“Predicting the physical structure of viruses from genomic data”

Antoni Luque
San Diego State University

INTRODUCTION

1. What are you trying to do? Obtain high-resolution molecular models of the physical structure of icosahedral viruses from genomic data using biophysical models and computational methods.

2. How is it done today? Imaging experimental methods, such as cryo-electron microscopy and X-ray, are currently employed to obtain high-resolution three-dimensional models of the physical structure of viruses. These methods require expensive technology and training, can take months or years for a single virus, and can be only applied when the viral host is known, the virus can be produced at high concentrations, and the viral particles can resist the required physico-chemical steps in the protocols. In the case of X-ray, it also requires the formation of crystals, which is an unpredictable process that is virus-dependent.

3. What is new in your approach? My approach relies on several findings lead by my research lab. We have recently developed a new mathematical framework to characterize icosahedral capsids published in Nature Communications (Twarock and Luque, 2019). We have also strong preliminary showing strong allometric relationships for different viral structural lineages and showing that single structural genes, like the major capsid protein, alone can predict the architecture of viruses. Our current models for the HK97 viral structural lineage show that we can predict the architecture of these viruses with 80% accuracy from the major capsid protein and we can reach 95% accuracy when including additional genomic information. Our hypothesis is that the other viral structural lineages will display analogous relationships that are specialized in each case. Thus our current results could be expanded to the rest of the virosphere.

4. Who cares? Structural biologists, ecologists, virologists.

5. What difference will it make? The platform proposed here would facilitate the prediction of the molecular physical structure of viruses as soon as there is genomic information available. That would provide an important leap when screening for emerging infectious diseases and will guide and complement the high-resolution methods to reconstruct viral capsids. It will also provide an estimate of the structure for viruses that have not been cultured, which would facilitate their isolation and host targeting. This would significantly expand our ability to understand why the physical structure of the 10^31 viruses on the planet has adapted to different environments and how they keep changing as climate and environmental conditions change. The information necessary to achieve this is currently available from most environments through sequencing, but we do not have the bioinformatic tools to translate that genomic information into specific viral structures.

6. What are the risks and payoffs? Our preliminary technology is well established for the HK97 viral lineage, in particular, for tailed phages. These are the most abundant viruses in the world and are also involved in infectious disease, carrying toxins and antibiotic resistance genes in pathogenic bacteria. The development of our bioinformatic pipeline and refinement of our methods for this group of viruses would be already an important advance in the field. If our hypothesis is true, that would mean that we could rapidly extend our technology to other icosahedral capsids, which represent the majority of viral taxa in the virosphere. This would also lay out the ground to approach non-icosahedral structures. In particular, elongated and conical structures would be an immediate target because they share similarities with

icosahedral capsids. The main risk would be that our hypothesis is not true. That, however, would open very interesting physical and biological questions: how is it possible that the fold of the major capsid protein of viral structural lineages has been conserved while the capsid structure is uncorrelated? What are the selection pressures that would be in play for the major capsid protein separately from the actual assembly, mechanical stability, and architecture of viral capsids?

7. How much will it cost? The development of the basic pipeline for tailed phages will cost $100,000. The expansion to the other ten icosahedral viral structural lineages will cost $500,000. This will involve two PhD students, a handful of undergraduate students, and summer work from two faculty. The gateway will be funded separately by the NSF XSEDE program.

8. How long will it take? The basic pipeline for tailed phages will be implemented in the first year and a half. Extending the bioinformatic pipeline to the other viral lineages will require the full three years of the grant contract.

9. What are the midterm and final check points for success? The first milestone will be the release of the gateway to predict the physical structures of tailed phages. The second milestone will be to test the initial pipeline with new high-resolution tailed phage structures reconstructed by our collaborator Simon White at the University of Connecticut. The third milestone will be to receive the approval for usability from our external committee formed by a structural biologist, ecologist, and virologist. The final milestone will be to disseminate the gateway to reach 100 users by the end of the grant.