Zika Virus: Discovering Effective Protease Inhibitors via Template-Based Modeling

Yongseok Choi

Department of Chemistry / School of Materials Science and Engineering, Gwangju Institute of Science and Technology, 123 Cheomdangwagi-ro, Buk-gu, Gwangju 61005, Korea.

Received Date; Accepted Date (to be inserted by the publisher after your manuscript is accepted)

Abstract: The Zika virus, which is a member of the flavivirus genus, poses a serious threat to humanity because there is no vaccine or cure. Zika is suspected to cause microcephaly, and it is rapidly spreading throughout parts of Brazil. Surprisingly, there are no known protein structures for the virus which are essential for drug and vaccine development. This paper investigates the Zika virus's nonstructural proteins with template-based modeling by using GalaxyTBM/Refine/SC. GalaxyDock was used to examine the effectiveness of various known serine protease inhibitors in inhibiting the Zika viral protease. In testing five inhibitors, Kunitz soybean trypsin inhibitor showed the strongest binding affinity (-10.082 kcal/mol). This paper provides a rudimentary foundation for further drug discovery research regarding the Zika virus.

Keywords: Zika virus (ZIKV), Template-based modeling, Kunitz soybean trypsin inhibitor (STI)

Introduction

Zika virus (ZIKV), which is part of the flavivirus genus, is an arthropod-borne virus that includes well-known viruses, such as West Nile virus and Dengue virus (DENV).[1] It has been linked to microcephaly and is receiving more attention since mid-2015 due to a large number of microcephaly occurrences in parts of Brazil.[2] Flavivirus virions are approximately 500 Å in diameter and are surrounded by 180 copies of two glycoproteins. These proteins cover a host-derived lipid bilayer which encapsulates the virus capsid protein along with its single, positive-strand RNA genome.[3] ZIKV genome encodes a single polyprotein that is cut into at least ten structural and nonstructural (NS) proteins co- and posttranslatinally where C terminus region codes seven nonstructural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, NS5).[4]

NS3 is an enzyme central to flavivirus replication and polyprotein processing as it is composed of two distinct functional domains: the protease domain (NS3Pro, requires cofactor NS2B to be fully active) and the helicase domain (NS3Hel).[5] In the past, HIV protease has been the target for antiviral drug design, and numerous inhibitors have been successfully developed that are currently used in treating AIDS.[6] NS3 may therefore be a good target for drug design, and the structural information of NS3 is indispensable to this effort. Regrettably, there are no reported crystal structures for any of the ZIKV nonstructural proteins. However, there are numerous known crystal structures for other flavivirus nonstructural proteins, and there are even structures already bound to inhibitors. With the complete sequence of the ZIKV RNA, it is possible to model the ZIKV nonstructural proteins by template-based modeling, and it is also possible to compare the inhibitor affinity between the modeled ZIKV structure and the known flavivirus structure already bound to the inhibitor. Aprotinin, also known as bovine pancreatic trypsin inhibitor, will be investigated in this paper to compare the binding affinity for ZIKV NS3Pro and DENV-3 protease (PDB: 3U1J), which has aprotinin already bound to it.[7]

Unfortunately, in spite of its binding effectiveness, aprotinin was banned through out the world in 2008 due to serious side-effects, including death.[8] In an effort to find safer alternatives, four typical serine protease inhibitors were surveyed including 4-(2-aminoethyl) benzenesulfonyl fluoride (AEBSF), Kunitz soybean trypsin inhibitor (STI), phenylmethylsulfonyl fluoride (PMSF) and leupeptin. Out of the four inhibitors, Kunitz STI showed the strongest binding affinity to ZIKV NS3Pro.

Theory and Computational Method

The complete genome of Zika virus strain MR766 has been reported (Genbank: AY632535.2). The nonstructural protein coding region was separated from the rest and was further cut into seven nonstructural protein coding sequences using the proposed cleavage sites.[9] Each of the seven sequences were given as input to GalaxyTBM science application in EDISON webpage (all other tools are also available on the EDISON webpage at https://www.edison.re.kr).[10] As the version of GalaxyTBM in EDISON

does not include the final refinement of protein structures, GalaxyTBM outputs were further refined by GalaxyRefine.[11]

In order to investigate ZIKV NS3Pro (residues 1 to 167), the best refined structure (the first entry in the GalaxyRefine output) was further optimized using GalaxySC.[11] GalaxyDock was used to compare the binding energies for various inhibitors.[12] To compare the energy difference for aprotinin binding, ZIKV NS3Pro and DENV-3 protease were given as input structures along with aprotinin that was originally bound to DENV-3 protease. Only residues 14 through 16 of aprotinin were used due to ligand size limitation of the GalaxyDock application. Whereas DENV-3 protease already had DENV-3 NS2B cofactor present on the crystal structure, ZIKV NS2B was not attached to ZIKV NS3Pro as docking is not accurate for long ligands.

The binding energy information obtained above was used as a guideline to compare the binding affinities of other serine protease inhibitors. AEBSF and PMSF were prepared using Maestro v10.4 as there are no tools available in EDISON to model these structures.[13] Kunitz STI and leupeptin were obtained from PDB structures (1AVU and 1TL9 respectively). Only residues interacting directly with the active site (residues 62 to 64) were used for Kunitz STI as GalaxyDock has a limit on the size of the docking substance.[14] The four inhibitors were given as input to GalaxyDock along with ZIKV NS3Pro model.

Results and Discussion

Structures for the six nonstructural protein models are given in Figure 1. NS2A was excluded as it was modeled as a linear molecule which is an inaccurate representation of a transmembrane protein.[4] Although linear, NS2B showed a ß-hairpin motif. The NS3Hel domain and the NS3Pro domain are clearly present in the model and are connected by a linker region. Enlarged NS3Pro model (Figure 2) shows the conserved His51, Asp75, and Ser135 residues which form the catalytic triad, showing that ZIKV NS3Pro is part of the trypsin serine protease superfamily (His 51 is incorrectly aligned). Figure 3 shows DENV-3 protease bound to aprotinin as present in the crystal structure. Gly133, Thr134, and Ser135 form the oxyanion hole and Asp129, Gly151, and Gly153 form the S1 binding pocket similar to other serine proteases.

Docking of aprotinin to ZIKV NS3Pro model was successful with aprotinin Lys15 being correctly inserted into the S1 binding pocket and the Lys15 oxygen being properly aligned for stabilization by the oxyanion hole (insertion of the side-chain into the S1 pocket and the proper alignment of P1 oxygen for oxyanion stabilization was the primary consideration for selecting the structural output provided by GalaxyDock). However, the ZIKV NS3Pro Ser135 side-chain oxygen and Gly133 nitrogen is not in the

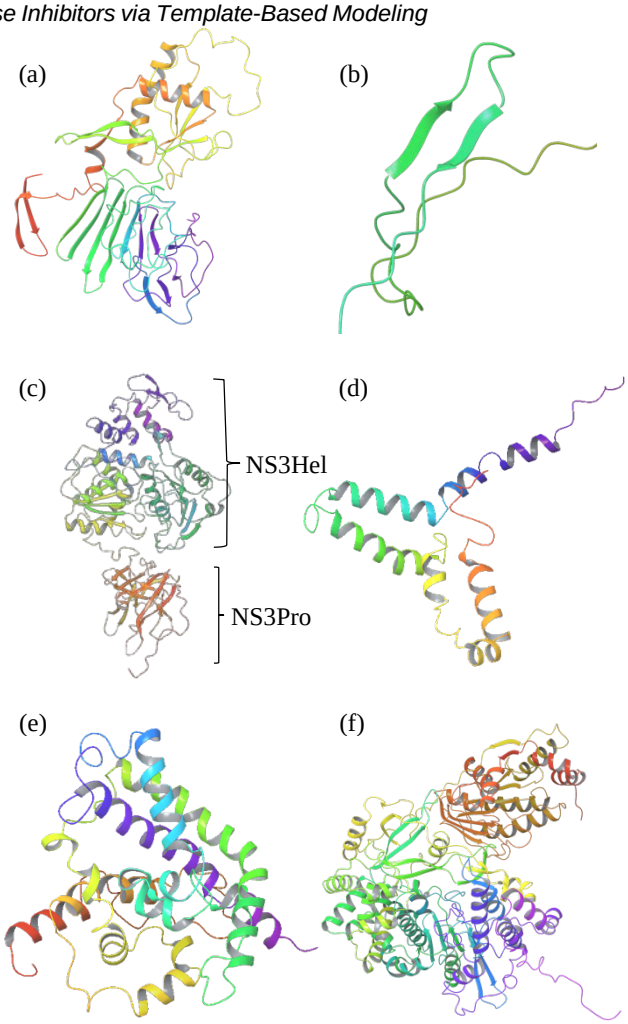

Figure 1. Model structures for six nonstructural proteins of ZIKV. (a) ZIKV NS1; (b) ZIKV NS2B, ß-hairpin region; (c) ZIKV NS3, two functional domains are clearly modeled; (d) ZIKV NS4A; (e) ZIKV NS4B; (f) ZIKV NS5. Protein secondary structure colored by rainbow spectrum from N-terminus (red) to C-terminus (purple). Figure generated by Maestro v10.4.

optimal position to participate in a chemical reaction. The AutoDock binding energy provided by GalaxyDock is -8.508 kcal/mol for DENV-3 protease and -8.114 kcal/mol for ZIKV NS3Pro model, resulting in a 0.394 kcal/mol difference.

NS2B, bound to DENV-3 protease and missing from ZIKV NS3Pro model, acts as a cofactor by stabilizing the protease, fulfilling a chaperone-like role and by being a part of the active site interacting directly with the substrate through the ß-hairpin motif.[5,7] However, even without NS2B, ZIKV NS3Pro interaction with aprotinin is comparable to that of DENV-3 protease and aprotinin, suggesting that aprotinin could bind more tightly with ZIKV NS3Pro when it is activated by NS2B. However, further investigation of the rest of the protease-inhibitor interaction must be carried

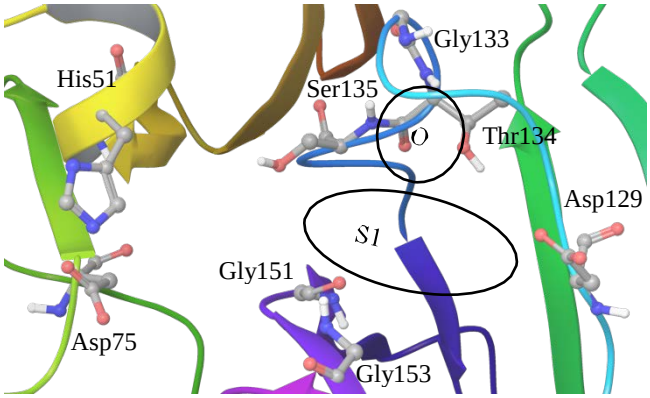

Figure 2. Catalytic triad, the oxyanion hole and the S1 pocket of modeled ZIKV NS3Pro. His51 side-chain, Gly133 nitrogen, and Ser135 side-chain is not properly aligned to participate in a chemical reaction. Carbon, oxygen, nitrogen, and hydrogen are colored grey, red, blue, and white respectively. Protein secondary structure coloring is the same as Figure 1. S1 pocket and oxyanion hole is represented by an ellipse with S1 and O written inside respectively. Figure generated by Maestro v10.4.

out with crystal structures to draw concrete conclusions.

The AutoDock binding energy provided by GalaxyDock for AEBSF, Kunitz STI, PMSF and leupeptin were -6.959, -10.082, -5.404, -7.888 kcal/mol respectively. In a past study, AEBSF was shown to exhibit 20% to 30% inhibition when compared to aprotinin for DENV-2 viral protease.[15] The binding energy difference between aprotinin and AEBSF for ZIKV NS3Pro model also shows that AEBSF will be less effective when compared to aprotinin as well, in agreement with the past study.[15] In the same paper, PMSF was shown to have no inhibitory effect on DENV-2 viral protease. This is also in agreement with the results presented in this paper as PMSF shows the lowest binding affinity to ZIKV NS3Pro model. The binding affinity difference can be attributed to the existence of an amino group. Whereas AEBSF has an amino group at the para-position that can be stabilized by the S1 pocket, PMSF has no such group. The stabilization of AEBSF due to the presence of the amino group shows the importance of the S1 pocket for binding to NS3Pro.

Leupeptin showed similar binding affinity when compared to aprotinin and Kunitz STI showed the highest binding affinity. This is in contrast with the above study[15] of DENV-2 where both inhibitors showed no inhibitory effect on the viral protease. Although further study is needed, Kunitz STI derived from soybean could be an effective drug against ZIKV NS3Pro with minimum side-effects as it is a natural product contained in soybeans. Kunitz STI bound to ZIKV NS3Pro model is shown in Figure 4.

Conclusion

The seriousness of recent microcephaly cases and other

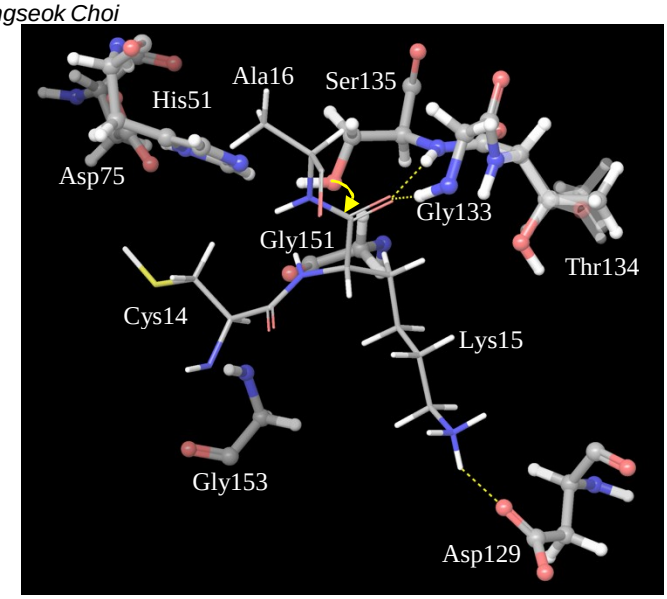

Figure 3. Aprotinin (residues 14 to 16) bound to DENV-3 protease. Aprotinin is correctly positioned within the S1 pocket and the aprotinin Lys15 oxygen is aligned to be stabilized by DENV-3 Gly133 and Ser135 during a chemical reaction. Yellow arrow indicates the nucleophilic attack of Ser135 to Lys15. H-bonds between the ligand and the enzyme are shown by yellow dotted lines. Follows the same coloring scheme as Figure 2. The enzyme is shown in ball & stick model and the ligand is shown in thin tube model. Blurred DENV-3 protease Thr134 is directly from the crystal structure. Figure generated by Maestro v10.4.

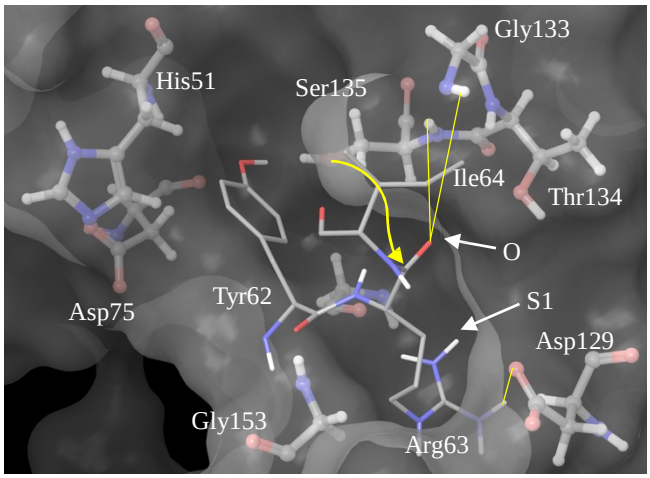

Figure 4. Kunitz STI (residues 62 to 64) bound to NS3Pro model. Surface of the enzyme is shown in light grey. Coloring scheme and modeling follows Figure 3. Although ZIKV NS3Pro His51 side-chain, Ser135 oxygen and Gly133 nitrogen is not properly aligned, STI Arg63 is correctly inserted into the S1 pocket, and the Arg63 oxygen is correctly aligned to be stabilized by Gly133 and Ser135. Yellow arrow indicates the nucleophilic attack of Ser135 to Arg63. Yellow lines indicate possible H-bonds. ZIKV NS3Pro Gly151 is not mentioned in the figure to avoid cluttering. S1 and O represents ZIKV NS3Pro S1 pocket, and the oxyanion hole respectively. Figure generated by Maestro v10.4.

neurological disorders reported in Brazil has been reconfirmed when the World Health Organization recently issued a statement of Public Health Emergency of International Concern.[16] This paper provides the structural models for the nonstructural proteins of ZIKV and takes a first look at the ZIKV NS3Pro model and its interaction with various serine protease inhibitors in an effort to pave the way for further research related to inhibiting ZIKV NS3Pro.

At this point, it must be stressed that all the results in this paper are based on structures that were obtained by template-based modeling. However, the accuracy of the Galaxy tools has been proven, and the accurate modeling of the bipartite ZIKV NS3 and the active site of ZIKV NS3Pro is yet another example of how computational methods are making meaningful progress to complement experimental results.[17] In vitro and in vivo validation of the the inhibitory effect of Kunitz STI is needed and the structures of the six nonstructural proteins presented here also require experimental validation. Further investigation for the effectiveness of Kunitz STI could be carried out by investigating other binding pockets on the enzyme.

All data and figures used in this paper are open for access at http://bit.ly/1QRr7uc.

Acknowledgments. This work has been supported by the project EDISON (EDucation-research Integration through Simulation On the Net), Chemistry and by the Molecular Modeling Laboratory in School of Materials Science and Engineering at Gwangju Institute of Science and Technology (GIST). Special thanks goes out to Ellis Lee at GIST for all the support he has provided.

References

1. M. D. Fernandez-Garcia, M. Mazzon, M. Jacobs, A. Amara, *Cell Host Microbe* **2009**, *5* (4), 318–328 DOI: 10.1016/j.chom.2009.04.001.
2. M. W. Report, R. B. Martines, J. Bhatnagar, M. K. Keating, L. Silva-flannery, J. Gary, C. Goldsmith, G. Hale, J. Ritter, D. Rollin et al., Notes from the Field **2016**, *65* (6), 2015–2016.
3. S. Mukhopadhyay, R. J. Kuhn, M. G. Rossmann, *Nat. Rev. Microbiol.* **2005**, *3* (1), 13–22 DOI: 10.1038/nrmicro1067.
4. B. D. Lindenbach, C. M. Rice, *Adv. Virus Res.* **2003**, *59* (FEBRUARY 2003), 23–61 DOI: 10.1016/S0065-3527(03)59002-9.
5. M. Bollati, K. Alvarez, R. Assenberg, C. Baronti, B. Canard, S. Cook, B. Coutard, E. Decroly, X. de Lamballerie, E. A. Gould et al., *Antiviral Res.* **2010**, *87* (2), 125–148 DOI: 10.1016/j.antiviral.2009.11.009.
6. L. Menéndez-Arias, *Antiviral Res.* **2010**, *85* (1), 210–231 DOI: 10.1016/j.antiviral.2009.07.006.
7. C. G. Noble, C. C. Seh, A. T. Chao, P. Y. Shi, *J. Virol.* **2012**, *86* (1), 438–446 DOI: 10.1128/JVI.06225-11.
8. M. Brecher, J. Zhang, H. Li, *Virol. Sin.* **2013**, *28* (6), 326–336 DOI: 10.1007/s12250-013-3390-x.
9. G. Kuno, G. J. J. Chang, *Arch. Virol.* **2007**, *152* (4), 687–696 DOI: 10.1007/s00705-006-0903-z.
10. J. Ko, H. Park, C. Seok, *BMC Bioinformatics* **2012**, *13* (1), 198 DOI: 10.1186/1471-2105-13-198.
11. L. Heo, H. Park, C. Seok, *Nucleic Acids Res.* **2013**, *41* (Web Server issue), 384–388 DOI: 10.1093/nar/gkt458.
12. W. H. Shin, C. Seok, *J. Chem. Inf. Model.* **2012**, *52* (12), 3225–3232 DOI: 10.1021/ci300342z.
13. Schrödinger Release 2015-4: Maestro, version 10.4, Schrödinger, LLC, New York, NY, **2015**.
14. H. K. Song, S. W. Suh, *J. Mol. Biol.* **1998**, *275* (2), 347–363 DOI: 10.1006/jmbi.1997.1469.
15. D. Leung, K. Schroder, H. White, N. X. Fang, M. J. Stoermer, G. Abbenante, J. L. Martin, P. R. Young, D. P. Fairlie, *J. Biol. Chem.* **2001**, *276* (49), 45762–45771 DOI: 10.1074/jbc.M107360200.
16. World Health Organization [WHO]. WHO Director-General summarizes the outcome of the Emergency Committee regarding clusters of microcephaly and Guillain-Barré syndrome **2016**.
17. J. Ko, H. Park, L. Heo, C. Seok, *Nucleic Acids Res.* **2012**, *40* (W1), 294–297 DOI: 10.1093/nar/gks493.