An *in silico* Model for the Evolution of Immune System Evasion

Dhruv B. Pai[1], Max Wolf[2], Yuri Wolf[3]

[1]Montgomery Blair High School, Silver Spring, MD. [2]Harvard Medical School, Boston, MA.
[3]National Institutes of Health, Bethesda, MD.

Petrovax

SMTB

ZIMIN FOUNDATION

Abstract

Understanding evolutionary trajectories of viruses is vital for the proactive development of vaccines. Unfortunately, traditional methods for processing viral capsids suffer from a number of drawbacks. *In silico* models can provide a complement to longstanding approaches.

In this study, we develop a 5x5 lattice simplification of the viral capsid and simulate its evolution using a Markov Chain Monte Carlo (MCMC) model. We demonstrate the evolution of a population of viruses away from antibody binding and towards capsid stability. We also show that this simple, stochastic model can result in a wide diversity of evolutionary trajectories and escape curves.

Our results have implications for the development of therapeutics for novel strains of rapidly-mutating viruses, like COVID-19 or Influenza. In the future, we hope to increase the biological accuracy of our model by implementing non-weak selection regimes and three dimensional lattice capsids.

Introduction

The varying evolutionary patterns of viruses have a number of implications for medicine, particularly in the development of vaccines. A greater understanding of the mutation of viral surface proteins could facilitate novel therapeutics for disease. Unfortunately, the specific mechanism of the evolution of viruses to escape selective pressures, such as the immune system or medications, is not well understood. As a result, it is nearly impossible to predict, for example, the glycoprotein configuration of the influenza virus even one month into the future. Vaccines to conditions caused by rapidly mutating viruses are therefore necessarily reactive, and a deeper understanding of viral capsule evolution could finally enable a proactive approach.

The study of viral capsids in order to determine evolutionary trajectories has two major variants: wild-type or lab. In wild-type studies, the virus capsid is analyzed directly from real-world patients infected with the virus, for example COVID-19. The tremendous number of patients can yield significant volumes of data, yet there remain fundamental gaps that render evolutionary predictions difficult. On the other hand, lab studies in model organisms provide a controlled, high-resolution environment, but these studies are tremendously expensive and take many years to come to fruition. Fortunately, *in silico* simulations can approximate the evolution of viruses under environmental pressures robustly, providing an effective complement to the aforementioned approaches. The central issue lies in creating a simulation model that simplifies the problem enough to perform Monte Carlo simulations at a large scale efficiency while maintaining biological characteristics of the original host-virus system.

Lattice evolution has emerged as promising approximation for the evolution of complicated structures, such as proteins. Strauss & Strauss (2001) provided a theoretical and empirical basis for the lattice structure formed by viral evolution of capsids. The regular organization formed by many viral capsids, such as those of alpha viruses, is a result of convergent evolution to a structure that is maximally stable.

In this project, we present a simplification of virus and antibody structure using a 2D lattice structure. We then analyze the evolutionary trajectories of our viral capsids using a Markov Chain Monte Carlo (MCMC) model. We compare the effect of nonlinear constraints on viral capsule proteins on the fixation of mutations in a population of viruses. Finally, we perform sensitivity analysis to demonstrate how our hyperparameters are reflective of the diversity of viral mutation and fixation rates in the real world.

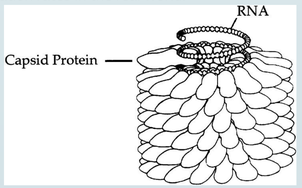
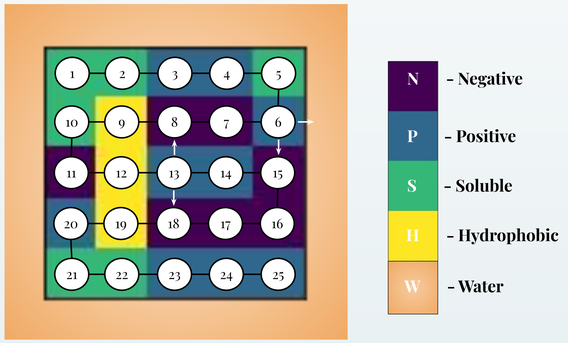


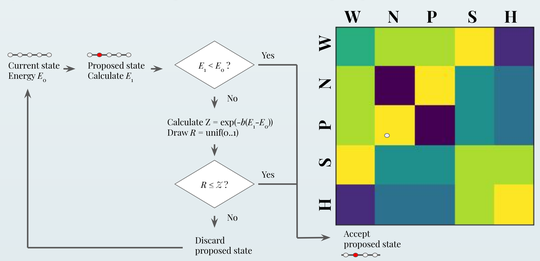
Figure 1: Regular viral capsid lattice for the Helical Tobacco Mosaic Virus Virion. Image credit Strauss & Strauss (2001).

Methodology

We used a lattice to model the viral capsid. Our setup is visualized in the figure below.



We began by generating possible viral ligands that minimize the folding energy of the protein. We did this using the Metropolis Hastings algorithm as follows



For phase two, we used Kimura's population genetics model to determine fixation of mutations. We simulated a population of N=1000 viruses. The fixation equations are then defined as:

$$s = -b * ((1-a) * dE_f + a * dE_b)$$

$$p = \frac{1 - e^{-s}}{1 - e^{-N \cdot s}}$$

For phase three, we sought to make our model more reflective of biological situations. We did this in the following ways

- Ignored selection for antibody binding if $E_b > 0$
- Antibody binding to all four sites of snake

Figure 2a: Visualization of optimal host antibody for the ligand, after adjustment under Kimura selection model.

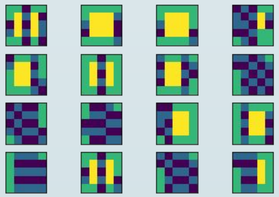


Figure 2b: Visualization of 16 optimal ligands generated by Metropolis-Hastings algorithm.

Results

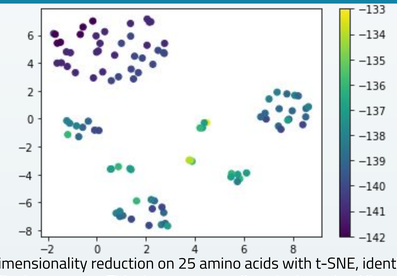


Figure 3: Dimensionality reduction on 25 amino acids with t-SNE, identifying clear clusters in final structures. This demonstrates diversity in optimal ligands.

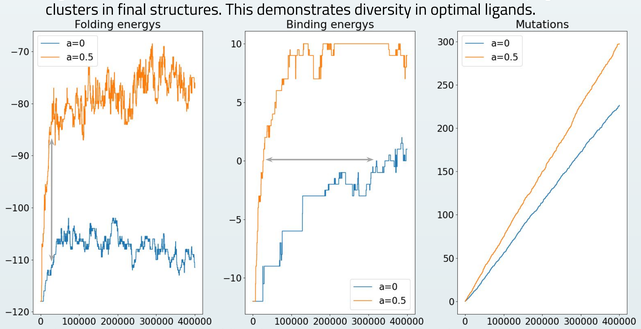


Figure 4: Comparison of folding, binding, and mutations for phase 2, median across 50 runs. The different metrics used for curve evaluation are also highlighted.

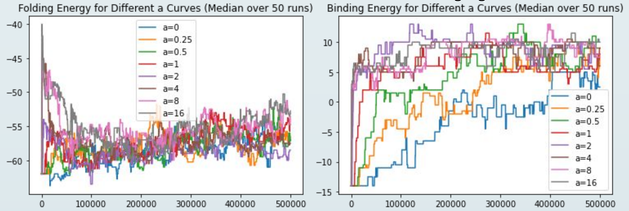


Figure 5: Comparison of different *a* values and their effect on binding escape and folding energy gain.

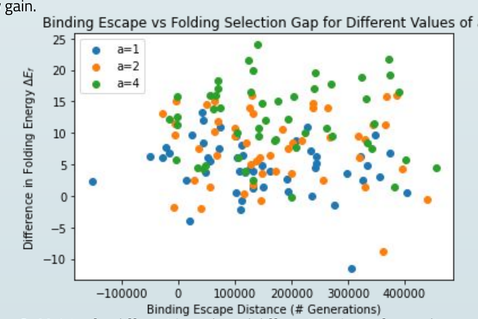


Figure 6: Metrics for different ligands and different t values of *a*, median across 50 runs. Higher *a* values tended to have a more significant binding escape but also folding loss.

Conclusion

- Successful implementation of viral escape with a 5x5 lattice
- Generated viable optimal protein capsules
- Modeled evolution-selection escape, balancing folding energy with antibody binding energy
- Different values of a affect metrics of escape
- Neutral escape impossible in 4 binding site case
- *In silico* model of viral evolutionary trajectories

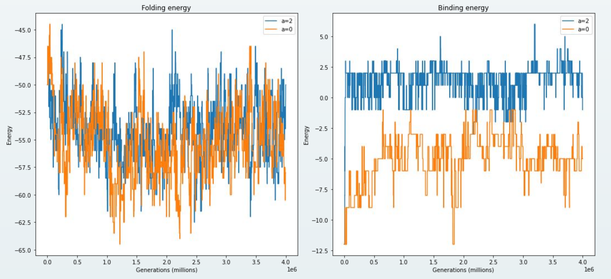


Figure 7: Folding and binding energy for four binding sites, median across 10 runs. Despite the long simulation time and number of runs, the signal remains very noisy, suggesting that a viral solution may take more mutation cycles to converge.

Discussion

- In the first phase, we successfully evolved proteins from a random starting sequence of amino acids towards a global minima for folding energy. The optimal capsids proposed are reflective of what is know about the structure of viral ligands.
- We noticed some important patterns in the final ligands created through our model. When we performed dimensionality reduction on different sequences, we found distinct clusters of optimal ligands. This suggests that there are broad categories of viral capsids that have folding energies at "almost global" minima with significant structural differences.
- We looked at the median mutation behavior across 50 runs to denoise the signal. We saw a clear binding escape accompanied by a loss in folding energy, and only fluctuations after the initial escape.
- We found that different values of *a* significantly affect the speed of antibody escape, but do not change long term folding or binding energy behavior. The neutral curve *a*=0, reaches 0 binding energy much slower than the instantaneous *a*=8 or *a*=16.
- Using our metrics of escape gap and folding energy gain, we compared median values for different starting ligands and found that the initial structure had a large influence on the speed of escape. Though higher *a* values tended to have faster escape, this benefit was not homogenous across starting ligands.
- We finally show that the 4-site binding behavior is very noisy and complex. The neutral curve never reaches E_b=0 whereas the selection curve reaches this point almost instantly. This is a potential topic of future investigation.

Selected References

Kimura, M. (1983). The neutral theory of molecular evolution. Cambridge University Press.

McCandlish, D. M., Epstein, C. L., & Plotkin, J. B. (2015). Formal properties of the probability of fixation: identities, inequalities and approximations. Theoretical population biology, 99, 98-113.

Strauss, J. H., & Strauss, E. G. (2001). Virus evolution: how does an enveloped virus make a regular structure?. Cell, 105(1), 5.