Illustrating and homology modeling the proteins of the Zika virus

Citation for published version:

Digital Object Identifier (DOI):
10.12688/f1000research.8213.2

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
F1000Research

Publisher Rights Statement:
Copyright: © 2016 Ekins S et al. This is an open access article distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
RESEARCH NOTE

Illustrating and homology modeling the proteins of the Zika virus [version 2; referees: 2 approved]

Sean Ekins¹-³, John Liebler⁴, Bruno J. Neves⁵, Warren G. Lewis⁶, Megan Coffee⁷, Rachelle Bienstock⁸, Christopher Southan⁹, Carolina H. Andrade⁵

¹Collaborations in Chemistry, Fuquay-Varina, NC, USA
²Collaborations Pharmaceuticals Inc., Fuquay-Varina, NC, USA
³Collaborative Drug Discovery Inc, Burlington, CA, USA
⁴Art of the Cell, Guilford, CT, USA
⁵LabMol - Laboratory for Molecular Modeling and Drug Design, Faculty of Pharmacy, Federal University of Goias, GO, Brazil
⁶Department of Medicine, Washington University School of Medicine, St Louis, MO, USA
⁷The International Rescue Committee, New York, NY, USA
⁸RJB Computational Modeling LLC, Chapel Hill, NC, USA
⁹Centre for Integrative Physiology, University of Edinburgh, Edinburgh, UK

Abstract

The Zika virus (ZIKV) is a flavivirus of the family Flaviviridae, which is similar to dengue virus, yellow fever and West Nile virus. Recent outbreaks in South America, Latin America, the Caribbean and in particular Brazil have led to concern for the spread of the disease and potential to cause Guillain-Barré syndrome and microcephaly. Although ZIKV has been known of for over 60 years there is very little in the way of knowledge of the virus with few publications and no crystal structures. No antivirals have been tested against it either in vitro or in vivo. ZIKV therefore epitomizes a neglected disease. Several suggested steps have been proposed which could be taken to initiate ZIKV antiviral drug discovery using both high throughput screens as well as structure-based design based on homology models for the key proteins. We now describe preliminary homology models created for NS5, FtsJ, NS4B, NS4A, HELICc, DEXDc, peptidase S7, NS2B, NS2A, NS1, E stem, glycoprotein M, propeptide, capsid and glycoprotein E using SWISS-MODEL. Eleven out of 15 models pass our model quality criteria for their further use. While a ZIKV glycoprotein E homology model was initially described in the immature conformation as a trimer, we now describe the mature dimer conformer which allowed the construction of an illustration of the complete virion. By comparing illustrations of ZIKV based on this new homology model and the dengue virus crystal structure we propose potential differences that could be exploited for antiviral and vaccine design. The prediction of sites for glycosylation on this protein may also be useful in this regard. While we await a cryo-EM structure of ZIKV and eventual crystal structures of the individual proteins, these homology models provide the community with a starting point for structure-based design of drugs and vaccines as well as a for computational virtual screening.
Introduction

All flaviviruses are spherical and contain a genome of approximately 11kb that functions as mRNA and encodes a polyprotein that leads to 10 proteins\(^1\). Examples include dengue virus, yellow fever and West Nile virus\(^2\). The recent pandemic of ZIKV occurring in South America, Latin America, the Caribbean and in particular Brazil spread by the *Aedes* mosquito has awakened dormant interest in this flavivirus which is a mild dengue-like disease\(^3\). However several documented cases of Guillain-Barré syndrome and other neurologic conditions represent important complications of the disease. In recent weeks the extent of the disease has also become apparent as new discoveries and announcements are made almost daily. Though clearly we have a considerable number of significant gaps in our knowledge which need addressing\(^4\).

The most concerning issue however is microcephaly observed in women who had ZIKV during pregnancy. There have been multiple cases of ZIKV found in fetal or newborn brain tissue that had signs of prenatal damage. The virus seems to have neurotropism in fetal brains, which may account for the presumed association between the infection and microcephaly\(^5,6\). The fetus in the recent case study had microcephaly with calcifications and ZIKV was found in the brain\(^7\). The ZIKV strain was identified as from French Polynesia (GenBank accession number KJ776791) and several polymorphisms were noted in the NS1, NS4B and FtsJ like methyltransferase regions. While the findings are not absolute proof that ZIKV causes microcephaly, the evidence from this case report strengthens the linkage\(^8\). Experts involved in the decision on the World Health Organization determined Public Health Emergency of International Concern (PHEIC) recommended the need for more research into the microcephaly link and need for an animal model to be developed. This group also interestingly called for open data sharing\(^9\). Early work 45 years ago in inoculated newborn mice showed that ZIKV had neurotoxic effects, enlarging astroglial cells and destroying pyriform cells. At the same time virus formation within the endoplasmic reticulum was also visualized\(^10\). We are not aware of any studies of effects of ZIKV on human brain or brain cells. Localization of such viruses to the brain is not unusual for flaviviruses i.e. West Nile virus and this tropism may arise from viral binding to glycosaminoglycans, as has been observed for dengue virus in human microvascular endothelial cells\(^11\). Heparan sulfate and the C-type lectin DC-SIGN (dendritic cell-specific intercellular adhesion molecule 3-grabbing nonintegrin) are well characterized adhesion structures for flaviviruses on cells. Interfering with glycan binding is one potential approach to preventing virus entry. Another is to acidify the endosome as has been demonstrated \textit{in vitro} with chloroquine for dengue virus infection\(^12\). Several entry and adhesion factors, including DC-SIGN, Tyro3, and AXL as well as others, have been shown to permit ZIKV entry in human skin cells\(^13\).

The routes for transmission of ZIKV besides mosquito are of some concern. Recent US CDC guidance to pregnant women describes precautions against sexual transmission of ZIKV\(^14\) and that the virus can persist for up to 12 weeks\(^15\). Possible ZIKV transmission through blood transfusion in French Polynesia was described by detecting the virus in 3% of asymptomatic blood donors\(^16\). Given how widespread ZIKV has become, there is a risk of depleting the blood supply, if donation after potential virus exposure is deferred. Methods have also been developed to inactivate ZIKV in plasma using amotosalen and UVA illumination\(^17\). There are issues with detection of ZIKV as a false positive dengue NS1 antigen test in a traveler to Switzerland was found to have the virus later. Therefore, cross-reactivity appears to be an issue in detection\(^18\) and this also suggests the need for better diagnostics to be developed.

Structural knowledge of the ZIKV proteins may allow us to understand exposed epitopes which will facilitate the development of specific diagnostic reagents that differentiate it from dengue and other flaviviruses. Furthermore, open sharing of the three-dimensional arrangement of viral surface proteins could allow the mapping of potential neutralizing epitopes, guiding efforts to rationally design effective vaccines. We recently developed a preliminary model for ZIKV glycoprotein E based on the dengue virus glycoprotein E late stage fusion intermediate as a trimer\(^19\). We now provide homology models of the glycoprotein E based on a dimer structure as well as attempts at modeling the other proteins in ZIKV. We have also investigated the likely glycosylation sites of the ZIKV envelope glycoprotein. Glycosylation may obstruct the binding of antibodies, or block access to potential underlying peptide antigens, so glycans may be an important consideration in diagnostic and vaccine development. Previously, glycosylation analysis using several computational tools predicted mammalian N-linked glycosylation at Asn-154 in most ZIKV strains\(^20\), a site known to be important in other flaviviruses. N-glycosylation of the dengue envelope glycoprotein at two sites has been shown to mediate interactions with DC-SIGN\(^21\). Insect cell expression of dengue shows both high-mannose and complex glycans\(^22\). However, flavivirus glycosylation in model systems such as insect cell culture and mammalian tumor cell lines may not represent the true infective insect and mammalian glycoprotein. Mammalian O-linked glycosylation was predicted at Thr-245 and Thr-381 in some isolates\(^23\). Yet, reliable prediction is hampered by a lack of O-glycosylation consensus sequences and preferences for O-glycosylation driven by structural characteristics\(^24\).

Understanding the three-dimensional structure of antigenic ZIKV proteins may help accelerate the development of antibodies for diagnostics and rationally designed vaccines. In addition, the comparison of the assembled surface glycoprotein of ZIKV with that of dengue virus may help understand the accessible epitopes for the development of anti-flaviral vaccines in general. There is considerable prior work including structure-based design and virtual screening for dengue, yellow fever and other related...
flaviviruses to develop antivirals to target envelope glycoproteins\textsuperscript{[39–45]}, guanylyltransferase\textsuperscript{[6]}, capsid protein NS3 helicase, NS2B-NS3 protease and NS5 polymerase\textsuperscript{[46,47]}, as well as whole cell screens\textsuperscript{[48]} which have produced many molecules potentially useful against ZIKV \textit{in vitro}. Early work has only tested a small number of FDA-approved drugs against ZIKV including (EC\textsubscript{50} in parentheses) interferon (34.3 IU/ml), ribavirin (143 ug/ml), 6-azauridine (1.5 ug/ml) and glycyrrhizin (384 ug/ml)\textsuperscript{[49]}. A recent paper by Hamel \textit{et al.}, from 2015 also showed interferon inhibited ZIKV replication in primary skin fibroblasts\textsuperscript{[50]}. However, the use of compounds against ZIKV should take into account the treatment of pregnant women, and many of the potential options are unsuitable for use in pregnancy because of toxicity and/or teratogenicity. Despite limited human data, the available data in animal models suggests caution. Azuridine is highly toxic to the fetus in model systems (for example 32). Ribavirin is not recommended for use in pregnancy due to embryotoxic and teratogenic effects\textsuperscript{[33]}. Interferon is a potential abortifacient\textsuperscript{[51]}. We also need to consider the treatment of fetuses as well as children (that might become infected after birth) and the relatively small subselection of FDA approved drugs that are approved for pediatric use\textsuperscript{[45]}. Therefore, alternative potential drugs for ZIKV are needed. The risks of medication use in pregnancy are notable. In particular, these include teratogenicity concerns. There is also the issue that the disease does not usually pose a direct risk to pregnant women themselves, so it’s important any drug, which will not be improving their own health, does not damage their health. Pregnancy can also create increased risks and liver problems, a particular concern with any new drug as it can also affect drug distribution. In pregnancy, a ZIKV infection at 13 weeks gestation was coupled with persistent virus in a fetus at 32 weeks\textsuperscript{33}. Treatment of symptomatic pregnant women may reduce risks of transmission to the fetus. A potential drug for ZIKV could also protect fetuses from damage by reducing transmission in the general population. If the drug appeared to reduce the duration of symptoms (which though mild can be annoying) and in turn reduced viral load, reducing the chance of transmission, this could benefit them. For example, cholera patients are often given antibiotics to reduce transmission. Also those with the flu are prescribed Tamiflu/Oseltamivir to reduce the duration of symptoms, minimize the severity of symptoms, but its use might also reduce transmission of flu in the general population. Since ZIKV has been found in semen many weeks after symptoms resolve\textsuperscript{12}, treatment of male partners may reduce viral load and reduce the long term risk of transmission.

To help accelerate drug discovery through computational analysis, we have now developed homology models of ZIKV proteins that may serve as potential drug and vaccine targets. To complement high-throughput screening efforts, we could perform virtual screening against the proteins in ZIKV. While there are crystal structures for proteins from dengue\textsuperscript{[34–37]}, yellow fever, West Nile virus and other flaviviruses\textsuperscript{[38–44]} there are (to date) none for ZIKV. Therefore we are limited to generating homology models, although the close evolutionary relationships between flavivirus and their component proteins and genomes represents a valid approach\textsuperscript{[45]}.

### Methods

#### Protein sequence alignment

As a prelude to modeling we assessed what 3D structural data had significant identity scores to a representative ZIKV polyprotein. For this we chose UniProtKB Q32ZE1_9FLAV. While we have requested the promotion of this entry to the Swiss-Prot expert review level it has been selected as the representative sequence for the UniRef90_Q32ZE1 entry that currently clusters 108 ZIKV individual sequence entries at 90\% (or above) amino acid identity. We then performed a BLAST search of this against Protein Data Bank (PDB) sequence entries. Protein BLAST analysis was also performed for each ZIKV protein sequence\textsuperscript{[46]} to identify the closest proteins\textsuperscript{[52]} and understand potential evolution.

#### Homology modeling

The amino acid sequences of ZIKV strain (GenBank accession number KJ776791\textsuperscript{48}) were retrieved from the GenBank database\textsuperscript{49} and used as targets for homology modelling using the SWISS-MODEL server\textsuperscript{[50,51]}. The latter performed the target-template sequence alignment after searching the putative X-ray template proteins in PDB for generating the 3D models for all target sequences. The best homology models were selected according to Global Model Quality Estimation (GMQE) and QMEAN statistical parameters. GMQE is a quality estimation which combines properties from the target-template alignment. The quality estimate ranges between 0 and 1 with higher values for better models. QMEAN4 scoring function consisting of a linear combination of four structural descriptors as described elsewhere in more detail\textsuperscript{[52,53]}. The pseudo energies returned from the four descriptors are related to what we would expect from high resolution X-ray structures of similar size using a Z-score scheme. Further, built models were exported to the SAVES server Version 4\textsuperscript{44} and their overall stereochemical quality, including backbone torsional angles through the Ramachandran plot, was checked according to PROCHECK\textsuperscript{54}. Lastly, each model was submitted to an energy minimization protocol, using the Smart Minimizer algorithm in Discovery Studio version 4.1 (Biotivia, San Diego, CA).

#### Site of glycosylation prediction

Mammalian N-glycosylation sites were predicted for glycoprotein E by submitting the sequence to web-based tools namely N-GlycoSite\textsuperscript{56}, GlycoEP\textsuperscript{57,58} and NetNGlyc Version 1.0\textsuperscript{59}.

#### Illustration for Zika virion and animation

The Zika virion illustrations were created by combining the homology model of the envelope ZIKV glycoprotein E with the symmetry data from the dengue virus envelope. PDB ID:1K4R\textsuperscript{60} contains the coordinates for three copies of the protein subunit of the dengue virus envelope, along with the symmetry data necessary to create the 180-subunit icosahedral structure of the complete viral envelope. The PyMOL Molecular Graphics System, Version 1.7.6.0. Schrödinger, LLC. was used to export the surface models of the three proteins in .obj format. Then they were imported into Lightwave 3D (NewTek, San Antonio, TX) where the symmetry data was used to instance copies of the model into the icosahedral envelope. The entire structure was copied several times.
and lighting applied as a surfacing effect to create a visually pleasing composition, and the image rendered out.

The next step was to import into Pymol the homology model of the ZIKV envelope protein which was homology modeled using PDB ID: 3P54 (from Japanese Encephalitis Virus) as a template. A surface model of this protein was exported from Pymol as an .obj, and imported into Lightwave in place of the dengue model, using the same symmetry operators to create the envelope array. Everything else about the picture was left the same (color, composition, lighting, etc...) to allow the structural differences to be more apparent, and that image rendered out as well.

The last step was to overlay the detailed area of the two images and create an animated gif to flip back and forth between the two images of ZIKV and dengue, again to allow the differences to be more clearly seen. These results are in general accordance with whole sequence analyses.

ZIKV glycoprotein E homology model conformation comparison

The immature and mature (this study) homology models for glycoprotein E were compared using the ‘align and superimpose’ protocol in Discovery Studio Version 4.1 (Biovia, San Diego, CA).

Results

Sequence alignment across flaviviruses

A BLAST search of ZIKV polyprotein against PDB sequence entries shows the highest scoring matches with 55–70% sequence identity (Supplementary material S1). Protein BLAST analysis of the individual ZIKV protein sequences show that many of the proteins are similar to the same protein from Spondweni virus in 12 out of 15 cases (Table 1). Exceptions were: FtsJ which was closer to the Murray Valley encephalitis virus, NS1 which was more similar to dengue virus 3, and glycoprotein E which was closest to the dengue virus 1 protein. These results are in general accordance with whole sequence analyses.

Homology modeling

The SWISS-MODEL server was used to generate alignments (Supplementary material S2) and homology models for all Zika proteins (Table 2, Figure 1, Supplementary material S3). First, we selected suitable template protein structures in PDB, observing the following criteria: the template should have a high coverage (i.e., > 65% of target aligned to template) and sequence identity >30%. Then, we used GMQE and QMEAN4 scoring function as an initial
Table 2. Summary of ZIKV homology models and statistical validation. The global and per-residue model quality has been assessed using the QMEAN scoring function\(^{53}\). For improved performance, weights of the individual QMEAN terms have been trained specifically for SWISS-MODEL\(^{50,51,71-73}\). GMQE = Global Model Quality Estimation, QMEAN4 is a scoring function consisting of a linear combination of four structural descriptors as described elsewhere in more detail\(^{52,53}\).

<table>
<thead>
<tr>
<th>Zika Protein</th>
<th>Template</th>
<th>Coverage</th>
<th>Sequence Identity</th>
<th>GMQE</th>
<th>MEAN4</th>
<th>PROCHECK analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Organism</td>
<td>Protein</td>
<td>PDB code</td>
<td>Identity</td>
<td></td>
<td>Most favored regions</td>
</tr>
<tr>
<td>NS5</td>
<td>Japanese encephalitis virus</td>
<td>NS5(^{61})</td>
<td>4K6M</td>
<td>100%</td>
<td>53%</td>
<td>0.89</td>
</tr>
<tr>
<td>FtsJ</td>
<td>West Nile virus methyltransferase(^{62})</td>
<td>2OY0</td>
<td>99%</td>
<td>54%</td>
<td>0.94</td>
<td>-1.23</td>
</tr>
<tr>
<td>HELICc</td>
<td>Dengue virus helicase/nucleoside triphosphatase catalytic domain(^{63})</td>
<td>2BHR</td>
<td>100%</td>
<td>55%</td>
<td>0.94</td>
<td>-1.60</td>
</tr>
<tr>
<td>DEXDc</td>
<td>Murray Valley encephalitis virus RNA helicase(^{64})</td>
<td>2V8O</td>
<td>100%</td>
<td>54%</td>
<td>0.93</td>
<td>-0.92</td>
</tr>
<tr>
<td>Peptidase S7</td>
<td>West Nile virus Ns2B-Ns3 protease(^{65})</td>
<td>2YOL</td>
<td>100%</td>
<td>53%</td>
<td>0.92</td>
<td>-0.25</td>
</tr>
<tr>
<td>NS1</td>
<td>West Nile virus non-structural protein (^{16})</td>
<td>4O6D</td>
<td>99%</td>
<td>48%</td>
<td>0.77</td>
<td>-4.09</td>
</tr>
<tr>
<td>E Stem</td>
<td>Dengue virus capsid protein(^{67})</td>
<td>3J2P</td>
<td>100%</td>
<td>46%</td>
<td>0.74</td>
<td>-8.02</td>
</tr>
<tr>
<td>Glycoprotein M</td>
<td>Dengue virus cryo-EM structure of virus(^{68})</td>
<td>3J27</td>
<td>100%</td>
<td>40%</td>
<td>0.73</td>
<td>-6.90</td>
</tr>
<tr>
<td>Propeptide</td>
<td>Dengue virus precursor membrane protein-envelope(^{69})</td>
<td>3C5X</td>
<td>87%</td>
<td>47%</td>
<td>0.74</td>
<td>-0.70</td>
</tr>
<tr>
<td>Capsid</td>
<td>West Nile virus core (C) protein(^{70})</td>
<td>1SFK</td>
<td>65%</td>
<td>42%</td>
<td>0.50</td>
<td>-3.54</td>
</tr>
<tr>
<td>Glycoprotein E</td>
<td>Japanese encephalitis virus envelope glycoprotein(^{70})</td>
<td>3P54</td>
<td>99%</td>
<td>47%</td>
<td>0.81</td>
<td>-3.76</td>
</tr>
<tr>
<td>NS4A</td>
<td>—</td>
<td>—</td>
<td>20%</td>
<td>29%</td>
<td>0.07</td>
<td>-1.75</td>
</tr>
<tr>
<td>NS2B</td>
<td>—</td>
<td>—</td>
<td>37%</td>
<td>48%</td>
<td>0.21</td>
<td>-0.49</td>
</tr>
<tr>
<td>NS2A</td>
<td>—</td>
<td>—</td>
<td>16%</td>
<td>31%</td>
<td>0.04</td>
<td>-3.10</td>
</tr>
<tr>
<td>NS4B</td>
<td>—</td>
<td>—</td>
<td>14%</td>
<td>30%</td>
<td>0.03</td>
<td>-2.81</td>
</tr>
</tbody>
</table>
criteria to discriminate good from bad models. Acceptable alignment values and higher GMQE and QMEAN4 scores were obtained during modeling, suggesting statistically acceptable homology models were generated for 11 proteins: NS5, FtsJ, HELICc, DEXDc, peptidase S7, NS1, E stem, glycoprotein M, propeptide, capsid, and glycoprotein E (Table 2, Figure 1 and Figure 2). The Ramachandran plots for these 11 proteins provide further evidence of their acceptability (Figure 2). On the other hand, because of low GMQE scores and of low coverage observed in X-ray template proteins available in the PDB, homology models for NS4B, NS4A, NS2B, and NS2A proteins appeared to have limitations regarding active sites and epitopes and they could not be validated.

After building of homology models, we performed an additional validation in order to explore stereochemical quality of dihedral angles phi against psi of amino acid residues in modeled structures and identify sterically allowed regions for these angles using PROCHECK analysis. The results shown in Table 2 and Figure 2 reveal that 58.4—70.3% residues of the modeled proteins are within the most favored regions (red), 27.1—43.3% residues of modeled proteins are within the additional allowed regions (yellow), 1.5—6.3% residues of modeled proteins are within the generously allowed regions (beige), and only 0.0—6.1% residues of modeled proteins are within the disallowed regions (white). These results showed that the overall stereochemical properties of the
Figure 2. Ramachandran plots for ZIKV NS5 (A), FtsJ (B), HELICc (C), DEXDc (D), Peptidase S7 (E), NS1 (F), E Stem (G), Glycoprotein M (H), Propeptide (I), Capsid (J), and Glycoprotein E (K) obtained by PROCHECK, showing the dihedral angles Psi and Phi of amino acid residues. Red represents most favored regions; yellow represents additional allowed regions; beige represents generously allowed regions; and white areas are disallowed regions.
generated models were highly reliable and the models could be useful to future molecular modeling studies.

Site of glycosylation prediction
Several web-based tools were used for N-glycosylation site predictions as it provides a more thorough approach. N-GlycoSite suggested N154 as a single N-glycosylation site matching the N-X-S/T/C consensus sequence. The same site was identified by GlycoEP using BPP settings giving a score of 0.65/1.00. NetNGlyc also gave the same predicted site, with a jury agreement of 6/9.

Zika virion compared to dengue virion
A qualitative analysis of the Zika virion (which was constructed based on the dengue cryo-EM virion) indicates that Zika appears to have slightly more raised ‘pimples’ on the surface. The glycoprotein E dimer in ZIKV also has a narrower ‘letter-box’ groove while the dengue virion has a bigger ‘pore’ between the intersection of 5 dimers (5 fold axis). These differences are considerably more apparent in the animation (Supplementary material S4). It is important to note that the differences may also be artefacts of the homology modeling approach and template used for modeling ZIKV glycoprotein E.

ZIKV glycoprotein E homology model conformation comparison
The homology models developed using two different templates namely the immature protein which was based on the dengue crystal structure 4gsx as a template and the mature protein which was based on PDB ID:3P54 from Japanese encephalitis virus showed a large difference (RMSD 13.47Å) (Figure 4). These proteins also demonstrate differences around the pocket used centered on the residues 270-277.

Discussion
The genus Flavivirus consists of 70 viruses many of which can cause severe human disease. There have been few sequence analyses of ZIKV previously in comparison to other flaviviruses. The genus Flavivirus produces a monophyletic tree with ZIKV being closest to Spondweni virus while mosquito borne, tick borne and no-vector viruses cluster separately. A BLAST analysis of all the ZIKV proteins in this study suggests for 12 of 15, their closest protein is in Spondweni virus (Table 1). More often strain sequences are compared within ZIKV and these showed variations in the NS5 gene and glycoprotein E. This is important as it would suggest perhaps targeting other proteins would have less issue with resistance or variability due to the strain of ZIKV.

If we are to address ZIKV in the short term while we await a vaccine we need to rapidly identify an antiviral, and preferably one that can be used against other related flaviviruses. Ideally we would need to treat pregnant women and provide them with prophylaxis that was safe to them and their fetus. Such an antiviral could also be used to reduce transmission in the population in general (by reducing viral load and symptoms and/or duration). As noted a decade ago and is still is true today, no antiviral drug is approved for any flavivirus to date. It has been suggested that one of the ways to target these viruses is to interfere with the NS2B/NS3 protease complex. Understanding of flavivirus proteins and other RNA viruses has benefited from the EU funded project VIZIER, in particular several West Nile virus, dengue virus and other flavivirus structures of NS3 or NS5 were solved during this project and allosteric inhibitor sites were identified on NS5. Multiple pharmaceutical companies have worked on this target for HCV leading to clinical candidates like IDX320, danoprevir (ITMN-191/R7227), GS-9256 and others. The only HCV protease targeting FDA approved drug is simeprevir, TMC435 and its use is avoided in pregnancy. Other HCV protease compounds are in clinical trials or submitted for FDA approval including Ledipasvir (formerly GS-5885). Testing these molecules against ZIKV in vitro would be useful.

We recently described 6 steps which could be taken to kick start research on ZIKV, one of which was to develop homology models for ZIKV proteins that are similar to those targeted by molecules that are also active against the dengue virus. Such an approach would then enable docking of compound libraries of known

---

**Figure 3.** Comparison of Zika and dengue virion illustrations.
antivirals, FDA approved drugs or other compounds. Ideally generating homology models with a single tool may not be enough. In particular, for those proteins with low sequence identity the use of servers and methods that use threading may be worthwhile (e.g. I-TASSER). However these methods are generally only accessible to academics while others are required to license the technologies. This is ironic as these technologies were developed in most cases with NIH and NSF funds. An alternative commercial homology modeling approach (MODELLER) was also used and generated a NS5 homology model and the top hit was also the Japanese encephalitis virus RdRp domain (PDB ID: 4HDH) compared with PDB ID: 4K6M from SWISS-MODEL. 4HDH also includes the ATP and zinc metal where the catalytic centers are. The dengue virus 3 polymerase (PDB ID: 4HHJ) has very high sequence homology and comes up as a potential target in MODELLER, which illustrates that all these viral RNA dependent polymerases are very similar.

While it is likely that the eventual availability of crystal structures of ZIKV proteins would improve the results of docking, the homology models described here (Figure 1, Figure 2, Supplementary material S2) represent a starting point that can be used to help prioritize compounds for testing as described previously. Proteins with templates above 25–40% sequence identity might suggest the proteins are related while below this is a twilight zone. Homology modeling is thought to fill in the gaps between proteins with x-ray structures and those with none. Experimental testing of homology models and crystal structures indicate that a similar enrichment rate can be achieved when identifying active compounds in a set decoys. Others have also described homology models that may be an excellent alternative when crystal structures are unavailable for human GPCRs, and have led to the first identification of inhibitors of the Mycobacterium tuberculosis Topoisomerase I after virtual screening prior to the crystal structure becoming available. Certainly there are still considerable challenges using homology models such as prediction of the correct binding pose but there are plenty of success stories. While databases of homology models exist like MODBASE and SWISS-MODEL, neither of these have any ZIKV protein homology models at the time of writing. There are many structural genomics initiatives and yet it would seem there are few if any continuing the work of VIZIER working on flaviviruses or emerging viruses.

Availability of structures are important as the structure of the ZIKV glycoprotein could be useful for design of antibodies selective for the virus which will be critical for the development of diagnostics, and understanding antibody binding also for the use of IV immunoglobulin in pregnancy and the organization of the epitopes on viral proteins may facilitate early work in vaccine development. There are further implications for understanding the antibody binding epitopes, which are sometimes shared between different flaviviruses. Broadly protective vaccines for flaviviruses may allow the simultaneous targeting of ZIKV and related viruses such as dengue.

Understanding glycosylation is therefore important. To date Asn-154 is mentioned in Faye et al., as a glycosylation site, as are Thr-170 (mucin type O-linked glycosylation) and other mucin sites at Thr-245 and Thr-381. Other probably O-GlcNAC attachment sites (Ser-142, Ser-227, Thr-231, Ser-304, Thr-366, and Thr-381) were also predicted. Our analysis of N-glycosylation with 3 different websites suggest Asn-154 also as a likely site of N-glycosylation in agreement with dengue virus.

Cryo-EM has been used to show how glycoprotein E dimers arrange on the surface of virions for flaviviruses including tick-borne encephalitis virus, West Nile virus, dengue virus 1, and dengue virus 4 (Supplementary material S3). The early work on the tick-borne encephalitis virus suggested 30 dimers on the surface and also pointed to how the glycoprotein E dimers can reorganize under low pH to form a trimmer. The packing of the dimers in the dengue virion is different to tick-borne encephalitis with the glycoprotein E dimers showing 30 ‘herringbone rafts’ each containing three dimers to result in 180 copies of the protein. West Nile virus again has a different arrangement with 60 trimers shown in the structure of the immature virus (Supplementary material S4). Even between the dengue serotypes 1, 2 and 4 for which there are cryo-EM structures, it is apparent while the rafts are very similar (as are the sequence identities [60%]) there is a different charge distribution of the surface of each. Dengue serotype 2 had larger continuous patches of positive charges which was proposed to enable improved binding to heparan sulfate. This might also be the case for ZIKV in that the charge pattern is again different and could be key for vaccine development. The availability of virion structures makes it feasible to understand structure function of the complete virus such as assessment of membrane curvature and how organization of membrane proteins affects this.

A model of the Zika virion was constructed as an illustration using the homology model of the glycoprotein E dimer (Figure 3). While the combined protein sequence of glycoprotein E and the immunoglobulin like domain is closest to dengue virus 1 (57 percent identity, Table 1) the closest template was for the crystal structure of the Japanese encephalitis virus envelope protein (53.12 percent identity, Table 2). This would suggest the virion should more closely resemble that of dengue virus 1, while producing a homology model based on a more distant virus might not be ideal. The homology model of glycoprotein E developed for the mature conformation in this study is significantly different from that developed previously for the immature conformation (Figure 4). The proposed binding site centered around residues 270-277 appears shallower in the mature conformation and this would certainly affect the kinds of molecules that it could interact with. It might also point to the need to interfere with the immature conformation as preferable.
versus the mature conformation. Ultimately perhaps this model of the Zika virion could help us understand how drugs could access the virus. Viruses affecting pregnancy, like say Varicella which causes microcephaly and other developmental problems\cite{104,105}, are often treated with IV immunoglobulin, i.e. antibodies, as well as antivirals to reduce the effect of the virus (or to avoid infection if given soon after exposure). The models could help us design combination approaches possibly targeting multiple proteins that might prevent drug resistance from occurring also.

Does having the homology models and the virion illustration help understand function? Well, the surface charge pattern might be inferred from the homology model and could be compared with dengue and other filoviruses for which there are cryo-EM structures. This may in turn present opportunities for vaccine design by indicating accessible surfaces and properties, allowing mapping of epitopes, design of accessible fragments and peptides for vaccine/diagnostic design. Vaccines themselves might be the only way to avoid the inevitable, otherwise, simply reducing the spread of ZIKV would just delay it. Women ultimately may just want to ‘get it over with’ and have ZIKV before they get pregnant and hope there is lasting immunity.

In summary, in the absence of crystal structures for any of the proteins comprising the ZIKV, we are left to attempt to construct homology models which we have done using the freely available SWISS-MODEL server. Further preparation of these models required freely available and commercial tools. In the case of the ZIKV glycoprotein E homology model, this has the added benefit of enabling the construction of a full virion. By comparing the Zika virion to the existing structures for other flaviviruses we can see similarities and differences on the surface (Supplementary material S4, Supplementary material S5). This relatively crude approach could help to understand how we might develop antivirals and vaccines against it. In addition we now provide homology models as a starting point for (small and large scale) docking studies and further evaluation which may complement other modeling efforts for ZIKV\cite{106}. Ultimately the results of their use can be compared with using ZIKV crystal structures once generated.

Notes added while in review
Since the initial publication of this article several cryo-EM and crystal structures have been published for ZIKV, including candidate targets for inhibitor medicinal chemistry optimization and vaccine design\cite{107-109}. Detailed comparisons between our homology models and the experimental structures will be the subject of future comparisons with the homology models described herein. It should be noted that the prediction for the site of glycosylation of glycoprotein E was experimentally verified by Sirohi et al.\cite{110}.

We would also point out the utility of this work through the World Community Grid we have started the OpenZika project\cite{111-114} which is using these homology models, template structures and crystal structures to dock millions of molecules (with AutoDock Vina\cite{116-118}) to identify compounds for testing against Zika with collaborators. Both the docking and activity screening results (http://openzika.ufg.br/experiments/) will be open to the community and published in due course. An obvious utility of this entire exercise is that, by testing, iterating and tweaking the methodologies each time (e.g. between Ebola and ZIKV) the open science community becomes better prepared for the next global pathogen emergency.

Author contributions
All authors contributed to the collaborative writing of this project. SE conceived and designed the experiments. SE, JL, BHN, WGL, CS and CHA carried out the research.

Competing interests
S.E. works for Collaborations in Chemistry, Collaborations Pharmaceuticals, Inc. and Collaborative Drug Discovery, Inc.

Grant information
CS was supported by Wellcome Trust Grant (to the IUPHAR/BPS Guide to PHARMACOLOGY) Number 099156/Z/12/Z.

I confirm that the funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgments
Dr’s Priscilla Yang, Andrew Marsh, Derek Gatherer, Lucio Freitas-Junior, Daniel Mietchen, Joel S Freundlich, Jair Siqueira-Neto, Antony J. Williams, Alex Perryman and Mr. Tom Stratton are thanked for their helpful discussions and tweets. Biovia is kindly acknowledged for providing Discovery Studio to SE.

Supplementary material
Supplementary material S1 (Alignments of ZIKV to other proteins), S2 (Alignment of individual ZIKV proteins for homology models), S4 (ZIKV versus dengue virion animation), and S5 (Published flavivirus cryo-EM structures [not to scale]).

Click here to access the data.

Supplementary material S3. PDB files for ZIKV homology models.
Click here to access the data.

PubMed Abstract | Publisher Full Text


PubMed Abstract


PubMed Abstract | Publisher Full Text


PubMed Abstract | Publisher Full Text

77. Arcen, VIZEIR. Reference Source


PubMed Abstract | Publisher Full Text | Free Full Text


PubMed Abstract | Publisher Full Text


PubMed Abstract | Publisher Full Text


PubMed Abstract | Publisher Full Text | Free Full Text


PubMed Abstract | Publisher Full Text


PubMed Abstract | Publisher Full Text | Free Full Text


PubMed Abstract | Publisher Full Text


PubMed Abstract | Publisher Full Text | Free Full Text

86. Zhang Y: I-TASSER. Reference Source


PubMed Abstract | Publisher Full Text | Free Full Text


PubMed Abstract | Publisher Full Text | Free Full Text


PubMed Abstract | Publisher Full Text | Free Full Text


PubMed Abstract | Publisher Full Text | Free Full Text


PubMed Abstract | Publisher Full Text | Free Full Text


PubMed Abstract | Publisher Full Text | Free Full Text


PubMed Abstract | Publisher Full Text | Free Full Text


111. Anon: EMDataBank.

Open Peer Review

Current Referee Status:  

Version 2

Referee Report 27 September 2016

doi:10.5256/f1000research.10258.r16629

 Bruno Botta  
Department of Chemistry and Pharmaceutical Technology, Sapienza University of Rome, Rome, Italy

The revised manuscript can be indexed. It has been improved accordingly.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.

Referee Report 16 September 2016

doi:10.5256/f1000research.10258.r16006

Andras Fiser  
Albert Einstein College of Medicine, Bronx, NY, USA

The Authors clarified many of the points I raised and I approve the manuscript.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.

Version 1

Referee Report 02 August 2016

doi:10.5256/f1000research.8833.r15364

Bruno Botta  
Department of Chemistry and Pharmaceutical Technology, Sapienza University of Rome, Rome, Italy

The work of Sean Ekins and colleagues describes an attempt to predict the 3D structure of a number of Zika Virus proteins by homology modeling.
Infections by Zika virus are seriously preoccupying the population and governments of endemic countries, with particular attention to pregnant women. Indeed, the most serious complications of infections by Zika virus have been observed in pregnant women (i.e. microencephalic fetus). However, the evidence that pregnant women are most affected by Zika virus infection, set some difficulties in developing focused therapeutic strategies, as nicely discussed by the authors. However, in this respect, which is the scenario proposed by authors for a candidate inhibitor of Zika virus replication. Why this molecule should be useful, and how pregnant women could deal with the administration of the drug?

Overall, the work is well written and organized in a clear and rational way, even if no experimental validation of the work carried out in silico is proposed. Accordingly, conclusions of this work are merely speculative. My major concern is on the usability of the protein structures for drug design or virtual screening, as claimed by the authors in the abstract and discussion. Indeed, sequence identity between template and target sequences is rather low in some cases (< 60 %). Moreover, the atomistic detail of the active site and its conformation may vary noticeably in the models depending on the program used, the force field, the level of refinement (algorithm, steps, solvent model, …) and the quality of the template structure. All these variables should be taken into consideration by the authors, and the refinement of 3D models should be at least attempted or discussed in deeper details. In my personal opinion, details on the conservation rate, sequence identity and structural similarity of the binding sites would be more helpful for drug design purposes.

Finally, an experimental validation of at least one of the structures modeled by authors (i.e. the most promising target for drug designing studies) should be provided. In fact, analysis with PROCHECK or the use of numerical scores is not sufficient to validate a 3D structure generated by homology modeling. There is a high risk that molecular docking towards a low-resolution structure could provide unrealistic results. The paper can be accepted after revision.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

**Competing Interests:** No competing interests were disclosed.
eradicate the virus as soon as possible so that the fetus perhaps is not exposed to the virus.

Overall, the work is well written and organized in a clear and rational way, even if no experimental validation of the work carried out in silico is proposed.

Response: This work was initiated in Feb 2016, months before the first crystal structures and experimental in vitro systems were published. Our goal was to build protein models that we would eventually use and which others could use in parallel, to start docking-based virtual screening and identify VS and experimental hits, to initiate the drug discovery cascade.

Accordingly, conclusions of this work are merely speculative.

Response: A modeling study, by definition is speculative and does not claim otherwise. However, in good faith, we provided models which did not exist. The fact that they have since gone on to form the basis of the OpenZika Project (manuscript submitted, http://openzika.ufg.br) and have been used to suggest compounds which are now being tested argues for their utility. Notwithstanding, we acknowledge that Zika protease, helicase and glycoprotein E experimental structures have now been elucidated and deposited in PDB.

My major concern is on the usability of the protein structures for drug design or virtual screening, as claimed by the authors in the abstract and discussion. Indeed, sequence identity between template and target sequences is rather low in some cases (< 60 %).

Response – As expected for the enzyme functions, the local active site identity and similarity scores are higher than indicated by the full-length identity %. Thus, a reasonable pocket model for docking is the primary goal, not the structural accuracy for the whole protein. We would also add that any team using our models would doubtless take a pragmatic approach that would include running various types of internal controls for virtual screening studies and confirmatory screening experiments (e.g. running the Dengue virus templates alone and checking for enrichment of established Dengue actives against the purified proteins). They would also probably prioritize their effort on the models with highest template similarity. Note also that, the first reviewer Dr. Fiser accepts that our similarity values are high enough for reliable modelling.

Moreover, the atomistic detail of the active site and its conformation may vary noticeably in the models depending on the program used, the force field, the level of refinement (algorithm, steps, solvent model, …) and the quality of the template structure.

Response – We agree that this is one of the challenges with this approach, but there were no other options without a single crystal structure for Zika protein when we started this work and submitted it.

All these variables should be taken into consideration by the authors, and the refinement of 3D models should be at least attempted or discussed in deeper details. In my personal opinion, details on the conservation rate, sequence identity and structural similarity of the binding sites would be more helpful for drug design purposes.
Response: We agree with your point. We are now refining the 3D models using molecular dynamics simulations. However, this will be the subject of work on going, taking an in depth approach to each structure and comparison to crystal structures when available. Our main goal in this paper which we submitted in Feb 2016 (and made available early March) was to provide the scientific community with the first 3D models of Zika proteins.

Finally, an experimental validation of at least one of the structures modeled by authors (i.e. the most promising target for drug designing studies) should be provided. In fact, analysis with PROCHECK or the use of numerical scores is not sufficient to validate a 3D structure generated by homology modeling.

Response – We agree that the experimental validation of the protein structures is of upmost importance in drug discovery programs, but this was not the scope of this paper. Again, our main goal was to generate homology models for the key proteins, which none crystal was available at the moment of the submission, and these models could be taken to initiate ZIKV antiviral drug discovery using both high throughput screens as well as structure-based design. Moreover, these structures were the core to initiate the World Community Grid project called OpenZika (https://www.worldcommunitygrid.org/research/zika/overview.do), which is a global research project to accelerate the discovery of an antiviral against the Zika virus.

PROCHECK was originally developed to evaluate the stereochemical quality of experimentally determined protein structures (Laskowski R A, MacArthur M W, Moss D S, Thornton J M (1993). J. App. Cryst., 26, 283-291; Morris A L, MacArthur M W, Hutchinson E G & Thornton J M (1992). Proteins, 12, 345-364). However, PROCHECK has been widely used and reported in the literature to check the stereochemical quality of homology models, to provide numerical scores that can evaluate the statistical quality of the generated 3D protein models (See bellow the list of summarized most recent literature). It is a quality assessment criterion and is not related to experimental validation.

Papers reporting the use of PROCHECK for assessing the stereochemical quality of homology models:

There is a high risk that molecular docking towards a low-resolution structure could provide unrealistic results.

Response – Yes we absolutely agree, but when we started this we had no crystal structures we had to start somewhere as a means to start to filter compounds rather than random HTS.
Our work has also correctly predicted the glycosylation site for glycoprotein E one of the first structures crystallized. Moreover, we have compared our models with the crystallographic structures, when they were made available. The RMSD values ranged from 0.72 to 1.8 Å, for the structures of NS1, NS3, glycoprotein E and NS2/NS3 proteins.
We have also optimized the 3D models and the MolProbit scores ranged from 1.28 to 2.81 Å. MolProbit is a score that combines the clashscore, rotamer, and Ramachandran evaluations into a single score, normalized to be on the same scale as X-ray resolution. Therefore, we are optimistic that our models are reliable to start a docking-based virtual screening program in the search of a new antiviral drug, which we already initiated.

The paper can be accepted after revision.

Response – Thank you!

**Competing Interests:** No competing interests were disclosed.
3. However the information on templates should be reported in the table, instead listing them extensively in the text.

4. Template selection was purely based on sequence similarity requirements. Given how influential this step and how few cases require attention in this specific study, one could imagine to perform a more detailed quality check of potential templates. For instance, if several templates are available in the same sequence identity range, one should consider picking the one with the best resolution etc. Figure 4 exposes this dilemma where two isoforms of the same template can result in models with very significant differences.

5. Energy refinement rarely, if ever, improves model quality and its use here should be better justified.

6. Some statements require attention e.g. in Abstract: “Eleven out of 15 models pass our criteria for selection”. What selection? This must be referring to quality or accuracy of models, and should be rephrased accordingly.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

**Competing Interests:** No competing interests were disclosed.

---

**Author Response 25 Aug 2016**

Sean Ekins, Collaborations in Chemistry, USA

Dear Prof Fiser, thank you for taking the time to review and constructively comment on this manuscript.

The paper describes homology models of 15 proteins encoded in the genome of Zika virus that are built by the SWISSmodel web server. A glycolysis site was also identified. Using the models of Zika glycoprotein E a complete structural model of the virion was constructed.

The practical results of the work are the 11 high quality homology models that could be used in the future for structure based drug development. I believe an interested researcher will initiate time consuming follow up studies with these models only if there is a substantial added value.

**Response:** Thank you! We would also point out the utility of this work is not only reflected in the high level of accesses, downloads and citations, but also for initiating the World Community Grid as the OpenZika project, which was launched in May, 2016 (https://www.worldcommunitygrid.org/research/zika/overview.do)

These models passed quality assessment criteria according to the authors. However a number of successful models (passing similarly well the same quality requirements) can be generated that will differ in small but essential details by e.g. side chain placement, or loop conformations. However, these differences can have a dramatic effect on the outcome of subsequent drug docking trials. These alternative models can be obtained when different softwares are used, as the authors allude to it, e.g. using Modeller vs SWISSModel, as these use different force fields and restraints to generate models. Therefore, the authors should consider providing a more insightful result by running several different modeling programs or using alternative templates (see 4., below), and
comparing the results, identifying similarly modeled parts of these models and providing a set of possible solutions for subsequent studies.

Response: We would agree that certainly an exhaustive analysis of different methods was not undertaken. That was not our goal. We attempted to build models that we could then put into practice and provide to the community for them to use knowing the possible caveats.

A battery of model quality checks were performed and additional energy minimization. However the overall high sequence identities between the target proteins and their respective templates (55% and up,) ensure that these models are highly reliable.

Response: Thank you!

1. Therefore the extensive reporting on Ramachandran plots (Figure 3) is not adding much to the results, it can be transferred to supplementary material.

   Response: Actually many publications use these plots and they may provide some insights on model quality so we have opted to retain them in the body of the manuscript, as the F1000Research journal does not have a limit for figures.

2. Similarly, Table 2 can be shrunk by eliminating many details of PROCHECK results.

   Response: This is an internet publication we are not constrained by pages or page charges and these numbers are informative to readers.

3. However the information on templates should be reported in the table, instead listing them extensively in the text.

   Response: We agree and we believe that it is neater and more accessible in a table. Therefore, we have added the information on templates in Table 2 and removed this information from the text.

4. Template selection was purely based on sequence similarity requirements. Given how influential this step and how few cases require attention in this specific study, one could imagine to perform a more detailed quality check of potential templates. For instance, if several templates are available in the same sequence identity range, one should consider picking the one with the best resolution etc. Figure 4 exposes this dilemma where two isoforms of the same template can results models with very significant differences.

   Response: We totally agree – some decisions had to be made as a trade-off to get the models built in a reasonable time to put into use. We could have spent months building models by which time the x-ray structures would have been out. We had models of the proteins and the virion at least 1 month before the first cryo-EM structure. One month is a long time in drug discovery, in which we could be finding molecules for testing.

5. Energy refinement rarely, if ever, improves model quality and its use here should be better justified.
Response: We totally agree with you, Dr. Fiser. Energy minimization does not improve model quality. The word “refined” in the paper was a mistake and we have corrected it now. We are using some servers, such as KoBaMIN server, to perform structure refinement and to improve models quality, but this will be subject of another manuscript. Our main goal in this first paper was to provide the scientific community with homology models for the key proteins, which none crystal was available at the moment of the submission of this paper, and these models could be taken to initiate ZIKV antiviral drug discovery using both high throughput screens as well as structure-based design. Moreover, these structures were the core to initiate the World Community Grid project called OpenZika (https://www.worldcommunitygrid.org/research/zika/overview.do), which is a global research project to accelerate the discovery of an antiviral against the Zika virus.

6. Some statements require attention e.g in Abstract: “Eleven out of 15 models pass our criteria for selection”. What selection? This must be referring to quality or accuracy of models, and should be rephrased accordingly.

Response: OK Thank you! Changed to - Eleven out of 15 models pass our model quality criteria for their further use.

**Competing Interests:** No competing interests were disclosed.