Sequence signatures in envelope protein may determine whether flaviviruses produce hemorrhagic or encephalitic syndromes

Winona C. Barker · Raja Mazumder ·
Sona Vasudevan · Jose-Luis Sagripanti ·
Cathy H. Wu

Abstract We analyzed the envelope proteins in pathogenic flaviviruses to determine whether there are sequence signatures associated with the tendency of viruses to produce hemorrhagic disease (H-viruses) or encephalitis (E-viruses). We found that, at the position corresponding to the glycosylated Asn-67 in dengue virus, asparagine (Asn) occurs in all seven viral species that cause hemorrhagic disease in humans. Furthermore, Asn was extremely rare at position 67 in six flaviviruses that cause encephalitis, being replaced by Asp in four of them. Of the 3,246 sequences from H- and E-viruses, we found that 2,916 sequences (90%) contained Asn in position 67 for H-viruses or Asp in position 67 for E-viruses. The change from Asn-67 that is prevalent in H-viruses to Asp-67 (common in E-viruses) contributes to a stronger electrostatically negative surface in the E-viruses as compared to the H-viruses. These findings should help predicting the disease potential of emerging and re-emerging flaviviruses and understanding the relationship between protein structure and disease outcome.

Keywords Flavivirus · Envelope protein ·
Hemorrhagic disease · Encephalitis · Sequence signature ·
Dengue · Yellow fever · West Nile · St. Louis encephalitis ·
Tick-borne encephalitis · Japanese encephalitis ·
Kyasanur forest disease · Omsk hemorrhagic fever

W. C. Barker · R. Mazumder · S. Vasudevan · C. H. Wu
Department of Biochemistry and Molecular & Cellular Biology,
Georgetown University Medical Center, Washington,
DC 20007, USA

J.-L. Sagripanti (✉)
US Army Edgewood Chemical Biological Center,
ATTN: AMSRD-ECB-RT, Aberdeen Proving Ground,
MD 21010-54242, USA
e-mail: joseluis.sagripanti@us.army.mil

Introduction

Flaviviruses are classified into three main groups: tick-borne flaviviruses, mosquito-borne flaviviruses, and flaviviruses with no known vector [1]. Flaviviruses naturally pathogenic to humans are found among both the tick-borne and mosquito-borne groups. The major human pathologies caused by these viruses can be grouped into two types: (a) encephalitis and (b) hemorrhagic disease. We refer here to the corresponding viruses as E-viruses or H-viruses, respectively. Some H-viruses also elicit neurologic symptoms [2]. In contrast, hemorrhagic symptoms in infections with E-viruses are reported rarely [3]. On a phylogenetic tree derived from amino acid sequences, H-virus species are intermixed with closely related E-virus species and nonpathogenic species [4, 5], yet there are lineages among the major pathogens that are associated with more or less severe outcomes [6, 7]. Although there are some correlations between phylogeny and disease association, epidemiology, and clinical manifestations [4–6], the underlying genetic characteristics that determine infectivity, pathogenicity, and virulence are not well understood.

The envelope protein of flaviviruses is responsible for phenotypic and immunogenic properties of the virion and is believed to lead the virus entry into cells and, hence, has critical roles in pathogenesis and immune evasion [2]. Virus binds to cell surface receptors and undergoes endocytosis. Acidification of the interior of the endosome induces an irreversible conformational change in the envelope protein that exposes its fusion domain and causes the envelope protein to transition from dimeric association

to trimeric. Viral and vesicle membranes fuse and virus is released into the cytoplasm [8].

In flavivirus envelope proteins, 72 amino acid residues (out of ~500) are completely conserved in viruses producing either hemorrhagic disease (H-viruses) or encephalitis (E-viruses) [9]. Since there is no suitable animal model for flaviviral hemorrhagic disease, it is difficult to study correlations between amino acid changes in the sequences of different viral strains and the ability of these strains to produce hemorrhagic disease. Therefore, we analyzed the nonconserved residues for amino acid signatures that are present in the envelope protein sequences of flaviviruses to identify residues that can be associated specifically to H-viruses or to E-viruses.

Materials and methods

Table 1 shows the virus names, abbreviations, classification as E-virus or H-virus, and UniProt Knowledgebase (UniProtKB) identifiers for representative sequences for the pathogenic flaviviruses. The taxonomic groups were defined based on the ICTVdB nomenclature (http://www.ncbi.nlm.nih.gov/ICTVdb/Ictv/fs_flavi.htm) of the International Committee on Taxonomy of Viruses [10], and on current research indicating that Alkhumra virus is a subtype of Kyasanur forest disease virus (KFDV) [11] and louping ill and related viruses are subtypes of tick-borne encephalitis virus (TBEV) [4]. The designation as H-virus or E-virus in Table 1 refers to the pathology encountered in the most severe cases.

Sequences were aligned using ClustalW [12] and then manually verified. The phylogenetic tree was generated using the neighbor-joining algorithm as implemented in MEGA [13].

Pairwise profile hidden-Markov-model (pHMM) sequence logos were created using the web tool available at http://www.sanger.ac.uk/Software/analysis/logomat-p [14]. A pHMM specifies position-specific amino acid distributions and insertion and deletion probabilities to describe a sequence family, displaying colored, one-letter amino acid symbols in a stack representing each position. The stack heights represent the information content of each position and are determined by the deviation of the position's amino acid frequencies from the background frequencies; the colors are based on chemical properties of the amino acids. A comprehensive review and details on HMM logos and the use of profile HMMs for visualizing sequence features in two

Table 1 Envelope proteins with their source viruses and taxonomic groups within the genus *Flavivirus*

Virus group and name	Virus abbreviation	UniProtKB accession/ID	Residue range[a]
Tick-borne viruses			
Mammalian tick-borne virus group			
Tick-borne Powassan virus[E]	POWV	Q04538/POLG_POWVL	279–775
Kyasanur forest disease virus[H]	KFDV	Q14F58/Q14F58_9FLAV	282–777
Alkhumra hemorrhagic fever virus[H]	AHFV	Q91B85/Q91B85_9FLAV	282–777
Omsk hemorrhagic fever virus[H]	OHFV	Q7T6D2/Q7T6D2_9FLAV	281–776
Tick-borne encephalitis virus[E]	TBEV	P14336/POLG_TBEVW	281–776
Louping ill virus[E]	LIV	P22338/POLG_LIV	281–776
Mosquito-borne viruses			
Dengue virus group			
Dengue virus serotype 1[H]	DENV1	P33478/POLG_DEN1S	281–774
Dengue virus serotype 2[H]	DENV2	P12823/POLG_DEN2P	281–775
Dengue virus serotype 3[H]	DENV3	P27915/POLG_DEN3P	281–773
Dengue virus serotype 4[H]	DENV4	P09866/POLG_DEN4D	280–774
Japanese encephalitis virus group			
Japanese encephalitis virus[E]	JAEV	P32886/POLG_JAEVJ	295–794
Murray Valley encephalitis virus[E]	MVEV	P05769/POLG_MVEV5	293–793
St. Louis encephalitis virus[E]	SLEV	P09732/POLG_STEVM	276–814
West Nile virus[E]	WNV	P06935/POLG_WNV	291–787
Kunjin virus[E]	KUNV	P14335/POLG_KUNJM	291–791
Yellow fever virus group			
Yellow fever virus[H]	YFV	P03314/POLG_YEFV1	286–778

[a] The numbers indicate the starting and ending location of the residues corresponding to the envelope protein in the sequence of the genome polyprotein

E E-virus—encephalitic disease-causing virus, *H* H-virus—hemorrhagic disease-causing virus

groups of protein sequences can be found in the literature [14, 15].

Electrostatic distribution calculations were done using the Adaptive Poisson–Boltzmann Solver (APBS) with default parameters [16]. Homology models of the post-fusion forms of Omsk hemorrhagic fever virus (OHFV, accession Q7T6D2), West Nile virus (WNV, Q91R02), and TBEV (P14336) were built with Swiss-PdbViewer [17] using the available post-fusion structures of dengue virus (DENV) and TBEV (PDB identifiers 1OK8 and 1URZ, respectively) and pre-fusion structures of WNV and TBEV as templates (PDB identifiers 2HG0 and ISVB, respectively).

Amino acid positions and domains mentioned herein are based on the dengue virus serotype 2 (DENV2) envelope protein sequence (P12823, range = 281–775; position 1 of envelope protein is position 281 of P12823 and position 67 of envelope protein is position 347 of P12823); protein accession numbers are from the UniProtKB [18]. UniProt Knowledgebase Release 11.2 was used for scanning the Asn-67 site in all Flaviviridae sequences in the database. The sequence P12823 (envelope protein range = 304–421; ID = POLG_DEN2P) was used as a query to BLAST (using default cutoff) [19] against the 58,752 Flaviviridae sequences. This analysis resulted in ∼4000 hits. The BLAST pairwise alignments were then scanned to determine the amino acid residues present in the position corresponding to Asn-67 in DENV2.

Results

Our own phylogenetic analysis (not shown) using the genome polyproteins of flaviviruses listed in Table 1 supported previous observations that the mosquito-borne and tick-borne human flaviviruses cluster separately and that the ability to cause hemorrhagic or encephalitic disease is not monophyletic in origin [4, 5]. Both Asn and Asp occur at position 67 of the envelope protein among tick-borne and mosquito-borne flaviviruses and among pathogenic and nonpathogenic flaviviruses. Among three major branches of mosquito-borne viruses (dengue, YFV, and JEV), Asn-67 occurs in two branches (dengue and YFV) where viruses are associated with hemorrhagic disease, and Asn-67 does not occur on the third branch where several of the viruses (JEV, WNV, Kunjin, and Murray valley encephalitis) are associated with encephalitic disease. Among tick-borne viruses, one major branch includes both hemorrhagic (OHFV and KFDV) and encephalitogenic (TBEV) flaviviruses, and OHFV (an H-virus) is more closely related overall to TBEV (an E-virus) than to KFDV (another H-virus).

From phylogenetic analysis it is also evident that the distances between the DENV types justify treating them as separate species; therefore, we have omitted from this study 140 dengue virus sequences in UniProt for which the type was not specified. On the other hand, the available sequence data indicate that Alkhumra virus, louping ill virus, and Kunjin virus are properly considered subtypes of KFDV, TBEV, and WNV, respectively. Therefore, we did not treat Alkhumra virus, louping ill virus, and Kunjin virus as separate species. For analysis, we combined the sequence available for Alkhumra virus with KFDV, the sequences for louping ill virus with TBEV, and the sequence reported for Kunjin virus with WNV.

Exploratory studies

We first compared seven envelope protein sequences from H-viruses and seven from closely related viruses that do not cause hemorrhagic disease. Of these, six cause encephalitic disease in humans. Langat virus is not generally associated with human disease, but in extremely rare cases it has caused low-level neurological disease symptoms [20]. Alignments and profiles were constructed for H-virus and non-hemorrhagic virus sequences. We visually examined the aligned pair of HMM profile logos (Fig. 1a, b) and compared the apparent differences against the individual alignments. We observed that Asn occurs at position 67 in all H-virus sequences and in only one of the E-virus sequences analyzed (Fig. 1a). Asn-67 is completely conserved in all the four dengue serotypes (DENV1-4) and is glycosylated in DENV envelope protein, but not in other flaviviral envelope proteins [21].

The prevalence of Asn at position 67 in all seven H-virus sequences initially analyzed, together with the biological role of this region (see "Discussion"), prompted more extensive study. Comparison of the three-dimensional molecular structures of the post-fusion forms of envelope proteins from an H-virus (DENV2) and an E-virus (TBEV) revealed that the orientation of this specific residue at position 67 is different in the two structures (Fig. 1c).

Examination of alignments of envelope protein sequences in UniProtKB from flaviviral species not represented in the initial set revealed that Asp and Asn are common amino acids at position 67, occurring also in flaviviruses that are not pathogenic or cause only mild symptoms in humans, for example, Asp-67 in Usutu virus [22], and Zika virus [23], and Asn-67 in Royal Farm virus and Saumarez Reef virus [4]. Since we could not classify such viruses as either hemorrhagic or encephalitogenic, we limited further studies to the known human pathogens listed in Table 1. We tabulated amino acid types at position 67 and compared surface electrostatic distributions in the vicinity of residue 67 in E-viruses and in H-viruses, as electrostatic properties of a protein play a major role in
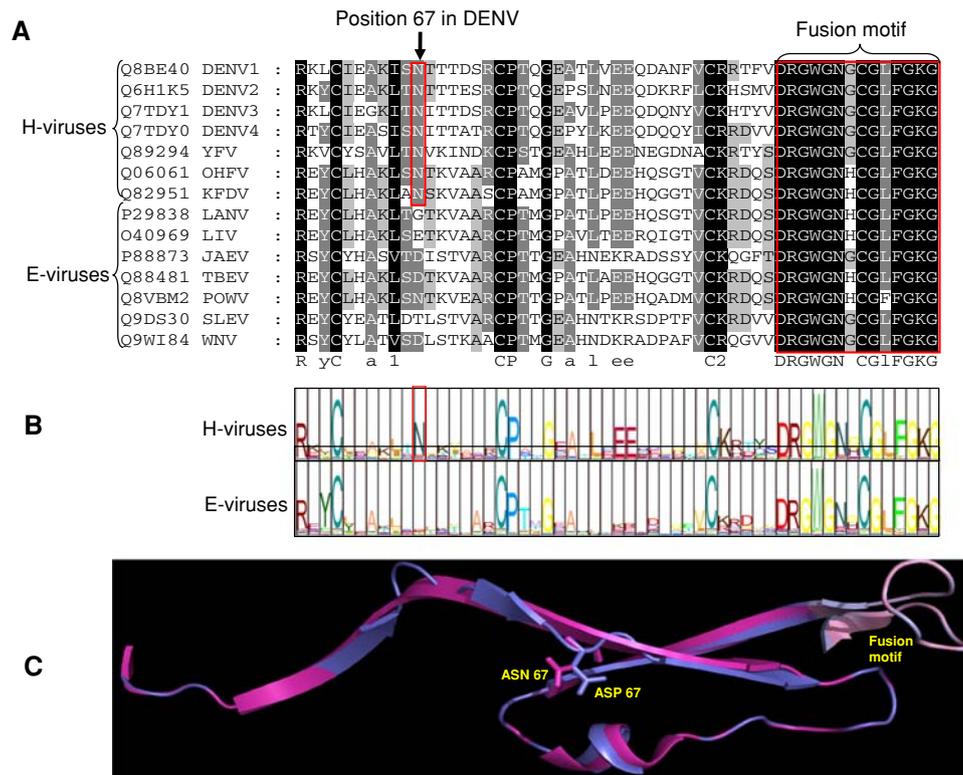
Fig. 1 A potential hemorrhagic signature site (Asn) in envelope proteins of hemorrhagic flaviviruses. Amino acid numbering is based on DENV2 envelope protein. **a** Section of aligned flavivirus envelope protein sequences corresponding to residues 50–111 of DENV2 envelope protein. Sequence accessions (UniProtKB) are: (H-virus) DENV1, Q8BE40; DENV2, Q6H1K5; DENV3, Q7TDY1; DENV4, Q7TDY0; KFDV, Q82951; OHFV, Q06061; and (non-H) LIV, O40969; TBEV, Q88481; WNV, Q9WI84; JAEV, P88873; SLEV, Q9DS30; POWV, Q8VBM2; LGTV, P29838. **b** Section of pairwise profile HMM logos. The hemorrhagic and non-hemorrhagic viruses used to create the logo are as listed for (**a**). Charged residues are colored in shades of red (positively charged) and brown (negatively charged); polar, uncharged residues are in shades of blue and blue-green; aliphatic residues are in shades of yellow, tan, and orange, except methionine, which is lavender; and aromatic residues are in shades of green. **c** Ribbon representation of the post-fusion structure of the envelope protein in the region of Asn-67 in hemorrhagic fever causing virus (DENV2; PDB ID: 1OK8—in magenta) and equivalent Asp residue from an encephalitis-causing virus (TBEV; modeled based on structure 1OK8—in blue). The fusion motifs are shown in lighter shade

defining protein function and the mechanisms of protein–protein and protein–ligand interactions [24, 25].

Amino acid composition at position 67 in pathogenic flaviviruses

We extracted 3,246 available envelope protein sequences from UniProtKB, release 11.2, corresponding to the pathogenic flaviviruses shown in Table 2. Viruses were grouped by disease syndrome (hemorrhagic fever or encephalitis) and, within each group, ordered by their relative incidence as human pathogens. Those viruses shown in bold type in Table 2 are important human pathogens causing 100 or more reported cases annually. Performing BLAST searches of the UniProtKB [18] and multiple sequence alignments, we investigated the amino acid composition of the Asn-67 site among the pathogenic flaviviruses for which sequences were available (Table 2).

Envelope protein position 67 is predominantly Asn in H-viruses

We observed Asn in position 67 of the envelope protein in 1,883 out of 3,246 sequences from pathogenic flaviviruses, occurring in 1,850 (93%) of 2,000 H-virus sequences and in only 33 E-virus sequences. These calculations exclude the 140 dengue sequences of unidentified type (all of which also have Asn-67). Of the 1,794 available DENV sequences, only 8 had Asp-67 instead of Asn-67, but the source of the clones could explain the absence of Asn (see "Discussion" below). Asn was found in 98–100% of sequences in five (DENV1–4 and KFDV) out of eight H-viruses. The most significant exception was yellow fever virus (YFV), in which 30% of sequences had Asn-67 and 70% had His-67. To statistically compensate for large differences between the numbers of sequences generally available for viruses that are serious public health problems (like DENV) and

Table 2 The amino acids occurring at position 67 in flaviviruses associated either with hemorrhagic fever or encephalitis

Disease	Virus species[a]	# Seq[b]	Amino acid at positon 67[c]						
			Asn (N)	His (H)	Asp (D)	Gly (G)	Glu (E)	Thr (T)	Ser (S)
Hemorrhagic fever	**Dengue 1**	528	98%		2%[d]				
	Dengue 2	633	100%						
	Dengue 3	387	100%						
	Dengue 4	246	100%						
	Yellow fever	199	30%	70%					
	Kyasanur forest disease	3	67%						33%[e]
	Omsk hemorrhagic fever	4	75%		25%[f]				
Encephalitis	**Japanese encephalitis**	313			100%				
	Tick-borne encephalitis	157	1%[g]		86%	1%[h]	12%[i]		
	West Nile	621			99%		1%[j]		
	St. Louis encephalitis	122						100%	
	Murray Valley encephalitis	1			100%				
	Powassan/deer tick	32	100%						

[a] Differences among dengue serotypes justify their analysis as separate species. In bold are those viruses of public health impact (causing 100 or more reported cases annually), which have elicited a relatively large number of studies and sequences

[b] The number of available sequences that were extracted from UniProtKB, release 11.2 is indicated. Amino acid abbreviations: *N* asparagine (Asn), *H* histidine (His), *D* aspartic acid (Asp), *G* glycine (Gly), *E* glutamic acid (Glu), *T* threonine (Thr), *S* serine (Ser)

[c] Position 67 refers to the amino acid corresponding to dengue 2 envelope protein position 67 (position 347 of the genome polyprotein sequence P12823)

[d] Seven sequences isolated during dengue outbreaks in Myanmar in 2000–2002: one from mosquito (Q2XRT6) and six from patients (Q2XRM3, Q2XS06, Q2XRU0, Q2XRK0, Q2XRS1, Q2XRR1) and one sequence (Q6TF66) with no publication or description of the source other than its isolation in 2002 in the Philippines

[e] Sequence from Alkhumra hemorrhagic fever virus variant of KFDV (Q91B85)

[f] Sequence from strain Bogolubovka (Q14F59)

[g] Sequence from strain RK1424 (O92342)

[h] Sequences from strain T-blood (Q9YJY8) and Est2546 (Q27JN3)

[i] Includes 18 sequences from louping ill virus subtypes of TBEV and one (Q88474) that, although isolated from Turkish sheep, was overall more similar to Central European strains of TBEV than to louping ill subtypes [38]

[j] Includes the sequence from Kunjin virus variant of WNV (P06935) and WNV sequences from strain MgAn 786/6/1995 (Q2NMF1), strain Madagascar-AnMg98 (Q0H7F2), and isolate LEIV-Krnd88-190 (Q80DN8)

those that are less so (like KFDV), we performed statistical analysis of sequences with and without Asn-67 for each viral species. A Chi-square test from a 2 × 2 table (H-virus or E-virus vs. sum of percentages in each species with Asn site or not) showed that the association of Asn-67 with viruses that produce hemorrhagic syndrome is highly significant (Chi-square = 539.8 and *P*-value < 0.00001).

Envelope protein position 67 is usually Asp in E-viruses

Of 1,246 E-virus sequences examined, Asp-67 was found in 1,066 (86%), including 100% of Japanese encephalitis virus, 86% of TBEV (including its subtypes), and 99% of WNV/Kunjin as well as the single representative of Murray Valley encephalitis virus. Among E-viruses, the most common substitution was the chemically similar amino acid Glu (in 23 of the 1,246 sequences). The Glu substitution was characteristic of the louping ill subtypes of TBEV and also found in 3 of 620 WNV sequences analyzed as well as in the single sequence available from the Kunjin subtype. Of 139 TBEV envelope protein sequences not from louping ill subtypes, 135 (97%) have Asp-67, two have Gly-67, one has Glu-67, and one has Asn-67. St. Louis encephalitis virus was the only major E-virus that did not have Asp-67. A Chi-square test from a 2 × 2 table (E-virus or H-virus vs. having Asp site or not) indicated that the association between Asp-67 and E-virus was statistically significant (Chi-square = 543 and *P*-value < 0.00001).

Electrostatic charge distribution

Differences in electrostatic charge distribution could contribute to the differences in the pathological manifestations of H-viruses and E-viruses. The crystal structure of DENV (PDB-ID 1OK8) in post-fusion form is available. However,

only the pre-fusion form of WNV (2HG0) is available. Since the envelope protein undergoes a major pH-driven conformational change between the pre- and the post-fusion forms, we modeled the post-fusion forms based on available structures from DENV [26]. We compared the electrostatic distribution of the envelope protein from two H-viruses, mosquito-borne DENV2 (P12823) and tick-borne (Q7T6D2), and two E-viruses, TBEV (P14336) and mosquito-borne WNV (Q91R02). In each case, the viruses with similar pathology are distant on the phylogenetic tree whereas OHFV and TBEV are closely related. The results depicted in Fig. 2 indicate that the H- and E-viruses have a distinct electrostatic disposition in the immediate vicinity of residue 67, with a stronger electrostatically negative surface in the E-viruses as compared to the H-viruses.

Discussion

We found a significant correlation between the identity of the residue at position 67 of the sequence of pathogenic flavivirus envelope protein and the nature of the disease caused by the virus. Since H-viruses and E-viruses are intermixed on the tick-borne branch of the flavivirus phylogenetic tree, phylogeny alone cannot explain the differential occurrence of amino acids at position 67. Out of 3,246 sequences from H- and E-viruses, 2,916 sequences (90%) contained Asn in position 67 for virus producing hemorrhagic syndrome or Asp in position 67 for virus producing encephalitic syndrome. Only two flaviviral species that have clear pathology and abundant sequence data did not fit the observed pattern. St. Louis encephalitis virus was the only major E-virus missing Asp-67, having the unique substitution Thr-67 in all 122 sequences surveyed. The only exception to the vast preponderance of Asn-67 in H-viruses was YFV, in which 59 sequences had Asn-67, typical of South American genotype I (found in Brazil) [26], and 140 had His-67. Asn-67 in those 59 YFV sequences is not part of a glycosylation motif. The glycosylation site corresponding to Asn-153 in dengue is also missing in YFV. Thus, neither envelope protein glycosylation nor presence of Asn at position 67 is essential for yellow fever virus to produce hemorrhagic symptoms. It must be mentioned that His, found in several YFV strains, is expected to contribute positive electrostatic distributions similar to Asn because of similarities in their properties. Conversely, the residues Asp and Glu occurring at position 67 in the E-viruses are both charged acidic residues that

Fig. 2 Electrostatic surface potentials of the post-fusion forms of the region around position 67 of the envelope protein of: (**a**) DENV2 (derived from X-ray crystallographic data from PDB-ID 1OK8), (**b**) OHFV (modeled based on structure 1OK8), (**c**) WNV (modeled based on structure of 1OK8), and (**d**) TBEV (modeled based on structure 1OK8). Blue, white, and red regions correspond to positive, neutral, and negative electrostatic potentials, respectively. Surface potentials were calculated covering a range between $-10kT/e$ to $+10kT/e$ (k is the Boltzmann constant, T is the temperature, and e is the unit charge). Amino acid numbering is based on DENV2 envelope protein
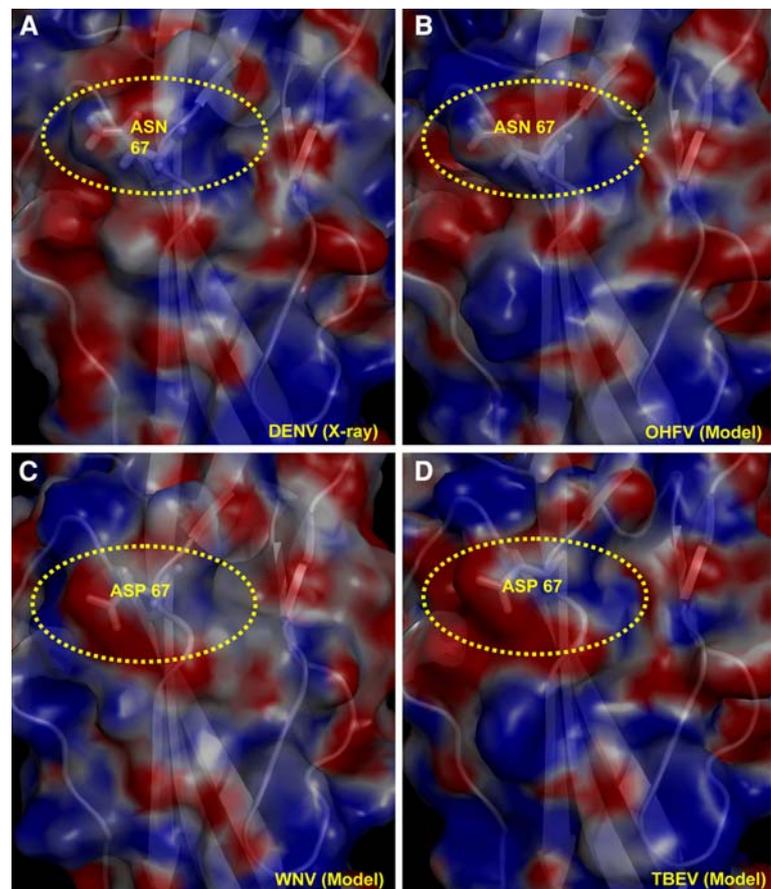
contribute to a similar electrostatically negative distribution around position 67.

Additional exceptions to the amino acid found preponderantly in position 67 fall into one or more of three categories: (1) isolated occurrence of an atypical amino acid within a species; (2) sequence data for a species are sparse; and (3) the virus rarely causes human disease. As noted in "Results" above, DENV1, WNV, and TBEV present examples of isolated atypical amino acids. Of 1,934 surveyed DENV sequences, 8 (all type 1) have Asp-67. Seven of these were isolated from Myanmar in 2000–2002, one from mosquito and six from infected patients, two of whom had hemorrhagic symptoms [27]. However, in each case, other clones from the same patient had Asn-67, so the association of Asn with ability to produce hemorrhagic disease is not contradicted in these cases. The one other dengue sequence with Asp-67 has no publication or description of the source other than its isolation in 2002 in the Philippines. Well-annotated sequence data extracted from viruses isolated from individual patients without passage in cell culture or animal hosts—thus preventing the accumulation of additional sequence changes [7]—will be critical to definitively establish or not a correlation between sequence changes and disease in those atypical cases.

Viruses for which data are sparse include Alkhumra, OHFV, and Powassan. The one available sequence of the Alkhumra genotype of KFDV [11, 28], which has been inappropriately called Alkhurma virus [29], underwent multiple serial passages during which mutation may have occurred. The recently reported [4] OHFV sequence (Q14F59) from strain Bogolubovka not only has Asp-67 but is unexpectedly over 3% different from complete envelope protein sequences described as being from strain "Bogoluvovska" (Q7T6D2, Q06061) and strain Kublin (Q6JJM0), which are more than 99% identical.

Because there are limited stocks of Powassan/deer tick virus, many of the 32 available sequences are from the same few original sources. In particular, Ebel et al. [30] obtained 11 of their samples from Kuno et al. [31]. Louping ill virus, Kunjin virus, and Powassan virus rarely infect humans and do not produce epidemic outbreaks. However, severe Powassan encephalitis with varying fatality rates has been reported [6, 32]. This, together with the presence of Asn-67 in Powassan envelope protein sequences, suggests that Powassan virus may have the potential of causing hemorrhagic disease. This observation suggests that any virus with Asn-67, such as Powassan, should be assumed to be potentially capable of producing hemorrhagic disease, even if few reports of disease in humans have been accumulated. On the other hand, flaviviruses with Asp-67 may be more likely to be encephalitogenic should they become pathogenic. For a virus to infect humans it has to perform several complex steps, including replication in vector and host, transmission from vector to host, cell attachment, cell lysis, etc. Each of these steps is facilitated by different proteins and different motifs and domains within these proteins. Some viruses may have the genetic capability for some but not all of these processes. Such a virus could mutate and become virulent in humans. An examination of the identity of the amino acid in position 67 of the envelope protein may thus be more effective for predicting its disease potential than is its relationship to other pathogenic viruses.

Asn-67 is located in the relatively less-studied domain II of the envelope protein (domain III is the most studied because of its involvement in the virus–host–cell receptor interaction). In a previous computational analysis of flavivirus envelope proteins [9], we have shown that DENV Asn-67 is under negative selection pressure (thereby might be providing physiological advantage to the virus), is in one of the top five high-affinity MHC-II binding 9-mer peptides (T_h-epitope), and is exposed in both the dimer (pre-fusion) and trimer (post-fusion). In DENV, less than 10% of the residues are exposed in both dimer and trimer. It is reasonable to conclude that Asn-67 confers an adaptive advantage to DENV and that changes in this amino acid are eliminated quickly. Indeed, mutants of DENV2 where Asn-67 in envelope protein was replaced by Gln did not grow in cultured mammalian cells [21, 33]. These mutants produced but did not release virus particles and, therefore, did not propagate in mammalian cells. The mouse-adapted Mochizuki strain of DENV1, which is not pathogenic for humans, has lost the Asn-67 glycosylation site because of a Thr-69 to Ile change [34].

A second glycosylation site, Asn-153, is found in dengue and most other flaviviruses in addition to glycosylation sites in the prM protein that could influence pathogenicity. Though missing the Asn-67 glycosylation site, more virulent strains of WNV are glycosylated at the Asn-153 equivalent site while the envelope proteins of less virulent strains are not [35, 36]. Recent studies on Japanese encephalitis virus have shown that a single glycosylation site in the prM protein is critical for viral biogenesis and pathogenicity in mice [37]. Whether the missing glycosylation at position 67 in other flaviviruses is similarly compensated needs further investigation.

We do not know the biological advantage of Asn-67 in viruses where this position is not glycosylated (as in KFDV, YFV, OHFV, and Powassan virus). Both the variety of amino acids found among flaviviruses at this position and the study where Asp-67 was found among multiple samples cloned from individual dengue patients [27] indicate that changes at this position do occur. Therefore, selective pressures must be responsible for the uniformity of amino acid identity within most pathogenic

flavivirus species. These pressures may or may not be related to disease manifestation; pathological propensity could be a "side effect".

The striking electrostatic distribution differences between closely related H- and E-viruses may contribute to the phenotype since electrostatic forces influence the nature of interacting partners. We speculate that mechanisms adopted by H- and E-viruses are distinct and electrostatically controlled. The amino acid sequences from the two tick-borne viruses, TBEV (E-virus) and OHFV (H-virus) (P14336 and Q7T6D2, respectively) in our comparisons are 93% identical and, surprisingly, have distinct electrostatic distributions. In contrast, the two H-viruses, DENV2 and OHFV, although distant phylogenetically, have similar electrostatic distributions, with both having negative patches near Asn-67 (appearing red in Fig. 2).

We suggest that any flavivirus could potentially develop the ability to cause hemorrhagic or encephalitic symptoms. We found that flaviviruses produce either hemorrhagic syndrome or encephalitis in statistically significant correlation with the identity of the amino acid at the position corresponding to the glycosylated Asn-67 of DENV envelope protein. Asn-67 was highly favored in hemorrhagic fever viruses and Asp was highly favored in viruses that cause encephalitis. This association between disease outcome and amino acid in position 67 appears to be correlated with the electrostatic distribution in that region of the envelope protein. These findings should assist in predicting the disease potential of emerging and re-emerging flaviviruses and in understanding the relationship between protein structure and disease outcome.

References

1. S. Cook, E.C. Holmes, Arch. Virol. **151**, 309 (2006). doi:10.1007/s00705-005-0626-6
2. D.J. Gubler, G. Kuno, L. Markoff, in *Fields Virology*, ed. by D.M. Knipe, P.M. Howley, D.E. Griffin, R.A. Lamb, M.A. Martin, B. Roizman, S.E. Straus (Lippincott Williams and Wilkins, Philadelphia, 2007), p. 1153
3. V.A. Ternovoi, G.P. Kurzhukov, Y.V. Sokolov, G.Y. Ivanov, V.A. Ivanisenko, A.V. Loktev, R.W. Ryder, S.V. Netesov, V.B. Loktev, Emerg. Infect. Dis. **9**, 743 (2003)
4. G. Grard, G. Moureau, R.N. Charrel, J.-J. Lemasson, J.-P. Gonzalez, P. Gallian, T.S. Gritsun, E.C. Holmes, E.A. Gould, X. de Lamballerie, Virology **361**, 80 (2007). doi:10.1016/j.virol.2006.09.015
5. M.W. Gaunt, A.A. Sall, X. de Lamballerie, A.K.I. Falconar, T.I. Dzhivanian, E.A. Gould, J. Gen. Virol. **82**, 1867 (2001)
6. T.S. Gritsun, V.A. Lashkevich, E.A. Gould, Antiviral Res. **57**, 129 (2003). doi:10.1016/S0166-3542(02)00206-1
7. K.C. Leitmeyer, D.W. Vaughn, D.M. Watts, R. Salas, I. Villalobos de Chacon, C. Ramos, R. Rico-Hesse, J. Virol. **73**, 4738 (1999)
8. K. Clyde, J.L. Kyle, E. Harris, J. Virol. **80**, 11418 (2006). doi:10.1128/JVI.01257-06
9. R. Mazumder, Z.-Z. Hu, C.R. Vinayaka, J.-L. Sagripanti, S.D.W. Frost, S.L. Kosakovsky Pond, C.H. Wu, Virus Genes **35**, 175 (2007). doi:10.1007/s11262-007-0103-2
10. M.H. van Regenmortel, M.A. Mayo, C.M. Fauquet, J. Maniloff, Arch. Virol. **145**, 2227 (2000). doi:10.1007/s007050070053
11. R.N. Charrel, A.M. Zaki, H. Attoui, M. Fakeeh, F. Billoir, A.I. Yousef, R. de Chesse, P. De Micco, E.A. Gould, X. de Lamballerie, Biochem. Biophys. Res. Commun. **287**, 455 (2001). doi:10.1006/bbrc.2001.5610
12. M.A. Larkin, G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson, D.G. Higgins, Bioinformatics **23**, 2947 (2007). doi:10.1093/bioinformatics/btm404
13. K. Tamura, J. Dudley, M. Nei, S. Kumar, Mol. Biol. Evol. **24**, 1596 (2007). doi:10.1093/molbev/msm092
14. B. Schuster-Böckler, A. Bateman, Bioinformatics **21**, 2912 (2005). doi:10.1093/bioinformatics/bti434
15. L.R. Rabiner, Proc. IEEE **77**, 257 (1989). doi:10.1109/5.18626
16. N.A. Baker, D. Sept, S. Joseph, M.J. Holst, J.A. McCammon, Proc. Natl Acad. Sci. USA **98**, 10037 (2001). doi:10.1073/pnas.181342398
17. N. Guex, M.C. Peitsch, Electrophoresis **18**, 2714 (1997). doi:10.1002/elps.1150181505
18. UniProt Consortium, Nucleic Acids Res. **36**, D190 (2008). doi:10.1093/nar/gkn141
19. J. Ye, S. McGinnis, T.L. Madden, Nucleic Acids Res. **34**, W6 (2006). doi:10.1093/nar/gkl164
20. A.A. Rumyantsev, B.R. Murphy, A.G. Pletnev, J. Virol. **80**, 1427 (2006). doi:10.1128/JVI.80.3.1427-1439.2006
21. J.E. Bryant, A.E. Calvert, K. Mesesan, M.B. Crabtree, K.E. Volpe, S. Silengo, R.M. Kinney, C.Y.-H. Huang, B.R. Miller, J.T. Roehrig, Virology **366**, 415 (2007). doi:10.1016/j.virol.2007.05.007
22. H. Weissenböck, J. Kolodziejek, A. Url, H. Lussy, B. Rebel-Bauder, N. Nowotny, Emerg. Infect. Dis. **8**, 652 (2002)
23. R.S. Lanciotti, O.L. Kosoy, J.J. Laven, J.O. Velez, A.J. Lambert, A.J. Johnson, S.M. Stanfield, M.R. Duffy, Emerg. Infect. Dis. **14**, 1232 (2008). doi:10.3201/eid1408.080287
24. N. Sinha, S.J. Smith-Gill, Curr. Protein Pept. Sci. **3**, 601 (2002). doi:10.2174/1389203023380431
25. D.R. Livesay, D. La, Protein Sci. **14**, 1158 (2005). doi:10.1110/ps.041221105
26. Y. Modis, S. Ogata, D. Clements, S.C. Harrison, Nature **427**, 313 (2004). doi:10.1038/nature02165
27. J. Aaskov, K. Buzacott, H.M. Thu, K. Lowry, E.C. Holmes, Science **311**, 236 (2006). doi:10.1126/science.1115030
28. R.N. Charrel, A.M. Zaki, S. Fagbo, X. de Lamballerie, J. Infect. **52**, 463 (2006). doi:10.1016/j.jinf.2005.08.011
29. T.A. Madani, J. Infect. **51**, 91 (2005). doi:10.1016/j.jinf.2004.11.012
30. G.D. Ebel, A. Spielman, S.R. Telford III, J. Gen. Virol. **82**, 1657 (2001)
31. G. Kuno, H. Artsob, N. Karabatsos, K.R. Tsuchiya, G.J.J. Chang, Am. J. Trop. Med. Hyg. **65**, 671 (2001)

32. D.W.C. Beasley, M.T. Suderman, M.R. Holbrook, A.D.T. Barrett, Virus Res. **79**, 81 (2001). doi:10.1016/S0168-1702(01)00330-6

33. J.A. Mondotte, P.-Y. Lozach, A. Amara, A.V. Gamarnik, J. Virol. **81**, 7136 (2007). doi:10.1128/JVI.00116-07

34. H. Ishak, T. Takegami, K. Kamimura, H. Funada, Microbiol. Immunol. **45**, 327 (2001)

35. S.L. Hanna, T.C. Pierson, M.D. Sanchez, A.A. Ahmed, M.M. Murtadha, R.W. Doms, J. Virol. **79**, 13262 (2005). doi:10.1128/JVI.79.21.13262-13274.2005

36. D.W.C. Beasley, M.C. Whiteman, S. Zhang, C.Y.-H. Huang, B.S. Schneider, D.R. Smith, G.D. Gromowski, S. Higgs, R.M. Kinney, A.D.T. Barrett, J. Virol. **79**, 8339 (2005). doi:10.1128/JVI.79.13.8339-8347.2005

37. J.-M. Kim, S.-I. Yun, B.-H. Song, Y.-S. Hahn, C.-H. Lee, H.-W. Oh, Y.-M. Lee, J. Virol. **82**, 7846 (2008). doi:10.1128/JVI.00789-08

38. J.E. Whitby, S.N. Whitby, A.D. Jennings, J.R. Stephenson, A.D. Barrett, J. Gen. Virol. **74**(Pt 5), 921 (1993). doi:10.1099/0022-1317-74-5-921