COMBINING PROTEIN INTERACTION AND FUNCTIONALITY CLASSIFICATION IN NS3 TO DETERMINE SPECIFIC ANTIVIRAL TARGETS IN DENGUE

by

Lamya Alomair
A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Bioinformatics and Computational Biology

Committee:

_____	Dr. M. Saleet Jafri, Dissertation Director
_____	Dr. Aarthi Narayanan, Committee Member
_____	Dr. Iosif Vaisman, Committee Member
_____	Dr. Iosif Vaisman, Acting Director, School of Systems Biology
_____	Dr. Donna M. Fox, Associate Dean, Office of Student Affairs & Special Programs, College of Science
_____	Dr. Peggy Agouris, Dean, College of Science
Date: _____	Spring Semester 2017 George Mason University Fairfax, VA

Combining Protein Interactions and Functionality Classification in NS3 to Determine Specific Antiviral Targets in Dengue

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

By

Lamya Alomair
Master of health informatics
King Saud bin Abdulaziz University for Health Sciences, 2008

Director: M. Saleet Jafri, Professor
School of Systems Biology, College of Science

Spring Semester 2017
George Mason University
Fairfax, VA

Dedication

This dissertation is dedicated to the soul of my dad Abdulaziz Alomair and to my mother Noura Alomair for dedicating themselves to raise and educate me, to my husband Saleh Alammari for his love and support, to my lovely sons Yazan & Abdulaziz, and my lovely daughters Shadan & Lamar Alammari.

Acknowledgements

During my journey of seven years as a PhD student I crossed roads with many people who enlightened my path and to whom I am greatly indebted. Firstly, I am grateful to my advisor Dr. M. Saleet Jafri whom I saw as both the respected dissertation director and the dear brother. It is because of his unlimited support and valuable guidance that I made it to complete this research. I would like to thank you for encouraging my research and for allowing me to grow as a researcher. Your advice on both research as well as on my career have been invaluable.

I would like to extend my sincere thanks to my dissertation committee members Dr. Aarthi Narayanan and Dr. Iosif Vaisman for their guidance and support throughout my research. I am thankful to Dr. Saleet's lab members Aman Ullah, Sangeeta Shukla, Nasrin Afzal, Roshan Paudel and Sarita Limbu who never hesitated to provide help and feedback. Many thanks to Imed Ben Chouikha for always being there for me with lots of support. I also thank Diane St. Germain, Andrea Nikoi and Melissa Hayes for providing me with the administrative help and information whenever needed.

I would like to acknowledge honorary with especially thank to my mentor Dr.Tariq Alsheddi, He is my best role model for a scientist and life .

My thanks to my father Abdulaziz Alomair whose soul kept shining down on me with encouragement, to my mother Noura Alomair who never stopped praying for me, to my brothers Khalid and Waleed, to my sisters Khouloud, Amal, Taghreed, Nada, and Saba. My utmost thanks to my husband Saleh for his unconditional support during my trip to the United States and to my lovely kids.

Finally, my thanks with love to my dear friends and my extended family who filled my journey with their continuous encouragement.

Table of Contents

List of Tables

List of Figures

Abstract

COMBINING PROTEIN INTERACTIONS AND FUNCTIONALITY CLASSIFICATION IN NS3 TO DETERMINE SPECIFIC ANTIVIRAL COMPOUNDS FOR DENGUE VIRUS.

Lamya Alomair, Ph.D.

George Mason University, 2017

Dissertation Director: Dr. M. Saleet Jafri

Dengue virus (DENV) is a serious worldwide health concern putting about 2.5 billion people in more than 100 countries at-risk Dengue is a member of the flaviviridae family, is transmitted to human via mosquitos. Dengue is a deadly viral disease. Unfortunately, there are no vaccines or antiviral that can prevent this infection and that is why researchers are diligently working to find cures. The DENV genome codes for multiple nonstructural proteins one of which is the NS3 enzyme that participates in different steps of the viral life cycle including viral replication, viral RNA genome synthesis and host immune mechanism. Recent studies suggest the role of fatty acid biogenesis during DENV infection, including posttranslational protein modification. Phosphorylation is among the protein post translational modifications and plays essential roles in protein folding, interactions, signal transduction, survival and apoptosis.

In silico study provides a powerful approach to gain a better understanding of the biological systems at the gene level. NS3 has the potential to be phosphorylated by any of the ~500 human kinases. We predicted potential kinases that might phosphorylate NS3 and calculated Dena ranking score using neural network and other machine learning based webserver programs. These scores enabled us to investigate and identify the top kinases that phosphorylate DENV NS3. We hypothesize that preventing the phosphorylation of NS3 may interrupt the viral replication and participate in antiviral evasion. Using multiple sequence alignment bioinformatics tools we verified the results of the highly conserved residues and the residues around active sites whose phosphorylation may have a potential effect on viral replication. We further verified the results with multiple bioinformatics tools. Moreover, we included the Zika virus in our research and analysis taking into consideration the facts that Zika is related to the dengue virus because it belongs to the same Flavivirus genus affecting humans which might lead to a lot of similarities between Zika and Dengue, and that Zika is available for *in vitro* testing.

Our studies propose that the Host-Mediated Phosphorylation of NS3 would affect its capability to interact with NS5 and knocking out one of the interacting proteins may inhibit viral replication. These results will open new doors for further investigation and future work is expected to help identify the key inhibition mechanisms.

Chapter 1: Introduction

Abstract

Dengue fever, also known as break bone fever, is an illness caused by the Dengue virus. It is transmitted to humans by a female Aedes mosquito's bite. There are four serotypes of Dengue virus, namely DENV-1, DENV-2, DENV-3, and DENV-4. The symptoms of Dengue fever range from a mild fever to severe conditions of dengue hemorrhagic fever that may be life threatening in some cases.

Dengue virus is mainly concentrated in tropical and subtropical regions. The name Dengue is derived from the Swahili sentence, *ka dinga pepo*, describing a cramp-like seizure caused by an evil spirit. Nonetheless, increased air travel, and unplanned urbanization, have led to an increase in the rate of infection globally. The disease rate has increased in recent eras, which has led to increased outbreaks of Dengue fever worldwide, putting nearly a third of the human population at high risk of infection. Moreover, the development of a successful vaccine has remained elusive with challenges implementing it in the field.

Dengue virus is a small virus that carries a single strand of RNA as its genome. The genome encodes only ten proteins. However, there are still plenty of gaps in understanding the Dengue virus (DENV) pathogenesis, mainly due to the lack of a suitable animal model.

Currently, the only effective way to prevent dengue virus transmission is to combat the virus carrying mosquitoes.

It is essential that we learn more about the pathogenesis, and mechanism, of transmission of the disease, in order to develop a vaccine and cure for the Dengue virus infection.

Historical Background

The Dengue virus (DENV) is a mosquito-borne infection that causes a severe illness and, sometimes, a potential deadly complication called Dengue hemorrhagic fever (DHF). Dengue virus is a member of the Flaviviradae family that also includes the West Nile virus (WNV) and yellow fever virus, amongst others. Dengue virus is transmitted to humans via two types of mosquitoes: *Aedes albopictus* and *Aedes aegypti*. There are four serotypes of Dengue virus: DENV-1, DENV-2, DENV-3, and DENV-4 (1).

The prevalence of these serotypes has grown dramatically around the world in recent decades. This global growth of Dengue virus is a serious health threat since there are neither specific treatments to the Dengue infection, nor an effective vaccine to prevent it. It is estimated that 2.5 billion people – over 40% of the world's population – are currently at risk of being infected. The World Health Organization (WHO) projects that 50 to 100 million dengue infections will occur worldwide every year (2).

The number of affected countries increased from 9 countries in 1970 to over 100 countries in 2010. Some researchers think that the socioeconomic impact of World War II, and dependency on ocean going vessels, planted the seed for the current dengue pandemic

(3). The origins of the word dengue are not clear from a historical point of view, but one theory posits that the origin is derived from the Swahili (tribes living on the coast or coastal area of East Africa) phrase, “Ka-dinga pepo”, meaning “cramp-like seizure caused by an evil spirit” (4). The first case reported in Europe was in France before the virus was discovered in other European countries (5).

The map below shows the spread of the Dengue virus during the period from January to March 2017 (Figure 1). Dengue has been endemic along the regions near the equator including Asia, Africa, Central America, and South America. It is important to note that the United States is an immediate neighbor to some of these countries and is not free from the threat of Dengue virus and, as such, greater efforts must be spent to understand all aspects related to the dengue infection in the hope of identifying a good treatment or putting in place a prevention system (3).

Figure 1: Dengue transmission risk in the past three months. The map shows the resent reported dengue cases in the world. There were 813 Alerts reported in the period January-March 2017 (44). The map shows that recent cases are recorded mainly in Latin America and Asia.

Dengue research falls into two main categories: 1) Some research follows the spread of the virus, and the factors leading to its transmission, such as weather and the presence of favorable environment for carrier mosquito. 2) Most of the research on dengue virus focuses on understanding how dengue virus infects the cell, what kind of cells it targets, and determining the mechanism of the viral replication and the production of dengue virus inside the cell (8, 20, 21, 22, 26).

Molecular Biology of Dengue:

Dengue virus has a single-stranded RNA genome with positive polarity, a 5′ m7GpppG cap, and a conserved 3′-terminal stem loop (SL) that is linked to proposed

functions in viral RNA transcription and translation (50). This genome is located inside the virus and is enclosed by the nucleocapsid, which, in turn, is enclosed within a viral envelope. The viral envelope has E&M proteins protruding from it. Dengue virus encodes a single polyprotein precursor consisting of 3391 amino acid residues, It has three virus-coded proteins, designated capsid (C) protein, membrane (M) protein, and envelope (E) glycoprotein, And 7 nonstructural proteins Table (1) (NS1, NS2A, NS2B, NS3, NS4A, NS4B, NS5) (51).

Table 1: Dengue non-structural proteins. Description of known functions of Dengue non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, NS5) and their positions in Dengue genome, length and function.

	Position	**Length**	**Function**
NS1	776-1127	352	Plays a role in viral RNA replication complex.
NS2A	1128-1345	218	Forms part of RNA replication complex
NS2B	1346-1475	130	Co-factor for NS3 protease
NS3	1476-2093	618	Serine protease,RNA helicase and RTPase/NTPase
NS4A	2094-2220	127	Possibly induces membrane alterations .
NS4B	2244-2491	248	Possibly blocks IFN-α or IFN-β induced signal transduction
NS5	2492-3391	900	Methyltransfease ,RNA-Dependent RNA polymerase

Dynamics between Dengue Virus and the Human Cells

Dengue Pathogenesis

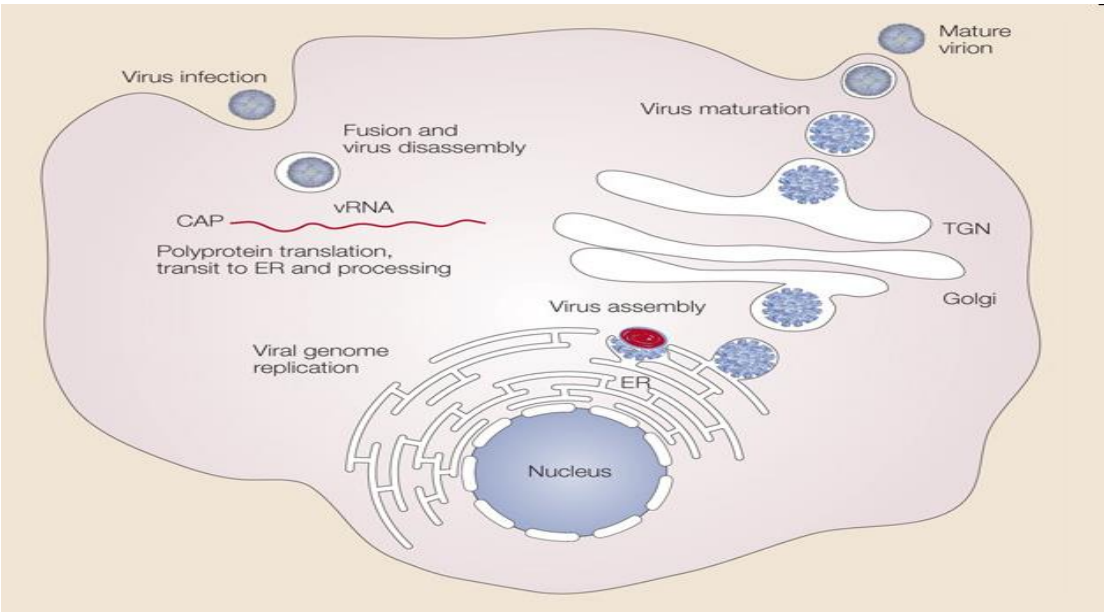

Figure 2: Dengue Pathogenesis. The replication cycle of dengue virus in the human cell starts by the viral infection after which the viral RNA (vRNA) reaches the nucleus for vial genome replication. After its assembly the virus goes through a maturation step then leaves the cell to proceed with other cell infections. Adapted from " A structural perspective of the flavivirus life cycle," by S. Mukhopadhyay and R. J. Kuhn, 2005, *Nature Reviews Microbiology* **3**, 13–22. doi:10.1038/nrmicro106. Copyright Macmillan Publishers Limited 2017.

The pathogenesis of dengue or the Dengue virus lifecycle begins when the dengue virus enters the surface of a host cell by receptor mediate; this process is called endocytosis. The reduced-pH of the endosome, cause the viral envelop of glycoprotein to undergo conformational rearrangement exposing the fusion loop and allowing individual glycoproteins to move freely in the plane of the viral membrane.

But before being released into the cytoplasm, the glycoproteins form trimers and they were exposed fusion loop insert into the host cell membrane the envelope glycoprotein then falls back on itself directing its carboxyl-terminus which is anchored in the viral membrane towards the fusion loops and a cellular membrane this causes the two membranes bend toward each other until the outer bilayer leaflet fuse.

The viral RNA (vRNA) is translated into a single polypeptide that is cut into ten proteins, and the viral genome is replicated. Virus assembly occurs on the surface of the endoplasmic reticulum (ER) when the structural proteins and newly synthesized RNA bud into the ER. The immature viral particles are transported through the trans-Golgi network (TGN), where they mature and convert into their infectious form. The mature viruses are then released from the cell and can go on to infect other cells Figure (2) (45).

Characteristics of the Dengue Virus Structure

Dengue virus genome encodes 10 viral proteins including 3 structural proteins: the capsid (C) protein, the envelope (E) glycoprotein, and the membrane (M) protein, responsible for viral structure and viral attachment to the host cell, and 7 non-structural proteins: NS1, NS2A, NS2B, NS3, NS4A, NS4B, NS5 (see Table 1) that are involved in viral replication as well as other cellular functions (22, 27) (Figure 3).

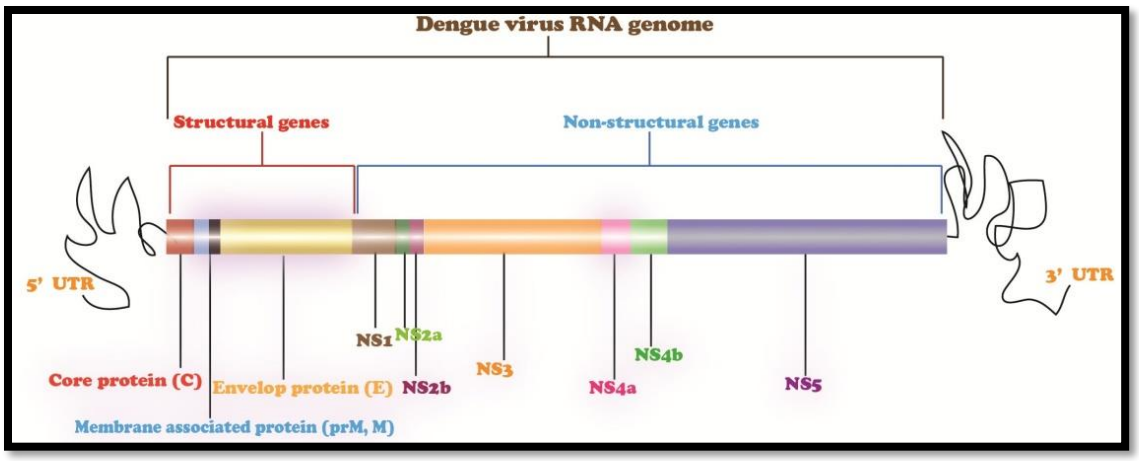

Figure 3: Dengue virus genome. The dengue virus genome encodes three structural (capsid [C], membrane [M], and envelope [E]) and seven nonstructural (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5) proteins. Adapted from " A structural perspective of the flavivirus life cycle," by S. Mukhopadhyay and R. J. Kuhn, 2005, *Nature Reviews Microbiology* **3**, 13–22. doi:10.1038/nrmicro106. Copyright Macmillan Publishers Limited 2017.

The dengue evolution, and the changes in the dengue viral genome over time, were the target of recent researches, in an attempt to investigate the genomic variations that might have occurred and helped the virus elude the immune system. The efforts for decade of deep working from scientist t and research tend to understand the mechanism the virus uses against the immune system's detection, and its reaction to viral infection, are targeted to help develop new treatments to cure the dengue viral infection (28, 29).

It has been shown that some particular viral sequences are associated with more severe dengue symptoms. In addition, certain dengue virus sequence variations may produce more deadly viruses with a greater potential for epidemic spread (10, 30, 31).

Moreover, scientists want to know how the dengue virus causes a breakdown of the human body that includes bleeding and vascular leakage that occurs with severe dengue infections (33). Diets rich in calcium have been shown to help the body decrease the length of the dengue illness (32).

Dynamics between Dengue Virus and the Human Cells

In addition to what was mentioned earlier about the Dengue virus' techniques to evade the immune system, the immune system itself facilitates infection while reacting to the infection during viral production. A study of the immune responses to Dengue virus showed that the inflammatory pathways and the multiple immune mechanisms that are triggered in the host cells to control the viral production are a combination of, 1- pathways that are critical for protection from infection, and 2- some mediators produced during the response phase that increase the severity of the disease (28).

When viruses like Dengue infect a host cell, viral proteins, such as NS3 and NS5, enter the cell and migrate to the endoplasmic reticulum (ER) membrane, which is the site of protein synthesis in the host cell. There, they utilize ER chaperones, such as calreticulin and calnexin, to help make viral proteins that are exported in lipid membrane droplets to form new viral particles that exit the cell. Knocking out calreticulin, therefore, interferes with viral replication (34, 35, 36).

Furthermore, other research look for the Fatty acid synthase (FASN) inhibition reduces Dengue viral replication. This is because inhibition of FASN induces the ER stress response (21). The mistargeting of FASN might be a result of the ER chaperonins being

busy making viral proteins (36).

When a cell gets infected, it invokes the ER stress mechanism, known as Unfolded Protein Response (UPR), which leads to apoptosis (22) And Apoptosis is important in the immune system, and plays significant roles in the control of the immune response (46). The virus counteracts it by activating other pathways in the infected cell. For example, it might inhibit CamKII, whose activation increases mitochondrial uptake of calcium, which in turn leads to apoptosis of the cell (37).

An increase in calcium in the diet decreases the length of dengue illness. Moreover, low calcium seems to occur with Dengue infection and the use of calcium supplements, anecdotally, seems to lessen the severity of the infection (40). It is possible that the Dengue virus tries to keep calcium low to prevent apoptosis, while increasing cellular calcium leads to apoptosis before the cell can make more virus particles. There might be some interaction of Dengue with calcium channels, such as the P/Q-type channels, that limit the cellular calcium levels and might also contribute to this.

The NS3 Protein in Dengue Virus

The NS3 protein of Dengue virus is known for multiple enzymatic activities. It is the second largest nonstructural protein of the virus. In its N-terminal region it has a serine proteinase, and an NTPase-helicase in its remaining 70% from the C-terminus end (23, 24, 27).

As mentioned previously, the DENV replication requires the expression of FASN. It is believed that NS3 increases the cellular fatty acid synthesis, and is also responsible for

redistribution of fatty acid synthase to specific sites of viral replication (21). The specific mechanism is not yet completely understood. It could be that NS3 directly binds and modifies FASN activity or, alternatively, NS3 interacts indirectly with FASN through another intermediate cellular protein.

One of the key components in DENV replication is NS5 of filoviruses. NS5 has been studied extensively, and its phosphoprotein has been shown to be critical; consequently, other multifunctional proteins might be regulated similarly (48). Recent studies suggested that the phosphorylation state of NS5 controls the association/dissociation of NS3 with NS5 and that this controls viral replication. NS5 phosphorylation is necessary for its binding to NS3 and this affects its localization (47). Blocking phosphorylation of NS5 prevents the dissociation of NS3 and NS5 and thus prevents the NS5 from entering the nucleus and keeps it mainly localized in the cytoplasm outside the nucleus.

Computational Studies

Because of the lack of useful animal model systems and the complexity of the dynamics involved between the virus and the host cells, there are several computational approaches used in the field of bioinformatics to understand the Dengue virus and the behavior of its major proteins. The conserved genes among the four serotypes of Dengue virus are believed to be the best targets for viral inhibition, thus more focus has been given on the NS3 gene. A phylogenetic trees method of study has been also used to study other

potential targets to determine specific regions of the NS3 sequence as potential regions of antiviral agents (43).

Methods, such as molecular docking, have been used to design multiple cyclic peptide inhibitors of Dengue virus NS3-NS2 Protease. Weak interactions were observed that suggested some of the designed inhibitors might be a potential inhibitor to the NS3 protease activities (39).

There were other attempts to predict the protein-protein interactions between Dengue virus proteins and their hosts in humans and in insects, computationally. Research results were limited to potential hypothesis for further experimental investigation (41). Some results, however, identified some new interactions in Dengue, such as zinc finger protein, calcium binding protein and other human proteins (42), that will need to be followed and confirmed.

Additional interactions such as phosphorylation of DENV NS5 were studied in silico and it had been predicted that the kinase PKC has a restriction ability that modulates the viral replication in host cell and prevents its outburst. On the other hand, the viral replication increases when the DENV NS5 phosphorylation by PKC gets inhibited (49).

Research Objectives

As previously mentioned, it is very important that we better understand all aspects related to the dengue viral production, replication, and the mechanisms the virus follows to evade and counteract the immune system. Our aim is to identify a collection of proteins whose interactions are similar to the NS3 and, based on that, identify potential targets to

inhibit the viral activities. The ultimate goal is to constitute the basis for determining specific compounds that can block major interactions needed by the dengue for viral replication. We propose a multifaceted computational approach to understand the viral activities by studying the protein interactions and by identifying several major protein fragments significantly similar to NS3. Then use this to determine their protein-protein interactions, clarify the functionality of the associated parts of dengue, and project the findings of protein interactions to the Dengue virus. The topics investigated in this dissertation are partitioned as follows:

- Chapter 1 presents a literature review on the virus and the related research. It includes a brief historical background on the Dengue virus and its spread, the dynamics of the virus interaction with the human cell including characteristics of its non-structural protein NS3, and a literature review of results and limitations pertaining to the computational research on Dengue virus.
- In chapter 2 we applied homology-base method to predict protein-protein interactions between NS3 and human proteins. Particularly, we looked at major protein interaction of similar proteins to NS3 and identified the functional similarities.
- In chapter 3 we selected top reliable prediction tools from the literature and we used them to predict and classify the top kinases that can potentially phosphorylate NS3. We identified a cohort of 8 major kinases in 19 different targeted residues and checked their individual functionalities.

- Chapter 4 was about an in silico molecular dynamics simulation of the phosphorylation of Dengue NS3 and the resulting conformational changes to investigate possible significant changes in the structure.
- Chapter 5 presents a conclusion and future work.

References

1. Gubler DJ. The global emergence/resurgence of arboviral diseases as public health problems. Archives of medical research. 2002 Aug 31;33(4):330-42.

2. Gubler DJ. TMH Dengue, Urbanization and Globalization: The Unholy Trinity of the 21 st Century. Tropical medicine and health. 2011;39(4 Supplement):3-11.

3. Gubler DJ. Dengue and dengue hemorrhagic fever. Clinical microbiology reviews. 1998 Jul 1;11(3):480-96.

4. Rigau-Pérez JG. The early use of break-bone fever (Quebranta huesos, 1771) and dengue (1801) in Spanish. The American journal of tropical medicine and hygiene. 1998 Aug 1;59(2):272-4.

5. La Ruche G, Souares Y, Armengaud A, Peloux-Petiot F, Delaunay P, Desprès P, Lenglet A, Jourdain F, Leparc-Goffart I, Charlet F, Ollier L. First two autochthonous dengue virus infections in metropolitan France, September 2010. Euro surveill. 2010 Sep 30;15(39):19676.

6. Wu PC, Guo HR, Lung SC, Lin CY, Su HJ. Weather as an effective predictor for occurrence of dengue fever in Taiwan. Acta tropica. 2007 Jul 31;103(1):50-7.

7. Little JL, Wheeler FB, Fels DR, Koumenis C, Kridel SJ. Inhibition of fatty acid synthase induces endoplasmic reticulum stress in tumor cells. Cancer research. 2007 Feb 1;67(3):1262-9.

8. World Health Organization (WHO). Dengue and severe dengue. [Internet]. 2017 [cited 2017 Jan 17]. Available from: http://www.who.int/mediacentre/factsheets/fs117/en/

9. Skrabanek L, Saini HK, Bader GD, Enright AJ. Computational prediction of protein–protein interactions. Molecular biotechnology. 2008 Jan 1;38(1):1-7.

10. Chen C, McGarvey PB, Huang H, Wu CH. Protein bioinformatics infrastructure for the integration and analysis of multiple high-throughput "omics" data. Advances in bioinformatics. 2010 Mar 29;2010.

11. Xenarios I, Salwinski L, Duan XJ, Higney P, Kim SM, Eisenberg D. DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. Nucleic acids research. 2002 Jan 1;30(1):303-5.

12. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009 May 1;25(9):1189-91.

13. Chitale M, Khan IK, Kihara D. In-depth performance evaluation of PFP and ESG sequence-based function prediction methods in CAFA 2011 experiment. BMC bioinformatics. 2013 Feb 28;14(3):S2.

14. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA. Gene Ontology: tool for the unification of biology. Nature genetics. 2000 May 1;25(1):25-9.

15. Wass MN, Barton G, Sternberg MJ. CombFunc: predicting protein function using heterogeneous data sources. Nucleic acids research. 2012 Jul 1;40(W1):W466-70.

16. Laskowski RA, Watson JD, Thornton JM. ProFunc: a server for predicting protein function from 3D structure. Nucleic acids research. 2005 Jul 1;33(suppl 2):W89-93.

17. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. Nature protocols. 2010 Apr 1;5(4):725-38.

18. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. Nucleic acids research. 2000 Jan 1;28(1):235-42.

19. Pickett BE, Greer DS, Zhang Y, Stewart L, Zhou L, Sun G, Gu Z, Kumar S, Zaremba S, Larsen CN, Jen W. Virus pathogen database and analysis resource (ViPR): a comprehensive bioinformatics database and analysis resource for the coronavirus research community. Viruses. 2012 Nov 19;4(11):3209-26.

20. Bente DA, Rico-Hesse R. Models of dengue virus infection. Drug Discovery Today: Disease Models. 2006 May 31;3(1):97-103.

21. Heaton NS, Perera R, Berger KL, Khadka S, LaCount DJ, Kuhn RJ, Randall G. Dengue virus nonstructural protein 3 redistributes fatty acid synthase to sites of viral replication and increases cellular fatty acid synthesis. Proceedings of the National Academy of Sciences. 2010 Oct 5;107(40):17345-50.

22. Umareddy I, Pluquet O, Wang QY, Vasudevan SG, Chevet E, Gu F. Dengue virus serotype infection specifies the activation of the unfolded protein response. Virology journal. 2007 Sep 24;4(1):91.

23. Luo D, Wei N, Doan DN, Paradkar PN, Chong Y, Davidson AD, Kotaka M, Lescar J, Vasudevan SG. Flexibility between the protease and helicase domains of the dengue virus NS3 protein conferred by the linker region and its functional implications. Journal of biological chemistry. 2010 Jun 11;285(24):18817-27.

24. Luo D, Xu T, Hunke C, Grüber G, Vasudevan SG, Lescar J. Crystal structure of the NS3 protease-helicase from dengue virus. Journal of virology. 2008 Jan 1;82(1):173-83.

25. Guzman MG, Halstead SB, Artsob H, Buchy P, Farrar J, Gubler DJ, Hunsperger E, Kroeger A, Margolis HS, Martínez E, Nathan MB. Dengue: a continuing global threat. Nature Reviews Microbiology. 2010 Dec 1;8:S7-16.

26. Gebhard LG, Filomatori CV, Gamarnik AV. Functional RNA elements in the dengue virus genome. Viruses. 2011 Sep 15;3(9):1739-56.

27. Idrees S, Ashfaq UA. A brief review on dengue molecular virology, diagnosis, treatment and prevalence in Pakistan. Genetic vaccines and therapy. 2012 Aug 28;10(1):6.

28. Suaya JA, Shepard DS, Beatty ME. Dengue burden of disease and cost of illness: TDR/SWG/08. Geneva: World Health Organization on behalf of the Special Programme for Research and Training in Tropical Diseases [Internet]. 2007 [cited 2017 Jan 10]. 54–60. Available from: http://www.who.int/tdr/publications/documents/swg_dengue_2.pdf

29. Bennett SN, Holmes EC, Chirivella M, Rodriguez DM, Beltran M, Vorndam V, Gubler DJ, McMillan WO. Selection-driven evolution of emergent dengue virus. Molecular biology and evolution. 2003 Oct 1;20(10):1650-8.

30. Fried JR, Gibbons RV, Kalayanarooj S, Thomas SJ, Srikiatkhachorn A, Yoon IK, Jarman RG, Green S, Rothman AL, Cummings DA. Serotype-specific differences in the risk of dengue hemorrhagic fever: an analysis of data collected in Bangkok, Thailand from 1994 to 2006. PLoS Negl Trop Dis. 2010 Mar 2;4(3):e617.

31. Lorono-Pino MA, Farfan-Ale JA, Zapata-Peraza AL, Rosado-Paredes EP, Flores-Flores LF, GARCÍA-REJÓN JE, DÍAZ FJ, Blitvich BJ, Andrade-Narvaez M, Jimenez-Rios E, Blair CD. Introduction of the American/Asian genotype of dengue 2 virus into the Yucatan State of Mexico. The American journal of tropical medicine and hygiene. 2004 Oct 1;71(4):485-92.

32. Cabrera-Cortina JI, Sánchez-Valdéz E, Cedas-DeLezama D, Ramírez-González MD. Oral calcium administration attenuates thrombocytopenia in patients with dengue fever. Report of a pilot study. InProc West Pharmacol Soc 2008 (Vol. 51, pp. 38-41).

33. Chunhakan S, Butthep P, Yoksan S, Tangnararatchakit K, Chuansumrit A. Vascular leakage in dengue hemorrhagic fever is associated with dengue infected monocytes, monocyte activation/exhaustion, and cytokines production. International journal of vascular medicine. 2015 Feb 5;2015.

34. Limjindaporn T, Wongwiwat W, Noisakran S, Srisawat C, Netsawang J, Puttikhunt C, Kasinrerk W, Avirutnan P, Thiemmeca S, Sriburi R, Sittisombut N. Interaction of dengue

virus envelope protein with endoplasmic reticulum-resident chaperones facilitates dengue virus production. Biochemical and biophysical research communications. 2009 Feb 6;379(2):196-200.

35. Khadka S, Vangeloff AD, Zhang C, Siddavatam P, Heaton NS, Wang L, Sengupta R, Sahasrabudhe S, Randall G, Gribskov M, Kuhn RJ. A physical interaction network of dengue virus and human proteins. Molecular & Cellular Proteomics. 2011 Dec 1;10(12):M111-012187.

36. Maruri-Avidal L, López S, Arias CF. Endoplasmic reticulum chaperones are involved in the morphogenesis of rotavirus infectious particles. Journal of virology. 2008 Jun 1;82(11):5368-80.

37. Ozcan L, Tabas I. Pivotal role of calcium/calmodulin-dependent protein kinase II in ER stress-induced apoptosis. Cell Cycle. 2010 Jan 15; 9(2): 223-224.

38. Oyadomari S, Takeda K, Takiguchi M, Gotoh T, Matsumoto M, Wada I, Akira S, Araki E, Mori M. Nitric oxide-induced apoptosis in pancreatic β cells is mediated by the endoplasmic reticulum stress pathway. Proceedings of the National Academy of Sciences. 2001 Sep 11;98(19):10845-50.

39. Tambunan US, Alamudi S. Designing cyclic peptide inhibitor of dengue virus NS3-NS2B protease by using molecular docking approach. Bioinformation. 2010 Jan 1;5(6):250-4.

40. Sánchez-Valdéz E, Delgado-Aradillas M, Torres-Martínez JA, Torres-Benítez JM. Clinical response in patients with dengue fever to oral calcium plus vitamin D administration: study of 5 cases. InProc West Pharmacol Soc 2009 (Vol. 52, pp. 14-7).

41. Doolittle JM, Gomez SM. Mapping protein interactions between Dengue virus and its human and insect hosts. PLoS Negl Trop Dis. 2011 Feb 15;5(2):e954.

42. Mairiang D, Zhang H, Sodja A, Murali T, Suriyaphol P, Malasit P, Limjindaporn T, Finley Jr RL. Identification of new protein interactions between dengue fever virus and its hosts, human and mosquito. PLoS One. 2013 Jan 11;8(1):e53535.

43. Ayub A, Ashfaq UA, Idrees S, Haque A. Global Consensus Sequence Development and Analysis of Dengue NS3 Conserved Domains. BioResearch open access. 2013 Oct 1;2(5):392-6.

44. http://www.healthmap.org/dengue/en/ [Cited 27 March 2017]

45. Mukhopadhyay, S., Kuhn, R. J., & Rossmann, M. G. A structural perspective of the flavivirus life cycle. *Nature Reviews Microbiology* **3**, 13–22 (2005). doi:10.1038/nrmicro106

46. Ekert PG, Vaux DL . Apoptosis and immune system . Br Med Bull (1997) 53 (3): 591-603.

47. Kapoor M, Zhang L, Ramachandra M, Kusukawa J, Ebner KE, Padmanabhan R. 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. 1995 Aug 11;270(32):19100-6.

48. Ying Huang , Kirk Staschke, Raffaele De Francesco, Seng-Lai Tan, Phosphorylation of hepatitis C virus NS5A nonstructural protein: A new paradigm for phosphorylation-dependent viral RNA replication, Virology,Volume 364, Issue 1, 20 July 2007, Pages 1–9

49. Noppakunmongkolchai W, Poyomtip T, Jittawuttipoka T, Luplertlop N, Sakuntabhai A, Chimnaronk S, Jirawatnotai S, Tohtong R. Inhibition of protein kinase C promotes dengue virus replication. Virology journal. 2016 Mar 1;13(1):35.

50. Raul C. Gomila, Glover W. Martin III, Lee Gehrk, NF90 Binds the Dengue Virus RNA 39 Terminus and Is a Positive Regulator of Dengue Virus Replication, PLoS One. 2011; 6(2): e16687.

51. Alvarez DE, Lodeiro MF, Filomatori CV, Fucito S, Mondotte JA, Gamarnik AV., Structural and functional analysis of dengue virus RNA. Novartis Found Symp. 2006;277:120-32; discussion 132-5, 251-3.

Chapter 2: Computational Prediction of Dengue NS3 Protein-Protein Interaction and Function

Abstract

Dengue virus (DENV) NS3 protein has a pivotal function in flavivirus RNA replication and viral protein maturation. Both the number of data sources in molecular biology, and their content, are increasing rapidly. Yet, the protein-protein interaction in Dengue remains poorly understood. Our approach presented a method to find out the significant protein-protein interactions between Dengue virus (DENV) proteins -NS3 in particular - and the human host proteins. We applied homology-base method to predict protein-protein interactions between NS3 and human proteins using Database of Interacting Proteins (DIP) database search.

Our approach is based on sequence similarities between our protein molecule sequence and other proteins or molecules in the DIP database as methods for protein structure and function analysis. This was the starting point of our project to explore the most sequence related proteins, then look to interaction networks and find out more about the communication that comes in the second level in the interaction links.

Introduction

Dengue is a mosquito-borne infection that causes a severe flu-like illness, and sometimes a potentially lethal complication called Dengue haemorrhagic fever. It belongs to the family *Flaviviradae*, genus *Flavivirus.* It is transmitted to humans via *Aedes Aegypti* (yellow fever mosquito). There are four serotypes of Dengue: DENV-1, DENV-2, DENV-3 and DENV-4 (1). The prevalence of these serotypes has grown dramatically around the world in recent decades. Some 2.5 billion people - two fifths of the world's population - are now at risk of contracting Dengue. The World Health Organization (WHO) currently estimates there may be 50 to 100 million Dengue infections worldwide every year (2).

Dengue virus type 2 encodes a single polyprotein precursor consisting of 3391 amino acid residues that is processed to at least 10 mature proteins by host and viral proteases. The NS3 protein contains a domain commonly found in cellular serine proteinases that, in cooperation with NS2B, is involved in polyprotein processing. In addition, NS3 and NS5 proteins contain conserved motifs found in several RNA helicases and RNA dependent RNA polymerases, respectively. Both enzymatic activities have been suggested to be involved in viral RNA replication. Furthermore, most of the cellular functions are passed out through physical interactions between deferent proteins. The discovery of the protein-protein interaction can consequently help to better understand the workings of the cell’s cellular functions. Essentially, there has been a considerable amount of work recently on developing high-throughput methods to produce more complete protein-protein interactions (3).

Major Protein Interactions of Proteins Similar to NS3

To identify important proteins that have substantial fragments with a high significant similarity to NS3 in Dengue, and study their protein-protein interaction with human, The interaction that we predicted was by using DIP (Database of Interacting Proteins) tools . DIP is a database of interacting protein are taken from large scale yeast-tow hybrid experiments.

DIP provides us with the desired list, which is important in checking previously documented findings related to those proteins. The behavior of significantly similar proteins to NS3 might shed some understanding of how the NS3 itself interacts.

After identifying all proteins that have substantial fragments with a high significant similarity to NS3 in Dengue types 1, 3, and 4, their protein-protein interactions with human proteins will be studied. In this step, we are extending the research done to the other serotypes of Dengue using the same techniques. The results will be used to study the difference between NS3 in Dengue type 2, and NS3 in Dengue 1, 3, and 4.

Materials and Methodology

Advancements in Bioinformatics, along with the availability of several software packages for large-scale prediction of protein-protein interaction and sequence alignments of the whole genome and the development of several functional databases that include function prediction methods to cope with the increasing data generated by high throughput sequencing experiments (4), makes it possible for us to study the viral protein from different levels. The following is a list of the main databases and computational tools used in our research:

1. DIP, a Database of Interacting Proteins (5), that hosts experimentally determined interactions between proteins.
2. Jalview, used to visualize multiple sequence alignments (6).
3. Kihara-PFP, which is a sequence similarity-based protein function prediction that will be used to predict the Gene Ontology (GO) annotations for query sequences beyond what can be found by conventional database searches, such as BLAST. Kihara takes into account sequences with weak similarities, as well as GO term associations observed in known annotations (7). (It is noteworthy to mention at this point that we discovered a mistake in the Kihara output. The issue was communicated to the Kihara team who fixed the code within a couple of days).
4. The Gene Ontology (GO) database presents a system of terms describing biological processes, cellular components, and molecular function and annotation data (8).
5. Moreover, other protein function prediction software, such as CombFunc (9), Profunc (10), and I-Tasser (Iterative Threading Assembly Refinement) (11), which

is based on a hierarchical algorithm for both protein structure and function prediction, have been used in our study.

The complete nucleotide sequences NS3 of Dengue virus type 1, 2, 3, 4, (each of about 618 amino acid residues) (Table 1), were extracted from the biological database of the National Centre for Biotechnology Information (NCBI) (12). These sequences construct the main input data that will be used later on with other tools to identify protein interaction and functional characteristics.

Table 1. Dengue virus serotype sequences used in this study

Name	**Description /Accession**
NS3 Dengue 1	nonstructural protein 3 [Dengue virus 1]/NP_722463.1
NS3 Dengue 2	non-structural protein 3, p [Dengue virus 2]/ NP_739587.2
NS3 Dengue 3	Nonstructural protein NS3 [Dengue virus 3]/ YP_001531172.2
NS3 Dengue 4	NS3 protein [Dengue virus 4]/NP_740321.1

The Virus Pathogen Database and Analysis Resource (ViPR) was used to retrieve multiple dengue virus sequences. ViPR is a good resource that provides access to sequence, annotations, immune epitopes, 3D structures, host factors, in addition to facilitating multiple sequence alignments, phylogenetic inferences, sequence variation determinations, BLAST comparisons, and other statistical analyses (13).

Protein-Protein interactions

Molecular interaction information is critical, and protein-protein interactions play a key role in all cellular processes, ranging from cellular division to apoptosis. With the advent of new technologies and computational methods, large sets of pair-wise physical protein-protein interactions and protein complexes have become available in public databases such as DIP. All the DIP data can be accessed online in both interactive and batch modes. The interactive, Web-based interface allows users to query the database for a specific protein based on its name, annotation, or species of origin.

Upon determining the protein-protein interactions from DIP, wherein we identify interactions between the NS3 protein molecule and other proteins or molecules that have similar sequences in order to track the interaction network, we proceed to investigate the functional prediction of different proteins involved in the network by using Protein Function Prediction (PFP). PFP is an automated function prediction tool based on sequence similarity that provides a three branch annotation for a given query sequence in biological process, molecular function, and cellular component.

The predicted functional properties from PFP were validated and compared to output from CombFunc, which is a Gene Ontology (GO) based function prediction server. CombFunc outputs data that can be associated with protein function by running several analyses for the input query sequence. It then combines the data using a Support Vector Machine (VSM) learning approach leading to the function prediction.

To identify all proteins that have substantial fragments with high significant similarity to NS3 in Dengue type 2, and study their protein-protein interaction with human,

we used DIP. The results in Table 2 show the list of proteins similar to NS3, and whose interaction network is available in the DIP database, since we used the BLAST tool available within DIP. The output provides the "Node", which refers to a protein in the tree that will be generated later to show the protein interaction network (Figure 1), the E-VAL showing the level of significance of the similarity between the protein found and NS3, the SwissProt and Refseq, and finally the name or description of the protein.

Table 2: Results from using BLAST NS3 Dengue type 2 sequence within DIP. The results are returned as a list of similar proteins to NS3 Dengue type 2.

Node	E-VAL	SwissProt	RefSeq	Protein Name/Description
DIP:3433N	0.029	O26051	NP_208313	ATP-DEPENDENT DNA HELICASE RECG HP1523
DIP:9937N	0.15	P43329	NP_415931	ATP-dependent helicase hrpA
DIP:153N	0.21	P25323	gi:102256	myosin-light-chain kinase A
DIP:18527N	1.6	Q9I7M1	NP_652666	CG18806-PA
DIP:6722N	1.9	Q04217	NP_013847	ATP-dependent RNA helicase DHR1
DIP:21737N	2.1	------	NP_572947	CG32604-PA open reading frame
DIP:17978N	2.9	------	NP_609946	CG10689-PA open reading frame
DIP:643N	3.2	P17948	NP_002010	protein-tyrosine kinase flt precursor
DIP:2668N	3.4	P20447	NP_011437	ATP-dependent RNA helicase DBP3
DIP:4516N	3.6	Q12099	NP_010304	translation initiation factor eIF-4A homolog FAL1
DIP:6276N	7.1	P13186	NP_013197	serine/threonine-specific protein kinase KIN2
DIP:27319N	9.2	O02425	NP_741632	spectrin beta chain

To further analyze the findings from the previous step, which gave us similar proteins to NS3 Dengue type 2 virus, we propose to investigate the interactions of each protein, since proteins in its network might interact with NS3.

We started with the best hit and investigated the protein-protein interaction of "ATP-DEPENDENT DNA HELICASE RECG HP1523". Figure 1 (A) shows the protein-protein interaction tree of a selected portion of interest from the protein-protein interaction

network as generated from DIP, wherein the nodes correspond to proteins and the edges correspond to interactions. We added Figure 3 (B) to indicate the specific protein names involved in the sub-network of interest.

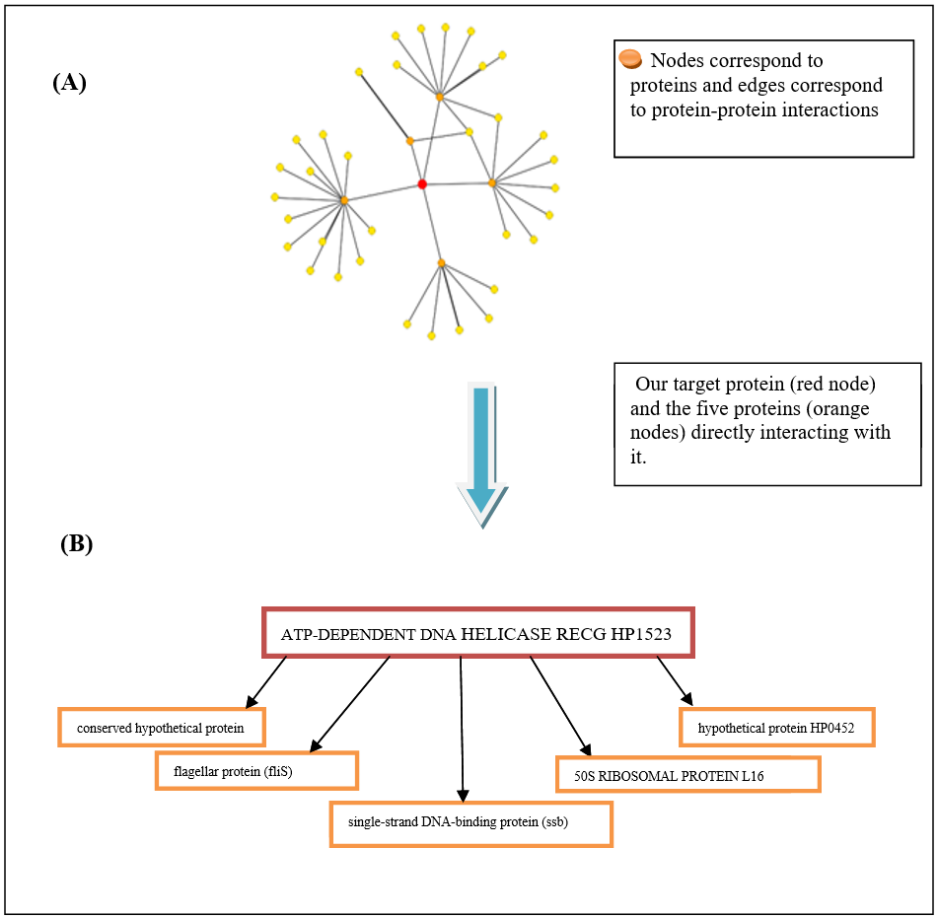

Figure 1: (A) map of protein – protein interactions for ATP-DEPENDENT DNA HELICASE RECG HP1523 from DIP server and our elaboration of the map. (B) Specific proteins involved in the network.

At a second level of the search, we performed a BLAST search comparing each of the proteins interacting with "ATP-DEPENDENT DNA HELICASE RECG HP1523" with the human and NS3 genomes. The results show, in particular, that this protein interacts with flagellar protein (fliS) HP0753. The path of our search suggests that proteins that are similar to flagellar protein most likely interact with NS3. The Blast search results of flagellar protein presented in the table below give a list of human proteins that are significantly similar to the flagellar protein and consequently are potential candidates to interact with NS3. Interestingly, the list shows zinc finger, a BTB domain-containing protein 20 isoform 2 (Homo sapiens) and calcium/calmodulin-dependent protein kinase type II which are themselves potential candidates to interact with NS3 and they might affect the viral replication (13).

Table 3: Blast search of proteins similar to (fliS) HP075 in human. The results are potential candidates of interaction with NS3. Hits show, particularly, the zinc fingers and Kinase type II.

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	
NP_078858.4	protocadherin Fat 4 precursor [Homo sapiens]	30.0	30.0	23%	0.57	47%	
NP_001157815.1	zinc finger and BTB domain-containing protein 20 isoform 2 [Homo sapiens] >ref|NP_056457.3| zinc finger and BTB domain-containing protein 20 isoform 2 [Homo sapiens]	29.3	29.3	38%	0.95	36%	
NP_001157814.1	zinc finger and BTB domain-containing protein 20 isoform 1 [Homo sapiens]	29.3	29.3	38%	0.95	36%	
NP_001245346.1	rho GTPase-activating protein 20 isoform 4 [Homo sapiens] >ref|NP_001245347.1| rho GTPase-activating protein 20 isoform 4 [Homo sapiens]	29.3	29.3	70%	1.1	22%	
	rho GTPase-activating protein 20 isoform 1 [Homo sapiens]	29.3	29.3	70%	1.1	22%	
NP_036602.1	transportin-3 isoform 1 [Homo sapiens]	26.6	26.6	38%	7.7	31%	
NP_079160.1	L-2-hydroxyglutarate dehydrogenase, mitochondrial precursor [Homo sapiens]	26.6	26.6	52%	8.1	29%	
NP_001177957.2	transportin-3 isoform 2 [Homo sapiens]	26.6	26.6	38%	8.5	31%	
NP_001895.1	catenin beta-1 [Homo sapiens] >ref|NP_001091680.1| catenin beta-1 [Homo sapiens] >ref|NP_001091679.1| catenin beta-1 [Homo sapiens]	26.6	26.6	43%	8.6	31%	
NP_001212.2	calcium/calmodulin-dependent protein kinase type II subunit delta isoform 3 [Homo sapiens]	26.6	26.6	42%	9.2	28%	
NP_001190194.1	interleukin-17 receptor C isoform 6 precursor [Homo sapiens]	26.6	26.6	30%	9.6	33%	
NP_116121.2	interleukin-17 receptor C isoform 3 precursor [Homo sapiens]	26.6	26.6	30%	9.6	33%	
NP_742126.1	calcium/calmodulin-dependent protein kinase type II subunit delta isoform 2 [Homo sapiens]	26.2	26.2	42%	9.9	28%	
NP_742113.1	calcium/calmodulin-dependent protein kinase type II subunit delta isoform 1 [Homo sapiens] >ref|NP_742125.1| calcium/calmodulin-dependent protein kinase type II subunit delta isoform 1 [Homo sapiens]	26.2	26.2	42%	9.9	28%	

We applied the same approach described above to the serotypes of Dengue 1 NS3, Dengue 3 NS3, and Dengue 4 NS3. The detailed results can be found in the appendix.

Functional Similarity

The second step is to clarify the functional similarity to narrow down our focus in Protein Function Prediction. Protein function similarity can also be used for protein interaction prediction. We run PFP for NS3 Dengue type 2. The sequence similarity, based on the protein function prediction server, generated a primary lists of predictions taking into account weakly similar sequences as well as GO term associations observed in known annotations. The results gave a total of 277 predictions divided as follows:

1. Molecular Function Terms 25 Predictions.
2. Biological Process Terms 193 Predictions.
3. Cellular Component Terms 59 Predictions.

Table 4 shows a sample output from PFP, a protein function prediction software, that gives a list of protein with similar functionalities as the NS3. The whole list is provided in the appendix.

Table 4: PFP output for NS3 Dengue. The PFP results on the table are a sample output showing a list of protein with similar functionalities as the NS3 along with Biological Process and Cellular Component similar proteins.

Molecular Function Terms		
PFP Score	**Term**	**Description**
112713.25	GO:0000166 [+]	nucleotide binding
95796.21	GO:0004386 [+]	helicase activity
94325.27	GO:0003676 [+]	nucleic acid binding
78705.82	GO:0005488 [+]	Binding
27484.17	GO:0003824 [+]	catalytic activity
Biological Process Terms		
PFP Score	**Term**	**Description**
24552.95	GO:0090304	nucleic acid metabolic process
17904.53	GO:0016070 [+]	RNA metabolic process
6610.14	GO:0075512 [+]	viral entry into host cell via clathrin-mediated endocytosis
4173.47	GO:0046718 [+]	viral entry into host cell
7005.70	GO:0019062 [+]	viral attachment to host cell
9054.11	GO:0006508 [+]	Proteolysis
Cellular Component Terms		
PFP Score	**Term**	**Description**
24191.69	GO:0005622 [+]	Intracellular
23741.71	GO:0044464	cell part
23741.27	GO:0005623	Cell
23676.09	GO:0044424	intracellular part
23140.93	GO:0005737 [+]	Cytoplasm

The second tool of choice to be used as a function prediction serves is CombFunc. It is a Gene Ontology (GO) based server that accepts a protein sequence and runs multiple analyses for the query sequence to obtain data that can be associated with protein function. This will be in turn used for prediction through machine learning. The results from CombFunc are shown in Table 5, below.

Table 5: ComFun Results for NS3 Dengue 2. The table shows seven functions resulting from a similarity based protein function prediction.

Molecular Function Predictions

rm	Description	Number	SVM Probability
05515	Protein binding	10	0.366
03824	Catalytic activity	10	0.351

Biological Process Predictions

GO Term	Description	Number	SVM Probability
GO:0075512	clathrin-mediated endocytosis of virus by host cell	10	0.418
GO:0039654	Fusion of virus membrane with host endosome membrane	10	0.418
GO:0046718	Viral entry into host cell	10	0.418
GO:0019062	Virion attachment to host cell	10	0.418
GO:0006508	Proteolysis	10	0.418
GO:0019048	Modulation by virus of host morphology or physiology	10	0.418
GO:0001172	transcription, RNA-templated	10	0.412

The results from the above prediction servers were compared and cross validated to ensure the main predicted proteins matched. For the extension of the study, to the rest of the dengue serotypes, we applied the same approach and tools used for NS3 Dengue

virus 2 to study function similarities associated in NS3 Dengue 1, 3, and 4 and to analyze the output we needed to combine all the results together as described next.

In order to investigate the combined results related to the four serotypes, we combined all outputs from the previous study that were originally applied to each serotype separately using Microsoft Access databases with the aim of exploring the differences between the four serotypes. Table 6 presents a sample output. The full combined result is in the appendix.

Table 6: Combined output from PFP for NS3 Dengue types 1,2,3,4.

ID	Description	d1.score	d2.score	d3.score	d4.score
GO:0009535	chloroplast thylakoid membrane		2544.83		
GO:0009058	biosynthetic process		5307.44		
GO:0043901	negative regulation of multi-organism process		3039.72	2692.85	2930.27
GO:0002697	regulation of immune effector process		3033.06	2716.51	2927.73
GO:0050688	regulation of defense response to virus		2981.84	2698.08	2880.52
GO:0002698	negative regulation of immune effector process		2858.79	2532.57	2755.86
GO:0032259	Methylation		2798.36		
GO:0044249	cellular biosynthetic process		2679.18		
GO:0044421	extracellular region part		2823.54		
GO:0017171	serine hydrolase activity		15968.02		11989.41
GO:0044434	chloroplast part		2580.78		
GO:0036265	cellular biosynthetic process		2782.55		
GO:0031984	organelle subcompartment		2447.33		
GO:0031668	cellular response to extracellular stimulus			3442.22	

In Table 6, the functions corresponding to a particular type will show the PFP score. For instance, “negative regulation of multi-organism process” (line 3 of the table) is predicted in types 2, 3, and 4 but not in type 1. Accordingly, regulation of defense response to virus and regulation of immune effector process is predicted in types 1,2, 3, and 4

We use a protein function prediction tool, PFP, to target the functions that are specific to one single serotype of NS3 dengue 2 but not to any other types. The results on

Table 7 suggest that the predicted functions associated with Dengue type 2, but not types 1, 3, and 4, are related to biosynthetic process and chloroplast. Further investigation of the results is required to determine if this is related to sequence variation with Dengue type 2.

Table 7: PFP Predicted functions for NS3 Dengue type 2

ID	Description	Score
GO:0009058	biosynthetic process	5307.44
GO:0033645	host cell endomembrane system	3331.8
GO:0044421	extracellular region part	2823.54
GO:0032259	Methylation	2798.36
GO:0036265	cellular biosynthetic process	2782.55
GO:0044249	cellular biosynthetic process	2679.18
GO:0009570	chloroplast stroma	2638.61
GO:0044434	chloroplast part	2580.78
GO:0009535	chloroplast thylakoid membrane	2544.83
GO:0055035	plastid thylakoid membrane	2542.08
GO:0009534	chloroplast thylakoid	2517.79
GO:0031976	plastid thylakoid	2514.95
GO:0033655	host cell cytoplasm part	2487.42
GO:0031984	organelle subcompartment	2447.33

Conclusion

Dengue virus (DENV) NS3 protein interaction returned a high number of significant BLAST hits that led us to further investigate the network of interactions. We found several interacting proteins whose connections can lead us to a better understanding of NS3 protein interactions with infected cells including Zinc finger and CamKII. Knocking out one of the interacting proteins that we found may inhibit viral replication. These findings open the door for further investigation and future work is expected to lead to the identification of key proteins that will cause the inhibition.

References

1. Gubler DJ. The global emergence/resurgence of arboviral diseases as public health problems. Archives of medical research. 2002 Aug 31;33(4):330-42.

2. Gubler DJ. TMH Dengue, Urbanization and Globalization: The Unholy Trinity of the 21 st Century. Tropical medicine and health. 2011;39(4 Supplement):3-11.

3. Kapoor M, Zhang L, Ramachandra M, Kusukawa J, Ebner KE, Padmanabhan R. 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. 1995 Aug 11;270(32):19100-6.

4. Chen C, McGarvey PB, Huang H, Wu CH. Protein bioinformatics infrastructure for the integration and analysis of multiple high-throughput "omics" data. Advances in bioinformatics. 2010 Mar 29;2010.

5. Xenarios I, Salwinski L, Duan XJ, Higney P, Kim SM, Eisenberg D. DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. Nucleic acids research. 2002 Jan 1;30(1):303-5.

6. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009 May 1;25(9):1189-91.

7. Chitale M, Khan IK, Kihara D. In-depth performance evaluation of PFP and ESG sequence-based function prediction methods in CAFA 2011 experiment. BMC bioinformatics. 2013 Feb 28;14(3):S2.

8. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA. Gene Ontology: tool for the unification of biology. Nature genetics. 2000 May 1;25(1):25-9.

9. Wass MN, Barton G, Sternberg MJ. CombFunc: predicting protein function using heterogeneous data sources. Nucleic acids research. 2012 Jul 1;40(W1):W466-70.

10. Laskowski RA, Watson JD, Thornton JM. ProFunc: a server for predicting protein function from 3D structure. Nucleic acids research. 2005 Jul 1;33(suppl 2):W89-93.

11. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. Nature protocols. 2010 Apr 1;5(4):725-38.

12. http://www.ncbi.nlm.nih.gov [Cited 2016 Oct 22]

13. Pickett BE, Greer DS, Zhang Y, Stewart L, Zhou L, Sun G, Gu Z, Kumar S, Zaremba S, Larsen CN, Jen W. Virus pathogen database and analysis resource (ViPR): a comprehensive bioinformatics database and analysis resource for the coronavirus research community. Viruses. 2012 Nov 19;4(11):3209-26.

14. Teo CS, Chu JJ. Cellular vimentin regulates construction of dengue virus replication complexes through interaction with NS4A protein. Journal of virology. 2014 Feb 15;88(4):1897-

Chapter 3: Predicting and Finding the Top Kinase that phosphorylate Dengue NS3

Abstract

Dengue virus (DENV) is a serious worldwide health concern that puts approximately 2.5 billion inhabitants of about 100 countries at risk. DENV, from the flaviviridea family, is transmitted to humans via mosquitos. The DENV genome codes for multiple nonstructural proteins, one of which is the NS3 enzyme that participates in multiple steps of the virus life cycle including replication, RNA genome synthesis, and host immune mechanism. Recent studies have suggested that the phosphorylation state of NS5 controls the association/dissociation of NS3 with NS5 and that this controls viral replication. The phosphorylation of NS3 might also play a role in the control of DENV replication. Therefore, we perform an *in silico* study of the phosphorylation of NS3. An *in silico* study is needed for a better understanding of the biological systems at the gene level. NS3 has the potential to be phosphorylated by any of the ~500 human kinases. We predicted and identified the potential kinases, and calculated their score using neural networks and other machine learning based programs. We investigated and identified the top kinases that phosphorylate DENV NS3, and that may interrupt the viral replication, and participate in antiviral evasion. Using multiple sequence alignment bioinformatics tools, we verified the results of the highly conserved residues, and the residues around active sites

whose phosphorylation may have a potential effect on viral replication. We further verified the results with multiple bioinformatics tools. Our studies propose that the Host-Mediated Phosphorylation of NS3 would affect its capability to interact with NS5, and knocking out one of the interacting proteins may inhibit viral replication. These results open new doors for further investigation, and future work is expected to help identify the key inhibition mechanisms.

Predicting the relationship between protein-protein interactions is challenging for any kind of protein, but it is especially challenging for molecules that carry out multiple cellular functions. Understanding the interactions at the molecular-level will allow scientists to understand how the viral infection works and, thereafter, find what protein function knockouts might inhibit its replication. Protein phosphorylation, which can play the role of a molecular rheostat or a switch for protein functions, is deemed to be the most frequent post-translational modification.

We propose a detailed study of the phosphorylation of the multifunctional Dengue virus protein NS3 with the aim of designing antiviral inhibitors by shedding more light on the effects of phosphorylation.

Introduction

Researchers are still in need of a clear understanding of the basic gene level biological mechanisms of the Dengue virus, in order to develop an effective therapeutic and prophylactic treatment for this important human pathogen. Such treatment would trigger complex interactions between the human cellular components and those of the virus.

One of the key components in DENV replication is NS5 of filoviruses. NS5 has been studied extensively, and its phosphoprotein has been shown to be critical; consequently, other multifunctional proteins might be regulated similarly.

The second largest key component in DENV replication machinery is the nonstructural protein 3 (NS3). The multifunctional enzyme, NS3, performs various actions during the viral replication, and plays an important role in antiviral evasion (1). The protein function is regulated by the post translational modification that a given protein undergoes. One such post translational modification is phosphorylation, which is involved in numerous interactions, such as protein-protein interactions, intracellular localization, signal transduction, protein folding, transcription regulation, and cell cycle, including cell survival and apoptosis.

The fact that protein phosphorylation is an important regulator of diverse intracellular processes, and that our previous study of protein-protein interaction for Dengue NS3 showed interesting kinases involved in the process, led us to further investigate kinases involved in the NS3 phosphorylation.

Accumulating evidence in the literature, our previous studies included both the protein-protein interactions and their functional similarities, as well as a high number of

significant Blast hits when checking the DENV NS3 protein interactions, stressed the need for a closer investigation of the complex network of interactions. We also found several interacting proteins whose connections can lead us to a better understanding of the mechanism governing the NS3 protein interaction with the infected cell. These findings open the door towards identifying the key proteins that may cause the inhibition of viral replication.

The family Flaviviridae comprises four main genera, *Flavivirus*, *Hepacivirus*, *Pestivirus* and a recent genus, *Pegivirus* (2). It has been shown that it is possible for NS3 to get phosphorylated in the HCV replication cells at the subgenomic level through phospho-specific staining and dephosphorylation assay (3).

Earlier studies showed that, within the cytoplasm of the infected cells, DENV NS5 interact with NS3, and the interplay between those two proteins depends on the phosphorylation status of NS5 (4).

For its part, NS3 has an interesting polypeptide processing function as an enzyme during which NS3 may be phosphorylated, which presents important ground for investigation. Indeed, no study to date has confirmed nor clarified the phosphorylation status of DENV NS3, and the identification of the mediator kinases involved during the phosphorylation of specific NS3 residues would be a strong interesting therapeutic objective (5).

Our research is based on previous findings, and on the need to identify new ideas that confirm the NS3 phosphorylation, identify the residues involved in the phosphorylation, and identify the top mediator kinases of that reaction. The results are

believed to break ground towards the *in silico* testing of the predicted protein phosphorylation sites, their residues, and the phosphorylating human kinases classified in our study based on their probabilities; this would then be extended to the in vitro conformation and forming of therapeutic drugs once the targets for inhibitions are identified. Such inhibition should aim to target, in particular, the phosphorylation of DENV NS3 which, like the phosphorylation of NS5, might have a crucial role for their interactions within the cytoplasm of the infected.

Materials and Methodology

The understanding of the genomic dynamics related to DENV requires the identification of the phosphorylation sites that are made possible in our post-genomic era wherein we can dissect protein functions and identify, using bioinformatics tools and databases, regulatory roles and complex interactions. A complete clarification of the phosphorylation mechanism makes it necessary to identify the involved kinases. The role of kinases has been a focus of studies over the last decade. Kinases can catalyze the hydrolysis of adenosine triphosphate (ATP), which, in turn, transfer a phosphate group to the appropriate serine (S) / threonine (T) or tyrosine (Y) residue (6). It is becoming increasingly important in the study of biological roles that the large number of predicted phosphorylation sites are systematically prioritized based on functional sites. It has been shown that phosphorylation sites, whose functions are known, have a higher likelihood of conservation. We combined the prediction tools to identify the top kinases with multiple sequence alignment tools to check the conservation sites. The prediction tools were chosen

after a careful review of the literature including studies classifying the different available tools. Our selected tools, described in the following paragraph, are rated as the most efficient, and they use different machine learning algorithms. The differentiation of the underlying algorithms in our selected tools will be used as our additional verification method of the results across different prediction bioinformatics software packages.

Predictions tools for phosphorylation sites

A variety of tools has been developed for *in silico* prediction of protein phosphorylation sites. Few research studies contrasted most of the available protein phosphorylation prediction tools. There are more than 40 software packages that predict the phosphorylation sites. About three fourths of those tools are kinase-specific; they provide the name of the kinase or kinase family involved in the phosphorylation. Because of the poor specificity and sensitivity of motifs, those tools use more than 15 different direct and combined machine learning (ML) algorithms in order to provide accurate predicted sites. Among the ML algorithms implemented in the prediction tools are the decision trees, genetic algorithms, support vector machines (SVMs), and artificial neural and position-specific scoring matrices (PSSMs) (7).

Our interest was on the kinase specific tools listed below in Table 1, so other tools were not considered in this study.

Table 1: Kinase specific tools

Author	Software	Website
Plewczyński *et al.* (2008)	AutoMotif	NA
Yu *et al.* (2010)	BAE	NA
Dang *et al.* (2008)	CRPhos	www.ptools.ua.ac.be/CRPhos
Iakoucheva *et al.* (2004)	DISPHOS	www.dabi.temple.edu/disphos
Tang *et al.* (2007)	GANNPhos	NA
Zhou *et al.* (2004)	GPS 1.0	gps.biocuckoo.org
Xue *et al.* (2008)	GPS 2.0	gps.biocuckoo.org
Xue *et al.* (2011)	GPS 2.1	gps.biocuckoo.org
Wang *et al.* (2008a)	IEPP	NA
Huang *et al.* (2005a)	KinasePhos 1.0	kinasephos.mbc.nctu.edu.tw
Wong *et al.* (2007)	KinasePhos 2.0	kinasephos2.mbc.nctu.edu.tw
Wan *et al.* (2008)	MetaPredPS	metapred.biolead.org/MetaPredPS
Gao and Xu (2010)	Musite	musite.sourceforge.net
Blom *et al.* (1999)	NetPhos	cbs.dtu.dk/services/NetPhos
Hjerrild *et al.* (2004)	NetPhosK	cbs.dtu.dk/services/NetPhos
Ingrell *et al.* (2007)	NetPhosYeast	cbs.dtu.dk/services/NetPhosYeast
Linding *et al.* (2007)	NetworKIN	networkin.info
Sobolev *et al.* (2010)	PAAS	NA
Durek *et al.* (2009)	Phos3D	phos3d.mpimp-golm.mpg.de
Li *et al.* (2008b)	PhoScan	bioinfo.au.tsinghua.edu.cn/phoscan
Gnad *et al.* (2007)	PHOSIDA	phosida.de
Koenig and Grabe (2004)	PHOSITE	NA
Heazlewood *et al.* (2008)	PhosPhAt	phosphat.mpimp-golm.mpg.de

Neuberger *et al.* (2007)	pkaPS	mendel.imp.ac.at/sat/pkaPS
Jung *et al.* (2010)	PostMod	pbil.kaist.ac.kr/PostMod
Biswas *et al.* (2010)	PPRED	ashiskb.info/research/ppred
Xue *et al.* (2006)	PPSP	ppsp.biocuckoo.org
Brinkworth *et al.* (2003)	Predikin 1.0	predikin.biosci.uq.edu.au
Saunders *et al.* (2008)	Predikin 2.0	predikin.biosci.uq.edu.au
Kim *et al.* (2004)	PredPhospho	NA
Berry *et al.* (2004)	rBBFNN	NA
Yaffe *et al.* (2001)	Scansite	scansite.mit.edu
Yoo *et al.* (2008)	SiteSeek	NA
Li *et al.* (2008a)	SMALI	lilab.uwo.ca/SMALI.htm

We focused on GPS 3.0, NetPhos 3.1 and Scansite 3, which were listed in a recent study as among the most reliable tools(8). The tools use multiple databases of phosphorylation sites that have already been established, such as Phospho.ELM (9) and PHOSIDA (10) for general phosphorylation, and the specific human proteins database, PhosphoSitePlus (11).

The experimental errors, however, are not completely negligible in most of the databases, as the phosphorylation sites are identified by using high throughput tools. Our *in silico* study combines the results of prediction tools with the multiple sequence alignment tools, the biological mechanisms, and supports in vitro findings from the literature, to come up with the top kinases.

Protein phosphorylation prediction software analysis

A benchmark study was conducted to determine the best performance characteristics among the leading protein phosphorylation prediction software, as the growth of sophistication in the collection of genomic data and the recent developments of computational bioinformatics gave us the opportunity for such selection in Table 1. Conclusively, and after a review of comparisons and categorizations of the available methods for computational phosphorylation site prediction tools with different machine learning techniques, we selected as phosphorylation prediction tools: 1. GPS, which ranked among the top tools, by predicting all phosphorylation sites with a fidelity of 83%; 2. NetPhos, which came in second place; and 3. ScanSite, which, in addition to making the top tools, is also widely used in the literature, (see Table 2).

Table 2: Selected phosphorylation site prediction tools

Name	***Technique***	***Reference***	***Website***
GPS 3.0	PSSM, GA	Xue *et al.* (2011) (12)	ps.biocuckoo.org
NetPhos3	ANN	Blom *et al.* (1999) (13)	bs.dtu.dk/services/NetPhos
Scansite 3.	PSSM	Yaffe *et al.* (2001) (14)	scansite.mit.edu

These three *in silico* tools for predicting protein phosphorylation sites, take a protein sequence, and output both the score measuring the likelihood that a residue (serine,

threonine or tyrosine) in the inputted sequence is a phosphorylation site, and the corresponding kinase.

We analyzed the amino acid sequence of (DENV2) NS3 using these three *in silico* tools for predicting protein phosphorylation sites and located possible human kinases that may phosphorylate amino acids on NS3.

Multiple sequence alignment

We took into consideration the results from several studies focusing on the conservation of phosphorylated sites. It was confirmed that phosphoproteins are subject to more conservation in evolution compared to their non-phosphorylated proteins. Thus, our approach in analyzing the predicted phosphorylated sites was to prioritize those sites that were conserved in the NS3 sequences. (15,16).

We ran multiple sequence alignments to identify the conserved sites among NS3 Dengue2 virus sequences. We then looked at phosphorylation core sites that were predicted from the protein phosphorylation prediction tools and checked if they belong to conserved sites in the multiple sequence alignment. We ran five multiple sequence alignments with a total of 971 NS3 DENV 2 sequences, 1533 NS3 DENV 1 sequences, 961 NS3 DENV 3 sequences, and 177 NS3 DENV 4 sequences.

The fact that Zika virus is related to dengue virus, Zika belongs to the same Flavivirus genus affecting humans might lead to a lot of similarities with the Dengue virus (37); and because of the availability of the Zika virus for *in vitro* testing, we included it in our research. Thus, 163 NS3 Zika virus sequences were added to the multiple sequence

alignment and to the rest of the methods to be included in this study. All sequences were extracted from the National Center for Biotechnology Information (NCBI).

Results and discussion

We analyzed the predicted protein phosphorylated sites and their corresponding residue from the output of the GPS, NetPhos, and Scansite tools, by looking at the positions of the residues, the amino acids in those positions, the predicted kinases, the peptides, and the scores associated to the predicted kinases. The results are as follows:

GPS Result for Dengue NS3

Based on calculated false positive rates, GPS uses three levels of thresholds: high, medium and low with a corresponding percentage of 2%, 6% and 10%, respectively. We ran GPS 3.0 with a medium threshold, which is consistent with the literature. Table 3, below, shows a sample output with the 24 top scores of the 1489 predicted phosphorylated sites in NS3.

Table 3: A sample GPS3 output of Dengue NS3 with the 24 predicted phosphorylation with top scores

Position	Code	Kinase	Peptide	Score
532	T	Other/TTK/TTK	LRGEARKTFVDLMRR	53.364
407	T	Other/TTK/TTK	NDWDFVVTTDISEMG	49.455
271	S	CMGC/MAPK/JNK/JNK2	TFTMRLLSPVRVPNY	31.353
500	T	CMGC/MAPK/JNK/JNK2	MLLDNINTPEGIIPS	29.941

137	S	CMGC/MAPK/JNK/JNK2	FSPGTSGSPIIDKKG	28.765
317	T	CMGC/MAPK/JNK/JNK2	AGIFMTATPPGSRDP	28.265
131	S	CMGC/MAPK/JNK/JNK2	GAVSLDFSPGTSGSP	27.971
45	T	Atypical/PDHK/PDHK/PDK1	AGVYKEGTFHTMWHV	27
189	T	STE/STE20/PAKA	IFRKRRLTIMDLHPG	26.407
244	T	CMGC/MAPK/JNK/JNK2	GLPIRYQTPAIRAEH	25.647
317	T	CMGC/MAPK/ERK/Erk1	AGIFMTATPPGSRDP	25.208
317	T	Atypical/PIKK/FRAP	AGIFMTATPPGSRDP	24.917
358	T	STE/STE20/MST/MST1	VTDFKGKTVWFVPSI	23.935
9	S	CMGC/MAPK/JNK/JNK2	GVLWDVPSPPPMGKA	23.059
131	S	CMGC/CDK/CDK5	GAVSLDFSPGTSGSP	22.808
271	S	CMGC/MAPK/ERK/Erk1	TFTMRLLSPVRVPNY	22.34
271	S	CMGC/CDK/CDK5/CDK5	TFTMRLLSPVRVPNY	22.148
601	Y	TK/Eph/EphB1	RWLDARIYSDPLALK	20
131	S	CMGC/CDK/CDK2/CDK2	GAVSLDFSPGTSGSP	18.322
601	Y	TK/Axl/MER	RWLDARIYSDPLALK	17.5
271	S	CMGC/CDK/CDK2/CDK2	TFTMRLLSPVRVPNY	16.51
9	S	CMGC/CDK/CDK2/CDC28	GVLWDVPSPPPMGKA	16.472
131	S	CMGC/CDK/CDK5/CDK5	GAVSLDFSPGTSGSP	15.827
131	S	CMGC/CDK/CDK2/CDC28	GAVSLDFSPGTSGSP	15.722

We looked at each kinase that appeared (at least once) in the GPS3 output for DENV 2 NS3 and its highest score in the table (with all 1489 predictions). Results are shown in Figure 1.

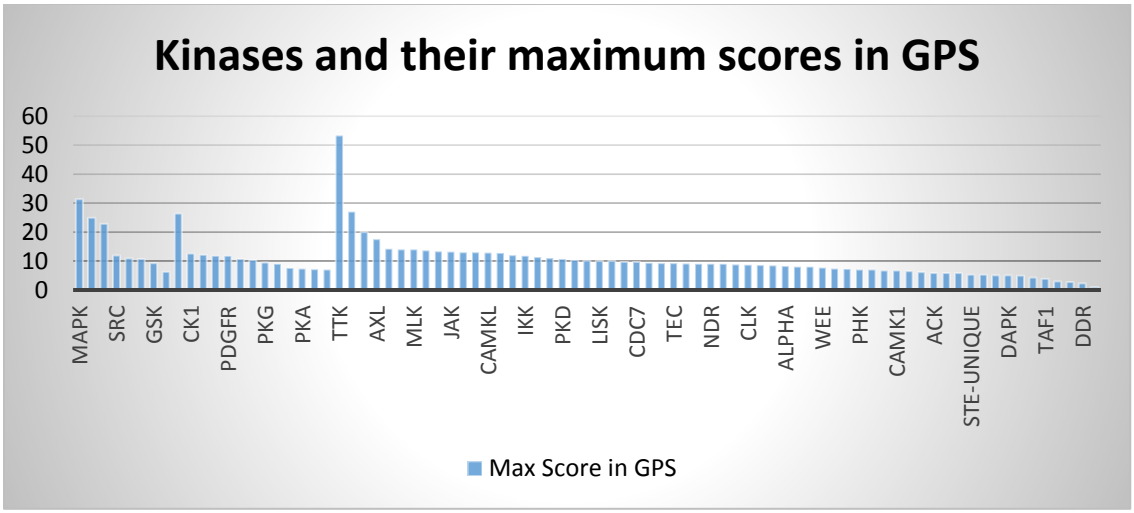

Figure 1: Kinases and their maximum scores in GPS. A single kinase that is predicted by GPS to be a potential candidate to phosphorylate NS3 might appear at multiple sites. GPS associates different scores to each phosphorylation by a kinase at different residues. The graph shows the 86 kinases that are predicted by GPS to phosphorylate NS3 and only their maximum score.

The previous graph shows a peak score between INSR and PDHK. A magnified portion of the graph confirms that the highest ranked kinase using GPS3 is TTK. The results were used along with outputs from the other prediction tools, and only the kinases that were common to all tools were analyzed. Accordingly, even though the score was high, the TTK result was considered inconsistent with the other software tools and, thus, we did not consider it for this study. It may be used in future investigations.

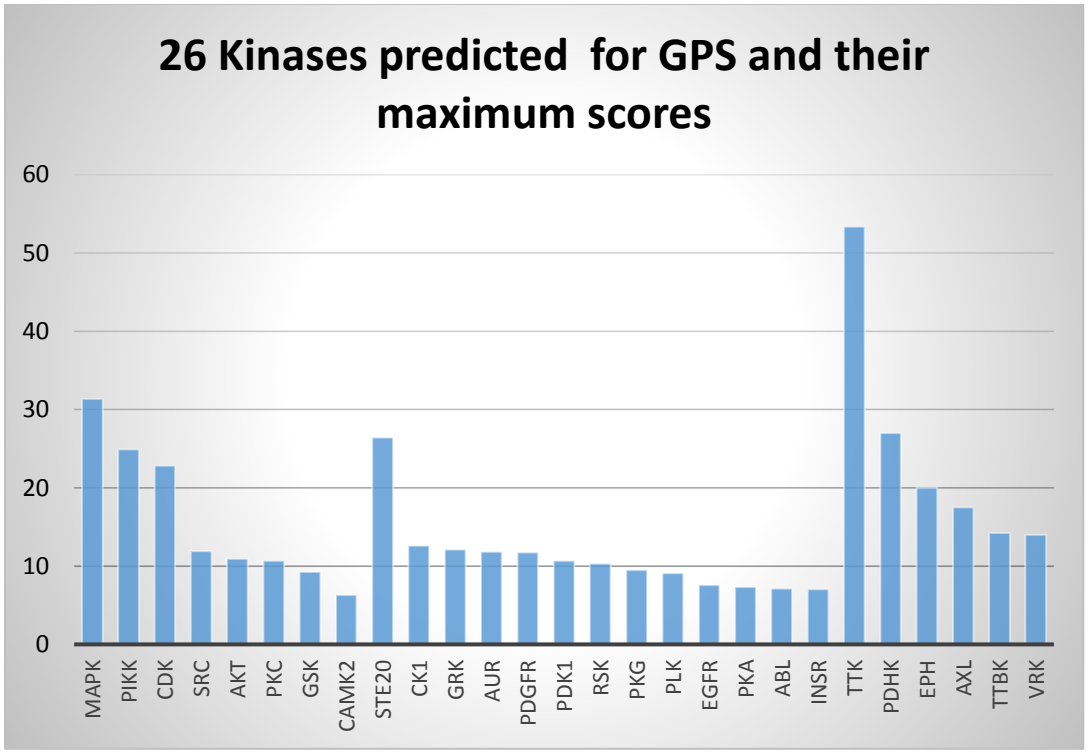

Figure 2: Magnified scores figure of Kinases and their maximum scores in GPS. A single kinase that is predicted by GPS to be a potential candidate to phosphorylate NS3 might appear at multiple sites. GPS associates different scores to each phosphorylation by a kinase at different residues. The graph shows 26 kinases that are predicted by GPS to phosphorylate NS3 and only their maximum score

GPS Result for Zika NS3

As previously mentioned, Zika was included in the analysis therefore we used GPS for Zika NS3 with the consistent medium threshold. A sample output with the 24 top scores of the predicted phosphorylated sites in Zika NS3 is shown below in Table 4.

Table 4: A sample GPS3 output of Zika NS3 with the 24 top scores

Position	Code	Kinase	Peptide	Score	Position
246	T	STE/STE20/MST/MST1	LPVRYMTTAVNVTHS	43.097	2.612
318	T	STE/STE20/MST/MST1	AAIFMTATPPGTRDA	36.097	2.612
608	S	AGC/PKC/PKCa/PRKCA	SDHAALKSFKEFAAG	29.727	5.842
174	T	CMGC/MAPK/JNK/JNK2	QGKREEETPVECFEP	29.5	6.118
137	S	CMGC/MAPK/JNK/JNK2	YPAGTSGSPILDKCG	29.059	6.118
318	T	CMGC/MAPK/JNK/JNK2	AAIFMTATPPGTRDA	28.265	6.118
608	S	Atypical/PDHK/PDHK	SDHAALKSFKEFAAG	28.05	2.805
174	T	CK1/CK1/CK1-D/CSNK1D	QGKREEETPVECFEP	27.45	8.142
331	S	CMGC/MAPK/JNK/JNK2	DAFPDSNSPIMDTEV	24.882	6.118
318	T	Atypical/PIKK/FRAP	AAIFMTATPPGTRDA	24.5	8.145
219	T	STE/STE20/MST/MST1	AIKKRLRTVILAPTR	23.871	2.612
318	T	CMGC/MAPK/ERK/Erk1	AAIFMTATPPGTRDA	23.854	10.326
608	S	AGC/GRK/BARK/BARK1	SDHAALKSFKEFAAG	23.024	3.384
1	S	CK1/VRK/VRK2	*******SGALWDVP	21.5	4
531	T	STE/STE20/PAKA	LRTEQRKTFVELMKR	19.372	6.397
1	S	Other/PLK	*******SGALWDVP	17.058	8.093
331	S	Atypical/PIKK/FRAP	DAFPDSNSPIMDTEV	16.556	8.145
174	T	CMGC/CDK/CDK5/CDK5	QGKREEETPVECFEP	16.432	7.867
68	Y	TK/Trk/TRKA	GEGRLDPYWGDVKQD	15.8	11.95
174	T	CMGC/CDK/CDK2/CDC28	QGKREEETPVECFEP	15.694	9.411
190	T	STE/STE20/PAKA	MLKKKQLTVLDLHPG	15.605	6.397
331	S	CMGC/MAPK/ERK/Erk1	DAFPDSNSPIMDTEV	15.549	10.326
219	T	CAMK/PIM/PIM1	AIKKRLRTVILAPTR	15.538	5.965
608	S	AGC/GRK/BARK	SDHAALKSFKEFAAG	15.512	3.806
452	S	Other/CDC7	PMPVTHASAAQRRGR	15.333	9.583

NetPhos Result for Dengue NS3

The second tool, NetPhos, was run on DENV NS3. The table below shows the predicted phosphorylation sites, the amino acids, the peptide, the score, and the phosphorylating kinase.

Table (5) shows a sample output with the 24 top scores of the (953) predicted phosphorylated sites in NS3.

Table 5: A sample NetPhos Dengue ns3 output with the 24 top scores.

Position	Code	Peptide	score	Kinase
266	T	HATFTMRLL	0.854	PKC
161	Y	RSGAYVSAI	0.82	Unsp
189	T	KRRLTIMDL	0.817	PKA
131	S	SLDFSPGTS	0.772	Unsp
23	Y	EDGAYRIKQ	0.751	Unsp
200	T	GAGKTKRYL	0.732	PKC
189	T	KRRLTIMDL	0.725	PKB
218	T	RGLRTLILA	0.714	PKC
266	T	HATFTMRLL	0.707	Unsp
135	S	SPGTSGSPI	0.685	Unsp
168	T	AIAQTEKSI	0.678	PKC
131	S	SLDFSPGTS	0.677	cdk5
301	S	RGYISTRVE	0.671	PKC
602	S	ARIYSDPLA	0.66	PKA
271	S	MRLLSPVRV	0.654	cdk5
244	T	IRYQTPAIR	0.653	cdk5

200	T	GAGKTKRYL	0.648	Unsp
317	T	FMTATPPGS	0.646	cdk5
389	T	LSRKTFDSE	0.642	Unsp
23	Y	EDGAYRIKQ	0.632	EGFR
435	T	PVILTDGEE	0.631	CKII
500	T	DNINTPEGI	0.63	Unsp
407	T	DFVVTTDIS	0.629	CKII
271	S	MRLLSPVRV	0.627	p38MAPK
358	T	FKGKTVWFV	0.627	PKC

As was previously done with GPS3, we looked separately at each kinase that appeared (at least once) in the NetPhos output for DENV NS3 and its highest score in the table (with all 953 predictions). Results are shown in Figure 3.

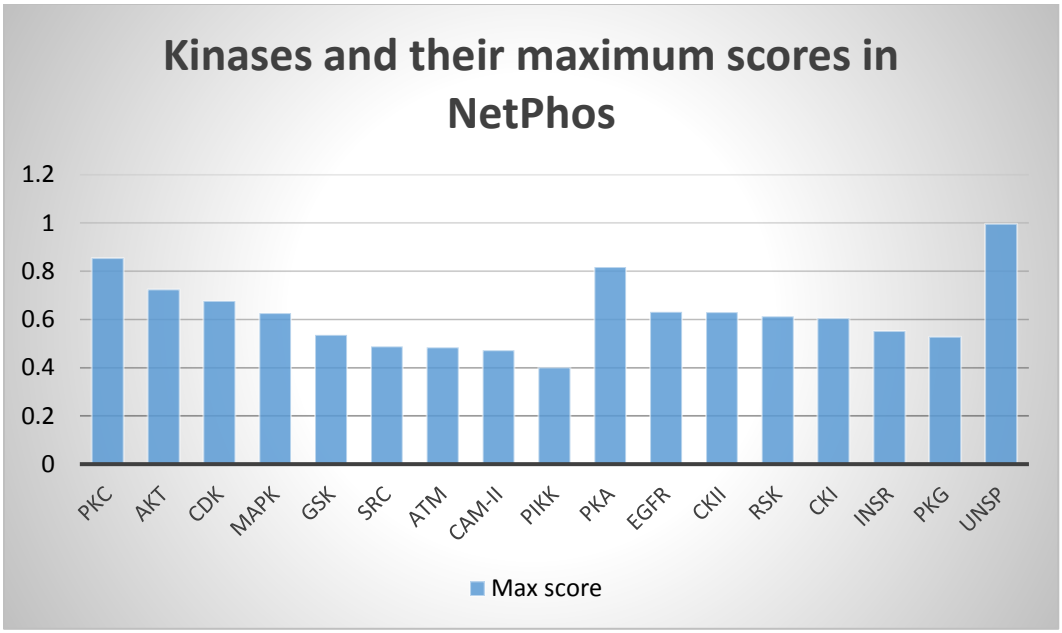

Figure 3: NetPhos output Dengue NS3. A single kinase that is predicted by NetPhos to be a potential candidate to phosphorylate NS3 might appear at multiple sites. NetPhos associates different scores to each phosphorylation by a kinase at different residues. The graph shows the 17 kinases that are predicted by NetPhos to phosphorylate NS3 and only their maximum score.

NetPhos Result for Zika NS3

Zika NS3 NetPhos Table (6) A sample output with the 24 top scores of the predicted phosphorylated sites in NS3.

Table 6: A sample NetPhos Zika ns3 output with the 24 top scores

Position	Code	Peptide	score	Kinase
303	T	GYISTRVEM	0.575	CKII
608	S	AALKSFKEF	0.57	PKG
480	T	GCAETDEGH	0.567	CKII
19	T	KGETTDGVY	0.564	unsp
322	T	TPPGTRDAF	0.563	PKC

245	T	VRYMTTAVN	0.562	PKG
346	S	ERAWSSGFD	0.559	PKA
318	T	FMTATPPGT	0.557	p38MAPK
53	T	MWHVTKGAA	0.555	PKC
18	T	KKGETTDGV	0.553	Unsp
246	T	RYMTTAVNV	0.549	Unsp
409	T	FVITTDISE	0.548	CKII
78	S	QDLVSYCGP	0.545	CKI
601	S	ARVCSDHAA	0.532	cdc2
347	S	RAWSSGFDW	0.524	PKA
23	Y	TDGVYRVMT	0.523	SRC
302	S	RGYISTRVE	0.522	Unsp
356	S	VTDHSGKTV	0.522	PKC
166	T	VSAITQGKR	0.521	PKC
584	Y	VWTKYGEKR	0.521	EGFR
251	T	AVNVTHSGT	0.519	CKI
506	S	GLIASLYRP	0.519	cdc2
137	S	GTSGSPILD	0.518	p38MAPK
182	S	CFEPSMLKK	0.518	cdc2
303	T	GYISTRVEM	0.575	CKII
608	S	AALKSFKEF	0.57	PKG

ScanSite Results for Dengue NS3:

The third tool we run was ScanSite for DENV NS3. Part of the 108 predicted output is given in Table 7.

Table 7: A sample Scansite Dengue ns3 output with the 24 top scores

Score	Motifs	Position	Peptide	Kinase
0.744	Basophilic serine/threonine kinase (Baso_ST_kin)	T532	LRGEARKtFVDLMRR	AURKB
0.731	Basophilic serine/threonine kinase (Baso_ST_kin)	T252	PAIRAEHtGREIVDL	PRKAA1
0.712	Basophilic serine/threonine kinase (Baso_ST_kin)	S452	GPMPVTHsSAAQRRG	PRKAA1
0.709	Basophilic serine/threonine kinase (Baso_ST_kin)	S602	WLDARIYsDPLALKE	AKT1
0.687	Basophilic serine/threonine kinase (Baso_ST_kin)	S271	TFTMRLLsPVRVPNY	AKT1
0.682	Kinase binding site group (Kin_bind)	T45	AGVYKEGtFHTMWHV	PDPK1
0.639	Proline-dependent serine/threonine kinase group (Pro_ST_kin)	S9	GVLWDVPsPPPMGKA	CDK5

0.638	Proline-dependent serine/threonine kinase group (Pro_ST_kin)	T317	AGIFMTAtPPGSRDP	CDC2
0.638	Tyrosine kinase group (Y_kin)	Y472	PKNENDQyIYMGEPL	INSR
0.633	Proline-dependent serine/threonine kinase group (Pro_ST_kin)	T317	AGIFMTAtPPGSRDP	MAPK3
0.627	Proline-dependent serine/threonine kinase group (Pro_ST_kin)	T244	GLPIRYQtPAIRAEH	CDK1
0.609	Proline-dependent serine/threonine kinase group (Pro_ST_kin)	T244	GLPIRYQtPAIRAEH	CDC2
0.606	Phosphotyrosine binding group (PTB)	Y472	PKNENDQyIYMGEPL	SHC1
0.606	Proline-dependent serine/threonine kinase group (Pro_ST_kin)	T244	GLPIRYQtPAIRAEH	CDK1
0.606	Proline-dependent serine/threonine kinase group (Pro_ST_kin)	S135	LDFSPGTsGSPIIDK	MAPK3
0.604	Basophilic serine/threonine kinase (Baso_ST_kin)	T389	VIQLSRKtFDSEYVK	PRKACG
0.602	Proline-dependent serine/threonine kinase group (Pro_ST_kin)	T317	AGIFMTAtPPGSRDP	CDK5
0.598	Proline-dependent serine/threonine kinase group (Pro_ST_kin)	S137	FSPGTSGsPIIDKKG	CDK1
0.595	Basophilic serine/threonine kinase (Baso_ST_kin)	T358	VTDFKGKtVWFVPSI	CAMK2G

0.595	Tyrosine kinase group (Y_kin)	Y601	RWLDARIySDPLALK	ABL1
0.588	Proline-dependent serine/threonine kinase group (Pro_ST_kin)	T500	MLLDNINtPEGIIPS	MAPK3
0.583	Basophilic serine/threonine kinase (Baso_ST_kin)	T189	IFRKRRLtIMDLHPG	PRKAA1
0.583	Proline-dependent serine/threonine kinase group (Pro_ST_kin)	S9	GVLWDVPsPPPMGKA	MAPK3
0.581	DNA damage kinase group (DNA_dam_kin)	S171	AIAQTEKsIEDNPEI	PRKDC

Similar to the previously completed GPS3 and NetPhos runs, we looked at each kinase that appeared (at least once) in ScanSite output for DENV NS3, and its highest score in the table (with all 108 predictions). Results are shown in Figure 4.

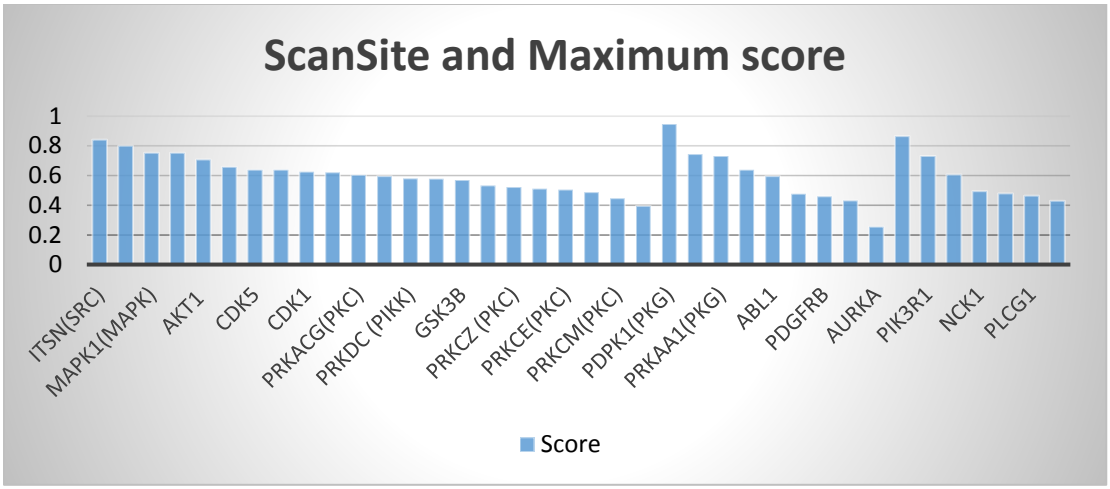

Figure 4: ScanSite output Dengue NS3and Maximum score , A single kinase that is predicted by ScanSite to be a potential candidate to phosphorylate NS3 might appear at multiple sites. ScanSite associates different scores to each phosphorylation by a kinase at different residues. The graph shows the 38 kinases that are predicted by ScanSite to phosphorylate NS3 and only their maximum score

ScanSite Results for Zika NS3

Zika NS3 ScanSite, Table 6, shows a sample output with the 24 top scores of the predicted phosphorylated sites in NS3.

Table 8: A sample output with the 24 top scores

Score	Motifs	Position	Peptide	Kinase
0.79	Clk2 Kinase (Clk2_Kin)	S346	EVPERAWsSGFDWVT	CLK2
0.73	Aurora B (AuroB)	T190	MLKKKQLtVLDLHPG	AURKB
0.72	Aurora B (AuroB)	S346	EVPERAWsSGFDWVT	AURKB
0.7	AMP_Kinase (AMPK)	T246	LPVRYMTtAVNVTHS	PRKAA1
0.68	Akt Kinase (Akt_Kin)	S346	EVPERAWsSGFDWVT	AKT1
0.66	Cdc2 Kinase (Cdc2_Kin)	T318	AAIFMTAtPPGTRDA	CDC2
0.66	Erk1 Kinase (Erk1_Kin)	T322	MTATPPGtRDAFPDS	MAPK3
0.65	Erk1 Kinase (Erk1_Kin)	T318	AAIFMTAtPPGTRDA	MAPK3

0.65	Aurora A (AuroA)	S346	EVPERAWsSGFDWVT	AURKA
0.65	Erk1 Kinase (Erk1_Kin)	S135	LDYPAGTsGSPILDK	MAPK3
0.64	AMP_Kinase (AMPK)	T190	MLKKKQLtVLDLHPG	PRKAA1
0.63	Cdk5 Kinase (Cdk5_Kin)	T318	AAIFMTAtPPGTRDA	CDK5
0.62	14-3-3 Mode 1 (1433_m1)	T245	GLPVRYMtTAVNVTH	YWHAZ
0.62	DNA PK (DNA_PK)	T322	MTATPPGtRDAFPDS	PRKDC
0.61	DNA PK (DNA_PK)	T166	GSYVSAItQGKREEE	PRKDC
0.61	DNA PK (DNA_PK)	T336	SNSPIMDtEVEVPER	PRKDC
0.61	Protein Kinase A (PKA_Kin)	T531	LRTEQRKtFVELMKR	PRKACG
0.6	GSK3 Kinase (GSK3_Kin)	T322	MTATPPGtRDAFPDS	GSK3A
0.6	Shc PTB (Shc_PTB)	Y161	VVIKNGSyVSAITQG	SHC1
0.6	Protein Kinase A (PKA_Kin)	T390	VIQLSRKtFETEFQK	PRKACG
0.6	CDK1 motif 1 - [ST]Px[KR]x (CDK1_1)	T318	AAIFMTAtPPGTRDA	CDK1
0.59	GSK3-improved (GSK3b)	T290	IMDEAHFtDPSSIAA	GSK3B
0.59	CDK1 motif 1 - [ST]Px[KR]x (CDK1_1)	S137	YPAGTSGsPILDKCG	CDK1
0.59	Protein Kinase A (PKA_Kin)	S160	GVVIKNGsYVSAITQ	PRKACG

Dengue and Zika NS3 top kinases

As previously mentioned, Zika was included in the analysis; and the list of its top predicted kinases match those of the DENV NS3, as will be explained further in the Analysis of the Results section.

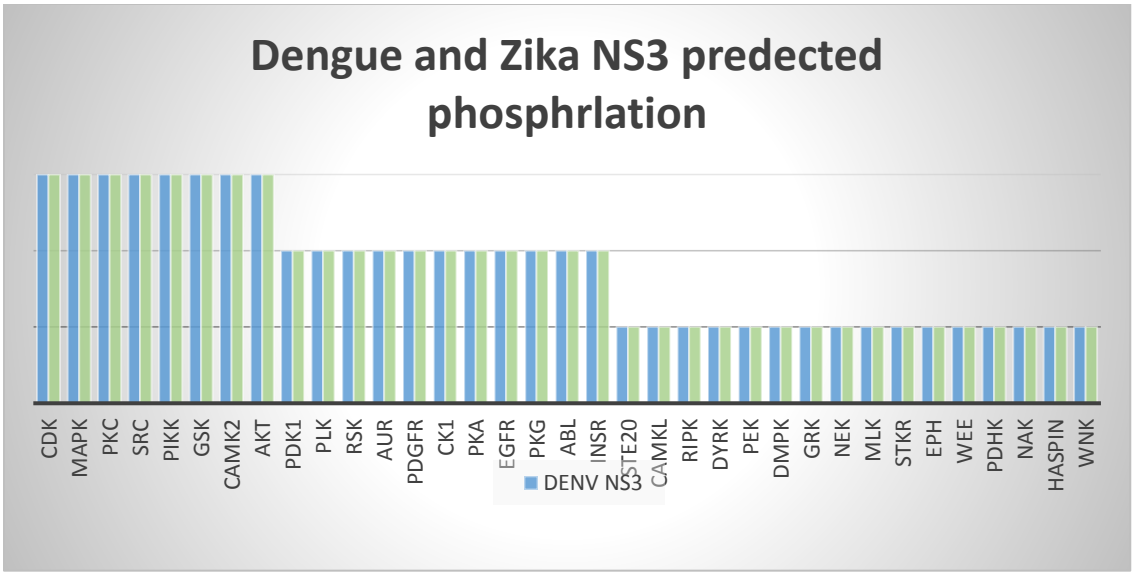

Figure 5: The figure shows the kinases in Dengue NS3 and Zika NS3 that are predicted by all three tools (GPS, NetPhos, and ScanSite). In the left we find the top kinases CDK, MAPK, PKC, CAMK2, SRC, PIKK, GSK, and AKT.

Analysis of the Result

Our investigation of possible protein phosphorylation sites revealed many potential candidate positions on DENV2 NS3, as well as several corresponding human kinases in all three tools (GPS, NetPhos, and ScanSite), along with their score and residue position in the sequence. The identification of the top kinases required a three-tier analysis (see Figure 6).

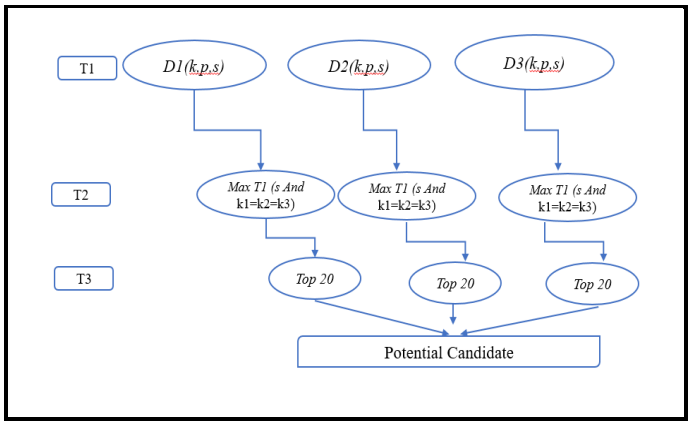

Figure 6: The three Tiers. Tier 1: All the kinases that were predicted to phosphorylate NS3.Tier 2: Max score predicted to phosphorylate NS3. Tier 3: The top 20 kinases that were predicted, by GPS, NetPhos, and Scansite, to phosphorylate NS3.

In the first tier, we used our selected tools, GPS, NetPhos, and Scansite and predicted a total 1489, 953, and 108 possible phosphorylating kinases, respectively. In the second tier, we considered the predicted kinases that have a combination of a high score along with a hit in each of the three tools. The number of kinases went down to 61, 76, and 64, respectively, from the three tools listed above. Lastly, in the third tier, we took the top 20 kinases from the second tier, Look to (Table 9).

Table 9: GPS, NetPhos, and ScanSite tier's outputs.

	Number of prediction Tier 1	Number of prediction Tier 2	Number of prediction Tier 3
GPS 3.0	1489	61	20
NetPhos3.1	953	76	20
Scansite 3	108	64	20

Not only did the above analysis find the top candidate kinases for DENV NS3 phosphorylation, but it also revealed that a top cohort of predicted kinases found in DENV NS3 is also found to phosphorylate the Zika NS3 residues; namely, the kinases CDK, MAPK, PKC, CAMK2, SRC, PIKK, GSK, and AKT.

Table 10: Top Kinases

	GPS			Netphos			ScanSite	
Kinase	Position	Score	Kinase	Postion	Score	Kinase	Position	Score
MAPK	9	23.06	cdk5	9	0.59	CDK5	9	0.64
SRC	23	10.1	SRC	23	0.49	PRKCZ(PKC)	45	0.52
PIKK	34	9.574	PKC	122	0.55	PRKCE(PKC)	68	0.49
PKC	45	9.564	cdk5	131	0.68	GSK3A	131	0.56
PIKK	127	9.447	GSK3	137	0.49	MAPK3	135	0.61
MAPK	131	27.97	PKC	163	0.53	CDK1	137	0.6
MAPK	134	11.29	PKC	168	0.68	PRKDC(PIKK)	171	0.58
MAPK	137	28.77	PKB(AKT)	189	0.73	CAMK2G	189	0.53
AKT	189	10.93	PKC	200	0.73	CDK1	244	0.63
MAPK	244	25.65	PKC	218	0.71	AKT1	271	0.69
MAPK	271	31.35	cdk5	244	0.65	PRKCD(PKC)	293	0.49
MAPK	317	28.27	PKC	266	0.85	MAPK3	317	0.63
PKC	364	9.677	cdk5	271	0.65	GSK3A	321	0.58
PKC	389	10.67	PKC	293	0.53	CAMK2G	358	0.6
PKC	453	9.948	PKC	301	0.67	PRKCA(PKC)	364	0.53
SRC	472	11.9	cdk5	317	0.65	CAMK2G	386	0.56
MAPK	500	29.94	PKC	352	0.6	PRKACG(PKC)	389	0.6
PKC	507	9.586	PKC	358	0.63	MAPK3	500	0.59
SRC	523	10.3	PKC	386	0.61	AKT1	602	0.71

Existing Research on Predicted Kinases:

It has been shown that viral infections activate various cellular signaling pathways, which can affect cellular function and virus replication. We ran our findings of the top kinases against the literature to investigate their possible involvement in signaling during the dengue viral replication. Below is a list of each of the kinases and what had been proven in both *in silico* and *in vitro* studies related to it:

- CDK: These cyclin-dependent kinases are proteins that perform their enzymatic activity only in the presence of an additional domain from a cyclin subunit (17). It has been shown that the bisindolylmaleimide I-HCl inhibits cyclin-dependent kinase (CDK) and, thus, inhibits DENV 2 NS5 (18). CDK was one of the Protein kinases predicted to phosphorylate NS5 (19).

- MAPK: In order to process and respond to external stimuli, the cells use Mitogen-activated protein kinases (MAPK) in their signaling pathways (20). Among the MAPKs, there are the JNK and p38 kinases, whose pathways are activated during DENV infection in macrophages. It has been clearly indicated that the inhibitors of JNK and p38 pathways reduce the viral activity (24,25). Moreover, the inhibition of JNK by quercetin was found to be useful in treating cardiovascular diseases related to vascular smooth muscle cells (VSMC) growth as it led to cell apoptosis (26). In fact, among the associated antiviral activities, the flavonoid quercetin has been reported to significantly reduce dengue DENV serotype 2 levels by 67% (21). Even though computational studies using molecular docking strongly suggested that quercetin is an inhibitor for dengue NS2B-NS3, the mechanism is still unknown (22,23).
- PKC: The fact that the inhibition of the protein kinase C (PKC) activity shows an increase in viral replication, suggests that the stimulation of this kinase's activity in the host cells is involved in suppressing the DENV replication. Recent studies suggest that NS5 phosphorylation by PKC may trigger a mechanism that modulates and affects dengue viral replication (19).
- CAMK2: The knockdown of the kinases Cdc42, p70 S6, and CaMKII, a Calcium/calmodulin-dependent protein kinase II, proves that the major reduction of DENV 2 resulted from the knockdown of the CamKII gamma gene (28).
- SRC: It has been established in previous studies that, from the SRC family of kinases (SFK), SRC, which is a proto-oncogene tyrosine-protein kinase, is, along

with the kinase, Fyn, a major member of the family involved in DENV infection (29,30,31).

- GSK: This kinase is believed to be (with CKII, proline-directed kinases, such as the CDKs, MAPKs, and glycogen synthase kinase 3 (GSK3)) a member of the CMGC kinase family, and has been found to involved in the phosphorylation of the Hepatitis C Virus NS5A (32).
- AKT: Akt (also known as protein kinase B or PKB), is a main player in multiple functional regulations in the cell. It plays a critical role in metabolism, proliferation, transcription, protein synthesis, growth, and survival (33). Quercetin, which is believed to reduce dengue virus replication, inhibits the phosphatidylinositol 3-kinase (PI3-K)/Akt pathway (34).
- PIKK: The (PIKK) family can also phosphorylate the Akt signaling pathway(33).

Multiple sequence alignment Results

The positions implicated with the top kinases need to be conserved for us to consider them relevant results. Multiple Sequence Alignment help us in our study by allowing us to visualize the most conserved regions, We used Multiple Sequence Alignment to identify the conservation among NS3 dengue virus sequences, Based on the previously mentioned confirmation that phosphoproteins are subject to more conservation in evolution compared to their non-phosphorylated proteins. We ran five multiple sequence alignments with a total of 971 NS3 DENV 2 sequences, 1533 NS3 DENV 1 sequences, 961 NS3 DENV 3 sequences, and 177 NS3 DENV 4 sequences.

In Figure 7, and the magnified view in Figure 8 we present sequence logos, which is graphical representation of alignments of a total 971 NS3 DENV 2 sequences were retrieved from the National Center for Biotechnology Information (NCBI) and were aligned and analyzed with MSA software, Jalview(35). In the sequence logos, The predicted phosphorylation sites are shown in red for Threonine THR T, and green for Serine SER S. The MSA output showed that the results appear to be conserved,

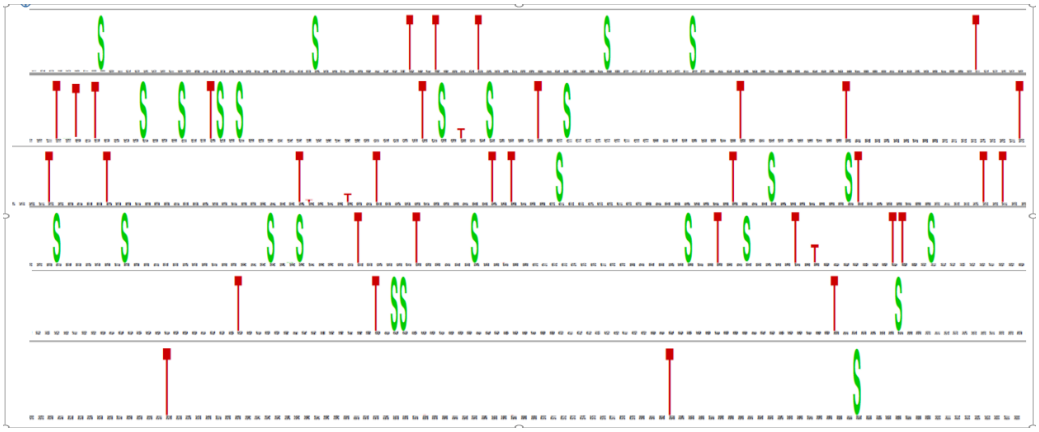

Figure 7: Multiple Sequence Alignment for 971 sequences of the Dengue Serotype 2 NS3. a graphics prepared from alignments, sites are shown in red for Threonine THR T, and green for Serine SER S.

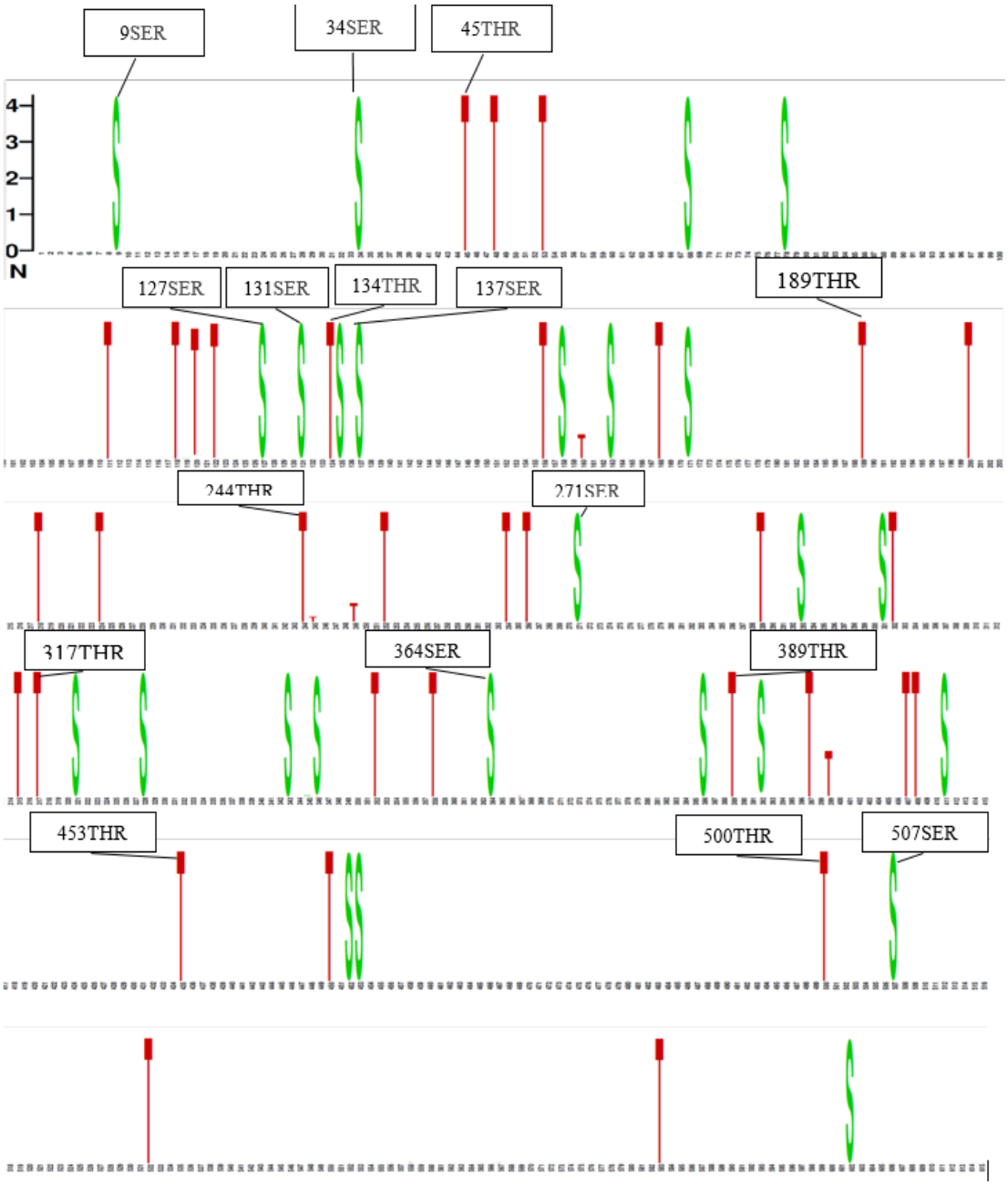

Figure 8: Magnified Multiple Sequence Alignment for 971 sequences of the Dengue Serotype 2 NS3. And the potential residues (9SER, 34SER, 45THR, 127SER, 131SER, 134STHR, 137SER, 189THR, 244THR, 271SER, 317THR, 364SER, 389THR, 453THR, 500THR, 507SER) sites are shown in red for Threonine THR T, and green for Serine SER S are visualize that they are high conversation residue.

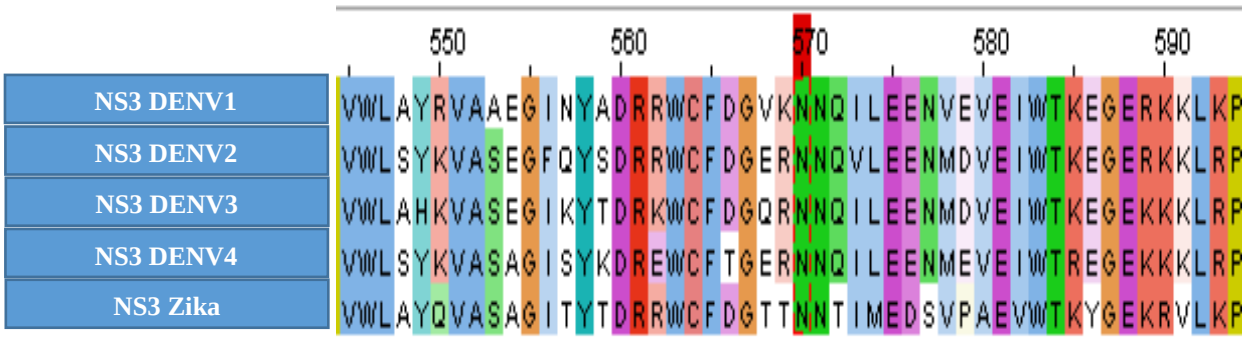

Figure 9: Multiple Sequence Alignment for NS3 4 serotypes of Dengue and Zika, shows conservation at position 570 and as well as neighboring positions in the 580-590 segments. NS3-NS5 interaction ELISA that the NS3 peptide spanning residues 566–585 disrupts NS3-NS5 interaction but not the null-peptide bearing the N570A mutation.

The C-Terminal domain of NS3 is in the fragment (180-618). Previous studies have focused on the NS3 peptide spanning the 566-585 residues and indicated its importance to disrupt the NS3-NS5 interactions while testing mutations at the N570A residue. Our multiple sequence alignment applied to DENV 1 NS3, DENV 2 NS3, DENV 3 NS3, DENV 4 NS3 And Zika NS3 shows the conservation at position 570 as well as neighboring positions in the 580-590 span. The phosphorylation site THR 583 that we predicted with our three tools belongs to the peptide of focus and is shown to be conserved (figure 9).

In Silico Molecular Visualization Of the Dengue NS3 Phosphorylation Sites and Conformational Changes.

Introduction

The NS3 protein of the Dengue virus is known for multiple enzymatic activities. It is the second largest nonstructural viral protein.

The protein function is regulated by the post translational modification that a given protein undergoes. One such post translational modification is phosphorylation, which is involved in numerous interactions, such as protein-protein interactions, intracellular localization, signal transduction, protein folding, transcription regulation, and cell cycle, including cell survival and apoptosis (38).

To understand the potential effects of amino acid serine/threonine phosphorylation in NS3 (39), we have to predict and identify specific phosphorylation sites within an *in silico* analysis, which will allow us to measure the potential effect of phosphorylation. A such *in silico* simulation requires the presence of the NS3 structure to be investigated before and after phosphorylation. The visualization of the mechanism plays a very important role in molecular dynamics, as it assists when observing the behavior of a protein to gain a better understanding.

The location of the predicted phosphorylation site within the structure of the protein is very important for protein folding, and when we look at the protein structure we can recognize whether the phosphorylation predicted site is close to an important

location on the protein, such as a functional site, an active site, or near the ATP-binding site (40). We use the NS3 structure (PDB 2VBC) prior to phosphorylation as the initial NS3 state to be compared to the post phosphorylation structures as Wild and mutant protein structure. Figure 1 shows the initial structure of NS3or the Wild NS3 protein structure. .

Molecular dynamics (MD) simulations of proteins offer the opportunity to simulate the motion of an atomic molecular assembly and to determine the net force and acceleration of each atom using Newtonian dynamics (41).

An accurate MD simulation begins with a realistic atomistic structure as a strong starting protein configuration, which is generated through experimental studies. Since many biological processes happen in aqueous solutions, solvation of the protein in MD is a very important step in mimicking the realistic environment; and its effects are crucial for identifying the final molecular conformation, as well as the binding energies and electronic properties of the molecule.

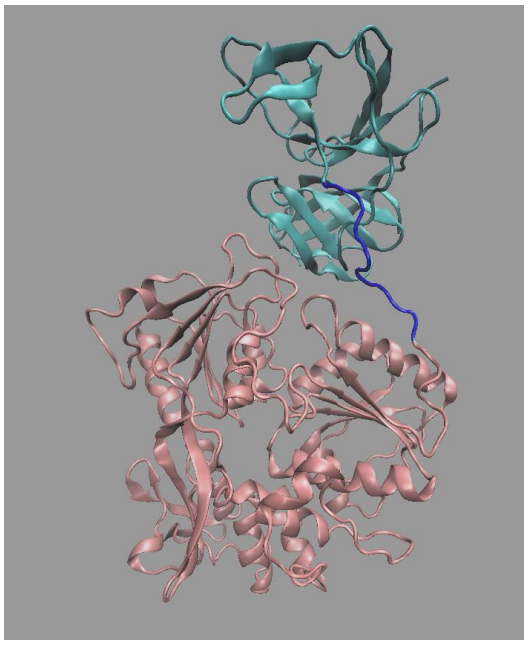

Figure 1: Crystal structure of the NS3 Protease-Helicase from Dengue Virus *(PDB entry 2VBC).* The domain (N-terminal) of the NS3 (residues 20-168) shown in cyan, the Linker (residues 169-179) shown in blue, and the Helicase Domain (C-terminal) of NS3 (NS3Hel, residues 180-618) shown in pink.

Material and Methods

The knowledge of Molecular Dynamics (MD) gives us a basic understanding of the molecular mechanisms and the structural information of the protein.

We utilized the Visual Molecular Dynamics (VMD)(42) program to visualize the 3D structure of Dengue NS3 before and after phosphorylation, to investigate the effect of the kinases predicted to phosphorylate NS3 on their associated residues and sites. VMD may be the most popular drawing method to view the overall architecture of a protein. VMD uses crystallographic structures as templates in the homology modeling process. A resulting structure from an MD simulation shows the favorable low energy conformation of the molecules after phosphorylation, which might indicate relevant changes in the form and, thus, in the function. Future simulations in this direction might help further prioritize the list of the top kinases for *in vitro* testing.

The 3D structures of DENV NS3 were collected from the protein databank (PDB ID: 2VBC) (3). Figure 1 shows the crystallized structure of the NS3 protease-helicase from the dengue virus (PDB 2VC). The ribbon represents the structure of the alpha-helix, while arrows represent the structure of the beta-sheets, and all other structures, as a tube.

The above NS3 structure is our base structure, modeled after a protein structure file containing information about all the connectivity of the protein, including residue information and name, residue types, charges, masses, the structure terms' bond, and their angles. That information is required to apply a force field to a molecular system within the simulation. This is the structure to be solvated in water (within the simulation) to mimic the aqueous solution in which the phosphorylation will be implemented in future work.

Our aim, for now, is to locate the different residues where phosphorylation is predicted, and suggest, based on the structure around them, what to prioritize in the case of an *in vitro* test.

Result and discussion

We applied GPS, NetPhos, and ScansSite to predict the potential phosphorylated sites of the NS3 protein. Our study showed some major candidate locations in the N-terminal and the C-terminal regions of NS3. The following regions appeared to constitute good targets for the phosphorylation: positions 9, 23, 34, 45, 127, 131, 134, 137, 189, 244, 277, 317, 364, 389, 453, 472, 500, 507, and 523 (Table 1).

Table 1: Potentials phosphorylation Positions

Position	ResID	Kinase
9	S, Serine, SER	MAPK, CDK5
23	Y, Tyrosine, TYR	SRC,
34	T, Threonine, THR	PIKK,
45	T, Threonine, THR	PKC, PRKCZ (PKC)
127	S, Serine, SER	PIKK
131	S, Serine, SER	MAPK, CDK5, GSK3A
134	T, Threonine, THR	MAPK
137	S, Serine, SER	MAPK, GSK3, CDK1
189	T, Threonine, THR	AKT, PKB(AKT), CAMK2G
244	T, Threonine, THR	MAPK, CDK5, CDK1
271	S, Serine, SER	MAPK, CDK5, AKT1
317	T, Threonine, THR	MAPK,CDK5,MAPK3
364	S, Serine, SER	PKC, PRKCA(PKC)
389	T, Threonine, THR	PKC, PRKACG(PKC)
453	S, Serine, SER	PKC
472	Y, Tyrosine, TYR	SRC
500	T, Threonine, THR	MAPK, MAPK3
507	T, Threonine, THR	PKC
523	Y, Tyrosine, TYR	SRC

To draw conclusions about the phosphorylation effects in the above locations, we visualized the structure, while taking into consideration whether the location of the residue of interest was internal or peripheral to the structure and whether a phosphorylation of the specific residue at that location might have a big effect on the system energy. Following, we present a sample of nine residues with the appropriate recommendation for *in vitro* testing.

Tyrosine (TYR 23 Y)

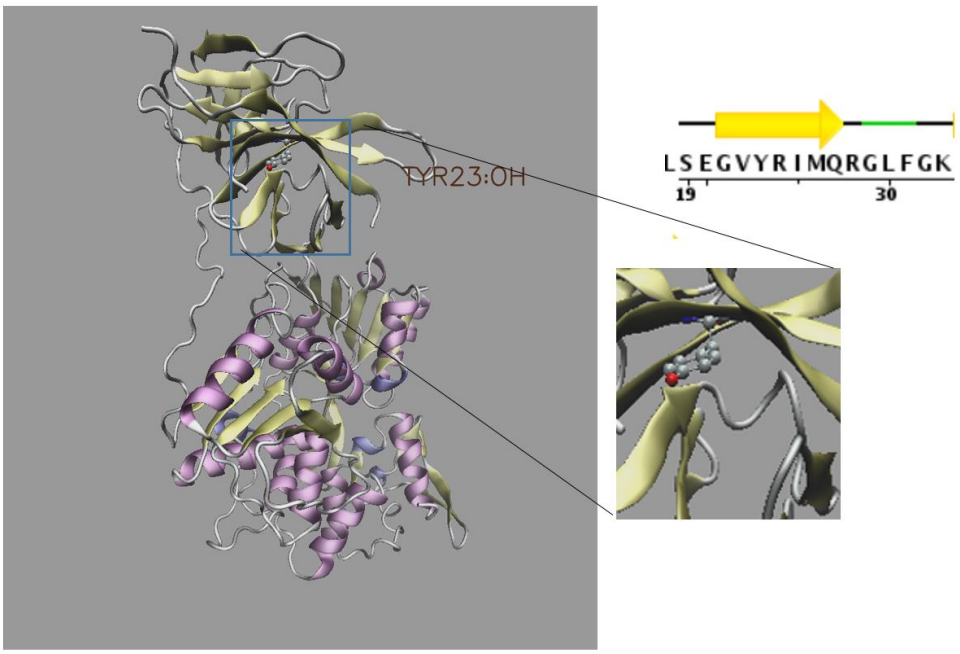

Figure 2: The residue Tyrosine (TYR 23 Y). Structure of DENV NS3 (PDB 2VBC) showing the TYR 23 residue. On the right the specific residue (secondary and tertiary) where phosphorylation can occur is given in a magnified view. On the left its position within the 3D NS3 structure is shown.

In Figure 2, above, the residue Tyrosine (TYR 23 Y), which was predicted by both GPS and NetPhos to be phosphorylated with Kinase SRC, is located on the beta-sheet in the N-terminal. The location is relatively internal, and the phosphorylation of TYR 23 Y might have a significant effect on the protein energy. This is a good candidate to prioritize for further study through MD simulations and *in vitro* testing.

Threonine (THR 34 T)

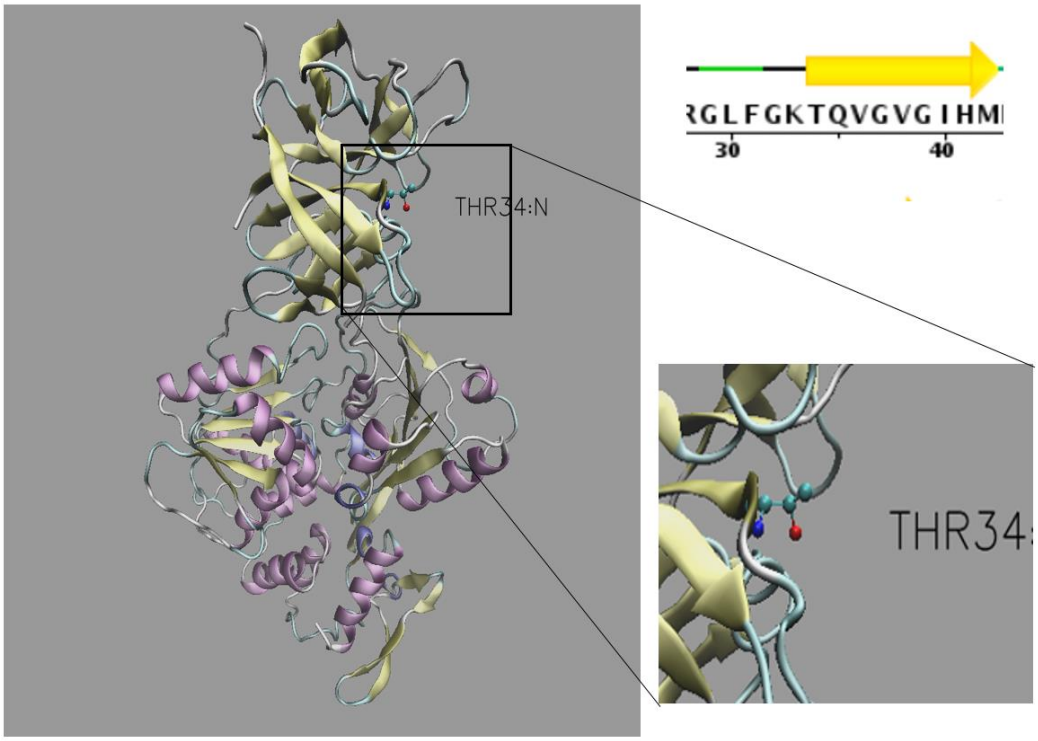

Figure 3: The residue Threonine (THR 34 T). Structure of DENV NS3 (PDB 2VBC) showing the THR 34 residue. On the right the specific residue (secondary and tertiary) where phosphorylation can occur is given in a magnified view. On the left its position within the 3D NS3 structure is shown.

In Figure 3, the residue Threonine (THR 34 T), which was predicted by GPS to only be phosphorylated with Kinase PIKK, is located on the beta-sheet in the N-terminal. The location is relatively external, and the phosphorylation of THR 34 T might not have a significant effect on the protein structure. This is not a good candidate to prioritize for further study through MD simulations and *in vitro* testing.

Threonine (THR 127 T)

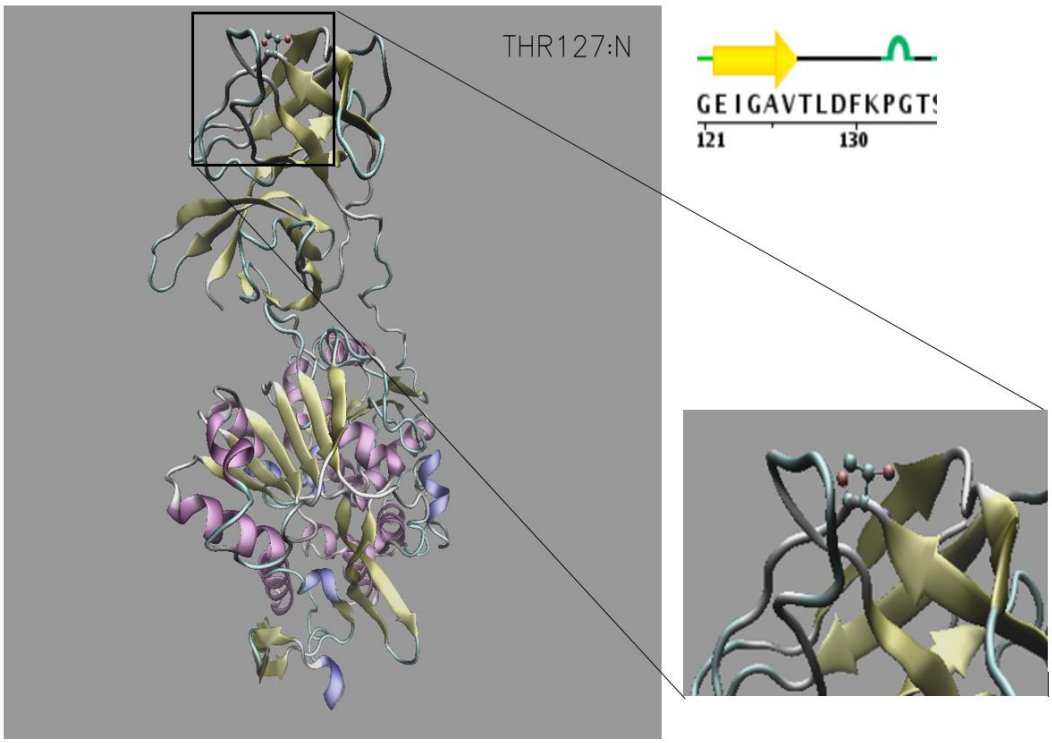

Figure 4: The residue Threonine (THR 127 T). Structure of DENV NS3 (PDB 2VBC) showing the THR 127 residue. On the right the specific residue (secondary and tertiary) where phosphorylation can occur is given in a magnified view. On the left its position within the 3D NS3 structure is shown.

In Figure 4, the residue Threonine (THR 127 T), which was predicted by GPS to only be phosphorylated with Kinase PIKK, is located on the beta-sheet in the N-terminal. The location is relatively external, and the phosphorylation of THR 127 T might not have a significant effect on the protein energy. This is not a good candidate to prioritize for further study through MD simulations and *in vitro* testing.

Threonine (THR 134 T)

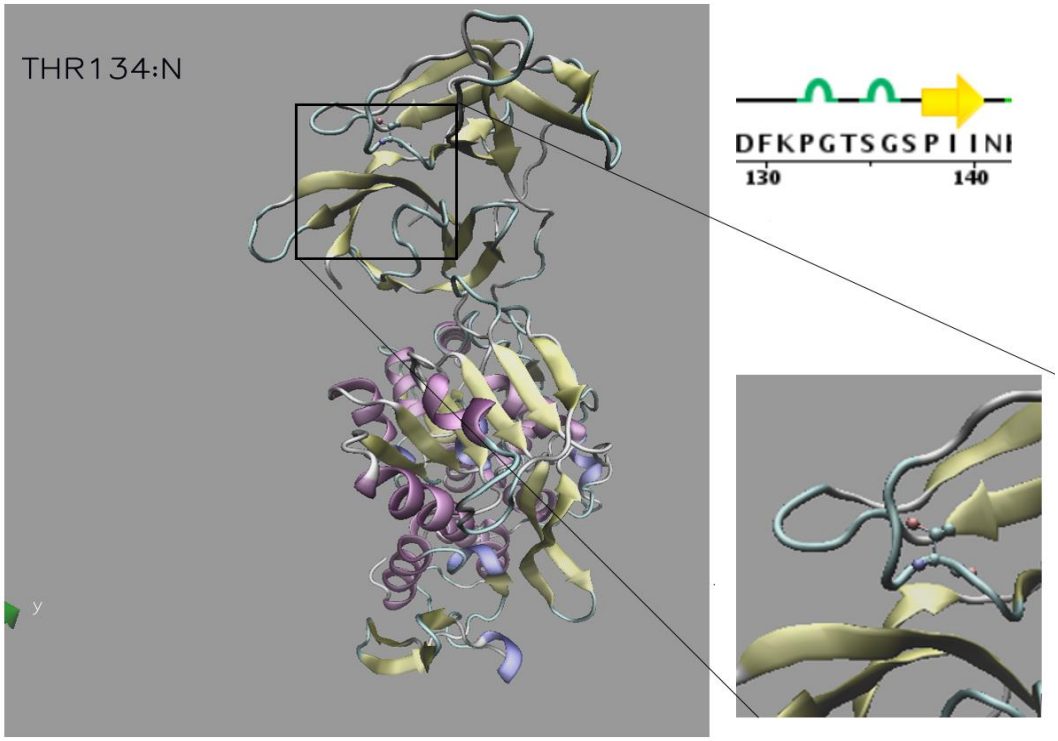

Figure 5 The residue Threonine (THR 134 T). Structure of DENV NS3 (PDB 2VBC) showing the THR 134 residue. On the right the specific residue (secondary and tertiary) where phosphorylation can occur is given in a magnified view. On the left its position within the 3D NS3 structure is shown.

In Figure 5, the residue Threonine (THR 134 T), which was predicted by GPS to only be phosphorylated with Kinase MAPK, is located on the beta-sheet in the N-terminal. The location is relatively internal, and the phosphorylation of THR 134 might have a significant effect on the protein energy. This is a good candidate to prioritize for further study through MD simulations and *in vitro* testing.

Serine (SER 137 S)

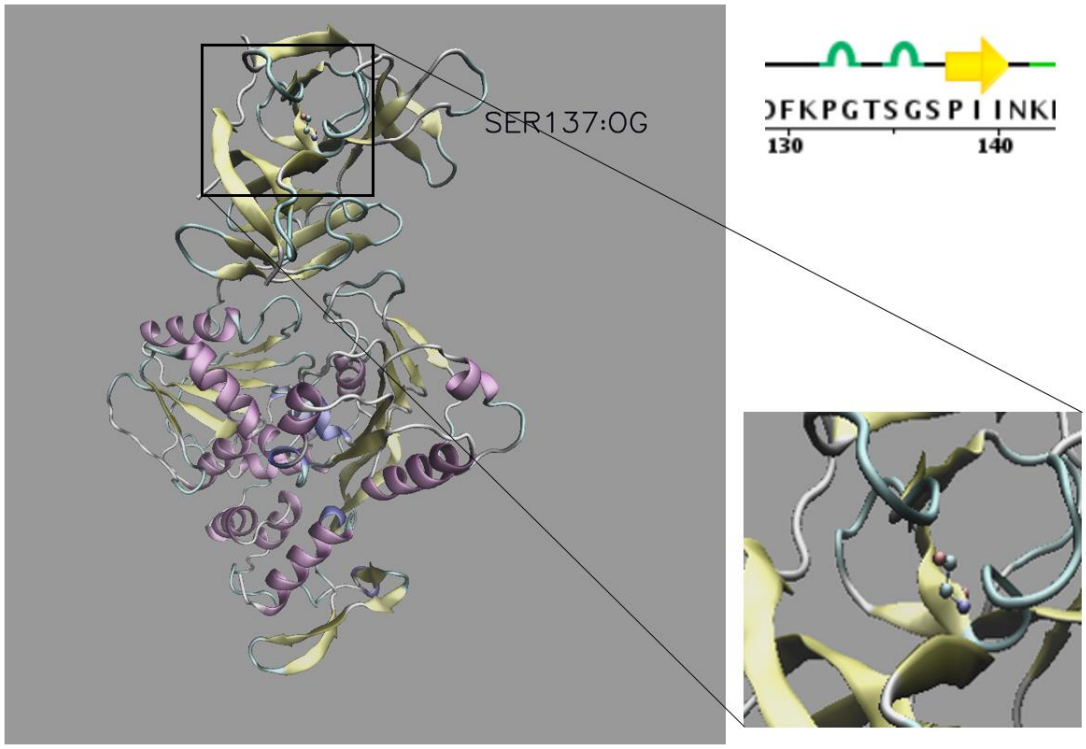

Figure 6: The residue Serine (SER 137 S). Structure of DENV NS3 (PDB 2VBC) showing the SER 137 residue. On the right the specific residue (secondary and tertiary) where phosphorylation can occur is given in a magnified view. On the left its position within the 3D NS3 structure is shown.

In Figure 7, the residue Threonine (THR 189 T), which was predicted by GPS, NetPhos, and Scansite to be phosphorylated with Kinase AKT, PKB(AKT), CAMK2G, is located on the beta-sheet in the C-terminal. The location is relatively internal, and the phosphorylation of THR 189 T might have a significant effect on the protein energy. This is a good candidate to prioritize for further study through MD simulations and *in vitro* testing.

Threonine (THR 189 T)

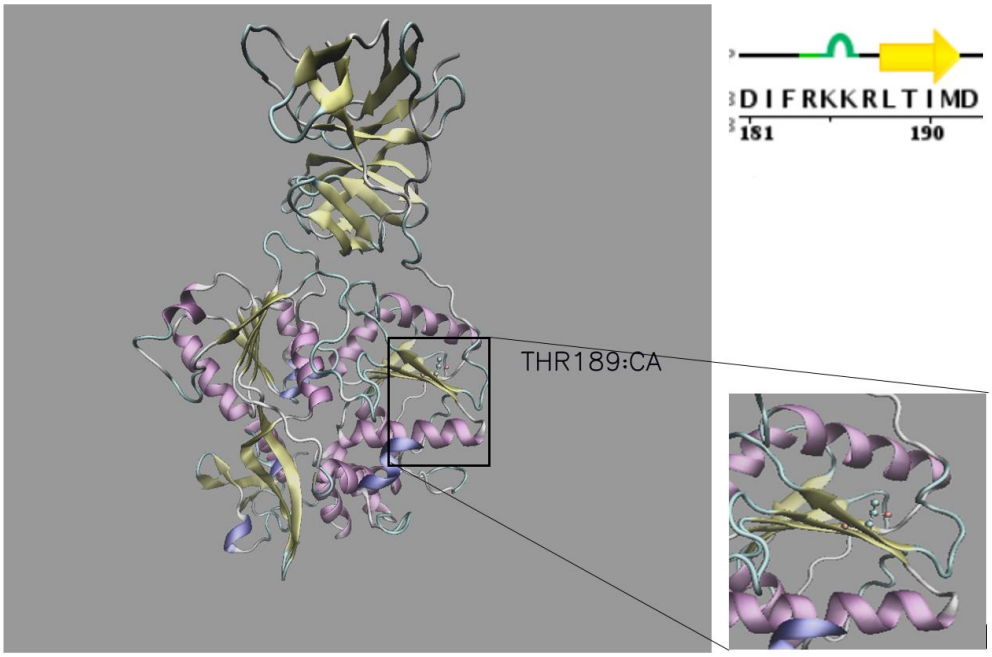

Figure 7: The residue Threonine (THR 189 T). Structure of DENV NS3 (PDB 2VBC) showing the THR 189 residue. On the right the specific residue (secondary and tertiary) where phosphorylation can occur is given in a magnified view. On the left its position within the 3D NS3 structure is shown.

In Figure 7, the residue Threonine (THR 189 T), which was predicted by GPS, NetPhos, and Scansite to be phosphorylated with Kinase AKT, PKB(AKT), CAMK2G, is located on the beta-sheet in the C-terminal. The location is relatively internal, and the phosphorylation of THR 189 T might have a significant effect on the protein energy. This is a good candidate to prioritize for further study through MD simulations and *in vitro* testing.

Threonine (THR 244 T)

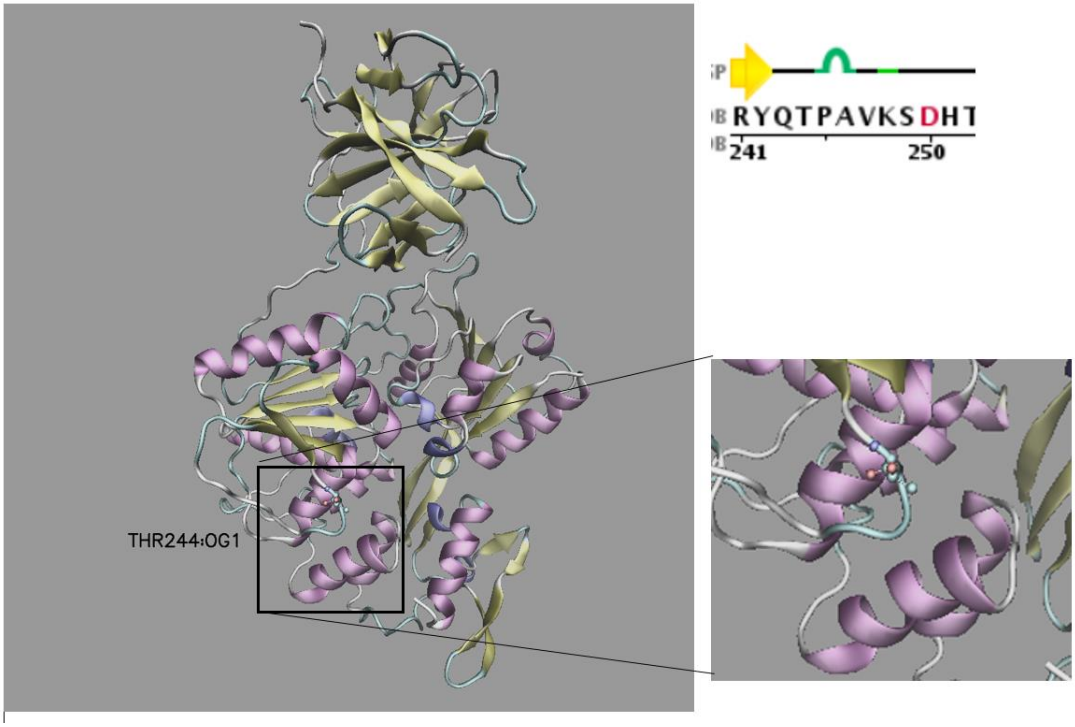

Figure 8: The residue Threonine (THR 244 T). Structure of DENV NS3 (PDB 2VBC) showing the THR 244 residue. On the right the specific residue (secondary and tertiary) where phosphorylation can occur is given in a magnified view. On the left its position within the 3D NS3 structure is shown.

In Figure 8, the residue Threonine (THR 244 T), which was predicted by GPS, NetPhos, Scansite to be phosphorylated with Kinase MAPK, CDK5, CDK1, is located on the beta-sheet in the C-terminal. The location is relatively internal, and the phosphorylation of THR 244 T might have a significant effect on the protein energy. This is a good candidate to prioritize for further study through MD simulations and *in vitro* testing.

Serine (SER 271 S)

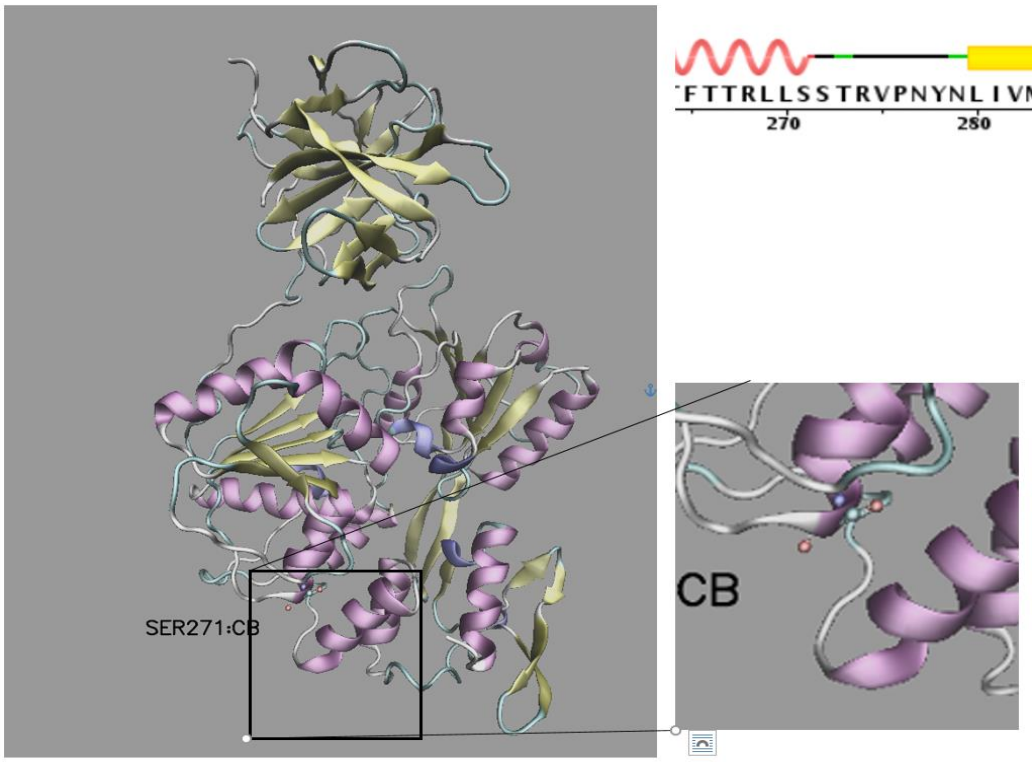

Figure 9: The residue Serine (SER 271 S), Structure of DENV NS3 (PDB 2VBC) showing the SER 271 residue. On the right the specific residue (secondary and tertiary) where phosphorylation can occur is given in a magnified view. On the left its position within the 3D NS3 structure is shown.

In Figure 9, the residue Serine (SER 271 S), which was predicted by GPS, NetPhos, Scansite to be phosphorylated with Kinase MAPK, cdk5, AKT is located on the beta-sheet in the C-terminal. The location is relatively internal and the phosphorylation of SER 271 S might have a significant effect on the protein energy. This is a good candidate to prioritize for a further study through MD simulations and *in vitro* testing

Threonine (THR 317 T)

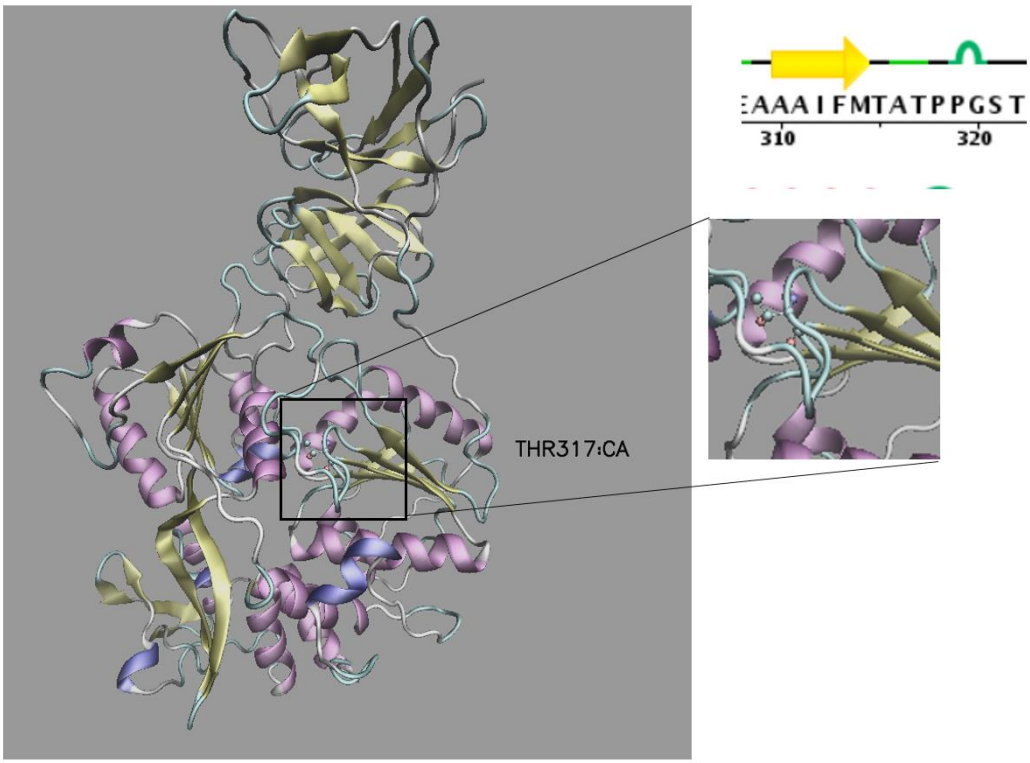

Figure 10: The residue Threonine (THR 317 T), Structure of DENV NS3 (PDB 2VBC) showing the THR 317 residue. On the right the specific residue (secondary and tertiary) where phosphorylation can occur is given in a magnified view. On the left its position within the 3D NS3 structure is shown.

In Figure 10, the residue Threonine (THR 317 T), which was predicted by GPS, NetPhos, Scansite to be phosphorylated with Kinase MAPK, cdk5, MAPK3, is located on the beta-sheet in the C-terminal. The location is relatively internal, and the phosphorylation of THR 317 T might have a significant effect on the protein energy. This is a good candidate to prioritize for further study through MD simulations and *in vitro* testing.

Threonine (THR 389 T)

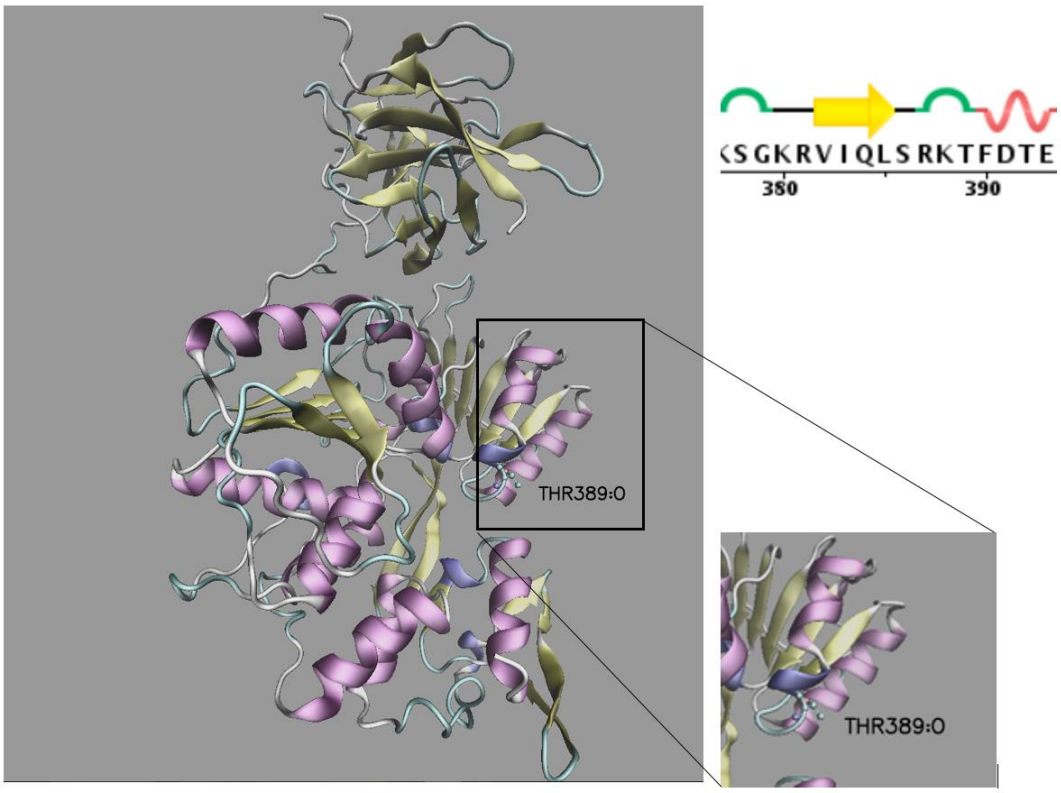

Figure 11: The residue Threonine (THR 389 T), Structure of DENV NS3 (PDB 2VBC) showing the THR 389 residue. On the right the specific residue (secondary and tertiary) where phosphorylation can occur is given in a magnified view. On the left its position within the 3D NS3 structure is shown.

In Figure 11, the residue Threonine (THR 389 T), which was predicted by GPS and ScanSite to be phosphorylated with Kinase PKC, is located on the beta-sheet in the C-terminal. The location is relatively internal, and the phosphorylation of THR 389 T might have a significant effect on the protein energy. This is a good candidate to prioritize for a further study through MD simulations and *in vitro* testing.

Solvent Accessible Surface Area (SASA)

The Solvent Accessible Surface Area calculation is a great tool that will allow us to evaluate how deep our top predicted residues are buried or peripheral to the NS3 structure. A high SASA score indicates that the residue is easily accessible and therefore its phosphorylation might have a very minimal effect on the NS3 folding. On the other hand a low SASA score informs that the residue is hardly accessible and inhibition of its phosphorylation might have a higher impact on the protein folding and consequently on the NS3 interactions –with other proteins – that are essential for the viral growth and replication. The table and the graph below show the NS3 SASA calculations. A cutoff of 0.2 was used and everything above that cutoff was considered easily accessible without perturbing the protein structure.

Table 2: Solvent Accessible Surface Area scores for the top residues.

Name	RESID position	SASA	SASA (SAR/SASMax)
Y TYR Tyrosine	23	24.93379	0.071101261
Y TYR Tyrosine	34	83.98074	0.23948
T THR Threonine	45	1.932205	0.007934
S SER Serine	127	89.43163	0.400321
S SER Serine	131	109.9993	0.493182
T THR Threonine	134	0.241526	0.000992
S SER Serine	137	0.724577	0.003249
T THR Threonine	189	17.84208	0.073258
Y TYR Tyrosine	244	69.13616	0.197149
S SER Serine	271	40.97129	0.183695
T THR Threonine	317	7.452904	0.031775
S SER Serine	364	22.59666	0.101312
T THR Threonine	389	31.74521	0.130344
S SER Serine	453	4.959524	0.022236
Y TYR Tyrosine	472	0	0.00000
T THR Threonine	500	38.85342	0.15953
S SER Serine	507	24.91948	0.111726
Y TYR Tyrosine	523	21.25426	0.060609

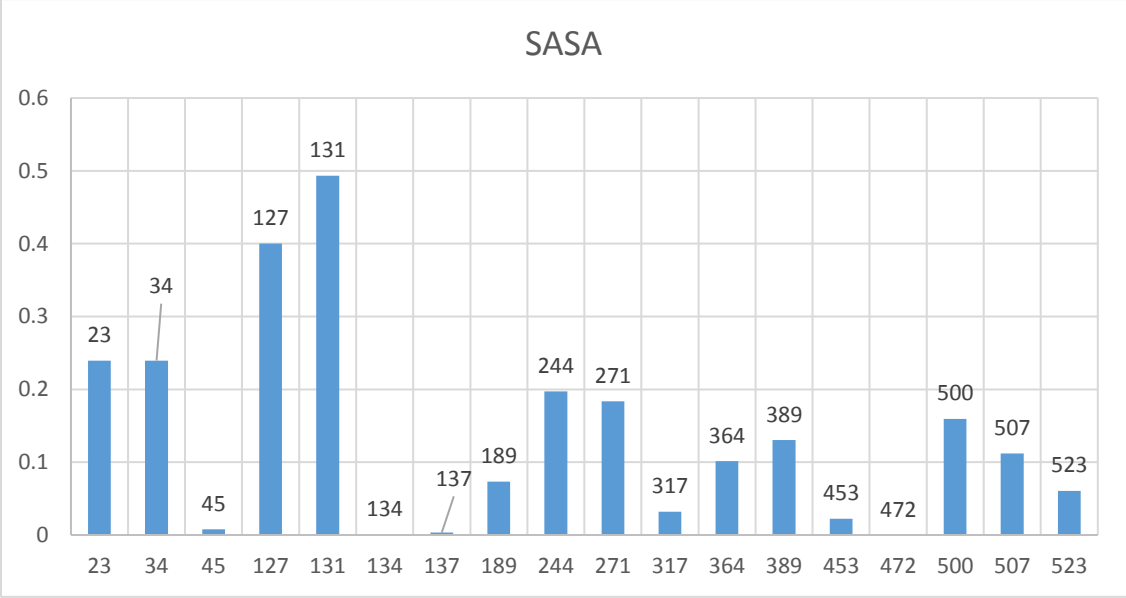

Figure 12: A graphical representation of the Solvent Accessible Surface Area scores for the top residues.

From the figure above we can see for instance that the residue Threonine (THR 317 T), which was predicted by GPS and ScanSite to be phosphorylated with Kinase PKC, and which is located on the beta-sheet in the C-terminal has a very low SASA 0.031775. The location is relatively internal, and the phosphorylation of THR 317 T might have a significant effect on the protein energy. This is a good candidate to prioritize for a further study through MD simulations and in vitro testing. Other good candidates we identified are Threonine (THR 389 T), Serine (SER 271 S), Threonine (THR 244 T), Threonine (THR 189 T), Serine (SER 137 S), and Tyrosine (TYR 23 Y). On the other hand, the residues Threonine (THR 131 T), Threonine (THR 127 T) and Threonine (THR 34 T) were relatively peripheral and are not recommended for prioritization.

Conclusion

In this study, we further investigated, using computational bioinformatics tools, the suggestion that the phosphorylation state of NS5 controls the association/disassociation of NS3 with NS5. We looked at the top kinases that are strong candidates to control the phosphorylation of NS3 aimed at describing the mechanism by which these phosphorylations control the viral replications. Inhibition of responsible phosphorylation kinases might stop the viral replications, not only in the Dengue virus, but potentially in similar viruses in the flaviviridea family such as the Zika virus.

Our three Tiers method started by finding all kinases predicted to phosphorylate DENV NS3, then selecting those with maximum scores, to finally identifying the name and position of the top kinases that were predicted by GPS, NetPhos, and Scansite to phosphorylate DENV NS3.

The identified kinases, namely CDK, MAPK, PKC, CAMK2, SRC, PIKK, GSK, and AKT, are all strong players in a still unclear mechanism of dengue viral replication and suppression as confirmed by several scientific research articles. The findings also support our suggestion that the host-mediated phosphorylation of NS3 would affect its capability to interact with NS5 and that knocking out one of the interacting proteins may inhibit viral replication.

The complexity of the networks, and the dynamics involved in the phosphorylation of important residues, whose knockout might inhibit viral replication, puts a heavy computational challenge on researchers. In our study, we suggested a way to find and prioritize target residues for phosphorylation that might have an important role in the viral

functions by visually investigating the phosphorylation sites - using visual molecular dynamics tools - and targeting our top predicted proteins whose phosphorylation happens deep inside the NS3 protein folded structure. For instance, the residue Threonine (THR 317 T), which was predicted by GPS and ScanSite to be phosphorylated with Kinase MAPK, is located on the beta-sheet in the C-terminal. The location is relatively internal, and the phosphorylation of THR 389 T might have a significant effect on the protein energy. This is a good candidate to prioritize for a further study through MD simulations and in vitro testing. Other good candidates we identified are Threonine (THR 389 T), Serine (SER 271 S), Threonine (THR 244 T), Threonine (THR 189 T), Serine (SER 137 S), and Tyrosine (TYR 23 Y). On the other hand, the residues Threonine (THR 134 T), Threonine (THR 127 T), Threonine (THR 34 T) were relatively peripherical and are not recommended for prioritization.

References

1.Natarajan S. NS3 protease from flavivirus as a target for designing antiviral inhibitors against dengue virus. Genetics and molecular biology. 2010;33(2):214-9.

2. Papageorgiou L, Loukatou S, Sofia K, Maroulis D, Vlachakis D. An updated evolutionary study of Flaviviridae NS3 helicase and NS5 RNA-dependent RNA polymerase reveals novel invariable motifs as potential pharmacological targets. Molecular BioSystems. 2016;12(7):2080-93.

3. Liefhebber JM, Hensbergen PJ, Deelder AM, Spaan WJ, van Leeuwen HC. Characterization of hepatitis C virus NS3 modifications in the context of replication. Journal of General Virology. 2010 Apr 1;91(4):1013-8.

4. Kapoor M, Zhang L, Ramachandra M, Kusukawa J, Ebner KE, Padmanabhan R. 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. 1995 Aug 11;270(32):19100-6.

5. Keck F, Ataey P, Amaya M, Bailey C, Narayanan A. Phosphorylation of single stranded RNA virus proteins and potential for novel therapeutic strategies. Viruses. 2015 Oct 12;7(10):5257-73.

6. Cieśla J, Frączyk T, Rode W. Phosphorylation of basic amino acid residues in proteins: important but easily missed. Acta Biochimica Polonica. 2011;58(2):137-48.

7. Trost B, Kusalik A. Computational prediction of eukaryotic phosphorylation sites. Bioinformatics. 2011 Nov 1;27(21):2927-35.

8. Vlachakis D, Bencurova E, Papageorgiou L, Bhide M, Kossida S. Protein phosphorylation prediction: limitations, merits and pitfalls. Journal of Molecular Biochemistry. 2015 Jul 19;4(2).

9. Dinkel H, Chica C, Via A, Gould CM, Jensen LJ, Gibson TJ, Diella F. Phospho. ELM: a database of phosphorylation sites—update 2011. Nucleic acids research. 2011 Jan 1;39(suppl 1):D261-7.

10. Gnad F, Ren S, Cox J, Olsen JV, Macek B, Oroshi M, Mann M. PHOSIDA (phosphorylation site database): management, structural and evolutionary investigation, and prediction of phosphosites. Genome biology. 2007 Nov 26;8(11):R250.

11. Hornbeck PV, Chabra I, Kornhauser JM, Skrzypek E, Zhang B. PhosphoSite: A bioinformatics resource dedicated to physiological protein phosphorylation. Proteomics. 2004 Jun 1;4(6):1551-61.

12. Xue Y, Ren J, Cao J, Liu Z. Computational prediction of post-translational modification sites in proteins. INTECH Open Access Publisher; 2011 Dec 2.

13. Blom N, Gammeltoft S, Brunak S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. Journal of molecular biology. 1999 Dec 17;294(5):1351-62.

14. Yaffe MB, Leparc GG, Lai J, Obata T, Volinia S, Cantley LC. A motif-based profile scanning approach for genome-wide prediction of signaling pathways. Nature biotechnology. 2001 Apr 1;19(4):348-53.

15. Boekhorst J, van Breukelen B, Heck AJ, Snel B. Comparative phosphoproteomics reveals evolutionary and functional conservation of phosphorylation across eukaryotes. Genome biology. 2008 Oct 1;9(10):R144.

16. Macek B, Gnad F, Soufi B, Kumar C, Olsen JV, Mijakovic I, Mann M. Phosphoproteome analysis of E. coli reveals evolutionary conservation of bacterial Ser/Thr/Tyr phosphorylation. Molecular & cellular proteomics. 2008 Feb 1;7(2):299-307.

17. Malumbres M. Cyclin-dependent kinases. Genome biology. 2014 Jun 30;15(6):122.

18. Reed KE, Gorbalenya AE, Rice CM. The NS5A/NS5 proteins of viruses from three genera of the family Flaviviridae are phosphorylated by associated serine/threonine kinases. Journal of virology. 1998 Jul 1;72(7):6199-206.

19. Noppakunmongkolchai W, Poyomtip T, Jittawuttipoka T, Luplertlop N, Sakuntabhai A, Chimnaronk S, Jirawatnotai S, Tohtong R. Inhibition of protein kinase C promotes dengue virus replication. Virology journal. 2016 Mar 1;13(1):35.

20. M Raman, W Chen and MH Cobb. Differential regulation and properties of MAPKs. Oncogene (2007) 26, 3100–3112.

21. Keivan Zandi, Boon-Teong Teoh,Sing-Sin Sam, Pooi-Fong Wong, Mohd Rais Mustafa, and Sazaly Abu Bakar, Antiviral activity of four types of bioflavonoid against dengue virus type-2. Virol J. 2011; 8: 560. Published online 2011 Dec 28. doi: 10.1186/1743-422X-8-560.

22. Padmanaban Senthilvel, Pandian Lavanya, Kalavathi Murugan Kumar, Rayapadi Swetha, Parimelzaghan Anitha, Susmita Bag, Sundaramoorthy Sarveswari, Vijayaparthasarathi Vijayakumar, Sudha Ramaiah, and Anand Anbarasu. Flavonoid from Carica papaya inhibits NS2B-NS3 protease and prevents Dengue 2 viral assembly. Bioinformation. 2013; 9(18): 889–895.

23. Lorena Ramos Freitas de Sousa, Hongmei Wu, Liliane Nebo, João Batista Fernandes, Maria Fátima das Graças Fernandes da Silva, Werner Kiefer, Manuel Kanitz, Jochen Bodem, Wibke E. Diederich. Flavonoids as noncompetitive inhibitors of Dengue virus NS2B-NS3 protease: Inhibition kinetics and docking studies. http://doi.org/10.1016/j.bmc.2014.12.015

24. Ceballos-Olvera I, Chávez-Salinas S, Medina F, Ludert JE, del Angel RM. JNK phosphorylation, induced during dengue virus infection, is important for viral infection and requires the presence of cholesterol. Virology. 2010 Jan 5;396(1):30-6.

25. Gopinathan Pillai Sreekanth a, b, Aporn Chuncharunee a, Boonyarit Cheunsuchon c,

Sansanee Noisakran d, Pa-thai Yenchitsomanus b, Thawornchai Limjindaporn a, b, *. JNK1/2 inhibitor reduces dengue virus-induced liver injury. http://doi.org/10.1016/j.antiviral.2017.02.003

26. Yoshizumi M, Tsuchiya K, Kirima K, Kyaw M, Suzaki Y, Tamaki T. Quercetin inhibits Shc- and phosphatidylinositol 3-kinase-mediated c-Jun N-terminal kinase activation by angiotensin II in cultured rat aortic smooth muscle cells. Molecular Pharmacology. 2001 Oct 1;60(4):656-65.

27. Perez-Vizcaino F, Bishop-Bailley D, Lodi F, Duarte J, Cogolludo A, Moreno L, Bosca L, Mitchell JA, Warner TD. The flavonoid quercetin induces apoptosis and inhibits JNK activation in intimal vascular smooth muscle cells. Biochem Biophys Res Commun. 2006 Aug 4;346(3):919-25. Epub 2006 Jun 9.

28. Teo CS, Chu JJ. Cellular vimentin regulates construction of dengue virus replication complexes through interaction with NS4A protein. Journal of virology. 2014 Feb 15;88(4):1897-913.

29. Kumar R, Agrawal T, Khan NA, Nakayama Y, Medigeshi GR. Identification and characterization of the role of c-terminal Src kinase in dengue virus replication. Scientific Reports. 2016;6.

30. Chu JJ, Yang PL. c-Src protein kinase inhibitors block assembly and maturation of dengue virus. Proceedings of the National Academy of Sciences. 2007 Feb 27;104(9):3520-5.

31. de Wispelaere M, LaCroix AJ, Yang PL. The small molecules AZD0530 and dasatinib inhibit dengue virus RNA replication via Fyn kinase. Journal of virology. 2013 Jul 1;87(13):7367-81.

32. Reed KE, Xu J, Rice CM. Phosphorylation of the hepatitis C virus NS5A protein in vitro and in vivo: properties of the NS5A-associated kinase. Journal of Virology. 1997 Oct 1;71(10):7187-97.

33. PI3K / Akt Signaling Resources [Internet]. Cell Signaling Technology. 2017 [cited 10 January 2017]. Available from: https://www.cellsignal.com/contents/science-pathway-research/pi3k-akt-signaling-resources/science-pathways-akt.

34. Rather A. Rafiq, Afnan Quadri, Lone A. Nazir, Kaiser Peerzada, Bashir A. Ganai, Sheikh A. Tasduq, A Potent Inhibitor of Phosphoinositide 3-Kinase (PI3K) and Mitogen Activated Protein (MAP) Kinase Signalling, Quercetin (3, 3', 4', 5, 7-Pentahydroxyflavone) Promotes Cell Death in Ultraviolet (UV)-B-Irradiated B16F10 Melanoma Cells, Published: July 6, 2015,http://dx.doi.org/10.1371/journal.pone.0131253 .

35. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009 May 1;25(9):1189-91.

36. Moon Y. F. Tay, Wuan Geok Saw, Yongqian Zhao, Kitti W. K. Chan, Daljit Singh, Yuwen Chong, Jade K. Forwood, Eng Eong Ooi, Gerhard Grüber,Julien Lescar, Dahai Luo,and Subhash G. Vasudevan, The C-terminal 50 Amino Acid Residues of Dengue NS3 Protein Are Important for NS3-NS5 Interaction and Viral Replication, J Biol Chem. 2015 Jan 23; 290(4): 2379–2394.

37.Giovanna Barba-Spaeth, Wanwisa Dejnirattisai, Alexander Rouvinski, Marie-Christine Vaney, Iris Medits, Arvind Sharma, Etienne Simon-Lorière, Anavaj Sakuntabhai, Van-Mai Cao-Lormeau, Ahmed Haouz, Patrick England, Karin Stiasny, Juthathip Mongkolsapaya, Franz X. Heinz, Gavin

R. Screaton & Félix A. Rey , Structural basis of potent Zika–dengue virus antibody cross-neutralization, Nature 536,48–53(04 August 2016) doi:10.1038/nature18938

38. Sudhakaran Prabakaran, Guy Lippens, Hanno Steen and Jeremy Gunawardena, Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding, WIREs Syst Biol Med2012. doi: 10.1002/wsbm.1185

39. Erbin He, Guanghui Yan, Jian Zhang, Jun Wang, and Wenfei Li, Effects of phosphorylation on the intrinsic propensity of backbone conformations of serine/threonine, J Biol Phys. 2016 Mar; 42(2): 247–258. Published online 2016 Jan 12. doi: 10.1007/s10867-015-9405-0.

40. Maximova T, Moffatt R, Ma B, Nussinov R, Shehu A. Principles and overview of sampling methods for modeling macromolecular structure and dynamics. PLoS Comput Biol. 2016 Apr 28;12(4):e1004619.

41. Luo D, Xu T, Hunke C, Grüber G, Vasudevan SG, Lescar J. Crystal structure of the NS3 protease-helicase from dengue virus. Journal of virology. 2008 Jan 1;82(1):173-83.

42. William Humphrey, Andrew Dalke, Klaus Schulten, VMD: Visual molecular dynamics, Journal of Molecular Graphics, Volume 14, Issue 1, February 1996, Pages 33-38

Chapter 4: Conclusion and Future Work

Conclusion

The first recorded case of probable dengue infection goes back to 1779, 247 years ago. More than 1,004 books and reports have been published on Dengue, in addition to 23,945 full text journal and articles, 17,000 scientific and medical abstracts/citations, 26,100 DNA and RNA sequences, and 23980 protein sequences. And yet, even with all the *in silico* and *in vitro* work, there is still no specific antiviral treatment currently available for dengue fever (1).

It is very important that we better understand all aspects related to the dengue viral production, replication, and the mechanisms the virus follows to evade and counteract the immune system(2). We identified a collection of proteins whose interactions are similar to the NS3 and, based on that, identified potential targets to inhibit the viral activities. We propose a multifaceted computational approach to better understand the viral activities by studying the protein interactions and identifying several major protein fragments significantly similar to NS3. This could then be used to determine the protein-protein interactions, clarify the functionality of the associated parts of dengue, and project the findings of protein interactions to the Dengue virus(4).

Dengue virus NS3 protein interaction returned a high number of significant BLAST hits, which led us to further investigate the network of interactions. We found several interacting proteins, whose connections can lead us to a better understanding of NS3 protein interactions with infected cells. NS3 makes viral proteins that are exported in lipid membrane droplets to form new viral particles. The viral particles then exit the cell to infect the whole body. Knocking out one of the interacting proteins that we found may inhibit viral replication.

When we further investigated, through the use of computational bioinformatics tools, the suggestion that the phosphorylation state of NS5 controls the association/disassociation of NS3 with NS5, we looked at the top kinases that were strong candidates to control the phosphorylation of NS3, with the aim of describing the mechanism by which the phosphorylation controls viral replications. Inhibition of responsible phosphorylation kinases might stop viral replications, not only in Dengue virus, but, potentially, in similar viruses in the flaviviridea family, such as Zika.

We conclude our study by suggesting a way to find and prioritize target residues for phosphorylation that might have an important role in viral functions. We used phosphorylation simulations, along with visual molecular dynamic modeling, to identify protein structures that favor a low energy conformation of the molecule after phosphorylation. The relevant changes in the form might indicate a relevant change in the function. The complexity of the networks and the dynamics involved in the phosphorylation of important residues, whose knockout might inhibit viral replication, puts

a heavy computational challenge on researchers. Our study offers an alternative way to identify significant candidate proteins for *in vitro* testing.

This is the first research to study the phosphorylation of Dengue NS3 with such depth, as we aim to push researchers closer to discovering a drug for the Dengue virus. The research findings can be summarized as follows:

- Chapter 1: The literature review constituted the main trigger of our focus on the host-mediated NS3 phosphorylation whose knockout might play a crucial role in the inhibition of viral replication.
- Chapter 2: Dengue virus (DENV) NS3 protein interaction returned a high number of significant BLAST hits that led us to further investigate the network of interactions. We found several interacting proteins whose connections can lead to a better understanding of NS3 protein interactions with infected cells including Zinc finger and CamKII. Knocking out one of the interacting proteins that we found may inhibit viral replication. These findings open the door for further investigation to identify key proteins that will cause the inhibition.
- Chapter 3: We further investigated, through the use of computational bioinformatics tools, the suggestion that the phosphorylation state of NS5 controls the association/disassociation of NS3 with NS5. We looked at the top kinases that are strong candidates to control the phosphorylation of NS3 aimed at describing the mechanism by which these phosphorylations control the viral replications. Inhibition of responsible phosphorylation kinases might stop the viral replications, not only in the Dengue virus, but potentially in similar viruses in the flaviviridea

family such as the Zika virus. Our three Tiers method identified kinases, namely CDK, MAPK, PKC, CAMK2, SRC, PIKK, GSK, and AKT, that are top candidates to phosphorylate NS3 and it turned out that they are all strong players in dengue viral replication and suppression as confirmed by our research of several scientific articles. The findings also support our suggestion that the host-mediated phosphorylation of NS3 would affect its capability to interact with NS5 and that knocking out one of the interacting proteins may inhibit viral replication.

- Chapter 4: The complexity of the networks, and the dynamics involved in the phosphorylation of important residues, whose knockout might inhibit viral replication, puts a heavy computational challenge on researchers. In our study, we suggested a way to find and prioritize target residues for phosphorylation that might have an important role in the viral functions by visually investigating the phosphorylation sites - using visual molecular dynamics tools - and targeting our top predicted proteins whose phosphorylation happens deep inside the NS3 protein folded structure. For instance, the residue Threonine (THR 317 T), which was predicted by GPS and ScanSite to be phosphorylated with Kinase MAPK, is located on the beta-sheet in the C-terminal. The location is relatively internal, and the phosphorylation of THR 389 T might have a significant effect on the protein energy. This is a good candidate to prioritize for a further study through MD simulations and in vitro testing. Other good candidates we identified are Threonine (THR 389 T), Serine (SER 271 S), Threonine (THR 244 T), Threonine (THR 189 T), Serine (SER 137 S), and Tyrosine (TYR 23 Y). On the other hand,

the residues Threonine (THR 134 T), Threonine (THR 127 T), Threonine (THR 34 T) were relatively peripherical and are not recommended for prioritization.

Future Work

We believe that after identifying a strong cohort of kinases that phosphorylate NS3, but before getting to *in vitro* testing, a deeper investigation of the *in silico* description of the NS3 residues might provide a better prioritization of the target proteins for *in vitro* testing, or provide enough evidence about the mechanism to bypass the *in vitro* confirmation. This investigation is to be completed by running multiple simulations, checking the force fields and energies, and by adding additional components to analyze the phosphorylation effects, such as measuring the solvate-accessible surface area (SASA) around each site of interest (4).

References

1. National Center for Biotechnology Information, U.S. National Library of Medicine. Dengue. 2017 [cited 2017 Jan 20]. Accessed from https://www.ncbi.nlm.nih.gov/gquery/?term=dengue
2. Rath D, Amlinger L, Rath A, Lundgren M. The CRISPR-Cas immune system: biology, mechanisms and applications. Biochimie. 2015 Oct 31;117:119-28.
3. Jeffrey A. Ubersax & James E. Ferrell Jr, Mechanisms of specificity in protein phosphorylation, Nature Reviews Molecular Cell Biology 8, 530-541 (July 2007) | doi:10.1038/nrm2203.
4. Folador EL, Hassan SS, Lemke N, Barh D, Silva A, Ferreira RS, Azevedo V. An improved interolog mapping-based computational prediction of protein–protein interactions with increased network coverage. Integrative Biology. 2014;6(11):1080-7.
5. Durham E, Dorr B, Woetzel N, Staritzbichler R, Meiler J. Solvent accessible surface area approximations for rapid and accurate protein structure prediction. Journal of molecular modeling. 2009 Sep 1;15(9):1093-108.

Biography

Lamya Alomair grew up in Riyadh Saudi Ariba. Received her Bachelor computer & Information's Sciences from King Saud University, she learned all about the art of analysis, managed to view data from different perspectives and learned about the transition of data from information to knowledge to assist physicians in making critical decisions based on accurate data. She started as a Programmer Analyst and then, I moved on to be a Database Specialist where I broadened my horizons about Oracle as a Database Administrator. She was a part of the first group to admitted to the First Master program in Health Informatics in Saudi Arabia back in 2005 at King Saud Bin Abdulaziz University for Health Sciences. She is a PhD student in the Bioinformatics and Computational Biology program, School of Systems Biology, George Mason University since Fall 2010.