

# SCIENTIFIC REPORTS



OPEN

## A molecular dynamics simulation study decodes the Zika virus NS5 methyltransferase bound to SAH and RNA analogue

Chih-Hung Chuang<sup>5</sup>, Shean-jaw Chiou<sup>1</sup>, Tian-Lu Cheng<sup>3,5</sup> & Yeng-Tseng Wang<sup>1,2,3,4,5</sup>

Since 2015, widespread Zika virus outbreaks in Central and South America have caused increases in microcephaly cases, and this acute problem requires urgent attention. We employed molecular dynamics and Gaussian accelerated molecular dynamics techniques to investigate the structure of Zika NS5 protein with S-adenosyl-L-homocysteine (SAH) and an RNA analogue, namely 7-methylguanosine 5'-triphosphate (m7GTP). For the binding motif of Zika virus NS5 protein and SAH, we suggest that the four Zika NS5 substructures (residue orders: 101–112, 54–86, 127–136 and 146–161) and the residues (Ser56, Gly81, Arg84, Trp87, Thr104, Gly106, Gly107, His110, Asp146, Ile147, and Gly148) might be responsible for the selectivity of the new Zika virus drugs. For the binding motif of Zika NS5 protein and m7GTP, we suggest that the three Zika NS5 substructures (residue orders: 11–31, 146–161 and 207–218) and the residues (Asn17, Phe24, Lys28, Lys29, Ser150, Arg213, and Ser215) might be responsible for the selectivity of the new Zika virus drugs.

Zika virus belongs to the *Flavivirus* genus and is transmitted by the *Aedes aegypti* and *Aedes albopictus* mosquitoes. Since 2015, widespread Zika virus outbreaks in Central and South America have caused increases in microcephaly cases. Microcephaly is a birth defect that results in babies born with parts of the brain and skull missing<sup>1</sup>. The Zika virus contains a positive single strand RNA genome of approximately 11 kb that encodes a polyprotein, which leads to three structural proteins, namely the envelope (E), membrane precursor (PrM), and capsid (C), and seven nonstructural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5)<sup>2</sup>. The three structural proteins contribute to the viral particles and the seven nonstructural proteins contribute to viral replication. The E protein is a major component of the viral surface and is involved with aspects of viral replication such as membrane fusion and host cell binding<sup>2</sup>. The NS1, NS3, and NS5 nonstructural proteins are large and highly-conserved proteins, whereas the NS2A, NS2B, NS4A, and NS4B nonstructural proteins are small and hydrophobic<sup>3</sup>. Grant *et al.* showed that NS5 is a promising target for vaccines and drugs against Zika and related viruses<sup>4</sup>. The 103-kDa NS5 protein is the largest viral protein whose C-terminal portion has RNA-dependent RNA polymerase activity and N-terminal RNA cap-processing activity (methyltransferase domain)<sup>5,6</sup>. In addition, Zika NS5 can promote the proteasomal degradation of host signal transducer and can activate transcription 2 (STAT2) to inhibit type I interferon (IFN) signalling and thus antagonize the host antiviral response<sup>4</sup>. The full-length Zika NS5 proteins were resolved by Li *et al.* (Protein Data Bank (PDB) ID: 5u0b; NS5 bound to S-adenosyl-L-homocysteine (SAH))<sup>7</sup>. The x-ray structure is the initial target used in our binding mechanism studies.

Computational simulations have been applied to study Zika virus-related proteins<sup>8</sup>. Unfortunately, these modelling structures contain insufficient inhibitor-protein interaction information. Understanding the conformational dynamics of inhibitor-protein structures, and the energy produced by inhibitor-protein interactions is crucial for translating conformational samplings into functional efficacies. The process of inhibitor-protein binding occurs in a microsecond, making detailed conformational states difficult to achieve with experimental data.

<sup>1</sup>Department of Biochemistry, School of Medicine, College of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan. <sup>2</sup>Center for Biomarkers and Biotech Drugs, Kaohsiung Medical University, Kaohsiung, Taiwan. <sup>3</sup>Graduate Institute of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan. <sup>4</sup>Department of Medical Research, Kaohsiung Medical University Hospital, Kaohsiung, Taiwan. <sup>5</sup>Department of Medical Laboratory Science and Biotechnology, Kaohsiung Medical University Hospital, Kaohsiung, Taiwan. Correspondence and requests for materials should be addressed to Y.-T.W. (email: [c00jsw00@kmu.edu.tw](mailto:c00jsw00@kmu.edu.tw))

Name	IC50 (μM)	ΔGbind(kcal/mol) (experimental)	ΔGbind(kcal/mol) (Our GaMD)	ΔGbind(kcal/mol) (WHAM/umbrella sampling)
SAH	0.43	8.82	8.42 ± 1.98	7.54
M7GTP	184.00	5.17	5.21 ± 1.21	5.91

**Table 1.** Experimental data of the Zika NS5 protein with SAH and RNA analogues<sup>21</sup>. \*The 2D structures are shown in Fig. S25.

Currently, two major issues occur due to (1) the conformational states of protein receptors between binding and unbinding states and (2) the pathway of the inhibitor-protein interaction process.

All-atoms molecular dynamics (MD) simulations remain limited to the conformational ensembles obtained from a single long-time-scale conventional molecular dynamics (cMD) simulation due to the possible energy barriers between various intermediate states. Therefore, a multiscale simulation method that combines an enhanced sampling technique, which can take samples at various intermediate states, with an all-atoms simulation is required. Enhanced sampling techniques have been successfully applied to calculate the free-energy profiles<sup>9</sup> and to perform conformational sampling through accelerated molecular dynamics (aMD)<sup>10</sup>. These enhanced sampling methods can provide key insights into the free-energy profiles and intermediate protein structures. In general, the disadvantage of enhanced sampling techniques is the requirement to predefine protein structures, potential energies, additional energies, and reaction coordinates. Simulations using aMD or Gaussian aMD (GaMD) are enhanced sampling methods that can avoid such requirements. Through aMD, a boost of potential energy is added to the potential energy surface system resulting in a decreased energy barrier, which enables the acceleration of the transitions between low-energy states<sup>10–12</sup>. This method has been successfully applied to simulations of biological systems, and hundreds-of-nanoseconds aMD simulations can yield the same results as millisecond cMD simulations<sup>13–16</sup>.

In recent studies, the aMD method has produced substantial energetic noise during reweighting<sup>17</sup>. In aMD simulations, the applied boost potential is typically of the order of tens-to-hundreds of kilocalories per mole, which is usually greater than those of other enhanced sampling methods. Accurately reweighting aMD simulations has been problematic, especially for large protein molecules<sup>18</sup>. By introducing GaMD, Miao *et al.* provided an approach to improving the aMD method. The boost potential of the GaMD method follows a near-Gaussian distribution, for which the cumulated second-order expansion improves the reweighting of aMD simulations<sup>19</sup>. The reweighted free-energy profiles yielded by GaMD are in close agreement with long-time-scale cMD simulations<sup>20</sup>.

In the present study, the experimental biological binding affinities<sup>21</sup> (IC50) of SAH and 7-methylguanosine 5'-triphosphate (m7GTP) were compared using GaMD simulations; these compounds are listed in Table 1. The transfer function ( $\Delta G_{\text{bind}} = -RT \ln(\text{IC}_{50})$ ) is used to convert the IC50 values into experimental  $\Delta G_{\text{bind}}$  values, which are listed in Table 1. We applied GaMD to simulate the Zika virus NS5 protein with SAH and m7GTP. Starting from the full-length of the Zika NS5 x-ray structure, the interaction of the two molecules with the Zika virus NS5 protein was studied. We demonstrated that GaMD simulation enables a detailed analysis of the interaction of these two molecules with the Zika NS5 protein.

## Theoretical Calculations Methods

**Gaussian accelerated molecular dynamics.** GaMD is an enhanced conformational sampling method for biomolecules that adds a harmonic boost potential to smooth the system potential energy surface<sup>19</sup>. When the system potential ( $V$ ) is lower than a referenced energy ( $E$ ), a harmonic boost potential ( $\Delta V$ ) is added as follows:

$$\Delta V = \frac{1}{2}K(E - V)^2, \text{ if } V < E \quad (1)$$

where  $K$  is a harmonic force constant. The modified system potential ( $V^*$ ) is given by

$$V^* = V + \frac{1}{2}K(E - V)^2, \text{ if } V < E \quad (2)$$

If the system potential ( $V$ ) is greater than a referenced energy ( $E$ ), the harmonic boost potential ( $\Delta V$ ) is equal to zero. By smoothing the potential energy surface for overcoming intermedia energy barriers, the boost potential satisfies the following step. For two potential energy values  $V_1$  and  $V_2$ , assume that  $V_1 < V_2$  and the biased  $V_1^* < V_2^*$ . By replacing  $V^*$  with equation (2), the relationship is expressed as follows:

$$E < \frac{1}{2}(V_1 + V_2) + \frac{1}{K} \quad (3)$$

Step (1) If  $V_1 < V_2$ , the potential difference on the smoothed energy surface should be smaller than that of the original energy surface. By replacing  $V^*$  with equation (2), the relationship is expressed as

$$E > \frac{1}{2}(V_1 + V_2) \quad (4)$$

Step (2) Combing equations (3) and (4), and the relationship  $V_{\text{min}} \leq V_1 < V_2 \leq V_{\text{max}}$ , we can derive

$$V_{max} \leq E \leq V_{min} + \frac{1}{K} \quad (5)$$

where  $V_{min}$  and  $V_{max}$  are the minimum and maximum potential energies.

Step (3) By equation (5), we can obtain

$$\frac{1}{K} \leq \frac{1}{V_{max} - V_{min}} \quad (6)$$

where  $K$  constant is defined as

$$K = K0 \left( \frac{1}{V_{max} - V_{min}} \right), 0 < K0 \leq 1 \quad (7)$$

and  $K0$  is the magnitude of the applied boost potential.

Step (4) The standard deviation (SD) of  $\Delta V$  must be sufficiently small to ensure accurate reweighting<sup>22</sup>.

$$\sigma_{\Delta V} = \sqrt{\left( \frac{\partial \Delta V}{\partial V} \Big|_{V = V_{ave}} \right)^2 \sigma_V^2} = K(E - V_{ave})\sigma_V \leq \sigma_0 \quad (8)$$

where the  $V_{ave}$  and  $\sigma_V$  are the average and SD of the potential energies, and  $\sigma_{\Delta V}$  is the SD of  $\Delta V$  with  $\sigma_0$  as a user-specified upper limit for accurate reweighting of potential energies. In our simulations, the SDs of the total potential and dihedral potential boosts are 10 kcal/mol.

Step (5) To extend step (2), if  $E = V_{max}$ , we can use equation (5) to obtain

$$K0 \leq \frac{\sigma_0 V_{max} - V_{min}}{\sigma_V V_{max} - V_{ave}} \quad (9)$$

According to equations (21) and (19),  $K0$  can be defined as:

$$K0 = \min \left\{ 1.0, \frac{\sigma_0 V_{max} - V_{min}}{\sigma_V V_{max} - V_{ave}} \right\} \quad (10)$$

Step (6) To extend step (2), if  $E = V_{min} + 1/k$ , we can use equation (8) to obtain

$$K0 \geq \left( 1 - \frac{\sigma_0}{\sigma_V} \right) \frac{V_{max} - V_{min}}{V_{max} - V_{ave}} \quad (11)$$

Step (7) GaMD provides the total potential boost, dihedral potential boost, and dual potential boost in order to accelerate the molecular simulations. The boost potential ( $\Delta V$ ) is given as follows:

$$\Delta V = \frac{1}{2} K0 \frac{1}{V_{max} - V_{min}} (E - V)^2, \text{ if } V < E \quad (12)$$

where  $K0$  is the magnitude of the applied boost potential, and  $V_{min}$  and  $V_{max}$  are the minimum and maximum potential energies of the system. Initially,  $K0$  is equal to 1.0, and  $V_{max}$  and  $V_{min}$  are obtained through cMD simulations. The distribution and anharmonicity of the GaMD method were applied to the alanine dipeptide, chignolin, and lysozyme simulations to characterize the extent to which  $\Delta V$  follows a Gaussian distribution<sup>19</sup>.

**Reweight free-energy calculations for GaMD simulations.** The probability distribution of the selected reaction<sup>22</sup> coordinate  $A(r)$  is defined as  $P^*(A)$ , where  $r$  can be distance, angle, root-mean-square deviation, or other factors. Using the GaMD boost energies of each reaction coordinate,  $P^*(A)$  can be reweighted and defined as

$$P(A_j) = P^*(A_j) \frac{\langle e^{\beta \Delta V(r)} \rangle_j}{\sum_{j=1}^M \langle e^{\beta \Delta V(r)} \rangle_j}, J = 1 \sim M \quad (13)$$

where  $M$  is the number of bins,  $\beta$  is equal to the KBT, and  $\langle e^{\beta \Delta V(r)} \rangle_j$  is the ensemble-average factor of the  $j$ th bin. For reducing the energetic noise, the ensemble-average factor can be defined as follows:

$$\langle e^{\beta \Delta V(r)} \rangle = \exp \left\{ \sum_{K=1}^{\infty} \frac{\beta^K}{K!} C_K \right\} \quad (14)$$

According to equation (14), the first three cumulants can be defined as follows:

$$\begin{aligned} C1 &= \Delta V, \\ C2 &= \Delta V^2 - \Delta V^2, \\ C3 &= \Delta V^3 - 3\Delta V^2 \Delta V + 2\Delta V^2 \end{aligned} \quad (15)$$

The reweighted free energies can then be calculated by

$$F(A_j) = -\frac{1}{\beta} \ln P(A_j) \quad (16)$$

**Free-energy calculations (WHAM) for umbrella sampling simulations.** A harmonic potential was applied to the stretching<sup>23</sup> constraints, i.e. the distance constraints between the center of mass of the ligands and the binding pockets (Figs S1 and S7) with a force constant of 10.0 kcal/mol. The RC1 reaction coordinates follow the distance from the centre of the SAH mass to the centre of mass defined by the residues that shape the binding pocket (Gly86, Trp87, Thr104, Lys105, Asp131, Val132, Asp146, and Ile147; Fig. S1). The RC2 reaction coordinates follow the distance from the centre of the m7GTP mass to the centre of mass defined by the residues that shape the binding pocket (Lys13, Leu16, Asn17, Met19, Ser150, Ser151, and Ser215; Fig. S7). The RC1 value varied from 3.0 to 24.0 Å in 1 Å increments. The RC2 value varied from 6.0 to 22.0 Å in 1 Å increments. The MD simulations for PMF determination were performed with an initial 5-ns equilibration followed by 10-ns sampling at a given reaction coordinate value. Moreover, the umbrella sampling simulations were performed with GENESIS 1.2.0 software<sup>24</sup>. Then the free energy profiles (PMF) were analysed with the WHAM software<sup>25</sup>.

**GaMD simulation of the Zika virus NS5 protein system.** First, we modified a partial length of the Zika NS5 protein structure (PDB ID: 5kqs), and we used PyMOL software to modify the ligand (m7GDP to m7GTP). Second, we aligned the partial length of the Zika NS5 protein structure and the full length of the Zika NS5 protein structure (PDB ID: 5u0b). We then inserted the ligand (m7GTP) into the full-length Zika NS5 protein structure. The initial complex structures are Zika NS5 with SAH and Zika NS5 with m7GTP. These complex structures were generated and then inserted into TIP3P solvent molecules. The size of the complex structures was approximately 10.00 × 11.00 × 11.00 nm<sup>3</sup>. These initial complexes were then simulated using the AMBER 16 package with the AMBER FF12SB all-hydrogen amino acid and amber GAFF parameters. The geometries of the SAH and m7GTP were fully optimized, and their electrostatic potentials were obtained using a single-point calculation at the Hartree–Fock level with the 6–31 G(d,p) basis set using the GAUSSIAN 09 program<sup>26</sup>. Subsequently, partial charges were obtained by employing the restrained electrostatic potential procedure using the Antechamber package. All cMD simulations were performed in the isothermal–isobaric (NPT) assembly with a simulation temperature of 310 K, unless stated otherwise, by using the Verlet integrator with an integration time step of 0.002 ps and SHAKE constraints<sup>27</sup> for all covalent bonds involving hydrogen atoms. In the electrostatic interactions, atom-based truncation was performed using the PME<sup>28</sup> method, and the switch van der Waals function was used with a 2.00 nm cut-off for atom-pair lists. These complex structures were minimized for 100,000 conjugate gradient steps and then subjected to a 100-ns, isothermal, constant-volume MD simulation. Moreover, the final structures from these simulations were used in five dependent 5000-ns GaMD simulation calculations, and these structures were used in the umbrella sampling simulations.

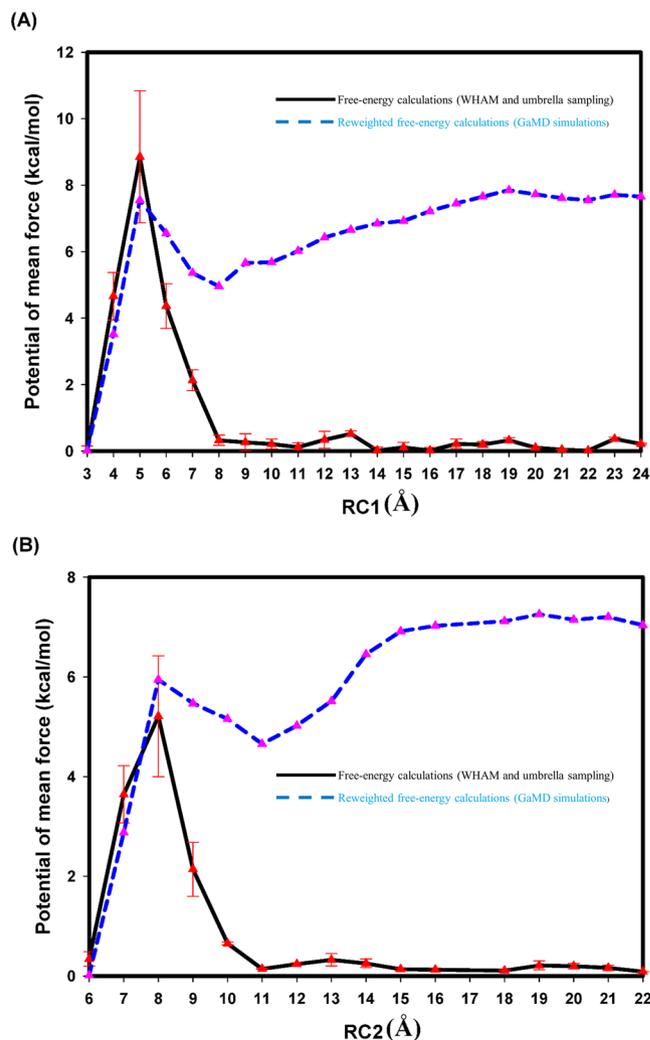
**Reweight Free-energy calculation through potential of mean force and preresidue displacement.** For the Zika NS5 protein with SAH (RC1 = 3–8 Å) and m7GTP (RC2 = 6–11 Å), all residues were obtained using preresidue displacement calculations. The preresidue displacement calculations were applied to the major barriers: RC1 = 3–8 Å and RC2 = 6–11 Å. The reaction coordinate profiles and preresidue displacement calculations were analysed using the AmberTools 16 and VMD software packages. The reaction coordinates profiles were calculated for the reaction coordinates used for the potential of mean force (PMF) calculations. The PyReweighting toolkit<sup>22</sup> was used to reweight the GaMD simulations for the PMF profile calculations and to examine the boost potential distributions. One-dimensional PMF profiles were also constructed using the reaction coordinates for the Zika NS5 protein with a bin size of 1.0 Å. When the number of simulation frames within a bin was lower than a certain limit, the bin was insufficiently sampled and thus was excluded for reweighting.

## Results and Discussion

**Potential of mean force calculation for Zika virus NS5 proteins: SAH- and m7GTP-complex GaMD simulations.** Following the original paper “Reweight free-energy calculations (PyReweighting-1D.py),”<sup>22</sup> we used five individual and independent 5000-ns GaMD simulations to estimate the error bar and average of the PMF, a measure of free energy, using the second-order cumulant expansion method<sup>22</sup>. Our results are listed in Table 1 and Fig. 1. Miao *et al.*<sup>22</sup> indicated that energy barriers were underestimated in the PyReweighting-1D.py program, leading to obvious fluctuations in the reweighted free-energy calculations. Our PMF calculations conformed to the results<sup>22</sup>. For the Zika virus NS5 protein and SAH complex, there is a major energy barrier of 8.42 ± 1.98 kcal/mol at RC1 = 3–8 Å. When RC1 = 5 Å, the complex reaches the top of the energy barrier. For the Zika virus NS5 protein and m7GTP complex, there is a major energy barrier of 5.21 ± 1.21 kcal/mol at RC2 = 6–11 Å. When the RC2 = 8 Å, the complex reaches the top of the energy barrier. Our energy barrier predictions were relatively close to the experimental data.

**Potential of mean force calculation for Zika virus NS5 proteins: SAH- and m7GTP-complex umbrella sampling simulations.** Our results are listed in Fig. 1. For the Zika virus NS5 protein and SAH complex, there is a major energy barrier of 7.54 kcal/mol at RC1 = 3–8 Å. When RC1 = 5 Å, the complex reaches the top of the energy barrier. For the Zika virus NS5 protein and m7GTP complex, there is a major energy barrier of 5.91 kcal/mol at RC2 = 6–11 Å. When the RC2 = 8 Å, the complex reaches the top of the energy barrier. Our energy barrier predictions were relatively close to our reweighted free-energy calculations.

**Functionally important residues and preresidue displacement.** The identification of functionally important residues can provide clear insight into the structural aspects of the Zika NS5 proteins. In this work, the



**Figure 1.** Free-energy profiles (PMF) at the reaction coordinates. The PMF profiles were calculated using 5000-ns GaMD simulations and umbrella sampling simulations. **(A)** Zika NS5 protein with SAH. **(B)** Zika NS5 protein with m7GTP.

structure-based approach was applied to identify the functionally important residues of the major energy barriers at  $RC1 = 3-8$  and  $RC2 = 6-11$  Å. From the complex structures at  $RC1 = 3-8$  and  $RC2 = 6-11$  Å, the important residues and pharmacophore regions were analysed using the Ligandscout program. Because there were many snapshots, residues with probability of  $>0.5$  were selected for the binding mode analysis. Our results are listed in Tables 2 and 3 (Figs S1–S12 present the representative snapshots at  $RC1 = 3-8$  and  $RC2 = 6-11$  Å). The preresidue displacement calculations in Figs 2 and 3 show that the partial length of the Zika NS5 protein (residue: 5–260) can affect the binding ability of the two analogues. For the preresidue displacement analysis of the SAH, the four Zika NS5 substructures (residue orders: 101–112, 54–86, 127–136, and 146–161) had clear fluctuations; the results are presented in Table 4 and Figs S13–S17. For the preresidue displacement analysis of m7GTP, the three Zika NS5 substructures (residue orders: 11–31, 146–161, and 207–218) had clear fluctuations; the results are displayed in Table 5 and Figs S18–S22.

**Decoding the binding of SAH with Zika virus NS5 proteins.** Our PMF profiles revealed that SAH must overcome a major energy barrier of 8.42 kcal/mol at  $RC1 = 3-8$  Å; then, this molecule can bind with the binding pockets of the Zika NS5 protein (Gly86, Trp87, Thr104, Lys105, Asp131, Val132, Asp146, and Ile147). In addition SAH must overcome the four Zika NS5 substructures (residue orders: 101–112, 54–86, 127–136, and 146–161). At  $RC1 = 8$  Å, SAH interacts with the Gly81, Asp146, Ser150, Lys182, and Arg213 residues, causing fluctuations in the four Zika NS5 substructures (residue orders: 101–112, 54–86, 127–136, and 146–161).

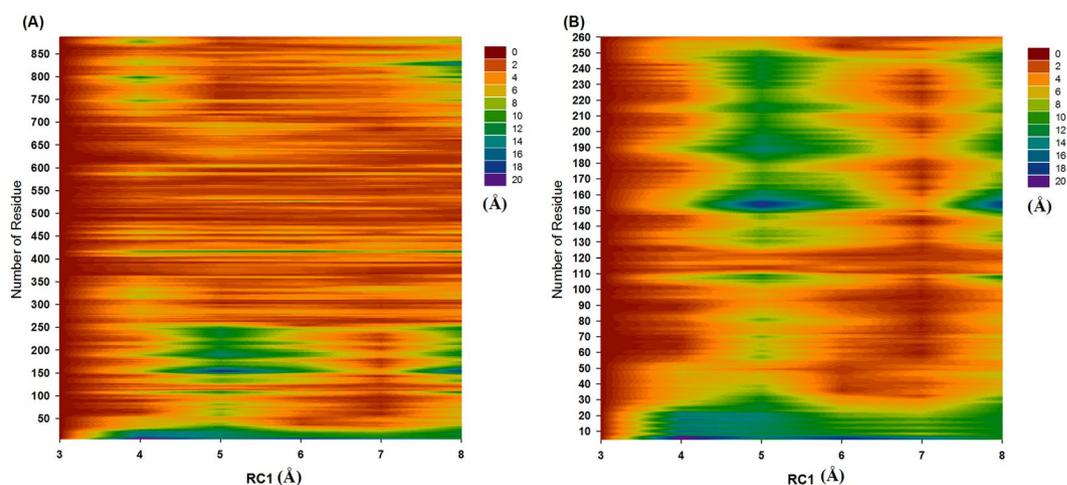
At  $RC1 = 7$  Å, SAH interacts with the Ser56, Arg84, His110, Asp146, and Ile147 residues, causing fluctuations in one Zika NS5 substructure (residue orders: 146–161). At  $RC1 = 6$  Å, SAH interacts with the Ser56, Arg84, Trp87, Asp146, Ile147, and Gly148 residues, causing fluctuations in the four Zika NS5 substructures (residue orders: 101–112, 54–86, 127–136, and 146–161). At  $RC1 = 5$  Å, SAH interacts with the Ser56, Arg84, Thr104, Asp146, Ile147, and Gly148 residues, causing fluctuations in the four Zika NS5 substructures (residue orders: 101–112, 54–86, 127–136, and 146–161). At  $RC1 = 4$  Å, SAH interacts with the Ser56, Arg84, Thr104, Gly106, Gly107, Asp146, and

RC1 (Å)	Eleostatic	Van der Waals	Hydrogen bonding
3	Trp87 and Ile147	Trp104, Asp146 and ILE147	Gly86, Lys105, Asp131 and Val132
4	Arg84 and Asp146	ILE147	Ser56, Arg84, Thr104, Gly106, Gly107 and Asp146
5	Arg84 and Asp146	ILE147	Ser56, Arg84, Thr104, Asp146 and Gly148
6	Arg84, Trp87 and Asp146	ILE147	Ser56, Arg84, Asp146 and Gly148
7	Arg84 and Asp146	ILE147	Ser56, Arg84, His110 and Asp146
8	Asp146, Lys182 and Arg213		Gly81, Asp146, Ser150 and Arg213

**Table 2.** Analysis of the Zika virus NS5 protein and SAH complex binding modes at RC1 = 3–8 Å.

RC2 (Å)	Eleostatic	Van der Waals	Hydrogen bonding
6			Lys13, Leu16, Asn17, Met19, Ser150, Ser151 and Ser215
7	Lys28 and Arg213		Asn17, Lys28, Ser150, Arg213 and Ser215
8	Lys28 and Arg213		Asn17, Lys28, Arg213 and Ser215
9	Phe24, Lys28 and Arg213		Lys28, Arg213 and Ser215
10	Lys28 and Arg213		Lys28, Arg213 and Ser215
11	Lys28, Lys29 and Arg213		Lys28 and Arg213

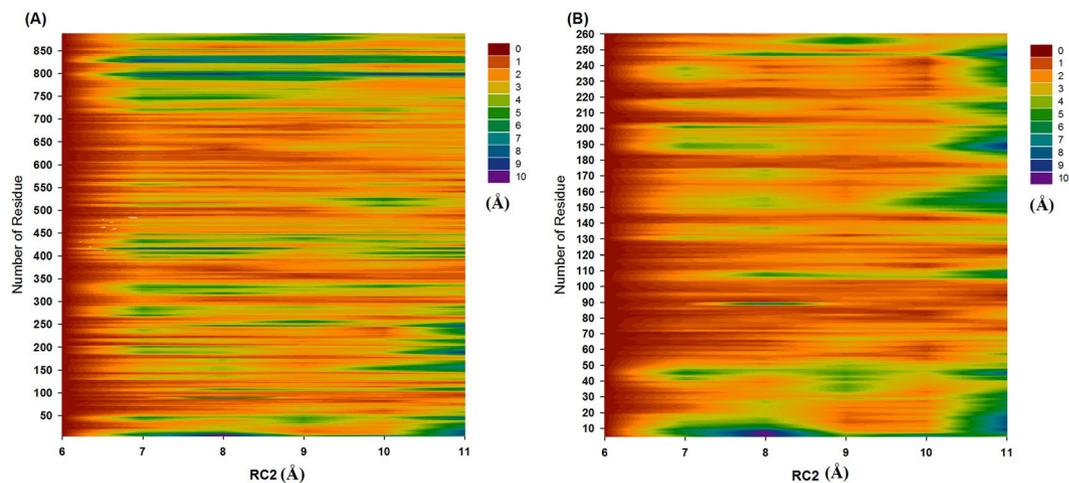
**Table 3.** Analysis of the Zika virus NS5 protein and m7GTP complex binding modes at RC2 = 6–11 Å.



**Figure 2.** Analysis of the prerésidue displacement of SAH. (A) Full length of the Zika NS5 protein and (B) partial length of the Zika NS5 protein (residue orders: 5–260).

ILE147 residues, causing fluctuations in three Zika NS5 substructures (residue orders: 101–112, 127–136, and 146–161). Our predicted binding mechanisms indicated that the four Zika NS5 substructures (residue orders: 101–112, 54–86, 127–136, and 146–161) and the Ser56, Gly81, Arg84, Trp87, Thr104, Gly106, Gly107, His110, Asp146, Ile147, and Gly148 residues affect the ability of the full-length Zika NS5 protein to bind with SAH.

**Decoding the binding of m7GTP with Zika virus NS5 proteins.** Our PMF profiles revealed that m7GTP must overcome a major energy barrier of 5.21 kcal/mol at RC2 = 6–11 Å; then, this molecule can bind with the binding pockets of the Zika NS5 protein (Lys13, Leu16, Asn17, Met19, Ser150, Ser151, and Ser215). In addition, m7GTP must overcome three Zika NS5 substructures (residue orders: 11–31, 146–161, and 207–218). At RC2 = 11 Å, m7GTP interacts with the Lys28, Lys29, and Arg213 residues, causing fluctuations in the three Zika NS5 substructures (residue orders: 11–31, 146–161, and 207–218). At RC2 = 10 Å, m7GTP interacts with the Lys28, Arg213, and Ser215 residues, causing fluctuations in the three Zika NS5 substructures (residue orders: 11–31, 146–161, and 207–218). At RC2 = 9 Å, m7GTP interacts with the Phe24, Lys28, Arg213, and Ser215 residues, causing fluctuations in the three Zika NS5 substructures (residue orders: 11–31, 146–161, and 207–218). At RC2 = 8 Å, m7GTP interacts with the Asn17, Lys28, Arg213, and Ser215 residues, causing no Zika NS5 substructure fluctuations. At RC2 = 7 Å, m7GTP interacts with the Asn17, Lys28, Ser150, Arg213, and Ser215 residues, causing fluctuations in the three Zika NS5 substructures (residue orders: 11–31, 146–161, and 207–218). Our predicted binding mechanisms indicated that the three Zika NS5 substructures (residue orders: 11–31, 146–161, and 207–218) and the Asn17, Phe24, Lys28, Lys29, Ser150, Arg213, and Ser215 residues affect the ability of the full-length Zika NS5 protein to bind with m7GTP.



**Figure 3.** Analysis of the preresidue displacement of m7GTP. **(A)** Full length of the Zika NS5 protein and **(B)** partial length of the Zika NS5 protein (residue orders: 5–260).

Range RC1 (Å)	residues
3–4	Residue:101–112, 127–136 and 146–161
4–5	Residue:101–112, 54–86, 127–136 and 146–161
5–6	Residue:101–112, 54–86, 127–136 and 146–161
6–7	Residue: 146–161
7–8	Residue:101–112, 54–86, 127–136 and 146–161

**Table 4.** Analysis of the Zika virus NS5 protein and SAH complex preresidue displacements at RC1 = 3–8 Å.

Range RC2 (Å)	residues
6–7	Residue:11–31, 146–161 and 207–218
7–8	
8–9	Residue:11–31, 146–161 and 207–218
9–10	Residue:11–31, 146–161 and 207–218
10–11	Residue:11–31, 146–161 and 207–218

**Table 5.** Analysis of the Zika virus NS5 protein and m7GTP complex preresidue displacements at RC2 = 6–11 Å.

**Comparing the SAH binding residues of the Zika virus NS5 proteins with those of other *Flavivirus* NS5 proteins.** The Ser56, Gly81, Arg84, Trp87, Thr104, Gly106, Gly107, His110, Asp146, Ile147, and Gly148 residues (PDB ID: 5U0B) were selected for comparison with other Zika NS5 proteins (PDB IDs: 5GOZ, 5GP1, 5KQR, 5KQS, 5M5B, 5TFR, 5TMH, 5ULP, 5VIM, 5WXB, 5WZ1, and 5WZ2). Our results indicated that these residues were conserved among the Zika NS5 proteins, and the results are shown in Fig. S23. The Ser56, Gly81, Arg84, Trp87, Thr104, Gly106, Gly107, His110, Asp146, Ile147, and Gly148 residues (PDB ID: 5U0B) were selected for comparison with the other *Flavivirus* NS5 proteins (PDB IDs: 3EVF (Yellow fever), 4V0R (Dengue fever), 2HKS (West Nile fever) and 4K6M (Japanese encephalitis)). Our results indicated that these residues were conserved among the NS5 proteins, and the results are shown in Fig. S24.

**Comparing the m7GTP binding residues of the Zika virus NS5 proteins with those of other *Flavivirus* NS5 proteins.** The Asn17, Phe24, Lys28, Lys29, Ser150, Arg213, and Ser215 residues (PDB ID: 5U0B) were selected for comparison with other Zika NS5 proteins (PDB IDs: 5GOZ, 5GP1, 5KQR, 5KQS, 5M5B, 5TFR, 5TMH, 5ULP, 5VIM, 5WXB, 5WZ1, and 5WZ2). Our results indicated that these residues were conserved among the Zika NS5 proteins, and the results are shown in Fig. S23. The Asn17, Phe24, Lys28, Lys29, Ser150, Arg213, and Ser215 residues (PDB ID: 5U0B) were selected for comparison with the other *Flavivirus* NS5 proteins (PDB IDs: 3EVF (Yellow fever), 4V0R (Dengue fever), 2HKS (West Nile fever) and 4K6M (Japanese encephalitis)). The results are shown in Fig. S24. Except for two residues (Lys28 and Lys29), the residues were conserved among the NS5 proteins. The two residues of the West Nile fever, Japanese encephalitis, and Yellow fever viruses were Arg-Lys, Arg-Arg, and Lys-Arg, respectively. The Arg and Lys residues have similar structures and isoelectric points. Thus, we think that these residue differences might have a reduced impact on the binding affinity of NS5 proteins with m7GTP.

## Conclusions

In this article, we used full-length Zika NS5 proteins (PDB ID: 5u0b; NS5 bound to SAH and m7GTP) as our initial structures. We employed 100-ns cMD simulations to optimize the two Zika virus NS5 protein complex structures. The RC1 reaction coordinates were defined as the distance between the centre of mass of SAH and the centre of mass of the binding pocket (Gly86, Trp87, Thr104, Lys105, Asp131, Val132, Asp146, and Ile147; Fig. S1). The RC2 reaction coordinates were defined as the distance between the centre of mass of m7GTP and the centre of mass of the binding pocket (Lys13, Leu16, Asn17, Met19, Ser150, Ser151, and Ser215). Then, we performed GaMD, preresidue displacements, and PMF calculations to predict the binding mechanisms of these molecules with Zika virus NS5 proteins. For the Zika virus NS5 protein and SAH complex, there is a major energy barrier of  $8.42 \pm 1.98$  kcal/mol RC1 = 3–8 Å. For the Zika virus NS5 protein and m7GTP complex, there is a major energy barrier of  $5.21 \pm 1.21$  kcal/mol at RC2 = 6–11 Å. Our energy barrier predictions were similar to the experimental data (Table 1). Moreover, we used the WHAM/umbrella sampling methods to check our reweighted free energy calculations. The energy barrier predictions were relatively close to our reweighted free-energy calculations. For the Zika NS5 protein and SAH complex, our results indicated that the four Zika NS5 substructures (residue orders: 101–112, 54–86, 127–136, and 146–161) and the Ser56, Gly81, Arg84, Trp87, Thr104, Gly106, Gly107, His110, Asp146, Ile147, and Gly148 residues affect the ability SAH to bind with the full-length Zika NS5 protein. For the Zika NS5 protein and m7GTP complex, our results indicated that three Zika NS5 substructures (residue orders: 11–31, 146–161, and 207–218) and the Asn17, Phe24, Lys28, Lys29, Ser150, Arg213, and Ser215 residues influenced Zika NS5 and m7GTP binding through the entrapment of this RNA analogue. Our results also indicated that the two molecules have different binding processes. For the Zika virus NS5 protein and SAH binding motif, our results suggested that the four Zika NS5 substructures (residue orders: 101–112, 54–86, 127–136, and 146–161) and the Ser56, Gly81, Arg84, Trp87, Thr104, Gly106, Gly107, His110, Asp146, Ile147, and Gly148 residues might be responsible for the selectivity of the new Zika virus drug, whereas for the Zika NS5 protein and m7GTP binding motif, our results suggested that three Zika NS5 substructures (residue orders: 11–31, 146–161, and 207–218) and the Asn17, Phe24, Lys28, Lys29, Ser150, Arg213, and Ser215 residues were responsible for the drug selectivity. We also compared our predicted binding residues for SAH and m7GTP with Zika virus with those of other *Flavivirus* NS5 proteins. The results revealed that these residues were conserved among the majority of NS5 proteins.

## References

- Mlera, L., Melik, W. & Bloom, M. E. The role of viral persistence in flavivirus biology. *Pathogens and Disease* **71**, 137–163, <https://doi.org/10.1111/2049-632x.12178> (2014).
- Faye, O. *et al.* Molecular Evolution of Zika Virus during Its Emergence in the 20th Century. *PLoS Negl Trop Dis* **8**, e2636, <https://doi.org/10.1371/journal.pntd.0002636> (2014).
- Chambers, T. J., Hahn, C. S., Galler, R. & Rice, C. M. Flavivirus Genome Organization, Expression, and Replication. *Annual Review of Microbiology* **44**, 649–688, <https://doi.org/10.1146/annurev.mi.44.100190.003245> (1990).
- Grant, A. *et al.* Zika Virus Targets Human STAT2 to Inhibit Type I Interferon Signaling. *Cell Host & Microbe* **19**, 882–890, <https://doi.org/10.1016/j.chom.2016.05.009>.
- Kapoor, M. *et al.* Association between NS3 and NS5 Proteins of Dengue Virus Type 2 in the Putative RNA Replicase Is Linked to Differential Phosphorylation of NS5. *Journal of Biological Chemistry* **270**, 19100–19106, <https://doi.org/10.1074/jbc.270.32.19100> (1995).
- Saiz, J.-C. *et al.* Zika Virus: the Latest Newcomer. *Frontiers in Microbiology* **7**, <https://doi.org/10.3389/fmicb.2016.00496> (2016).
- Zhao, B. *et al.* Structure and function of the Zika virus full-length NS5 protein. *Nature Communications* **8**, 14762, <https://doi.org/10.1038/ncomms14762> (2017). <https://www.nature.com/articles/ncomms14762#supplementary-information>
- Ekins, S. *et al.* Illustrating and homology modeling the proteins of the Zika virus [version 1; referees: 2 approved with reservations]. Vol. 5 (2016).
- Johnston, J. M. & Filizola, M. Showcasing modern molecular dynamics simulations of membrane proteins through G protein-coupled receptors. *Current Opinion in Structural Biology* **21**, 552–558, <https://doi.org/10.1016/j.sbi.2011.06.008> (2011).
- Miao, Y., Nichols, S. E. & McCammon, J. A. Free energy landscape of G-protein coupled receptors, explored by accelerated molecular dynamics. *Physical Chemistry Chemical Physics* **16**, 6398–6406, <https://doi.org/10.1039/C3CP53962H> (2014).
- Markwick, P. R. L. & McCammon, J. A. Studying functional dynamics in bio-molecules using accelerated molecular dynamics. *Physical Chemistry Chemical Physics* **13**, 20053–20065, <https://doi.org/10.1039/C1CP22100K> (2011).
- Hamelberg, D., de Oliveira, C. A. F. & McCammon, J. A. Sampling of slow diffusive conformational transitions with accelerated molecular dynamics. *The Journal of Chemical Physics* **127**, 155102, <https://doi.org/10.1063/1.2789432> (2007).
- Pierce, L. C. T. & Salomon-Ferrer, R. Augusto F. de Oliveira, C., McCammon, J. A. & Walker, R. C. Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics. *Journal of Chemical Theory and Computation* **8**, 2997–3002, <https://doi.org/10.1021/ct300284c> (2012).
- Gasper, P. M., Fuglestad, B., Komives, E. A., Markwick, P. R. L. & McCammon, J. A. Allosteric networks in thrombin distinguish procoagulant vs. anticoagulant activities. *Proceedings of the National Academy of Sciences* **109**, 21216–21222, <https://doi.org/10.1073/pnas.1218414109> (2012).
- Wang, Y., Markwick, P. R. L., de Oliveira, C. A. F. & McCammon, J. A. Enhanced Lipid Diffusion and Mixing in Accelerated Molecular Dynamics. *Journal of Chemical Theory and Computation* **7**, 3199–3207, <https://doi.org/10.1021/ct200430c> (2011).
- Markwick, P. R. L., Pierce, L. C. T., Goodin, D. B. & McCammon, J. A. Adaptive Accelerated Molecular Dynamics (Ad-AMD) Revealing the Molecular Plasticity of P450cam. *The Journal of Physical Chemistry Letters* **2**, 158–164, <https://doi.org/10.1021/jz101462n> (2011).
- Shen, T. & Hamelberg, D. A statistical analysis of the precision of reweighting-based simulations. *The Journal of Chemical Physics* **129**, 034103, <https://doi.org/10.1063/1.2944250> (2008).
- Kappel, K., Miao, Y. & McCammon, J. A. Accelerated molecular dynamics simulations of ligand binding to a muscarinic G-protein-coupled receptor. *Quarterly Reviews of Biophysics* **48**, 479–487 (2015).
- Miao, Y., Feher, V. A. & McCammon, J. A. Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *Journal of Chemical Theory and Computation* **11**, 3584–3595, <https://doi.org/10.1021/acs.jctc.5b00436> (2015).
- Miao, Y., Feixas, F., Eun, C. & McCammon, J. A. Accelerated molecular dynamics simulations of protein folding. *Journal of Computational Chemistry* **36**, 1536–1549, <https://doi.org/10.1002/jcc.23964> (2015).
- Coutard, B. *et al.* Zika Virus Methyltransferase: Structure and Functions for Drug Design Perspectives. *Journal of Virology* **91**, <https://doi.org/10.1128/jvi.02202-16> (2017).

22. Miao, Y. *et al.* Improved Reweighting of Accelerated Molecular Dynamics Simulations for Free Energy Calculation. *Journal of Chemical Theory and Computation* **10**, 2677–2689, <https://doi.org/10.1021/ct500090q> (2014).
23. Al-Hasani, R. & Bruchas, M. R. Molecular Mechanisms of Opioid Receptor-Dependent Signaling and Behavior. *Anesthesiology* **115**, 1363–1381, <https://doi.org/10.1097/ALN.0b013e318238bba6> (2011).
24. Kobayashi, C. *et al.* GENESIS 1.1: A hybrid-parallel molecular dynamics simulator with enhanced sampling algorithms on multiple computational platforms. *Journal of Computational Chemistry* **38**, 2193–2206, <https://doi.org/10.1002/jcc.24874> (2017).
25. Grossfield, A. WHAM: an implementation of the weighted histogram analysis method. <http://membrane.urmc.rochester.edu/content/wham/> Version 2.09.1 (2017).
26. Frisch, M. J. *et al.* (Wallingford CT, 2009).
27. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **23**, 327–341, [https://doi.org/10.1016/0021-9991\(77\)90098-5](https://doi.org/10.1016/0021-9991(77)90098-5) (1977).
28. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N [center-dot] log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **98**, 10089–10092 (1993).

## Acknowledgements

This work was supported by the Kaohsiung Medical University “Aim for the Top 500 Universities Grant” (KMU-TP105C06 and KMU-TP105C14), Kaohsiung, Taiwan. This work also supported by the Republic of China and the National Science Council of the Republic of China, Taiwan (105–2113-M-037–009-).

## Author Contributions

Dr. Yeng-Tseng Wang initiated the research, performed the simulations, and wrote the manuscript. Dr. Chih-Hung Chuang supervised the study and refined the manuscript and the supplementary information. Dr. Shean-jaw Chiou gave a technical support for replying the reviewer’s comments and he also re-wrote the revised Supplementary Information. Dr. Tian-Lu Cheng gave a technical support for replying the reviewer’s comments.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-24775-4>.

**Competing Interests:** The authors declare no competing interests.

**Publisher’s note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018