West Nile Virus T-Cell Ligand Sequences Shared with Other Flaviviruses: a Multitude of Variant Sequences as Potential Altered Peptide Ligands

Keun-Ok Jung,* Asif M. Khan,a,b,c Benjamin Yong Liang Tan,b Yongli Hu,c,d Gregory G. Simon,a Eduardo J. M. Nascimento,d François Lemonnier,* Vladimir Brusic,* Olivo Miotto,b Tan Wee Tan,b Ernesto T. A. Marques,d,h Rafael Dhalia,b Jerome Salmon,a,i and J. Thomas Augusta,b

Department of Pharmacology and Molecular Sciences, The Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; Perdana University Graduate School of Medicine, Serdang, Selangor Darul Ehsan, Malaysia; Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore; Center for Vaccine Research, University of Pittsburgh, Pittsburgh, Pennsylvania, USA; Unité Immunité Cellulaire Antivirale, Institut Pasteur, Paris, France; Cancer Vaccine Center, Dana-Farber Cancer Institute, Boston, Massachusetts, USA; Mahidol-Oxford Research Unit, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand; Virology and Experimental Therapy Laboratory, Argeu Magalhães Research Center, Recife, Pernambuco, Brazil; and Unité des Rétrovirus Endogènes et Éléments Rétrovirales des Eucaryotes Supérieurs, Institut Gustave Roussy, Villejuif, France

Phylogenetic relatedness and cocirculation of several major human pathogen flaviviruses are recognized as a possible cause of deleterious immune responses to mixed infection or immunization and call for a greater understanding of the inter-Flavivirus protein homologies. This study focused on the identification of human leukocyte antigen (HLA)-restricted West Nile virus (WNV) T-cell ligands and characterization of their distribution in reported sequence data of WNV and other flaviviruses. H-2-deficient mice transgenic for either A2, A24, B7, DR2, DR3, or DR4 HLA alleles were immunized with overlapping peptides of the WNV proteome, and peptide-specific T-cell activation was measured by gamma interferon (IFN-γ) enzyme-linked immunosorbent spot (ELISpot) assays. Approximately 30% (137) of the WNV proteome peptides were identified as HLA-restricted T-cell ligands. The majority of these ligands were conserved in ~88% of analyzed WNV sequences. Notably, only 51 were WNV specific, and the remaining 86, chiefly of E, NS3, and NS5, shared an identity of nine or more consecutive amino acids with sequences of 64 other flaviviruses, including several major human pathogens. Many of the shared ligands had an incidence of >50% in the analyzed sequences of one or more of six major flaviviruses. The multitude of WNV sequences shared with other flaviviruses as interspecies variants highlights the possible hazard of defective T-cell activation by altered peptide ligands in the event of dual exposure to WNV and other flaviviruses, by either infection or immunization. The data suggest the possible preferred use of sequences that are pathogen specific with minimum interspecies sequence homology for the design of Flavivirus vaccines.

West Nile virus (WNV), a member of the genus Flavivirus, is a mosquito-borne RNA pathogen closely related to numerous other flaviviruses of the Japanese encephalitis (JE) group, and to a lesser degree to Dengue virus (DENV), Yellow fever virus (YFV), and several other flaviviruses of about 70 different species (5, 18). The genome is a positive-sense, single-stranded RNA of approximately 10,293 nucleotides encoding a polyprotein of approximately 3,430 amino acids (aa) that is cleaved to produce three structural proteins, capsid (C), precursor membrane (prM), and envelope (E), and seven nonstructural (NS) proteins, NS1, 2a, 2b, 3a, 3a, 3b, and 3c (24, 37). Originally isolated in Africa in 1937 (62), WNV has become an increasingly important human pathogen widely endemic in Africa, Asia, Europe, and North America and a significant agent of viral encephalitis (28, 38, 45). WNV and several of the major human pathogen flaviviruses are known to cocirculate (28, 71); for example, WNV and St. Louis encephalitis virus (LEV) in North America; WNV, DENV, and Japanese encephalitis virus (JEV) around the Indian subcontinent and portions of Southeast Asia (SEA); and WNV, DENV, JEV, and Murray Valley encephalitis virus (MVE) in neighboring places of SEA and Australasia.

The widespread colocalization of WNV with other flaviviruses calls for a greater understanding of the phylogenetic relatedness and possible pathophysiologic associations of flaviviruses. We previously reported a large-scale analysis of the conservation and variability of overlapping nonamers, the typical length of human leukocyte antigen (HLA) class I or class II binding cores of T-cell epitopes (47), of the 2,746 WNV protein sequences collected from the NCBI Entrez Protein Database (17). Notably, of the 88 completely conserved sequence fragments identified, representing 34% of the WNV proteome, 67 were also present in many other flaviviruses with identities of nine or more amino acids. These sequence homologies called attention to a broad inter-Flavivirus risk of altered peptide ligands (APL) (58), T-cell epitopes with one or more amino acid differences, that possibly would result in modified immune responses in the event of exposure to multiple Flavivirus pathogens.

This study focused on analyses of the diversity and presence in

Received 20 January 2012 Accepted 27 April 2012
Published ahead of print 9 May 2012
Address correspondence to J. Thomas August, taugust@jhmi.edu.
K-O.J. and A.M.K. contributed equally to this article.
Supplemental material for this article may be found at http://jvi.asm.org/.
Copyright © 2012, American Society for Microbiology. All Rights Reserved.
doi:10.1128/JVI.00166-12
other flaviviruses of WNV peptide sequences that are HLA restricted and T-cell receptor (TCR) engaged (T-cell ligands). Several reports cite the role of CD8+ cytolytic T lymphocytes (CTL) and CD4+ helper T lymphocytes (HTL) in the immune response to WNV infection (3, 8, 42, 46); however, knowledge of the identities and properties of WNV T-cell epitopes in the human host is limited, and few have been reported to date (19–21, 32, 42, 61). Thus, for this study, large-scale identification of WNV T-cell ligands was performed by use of HLA transgenic (Tg) mice, a leading animal model system for analysis of HLA-restricted T-cell responses to human pathogens (4, 12, 25, 40, 41, 43, 49, 63, 65). Six mice transgenic for prototype HLA proteins, 3 class I and 3 class II, were immunized with 452 peptides covering the entire WNV proteome and peptide-specific T-cell responses were assayed by gamma interferon (IFN-γ) enzyme-linked immunosorbent spot (ELISpot) assay. Characterization of the identified sequences included their evolutionary conservation and variability in reported sequence data of WNV and other flaviviruses. The results provide details of protein sequence homologies of flaviviruses that are relevant to possible consequences of Flavivirus infection and immunization and to the design of a new generation of vaccines that avoid interspecies protein homologies.

MATERIALS AND METHODS

Ethics statement. This study was performed in strict accordance to the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. The protocol was approved by the Johns Hopkins University (JHU) Institutional Animal Care & Use Committee (MO06M106).

HLA transgenic mice. WNV T-cell ligand sequences were identified by use of HLA transgenic mice, a leading animal model system for analysis of HLA-restricted T-cell responses to human pathogens (4, 12, 25, 40, 43, 49, 63, 65), as follows: H-2 class II deficient, HLA-DR2 (DRB1*1501) (68), HLA-DR3 (DRB1*0301) (29, 64), and HLA-DR4 (DRB1*0401)/human CD4 (huCD4) (6, 10); and H-2 class I deficient, HLA-A2.1 (A*0201) (HHD monochain) (44), HLA-A24 (A*2402)/huCD8α (F. Lemonnier unpublished data), and HLA-B7 (B*0702) (51). HLA-DR2 transgenic mice express chimeric molecules, with α1 and β1 domains encoded by the HLA-DRα1*0101 and -DRB1*1501 sequences and the other domains encoded by I-Eα and I-Eβ sequences from the I-E promoters (68). HLA-DR3 transgenic mice express the full-length HLA-DRA1*0101 and -DRB1*0301 sequences (29, 64). HLA-DR4/huCD4 transgenic mice express the full-length HLA-DRA1*0101 and -DRB1*0401 sequences from the I-Eα promoter and the human CD4 sequence from the murine CD3δ promoter (6, 10). HLA-DR2 and -DR3 mice have a homologous deletion of the murine H-2 class II region, and HLA-DR4/huCD4 mice are deficient for I-Aβ and I-Eα. HLA-DR2 mice have a predominant C57BL/6 background, and HLA-DR3 and -DR4/huCD4 mice have mixed backgrounds (B6, B10.H2b, and DBA/1J, 129/Sv, C57BL/6, respectively). HLA-A2.1 (HHD) transgenic mice express a chimeric monochain containing the HLA-A*0201 α1 and α2 domains and the murine H-2Dα α3 domain linked to human β2-microglobulin (huβ2-m) from the HLA-A2.1 promoter and are deficient for H-2D and murine β2-m (mouse β2-m) (44). HLA-A24/huCD8α mice express the full-length HLA-A*2402, huβ2-m, and huCD8α sequences and are deficient for H-2K, H-2D, and huβ2-m (Lemonnier et al., unpublished). HLA-B7 mice express a chimeric heavy chain with the HLA-B*0702 α1 and α2 domains and the H-2Kd α3 domain from the HLA-B7 promoter and are deficient for H-2K and H-2D (51). The three HLA class I transgenic strains have been back-crossed for 6 to 12 generations on the C57BL/6 genetic background (F. Lemonnier, Pasteur Institute). Animals were bred and maintained at the Johns Hopkins School of Medicine Research Animal Resources facilities. Specific-pathogen-free (SPF) transgenic mice were derived through in vivo immunization and to the transfer of outbred foster females (K. Thompson and J. Watson, The Johns Hopkins School of Medicine).

Synthetic WNV peptides. A library of 452 overlapping peptides (see Table S6A to D in the supplemental material) covering the entire WNV proteome (NY99-flamingo 382-99 strain), each 12 to 20 amino acids in length with an overlap of 9 to 11 residues (mainly of 10) (>80% purity), was obtained as lyophilized powders from the Biodefense and Emerging Infections Research Resources Repository, NIAID, NIH (BEI Resources). Each peptide was dissolved in 100% dimethyl sulfoxide (DMSO) and constituted to 20% with sterile filtered water. The final concentration of each peptide was 2 μg/μl. Dissolved peptides were stored at −20°C. Multiple studies have shown that peptides of these lengths are effective as immunogens (2, 9, 20).

Large-scale peptide pool immunization. The 6 HLA transgenic mice were each immunized with the WNV peptides by use of a peptide pool protocol for large-scale T-cell ligand identification (50). The peptides were divided into 4 immunization pools containing 1 μg of each peptide in groups of about 100 peptides each, as follows: pool 1, 88 peptides spanning the prM and E proteins; pool 2, 107 peptides spanning the C, NS1, NS2a, and NS2b proteins; pool 3, 135 peptides spanning NS3, NS4a; and NS4b proteins; pool 4, 122 peptides spanning the NS5 protein (see Table S6A to D in the supplemental material). The criteria for identification of positive pools of peptides or individual peptides were based on the following combination of equations: (i) mean number of spots (peptides) − 2 · standard deviation (SD) > mean number of spots (background); (ii) mean number of spots (peptides) > mean number of spots (background) + 2 SD; (iii) mean number of spots (peptides) − mean number of spots (background) > 10 per million cells.

Day 1. Each pool was mixed with 50 μl zymosan, 10 mg/ml (Sigma-Aldrich Co, St. Louis, MO) in phosphate-buffered saline (PBS) as the adjuvant and administered subcutaneously at the base of the tail to groups of 9 to 12 mice of each transgenic strain (7, 11). Mice immunized with zymosan without peptides served as negative controls.

Days 15 to 19. After one immunization, two mice of each Tg strain on a given day were sacrificed, and their splenocytes were selectively depleted of CD8 or CD4 T cells for HLA class II and class I transgenic strains, respectively, and deconvolution analyses of the T-cell responses to the peptide pools (10 peptides, 1 μg peptide per pool) were performed by IFN-γ ELISpot assays. On the following day 2, for each Tg strain, additional mice, the number depending on the number of possible positive wells selected, were sacrificed, and the individual WNV peptide immunogens identified by the deconvolution analyses were individually tested by ELISpot assay. Experimental values for the validation assays reported herein (see Table S1 in the supplemental material) were obtained with a peptide concentration of 10 μg/ml and a minimum of 3 assays in duplicate with different immunized mice. Control mice were sacrificed on the same day.

Day 21. The remaining mice were immunized a second time on day 21 with the original peptide pool without zymosan.

Day 35. The mice immunized on day 21 and corresponding control mice were sacrificed, and splenocyte T-cell responses to individual WNV peptides were further assessed by ELISpot assay with peptide concentrations of 10, 1, and 0.1 μg/ml. Positive-control dengue virus peptides immunogenic in the relevant transgenic mouse were included in each immunization protocol to evaluate the responses of the individual immunized mouse.

DNA immunization. The WNV strain NY99-flamingo 382-99 NS3 sequence (NCBI Entrez nucleotide database accession number AF196835) with Nhel and Kpnl sites was optimized using the Leto 1.0 software, synthesized by GeneArt Inc. (Toronto, Canada), and inserted into the pF4 vector (15). A pF4 vector encoding the dengue virus type 2 (strain 16681) prM and E antigens (p-DENV-prM-E) (J. Salmon, unpublished data) was used as a control. Four 6- to 8-week-old female HLA-DR2 transgenic mice were immunized subcutaneously at the base of the tail,
twice at 3-week intervals, with 50 μg of the endotoxin-free DNA plasmid p-WNV-NS3, or p-DENV-prM-E. The mice were sacrificed on day 42, and the CD4 T-cell responses to the WNV NS3 peptides were assessed by IFN-γ ELISPOT assays.

**IFN-γ ELISPOT assays.** *Ex vivo* IFN-γ ELISPOT assays were performed using mouse IFN-γ ELISPOT sets (BD Biosciences, San Jose, CA) following the manufacturer’s recommendations. Briefly, 96-well ELISPOT plates were coated with anti-IFN-γ antibody (5 μg/mL) by incubation at 4°C overnight and then blocked with RPMI 1640 medium containing 10% heat-inactivated fetal calf serum, 2 mM L-glutamine, 100 μg/mL streptomycin, and 100 U penicillin, for 2 h at room temperature. Freshly isolated splenocytes from HLA class II and HLA class I transgenic mice were depleted of CD8 and CD4 T cells, respectively, using magnetic beads according to the manufacturer’s protocol (Miltenyi Biotech, Auburn, CA). Flow cytometry analysis of the depleted cells indicated this method routinely achieved >95% depletion of the targeted cells. CD8- or CD4-depleted splenocytes (100 μL containing 0.5 × 10⁸ to 1.0 × 10⁸ cells/well) were plated together with WNV peptides. The final concentration of each peptide was 10 μg/mL in the peptide matrix pool and individual peptide validation assays, and 10, 1, and 0.1 μg/mL in the titration assays. Each peptide preparation was tested in duplicate wells. Cells plated without peptide (medium alone) served as negative controls, and concanavalin A (2.5 μg/mL; Sigma-Aldrich, St. Louis, MO) and known HLA-restricted peptides from dengue virus serotype 3 were included as positive controls. The cells were incubated at 37°C, 5% CO₂, for 16 h. The plates were washed and incubated with biotinylated anti-IFN-γ antibody for 2 h at room temperature, followed by horseradish peroxidase (HRP)-conjugated streptavidin for 1 h at room temperature. Detection was performed with AEC substrate (Calbiochem, San Diego, CA) following the manufacturer’s instructions. IFN-γ spot-forming cells (SFC) were counted using the Immunospot Series 3B Analyzer ELISPOT reader and Immunospot software version 3.0 (Cellular Technologies, Shaker Heights, OH). Experimental values were expressed as the mean numbers of SFC/10⁶ CD8- or CD4-depleted splenocytes ± SD, after subtraction of values from negative controls (background). Positive ELISPOT responses were defined as values above 10 and above the background plus 2 SD. Each ELISPOT-positive response was confirmed by three assays: by matrix screening, individually by the validation assay with the individual peptide, and by peptide titration. The apparent binding avidity of the WNV peptides to either the HLA or TCR molecules in the IFN-γ ELISPOT assay was measured by titration of the peptides over a 100-fold range of concentrations (10, 1, and 0.1 μg/mL).

**Prediction of putative minimal nonamer T-cell epitopes within the ELISPOT-positive peptides.** The minimal T-cell epitopes within the 15 to 19 amino acids of the ELISPOT-positive peptides (WNV T-cell ligands) were predicted by use of NetMHC 3.2 (http://www.cbs.dtu.dk/services/NetMHC-3.2/) (26) and NetMHCIIpan 3.0 (http://www.cbs.dtu.dk/services/NetMHCIIpan/) (39) immunoinformatics Web servers for HLA class I (A*0201, A*2402, and B*0702) and class II (DRB1*1501, DRB1*0301, and DRB1*0401) alleles, respectively. These servers were retrieved from the NCBI Entrez Protein Database (1, 70) through the immunoinformatics Web servers for HLA alleles, respectively. These servers were retrieved from the NCBI Entrez Protein Database (1, 70) through the NCBI Taxonomy Browser application (Taxonomy ID 11082) as described by Koo et al. (17). These included over 100 aligned sequences of each protein (C, 264; prM, 417; E, 927; NS1, 164; NS2a, 143; NS2b, 146; NS3, 146; NS4a, 142; NS4b, 141; and NS5, 256).

**Entropy analysis of WNV T-cell ligand sequences.** The evolutionary conservation and variability of the identified T-cell ligand regions in the alignments of the recorded WNV sequences were measured by use of a modified Shannon’s entropy computation function (16, 56) in the Antigenic Variability Analyzer (AVANA) software (http://sourceforge.net/projects/avana/) (33). The entropy of a ligand sequence (15 to 19 aa) position in the alignment was measured as the average of the entropies of the (7 to 11) nonamer positions within the ligand sequence position. AVANA was also used to study the incidence of the individual ligand sequence and its intraspecies variants in the corresponding WNV protein alignments. For a given WNV T-cell ligand sequence present in the alignment, the intraspecies variants are defined as peptides from that region of the alignment that differed by at least 1 amino acid from the ligand.

**T-cell ligand sequence homologies with other flaviviruses.** Homologies of the WNV (Taxonomy ID 11082) T-cell ligand sequences were identified by performing a BLASTP (31) search against all protein sequences of all flaviviruses (Taxonomy ID 11051) present in the NCBI “nr” database (as of January 2009). The parameters set for the search by use of the NCBI BLAST Web server (http://blast.ncbi.nlm.nih.gov) were as follows: set limit by Entrez query to “txid11051[Organism:exp] NOT txid11082[Organism:exp] NOT txid81077[ORGN],” disabled the option “automatically adjust parameters for short sequences,” disabled “low-complexity” filter, and set maximum number of aligned sequences to be displayed to 20,000, expect threshold to 200,000, word size to 2, matrix to “PAM30,” gap costs to “Existence: 9, Extension: 1,” and compositional adjustments to “no adjustment.” Artificial sequence hits were removed by the “NOT txid81077[ORGN]” keyword.

The incidence of each shared WNV ligand was analyzed in reported sequences of the following six major flaviviruses that had sufficient published data (20 or more sequences corresponding to the ligand): LEV, JEV, *Tick-borne encephalitis virus* (TBEV), Powassan encephalitis virus (PV), YFV, and DENV (one or more of the four dengue virus serotypes). The incidence was determined by use of a BLASTP search performed at the NCBI BLAST Web server with a similar setting as above, except that the Entrez query was in the form of “txid of virus>[Organism:exp] NOT txid81077[ORGN]”, where “txid of virus” was “txid11080” for LEV, “txid11072” for JEV, “txid11084” for TBEV, “txid11083” for PV, “txid11089” for YFV, and “txid12637” for DENV. The incidence of a shared ligand sequence in these other flaviviruses was determined for both full-length and partial identity matches. This was done for partial matches by dividing the number of BLAST hits that had nine or more, but not full-length, consecutive amino acid matches to the ligand sequence by the total number of hits from the species for the ligand. Similarly, the incidence of full-length matches was computed by determining the number of hits with nine or more matches that were full length.

**RESULTS**

**Identification of HLA-restricted T-cell ligand sequences of the WNV proteome.** Six HLA transgenic (Tg), H-2-deficient mice, each expressing one of the predominant HLA class I (A2, A24, or B7) or class II (DR2, DR3, or DR4) alleles, were each immunized with 4 peptide pools (88 to 135 peptides per pool) that contained a total of 452 overlapping peptides, 12 to 20 amino acids each, comprising the entire WNV proteome. Each pool was mixed with 50 μl zymosan, 10 mg/ml, as the adjuvant. Splenocytes of the immunized mice, selectively depleted of murine CD4 and CD8 T cells for HLA class I or class II alleles, respectively, were analyzed by IFN-γ ELISPOT assays and a peptide matrix deconvolution approach (see Materials and Methods). A total of 137 peptides, ~30% of the 452 total, were found to activate T cells of one or more of the immunized Tg mice (Table 1; also see Table S1 in the supplemental material), the majority with high apparent avidity in the ELISPOT assays at peptide concentrations of 1.0 or 0.1 μg/mL, except for splenocytes of the B7 Tg mice (see Table S2). The efficacy of the peptide immunization approach was further established by comparison with naked DNA immunization of HLA-
DR2 Tg mice with a plasmid expression vector encoding the WNV NS3 protein. All but two (14/16: ~88%) of the peptide-specific T-cell responses following peptide immunization were also detected after DNA immunization (see Fig. S1 in the supplemental material).

Nearly one-third of the 137 WNV T-cell ligands were immunogenic in 2 or more Tg strains, both class I and class II, resulting in a total of 200 HLA-restricted T-cell ligand responses, 74 class I (40 A2, 24 A24, and 10 B7) and 126 class II (50 DR2, 38 DR3, and 38 DR4). T-cell responses to the same peptide by 2 or more Tg mice can be attributed to overlapping epitopes in the same peptide and/or to promiscuous epitope presentation by multiple HLA alleles (55). The allele-specific T-cell responses were lowest for HLA-B7 (~2%) and from ~5% to 11% for the remaining alleles. NS3 (~55%) and E (~37%) had the greatest fractions of HLA class I and class II ligand sequences, while C (~7%) had the least. About half of the ligand sequences were in clustered localizations of 3 or more overlapping ligand peptides each of the peptides overlapped with their adjacent peptides by 9 or more amino acids (see Table S1 in the supplemental material). This included all of the 6 prM ligand peptides forming a single cluster, 15 E peptides of 4 clusters, 3 NS1 peptides of 1 cluster, 21 NS3 peptides of 5 clusters, and 6 NS5 peptides of 2 clusters.

A correlation of the 137 HLA-restricted T-cell ligand sequences with reports of human T-cell epitopes (19–21, 32, 42, 61) showed that 22 of the ligand sequences matched 9 or more consecutive amino acids of 28 reported T-cell epitope peptides, both class I and class II (see Table S1 in the supplemental material). Additionally, the 137 ligand sequences were analyzed for candidate minimal HLA class I and II binding motifs within the 15 to 19 amino acids by use of NetMHC 3.2 and NetMHCIIpan 2.0 algorithms (26, 39). A total of 76 (38%) of the 200 positive HLA-restricted ELISpot responses corresponded to at least one predicted nonamer epitope with a concordant predicted HLA restriction (see Table S1). This relatively low extent of correspondence among the discovered HLA-restricted immunogenic peptides and predicted HLA binding motifs is comparable to those reported for other pathogens (27) and can be attributed, at least in part, to the fact that HLA binding prediction does not account for the pivotal role of the TCR in the immunogenicity of a peptide (67).

**Evolutionary conservation and variants of the WNV T-cell ligand sequences.** The identification of ligand sequences was conducted with the peptides of the NY99-flamingo 382-99 WNV strain. The incidence (% occurrence) of these ligands in the aligned sequences of reported WNV was analyzed. Further, entropy, a measure of overall diversity (see Materials and Methods), was determined for the positions in the alignment that contained the ligands. Five of the 137 T-cell ligands, of NS3 and NS5, were completely conserved in all of the analyzed WNV sequences, with a 100% incidence and zero entropy (see Table S3 in the supplemental material). The other 132 ligands had an incidence of 69 to 99% of the analyzed sequences (alignment entropies of <0.1 to 1.6), with the remaining 1 to 31% of the analyzed sequences containing one or more amino acids that differed from the ligands. Only seven of the ligands, with alignment entropies of 0.7 to 1.6, had an individual variant of incidence of >10% (~11 to 23%) of the analyzed WNV sequences (Table 2). Most of the variants of the ligand sequences were minor with negligible to low individual incidences (~1 to 9%) in the analyzed sequences (see Table S3). For example, about ~6% of the WNV sequences at the alignment position E 62 to 77 (entropy of 0.5) contained 16 intraspecies variants, each with a negligible incidence of ~1% or less (see Table S4).

### TABLE 1 Distribution of HLA-restricted ELISpot responses in WNV proteins

<table>
<thead>
<tr>
<th>WNV protein</th>
<th>Protein size (aa)</th>
<th>No. of overlapping peptides</th>
<th>No. (~%) of ELISpot-positive peptides</th>
<th>HLA-restricted T-cell activation</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>123</td>
<td>15</td>
<td>1 (7)</td>
<td>A2 A24 B7 DR2 DR3 DR4 Total</td>
</tr>
<tr>
<td>prM</td>
<td>167</td>
<td>21</td>
<td>6 (29)</td>
<td>0 0 1 0 1 6 4 4 17</td>
</tr>
<tr>
<td>E</td>
<td>501</td>
<td>67</td>
<td>25 (37)</td>
<td>8 6 2 8 12 6 42</td>
</tr>
<tr>
<td>NS1</td>
<td>352</td>
<td>46</td>
<td>10 (22)</td>
<td>2 1 0 1 3 5 12</td>
</tr>
<tr>
<td>NS2a</td>
<td>231</td>
<td>30</td>
<td>10 (33)</td>
<td>0 3 0 8 3 3 17</td>
</tr>
<tr>
<td>NS2b</td>
<td>131</td>
<td>16</td>
<td>6 (38)</td>
<td>1 3 1 2 0 2 9</td>
</tr>
<tr>
<td>NS3</td>
<td>619</td>
<td>84</td>
<td>46 (55)</td>
<td>26 0 1 16 11 4 58</td>
</tr>
<tr>
<td>NS4a</td>
<td>149</td>
<td>18</td>
<td>5 (28)</td>
<td>0 3 0 1 1 2 7</td>
</tr>
<tr>
<td>NS4b</td>
<td>255</td>
<td>33</td>
<td>9 (27)</td>
<td>0 1 0 3 3 5 12</td>
</tr>
<tr>
<td>NS5</td>
<td>905</td>
<td>122</td>
<td>19 (16)</td>
<td>2 5 4 5 2 7 25</td>
</tr>
<tr>
<td>Total</td>
<td>3,433</td>
<td>452</td>
<td>137 (30)</td>
<td>40 24 10 50 38 38 200</td>
</tr>
</tbody>
</table>

Percentage of positive ligand sequences per complete proteome array

<table>
<thead>
<tr>
<th></th>
<th>~9%</th>
<th>~5%</th>
<th>~2%</th>
<th>~11%</th>
<th>~8%</th>
<th>~8%</th>
</tr>
</thead>
</table>

* As identified by IFN-γ ELISpot assay in HLA Tg mice immunized with WNV overlapping peptides.

* Corresponding to the WNV flamingo strain [NCBI Entrez Protein Database accession number AAF20092.2]
among sequences of 26 other flaviviruses, most commonly E 99 to 113, NS5 598 to 615, and NS5 771 to 788, which individually were present in 11 to 14 other flaviviruses (Fig. 3). The majority of the shared ligands were in six major reported flaviviruses: the Japanese encephalitis group viruses (LEV and JEV), tick-borne encephalitis group viruses (TBEV and PV), YFV, and the four dengue virus serotypes (DENV1 to 4) (see Table S5 in the supplemental material). LEV and JEV, as expected, had the greatest number of partial and full-length identity matches, followed by DENV with mostly partial matches. Many of the full-length and partial matches among the flaviviruses were of high incidence, present in ~90 to 100% of the analyzed sequences, with low incidence matches mostly for DENV. Each shared ligand (either of full-length or partial identity) with an incidence less than 100% of the analyzed viral sequences had the remaining fraction, which lacked nine or more consecutive amino acid identity, highly similar to the shared ligand, with mostly one to two amino acid differences (data not shown). Both of the shared and highly similar sequences of the other flaviviruses are potential interspecies TCR altered peptide ligands of WNV, with one or more amino acid differences in the region corresponding to or flanking the WNV T-cell ligand sequence.

**DISCUSSION**

Sequence homologies among proteins of flaviviruses are expected because of the genetic relationships that established the phylogeny of these viruses. Nevertheless, these homologies have exceptional relevance when viewed in the context of the immunity of T-cell epitopes. WNV has particular pertinence because of its remarkable evolutionary conservation (17). Eighty-eight sequences of nine or more amino acids, representing 34% of the WNV proteome, were completely conserved in recorded WNV, and the remaining regions of the proteome generally had less than 10% of the sequences as variants of the predominant sequence at the position, hence the conserved nature of the majority of the 137 HLA-restricted WNV ligands and the low incidence of variants that differed from the ligand sequences. The high conservation of WNV is favorable for a satisfactory host immune response following natural infection or vaccination, because the low cumulative and negligible individual incidences of variant sequences provide little chance of immune escape by the proliferation of variants of the founder virus.

However, a negative aspect of the WNV evolutionary conservation is the extensive protein sequence homology with as many as 64 other viruses of the genus *Flavivirus*. Only 51 of the 137 ligand sequences were specific to WNV; the remaining 86 were shared with other flaviviruses, as sequences of either full-length or partial identity. The most frequently shared were the highly conserved sequences of E, NS1, NS3, and NS5, present in several major flaviviruses, including LEV, JEV, TBEV, PV, YFV, and DENV. Many of the shared sequences had high incidences, >50% of the analyzed sequences of one or more of the six major flaviviruses, particularly the closely related LEV and JEV. Further, shared sequences that did not occur in all (100%) of the analyzed sequences of the other flaviviruses had a highly similar remaining fraction, with mostly one to two amino acid differences. Both the shared and highly similar sequences of the other flaviviruses, with one or more amino acid differences corresponding to the WNV T-cell ligand and/or flanking sequences, represent potential interspecies TCR altered peptide ligands of WNV. Others (19, 72) have shown that interspecies variants of WNV T-cell ligands can be epitopes with the same HLA restriction, and Moran et al. (36) reported that T-cell clones from an acute dengue patient responded to epitope variants from other flaviviruses and also suggested that this may be of pathological significance, particularly in geographical regions where related viruses cocirculate.

This demonstration that a large number of WNV HLA-restricted T-cell ligand sequences shared 9 or more consecutive amino acid identities with multiple other flaviviruses serves to display the full extent of potential immune cross-reactivity among
Flavivirus proteins. It is clear that exposure to multiple flaviviruses by infection or immunization may risk immune responses to a large number of cross-reactive T cell epitopes that, as altered peptide ligands (APL), may significantly affect the immune responses to the pathogens. The concept of “original antigenic sin” in the role of cross-reactive immune responses was initially applied to serum antibody responses (39, 40) and later extended to T-cell responses (41, 42). It now has been extensively studied with respect to a variety of immune responses, including thymic education, T-cell memory, pathogen virulence, and autoimmune dis-
ease (43, 44, 45, 46, 47). In the context of T-cell responses, this concept relates to the role of cross-reactive T-cell epitopes resulting from dual exposure to viruses with sequence homology and epitope sequences that contain one or more amino acid substitutions at TCR contact residues as compared to the epitope of a previous pathogen or immunogen. Memory T cells selectively engaged by a variant epitope sequence may exhibit an impaired immune response, depending on the positions and types of amino acid substitutions surrounding or within T-cell epitope cores and the effect of these changes on the affinity of the interaction (48, 49). Protein structural analysis has provided evidence for extensive epitope interresidue interactions whereby an amino acid substitution at an HLA anchor site of epitope binding to a memory T cell can diminish or modify the TCR response to this epitope and block amplification of the immune response (13,48, 66). T-cell anergy may also result from partial activation, such as from impaired costimulation of T cells by antigen-presenting cells (APC) (57, 60). The effect in humans of APL inhibition or modification of T-cell immune responses has been widely recognized (14, 53, 58, 59). A recognized example is the immunopathology observed during secondary DENV infection, where misdirected immune responses of cross-reactive T cells are suggested to contribute to the severe dengue hemorrhagic fever (30, 34, 35, 52, 54, 69).

The selection of evolutionarily conserved protein sequences has widely been considered important to vaccine design in order to limit the selective loss of immunity resulting from mutation and protein modifications. However, as shown herein, sequences conserved in the evolution of viruses can be present in many different forms in viruses of related species. In this case, such conserved sequences could be recognized by the TCR of cells previously activated through infection or vaccination by a homologous sequence of another virus. Thus, we propose that the selection of conserved sequences that are also virus specific should have precedence in vaccine design.

ACKNOWLEDGMENTS

We are grateful to Lars Fugger (Weatherall Institute of Molecular Medicine, Oxford, United Kingdom) and Arthur Vanderbark (Oregon Health and Science University, Portland, OR), Chella S. David (Mayo Clinic, Rochester, MN), and Grete Sønderstrup (Stanford University School of Medicine) for providing HLA-DR2, -DR3, and -DR4 transgenic mice, respectively. Overlapping WNV peptide arrays were obtained through the NIH Biodefense and Emerging Infectious Disease Research Resources Repository, NIAID, NIH. We thank Paul Nordstrom August, Natascha May Thevasagayam, and Rashmi Sukumaran for their help with the manuscript.

No competing interests exist.

REFERENCES


