In Silico Vaccine Design through Empirical Fitness Landscapes and Population Dynamics

PI: Professor Andrew Ferguson (corresponding author)
Materials Science and Engineering
University of Illinois at Urbana-Champaign
alf@illinois.edu

Collaborators: Greg Hart
Physics
University of Illinois at Urbana-Champaign
grhart2@illinois.edu

Chin-Yu Chester Cheng
Computer Science
University of Illinois at Urbana-Champaign
ccheng32@illinois.edu

Executive Summary

Hepatitis C virus (HCV) affects 170 million people worldwide, and kills 700,000 annually. Vaccination provides the most realistic and cost-effective hope of controlling this epidemic, but no vaccine is available. Computational models can offer rational precepts to inform and accelerate vaccine design. We have developed a computational tool to translate databases of viral sequences into "fitness landscapes" mapping the replicative capacity of the virus as a function of its genome. These landscapes represent the mutational playing field over which the virus evolves. By integrating these landscapes with agent-based models of viral mutation and host immune response, we have explicitly modeled the host-pathogen dynamics over its empirically-defined fitness landscape. Using this simulator, we have employed the hardware resources of Blue Waters to perform computational screening of candidate vaccine components to identify those best able to cripple viral fitness and block immune escape. These findings can inform next-generation HCV vaccine design.

Description of Research Activities and Results

Key Challenges

Hepatitis C virus (HCV) continues to pose a global threat to public health. Despite the availability of efficacious drug treatments in the developed world, the high cost of these therapies make them effectively unavailable in the developing world where the preponderance of infections occur. A prophylactic vaccine represents the most cost-effective and realistic strategy to combat the epidemic, but despite 25 years of research one is still not available. A challenge in vaccine design is the identification of promising targets within the virus that can be targeted by a vaccine that simultaneously cripple viral fitness and are not subject to facile mutational escape. Computational models of viral infection and the host immune response can systematically

identify promising targets that may be translated into rational precepts for experimental development and testing of HCV vaccines.

Why it Matters

The development and validation of robust and accurate computational simulators of viral mutation and host immune pressure can serve to both illuminate the molecular mechanisms of viral behavior, and identify vulnerable targets with translational impact in guiding experimental vaccine design. The results of this work enabled by the Blue Waters Exploratory Allocation have demonstrated proof-of-principle of our immune simulator, validated our model predictions against experimental data, and enabled computational evaluation of candidate vaccine formulations against one of ten HCV proteins. These results have enabled the development of one publication to be submitted and set the stage for future proposals to federal funding bodies to support this work, and to Blue Waters to support the development of similar simulators for the remaining nine HCV proteins.

Why Blue Waters

The simulations of the viral mutational evolution over our previously parameterized viral fitness landscapes is implemented via an agent based model comprising N=50,000 distinct viral sequences, each of which is described by a m≈300 vector characterizing the viral genome. Each virus in the ensemble produces p=10 daughter virions in each mutation step, and we subsequently downselect the viral population from pN back down to N according to a stochastic fitness-weighted procedure in which sequences that reside in fitter regions of the landscape possess a higher chance of survival. These are known as Wright-Fisher dynamics, and provide a good description of viral mutation-selection.

The host immune response is described by a set of ordinary differential equations modeling the dynamics of the host T-cells as they recognize the virus, activate, mature, proliferate, and die. The coupling to the viral dynamics occurs through one term imposing a penalty on the fitness of viral strains that are recognized and attacked by particular members of the T-cell population, and through a recognition term in which T-cells that recognize particular viral strains are primed to activate and proliferate. The relatively small T-cell populations within our control volume mean that fluctuations are important, and we implement a stochastic integration protocol via Gillespie dynamics to explicitly capture these effects.

The agent-based model for the viral dynamics carry a high memory and computational cost due to the size of the viral population and the fitness-weighted selection protocol. Moreover, the stochastic nature of the viral and host dynamics means that multiple K=100 replicas must be run for each realization. Finally, screening over multiple viral candidates entails a separate suite of K=100 runs for each of C=35 infected hosts representative of the diversity of persons in the North American population for each vaccine candidate to be considered. Accordingly, we require the scale and parallelism available within Blue Waters to support the computational intensity of each single simulation, and also achieve the range of simulations to evaluate large numbers of vaccine

candidates in a variety of hosts. Furthermore, the volume of data generated is also significant as the viral sequences present at each time point must be written to disc. Empirical timing estimates show that to evaluate a single vaccine candidate in the C=35 hosts performing 150-generation runs with K=100 replicas requires around 525,000 CPU-hours and will produce a total of around 3.9 TB of data.

Accomplishments

We have demonstrated proof-of-principle of the technical feasibility of our approach, experimental validation of its predictions against published clinical time courses of viral evolution in infected hosts, and used our approach to evaluate and design vaccine candidates for two hosts. First, we successfully ported our code to Blue Waters and demonstrated its performance at scale using OpenMP, MPI, CUDA, and Boost libraries called from within our C++ code. Second, we used our simulator to predict the mutational time evolution of the virus within seven infected hosts for which clinical longitudinal data were available tracking the particular viral mutants observed in blood samples drawn as a function of time. In six of the seven cases, the predictions of our model were in excellent agreement with the clinical data, with statistically significant p-values ($p < 10^{-6}$). In the remaining case, the clinically observed T-cell responses are inconsistent with the reported immunological haplotype of the patient, suggesting that the patient had been immunologically mischaracterized in the clinical data. Third, we considered two representative hosts – 0684MX and BR554 – and used our simulator to predict the efficacy of the ensemble of all possible vaccine candidates consistent with the immunological genotypes of the hosts. In each case we identified a number of promising vaccine candidates that led to strong and durable responses by priming T-cells that imposed strong fitness penalties upon the viral population for long periods to time (**Fig. 1**). Interestingly, we also found vaccine candidates that led to *poorer* immune responses compared to no vaccination by priming the "wrong" T-cell responses to attack regions of the virus from which mutational escape is facile. These results lay the foundations for large-scale simulations of vaccine candidates for all representative hosts in the North American population for the particular protein considered, and for extending this work to ten HCV proteins.

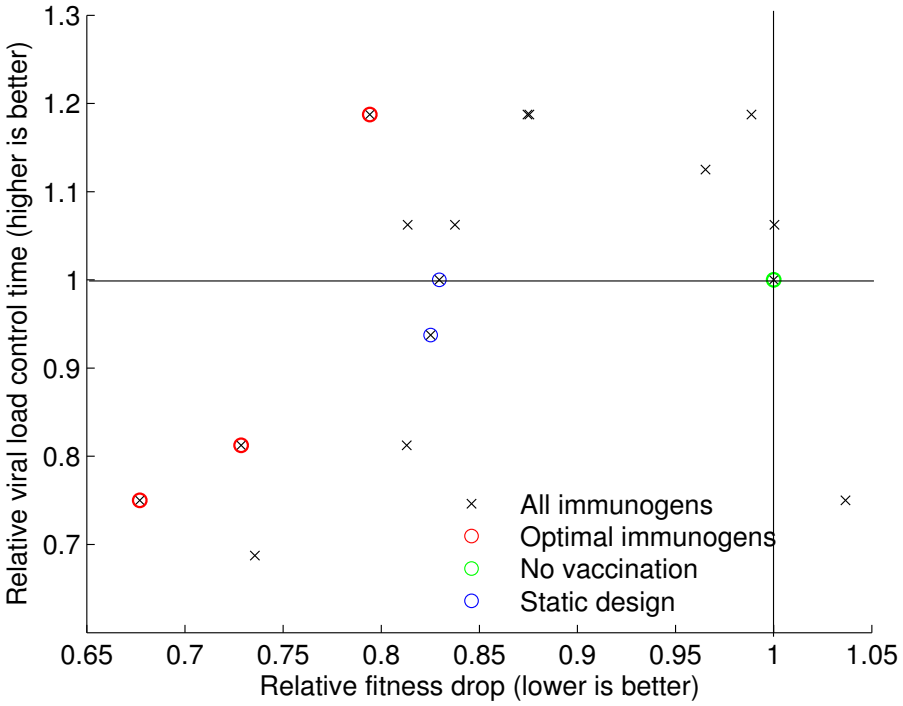

Figure 1. Computational evaluation of inoculation of patient 0684MX with the 15 vaccine candidates targeting the NS5B protein that are consistent with the host immunological genotype. The black crosses characterize the immune response of the host in terms of the mean fitness drop imposed upon the viral population and control time, both reported relative to the response in the absence of vaccination (green circle). The red circles indicate the Pareto optimal vaccine candidates evaluated according to these two measures. The blue circles indicate the vaccine candidate identified by the landscape alone, without considering the host-pathogen dynamics. Those vaccine candidates residing in the upper-left quadrant provide superior strength and length of control relative to no vaccination, and are good candidates for experimental testing.

List of publications, data sets associated with this work

G.R. Hart and A.L. Ferguson "Computational design of hepatitis C virus immunogens from host-pathogen dynamics over empirical viral fitness landscapes" (to be submitted, 2018)