

Published in final edited form as:

Int J Bioinform Res Appl. 2010 ; 6(2): 134–146.

Flavitrack analysis of the structure and function of West Nile non-structural proteins

Petr Danecek^{*,†} and C. H. Schein^{*,†,‡,§}

C. H. Schein: chschein@utmb.edu

^{*}Sealy Center for Structural Biology and Molecular Biophysics, University of Texas Medical Branch, 301 University Blvd, Galveston TX 77555-0857

[†]Sealy Center for Vaccine Development, University of Texas Medical Branch, 301 University Blvd, Galveston TX 77555-0857

[‡]Departments of Biochemistry and Molecular Biology, University of Texas Medical Branch, 301 University Blvd, Galveston TX 77555-0857

[§]Departments of Microbiology and Immunology, University of Texas Medical Branch, 301 University Blvd, Galveston TX 77555-0857

Abstract

The Flavitrack database groups Flaviviruses, evolutionarily related organisms with high subtype variability, according to their phenotypes. Here, PCPmer tools were used to calculate consensus sequences based on conservation of physicochemical properties (PCP) for 919 sequences of NS2a, a non-structural protein involved in preventing host interferon response to infection. Conserved PCP-motifs were detected, primarily in the N-terminal half of NS2a. One model structure, based on a nuclear receptor, groups residues essential for West Nile infectivity (I59, V61, and M103) in a pocket on the protein surface. These methods will aid in the design of vaccines and specific therapies against Flaviviruses.

Keywords

Consensus sequence; Flavitrack; NS2a; PCPmer; West Nile virus

1 Introduction

Flaviviruses (FV) are important human and animal pathogens. They are transmitted mainly by mosquitoes and ticks and can cause various diseases, including encephalitis and hemorrhagic fever [15,8]. Flaviviruses are single-stranded positive-sense RNA viruses with three structural proteins (C, prM, and E) and several nonstructural proteins (NS1, NS2a, NS2b, NS3, NS4a, 2K, NS4b, and NS5) (Figure 1) [27]. Many, including Yellow Fever (YF), Japanese encephalitis (JBE), and Dengue (DV1 to DV4) are endemic throughout the tropics and subtropical region of the world. In addition, West Nile virus (WN) has recently become disseminated across the continental United States [11,9]. Although most cases of WN and other FV are relatively mild, the disease can progress to a fatal encephalitis, which, according to statistics from the Center for Disease Control, has taken the lives of more than 1000 individuals in the US since the first infections were detected in 1999 [1]. The rapid spread of the disease, and the high mortality rates for severe infections (ranging from 3-14%), has brought into focus the need to better understand the mechanisms of pathogenesis. The Flavitrack database [2] was constructed to allow statistically valid comparisons and particularly to define sequence features

that correlate with disease severity [23]. Better understanding of these features can aid in the design of multivalent vaccines against FV [6,18,10].

Here, we illustrate some of the methods we have developed to use the data in Flavitrack to determine consensus sequences, and generate information that can aid in understanding the effects of known mutants and sequence variants on pathogenesis. The example we chose is the NS2a protein, a small (231 amino acids in Kunjin, a strain of West Nile that is prevalent in Australia) protein that is cleaved from the middle of the virus polyprotein. Although the structure and function is not known, there is evidence that NS2a plays a role in the formation of virus-induced membranes and in preventing the production of interferon after virus infection [21,17]. Interferons (IFN) are proteins produced soon after virus infection that play a major role in the humoral defense against viral replication. Although several FV non-structural proteins have been implicated in preventing IFN induction, a single point mutation in NS2a restores the IFN response [20]. Several other sequence variants in this protein also attenuate the virus by affecting encapsidation, as well as host cell responses [28,17]. This protein is thus a candidate to use in design of antiviral drugs. No experimental structure has been determined for NS2a, nor is its function within the virus life cycle known.

NS2a is the most variable of all FV proteins, according to our analysis, and its sequence identity ranges for mosquito-borne flaviviruses from 20 to 65% (Figure 2). In this work we attempted to identify physicochemical motifs in NS2a common to the mosquito-borne viruses. We first calculated consensus sequences based on conserved physicochemical properties for each of the virus groups separately using the methodology outlined in Section 2.3. The consensus sequences were then analyzed for regions with conserved physicochemical properties, as described in Section 2.2. We also prepared models for this protein. The fold recognition servers [14] returned no significant match, but some of the results suggested that the known mutants may function together in preventing the IFN response. The models and physicochemical motifs are a starting point for structural and mutagenesis studies to explore the function of this protein.

2 Methods

2.1 The Flavitrack Database

The Flavitrack database [2] was established to aid vaccine development efforts against Flaviviruses by providing easy-to-use tools for data retrieval and sequence analysis. Currently, the database contains more than 1400 complete genomic sequences. The data are continually downloaded from the NCBI website and then manually annotated with a “license plate” that encodes information about the disease type, fatality, host, etc. This enables easy interpretation of large multiple sequence alignments. The polyprotein annotations (with respect to the beginning and end of the proteins shown in Figure 1) are added, and, where necessary, the annotations from the NCBI are corrected. Genomic sequences with a desired phenotype can be rapidly selected using the search page of Flavitrack. Here, we used it to filter 919 wild-type NS2a protein sequences of 8 mosquito-borne flavivirus groups: DV1 to DV4, YF, WN, JBE and SLE (St. Louis encephalitis virus). Multiple sequence alignments were then generated with the Clustalw 2.0.3 program with default parameters [16].

2.2 Relative Entropy: a Measure of Sequence Conservation

The aligned sequences of the virus groups were analyzed for conserved/variable regions by calculating the relative entropy (also known as the Kullback-Leibler divergence) according to the five quantitative descriptors E1 to E5 [30]. In this approach, each of the 20 naturally occurring amino acids is represented as a point in a five dimensional space, where the five dimensions roughly correspond to hydrophobicity/hydrophylicity (E1); size (E2); alpha-helix propensity (E3). The property E4 is partially related to the partial specific volume, number of

codons and relative abundance of the amino acids; and E5 correlates weakly with β -strand propensity.

The variability of a given column of the multiple alignment (K) was determined by a weighted sum of Kullback-Leibler divergences (K_p) calculated for each property E1 to E5:

$$K = \sum_{p=E1}^{E5} s_p K_p$$

$$= \sum_{p=E1}^{E5} s_p \sum_{i=1}^n Q_p(i) \log_2 \frac{Q_p(i)}{P_p(i)}, \quad (1)$$

where Q_p are the discrete probability distributions of the five property values observed in the alignment and P_p are the probability distributions of a random sample based on the amino acid distribution in the Flavitrack database. In this equation, the index p iterates over the five properties E1 to E5 and the index i over the discrete probability distribution bins (the 20 amino acids were grouped into 5 bins for each vector, so $n = 5$). Thus $Q_p(i)$ is the fraction of the component p observed in the bin i and $P_p(i)$ is the corresponding background frequency. In other words, $Q_p(i)$ gives the frequency at which a group of amino acids that are most closely matched in properties (according to p) occurs in a given column, and $P_p(i)$ is the corresponding expected random frequency calculated over the whole database.

We introduced a modification to the original approach implemented in the PCPMer software suite [29,4] and used a scaled sum of the relative entropies K_p for the total entropy of a given column rather than their maximum. We scale the individual dimensions of the property space with scale factors s_p to yield a “uniform” space, in which all five properties have equal significance. That is, we scaled the five dimensions to obtain the maximum possible entropy of 1 for each of the five properties. This scaling does not affect the results for positions with high entropies (such as conserved Cysteines and Tryptophans), but improves results for lower entropy positions with more common amino acids such as Leucine or Isoleucine.

2.3 Consensus Sequences

The Flavitrack database does not contain the same number of sequences for each group of FV. For example, the multiple sequence alignment which was analysed in this study has 135 WN sequences but only 45 for JBE (Figure 6). Simply aligning all the sequences and calculating relative entropies for this alignment would bias the selection of motifs and conserved residues toward viral groups with the largest number of representatives. One method to overcome this problem is to select individual sequences for each FV group that are to a large extent identical to one another. However, this also masks single residue variability. Here, consensus sequences that reflect the conserved physicochemical properties of each viral group were constructed, and these were then used to evaluate the total relative entropy of the whole alignment.

One problem that may arise is that the same amino acid could be randomly selected at some highly variable position for all groups and the total relative entropy would be then mistakenly considered as that of an absolutely conserved position. For example, in three distinct alignments where a given position was quite random, an Ala residue may be designated. If only the consensus sequences were read into PCPMer, then the latter program would assume that the Ala at this position was quite conserved, as it occurred in three sequences. To prevent this happening, the relative entropies determined when the groups' consensus sequences were calculated are read into PCPMer as well. If the average of these subtype entropies was lower

than the entropy at that position calculated for the alignment of the consensus sequences, the lower value was taken instead.

To determine a physicochemical property consensus sequence for each group or alignment of sequences, each position (column of the original alignment) was represented by an amino acid that best approximated the average value of the properties summed over all the amino acids that occur in the column. More specifically, for every position of the multiple sequence alignment we calculated the average property values as

$$\bar{E}^p = \sum_{j=1}^m \frac{E_j^p}{m}, \quad (2)$$

where m is the number of amino acids in the given column of the alignment; $p = E1, \dots, E5$; and E_j^p are the five property values of the amino acid at that column of the j -th sequence. The consensus amino acid was that with the least Euclidean distance from the average. Because this geometrical minimization approach might select amino acids which do not appear in the corresponding column of the alignment, we restricted the consensus to the set of amino acids occurring naturally at that position. This restriction is to prevent losing information that may not be captured by the quantitative descriptors.

The five-dimensional property space was biased individually for each column of the alignment to make more significant properties that are most conserved in that column. Except for the bias, the Euclidean distance of an amino acid A_a from \bar{E}^p representing average properties was defined in the usual way

$$|A_a, \bar{E}| = \sqrt{\sum_{p=E1}^{E5} b_p |E_a^p - \bar{E}^p|^2}. \quad (3)$$

The scale factors were calculated as $b_p = s_p K_p / K$, see Equation 1.

2.4 Protein Modeling

We used the basic local alignment search tool (BLAST) [5] and the fold recognition server 3D-PSSM [14] to search for homologues of known function and structure. The NS2a protein sequence was aligned with proteins suggested by the fold recognition server results. Homology modeling was carried out with the MPACK program suite with default parameters [12,7,26,3].

3 Results

3.1 Deriving Conserved Motifs for NS2a

Of all the FV proteins, the NS2a protein is the most variable across the family. As shown in Figure 2, NS2a has very low sequence identity between flavivirus groups. The protein is much more conserved in the tick borne viruses and in the subgroup of mosquito borne viruses causing encephalitic diseases (WN, JBE and SLE), than in mosquito borne overall. Despite the high variability, there are areas of overall physicochemical property conservation, such as for example the positively charged amino acids essential for cleavage by the NS3 protease [24],

or the repeated segments of hydrophobic amino acids, which are rich in Leu and Val residues (Figure 3).

The PCPmer statistical analysis revealed several regions of conserved physicochemical properties common to the group of mosquito borne viruses, as depicted by the thick black and grey lines in Figure 6. The residues D73 and M108, which when mutated resulted in reduction in genome replication and virus attenuation [28], were found to lie within conserved motifs. Another residue, I59, which is critical for NS2a function and its substitution blocked production of virus particles [19,17], fell outside conserved motifs. Other potentially important residues which are not contained in conserved motifs but are significantly conserved throughout FV in terms of physicochemical properties throughout the alignment are L18, K31, L44, V61, I129, R175, V181, K196-K198, and P215.

3.2 Possible Models for NS2a

The structure of the NS2a protein is not known and BLAST search revealed no close homologue of known function or structure. To find an appropriate template, the sequences of the NS2a protein from representatives of three groups of flaviviruses (YF, DV2 and WN, see Figure 3) were individually submitted to the 3D-PSSM fold recognition server. As the three proteins are only about 20-40% identical to one another, we reasoned that any templates that were returned for all three proteins were likely to represent a starting point for modeling the proteins. While no significant match was found, the server returned the same PDB entries as potential templates for all three proteins, although not in the same order of significance. Several templates were selected based on the coverage and percent identity to the template. All models suggested that the protein is composed of multiple helices, but the models differed in their spatial relationship to one another. In our models, the conserved hydrophobic repeats (11-18, 33-44, 51-61, 74-81, 89-95, 105-114, 136-143, 176-184, as indicated in Figure 6) are predominantly in helical regions. Most of these regions are also sufficiently conserved that they are detected as physicochemical motifs (Figure 6).

One possible model, illustrated here, was based on an alignment from 3D-PSSM with the PDB file 1QKM, a structure for a nuclear receptor protein (Figure 3), which placed three important residues close to each other (Figure 3.2). The average E-value was 4.86 with 18% identity and the template covered 230 residues. The backbone RMSD to the template for the aligned regions was 4 Å (due to the flexible loop regions). Another possible model, shown in Figure 3.2, was based on an alignment with the PDB file 1A5T. The average E-value was 5.14 with 20% identity and the template covered 158 residues. The backbone RMSD to the template for the aligned regions was 1 Å. This model did not cover the residues A30, I59 and V61. It is likely that other models for this template will be selected in the future when we have more experimental biophysical data.

4 Discussion

A methodology is described here for comparing genomes of the Flaviviruses, closely related organisms which exhibit high subtype variability. We modified existing methods for physicochemical properties conservation analysis [29] and formulated a new method for determining significantly conserved residues by means of the E1-E5 property vectors (Equations 1-3). We then applied these methods to the NS2a protein, the most variable of the FV proteins (Figures 2-6). Based on analysis of the sequence, known natural mutations (Table 1) and fold recognition results, we prepared models using the MPACK modeling suite (Figures 3.2, 3.2).

One of the goals of our work is to include all available sequence data, so as best to relate sequence changes to phenotype. Typically, only a few reference sequences, usually chosen at

random from all the virus sequences in a group, are compared to one another to determine relationships between the FV groups. However, this can give misleading results. We chose instead to analyze each group separately and then compare the consensus sequences, designed based on the statistical analysis of conserved physicochemical properties, to one another.

While the NS2a protein is variable throughout both the mosquito-borne and hemorrhagic groups, it is quite conserved within the group of mosquito-borne viruses causing encephalitic diseases (WN, JBE, SLE; Figures 2-6). A few highly conserved positions in the protein correlate with its ability to interfere with interferon production by the host cell. For example, most of the mutations that are known to affect NS2a's function map to the relatively more conserved N-terminus. Further, even though there is considerable diversity at the amino acid level, comparison by either amino acid polarity (Figure 3) or according to the E1-E5 property vectors (Figure 6) shows that the overall physicochemical parameters are conserved across the sequence to some degree, and it is likely that these properties account for the protein's functions.

The models shown were based on the structures 1QKM and 1A5T as templates. Although the functional significance of these models cannot at the present time be determined without experimental evidence, the overall structure of the model based on 1QKM (Figure 3.2) is consistent with what is known about the NS2a protein from sequence analysis, and data from mutations and viral adaptation revertants [17]. Although the sequence is rich in conserved hydrophobic residues, the structure based on the model 1QKM suggests that the protein could fold so as to expose conserved positively charged residues, especially those that are cleavage sites for polyprotein processing, in loop regions. The model also brings together three important residues, at different areas of the primary sequence. However, two other important residues, A30 and T149, fall outside this area. The models can be further tested by mutations of other residues in the vicinity of these functional point mutations.

In conclusion, we have established tools to determine consensus sequences according to physicochemical properties of FV proteins. As many of the essential NS proteins do not closely resemble any known protein, bioinformatic analysis of their conserved residues and structural modeling are starting points to define their functions in virus replication.

Acknowledgments

This work was supported by the NIH grant AI064913. Breanne Yingling, UTMB summer undergraduate researcher, did mutation analysis for this paper.

References and Notes

1. CDC West Nile virus homepage.
2. Flavitrack. <http://carnot.utmb.edu/flavitrack/>
3. MPACK - Homology Modeling Package. <http://curie.utmb.edu/mpack/>
4. PCPMer - Physical Chemical Property Based Motif Analyzer. <http://landau.utmb.edu:8080/WebPCPMer/>
5. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *Journal of Molecular Biology* October;1990 215(3):403–10. PMID: 2231712. [PubMed: 2231712]
6. Chang, GwongJen J.; Kuno, Goro; Purdy, David E.; Davis, Brent S. Recent advancement in flavivirus vaccine development. *Expert Review of Vaccines* April;2004 3(2):199–220. PMID: 15056045. [PubMed: 15056045]
7. Cummins, Scott F.; Xie, Fang; de Vries, Melissa R.; Annangudi, Suresh P.; Misra, Milind; Degnan, Bernard M.; Sweedler, Jonathan V.; Nagle, Gregg T.; Schein, Catherine H. Aplysia temptin - the 'glue' in the water-borne attractin pheromone complex. *The FEBS Journal* October;2007 274(20):5425–37. PMID: 17894821. [PubMed: 17894821]

8. Gould EA, Solomon T. Pathogenic flaviviruses. *Lancet* February;2008 371(9611):500–9. PMID: 18262042. [PubMed: 18262042]
9. Gubler, Duane J. The continuing spread of West Nile virus in the western hemisphere. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* October; 2007 45(8):1039–46. PMID: 17879923. [PubMed: 17879923]
10. Guy, Bruno; Almond, Jeffrey W. Towards a dengue vaccine: progress to date and remaining challenges. *Comparative Immunology, Microbiology and Infectious Diseases* March;2008 31(23): 239–52. PMID: 17889365.
11. Hayes, Edward B.; Gubler, Duane J. West Nile virus: epidemiology and clinical features of an emerging epidemic in the United States. *Annual Review of Medicine* 2006;57:181–94. PMID: 16409144.
12. Ivanciuc, Ovidiu; Oezguen, Numan; Mathura, Venkatarajan S.; Schein, Catherine H.; Xu, Yuan; Braun, Werner. Using property based sequence motifs and 3D modeling to determine structure and functional regions of proteins. *Current Medicinal Chemistry* March;2004 11(5):583–93. PMID: 15032606. [PubMed: 15032606]
13. Jia, Yongqing; Moudy, Robin M.; Dupuis, Alan P.; Ngo, Kiet A.; Maffei, Joseph G.; Jerzak, Greta VS.; Franke, Mary A.; Kauffman, Elizabeth B.; Kramer, Laura D. Characterization of a small plaque variant of West Nile virus isolated in New York in 2000. *Virology* October;2007 367(2):339–47. PMID: 17617432. [PubMed: 17617432]
14. Kelley LA, MacCallum RM, Sternberg MJ. Enhanced genome annotation using structural profiles in the program 3D-PSSM. *Journal of Molecular Biology* June;2000 299(2):499–520. PMID: 10860755. [PubMed: 10860755]
15. Kramer, Laura D.; Li, Jun; Shi, PeiYong. West Nile virus. *Lancet Neurology* February;2007 6(2): 171–81. PMID: 17239804. [PubMed: 17239804]
16. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. Clustal W and Clustal X version 2.0. *Bioinformatics* (Oxford, England) November;2007 23(21):2947–8. PMID: 17846036.
17. Leung, Jason Y.; Pijlman, Gorben P.; Kondratieva, Natasha; Hyde, Jennifer; Mackenzie, Jason M.; Khromykh, Alexander A. Role of nonstructural protein NS2A in flavivirus assembly. *Journal of Virology* May;2008 82(10):4731–41. PMID: 18337583. [PubMed: 18337583]
18. Lieberman, Michael M.; Clements, David E.; Ogata, Steven; Wang, Gordon; Corpuz, Gloria; Wong, Teri; Martyak, Tim; Gilson, Lynne; Collier, BethAnn; Leung, Julia; Watts, Douglas M.; Tesh, Robert B.; Siirin, Marina; Travassos da Rosa, Amelia; Humphreys, Tom; Weeks-Levy, Carolyn. Preparation and immunogenic properties of a recombinant West Nile subunit vaccine. *Vaccine* 2007;25(3):414–23. PMID: 16996661. [PubMed: 16996661]
19. Liu, Wen Jun; Chen, Hua Bo; Khromykh, Alexander A. Molecular and functional analyses of Kunjin virus infectious cDNA clones demonstrate the essential roles for NS2A in virus assembly and for a nonconservative residue in NS3 in RNA replication. *Journal of Virology* July;2003 77(14):7804–13. PMID: 12829820. [PubMed: 12829820]
20. Liu, Wen Jun; Wang, Xiang Ju; Clark, David C.; Lobigs, Mario; Hall, Roy A.; Khromykh, Alexander A. A single amino acid substitution in the West Nile virus nonstructural protein NS2A disables its ability to inhibit alpha/beta interferon induction and attenuates virus virulence in mice. *Journal of Virology* March;2006 80(5):2396–404. PMID: 16474146. [PubMed: 16474146]
21. Liu, Wen Jun; Wang, Xiang Ju; Mokhonov, Vladislav V.; Shi, PeiYong; Randall, Richard; Khromykh, Alexander A. Inhibition of interferon signaling by the New York 99 strain and Kunjin subtype of West Nile virus involves blockage of STAT1 and STAT2 activation by nonstructural proteins. *Journal of Virology* February;2005 79(3):1934–42. PMID: 15650219. [PubMed: 15650219]
22. Mackenzie JM, Khromykh AA, Jones MK, Westaway EG. Subcellular localization and some biochemical properties of the flavivirus Kunjin nonstructural proteins NS2A and NS4A. *Virology* June;1998 245(2):203–15. PMID: 9636360. [PubMed: 9636360]
23. Misra, Milind; Schein, Catherine H. Flavitrack: an annotated database of flavivirus sequences. *Bioinformatics* (Oxford, England) October;2007 23(19):2645–7. PMID: 17660525.

24. Nestorowicz A, Chambers TJ, Rice CM. Mutagenesis of the yellow fever virus NS2A/2B cleavage site: effects on proteolytic processing, viral replication, and evidence for alternative processing of the NS2A protein. *Virology* February;1994 199(1):114–23. PMID: 8116234. [PubMed: 8116234]
25. Muñoz Jordan, Jorge L.; Sánchez-Burgos, Gilma G.; Laurent-Rolle, Maudry; García-Sastre, Adolfo. Inhibition of interferon signaling by dengue virus. *Proceedings of the National Academy of Sciences of the United States of America* November;2003 100(24):14333–8. PMID: 14612562. [PubMed: 14612562]
26. Oezguen, Numan; Zhou, Bin; Negi, Surendra S.; Ivanciuc, Ovidiu; Schein, Catherine H.; Labesse, Gilles; Braun, Werner. Comprehensive 3D-modeling of allergenic proteins and amino acid composition of potential conformational IgE epitopes. *Molecular Immunology* August;2008 45(14):3740–7. PMID: 18621419. [PubMed: 18621419]
27. Roehrig, John T. Antigenic structure of flavivirus proteins. *Advances in Virus Research* 2003;59:141–75. PMID: 14696329. [PubMed: 14696329]
28. Rossi, Shannan L.; Fayzulin, Rafik; Dewsbury, Nathan; Bourne, Nigel; Mason, Peter W. Mutations in West Nile virus nonstructural proteins that facilitate replicon persistence in vitro attenuate virus replication in vitro and in vivo. *Virology* July;2007 364(1):184–95. PMID: 17382364. [PubMed: 17382364]
29. Schein, Catherine H.; Zhou, Bin; Braun, Werner. Stereophyicochemical variability plots highlight conserved antigenic areas in Flaviviruses. *Virology Journal* 2005;2:40. PMID: 15845145. [PubMed: 15845145]
30. Venkatarajan, Mathura S.; Braun, Werner. New quantitative descriptors of amino acids based on multidimensional scaling of a large number of physical–chemical properties. *Journal of Molecular Modeling* December;2001 7(12):445–453.

Biographical notes

Petr Danecek received his Ph.D. degree in biophysics in 2008 from the Charles University in Prague, Czech Republic. He is currently working as a postdoctoral researcher at the University of Texas Medical Branch, USA.

Catherine H. Schein, PhD, Associate Professor, is a computational biologist with training in biochemistry, chemical engineering and microbiology. Her current work is in databases for analyzing epitopes of allergens (Structural database of allergenic proteins, SDAP) and viral proteins (Flavitrack), sequence analysis and decomposition (PCPmer), drug design (inhibitors of bacterial toxins and protein aggregation), protein modeling, NMR structure analysis and production of recombinant proteins for structural, functional and clinical studies. She has published over 80 papers and reviews, holds several patents and edited the book *Nuclease Methods and Protocols* (Humana Press, 2001).

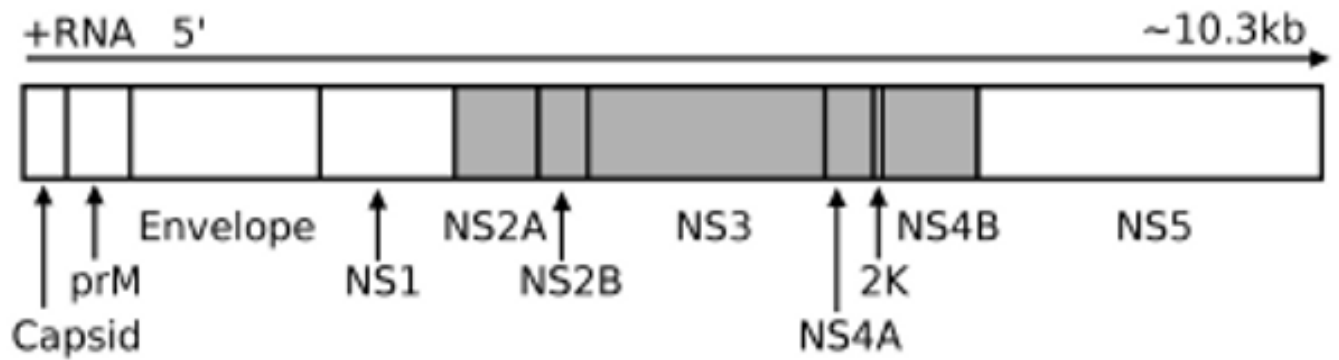
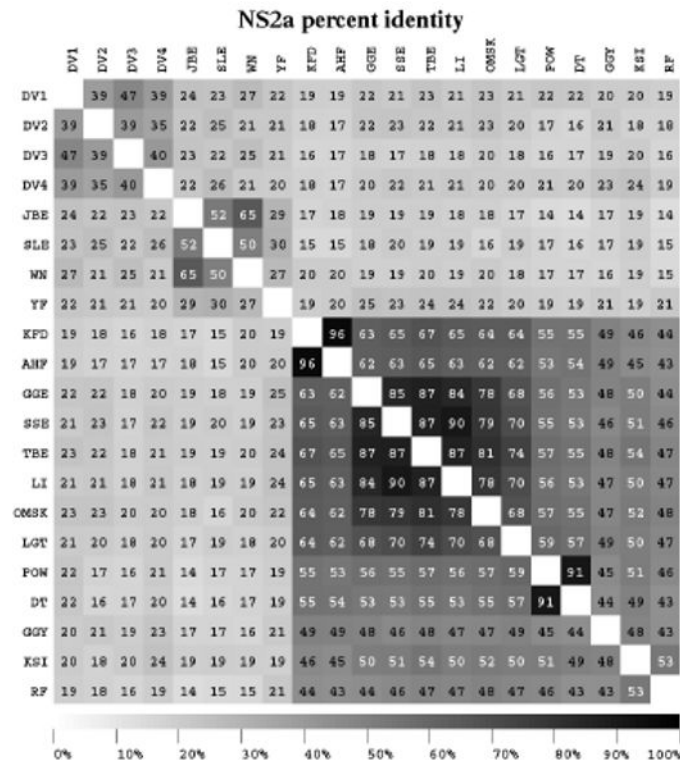


Figure 1.

Flaviviruses are enveloped viruses with a single-stranded, positive-sense RNA genome. The 10.3kb genome is translated as a single polyprotein, which is cleaved into three structural proteins (capsid, membrane and envelope) and several non-structural proteins, the exact function and structure of which is generally not known. The shaded NS proteins have been implicated in preventing host cell IFN response [25]. NS3 is a protease that cleaves several of the NS proteins.

**Figure 2.**

Analysis of the interspecies conservation of the NS2a protein, in terms of sequence identity between FV sequences. The protein is much more conserved in the tick borne flaviviruses (KFD-RF in the above table) and the mosquito borne encephalitic viruses (JBE, SLE and WN) than in the mosquito borne overall (DV1-YF). See the Flavitrack website for virus abbreviations [2]. The length of the shorter sequence was used in the denominator in the percent identity calculations.

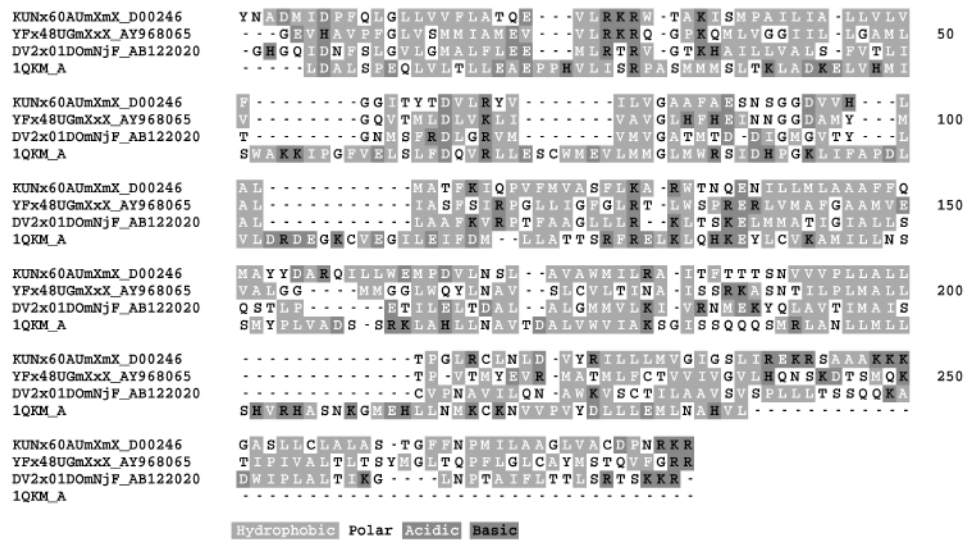


Figure 3.

Alignment of the Kunjin virus NS2a sequence (accession D00246) to the template (1QKM) as suggested by the fold-recognition server 3D-PSSM. Representative sequences for Yellow Fever and Dengue subtype 2 have been included to show the overall variability of the NS2a protein across the mosquito borne FV. The sequences are all 18-19% identical to the 1QKM template for the given alignment. The conserved positively charged amino acids (for example, at residues 25-27 and at the C-terminus) are essential for cleavage of NS2a from NS1 and NS2b [24].

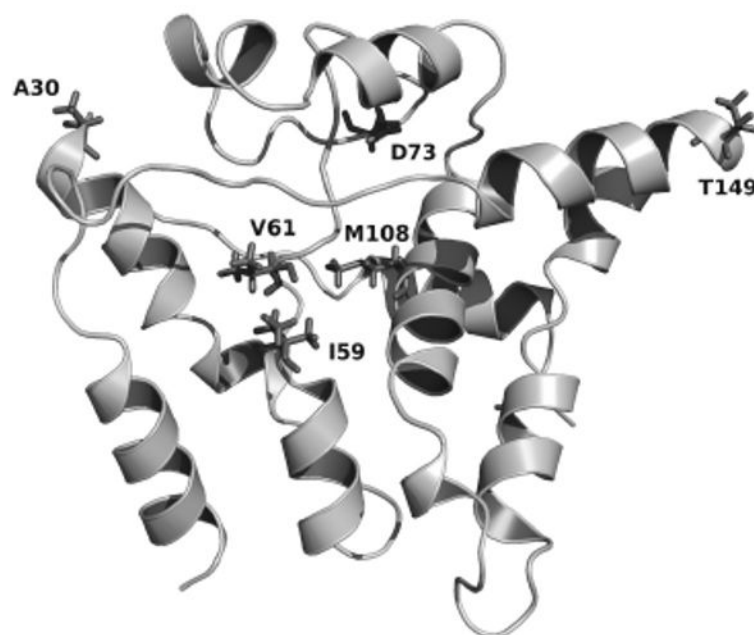


Figure 4.

One possible model of the NS2a protein of Kunjin (WN) virus, based on PDB structure 1QKM, showing the position of residues known to affect the function of the protein. Three of the residues (I59, V61, M108) group together in a negatively charged pocket. The overall positive surface charge could account for the RNA binding ability of NS2a [22].

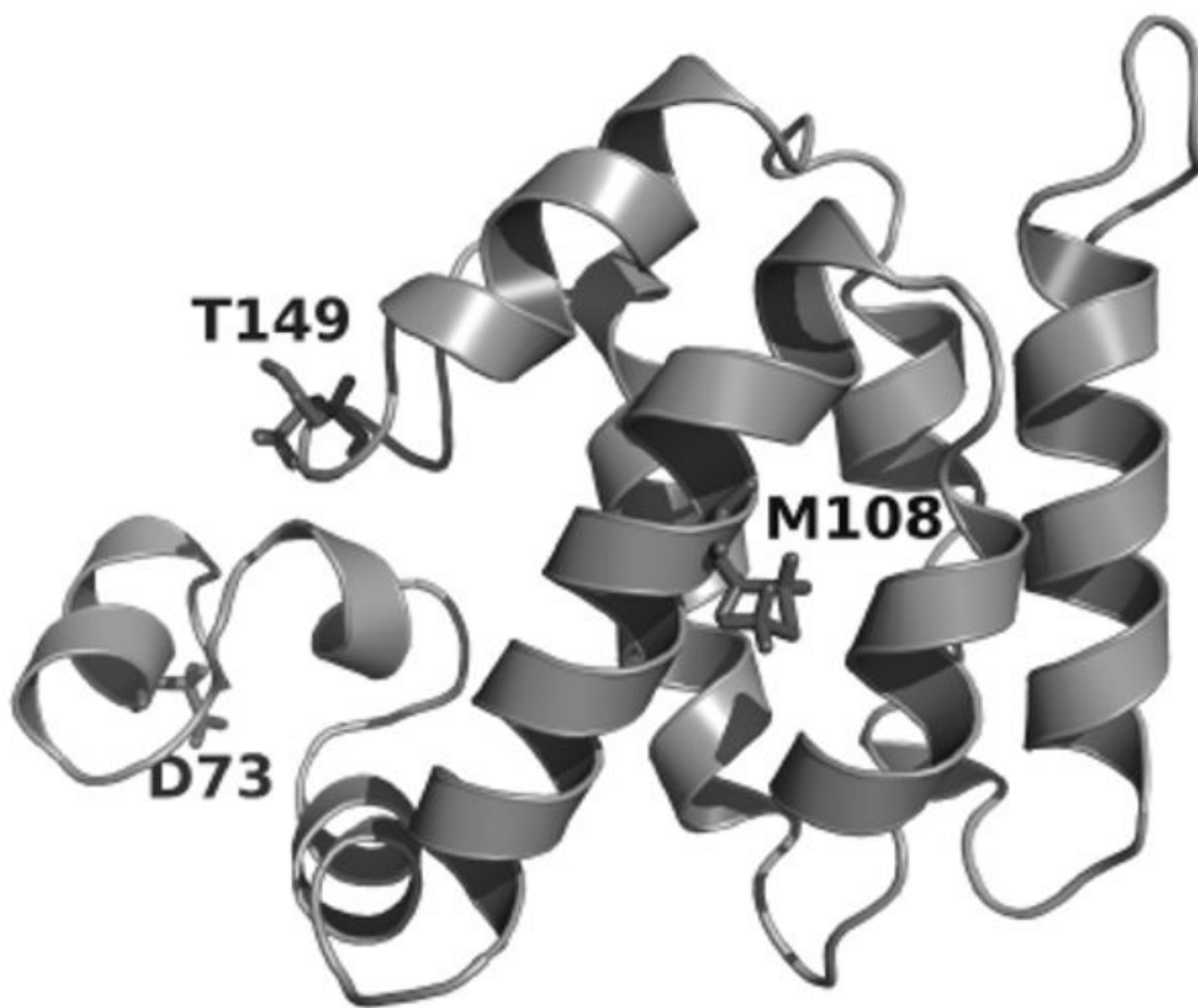


Figure 5.
Another possible model of the NS2a protein of Kunjin (WN) virus, based on PDB structure 1A5T.

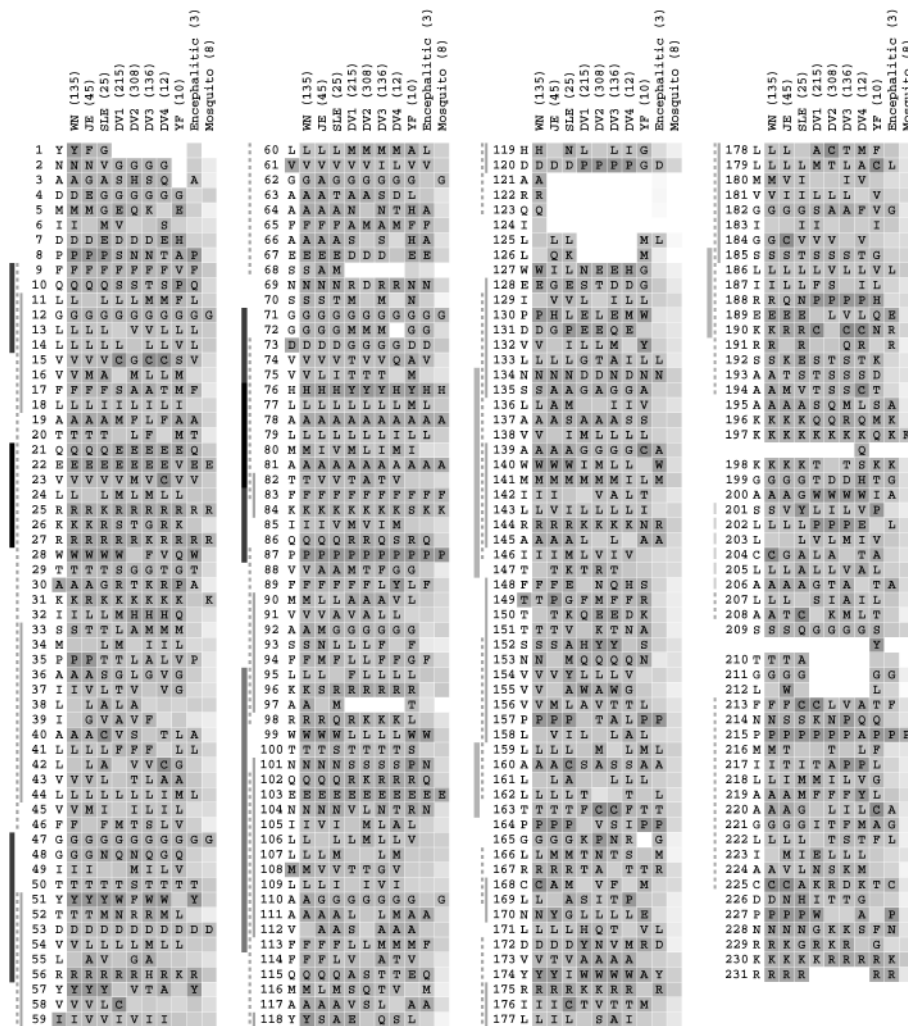


Figure 6.

The consensus sequences of NS2a protein for different groupings of FV are depicted in terms of relative entropies, calculated as described in Methods. Each of the columns corresponds to a whole multiple sequence alignment, where the degree of conservation of the consensus residue is indicated by the color shading. Darker colors correspond to higher entropies; the predominant amino acid is given for significantly conserved positions. Blank shaded fields correspond to variable positions, white areas indicate insertion/deletion between the groups. The “mosquito” column is an alignment of the consensus sequences for 8 groups of FV that are spread by mosquitoes; the “encephalitic” column is from an alignment of the three virus groups causing encephalitic disease (WN, JBE and SLE). The dotted lines indicate helices as predicted by the 3D-PSSM server and thin solid lines indicate helices in our model. The thick black and grey lines are motifs of the “mosquito” alignment as determined by the PCPmer software suite, with modifications as noted in Methods. The lighter lines indicate less significant motifs. The relative entropy cutoff value for the motifs in the more conserved N-terminus (0.45, where 0.5 would be a completely conserved Leu residue, and 1 a completely conserved Cys) was higher than that used for the relatively less conserved C-terminus (0.35). The positions are numbered according to the WN sequence with the accession AF202541. The residues A30, I59, V61, D73, and M108 are known to play a role in the ability of NS2a to block host cell IFN response or prevent virus encapsidation (Table 1).

Table 1

Mutations in NS2a interfere with WN virus growth and restore IFN response.

Virus	Mutation	Effect
Kunjin	A30P	Elicits more rapid induction and higher levels of synthesis of IFN- α/β [20].
Kunjin	I59N	The substitution blocks production of virus particles [19,17].
WNV	V61A	Less efficient replication (with simultaneous mutation prM P54S) [13].
WNV	D73H	Poor replication in vitro, attenuated phenotype in vivo [28].
WNV	M108K	Poor replication in vitro, attenuated phenotype in vivo [28].
Kunjin	T149P	Compensatory mutation, rescues the A30P defect [17].