Identification of a Zika NS2B epitope as a biomarker for 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[b], Clarice N.L. Morais[b], Marli T. Cordeiro[b], Roberto D. Lins[2,&], Ernesto T.A. Marques[b,g,&] and Thomas Jaenisch[c,h,&*]

[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.] *Department of Infectious Diseases and Microbiology, University of Pittsburgh, Pittsburgh, PA, United States of America*
[h.] *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

Additional Materials and methods.

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

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

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

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

Table S5. Contingency table analysis of PLS-DA model in Figure 2.

Table S6. Contingency table analysis of PLS-DA model in Figure 3 A-B.

Table S7. Contingency table analysis of PLS-DA model in Figure 3 C-D.

Table S8. Contingency table analysis of PLS-DA model in Figure 3 E-F.

Table S9. Confusion Matrix analysis for the NS2B-Grafted protein ELISA assay using 1.16 as cut-off (100% specificity and 85.71% sensitivity).

Table S10. Samples classification according to the NS2B-Grafted protein ELISA assay.

Figure S1. Antibody binding curves to the NS2B peptide by MST.

Figure S2. Superposition of the folded peptide to its native counterpart as in PD ID 5H6V.

Figure S3. Flowchart of antigen modeling protocol.

Figure S4. Sequence alignment among the NS2b-Grafted protein, the native scaffold and the ZIKV polyprotein.

Figure S5. Chromatographic purification and SDS-PAGE of the NS2B-Grafted protein.

Figure S6. 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**.

$$\mathbf{X = TP^T + E_x} \tag{2}$$

$$\mathbf{y = Uq^T + e_y} \tag{3}$$

$$\mathbf{U = TW} \tag{4}$$

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.

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

$$\hat{y} = X_{new}b \tag{6}$$

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						ZIKV Phenotype	
	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**		**Acute**	**Convalescent**
								ZIKV	**DENV**	**(< 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.

				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

[a] Acute Disseminated Encephalomyelitis;

[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

Table S5. Contingency table analysis of PLS-DA model in Figure 2.

		Actual class	
		Convalescent ZIKV+, DENV- Convalescent ZIKV+, DENV low, Convalescent ZIKV+, DENV+	NeuroZIKV, DENV+ NeuroZIKV, DENV- ZIKV-, DENV-
Predicted as	Convalescent ZIKV+, DENV- Convalescent ZIKV+, DENV low Convalescent ZIKV+,DENV+	28	9
	NeuroZIKV, DENV+ NeuroZIKV, DENV- ZIKV-, DENV-	6	34

Table S6. Contingency table analysis of PLS-DA model in Figure 3 A-B.

		Actual class	
		Convalescent ZIKV+, DENV- Convalescent ZIKV+, DENV low, Convalescent ZIKV+, DENV+	NeuroZIKV, DENV+ NeuroZIKV, DENV- ZIKV-, DENV-
Predicted as	Convalescent ZIKV+, DENV- Convalescent ZIKV+, DENV low Convalescent ZIKV+,DENV+	32	1
	NeuroZIKV, DENV+ NeuroZIKV, DENV- ZIKV-, DENV-	2	18

Table S7. Contingency table analysis of PLS-DA model in Figure 3 C-D.

		Actual class	
		Convalescent ZIKV+, DENV-	NeuroZIKV, DENV+ NeuroZIKV, DENV-
Predicted as	Convalescent ZIKV+, DENV-	11	10
	NeuroZIKV, DENV+ NeuroZIKV, DENV-	3	25

Table S8. Contingency table analysis of PLS-DA model in Figure 3 E-F.

		Actual class	
		NeuroZIKV, DENV+	NeuroZIKV, DENV-
Predicted as	NeuroZIKV, DENV+	21	2
	NeuroZIKV, DENV-	3	9

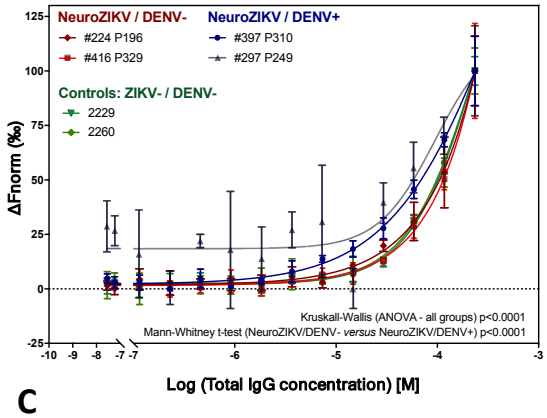
Table S9. Confusion Matrix analysis for the NS2B-Grafted protein ELISA assay using 1.16 as cut-off (100% specificity and 85.71% sensitivity).

		Actual class	
		Positive	Negative
Predicted as	Positive	6	0
	Negative	1	16

Table S10. Samples' classification according to the NS2B-Grafted ELISA assay.

		Samples	
		MC - Mothers	**MC - Newborn**
Predicted as	Positive	0	1
	Negative	21	20

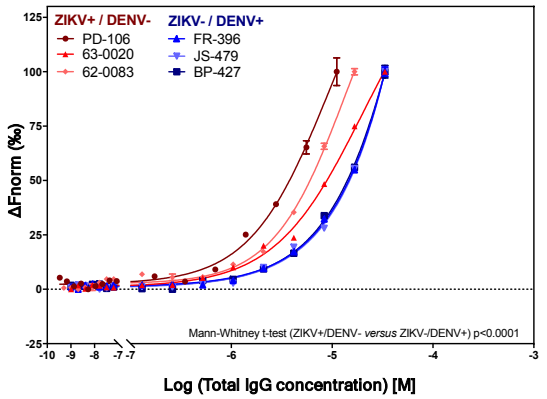
A



B

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

C



D

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

Figure S1. 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.

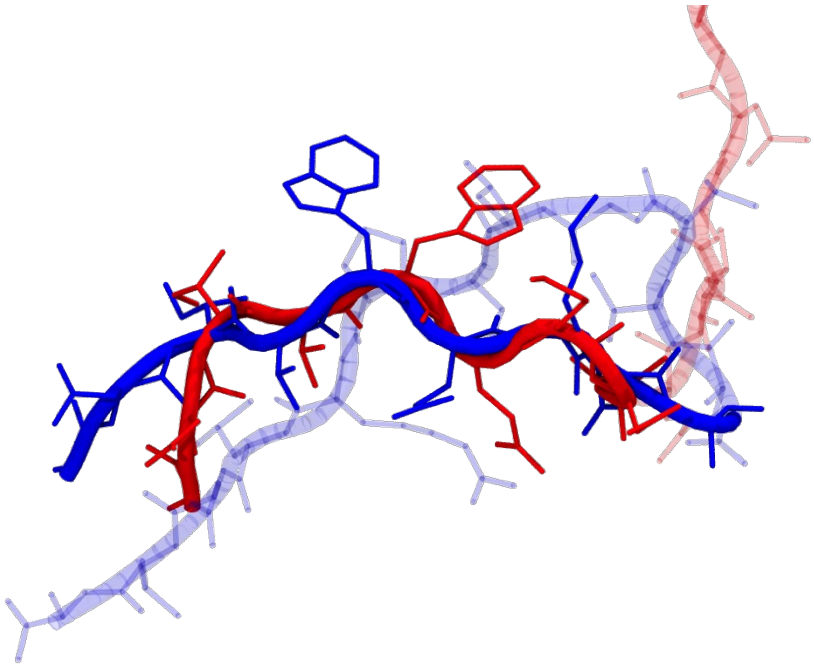


Figure S2. Superposition of the folded peptide (blue) to its native counterpart (red) as in PD ID 5H6V.

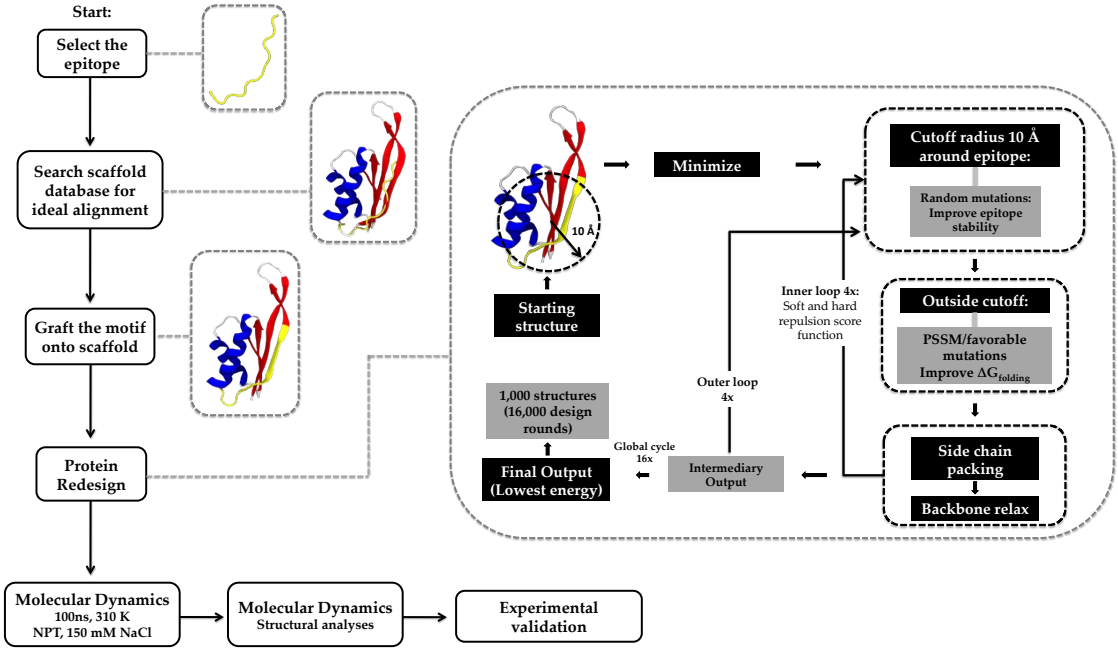


Figure S3. Flowchart of the antigen modeling protocol using the MotifGraft strategy followed by redesign of the protein to increase epitope conformation and protein folding stability, Molecular Dynamics Analyses and finally experimental validation by means of ELISA assay.

A

```
NS2b-Grafted       1 NVVYIGKKPVMNYVLAVLTQLNNGTEVVIKARGKAISKAVDVAEMVIKRFKKDAEVTGNSPRLDKITNPDGRETNVSTIE
Native-Scaffold    1 NVVLIGKKPVMNYVVAVLTQLTSNDEVIIKARGKAINKAVDVAEMIRNRFIKDIKIKKIEIGTDKVKNPDGREVNVSTIE

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 KFKKDLAAMLRIINARKEKKRRGADTSVGIVGLLLTTAMA
```

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--------------------------------KKPVMNYVLAVLTQLNNGTEVV----------------IKARG
ZIKV-E          1 -IRCIGVSNRDFVEGMSGGTWVDVVLEHGGCVTVMAQDKPTVDIELVTTTVSNMAEVRSYCYEASISDMASDSRCPTQG

NS2b-Grafted   34 KAISKAVDVAEMVIKR-------------TEKDA-----------------------------------------------
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--------DAEVTGNSPRLDKITNPDGR---------------
ZIKV-NS1      241 IPKSLAGPLSHHNTREGYRTQMKGPWHSEELEIRFEECPGTKVHVEETCGTRGPSLRSTTASGRVIEEWCCRECTMPPL

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

G

```
NS2b-Grafted    1 NVVYL------------GKKPVMNYVLAVLTQLNNG---------TEVVIKARGKAISKAVDVAEM--------------
ZIKV-NS2A       1 GVLVILLMVQEGLKKRMTTKIIISTSMAVLVAMILGGFSMSDLAKLAILMGATFAEMNTGGDVAHLALIAAFKVRPALLV

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

H

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

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

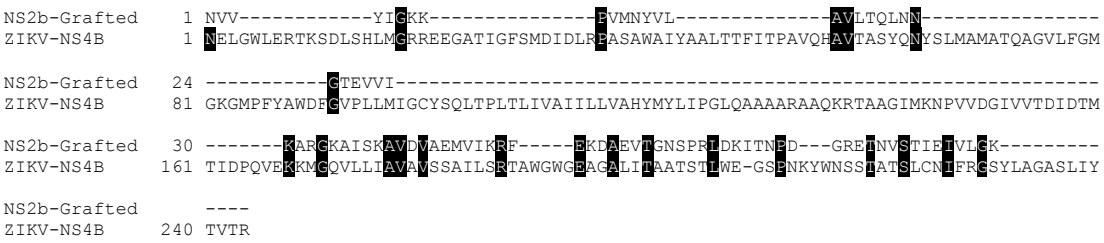
I

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

NS2b-Grafted    6 GK-------------------------------------------------------KEVMN--------------------
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-------------------IKRFEK
ZIKV-NS4A       1 VMEALGTLPGHMTERFQEAIDN-LAVLMRAETGSRP-YKAAAAQLPETLETIMLLGLLGTVSLGIFFVLMRNKGIGKMGF

NS2b-Grafted   53 DAEVTGNS-----------------------------ERLDKITNPDGRETNVSTIEIVLGK------
ZIKV-NS4A      79 GMVTLGASAWLMWLSEIEPARIACVLIVVFLLLVVLIEEPEKQRSPQDNQMAI-ILMVAVGLLGLITA
```

K



L

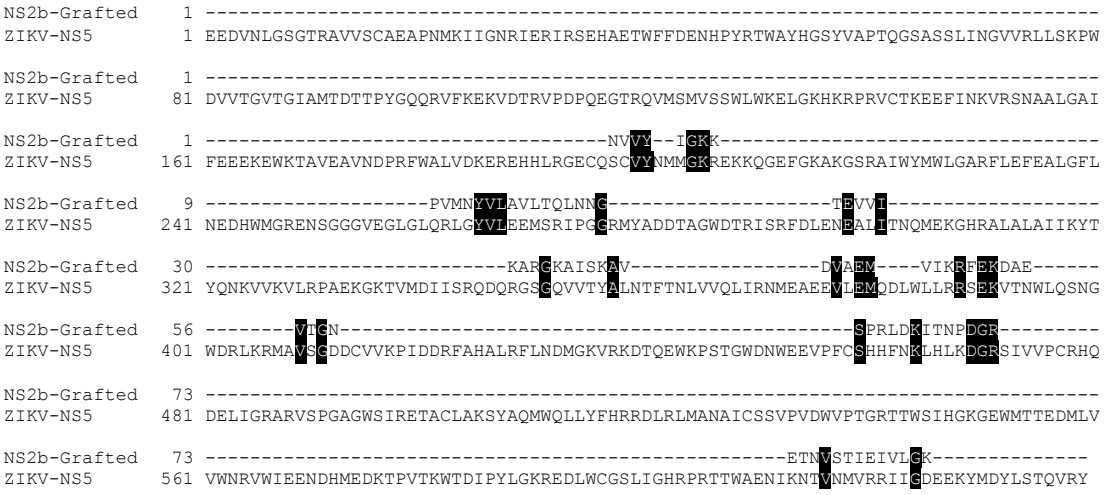


Figure S4. Alignment of the NS2b-Grafted protein with (A) the native scaffold; (B) ZIKV Capsid protein; (C) ZIKV PrM protein; (D) ZIKV Membrane protein; (E) ZIKV Envelope protein; (F) ZIKV-NS1 protein; (G) ZIKV-NS2A protein; (H) ZIKV-NS2B protein; (I) ZIKV-NS3 protein; (J) ZIKV-NS4A protein; (K) ZIKV-NS4B protein and (L) ZIKV-NS5 protein. Identical amino acids are black shaded, whereas the epitope is highlighted in yellow.

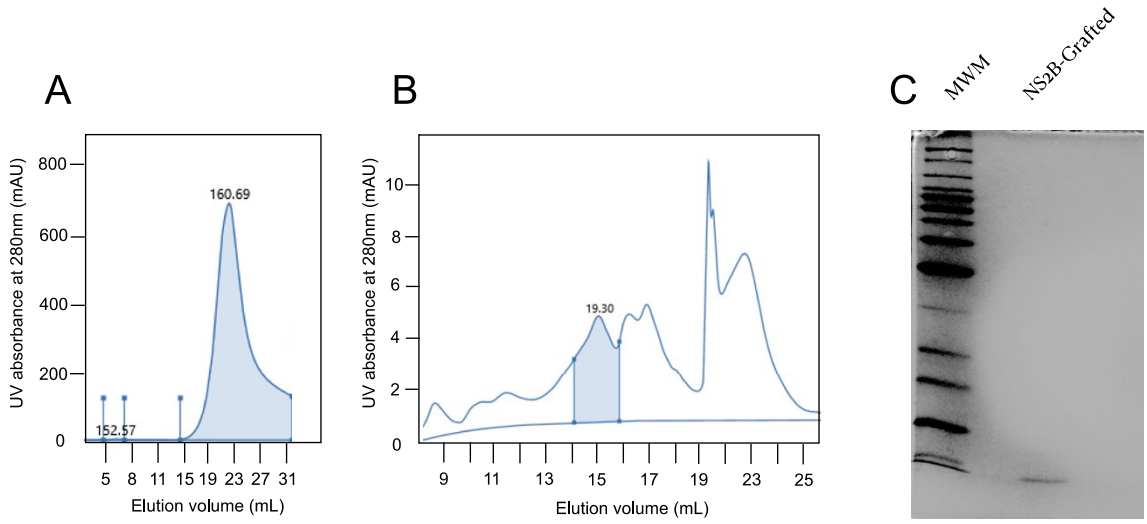


Figure S5. Chromatographic purification and SDS-PAGE of the NS2B-Grafted protein. The NS2B-Grafted protein was submitted to a two-step purification including affinity chromatography (A) followed by size exclusion chromatography (B). UV absorbance at 280 nm (mAU) was monitored over time, and fractions 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).

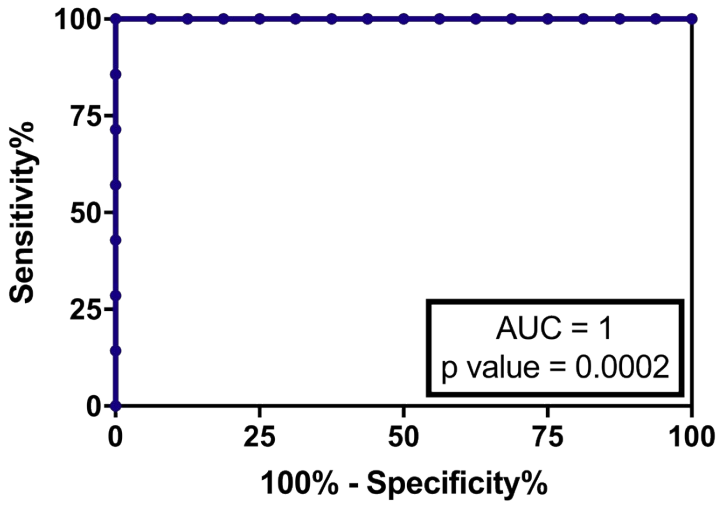


Figure S6. ROC curve analysis of the NS2B-Grafted protein ELISA. One hundred and sixty (160) cryopreserved serum samples were previously classified as acute or convalescent samples, according to serological data (presence of IgM and IgG antibodies against ZIKV and or DENV-1,2,3,4 through ELISA and PRNT assays) and classified into the following groups: ZIKV IgG positive samples (N=22), ZIKV IgM positive samples (N=20), DENV IgG positive samples (N=61), ZIKV and DENV IgG positive samples (N=20) and flavivirus naïve samples (N=37). Testing of this well-characterized sample set allowed determination of accuracy of the NS2B-Grafted protein ELISA. 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.