Zika virus therapeutic lead compounds discovery using chemoinformatics approaches

Abstract

The Zika virus disease (ZVD) currently spreading around the globe has no known treatment available. A database of Zika-related genes (N=69) was subjected to chemoinformatics approaches using the canSAR protein annotation tool. Ligand-based druggability analysis of the database of genes identified thirty-five druggable targets encompassing adhesion molecules, cytokines, enzymes, growth factors and receptors. Eight of these proteins (CCR5, HLA-DRB1, IL6, LTA, PLAT, PPIB, TNF and VEGFA) are current drug targets. Active (IC50<100nM) Rule of Five (RO5) compliant and toxicophore negative CHEMBL compounds were identified for nine of these Zika virus lead targets (AXL, CASP1, CCL2, CCR5, CTSS, CXCL8, E1F2AK2, NPM1 and TYRO3). The AXL lead gene is a target for an FDA approved antineoplastic kinase inhibitor (Crizotinib/Xalkor). Another FDA approved kinase inhibitor for neoplasms; (Gefitinib) targets the E1F2AK2 lead protein. The protein targets (AXL, CASP1, CCR5, CTSS and RAF1) had highly bioactive compounds available (IC50<1nM). CHEMBL compounds with unique protein target specificity were identified for AXL, CCL2, CCR5, CASP1, CTSS, CXCL8/IL8, NPM1/ALK and TYRO3. The RO5 compliant and toxicophore-negative lead compounds identified in the study offer a rationale for rapid verification in Zika virus cell culture models for drug discovery.

Keywords: canSAR, chemoinformatics, druggable targets, guillain-barré syndrome, ligand binding, microcephaly, zika virus

Abbreviations: canSAR, integrated cancer drug discovery platform; CDC, us centers for disease control; CHEMBL, database of bioactive compounds at european bioinformatics institute of the european molecular biology laboratory (EMBL); FDA, federal drug administration; HUGO, human genome nomenclature committee; RO5, pfizer’s rule of five; IC50, inhibitory concentration 50%; ZVD, zika virus disease; ZIKV, zika virus; PDB, protein database

Introduction

The Zika virus (ZIKV), a member of the virus family Flaviviridae and genus Flavivirus, is related to dengue, yellow fever, Japanese encephalitis, West Nile and Sporvendi virus.1,2 It is primarily transmitted by female Aedes aegypti mosquitoes.1 Initially identified in Uganda,10 the ZIKV is rapidly spreading around the globe11–14 and is declared as a global public emergency by the World Health Organization (WHO report, 5 February, 2016). The flu-like symptoms are largely minor in most individuals and include fever, rash, arthralgia and conjunctivitis. However, in pregnant women the Zika virus disease (ZVD) is a cause for major concern as it is strongly suspected to be associated with the neurological disorders Guillain-Barré syndrome and microcephaly.15–19 While a strong causal link between the ZIKV and these disorders is yet to be established, a recent case controlled study provides a first evidence for association of the ZIKV with Guillain-Barré syndrome.20 Possible links to other neurological disorders are beginning to emerge.21 Further, the detection of the ZIKV in diverse body fluids including amniotic fluid,22 breast milk23 and semen24,25 increases the risk of transmission.

Currently, no treatment or diagnosis for ZVD is available. Relatively few publications exist on ZIKV and most of them revolved around case studies.26 The molecular cloning of the ZIKV and protein motifs and domains characterization,27 can facilitate rapid diagnostics and therapeutics based on the viral targets. As an obvious approach, vaccines are being explored around the globe. However, additional therapeutic approaches including repurposing/rescheduling of drugs are needed to combat this global emergency.

Chemoinformatics is emerging to be a powerful approach to harness the target protein structures into chemical leads.24 The protein 3 D structures can be subjected to ligand-based druggability analysis and bioactive hit compounds can be identified from the chEMBL repository.28 The hit compounds can be subjected to Pfizer’s Rule of Five (RO5) to identify drug-like orally bioavailable small molecular weight lead compounds with no toxicophores.29–32 The lead compounds can be rapidly tested in cell culture models, enabling accelerated drug discovery. Utilizing such an approach, recently lead compounds for Ebola virus, were discovered.33,34

A recently described database of ZIKV related genes35,36 were subjected to chemoinformatics analysis using the canSAR integrated protein annotation tool.37 Using ligand-based druggability analysis scores, RO5 compliant active hit compounds (IC50<100nM) for 35 ZIKV targets were identified. For 9 of these targets encompassing enzymes, cytokines, cytokine/ kinase receptors and phosphoprotein, toxicophore negative lead compounds emerged. The chEMBL compounds with unique target specificity were also identified for eight target proteins (AXL, CCL2, CCR5, CASP1, CTSS, CXCL8, NPM1 and TYRO3) encompassing enzymes, kinase receptors, phosphoproteins, cytokines and cytokine receptors. The drug-like compounds discovered in the study provide a rationale for verification in ZIKV assay systems for therapeutic leads discovery.

Methods

The bioinformatics and proteomics tools used in the study have been described in detail.33–38 The protein annotation and chemoinformatics analysis of the ZIKV-related genes was performed using the canSAR
2.0 integrated knowledgebase, a publically available database. The Browse canSAR analysis tool was used and the ZIKV-associated proteins were batch analyzed for protein annotations, 3D structures, compounds and bioactivity-related information. The Swiss Protein Database was used to obtain protein 3D structure template models related information. The chemical structures were obtained from the chEMBL database. Comprehensive gene annotation and phenotype analysis for the ZIKV-associated genes was established using the GeneCards Suite comprising the GeneAlaCart, GeneAnalytics and the VarElect tools, and the UniProt database and The Drug Bank. Protein expression was verified using the Human Protein Map, the Human Protein Atlas, Proteomics DB and the Multi Omics Protein Expression Database.

The hit compounds were subjected to diverse filters from the canSAR bioactivity analysis tool such as activity and assay types, concentrations, molecular weight, RO5 violations, prediction of oral bioavailability and toxicophores. Putative drug hits were filtered from the canSAR datasets for the ZIKV-associated genes using Lipinski’s rule of five (also known as Pfizer’s rule of five). RO5. The RO5 is a rule of thumb to evaluate druggableness and to determine whether a compound with a certain pharmacological or biological activity possesses properties that would make it a likely orally active drug in humans. For lead identification, highest stringency was chosen for the RO5 violation (value=0). Drugs with IC50 values, inhibitory activities and Ki values are chosen for the canSAR output. Toxicophore negative output was chosen to filter the hits for toxicity associated compound structures and was verified from the chEMBL database.

Results

Chemoinformatics analysis of the ZIKV-associated genes

Relatively little information is available for the molecular targets involved in the ZIKV infection. Two recent reports have identified putative candidate genes in the host cell genome for the ZIKV. In one study using bioinformatics approaches, 55 host cell targets were identified for the Zika-related viruses. Based on these targets, a pipeline of 79 FDA approved drugs was predicted for therapeutic use in the ZVD. Another study, using a ZIKV infected cell culture model, additional ZIKV-associated genes were described. In an effort to establish druggableness of these genes, a working database of 69 genes was generated and was subjected to protein annotation analysis from the canSAR bioinformatics tool (Figure 1A). Major classes of druggable proteins identified included 1) enzymes encompassing deaminase (APOBEC3), helicases (DDX42, DDX58, IFIHI), peptidases (CASPI, CTSS, FURIN, PLAT), oleno adenylate synthases (OAS1, OAS2, OAS3), GTPase (KRAS), phosphorylase (PPIB), protein kinases (EIF2AK2, RAF1, TYRO3), synthases (PT1, OAS1, OAS2, OAS3), GTPase (KRAS), phosphorylase (PPIB), synthase (PTS); 2) secreted proteins including cytokines (IL1B, IL6, IL8, IL11, LTA, TNF), chemokines (CCL2, CCL5, CXCL10, CXCL11), CD antigens (CD40LG, CD209), enzyme (OAS1), clotting-related (PLAT, THPO), growth factors (THPO, VEGFA), Mannose-binding (MBL2) and Ubiquitin-like protein (ISG15); 3) receptors including tyrosine kinase (AXL, TYRO3), phosphatase (PTPN11), chemokine (CCR5), toll like (TLR3), hepatitis A virus (HAVCR1) and immunoglobulin (FCGR2A), 4) adhesion molecule (VCAM1) and interferon-related (A2M2, EIF2AK2, IFIH1, IFITM1, IFITM2, IFITM3, IFIT4, ISG20, MX1,TMEM173). Other classes of the proteins included adhesion molecule (VCAM1), antiviral signaling (MAVS), cell cycle-associated (CAPRIN1), immunoglobulin receptor (FCGR2A), lympohtoxin (LTA), major histocompatibility complex (HLA-A, HLA-DRB, MICB), neurological atrophy-related (ATN1), nucleotide binding (ABC7, CEBB1, MX1), receptor antagonist (IL1RN) and virus receptors (HAVCR2, IVNS1ABP). Eight of the ZIKV-associated proteins are current drug targets (CCR5, HLA-DRB1, IL6, LTA, PLAT, PPIB, TNF and VEGFA).

![Figure 1](image)

**Figure 1** Ligand-based druggability analysis of the ZIKV candidate genes.

The canSAR protein annotation tool was batch analyzed using the list of 69 ZIKV-associated genes. Major classes of the protein families are shown. Ligand-based druggability prediction (>60% and >90% confidence scores) is shown. Genes with active compounds and RO5 compliant compounds are shown. The numbers above the bars indicate the number of genes under each category.

The presence of druggable proteins such as enzymes, cytokines/chemokines and receptors among the ZIKV-associated proteins raises the possibility of small molecular weight ligands discovery. Hence, the ZIKV database of genes was subjected to ligand-based druggability analysis using the canSAR tool (Figure 1B). Among the 69 ZIKV-associated genes, 31 proteins had 3D structures with ligands information available. Using a ligand-based druggability score cutoff (>60%), 35 ZIKV-associated genes were predicted to be druggable. Eight of the ZIKV-associated proteins including an adhesion molecule (VCAM1), enzymes (CASPI, CTSS, RAF1 and TYRO3), a chemokine (CCR5), thrombin (THBD), and a proteoglycan (CSPG5) had a very high reliability score (>90%) for ligand-based druggability. This makes them ideal candidates for drug discovery efforts.

These proteins provided an initial working database of ZIKV-associated proteins for lead drug discovery. The complete canSAR output is shown in Supplemental Tables S1 & S2. A comprehensive bioinformatics analysis of the ZIKV-associated database of 69 genes is shown in Supplemental Tables S3. Bioactive hit compounds (<100nM) were identified for 17/69 ZIKV-associated genes. Fourteen of these proteins had RO5 compliant (small molecular weight and oral bioavailability) active hit compounds (<100nM). These hits provided a framework for lead compounds discovery for the ZIKV drug discovery efforts.

**ZIKV lead compounds discovery**

Among the database of ZIKV-associated 69 genes, 17 genes for which active hit compounds (<100nM) were available provided a basis for lead compound discovery. These genes included an adhesion molecule (VCAM1), enzymes (CASPI, CTSS, EIF2AK2, FURIN, PPIB, RAF1 and TYRO3), growth factor (VEGFA), chemokines (CCL2, CCL5), receptors (AXL, CCR5), cytokines (CXCL8, TNF), a phosphoprotein (NPM1) and a thrombin (PLAT). The RO5 compliant bioactive hit compounds for the 14 ZIKV-associated proteins were
further characterized to identify the most active compounds (Figure 2) (Supplemental Tables S4). Bioactive compounds (<1 nM) were identified for three enzymes (CASP1, CTSS, RAPE1), a chemokine receptor (CCR5) and a tyrosine kinase receptor (AXL). Further, a set of the most potent lead compounds (<0.1 nM) were also identified from the canSAR library of compounds for the CCR5 (receptor) and CTSS (enzyme) target genes.

**Figure 2** ZIKV active ROS compliant lead compounds.

From the ROS compliant active lead compounds for the ZIKV associated genes identified using the canSAR bioactivity profiling tool, the output was filtered using ROS value (zero), molecular weight (<500), alogP (<5) and activity (IC50/inhibition). The number of compounds identified for each gene is shown. Based on the potency of compounds, the numbers are color-coded.

**Drug leads for the ZIKV assays**

Compounds with toxicophores offer a poor rationale for drug development. Potential ZIKV clinical trial patients would encompass pregnant women for whom toxic side effects would be a major challenge. For an effective ZIKV therapeutics, the toxicity should be minimal. This can be accomplished by eliminating the hits containing toxicophores. The canSAR database has filters to eliminate toxicophores for the hit compounds. Use of these filters helped to reduce the ROS compliant hits to 26 putative lead compounds for testing in the laboratory (Table 1). Nine lead genes were identified which included CASP1, CTSS, EIF2AK2 (enzymes), CCL2 (chemokine), CXCL8 (cytokine), CCR5 (chemokine receptor), NPM1 (phosphoprotein), AXL (tyrosine kinase receptor) and TYRO3 (protein kinase receptor). CHEMBL compounds with unique protein target specificity were identified for the proteins AXL, CCL2, CCR5, CASP1, CTSS, CXCL8/L, NPM1/ALK and TYRO3. A lead compound that targets the AXL, receptor tyrosine kinase gene (CHEMBL 601789) is an FDA approved drug (Crizotinib/Xalkori) for neoplasms. Additional highly selective chEMBL compounds (CHEMBL2023349, CHEMBL3545236) for the AXL protein are also available and are in clinical trials for AML and non-small cell lung cancer. Another anticancer FDA approved kinase inhibitor (Gefitinib) targets the EIF2AK2 lead protein (CHEMBL939).

(Table 1) also shows the protein expression data for the ZIKV lead genes in the Zika target tissues, the dermal fibroblasts, keratinocytes and immature dendritic cells. Target gene expression was seen in keratinocytes (AXL, CXCL8, EIF2AK2, NPM1, TYRO3), fibroblasts (AXL, CASP1, CXCL8, NPM1, CCL2) and immature dendritic cells (CCR5, CCL2). As these tissues, in particular the dermal keratinocytes and fibroblasts, provide the first line of defense against ZIKV infection, drugs targeting these nine proteins offer a therapeutic potential.

**Implication of the ZIKV lead genes in other disorders**

ZIKV is related to other *flaviviruses* and ZVD is suspected to be associated with two neurological disorders, Guillain-Barré syndrome and microcephaly. Hence, the VarElect bioinformatics tool from the GeneCards Suite was used to analyze the nine ZIKV lead genes with active compounds and a phenotype association analysis was performed (Supplemental Tables S5). The ZIKV lead genes were found to be associated with dengue (CASP1, CCL2, CXCL8, EIF2AK2), Japanese encephalitis (EIF2AK2, NPM1), West Nile (CCR5, EIF2AK2) and yellow fever (CCR5) viruses. In addition, association was seen for microcephaly (EIF2AK2) and Guillain-Barré syndrome (CCR5). Further, the lead genes were implicated with virus attachment (AXL, CASP1, CCL2, CCR5, CTSS, CXCL8, EIF2AK2, NPM1, TYRO3), RNA virus infection (CASP1, CCL2), virus replication (CASP1, CXCL8, EIF2AK2, NPM1) and antiviral effects (CASP1, CCL2, CCR5, CXCL8, EIF2AK2, NPM1). These results suggest that the lead drugs from the study can be tested against multiple Zika-related viruses.

**Discussion**

ZIKV infection is a global healthcare emergency. Currently, no treatment is available for the disease. Various options are available to explore for the therapeutics of ZVD, such as targeting the ZIKV specific genes, development of neutralizing antibodies, vaccines and targeting host specific genes whose function may be required for virus entry, adsorption or replication. Efforts are underway for vaccines development around the globe. If the ZIKV acquires mutations, the efficacy of the vaccines may be compromised. Alternative therapeutic approaches are needed to address ZVD. Repurposing and rescheduling of current drugs is an effective approach.

Until recently, ZIKV infection has been a largely neglected tropical disease and publications on the molecular target(s) of the ZIKV have been nonexistent. However, recently efforts to identify host cell targets as well as to develop a rationale for repurposed drugs are beginning to shed light. Using bioinformatics approaches and the protein sequence homology across diverse *flaviviruses*, a database of candidate ZIKV-related genes and a pipeline of FDA approved drugs were developed. In another study, Hamel et al. showed that human dermal fibroblasts, epidermal keratinocytes and immature dendritic cells are permissive to the most recent ZIKV isolate, responsible for the epidemic in French Polynesia. This study also identified several entry and/or adhesion molecules including DC-SIGN, AXL and TYRO3 permitted ZIKV entry. Blocking of AXL gene expression, either using neutralizing antibodies or siRNA, abrogated the ZIKV infection, which suggested a drug therapy potential for the AXL gene. The list of genes from these two studies provided a framework for exploring druggability of the ZIKV associated genes in the human genome. From a database of 69 ZIKV-related genes, nine genes (AXL, CASP1, CCL2, CCR5, CTSS, CXCL8, EIF2AK2, NPM1 and TYRO3) were identified with active drug-like compounds (low molecular weight, orally bioavailable, no toxicophore, RO 5 compliant). The genes AXL Receptor Tyrosine Kinase/AXL and Eukaryotic Translation Initiation Factor 2-Alpha Kinase 2(EIF2AK2) are targets for the FDA approved drugs Crizotinib/Xalkor and Gefitinib respectively. In view of the knockout results on the AXL gene in the abrogation of ZIKV infection it is tempting to postulate a therapeutic usefulness of the Crizotinib/Xalkor drug for ZVD.

Among the nine ZIKV lead proteins, the eukaryotic translation initiation factor 2 alpha kinase 2 (EIF2AK2) gene was found to be associated with multiple variant phenotypes (RNA-helicase, *Aedes* mosquito, single strand RNA, dengue, Japanese encephalitis, West
Nile, yellow fever, microcephaly, virus attachment/repllication and antiviral). This gene is an IFN-induced dsRNA-dependent serine/threonine-protein kinase, which plays a key role in the innate immune response to viral infection and is also involved in the regulation of signal transduction, apoptosis, cell proliferation and differentiation. It exerts its antiviral activity on a wide range of DNA and RNA viruses including hepatitis C virus (HCV), hepatitis B virus (HBV), measles virus (MV) and herpes simplex virus 1 (HHV-1). The first FDA approved selective inhibitor of the EGFR tyrosine kinase for neoplasms, (Gefitinib /EGFR inhibitor) targets the EGF2AK2 protein CHEMBL939. This drug offers an additional potential for therapeutic evaluation against the ZIKV infection. Another interesting target emerging from this study is the protein Tyrosine Kinase| TYRO3. This gene acts as a receptor for lassa virus and lymphocytic choriomeningitis virus, possibly through Growth Arrest-Specific 6/GAS6 binding to phosphatidyl-serine at the surface of the virion envelope. TYRO3 also acts as a receptor for ebola virus, possibly through GAS6 binding to phosphatidyl-serine at the surface of the virion envelope.

Hammel et al. have recently showed that the ZIKV entry is mediated by CD209, AXL, Tyro3 proteins and to a lesser extent, by TIM-1. They hypothesized that a cooperative interaction exists between both AXL and TIM-1 receptors. The TIM-1 acting as an attachment factor binds to ZIKV and transfers them to AXL, which internalizes in the ZIKV. The availability of lead compounds for AXL and TYRO3 from the current study allows verification of this premise. The chemokine co-receptor CCR5 is required for HIV and West Nile viruses to enter the host cells. Thus, it is reasonable to hypothesize that the lead compound against the CCR5 identified in the study might prevent ZIKV entry into the susceptible cells. Furthermore, the infection of skin fibroblasts by ZIKV resulted in the upregulation of the inflammatory chemokine CCL2 expression. The inflammation, which often accompanies the ZVD could be targeted by the CCL2 lead compound. The flaviviruses, such as West Nile and Dengue, infect the skin keratinocytes. The ZIKV and Dengue virus infection induce the appearance of apoptotic cells in the epidermis of infected human skin explants. Thus, inhibitors of apoptosis such as the CASP1 lead compound identified in the study could offer a therapeutic advantage against the flaviviruses. The neurological disorders such as Guillain-Barré syndrome and microcephaly have a complex etiology and their precise molecular targets are unknown. Currently, model systems are not available for either of these disorders. Verification of the lead compounds discovered in this study awaits future development of such a model system.

The database of ZIKV-related genes and the list of hit compounds generated in this study provide a starting point for accelerated drug discovery for ZVD. Efforts are underway to develop mouse models for the ZIKV infection. However the cell culture model described by Hamel et al., provides an in vitro model system to rapidly verify these compounds. A major challenge facing ZIKV therapeutics is that the main cohort of ZVD patients involves pregnant women. The toxicity and side effects of any therapeutics including the vaccines need to be evaluated carefully during pregnancy. Specific guidelines are available for testing clinical candidate drugs in pregnant women. Antiviral, antidiabetic and anticancer drugs are being used during pregnancy.

**Conclusion**

A database of 69 ZIKV-related genes provided a starting point for ZIKV drug discovery. A pipeline of 26 ZIKV therapeutic drug-like RO5 compliant, toxicophore negative lead compounds targeting nine ZIKV-related host cell proteins were discovered using chemoinformatics approaches. Two of the lead compounds for the targets AXL and EGF2AK2 are FDA approved drugs for cancer. The AXL lead compound can be tested in the cell culture model of Hamel et al., to verify the abrogation of ZIKV infection by knockout experiments with the AXL gene. The drug-like compounds discovered in the study provide a rationale for rapid testing in a ZIKV infected cell culture model and for investigation of the structure-activity relationship for lead verification and optimization.

**Acknowledgements**

This work was supported in part by the Genomics of Cancer Fund, Florida Atlantic University Foundation. I thank the canSAR database for data access and Je anne Narayanan for editorial assistance.

**Conflict of interest**

The author declares no conflict of interest.

**References**


---

**Citation** Narayanan R. Zika virus therapeutic lead compounds discovery using chemoinformatics approaches. *MOJ Proteomics Bioinform*. 2016;3(3):68–73.

DOI: 10.15406/mojpb.2016.03.00088
screen of approved drugs. 


