



# Deep Mutational Scanning Comprehensively Maps How Zika Envelope Protein Mutations Affect Viral Growth and Antibody Escape

Marion Sourisseau,<sup>a</sup>  Daniel J. P. Lawrence,<sup>b,c</sup> Megan C. Schwarz,<sup>a</sup> Carina H. Storrs,<sup>a</sup> Ethan C. Veit,<sup>a</sup>  Jesse D. Bloom,<sup>b,c,d</sup>  Matthew J. Evans<sup>a</sup>

<sup>a</sup>Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, New York, USA

<sup>b</sup>Basic Sciences Division and Computational Biology Program, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA

<sup>c</sup>Department of Genome Sciences, University of Washington, Seattle, Washington, USA

<sup>d</sup>Howard Hughes Medical Institute, Seattle, Washington, USA

**ABSTRACT** Functional constraints on viral proteins are often assessed by examining sequence conservation among natural strains, but this approach is relatively ineffective for Zika virus because all known sequences are highly similar. Here, we take an alternative approach to map functional constraints on Zika virus's envelope (E) protein by using deep mutational scanning to measure how all amino acid mutations to the E protein affect viral growth in cell culture. The resulting sequence-function map is consistent with existing knowledge about E protein structure and function but also provides insight into mutation-level constraints in many regions of the protein that have not been well characterized in prior functional work. In addition, we extend our approach to completely map how mutations affect viral neutralization by two monoclonal antibodies, thereby precisely defining their functional epitopes. Overall, our study provides a valuable resource for understanding the effects of mutations to this important viral protein and also offers a roadmap for future work to map functional and antigenic selection to Zika virus at high resolution.

**IMPORTANCE** Zika virus has recently been shown to be associated with severe birth defects. The virus's E protein mediates its ability to infect cells and is also the primary target of the antibodies that are elicited by natural infection and vaccines that are being developed against the virus. Therefore, determining the effects of mutations to this protein is important for understanding its function, its susceptibility to vaccine-mediated immunity, and its potential for future evolution. We completely mapped how amino acid mutations to the E protein affected the virus's ability to grow in cells in the laboratory and escape from several antibodies. The resulting maps relate changes in the E protein's sequence to changes in viral function and therefore provide a valuable complement to existing maps of the physical structure of the protein.

**KEYWORDS** Zika virus, glycoproteins, mutagenesis, neutralizing antibodies, virus entry

Zika virus (ZIKV) became the subject of intense interest after recent human outbreaks in the Yap Islands, French Polynesia, and Brazil (1–3) were associated with severe birth defects and neurological disease (4, 5). ZIKV is a member of the genus *Flavivirus* and is closely related to the dengue, West Nile, yellow fever, and Japanese encephalitis viruses (6). Like all viruses in this genus, ZIKV has an approximately 11.8-kb, capped, positive-sense, single-stranded RNA genome. This RNA comprises a single open reading frame that is translated into an approximately 3,432-amino-acid polyprotein that is

**Citation** Sourisseau M, Lawrence DJP, Schwarz MC, Storrs CH, Veit EC, Bloom JD, Evans MJ. 2019. Deep mutational scanning comprehensively maps how Zika envelope protein mutations affect viral growth and antibody escape. *J Virol* 93:e01291-19. <https://doi.org/10.1128/JVI.01291-19>.

**Editor** Julie K. Pfeiffer, University of Texas Southwestern Medical Center

**Copyright** © 2019 American Society for Microbiology. All Rights Reserved.

Address correspondence to Jesse D. Bloom, [jbloom@fredhutch.org](mailto:jbloom@fredhutch.org), or Matthew J. Evans, [matthew.evans@mssm.edu](mailto:matthew.evans@mssm.edu).

M.S. and D.J.P.L. contributed equally to this work.

**Received** 4 August 2019

**Accepted** 6 September 2019

**Accepted manuscript posted online** 11 September 2019

**Published** 13 November 2019

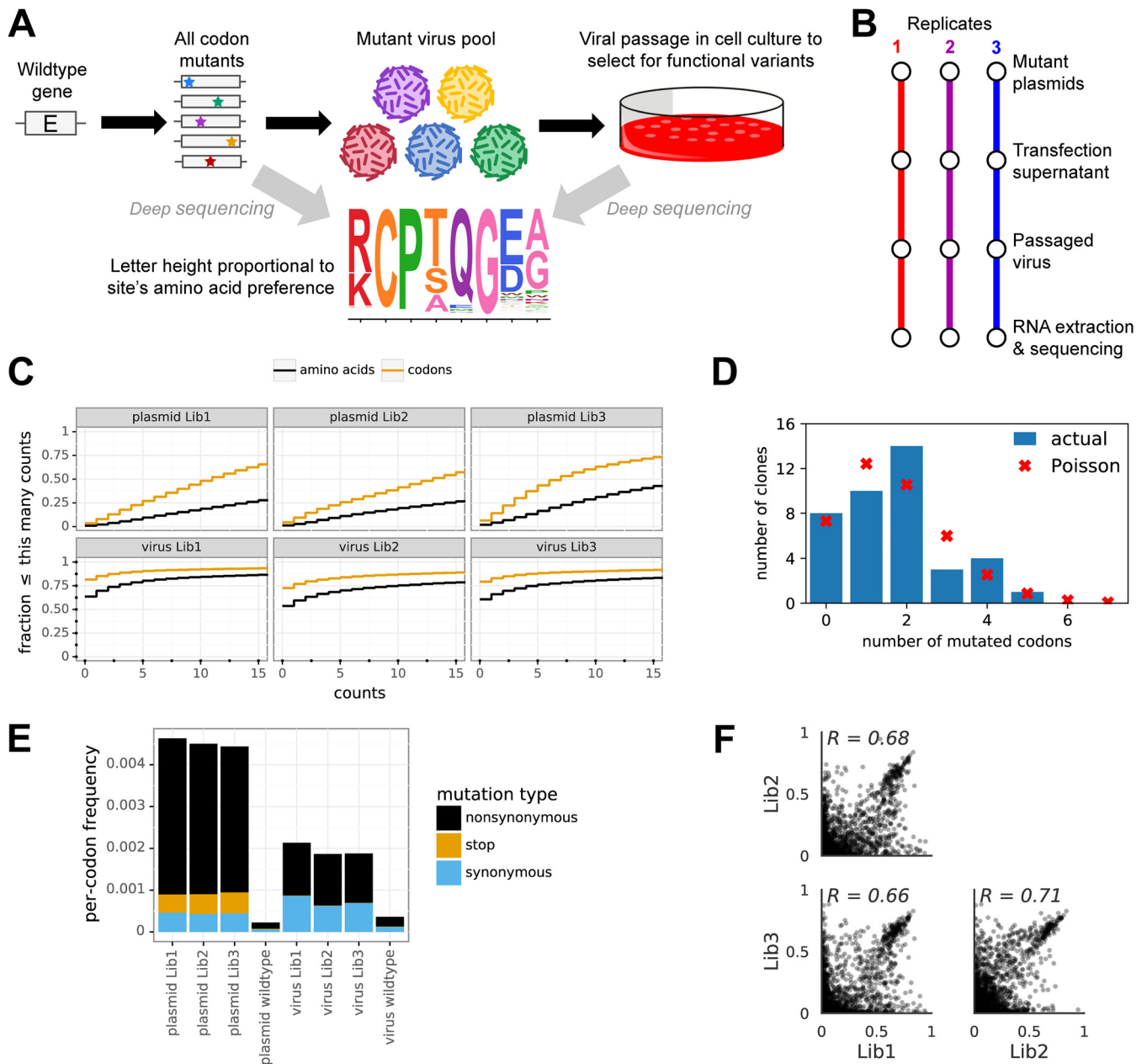
processed by host and viral proteases into three structural and seven nonstructural proteins. The mature infectious virion is comprised of a single copy of the RNA genome, surrounded by a capsid (C) protein shell and a lipid envelope bearing 180 copies each of the membrane (M) and envelope (E) proteins (7).

Immature ZIKV particles bearing 60 heterotrimeric premembrane (prM) and E protein spikes assemble at the endoplasmic reticulum and transit through the *trans*-Golgi network, where E proteins undergo rearrangement into antiparallel dimers arranged in a herringbone pattern characteristic of flaviviruses (7). This rearrangement exposes the prM cleavage site, which is cleaved by a furinlike protease. Upon release of the virion into the more neutral pH of the extracellular environment, the pr peptide dissociates from the M and E protein, resulting in the mature virion. The E protein in the mature virion is folded into three domains (8, 9). Domain I (DI) is comprised of a central  $\beta$ -barrel and a highly flexible loop that bears the sole E protein N-linked glycosylation site, N154. This glycan is highly conserved between most flaviviruses but is absent from some ZIKV isolates, particularly those of the African lineage (10, 11). Domain II (DII) forms an elongated fingerlike shape that is the major interaction surface between monomers of an E dimer. The fusion loop is located at the region of DII that is distal to domains I and III. The major feature of domain III (DIII) is an immunoglobulinlike region that is thought to contribute to host cell binding (12, 13). However, an E protein binding receptor has not yet been identified, and thus, the viral determinants for such interactions are not defined.

In order to maintain viral fitness, the E protein must retain the capacity to undergo the complex series of protein-protein interactions and conformational changes to form mature viral particles. Once these viral particles are released from a cell, the E protein needs to functionally interact with both insect and mammalian target cells in order to mediate viral binding and fusion. At the same time, the E protein is the primary target of neutralizing antibodies. As such, this protein is subjected to an array of evolutionary pressures. Any viral antigenic variation is expected to be due largely to mutations in the E protein. For many viral proteins, it is possible to identify functionally constrained motifs and sites with capacity for evolutionary change by examining conservation in alignments of natural viral variants. However, because ZIKV has only been carefully studied very recently, nearly all available sequences are from recent viruses. Furthermore, simply comparing viral sequences shows little difference across both older African and more recent Asian lineages, with even the most divergent protein sequences being >95% identical (14). Inspection of natural sequences is therefore not an effective method to identify constraint and selection of the E protein.

An alternative way to assess the mutational constraints on a protein is to use deep mutational scanning to measure the effects of all amino acid mutations of a protein on viral growth in cell culture and antigenic recognition by antibodies. Deep mutational scanning is a relatively new approach that involves creating large libraries of mutants of a gene, imposing a functional selection, and using deep sequencing to estimate the effect of every mutation from its change in frequency after selection (15). This approach has been used to assess the effects of all viral mutations on the function and antigenicity of key proteins from influenza virus (16–18) and HIV (19, 20). Two studies have applied deep mutational scanning to ZIKV (21, 22), but these studies did not systematically map the mutational tolerance or antigenicity of all amino acid mutations. Instead, they used libraries comprised of a subset of E protein variants to identify mutations that differentially affect viral growth in mammalian and insect cells.

Here, we applied deep mutational scanning to measure the effects of all amino acid mutations to the ZIKV E protein on viral growth in cell culture. The resulting sequence-function map identifies which sites in E protein are amenable to mutational change and which ones are under functional constraint. In addition, we completely map mutations that alter the antigenicity of E protein with respect to antibodies targeting two distinct regions. Overall, our results provide a map of the fine-grained constraints on the E protein and demonstrate a powerful platform for further functional and antigenic investigations of this protein.



**FIG 1** Deep mutational scanning of ZIKV E protein. (A) Overview of our experimental approach. We generated plasmid libraries with all codon mutations to the E protein gene and then used these plasmids to generate libraries of mutant viruses that we passaged in cell culture at an MOI of 0.01 infectious unit per cell. By using deep sequencing to measure the frequency of each mutation in the initial plasmid library versus its frequency in the passaged mutant viruses, we determined the amino acid preference of each site. (B) The deep mutational scanning was performed in full biological triplicate, with each replicate beginning with the independent generation of a mutant library. (C) Fractions of all amino acid or codon mutations observed at least the indicated number of times in the original plasmid libraries (Lib1, Lib2, and Lib3) and the viruses that have undergone functional selection. (D) Sanger sequencing of 40 clones in the plasmid library showed that the numbers of codon mutations per clone were roughly Poisson distributed. (E) Frequencies of each type of codon mutation averaged across the entire gene in the original plasmid libraries, the passaged virus libraries, and wild-type plasmid and virus controls. (F) The correlations between the amino acid preferences measured in each biological replicate. Each point represents a preference for a specific amino acid at a specific site, and the  $R$  values show the Pearson correlation coefficients.

(This article was submitted to an online preprint archive [23].)

## RESULTS

**Deep mutational scanning to measure the effects of all amino acid mutations to ZIKV E protein on viral growth in cell culture.** To measure the effects of all amino acid mutations to the ZIKV E protein, we followed the experimental scheme illustrated in Fig. 1A. First, we used a previously developed PCR-based technique (24, 25) to create

a library of all codon mutants of the full-length E protein-coding gene. We created this library in our previously reported infectious clone of the ZIKV prototype strain MR766 (26), choosing this clone because the plasmid is stable in bacteria (which facilitated the construction of the mutant library) and because the virus grows to high titers (which helped avoid bottlenecks during generation of the mutant virus library). We created three independent E protein mutant libraries, which were handled fully independently in all subsequent steps to provide true biological replicates (Fig. 1B). Each library replicate was comprised of  $>10^6$  unique plasmid transformants. There are 504 codons in the MR766 E gene, and thus, with 19 alternative amino acids, there are 9,576 possible amino acid mutations and 31,752 possible mutant codon sequences if all nucleotide positions in each codon are randomized. Deep sequencing of the plasmid library showed that virtually all of these mutations were observed multiple times in each replicate library (Fig. 1C). Our mutagenesis method is expected to introduce a Poisson number of codon mutations per variant (24, 25), and Sanger sequencing of 40 plasmids verified that the number of mutations per variant followed a roughly Poisson distribution with a mean of 1.7 (Fig. 1D), meaning that our libraries measure the average effect of each amino acid mutation alone and in several closely related backgrounds.

To generate mutant viruses from the mutant plasmids (Fig. 1A), we transfected 293T cells with the plasmid DNA libraries to produce pools of viruses with genomes encoding all the E protein mutants. Cells transfected with the mutant plasmid libraries produced titers of  $9 \times 10^4$  to  $5 \times 10^5$  infectious units per ml after 48 h, which was 240- to 1,280-fold less than the titers obtained with the wild-type ZIKV genome. To select for the functional variants in our virus libraries, we infected Vero cells with the transfection supernatants at a multiplicity of infection (MOI) of 0.01, using  $10^8$  cells to maintain the diversity of  $10^6$  independent E protein mutants. A low MOI was required to link the mutant viral genome with its respective virion, as virus produced from transfected 293T cells may have a genotype-phenotype mismatch (i.e., the E protein variants that compose the virion are likely to be mixtures and may not be the same as the one encoded on the packaged genome). Low-MOI treatment also imposed a selective barrier: viruses underwent multicycle replication in proportion to the functionality of their E protein. At 24 h postinfection, we washed the infected cells to remove input virus and then let the infection continue for two additional days. At this point, viral variants with high fitness would have undergone 3 or 4 rounds of additional infection and, thus, would be enriched in the population.

We then used deep sequencing to quantify the frequency of each mutation in the mutant viruses relative to its frequency in the initial plasmid mutant libraries. To sequence the passaged viruses, we extracted and reverse transcribed RNA from the infected Vero cells. In order to ensure high sequencing accuracy, we used a previously described barcoded-subamplicon sequencing approach (16, 17), which uses random nucleotide barcodes attached to template molecules to correct for sequencing errors. As shown by the results in Fig. 1E, the initial plasmid libraries had high rates of mutations, which included nonsynonymous, stop codon, and synonymous mutations. After viral passaging, the number of mutations decreased sharply. In particular, stop codons (which are expected to be uniformly deleterious) were purged, and nonsynonymous mutations (many of which will be deleterious) decreased in frequency. Overall, slightly less than half of all amino acid mutations were still observed in the mutant viruses after passaging (Fig. 1C). The results in Fig. 1E also show that the apparent rate of mutations in wild-type plasmid or virus generated from wild-type plasmid was low, indicating that sequencing, reverse-transcription, and viral-replication errors made only modest contributions to the mutations observed in the sequencing.

We used the sequencing data to measure the "preference" value of each site in E protein for each amino acid. These preferences represent enrichments of each amino acid at each site after selection for viral growth, normalized to the abundance of the wild-type codon at each site. The data in Fig. 1F show the correlations among all 10,080 amino acid preference values (there are 504 E protein codons multiplied by 20 unique amino acids for 10,080 preference values) between each biological replicate of the

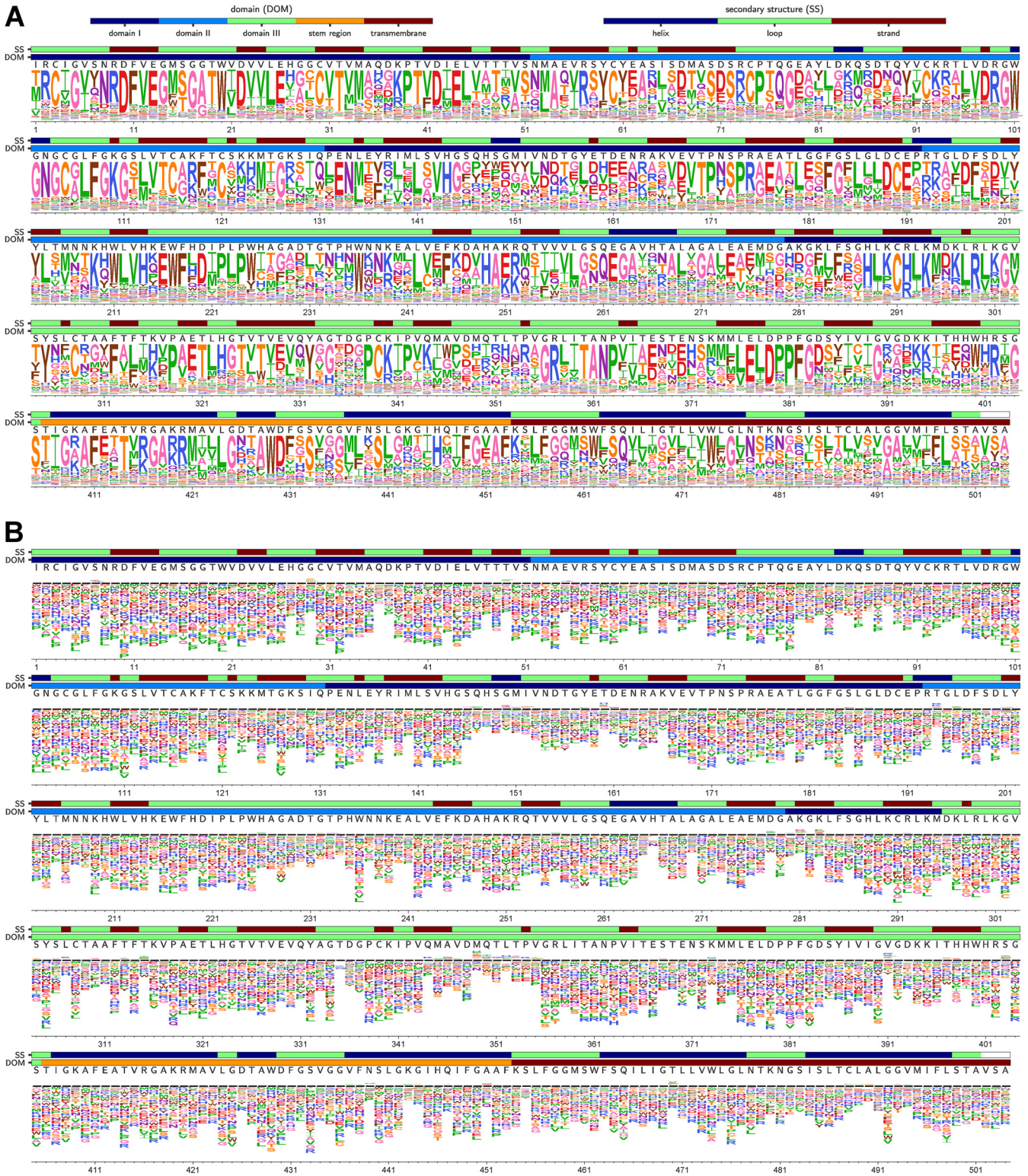
experiment. Although there is some noise, the preferences were strongly correlated among replicates. For the remainder of this paper, we use the average values of the three replicates. The across-replicate averages of the amino acid preferences for all sites in E are shown in Fig. 2A. Visual inspection of this figure reveals that the wild-type amino acid was usually, but not always, the most preferred amino acid at a site. It is also apparent that mutational tolerance varies widely across the E protein, with some sites strongly preferring a specific amino acid (e.g., site 10), while other sites tolerated many amino acids (e.g., site 36). The amino acid preference data in Fig. 2A can also be mathematically transformed into a measurement of the effect of each mutant amino acid relative to that of the wild-type identity at that site in MR766. These mutational effects are shown in Fig. 2B; as is apparent from inspection of this figure, mutations were usually deleterious (below the black line), but this was much truer for some sites than for others. For instance, most mutations were strongly deleterious at the mutationally intolerant site 10, whereas many mutations only had small effects at the mutationally tolerant site 36. Overall, these maps provide a wealth of information about the sequence-function relationships in E protein.

**Comparing mutational effects measured in the laboratory to amino acid frequencies in nature.** We next examined how the amino acid preferences measured in our screen compared to the frequencies of amino acids at each site in an alignment of naturally occurring ZIKV E proteins. Note that these natural sequences represent consensus sequences of the viral population in the hosts from which they are isolated and so do not account for minority variants that arise within a host. However, they do reflect those mutations that have reached high frequency within infected hosts. Among naturally occurring E proteins, only a single amino acid is observed at most sites among these natural sequences (Fig. 3A), with a small minority of sites having two different amino acids observed in naturally occurring sequences (e.g., site 120 in Fig. 3A). As a result, the correlation between our experimentally measured amino acid preferences and the amino acid frequencies in nature was modest, although still highly significant (Fig. 3B). In particular, it is obvious from the results shown in Fig. 3B that although our experiments measure a continuum of amino acid preferences spanning from 0 to about 0.8, the frequencies of amino acids among natural E protein sequences are usually either nearly zero or nearly one at each site.

There are two possible reasons why the natural amino acid frequencies are usually at the extremes of 0 or 1: either ZIKV E protein is intolerant of mutations in nature or natural evolution has not sampled many of the tolerable changes. The latter explanation seems plausible because the vast majority of ZIKV sequences were collected in just the last few years from closely related strains. To distinguish between these possibilities, we compared two different regions of E protein that, based on structural considerations and comparison to other flaviviruses, are thought to have widely different levels of functional constraint: the conserved fusion loop and the hypervariable glycan loop (27). Both regions exhibit very little variation among natural ZIKV sequences, but they differ greatly in the mutational tolerance measured in our deep mutational scanning (Fig. 3C). Our experiments suggest that most sites in the fusion loop strongly prefer the single amino acid that is most common in nature but that most sites in the hypervariable glycan loop can tolerate many amino acids. These results support the notion that the high conservation of E in nature is due at least in part to the limited sampling of evolutionary space among sequenced isolates and suggest that deep mutational scanning is indeed a useful alternative to measure the inherent mutational tolerance of each site. An important additional point, however, is that our experimental selections in Vero cells are almost certainly more permissive than natural selection in mosquito and primate hosts—this point is discussed in more detail in Discussion.

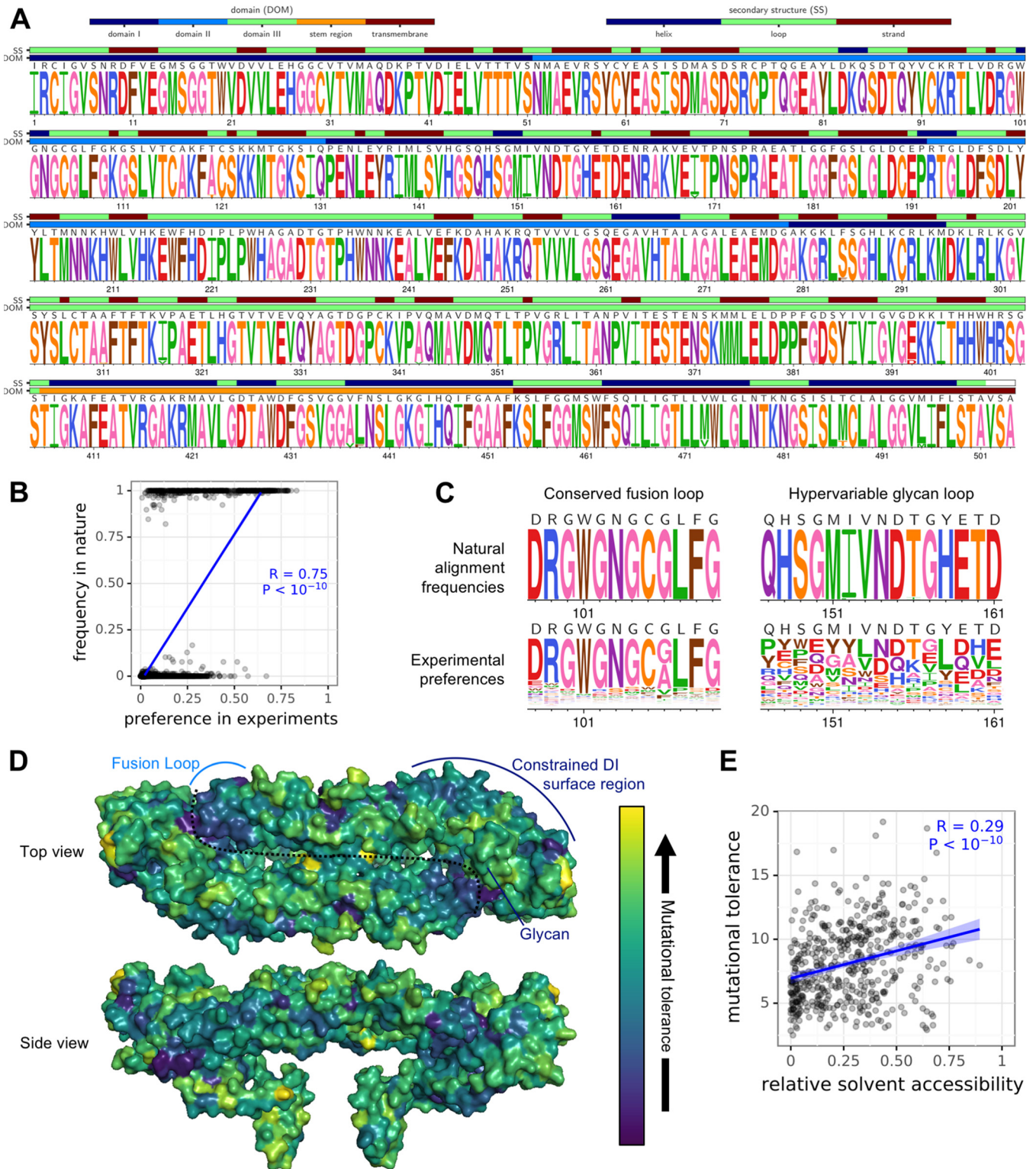
We next used the experimentally measured amino acid preferences to calculate the mutational tolerance of each site in the E protein and mapped these mutational tolerance values onto the protein's structure (Fig. 3D). Mutational tolerance varies widely across the protein. At the most general level, sites on the surface of the E protein tend to be more mutationally tolerant than sites in the core—a trend that has also been





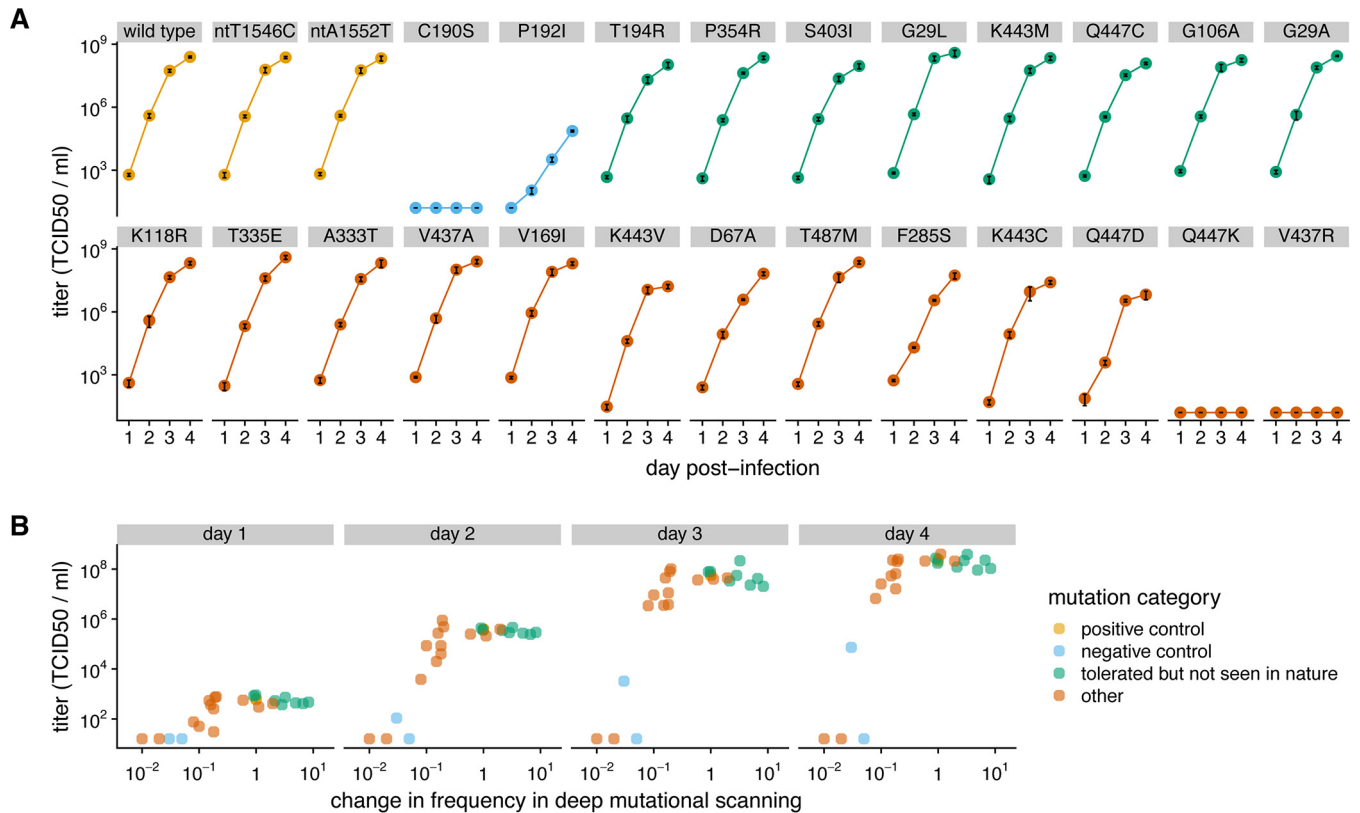
**FIG 2** Effects of all amino acid mutations at each site in the E protein as measured in the deep mutational scanning. (A) The preference of each site for each amino acid. The height of each letter is proportional to the enrichment of that amino acid after selection relative to its frequency before selection, so taller letters indicate more strongly preferred amino acids. The letters immediately above each logo stack indicate the wild-type amino acid in the parental MR766 E protein, the first color bar (DOM) indicates the domain of the E protein as defined in Dai et al. (27), and the second color bar (SS) indicates the secondary structure of that region as defined using DSSP (49) on PDB structure 5IRE (7). (B) An alternative representation of the same data shown in panel A. In this representation, the height of each letter represents the estimated effect of that mutation, where letters above the black line indicate favorable mutations (there are relatively few of these) and letters below the black line indicate unfavorable mutations. The effect of a mutant amino acid is equal to the logarithm of the ratio of its preference divided by the preference for the wild-type amino acid at that site.





**FIG 3** Comparison of deep mutational scanning measurements to variation in nature. (A) The frequencies of amino acids at each site in an alignment of naturally occurring E protein variants. The height of each letter is proportional to the frequency of that amino acid in the sequence alignment. (B) Correlation between amino acid preferences measured in the experiments and the frequencies of amino acids in the natural alignment. (C) Direct comparison of natural amino acid frequencies and experimentally measured amino acid preferences in the conserved fusion loop and hypervariable glycan loop as defined in Dai et al. (27). (D) Experimentally measured mutational tolerance mapped on the structure of E protein in PDB structure 5IRE (7). The mutational tolerance of each site is quantified as the Shannon entropy of the amino acid preferences shown in Fig. 2A, such that sites that tolerate many amino acids have a high mutational tolerance and sites that strongly prefer a single amino acid have a low mutational tolerance. This mutational tolerance metric was then used to color the structure of the E protein. The structure of an E protein dimer is shown, with each monomer separated by a dotted line. Key features are marked, including the glycan and highly constrained fusion loop and dimer-distal domain I (DI) region. (E) Correlation between the relative solvent accessibility and mutational tolerance

(Continued on next page)



**FIG 4** Growth curves of selected mutant viruses. (A) Vero cells were infected with indicated mutant viruses at an MOI of 0.01, and the titers of virus in the supernatant were determined at 1, 2, 3, and 4 days postinfection. The growth curves were performed in quadruplicate, with two biological replicates whose titers were determined twice, and the black lines inside the points show standard errors. Mutant categories, as described in Results, are indicated by color as shown in the key in panel B. (B) Relationships among the effect of each mutation in the deep mutational scanning and its titers in the growth curves grouped by day postinfection. The x axis represents the fold change in frequency during the deep mutational scanning, which is the ratio of the amino acid preference of the mutant amino acid relative to that of the wild type.

observed for many other proteins (28–30). To quantify this trend, we computed the correlation between a site's solvent accessibility and mutational tolerance and found that the correlation was modest but highly significant (Fig. 3E). Strikingly, nearly the entire outward-facing surface of the E protein demonstrated a high degree of mutational tolerance. Exceptions included the fusion loop region, which as pointed out above, tolerated nearly no changes, and the dimer-distal area of the DI region, which also appeared intolerant to mutation (Fig. 3D).

**Validation of deep mutational scanning measurements with growth curves of individual viral mutants.** To validate that the high-throughput deep mutational scanning measurements actually reflected mutational effects on viral growth, we generated a subset of E protein mutants and tested their growth individually. To do so, we cloned each mutation into the parental plasmid, produced homogenous stocks of these viruses by transfection of 293T cells, and then performed multicycle growth curves starting at an MOI of 0.01 infectious unit per cell in Vero cells as described in the legend to Fig. 4A (26, 31).

For these growth curves, we selected several groups of mutants. As positive controls, we used the wild-type virus and two synonymous mutants whose mutations are not expected to affect viral fitness, and all of these viruses grew well (Fig. 4A). As

### FIG 3 Legend (Continued)

of each site. Relative solvent accessibility is computed with DSSP (49) using a single monomer of PDB structure [5IRE](#), normalizing to the absolute solvent accessibilities in Tien et al. (50). Mutational tolerance is computed as the number of effective amino acids, which is the exponential of the Shannon entropy of the preferences.



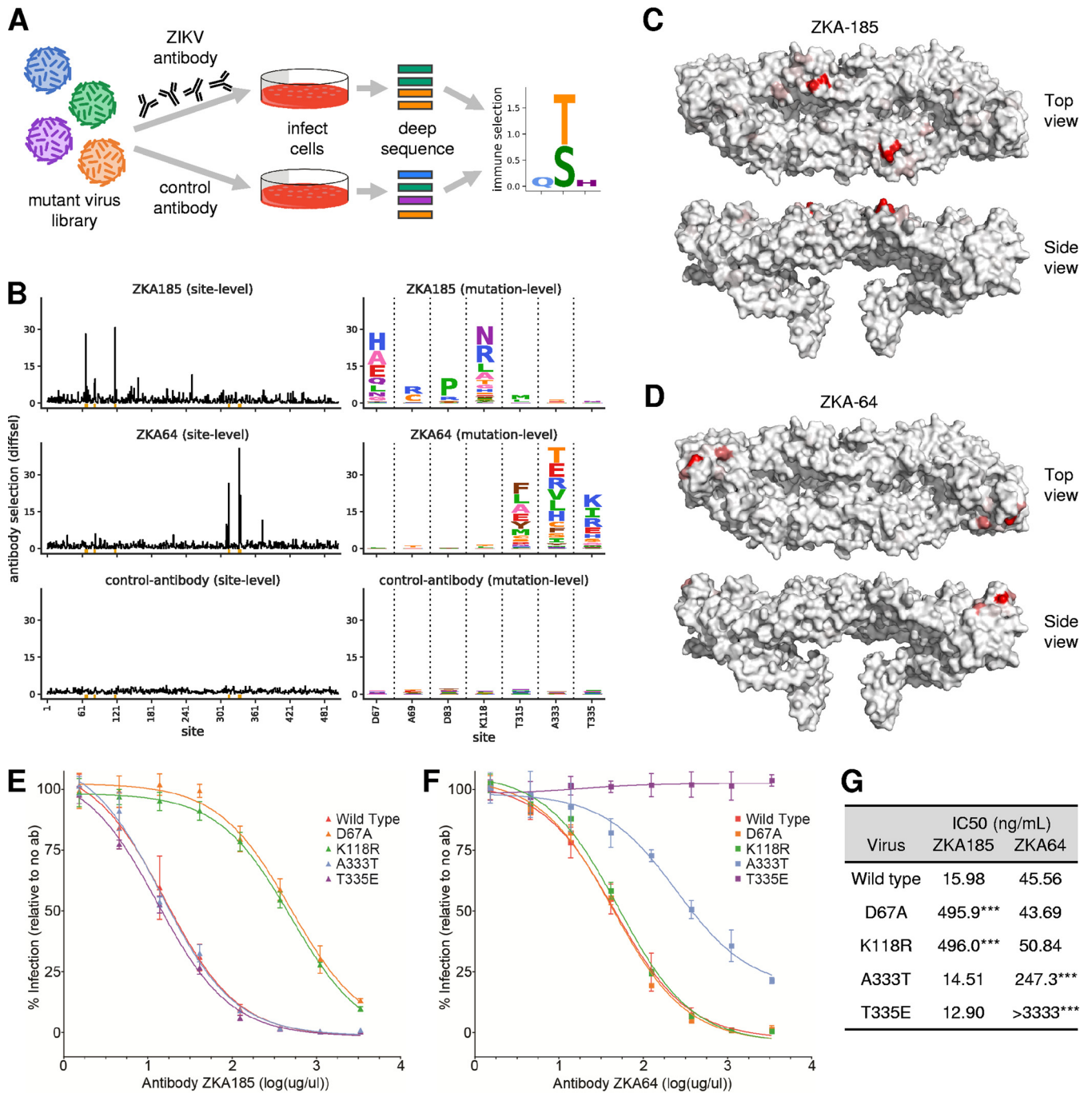
negative controls, we selected a C190S mutation (bearing a change of C to S at position 190), as this cysteine residue is absolutely conserved in all flavivirus sequences, and a P192I mutation, as this residue is a proline in all but the most divergent tick-borne encephalitis and yellow fever viruses, where it is a valine. Indeed, C190S produced no detectable virus, and the growth of P192I was highly impaired (Fig. 4A). We also selected eight mutations that our deep mutational scanning suggests should be well tolerated but that are not observed among natural ZIKV E protein sequences. All eight of these mutants grew well (Fig. 4A), supporting the notion that the deep mutational scanning can identify well-tolerated mutations that are not observed among natural sequences. We generated 13 additional mutants that were expected to (and did) have a range of effects on viral growth (Fig. 4A).

Overall, these growth curves showed a good relationship between the measurements in the deep mutational scanning and the growth of the individual mutants in Vero cells (Fig. 4B). Specifically, mutations that decreased by >10-fold in the deep mutational scanning strongly impaired viral growth, whereas mutations that had wild-type or better scores in the deep mutational scanning always grew well in the growth curves, even if these mutations are not observed among natural sequences. However, mutations that appeared to be just mildly attenuated in the deep mutational scanning (selected at a frequency within 10-fold of the wild-type frequency) often grew about as well as the wild type, suggesting that there was some noise in the high-throughput measurements and caution should be used in calling mutations deleterious if they were only mildly depleted in the high-throughput experiments.

**Mapping mutations that escape neutralizing antibodies.** Deep mutational scanning can be extended to completely map viral mutations that escape antibody neutralization, in a technique termed mutational antigenic profiling (Fig. 5A presents an outline of this approach) (18, 25). To demonstrate that our MR766 libraries can be used in this manner, we selected our libraries with two previously described patient-derived antibodies that efficiently inhibit ZIKV infection. One antibody, ZKA64, was shown to bind a recombinant form of domain III of the E protein, while the other, ZKA185, was originally described as a neutralizing, nonbinding antibody because it did not interact with purified full-length E protein or the E DIII domain (14).

We began with the virus libraries that had been passaged at a low MOI to ensure a link between genotype and phenotype. We incubated these mutant virus populations for 1 h with a concentration of antibody that neutralized roughly 99.9% of the infectious titer of wild-type virus (99.9% inhibitory concentration [ $IC_{99.9}$ ]). This quantity of antibody leads to a strong signal under the assumption that about 0.5% to 1% of all possible amino acid mutations mediate escape (this is expected if the typical antibody epitope contains 2 to 8 residues and about half the amino acid substitutions at these sites mediate escape). We then added this antibody-virus mixture to Vero cells and incubated them for 24 h. We next washed the cultures to remove unbound input virus, extracted total RNA from infected cells, and reverse transcribed and amplified the E protein gene for deep sequencing. We simultaneously performed a control mock neutralization with the hepatitis C virus (HCV)-neutralizing AR3A monoclonal antibody (42), which has no impact on ZIKV infection. Quantification of the viral RNA remaining after neutralization by reverse transcription-quantitative PCR (qRT-PCR) indicated that 1.5% to 2.6% of the mutant virus pool escaped antibody neutralization (data not shown).

Each ZIKV E antibody strongly selected mutations at just a few sites in the E protein, whereas the control antibody did not select any mutations (Fig. 5B). The different anti-ZIKV antibodies selected distinct mutations, suggesting that the mutations directly impact the antibody epitope rather than altering E protein conformation to generally increase neutralization sensitivity. We mapped the sites of strongly selected mutations onto the structure of the E protein and found that the most strongly selected mutations clustered near each other (Fig. 5C). The ZKA64-selected mutations were all found in the DIII domain of the E protein, as previously predicted (14). One of these residues, T315,



**FIG 5** Mutational antigenic profiling to map antibody escape mutations. (A) Schematic of the workflow. The mutant virus libraries were incubated with ZIKV neutralizing antibodies or an irrelevant antibody and infected into cells, and the resulting viral RNAs were deep sequenced to identify which mutations survive immune selection. The data can be analyzed to compute the differential selection for each mutation, which is the logarithm of its enrichment in the immune selection relative to that in a mock-selection control. Hypothetical results are shown, with larger letters indicating more strongly selected mutations. (B) The overall selection at each site (line plots) and the selection for specific mutations at strongly selected sites (logo plots) for two anti-ZIKV antibodies, ZKA185 and ZKA64, as well as the negative-control antibody that does not target ZIKV. diffsel, differential selection. (C, D) The overall strength of immune selection at each site mapped on the structure of the E protein, with red coloring indicating residues with differential mutant selection, using the indicated antibodies. (E, F) Neutralization curves validated the idea that strongly selected mutations indeed greatly reduced viral neutralization for each antibody (ab). Error bars show standard errors. (G) IC<sub>50</sub>s computed from the neutralization curves. The value of >3,333 indicates that 50% inhibition was not observed at the highest concentration of antibody tested. \*\*\*, *P* < 0.001 by one-way analysis of variance (ANOVA) compared to the results for the wild type.

is in the ABCD sheet epitope, and the A333 and T335 residues are in the lateral ridge epitope. Conversely, ZKA185-selected mutations all occurred at sites in the DII E protein domain. One of these sites, D67, was recently shown (32) to be a location for a single ZKA185 resistance mutation, D67E, which our results confirm.

To validate the antibody escape mutants, we chose two mutations expected to strongly affect sensitivity to each anti-ZIKV antibody (A333T and T335E for ZKA64, and D67A and K118R for ZKA185). We performed neutralization assays against viruses engineered to contain these mutations, using a previously described approach (14, 22, 33). As expected from the mutational antigenic profiling, each mutation greatly reduced neutralization sensitivity to the antibody that selected it, but not to the other antibody (Fig. 5E to G). These results show that mutational antigenic profiling with our viral libraries can accurately map antibody escape mutations.

## DISCUSSION

We have completely mapped the effects of all amino acid mutations to the ZIKV E protein on viral growth in cell culture. The resulting sequence-function map can be used to assess the probable effect of every possible amino acid mutation—something that is valuable both for understanding protein function and for assessing the probable impact of mutations observed in natural sequences.

Some aspects of this sequence-function map can be interpreted in light of what is already known about E protein's structure and function. For instance, disulfide bonds often play an important role in stabilizing proteins. Consistent with the structural importance of these bonds, our results show that the first 12 cysteines in the E protein, all of which form disulfide bonds (7), are highly intolerant of change to any other amino acid. In contrast, the last cysteine in the protein's sequence (C488) is not involved in a disulfide bond, and this site is highly tolerant of mutations. Similarly, ionization of histidine residues is important for the low-pH conformational change in the E protein that helps mediate membrane fusion. There are 16 histidines in the ZIKV MR766 E protein, and our map shows that five of them were strongly selected during the deep mutational scanning (H144, H214, H249, H288, and H446). These five histidines are all ones that studies of other flavivirus E proteins have suggested are important for fusion: mutating H249 and H288 led to a loss of infectivity in yellow fever virus (34), while West Nile virus was strongly impaired by mutating H144 and H446 and more moderately impaired by mutating H214 and H446 (35). We can also rationalize some of the sequence-function map in more general terms. For instance, sites on the surface of the E protein tend to be more mutationally tolerant than sites in the core, a trend that has also been observed for many other proteins (28, 29) and presumably reflects the increased constraints on residue packing in the interior of a protein. In addition, certain functionally important regions of the protein are relatively mutationally intolerant despite being on the core. These regions include the fusion loop and a surface region extending from near the end of domain I to the beginning of domain III that has been implicated as a flavivirus receptor-binding site (12, 13).

But the greatest value of our sequence-function map is not that it confirms the expected selection at the few E protein residues known to be important for some specific biological function but, rather, that it provides a means to assess the likely impact of the vast numbers of mutations for which there is not prior functional information. The relationship between structure and function is complex, and furthermore, mutations to a protein can affect a myriad of E protein traits, ranging from translation to folding to assembly to entry. Using our map, we can make observations about mutational effects that are virtually impossible to rationalize from current molecular knowledge: for instance, we can hypothesize no clear explanation of the constraints at site 447 that explain why the virus is completely tolerant of Q447C, is substantially attenuated by Q447D, and is completely inactivated by Q447K, and yet, both our map and our validation experiments show that this is the case. Because the effects of many other mutations are similarly difficult to rationalize, our map forms a useful reference for studies of E protein evolution and function.

The key to the technical success of our experiments was the use of a highly efficient plasmid-based reverse-genetics system to generate the mutant virus library (26). Flavivirus genomes are notoriously difficult to handle in traditional plasmid-based cloning systems, since they often contain nucleotide sequences that are toxic in bacteria. The



plasmid used in our experiment has an intron inserted into the most toxic region, which eliminates the bacterial toxicity but is spliced in mammalian cells and so does not affect the actual generated virus (26). By using this plasmid and also choosing a viral strain (MR766) that grows to high titers in cell culture, we were able to generate viral libraries that effectively sampled all amino acid mutations to the E protein. This library diversity contrasts with those of two other recent ZIKV deep mutational scanning studies (21, 22): both of these studies revealed important biological insights about host adaptation, but neither produced a complete map of amino acid-level selection, due to an inability to generate and maintain the viral library diversity to sample all amino acids. Of course, our experiments were still subject to noise, probably because even our high-efficiency reverse genetics had some bottlenecks of viral diversity. Our validation experiments with single-mutant growth curves suggested that mutations that are strongly depleted in the deep mutational scanning (decrease of  $>10$ -fold, equivalent to a mutational effect of  $<\log_2 0.1 = -3.3$  in the raw data provided in Data Set S3 in the supplemental material) are consistently deleterious for viral growth. Likewise, mutations whose preference the deep mutational scanning suggests is comparable to that of the wild type are consistently well tolerated. However, caution should be used in interpreting the effects of mutations that the deep mutational scanning suggests are just mildly deleterious, since these modest effects fall on the boundary between the signal and noise of the high-throughput measurements.

A major point to keep in mind is that our study mapped how mutations affect the growth of a single strain of ZIKV in Vero cells, meaning that there are some major caveats that must be kept in mind when interpreting the results. First, the MR766 ZIKV strain we used was isolated in 1947 and was subsequently passaged in the laboratory (10). Therefore, the effects of some mutations could differ between this strain and more contemporary ZIKV strains. Indeed, prior deep mutational scanning studies of other viruses have shown that mutational effects do shift somewhat among different strains of the same virus (19, 36, 37), although these shifts tend to be relatively small when the strains are closely related, as is the case for all known ZIKV strains. Second, Vero cells are a relatively permissive monkey-derived cell line, whereas in actuality, ZIKV evolves under selection to efficiently replicate in humans and mosquitoes. Indeed, prior deep mutational scanning studies of ZIKV (21, 22) and other viruses (38, 39) have identified numerous mutations with effects that depend on the host species and level of innate immune function in the cells used to grow the virus. In addition, we have analyzed selection at the level of amino acids, but at some sites, there is also likely to be selection on specific codons due to their effect on RNA structure or translation. Therefore, our map should be interpreted as providing a baseline estimate of the functional effects of mutations to the E protein in a relatively permissive context.

However, the fact that the cell culture model used in our experiment is somewhat artificial represents an opportunity as well as a caveat. In particular, it provides the capacity to directly manipulate the selective pressure to map how mutations affect some particular relevant process. In our case, we have used this fine-grained level of control to perform the first complete mapping of ZIKV antibody escape mutants by selecting the virus library in the presence and absence of monoclonal antibodies. This work is of special interest because the conformational flexibility of the E protein means that antibody neutralization mechanisms can be complex (40), with some neutralizing antibodies not even binding to recombinant E protein (14).

We mapped all viral mutations that affected neutralization by two antibodies, ZKA64 and ZKA185 (14). Both antibodies strongly selected mutations at just a few spatially proximal sites, thereby defining the functional epitopes. For ZKA64, this functional epitope was in domain III of the E protein, consistent with the original report that it bound to this domain (14). The ZKA185 antibody does not bind to recombinant E protein, so the original paper describing this antibody was unable to map its epitope (14), although later work found a single neutralization-resistant variant with mutations at sites 67 and 84 in domain II of the E protein (32). Consistent with this later study, we found that ZKA185 exerted strong selection at a few sites in domain II, including site 67

and, to a much lesser degree, site 84. Notably, the sites selected by ZKA185 are well separated on the linear E sequence but tightly grouped on the protein structure. Overall, we think the ability to use our libraries to completely map ZIKV antibody escape mutations could prove very useful. Antibody escape mutants are traditionally isolated by passaging virus in the presence of neutralizing antibody, but each such experiment typically only isolates one or a few of the possible escape mutants. In contrast, our experiments completely map all amino acid mutations that affect neutralization and, thus, provide a full picture of the functional epitope.

Overall, our work provides a complete map of how E protein amino acid mutations affect viral growth in a permissive cell line and viral escape from two monoclonal antibodies. This map in and of itself is a valuable complement to structural maps of the E protein and will be useful for interpreting and designing mutagenesis studies to probe the function and antigenicity of the protein. In addition, our work provides a template for future studies to map mutations that affect more specific viral phenotypes, such the ability to replicate in specific cell lines or animals/mosquitos or to escape from antigenic recognition by other antibodies or even polyclonal antisera.

## MATERIALS AND METHODS

**Cell lines and culture.** 293T cells (provided by Charles M. Rice, Rockefeller University) and Vero cells (ATCC CCL-81; ATCC, Manassas, VA) were grown in Dulbecco's modified Eagle's medium (DMEM) (Gibco BRL; Life Technologies, Gaithersburg, MD) with 10% fetal bovine serum (FBS) (Gibco BRL; Life Technologies).

**Antibodies.** ZKA185 and ZKA64 are human monoclonal neutralizing antibodies targeting ZIKV E protein (Ab00835-15.0 and Ab00779-15.0, respectively; Absolute Antibody, Oxford, UK) (14). The humanized monoclonal antibody D1-4G2-4-15 (4G2) (Absolute Antibody, Oxford, UK) is a broadly reactive flavivirus antibody that binds to an epitope at the fusion loop domain of the E protein (41). Rabbit anti-NS3 antibody was raised against a peptide consisting of amino acid residues 456 to 469 of the ZIKV MR766 sequence (GenBank accession number [AAV34151](#)). The anti-HCV E2 monoclonal antibody (clone AR3A) used as a control for the mutational antigenic profiling was provided by Mansun Law (Scripps Research Institute) (42).

**Creation of mutant plasmid libraries.** The mutant plasmid libraries were cloned into our previously described single-plasmid reverse genetics system for ZIKV strain MR766 (sequence is available under GenBank accession number [KX830961](#)) (26). In order to enable efficient cloning into this plasmid backbone, we designed a recipient plasmid in which a 1.65-kb region of this plasmid, including the entire E protein coding sequence, was replaced by enhanced green fluorescent protein (eGFP) flanked by NotI sites and named this plasmid ZIKV\_MR766\_int\_GFP ("int" stands for the fact that the clone contains an intron to reduce toxicity in bacteria [26]). The sequence of ZIKV\_MR766\_int\_GFP in GenBank format is at [https://github.com/jbloomb/lab/ZIKV\\_DMS\\_with\\_EvansLab/blob/master/data/1726\\_ZIKV\\_MR766\\_int\\_GFP.gb](https://github.com/jbloomb/lab/ZIKV_DMS_with_EvansLab/blob/master/data/1726_ZIKV_MR766_int_GFP.gb).

We created codon mutant libraries of the E gene using a previously described PCR mutagenesis approach (24). We first designed mutagenic primers that tiled all codons in the E gene. These primers were designed, using the computer scripts at <https://github.com/jbloomb/lab/CodonTilingPrimers>, to have an average melting temperature of ~65°C, as described in reference 25. The primer sequences are in Data Set S1 in the supplemental material. We performed the mutagenesis PCR using these primers and the following end primers flanking the E gene: 5'-GCCATTGCCTGGCTTTGGGAAGC-3' and 5'-TGGTACTTGTACCGGTCCCTCCAGGC-3'. Two rounds of mutagenic PCR were performed in triplicate exactly as described in reference 24. After NotI digestion and purification of the recipient plasmid backbone ZIKV\_MR766\_int\_GFP, mutagenized E gene amplicons were cloned into the recipient MR766 genome using Gibson assembly and transformed into 10-beta electrocompetent *Escherichia coli* cells (NEB). Transformants were plated on LB medium supplemented with carbenicillin. Each of the triplicate libraries contained  $>1 \times 10^6$  unique transformants as estimated by plating a dilution of the library transformations. Fifteen individual colonies of each library were Sanger sequenced, and the results are summarized in Fig. 1. Each codon mutant library was prepared from a pool of plated transformants resuspended in LB medium by using the HiSpeed maxiprep kit (Qiagen). In order to help ensure greater plasmid stability, the bacterial growth steps were performed at 30°C.

**Cloning single mutant ZIKV.** Individual amino acid mutants were generated by PCR, using previously described techniques (26, 43). Briefly, two rounds of PCR were performed. The first round involved two reactions: one reaction with a forward oligonucleotide (sequences available upon request) that overlaps a unique restriction site in the ZIKV strain MR766 plasmid (26) and a reverse primer that encodes the desired mutation, and the second reaction with a forward oligonucleotide also bearing the desired mutation and a universal reverse oligonucleotide that spans another unique restriction site. These products were then combined as the template in a third PCR with only the outside universal oligonucleotides. This product was cloned in the parental plasmid using the In-Fusion cloning kit (Clontech, Mountain View, CA). All PCR-amplified sequences, cloning sites, and desired mutations were verified by Sanger sequencing.

**Rescue of ZIKV E libraries or WT or single mutant ZIKV.** Viruses and ZIKV E libraries were produced by transfecting plasmids into 293T cells as previously reported (26, 33). Briefly, prior to transfection, 6-well plates and 24-well plates were seeded overnight with  $1.2 \times 10^6$  or  $0.3 \times 10^6$  cells/well, respectively. For each library replicate, 6 wells of a 6-well plate were transfected using 1  $\mu$ g of plasmid DNA per well and *TransIT-LT1* transfection reagent (Mirus Bio, Madison WI) according to the manufacturer's recommendations. The viral supernatants were collected at day 2 posttransfection, pooled, filtered (pore size, 0.45  $\mu$ m) to remove cellular debris, and stored at  $-80^\circ\text{C}$ . For each defined mutant virus, 1 well of a 24-well plate was transfected using 1  $\mu$ g of plasmid DNA and *TransIT-LT1* transfection reagent (Mirus Bio, Madison WI) according to the manufacturer's recommendations. The viral supernatants were collected at day 2 posttransfection, filtered (pore size, 0.45  $\mu$ m), and stored at  $-80^\circ\text{C}$ .

**Titration of infectious viruses and ZIKV E libraries.** For screening, the infectious titers of rescued wild-type virus, as a positive control, and mutant libraries were determined by cytopathic effect (CPE)-based limiting dilution assay on Vero cells, based on previously described techniques (26, 33). One day prior to infection, 96-well plates were seeded at a density of  $10^4$  cells/well. They were infected with 100  $\mu$ l of virus serially diluted in DMEM with 2% FBS (8 wells per dilution). Infection was scored by detection of ZIKV-induced CPE, and 50% tissue culture infective dose (TCID<sub>50</sub>) units were calculated according to the method of Reed and Muench (44). The titers of individually rescued mutants were determined by immunostaining with the pan-flavivirus E protein-reactive 4G2 antibody and flow cytometry, as previously described (33). Briefly, 1 day prior to infection, 24-well plates were seeded at a density of  $1.25 \times 10^5$  cells/well. The next day, 250  $\mu$ l of serially diluted transfection supernatants in DMEM with 2% FBS was added to each well (in triplicate). Staining was performed at 24 h postinfection to avoid any spread. Only the viral dilutions leading to 30% to 1% infected cells were used to calculate the infectious titers. Infectious units per ml were calculated using the following formula: percentage of infectious events  $\times$  number of cells/volume (ml) of viral inoculum.

**Selection of ZIKV E libraries and control WT ZIKV.** Ten 15-cm-diameter petri dishes per replicate or WT control with  $10^7$  cells on the day of infection were infected at an MOI of 0.01 infectious unit per cell. The viral supernatant was collected at days 2 and 3 postinfection, filtered (pore size, 0.45  $\mu$ m) and stored at  $-80^\circ\text{C}$ . Infected cells were collected at day 3 postinfection, and total RNA was extracted using the RNeasy plus kit (Qiagen, Valencia, CA) according to the manufacturer's recommendations.

**Growth curves.** Vero cells ( $1 \times 10^5$ ) were seeded in a well of a 24-well plate 1 day prior to infection, in triplicate, and then infected at an MOI of 0.01 infectious unit per cell in DMEM with 2% FBS. Virus was removed 6 h later, and cells were washed twice with phosphate-buffered saline (PBS) before 1 ml of DMEM with 2% FBS was added. Supernatants were collected and filtered (pore size, 0.45  $\mu$ m) daily for the next 4 days. Supernatant infectivity was determined by TCID<sub>50</sub> assay as described above.

**Measurement of the IC<sub>99.9</sub>.** Twenty-four-well plates were seeded 1 day prior to infection at a density of  $1.25 \times 10^5$  cells/well in 500  $\mu$ l DMEM supplemented with 2% FBS. Twofold serial dilutions of antibody were incubated at  $37^\circ\text{C}$  for 1 h in 500  $\mu$ l medium with  $2.5 \times 10^5$  infectious units of control WT virus that had been selected on Vero cells, in triplicate. The culture medium on the above-mentioned Vero cells was then replaced with the virus-antibody dilutions for 24 h. Cells were collected after extensive washes with PBS. Total RNA was extracted with the RNeasy plus kit (Qiagen). Viral RNA quantification was performed on 20 ng of RNA by ZIKV-specific qRT-PCR, as described below. The IC<sub>99.9</sub> was calculated for each antibody using Prism software (GraphPad Software) (dose-response, three parameters). For ZKA64 and ZKA185, the IC<sub>99.9</sub> values were 531 and 274 ng/ml, respectively.

**Selection of antibody escape mutants.** One day prior to infection, 6-well plates were seeded at a density of  $5 \times 10^5$  cells/well in 2 ml DMEM supplemented with 2% FBS. Both ZKA64 and ZKA185, at the IC<sub>99.9</sub> concentrations stated above, and AR3A, at 1  $\mu$ g/ml, were incubated at  $37^\circ\text{C}$  for 1 h with  $10^6$  infectious units of ZIKV E selected libraries or control WT virus in 2 ml of DMEM with 2% FBS. Each antibody mixture was then added to the cells. At 24 h postinfection, cells were extensively washed to remove the inoculum, and RNA was extracted from the cells using the RNeasy plus kit (Qiagen) for deep sequencing.

**Neutralization assays.** One day prior to infection, 24-well plates were seeded with  $1.25 \times 10^5$  Vero cells per well. Threefold serial dilutions of antibodies were incubated with  $2.5 \times 10^4$  (MOI of 0.1) infectious units of virus at  $37^\circ\text{C}$  for 30 min and then used to infect Vero cells. At 24 h postinfection, cells were collected and analyzed by flow cytometry, as described above, to determine the percentage of infected cells. The percentage of infected cells in each well was normalized relative to the percentage of infected cells in the absence of antibody. Analysis of dose-response curves was performed with Prism software (GraphPad Software) (dose-response, three parameters) to calculate the apparent IC<sub>50</sub> for each antibody.

**ZIKV-specific qRT-PCR.** Viral RNA quantification was performed by ZIKV-specific qRT-PCR, as previously described (26). Total RNA was extracted with the RNeasy plus kit (Qiagen, Valencia, CA) according to the manufacturer's recommendations. For quantifying ZIKV RNA, 20 ng of total RNA was reverse transcribed *in vitro* using a high-capacity cDNA reverse transcription kit (Thermo Fisher Scientific, Waltham MA). Quantitative PCR was performed with ZIKV-specific primers (5'-TTGGTCATGACTGCTG ATTGC and 5'-CCYTCCACRAAGTCYCTATTGC) and the LightCycler 480 SYBR green I master mix (Roche Applied Science) using the LightCycler 480 II real-time PCR system (Roche Applied Science). Sequences of the primers targeting ZIKV were adapted from previously published sequences (45). Quantification of ZIKV RNA copies per microgram of total RNA was performed against a standard curve of *in vitro*-transcribed MR766 ZIKV RNA using the T7 RiboMAX express large-scale RNA production system (Promega).



**Flow cytometry.** Cells were fixed in 4% paraformaldehyde and stained with 4G2 (E protein) or NS3 antibody and goat-anti-human or donkey anti-rabbit secondary antibody, respectively, conjugated to Alexa Fluor 647 (Thermo Fisher Scientific, Waltham MA), using previously described methods (46). Data were acquired on a Fortessa flow cytometer (Becton Dickinson, Franklin Lakes, NJ) and analyzed using FlowJo software (Tree Star, USA).

**Deep sequencing using barcoded-subamplicon sequencing.** For each sample, viral RNA was extracted with the RNeasy plus kit (Qiagen, Valencia, CA) according to the manufacturer's recommendations. The E gene was then reverse transcribed using AccuScript reverse transcriptase and primers flanking the E coding sequence (5'-GCCATTGCTGGCTTTGGGAAGC-3' and 5'-TGGTACTTGACCGGTCCCTCCAGGC-3').

To obtain high sequencing accuracy, we used the barcoded-subamplicon sequencing described in reference 17 (see also [https://jbloomlab.github.io/dms\\_tools2/bcsubamp.html](https://jbloomlab.github.io/dms_tools2/bcsubamp.html)). We generated five subamplicons of approximately 430 bp that tile the length of the E coding sequence. Each primer used for amplification adds an N8 randomized barcode and Illumina adapter sequences. These primers are provided in Data Set S2. Round 1 PCRs were set up using 12  $\mu$ l of 2 $\times$  KOD hot start master mix, 4 ng of cDNA template, 2  $\mu$ l of forward primer diluted to 5  $\mu$ M, and 2  $\mu$ l of reverse primer diluted to 5  $\mu$ M. The cycling program used for the round 1 PCR was as follows: step 1, 95°C for 2 min; step 2, 95°C for 20 s; step 3, 70°C for 1 s; step 4, 50°C for 20 s; step 5, 70°C for 20 s; step 6, go to step 2 (9 times); step 7, 95°C for 1 min; and step 8, 4°C hold. The final denaturation step ensures that double-stranded DNA molecules entering round 2 PCR will contain two uniquely barcoded variants.

Subamplicons were purified using AMPure XP beads (1:1 bead-to-sample volume), diluted, and used as the template for a second round of PCR using  $\sim 8 \times 10^5$  single-stranded template molecules per subamplicon, per sample. The primers used for round 2 PCR add sample-specific indices and the Illumina cluster-generating sequences. The round 2 PCRs contained 20  $\mu$ l of 2 $\times$  KOD hot start master mix, template DNA diluted to  $8 \times 10^5$  single-stranded DNA (ssDNA) molecules per subamplicon, 4  $\mu$ l of forward primer diluted to 5  $\mu$ M, and 4  $\mu$ l of reverse primer diluted to 5  $\mu$ M. The cycling program used for round 2 PCR was as follows: step 1, 95°C for 2 min; step 2, 95°C for 20 s; step 3, 70°C for 1 s; step 4, 55°C for 20 s; step 5, 70°C for 20 s; step 6, go to step 2 (29 times); and step 7, 4°C hold. The round 2 subamplicons were pooled, size selected by gel extraction (Zymo), and subsequently purified with AMPure XP beads. Subamplicons were sequenced with an Illumina HiSeq 2500 using 2  $\times$  250-bp paired-end reads in rapid run mode.

The raw deep sequencing data have been deposited in the Sequence Read Archive under BioProject accession number [PRJNA530795](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA530795).

**Analysis of deep sequencing and data availability.** Computer code that performs all of the analyses of the deep sequencing data is on GitHub at [https://github.com/jbloomlab/ZIKV\\_DMS\\_with\\_EvansLab](https://github.com/jbloomlab/ZIKV_DMS_with_EvansLab). This repository includes a notebook that gives detailed statistics on read depth, mutation frequencies, etc., for all samples (see [https://github.com/jbloomlab/ZIKV\\_DMS\\_with\\_EvansLab/blob/master/results/summary/analysis\\_notebook.md](https://github.com/jbloomlab/ZIKV_DMS_with_EvansLab/blob/master/results/summary/analysis_notebook.md)).

Briefly, we used `dms_tools2` (47) ([https://jbloomlab.github.io/dms\\_tools2/](https://jbloomlab.github.io/dms_tools2/)), version 2.4.14, to process the deep sequencing data to count the occurrences of each mutation in each sample (see [https://jbloomlab.github.io/dms\\_tools2/bcsubamp.html](https://jbloomlab.github.io/dms_tools2/bcsubamp.html) for details on how the software does this). The amino acid preferences were computed from these counts using the approach described in reference 47 (see also [https://jbloomlab.github.io/dms\\_tools2/prefs.html](https://jbloomlab.github.io/dms_tools2/prefs.html)). The mutational effects are simply the log of the preference value for the mutant amino acid divided by the preference value for the wild-type amino acid. The differential selection in the antibody selections was computed using the approach described in reference 18 (see also [https://jbloomlab.github.io/dms\\_tools2/diffsel.html](https://jbloomlab.github.io/dms_tools2/diffsel.html)), computing selection for each antibody relative to that under the no-antibody control condition. The figures in the paper show the results for just the positive differential selection. For both the amino acid preferences and antibody differential selection, the paper reports the average value of the three experimental replicates. To make the logo plots of the antibody selection in Fig. 5B, we used the program `dmslogo`, version 0.2.3 (48).

The raw numerical values of the counts of mutations in each sample, the amino acid preferences, and the differential selection are provided in CSV files at [https://github.com/jbloomlab/ZIKV\\_DMS\\_with\\_EvansLab](https://github.com/jbloomlab/ZIKV_DMS_with_EvansLab). In addition, Data Set S3 gives these numerical values in Excel format.

**Sanger sequencing of passaged viruses.** We verified that the sequences of individually rescued mutant viruses that displayed impaired growth characteristics were stable throughout the course of the above-described experiments. To do so, total RNA from 100  $\mu$ l of supernatant from the day 4 postinfection growth curve was extracted with the EZNA viral RNA kit (Omega Bio-Tek) according to the manufacturer's recommendations. The extracted RNA was used as a template for random-hexamer-primed cDNA synthesis using the SuperScript III first-strand synthesis system (Thermo Fisher Scientific, Waltham, MA). This cDNA was used for PCR using the Expand high-fidelity PCR system (Roche Life Sciences, Indianapolis, IN) with oligonucleotides flanking each mutant region and bulk PCR products subjected to Sanger sequencing. No reversions or otherwise-unintended mutations were found in any virus following the growth curve experiments.

**Alignment of naturally occurring E protein variants.** As shown by the results in Fig. 3A to C, we analyzed an alignment of naturally occurring ZIKV E protein variants. To create this alignment, we downloaded from the NCBI Virus Variation database (<https://www.ncbi.nlm.nih.gov/genome/viruses/variation/>) all ZIKV E protein sequences present as of 26 March 2019, collapsing identical nucleotide sequences. We then added the E region of our wild-type MR766 sequence to this alignment and removed any sequences that were incomplete or redundant at the codon level. We also manually removed sequences that appeared to be derived from an experiment or otherwise spurious. The resulting

alignment is available at [https://github.com/jbloomlab/ZIKV\\_DMS\\_with\\_EvansLab/blob/master/data/E\\_alignment.fasta](https://github.com/jbloomlab/ZIKV_DMS_with_EvansLab/blob/master/data/E_alignment.fasta).

**Structural analyses.** Mutational tolerance and antibody escape were mapped onto the surface representation of the E protein dimer of PDB structure 5IRE using PyMol. Mutational tolerance was quantified by the Shannon entropy. The structure in Fig. 3D is colored by site from the lowest to highest Shannon entropy as indicated by the color scale bar in that panel. Sites of escape from monoclonal antibodies were mapped on a color gradient from white to red according to the positive-site differential selection (averaged across triplicates) and used to color the structures in Fig. 5C and D.

The data in Fig. 3B and E show Pearson correlation coefficients (*R* values) and the associated *P* values as calculated using Python's `scipy.stats.pearsonr` function.

**Data availability.** The raw deep sequencing data have been deposited in the Sequence Read Archive under BioProject accession number PRJNA530795. The sequence of the plasmid encoding the wild-type MR766 ZIKV strain was previously deposited in GenBank under accession number KX830961.1. The computer code used to analyze the deep sequencing data to generate the results in the paper is available on GitHub at [https://github.com/jbloomlab/ZIKV\\_DMS\\_with\\_EvansLab/](https://github.com/jbloomlab/ZIKV_DMS_with_EvansLab/).

## SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/JVI.01291-19>.

**SUPPLEMENTAL FILE 1**, XLS file, 0.1 MB.

**SUPPLEMENTAL FILE 2**, XLSX file, 0.01 MB.

**SUPPLEMENTAL FILE 3**, XLSX file, 1.3 MB.

## ACKNOWLEDGMENTS

We thank Adam Dingens for assistance with data analysis, Charles Rice (Rockefeller University) for 293T cells, and Mansun Law (Scripps) for the HCV AR3A negative-control antibody. We thank the Fred Hutch Genomics Core for performing the deep sequencing and the Dean's Flow Cytometry CORE at the Icahn School of Medicine at Mount Sinai for technical assistance.

This work was supported by grants number R21 AI133649 (to M.J.E. and J.D.B.), R21 AI140196 (to M.J.E.), and R01 AI141707 (to J.D.B.). J.D.B. is an Investigator of the Howard Hughes Medical Institute. M.J.E. and J.D.B. hold Investigator in Pathogenesis of Infectious Disease Awards from the Burroughs Wellcome Fund.

## REFERENCES

- Duffy MR, Chen T-H, Hancock WT, Powers AM, Kool JL, Lanciotti RS, Pretrick M, Marfel M, Holzbauer S, Dubray C, Guillaumot L, Griggs A, Bel M, Lambert AJ, Laven J, Kosoy O, Panella A, Biggerstaff BJ, Fischer M, Hayes EB. 2009. Zika virus outbreak on Yap Island, Federated States of Micronesia. *N Engl J Med* 360:2536–2543. <https://doi.org/10.1056/NEJMoa0805715>.
- Musso D, Nilles EJ, Cao-Lormeau V-M. 2014. Rapid spread of emerging Zika virus in the Pacific area. *Clin Microbiol Infect* 20:O595–O596. <https://doi.org/10.1111/1469-0691.12707>.
- Campos GS, Bandeira AC, Sardi SI. 2015. Zika virus outbreak, Bahia, Brazil. *Emerg Infect Dis* 21:1885–1886. <https://doi.org/10.3201/eid2110.150847>.
- Broutet N, Krauer F, Riesen M, Khalakdina A, Almiron M, Aldighieri S, Espinal M, Low N, Dye C. 2016. Zika virus as a cause of neurologic disorders. *N Engl J Med* 374:1506–1509. <https://doi.org/10.1056/NEJMp1602708>.
- Mlakar J, Korva M, Tul N, Popović M, Poljšak-Prijatelj M, Mraz J, Kolenc M, Resman Rus K, Vesnaver Vipotnik T, Fabjan Vodusek V, Vizjak A, Pizem J, Petrovec M, Avšič Županc T. 2016. Zika virus associated with microcephaly. *N Engl J Med* 374:951–958. <https://doi.org/10.1056/NEJMoa1600651>.
- Lindenbach BD, Murray CL, Thiel H-J, Rice CM. 2013. Flaviviridae, p 712–746. *In* Knipe DM, Howley PM, Cohen JI, Griffin DE, Lamb RA, Martin MA, Racaniello VR, Roizman B (ed), *Fields virology*, 6th ed. Lippincott Williams & Wilkins, Philadelphia, PA.
- Sirohi D, Chen Z, Sun L, Klose T, Pierson TC, Rossmann MG, Kuhn RJ. 2016. The 3.8 Å resolution cryo-EM structure of Zika virus. *Science* 352:467–470. <https://doi.org/10.1126/science.aaf5316>.
- Modis Y, Ogata S, Clements D, Harrison SC. 2004. Structure of the dengue virus envelope protein after membrane fusion. *Nature* 427:313–319. <https://doi.org/10.1038/nature02165>.
- Rey FA, Heinz FX, Mandl C, Kunz C, Harrison SC. 1995. The envelope glycoprotein from tick-borne encephalitis virus at 2 Å resolution. *Nature* 375:291–298. <https://doi.org/10.1038/375291a0>.
- Haddow AD, Schuh AJ, Yasuda CY, Kasper MR, Heang V, Huy R, Guzman H, Tesh RB, Weaver SC. 2012. Genetic characterization of Zika virus strains: geographic expansion of the Asian lineage. *PLoS Negl Trop Dis* 6:e1477. <https://doi.org/10.1371/journal.pntd.0001477>.
- Nybakken GE, Oliphant T, Johnson S, Burke S, Diamond MS, Fremont DH. 2005. Structural basis of West Nile virus neutralization by a therapeutic antibody. *Nature* 437:764–769. <https://doi.org/10.1038/nature03956>.
- Thullier P, Demangel C, Bedouelle H, Mègret F, Joann A, Deubel V, Mazié JC, Lafaye P. 2001. Mapping of a dengue virus neutralizing epitope critical for the infectivity of all serotypes: insight into the neutralization mechanism. *J Gen Virol* 82:1885–1892. <https://doi.org/10.1099/0022-1317-82-8-1885>.
- Chen Y, Maguire T, Hileman RE, Fromm JR, Esko JD, Linhardt RJ, Marks RM. 1997. Dengue virus infectivity depends on envelope protein binding to target cell heparan sulfate. *Nat Med* 3:866–871. <https://doi.org/10.1038/nm0897-866>.
- Stettler K, Beltramello M, Espinosa DA, Graham V, Cassotta A, Bianchi S, Vanzetta F, Minola A, Jaconi S, Mele F, Foglierini M, Pedotti M, Simonelli L, Dowall S, Atkinson B, Percivalle E, Simmons CP, Varani L, Blum J, Baldanti F, Cameroni E, Hewson R, Harris E, Lanzavecchia A, Sallusto F, Corti D. 2016. Specificity, cross-reactivity, and function of antibodies elicited by Zika virus infection. *Science* 353:823–826. <https://doi.org/10.1126/science.aaf8505>.
- Fowler DM, Fields S. 2014. Deep mutational scanning: a new style of protein science. *Nat Methods* 11:801–807. <https://doi.org/10.1038/nmeth.3027>.
- Wu NC, Young AP, Al-Mawsawi LQ, Olson CA, Feng J, Qi H, Chen S-H, Lu I-H, Lin C-Y, Chin RG, Luan HH, Nguyen N, Nelson SF, Li X, Wu T-T, Sun R. 2014. High-throughput profiling of influenza A virus hemagglutinin

- gene at single-nucleotide resolution. *Sci Rep* 4:4942. <https://doi.org/10.1038/srep04942>.
17. Doud MB, Bloom JD. 2016. Accurate measurement of the effects of all amino acid mutations on influenza hemagglutinin. *Viruses* 8:155. <https://doi.org/10.3390/v8060155>.
  18. Doud MB, Hensley SE, Bloom JD. 2017. Complete mapping of viral escape from neutralizing antibodies. *PLoS Pathog* 13:e1006271. <https://doi.org/10.1371/journal.ppat.1006271>.
  19. Haddox HK, Dingens AS, Hilton SK, Overbaugh J, Bloom JD. 2018. Mapping mutational effects along the evolutionary landscape of HIV envelope. *Elife* 7:e34420. <https://doi.org/10.7554/eLife.34420>.
  20. Dingens AS, Arenz D, Weight H, Overbaugh J, Carrera J. 2019. An antigenic atlas of HIV-1 escape from broadly neutralizing antibodies distinguishes functional and structural epitopes. *Immunity* 50:520–532.e3. <https://doi.org/10.1016/j.immuni.2018.12.017>.
  21. Setoh YX, Amarilla AA, Peng NYG, Griffiths RE, Carrera J, Freney ME, Nakayama E, Ogawa S, Watterson D, Modhiran N, Nanyonga FE, Torres FJ, Slonchak A, Periasamy P, Prow NA, Tang B, Harrison J, Hobson-Peters J, Cuddihy T, Cooper-White J, Hall RA, Young PR, Mackenzie JM, Wolvetang E, Bloom JD, Suhrbier A, Khromykh AA. 2019. Determinants of Zika virus host tropism uncovered by deep mutational scanning. *Nat Microbiol* 4:876–887. <https://doi.org/10.1038/s41564-019-0399-4>.
  22. Gong D, Zhang T-H, Zhao D, Du Y, Chapa TJ, Shi Y, Wang L, Contreras D, Zeng G, Shi P-Y, Wu T-T, Arumugaswami V, Sun R. 2018. High-throughput fitness profiling of Zika Virus E protein reveals different roles for glycosylation during infection of mammalian and mosquito cells. *iScience* 1:97–111. <https://doi.org/10.1016/j.isci.2018.02.005>.
  23. Sourisseau M, Lawrence DJP, Schwarz MC, Storrs CH, Veit EC, Bloom JD, Evans MJ. 2019. Deep mutational scanning comprehensively maps how Zika envelope protein mutations affect viral growth and antibody escape. *bioRxiv* <https://doi.org/10.1101/725556>.
  24. Bloom JD. 2014. An experimentally determined evolutionary model dramatically improves phylogenetic fit. *Mol Biol Evol* 31:1956–1978. <https://doi.org/10.1093/molbev/msu173>.
  25. Dingens AS, Haddox HK, Overbaugh J, Bloom JD. 2017. Comprehensive mapping of HIV-1 escape from a broadly neutralizing antibody. *Cell Host Microbe* 21:777–787.e4. <https://doi.org/10.1016/j.chom.2017.05.003>.
  26. Schwarz MC, Sourisseau M, Espino MM, Gray ES, Chambers MT, Tortorella D, Evans MJ. 2016. Rescue of the 1947 Zika virus prototype strain with a cytomegalovirus promoter-driven cDNA clone. *mSphere* 1:e00246-16. <https://doi.org/10.1128/mSphere.00246-16>.
  27. Dai L, Song J, Lu X, Deng Y-Q, Musyoki AM, Cheng H, Zhang Y, Yuan Y, Song H, Haywood J, Xiao H, Yan J, Shi Y, Qin C-F, Qi J, Gao GF. 2016. Structures of the Zika virus envelope protein and its complex with a flavivirus broadly protective antibody. *Cell Host Microbe* 19:696–704. <https://doi.org/10.1016/j.chom.2016.04.013>.
  28. Ramsey DC, Scherrer MP, Zhou T, Wilke CO. 2011. The relationship between relative solvent accessibility and evolutionary rate in protein evolution. *Genetics* 188:479–488. <https://doi.org/10.1534/genetics.111.128025>.
  29. Thyagarajan B, Bloom JD. 2014. The inherent mutational tolerance and antigenic evolvability of influenza hemagglutinin. *Elife* 3:e03300. <https://doi.org/10.7554/eLife.03300>.
  30. Haddox HK, Dingens AS, Bloom JD. 2016. Experimental estimation of the effects of all amino acid mutations to hiv's envelope protein on viral replication in cell culture. *PLoS Pathog* 12:e1006114. <https://doi.org/10.1371/journal.ppat.1006114>.
  31. Fulton BO, Sachs D, Schwarz MC, Palese P, Evans MJ. 2017. Transposon mutagenesis of the Zika virus genome highlights regions essential for RNA replication and restricted for immune evasion. *J Virol* 91:e00698-17. <https://doi.org/10.1128/JVI.00698-17>.
  32. Wang J, Bardelli M, Espinosa DA, Pedotti M, Ng T-S, Bianchi S, Simonelli L, Lim EXY, Foglierini M, Zatta F, Jaconi S, Beltramello M, Cameroni E, Fibriansah G, Shi J, Barca T, Pagani I, Rubio A, Broccoli V, Vicenzi E, Graham V, Pullan S, Dowall S, Hewson R, Jurt S, Zerbe O, Stettler K, Lanzavecchia A, Sallusto F, Cavalli A, Harris E, Lok S-M, Varani L, Corti D. 2017. A human bi-specific antibody against Zika virus with high therapeutic potential. *Cell* 171:229–241.e15. <https://doi.org/10.1016/j.cell.2017.09.002>.
  33. Chambers MT, Schwarz MC, Sourisseau M, Gray ES, Evans MJ. 2018. Probing Zika virus neutralization determinants with glycoprotein mutants bearing linear epitope insertions. *J Virol* 92:e00505-18. <https://doi.org/10.1128/JVI.00505-18>.
  34. Fritz R, Stiasny K, Heinz FX. 2008. Identification of specific histidines as pH sensors in flavivirus membrane fusion. *J Cell Biol* 183:353–361. <https://doi.org/10.1083/jcb.200806081>.
  35. Nelson S, Poddar S, Lin T-Y, Pierson TC. 2009. Protonation of individual histidine residues is not required for the pH-dependent entry of West Nile virus: evaluation of the “histidine switch” hypothesis. *J Virol* 83:12631–12635. <https://doi.org/10.1128/JVI.01072-09>.
  36. Doud MB, Ashenberg O, Bloom JD. 2015. Site-specific amino acid preferences are mostly conserved in two closely related protein homologs. *Mol Biol Evol* 32:2944–2960. <https://doi.org/10.1093/molbev/msv167>.
  37. Lee JM, Huddleston J, Doud MB, Hooper KA, Wu NC, Bedford T, Bloom JD. 2018. Deep mutational scanning of hemagglutinin helps predict evolutionary fates of human H3N2 influenza variants. *Proc Natl Acad Sci U S A* 115:E8276–E8285. <https://doi.org/10.1073/pnas.1806133115>.
  38. Wu NC, Young AP, Al-Mawsawi LQ, Olson CA, Feng J, Qi H, Luan HH, Li X, Wu T-T, Sun R. 2014. High-throughput identification of loss-of-function mutations for anti-interferon activity in the influenza A virus NS segment. *J Virol* 88:10157–10164. <https://doi.org/10.1128/JVI.01494-14>.
  39. Soh YS, Moncla LH, Eguia R, Bedford T, Bloom JD. 2019. Comprehensive mapping of adaptation of the avian influenza polymerase protein PB2 to humans. *Elife* 8:e45079. <https://doi.org/10.7554/eLife.45079>.
  40. Goo L, VanBlargan LA, Dowd KA, Diamond MS, Pierson TC. 2017. A single mutation in the envelope protein modulates flavivirus antigenicity, stability, and pathogenesis. *PLoS Pathog* 13:e1006178. <https://doi.org/10.1371/journal.ppat.1006178>.
  41. Crill WD, Chang G-J. 2004. Localization and characterization of flavivirus envelope glycoprotein cross-reactive epitopes. *J Virol* 78:13975–13986. <https://doi.org/10.1128/JVI.78.24.13975-13986.2004>.
  42. Law M, Maruyama T, Lewis J, Giang E, Tarr AW, Stamatakis Z, Gastaminza P, Chisari FV, Jones IM, Fox RI, Ball JK, McKeating JA, Kneteman NM, Burton DR. 2008. Broadly neutralizing antibodies protect against hepatitis C virus quasispecies challenge. *Nat Med* 14:25–27. <https://doi.org/10.1038/nm1698>.
  43. Michta ML, Hopcraft SE, Narbus CM, Kratovac Z, Israelow B, Sourisseau M, Evans MJ. 2010. Species-specific regions of occludin required by hepatitis C virus for cell entry. *J Virol* 84:11696–11708. <https://doi.org/10.1128/JVI.01555-10>.
  44. Reed LJ, Muench H. 1938. A simple method of estimating fifty per cent endpoints. *Am J Epidemiol* 27:493–497. <https://doi.org/10.1093/oxfordjournals.aje.a118408>.
  45. Lanciotti RS, Kosoy OL, Laven JJ, Velez JO, Lambert AJ, Johnson AJ, Stanfield SM, Duffy MR. 2008. Genetic and serologic properties of Zika virus associated with an epidemic, Yap State, Micronesia, 2007. *Emerg Infect Dis* 14:1232–1239. <https://doi.org/10.3201/eid1408.080287>.
  46. Sourisseau M, Michta ML, Zony C, Israelow B, Hopcraft SE, Narbus CM, Parra Martín A, Evans MJ. 2013. Temporal analysis of hepatitis C virus cell entry with occludin directed blocking antibodies. *PLoS Pathog* 9:e1003244. <https://doi.org/10.1371/journal.ppat.1003244>.
  47. Bloom JD. 2015. Software for the analysis and visualization of deep mutational scanning data. *BMC Bioinformatics* 16:168. <https://doi.org/10.1186/s12859-015-0590-4>.
  48. Lee JM, Eguia R, Zost SJ, Choudhary S, Wilson PC, Bedford T, Stevens-Ayers T, Boeckh M, Hurt A, Lakdawala SS, Hensley SE, Bloom JD. 2019. Mapping person-to-person variation in viral mutations that escape polyclonal serum targeting influenza hemagglutinin. *bioRxiv* <https://doi.org/10.1101/670497>.
  49. Touw WG, Baakman C, Black J, Te Beek TAH, Krieger E, Joosten RP, Vriend G. 2015. A series of PDB-related databanks for everyday needs. *Nucleic Acids Res* 43:D364–D368. <https://doi.org/10.1093/nar/gku1028>.
  50. Tien MZ, Meyer AG, Sydykova DK, Spielman SJ, Wilke CO. 2013. Maximum allowed solvent accessibility of residues in proteins. *PLoS One* 8:e80635. <https://doi.org/10.1371/journal.pone.0080635>.