TITLE: Bioinformatics analysis of mutations sheds light on the evolution of Dengue NS1 protein with implications in the identification of potential functional and druggable sites

Publication: Sharma, A., Krishna, S., & Sowdhamini, R. (2023). Bioinformatics Analysis of Mutations Sheds Light on the Evolution of Dengue NS1 Protein With Implications in the Identification of Potential Functional and Druggable Sites. *Molecular Biology and Evolution*, *40*(3). https://doi.org/10.1093/molbev/msad033

(Below text is taken from our published work and formatted as per the mentioned format. Supplementery data is also added after 'Impact of research in the advancement of knowledge or benefit of mankind')

Introduction:

Dengue is one of the most infectious and widespread vector-borne diseases, known to infect millions of people every year across the globe (Malavige et al. 2004). It is caused by single-stranded positive sense (+ssRNA) and non-segmented RNA virus belonging to the Flaviviridae family. Clinical symptoms range from mild to severe, which include dengue hemorrhagic fever (DHF) and dengue shock syndrome (DSS) (Malavige et al. 2004). Mild cases are treatable by available medication for the fever and fluid intake, but severe cases are still challenging and difficult to treat (Malavige et al. 2004). The main reason is the unavailability of effective vaccines, therapy, or antiviral drugs. Vaccines currently under clinical trials have under-performed on efficacy and effectiveness, which can be linked to the complex immune responses generated by heterotypic dengue viruses, such as antibody-dependent enhancement (ADE) (Boonnak et al. 2008; Beltramello et al. 2010; Screaton et al. 2015; Neufeldt et al. 2018) and cytokine storm. The size of the dengue virus genome is approx. ~ 11kb and is composed of three structural (Capsid, PrM, and Envelope) and seven non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, NS5) (Neufeldt et al. 2018).

Mutation rates in RNA viruses are generally high due to their error-prone RNA-dependent RNA polymerase-mediated replication (Drake 1993; Drake and Holland 1999; Duffy 2018), leading to high genetic diversity among RNA viruses (Sanjuán and Domingo-Calap 2021). Dengue serotypes share approx. 70% sequence similarity with each other (Weaver and Vasilakis 2009; Neufeldt et al. 2018). NS1 (Non-structural protein 1) is a 350 amino acids long (45-55kD) glycoprotein, which exists in multiple structural states such as monomer, dimer, and hexamer (Gutsche et al. 2011; Akey et al. 2014; Watterson et al. 2016). Many non-structural proteins (NS3 and NS5), along with NS1, are highly conserved in nature (Neufeldt et al. 2018). The present study mainly focuses on the possible reasons for NS1 protein conservation.

NS1 is a critical protein that plays an essential role in viral replication and disease pathogenesis (Akey et al. 2014; Fan et al. 2014; Rastogi et al. 2016). Different states of NS1 are reported to perform a diverse range of functions. NS1 dimer is the predominant state within the host cell, which is critical for viral replication, and modulating viral production through their interaction with other dengue proteins [structural proteins PrM and PrE (Scaturro et al. 2015)]. It is reported to interact with several proteins (Khadka et al. 2011; Silva et al. 2013; Karyala et al. 2016; Hafirassou et al. 2017) belonging to various pathways, such as innate immunity, ERAD pathway, OST complex, and DNA repair. On the other hand, hexamer (sNS1) is secreted outside the cell as a lipid cargo (Screaton et al. 2015), where it is recognized by host immune system. Antibodies generated against sNS1 act as diagnostic markers. However, some of the NS1 antibodies have been identified as auto-antibodies. These auto-antibodies interact with the host-proteins in endothelial cells and

cause plasma leakage (Lin et al. 2003; Avirutnan et al. 2006; Cheng et al. 2009; Malavige and Ogg 2017; Jayathilaka et al. 2018). sNS1 interferes with complement systems proteins and activates CD4 monocyte, leading to the release of inflammatory cytokines (Modhiran et al. 2015; Neufeldt et al. 2018). Therefore, understanding the basis of NS1 sequence conservation would help in identifying immunogenic regions and druggable sites for therapeutic targeting.

OBJECTIVES:

1.1 To understand why NS, a clinically and pathogenically important Dengue protein is highly conserved in nature?

Our initial hypothesis was based on multiple quaternary states of NS1, which might increase the number of residues that are not free to mutate. Such residues with functionally important regions could play an essential role in NS1 conservation. These residues can be created by forming new interfaces or structural rearrangement during the multimerisation process. For more understanding, NS1 sequences from complete viral protein sequences available in two public databases: NCBI (Brister et al. 2015) and VIPER databases (Pickett et al. 2012) were extracted and analyzed.

1.2 Development of sequence and structure analysis pipeline for predicting protein-protein interaction and druggable sites for therapeutically important viral proteins.

We finally showed the implication of our analysis in predicting possible functional (protein-protein interaction) sites used by NS1 protein and computationally validated them by blind docking. A solvent-exposed druggable site containing known critical residues was identified, and small molecule inhibitors were screened against them. MD simulations showed that some of these small molecules form stable interactions with NS1. The effect of mutation on other NS1 proteins (belonging to the family *flavivridae* and the influenza virus) showed similar results compared to the Dengue NS1 protein. This provides a new direction to apply the afore-mentioned approach in predicting the druggable site on other viral proteins.

Methods:

2.1 Data retrieval, processing, and Sequence analysis:

Complete viral protein sequences were downloaded from two publicly available databases, NCBI (Brister et al. 2015) and Viper (Pickett et al. 2012), respectively. NS1 sequences were retrieved and filtered to obtain a set of complete and unique sequences. Shannon entropy (Litwin and Jores 1992) was calculated to understand the diversity at each amino acid position. This study's reference strains were as follows: DENV1: Accession number KM204119.1, DENV2: Accession number U87411.1, DENV3: Accession number KU050695.1 and DENV4: Accession number KR011349.2 were retrieved from NCBI. Subsequently, an intense literature survey was done to identify structurally and functionally critical residues. A Python script was written to analyze the temporal and geographical distribution of NS1.

2.2 Model generations for NS1 protein of all Dengue serotypes:

a) Homology modeling for disordered loop regions in DENV 2 NS1 crystal structure:

The solved structure of Dengue serotype 2 was downloaded from PDB (Berman et al. 2000) (PDB ID: 406B) (Akey et al. 2014). In this model, the loop from the wing domain was incomplete. Therefore, the complete model was generated by modeling loops using multi-template homology modeling by MODELLER 9.22 (Eswar et al. 2006). Blastp (Altschul et al. 1990) was performed on the PDB database to find the closest homolog. The closest predicted homolog 'West Nile virus NS1' also had similar unresolved loop regions. Therefore, the second closest homologue, 'Zika virus NS1' (PDB ID: -5GS6) (Xu et al. 2016), with query coverage 99% and sequence identity of 53.60%, was considered as template. Corresponding missing loop regions were selected from the Zika structure, and homology models were obtained. Individual sequence identity for loop1 (9-14), loop2 (109-132), and loop3 (160-167) were 40%, 35%, and 75%, respectively. Models were filtered according to the DOPE score. The top three predicted models were then validated using SAVES 5.0 and ProSA server (Wiederstein and Sippl 2007). The final selected model was simulated at 300K temperature, 1 bar pressure for 100 ns in an OPLS force field (Jorgensen et al. 1996). GROMACS (Berendsen et al. 1995) was used for Molecular dynamic simulations.

b) Homology modeling of DENV 1, 3, and 4 NS1 proteins:

MODELLER 9.22 was used for homology modeling using the complete DENV2 NS1 model as a template. Obtained models were selected and validated as mentioned above.

c) NS1 Hexamer

The solved structure of DENV2 NS1 (PDB: 4O6B) is an asymmetric unit. Therefore, the hexameric conformation was constructed by applying the crystallographic 3-fold to PDB coordinates. For hexamer modelling of the rest of dengue serotype NS1, coordinates in 4O6B were used as a reference and the respective dimer model as a starting asymmetric unit. PyMol 2.4.0 was used to generate these bio-assemblies.

2.3 Relative solvent accessibility (RSA) calculation:

RSA of NS1 dimer for all four serotypes was calculated using JOY (Mizuguchi et al. 1998) from an in-house IWS server (http://caps.ncbs.res.in/iws/). Residues with a total side chain RSA value equal to or less than 7.0 were classified as buried residues, whereas residues with RSA value greater than 7.0 were considered accessible residues.

2.4 Effect of mutations on NS1 stability:

The effect of mutation on NS1 structure stability was predicted using FoldX 5.0 (Schymkowitz et al. 2005). Three types of mutations lists were created from the sequence dataset:

a) List of single mutations per amino acid position: This list consists mutations present in a residue position of all strains in comparison to reference strain.

b) Comprehensive list for virtual saturation mutagenesis: where each amino acid residue in the NS1 sequence was changed to the rest of 19 amino acid residues (no permutation and combination has been tried).

c) List of combinations of mutations present in a strain in comparison to reference strain: To achieve this, NS1 sequences having metadata about sample type, isolation date, region, and disease status were filtered from the complete set of sequences. This ensured that sequences under analysis are representative of the viral population circulating in the human population. Identical sequences (100% sequence similarity) were clustered, and were categorized into high (s>50), medium(20<s>50), low(1<s>20) and one-time (s=1), based on their cluster size (s). Representative sequence from each cluster was compared with reference sequence and list of mutation combinations was generated.

To predict effect of mutations on protein stability, each mutation list was passed to FoldX. It first repairs the reference protein structure, and then calculates energy difference between wild-type and mutated structures (ΔΔG). Calculation was performed in triplicates and mean ΔΔG score was used to classify mutations. Mutations are classified into three categories, stabilizing (ΔΔG < - 0.46 kcal), destabilizing (ΔΔG > - 0.46 kcal) and neutral (−0.46 kcal/mol < ΔΔG ≤ +0.46 kcal/mol). For detailed analysis, destabilizing and stabilizing groups were further divided into three sub-groups based on ΔΔG scores. Highly stabilizing: ΔΔG < −1.84 kcal/mol, stabilizing: −1.84 kcal/mol ≤ ΔΔG < −0.92 kcal/mol, slightly stabilizing: −0.92 kcal/mol ≤ ΔΔG < −0.46 kcal/mol, slightly destabilizing: + 0.46 kcal/mol < ΔΔG ≤ +0.92 kcal/mol, Destabilizing: + 0.92 kcal/mol < ΔΔG ≤ + 1.84 kcal/mol, highly destabilizing: ΔΔG > + 1.84 kcal/mol).

In case of virtual saturation mutagenesis, all stabilizing and neutral mutations were labeled as 1. On the other hand, all destabilizing mutations for a residue position were labelled as 1. Total aggregate score for each position was calculated and termed as Mutational capacity per residue positions (MC/RP). A Python script was used for calculation and plotting.

2.5 Molecular dynamic simulations:

MD simulation were used to validate some of the top predicted mutations/combinations. Simulation (in triplicate) were performed using GROMACS-5.2.2 (Berendsen et al. 1995). A standard procedure was adopted, including defining an orthorhombic box around the protein of interest with 10Å as a buffer distance. The box was solvated using a generic equilibrated 3-point solvent model (spc216.gro), and the system was neutralized by adding a counterpart ion. Both solvated and electro-neutral systems were minimized to avoid steric clashes energy before the production run, using the steepest descent minimization algorithm. Force <1000 kJ/mol/nm was considered as the cut-off for the energy minimization step. It was followed by two-phase step where the temperature and pressure of assembly were brought sequentially in optimum to the main production run. In the first phase, short MD simulation run of 100 ps was done under an NVT ensemble using a modified Berendsen thermostat with constraints in bonds involving H-bond. This was followed by a 100 ps-run at 300K and 1 atmospheric pressure using an NPT ensemble. Parrinello-Rahman barostat was used for pressure coupling. After relaxation, production MD was run in an NPT ensemble using an OPLS force field. For simulations, default parameters of leap-frog integrator were used. The temperature and pressure were kept at 300K and 1 bar, respectively. The production MD was run for 100 nanoseconds. 'rms' an utility of GROMACS was used to calculate RMSD (Root mean square deviation).

2.6 Amino acid interaction change analysis

Each scenario (wild, single mutation, and combination of mutations) was simulated for 100ns (method 2.5), and the RMSD of each trajectory was calculated. A uniform stabilization time for all trajectories were determined, and trajectories were split from that point to the end of the simulation (100ns). Conformation at every nanosecond was obtained, and amino acid interaction around residue/mutation of interest was identified using Schrödinger (Schrödinger release 2022-3: BioLuminate, Schrödinger, LLC, New York, NY) script:protein_interaction_analysis.py. Command $SCHRODINGER/run protein_interaction_analysis.py -structure {0} -group1 {1} -group2 {2} -outfile {3} was used to identify changes in amino acid interaction, where group1 represents residue of interest and group2 represents rest of the residues. Default parameters were used to predict H-bonds, van der Waal collisions, pi-pi stacking, and salt-bridge formation. Changes induced by a mutation in a single chain of NS1 dimer were considered during analysis. This was done to uniformly compared changes between monomer and dimer state. In the dimer case, the exact number of any interaction formation can be obtained by multiplying the mentioned value by 2.

2.7 Protein-protein docking

Blind docking was performed using a HDOCK server (Yan et al. 2017). It predicts the binding complexes between two proteins using template-based modelling and *ab-initio* docking as hybrid docking algorithms. It provides an HDOCK score which can be used to rank the models. Docking was performed using default parameters.

2.8 Virtual Screening of inhibitors

Drugbank database (Wishart et al. 2018) was first downloaded in SDF format (Standard Delay Format). Virtual screening has been done for Dengue serotype 2 NS1 protein only. Downloaded compounds were prepared for screening using the Ligprep module in Schrödinger (Madhavi Sastry et al. 2013). OPLS3e (Harder et al. 2016) force field and pH 7.4 was used for ligand preparation. Specified chirality was retained, and one structure per ligand was specified as an output. Protein was prepared for docking using Protein-preparation wizard (Madhavi Sastry et al. 2013) in Maestro Schrödinger (Schrödinger Release 2022-3: Maestro, Schrödinger, LLC, New York, NY, 2021.) Docking is performed on a site having conserved solvent-exposed residues with zero mutational capacity and functional relevance. A receptor grid around the docking region was specified and generated using a receptor-grid generation module in Glide (Halgren et al. 2004). Residues from predicted sites were specified, and rotatable bonds across the site (if any) were enabled during grid generation. Protein-ligand docking was performed using a glide docking module in the Schrödinger suite. We followed 3-step docking as mentioned in suit: a) High-throughput virtual screening (HTVS) to narrow the list of potential ligands. According to their docking score top, 10 percent ligands were taken further for b) Standard Precision (SP) mode, and finally, the compound obtained from the SP step with a docking score less than 5 kcal/mol was screened by c). Extra Precision mode (XP) (Friesner et al. 2006). MM-GBSA (Molecular Mechanics energies combined with Generalized Born and Surface Area continuum solvation) score was then calculated for the top 20% of compounds obtained from the XP step. After MM-GBSA calculations, the protein-ligand interaction fingerprint was generated using the SIFT module from Schrödinger. The screened hits were clustered based on the fingerprint and Tanimoto coefficient (Halgren 2007; Wang et al. 2015). Potential compounds were selected according to their interaction with critical residues, mode of action, MM-GBSA score, XP score, and cluster size. Top six compounds were identified as potential inhibitor candidates, and were taken forward for the MD simulation run. To streamline the virtual screening pipeline, we have used Desmond, a utility from the Schrödinger suite, to perform the simulation runs. The pipeline for MD simulation was as followed in our previous study (Sharma et al. 2020).

2.9 Statistics

Odds ratios calculate any association between two categorical features (Szumilas 2010), and were calculated using a Python script for statistical correlation. The significance of the correlation was determined by p-value. High and positive odds ratios with significant p-values suggested a strong association between the two features.

Results

3.1 Generation of NS1 monomer, dimer, and Hexamer models

The structure of the DENV2 NS1 dimer is available in the protein data bank, ID 4O6B. However, the structure has some unresolved loop regions, which were modeled using multi-template homology modeling (Figure 1a) and Zika virus NS1 structure (PDB ID 5GS6) as a template (please see Methods for details). Predicted models were sorted according to the DOPE score, and then validated by two web servers: ProSA and SAVES 5.0 (data not shown). MD simulations were performed in triplicates for the best-predicted model in the OPLS force field for 100 nanoseconds(ns) duration. RMSD (Root mean square deviation) (Figure 1b) was calculated from the MD simulation trajectory. The model was found to be stable throughout the simulation run. Dengue sequences approved by WHO and CDC were used as references (Añez et al. 2016) to model NS1 of other dengue serotypes. The above-modelled DENV2 NS1 model was used as a template, and the protocol (Methods 2.2b) was followed for the generation, validation, and selection of the best NS1 dimer model for each dengue serotype (Supplementary Figures S1). NS1 is majorly present in the dimer state within the cell, but for detailed analysis, a monomer unit from the dimer was also extracted. MD simulation run (100ns) was performed to relax the dimer interface in the monomeric unit (Supplementary Figures S1). For the hexameric units, lipid-free models were constructed by applying the crystallographic 3-fold symmetry on the deposited DENV2 NS1 dimer coordinates (PDB ID 4O6B) (Figure 1c). A similar approach (Method 2.2b) was used to construct the hexameric model for the rest of the serotypes (Supplementary Figures S1).

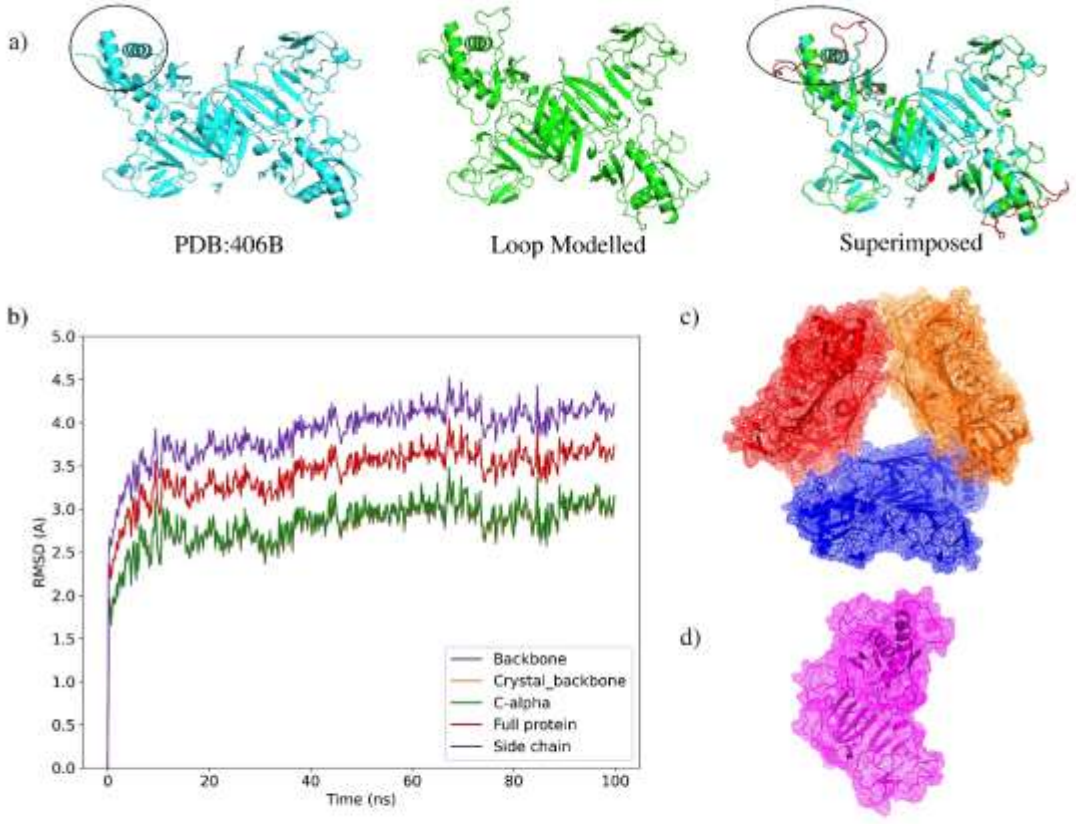

Figure 1: Loop modeling of DENV2 NS1 followed by validation through MD simulation.
a) The unresolved loop regions in solved structure of the DENV2 (PDB:406B) (black circle) were modelled using Zika NS1 as template. Panel shows NS1 structure, loop modelled NS1, and superimposed structure between 4o6b and modelled NS1. Missing region (black circle) and modelled loops are shown in red.
b) The stability of modelled NS1 is validated by MD simulation run for 100ns (n=3). RMSD (Root mean square deviation) of modelled structure (Full protein) and structural components (Backbone, crystal backbone, C-alpha atoms, and side chains) for complete simulation trajectory was plotted, where x-axis represents time (ns) and y-axis represent RMSD (Å). Stabilization of RMSD trends suggests the modelled structure's stability.
c) The DENV2 hexamer model was constructed by applying the crystallographic 3-fold to the deposit DENV2 NS1 dimer coordinates. Each color represents one dimer unit.
d) Monomer unit was extracted from modelled NS1 structure, and simulated for 100ns (n=3) to relax NS1 dimer interface.

Table 1: Sequences used in the analysis. Two publicly available databases (NCBI and VIPER) were used to extract and process sequences.

Serotypes	Sequences in the NCBI database	Sequences in the VIPER database	COMBINE (Unique accession Number)
DENV1	1729	2165	2555
DENV2	1401	1589	1931
DENV3	871	997	1181
DENV4	344	254	420

3.2 Sequence analysis

To perform sequence analysis, DENV1 (n=2555), DENV2 (n=1931), DENV3 (n=1181), and DENV4 (n=420) sequences were extracted from NCBI and VIPER database, and used for analysis (Table 1). Each sequence was compared with reference sequence and mutations per residue positions were identified. Number of mutations were found in DENV1 (n=364), DENV2 (n=741), DENV3 (n=956), and DENV4 (n=794) sequence dataset. Given the large dataset taken for DENV1, there was relatively fewer mutations due to the geographical and temporal confining of sequences. On the other hand, sequences of other dengue serotypes belonged to diverse geographical locations and periods (data not shown). Dengue NS1 is one of the most conserved non-structural proteins of the dengue virus. To understand the extent of diversity within NS1 protein, Shannon entropy (SE) score per residue position (Figure 2a, 2c) was calculated. Similar SE scores among Dengue serotypes suggests similarity in their sequence conservation. Low SE scores reflect high conservation at that position. The DENV2 NS1 SE scores were represented on the DENV2 dimer model (Figure 2b).

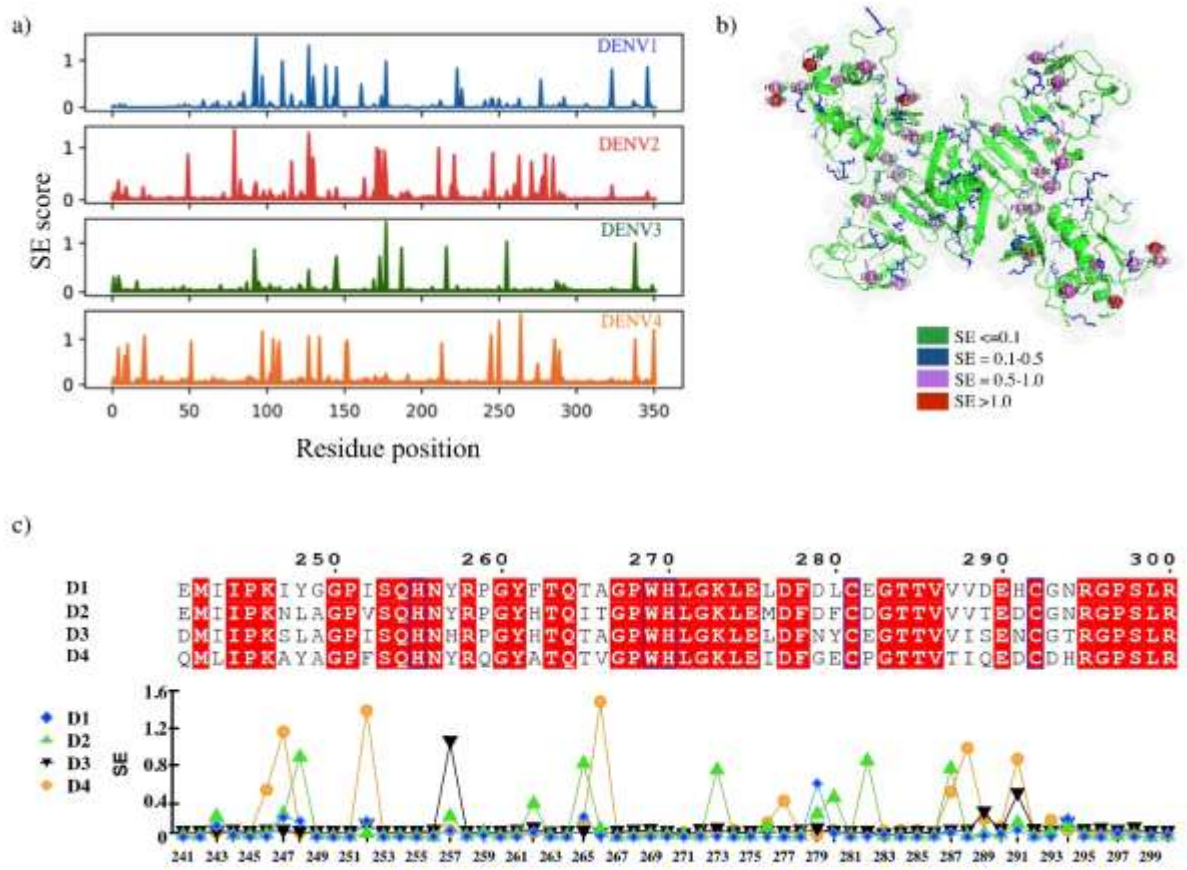

Figure 2: Shannon entropy (SE) scores of each residue position of NS1 protein for all Dengue serotypes.

a) Shannon entropy (SE) scores (y-axis) per residue positions (x-axis) were calculated for all Dengue serotype NS1 proteins. SE scores represent the frequency of different amino acid types at that position. It ranges from 0 (only one residue is present at that position) to 4.322 (all 20 residues are equally represented in that position). Majority of NS1 residues were observed to have low SE scores. Low SE scores suggest conserved residues.
b) The SE scores of DENV2 NS1 are represented in detail on modelled structure. Region of structure are marked as per SE score <=0.1 (fully conserved), between 0.1 to 0.5 (Blue), between 0.5 to 1.0 (Magenta), and >0.1 (Red). SE score 0.1 was identified as cut-off, and positions were categorized into observed-conserved (SE < 0.1) and variable (SE >0.1) regions.
c) A segment of NS1 protein sequence (240aa-300aa) and respective SE score per position is shown. This represents the basis of our sequence analysis, where we attempt to understand possible reasons of varied SE scores across whole protein. Here, red represents identical residues among serotypes. The Shannon entropy scores per residue position are shown as a line graph. D1 represents Dengue serotype 1, D2 represents Dengue serotype 2, D3 represents Dengue serotype 3, and D4 represents Dengue serotype 4.

A frequency distribution analysis of SE scores was performed. SE score of 0.1 was identified as the cut-off (Supplementary Figures S2) from distribution analysis to understand the diversity among residue positions. This led to the classification of residues into two categories: (a) Observed low SE (SE score $\leq$ 0.1), suggests conserved positions, (Figure 2b) and (b) Observed high SE (SE score > 0.1), represents comparatively variable positions. Residues with low SE were more in number as compared to diverse positions (Figure 2a). The possible reason behind residue conservation could be its role played in protein stability, functionality, or both. In this study, we focused on the role of structurally critical residues in NS1 sequence conservation. Therefore, we asked what will be the effect of observed mutations (obtained from NS1 sequences) on NS1 structure stability.

3.3 Majority of single mutations are destabilizing in nature

To predict the effect of identified mutations on protein stability, FoldX was used and free energy change between ($\Delta\Delta G$) wild-type and mutated protein structure ($\Delta\Delta G$ = ΔG mutant - ΔG wild-type) was calculated. The $\Delta\Delta G$ score

determines the effect of mutations on protein stability, and based on the score, each mutation was classified into three categories (neutral, stabilizing, and destabilizing). In the case of DENV2 dimer, 58.8% of mutations were predicted to be destabilizing in nature, further divided into subgroups (39%-highly destabilizing, 11.2% destabilizing, and 8.6%-slightly destabilizing) followed by neutral (20.9%). In case of stabilizing categories, 20.3% of mutations were further classified into subgroups i.e., 8.1%-slightly stabilizing, 7.7%-stabilizing and 4.5%-Highly stabilizing (Figure 3a). Similar analysis was performed on all NS1 states of each dengue serotype (Supplementary Figure S3a). The majority of the single mutations were predicted as destabilizing, followed by neutral and stabilizing in nature (Figure 3a and S3a).

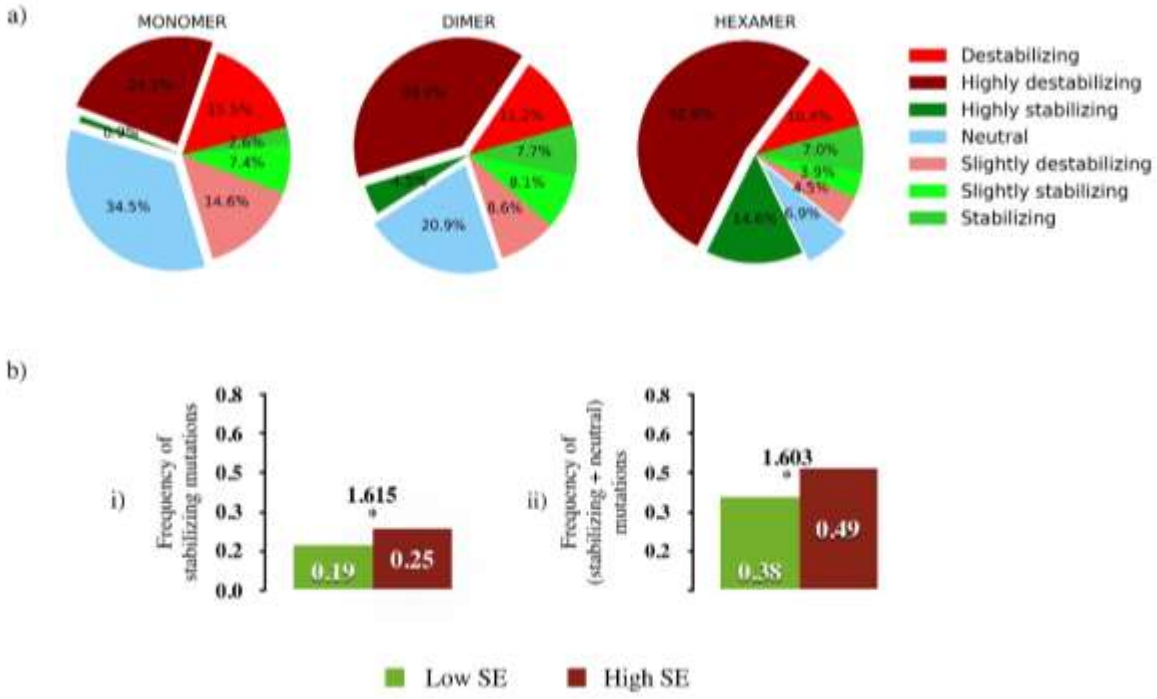

Figure 3: The effect of know single mutation on NS1 protein stability.
a) Destabilizing mutations constitute the majority of known single mutations in DENV2 NS1. Mutations were classified based on free energy difference (ΔΔG) calculated by FoldX 5.0. Three pie charts representing the frequency of mutations in each category for NS1 state (Monomer, Dimer, and Hexamer) are shown.
b) The frequency of stabilizing and neutral mutations was significantly higher in observed-variable (high SE) vs. observed-conserved (low SE) residue positions. The numbers above pairs of bars are the odds ratios, and stars below them denote statistical significance (p-value) calculated according to one-tail Fisher's exact test (* - $0.01 < P < 0.05$, ** − $0.001 < P < 0.01$, *** − $P < 0.0001$ and no star for non-significant p-value).

Frequency distribution of ΔΔG of mutations within observed-conserved and variable regions was performed. The frequency of stabilizing mutations was higher in observed-variable region (high SE positions) with an odd ratio of 1.615 in DENV2 NS1 dimer (Figure 3a-i). High odd ratios suggest that the probability of finding a stabilizing mutation in the observed-variable (high SE) position is 1.6 times higher compared to observed-conserved (low SE) positions. p-value for the DENV1 odds ratio was not significant. This could be due to the presence of less number of mutations in DENV1 dataset (Supplementary Table 1). Stabilizing mutations suggest an advantageous structural change in the protein. Also, these mutations could be functionally important based on their presence or role in protein-protein interaction regions. Predicting the effect of these mutations on NS1 functionality is not considered for this study. Therefore, stabilizing and neutral fractions were considered as a complete set of functionally important mutations. The probability of finding these (neutral+stabilizing) mutations in observed low *versus* high SE positions was calculated. Similar to stabilizing mutations, the frequency of these stabilizing mutations was higher in the variable (high SE) positions with significant odd ratios (Figure 3b-ii). There was similar association between stabilizing, stabilizing+neutral mutations and variable positions but with a varied extent of enrichment (different odd ratio values) for other dengue serotypes

(Supplementary Table S2). This suggests that nature of effect of mutation on protein stability can be one of the factors result in formation of observed-conserved/variable positions.

On the other hand, destabilizing mutations were also present in observed-variable regions. Past studies suggest that the destabilizing effect of one mutation can be neutralized by a nearby compensatory substitution (Friedrich et al. 2004; Knies et al. 2008). Therefore, to understand the presence of destabilizing mutations in the observed-variable fraction, we hypothesized that the emergence of these mutations is due to the presence of some compensatory mutation which are nearby or compensatory effect by whole combination of mutations. A systematic study of each possible combination of mutations is computationally intensive. Therefore, we only focused on the most frequent combinations present in our DENV2 NS1 sequence dataset (Method 2.4c).

3.4 Structural changes caused by a combination of mutations lead to the selection of destabilizing mutations in the viral population

In the previous section, majority of observed single mutations were destabilizing in nature (Figure 3a). However, most of the combinations of mutations (which refer to simultaneous multiple mutations) were predicted as stabilizing in dimer and hexamer states (Figure 4a). This leads to the possibility of compensatory behavior within combination of mutations. Interestingly, this trend was reversed in the case of monomer, and majority of combinations were predicted as destabilizing in nature (Figure 4a). Furthermore, highly frequent mutation combinations in viral population tend to be stabilizing in dimer and hexamer (Figure 4a) states in comparison to single occurrence cases. This suggests that in most cases, combination of mutations which maintain structural integrity of NS1 functional states are selected. MD simulations were performed to further confirm the differences observed between effect of combination of mutations on monomer vs. dimer state.

Two of the high-frequency combinations (Supplementary Table 4) a) comb165 (H130Y, H51Q, S81A, T266A, V178A, and V287I) b) comb167 (D282E, I213M, I265T, K173R, L248F, N223S, and T266A) within DENV2 dataset (method 2.4) were selected for analysis (Figure 4b). Comb165 was predicted to be ‘slightly stabilizing’ in monomer ($\Delta\Delta G$=-0.806704 kcal/mol) (Figure 4c-i) and ‘highly stabilizing’ in both dimer ($\Delta\Delta G$=-5.27766 kcal/mol) (Figure 4d-i) and hexamer states ($\Delta\Delta G$=-13.8669 kcal/mol) (Supplementary Figure S4d). On the other side, comb167 was predicted as ‘highly destabilizing’ in monomer ($\Delta\Delta G$=4.15687 kcal/mol) (Figure 4c-ii), but ‘highly stabilizing’ in both dimer ($\Delta\Delta G$=-2.64617 kcal/mol) (Figure 4d-ii) and hexamer ($\Delta\Delta G$=-7.04378 kcal/mol) (Supplementary Figure S4d) states. Individual mutations within these combinations had varied effects on NS1 stability (Figure 4c-d, Supplementary Table 5).

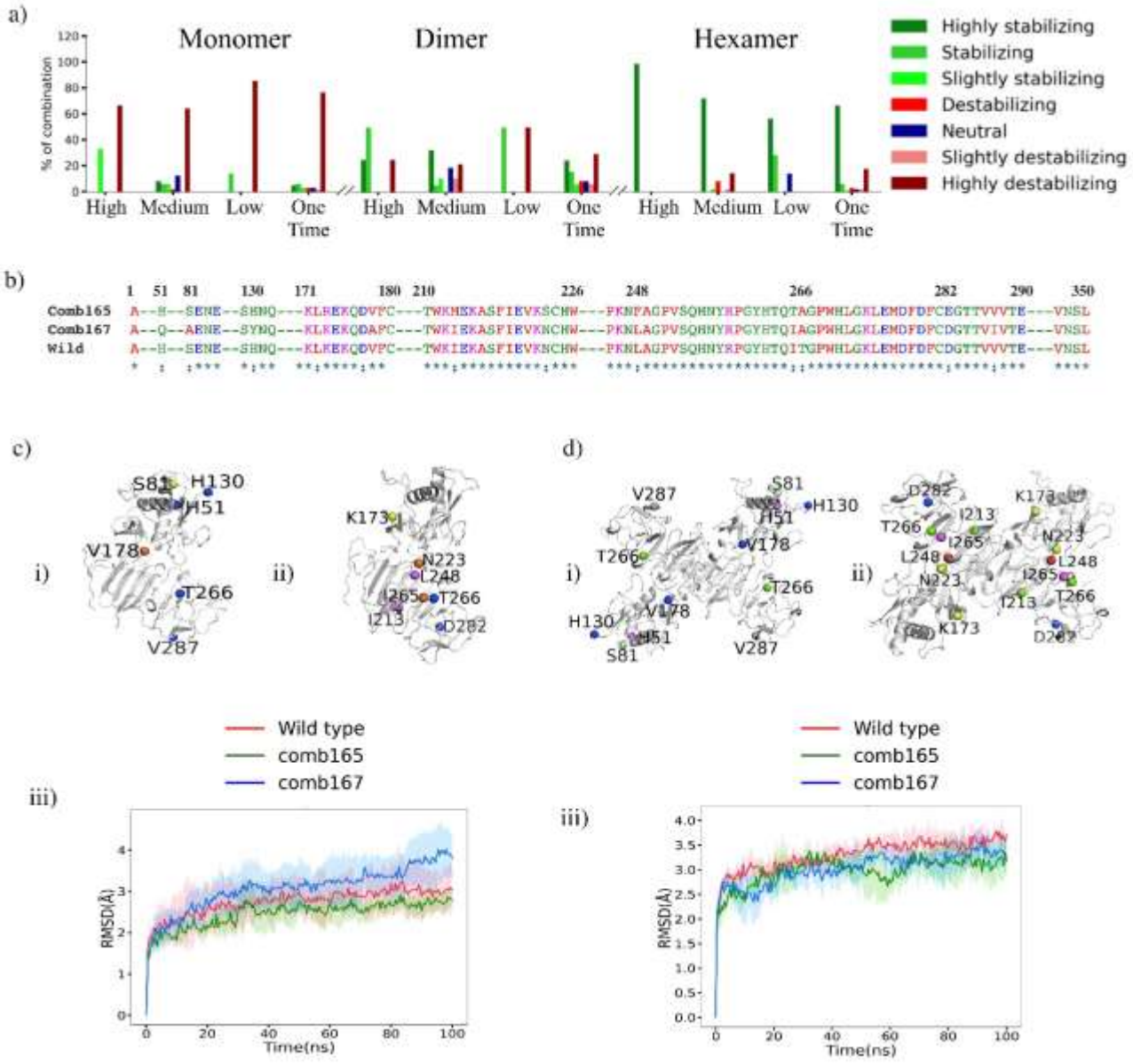

Figure 4: Varied effect of combination of mutations on NS1 states (Monomer vs Dimer) stability.
a) Most combinations of mutations were predicted as stabilizing in dimer and hexamer states, and destabilizing in monomer state. Mutation combinations were grouped into high (s>50), medium (20<s>50), low (1<s>20) and one-time (s=1) (x-axis), based on their observed frequency (cluster size(s)) in sequences from patient samples. Y-axis represents percentage of combinations having varied effect on stability (bar) within each group across NS1 states. Color of individual bars represents the effect of mutation combination on NS1 stability.
b) Multiple sequence alignment highlighting mutations present in comb165 and comb167 compared to wild type amino acid sequence. '*' represents sequence conservation, ':' represents mutations.
c) & d) MD simulation runs (n=3) were performed to validate observed effect of mutations combinations, i) comb165, and ii) comb167 on DENV2 NS1 monomer and dimer state, respectively. Two mutation combinations are represented on DENV2 NS1 monomer (c-i & ii) and dimer state (d-i & ii) structures. Stabilizing (green), neutral (blue) and destabilizing (red) mutations are shown in different colors. Root mean square deviation (RMSD) of protein structure conformer at every nanosecond was compared with starting structure of simulation. It is represented as line plot for both NS1 monomer (c-iii) and dimer (d-iii), where x-axis represents time (ns) and y-axis represents RMSD (). Solid line showed average RMSD after three runs, and the shaded region around the solid line represents standard deviation in RMSD value at that time point. RMSD values as well as trends for stabilizing mutation combination were comparable with wild-type. For destabilizing combination (comb165 in monomer), average RMSD was higher, and trend was not stabilized during whole simulation run.

MD simulation runs of 100ns were performed in triplicate for each combination, and protein structure conformers at each nanosecond were sampled. Root mean square deviation (RMSD) between each conformer and reference structure was calculated (Figure 4c-iii & Figure 4d-iii). The change in RMSD trend between wild-type and mutated structure was used to analyze the effect of mutation combinations on structure stability. Monomer with comb167 had higher RMSD in comparison to wild-type, and RMSD trajectory was not stabilized within 100ns (Figure 4c-iii). Both

observations suggested that comb167 was indeed destabilizing the monomer. On the other hand, stabilizing combination (comb165) showed similar RMSD trends compared to wild-type (Figure 4c-iii). In the dimer case, both combinations' RMSD trajectories were stabilized (Figure 4d-iii), and higher stabilization combination (comb165) had lower average RMSD (2.96 Å) in comparison to both wild type (3.30Å) and comb167 (3.02Å). These results were in line with the FoldX predictions.

FoldX considers multiple parameters including various interactions, solvation energies, and entropies to predict the effect of mutation on protein structure. However, in this study, we only focused on the changes induced in amino acid interactions around a mutated position. In order to achieve this, individual mutations in each combination (monomer and dimer) were simulated for 100ns as triplicate runs. RMSD trends were analyzed, and 60ns was considered as uniform time-point for trajectory stabilization. Protein conformation at every nanosecond (60-100ns) was extracted, and changes in the amino acid interactions across mutated position were calculated using the Schrodinger suite (method 2.6). During analysis, changes in the number of hydrogen bonds (H-bond) were frequently observed, and in some cases, the formation of salt-bridge and pi-pi stacking was observed. For example, in case of K173R mutation, changes in the number of H-bond formations were evident in both monomer and dimer case (Supplementary Figure S4-a). Changes induced around mutated residue in a single chain of dimer were considered for analysis to keep uniformity during comparison between monomeric and dimeric states.

Comb165 consists of only one destabilizing mutation H51Q (Figure 4b-i, figure 4c-i), while the rest were either neutral or stabilizing (Supplementary Table 5). On the other hand, comb167 showed interesting observation where it is predicted as stabilizing in dimer but highly destabilizing in the monomer state. In monomer, comb167 consists of multiple individual destabilizing mutations (Supplementary Table 5) such as N223S, L248F, I213, and I265T (Figure c-ii), whereas in dimer, two of these mutations (I213M and N223S) were stabilizing in nature (Supplementary Table 5). A pair of mutations can be considered compensatory if individual mutations have the opposite effect on protein stability, and are spatially close to each other (Davis et al. 2009). In this context, two mutation pairs with possible compensatory behavior, a) N223S (stabilizing) and L248F (destabilizing), b) I265T (destabilizing) and T266A (stabilizing) were observed. The N223S-L248F pair was spatially close (~4.6Å apart), whereas I265T-T266A was linearly present one after another. Within comb167, the number of H-bonds present in L248F proximity decreases compared to wild and single mutation conditions (Supplementary Figure S4b-i), and an opposite trend was observed for N223S (Supplementary Figure S4b-ii). Here, bond formation across L248F-N223S suggested that the changes resulting in the loss of bonds around one mutated position (L248F) could be compensated by increasing the number of bonds around another position 'N223S', present within the same region.

To further confirm compensatory behavior between these two mutations, a double mutant (L248F and N223S) was modeled and simulated for 100ns as triplicate runs. Presence of L248F mutation in N223S background results in loss of H-bonds across residue position 223 (Supplementary Figure S4c-ii). In the case of position 248 (Supplementary Figure S4c-i), increased retention in H-bond formation was observed in one simulation run, whereas for other two runs, changes

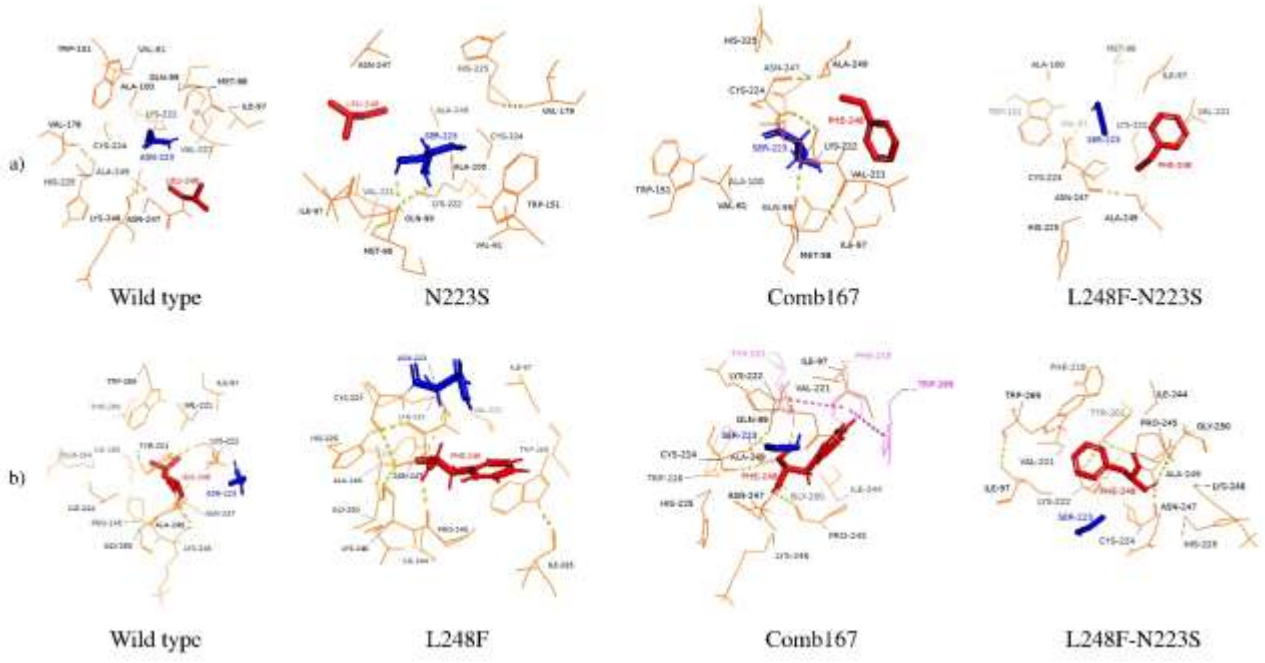

Figure 5: Change in amino acid interaction of a mutated residue in single vs combination of mutations.
a) & b) Changes in amino acid interactions around 5Å radius. Residue a) 223, and b) 248 is shown in various conditions. Yellow dotted line represents H-bond and magenta lines represent pi-pi interaction. This is a snapshot of interactions at the end of the simulation run. Changes in type of residues, and formation and breaking of new bonds can be observed in the snapshots.

were comparable to single mutation. This compensation in retention of H-bond formation within double mutants demonstrates the influence of mutations on interaction network of each other. Such influence could be compensatory over a broad range within combination of mutations. Figure 5 represents a snapshot of amino acid change around wild vs single vs double vs combination of mutations in 223 (Figure 5-i) and 248 (Figure 5-ii) residue position respectively. A similar analysis was performed for I265T and T266A mutation pair (Supplementary Figure S4c). The effect of I265T mutation on the T266A environment was apparent from the difference in average number of H-bonds formed around T266A in I265T absent (~0.5) (comb165) vs present (~1.5) condition (comb167) (Supplementary Figure S4c-i). However, the effect of T266A mutation w.r.t to bond formation around I265T was not observed (Supplementary Figure S4c-ii). There are multiple factors which can contribute to mutation stabilization, therefore, compensatory behavior between this pair could be due to other mechanisms/parameters which are not considered in this study.

The observed results demonstrate that the reason for the varied effects between combination *vs.* individual mutations on NS1 protein stability could be based on the changes observed in the local environment around the mutated residue (Figure 5a&b) (Supplementary Figure S4). Our results suggest that combinations of mutations play a role in the emergence and selection of destabilizing mutations in the viral population. Studies have shown that mutations have long-range effect on protein conformational change and stability (Unal et al. 2013; Bigman and Levy 2018). Therefore, the presence of compensatory mutations (immediately close to each other) can be one of the mechanisms to neutralize the destabilizing effect. Further investigation is required to uncover other probable mechanisms involved in neutralization of a destabilizing mutation in a combination of mutations.

During analysis of identified single mutations, an interesting observation was made in DENV2 NS1 dimer. Every mutation at several residue positions (such as A122, M98, M124) was predicted as destabilizing in nature. Buried regions of protein are more sensitive to mutations compared to solvent-exposed regions (Bhasin and Varadarajan 2021).

Therefore, solvent accessibility status of these residues could explain the observation. To determine their accessibility status, relative solvent accessibility (RSA) was calculated (Method 2.3). A122 (RSA =30.9) and M124 (RSA = 62.6) were found to be solvent-exposed, whereas M98 was buried (RSA=0.8). Similar residues (both exposed and buried) were also identified in other dengue serotypes. This suggests that irrespective of their accessibility, these residue positions could be structurally critical, resulting in every mutation (till date) to be destabilizing in nature. Further, Figure 3a clearly shows the change in distribution of destabilizing and stabilizing mutations among different NS1 states. This could be due to NS1 state-dependent effect of mutations on stability. Therefore, understanding the distribution of such sites across NS1 states, and their association with observed-conserved vs. observed-variable positions would unravel the role of NS1 protein states in its diversity.

3.5 Prediction of the virtual-conserved and -variable region in DENV NS1 and their association with sequence variability

Even if the trends of the multiple mutations from patient samples indicate an overall stabilizing effect on NS1, we wanted to ensure that our observations are not limited or biased by the availability of patient samples. Hence, FoldX was used next to perform an extensive virtual saturation mutagenesis (Chauhan and Sowdhamini 2022), for predicting mutationally impermeable regions, where each residue position in the reference structure was sequentially and individually mutated to the rest of the 19 amino acids. The output of each substitution was characterized on their $\Delta\Delta G$ score (Method 2.4). For normalization, each destabilization mutation was labeled as 0, while neutral and stabilization mutations were labeled as 1. Normalized scores for every mutation on reference position were then aggregated, and a final score depicting MC/RP was obtained (Figure 6a). MC/RP score ranged from 0 to 19. MC/RP score equal to zero suggests all changes are destabilizing, and thereby, anticipated to not accommodate any mutation. Such residues were termed as 'virtual-conserved'. On the other side, MC/RP score equal to 19 for a residue suggests all changes are either neutral or stabilizing or both, and were labelled as virtual-variable residues.

The Pearson correlation coefficient was calculated to check variation in MC/RP scores among NS1 states. The correlation coefficient of the MC/RP score of monomer with dimer and hexamer states was 0.13 and 0.07, respectively, whereas the correlation coefficient between dimer and hexamer state was 0.95. This correlation suggested variation in MC/RP scores of the same residue positions across NS1 states, especially from monomer to higher order. Therefore, a detailed comparison between DENV2 NS1 states was carried out. Out of 350 residues in NS1, 70, 88, and 105 residues, were predicted as virtual-conserved in monomer, dimer, and hexamer states, respectively (Figure 6b-ii). 24 residues G4, C56, M68, D93, I97, I136, C144, G200, W211, W266, M242, G250, G267, P268, F280, G283, C292, L308, W312, C313, P320, P321, L322, G329 were common among all three states (Figure 6b-i). Both solvent-exposed (G4, C56, D93, I97, W211, F280, G283, L308, P320, G329) as well as buried (M68, I136, C144, G200, W211, W266, M242, G250, G267, P268, C292, W312, C313, P321, L322) residues were present in this fraction, which follows the observation found in Section 1.4. Comparable results were obtained for the rest of the Dengue serotypes (Supplementary Figure S3b).

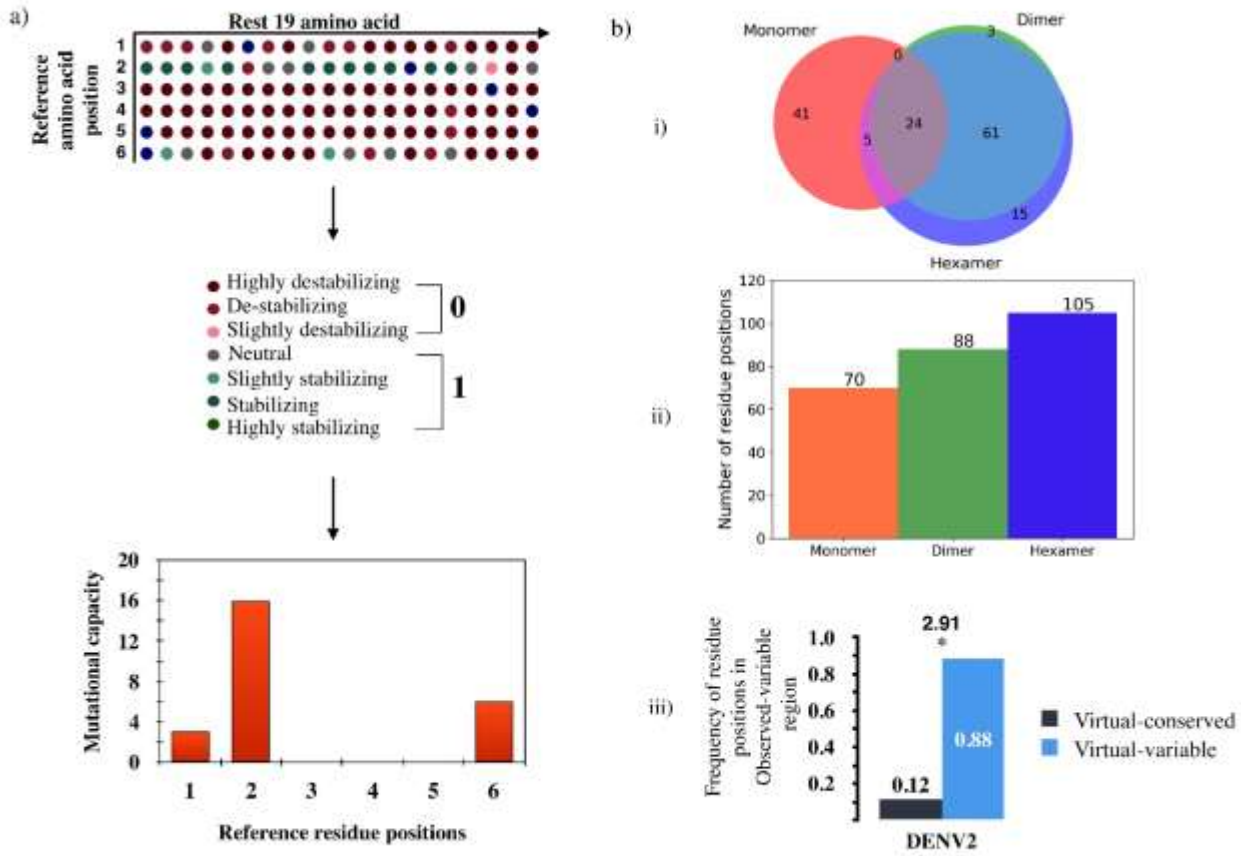

Figure 6: Prediction and enrichment of virtual-conserved/variable positions in observed-conserved/variable position of DENV2 NS1 protein. a) Workflow used to predict the mutational capacity of a reference residue position for DENV2 NS1 protein. Firstly, each reference residue position was mutated sequentially to the rest of the 19 amino acids, and mutation were classified based on their ΔΔG scores. To normalize the scores, destabilizing mutations were labeled as 0, whereas stabilizing and natural were labeled as 1 and aggregated (each bar) to calculate MC/RP score. The x-axis represents reference residue positions, whereas y-axis represents the mutational capacity score (MC/RP).
b) The virtual-conserved residues were found to vary among different DENV2 NS1 states, represented by a Venn diagram (b-i). The number of such residues was observed to increase with increase in NS1 state (b-ii). The enrichment of virtual-conserved residues in observed-variable (high SE) vs. observed-conserved (low SE) region was found through frequency distribution analysis (b-iii). Here, each bar represents the frequency of observed – variable residues in virtual-conserved vs virtual-variable positions. The number above the pair of bars is the odds ratio, and stars below them denote statistical significance (p-value) calculated according to one-tail Fisher's exact test (* - $0.01 < P < 0.05$, ** − $0.001 < P < 0.01$, *** − $P < 0.0001$ and no star for non-significant p-value).

The MC/RP scores of the NS1 dimer state were used for further analysis in this study. It was selected because (a) the dimer state of NS1 is predominant within the cell, (b) we assumed that immediate dimerization of NS1 monomer could act as an initial purifying step, and any mutation resulting in misfolding of monomer will be instantly degraded, (c) any mutation in monomer affecting dimer formation will result in no viral replication, d) most frequent mutation combinations were stabilizing in dimer (Figure 4a), and (e) dimer acts as the basic unit for hexamer formation. MC/RP scores of dimer and hexamer are also highly correlated (0.95). Therefore, if MC/RP score is really a good predictor of sequence diversity, then the probability of mutations occurring in NS1 protein would depend on the MC/RP score (virtual-conserved/variable) of the NS1 dimer state.

Validation of predicted MC/RP scores before their usage for further analysis is essential. Therefore, we assumed that the predicted MC/RP scores will align with observed-critical residues. Specifically, experimentally proved 'structurally critical residues' were expected to be virtual-conserved type. An extensive literature survey was carried out and a set of known critical residues of the DENV2 NS1 protein (Pryor and Wright 1993; Pryor and Wright 1994; Wallis et al. 2004; Somnuke et al. 2011; Scaturro et al. 2015) were identified (Table 2; Supplementary Table 3). MC/RP score

of zero was observed for eight Cysteine residues important in structural stability and dimer formation. This observation suggests that other virtual-conserved residues might be structurally critical.

Table 2: Some of the known critical residues of DENV2 NS1 protein and their respective Shannon entropy (SE), relative solvent accessibility (RSA), and mutational capacity per residue position (MC/RP) scores in the dimer state.

Importance	Amino Acid residue	Shannon entropy (SE)	Relative solvent accessibility (RSA)	MC/RP scores (dimer)	Reference
Structure integrity	**C5**	0.01	55.2	0	(Pryor and Wright 1993; Wallis et al. 2004; Akey et al. 2014; Edeling et al. 2014; Scaturro et al. 2015)
	C56	0.01	22	0	
	C180	0.01	1.9	0	
	C292	0.01	2.2	0	
	C313	0.01	1.1	0	
	C314	0.012	0	0	
	C317	0.012	0	0	
	C330	0.012	12.8	0	
critical for viral replication	**Y159**	0.00	43.4	2	(Scaturro et al. 2015)
	G160	0.01	6.1	0	
	W8	0.03	88.9	5	
	F161	0.01	85.3	3	
	G162	0.02	123.4	0	
	P320	0.024	51.2	0	
	P321	0.012	0	0	
Less effect on viral replication but huge effect on viral production	**S115**	0.03	108.7	19	
	W116	0.01	63.3	9	(Scaturro et al. 2015)
	D181	0.03	13.2	13	
	T302	0.02	44.2	14	
Important glycosylation site	**N131**	0.02	31.3	9	(Pryor and Wright 1994)
	N208	0.03	66.5	17	

On the other hand, for functionally critical residues (effect only viral replication or viral production) (Table 2), MC/RP scores were within the range of 0 and 19. Observed-conserved type and varied MC/RP scores of such positions were expected because of purifying selection toward maintaining their functionality. In this fraction, in addition, some residues (G160, G162, P320, and P321) were also grouped within virtual-conserved type. As the mechanism of action for these residues is unknown, it is quite possible that there is a replication inhibition by structural changes or backbone conformational constraints caused by them (Scaturro et al. 2015). The emergence of known structurally critical residues in our predicted set of virtual-conserved residues validates our approach.

To understand the role of the virtual-conserved residues in NS1 diversity, we hypothesized that there would be a significant enrichment of these residues in observed-conserved (low SE) regions. The frequency of virtual-conserved residues in observed-conserved regions, and virtual-variable residues in observed-variable regions, was higher with significant odd ratios (Figure 6b-iii & Supplementary Table 1). Figure 6b-ii shows an increase in the formation of virtual-conserved residues across NS1 states (from monomer to hexamer). Therefore, the quaternary arrangement of NS1 affects its protein evolution by increasing the number of virtual-conserved sites, and thereby restricting overall mutational space.

There is a significant association between predicted (virtual-conserved or variable) and observed (low or high SE) sequence diversity. However, a few contradictory results were observed in analysis. The proportion of virtual-variable residues in observed-conserved regions was considerably high (Supplementary Table 1). Such observations might be due to the reasons that (a) these residues are functionally critical (Table 2), and (b) Dengue virus is in the early stage of evolution, and NS1 protein has not diverged significantly. Therefore, under the assumption that these residues have no functional importance, such positions would be likely to accumulate mutations, and can act as mutational hotspot regions in future. Our analysis also predicted a few observed-variable regions (high SE) as virtual-conserved residues (6 out of 52 high SE residues in DENV2 dimer). These two features are contradictory to each other. Possible reasons could be: (a) some of the mutations are functionally important, and therefore, selected at the expense of structure stability, (b) it can be considered as a limitation in FoldX prediction accuracy (Buß et al. 2018), and (c) presence of compensatory behavior in combination of mutations (Section 1.4).

3.6 Implementing features of NS1 protein evolution in the identification of critical residues for protein structure/function and druggable sites

Studies indicate the role of understanding protein evolution in predicting or identifying functional and druggable sites (Ghadermarzi et al. 2019). Therefore, we applied the pattern of mutation accumulation and associations learnt from this study to predict critical residues involved in NS1 protein structure and function.

A. Prediction of possible critical residues in dimer interface and motifs for protein-protein interaction

The emergence of experimentally proved critical residues in the set of virtual-conserved residue positions (Section 1.5) suggested that MC/RP scores can be used to predict unknown critical residues in NS1 protein. We augmented our approach with literature studies and searched for regions with sequence conservation. Sequence conservation suggests that a residue is important structurally or functionally or both. Therefore, using MC/RP scores as an additional filter over sequence conservation will be beneficial in narrowing down the search. Figure 7a shows the complete set of virtual-conserved residues in DENV2 NS1.

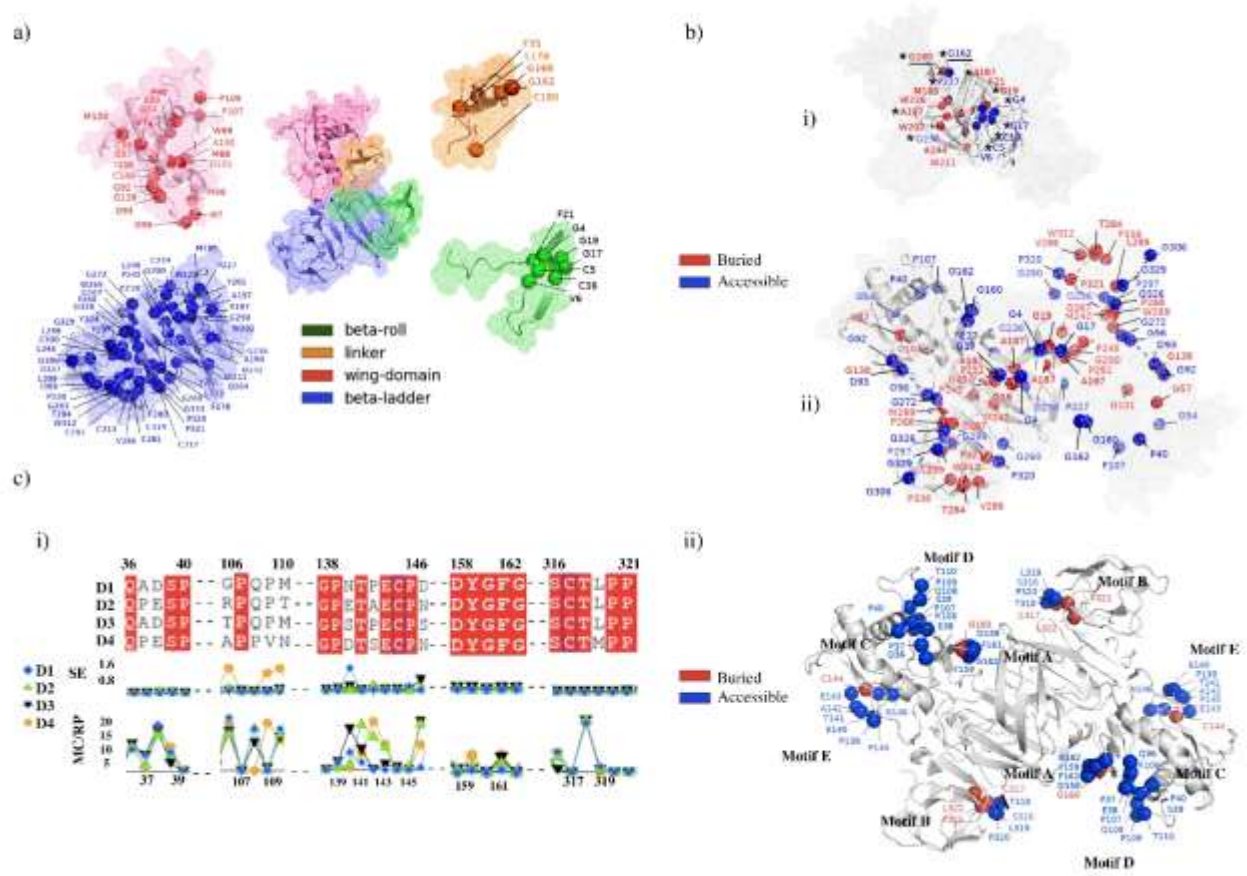

Figure 7: Critical residues in the dimer interface of DENV2 NS1 protein and interaction motifs.
a) Representation of all virtual-conserved position in different domains of DENV2 NS1 protein in dimer state.
b) Representation of virtual-conserved residues in DENV2 NS1 dimer interface (b-i). * represents conserved position across all serotypes and '_' represents already known important residue position in DENV2 NS1 protein. Set of sequentially conserved, observed and virtual-conserved residue positions (b-ii) among all dengue serotypes are shown in NS1 dimer backbone. Red represents buried residues, whereas blue represents accessible residues.
c) Representation of predicted protein-protein interaction sites. Conservation and importance of residues (c-i) within each site is shown by sequence alignment (across dengue serotypes), Shannon entropy (SE) and MC/RP scores (line plots). Low SE (SE <0.1) indicates observed-conserved residues. MC/RP scores equal to zero indicate virtual-conserved residues. Identified sites are marked on DENV2 dimer backbone where buried (red) and solvent-accessible (blue) residues are marked.
.

Apart from known structurally critical Cysteine residues (Pryor and Wright 1993; Wallis et al. 2004; Scaturro et al. 2015), 39 observed and virtual-conserved positions (G4, G17, G19, P40, G54, G57, G92, D93, G96, G101, P107, G138, G160, G162, A187, A197, P227, G236, M242, P245, G250, P251, G260, G267, P268, W269, G272, T284, V286, G296, P297, L299, G306, W312, P320, P321, G326, G329, P338; also please see Table 3) were identified in all four serotypes (Figure 7b-ii). NS1 dimer interface is critical for viral replication (Rastogi et al. 2016), therefore, interface residues were first identified, and virtual-conserved positions were filtered. Multiple such positions were found in DENV1 (n=16), DENV2 (n=18), DENV3 (n=21), and DENV4 (n=16) NS1 dimer interfaces (Supplementary Table 6). 11 residues (G4, C5, C16, G17, G19, G160, G162, A187, A197, P227, and G236) were conserved across all dengue serotypes (Supplementary Table 6) (Figure 7b-i). G160 and G162 (Figure 7b-i (Underlined residues)) are critical for viral replication (Table 2); other nine residue positions could have structural or functional importance. These residues can be considered as a starting point to identify protein-protein interaction motifs or druggable sites.

NS1 has a huge array of host interacting proteins (Karyala et al. 2016). A protein-protein interaction motif should be stable and contain exposed and reactive amino acid residues. Therefore, we searched for sites with observed and virtual-conserved sites, high RSA score (solvent-exposed), and reactive residues. Motif A (158-DYGFG-162) on dimer

interface satisfied the selection criteria. It contained reactive residues, virtual-conserved residues (D158, Y158, and F161), and three known functionally critical residues (G160, F161, and G162) [Table 2]. Interestingly, a number of Proline residues belonged to virtual-conserved set and were identified on the NS1 periphery (Figure 7b-ii). Past studies suggest Proline-rich motifs (Kay et al. 2000; Krieger et al. 2005; Pagano 2013) as interaction motifs for SH3-domain containing proteins and functionally important. SH3 domain containing Lyn (Li et al. 2020) and Src kinases (H Chu and Yang 2006; Pagano 2013) are involved in dengue and Zika virus maturation during infection. Therefore, NS1 protein might utilize these Proline-rich motifs to interact with SH3-domain-containing proteins. Motifs consisting of Proline, and other residues satisfying our selection criteria were searched for (Supplementary Table 7), and four such Proline-rich motifs were identified (Figure 7c-i). Motif B (316-SCTLPPL-322) present on beta-ladder domain, was both evolutionarily (observed low SE) and sequentially conserved (expect for L319) among each serotype. It contains some of the known critical residues (P320, P321, and C317) (Scaturro et al. 2015). On the wing domain, motif C (36-Q(P/X)XSP-40) is an evolutionarily conserved site but with different composition in dengue serotypes (X represent variable residue). Motif D (106-XPX(P/X)X-110) is sequentially variable, however, it can be a potential interacting motif due to the presence of Proline residues and close localization with dimer interface. Motif E (139-PXT(P/X)ECPX-146) is present near the junction of the wing and beta-ladder domain. It consists of two Proline and one Cysteine residue in all four serotypes. All the above sites are represented on DENV2 NS1 backbone in Figure 7c-ii.

Table 3: Predicted structurally important and conserved residue positions across all four dengue serotypes.

	Residue Name				**Shannon Entropy (SE)**				**Relative solvent accessibility (RSA)**			
AA Num	**D1**	**D2**	**D3**	**D4**	**D1**	**D2**	**D3**	**D4**	**D1**	**D2**	**D3**	**D4**
4	GLY	GLY	GLY	GLY	0.005	0.012	0	0	40.8	48.8	35.3	45.2
17	GLY	GLY	GLY	GLY	0	0.017	0.033	0.048	56.1	66	82	63.8
19	GLY	GLY	GLY	GLY	0	0.013	0.033	0.044	0.3	1.1	0.6	0.2
40	PRO	PRO	PRO	PRO	0.006	0.012	0.033	0.048	8.7	8.9	11.8	13.4
54	GLY	GLY	GLY	GLY	0.005	0.028	0	0.024	88.9	92.4	91.1	77.4
57	GLY	GLY	GLY	GLY	0.001	0.012	0.043	0.048	0	0	0	0
92	GLY	GLY	GLY	GLY	0.01	0.012	0.043	0.048	53.3	54.8	46.3	52.7
93	ASP	ASP	ASP	ASP	0.115	0.087	0.052	0.109	76.8	71.3	76.7	70.8
96	GLY	GLY	GLY	GLY	0.018	0.012	0.052	0.048	20	17.7	15.7	19.4
101	GLY	GLY	GLY	GLY	0	0.012	0.051	0.048	0	0	0	0
107	PRO	PRO	PRO	PRO	0	0.012	0.07	0.085	37.7	32.2	33.2	28.5
138	GLY	GLY	GLY	GLY	0	0	0.043	0.052	7	11.6	2.2	14.1
160	GLY	GLY	GLY	GLY	0.005	0.012	0.01	0.028	10.3	6.1	5.4	10.5

162	GLY	GLY	GLY	GLY	0	0.018	0.049	0.071	112.3	111	110.8	111.2
187	ALA	ALA	ALA	ALA	0.01	0.013	0.052	0.048	0.4	0.2	1.6	1.6
197	ALA	ALA	ALA	ALA	0	0.013	0.062	0.048	2.9	4.1	2.7	6.9
227	PRO	PRO	PRO	PRO	0.005	0.02	0.052	0.048	10.8	13.5	2.4	13
236	GLY	GLY	GLY	GLY	0	0.074	0.069	0.048	125.2	126.5	126	124.1
242	MET	MET	MET	MET	0	0.012	0.051	0.044	0.9	0.8	0.9	0.2
245	PRO	PRO	PRO	PRO	0	0.019	0.052	0.048	7	9.4	9.4	6.4
250	GLY	GLY	GLY	GLY	0	0.012	0.052	0.048	0	0	0	0
251	PRO	PRO	PRO	PRO	0	0.012	0.052	0.048	1.4	0.9	0.3	1.2
260	GLY	GLY	GLY	GLY	0.001	0.012	0.052	0.048	36.7	43.3	51.1	56.1
267	GLY	GLY	GLY	GLY	0	0	0.052	0.048	0	0	0	0
268	PRO	PRO	PRO	PRO	0	0.012	0.062	0.048	0	0	0	0.2
269	TRP	TRP	TRP	TRP	0	0.025	0.07	0.048	3.3	5.3	7.5	4.6
272	GLY	GLY	GLY	GLY	0.009	0.013	0.07	0.044	21.4	23.5	21.7	21.7
284	THR	THR	THR	THR	0	0.012	0.043	0.073	0	0	0	0
286	VAL	VAL	VAL	VAL	0.005	0.012	0.043	0.048	0	0	0	1.8
296	GLY	GLY	GLY	GLY	0	0.012	0.081	0.048	36.5	37.6	36.2	41.8
297	PRO	PRO	PRO	PRO	0.04	0.012	0.062	0.048	36	42.8	47.5	30.7
299	LEU	LEU	LEU	LEU	0.011	0	0.051	0.048	1.3	3.2	2.1	0
306	GLY	GLY	GLY	GLY	0.001	0	0.049	0.048	122.4	111.1	83.8	115.6
312	TRP	TRP	TRP	TRP	0.006	0.018	0.051	0.052	0	0	0	0
320	PRO	PRO	PRO	PRO	0.001	0.024	0.043	0.048	52.5	51.2	52.1	48.1
321	PRO	PRO	PRO	PRO	0.001	0.012	0.062	0.048	0	0	0	0.1
326	GLY	GLY	GLY	GLY	0.011	0.012	0.052	0.048	20	20	18.9	20
329	GLY	GLY	GLY	GLY	0.001	0.012	0.062	0.073	61.4	67.4	62.2	65.6
338	PRO	PRO	PRO	PRO	0	0.012	0.052	0.048	0.1	0.2	0	0

1 AA Num: Amino acid number, D1: Dengue serotype 1, D2: Dengue serotype 2, D3: Dengue serotype 3 and D4: Dengue serotype 4.

Residues having RSA score less than 7 are buried residues whereas positions with RSA score more than 7 are solvent-exposed. This table represents conserved positions (same sequence as well as less diversity [low SE]) across Dengue serotypes.

Computational validation of these motifs was carried out by protein-protein docking. SH3 domain kinases abl, cortactin, yes, Lyn and src were considered as ligand proteins. Profilin (Kay et al. 2000), an active actin-modulator which binds to Proline-rich motif, was also considered for docking (Table 4). DENV2 NS1 dimer was used as a receptor protein, and blind docking was performed using the HDOCK server (Yan et al. 2017). Blind docking was done to check the unbiased preference of ligand proteins for NS1 regions. This approach enables one to ask how frequently motifs of interest will be sampled without any prior knowledge during docking. High sampling suggests a given motif would likely be an interacting motif within the host cell. In order to achieve this, top 100 models predicted by HDOCK were considered for analysis. Models were classified into no-motif, specific-motif, and multiple motif binders based on the type of NS1 residues presented within 5Å of the SH3 domain. The majority of models were motif binders where the ligand protein was either binding specifically to a single motif or simultaneously to multiple motifs (Table 4).

Table 4: Protein-protein docking against some of the SH3 domain proteins and Profilin with DENV2 NS1 as a receptor protein

Protein Name	No Motif	Specific Motifs					Multiple Motifs	Predicted top model					
		A	B	C	D	E		Model No	Motif Description	HDOCK Score	PPCHECK (kcal/mol) TE	PPCHECK (kcal/mol) NE/R	MM-GBSA (kcal/mol)	Z-SCORE (DockScore)
Abl (5MO4)	17	21	12	1	1	8	40	3	Motif B & E	-254.11	-58.66	-0.55	-115.46	0.44
Cortactin (2DIX)	47	4	16	0	2	14	17	5	Motif E	-198.58	-138.33	-1.61	-63.15	2.19
Lyn (6NMW)	41	6	14	0	0	23	16	4	Motif E	-203.86	-127.60	-1.77	-92.98	1.46
Src (IFMK)	12	12	12	2	3	13	46	1	Motif B	-241.22	-45.59 & -65.87	-1.01 & -1.22	-124.34	0.40
Yes (2HDA)	33	3	15	0	0	21	18	96	Motif E	-166.21	-112.24	-1.68	-73.19	1.22
Profilin (2PBD)	23	6	14	2	2	25	28	9	Motif B	-226.07	-198.54	-2.33	-146.99	1.56

TE and NE/R represent calculated total and normalized energy per interface residues of protein-protein docked pose by PPCHECK, respectively.
'No motif' represents the total number of docked poses (out of 100) where no ligand residue was close within 5 Å distance of residues belonging to each motif. 'Specific Motifs' and 'Multiple Motifs' represent poses where the ligand's residues were within 5 Å distance of amino acid of a single motif and with multiple motifs at the same time, respectively. All distances were calculated w.r.t C-alpha atom of residues.
Src was found to interact with both chains ('A' and 'B') of NS1. Therefore, two energy values are shown in PPCHECK columns.

In order to select the best-predicted model/pose, several factors like ligand's interaction with motifs, HDOCK score, total stabilization energy of complex (kcal/mol) calculated by PPCHECK, MM-GBSA score (Genheden and Ryde 2015) and Z-score predicted by DOCKSCORE (Malhotra et al. 2015) were checked. Models were ranked based on interaction with domain residue > HDOCK score > Stabilization energy ~ MM-GBSA score > DOCKSCORE. Top poses (Supplementary Figure S5) and their respective scores are described in Table 4. Motif B was identified as interacting motif for Src kinase and Profilin (Table 4), whereas motif E was identified as an interacting motif for Cortactin, Lyn, and Yes kinase. Abl kinase was found to interact with both motifs B and E. Motif B contains some previously known critical residues; therefore, we propose motif E as a novel protein-protein interacting site for further experimental validation. Interestingly, motif A (without Proline residue) was not present in any top predicted poses, which is in accordance with literature. Selecting a top pose is always challenging and could be easily mistaken for the correct pose. Therefore, our

selected top poses should be considered with caution. The observations suggest that NS1 could have plasticity for the usage of interaction sites (Table 4). This property could easily allow NS1 protein to modulate functionality or sequester multiple host proteins simultaneously.

The introduction of MC/RP score as a parameter in druggable site selection criteria could be important in drug discovery. This feature suggests that in cases of unnatural evolutionary pressure (small inhibitors), a druggable site would ideally have less probability of mutating at virtual-conserved positions. Therefore, along with protein-protein interaction motif prediction and identification of possible mutational hotspot regions (Section 3.4), our sequence and structural analysis has direct application in identifying druggable sites.

B. Virtual screening against identified conserved druggable site dengue NS1 protein

The site 158-DYGFG-162 (Figure 8a) was identified as the target druggable site. This site was selected because of (a) evolutionary (Low SE and identical residues) conservation among all dengue serotypes (Table 5), (b) it contains three known functionally critical residues G160, F161, and G162 (Table 2), and two virtual-conserved residues (G160 and G162), (c) it is present on dimer interface and found as protein-protein interacting motif (Motif A) in many models during docking (Table 4), and (d) it is solvent-accessible as RSA scores for most residues were more than 8 (Table 5).

Table 5: SE, RSA and MC/RP score of the identified druggable site.

	Residue Name				Shannon Entropy (SE)				Relative Solvent Accessibility (RSA)				Mutational capacity per residue position (MC/RP)			
AA Num	**D1**	**D2**	**D3**	**D4**	**D1**	**D2**	**D3**	**D4**	**D1**	**D2**	**D3**	**D4**	**D1**	**D2**	**D3**	**D4**
158	ASP	ASP	ASP	ASP	0.01	0.03	0.06	0.05	50.8	42.3	40.5	42.7	0	1	1	0
159	TYR	TYR	TYR	TYR	0.01	0.00	0.06	0.05	26.5	22.9	21.0	27.7	1	2	1	6
160	GLY	GLY	GLY	GLY	0.01	0.01	0.01	0.03	10.3	6.1	5.4	10.5	0	0	0	0
161	PHE	PHE	PHE	PHE	0.00	0.01	0.05	0.09	45.5	46.1	48.0	44.0	1	3	5	6
162	GLY	GLY	GLY	GLY	0.00	0.02	0.05	0.07	112.3	111.0	110.8	111.2	0	0	0	0

D1: Dengue serotype 1, D2: Dengue serotype 2, D3: Dengue serotype 3 , D4:Dengue serotype 4,
MC/RP scores of NS1 dimer state of each dengue serotype were used for identifying the site.

An extensive virtual screening of small inhibitors (10,467 small-molecule inhibitors from Drugbank) against the identified druggable site was carried out. Molecules were first prepared for docking using the ligprep module. A receptor grid was generated over the site, and then docking was performed using a glide dock program (Halgren et al. 2004). Virtual screening was performed in three phases: a) High-throughput virtual screening (HTVS), b) Standard precision

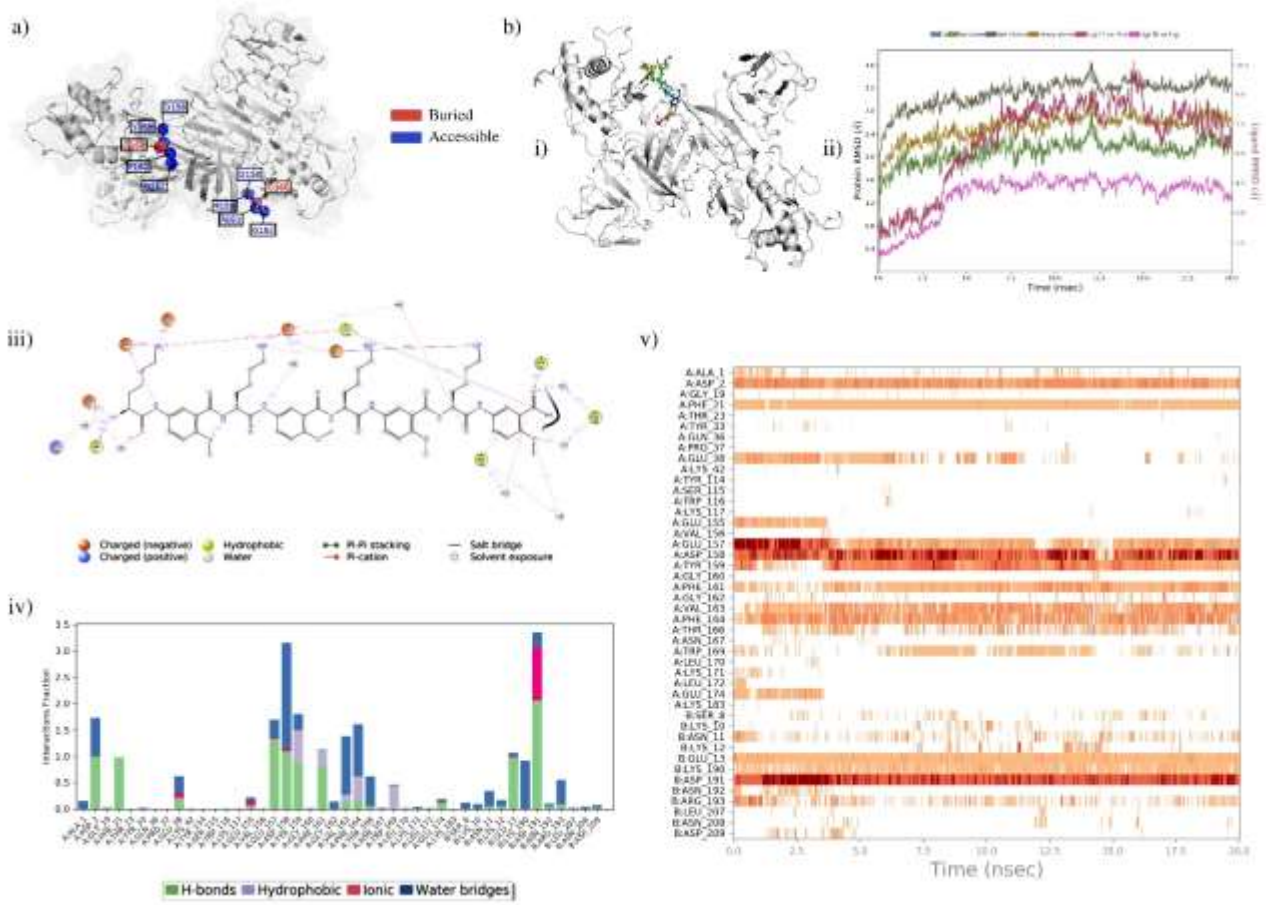

Figure 8: High-throughput screening of small molecule inhibitor against predicted druggable site.
a) Representation of identified druggable sites on the DENV2 dimer backbone. Red represents buried residues, and blue represents accessible residues. b) High-throughput screening of small molecule inhibitors and validation through set of systemic analyses. i) Represents protein-ligand complex of DENV2 NS1 and small molecule inhibitor **Delparantag** ii) Root mean square deviation (RMSD plot where the x-axis represents simulation time (ns) and the y-axis represents RMSD (A). iii) 2D-representation of the interaction of the ligand with NS1 residues throughout MD simulation iv) Stacked bar plot represents the fraction of time (y-axis) as well as nature of interaction (different stacks in each bar) between ligand and NS1 residues (x-axis) vi) Protein-ligand interaction represented as a timeline where the x-axis represents interaction made by ligand (orange) during simulation run and the y-axis represents interacting residues. White space represents no interaction between ligand and protein residues at that time point.

(SP), and c) Extra precision (XP), where molecules evaluated in one phase passed to the second phase (HTVS → SP → XP) for further evaluations. The top 10% of 9736 molecules screened in HTVS were passed to SP. The output of the SP step was further screened, and molecules with ΔG score less than -5 kcal/mol (522 molecules) were passed to the XP step. MM-GBSA calculations were performed on the top 20% of screened compounds (112) obtained from the XP step. To screen top hits, compounds showing interaction with at least three residues from the target site (Supplementary Figure S6a) were retained for further analysis. Filtered compounds were clustered based on their structural fingerprints and Tanimoto coefficient (Halgren et al. 2004) (Supplementary Figure S6a-b). Sixteen clusters with unique backbone and functional groups were obtained. Compounds were then ranked (from different clusters) based on their cluster size, MM-GBSA, and XP scores, and top 6 ranked compounds were selected as top hits (Table 6).

In order to check the ligand-protein complex stability of top hits, MD simulation run of 20ns was performed using Desmond (Bowers et al. 2006). Figure 8b shows the analysis to check ligand-protein complex stability of one of the top hit 'Delparantag' (an investigational drug for coronary artery disease) with DENV2 NS1. The interactions between protein and ligand (Figure 8b-i) were considered stable if (a) structural variations were less, (b) a high percentage of various interactions, such as hydrogen bonds, hydrophobic interactions, and salt-bridges, were formed during the run, and (c) protein-ligand complex is intact throughout the MD-trajectory. The RMSD plot assessed the stability of the complex

(Figure 8b-ii), and fluctuations in the protein regions were analyzed by RMSF (root mean square fluctuation) calculations. Residues in loop regions are expected to fluctuate more than residues in secondary structures such as helix and beta-sheet. A similar trend was observed as NS1 loop regions and regions interacting with Delparantag were fluctuating more than others (Supplementary Figure S6c). Protein-ligand contacts (Figure 8b-iii) were analyzed to confirm interaction of ligand with critical residues in the target site. In the case of Delparantag, stable interactions with protein residues D158, Y159, F161, and G162 were found throughout the simulation run (Figure 8b-iv). A similar analysis was performed on other top hits, and summarized in Table 6. Out of six, two drugs (Spectinomycin and Nadolol) are already FDA approved, three (Brilacidin, Delparantag, and Bradykinin) were in the investigational stage, whereas one (Lividomycin A) was in the experimental stage (Supplementary Figure S7-11). Brilacidin, Lividomycin A, and Bradykinin were observed to maintain stable interaction with the target site throughout the simulation.

Table 6: Top hits obtained after virtual screening against a conserved, mutationally impermeable, and functionally important site in the DENV2 NS1 protein.

Title	State	Function	XP score (Kcal/mol)	MMGBSA score (Kcal/mol)	Cluster - size	MD-simulation (20ns)	Interaction with target residues
Brilacidin	Investigational	Brilacidin is under investigation for the supportive care of Mucositis, Stomatitis, Mouth Diseases, and Head and Neck Neoplasms.	-8.08	-58.00	8-2	Stable	throughout the simulation
Delparantag	**Investigational**	**Delparantag has been used in trials studying the treatment of Angioplasty, Coronary Artery Disease (CAD), and Percutaneous Coronary Intervention.**	**-10.23**	**-54.09**	**11-3**	**Stable**	**throughout the simulation**
Lividomycin A	Experimental	Antibiotic	-11.80	-39.97	7-20	Stable	throughout the simulation
Bradykinin	Investigational	Bradykinin has been investigated for the basic science and treatment of Hypertension and type 2 diabetes.	-10.87	-36.66	10-17	Stable	throughout the simulation
Spectinomycin	Approved	Antibiotic	-7.43	-33.52	18-3	Stable	With one residue for the majority of the simulation
Nadolol	Approved	Nadolol is a nonselective beta adrenal receptor blocker used to lower blood pressure.	-7.60	-31.82	15-2	Stable	weak

Ligand 'Delparantag' (in bold) is explained in the text.

This study is focused on understanding the role of structural components of Dengue NS1 protein in its diversity. All the above results and correlations are in the context of Dengue NS1. Therefore, a similar analysis was next performed to test the effectiveness of our approach on NS1 proteins of other flaviviruses and influenza virus.

3.7 Extension of Dengue NS1 (sequence and structure) analysis to other NS1 proteins results in the identification of similar evolutionary trends

Complete NS1 protein sequences from other *flaviviruses* (Japanese encephalitis virus (JEV)=386, West Nile virus (WNV)=2672, Yellow fever virus (YFV) =438, and Zika=1060) and influenza virus (n=13095) were retrieved from VIPER database. The sequence of each of these proteins of known structure was considered as a reference sequence. Structures were refined in case of presence of missing regions (Method 2.2). SE scores per residue position were calculated and residue positions were categorized into observed low and high SE groups, respectively. The cut-off value for each protein was decided after analyzing the density distribution of their respective SE scores (Supplementary Figure S12). Further, single mutations present per residue position were identified, and their individual effect on protein stability was predicted (Method 2.4). Similar to the trend identified in Dengue virus, destabilizing mutations again dominated the spectrum, followed by neutral and stabilizing mutations (Supplementary Figure S13a). This similarity in trend suggested that NS1 proteins tends to remain conserved by occupying stable structural conformation. Such stable structural conformation might be vital for them to perform function and act as an antigenic protein.

Subsequently virtual-conserved and variable residues positions were predicted, and their association with low and high SE regions was determined. Except for Zika virus, odds ratios between observed-variable (high SE) and vritual-variable (Supplementary Table 2) residue positions were positive with significant p-value for each NS1 protein (Supplementary Figure S13b). This enrichment was again similar to observation found in DENV NS1 (Section 1.5). Further, the influenza virus NS1 had several variable positions with very high SE scores, and was thereby used to cross-validate the correlation between permeable and diverse regions (SE score>1). Out of 40 high SE (variable) positions only two residues were virtual-variable (Supplementary Table 8). These similar trends among Dengue NS1 and other NS1 proteins suggest that our analysis is applicable on other viral NS1 proteins, and can used to predict protein-protein interaction sites or screen/design drugs against druggable site**.** Interaction motifs will vary among different NS1 proteins. To check sequence conservation of our identified protein-protein interaction motif across *flaviviruses*, reference sequences of each NS1 were aligned, and motif conservation was analyzed. Interestingly, Proline residues within motifs were mostly conserved across flaviviruses (Supplementary Figure S13c). Motif B and Motif A which contain known critical residues are more conserved as compared to others.

Statistical Analysis:

In this study Odds ratios were calculated to find the association between two categorical features (Szumilas 2010). Odd ratios are calculated over 2X2 cross-tabulation that has frequency distribution of variables. Therefore, to check significance of the correlation (odd ratio) one-tailed Fisher's exact test was used. Fisher' exact test was used due to their extensive use in literature for checking significance of relationship between categorical variables. High and positive odds ratios with significant p-values ($p<0.05$) suggested a strong association between the two feature.

Discussion:

NS1 is a critical Dengue viral protein (Akey et al. 2014; Fan et al. 2014; Rastogi et al. 2016) and extensively studied as therapeutic target. Interestingly, NS1 is highly conserved as compared to other immunogenic protein such as envelope protein of Flaviviruses. The sequence analysis of NS1 protein distinguishes the residues into observed-conserved (low SE) and observed-variable (high SE) positions. The majority of the observed mutations were destabilizing with more enrichment in observed low SE regions (Section 3.3). Virtual saturation mutagenesis was carried out to predict virtual-conserved (low MC/RP score) and virtual-variable positions (MC/RP score more than 0). The significant enrichment of virtual-conserved positions in observed-conserved (low SE) regions (Section 3.5) showed their role in defining NS1 diversity. The proportion of these positions increased with the multimeric state of NS1 (monomer to dimer to hexamer), in accordance with our hypothesis. This suggests the restriction of mutational space owing to higher order quaternary states which renders more surface area of the protein as solvent-buried.

Due to the high mutation rate, multiple mutations can be observed in the viral proteins. This led to the study of combinations of mutations present in high frequency (Section 3.4) in patient samples. The differential effect of mutation combinations vs. individual mutations (Section 3.4) point to the possible mechanism for the emergence of destabilizing mutation in the viral population. The role of compensatory mutations was observed as a factor in the emergence of destabilizing mutations. Analysis of the combination of mutations showed avoidance of mutation toward dimer and hexametric interfaces. Apart from understanding the trends of NS1 evolution, the application of our sequence and structure analysis is demonstrated by predicting possible protein-protein interaction (Table 4) and druggable sites (Table 5). Along with interacting motifs, a conserved, structurally and functionally important site near the dimer interface was selected followed by identification of six potential small inhibitors. A similar trend in other NS1 proteins suggest the importance of maintenance of structural integrity by NS1 proteins in acting as antigenic proteins (Section 3.7).

We have considered some reference sequences, and the whole analysis was carried over them. These reference sequences are the building blocks of our analysis. Therefore, the repertoire of virtual-conserved/variable sites is subject to change with the reference sequence used. Unlike our virtual saturation mutagenesis, studying each combination is highly computationally extensive and challenging. This is another limitation of the study. Overcoming the limitations, our analysis conclusively shows the role of virtual-conserved residues in Dengue NS1 evolution. Such residues could help to identify possible protein-protein interaction motifs. These regions can further be confirmed experimentally. The study also uncovers the nature of virtual-variable sites that can accommodate mutations during future viral evolution.

IMPACT OF RESEARCH IN THE ADVANCEMENT OF KNOWLEDGE OR BENEFIT OF MANKIND:

4.1 NS1 protein conservation

The sequence and structure analysis of the NS1 protein detailed above enabled us to outline possible structural factors involved in its evolution. One primary reason for NS1 protein conservation could be its resistance to change the structure, since most observed single mutations were destabilizing in nature (Section 3.3). This restriction would be enhanced by the formation of new virtual-conserved positions (a position which ideally can accumulate mutation on linear sequence) across multimeric states (dimer and hexamer) (Section 3.5). Further analysis showed the role of a combination of mutations in NS1 diversity (Section 3.4), allowing destabilizing mutations (which constitute a major proportion) to emerge along with neutral or stabilizing mutations. The requirement of compensatory mutations could be a rate-limiting step in protein divergence. In such cases, a protein requires a counterbalancing mutation to neutralize the effect of the destabilizing mutation. Also, mutations in high-frequency combinations were sparsely present in dimeric (Figure 4c-i-ii) or hexameric interfaces (Supplementary Figure S4d-i-ii). Out of 16 unique positions in the high combination group, only two mutations ('T165S', 'V6I') were found in the dimer interface, while none of the positions were present in the hexamer interface. Interface residues for dimer and hexameric states were confirmed by PPCHECK (Sukhwal and Sowdhamini 2015).

To our best knowledge this study is first to perform such a detail analysis and showed a proper mechanism of conservation of such an important protein.

4.2 Development of structure and sequence analysis pipeline for predicting druggable sites

Identification of druggable site is a primary step toward designing inhibitors. Druggable site should consist important residues with structure/functional/both roles and should be accessible in nature. Over this, due to fast emergence of resistance among virus make finding new sites highly challenging in nature. Toward this, our study has suggested a possible pipeline where one can identify druggable sites (section 3.6) which would have high evolutionary pressure against mutation.

4.3. A Pan-Flaviviruses trend

This study showed a similar evolutionary trend of NS1 protein across all Flaviviruses (Section 3.7). This can be further used in development of pan-flavivirus based therapeutic agents or antibodies.

REFERENCES:

Akey DL, Brown WC, Dutta S, Konwerski J, Jose J, Jurkiw TJ, DelProposto J, Ogata CM, Skiniotis G, Kuhn RJ, et al. 2014. Flavivirus NS1 structures reveal surfaces for associations with membranes and the immune system. *Science* [Internet] 343:881–885. Available from: https://pubmed.ncbi.nlm.nih.gov/24505133/

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403–410.

Añez G, Heisey DA, Volkova E, Rios M. 2016. Complete Genome Sequences of Dengue Virus Type 1 to 4 Strains Used for the Development of CBER/FDA RNA Reference Reagents and WHO International Standard Candidates for Nucleic Acid Testing. *Genome Announc.* [Internet] 4. Available from: /pmc/articles/PMC4751306/

Avirutnan P, Punyadee N, Noisakran S, Komoltri C, Thiemmeca S, Auethavornanan K, Jairungsri A, Kanlaya R, Tangthawornchaikul N, Puttikhunt C, et al. 2006. Vascular leakage in severe dengue virus infections: A potential role for the nonstructural viral protein NS1 and complement. *J. Infect. Dis.* 193:1078–1088.

Beltramello M, Williams KL, Simmons CP, MacAgno A, Simonelli L, Quyen NTH, Sukupolvi-Petty S, Navarro-Sanchez E, Young PR, De Silva AM, et al. 2010. The Human Immune Response to Dengue Virus Is Dominated by Highly Cross-Reactive Antibodies Endowed with Neutralizing and Enhancing Activity. *Cell Host Microbe* [Internet] 8:271–283. Available from: /pmc/articles/PMC3884547/

Berendsen HJC, van der Spoel D, van Drunen R. 1995. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* 91:43–56.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. 2000. The Protein Data Bank. *Nucleic Acids Res.* 28:235–242.

Bhasin M, Varadarajan R. 2021. Prediction of Function Determining and Buried Residues Through Analysis of Saturation Mutagenesis Datasets. *Front. Mol. Biosci.* 8:1–11.

Bigman LS, Levy Y. 2018. Stability Effects of Protein Mutations: The Role of Long-Range Contacts. *J. Phys. Chem. B* 122:11450–11459.

Boonnak K, Slike BM, Burgess TH, Mason RM, Wu S-J, Sun P, Porter K, Rudiman IF, Yuwono D, Puthavathana P, et al. 2008. Role of Dendritic Cells in Antibody-Dependent Enhancement of Dengue Virus Infection. *J. Virol.* [Internet] 82:3939. Available from: /pmc/articles/PMC2292981/

Bowers KJ, Chow E, Xu H, Dror RO, Eastwood MP, Gregersen BA, Klepeis JL, Kolossvary I, Moraes MA, Sacerdoti FD, et al. 2006. Scalable algorithms for molecular dynamics simulations on commodity clusters. *Proc. 2006 ACM/IEEE Conf. Supercomput. SC'06*.

Brister JR, Ako-Adjei D, Bao Y, Blinkova O. 2015. NCBI viral genomes resource. *Nucleic Acids Res.* [Internet] 43:D571–D577. Available from: https://pubmed.ncbi.nlm.nih.gov/25428358/

Buß O, Rudat J, Ochsenreither K. 2018. FoldX as Protein Engineering Tool: Better Than Random Based Approaches? *Comput. Struct. Biotechnol. J.* [Internet] 16:25–33. Available from: https://doi.org/10.1016/j.csbj.2018.01.002

Chauhan PK, Sowdhamini R. 2022. LIM domain-wide comprehensive virtual mutagenesis provides structural rationale for cardiomyopathy mutations in CSRP3. *Sci. Reports 2022 121* [Internet] 12:1–11. Available from: https://www.nature.com/articles/s41598-022-07553-1

Cheng H-J, Lin C-F, Lei H-Y, Liu H-S, Yeh T-M, Luo Y-H, Linà Y-S. 2009. Proteomic Analysis of Endothelial Cell Autoantigens Recognized by Anti-Dengue Virus Nonstructural Protein 1 Antibodies. *Exp Biol Med* [Internet] 234:63–73. Available from: http://www.matrixscience.com/

Davis BH, Poon AFY, Whitlock MC. 2009. Compensatory mutations are repeatable and clustered within proteins. *Proc. R. Soc. B Biol. Sci.* 276:1823–1827.

Drake JW. 1993. Rates of spontaneous mutation among RNA viruses. *Proc. Natl. Acad. Sci. U. S. A.* 90:4171–4175.

Drake JW, Holland JJ. 1999. Mutation rates among RNA viruses. *Proc. Natl. Acad. Sci. U. S. A.* [Internet] 96:13910–13913. Available from: www.pnas.org

Duffy S. 2018. Why are RNA virus mutation rates so damn high? *PLoS Biol.* [Internet] 16. Available from: /pmc/articles/PMC6107253/

Edeling MA, Diamond MS, Fremont DH. 2014. Structural basis of flavivirus NS1 assembly and antibody recognition. *Proc. Natl. Acad. Sci. U. S. A.* 111:4285–4290.

Eswar N, Webb B, Marti- Renom MA, Madhusudhan MS, Eramian D, Shen M, Pieper U, Sali A. 2006. Comparative protein structure modeling using Modeller. *Curr. Protoc. Bioinforma.* [Internet] Chapter 5. Available from: https://pubmed.ncbi.nlm.nih.gov/18428767/

Fan J, Liu Y, Yuan Z. 2014. Critical role of Dengue Virus NS1 protein in viral replication. *Virol. Sin.* 29:162–169.

Friedrich TC, Frye CA, Yant LJ, O'Connor DH, Kriewaldt NA, Benson M, Vojnov L, Dodds EJ, Cullen C, Rudersdorf R, et al. 2004. Extraepitopic Compensatory Substitutions Partially Restore Fitness to Simian Immunodeficiency Virus Variants That Escape from an Immunodominant Cytotoxic-T-Lymphocyte Response. *J. Virol.* [Internet] 78:2581. Available from: /pmc/articles/PMC369222/

Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA, Sanschagrin PC, Mainz DT. 2006. Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.* 49:6177–6196.

Genheden S, Ryde U. 2015. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.* [Internet]

10:449. Available from: /pmc/articles/PMC4487606/

Ghadermarzi S, Li X, Li M, Kurgan L. 2019. Sequence-Derived Markers of Drug Targets and Potentially Druggable Human Proteins. *Front. Genet.* 10:1–18

Gutsche I, Coulibaly F, Voss JE, Salmon J, D'Alayer J, Ermonval M, Larquet E, Charneau P, Krey T, Mégret F, et al. 2011. Secreted dengue virus nonstructural protein NS1 is an atypical barrel-shaped high-density lipoprotein. *Proc. Natl. Acad. Sci. U. S. A.* [Internet] 108:8003–8008. Available from: https://pubmed.ncbi.nlm.nih.gov/21518917/

H Chu JJ, Yang PL. 2006. c-Src protein kinase inhibitors block assembly and maturation of dengue virus.

Hafirassou ML, Meertens L, Umaña-Diaz C, Labeau A, Dejarnac O, Bonnet-Madin L, Kümmerer BM, Delaugerre C, Roingeard P, Vidalain PO, et al. 2017. A Global Interactome Map of the Dengue Virus NS1 Identifies Virus Restriction and Dependency Host Factors. *Cell Rep.* [Internet] 21:3900–3913. Available from: https://pubmed.ncbi.nlm.nih.gov/29281836/

Halgren T. 2007. New method for fast and accurate binding-site identification and analysis. *Chem. Biol. Drug Des.* 69:146–148.

Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, Banks JL. 2004. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* [Internet] 47:1750–1759. Available from: https://pubmed.ncbi.nlm.nih.gov/15027866/

Harder E, Damm W, Maple J, Wu C, Reboul M, Xiang JY, Wang L, Lupyan D, Dahlgren MK, Knight JL, et al. 2016. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* [Internet] 12:281–296. Available from: https://pubmed.ncbi.nlm.nih.gov/26584231/

Jayathilaka D, Gomes L, Jeewandara C, Jayarathna GSB, Herath D, Perera PA, Fernando S, Wijewickrama A, Hardman CS, Ogg GS, et al. 2018. Role of NS1 antibodies in the pathogenesis of acute secondary dengue infection. *Nat. Commun.* [Internet] 9. Available from: http://dx.doi.org/10.1038/s41467-018-07667-z

Jorgensen WL, Maxwell DS, Tirado-Rives J. 1996. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* [Internet] 118:11225–11236. Available from: https://pubs.acs.org/doi/abs/10.1021/ja9621760

Karyala P, Metri R, Bathula C, Yelamanchi SK, Sahoo L, Arjunan S, Sastri NP, Chandra N. 2016. DenHunt - A Comprehensive Database of the Intricate Network of Dengue-Human Interactions. *PLoS Negl. Trop. Dis.* [Internet] 10:e0004965. Available from: https://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0004965

Kay BK, Williamson MP, Sudol M. 2000. The importance of being proline: the interaction of proline- rich motifs in signaling proteins with their cognate domains. *FASEB J.* 14:231–241.

Khadka S, Vangeloff AD, Zhang C, Siddavatam P, Heaton NS, Wang L, Sengupta R, Sahasrabudhe S, Randall G, Gribskov M, et al. 2011. A physical interaction network of dengue virus and human proteins. *Mol. Cell. Proteomics* 10.

Knies JL, Dang KK, Vision TJ, Hoffman NG, Swanstrom R, Burch CL. 2008. Compensatory Evolution in RNA Secondary Structures Increases Substitution Rate Variation among Sites. *Mol. Biol. Evol.* [Internet] 25:1778. Available from: /pmc/articles/PMC2734131/

Krieger F, Möglich A, Kiefhaber T. 2005. Effect of proline and glycine residues on dynamics and barriers of loop formation in polypeptide chains. *J. Am. Chem. Soc.* 127:3346–3352.

Li MY, Naik TS, Siu LYL, Acuto O, Spooner E, Wang P, Yang X, Lin Y, Bruzzone R, Ashour J, et al. 2020. Lyn kinase regulates egress of flaviviruses in autophagosome-derived organelles. *Nat. Commun.* [Internet] 11. Available from: http://dx.doi.org/10.1038/s41467-020-19028-w

Lin CF, Lei HY, Shiau AL, Liu CC, Liu HS, Yeh TM, Chen SH, Lin YS. 2003. Antibodies from dengue patient sera cross-react with endothelial cells and induce damage. *J. Med. Virol.* 69:82–90.

Litwin S, Jores R. 1992. Shannon Information as a Measure of Amino Acid Diversity. *Theor. Exp. Insights into Immunol.* [Internet]:279–287. Available from: https://link.springer.com/chapter/10.1007/978-3-642-76977-1_17

Madhavi Sastry G, Adzhigirey M, Day T, Annabhimoju R, Sherman W. 2013. Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments. *J. Comput. Aided. Mol. Des.* 27:221–234.

Malavige GN, Fernando S, Fernando DJ, Seneviratne SL. 2004. Dengue viral infections. *Postgrad. Med. J.* 80:588–601.

Malavige GN, Ogg GS. 2017. Pathogenesis of vascular leak in dengue virus infection. *Immunology* 151:261–269.

Malhotra S, Mathew OK, Sowdhamini R. 2015. DOCKSCORE: A webserver for ranking protein-protein docked poses. *BMC Bioinformatics* [Internet] 16:1–6. Available from: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-015-0572-6

Mizuguchi K, Deane CM, Blundell TL, Johnson MS, Overington JP. 1998. JOY: protein sequence-structure representation and analysis. *Bioinformatics* [Internet] 14:617–623. Available from: https://pubmed.ncbi.nlm.nih.gov/9730927/

Modhiran N, Watterson D, Muller DA, Panetta AK, Sester DP, Liu L, Hume DA, Stacey KJ, Young PR. 2015. Dengue virus NS1 protein activates cells via Toll-like receptor 4 and disrupts endothelial cell monolayer integrity. *Sci. Transl. Med.* 7.

Neufeldt CJ, Cortese M, Acosta EG, Bartenschlager R. 2018. Rewiring cellular networks by members of the Flaviviridae family. *Nat. Rev. Microbiol.* [Internet] 16:125–142. Available from: https://pubmed.ncbi.nlm.nih.gov/29430005/

Pagano MA. 2013. Viral proteins and Src family kinases: Mechanisms of pathogenicity from a liaison dangereuse. *World J. Virol.* 2:71.

Paranavitane SA, Gomes L, Kamaladasa A, Adikari TN, Wickramasinghe N, Jeewandara C, Shyamali NLA, Ogg GS, Malavige GN. 2014. Dengue NS1 antigen as a marker of severe clinical disease. *BMC Infect. Dis.* 14.

Pickett BE, Sadat EL, Zhang Y, Noronha JM, Squires RB, Hunt V, Liu M, Kumar S, Zaremba S, Gu Z, et al. 2012. ViPR: an open bioinformatics database and analysis resource for virology research. *Nucleic Acids Res.* [Internet] 40:D593. Available from: /pmc/articles/PMC3245011/

Płaszczyca A, Scaturroid P, Neufeldtid CJ, Corteseid M, Cerikan B, Ferla S, Brancale A, Pichlmair A, Bartenschlager R. 2019. A novel interaction between dengue virus nonstructural protein 1 and the NS4A-2K-4B precursor is required for viral RNA replication but not for formation of the membranous replication organelle. *PLOS Pathog.* [Internet] 15:e1007736. Available from: https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1007736

Pryor MJ, Wright PJ. 1993. The effects of site-directed mutagenesis on the dimerization and secretion of the NS1 protein specified by dengue virus. *Virology* [Internet] 194:769–780. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8389081

Pryor MJ, Wright PJ. 1994. Glycosylation mutants of dengue virus NS1 protein. *J. Gen. Virol.* [Internet] 75 (Pt 5):1183–1187. Available from: https://pubmed.ncbi.nlm.nih.gov/8176380/

Rastogi M, Sharma N, Singh SK. 2016. Flavivirus NS1: A multifaceted enigmatic viral protein. *Virol. J.* 13.

Sanjuán R, Domingo-Calap P. 2021. Genetic Diversity and Evolution of Viral Populations. *Encycl. Virol.* [Internet]:53. Available from: /pmc/articles/PMC7157443/

Scaturro P, Cortese M, Chatel-Chaix L, Fischl W, Bartenschlager R. 2015. Dengue Virus Non-structural Protein 1 Modulates Infectious Particle Production via Interaction with the Structural Proteins. *PLoS Pathog.* 11:1–32.

Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. 2005. The FoldX web server: an online force field. *Nucleic Acids Res.* [Internet] 33:W382. Available from: /pmc/articles/PMC1160148/

Screaton G, Mongkolsapaya J, Yacoub S, Roberts C. 2015. New insights into the immunopathology and control of dengue virus infection. *Nat. Rev. Immunol.* [Internet] 15:745–759. Available from: https://pubmed.ncbi.nlm.nih.gov/26603900/

Sharma A, Tiwari V, Sowdhamini R. 2020. Computational search for potential COVID-19 drugs from FDA-approved drugs and small molecules of natural origin identifies several anti-virals and plant products. *J. Biosci.* [Internet] 45:1–18. Available from: https://link.springer.com/article/10.1007/s12038-020-00069-8

Silva EM, Conde JN, Allonso D, Nogueira ML, Mohana-Borges R. 2013. Mapping the Interactions of Dengue Virus NS1 Protein with Human Liver Proteins Using a Yeast Two-Hybrid System: Identification of C1q as an Interacting Partner. *PLoS One* 8.

Somnuke P, Hauhart RE, Atkinson JP, Diamond MS, Avirutnan P. 2011. N-linked glycosylation of dengue virus NS1 protein modulates secretion, cell-surface expression, hexamer stability, and interactions with human complement. *Virology* 413:253–264.

Sukhwal A, Sowdhamini R. 2015. PPCheck: A Webserver for the Quantitative Analysis of Protein-Protein Interfaces and Prediction of Residue Hotspots. *Bioinform. Biol. Insights* [Internet] 9:141–151. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26448684

Szumilas M. 2010. Explaining Odds Ratios. *J. Can. Acad. Child Adolesc. Psychiatry* [Internet] 19:227. Available from: /pmc/articles/PMC2938757/

Unal H, Jagannathan R, Bhatnagar A, Tirupula K, Desnoyer R, Karnik SS. 2013. Long range effect of mutations on specific conformational changes in the extracellular loop 2 of angiotensin II type 1 receptor. *J. Biol. Chem.* 288:540–551.

Wallis TP, Huang CY, Nimkar SB, Young PR, Gorman JJ. 2004. Determination of the disulfide bond arrangement of dengue virus NS1 protein. *J. Biol. Chem.* [Internet] 279:20729–20741. Available from: https://pubmed.ncbi.nlm.nih.gov/14981082/

Wang L, Wu Y, Deng Y, Kim B, Pierce L, Krilov G, Lupyan D, Robinson S, Dahlgren MK, Greenwood J, et al. 2015. Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *J. Am. Chem. Soc.* 137:2695–2703.

Watterson D, Modhiran N, Young PR. 2016. The many faces of the flavivirus NS1 protein offer a multitude of options for inhibitor design. *Antiviral Res.* [Internet] 130:7–18. Available from: https://pubmed.ncbi.nlm.nih.gov/26944216/

Weaver SC, Vasilakis N. 2009. Molecular Evolution of Dengue Viruses: Contributions of Phylogenetics to Understanding the History and Epidemiology of the Preeminent Arboviral Disease. *Infect. Genet. Evol.* [Internet] 9:523. Available from: /pmc/articles/PMC3609037/

Wiederstein M, Sippl MJ. 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* [Internet] 35:W407–W410. Available from: https://academic.oup.com/nar/article/35/suppl_2/W407/2920938

Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Sajed T, Johnson D, Li C, Sayeeda Z, et al. 2018. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* [Internet] 46:D1074–D1082. Available from: https://pubmed.ncbi.nlm.nih.gov/29126136/

Xu X, Song H, Qi J, Liu Y, Wang H, Su C, Shi Y, Gao GF. 2016. Contribution of intertwined loop to membrane association revealed by Zika virus full-length NS1 structure. *EMBO J.* [Internet] 35:2170–2178. Available from: https://pubmed.ncbi.nlm.nih.gov/27578809/

Yan Y, Zhang D, Zhou P, Li B, Huang SY. 2017. HDOCK: a web server for protein–protein and protein–DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Res.* [Internet] 45:W365. Available from: /pmc/articles/PMC5793843/

Supplementary Tables

Supplementary Table 1: Distribution of number of residue positions with mutational capacity per residue position (MC/RP) score equal to 0 or more than 0 in low vs high SE region of NS1 dimer protein across all Dengue serotypes. Virtual-conserved positions tend to be less variable and vice-versa.

Dimer	Serotype	Category	Virtual-conserved (MC/RP =0)	Virtual-variable (MC/RP >0)
	DENV1	**Low SE (SE <=0.1)**	77	238
		High SE (SE >0.1)	3	31
	DENV2	**Low SE (SE <=0.1)**	82	216
		High SE (SE >0.1)	6	46
	DENV3	**Low SE (SE <=0.1)**	77	238
		High SE (SE >0.1)	2	33
	DENV4	**Low SE (SE <=0.1)**	73	216
		High SE (SE >0.1)	1	60

DENV1: Dengue serotype 1, DENV2: Dengue serotype 2, DENV3: Dengue serotype 3 and DENV4: Dengue serotype 4
Low and High SE residues represent observed-conserved and –variable positions respectively.

Supplementary Table 2: Calculated odd ratio for multiple features across various viral NS1 protein. Odd ratio more than 1 represents a positive association between two conditions present in testing features

SEROTYPE	FEATURES	Odd Ratios	CI (95%)	Z-score	P-value	Star
	Stabilization mutations in High SE positions	1.62	0.94-2.78	1.71	0.07	-
	(Stabilization + Neutral) in High SE positions	1.69	1.07-2.6	2.254	0.02	*
	De-stabilization mutations in Low SE positions	1.69	1.07-2.6	2.254	0.02	*
	MC/RP > 0 in high SE positions (Dimer)	3.45	1.02-11.584	2.005	0.045	*
DENV1						
	Stabilization mutations in High SE positions	1.615	1.09-2.48	2.409	0.01	*
	(Stabilization + Neutral) in High SE positions	1.603	1.153-2.22	2.8	0.005	**
	De-stabilization mutations in Low SE positions	1.603	1.153-2.22	2.8	0.005	**
	MC/RP > 0 in high SE positions (Dimer)	2.91	1.9-7.07	2.358	0.018	*
DENV2						
	Stabilization mutations in High SE positions	3.3	2.15-5.08	5.483	<0.0001	***
	(Stabilization + Neutral) in High SE positions	3.22	2.23-4.64	6.289	<0.0001	***
	De-stabilization mutations in Low SE positions	3.22	2.23-4.64	6.289	<0.0001	***
	MC/RP > 0 in high SE positions (Dimer)	5.33	1.25-22.76	2.264	0.02	*
DENV3						
	Stabilization mutations in High SE positions	2.3	1.59-3.54	4.2	<0.0001	***
	(Stabilization + Neutral) in High SE positions	2.2	1.62 - 3.1	4.886	<0.0001	***
	De-stabilization mutations in Low SE positions	2.2	1.62 - 3.1	4.886	<0.0001	***
	MC/RP > 0 in high SE positions (Dimer)	20.2	2.7-148.9	2.958	0.003	**
DENV4						
	MC/RP > 0 in high SE positions (Monomer)	2.325	1.016-5.31	2.0	0.045	*
H1N1						
	MC/RP > 0 in high SE positions (Dimer)	3.05	1.05-8.79	2.064	0.039	*
YFV						
	MC/RP > 0 in high SE positions (Dimer)	2.689	1.3-5.35	2.811	0.004	**
WNV						
	MC/RP > 0 in high SE positions (Dimer)	1.986	0.95-4.12	1.84	0.06	
ZIKA						
JEV	MC/RP > 0 in high SE positions (Dimer)	2.04	1.08-3.83	2.22	0.02	*

MC/RP > 0: represents virtual-variable residue positions.
P-value calculated according to one-tail Fisher's exact test (* - $0.01 < P < 0.05$, ** − $0.001 < P < 0.01$, *** − $P < 0.0001$ and no star for non-significant P-value). DENV1 : Dengue serotype 1, DENV2 : Dengue serotype 2, DENV3 : Dengue serotype 3 and DENV4 : Dengue serotype 4, H1N1 : Influenza, YFV : Yellow fever virus, WNV : West Nile virus, JEV : Japanese Encephalitis virus.

Supplementary Table 3: Shannon entropy (SE) and mutational capacity per residue position (MC/RP) scores of known critical residues of DENV2 NS1 protein.

	Residue Name				Shannon Entropy (SE)				MC/RP			
AA Num	**D1**	**D2**	**D3**	**D4**	**D1**	**D2**	**D3**	**D4**	**D1**	**D2**	**D3**	**D4**
3	SER	SER	MET	THR	0.03	0.13	0.28	0.20	16	16	1	11
5	CYS	CYS	CYS	CYS	0.01	0.01	0.03	0.04	0	0	0	0
9	TRP	TRP	TRP	TRP	0.00	0.03	0.03	0.05	10	5	4	11
11	GLY	ASN	GLY	GLY	0.01	0.22	0.04	0.07	15	18	7	17
12	ARG	LYS	LYS	LYS	0.01	0.06	0.05	0.96	9	6	8	8
33	TYR	TYR	TYR	TYR	0.00	0.01	0.03	0.07	4	2	0	1
56	CYS	CYS	CYS	CYS	0.00	0.01	0.04	0.05	0	0	14	0
56	CYS	CYS	CYS	CYS	0.00	0.01	0.04	0.05	0	0	14	0
115	SER	SER	SER	SER	0.00	0.03	0.04	0.05	11	19	15	19
116	TRP	TRP	TRP	TRP	0.00	0.01	0.01	0.02	7	9	11	4
131	ASN	ASN	ASN	ASN	0.00	0.02	0.04	0.05	19	9	16	17
137	ASP	ASP	ASP	ASP	0.01	0.02	0.04	0.05	0	1	0	0
151	TRP	TRP	TRP	TRP	0.01	0.01	0.05	0.05	5	4	7	8
154	TRP	LEU	TRP	LEU	0.00	0.02	0.05	0.99	1	2	3	2
159	TYR	TYR	TYR	TYR	0.01	0.00	0.06	0.05	1	2	1	6
160	GLY	GLY	GLY	GLY	0.01	0.01	0.01	0.03	0	0	0	0
161	PHE	PHE	PHE	PHE	0.00	0.01	0.05	0.09	1	3	5	6
162	GLY	GLY	GLY	GLY	0.00	0.02	0.05	0.07	0	0	0	0
169	TRP	TRP	TRP	TRP	0.01	0.01	0.05	0.05	7	6	9	7
180	CYS	CYS	CYS	CYS	0.01	0.01	0.05	0.07	0	0	0	0
181	ASP	ASP	ASP	ASP	0.01	0.03	0.05	0.08	4	13	9	8
190	LYS	LYS	LYS	LYS	0.01	0.04	0.05	0.04	4	6	6	10
197	ALA	ALA	ALA	ALA	0.00	0.01	0.06	0.05	0	0	0	0
208	ASN	ASN	ASN	ASN	0.01	0.03	0.07	0.07	19	17	15	16
253	SER	SER	SER	SER	0.00	0.01	0.05	0.05	1	4	2	3
285	THR	THR	THR	THR	0.01	0.01	0.05	0.07	11	14	11	14
292	CYS	CYS	CYS	CYS	0.00	0.01	0.07	0.05	0	0	0	0
292	CYS	CYS	CYS	CYS	0.00	0.01	0.07	0.05	0	0	0	0
302	THR	THR	THR	THR	0.01	0.02	0.04	0.07	16	14	17	17
312	TRP	TRP	TRP	TRP	0.01	0.02	0.05	0.05	0	0	0	0
313	CYS	CYS	CYS	CYS	0.00	0.01	0.04	0.05	0	0	0	0
313	CYS	CYS	CYS	CYS	0.00	0.01	0.04	0.05	0	0	0	0
318	THR	THR	THR	THR	0.00	0.01	0.05	0.05	17	18	18	17
320	PRO	PRO	PRO	PRO	0.00	0.02	0.04	0.05	0	0	0	0
321	PRO	PRO	PRO	PRO	0.00	0.01	0.06	0.05	0	0	0	0

D1: Dengue serotype 1, D2: Dengue serotype 2, D3: Dengue serotype 3 and D4: Dengue serotype

Supplementary Table 4: Top six combination of mutations in DENV2 NS1 protein with their number of occurrence, and ΔΔG score calculated by FoldX across monomer, dimer and hexamer states. Characterization of effect of mutations based on ΔΔG score is represented as Feature.

		MONOMER	**DIMER**	**HEXAMER**
Mutations	**Strain count**	**ΔΔG value / Feature**	**ΔΔG value / Feature**	**ΔΔG value / Feature**
D282E,I213M,I265T,K173R,L248F,N223S,T266A	167	4.15 / highly de-stabilizing	-2.64 / highly_stabilizing	-7.04 / highly_stabilizing
H130Y,H51Q,S81A,T266A,V178A,V287I	165	-0.80 / slightly stabilising	-5.27 / highly_stabilizing	-13.8 / highly_stabilizing
D282E,F280L,I213M,I265T K173R,K95R,L248F,N223S,,T266A	63	4.6/ highly de-stabilizing	-0.8 / slightly stabilising	-5.4 / highly_stabilizing
H51Q,S81A,T266A,V178A	60	-0.64 / slightly stabilising	-2.49 / highly_stabilizing	-10.63 / highly_stabilizing
D282E,I213M,I265T,K173R,L248F,N223S,T266A,V6 I	56	3.32/ highly de-stabilizing	-2.43 /highly_stabilizing	-7.51 / highly_stabilizing
D282E,I213M,I265T,K173R,L248F,N223S,T165S,T2 66A	54	3.86 / highly de-stabilising	-2.59 / highly_stabilizing	-3.22 / highly_stabilizing

AXB (D282E): 'A' ('D') represents wild amino acid residue, 'X'(282) represents amino acid position and 'B' ('E) represents mutation

Supplementary Table 5: It depicts ΔΔG score (kcal/mol) calculated by FoldX for two combination of mutations (comb165 and comb167) and individual mutations present in that combination for different DENV2 NS1 states. Characterization of effect of mutations on NS1 stability by ΔΔG score is represented as Feature.

		MONOMER	DIMER	HEXAMER
Mutations	**Strain count**	**ΔΔG score (kcal/mol) / Feature**	**ΔΔG score (kcal/mol) / Feature**	**ΔΔG score (kcal/mol) / Feature**
H130Y, H51Q, S81A, T266A, V178A, V287I	165	-0.80 / slightly stabilising	—5.27 / highly stabilising	--13.86 / highly stabilising
H51Q		-0.369/ Neutral	1.158 / de-stabilizing	3.70256 / highly de-stabilizing
S81A		-0.958/ stabilizing	-0.904 / slightly stabilising	-1.6628 / stabilizing
H130Y		-0.08/neutral	-0.377/ neutral	-0.717 / slightly stabilising
V178A		0.48/slightly destabilising	0.202 / neutral	0.564 / slightly destabilising
T266A		0.217/ neutral	-4.14 / highly_stabilizing	-10.9726 / highly_stabilising
V287I		-0.309/neutral	-0.63 / slightly stabilising	-1.40 / stabilizing
D282E, I213M, I265T, K173R, L248F, N223S, T266A	167	4.15 / highly de-stabilizing	-2.64617 / highly stabilising	-7.04378 / highly stabilising
K173R		-1.14 / stabilizing	-1.34 / stabilizing	-1.29 / stabilizing
I213M		1.569 / de-stabilizing	-2.66 / highly_stabilizing	-6.02 / highly_stabilizing
N223S		0.83 / slightly destabilising	-1.167 / stabilizing	0.263/neutral
L248F		1.33 / de-stabilizing	4.69 / highly de-stabilizing	6.5 /highly de-stabilizing
I265T		0.77 / slightly destabilising	1.09 / de-stabilizing	3.30614/highly de-stabilizing
T266A		0.21 / neutral	4.14 / highly_stabilizing	-10.9726/highly_stabilizing
D282E		0.22 / neutral	0.017 / neutral	-0.046 /neutral

AXB (D282E): 'A' ('D') represents wild amino acid residue, 'X'(282) represents amino acid position and 'B' ('E) represents mutation

Supplementary Table 6: Identification of virtual-conserved residues (MC/RP = 0) in dimer interface of all dengue serotypes.

AA_num	D1_Residue	AA_num	D2_Residue	AA_num	D3_Residue	AA_num	D4_Residue
4	**GLY**	**4**	**GLY**	**4**	**GLY**	**4**	**GLY**
5	**CYS**	**5**	**CYS**	**5**	**CYS**	**5**	**CYS**
		6	VAL				
16	**CYS**	**16**	**CYS**	**16**	**CYS**	**16**	**CYS**
17	**GLY**	**17**	**GLY**	**17**	**GLY**	**17**	**GLY**
19	**GLY**	**19**	**GLY**	**19**	**GLY**	**19**	**GLY**
21	PHE	21	PHE	21	PHE		
				26	VAL		
				33	TYR		
158	ASP					158	ASP
160	**GLY**	**160**	**GLY**	**160**	**GLY**	**160**	**GLY**
162	**GLY**	**162**	**GLY**	**162**	**GLY**	**162**	**GLY**
				179	LEU		
183	ARG						
		185	MET	185	MET	185	MET
187	**ALA**	**187**	**ALA**	**187**	**ALA**	**187**	**ALA**
		194	ALA				
197	**ALA**	**197**	**ALA**	**197**	**ALA**	**197**	**ALA**
202	TRP	202	TRP	202	TRP		
				207	LYS		
				209	GLY		
		211	TRP			211	TRP
226	TRP	226	TRP	226	TRP		
227	**PRO**	**227**	**PRO**	**227**	**PRO**	**227**	**PRO**
						232	LEU
236	**GLY**	**236**	**GLY**	**236**	**GLY**	**236**	**GLY**
				237	VAL	237	VAL
Total	16		18		21		16

MC/RP =Mutational capacity per residue position.
Bold residue positions and names represent virtual and sequentially conserved residues across serotypes.
Bold and underlined residues represents experimental proved critical residues in DENV2 serotype.
D1: Dengue serotype 1, D2: Dengue serotype 2 , D3:Dengue serotype 3 , D4:Dengue serotype 4

Supplementary Table 7: Residues presents in predicted protein-protein interaction sites with their residue name, Shannon entropy (SE), Relative solvent accessibility (RSA) and mutational capacity per residue position (MC/RP) scores across all four dengue serotypes. A number of virtual-conserved Proline residues were found in these sites across dengue serotypes.

AA Num	D1 Residue	D2 Residue	D3 Residue	D4 Residue	D1 SE	D2 SE	D3 SE	D4 SE	D1 RSA	D2 RSA	D3 RSA	D4 RSA	D1 MC/RP	D2 MC/RP	D3 MC/RP	D4 MC/RP
36	GLN	GLN	GLN	GLN	0.01	0.01	0.03	0.05	22.90	19.50	22.00	19.20	8.00	9.00	10.00	9.00
37	ALA	**PRO**	ALA	**PRO**	0.01	0.01	0.03	0.05	4.10	22.50	9.20	20.60	6.00	**2.00**	6.00	**2.00**
38	ASP	GLU	ASP	GLU	0.01	0.09	0.03	0.04	27.90	30.70	27.10	23.30	17.00	13.00	15.00	15.00
39	SER	SER	SER	SER	0.01	0.05	0.04	0.05	11.00	22.80	12.20	6.30	7.00	6.00	3.00	9.00
40	**PRO**	**PRO**	**PRO**	**PRO**	0.01	0.01	0.03	0.05	8.70	8.90	11.80	13.40	**0.00**	**0.00**	**0.00**	**0.00**
106	GLY	ARG	THR	ALA	0.07	0.16	0.06	1.02	90.60	51.40	74.30	68.70	19.00	13.00	14.00	17.00
107	**PRO**	**PRO**	**PRO**	**PRO**	0.00	0.01	0.07	0.09	37.70	32.20	33.20	28.50	0.00	0.00	0.00	0.00
108	GLN	GLN	GLN	PRO	0.00	0.01	0.04	0.05	36.60	30.20	32.40	22.50	15.00	15.00	10.00	0.00
109	**PRO**	**PRO**	**PRO**	VAL	0.00	0.01	0.04	0.81	13.50	21.20	14.70	37.50	**0.00**	**0.00**	**1.00**	17.00
110	MET	THR	MET	ASN	0.01	0.09	0.13	1.02	106.00	77.70	97.60	97.20	15.00	13.00	12.00	11.00
139	**PRO**	**PRO**	**PRO**	**PRO**	0.01	0.02	0.08	0.05	87.40	70.80	91.20	71.20	**3.00**	**1.00**	**3.00**	**1.00**
140	ASN	GLU	SER	ASP	0.87	0.02	0.15	0.08	78.30	52.90	61.70	67.80	7.00	2.00	16.00	1.00
141	THR	THR	THR	THR	0.00	0.01	0.04	0.05	24.70	37.60	26.20	43.50	3.00	18.00	8.00	7.00
142	**PRO**	ALA	**PRO**	SER	0.04	0.19	0.02	0.15	97.10	112.60	99.30	116.40	**0.00**	13.00	**0.00**	19.00
143	GLU	GLU	GLU	GLU	0.00	0.04	0.05	0.05	80.70	76.10	86.50	70.80	0.00	9.00	2.00	10.00
144	CYS	CYS	CYS	CYS	0.00	0.02	0.06	0.05	0.00	0.00	0.10	0.00	0.00	0.00	1.00	0.00
145	**PRO**	**PRO**	**PRO**	**PRO**	0.18	0.02	0.05	0.05	49.70	47.10	57.40	45.30	**0.00**	**1.00**	**0.00**	**0.00**
146	ASP	ASN	SER	ASN	0.03	0.07	0.42	0.05	37.00	28.30	26.00	33.10	1.00	18.00	19.00	10.00
316	SER	SER	SER	SER	0.01	0.01	0.05	0.02	90.10	87.90	83.90	84.40	7.00	5.00	1.00	7.00
317	CYS	CYS	CYS	CYS	0.00	0.01	0.05	0.05	0.00	0.00	0.00	0.60	0.00	0.00	0.00	0.00
318	THR	THR	THR	THR	0.00	0.01	0.05	0.05	46.30	46.10	47.00	33.80	17.00	18.00	18.00	17.00
319	LEU	LEU	LEU	MET	0.01	0.02	0.05	0.05	40.20	30.90	26.80	60.00	2.00	1.00	1.00	0.00
320	**PRO**	**PRO**	**PRO**	**PRO**	0.00	0.02	0.04	0.05	52.50	51.20	52.10	48.10	**0.00**	**0.00**	**0.00**	**0.00**
321	**PRO**	**PRO**	**PRO**	**PRO**	0.00	0.01	0.06	0.05	0.00	0.00	0.00	0.10	**0.00**	**0.00**	**0.00**	**0.00**
322	LEU	LEU	LEU	LEU	0.00	0.03	0.05	0.07	0.00	0.00	0.70	0.90	1.00	0.00	0.00	1.00

Bold represents Proline residues and their respective MC/RP scores.
D1: Dengue serotype 1, D2: Dengue serotype 2 , D3:Dengue serotype 3 , D4:Dengue serotype 4

Supplementary Table 8: Observed highly variable regions of Influenza NS1 protein and their Shannon entropy (SE), and mutational capacity per residue position (MC/RP) scores. Like Dengue NS1 protein, observed-variable regions of Influenza NS1 were also predicted to be virtual-variable (MC/RP >0) with two exceptions (A60 and M119)

AA_num	Wild aa	SE	MC/RP score	AA_num	Wild aa	SE	MC/RP score
6	V	1.19	14	91	T	1.38	17
21	R	1.22	11	95	L	1.23	9
22	F	1.27	3	98	M	1.19	1
23	A	1.19	2	101	D	1.28	4
25	Q	1.69	10	108	K	1.09	9
26	E	1.86	15	111	V	1.43	11
28	G	1.15	4	112	A	2.04	18
44	R	1.10	13	114	S	1.33	1
48	S	1.13	18	117	I	1.05	2
55	E	1.38	2	**119**	**M**	**1.03**	**0**
56	T	1.71	16	125	D	1.21	14
59	R	2.17	13	127	N	1.22	11
60	**A**	**1.56**	**0**	129	I	1.95	2
67	R	2.09	9	139	D	1.66	17
70	E	1.25	14	145	I	1.12	4
74	D	1.27	14	166	L	1.21	1
76	A	1.48	18	171	D	2.49	12
81	I	1.14	11	189	D	1.07	19
84	V	1.63	15	194	V	1.23	3
86	A	1.14	19	197	T	1.31	7
90	L	1.13	7				

Bold represents virtual-conserved and observe-variable residues.
AA_num: Amino acid number, Wild aa: Amino acid present in reference strain.

Supplementary Figures

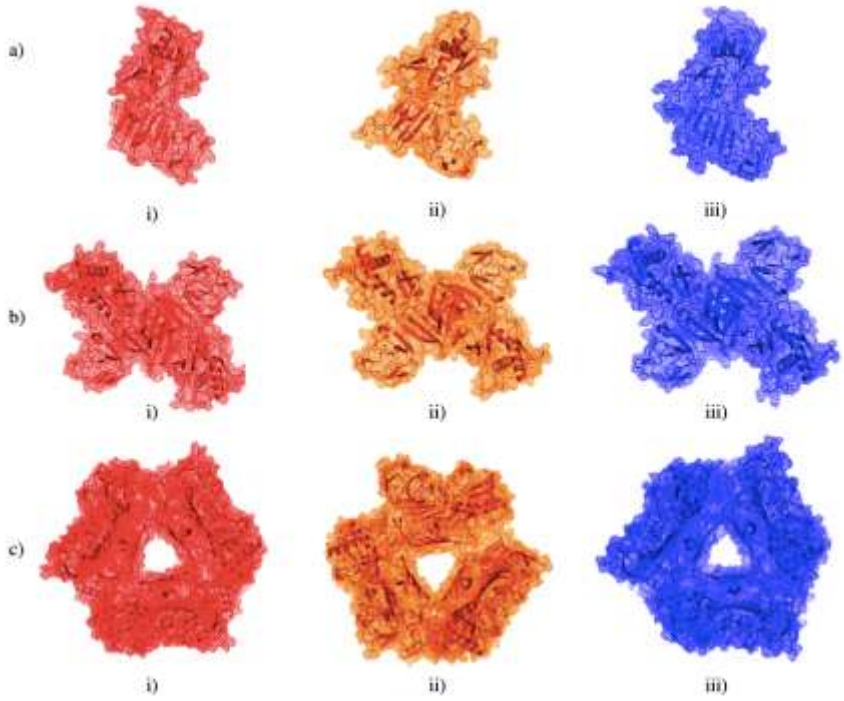

Figure S1: Monomer, dimer and hexamer models of Dengue serotypes through homology modelling by modeller9.2.
a) Monomer models of i) DENV1-NS1 ii) DENV3 -NS1 iii) DENV4 NS1
b) Dimer models of i) DENV1-NS1 ii) DENV3-NS1 iii) DENV4-NS1
c) Hexamer models of i) DENV1-NS1 ii) DENV3 -NS1 iii) DENV4-NS1
Dengue NS1 protein structure was used as template from above models.

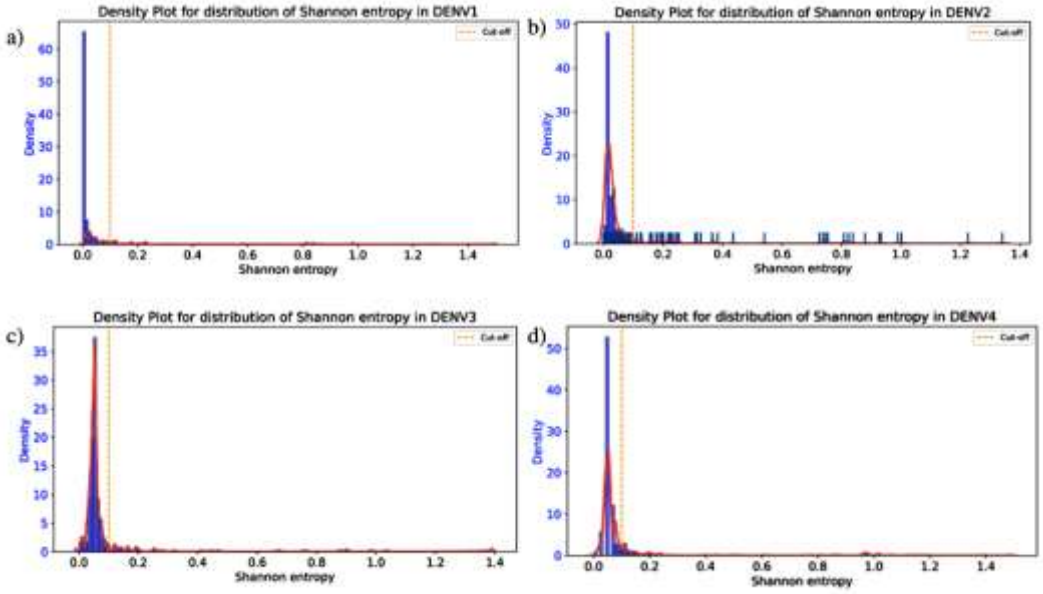

Figure S2: Density distribution analysis of Shannon entropy (SE) scores calculated for each dengue serotype.
a). DENV1 b). DENV2 c). DENV3 d). DENV4
X-axis represents SE score values and Y-axis represents density distribution of that particular SE score. Cut-off values used in this study to differentiate between low SE and high SE residue positions is shown in dotted vertical orange line.

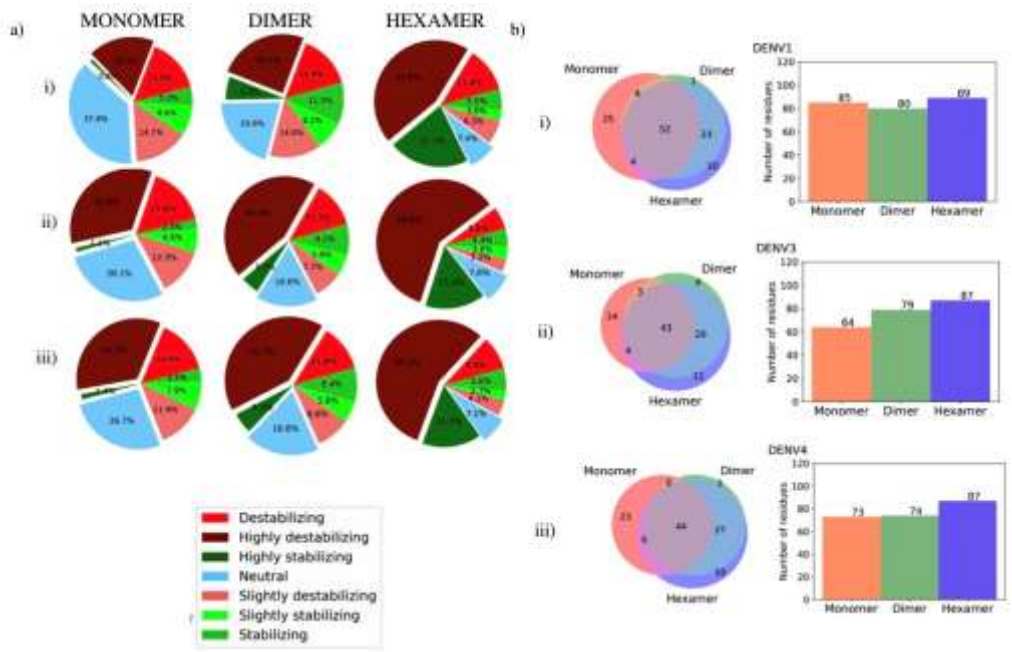

Figure S3: Effect of known single mutations on NS1 protein stability and prediction of virtual-conserved and -variable residues in Dengue serotypes.

a) The distribution of effect of known single mutation on i) DENV1 ii) DENV3 iii) DENV4 NS1 structure stability is represented as pie charts. Classification of effect of mutations was based on free energy ($\Delta\Delta G$) difference calculated by FoldX 5.0. Each sub-panel consists three pie-chart which represents distribution of classes across monomer, dimer and hexamer states.

b) It represents distribution of virtual-conserved positions in i) DENV1 ii) DENV3 iii) DENV4 NS1 protein respectively. Venn-diagram (left panel) represents overlap and uniqueness of such region among NS1 states (monomer, dimer and hexamer) whereas Bar chart (right panel) represents changes in total number of virtual-conserved positions across three states.

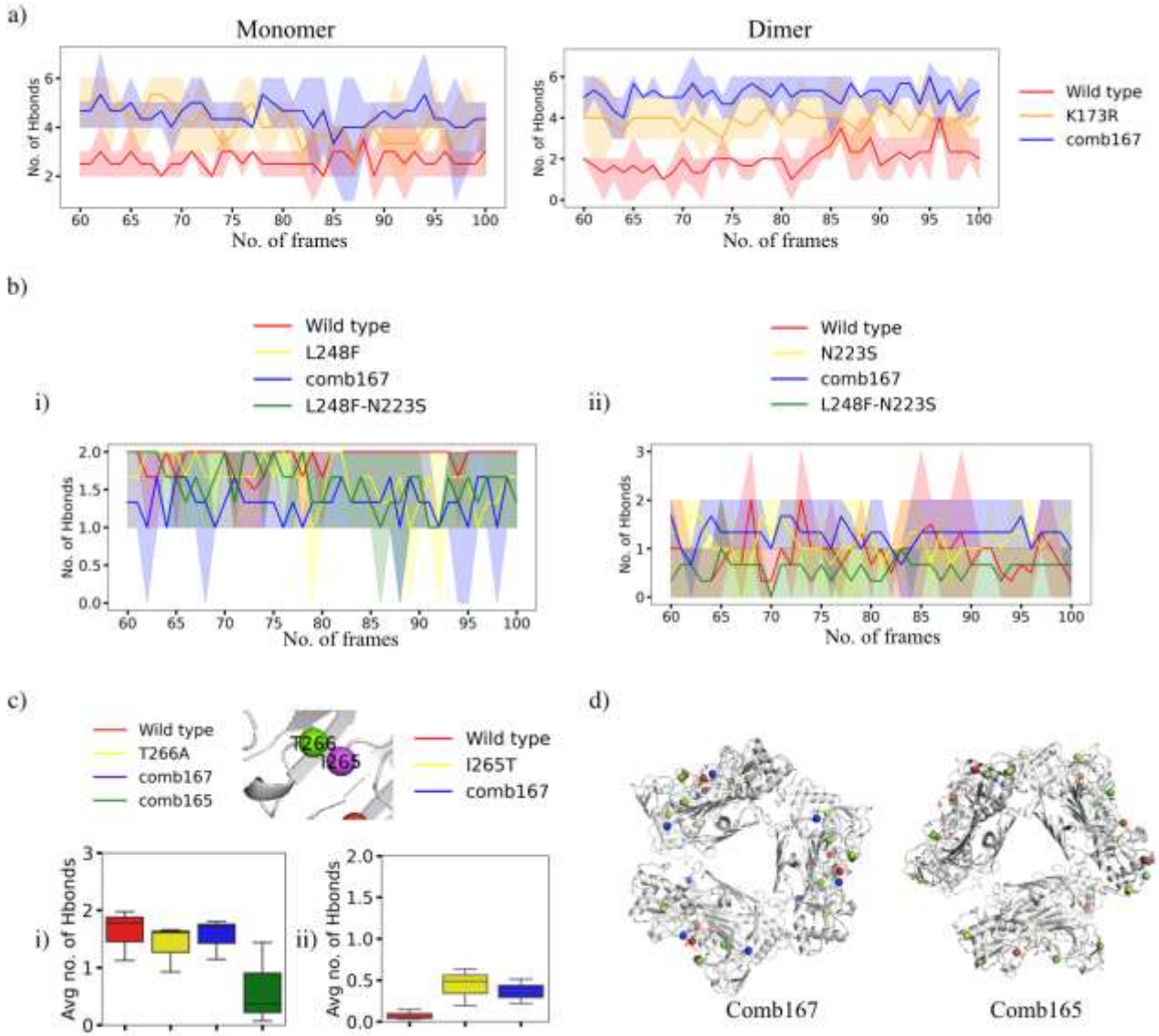

Figure S4: Amino acid interaction analysis of residue positions in wild type vs single vs combination of mutation across NS1 states. This figure also showed quantification of effect of individual mutations present in two pairs of compensatory mutations.

a) Change in amino acid interaction across 5Å of residues position 173 in wild-type (K173, red) vs single mutations (K173R, orange) vs combination of mutations (comb167, blue) in DENV2 NS1 monomer and dimer, respectively. Line chart representing changes in number of hydrogen bonds (h-bonds) formed across 60-100 ns of simulation trajectory (N=3). Formation of h-bonds (solid lines) increases more evidently in dimer state.

b & c) Case study of two possible compensatory mutation pairs (b) L248F-N223S, and (c) I265T-T266A. A gradual decrease in formation of h-bonds seen around 248 position (b-i) in comb167 (green) in compared to wild-type (red), single (blue), and double-mutant (green). Similarly, an opposite trend was observed around 223 position (b-ii). H-bond formation increased around 5Å of residues position 223 in comb167 (blue) as compared to wild-type (red), single (yellow), and double-mutant (green). Other pair of mutation (I265T-T266A) present on linear sequence (c) with different individual effect on DENV2 NS1 stability (I265T- destabilizing and T266A-stabilizing). Changes in amino acid interaction across (5Å) in wild-type (red) vs single (yellow) vs combination of mutations (comb167 (blue) and comb165 (green)) is quantified. Box plot represents average number of h-bonds formed between 60-100ns of simulation trajectory of (c-i) T266A and (c-ii) I265T.

d) Combination of mutations, comb167 and comb165 are represented as dots on DENV2 NS1 hexamer state, respectively. Destabilizing (red), neutral (blue), and stabilizing (green) mutations represent in different color shades.

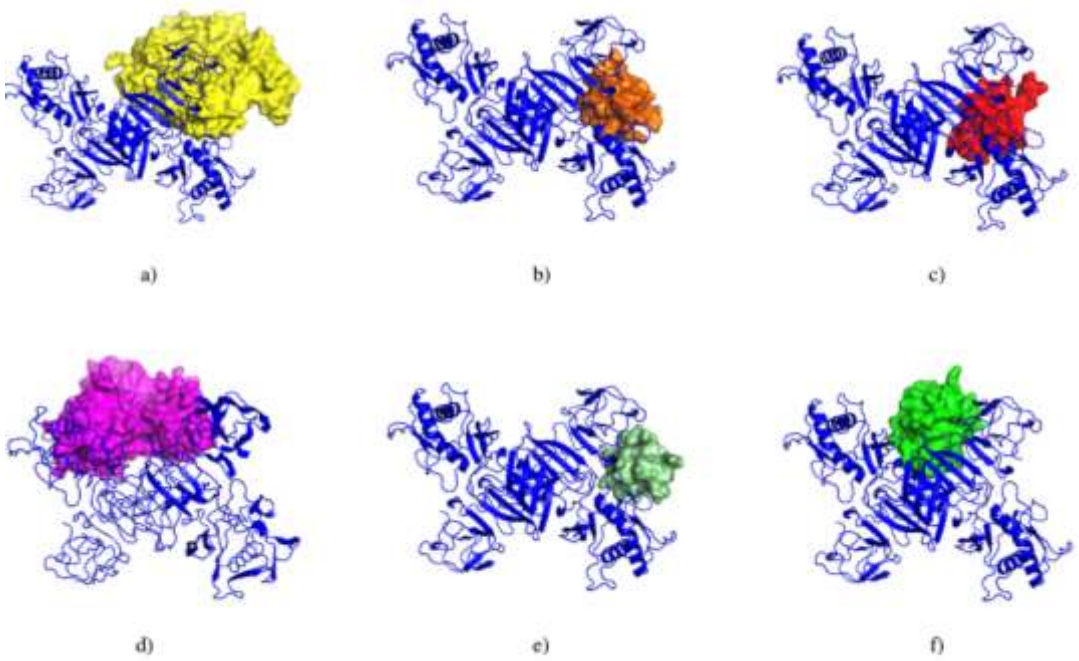

Figure S5: Best predicted models for protein-protein interaction between DENV2 NS1 and human protein ligands.
a). DENV2 NS1- Abl kinase b) DENV2 NS1-Cortactin c) DENV2 NS1- Lyn kinase
d) DENV2 NS1- Src kinase e) DENV2 NS1- Yes kinase f) DENV2 NS1 – Profillin.
Ligand proteins are known to interact with Proline-rich interaction motifs.

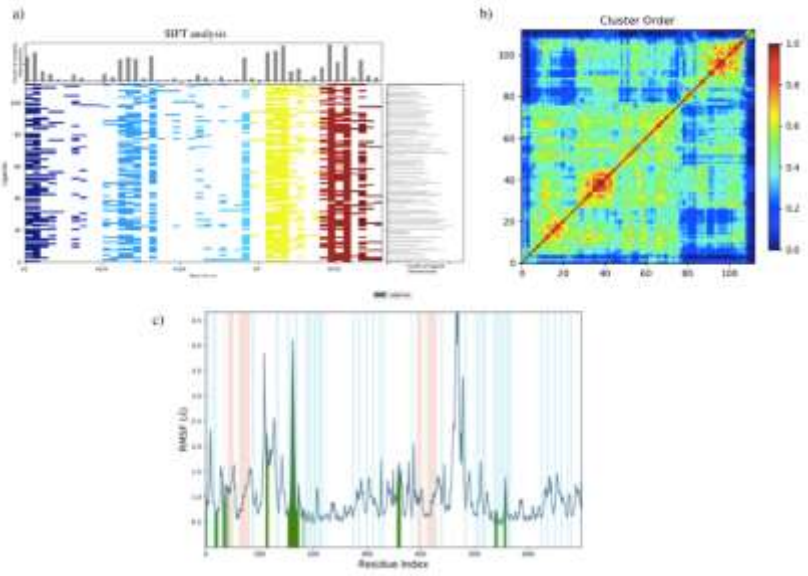

Figure S6:Clustering of screened small molecule inhibitors against identified drug-able site.
a) Structural interaction fingerprint analysis (SIFT) of protein-ligand analysis. Y-axes represents number of ligands and X-axis represent residue position in NS1 protein. Count of residue and ligand interaction is mentioned above and on-left side of plot respectively.
b) Representation of clustering based on tanimoto coefficient.
c) RMSF (root mean square fluctuation) plot showing fluctuation observed in different secondary structure of NS1 protein throughout the simulation. Loop regions (white space) constitute more fluctuations compared to helix (orange) and beta-sheets (blue). Green region represents the protein-ligand interacting regions.

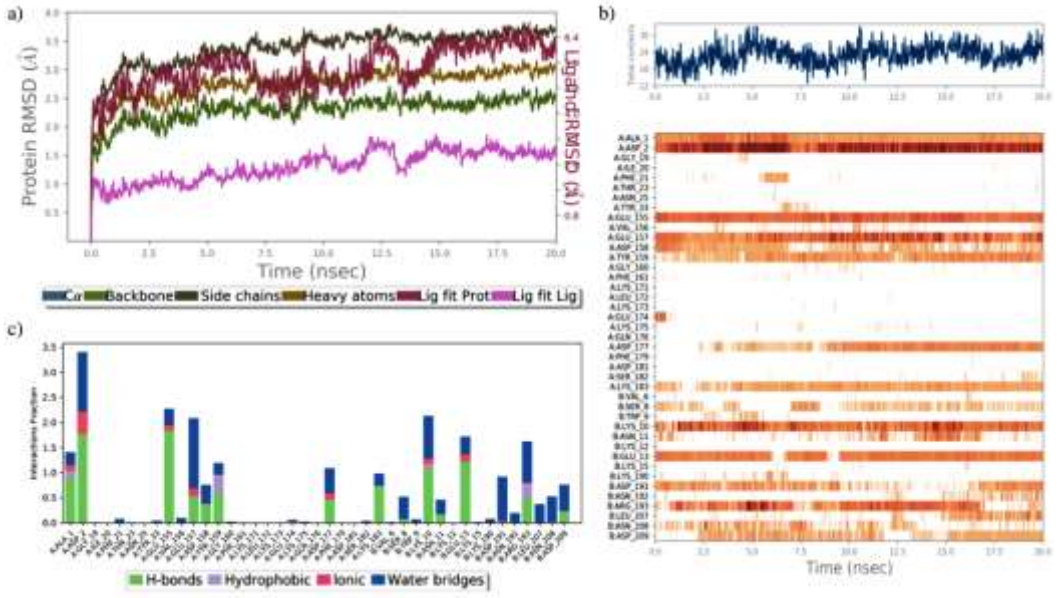

Figure S7: MD simulation analysis of DENV2 NS1 and Brilacidin protein-ligand complex.
a) RMSD (Root mean square deviation) plot represents fluctuation in NS1 protein structure throughout out the simulation run. RMSD is calculated with respect to reference structure (NS1 structure at time 0ns). Ligand RMSD represents how stable the ligand is with respect to protein and ligand pocket and has shown in Red dotted line.
b) Simulation Interactions Diagramme representing qualitative analysis of NS1 residues(X-axis) and Brilacidin interaction throughout the simulation run. Different subtype of interactions is shown in different stack colors whereas height of stacked-bar represents relative time period for which respective residue was interacting with ligand. Value 1 represents that residue was interacting for whole simulation run. Some residues found to have value>1 which suggested these residues might involve in multiple interactions of same subtype with different ligand atoms,
c) Timeline representation of protein-ligand interaction where x-axis represents interaction made by ligand during simulation run and y-axis represents interacting residue

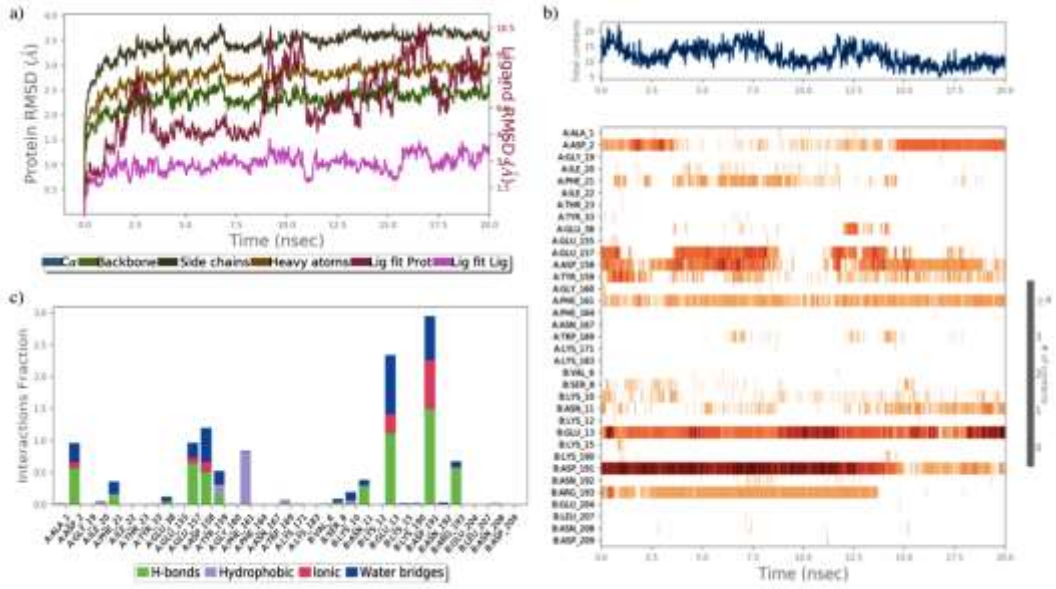

Figure S8: MD simulation analysis of DENV2 NS1 and Lividomycin A protein-ligand complex.
a) RMSD (Root mean square deviation) plot represents fluctuation in NS1 protein structure throughout out the simulation run. RMSD is calculated with respect to reference structure (NS1 structure at time 0ns). Ligand RMSD represents how stable the ligand is with respect to protein and ligand pocket and has shown in Red dotted line.
b) Simulation Interactions Diagramme representing qualitative analysis of NS1 residues(X-axis) and Lividomycin A interaction throughout the simulation run. Different subtype of interactions is shown in different stack colors whereas height of stacked-bar represents relative time period for which respective residue was interacting with ligand. Value 1 represents that residue was interacting for whole simulation run. Some residues found to have value>1 which suggested these residues might involve in multiple interactions of same subtype with different ligand atoms,
c) Timeline representation of protein-ligand interaction where x-axis represents interaction made by ligand during simulation run and y-axis represents interacting residue

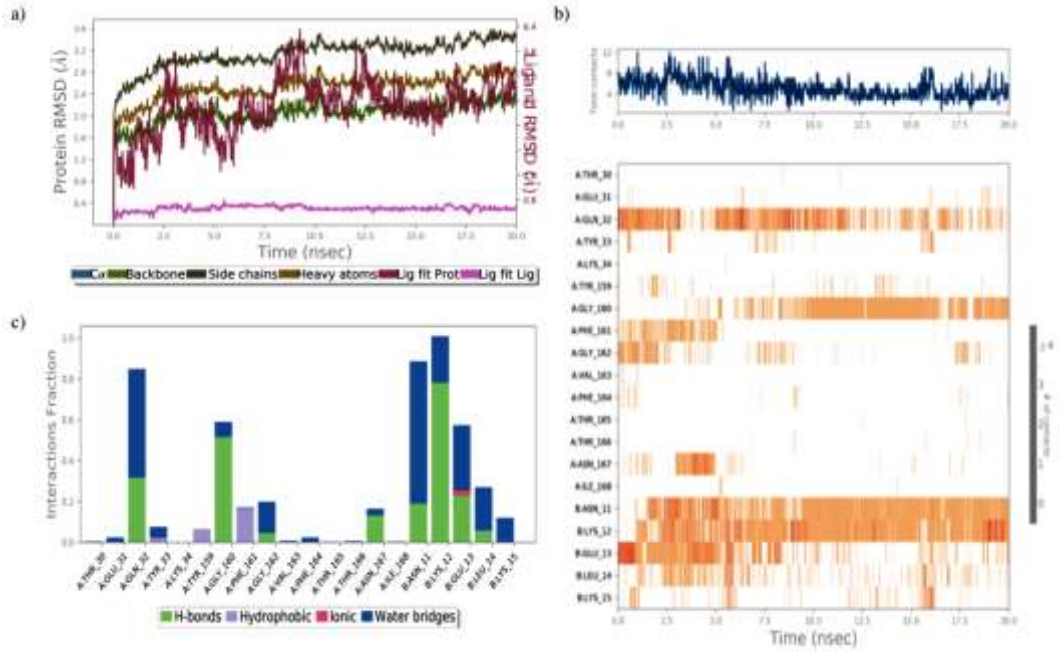

Figure S9: MD simulation analysis of DENV2 NS1 and Bradykinin protein-ligand complex.
a) RMSD (Root mean square deviation) plot represents fluctuation in NS1 protein structure throughout out the simulation run. RMSD is calculated with respect to reference structure (NS1 structure at time 0ns). Ligand RMSD represents how stable the ligand is with respect to protein and ligand pocket and has shown in Red dotted line.
b) Simulation Interactions Diagramme representing qualitative analysis of NS1 residues(X-axis) and Bradykinin interaction throughout the simulation run. Different subtype of interactions is shown in different stack colors whereas height of stacked-bar represents relative time period for which respective residue was interacting with ligand. Value 1 represents that residue was interacting for whole simulation run. Some residues found to have value>1 which suggested these residues might involve in multiple interactions of same subtype with different ligand atoms,
c) Timeline representation of protein-ligand interaction where x-axis represents interaction made by ligand during simulation run and y-axis represents interacting residue

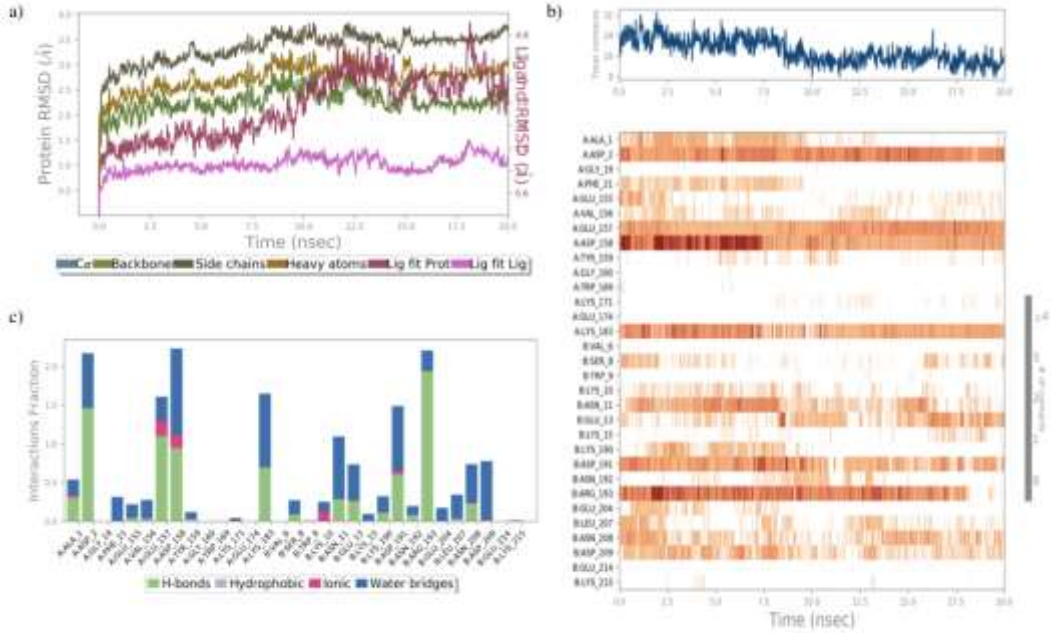

Figure S10: MD simulation analysis of DENV2 NS1 and Spectinomycin protein-ligand complex.
a) RMSD (Root mean square deviation) plot represents fluctuation in NS1 protein structure throughout out the simulation run. RMSD is calculated with respect to reference structure (NS1 structure at time 0ns). Ligand RMSD represents how stable the ligand is with respect to protein and ligand pocket and has shown in Red dotted line.
b) Simulation Interactions Diagramme representing qualitative analysis of NS1 residues(X-axis) and Spectinomycin interaction throughout the simulation run. Different subtype of interactions is shown in different stack colors whereas height of stacked-bar represents relative time period for which respective residue was interacting with ligand. Value 1 represents that residue was interacting for whole simulation run. Some residues found to have value>1 which suggested these residues might involve in multiple interactions of same subtype with different ligand atoms,
c) Timeline representation of protein-ligand interaction where x-axis represents interaction made by ligand during simulation run and y-axis represents interacting residue

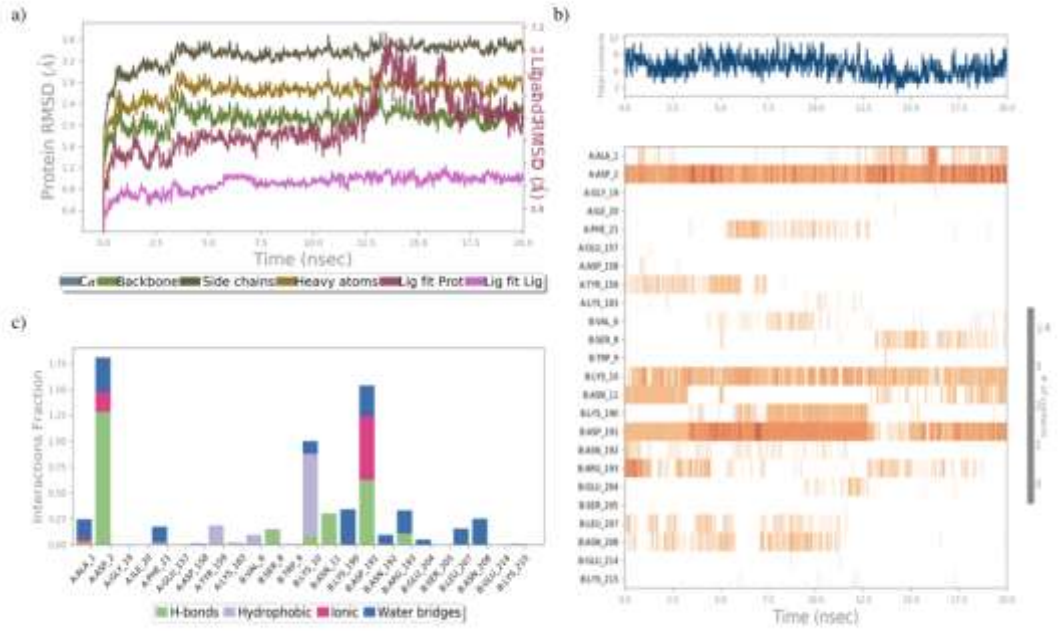

Figure S11: MD simulation analysis of DENV2 NS1 and Nadolol protein-ligand complex.
a) RMSD (Root mean square deviation) plot represents fluctuation in NS1 protein structure throughout out the simulation run. RMSD is calculated with respect to reference structure (NS1 structure at time 0ns). Ligand RMSD represents how stable the ligand is with respect to protein and ligand pocket and has shown in Red dotted line.
b) Simulation Interactions Diagramme representing qualitative analysis of NS1 residues(X-axis) and Nadolol interaction throughout the simulation run. Different subtype of interactions is shown in different stack colors whereas height of stacked-bar represents relative time period for which respective residue was interacting with ligand. Value 1 represents that residue was interacting for whole simulation run. Some residues found to have value>1 which suggested these residues might involve in multiple interactions of same subtype with different ligand atoms,
c) Timeline representation of protein-ligand interaction where x-axis represents interaction made by ligand during simulation run and y-axis represents interacting residue

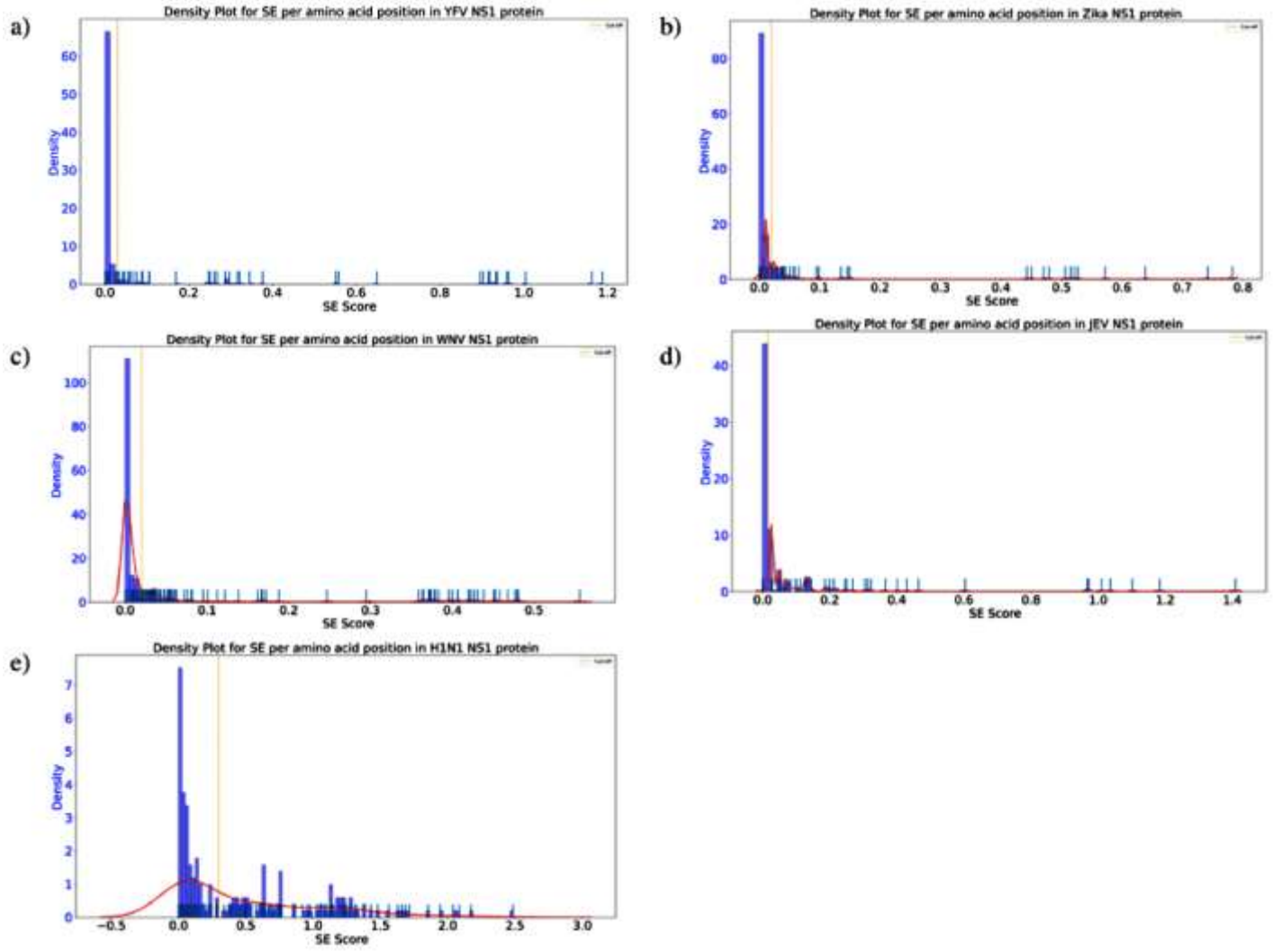

Figure S12:Density distribution analysis of SE scores for other NS1 proteins.
a). Yellow fever virus (YFV) b). Zika c). West nile virus (WNV) d). Japanese encephalitis virus (JEV) e) H1N1 (Human Influenza 1).
Cut-off SE score for categorizing residue positions in observed conserved (low SE) and variable (high SE) positions was determined by density distribution analysis of SE scores. **X-**axis represents SE score values and Y-axis represents density distribution of positons at a particular SE score. Cut-off values is shown in dotted vertical orange line. Cut-off was different for each NS1 protein, and selected based on dip observed in curved line (red). SE denotes 'Shannon entropy'.

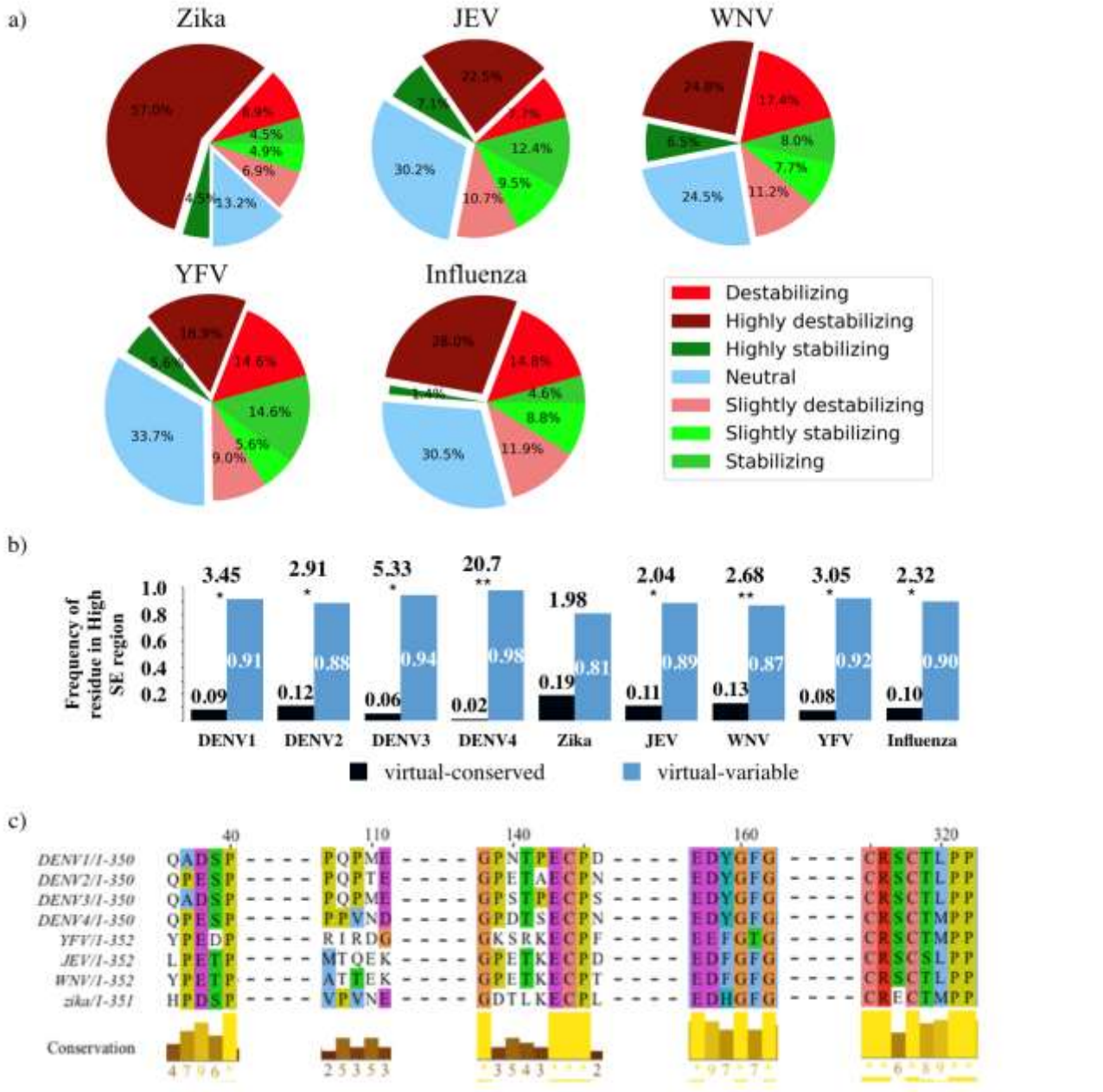

Figure S13:Similar sequence and structure analysis of other NS1 (flavivirus and influenza) proteins showed comparable trends to DENV NS1 protein.

a) The distribution of effect of identified single mutations on Zika, Japanese encephalitis virus (JEV), West nile virus (WNV), Yellow fever virus (YFV), and Influenza viruses is shown as pie chart. The classification of mutation's effect was performed based on free energy differences (ΔΔG) calculated from Foldx 5.0.

b) Enrichment virtual-variable (MC/RP=0) residue in observed-variable (high SE) positions was consistent for most of the NS1 proteins used in this study. Each bar represents frequency of high SE position in set of virtual-conserved vs virtual-variable positions. The numbers above pairs of bars are the odds ratios and stars below them represents statistical significance (P-value) calculated according to one-tail Fisher's exact test (* - $0.01 < P < 0.05$, ** – $0.001 < P < 0.01$, *** – $P < 0.0001$ and no star for non-significant P-value).

c) Predicted motifs in this study, and their conservation across flavivirus NS1 proteins. Reference sequences for each NS1 protein were extracted from their solved structure or literature. Sequence were then aligned using Clustal-omega. Sequence conservation is shown through a colour gradient used by Clustal, with conservation score mentioned at the bottom of alignment.

Acknowledgments:

We thank NCBS (TIFR) for its infrastructural facilities. We thank Ms Revathy Menon for help in improving the manuscript. This work was partly supported by the capacity component of the Indo Africa dengue sequencing to vaccine grant from Narayana Murthy (Infosys) and the NCBS core funds to SK. RS acknowledges funding and support provided by the JC Bose Fellowship (SB/S2/JBR/2021/000006) from the Science and Engineering Research Board, India, her Kiran Mazumdar Shaw Computational Biology Chair grant at the Institute of Bioinformatics and Applied Biotechnology, Bangalore (IBAB/MSCB/182/2022) and Bioinformatics Centre Grant funded by Department of Biotechnology, India (BT/PR40187/BTIS/137/9/2021).