



OPEN

Using Accelerated Molecular Dynamics Simulation to elucidate the effects of the T198F mutation on the molecular flexibility of the West Nile virus envelope protein

Renan Patrick da Penha Valente^{1,3}, Rafael Conceição de Souza¹, Gabriela de Medeiros Muniz¹, João Elias Vidueira Ferreira², Ricardo Morais de Miranda², Anderson Henrique Lima e Lima³ & João Lídio da Silva Gonçalves Vianez Junior¹✉

The envelope (E) protein is an important target for antibodies in flavivirus. Literature reports that the mutation T198F, located at the domain I-II hinge of the E protein, regulates viral breathing and increases the accessibility of a distal cryptic epitope located on the fusion loop, having a direct impact in the neutralization of West Nile virus (WNV). Our study aimed to describe, using accelerated molecular dynamics simulations, the effects of the T198F mutation in the flexibility of the E protein of WNV and to elucidate the mechanism that regulates epitope accessibility. The simulation results revealed that the mutation favors the formation of alternative hydrogen bonds, hampering the bending movement between domains I and II. We hypothesized that this is the mechanism by which the T198F mutation, located at the middle of the protein, locks the distal cryptic epitope near a single preferred conformation, rendering it more prone to recognition by antibodies.

West Nile virus (WNV) is a neurotropic human pathogen commonly found in the African region, as well as in Europe, Middle East, Americas and West Asia. The most frequent way of transmission to humans occurs by the bite of infected mosquitoes. Birds are considered its main reservoir, while humans and other mammals are considered incidental hosts^{1,2}. The virus was discovered in a febrile adult woman in the West Nile district of Uganda in 1937³. Since its discovery, many cases have been reported and today the virus represents a real threat to humans, equines and wildlife, mainly because no vaccine is available^{1,4,5}.

WNV is a member of the *flaviviridae* family, genus flavivirus. Like all flaviviruses, it has a single stranded positive sense RNA, encoding structural - Capsid (C), pre-membrane (prM) and Envelope (E) - and non-structural - NS1, NS2A, NS2B, NS3, NS4A, NS4B, NS5 - proteins^{6,7}. The E protein is responsible for virus entry^{8,9} and presents three structural domains (DI, DII and DIII)¹¹. There are 180 copies of the envelope protein arranged as antiparallel dimers which are distributed on the surface of mature virions, so it is a major target for neutralizing antibodies⁸.

Neutralization studies suggest that the T198F mutation alone can regulate WNV conformation dynamics (“viral breathing”), having a significant impact in the exposure of a cryptic epitope, modulating antibody recognition potency. This epitope is targeted by the monoclonal antibody E60 (and henceforth referred as epitope E60) and is located on the distal fusion loop (FL) of the envelope protein¹⁰.

Although Goo *et al.* hypothesized that the T198F mutation increased the accessibility of E60 due to changes in the conformational dynamics of the E protein, they could not infer the structural basis of this phenomena from their data¹⁰. The elucidation of the underlying mechanisms leading to alterations in molecular flexibility and antibody recognition can be achieved using molecular modeling methods, with techniques such as classical

¹Center for Technological Innovation, Evandro Chagas Institute, Ministry of Health, Ananindeua, PA, 67030-000, Brazil. ²Federal Institute of Education, Science and Technology of Pará, Tucuruí, PA, 68455695, Brazil. ³Laboratório de Planejamento e Desenvolvimento de Fármacos, Instituto de Ciências Exatas e Naturais, Universidade Federal do Pará, 66075-110, Belém, PA, Brazil. ✉e-mail: vianez.iec@gmail.com

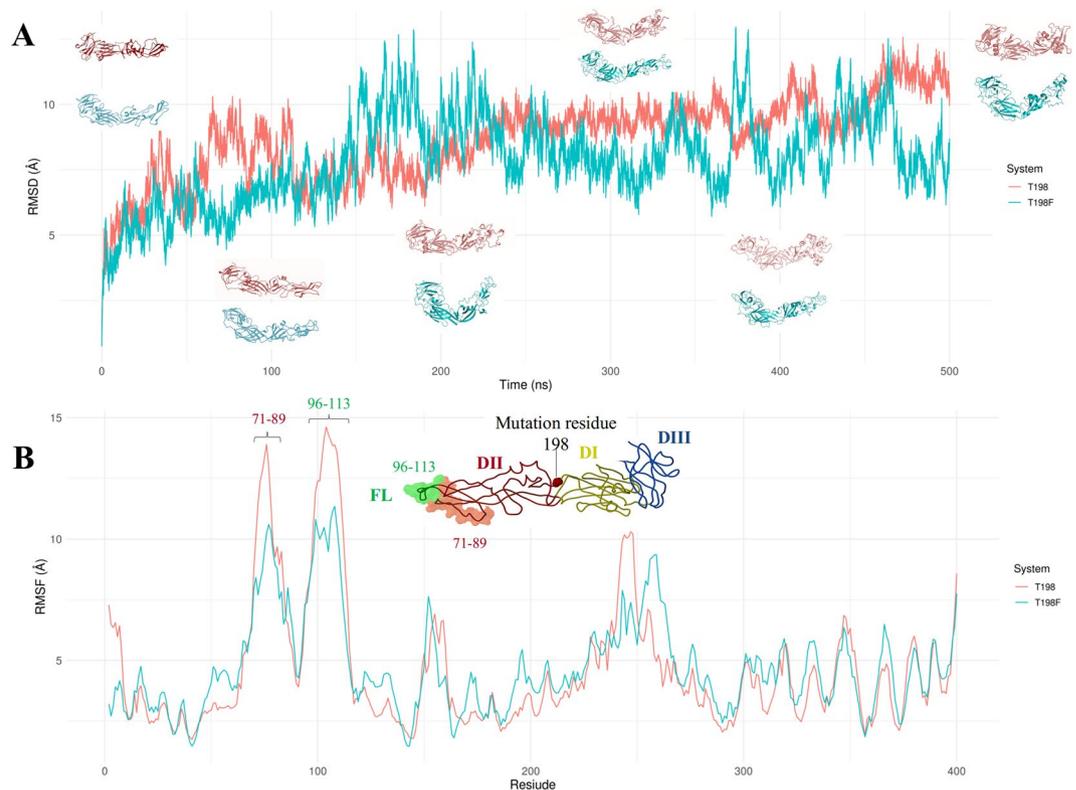


Figure 1. RMSD (A) and RMSF (B) during 500 ns of aMD simulation. In (A), changes in conformations are illustrated for different times (0, 100, 200, 300, 400 and 500 ns), showing that the T198F system presents lower RMSD values, suggesting limited changes in conformation. In (B), the residues with more fluctuations are displayed (71–89 and 96–113). It is also shown the illustration of the envelope protein, depicting DI (yellow), DII (red), DIII (blue) and the fusion loop (FL, green). The green shaded region comprises the residues 96–113 (FL) while the red shaded region comprises the residues 71–89.

Molecular Dynamics (cMD) and accelerated Molecular Dynamics (aMD) simulations^{11–13}. Regarding the conformations of protein macrostructures, aMD simulation scales have greater efficiency when compared to MD simulation, for example, an aMD simulation of 500 ns can be equivalent to about 1 million nanoseconds on the MD scales, depending on the system¹².

Our study aimed to investigate, using aMD simulations, the molecular flexibility of the envelope protein of WNV. Two systems, wild (T198) and mutant (T198F) were simulated in order to gain insight into the possible mechanisms that regulate the exposure of the cryptic epitope E60.

Results

Estimated parameters for aMD simulation. Classical molecular dynamic (cMD) simulation, one of the first steps necessary to simulate aMD, must give information about total potential energy (EPTOT) and dihedral angle energy (DIHED) since the average of these parameters is used as input in an aMD. In this respect, the variables EPTOT and DIHED remained stable during 10 ns of cMD simulation considering both wild and mutant systems of the envelope protein from WNV (See Fig. S1). The averages for EPTOT were -576560.0293 kcal/mol and -575954.3888 kcal/mol for T198 and T198F systems, respectively. The averages DIHED were 4863.0261 kcal/mol and 4902.0074 kcal/mol, respectively. Values for the calculated parameters used as input in aMD simulation and also the variables to calculate them can be found in Table S1.

Root Mean Square Deviation (RMSD) and Root Mean Square Fluctuation (RMSF) analysis indicated that the T198F mutation hampers the flexibility of the E protein, mainly in structural domain II. During 500 ns of aMD simulation, the RMSD analysis was used to measure conformational changes in the E protein, using as reference the structure obtained from the last frame of the 10 ns of cMD. On the other hand, RMSF was employed to quantify the fluctuation in each residue of both systems, allowing to identify the most flexible regions. The RMSD analysis shows that there are more structural deviations in the T198 system, suggesting reduced flexibility in the mutant system (Fig. 1). RMSF analysis allowed us to identify that the increased flexibility in the T198 system was mainly associated with two regions of the structural domain II (DII). One of the regions involves the residues 71–89, whose average RMSF values were about 15 Å for wild and 10 Å for the T198F system, respectively. The other region includes residues 96–113, whose average RMSF values were similar to the T198, with a small increase. These regions in DII include the fusion loop (FL) and E60.

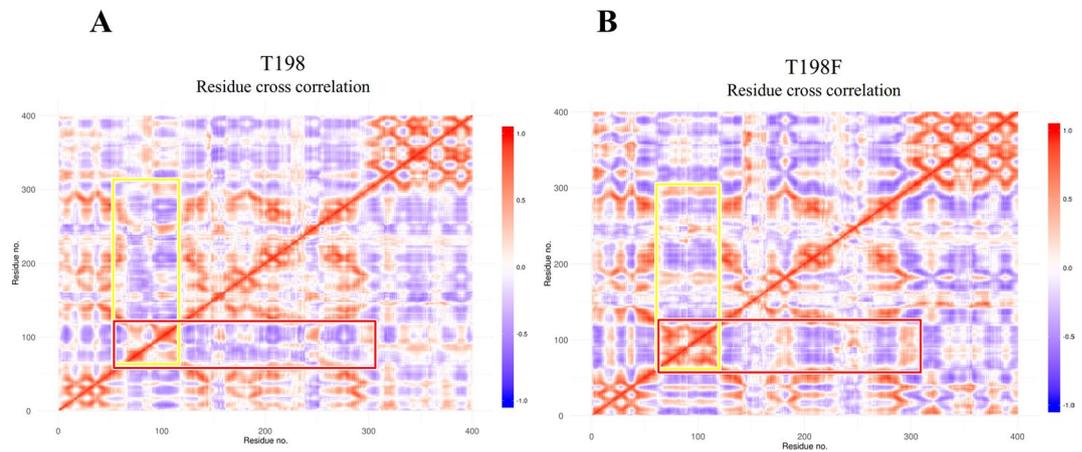


Figure 2. Cross-correlation map (prepared using R software¹⁴) for aMD simulation: (A) T198 system and (B) T198F system, region of colored rectangles include residues 71–89, 96–113 and 130–280.

Dynamical cross-correlation analysis shows that the T198F mutation decreases correlated movements in the E protein of WNV. A dynamic cross-correlation matrix (DCCM) was generated to identify correlated or anti-correlated movements¹² in both systems. From the resulting matrix, we could infer that residues 71–89 and 96–113 shows a considerable correlated movement with residues 130–280 (colored rectangles, Fig. 2A). In the mutated system, was observed a loss of those correlated movements in some regions, and significant increases of anti-correlated movements in the other system (Fig. 2B). The region 130–280 corresponds to residues that are part of DI–DII, while regions 93–113 and 71–89 corresponds to the fusion loop (FL) and to a loop adjacent to FL, respectively. In general, after mutation, the protein revealed an increase in correlated and anti-correlated movements, this should be related to exposure of the epitope.

PCA analysis indicated that the T198F mutation changes conformational minimum states. Based on principal components and free energy (Fig. S2, support information), histograms were projected to identify the preferable conformations of the two systems. Differences in the conformations observed in each basin depict a bending movement between DI and DII, which is particular of the E protein. This data corroborates the RMSF and cross-correlations analysis, suggesting a restriction in the flexibility of the T198F system, that quickly reaches a stable ensemble of conformations. Figure 3 presents minimal conformations for the systems T198 and T198F, the alignment between the average ensemble of conformations, clearly demonstrates that the mutation exposes FL, as well as the Epitope E60.

Hydrogen bond investigation explains the ensemble of preferential conformations. *In silico* investigation of hydrogen bonds were identified some key interactions to explain the minimum ensembles of conformations (Fig. 4). We considered as stable hydrogen bonding interactions those that were maintained for >50% of the total simulation time. The hydrogen bond interaction between the residues V356–T40 was the most effective in the mutant system, being established > 80% of the simulation. In general, the wild-type system presented more interactions than T198F: 2233 and 2077, respectively (difference = 156). Nevertheless, they were not so effective, since a part of them were not considered stable (>50% of the simulation time). The higher stability of alternative hydrogen bonds after mutation (Fig. 4) explains the loss of movements of the T198F system, leading to the observed changes in the conformational minimum states.

Discussion

Our simulations revealed that the envelope protein in its native state (T198 system) can explore the minimums of the conformational space with more difficulty in comparison to the mutated (T198F) system, which is less stable (greater molecular flexibility), agreeing with the hypothesis of altered flexibility supported by Goo *et al.*¹⁰. The changes in the flexibility induced by the T198F mutation were accentuated near the cryptic epitope E60, as was already suggested by them. Besides confirming what was suggested, the present computational simulation predicted that the increased exposure of E60 is due to the reduced flexibility in the T198F system. Even though the T198F mutation is located in the middle of the E protein and far from E60, it allowed the strengthening of particular hydrogen bonds (Fig. 4). This strengthening was able to restrict a bending movement that occurs between DI and DII. In this way, we propose that this is the mechanism by which the T198F mutation exposure the epitope E60 in a minimum ensemble of conformations, rendering this region more susceptible to antibody binding. To summarize, this study demonstrates how the alteration of a single residue in the hinge region of the E protein can influence the dynamics of a distal cryptic epitope, shedding light in the observed differences in neutralization efficiencies and have implications in future attempts of rational planning of vaccines and design antiviral compounds.

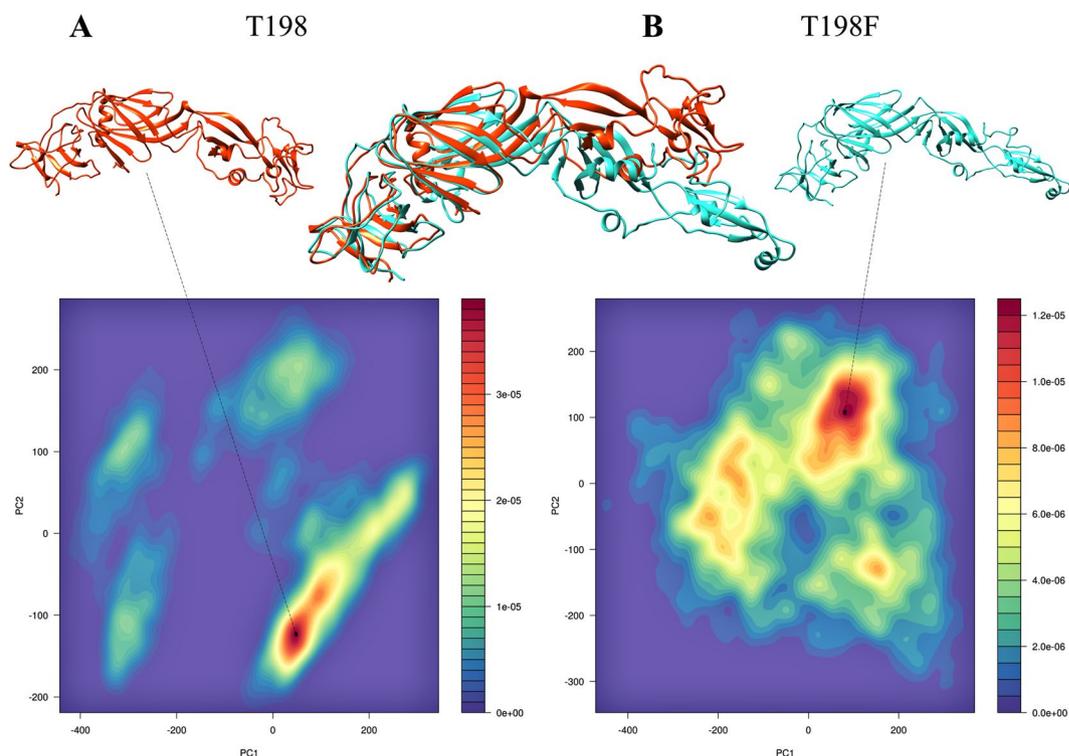


Figure 3. Histogram for T198 (A) and T198F (B) systems. In the middle is the alignment of the minimal ensemble conformations of the systems A and B. Protein images prepared using UCSF Chimera¹⁵ and PC1 vs PC2 using R software¹⁴.

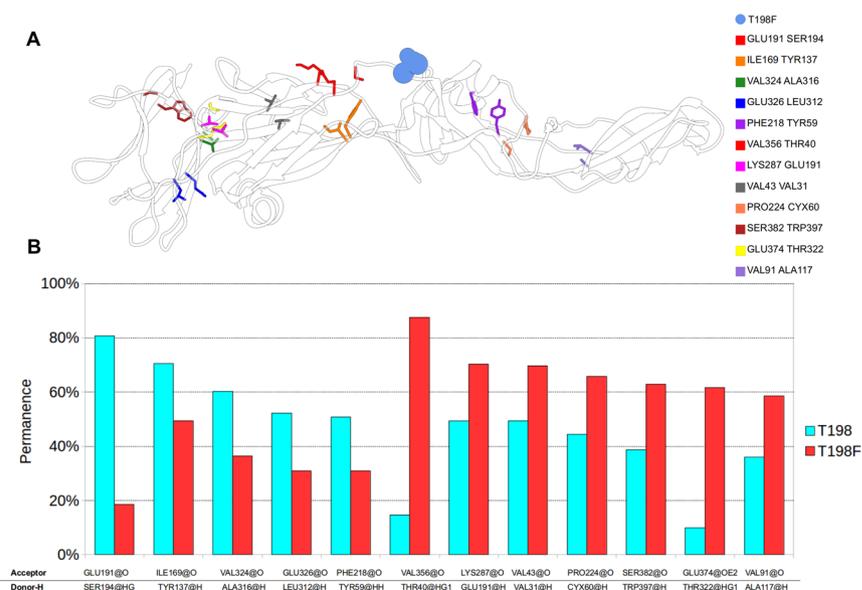


Figure 4. (A) Bonds that reduces the flexibility of the E protein of WNV in DI, DII and DIII (prepared in UCSF Chimera¹⁵), and (B) permanence in 500 ns of simulation.

Materials and Methods

System settings. The envelope protein of WNV used was retrieved from the protein data bank (PDB, www.rcsb.org)¹⁶ and identified by code 2HG0¹¹. Firstly, the amino acid protonation states were settled using H++ (<http://biophysics.cs.vt.edu>) webserver¹⁷. The force field employed to describe the protein was the ff99SB¹⁸ using tLeap module of AMBER 16¹⁹. The protein was solvated with TIP3P water molecules²⁰ and counter-ions were added in order to maintain the electroneutrality of the systems.

Classical and accelerated molecular dynamics simulations. Energy minimization, equilibration and system heating were carried out with AMBER 16 package as described previously by our group²¹. Firstly, a five-step minimization protocol was applied. During stages 1–3, all heavy atoms were restrained using a harmonic constant of 1000 kcal mol⁻¹ Å² while applying 5000 and 10000 steps of steepest-descent (SD) and conjugate-gradients (CG), respectively. In step 4, 30000 steps of SD and 35000 steps of CG were used, while maintaining only solute heavy atoms restrained. During step 5, 5000 steps of SD and 50000 steps of CG were applied with all atoms free to move. Afterward, a preliminary MD (2 ns) with all heavy atoms positions restrained (25 kcal mol⁻¹ Å) was performed to heat each system to 310 K in NVT employing Langevin dynamics as thermostat (collision frequency of 2 ps). SHAKE algorithm was employed to constraints all bonds involving hydrogen atoms and the Particle Mesh Ewald method²² was employed to calculate the electrostatic interactions of long distances with a cutoff radius of 10 Å.

The average dihedral and total potential energies were computed during a 10 ns classical molecular dynamics simulation and taken as reference for the aMD simulations. The aMD modifies the original dihedral potential ($V(r)$) by adding an increment, $\Delta V(r)$ when $V(r)$ is below a defined energy level E ^{19,23}, as in Eq. 1:

$$\Delta V(r) = \begin{cases} 0, & \frac{(E - V(r))^2}{\alpha + (E - V(r))} \cdot \frac{V(r) \geq E}{V(r) < E} \end{cases} \quad (1)$$

Where α modules the depth and roughness of the energy valleys in the modified potential.

The parameters E and α were the calculated average energy for the dihedral and total potential energy obtained at the end of 10 ns of simulation of classic MD^{19,23}.

Analysis of the simulations. To compute the root mean square deviation (RMSD) and the root mean square fluctuation (RMSF) for both trajectories, the CPPTRAJ module of AmberTools 16 was used, considering only alpha carbon atoms²⁴ according to equations described by Romanowska and coworkers in 2008 (Eq. 2 and Eq. 3):

$$\text{RMSD}(t_1, t_2) = \sqrt{\frac{1}{M} \sum_{i=1}^N m_i \|\vec{r}_i(t_1) - \vec{r}_i(t_2)\|^2} \quad (2)$$

$$\text{RMSF}(i) = \sqrt{\frac{1}{T} \sum_{t=1}^T \|\vec{r}_i(t) - \bar{\vec{r}}_i\|^2} \quad (3)$$

Principal component analysis (PCA) and free energy landscape (FEL). PCA is a well-known technique used to extract large-scale concerted movements in biomolecules. Briefly, a covariance matrix C_{ij} of the fluctuation of the alpha carbons is calculated and diagonalized, using the molecular dynamics trajectory as an input. The elements of the covariance matrix are:

$$C_{ij} = (x_i - x_j)(i, j = 1, 2, 3, \dots, 3N) \quad (4)$$

Where i and j denote all pairs of the 3N cartesian coordinates. x_i and x_j are instantaneous values of the i -th and j -th alpha carbon atom, respectively. N is the number of atoms considered and \bar{x}_i and \bar{x}_j represent the average value in all configurations obtained in the aMD²⁵. Dynamical cross-correlation matrices (DCCM), principal component and FEL analysis (PCA) were computed using the Bio3D package²⁶ in R software¹⁴.

Received: 17 October 2019; Accepted: 15 May 2020;

Published online: 15 June 2020

References

1. WHO. Chagas Disease (American Trypanosomiasis) [Internet] (2017).
2. Amraoui, F. *et al.* Culex pipiens, an Experimental Efficient Vector of West Nile and Rift Valley Fever Viruses in the Maghreb Region. *PLoS One*. **7**, e36757, <https://doi.org/10.1371/journal.pone.0036757> (2012).
3. Smithburn, K. C., Hughes, T. P., Burke, A. W. & Paul, J. H. A Neurotropic Virus Isolated from the Blood of a Native of Uganda 1. *Am J Trop Med Hyg* **s1-20**, 471–492, <https://doi.org/10.4269/ajtmh.1940.s1-20.471> (1940).
4. Heinz, F. X. & Stiasny, K. Flaviviruses and flavivirus vaccines. *Vaccine*. **30**, 4301–4306, <https://doi.org/10.1016/j.vaccine.2011.09.114> (2012).
5. Chang, G.-J. J., Kuno, G., Purdy, D. E. & Davis, B. S. Recent advancement in flavivirus vaccine development. *Expert Rev Vaccines* **3**, 199–220, <https://doi.org/10.1586/14760584.3.2.199> (2004).
6. Bollati, M. *et al.* Structure and functionality in flavivirus NS-proteins: perspectives for drug design. *Antiviral Res.* **2009/11/27** **87**, 125–148, <https://doi.org/10.1016/j.antiviral.2009.11.009> (2010).
7. Blazevic, J., Rouha, H., Bradt, V., Heinz, F. X. & Stiasny, K. Membrane Anchors of the Structural Flavivirus Proteins and Their Role in Virus Assembly. *J Virol*. **90**, 6365–6378, <https://doi.org/10.1128/jvi.00447-16> (2016).
8. Minarno, E. B. & Hariani, L. Gizi dan kesehatan: Perspektif Al-Qur'an dan sains. [https://doi.org/10.1016/0038-092X\(88\)90131-4](https://doi.org/10.1016/0038-092X(88)90131-4) (2008).
9. Zhang, Y. *et al.* Conformational changes of the flavivirus E glycoprotein. *Structure*. **12**, 1607–1618, <https://doi.org/10.1016/j.str.2004.06.019> (2004).
10. Goo, L., VanBlargan, L. A., Dowd, K. A., Diamond, M. S. & Pierson, T. C. A single mutation in the envelope protein modulates flavivirus antigenicity, stability, and pathogenesis. *PLoS Pathog*. **13**, e1006178 (2017).
11. Nybakken, G. E., Nelson, C. A., Chen, B. R., Diamond, M. S. & Fremont, D. H. Crystal structure of the West Nile virus envelope glycoprotein. *J Virol*. **2006/09/20** **80**, 11467–11474, <https://doi.org/10.1128/JVI.01125-06> (2006).

12. Romelia Salomon LP & RW. Accelerated Molecular Dynamics in AMBER Example Analysis of all-atom enhanced sampling method accelerated molecular dynamics (aMD) to investigate conformational changes in proteins that typically occur on the millisecond time scale. In: AMBER Advanced Tutorial 22 (2013).
13. Costa, C. H. S. *et al.* Computational study of conformational changes in human 3-hydroxy-3-methylglutaryl coenzyme reductase induced by substrate binding. *J Biomol Struct Dyn* **37**, 4374–4383, <https://doi.org/10.1080/07391102.2018.1549508> (2019).
14. R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria. doi:R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL, <http://www.R-project.org> (2017).
15. Pettersen, E. F. *et al.* UCSF Chimera - A visualization system for exploratory research and analysis. *J Comput Chem*; <https://doi.org/10.1002/jcc.20084> (2004).
16. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res* **28**, 235–242, <https://doi.org/10.1093/nar/28.1.235> (2000).
17. Anandakrishnan, R., Aguilar, B. & Onufriev, A. V. H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res* **40**, W537–W541, <https://doi.org/10.1093/nar/gks375> (2012).
18. Hornak, V. *et al.* Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **65**, 712–725, <https://doi.org/10.1002/prot.21123> (2006).
19. Case, D. A. *et al.* AMBER 2016 Reference Manual. *Univ Calif San Fr CA, USA*. 1–923 (2016).
20. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* **79**, 926–935, <https://doi.org/10.1063/1.445869> (1983).
21. Siqueira, A. S. *et al.* Comparative modeling and molecular dynamics suggest high carboxylase activity of the Cyanobium sp. CACIAM14 RbcL protein. *J Mol Model*. **22**. <https://doi.org/10.1007/s00894-016-2943-y> (2016).
22. Nishimura, Y. & Nakai, H. Parallel implementation of efficient charge–charge interaction evaluation scheme in periodic divide-and-conquer density-functional tight-binding calculations. *J Comput Chem* **39**, 105–116, <https://doi.org/10.1002/jcc.25086> (2018).
23. Gedeon, P. C., Thomas, J. R. & Madura, J. D. Accelerated molecular dynamics and protein conformational change: a theoretical and practical guide using a membrane embedded model neurotransmitter transporter. *Methods Mol Biol* **1215**, 253–287, https://doi.org/10.1007/978-1-4939-1465-4_12 (2015).
24. Roe, D. R. & Cheatham, T. E. PTRAJ and CPPTRAC: Software for processing and analysis of molecular dynamics trajectory data. *J Chem Theory Comput* **9**, 3084–3095, <https://doi.org/10.1021/ct400341p> (2013).
25. Yang, Y., Liu, H. & Yao, X. Understanding the molecular basis of MK2–p38 α signaling complex assembly: insights into protein–protein interaction by molecular dynamics and free energy studies. *Mol Biosyst.* **8**, 2106–2118, <https://doi.org/10.1039/C2MB25042J> (2012).
26. Grant, B. J., Rodrigues, A. P. C., ElSawy, K. M., McCammon, J. A. & Caves, L. S. D. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics*. **22**, 2695–2696, <https://doi.org/10.1093/bioinformatics/btl461> (2006).

Acknowledgements

Evandro Chagas Institute, Federal University of Pará, Federal Institute of Education, Science and Technology of Pará, FAPESPA and CNPq for cooperation and support to this study. This work was supported by Evandro Chagas Institute, Ministry of Health, Brazil. RPPV had a scholarship by Fundação Amazônia Paraense de Amparo à Pesquisa (FAPESPA) through the Institutional Scientific Initiation Scholarship Program (PIBIC) of the Evandro Chagas Institute. RCS and GBM had a scholarship by National Counsel of Technological and Scientific Development (CNPq). The funding bodies had no role in the design of the study, in collection, analysis, and interpretation of data, or in writing the manuscript.

Author contributions

Renan Patrick da Penha Valente conducted research with technical and intellectual support of Rafael Conceição de Souza and Gabriela de Medeiros Muniz running and programming scripts. João Elias Vidueira Ferreira, Ricardo Morais de Miranda, and Anderson Henrique Lima e Lima were professors who assisted in the interpretation of data and discussion. This research was guided by João Lídio da Silva Gonçalves Vianez Junior, participating in all the processes. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-66344-8>.

Correspondence and requests for materials should be addressed to J.L.d.S.G.V.J.

Reprints and permissions information is available at www.nature.com/reprints.

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) 2020