**From Innovation to Application**

**OpenZika: An IBM World Community Grid Project to Accelerate Zika Virus Drug Discovery**

Sean Ekins1\*, Alexander L. Perryman2\* and Carolina Horta Andrade3\*

## 1 Collaborations Pharmaceuticals, Inc., 5616 Hilltop Needmore Road, Fuquay-Varina, North Carolina 27526, United States.

2 Department of Pharmacology, Physiology and Neuroscience, Rutgers University–New Jersey Medical School, Newark, New Jersey 07103, United States.

**3**LabMol - Laboratory for Molecular Modeling and Drug Design, Faculdade de Farmácia, Universidade Fedral de Goiás, Goiânia, Goiás 74605-170, Brazil.

**Email:** Carolina Horta Andrade: [andradech@yahoo.com](mailto:andradech@yahoo.com), Alexander L. Perryman: [Alex.L.Perryman@njms.rutgers.edu](mailto:Alex.L.Perryman@njms.rutgers.edu), Sean Ekins: [ekinssean@yahoo.com](mailto:ekinssean@yahoo.com)

**Running title:** OpenZika

**Box**

Advantages

* Open Science could discover new antivirals using docking
* Narrows down compounds to test and saves time
* Free to use distributed computing

Disadvantages

* Concern around intellectual property ownership and whether companies will develop drugs coming from effort
* Need for experimental testing not removed
* Testing in vitro and in vivo is not free

**Abstract**

The Zika virus outbreak in the Americas has caused global concern, and to help this fight against Zika we launched the OpenZika project. OpenZika is an IBM World Community Grid Project that uses distributed computing on millions of computers and Android devices to run docking experiments, in order to dock millions of drug-like compounds against crystal structures and homology models of Zika proteins (and other related flavivirus targets). This will enable the identification of new virtual candidates that can then be tested *in vitro*, to advance the discovery and development of new antiviral drugs against the Zika virus. The docking data will be made openly accessible so that all members of the global research community can use it to further advance drug discovery studies for Zika and other viruses.

**Introduction**

The Zika virus (ZIKV) has emerged as a major public health threat to the Americas since 2015 [1]. We have previously suggested that it represents [2] an opportunity for scientific collaboration and open scientific exchange, which we have witnessed first-hand since then. The health of future generations may very well depend on the decisions we make in the coming months and our willingness to share our findings quickly, openly and collaborate to rapidly find a cure for this disease. Since February 1st, 2016, when the World Health Organization deemed the cluster of microcephaly cases, Guillain-Barré, and other neurological disorders reported as associated with ZIKV in Latin America and the Caribbean as constituting a Public Health Emergency of International Concern [3] (PHEIC), we have seen a rapid increase in publications. These have strengthened the link of these disorders with ZIKV [4-6], identified a potential entry receptor as AXL in neural stem cells [7] and demonstrated *in vitro* and *in vivo* models that could be used for testing potential small molecule therapeutics and vaccines [8].

At least one preliminary study has suggested that chloroquine may have some low micromolar activity *in vitro* [9], although much more active compounds would be desirable that are also safe during pregnancy. We [2] and others [10,11] have also described steps that could be taken to initiate a drug discovery program on ZIKV; for example, we proposed the use of computational approaches such as virtual screening of chemical libraries or focused screening to repurpose FDA and/or EU approved drugs [12] to accelerate the discovery of an anti-ZIKV drug. The initiation of an antiviral drug discovery program can be done using structure-based design, based on homology models of the key ZIKV proteins. With the lack of structural information regarding the proteins of ZIKV, we decided to build homology models for all the ZIKV proteins, based on close homologs such as Dengue. Freely available online software was used to create these homology models of NS5, FtsJ, HELICc, DEXDc, Peptidase S7, NS1, E Stem, Glycoprotein M, propeptide, capsid and glycoprotein E ZIKV proteins [13] (Table S1) as well as to illustrate the complete virion, which was made available on line on March 3rd, 2016. We also predicted the site of glycosylation of glycoprotein E as Asn154, and this was recently experimentally verified [14].

By the end of March 2016, there were no crystal structures or cryo-EM data of the ZIKV. Since then, we have to date seen 2 cryo-EM structures at 3.8 Å [14] (PDB 5IRE) and 3.7 Å [15] (PDB 5IZ7) and multiple crystal structures of NS1 (PDB 5IY3) [16], glycoprotein E in complex with an antibody [16] (PDB 5JHL and 5JHM) and the helicase domain of NS3 [17] (PDB 5JMT). These structures, alongside the developed homology models, represent potential starting points for structure-based drug design efforts, such as docking-based virtual screening, with the goal being to find molecules that are predicted to have high affinity with these proteins. These predictions will then be tested against the virus in cell-based assays and/or using individual protein-based assays. Obviously, there are millions of molecules available that can be assayed, but which ones are likely to work, and how should we prioritize them?

Since March, we have been preparing a new open collaborative project called OpenZika (Figure 1), for IBM’s World Community Grid (WCG, [worldcommunitygrid.org](https://www.worldcommunitygrid.org/)). WCG is an IBM philanthropic initiative that allows anyone with a computer (Mac, Linux, or Windows) or Android device (smartphone or tablet) to donate their devices’ unused processing power towards scientific research on health, poverty and sustainability. More than three million computers and mobile devices used by nearly 750,000 people and 470 institutions across 80 countries have contributed virtual supercomputing power for more than two-dozen vitally important projects on World Community Grid over the last 11 years, at a value of more than $100 million. WCG is enabled by Berkeley Open Infrastructure for Network Computing (BOINC) [18], and with support from the [National Science Foundation](http://www.nsf.gov/) [19]. WCG has helped researchers identify potential drugs or new chemical tools for other neglected and emerging diseases, such as Malaria (GO Fight Against Malaria) [20], Leishmaniasis (Drug Search for Leishmaniasis) [21], Schistosomiasis (Say No to Schistosoma) [22], and Ebola (Outsmart Ebola Together) [23].

On May 18th, 2016, we launched the OpenZika project [24,25] with the proposed virtual screening of ~20 million compounds that are in the ZINC [26,27] database (Figure 1), as well as all of the FDA approved drugs and the NIH clinical collection [28]. These screens are being performed with the AutoDock Vina software [29-31] and the homology models and crystal structures (Table S1), to discover novel candidate compounds that can potentially be developed into new drugs for curing ZIKV. These will be followed by additional Virtual Screens with the whole PubChem database (at most ~90 million compounds) [32], after their structures are prepared for docking (and filtered to remove compounds that contain too many rotatable bonds and those with atom types that AutoDock Vina does not recognize). We have already prepared the docking input files for 5.6 million compounds from ZINC (*i.e.,* the libraries that were previously used in the GO Fight Against Malaria project on World Community Grid), [30] which are currently being used in the initial set of virtual screens on OpenZika. These docking input files (“pdbqt” files) are also openly accessible (<http://zinc.docking.org/pdbqt/)>. Initially, compounds will be screened against the ZIKV homologs of drug targets that have been well-validated in research against dengue virus and hepatitis C virus, such as NS5 and Glycoprotein E. Both the mature and immature protein conformers of Glycoprotein E can be used as targets, to study how this affects the nature of the different hits identified. In addition, we are also docking compounds against the crystal structures of the virus templates that were used to create these homology models and additional crystal structures of proteins (from pathogens) that have high sequence identity to the ZIKV targets. These might allow us to identify broad-spectrum antivirals against multiple flaviviruses, such as dengue and yellow fever. In addition, docking against the crystal structure of a related protein from a different pathogen can sometimes discover novel hits against the pathogen of interest, without even using a homology model [33].

As well as applying docking-based filters, the promising candidate compounds identified on OpenZika will also be filtered using machine learning models, which we have developed for predicting some pharmacokinetic and toxicity properties, including metabolic stability [34-36] and hERG blockage [37,38]. These should be useful selection criteria to enhance the identification of promising compounds for subsequent tests by our collaborators in whole-cell ZIKV assays, in order to verify their antiviral activity for blocking ZIKV infection or replication. Since all OpenZika data will be in the public domain (we will release the docking results to the public as soon as they are completed), we and other labs can then advance the development of some of these new virtual hits into experimental hits, leads and drugs. This exemplifies open science, which should help scientists around the world as they address the long and arduous process of discovering and developing new drugs. Performing OpenZika on World Community Grid could identify compounds that not only target ZIKV but also other flaviruses. Screening millions of compounds against many different protein models in this way would take far more resources and time than any academic researcher could generally obtain or spend. Without the unique community of volunteers and tremendous resources provided by World Community Grid, this project would have been very difficult to initiate in a reasonable time frame.

The potential challenges we foresee will be finding laboratories with sufficient funding to pursue compounds, synthesize analogs, and develop target-based assays, to validate our predictions and generate SAR (Structure Activity Relationship) data to guide the process of developing the new hits into leads and then drugs. In addition, funding of ZIKV research will likely need to be sustained for several years if not longer (e.g., HIV research has been funded for decades). As with other WCG projects, once scientists identify experimentally validated leads, finding a company to license them and pursue them further will need incentives such as the FDA Tropical Disease Priority voucher, [39] which has a financial value on the open market [40].

By our and other groups working together and opening our work to the scientific community, many other researchers in the world will also be able to take promising molecular candidates forward, to accelerate progress towards defeating the ZIKV outbreak. We invite other interested researchers to join us, and we hope new volunteers will donate their dormant, spare computing cycles to this cause [41]. We expect to report the computational and experimental results of this collaboration in due course.

**Acknowledgements**

We kindly acknowledge all our collaborators on this project (<http://openzika.ufg.br)> and the wonderful support and encouragement of the IBM WCG team and the community of volunteers for donating their computer time.

Wrote the paper: SE ALP CHA

**Supporting Information**

A listing of all models which are already publically accessible are available in supporting information.

**References**

1. Faria NR, Azevedo Rdo S, Kraemer MU, Souza R, Cunha MS, Hill SC, et al. Zika virus in the Americas: Early epidemiological and genetic findings. Science. 2016; 352:345-9.

2. Ekins S, Mietchen D, Coffee M, Stratton TP, Freundlich JS, Freitas-Junior L, et al. Open drug discovery for the Zika virus. F1000Res. 2016; 5:150.

3. Anon. WHO Director-General summarizes the outcome of the Emergency Committee regarding clusters of microcephaly and Guillain-Barré syndrome. 2016. <http://who.int/mediacentre/news/statements/2016/emergency-committee-zika-microcephaly/en/>.

4. Rasmussen SA, Jamieson DJ, Honein MA, Petersen LR. Zika Virus and Birth Defects--Reviewing the Evidence for Causality. N Engl J Med. 2016; 374:1981-7.

5. Mlakar J, Korva M, Tul N, Popovic M, Poljsak-Prijatelj M, Mraz J, et al. Zika Virus Associated with Microcephaly. N Engl J Med. 2016; 374:951-8.

6. Cugola FR, Fernandes IR, Russo FB, Freitas BC, Dias JLM, Guimaraes KP, et al. The Brazilian zika virus strain causes birth defects in experimental models. Nature. 2016; In Press.

7. Nowakowski TJ, Pollen AA, Di Lullo E, Sandoval-Espinosa C, Bershteyn M, Kriegstein AR. Expression Analysis Highlights AXL as a Candidate Zika Virus Entry Receptor in Neural Stem Cells. Cell Stem Cell. 2016; 18:591-6.

8. Shan C, Xie X, Muruato AE, Rossi SL, Roundy CM, Azar SR, et al. An Infectious cDNA Clone of Zika Virus to Study Viral Virulence, Mosquito Transmission, and Antiviral Inhibitors. Cell Host Microbe. 2016.

9. Delvecchio R, Higa LM, Pezzuto P, Valadao AL, Garcez PP, Monteiro FL, et al. Chloroquine inhibits Zika Virus infection in different cellular models. bioRxiv. 2016.

10. Shan C, Xie X, Barrett ADT, Garcia-Blanco MA, Tesh RB, Fernando da Costa Vasconcelos P, et al. Zika Virus: Diagnosis, Therapeutics, and Vaccine. ACS Infect Dis. 2016; In Press.

11. Malone RW, Homan J, Callahan MV, Glasspool-Malone J, Damodaran L, Schneider Ade B, et al. Zika Virus: Medical Countermeasure Development Challenges. PLoS Negl Trop Dis. 2016; 10:e0004530.

12. Ekins S, Williams AJ, Krasowski MD, Freundlich JS. In silico repositioning of approved drugs for rare and neglected diseases. Drug Discov Today. 2011; 16:298-310.

13. Ekins S, Liebler J, Neves BJ, Lewis WG, Coffee M, Bienstock R, et al. Illustrating and homology modeling the proteins of the Zika virus. F1000Res. 2016; 5:275.

14. Sirohi D, Chen Z, Sun L, Klose T, Pierson TC, Rossmann MG, et al. The 3.8 A resolution cryo-EM structure of Zika virus. Science. 2016; 352:467-70.

15. Kostyuchenko VA, Lim EX, Zhang S, Fibriansah G, Ng TS, Ooi JS, et al. Structure of the thermally stable Zika virus. Nature. 2016; 533:425-8.

16. Song H, Qi J, Haywood J, Shi Y, Gao GF. Zika virus NS1 structure reveals diversity of electrostatic surfaces among flaviviruses. Nat Struct Mol Biol. 2016; 23:456-8.

17. Tian H, Ji X, Yang X, Xie W, Yang K, Chen C, et al. The crystal structure of Zika virus helicase: basis for antiviral drug design. Protein Cell. 2016.

18. Anon. Open-source software for volunteer computing 2016. <http://boinc.berkeley.edu/>.

19. Anon. National Science Foundation. 2016. <http://www.nsf.gov/>.

20. Anon. GO fight against malaria. 2016. <http://www.worldcommunitygrid.org/research/gfam/overview.do>.

21. Anon. Drug search for leishmaniasis. 2016. <http://www.worldcommunitygrid.org/research/dsfl/overview.do>.

22. Anon. Say no to schistosoma. 2016. <http://www.worldcommunitygrid.org/research/sn2s/overview.do>.

23. Anon. Outsmart Ebola together. 2016. <http://www.worldcommunitygrid.org/research/oet1/overview.do>.

24. Anon. Zika Treatment Search Launched, Fueled By IBM's World Community Grid. 2016. <http://www-03.ibm.com/press/us/en/pressrelease/49778.wss>.

25. Andrade CH. Help an international research team fight the zika virus. 2016. <http://citizenibm.com/2016/05/andrade_zika_wcgrid.html>.

26. Irwin JJ, Shoichet BK. ZINC--a free database of commercially available compounds for virtual screening. J Chem Inf Model. 2005; 45:177-82.

27. Irwin JJ, Sterling T, Mysinger MM, Bolstad ES, Coleman RG. ZINC: a free tool to discover chemistry for biology. J Chem Inf Model. 2012; 52:1757-68.

28. Anon. NIH Clinical Collection. 2014. <http://www.nihclinicalcollection.com/>.

29. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010; 31:455-61.

30. Perryman AL, Yu W, Wang X, Ekins S, Forli S, Li SG, et al. A Virtual Screen Discovers Novel, Fragment-Sized Inhibitors of Mycobacterium tuberculosis InhA. J Chem Inf Model. 2015.

31. Perryman AL, Santiago DN, Forli S, Santos-Martins D, Olson AJ. Virtual screening with AutoDock Vina and the common pharmacophore engine of a low diversity library of fragments and hits against the three allosteric sites of HIV integrase: participation in the SAMPL4 protein-ligand binding challenge. J Comput Aided Mol Des. 2014; 28:429-41.

32. Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, et al. PubChem Substance and Compound databases. Nucleic Acids Res. 2016; 44:D1202-13.

33. Kincaid VA, London N, Wangkanont K, Wesener DA, Marcus SA, Heroux A, et al. Virtual Screening for UDP-Galactopyranose Mutase Ligands Identifies a New Class of Antimycobacterial Agents. ACS Chem Biol. 2015; 10:2209-18.

34. Perryman AL, Stratton TP, Ekins S, Freundlich JS. Predicting Mouse Liver Microsomal Stability with "Pruned" Machine Learning Models and Public Data. Pharm Res. 2016; 33:433-49.

35. Clark AM, Dole K, Coulon-Spektor A, McNutt A, Grass G, Freundlich JS, et al. Open Source Bayesian Models. 1. Application to ADME/Tox and Drug Discovery Datasets. J Chem Inf Model. 2015; 55:1231-45.

36. Ekins S. Progress in computational toxicology. J Pharmacol Toxicol Methods. 2014; 69:115-40.

37. Braga RC, Alves VM, Silva MF, Muratov E, Fourches D, Tropsha A, et al. Tuning HERG out: antitarget QSAR models for drug development. Curr Top Med Chem. 2014; 14:1399-415.

38. Braga RC, Alves VM, Soilva MFB, Muratov E, Fourches D, Liao LM, et al. Pred-hERG: A Novel web-Accessible Computational Tool for Predicting Cardiac Toxicity. Mol Informatics. 2015; 34:698-701.

39. Text of the Adding Zika Virus to the FDA Priority Review Voucher Program Act. 2016. <https://www.govtrack.us/congress/bills/114/s2512/text>.

40. Ekins S, Wood J. Incentives for Starting Small Companies Focused on Rare and Neglected Diseases. Pharm Res. 2016; 33:809-15.

41. Andrade CH, Perryman AL, Ekins S. OpenZika. 2016. <https://www.worldcommunitygrid.org/research/zika/overview.do>.

**Figure Legends**

Figure 1. Workflow for the OpenZika project. OpenZika runs on IBM’s World Community Grid, which provides distributed computing for docking commercially available compounds into multiple ZIKV and related virus homology models and crystal structures using AutoDock Vina (AD Vina). This ultimately produces computational hits that score well against individual proteins, which can then be prioritized for *in vitro* testing.