Identification of a Zika NS2B epitope for which absence of IgG response is associated with severe neurological symptoms and the design of a biomarker capable of discriminatory diagnostics between severe and non-severe clinical phenotypes

Felix F. Loeffler[a,¶], Isabelle F.T. Viana[b,¶], Nico Fischer[c], Danilo F. Coêlho[b,d], Carolina S. Silva[e], Antônio F. Purificação Jr.[b], Catarina M.C.S. Araújo[b], Bruno H.S. Leite[b], Ricardo Durães-Carvalho[f], Tereza Magalhães[g], Clarice N.L. Morais[b], Marli T. Cordeiro[b], Roberto D. Lins[2,&], Ernesto T.A. Marques[b,h,&] and Thomas Jaenisch[c,i,&*]

a. Max Planck Institute of Colloids and Interfaces, Department of Biomolecular Systems, Potsdam, Germany.
b. Department of Virology, Aggeu Magalhães Institute, Oswaldo Cruz Foundation, Recife, PE, Brazil.
c. Section Clinical Tropical Medicine, Department of Infectious Diseases, Heidelberg University Hospital, Germany.
d. Department of Fundamental Chemistry, Federal University of Pernambuco, Recife, PE, Brazil.
e. Department of Chemical Engineering, Federal University of Pernambuco, Recife, PE, Brazil.
f. Laboratory of Virology, University of Campinas, Campinas, SP, Brazil.
g. Arthropod-borne and Infectious Diseases Laboratory (AIDL), Department of Microbiology, Immunology and Pathology, Colorado State University, Fort Collins, CO, United States of America.
h. Department of Infectious Diseases and Microbiology, University of Pittsburgh, Pittsburgh, PA, United States of America
i. German Centre for Infection Research (DZIF), Heidelberg Site, Heidelberg, Germany.
¶ These authors contributed equally to this work.
& These authors also contributed equally to this work.
*Corresponding author: thomas.jaenisch@urz.uni-heidelberg.de

Abstract

The identification of specific biomarkers for Zika infection and its clinical complications is fundamental to mitigate the infection spread, which has been associated with a broad range of neurological sequelae. We present the characterization of antibody responses in serum samples from individuals infected with Zika, presenting non-severe (classical) and severe (neurological disease) phenotypes, with high-density peptide arrays comprising the Zika NS1 and NS2B proteins. The data pinpoints one strongly IgG-targeted NS2B epitope in non-severe infections, which is absent in Zika patients, where infection progressed to the severe phenotype. This differential IgG profile between the studied groups was confirmed by multivariate data analysis. Molecular dynamics simulations and circular dichroism have shown that the peptide in solution presents itself in a sub-optimal conformation for antibody recognition, which led us to computationally engineer an artificial protein able to stabilize the NS2B epitope structure. The engineered protein was used to interrogate paired samples from mothers and their babies presenting Zika-associated microcephaly and confirmed the absence of NS2B IgG response in those samples. These findings suggest that the assessment of antibody responses to the herein identified NS2B epitope is a strong candidate biomarker for the diagnosis and prognosis of Zika-associated neurological disease.

Introduction

Zika virus (ZIKV) emerged in the Americas, causing an unprecedented epidemic of microcephaly in babies born to mothers infected during pregnancy, and neurological disease in adults following acute infection. Some uncertainty still remains regarding the time of introduction into the Americas; however, the virus likely entered Brazil in 2013,[1] with the first cases of microcephaly reported in 2015. The temporal correlation between ZIKV introduction in Brazil and the microcephaly epidemic led the Brazilian Government to hypothesize the presence of a causal association.[2, 3] Shortly after (1 February 2016), the World Health Organization (WHO) declared a 'Public Health Emergency of International Concern' for the clusters of microcephaly and other neurological disorders, which was only lifted in November 2016.[4]

The causal association between congenital ZIKV infection and microcephaly (now referred to as Congenital Zika Syndrome [CZS] due to its broad range of clinical manifestations) was accepted by WHO in 2016,[5, 6] and evidence continued to accumulate in the following years.[7] In addition to that, explosive outbreaks in large populations in Latin America revealed other severe neurological sequelae of ZIKV infection in children and adults, including Guillain-Barré syndrome: an immune-mediated demyelinating motor and sensory peripheral neuropathy leading to paralysis.[8] The ability of ZIKV to cause neurological disease is not unique among flaviviruses.[9] Several candidate neurovirulence mechanisms have been postulated, among them the glycosylation of the envelope protein,[10] the presence of neuronal receptors (such as AXL) only recognized by ZIKV, and the mutation S139N in the precursor membrane protein of ZIKV, which enhances neurovirulence possibly by creating a new receptor for progenitor cells.[8, 11] To date, however, the viral determinants of ZIKV neurovirulence and the immune components involved have not been fully unraveled.

Prior infection with dengue virus (DENV) has been suggested to be associated with more severe manifestations in ZIKV infections.[12] In countries like Brazil, more than 90% of the adult population has been previously exposed to DENV.[13] ZIKV and DENV are both members of the *Flaviviridae* family and exhibit considerable cross-reactivity in serological tests, which proves the close phylogenetic and antigenic relationship between these viruses [14-16]. High anti-DENV titers have been reported to be linked with protection from Zika,[17] whereas sub-neutralizing levels of anti-DENV have been shown to enhance ZIKV infection in vitro.[12, 18, 19]

The overlapping clinical syndromes and geographical distribution make differential diagnosis a major challenge, especially in situations where the causative viral agent cannot directly be detected and diagnosis is based on indirect serological tests. Highly specific serological tests, capable of discriminating between different flaviviruses and therefore providing accurate prognosis of severe disease, are so far not available. Several serological test platform candidates are currently under development and validation;[20-24] however, an accurate diagnostic and prognostic marker for ZIKV infection and its severe neurological manifestations has yet to be demonstrated.[25]

We aimed at characterizing important ZIKV-specific epitopes by analyzing serum samples taken from patients with different clinical outcomes after ZIKV infection in Recife, Northeast Brazil, the epicenter of the ZIKV outbreak in the Americas. With this objective, we explored a novel high-density peptide array technology to search the whole proteome of ZIKV for serological biomarkers of infection. Several platforms based on peptide arrays have been developed and applied for infectious disease research.[26, 27] While most reports qualify the arrays using the detection of linear epitopes binding well-characterized monoclonal antibodies, recently, we and others have used this technology to demonstrate the readout of antibody profiles against malaria in different patient sera,[28] as well as in Lyme disease patients, in the identification of ZIKV epitopes[29] and for general vaccine studies.[29-31] Using our novel solid material-based combinatorial synthesis method,[32, 33] the whole peptidome of ZIKV was produced on a unique single glass slide as overlapping 15-mer peptides.

Due to the complexity and multivariate nature of the data obtained with this new peptide array technology, multivariate analysis techniques rise as alternatives to explore the dataset and to recognize consistently patterned structures that correlate groups of individuals with different clinical outcomes and specific peptide sequences. In this scenario, exploratory and classification techniques were employed to screen our peptide array against ZIKV clinical samples and identify one epitope sequence as a potential biomarker for ZIKV infection and its neurological manifestation with statistical reliability. Following peptide identification, the peptide's structural and thermodynamic stability were assessed by molecular dynamics simulations and confirmed by circular dichroism analysis. To improve the performance of the identified epitope as an immunological marker, we computationally designed a new protein that mimics the natural structure of the peptide. The potential use of this protein as a prognostic marker for ZIKV infections with neurological complications was then confirmed.

Results and discussion

Identification of an immunoreactive peptide within the ZIKV NS2B protein by high-density peptide array and Multivariate Analysis

The customized peptide array containing 1,878 15-mer peptides covering the whole proteome of the ZIKV was first used to screen acute and convalescent ZIKV samples, stratified by the presence or absence of DENV background (Tables S1 and S2 and Figure S1), to determine the overall IgG-binding properties towards specific peptides. The samples were categorized into eight groups consisting of i) acute Zika patients without dengue infection history (n=14), ii) acute Zika patients with dengue infection history (n=17), iii) acute Zika patients with low antibody titers against dengue (n=3, this group presents negative results for Dengue PRNT and Dengue IgG capture ELISA, and was included due to the difficulty to define the comparison group of 'convalescent Zika patients without dengue infection history' since Dengue seroprevalence in Recife can be above 90%[13, 34]), iv) convalescent Zika patients without dengue infection history (n=14), v) convalescent Zika patients with dengue infection history (n=17), vi) convalescent Zika patients with low antibody titers against dengue (n=3), vii) Zika infections with neurological symptoms (NeuroZIKV) without dengue infection history (n=11), and viii) Zika infections with neurological symptoms (NeuroZIKV) with dengue infection history (n=24). Important to note that all of the NeuroZIKV serum samples used in the study have exhibited ZIKV IgG positive results in previous tests (IgG ELISA and PRNT). Serum samples from individuals not exposed to ZIKV or DENV were used as assay control group (n=8).

The median fluorescence intensity data obtained for each group showed specific IgG responses from patients with confirmed ZIKV infection in convalescent *versus* acute samples towards several peptides of the NS1 and NS2B proteins (for clarity purposes, only the data from NS1 and NS2 proteins is shown in Figure 1).

A second peptide array, composed of 401 15-mer peptides only covering the amino acid sequences of NS1, NS2A and NS2B proteins, was used to confirm these findings. The fluorescence intensity data of both arrays is shown in Figure 1A. Individual analysis of the median for each NS2B peptide revealed four strongly immunoreactive amino acid sequences, termed P1 (GDITWEKDAEVTGNS), P2 (GDITWEKDAEITGNS), P3 (TWEKDAEVTGNSPRL) and P4 (TWEKDAEITGNSPRL) (Figure 1B). The corresponding peptides are adjacent; consequently, their sequences are overlapping. Combining the two sequences led to the identification of a major NS2B epitope (GDITWEKDAEV(I)TGNSPRLDVA) located at the ZIKV poly-proteome position (1425–1445). Figure 1B shows differences among the median values of different groups of patients from a qualitative standpoint.

Our data show that strong IgG antibody response was elicited against the identified epitope in non-severe ZIKV cases and suggests that the antibody response against NS2B is absent in Zika infections with neurological symptoms (NeuroZIKV) in individuals not previously exposed to DENV. To strengthen this observation, Principal Component Analysis (PCA) was initially performed to assess the IgG response in the NS2B peptide array dataset. PCA is an unsupervised technique based on dataset's variance that can be employed as a pattern recognition tool (for further detail on PCA theory see associated content). The PCA analysis highlighted similarities and differences between NeuroZIKV and non-severe ZIKV cases, as well as identified the importance of each peptide signal related to these differences.

The scores plot (shown in Figure 2A) shows the relationship among objects analyzed (patients), in such way that the closer two patients are, more similarities are observed in terms of signal response against the antibody. The 2 first principal components (PCs) obtained from a 2-component PCA model are depicted in Figures 2A and 2B. In Figure 2A the scores scatter plot is shown for the control group (ZIKV- DENV-), convalescent Zika without Dengue infection history (ZIKV+, DENV- (group IV)), convalescent Zika with Dengue infection history (ZIKV+, DENV+ (group V)), convalescent Zika with low antibody titers against Dengue (ZIKV+, DENV low (group VI)) and Zika infection (acute and convalescent) with neurological symptoms without Dengue infection history (NeuroZIKV, DENV- (group VII)) patients. It is possible to observe that the majority of the samples in the NeuroZIKV, DENV- and the control group has negative score values for PC1, whereas non-severe ZIKV samples have positive score values of PC1. The loadings plot, depicted in Figure 2B, shows how each peptide array contributes for the differences observed in the scores plot. Positive loading values indicates higher intensity responses for patients with positive scores and lower intensities for patients with negative scores. In PC1, the difference accounts for 37% of total data variance and is mainly related to the following sequence peptides with positive values of PC1 loading plot (Figure 2B): GDITWEKDAEVTGNS, GDITWEKDAEITGNS, TWEKDAEVTGNSPRL, TWEKDAEITGNSPRL, KDAEVTGNSPRLDVA and KDAEITGNSPRLDVA, in which the first four are the peptides described in Figure 1B. All these peptides have in common the amino acid sequence KDAEV(I)TGNS.

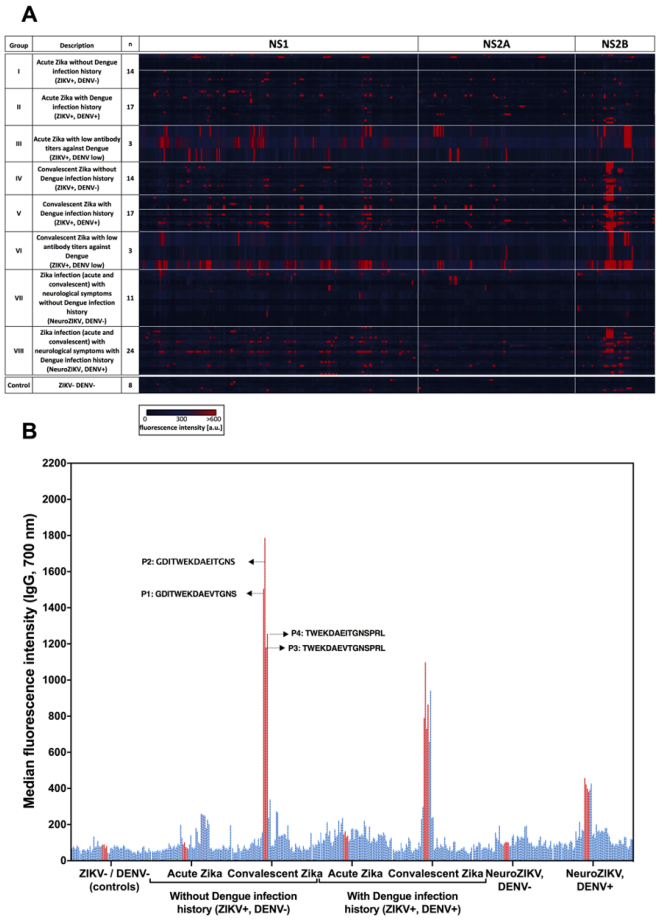


Figure 1. Peptide array screening and identification of putative ZIKV NS2B peptides. (A) Peptide array screening showing the selection of 401 peptides covering the ZIKV NS1 and NS2B proteins. The IgM and IgG antibody profiles were determined for a control group (n=8) and after incubation of acute and convalescent serum samples from 103 individuals according to the following groups: i) acute Zika patients without dengue infection history (n=14), ii) acute Zika patients with dengue infection history (n=17), iii) acute Zika patients with low antibody titers against dengue (n=3), iv) convalescent Zika patients without dengue infection history (n=14), v) convalescent Zika patients with dengue infection history (n=17), vi) convalescent Zika patients with low antibody titers against dengue (n=3), vii) Zika infections with neurological manifestations (NeuroZIKV) without dengue infection history (n=11), viii) NeuroZIKV infections with dengue infection history (n=24). (B) Identification of putative ZIKV NS2B epitopes. Immunoreactivity of the ZIKV NS2B peptides was individually assessed against the serum samples groups described in Figure 1A. Bars represent the median fluorescence intensity for individual peptides. The four peptides showing the highest fluorescence signals in the array are highlighted in red.

According to the loadings and scores plots of PCA model, the abovementioned peptides (positive loading values in PC1, Figure 2B) are positively correlated to non-severe ZIKV patients (positive scores in PC1, Figure 2A blue symbols) and negatively correlated to NeuroZIKV, DENV- patients (negative scores in PC1, Figure 2A red and grey symbols). Therefore, it is possible to state that non-severe ZIKV samples have higher fluorescence signal intensity for those peptides when compared to NeuroZIKV, DENV- samples. Moreover, no significant difference among non-severe ZIKV patients (groups IV, V and VI) or between NeuroZIKV, DENV- patients and the control group (Figure 2A red and grey symbols) were observed. It is important to note that there is no difference between the NeuroZIKV, DENV- and the control group, therefore the NeuroZIKV, DENV- signal response is only background noise. Although PC2 accounts for 12% of total data variance, its score scatter plot does not show distinctive features.

Our results from the PCA model indicate that the previously mentioned peptides are potential candidates to discriminate NeuroZIKV patients in the clinical outcomes. To test statistical difference among the groups in a multivariate way, the fluorescence signals were employed as classification inputs to build partial least squares discriminant analysis (PLS-DA) models. In contrast with PCA, PLS-DA is a supervised technique focused in classification and aims to maximize the covariance between the groups of patients and its IgG response profiles (for further detail on PLS-DA theory see associated content). In this case, the model is built in a way that a decision

function, shown as a threshold on scores plot of PLS-DA (red line Figure 2E), is set to discriminate patients belonging to different groups and which peptides are more important in that discrimination.

Four different PLS-DA models were built: (i) a global model to test the ability of the epitope of interest to discriminate all NeuroZIKV patients (groups VII, VIII) plus the control group from all non-severe ZIKV patients (groups IV, V and VI); (ii) a model to discriminate NeuroZIKV, DENV- (group VII) from non-severe ZIKV patients (groups IV, V and VI); (iii) a model to discriminate NeuroZIKV (groups VII, VIII) from non-severe ZIKV, DENV- (group IV) patients; and (iv) a model to discriminate NeuroZIKV patients according to their DENV background (group VII *versus* VIII).

The global model was built to investigate if the differences on IgG signal response for the given epitope will suffice to discriminate non-severe (groups IV, V and VI) from NeuroZIKV patients (groups VII and VIII), regardless their DENV background. The PLS-DA model shown in Figure 2C provided a value for the area under the curve (AUC) of the receiver operating characteristic (ROC) curve equal to 0.83 with sensitivity and specificity values for NeuroZIKV class of 79% and 85%, respectively. The variable importance in projection (VIP) scores plot is an important parameter to evaluate which peptides are more important to diagnosis prediction (Figure 2D) and must have values above the significance threshold. The peptides at data matrix positions 364-367 containing the sequence KDAEV(I)TGNS are the ones above the discriminant threshold and, therefore the most important variables to discriminate the two groups of patients. This result corroborates that the mentioned peptides are more likely to be an epitope from NS2B protein that can be used as a potential biomarker for NeuroZIKV infection.

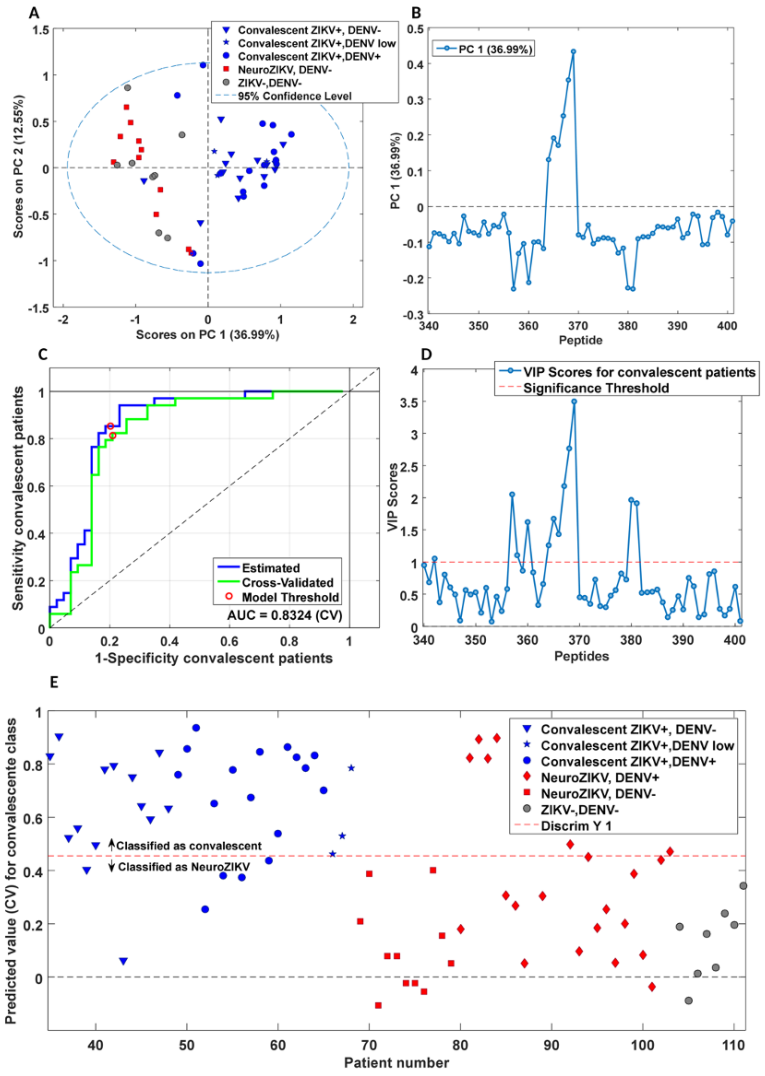


Figure 2. Summary of the statistical data analysis. (A) Scores and (B) loadings plots of 2-component PCA model including control group, convalescent ZIKV+ (DENV-, DENV+ and DEN low) and NeuroZIKV, DENV- patients; and PLS-DA results for the 1-latent variable global model, including (C) ROC curve, (D) VIP scores plot for PLS-DA model and (E) predicted values of samples.

Figure 2E shows the predicted scores for each patient, in which those who shows values above 0.45 (defined as the threshold by maximizing model sensitivity and specificity) will be classified as convalescent patients. It is worth mentioning that thirteen (9 NeuroZIKV DENV+ and 4 non-severe ZIKV+ DENV+) out of fifteen misclassified samples belong to patients with previous DENV infection (Figure 2E). This result suggests that antibodies from DENV previous infections may cross-react with the identified epitope.

To evaluate the influence of NeuroZIKV, DENV+ patients in the classification model performance, a 1-latent variable PLS-DA model was built to classify patients as either non-severe or control/NeuroZIKV, DENV-, without including NeuroZIKV, DENV+ patients. The obtained ROC curve (Figure 3A) provided an AUC value of approximately 0.98 at 95% confidence level. This improved classification model showed higher sensitivity and specificity values of 89% and 94%, respectively, for NeuroZIKV class. Evaluating the VIP scores plot for the classification model (Figure 3B), it is clear that the peptides containing the KDAEV(I)TGNS sequence appear again as above the significance threshold. This result shows that the identified epitope from NS2B protein can be used as a biomarker for the identification of ZIKV patients who have developed severe neurological symptoms.

Interestingly, Figure 1B shows that low IgG antibody response against the identified epitope was observed in NeuroZIKV patients that have been previously exposed to DENV. However, this response is apparently higher when compared to the control and NeuroZIKV groups without previous DENV infection. This finding suggests that DENV infection may elicit IgG response against the NS2B epitope. To further investigate this, an additional PLS-DA model was built (Figure 3C and 3D) including only non-severe ZIKV, DENV- and all NeuroZIKV (groups VII, VIII) patients.

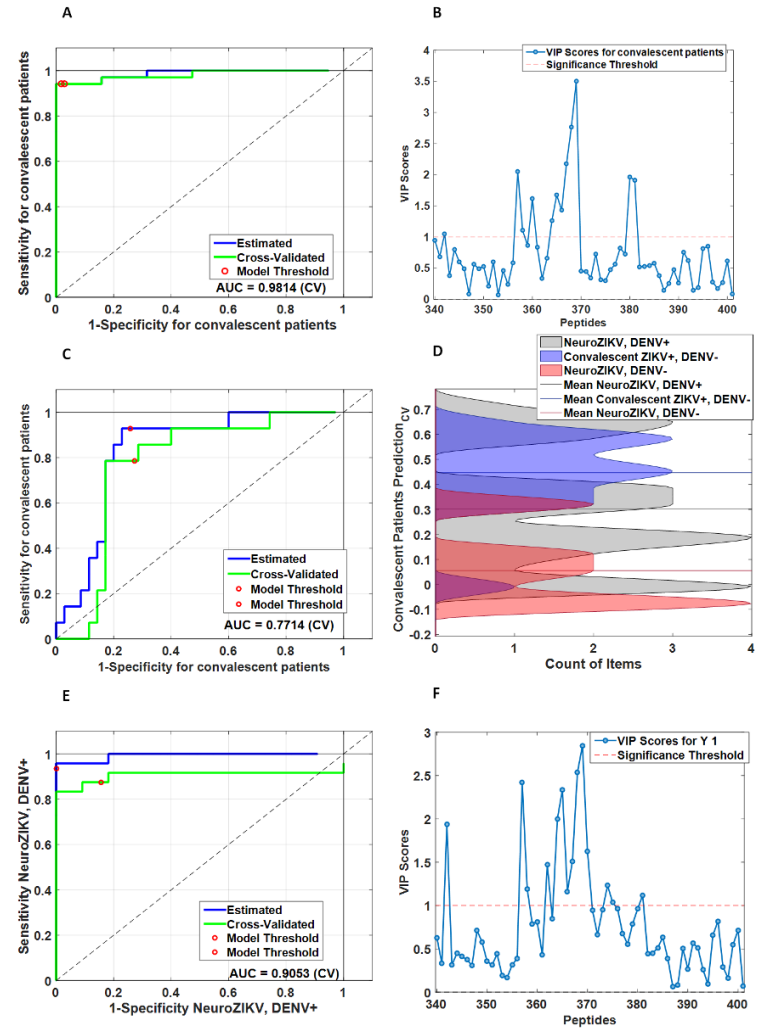


Figure 3. Summary of the additional PLS-DA models. (A) ROC curve and (B) VIP scores plot of PLS-DA to discriminate NeuroZIKV, DENV- from non-severe ZIKV patients (group VII vs groups IV, V and VI); (C) ROC curve and (D) Distribution classification histogram of patients on validation step for PLS-DA model to discriminate non-severe ZIKV, DENV- (group IV) from NeuroZIKV (groups VII, VIII) patients; (E) ROC curve and (F) VIP scores plot of PLS-DA to discriminate NeuroZIKV, DENV- from NeuroZIKV, DENV+ patients (group VII vs group VIII).

The ROC curve provided an AUC of 0.77 and model sensitivity and specificity for NeuroZIKV class were 71% and 78%, respectively, reflecting a model conflict to find a proper discrimination function. During model optimization step the patients' membership is predicted as a value that range from 0 to 1, with a threshold of 0.33 (defined by maximizing sensitivity and specificity). This means that patients that were predicted with a value higher than 0.33 belong to convalescent class, while values lower than that correspond patients not assigned to convalescent class. It is possible to observe that NeuroZIKV DENV+ patients (grey curve) are predicted along the whole range, while non-severe ZIKV, DENV- patients (blue curve) are systematically above the threshold and NeuroZIKV DENV- patients below the threshold. This result shows a large overlap of the three distributions, particularly NeuroZIKV DENV+ over the other two, which are partially separated from each other. This means that NeuroZIKV patients that were previously exposed to DENV also show significant fluorescence signal and cannot be unequivocally discriminated from non-severe ZIKV+, DENV- patients, in the current dataset. On the other hand, a model built considering only NeuroZIKV patients shows that NeuroZIKV DENV+ are statistically different from NeuroZIKV DENV- cases, and the latter has significantly lower responses against the abovementioned NS2B epitopes (Figure 3E and 3F). Those findings suggest that neurological manifestations of ZIKV infection are associated with reduced antibody response towards the identified NS2B epitope, except when DENV antibodies are present. Corroborating with this finding, microscale thermophoresis (MST) experiments have shown that the serum antibody half-maximal binding parameter (EC50) from the NeuroZIKV samples was remarkably low (EC50 values ca. 10^{-3} Molar), which were comparable to the background noise observed in the negative control samples (Supporting Figures S2A and S2B).

Molecular basis of the interaction between the identified NS2B epitope and ZIKV/DENV antibodies

Sequence comparison between the identified NS2B epitope and other flaviviruses showed that this region in the NS2B ZIKV protein presents fundamental differences in surface charge (Figure 4A). ZIKV exhibits a positively charged residue (K) in a position where all other flaviviruses exhibit a negatively charged residue (D or E) and an N instead of A, V, or S. In fact, differential charge distribution on the surface of ZIKV and DENV NS1 proteins has previously been documented and suggested as a candidate mechanism to be explored in the query for viral discrimination[29]. Experimental measurements of biomolecular association have shown that hydrogen bonds involving individual charged groups can contribute with up to 4 kcal mol^{-1}. This value is considerably lower in uncharged interactions with around 0.5 to 1.5 $kcal.mol^{-1}$. Residues of opposite charge in the ZIKV and DENV epitopes should therefore be central in the discrimination. Interestingly, if charge was assumed to be the main driving force behind discrimination between ZIKV and DENV proteins and related immune response, the expected values for energy association would translate in roughly one order of magnitude binding constant. Microscale thermophoresis (MST) experiments were conducted and have determined the difference in half-maximal binding parameter (EC50) between IgG antibodies in human serum and the identified epitope (Supporting Figures S2C and S2D) was precisely that as determined computationally. However, the measured affinity is more likely to be an apparent affinity (defined here as half-maximum binding parameter given the lack of a 1:1 ratio IgG-peptide) as a total serum polyclonal IgG were included in the experiment. Moreover, despite the punctual charge difference, the overall similarity between ZIKV and DENV NS2B epitope is ca. 62%. Since antibodies against this region appear to be elicited in both DENV and ZIKV non-severe infections, cross-reactivity with non-specific DENV antibodies could be expected, as evidenced in the above PLS-DA models.

It must be kept in mind that antibody responses from a peptide may be suboptimal, as the conformation of peptides may not necessarily reflect of that sequence in the native protein. Peptide flexibility is well known to compromise binding strength and enhance cross-reactivity, which could potentially explain the signals from NeuroZIKV patients previously infected by DENV. In fact, using a less sensitive technique, such as Enzyme-Linked ImmunoSorbent Assay (ELISA), than the peptide array, we have attempted to screen a larger number (N = 42) of serum samples from patients with ZIKV-associated neurological disease using the NS2B peptide to support the finding that IgG antibody response towards the identified epitope is not found in those individuals. However, despite assay optimization, the signal responses were as low as the background noise, which precluded further testing (data not shown for conciseness).

To understand the molecular-level features underlying the interactions between the identified NS2B epitope and the corresponding ZIKV antibodies, we performed sequence and structural analyses of the peptide, the latter by means of molecular dynamics simulations. From a structural point of view, a partial structure of ZIKV NS2B is found in complex with NS3 (Figure 4A), in which the N-terminal region (residues 49–67) of NS2B forms a β-strand packed within the N-terminal region of NS3.[35-38] Although the C-terminal part of NS2B (residues 68–96) also forms a β-hairpin, NMR results have indicated that the C-terminal region shows significant structural diversity, being highly flexible and often disordered.[36, 37] Mishra et al. have found that the lack of the residues K63, D64 drastically decreases the immunogenicity of the peptide.[39] Based on the high flexibility of the C-terminal region and on the finding of Mishra et al., the N-terminal portion of the peptide appears to be a promising site for antibody recognition.

Antigenic peptides exist in solution in a variety of conformations, and only conformations that are structurally similar to the one in the native protein will bind to the antibody.[40] To characterize the structural dynamics and conformational preferences of the 21-mer peptide based on NS2B address (comprising the N-terminal residues 57–GDITWEKDAEVTGNSPRLDVA-77 in the NS2B protein), a 1-μs molecular dynamics simulation was performed in solution. Structural stability was assessed by the positional root-mean-square deviation (RMSD) of all backbone atoms, as a function of simulation time (Figure 4B). The native crystal structure (PDB 5H6V[35]) was used as a reference structure. The peptide remains disordered and unstructured in the first half of simulation time, as evidenced by a large variation on RMSD values and the time-dependent profile of secondary structures (DSSP, Figure 4C). The RMSD converges to a value of approx. 1.28 nm after 0.6 μs, indicating that the peptide has assumed a preferred conformation. The molecule assumes a flexible conformation around a β-hairpin motif, where two β-strands are separated by a short loop (Figs 4C and 4D). The antigen-antibody recognition mechanism also requires shape complementarity;[41] therefore, highly flexible peptides are unable to efficiently bind to their respective antibodies.[42] A stable, close to its native-like conformation would allow the NS2B-based peptide to be recognized by the antibodies. The folded structure of the 21-mer peptide in solution, as evidenced by the simulation, shows that only parts of the N-terminal residues of the peptide are in a similar conformation as in its native counterpart (Figure S3), which can lower the ability of the peptide to bind to the antibody.

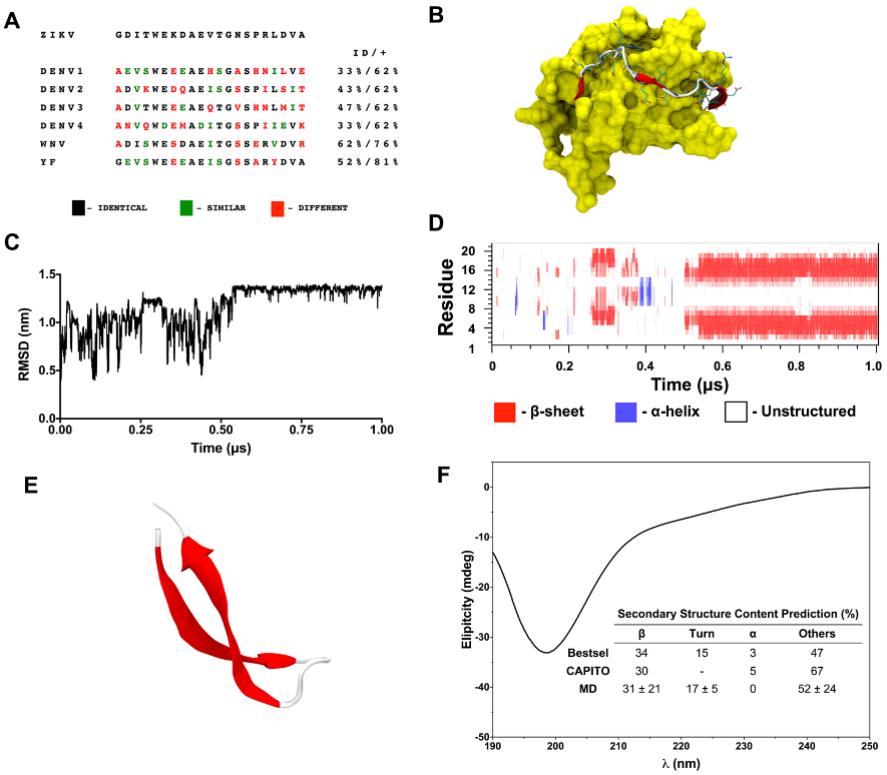


Figure 4. **Sequence and structural analyses of the identified NS2B peptide.** (A) Sequence comparison of the 21-mer peptide based on the ZIKV NS2B epitope with the most frequently occurring strains of DENV, WNV, and YFV (Residues are color-coded to highlight identity, chemical similarity, and differences. 'ID' stands for identity; a '+' indicates similarity degree). (B) Structure representation of ZIKV NS2B (cartoon and sticks) bound to NS3 (yellow surface) based on PDB ID 5H6V.[35] (C) Root-mean-square deviation of the NS2B-based peptide as a function of simulation time. The same region in PDB ID 5H6V[35] was used as a reference. (D) Secondary structure content of the peptide as a function of simulation time. (E) Representative structure of the folded peptide in cartoon (β strands are shown in red and unstructured regions in white). (F) Far-UV circular dichroism (CD) spectra of the NS2B peptide in acetate buffer collected at 25 ºC. Predicted secondary structure content based on experimental data is shown in the graph.

To further investigate the secondary structure content of the peptide, we have acquired the experimental circular dichroism (CD) spectrum (Figure 4E). The CD graph shows a strong negative band near 200 nm, characteristic of disordered polypeptides.[43] Moreover, to obtain a detailed structure information from the CD spectrum, we predicted the secondary structure components of the peptide using two different algorithms: BeStSel[44] and CAPITO[45]. Both algorithms predicted that the peptide's secondary structure is composed by ca. 30% of β-sheet and

ca. 70% of unstructured motifs (Turn + Others) (Figure 4E, inside table). This prediction is in agreement to the MD data, especially regarding β-sheet content (Figure 4E, inside table). A peptide in solution is characterized by an ensemble of conformers with different contents of secondary structure and the experimental CD spectrum represents an average of the contributions from different conformers.[46] Although the MD simulation suggests that the peptide may adopt a preferred β-hairpin conformation, the variation in its secondary content along the trajectory shows a significant level of conformational disorder. These results, along with the fact that the peptide exhibits a CD spectrum typical of disordered proteins, would imply that the unfolded, rather than the folded, conformers determine the shape of the CD spectra.[43] Together, the findings from MD and CD analyses support that the peptide in solution adopts a non-native like conformation, leading to a suboptimal affinity by specific ZIKV antibodies and explaining why the signal responses were not significantly different from background noise in the ELISA assays and the inability of the peptide array to fully discriminate NeuroZIKV patients with previous DENV infections from NeuroZIKV/DENV-.

Engineering an antigen carrying the NS2B epitope and absence of anti-NS2B IgG antibodies in mother-babies paired samples

Aiming to overcome the lack of sensibility and specificity of the NS2B peptide due to its high flexibility, we have performed the computational engineering of a synthetic, soluble and stable protein containing the target epitope. The new protein was designed to allow the exposure of the epitope in its native conformation, allowing its recognition by specific antibodies.

Based on Mishra et al.[39] and our findings, we chose to use the residues 62-EKDAEVTGNSPRL-74 of the NS2B protein as the epitope core. The structural motif of the NS2B epitope was aligned to the structure of 1250 proteins, yielding one structure with low RMSD and high exposure of the epitope. The scaffold used was the Mja10b protein (PDB ID: 1NH9),[47] a DNA-binding protein from the hyperthermophilic organism *Methanocaldococcus jannaschii*. In order to ensure absence of cross-reactivity in the following immunoassays, a BLASTp alignment of the engineered protein was performed against the entire ZIKV polyproteome showing no identity or even similarity to other regions than the carried-out epitope (data not shown for conciseness). The epitope was grafted onto the 3rd β-sheet (Figure 5A, highlighted in yellow) and after the redesign of the chimeric protein (NS2B-Grafted) the identity relative to native Mjab10 was 69%. An improvement of $\Delta\Delta G$ = -13,819 REU (Rosetta Energy Units) in the free energy of NS2B-grafted relative to the native one was also achieved, suggesting that the designed protein is a thermodynamically more stable structure. Subsequently, the structural stability of the designed protein was evaluated by means of a 100 ns MD simulation. The native scaffold was also evaluated for comparison purposes. The RMSD for the backbone atoms of both NS2B-grafted and native scaffold proteins relative to their respective initial structures were calculated over the entire trajectory (Figure 5B). While the native scaffold showed high RMSD values after 40 ns of simulation time (Figure 5B, black curve), the NS2B-grafted protein showed values around 0.2 nm throughout the trajectory (Figure 5B, red curve). This result suggests that the NS2B-grafted is structurally more stable, due to the favorable interactions designed during the modeling stage. The conformational stability of the epitope was also evaluated by comparing the epitope backbone in the chimera (Figure 5B, blue curve) with its native structure (Figure 4A). The RMSD of the epitope backbone converges to an average value of 0.17 nm, suggesting a native-like conformation. This similarity can be seen when the native motif is aligned to the epitope in the NS2B-grafted (Figure 5A, cyan structure). The mean quadratic fluctuation (RMSF) of the backbone atoms was also evaluated over the entire trajectory (Figure 5C). For NS2B-grafted protein the RMSF values remained low (under 0.3 nm), including the region where the epitope was allocated (Figure 5C, highlighted in yellow). The regions with greatest atomic fluctuations observed correspond to regions of loops and turns, which are expected to be intrinsically disordered. The low average values of RMSD and RMSF are indicative that the presence of the epitope does not destabilize the scaffold and the designed NS2B-grafted maintains its native-like structure. The solvent-accessible surface area (SASA) for the epitope residues was also calculated during the trajectory and compared with the exposure in the native structure (Figure 5D). Our data suggests that the epitope is exposed, similarly to its the native structure, and available for antibody recognition. Moreover, this finding includes residues K63 and D64 found by Mishra et al. to be critical for antigen-antibody binding.[39] In agreement with the RMSD and RMSF data, the DSSP for the NS2B-grafted protein indicates that all secondary structures were well maintained throughout the simulation (Figure 5E), including the epitope region (Figure 5E, highlighted in yellow).

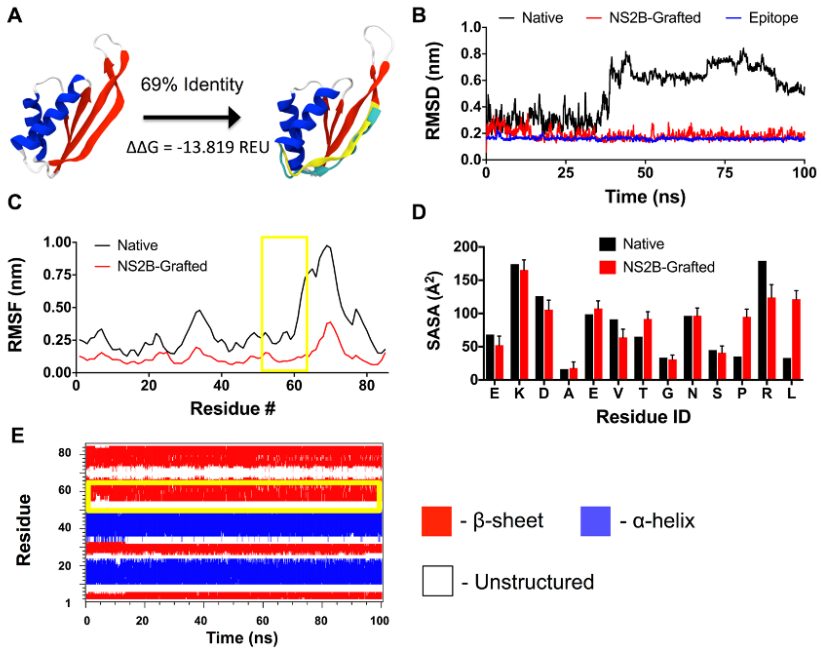


Figure 5. Structural characterization *in silico* of the engineered protein carrying the identified NS2B peptide. (A) Native structure of the scaffold protein (left) and structure of the NS2B-Grafted protein after design/molecular dynamics. The percentage indicates the sequence identity between the designed protein and the native scaffold. ΔΔG stands for the respective change of free energy of folding after the design protocol in Rosetta units (REU). (B) RMSD analysis for the backbone atoms for the Native scaffold (black), for the NS2B-Grafted designed protein (red) and comparison of the epitope with its native structure (blue). (C) Flexibility analysis using the RMSF values for the backbone atoms of the Native scaffold (black) and NS2B-Grafted (red). The region where the epitope was grafted is highlighted in yellow. (D) Solvent accessibility surface (SASA) indicating the exposure per residue of the epitope in its native structure (black) and in the designed protein (red). (E) Time-dependent secondary structure map indicating the maintenance of structural motifs during the MD simulation.

Following the engineering of a protein closely mimicking the natural structure of the NS2B epitope (Supplemental Figure S4), presence of IgG towards the NS2B epitope was assessed in a subset of 21 paired samples from mothers who delivered babies with ZIKV-associated microcephaly and their children through ELISA. In agreement with the computational analysis, the NS2B-grafted protein exhibited enhanced immunoreactivity when compared to the NS2B peptide from the peptide array and ELISA (latter not shown for conciseness) and allowed a clear distinction between negative and positive control samples (Figure 6A). Corroborating our findings, the anti-NS2B IgG antibody profile can be used to distinguish non-severe from severe (NeuroZIKV) infections, while the analysis of the paired samples confirmed that this IgG response is absent in the serum from mothers who delivered babies with ZIKV-associated microcephaly and their progeny (Figure 6B). On the other hand, the role of the absence of such antibodies in the development of neurological outcomes is beyond the scope of this work, and therefore remains as an open question.

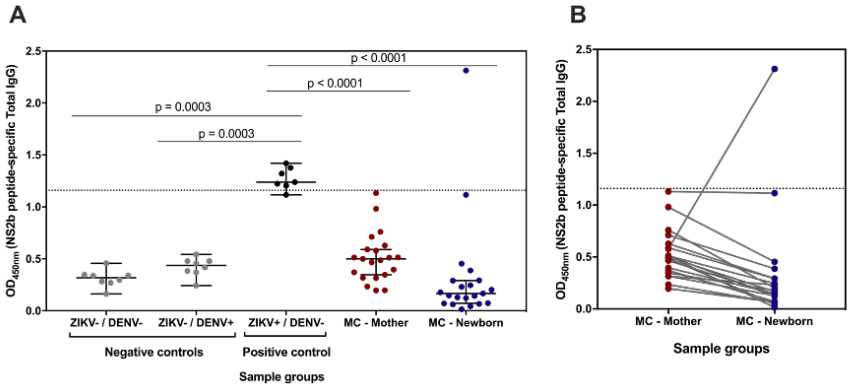


Figure 6. Detection of total IgG antibodies against the NS2B epitope engineered in the NS2B-Grafted protein in paired samples from mothers and babies with ZIKV-associated microcephaly through ELISA. (A) Scatter plot representation of NS2B-specific IgG antibodies detection in serum samples from ZIKV naïve (ZIKV-/DENV- (n=8) and ZIKV-/DENV+ (n=8), negative controls) individuals, ZIKV+/DENV- individuals (n=7, positive controls) and mothers who delivered babies with ZIKV-associated microcephaly and their progeny (n=21 pairs). The results are shown as mean values over all measurements and corresponding standard error of the mean (S.E.M.). Dots represent individual measurements. A dashed line indicates the median of the positive control group for comparison purposes. The p-values of two-tailed Mann-Whitney test are indicated above each group comparison. (B) Comparison of the detection of anti-NS2B-IgG in mother-children paired samples. Each sample was tested in duplicate and the mean values are shown. A dashed line indicates the assay cut-off as calculated from a ROC curve (Figure S5). Gray lines connect paired samples. MC stands for 'Microcephaly'.

Conclusions

We screened well-characterized serum samples from ZIKV-infected individuals in the Northeastern region of Brazil using high-density peptide arrays displaying the whole proteome of ZIKV. One linear epitope (composed of four adjacent and partially overlapping peptide sequences) from the ZIKV NS2B protein was identified as a ZIKV-specific sequence. We found that ZIKV infections elicited high-affinity IgG antibody response against the identified epitope especially in non-severe ZIKV cases, although previous DENV infections may elicit cross-reactive IgG response against the identified epitope, when presented as a peptide.

Furthermore, our data analysis has shown that a decreased antibody response to this epitope was associated with severe neurological manifestations after ZIKV infection (NeuroZIKV) compared with uncomplicated ZIKV infections. Interestingly, unlike what was observed in the uncomplicated ZIKV cases, where the IgG antibody response did not show any significant difference regardless of DENV background, the NeuroZIKV samples from patients previously infected with Dengue (NeuroZIKV / DENV+) showed higher IgG antibody responses against NS2B when compared to NeuroZIKV / DENV- samples. Nevertheless, even though our data indicates that DENV infection also induces cross-reactive IgG antibodies against the NS2B epitope, the global classification model highlighted the important contribution of peptides containing the amino acid sequence KDAEV(I)TGNS. Therefore, the IgG antibody response towards this sequence itself is capable to discriminate between the non-severe ZIKV and NeuroZIKV infection, regardless of their DENV history, with sensitivity and specificity rates of 79% and 85%, respectively.

Our computational and CD analysis for the NS2B peptide in solution indicate that conformational flexibility as the limiting factor for an unambiguously discrimination of NeuroZIKV patients, as this leads to a lower affinity by specific ZIKV antibodies as well as cross-reaction with non-specific DENV antibodies. Nevertheless, based on the native peptide motif, we designed a synthetic protein that mimics the natural structure of the NS2B epitope. Using this new NS2B-based antigen in an ELISA assay, we have shown that synthetic protein containing the NS2B epitope can be used as a molecular marker for discrimination between severe and non-severe ZIKV clinical phenotypes.

Experimental

Patient population

Sixty-eight (68) cryopreserved serum samples were collected from individuals with acute febrile illness enrolled in a prospective cohort study from May 2015 to May 2017 established in Recife, Brazil as part of the 'International Research Consortium on Dengue Risk Assessment, Management, and Surveillance' – IDAMS (ethics committee protocol number: 15580013.5.1001.5534).[48-50, ‡] The age of the individuals from whom serum samples were collected ranged from 9 to 57 years. Sample collection was performed on the first day of recruitment (day 1 – acute sample), corresponding to the period within the first 72 hours of the febrile phase, and during the convalescent phase (days 10 – 30 after recruitment – convalescent sample). Furthermore, 35 NeuroZIKV patients enrolled in the prospective cohort study 'Immunopathology of Dengue infection: viral and host genetic determinants associated to Dengue pathology' – PRONEX (ethics committee protocol number: 0009.0.095.000-11) were identified around day 1–14 after the onset of neurological symptoms. Nine of the selected samples were collected after this period, between day 60 and 180 after the onset of neurological symptoms. All samples were classified as convalescent samples, according to clinical and serological data. Moreover, 23 cryopreserved serum samples were obtained from subjects enrolled in the clinical cohort 'Prospective study for identification of clinical signs, virological and immunological factors that are predictive of severe Dengue' – Severe Dengue (ethics committee protocol number: 28309414.9.3001.5201) established in the Department of Virology, Aggeu Magalhães Institute, Oswaldo Cruz Foundation – FIOCRUZ, Recife, Brazil. Additionally, 21 paired samples from mothers who delivered babies with microcephaly and their progeny were obtained from subjects enrolled in the cohort study 'Neurological

manisfestations associated to Zika virus' (ethics committee protocol number: 51106115.8.0000.5190) established in the Department of Virology, Aggeu Magalhães Institute, Oswaldo Cruz Foundation – FIOCRUZ, Recife, Brazil. We also included naïve serum samples collected from European donors who had never travelled to tropical areas or who had visited Brazil for the first time. The characteristics of the patients are shown in Supporting Tables S1 and S2, while detailed information for each individual NeuroZIKV sample is described on the Supporting Table S3. Detailed description of the paired samples from mothers who delivered babies with ZIKV-associated microcephaly and their progeny is provided on Supporting Table S4.

Diagnostic testing of participants

All serum samples were tested following the procedures described above. Samples that, according to the molecular, serological and PRNT tests, were positive only for ZIKV were classified as ZIKV+/DENV-, while samples negative for all ZIKV tests and positive for at least one DENV serotype were classified as ZIKV-/DENV+. Samples that were negative for all tested viruses were classified as ZIKV-/DENV-. All patients were either DENV-confirmed or ZIKV-confirmed, or both. For molecular viral diagnosis, quantitative real-time PCR (qRT-PCR) for DENV and ZIKV was performed following modified previously reported protocols.[16, 51] As for serology, samples were assayed for anti-DENV IgM and IgG (Panbio Dengue Capture ELISA) and anti-ZIKV IgM (MAC-ELISA, Centers for Disease Control and Prevention [CDC])[52] through ELISA. Additionally, all samples were re-tested by indirect DENV ELISA and plaque reduction neutralization test (PRNT) for ZIKV as well as for each of the four DENV serotypes (DENV1–4) to rule out the possibility of undetected previous DENV exposure. Within the context of Northeast Brazil, most adults had previously been exposed to DENV, with seroprevalence rates ranging between 74.3% and 91.1%, also depending on socio-economic strata.[13] According to the assay results, each sample was classified as: (1) ZIKV+/DENV- if positive for ZIKV qRT-PCR and negative for DENV qRT-PCR in the acute phase and/or positive for anti-ZIKV IgM with titers > 2 times those for anti-DENV IgM in the convalescent phase in the MAC-ELISA, and negative for anti-DENV IgG in the acute phase in the Panbio assay; (2) ZIKV-/DENV- if negative for DENV and ZIKV qRT-PCRs in the acute phase, negative for anti-DENV and anti-ZIKV IgM in the acute phase in MAC-ELISA, and negative for anti-DENV IgG in both acute and convalescent phases in the Panbio assay [50]. The Panbio Dengue IgG Indirect ELISA was only performed for the adults and control subgroups. The assessment of virus-specific neutralizing antibodies was performed by PRNT.[50] Briefly, PRNT assays were performed in Vero cells with locally isolated DENV1–4 and ZIKV strains. PRNT positivity was calculated based on a 50% reduction in plaque counts ($PRNT_{50}$). Samples were considered positive when neutralizing antibody levels were ≥1:100 for ZIKV and ≥1:20 against at least one DENV serotype. Specific antibody titers for ZIKV and DENV were estimated using a four-parameter non-linear regression model. Acute and convalescent samples were assessed in parallel, and a convalescent/acute titer ratio >4 indicated acute infection. Important to note that all of the NeuroZIKV samples used in this study has presented anti-ZIKV IgG positive results through ZIKV IgG ELISA and ZIKV PRNT.

Peptide array technology

Whole proteomes of 15 different ZIKV strains (S1 Fig) were retrieved from the National Center for Biotechnology Information (NCBI) database and translated into overlapping 15-mer peptides (with a peptide-peptide overlap of 12 for DENV and 14 for ZIKV). The accession codes for the ZIKV proteomes from the NCBI were: ZikaUgandaStrainMR766-NIID, ZikaUgandaStrainMR766, ZikaChinaStrainVE_Ganxian, ZikaPhilStrainCPC-0740, ZikaThaiStrainSV0127-14, ZikaChinaStrainZJ03, ZikaHaitiStrain1225/2014, ZikaBrazilStrainSPH2015, ZikaBrazilStrainNatalRGN, ZikaDomRepStrainPD2, ZikaDomRepStrainPD1, ZikaChinaStrainGD01, ZikaBrazilStrainZKV2015, ZikaBrazilStrainSSABR1, and ZikaPuertoRicoStrainPRVABC59. The peptide sequences were aligned, duplicates deleted, and produced as high-density peptide arrays. In addition to that, a c-myc peptide (EQKLISEEDL) and a peptide derived from poliovirus (KEVPALTAVETGAT, reacts with serum due to comprehensive vaccination) were included as controls.

The arrays were generated in a laser-printer-based approach by PEPperPRINT GmbH (Heidelberg, Germany). Twenty different amino acid particle types were selectively deposited as spots on a glass slide coated with a poly(ethylene glycol) methacrylate/poly(methyl metacrylate) graft copolymer functionalized with a βAla-Asp-βAla linker, with each particle type embedding a different amino acid. When a layer of all twenty different particle types was completed, the particles were melted at 90 ºC, releasing the amino acids containing Fmoc-amino acid pentafluorophenyl esters that coupled to reactive groups on the functionalized glass slide. Following Merrifield's principle of orthogonal peptide synthesis, excess material was washed away, and the particle-deposition process was repeated.[32] In the end, each spot on the glass slide featured a different peptide with a freely chosen sequence.

Immunostaining of peptide arrays

Peptide microarrays were placed in incubation trays (PEPperPRINT GmbH; Heidelberg, Germany) and blocked for 30 minutes at room temperature under orbital shaking (140 rotations per minute [rpm]) with blocking buffer MB-070 (Rockland). Human sera were then diluted to 1:1000 in assay buffer (10% Rockland blocking buffer in phosphate-buffered saline (PBS), 0.05% Tween 20, pH 7.4) and incubated for 16 hours at 4 °C under orbital shaking (140 rpm). Peptide microarrays were washed three times with PBS supplemented with 0.05% Tween 20 (PBS-T), followed by incubation with the detection antibodies, goat anti-human IgG Fc specific DyLight 680 conjugated and goat anti-human IgM antibody DyLight 800 conjugated, diluted to 1:2500 and 1:5000, respectively in assay buffer (PBS-T with 10% Rockland blocking buffer), along with a 1:150 dilution of the control antibody (anti-c-myc antibody, PEPperPRINT, Germany), for one hour at room temperature under orbital shaking (140 rpm). No specific antibodies were added to interact with the poliovirus peptide control spots, since poliovirus vaccination in Brazil has been established across the country. The peptide microarrays were washed three times with PBS-T and rinsed with deionized water. After drying in a stream of air, fluorescence images were acquired using an Odyssey Imaging System (LICOR) at 700 and 800 nm, with a resolution of 21 µm and a scanning sensitivity of 7. Image analysis and quantification were performed with the PepSlide Analyzer software (Sicasys Software GmbH; Heidelberg, Germany). Background fluorescence was determined as the fluorescence measured when only the secondary antibodies anti-IgG DyLight 680 (1:2500) and anti-IgM DyLight 800 (1:5000) were added.

Statistical analysis

In the present work, multivariate analysis models were built to verify statistical differences among fluorescence intensities of patient groups. Before building multivariate models, pre-processing techniques were employed to correct unwanted variation inherent to the measurement. The data set was pre-processed using normalization to the maximum value of each patient for all variables (peptide fluorescence signals), baseline correction (Automatic Whittaker Filter with asymmetry of 0.001 and lambda of 100) and mean center. Multivariate data analyses were carried out with the IgG ZIKV data set in two forms. Initially, the exploratory analysis technique principal component analysis (PCA)[53] was employed to visualize data variance in a space of reduced dimensionality. Afterwards, a classification technique based on partial least squares discriminant analysis (PLS-DA)[54] was performed and the variable importance in projection (VIP scores) plots were employed to evaluate the most important NS2b peptides to discriminate between patients according their clinical outcomes. Due the small number of samples available and to include maximum variability to the model, leave-one-out cross-validation method was employed to validate and determine model complexity based on the true positive rate (or sensitivity (Sn)) and the true negative rate (or specificity (Sp)). All statistical analyses were performed in PLS_Toolbox 8.6 (Eigenvector Research Inc., USA) in MATLAB environment.

Engineering an antigen containing the ns2b epitope

The native motif of the NS2B epitope was retrieved from the crystallographic structure deposited in the Protein Data Bank (PDB ID: 5H6V).[35] The epitope was transplanted, *in silico*, to a carrier protein (scaffold), in order to maximize its exposure on its native conformation. This procedure was performed using the MotifGraft tool[55] present in the Rosetta v3.10 software.[56, 57] First, a library of scaffold proteins, composed by 1250 crystallographic structures, was assembled according to criteria described by Silva *et al*.[55] Then, the NS2B motif was structurally aligned to all proteins in the library aiming to determine which is the best carrier and region for transplantation. To accept an alignment, the root mean-square deviation (RMSD) between the backbone of the epitope and the scaffold protein should be lower or equal to 2.5 Å. The epitope side-chains are transplanted to the scaffold protein if the alignment criterion is met. After transplantation, the carrier protein is redesigned within a 10 Å radius around the epitope intending to increase favorable interactions and ensure the folding of the new protein. In order to allow only mutations that will result in a total stabilizing effect, a position-specific substitution matrix (PSSM) was used,[58] according to instructions described by Goldenzweig *et al*.[59] PSSM represents the logarithmic probability of looking any of the 20 possible amino acids in each position. Mutations for the most frequently observed amino acids generally increase protein stability (the consensus effect), whereas mutations never previously identified in natural diversity tend to decrease stability.[59-62] During the redesign process, combinatorial sequence optimization is carried out, with four iterations of sequence design, each one followed by side-chain and backbone minimization. At the end, the lowest energy structure is obtained according to the REF2015 energy function of the Rosetta.[63]

Computational characterization of structural and thermodynamic properties of the identified NS2B peptide and the designed NS2B-grafted antigen

Initial coordinates for the NS2B-based peptide were taken from the crystal structure of the NS2B-NS3 complex (PDB ID 5H6V) and centered into a cubic box of dimensions 8.0 x 8.0 x 8.0 nm. For the modeled antigen carrying the NS2B

epitope, the coordinates were taken from the lowest energy decoy from Rosetta design protocol and centered in a box of dimensions 12 x 12 x 12 nm. All simulations were carried out by explicitly solvating the biomolecules using the SPC water model.[64] Periodic boundary conditions were used in the x, y and z directions. Molecular dynamics simulation was carried out using the NPT ensemble, and all atomistic simulations were performed for 1 μs using the GROMOS 53A6 force field.[65] The LINCS method was used to constrain all bonds.[66] Solute and solvent temperatures were separately coupled to a thermostat using the velocity rescaling scheme[67] with a reference temperature of 310 K and a relaxation time of 0.2 ps. Pressure was kept constant at 1.013 bar by isotropic coordinate scaling using the Berendsen barostat[68] with a relaxation time of 0.2 ps and compressibility (κ_T) of 4.5 x 10^{-5} bar^{-1}. A short-range cut-off radius of 1.4 nm was used for all non-bonded interactions. Long-range electrostatic interactions were taken into account using the reaction field method[69] with ε=66 beyond the cut-off radius of 1.4 nm. Counter ions were added to ensure a neutral system net charge and a salt concentration of 0.150 mM. The systems were first energy-optimised using 5,000 steps of the steepest descent algorithm. Integration was carried out using an integration time step of 2 fs based on the leapfrog algorithm.[70] System thermalization was carried out in NVT ensemble for 100 ps, followed by a pressure equilibration in an NPT ensemble for 500 ps. All equilibration steps were performed with position constraints for all atoms. All simulations and analyses were performed using the GROMACS 4.6.5 simulation package.[71] The most representative conformer of each simulated biomolecule was selected by structural clustering analysis. The clusters were obtained using the GROMOS method[72] implemented in the *g_cluster* GROMACS tool. A cut-off of 0.2 nm was used for calculating the root-mean-square deviations (RMSD) of superimposed backbone and β-carbon atoms. For the peptide only the last 550 ns period of the trajectory was used for clustering, whereas for the modeled antigen containing the NS2B epitope the last 75 ns was used. Visual analyses were carried out using the Visual Molecular Dynamics (VMD) software version 1.8.7.[73]

Circular dichroism

The NS2B peptide was diluted in acetate buffer (0.5 mg/mL) to the concentration of 60μM. Circular dichroism data were collected on an Olis DSM17 Spectrometer. Scans were recorded in a 0.1 cm path length. Circular dichroism scans were carried out from 260 to 190 nm with 5 seconds averaging times, 1 nm step size, and 2 nm bandwidth at 25 °C. Spectra were corrected for a buffer blank and baseline molar ellipticity at 260 nm. Scan data were smoothed by the Stavistsky-Golay method.

Detection of anti-NS2B IgG antibodies through ELISA

The DNA sequence coding for the engineered NS2B-grafted protein was optimized for enhanced expression in prokaryotic systems, cloned into the pET14b (Novagen) expression vector and produced as a 6His-N-terminally tagged protein. The recombinant protein was purified by automated affinity chromatography followed by a size exclusion chromatography. High binding, half area 96-well polystyrene plates (Costar; Lowell, MA, USA) were coated overnight at 4 ºC with 1 μg/mL of the recombinant protein diluted in 0.2 M carbonate/bicarbonate buffer (Pierce, IL, USA). Plates were blocked with skimmed milk (Bio-Rad) at 5% (w/v) in PBS-T buffer [1X PBS with 0.05% (v/v) Tween 20] for 15 minutes at room temperature. Serum samples were diluted (1:50) in assay buffer [5% (w/v) skimmed milk in PBS-T] and added to plates. All samples were incubated for 2 hours at room temperature. Plates were washed five times with PBS-T, followed by incubation with the horseradish peroxidase (HRP)-linked antibody against total IgG (Jackson ImmunoResearch, 1:30 000 dilution) diluted in assay buffer for 1 hour at room temperature. After a second wash step, the reaction was developed by the addition of tetramethylbenzidine TMB-KPL substrate (Pierce, IL, USA) for 30 minutes at room temperature, followed by 1N HCl. Optical densities at a wavelength of 450nm (OD450nm) were read using a microplate spectrophotometer (BioTek; Winooski, VT, USA). A subset of ZIKV positive and Flavivirus naïve sera were used as positive and negative controls, respectively. The controls were run in each plate and used to determine the reproducibility of the assay.

Author Contributions

TJ, ETAM, RDL, IFTV and FFL designed the study. FFL and NF performed the peptide microarrays. DFC performed protein engineering and MD simulations. CSS and RDC performed detailed statistical analysis in the entire data set. IFTV and AFP produced the engineered protein and performed the CD measurements. IFTV, CMCSA and BHSL performed the MST and ELISA experiments. MTC, TM and CNLM provided the serum samples. IFTV, RDL and DFC wrote the manuscript with the input from all authors.

Conflicts of interest

A patent application for the epitope has been filed by TJ, NF, FFL, TM, and ETA. A patent application for the engineered protein carrying the identified NS2B epitope has been filed by RDL, DFC and IFTV. Otherwise, the authors have declared that they have no competing interests.

Acknowledgements

This study was supported by grants from the German Research Foundation (DFG) via the Heidelberg Karlsruhe Research Partnership (HEiKA to TJ and FFL), the European Commission (IDAMS FP7 281803, ZIKAlliance H2020 734548), the German Centre for Infection Research (DZIF), Heidelberg Site, CuraZika Foundation, FACEPE, NUQAAPE, CAPES, CNPq, INCT-FCx, FAPESP (2019/01255-9) and the German Federal Ministry of Education and Research (NanoMatFutur grant 13XP5050A). Computer allocation was partly granted by the Brazilian National Scientific Computing Center (LNCC). The company PEPperPRINT based in Heidelberg supported the design and printing of the peptide microarrays. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Notes and references

‡ Cohort Ethics Statement: the human serum samples reported by the authors in this article were provided by the 'International Research Consortium on Dengue Risk Assessment, Management, and Surveillance' – IDAMS, 'Immunopathology of Dengue infection: viral and host genetic determinants associated to Dengue pathology' – PRONEX, 'Prospective study for identification of clinical signs, virological and immunological factors that are predictive of severe Dengue' and 'Neurological manisfestations associated to Zika virus' established at FIOCRUZ, Recife, Brazil, after approval by the institutional ethical committee under the protocol numbers 15580013.5.1001.5534, 0009.0.095.000-11, 28309414.9.3001.5201, and 51106115.8.0000.5190, respectively.

1. N. R. Faria, R. Azevedo, M. U. G. Kraemer, R. Souza, M. S. Cunha, S. C. Hill, J. Theze, M. B. Bonsall, T. A. Bowden, I. Rissanen, I. M. Rocco, J. S. Nogueira, A. Y. Maeda, F. Vasami, F. L. L. Macedo, A. Suzuki, S. G. Rodrigues, A. C. R. Cruz, B. T. Nunes, D. B. A. Medeiros, D. S. G. Rodrigues, A. L. N. Queiroz, E. V. P. da Silva, D. F. Henriques, E. S. T. da Rosa, C. S. de Oliveira, L. C. Martins, H. B. Vasconcelos, L. M. N. Casseb, D. B. Simith, J. P. Messina, L. Abade, J. Lourenco, L. C. J. Alcantara, M. M. de Lima, M. Giovanetti, S. I. Hay, R. S. de Oliveira, P. D. S. Lemos, L. F. de Oliveira, C. P. S. de Lima, S. P. da Silva, J. M. de Vasconcelos, L. Franco, J. F. Cardoso, J. Vianez-Junior, D. Mir, G. Bello, E. Delatorre, K. Khan, M. Creatore, G. E. Coelho, W. K. de Oliveira, R. Tesh, O. G. Pybus, M. R. T. Nunes and P. F. C. Vasconcelos, *Science*, 2016, **352**, 345.
2. C. Brito, *Acta Med. Port.*, 2015, **28**, 679.
3. L. Schuler-Faccini, E. M. Ribeiro, I. M. Feitosa, D. D. Horovitz, D. P. Cavalcanti, A. Pessoa, M. J. Doriqui, J. I. Neri, J. M. Neto, H. Y. Wanderley, M. Cernach, A. S. El-Husny, M. V. Pone, C. L. Serao, M. T. Sanseverino and F. Brazilian Medical Genetics Society-Zika Embryopathy Task, *Morb. Mortal. Wkly. Rep.*, 2016, **65**, 59.
4. D. L. Heymann, A. Hodgson, A. A. Sall, D. O. Freedman, J. E. Staples, F. Althabe, K. Baruah, G. Mahmud, N. Kandun, P. F. Vasconcelos, S. Bino and K. U. Menon, *Lancet*, 2016, **387**, 719.
5. WHO, Zika Causality Statement, https://www.who.int/emergencies/zika-virus/causality/en/, (accessed 26.03.2019, 2019).
6. F. Krauer, M. Riesen, L. Reveiz, O. T. Oladapo, R. Martinez-Vega, T. V. Porgo, A. Haefliger, N. J. Broutet, N. Low and W. H. O. Z. C. W. Group, *PLOS Med.*, 2017, **14**, e1002203.
7. Thalia Velho Barreto de Araújo, Prof Ricardo Arraes de Alencar Ximenes, Demócrito de Barros Miranda-Filho, Wayner Vieira Souza, Ulisses Ramos Montarroyos, Ana Paula Lopes de Melo, Sandra Valongueiro, Maria de Fátima Pessoa Militão de Albuquerque, Cynthia Braga, Sinval Pinto Brandão Filho, Marli Tenório Cordeiro, Enrique Vazquez, Danielle di Cavalcanti Souza Cruz, Claudio Maierovitch Pessanha Henriques, Luciana Caroline Albuquerque Bezerra, Priscila Mayrelle da Silva Castanha, Rafael Dhalia, Ernesto Torres Azevedo Marques-Júnior, Prof Celina Maria Turchi Martelli, Prof Laura Cunha Rodrigues, o. b. of, i. f. t. M. E. R. Group, t. B. M. o. Health, t. P. A. H. Organization, I. d. M. I. P. F. Figueira and t. S. H. D. o. Pernambuco, *Lancet Infect. Dis.*, 2017, **18**, 328.
8. E. Dirlikov, J. V. Torres, R. B. Martines, S. Reagan-Steiner, G. V. Perez, A. Rivera, C. Major, D. Matos, J. Munoz-Jordan, W. J. Shieh, S. R. Zaki and T. M. Sharp, *Emerg. Infect. Dis.*, 2018, **24**, 114.
9. K. Dinnon, E. Gallichotte, E. Fritch, V. Menachery and R. Baric, *PLOS Negl. Trop. Dis.*, 2019, **13**, e0007212.
10. C. R. Fontes-Garfias, C. Shan, H. Luo, A. E. Muruato, D. B. A. Medeiros, E. Mays, X. Xie, J. Zou, C. M. Roundy, M. Wakamiya, S. L. Rossi, T. Wang, S. C. Weaver and P. Y. Shi, *Cell Rep.*, 2017, **21**, 1180.
11. L. Yuan, X. Y. Huang, Z. Y. Liu, F. Zhang, X. L. Zhu, J. Y. Yu, X. Ji, Y. P. Xu, G. Li, C. Li, H. J. Wang, Y. Q. Deng, M. Wu, M. L. Cheng, Q. Ye, D. Y. Xie, X. F. Li, X. Wang, W. Shi, B. Hu, P. Y. Shi, Z. Xu and C. F. Qin, *Science*, 2017, **358**, 933.
12. P. M. S. Castanha, E. J. M. Nascimento, C. Braga, M. T. Cordeiro, O. V. de Carvalho, L. R. de Mendonca, E. A. N. Azevedo, R. F. O. Franca, R. Dhalia and E. T. A. Marques, *J. Infect. Dis.*, 2017, **215**, 781.
13. C. Braga, C. F. Luna, C. M. Martelli, W. V. de Souza, M. T. Cordeiro, N. Alexander, F. de Albuquerque Mde, J. C. Junior and E. T. Marques, *Acta Trop.*, 2010, **113**, 234.

14. M. R. Duffy, T. H. Chen, W. T. Hancock, A. M. Powers, J. L. Kool, R. S. Lanciotti, M. Pretrick, M. Marfel, S. Holzbauer, C. Dubray, L. Guillaumot, A. Griggs, M. Bel, A. J. Lambert, J. Laven, O. Kosoy, A. Panella, B. J. Biggerstaff, M. Fischer and E. B. Hayes, *N. Engl. J. Med.*, 2009, **360**, 2536.
15. R. Buathong, L. Hermann, B. Thaisomboonsuk, W. Rutvisuttinunt, C. Klungthong, P. Chinnawirotpisan, W. Manasatienkij, A. Nisalak, S. Fernandez, I. K. Yoon, P. Akrasewi and T. Plipat, *Am. J. Trop. Med. Hyg.*, 2015, **93**, 380.
16. R. S. Lanciotti, O. L. Kosoy, J. J. Laven, J. O. Velez, A. J. Lambert, A. J. Johnson, S. M. Stanfield and M. R. Duffy, *Emerg. Infect. Dis.*, 2008, **14**, 1232.
17. I. Rodriguez-Barraquer, F. Costa, E. J. M. Nascimento, N. J. Nery, P. M. S. Castanha, G. A. Sacramento, J. Cruz, M. Carvalho, D. De Olivera, J. E. Hagan, H. Adhikarla, E. A. Wunder, Jr., D. F. Coelho, S. R. Azar, S. L. Rossi, N. Vasilakis, S. C. Weaver, G. S. Ribeiro, A. Balmaseda, E. Harris, M. L. Nogueira, M. G. Reis, E. T. A. Marques, D. A. T. Cummings and A. I. Ko, *Science*, 2019, **363**, 607.
18. W. Dejnirattisai, P. Supasa, W. Wongwiwat, A. Rouvinski, G. Barba-Spaeth, T. Duangchinda, A. Sakuntabhai, V. M. Cao-Lormeau, P. Malasit, F. A. Rey, J. Mongkolsapaya and G. R. Screaton, *Nat. Immunol.*, 2016, **17**, 1102.
19. L. Priyamvada, K. M. Quicke, W. H. Hudson, N. Onlamoon, J. Sewatanon, S. Edupuganti, K. Pattanapanyasat, K. Chokephaibulkit, M. J. Mulligan, P. C. Wilson, R. Ahmed, M. S. Suthar and J. Wrammert, *Proc. Natl. Acad. Sci. USA*, 2016, **113**, 7852.
20. J. A. Muller, M. Harms, A. Schubert, B. Mayer, S. Jansen, J. P. Herbeuval, D. Michel, T. Mertens, O. Vapalahti, J. Schmidt-Chanasit and J. Munch, *Med. Microbiol. Immunol.*, 2017, **206**, 175.
21. A. Sharma and S. K. Lal, *Front. Microbiol.*, 2017, **8**, 110.
22. A. Sakudo, A. Viswan, H. Chou, T. Sasaki, K. Ikuta and M. Nagatsu, *Mol. Med. Rep.*, 2016, **14**, 697.
23. M. Sanchez-Purra, M. Carre-Camps, H. de Puig, I. Bosch, L. Gehrke and K. Hamad-Schifferli, *ACS Infect. Dis.*, 2017, **3**, 767.
24. D. Focosi, F. Maggi and M. Pistello, *Clin. Infect. Dis.*, 2016, **63**, 227.
25. A. Chua, I. Prat, C. M. Nuebling, D. Wood and F. Moussy, *PLOS Negl. Trop. Dis.*, 2017, **11**, e0005269.
26. F. Breitling, A. Nesterov, V. Stadler, T. Felgenhauer and F. R. Bischoff, *Mol. Biosyst.*, 2009, **5**, 224.
27. C. Katz, L. Levy-Beladev, S. Rotem-Bamberger, T. Rito, S. G. D. Rudiger and A. Friedler, *Chem. Soc. Rev.*, 2011, **40**, 2131.
28. F. F. Loeffler, J. Pfeil and K. Heiss, *Methods Mol. Biol.*, 2016, **1403**, 569.
29. M. C. L. C. Freire, Pol-Fachin, L., Coêlho, D. F., Viana, I. F. T., Magalhães, T., Cordeiro, M. T., Fischer, N., Loeffler, F. F., Jaenisch, T., Franca, R. F., Marques, E. T. A., Lins, R. D., *ACS Omega*, 2017, **2**, 3913.
30. L. K. Weber, A. Palermo, J. Kugler, O. Armant, A. Isse, S. Rentschler, T. Jaenisch, J. Hubbuch, S. Dubel, A. Nesterov-Mueller, F. Breitling and F. F. Loeffler, *J. Immunol. Methods*, 2017, **443**, 45.
31. L. K. Weber, A. Isse, S. Rentschler, R. E. Kneusel, A. Palermo, J. Hubbuch, A. Nesterov-Mueller, F. Breitling and F. F. Loeffler, *Eng. Life Sci.*, 2017, **17**, 1078.
32. V. Stadler, T. Felgenhauer, M. Beyer, S. Fernandez, K. Leibe, S. Guttler, M. Groning, K. Konig, G. Torralba, M. Hausmann, V. Lindenstruth, A. Nesterov, I. Block, R. Pipkorn, A. Poustka, F. R. Bischoff and F. Breitling, *Angew. Chem. Int. Ed.*, 2008, **47**, 7132.
33. F. F. Loeffler, T. C. Foertsch, R. Popov, D. S. Mattes, M. Schlageter, M. Sedlmayr, B. Ridder, F. X. Dang, C. von Bojnicic-Kninski, L. K. Weber, A. Fischer, J. Greifenstein, V. Bykovskaya, I. Buliev, F. R. Bischoff, L. Hahn, M. A. Meier, S. Brase, A. K. Powell, T. S. Balaban, F. Breitling and A. Nesterov-Mueller, *Nat. Commun.*, 2016, **7**, 11844.
34. P. M. Castanha, M. T. Cordeiro, C. M. Martelli, W. V. Souza, E. T. Marques, Jr. and C. Braga, *Epidemiol. Infect.*, 2013, **141**, 1080.
35. Y. Li, Z. Zhang, W. W. Phoo, Y. R. Loh, W. Wang, S. Liu, M. W. Chen, A. W. Hung, T. H. Keller, D. Luo and C. Kang, *Structure*, 2017, **25**, 1242.
36. A. Roy, L. Lim, S. Srivastava, Y. Lu and J. Song, *PLOS One*, 2017, **12**, e0180632.
37. Z. Zhang, Y. Li, Y. R. Loh, W. W. Phoo, A. W. Hung, C. Kang and D. Luo, *Science*, 2016, **354**, 1597.
38. W. W. Phoo, Y. Li, Z. Zhang, M. Y. Lee, Y. R. Loh, Y. B. Tan, E. Y. Ng, J. Lescar, C. Kang and D. Luo, *Nat. Comm.*, 2016, **7**, 13410.
39. N. Mishra, A. Caciula, A. Price, R. Thakkar, J. Ng, L. V. Chauhan, K. Jain, X. Che, D. A. Espinosa, M. Montoya Cruz, A. Balmaseda, E. H. Sullivan, J. J. Patel, R. G. Jarman, J. L. Rakeman, C. T. Egan, C. Reusken, M. P. G. Koopmans, E. Harris, R. Tokarz, T. Briese and W. I. Lipkin, *MBio*, 2018, **9**.
40. D. H. Sachs, A. N. Schechter, A. Eastlake and C. B. Anfinsen, *Proc. Natl. Acad. Sci. USA*, 1972, **69**, 3790.
41. A. Datta-Mannan, A. Thangaraju, D. Leung, Y. Tang, D. R. Witcher, J. Lu and V. J. Wroblewski, *MAbs*, 2015, **7**, 483.
42. I. F. T. Viana, T. A. Soares, L. F. O. Lima, E. T. A. Marques, M. A. Krieger, R. Dhalia and R. D. Lins, *RSC Advances*, 2013, **3**, 11790.
43. V. Uversky and S. Longhi, *Instrumental Analysis of Intrinsically Disordered Proteins: Assessing Structure and Conformation*, Wiley, 2011.
44. A. Micsonai, F. Wien, É. Bulyáki, J. Kun, É. Moussong, Y.-H. Lee, Y. Goto, M. Réfrégiers and J. Kardos, *Nuc. Acids Res.*, 2018, **46**, W315.
45. C. Wiedemann, P. Bellstedt and M. Görlach, *Bioinformatics*, 2013, **29**, 1750.
46. A. Glättli, X. Daura, D. Seebach and W. F. van Gunsteren, *J. Am. Chem. Soc.*, 2002, **124**, 12972.
47. G. Wang, R. Guo, M. Bartlam, H. Yang, H. Xue, Y. Liu, L. Huang and Z. Rao, *Protein. Sci.*, 2003, **12**, 2815.

48. T. Jaenisch, Idams, A. Sakuntabhai, Denfree, A. Wilder-Smith and DengueTools, *PLOS Negl. Trop. Dis.*, 2013, **7**, e2320.
49. T. Jaenisch, D. T. Tam, N. T. Kieu, T. Van Ngoc, N. T. Nam, N. Van Kinh, S. Yacoub, N. Chanpheaktra, V. Kumar, L. L. See, J. Sathar, E. P. Sandoval, G. M. Alfaro, I. S. Laksono, Y. Mahendradhata, M. Sarker, F. Ahmed, A. Caprara, B. S. Benevides, E. T. Marques, T. Magalhaes, P. Brasil, M. Netto, A. Tami, S. E. Bethencourt, M. Guzman, C. Simmons, N. T. Quyen, L. Merson, N. T. Dung, D. Beck, M. Wirths, M. Wolbers, P. K. Lam, K. Rosenberger and B. Wills, *BMC Infect. Dis.*, 2016, **16**, 120.
50. T. Magalhaes, C. Braga, M. T. Cordeiro, A. L. S. Oliveira, P. M. S. Castanha, A. P. R. Maciel, N. M. L. Amancio, P. N. Gouveia, V. J. Peixoto-da-Silva, Jr., T. F. L. Peixoto, H. Britto, P. V. Lima, A. R. S. Lima, K. D. Rosenberger, T. Jaenisch and E. T. A. Marques, *PLOS Negl. Trop. Dis.*, 2017, **11**, e0006055.
51. R. S. Lanciotti, O. L. Kosoy, J. J. Laven, A. J. Panella, J. O. Velez, A. J. Lambert and G. L. Campbell, *Emerg. Infect. Dis.*, 2007, **13**, 764.
52. M. T. Cordeiro, C. A. Brito, L. J. Pena, P. M. Castanha, L. H. Gil, K. G. Lopes, R. Dhalia, J. A. Meneses, A. C. Ishigami, L. M. Mello, L. X. Alencar, K. M. Guarines, L. C. Rodrigues and E. T. Marques, *J. Infect. Dis.*, 2016, **214**, 1897.
53. R. Bro and A. K. Smilde, *Anal. Methods*, 2014, **6**, 2812.
54. R. G. Brereton and G. R. Lloyd, *J. Chemom.*, 2014, **28**, 213.
55. D.-A. Silva, B. E. Correia and E. Procko, in *Computational Design of Ligand Binding Proteins*, ed. B. L. Stoddard, Springer New York, New York, NY, 2016, pp. 285.
56. A. Leaver-Fay, M. Tyka, S. M. Lewis, O. F. Lange, J. Thompson, R. Jacak, K. W. Kaufman, P. D. Renfrew, C. A. Smith, W. Sheffler, I. W. Davis, S. Cooper, A. Treuille, D. J. Mandell, F. Richter, Y.-E. A. Ban, S. J. Fleishman, J. E. Corn, D. E. Kim, S. Lyskov, M. Berrondo, S. Mentzer, Z. Popović, J. J. Havranek, J. Karanicolas, R. Das, J. Meiler, T. Kortemme, J. J. Gray, B. Kuhlman, D. Baker and P. Bradley, in *Meth. Enzymol.*, eds. M. L. Johnson and L. Brand, Academic Press, 2011, vol. 487, pp. 545.
57. S. J. Fleishman, A. Leaver-Fay, J. E. Corn, E.-M. Strauch, S. D. Khare, N. Koga, J. Ashworth, P. Murphy, F. Richter, G. Lemmon, J. Meiler and D. Baker, *PLOS One*, 2011, **6**, e20161.
58. S. F. Altschul, E. M. Gertz, R. Agarwala, A. A. Schäffer and Y.-K. Yu, *Nuc. Acid Res.*, 2008, **37**, 815.
59. A. Goldenzweig, M. Goldsmith, Shannon E. Hill, O. Gertman, P. Laurino, Y. Ashani, O. Dym, T. Unger, S. Albeck, J. Prilusky, Raquel L. Lieberman, A. Aharoni, I. Silman, Joel L. Sussman, Dan S. Tawfik and Sarel J. Fleishman, *Mol. Cell*, 2016, **63**, 337.
60. B. Steipe, B. Schiller, A. Pluckthun and S. Steinbacher, *J. Mol. Biol.*, 1994, **240**, 188.
61. T. J. Magliery, *Curr. Opin. Struct. Biol.*, 2015, **33**, 161.
62. L. G. Nivon, S. Bjelic, C. King and D. Baker, *Proteins*, 2014, **82**, 858.
63. R. F. Alford, A. Leaver-Fay, J. R. Jeliazkov, M. J. O'Meara, F. P. DiMaio, H. Park, M. V. Shapovalov, P. D. Renfrew, V. K. Mulligan, K. Kappel, J. W. Labonte, M. S. Pacella, R. Bonneau, P. Bradley, R. L. Dunbrack, R. Das, D. Baker, B. Kuhlman, T. Kortemme and J. J. Gray, *J. Chem. Theory Comput.*, 2017, **13**, 3031.
64. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren and J. Hermans, in *Intermolecular Forces*, ed. B. Pullman, Springer Netherlands, 1981, vol. 14, ch. 21, pp. 331.
65. C. Oostenbrink, A. Villa, A. E. Mark and W. F. Van Gunsteren, *J. Compt. Chem.*, 2004, **25**, 1656.
66. B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, *J. Compt. Chem.*, 1997, **18**, 1463.
67. G. Bussi, D. Donadio and M. Parrinello, *The Journal of chemical physics*, 2007, **126**, 014101.
68. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *The Journal of chemical physics*, 1984, **81**, 3684.
69. I. G. Tironi, R. Sperb, P. E. Smith and W. F. van Gunsteren, *The Journal of chemical physics*, 1995, **102**, 5451.
70. R. W. Hockney, in *Methods in Computational Physics*, eds. B. Alder , S. Fernbach and M. Rotenberg, Academic Press, New York/London, 1970.
71. B. Hess, C. Kutzner, D. van der Spoel and E. Lindahl, *J. Chem. Theory Comput.*, 2008, **4**, 435.
72. X. Daura, B. Oliva, E. Querol, F. X. Aviles and O. Tapia, *Proteins*, 1996, **25**, 89.
73. W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graph.*, 1996, **14**, 33.

SUPPLEMENTAL INFORMATION

Identification of a Zika NS2B epitope for which absence of IgG response is associated with severe neurological symptoms and the design of a biomarker capable of discriminatory diagnostics between severe and non-severe clinical phenotypes

Felix F. Loeffler[a,¶], Isabelle F.T. Viana[b,¶], Nico Fischer[c], Danilo F. Coêlho[b,d], Carolina S. Silva[e], Antônio F. Purificação Jr.[b], Catarina M.C.S. Araújo[b], Bruno H.S. Leite[b], Ricardo Durães-Carvalho[f], Tereza Magalhães[g], Clarice N.L. Morais[b], Marli T. Cordeiro[b], Roberto D. Lins[2,&], Ernesto T.A. Marques[b,h,&] and Thomas Jaenisch[c,i,&*]

a. Max Planck Institute of Colloids and Interfaces, Department of Biomolecular Systems, Potsdam, Germany.
b. Department of Virology, Aggeu Magalhães Institute, Oswaldo Cruz Foundation, Recife, PE, Brazil.
c. Section Clinical Tropical Medicine, Department of Infectious Diseases, Heidelberg University Hospital, Germany.
d. Department of Fundamental Chemistry, Federal University of Pernambuco, Recife, PE, Brazil.
e. Department of Chemical Engineering, Federal University of Pernambuco, Recife, PE, Brazil.
f. Laboratory of Virology, University of Campinas, Campinas, SP, Brazil.
g. Arthropod-borne and Infectious Diseases Laboratory (AIDL), Department of Microbiology, Immunology and Pathology, Colorado State University, Fort Collins, CO, United States of America.
h. Department of Infectious Diseases and Microbiology, University of Pittsburgh, Pittsburgh, PA, United States of America
i. German Centre for Infection Research (DZIF), Heidelberg Site, Heidelberg, Germany.
¶ These authors contributed equally to this work.
& These authors also contributed equally to this work.
*Corresponding author: thomas.jaenisch@urz.uni-heidelberg.de

Supporting Information – Contents

Figure S3. Superposition of the folded peptide to its native counterpart as in PD ID 5H6V.

Figure S4. Flowchart of antigen modeling protocol.

Figure S5. Sequence alignment among the NS2b-Grafted protein, the native scaffold and the ZIKV polyprotein.

Figure S6. Chromatographic purification and SDS-PAGE of the NS2B-Grafted protein.

Figure S7. ROC curve analysis of the NS2B-Grafted protein.

ADDITIONAL MATERIALS AND METHODS

Antibody affinity determination by microscale thermophoresis (MST)

Twenty micromolar of commercially synthesized NS2B peptide (GenScript, United States) was labeled using RED fluorescent dye NT-647-NHS Labeling Kit (NanoTemper Technologies, Germany) according to manufacturer's instructions. Unreacted dye was removed by buffer-exchange chromatography (EC), and labeled peptides were eluted in 1X PBS buffer. The total protein content was measured using a colorimetric assay (Pierce BCA, Protein Assay Kit, United States) in all of the serum samples used. Binding assays were performed using MST standard coated capillaries (NanoTemper). MST measurements comparing ZIKV *versus* DENV-infected samples were conducted with 150 nM of labeled peptide and a series of 16 1:2 serial dilutions of human serum (where the highest concentration of total IgG was 33.25 μM) in 1X PBS, 0.05% Tween-20 buffer. MST measurements comparing NeuroZIKV samples with and without previous DENV infection were conducted with 300 nM of labeled peptide and a series of 16 1:2 serial dilutions of human serum (where the highest concentration of total IgG was 640 nM) in 1X PBS, 0.05% Tween-20 buffer. Measurements were performed at 25 ºC using the Monolith NT.115 (NanoTemper). The red excitation LED was set to 40% and laser power to medium. Laser on-time was set to 30 s, and laser off-time was set to 5 s. The EC50 was derived from three independent thermophoresis experiments. The quality of each MST run was assessed by performing a capillary scan before and after each run to check that the fluorescence between samples remained within 10% variation. The graphs were plotted using Prism 7.0 software.

Multivariate data analysis

Complex datasets composed by multivariate responses [1], such as the one produced by the peptide array technology, often require multivariate data analysis techniques to give a more comprehensive understanding of the data. To get a better assessment of the relevant information contained in the data, two techniques will be further detailed here.

Principal component analysis (PCA) is an unsupervised learning technique, which means that it does not require any prior information regarding the class which the objects analyzed belong. The main goal of PCA analysis is to represent data in a new space of reduced dimensionality in which the new axes, known as principal components (PCs), will be built as a linear combination of the original variables (IgG response profiles) and in the direction of data's maximum variance. In that sense, the first PC of the model will explain the maximum variance of the data, while the second, orthogonal to PC1, will explain the maximum variance that is not explained by PC1. Mathematically speaking, PCA is a decomposition technique in which the **X** matrix containing measured dataset (patients *versus* IgG response per peptide array) will be decomposed into two according Equation 1.

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathbf{T}} + \mathbf{E} \quad (1)$$

The scores matrix **T** contains the new coordinates of samples in the new vector space of reduced dimensionality, while the loading matrix $\mathbf{P}^{\mathbf{T}}$ contains the weights of each original variable in the new PCA model. The residual matrix **E** contains the data variability that is not included in the model. To a better understanding of PCA and output interpretation we suggest the following literature.[1-3]

In contrast with PCA, that is an unsupervised learning method, partial least squares discriminant analysis (PLS-DA) is supervised and needs prior information

regarding objects to be modeled, in this case the patients' membership. Unlike PCA, PLS-DA does not seek for the maximum variance in matrix **X**, but the maximum covariance existent between matrix **X** and a **y** vector.[4, 5] The latter consists in a binary vector containing objects membership (ones for objects belonging to the class and zeros for those that does not).[6] From a mathematical standpoint, a PLS-DA model is built according Equations 2 and 3. Although **T** and **U** are score matrices and **P** and **q** are loadings, **T** and $\mathbf{P}^T$ are different from the ones in Equation 1, since now those also must be good predictors for **y**; $\mathbf{E_x}$ and $\mathbf{e_y}$ are residual matrix and vector, respectively. The inner relationship between **X** and **y** is stablished by Equation 4 and ruled by the weight matrix **W**.

$$\mathbf{X = TP^T + E_x} \quad (2)$$

$$\mathbf{y = Uq^T + e_y} \quad (3)$$

$$\mathbf{U = TW} \quad (4)$$

Afterwards, it is possible to obtain the coefficient regression vector **b**, that will be used to predict the membership of the objects in a new $\mathbf{X_{new}}$ matrix as depicted in Equation 5 and 6. From **b** it is possible to obtain the variable importance in projection (VIP) scores that shows which original variables are more important in the discrimination model.[7] For a more comprehensive understanding of PLS-DA, the reader is encourage to read the referred literature.

$$\mathbf{b = W(P^TW)q^T} \quad (5)$$

$$\hat{\boldsymbol{y}} = \boldsymbol{X_{new}b} \quad (6)$$

Antigen modeling procedure

Aiming to design a synthetic, soluble and stable protein carrying the identified NS2B epitope, we have used the epitope-scaffolding strategy, in which the structural motif of the epitope is grafted into a larger protein structure. The new protein, in turn, will expose the epitope in its native conformation, allowing its recognition by specific antibodies. The stabilization of the epitope's conformation, by incorporating it within a larger structure, aims to reduce the entropic penalty of when using flexible peptide.[8]

The computational engineering of the new antigen was performed using the MotifGraft[8] tool included in the Rosetta software package.[9] The protocol (Figure S4) consists of first building a database of scaffold structures. Then, the scaffold library is computationally searched for possible graft locations, by means of structural alignments between the scaffold and the epitope motif. If the backbones superimpose with a RMSD ≤ 2.5 Å, and do not produce great steric clashes, the side chains of the epitope are transplanted to the corresponding positions on the scaffold (side-chain grafting). However, most natural proteins are only marginally stable,[10] and when removed from their natural context, many proteins fold incorrectly and aggregate.[11] Therefore, it is necessary to redesign the grafted protein to ensure correct folding. First, to keep the epitope in its conformation, we aim to introduce intra-molecular interactions/contacts that will stabilize it. Thus, all residues within a 10 Å radius around the epitope are randomly modified, while the epitope itself does not undergo mutation (only repackaging of its side chains). On the other hand, to keep the folding of the rest of the scaffold protein (i.e. outside the 10 Å radius), instead of introducing random mutations, we use a phylogenetic restriction. Based on a database of homologous sequences to the scaffold protein, a position-specific substitution matrix (PSSM) is created.[12] Each PSSM element represents the logarithmic probability of observing a certain amino acid at a specific position in the protein. Thus, during the scaffold

redesign, only amino acids with a favorable score are used. The phylogenetic restriction for the allowed sequence space is based on the well-known fact that, in general, deleterious mutations are eliminated by natural selection. In other words, the mutation for the most frequently observed amino acid often increases stability (the consensus effect).[11] At the end of each mutation step, the structure is minimized and its energy calculated using a score function.[13] As with free energy (ΔG), a conformation/sequence with a lower score is more favorable, and the objective is to obtain the sequence of amino acids with the lower energy to stabilize the protein.[10, 11, 13] Thus, the criterion for accepting a mutation is based on thermodynamics free energy (i.e. $\Delta\Delta G < 0$), in order to increase the stability of both the epitope and the protein as a whole. The structural dynamics of the final redesigned structure is then characterized by means of MD.

Table S1. Origin and characterization of the samples used in the peptide array.

				Serological characterization							
	DENV previous exposure	**Study cohort**	**ZIKV RT-PCR (Pos/ Neg)**	**Panbio Dengue IgM Capture ELISA**	**Panbio Dengue IgG Capture ELISA**	**CDC Zika IgM MAC-ELISA**	**Panbio Dengue IgG Indirect ELISA**	**PRNT**		**ZIKV Phenotype**	
								ZIKV	**DENV**	**Acute**	**Convalescent**
										(< 3 days)	**(> 10 days)**
Adults	Positive	IDAMS	34/0	Neg	Pos	Pos	Pos	Pos	Pos	17	17
	Negative	IDAMS	28/0	Neg	Neg	Pos	Neg	Neg or unspecific	Neg or unspecific	14	14
	Low positive	IDAMS	6/0	Neg	Neg	Pos	Neg	Undetermined	Pos (low titer)	3	3

										No disease
Controls (adults)	Brazilians	Severe dengue	0/10	Neg	Neg	Neg	Pos	Neg	Pos	8
	Europeans	Severe dengue	0/7	Neg	Neg	Neg	Neg	Neg	Neg	7

										NeuroZIKV
NeuroZIKV	Positive	PRONEX	10/17	Neg	Pos	Pos	-	Pos	Pos	24
	Negative	PRONEX	7/1	Neg	Neg	Neg	-	Neg	Neg	11
									Total	118

Table S2. Group stratification of the samples used in the MST assays.

[a] Acute Disseminated Encephalomyelitis;

				Serological characterization						
	Study cohort	ZIKV RT-PCR	Neurological diagnosis	Panbio Dengue IgM Capture ELISA	Panbio Dengue IgG Capture ELISA	CDC Zika IgM MAC-ELISA	Panbio Dengue IgG Indirect ELISA	PRNT		N =
								ZIKV	DENV	
DENV- ZIKV+	Severe dengue	Pos	-	Neg	Pos	Pos	Pos	Pos	Neg	3
DENV+ ZIKV-	Severe dengue	Neg	-	Pos	Pos	Neg	-	Neg	-	3
NeuroZIKV / DENV-	PRONEX	Pos	ADEM[a] / GBS[b]	Neg	Neg	Pos	Neg	Pos	Neg	2
NeuroZIKV / DENV+	PRONEX	Pos	GBS	Neg	Pos	Pos	Pos	Pos	Neg	2
Negative controls (ZIKV- / DENV-)	PRONEX	Neg	-	Neg	Neg	Neg	Neg	Neg	Neg	2
									Total	12

[b] Guillain-Barré Syndrome

Table S3. Epidemiological, serologic and diagnostic characterization of the severe ZIKV (NeuroZIKV) samples used in the peptide array.

Group	Patient Number	Sample ID	Neurological Diagnosis	Age	Biological Sex	Days of Symptoms	ZIKV Serology	ZIKV RT-PCR (Serum)	ZIKV RT-PCR (LCR)	DENV Serology
NeuroZIKV / DENV- (N = 8)	329	416	Acute disseminated encephalomyelitis	2	F	8	IgM+ IgG-	Pos	Pos	IgM- IgG-
	196	224	Guillain-Barré Syndrome	21	F	2	IgM+	Neg	NT	IgM- IgG-
	228	268	Neuritis	51	F	1	NT*	Neg	Pos	IgG-
	85	87	Guillain-Barré Syndrome	46	M	90	IgM-	Pos	NT	IgM- IgG-
	228	235	Neuritis	51	F	180	IgM-	Pos	NT	IgM- IgG-
	368	402	Optical Neuritis and demyelinating lesions	28	F	90	IgM-	Pos	NT	IgM- IgG-
	215	219	Post-arboviral infection dystonia	53	F	7	IgM-	Pos	NT	IgM- IgG-
	290	315	Meningoencephalitis	30	M	60	IgM-	Pos	NT	IgM- IgG-
NeuroZIKV / DENV+ (N = 27)	315	402	Encephalitis	18	M	60	IgM-	Neg	Neg	IgM- IgG+
	252	303	Acute disseminated encephalomyelitis	56	F	6	IgM-	Pos	Neg	IgM- IgG+
	310	397	Guillain-Barré Syndrome	28	F	NA**	IgM+	Pos	NT	IgM- IgG+
	300	383	Guillain-Barré Syndrome	68	F	120	IgM-	Pos	NT	IgM- IgG+
	282	358	Guillain-Barré Syndrome	26	M	14	IgM+	Pos	Pos	IgM- IgG+
	318	405	Guillain-Barré Syndrome	25	M	10	IgM+	Pos	NT	IgM- IgG+
	249	297	Guillain-Barré Syndrome	48	F	4	IgM+	Pos	Pos	IgM- IgG+
	280	356	Guillain-Barré Syndrome	35	M	1	IgM+	Pos	Pos	IgM- IgG+
	189	217	Guillain-Barré Syndrome	NA	F	2	IgM-	Neg	NT	IgM- IgG+
	191	219	Guillain-Barré Syndrome	NA	M	2	IgM-	Pos	NT	IgM- IgG+
	195	223	Peroneal paralysis and numbness	30	F	1	IgM+	Neg	NT	IgM- IgG+
	195	294	Peroneal paralysis and numbness	30	F	10	IgM+	NT	NT	IgM- IgG+
	208	238	Guillain-Barré Syndrome	42	F	2	IgM-	Pos	NT	IgM- IgG+
	208	293	Guillain-Barré Syndrome	42	F	9	IgM+	NT	NT	IgM- IgG+
	216	315	Guillain-Barré Syndrome	50	M	4	IgM+	Neg	NT	IgM- IgG+
	290	370	Meningoencephalitis	NA	F	NA	IgM+	Neg	NT	IgG+
	341	429	Optical Neuritis	NA	F	NA	IgM+	Neg	NT	IgG+
	296	379	Guillain-Barré Syndrome	NA	F	NA	IgM+	NT	Neg	IgG+
	331	418	Guillain-Barré Syndrome	NA	F	NA	IgM+	NT	NT	IgG+
	195	294	Guillain-Barré Syndrome	30	F	1	IgM+	Neg	NT	IgM- IgG+
	274	290	Optical Neuritis	43	M	72	IgM+	Neg	NT	IgM- IgG+
	167	171	Guillain-Barré Syndrome	41	M	60	IgM-	Pos	NT	IgM- IgG+
	322	350	Myasthenia gravis and convulsion	56	M	120	IgM-	NT	NT	IgM- IgG+
	343	431	Optical Neuritis	NA	F	NA	IgM+	NT	NT	IgM- IgG+
	314	401	Guillain-Barré Syndrome	57	F	NA	IgM+	NT	NT	IgM- IgG+
	294	377	Guillain-Barré Syndrome	74	M	NA	IgM+	NT	NT	IgM- IgG+
	313	400	Guillain-Barré Syndrome	62	M	NA	IgM+	NT	NT	IgM- IgG+

*NT = Not tested
**NA = Not available

Table S4. Laboratorial characterization of the paired serum samples from mothers who delivered babies with microcephaly and their progeny.

	MC* - Newborn											MC - Mother						
Paired Samples ID	Biological sex	ZIKV Diagnosis	IgM ZIKV in serum	qRT-PCR ZIKV in serum	IgM ZIKV in CSF**	qRT-PCR ZIKV in CSF	DENV1 PRNT	DENV2 PRNT	DENV3 PRNT	DENV4 PRNT	ZIKV PRNT	IgM ZIKV in serum	qRT-PCR ZIKV in serum	DENV1 PRNT	DENV2 PRNT	DENV3 PRNT	DENV4 PRNT	ZIKV PRNT
01-012-0-1	M	Yes	Neg	Neg	Neg	Pos	Neg	Neg	Pos	Pos	Pos	Neg	Neg	Neg	Neg	Pos	Pos	Pos
01-013-0-1	M	Yes	Neg	Neg	Neg	Pos	Neg	Neg	Pos	Pos	Neg	Neg	Neg	Neg	Neg	Pos	Pos	Neg
01-014-0-1	F	Yes	Neg	Neg	Neg	Pos	Neg	Neg	Pos	Pos	Neg	Neg	Neg	Neg	Neg	Pos	Pos	Neg
01-015-0-1	M	Yes	Neg	Neg	Neg	Pos	Neg	Neg	Pos	Pos	Neg	Neg	Neg	Neg	Neg	Pos	Pos	Neg
02-004-0-1	F	Yes	Neg	Pos	Neg	Pos	Neg	Neg	Pos	Pos	Pos	Neg	Neg	Neg	Neg	Pos	Pos	Pos
02-005-0-1	F	Yes	Neg	Neg	Neg	Pos	Neg	Neg	Pos	Pos	Pos	Neg	Neg	Neg	Neg	Pos	Pos	Pos
02-006-0-1	M	Yes	Neg	Neg	Neg	Pos	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg
02-008-0-1	M	Yes	Pos	Neg	Pos	Neg	Neg	Neg	Pos	Pos	Pos	Neg	Neg	Neg	Neg	Pos	Pos	Pos
02-018-0-1	F	Yes	Neg	Neg	Neg	Pos	Neg	Neg	Pos	Pos	Neg	Neg	Neg	Neg	Neg	Pos	Pos	Neg
03-012-0-1	M	Yes	Neg	Neg	Neg	Pos	Neg	Neg	Pos	Pos	Pos	Neg	Neg	Neg	Neg	Pos	Pos	Pos
04-003-0-1	F	Yes	Pos	Neg	Pos	Neg	Neg	Neg	Pos	Pos	Pos	Neg	Neg	Neg	Neg	Pos	Pos	Pos
04-004-0-1	F	Yes	Neg	Neg	Pos	Neg	Neg	Neg	Pos	Pos	Pos	Neg	Neg	Neg	Neg	Pos	Pos	Pos
04-006-0-1	M	Yes	Pos	Neg	Pos	Neg	Neg	Neg	Pos	Neg	Pos	Neg	Neg	Neg	Neg	Pos	Neg	Pos
04-007-0-1	F	Yes	Pos	Neg	Pos	Neg	Neg	Neg	Neg	Neg	Pos	Neg	Neg	Neg	Neg	Neg	Neg	Pos
04-008-0-1	F	Yes	Pos	Neg	Pos	Neg	Neg	Neg	Neg	Neg	Pos	Neg	Neg	Neg	Neg	Neg	Neg	Pos
04-019-0-1	M	Yes	Pos	Neg	Pos	Pos	Neg	Neg	Neg	Pos	Pos	Neg	Neg	Neg	Neg	Neg	Pos	Pos
04-020-0-1	F	Yes	Neg	Neg	Neg	Pos	Neg	Neg	Pos	Pos	Pos	Neg	Neg	Neg	Neg	Pos	Pos	Pos
06-001-0-1	M	Yes	Pos	Neg	Pos	Neg	Neg	Neg	Neg	Neg	Pos	Neg	Neg	Neg	Neg	Neg	Neg	Pos
08-001-0-1	M	Yes	Neg	Neg	Neg	Pos	Neg	Neg	Neg	Neg	Pos	Neg	Neg	Neg	Neg	Neg	Neg	Pos
08-003-0-1	F	Yes	Neg	Neg	Neg	Pos	Neg	Neg	Pos	Pos	Pos	Neg	Neg	Neg	Neg	Pos	Pos	Pos
08-007-0-1	M	Yes	Neg	Neg	Neg	Pos	Neg	Neg	Pos	Pos	Pos	Neg	Neg	Neg	Neg	Pos	Pos	Pos

*MC = Microcephaly
*CSF = Cerebrospinal fluid

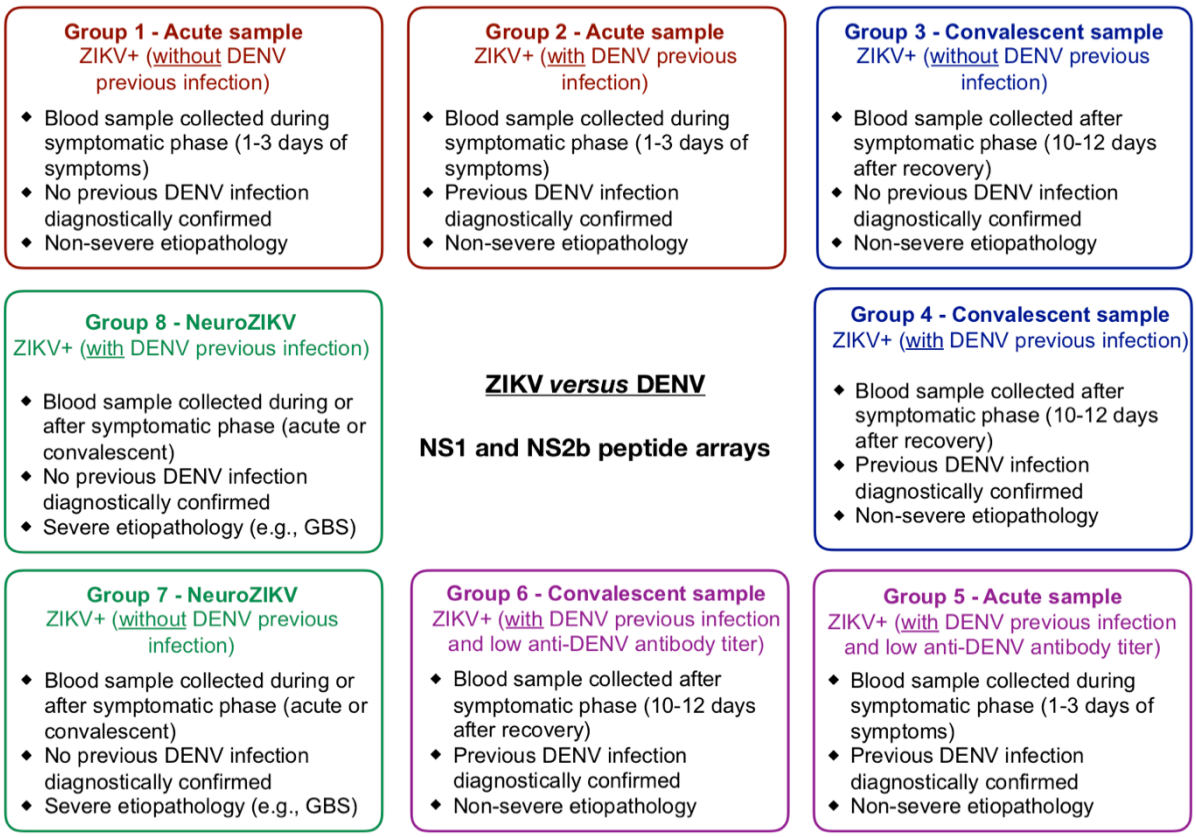
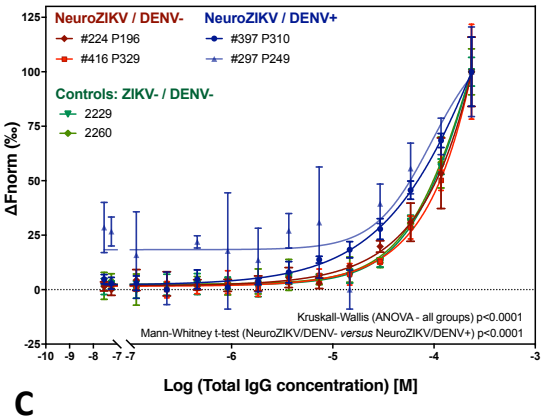


Figure S1. Sample set used in the peptide array, resulting in the identification of the NS2B peptide. One hundred and twenty well-characterized serum samples were divided into 8 groups, according to sample stratification molecular and serological tests. Comparison groups included: 1. ZIKV+ acute samples with and without previous DENV infection (coloured in red); 2. ZIKV+ convalescent samples with and without previous DENV infection (coloured in blue); 3. ZIKV+ acute and convalescent samples with low anti-DENV antibody titers (coloured in magenta); 4. NeuroZIKV samples with and without previous DENV infection (coloured in green).

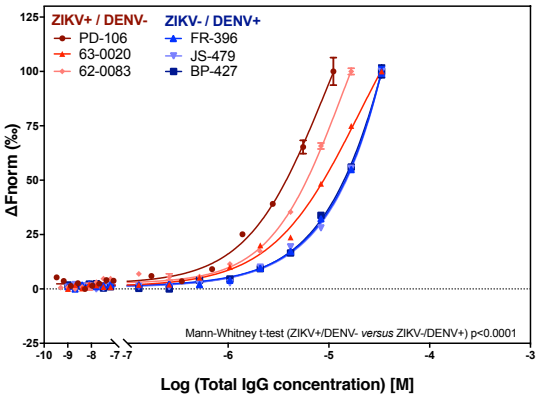
A



B

Group	Sample ID	EC50	R square
NeuroZIKV / DENV -	#224 P196	Undetermined	0.9968
	#416 P329	Undetermined	0.9972
NeuroZIKV / DENV +	#397 P310	4.71 x 10^{-4}	0.9970
	#297 P249	9.41 x 10^{-5}	0.8573
Controls: ZIKV- / DENV-	2229	2.83 x 10^{-4}	0.9973
	2260	2.58 x 10^{-4}	0.9963

C



D

Group	Sample ID	EC50	R square
ZIKV+ / DENV -	PD106	9.26 x 10^{-6}	0.9954
	62-0083	1.39 x 10^{-5}	0.9958
	62-0020	1.99 x 10^{-5}	0.9964
ZIKV- / DENV +	BP427	1.19 x 10^{-4}	0.9983
	JS479	5.57 x 10^{-4}	0.9978
	FR396	2.65 x 10^{-4}	0.9999

Figure S2. Antibody binding curves to the NS2B peptide by MST. (A) Serum binding curves comparison between NeuroZIKV samples with (NeuroZIKV / DENV+) and without previous DENV infection (NeuroZIKV / DENV-). MST data was acquired from serum samples from two subjects exhibiting severe ZIKV infection, with ZIKV RT-PCR and DENV IgG ELISA positive results (NeuroZIKV / DENV+, blue curves) and two subjects exhibiting severe ZIKV infection, with ZIKV RT-PCR positive result and DENV IgG ELISA negative result (NeuroZIKV / DENV-, red curves). Two samples from ZIKV and DENV naïve subjects were included as negative controls (green curves). Plots of the normalised fluorescence (ΔFnorm (‰)) vs. the concentration of total IgG are shown. Error bars represent standard deviations from three individual repeat experiments. The half-maximal binding parameter (EC50) was determined from a non-linear regression analysis. Kruskal-Wallis (ANOVA) test was performed and showed a significant difference among the tested groups (p value < 0.0001). Mann-Whitney t

test between NeuroZIKV/DENV- versus NeuroZIKA/DENV+ samples also revealed a significant difference (p value < 0.0001). (B) Summarizing table indicating the EC50 of anti-NS2B antibodies in the serum samples tested in A. (C) Serum binding curves comparison between ZIKV and DENV positive samples. MST data was acquired from serum samples from three subjects with confirmed single ZIKV infection (ZIKV+/DENV-, red curves) and three subjects with confirmed single DENV infection (ZIKV-/DENV+, blue curves). Plots of the normalised fluorescence (ΔFnorm (‰)) vs. the concentration of total IgG are shown. Error bars represent standard deviations from three individual repeat experiments. The half-maximal binding parameter (EC50) was determined from a non-linear regression analysis. Mann-Whitney t test was performed and showed that a significant difference between the ZIKV+ / DENV- and ZIKV- / DENV+) p-value < 0.0001. (D) Summarizing table indicating the EC50 of anti-NS2B antibodies in the serum samples tested in D.

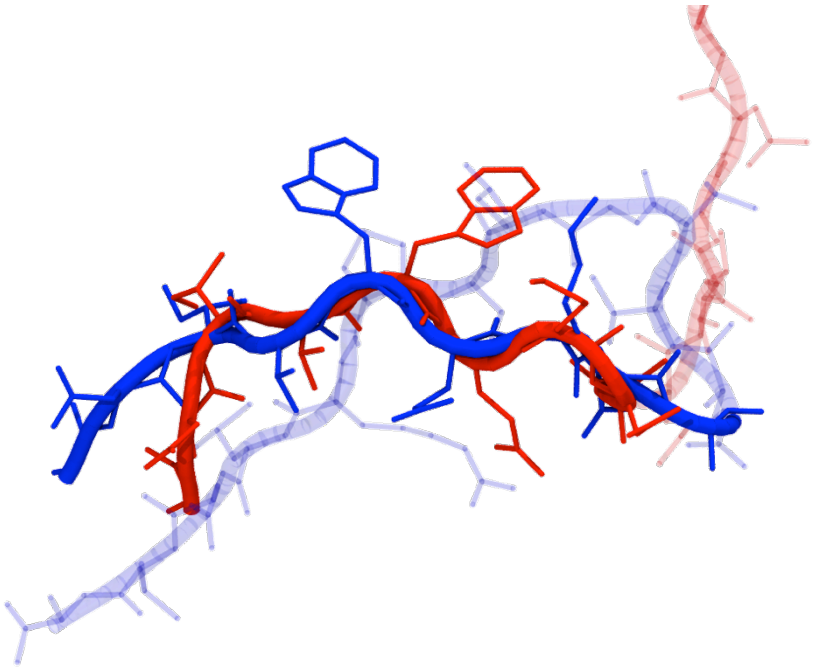


Figure S3. Superposition of the folded peptide (blue) to its native counterpart (red) as in PD ID 5H6V.

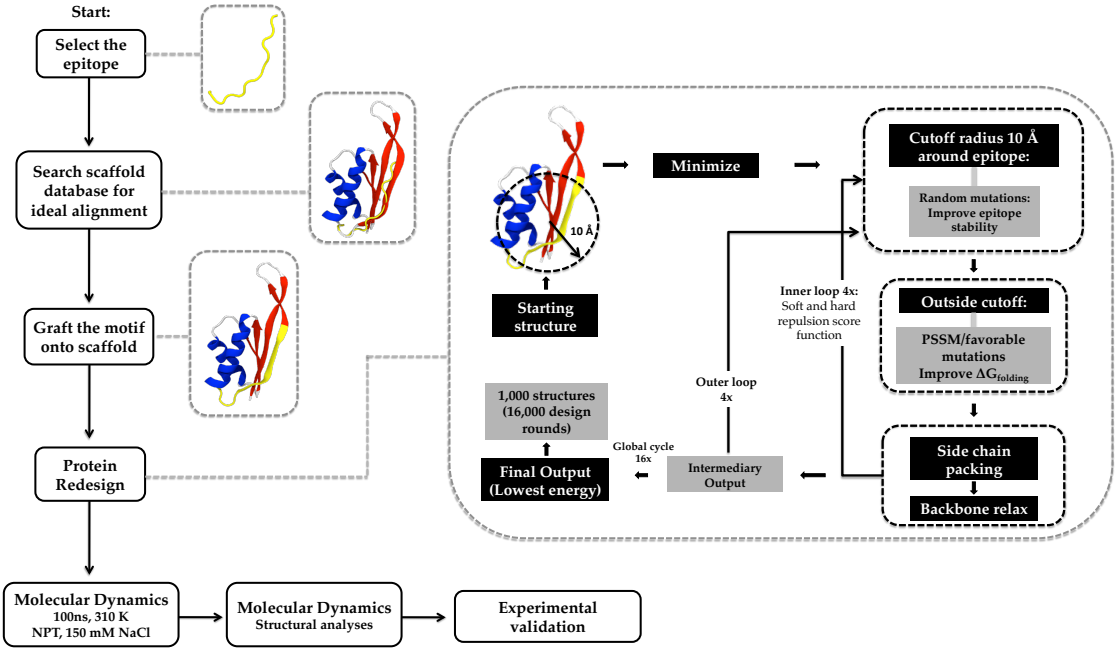


Figure S4. Flowchart of the antigen modeling protocol using the MotifGraft strategy followed by redesign of the protein to increase epitope conformation and protein folding stability, Molecular Dynamics Analyses and finally experimental validation by means of ELISA assay.

A

```
NS2b-Grafted       1 NVVYIGKKPVMNYVLAVLTQLNNGTEVVIKARGKAISKAVDVAEMVIKRFEKDAEVTGNSPRLDKITNPDGRETNVSTIE
Native-Scaffold    1 NVVLIGKKPVMNYVVAVLTQLTSNDEVTIKARGKAINKAVDVAEMIRNRFIKDIKIKKIEIGTDKVKNPDGREVNVSTIE

NS2b-Grafted      81 IVLGK
Native-Scaffold   81 IVLAK
```

B

```
NS2b-Grafted    1 ------NVVYIGKKPV----------------------MNYVLAVLTQLNNGTEVVIKARGKAISKAVDVAEM----VIK
ZIKV-Capsid     1 KKSGGFRIVNMLKRGVARVSPFGGLKRLPAGLLLGHGPIRMVLAILAFLRF---TAIKPSLGLINRWGSVGKKEAMEIIK

NS2b-Grafted   49 RFEKDAEVTGNSPRLDKITNPDGRETNVSTIEIVLGK---
ZIKV-Capsid    78 KFKKLLAAMLRIINARKEKKRRGADTSVGIVGLLLTTAMA
```

C

```
NS2b-Grafted    1 NVVYIGKKPVMNYVLAVLTQLNNGTEVVIKAR---GKAISKAVDVAEMVIKRFEKDAEVTGNSPRLDKITNPDGRE--TN
ZIKV-PrM        1 --AEVTRRGSAYYMY--LDRNDAGEAISFPTTLGMNKCYIQIMDLGHMC------DATMSYECPMLDEGVEPDDVDCWCN

NS2b-Grafted   76 VSTIEIVLGK-------------
ZIKV-PrM       71 TTSTWVVYGTCHHKKGEARRSRR
```

D

```
NS2b-Grafted    1 NVVYIGKKPVMNYVLAVLTQLNNGTEVVIKARGKAISKAVDVAEMVIKRFE---KDAEVTGNSPRLDKITNPDGRETNVS
ZIKV-M          1 ------------------VTLPSHSTRKLQTRSQTWLESREYTKHLIRVENWIFRNPGFALAAAAIAWLLGSSTSQKVIY

NS2b-Grafted   78 TIEIVLGK----
ZIKV-M         63 LVMILLIAPAYS
```

E

```
NS2b-Grafted    1 NVVYIG--------------------------------KKPVMNYVLAVLIQLNNGTEVV----------------IKARG
ZIKV-E          1 -IRCIGVSNRDFVEGMSGGTWVDVVLEHGGCVTVMAQDKPTVDIELVTTTVSNMAEVRSYCYEASISDMASDSRCPTQG

NS2b-Grafted   34 KAISKAVDVAEMVIKR-------------EEKDA-----------------------------------------------
ZIKV-E         79 EAYLDKQSDTQYVCKRTLVDRGWGNGCGLFGKGSLVTCAKFACSKKMTGKSIQPENLEYRIMLSVHGSQHSGMIVNDTGH

NS2b-Grafted   55 ---------EVTGNSPR-----------------------------------------------------------------
ZIKV-E        159 ETDENRAKVEITPNSPRAEATLGGFGSLGLDCEPRTGLDFSDLYYLTMNNKHWLVHKEWFHDIPLPWHAGADTGTPHWNN

NS2b-Grafted   63 ------------------------------------------------------LDKIT--------------------
ZIKV-E        239 KEALVEFKDAHAKRQTVVVLGTQEGAVHTALAGALEAEMDGAKGRLSSGHLKCRLKMDKLRLKGVSYSLCTAAFTFTKIP

NS2b-Grafted   68 ----------------------------------NPDGRETNVSTI---------------------------------
ZIKV-E        319 AETLHGTVTVEVQYAGTDGPCKVPAQMAVDMQTLTPVGRLITANPVITESTENSKMMLELDPPFGDSYIVIGVGEKKITH

NS2b-Grafted   80 ------------------------------------------------------------------EIVLGK--------
ZIKV-E        399 HWHRSGSTIGKAFEATVRGAKRMAVLGDTAWDFGSVGGALNSLGKGIHQIFGAAFKSLFGGMSWFSQILIGTLLMWLGLN

NS2b-Grafted      --------------------------
ZIKV-E        479 TKNGSISLMCLALGGVLIFLSTAVSA
```

F

```
NS2b-Grafted    1 ------------------------------------------------------------NVVYIGKKPVMNYVL---
ZIKV-NS1        1 GCSVDFSKKETRCGTGVFVYNDVEAWRDRYKYHPDSPRRLAAAVKQAWEDGICGISSVSRMENIMWRSVEGELNAILEEN

NS2b-Grafted   16 ----------------------AVLTQLNNG-----------------------------------------------
ZIKV-NS1       81 GVQLTVVVGSVKNPMWRGPQRLPVPVNELPHGWKAWGKSYFVRAAKTNNSFVVDGDTLKECPLKHRAWNSFLVEDHGFGV

NS2b-Grafted   25 --TEVVIKAR-----------GKAI-----------------------------------------------SKAVDVAEMV
ZIKV-NS1      161 FHTSVWLKVREDYSLECDPAVIGTAVKGKEAVHSDLGYWIESEKNDTWRLKRAHLIEMKTCEWPKSHTLWTDGIEESDLI

NS2b-Grafted   47 IK---------------------------------RFEK--------DAEVTGNSPRLDKIINPDGR---------------
ZIKV-NS1      241 IPKSLAGPLSHHNTREGYRTQMKGPWHSEELEIRFEECPGTKVHVEETCGTRGPSLRSATASGRVIEEWCCRECTMPPL

NS2b-Grafted   73 --------------------ETNVSTIEIVLGK
ZIKV-NS1      320 SFRAKDGCWYGMEIRPRKEPESNLVRSMVTAGS
```

G

```
NS2b-Grafted    1 NVVYI------------GKKPVMNYVLAVLTQLNNG---------TEVVIKARGKAISKAVDVAEM--------------
ZIKV-NS2A       1 GVLVLLLMVQEGLKKRMTTKIIISTSMAVLVAMILGGFSMSDLAKLAILMGATFAEMNTGGDVAHLALIAAFKVRPALLV

NS2b-Grafted   46 -------------------------VIKREKDAEVTGNSPRLD-----KITNPDGRETNVSTIEIVLGK
ZIKV-NS2A      81 SFIFRANWTPRESMLLALASCFLQTAISALEGDLMVLINGFALAWLAIRAMVVPRTDNITLAILAALTPL
```

H

```
NS2b-Grafted    1 NVVYIGKKPVMNYVLAVLTQLNNG-TEVVIKARGKAISKAVDVA---------EVVIKR-----PEKDAEVTGNSPRLD-
ZIKV-NS2B       1 SEVLTA--------VGLICALAGGFAKADIEMAGPMAAVGLLIVSYVVSGKSVDMYIERAGDITWEKDAEVTGNSPRLDV

NS2b-Grafted   65 --------KITNPDGRETNVSTIEIVL------------------------GK-
ZIKV-NS2B      73 ALDESGDFSLVEDDGPPMREIILKVVLMTICGMNPIAIPFAAGAWYVYVKTGKR
```

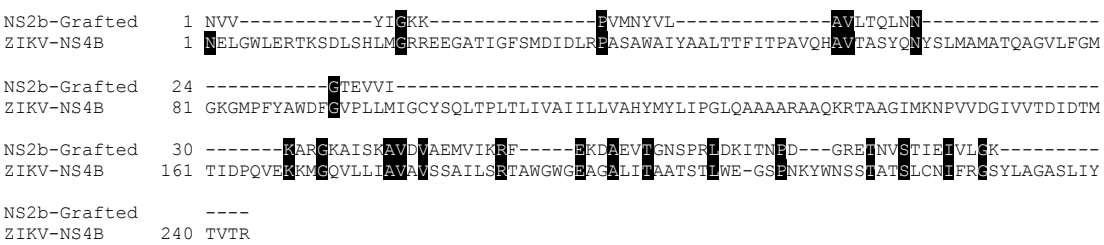
I

```
NS2b-Grafted    1 ------------------------------------------------------------------------------NVVYI
ZIKV-NS3        1 SGALWDVPAPKEVKKGETTDGVYRVMTRRLLGSTQVGVGVMQEGVFHTMWHVTKGSALRSGEGRLDPYWGDVKQDLVSYC

NS2b-Grafted    6 GK-----------------------------------------------------KPVMN--------------------
ZIKV-NS3       81 GPWKLDAAWDGHSEVQLLAVPPGERARNIQTLPGIFKTKDGDIGAVALDYPAGTSGSPILDKCGRVIGLYGNGVVIKNGS

NS2b-Grafted   13 ------------------------------------------------------------------------------
ZIKV-NS3      161 YVSAITQGRREEETPVECFEPSMLKKKQLTVLDLHPGAGKTRRVLPEIVREAIKTRLRTVILAPTRVVAAEMEEALRGLP

NS2b-Grafted   13 --YVLAVLTQLNNGTEVV-------------------------------------IKARG----------------------
ZIKV-NS3      241 VRYMTTAVNVTHSGTEIVDLMCHATFTSRLLQPIRVPNYNLYIMDEAHFTDPSSIAARGYISTRVEMGEAAAIFMTATPP

NS2b-Grafted   34 ----------------------KAISKAVD---------------------------------VAEMVIKRFEKDAEVT--
ZIKV-NS3      321 GTRDAFPDSNSPIMDTEVEVPERAWSSGFDWVTDHSGKTVWFVPSVRNGNEIAACLTKAGKRVIQLSRKTFETEFQKTKH

NS2b-Grafted   58 ------------------------------------------------------------GNSPR---------------
ZIKV-NS3      401 QEWDFVVTTDISEMGANFKADRVIDSRRCLKPVILDGERVILAGPMPVTHASAAQRRGRIGRNPNKPGDEYLYGGGCAET

NS2b-Grafted   63 -------------LDKITNPDG-------------------------RETNVSTI------------------------
ZIKV-NS3      481 DEDHAHWLEARMLLDNIYLQDGLIASLYRPEADKVAAIEGEFKLRTEQRKTFVELMKRGDLPVWLAYQVASAGITYTDRR

NS2b-Grafted   80 ------------------------------------------------EIVLGK-
ZIKV-NS3      561 WCFDGTTNNTIMEDSVPAEVWTRHGEKRVLKPRWMDARVCSDHAALKSFKEFAAGKR
```

J

```
NS2b-Grafted    1 NVVYIGKKP---------VMNYVLAVLTQLNNGTEVVIKARGKAISKAVDVAEMV-------------------KRFEK
ZIKV-NS4A       1 VMEALGTLPGHMTERFQEAIDN-LAVLMRAETGSRP-YKAAAAQLPETLETIMLLGLLGTVSLGIFFVLMRNKGIGKMGF

NS2b-Grafted   53 DAEVTGNS----------------------------PRLDKITNPDGRETNVSTIEIVLGK------
ZIKV-NS4A      79 GMVTLGASAWLMWLSEIEPARIACVLIVVFLLLVVLIPEPEKQRSPQDNQMAI-IIMVAVGLLGLITA
```

K



L

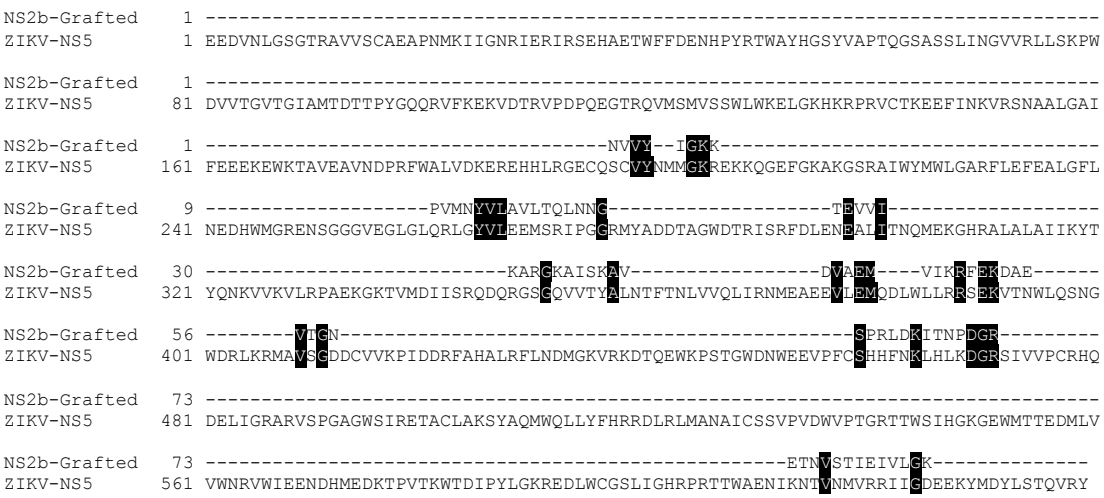


Figure S5. Alignment of the NS2b-Grafted protein with (A) the native scaffold; (B) ZIKV Capsid protein; (C) ZIKV PrM protein; (D) ZIKV Membrane protein; (E) ZIKV Envelope protein; (F) ZIKV-NS1 protein; (G) ZIKV-NS2A protein; (H) ZIKV-NS2B protein; (I) ZIKV-NS3 protein; (J) ZIKV-NS4A protein; (K) ZIKV-NS4B protein and (L) ZIKV-NS5 protein. Identical amino acids are black shaded, whereas the epitope is highlighted in yellow.

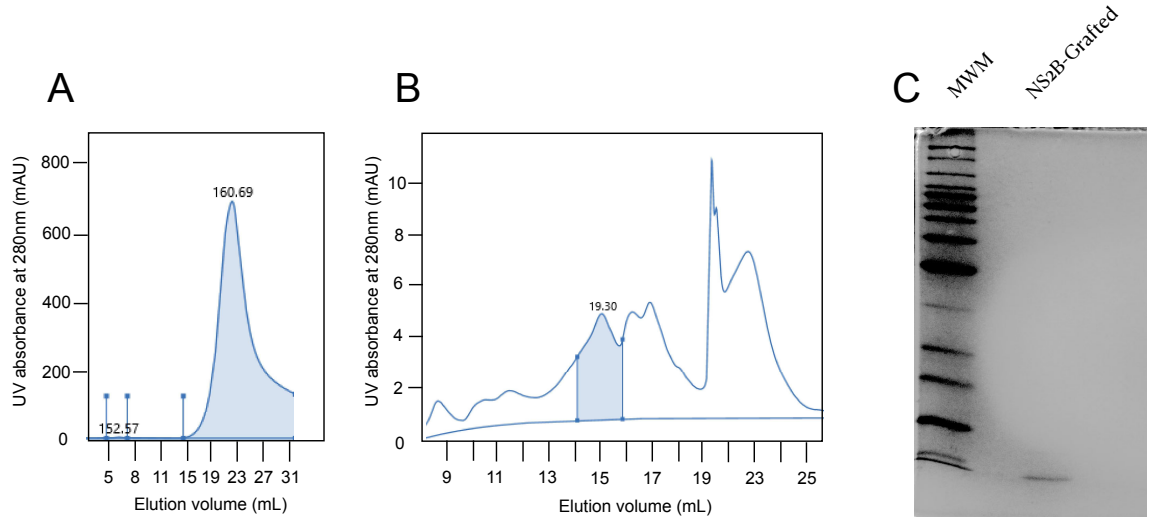


Figure S6. Chromatographic purification and SDS-PAGE of the NS2B-Grafted protein. The NS2B-Grafted protein was submitted to a two-step purification including affinity chromatography (A) followed by size exclusion chromatography (B). UV absorbance at 280 nm (mAU) was monitored over time, and fractions corresponding to the NS2B-Grafted protein (shaded areas in the graphs) were collected. Protein electrophoresis of the NS2B-Grafted protein shows the resultant protein presents a molecular weight of 11.5 kDa, while no aggregates are observed (C). The molecular weight marker (MWM) used was the Benchmark protein ladder (Thermo Fisher Scientific).

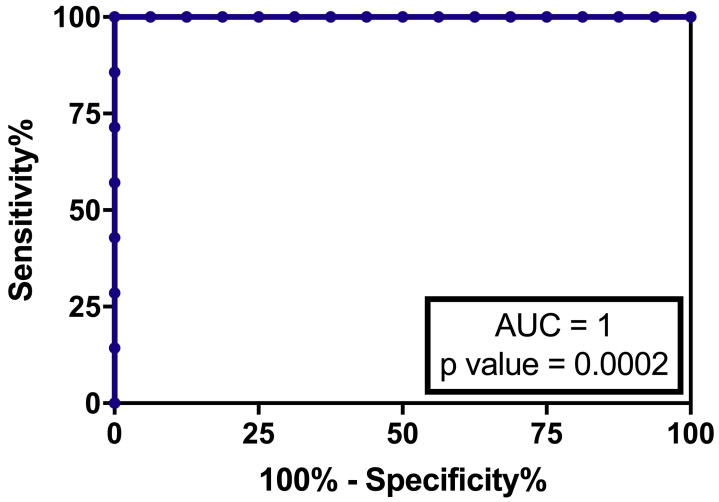


Figure S7. ROC curve analysis of the NS2B-Grafted protein. The paired results for sensitivity and specificity were plotted as points in a ROC space and the trade-off between these measures for different discrimination cut-offs are graphically represented. The cut-off of 1.16 was used for comparison purposes in the ELISA assay using sera from mothers and babies with ZIKV-associated microcephaly.

ASSOCIATED CONTENT REFERENCES

1. Bro, R.; Smilde, A. K., Principal component analysis. *Analytical Methods* **2014,** *6* (9), 2812-2831.
2. Ringnér, M., What is principal component analysis? *Nature Biotechnology* **2008,** *26* (3), 303-304.
3. Wold, S.; Esbensen, K.; Geladi, P., Principal component analysis. *Chemometrics and Intelligent Laboratory Systems* **1987,** *2* (1), 37-52.
4. Brereton, R. G.; Lloyd, G. R., Partial least squares discriminant analysis: taking the magic away. *Journal of Chemometrics* **2014,** *28* (4), 213-225.
5. Gromski, P. S.; Muhamadali, H.; Ellis, D. I.; Xu, Y.; Correa, E.; Turner, M. L.; Goodacre, R., A tutorial review: Metabolomics and partial least squares-discriminant analysis--a marriage of convenience or a shotgun wedding. *Anal Chim Acta* **2015,** *879*, 10-23.
6. Ståhle, L.; Wold, S., Partial least squares analysis with cross-validation for the two-class problem: A Monte Carlo study. *Journal of Chemometrics* **1987,** *1* (3), 185-196.
7. Farrés, M.; Platikanov, S.; Tsakovski, S.; Tauler, R., Comparison of the variable importance in projection (VIP) and of the selectivity ratio (SR) methods for variable selection and interpretation. *Journal of Chemometrics* **2015,** *29* (10), 528-536.
8. Silva, D.-A.; Correia, B. E.; Procko, E., Motif-Driven Design of Protein–Protein Interfaces. In *Computational Design of Ligand Binding Proteins*, Stoddard, B. L., Ed. Springer New York: New York, NY, 2016; pp 285-304.
9. Leaver-Fay, A.; Tyka, M.; Lewis, S. M.; Lange, O. F.; Thompson, J.; Jacak, R.; Kaufman, K. W.; Renfrew, P. D.; Smith, C. A.; Sheffler, W.; Davis, I. W.; Cooper, S.; Treuille, A.; Mandell, D. J.; Richter, F.; Ban, Y.-E. A.; Fleishman, S. J.; Corn, J. E.; Kim, D. E.; Lyskov, S.; Berrondo, M.; Mentzer, S.; Popović, Z.; Havranek, J. J.; Karanicolas, J.; Das, R.; Meiler, J.; Kortemme, T.; Gray, J. J.; Kuhlman, B.; Baker, D.; Bradley, P., Rosetta3: An Object-Oriented Software Suite for the Simulation and Design of Macromolecules. In *Methods in Enzymology*, Johnson, M. L.; Brand, L., Eds. Academic Press: 2011; Vol. 487, pp 545-574.
10. Magliery, T. J., Protein stability: computation, sequence statistics, and new experimental methods. *Current opinion in structural biology* **2015,** *33*, 161-168.

11. Goldenzweig, A.; Goldsmith, M.; Hill, Shannon E.; Gertman, O.; Laurino, P.; Ashani, Y.; Dym, O.; Unger, T.; Albeck, S.; Prilusky, J.; Lieberman, Raquel L.; Aharoni, A.; Silman, I.; Sussman, Joel L.; Tawfik, Dan S.; Fleishman, Sarel J., Automated Structure- and Sequence-Based Design of Proteins for High Bacterial Expression and Stability. *Molecular Cell* **2016,** *63* (2), 337-346.
12. Altschul, S. F.; Gertz, E. M.; Agarwala, R.; Schäffer, A. A.; Yu, Y.-K., PSI-BLAST pseudocounts and the minimum description length principle. *Nucleic Acids Research* **2008,** *37* (3), 815-824.
13. Alford, R. F.; Leaver-Fay, A.; Jeliazkov, J. R.; O'Meara, M. J.; DiMaio, F. P.; Park, H.; Shapovalov, M. V.; Renfrew, P. D.; Mulligan, V. K.; Kappel, K.; Labonte, J. W.; Pacella, M. S.; Bonneau, R.; Bradley, P.; Dunbrack, R. L.; Das, R.; Baker, D.; Kuhlman, B.; Kortemme, T.; Gray, J. J., The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *Journal of Chemical Theory and Computation* **2017,** *13* (6), 3031-3048.