteins, where there is a transmembrane segment, which is responsible for insertion of E1 or E2 into the lipid envelope [Garoff & Simons, 1974; Garoff & Soderlund, 1978]. Furthermore, some alphavirus-conserved sequences were highly conserved among eight strains, such as LAAQIEDLRRSIANLTFK at C positions 37-54, a putative coiled coil α-helix which is important for viral core assembly [Perera *et al*., 2001]; KPGKRQRMCMKLESD at C positions 95-109, which has been shown to bind to ribosomes and to lie within a region that binds genomic RNA [Strauss & Strauss, 1994; Wengler *et al*., 1992]; and VFGGVYPFMWGGAQCFC at E1 positions 80-96, which is thought to be involved in fusion of the viral envelope with cellular membranes to release the nucleocapsid into the cytoplasm of the infected cells [Strauss & Strauss, 1994; Takkinen, 1986]. A cluster diagram for the complete structural protein sequences was presented in figure 4. Two major genotypes were found in 8 WEEV strains. Genotype A included strains Fleming, California, and McMillan, while genotype B was composed of strains CBA-87, B11, 71V-1658, Mn520, and Mn548. This genotype grouping exactly matched pathotype grouping. In genotype B,

strains were clustered primarily by the year of isolation. The oldest strain CBA-87 (1958) occupied basal position, while the second oldest strain B11 (1961) was located in the next to CBA-87, most recent strains Mn548 (1984), and Mn520 (1981) occupied the terminal branches, and 71V-1658 (1971) occurred in the middle branch, indicating this genotype evolved overall as a single lineage since 1958. Also in genotype B, the WEEV strains were distributed in both North and South Americas, suggesting that some WEEV strains are broadly disturbed. This is different from other New World alphaviruses as EEEV and Venezuelan equine encephalitis virus (VEEV), in which the North and South America strains are genetically distinct [Weaver *et al*., 1992]. There are 45 Cys amino acids, potential disulphide bond formation sites and 7 Asn-Xaa-Ser/Thr sequences, potential N-glycosylation sites in the complete WEEV structural protein sequence. All Cys amino acids and Asn-Xaa-Ser/Thr sequences among eight strains were conserved. Although we could not identify consistent amino acid substitutions between two genotypes, there was a potential amino acid difference in some positions, including: C-57, C-89, C-250, E2-23, E2-59, 6K-31, E1-196, E1-349, and E1-374.

DISCUSSION

Much of the available information on the molecular details of the structure and the pathogenesis of the alphaviruses has come from extensive works with two members, SINV and Semliki Forest virus (SFV) [Atkins *et al.,* 1999; Caballero-Herrera & Nilsson, 2003; Davis *et al*., 1986; Gibbons *et al*, 2004; Lustig *et al*., 1988; Olmsted *et al*., 1984; Polo *et al.*, 1988; Polo & Johnston, 1991; Tuittila & Hinkkanen, 2003]. Relatively little is known

about the other alphaviruses, particularly WEEV. Despite their similarities at the molecular level, alphaviruses are diverse in the severity of the diseases they cause in humans and other vertebrates. WEEV can cause severe and sometimes fatal encephalitis in humans and are significant pathogens in horses [Reisen & Monath, 1988]. In contrast, SINV and SFV cause only mild fever and rash or an asymptomatic infection in humans [Shope, 1980]. In order to gain insight into molecular determinants of WEEV pathogenesis, we studied viral infectivity with genetic variation among eight different strains.

An animal model is crucial to study the pathogenesis of viruses. So far, there has not been a reliable infection animal model for WEEV. When we first tried to inoculate WEEV by intraperitoneally, no symptoms of infection of adult mice were noticed, which was different from Bianchi's reports [Bianchi *et al.,* 1993]. We chose BALB/c adult mice with the age of 7-20 weeks, whereas Bianchi used out-bred Swiss NIH adult mice with the age of 4-5 weeks. The difference of mouse strain and age could be the reason why we obtained different results. Research with C57 black mice indicated mice less than 8 weeks old were susceptible to intraperitoneal infectivity (unpublished results). Nevertheless, when adult mice were inoculated intranasally with WEEV at a similar or lower dose, all eight examined strains were 100% lethal to the mice in this model. Based on the time duration of death, the eight strains could be divided into two pathotypes. Interestingly, the more virulent pathotype consisted of WEEV strains isolated during the major epidemic in the 1930s and 1940s. The less virulent pathotype was composed of the more recent isolates of WEEV. Our results are consistent with a study by Bianchi *et al.*, 1993, who examined the virulence of two CBA South American strains, one of which was strain CBA87. This strain, isolated 25 years earlier

than CBA CIV180, developed higher brain titers and killed mice earlier when suckling mice were inoculated with CBA strain viruses by the intracranial or intraperitoneal routes of inoculation. Natural attenuation of WEEV strains over time is a trend that is observed when considering time to death as an indicator of virulence.

Virus virulence is reflected throughout the infectivity cycle, and manifests itself in virus entry to cells, viral replication, viral interaction with host cells, interferon production, immune response induced by viral infection and other factors. The structural proteins play an important role in the interaction between the virus and its surrounding environment. The WEEV structural proteins are potentially related to viral virulence. Thus, the complete structural protein sequence was chosen for analysis among eight different strains to elucidate WEEV infectivity at the molecular level. We found relatively few differences in the 26S structural protein genome between the different strains. More than 96% nucleotide and amino acid identity was found amongst them, which indicates the relative stability of WEEV within its natural environment. The small variable regions were concentrated in C, E1 and E2 proteins, while the 6K and E3 were relatively more conserved. Once the virions are exposed to the external environment, it is subject to a selective pressure exerted by antibodies of infected or immunized animals. As a result, E1 and E2 proteins were shown to be relatively more diverse. The divergence was not randomly distributed over C, E1 and E2. The segments in C, E1 or E2 including those important for WEEV basic biological functions such as E1/E2 insertion to a lipid envelop, C binding to ribosome, viral core assembly, and fusion of viral envelope with infected cells were found to be conserved. Furthermore, three-dimensional topography of the structural proteins of eight strains is likely similar since all of Cys amino

acids and potential N-glycosylation sites in the structural proteins were conserved. The cluster diagram divided the eight strains into two closely related genotypes, which exactly matched the pathotype grouping, suggesting genetic variation of structural proteins contributes to the infectivity changes. Besides structural proteins, there are four non-structural proteins (nsP1, nsP2, nsP3, and nsP4) encoded by 5' two-thirds of the WEEV genome, which are enzymes involved in viral replication. Thus, mutations in the non-structural proteins of viral genomes can play a significant role in virus attenuation by affecting viral replication. Attenuating mutations in non-structural proteins have been identified for SINV [Frolova *et al*., 2002], SFV [Tuittila *et al.*, 2000], and VEEV [Kinney *et al*., 1989]. In our study, we do not know whether difference in non-structural proteins will contribute to the differences in virus infectivity observed. Further study is required to identify their contribution to virulence.

Eight strains of WEEV differed in their virulence for mice. This difference in virulence was correlated to the variations in the amino acid compositions of structural proteins. How did WEEV genetic variation influence its infectivity? A general feature of alphavirus pathogenesis is that remarkably small genetic changes, even a single nucleotide or amino acid substitution in structural proteins, can have dramatic effects on viral virulence. For example, AR339 strain of SINV was not fatal in adult mice, but a neurovirulent strain that caused fatal encephalitis in adult mice only had four amino acid difference from AR339 in the structural proteins [Griffin *et al.*, 1989]. In the E2 of SINV, the residues at positions 55 and 172 determined the neurovirulence for mice of different ages and the efficiency of replication in nervous system tissue [Tucker *et al*., 1997]. Further studies showed a single

residue change from Arg to Ser at E2 position 114 in SINV was sufficient to attenuate virulence in newborn mice and accelerate penetration of BHK cells [Davis *et al*., 1986]. A similar result was obtained for VEEV, which showed only one residue substitution in E2 could change VEEV pathogenesis [Aronson *et al.*, 2000]. In our studies, we found there was a potential amino acid substitution in some positions between two groups, such as positions C-57, C-89, C-250, E2-23, E2-59, 6K-31, E1-196, E1-349, and E1-374, suggesting the potential relative importance of these amino acid variations in virulence changes. Further studies need to be undertaken to identify essential amino acids in WEEV infectivity and how these essential amino acids influence the infectivity.

REFERENCES

Atkins, G. J., Sheahan, B.J. & Liljestrom, P. (1999). The molecular pathogenesis of Semliki Forest virus: a model virus made useful? *J Gen Virol* **80**, 2287-2297.

Aronson, J. F., Grieder, F. B., Davis, N. L., Charles, P.C., Knott, T., Brown, K. & Johnston, R. E. (2000). A single-site mutant and revertants arising in vivo define early steps in the pathogenesis of Venezuelan equine encephalitis virus. *Virology* **270**, 111-123.

Bianchi, T. I., Aviles, G., Monath, T. P. & Sabattini, M. S. (1993). Western equine encephalomyelitis: virulence markers and their epidemiologic significance. *Am J Trop Med Hy* **49**, 322-328.

Caballero-Herrera, A. & Nilsson, L. (2003). Molecular dynamics simulations of the E1/E2 transmembrane domain of the Semliki Forest virus. *Biophys J* **85**, 3646-3658.

Davis, N. L., Fuller, F. J., Dougherty, W. G., Olmsted, R. A. & Johnston, R. E. (1986). A single nucleotide change in the E2 glycoprotein gene of Sindbis virus affects penetration rate in cell culture and virulence in neonatal mice. *Proc Natl Acad Sci USA* **83**, 6771-6775.

Frolova, E. I., Fayzulin, R. Z., Cook, S. H., Griffin, D. E., Rice, C. M. & Frolov, I. (2002). Roles of nonstructural protein nsP2 and Alpha/Beta interferons in determining the outcome of Sindbis virus infection. *J Virol* **76**, 11254-11264.

Gaedigk-Nitschko, K. & Schlesinger, M. J. (1990). The Sindbis virus 6K protein can be detected in virions and is acylated with fatty acid. *Virology* **175,** 274-281.

Garoff, H. & Simons, K. (1974). Location of the spike glycoproteins in the Semliki Forest virus membrane. *Proc Natl Acad Sci USA* **71**, 3988-3992.

Garoff, H. & Soderlund, H. (1978). The amphiphilic membrane glycoproteins of Semliki Forest virus are attached to the lipid bilayer by their COOH-terminal ends. *J Mol Biol* **124**, 535-549.

Gibbons, D. L., Ahn, A., Liao, M., Hammar, L., Cheng, R. H. & Kielian, M. (2004). Multistep regulation of membrane insertion of the fusion peptide of Semliki Forest virus. *J Virol* **78,** 3312-3318.

Greenway, T. E., Elridge, J. E., Ludwig, G., Staas, J. K., Smith, J. F., Giley, R. M. & Michalek, S. M. (1995). Enhancement of protective immune responses to Venezuelan equine encephalitis (VEE) virus with microencapsulated vaccine. *Vaccine* **13**, 1411-1420.

Griffin, D. E., Hahn, C. S., Jackson, A. C., Lustig, S., Strauss, E. G. & Strauss, J. H. (1989). The basis of Sindbis virus neurovirulence. In *Cell biology of virus entry, replication and pathogenesis,* pp. 387-396. Edited R.W. Compans, A. Helenius, & M.B. Oldstone. New York: Alan R. Liss, Inc.

Kinney, R. M., Johnson, B. J., Welch, J. B., Tsuchiya, K. R. & Trent, D. W. (1989). The full-length nucleotide sequences of the virulent Trinidad donkey strain of Venezuelan equine encephalitis virus and its attenuated vaccine derivative, strain TC-83.*Virology* **170**, 19-30.

Kramer, L. D. & Fallah, H. M. (1999). Genetic variation among isolates of western equine encephalomyelitis virus from California. *Am J Trop Med Hyg* **60**, 708-713.

Kumar, S., Tamura, K. & Nei, M. (2004). MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform* **5**, 150-63.

Liljestrom, P. & Garoff, H. (1991). Internally located cleavable signal sequences direct the formation of Semliki Forest virus membrane proteins from a polyprotein precursor. *J. Virol* **65**, 147-154.

Lusa, S., Garoff, H. & Liljestrom, P. (1991). Fate of the 6K membrane protein of Semliki Forest virus during virus assembly. *Virology* **185**, 843-846.

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

Netolitzky, D. J., Schmaltz, F. L., Parker, M. D., Rayner, G. A., Fisher, G. R., Trent, D. W., Bader, D. E. & Nagata, L. P. (2000). Complete genomic RNA sequence of western equine encephalitis virus and expression of the structural genes. *J. Gen. Virol* **81**, 151-159.

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

Perera, R., Owen, K.E., Tellinghuisen, T.L., Gorbalenya, A.E. & Kuhn, R.J. (2001). Alphavirus nucleocapsid protein contains a putative coiled coil alpha-helix important for core assembly. *J Virol* **75,** 1-10.

Polo, J.M., Davis, N.L., Rice, C.M., Huang, H.V. & Johnston, R.E. (1988). Molecular analysis of Sindbis virus pathogenesis in neonatal mice by using virus recombinants constructed in vitro. *J Virol* **62,** 2124-2133.

Polo, J. M. & Johnston, R.E. (1991). Mutational analysis of a virulence locus in the E2 glycoprotein gene of Sindbis virus. *J Virol* **65,** 6358-6361.

Reisen, W.K. & Monath, T.P. (1988). Western equine encephalomyelitis. In *The arboviruses: epidemiology and ecology,* vol. V, pp. 89-137. Edited by T.P. Monath. Boca Raton, FL: CRC Press.

Schlesinger, S. & Schlesinger, M.J. (1996). Togaviridae: The viruses and their replication. In *Fields Virology,* 3rd edn, pp. 825-841. Edited by B.N. Fields, D.M. Knipe & P.M. Howly, Philadelphia: Raven Press.

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

Strauss, J.H. & Strauss, E.G. (1988). Evolution of RNA viruses. *Annual Review of Microbiology* **42,** 657-683.

Strauss, J.H. & Strauss, E.G. (1994). The alphaviruses: gene expression, replication, and evolution. *Microbiological Review* **58,** 491-562.

Takkinen, K. (1986) Complete nucleotide sequence of the nonstructural protein genes of Semliki Forest virus. *Nucleic Acids Res.* **14,** 5667-5682.

Tucker, P.C., Lee, S.H., Bui, N., Martinie, D. & Griffin, D.E. (1997). Amino acid changes in the Sindbis virus E2 glycoprotein that increase neurovirulence improve entry into neuroblastoma cells. *J Virol* **71,** 6106-6112.

Tuittila, M. & Hinkkanen, A.E. (2003). Amino acid mutations in the replicase protein nsP3 of Semliki Forest virus cumulatively affect neurovirulence. *J Gen Virol* **84,** 1525-1533.

Tuittila, M.T., Santagati, M.G., Roytta, M., Maatta, J.A. & Hinkkanen, A.E. (2000). Replicase complex genes of Semliki Forest virus confer lethal neurovirulence. *J Virol* **74,** 4579-4589.

Weaver, S.C., Kang, W., Shirako, Y., Rumenapf, T., Strauss, E.G. & Strauss, J.H. (1997). Recombinational history and molecular evolution of western equine encephalomyelitis complex alphaviruses. *J Virol* **71,** 613-623.

Weaver, S.C., Rico-Hesse, R. & Scott, T.W. (1992). Genetic diversity and slow rates of evolution in New World alphaviruses. *Curr Top Microbiol Immunol* **176,** 99-117.

Wengler, G., Wurkner, D. & Wengler, G. (1992). Identification of a sequence element in the alphavirus core protein which mediates interaction of cores with ribosomes and the disassembly of cores. *Virology* **191,** 880-888.

Zlotnik, I., Peacock, S., Grant, D.P. & Batter-Hatton, D. (1972). The pathogenesis of western equine encephalitis virus (W.E.E.) in adult hamsters with special reference to the long and short term effects on the C.N.S. of the attenuated clone 15 variant. *Br J Exp Pathol* **53,** 59-77.

LEGENDS

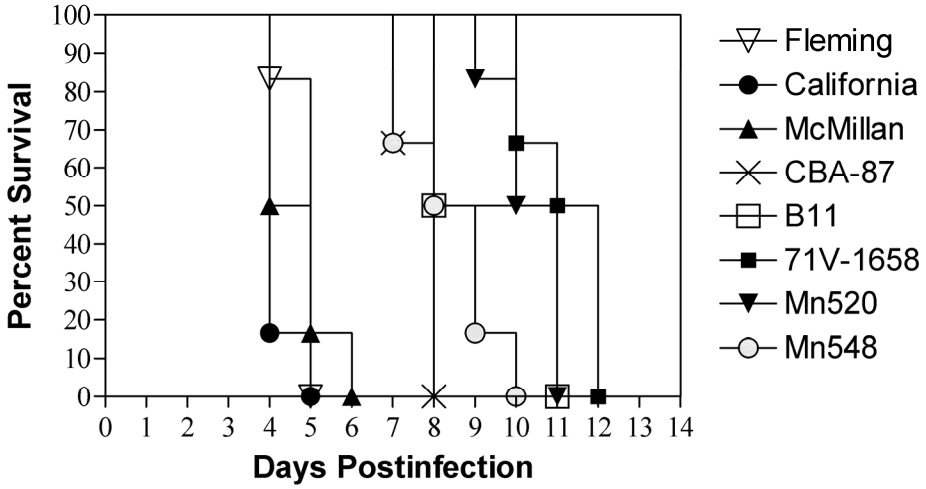

Fig. 1

Fig. 1. WEEV infectivity in a mouse model. Groups of 6 mice were inoculated intranasally with 50 μl of different strains of WEEV. The mice were monitored for 14 days, and the survival percentage was graphed.

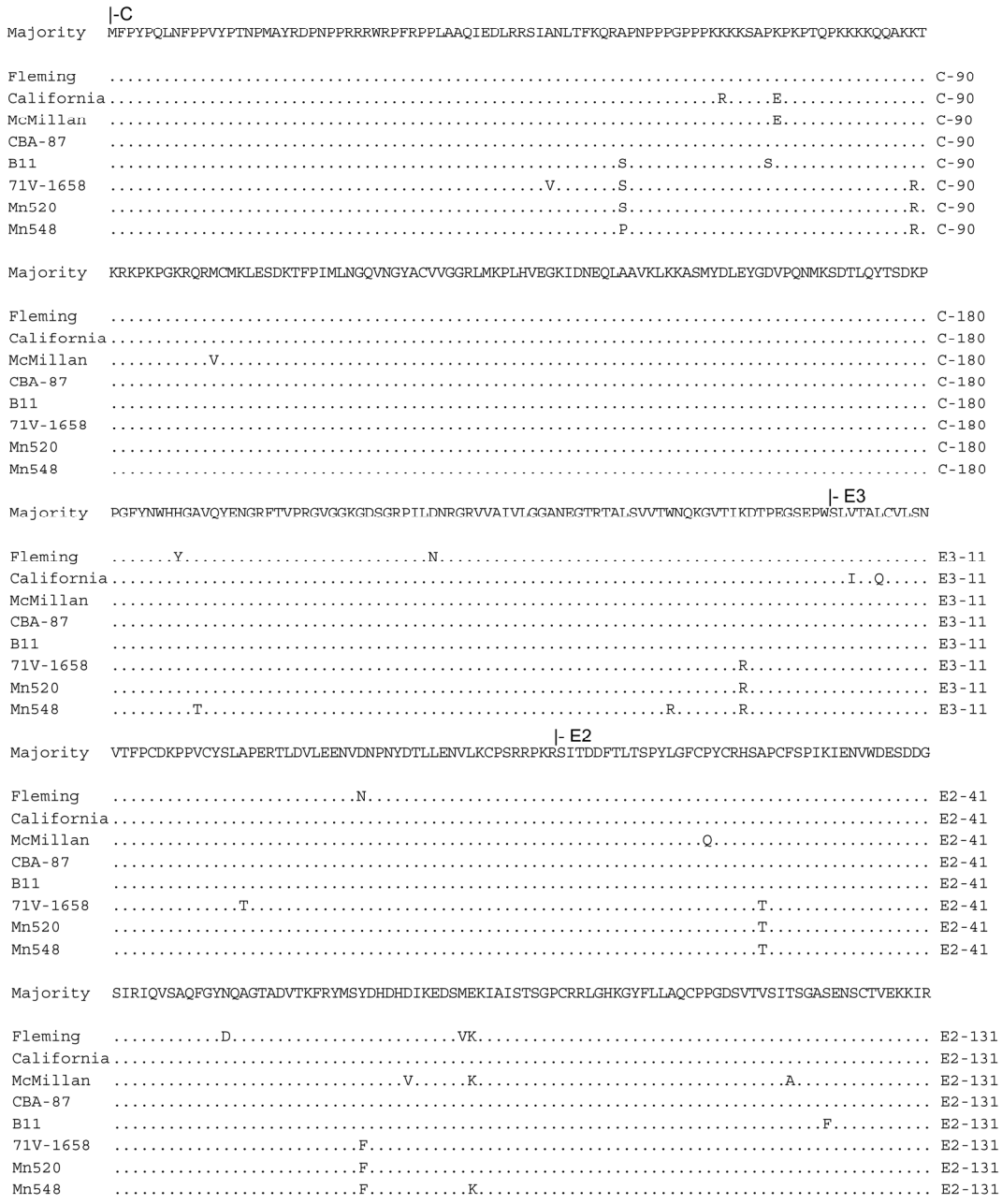
|-C
Majority MFPYPQLNFPPVYPTNPMAYRDPNPPRRRWRPFRPPLAAQIEDLRRSIANLTFKQRAPNPPPGPPPKKKKSAPKPKPTQPKKKKQQAKKT
Fleming C-90
California C-90
McMillan C-90
CBA-87 ..... C-90
B11 ..... C-90
71V-1658 ..... C-90
Mn520 ..... C-90
Mn548 ..... C-90
Majority KRKPKPGKRQRMCMKLESDKTFPIMLNGQVNGYACVVGGRLMKPLHVEGKIDNEQLAAVKLKKASMYDLEYGDVPQNMKSDTLQYTSDKP
Fleming C-180
California C-180
McMillan C-180
CBA-87 ..... C-180
B11 ..... C-180
71V-1658 ..... C-180
Mn520 ..... C-180
Mn548 ..... C-180
|- E3
Majority PGFYNWHHGAVQYENGRFTVPRGVGGKGDSGRPILDNRGRVVAIVLGGANEGTRTALSVVTWNQKGVTIKDTPEGSEPWSLVTALCVLSN
Fleming E3-11
California E3-11
McMillan E3-11
CBA-87 ..... E3-11
B11 ..... E3-11
71V-1658 ..... E3-11
Mn520 ..... E3-11
Mn548 ..... E3-11
|- E2
Majority VTFPCDKPPVCYSLAPERTLDVLEENVDNPNYDTLLENVLKCPSRRPKRSITDDFTLTSPYLGFCPYCRHSAPCFSPIKIENVWDESDDG
Fleming E2-41
California E2-41
McMillan E2-41
CBA-87 ..... E2-41
B11 ..... E2-41
71V-1658 ..... E2-41
Mn520 ..... E2-41
Mn548 ..... E2-41
Majority SIRIQVSAQFGYNQAGTADVTKFRYMSYDHDHDIKEDSMEKIAISTSGPCRRLGHKGYFLLAQCPPGDSVTVSITSGASENSCTVEKKIR
Fleming E2-131
California E2-131
McMillan E2-131
CBA-87 ..... E2-131
B11 ..... E2-131
71V-1658 ..... E2-131
Mn520 ..... E2-131
Mn548 ..... E2-131

Fig. 2-1

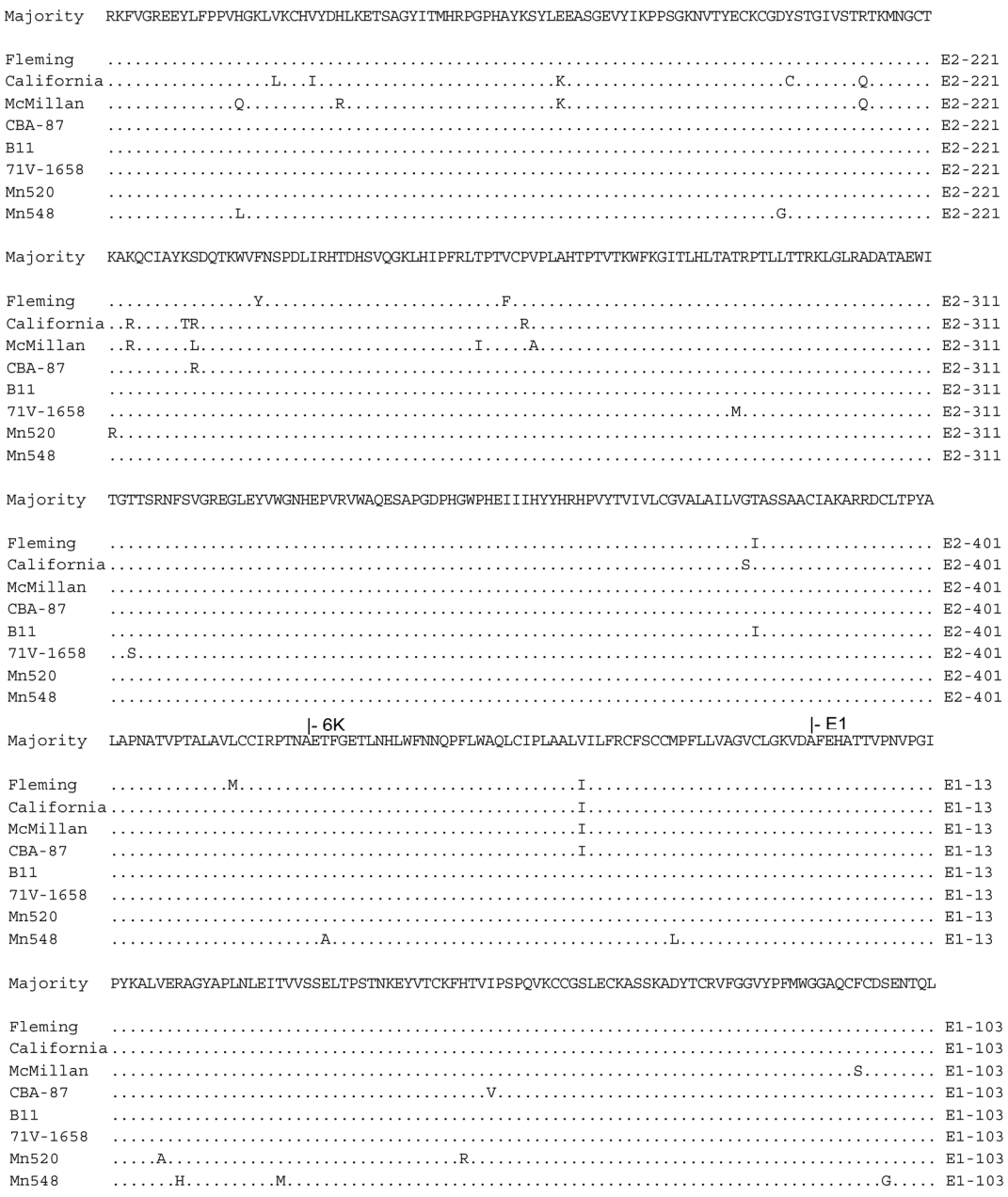

Fig. 2-2

Fig. 2. A multiple alignment of deduced amino acid sequences of eight WEEV strains.

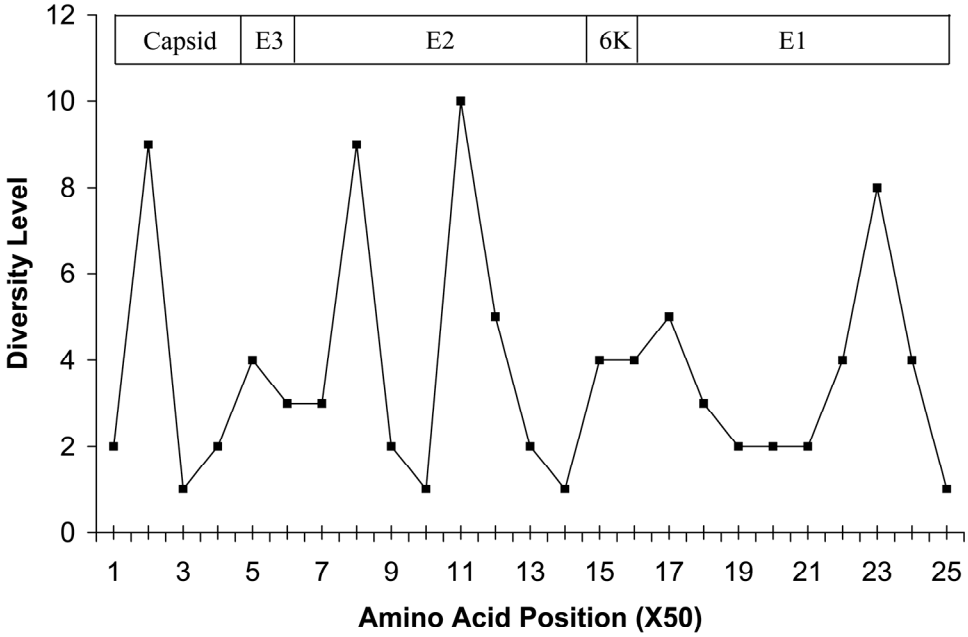

Fig. 3

Fig. 3. Sequence diversity level among different WEEV strains was designated by values, which were plotted in windows of 50 amino acids across the complete structural proteins.

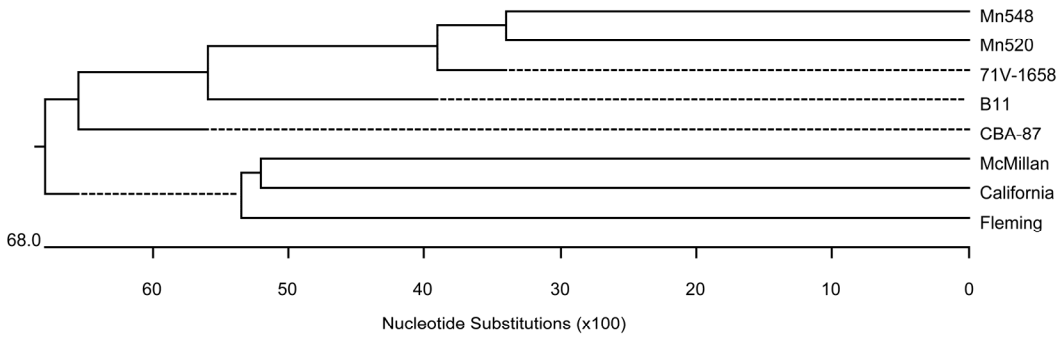

Fig.4

Fig. 4. A **c**luster diagram of WEEV structural proteins was established using the Jotun Hein algorithm of the MEGALIGN program in the DNASTAR package. Horizontal lines are proportional to the number of substitutions between branch points. The length of each pair of branches represents the distance between sequence pairs.

Table 1. Eight WEEV Virus Strains

WEEV Strain	Passage History	Source	Place and Year of Isolation
Fleming	SM5, V2	Human	California, unknown
California	GP/?, SM27, V2	Brain of Horse	California, 1930
McMillan	M2, SM2, V2	Human	Ontario, 1941
CBA-87	SM1, V2	Horse	Argentina, 1958
B11	?,V5, BHK2,		1961
71V-1658	P2, SM1, V2	Horse	Oregon, 1971
Mn520	TC, V2		Manitoba, 1981
Mn548	P2 TC, V2		Manitoba, 1984

SM, suckling mouse; V, Vero cells; GP, guinea pig; ?, unknown; M, mouse; TC: tissue culture

Table 2. Primers used for DNA sequencing of complete WEEV structural proteins

Primer name (genetic sense)	Nucleotide sequence (5'-3')
WEEVN2	GGCTGATGAAACCACTCCAC
WEEVN4	TCACGAGCGGAGCATCTGAG
WEEVN5	GGCATCACCCTCCACCTGAC
WEEVNH	ACGCCATACGCGCTTGCACCG
WEEVN7	CTATTGATCATGCAGTCGCA
WEEVN9	GAGGAGTGGGCGGGAAAGGC
WEEVN10	CTAAAACTCGATGTATTTCC
WEEVP1 (-)	CTGGGGAACGTCGCCATACT
WEEVP2 (-)	CGTTCTCCAGCAGCGTGTCG
WEEVP3 (-)	CTTCAAGTGATCGTAAACGT
WEEVP5 (-)	GTTCGACCAACGCCTTATAC
WEEVP6 (-)	AAGGGTGAAAAAGCGGCTGA
WEEVP8 (-)	TGGAAACTGCCGCCTGGAAT

Table 3. Percent nucleotide (lower left) and deduced amino acid (upper right) identities of the complete structural proteins among eight WEEV strains*

	Flem	Cali	McMi	CBA	B11	71V	Mn520	Mn548
Flem	------	96.7	97.7	98.4	98.1	97.8	97.7	97.4
Cali	98.6	------	98.1	98.4	97.8	97.7	97.6	97.1
McMi	98.5	99.1	------	98.4	97.9	97.7	97.7	97.4
CBA	97.8	98.2	98.0	------	98.7	98.5	98.5	98.0
B11	98.4	98.6	98.5	98.2	------	98.6	98.5	98.0
71V	97.3	97.5	97.5	97.1	98.3	------	99.3	98.7
Mn520	97.2	97.4	97.3	96.9	98.1	99.4	------	98.6
Mn548	97.1	97.3	97.3	96.8	98.0	99.3	99.2	------

* Comparison of 3711 aligned nucleotides and 1236 aligned amino acids. Virus abbreviations: Flem = Fleming; Cali = California; McMi = McMillan; CBA = CBA-87; 71V = 71V-1658.