Analysis of Mathematical Methods for Modeling the HIV Disease.

Orry Butler

June 10, 2022

Abstract

The contents of this paper include analysis of the dynamics, evolution, and protein interactions of the human immunodeficiency virus (HIV). Our research is presented based on data from several other papers, as well as data obtained from databases that house biological data such as protein structures and genome sequences. This is being done in hopes of accurately reproducing results found in multiple papers based on the mathematical modeling of the HIV disease. Through this procedure, we hope to come to a better understanding of mathematical models for biological diseases such as HIV.

1 Introduction

The human immunodeficiency virus (HIV) is a relatively new and dangerous disease that is currently affecting a worldwide population. HIV was originally acknowledged as a disease in 1981, when gay men in the United States started suddenly dying from unknown causes [4]. Since the disease's discovery, numerous studies have been conducted in order to model the disease's spread and its subsequent mortality rates. HIV is a disease that attacks the immune system and, if left unchecked, can turn into Acquired Immunodeficiency Syndrome (AIDS) [9]. While HIV is not often deadly, it becomes deadly in most cases when the disease progresses to AIDS. In fact, only 18 percent of people with AIDS, who do not receive treatment for it, survived longer than 6 years with the disease[7]. While this survival rate paints a bleak picture of the progression from HIV to AIDS, there has been quite a bit of work put into developing effective treatments of the HIV disease. There are many different treatment methods being studied today, but we will be looking into the two which are being studied in the paper we are attempting to reproduce. These two treatment methods are early antiretroviral therapy (ART) and pre-exposure prophylaxis (PrEP), which primarily focus on the early detection and treatment of HIV [5].

In order to look into HIV, we will be studying the disease in three different ways – the first of which is through differential equations. In this portion, we will be analyzing an SIR model, which studies the spread of HIV in the gay population of South Korea [5]. We will be studying this SIR model through methods such as phase diagrams and linearization of the set of equations. These two methods will allow us to see the dynamics of the HIV disease with and without treatments and assess how it spreads through populations.

The next aspect that we will be looking into is the evolution of HIV through topological data analysis. In order to perform topological data analysis of the HIV disease, we will attempt to model HIV evolution based on data presented in [1]. We will analyze both the horizontal and vertical evolution of the HIV disease in order to paint a picture of where the disease is derived from.

Lastly, we will be taking a look at 3-dimensional modeling of the envelope protein GP-120. We will create a model of the protein base on [2]. Then we will use two different modeling softwares YASARA and Chimera in order to model the disease and performs minimization on the protein to see how the proteins bond.

When we perform all of these tasks on concepts associated with the HIV disease, we will hopefully be able to better understand where the disease comes from, the dynamics of its spread, and the ways

that proteins help in disease bonding. The main points of our research are the following: analyzing differential equations modeling HIV in certain populations, performing Topological analysis on genome sequences in order to see evolutionary relations, and looking into protein structures of the disease and how they explain the dynamics of HIV's spread.

2 Differential Equation Modeling of HIV spread

While there are many different treatment methods being studied today, we will be researching the two that are being studied in the paper we are attempting to reproduce. These two treatment methods early antiretroviral therapy (ART) and pre-exposure prophylaxis (PrEP), which primarily focus on the early detection and treatment of HIV [5]. The researchers of the paper study the gay population of South Korea, and we are attempting to reproduce their findings using mathematical models with differential equations and numerical methods to study the spread of HIV through homosexual populations in South Korea [5]. It also takes into account these two treatment methods in studying the effectiveness of the treatments methods versus each other and when there is no treatment present. We will be attempting to linearize the differential equations, present graphs of numerical methods, and present phase diagrams to model the spread and treatment methods of the HIV disease from data that is included

2.1 Results

2.1.1 Linearization

In order to understand the linearlization, we first must look at the model and the data which was presented in the original work. The model and the data we will be using is presented and explained below.

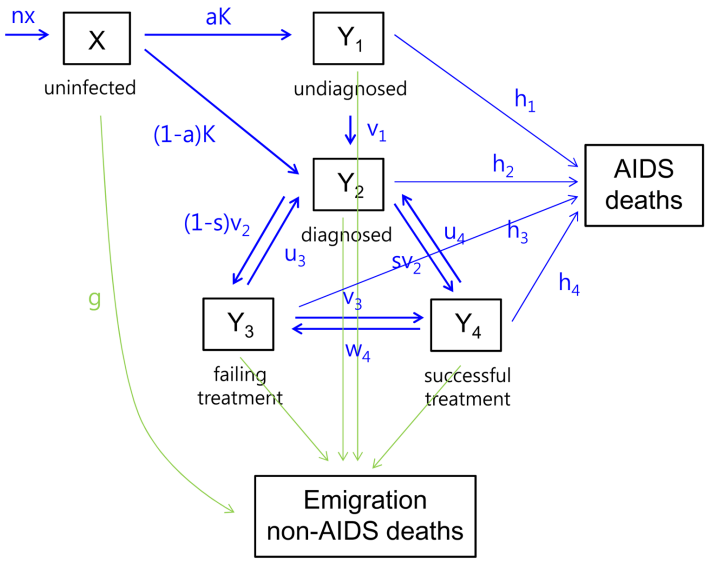


Figure 1: This figure is from Kim et al. [5] This figure depicts the compartment model presented in the paper.

The compartment model shows many different scenarios, such as the case where people who have HIV remain untreated, and cases where people who have HIV are treated or are failing treatment. After looking at this model, we can look at the differential equations modeling the spread of the disease and disease deaths, which is depicted in the paper. The equations we will be using can be seen below.

$$\frac{dX}{dt} = nx - (g + K)X$$
$$\frac{dY_1}{dt} = a\,KX - (g + h_1 + v_1)Y_1$$
$$\frac{dY_2}{dt} = (1 - a)KX + v_1Y_1 + u_3Y_3 + u_4Y_4 - (g + h_2 + v_2)Y_2$$
$$\frac{dY_3}{dt} = (1 - s)v_2Y_2 + w_4Y_4 - (g + h_3 + v_3 + u_3)Y_3$$
$$\frac{dY_4}{dt} = s\,v_2Y_2 + v_3Y_3 - (g + h_4 + u_4 + w_4)Y_4$$

$$K = \frac{\sum_{i=1}^{4} K_iY_i}{X + \sum_{i=1}^{4} Y_i}$$

$$K_1 = f_p\,f_u\,bc$$
$$K_2 = f_p\,f_u\,f_d\,bc$$
$$K_3 = f_p\,f_u\,f_d\,f_{tf}\,bc$$
$$K_4 = f_p\,f_u\,f_d\,f_{ts}\,bc$$

Figure 2: These equations are from Kim et al. [5] Differential equations depicting HIV spread and mortality amongst the male gay population in South Korea

In order to understand these equations, we first have to understand what the equations and variable represent. We can interpret these equations one at a time. The first equation $\frac{dX}{dt}$ is representative of the change in the uninfected population over time frame being accessed. The second equation $\frac{dY_1}{dt}$ represents the change in the indigenous population over time, which is also representative of how many people have contracted HIV but are unaware of it in a certain time period. Next, the third equation $\frac{dY_2}{dt}$ is representative of the number of people going from undiagnosed to diagnosed over a certain time period. The fourth equation $\frac{dY_3}{dt}$ is representative of the change in number of people failing treatment over time. These people failing treatment can either become people succeeding in treatment or go back to the diagnosed population over time. The last equation is the equation $\frac{dY_4}{dt}$, which is the change in successful treatments over time. The successful treatments in the model can go from successful treatments to either failing treatments or diagnosed. These equations are inserted into a matrix, called the Jacobian matrix–the first column contains the components of each equation with the variable X; the second column contains the elements with Y_1; the third column contains the elements Y_2; the fourth column contains elements of equations with Y_3; and the fifth column contains the elements of each equation corresponding to Y_4. When all equations are inputted, the produced Jacobian matrix will look like the following.

$$\begin{bmatrix} -g-\frac{\partial KX}{\partial X} & -\frac{\partial KX}{\partial Y1} & -\frac{\partial KX}{\partial Y2} & -\frac{\partial KX}{\partial Y3} & -\frac{\partial KX}{\partial Y4} \\ \alpha\frac{\partial KX}{\partial X} & -g - h_1 - v_1 + \alpha\frac{\partial KX}{\partial Y1} & \alpha\frac{\partial KX}{\partial Y2} & \alpha\frac{\partial KX}{\partial Y3} & \alpha\frac{\partial KX}{\partial Y4} \\ (1\text{-}\alpha)\frac{\partial KX}{\partial X} & v_1+(1\text{-}\alpha)\frac{\partial KX}{\partial Y1} & -g - h_2 - v_2+(1\text{-}\alpha)\frac{\partial KX}{\partial Y2} & u_3+(1\text{-}\alpha)\frac{\partial KX}{\partial Y4} & u_4+(1\text{-}\alpha)\frac{\partial KX}{\partial Y4} \\ 0 & 0 & (1-s)v_2 & -g - h_3 - v_3 - u_3 & w_4 \\ 0 & 0 & sv_2 & v_3 & -g - h_4 - u_4 - w_4 \end{bmatrix}$$

Table 1: The Jacobian Matrix for the Model done by Tianqi Li[11]

We then linearize the system by finding three real eigenvalues and two imaginary eigenvalues. We can find the eigenvectors corresponding to the eigenvalues and produce these two matrices. The matrix on the top represents the eigenvalues and the one on the bottom represents the corresponding eigenvectors that were solved for using the complex eigenvectors we found.

$$\begin{bmatrix} \text{-1.6636} & 0 & 0 & 0 & 0 \\ 0 & -0.0437 + 0.0265i & 0 & 0 & 0 \\ 0 & 0 & -0.0437 - 0.0265i & 0 & 0 \\ 0 & 0 & 0 & -0.1568 & 0 \\ 0 & 0 & 0 & 0 & 0.9996 \end{bmatrix}$$

$$\begin{bmatrix} \text{-0.0613} & 0.9442 & 0.9442 & -0.6005 & -0.0522 \\ 0.0366 & 0.0201 - 0.0515i & 0.0201 + 0.0515i & -0.2719 & 0.0360 \\ \text{-0.6617} & -0.0041 - 0.0624i & -0.0041 + 0.0624i & 0.0739 & -0.1528 \\ \text{-0.1036} & -0.0223 - 0.1085i & -0.0223 + 0.1085i & 0.2912 & -0.6269 \\ 0.7391 & -0.0521 - 0.2941i & -0.0521 + 0.2941i & 0.6893 & 0.7614 \end{bmatrix}$$

Table 2: The Eigenvalues and Eigenvector of the non-trivial point for Scenario 3 both tables done by Tianqi Li[11]

We can conclude that the system has both complex and non-complex eigenvalues–meaning that there will be a spiral at stable points, with the stable points corresponding to the real eigenvalues and the spiral corresponding to the complex eigenvalues. Thus, we can infer that the system of equations is in a spiral with stability points at the real eigenvalues.

2.1.2 Phase Diagrams

To produce the phase diagrams, we follow some procedures using our differential equations and data we have been using from [5]. It showed to be incredibly difficult to find a method to plot the phase diagrams considering that the system of equations is in 5 dimensions. Producing the phase diagrams for the scenario requires us to first combine the values for Y1 through Y4 in order to know the total number of infected. Then, we graph this computed value versus the uninfected population X to produce a phase diagram for the system of differential equations. The graph of the phase diagrams that we find will depict the behaviors of the system of differential equations.

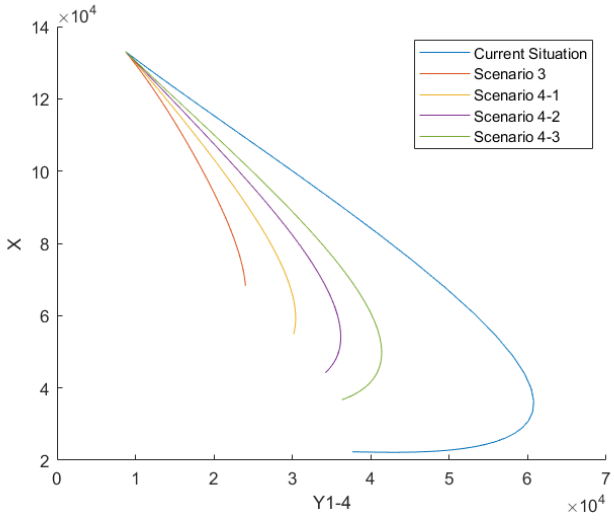


Figure 3: Phase Diagram across Scenarios

When we look at the graph of the phase diagrams, we can immediately see a few simple things. First, we can see the basic idea that we saw in our linearization–that there will be a spiral in the behavior of the differential equations. The graphs do not show a full spiral, mainly because of our difficulties in making a five dimensional phase diagram, but in many of the scenarios you can see the depicted lines starting to come back around to form a spiral due to the calculated complex eigenvalues. Similarly, in the cases involving Y, when the uninfected population is the highest you have a fairly low number

of infected populations or Y values. You can also see that in most of the cases where the graphs for the Y values goes upwards toward its maximum, those graphs start to decrease in Y values. This could mean many different things. For example, the Y values could be decreasing because the infected population is going from the infected population to the population of AIDS deaths. While this seems like a bleak picture, we can also see that where there are considerably few cases in the graph there are considerably more uninfected population. So, the best preventative measure for HIV/AIDS would be early detection, having safe sex, and other ways, such as the use of PREP, in order to prevent contracting the disease.

2.2 Conclusion

In conclusion, we used both phase diagrams and linearization to reproduce and arrive at conclusions based on the data presented in Kim et al[5]. In order to interpret the system of linear equations, we first looked into the linearization of the system. To perform linearization, we first found the Jacobian matrix of the system of equations and then found the eigenvalues and eigenvectors of the system. We then found that there were a pair of complex eigenvalues and 3 real eigenvalues. That system would be a stable spiral, meaning that the system of equations was stable based on the eigenvalues and eigenvectors that we found. We then looked into the phase diagrams that could be found from the same system of equations and variables. What we found from looking at the phase diagrams was that preventative treatments for HIV are the best method because in many of the cases where people contract HIV, even with treatment, those people would end up dying. When we look at all of the evaluations we performed on this set of differential equations, we can determine that the spread of HIV in the South Korean gay population is stable and that treatments have been fairly effective. While treatments have proven successful, the best way to prevent the spread of HIV and AIDS deaths is to take preventative methods, such as the drug PREP or performing safe sex in order to help prevent deaths from AIDS.

3 Evolutionary and Genomics modeling of HIV Virus

To look at the evolutionary modeling of the HIV disease, we will be looking at the paper [1]. We use both models and sequences from this paper in order to look into evolutionary models of the HIV disease and figure out what disease HIV evolves from. We will be attempting to reproduce the phlogenetic tree that is depicted in the paperby performing topological data analysis (TDA) in order to interpret both the horizontal and vertical evolution of the HIV Virus.

3.1 TDA Analysis

An important thing that we first need to look at, in order to perform TDA on the sequence, is to show the tree and all of the sequences highlighted in the paper [1] and show what we are attempting to reproduce. The phylogentetic tree depicted in the paper can be seen below and the phylogenetic, as well as all of the different diseases we will be looking intom to see the evolution.

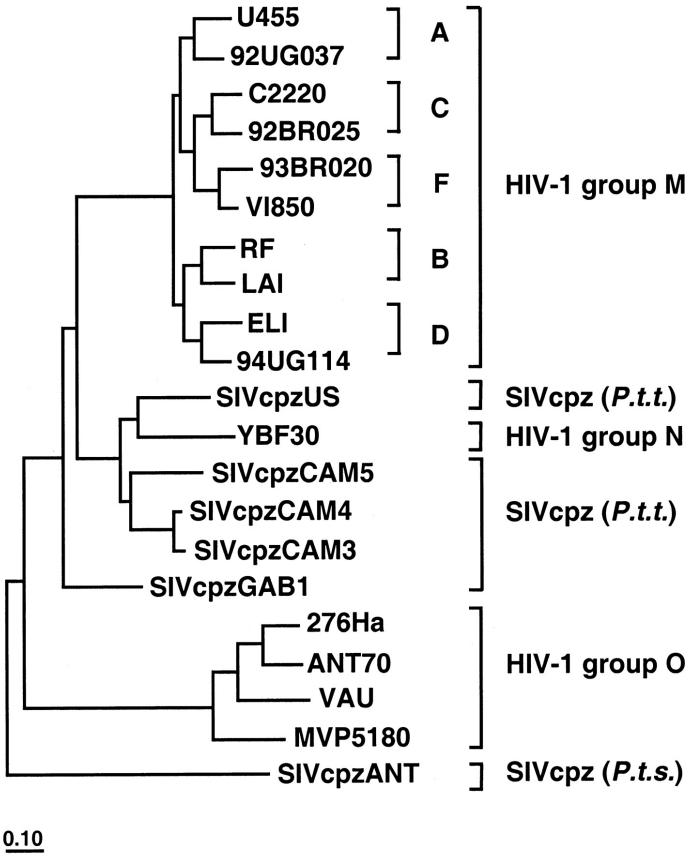


Figure 4: The phylogenetic tree produced by Hahn et al.[1]

When given this tree and these diseases that we intend to look at, we can find the sequences and the genetic distances of each sequence and put them into a matrix with all of the genetic distances. When computing the distance matrix, we used the MaximumLikelyhood method to find our genetic distances, as that method was used in the paper we are trying to replicate [1]. Now that we have a distance matrix, we can start to look at our topological data analysis. First, we can look at the method that we followed in order to perform TDA, which uses both R and MATLAB programming languages. We first use R and the calculate homology function on the genetic distance matrix to calculate the vietoris rips diameter of the distance matrix. When we perform this procedure, we come away with both 1 dimensional and 0 dimensional betti numbers and simplicies. We can then plot these values on a graph using the plot barcode R function. The graph computed from performing these procedures is shown below.

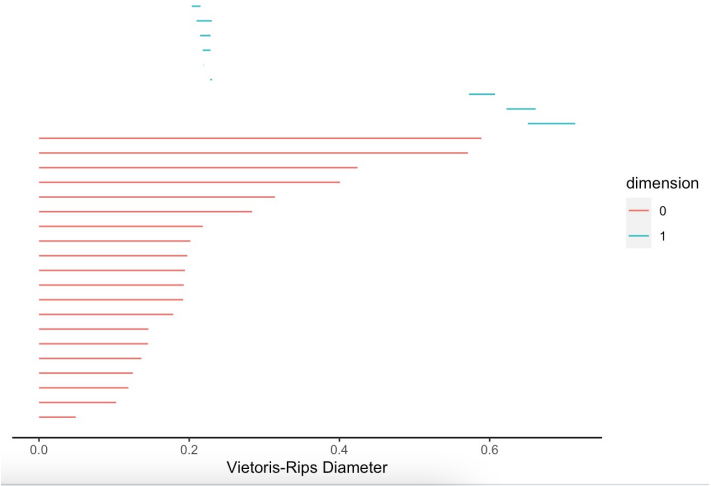
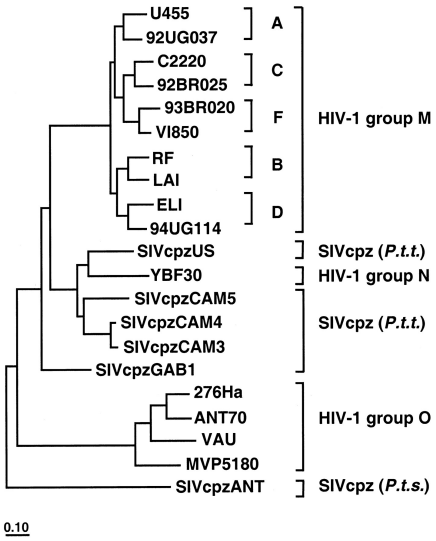


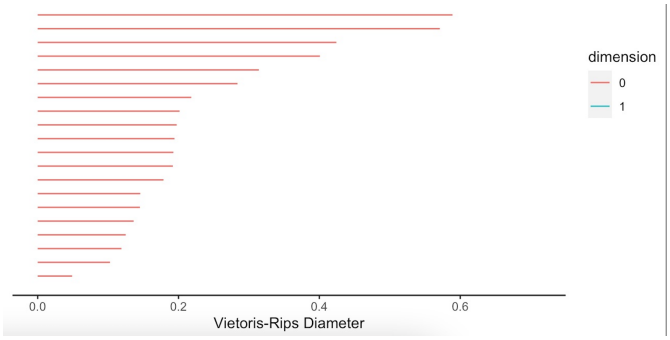
Figure 5: Barcode diagram from the computed distance matrix.

When we look at this graph, we can see that we have 20 0-dimensional lines and we have 5 1-dimensional lines. We interpret this plot is in that the 0-dimensional lines are indicative of the phylogenetic tree listed above and the 1-dimensional lines are indicative of recombination events in the genome.

We first take time to analyze the 0-dimensional lines and then the one dimensional lines. We can look at the 0-dimensional lines in a similar way that we look at phylogenetic tree as an interpretation of the vertical evolution of a disease. Below, you can see the bar code plot of 0-dimensions side by side with the phylogenetic tree[1].



(a) Phylogenetic tree from Hahn et al.[1]



(b) Plot of 0-dimensional lines

Figure 6: A figure with two subfigures

When we look at both of these figures side by side, we can start to interpret what the 0-dimensional barcode plot represents. We interpret the 0-dimensional barcode plot with the longest lines representing the segments of the phylogenetic tree such as SIVcpzANT, which has the least number of branches from the root to get to the branch. We can interpret the shortest segments as the portion of the phylogenetic tree with the most branches going from the root to the leaves. So we can interpret 21 0-dimensional lines to correspond to each of the leaves of the phylogenetic tree, showing us how the vertical evolution works. We can also see, through our zero dimensional topological data analysis and

the phylogenetic tree, that HIV is evolving from the disease SIV.

We will now look at the 1-dimensional components of the graph and interpret those. We can interpret the one dimensional lines depicted in the graph as the number of recombination events that occur in the evolution of the HIV virus. As you can see from the barcode plot shown below, there are 5 1-dimensional lines–meaning that there will be 5 recombination events in the evolution of the HIV virus.

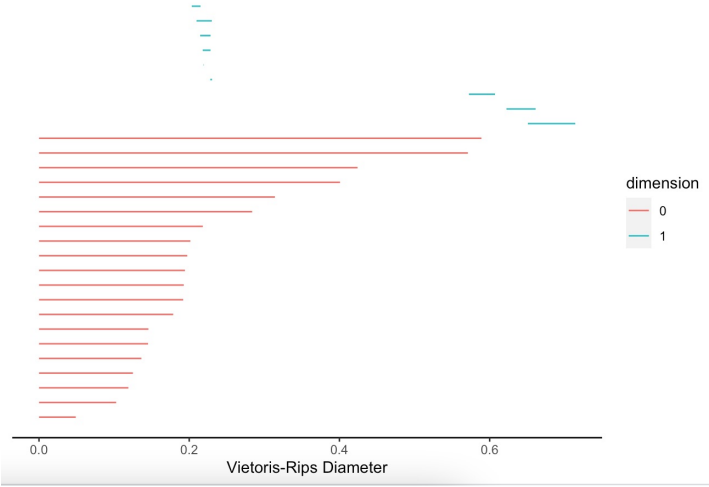


Figure 7: Barcode diagram from the computed distance matrix

There are other ways in which we can look into visualizing these recombination events, or the horizontal evolution. The first of which is through a Rips Diagram. We can make a rips diagram by using the calculate homology function in R on our distance matrix. Then, we plot the data using the function RipsDiag, which will give us a graph that looks like the one shown below.

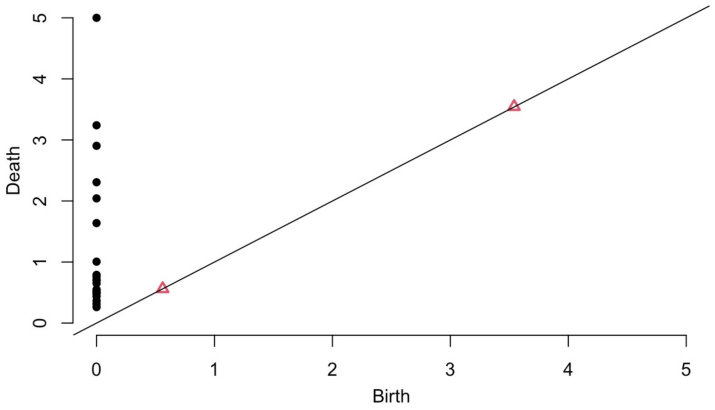


Figure 8: Rips diagram from the computed distance matrix

In this graph, the black dots represent single points of evolution. The red triangles are indicative of the recombination events. However, there is an issue in the data we are getting out of the rips diagram because HIV is known to have quite a few times with just HIV-1 having around 8-10 recombination events in its synthesis[8]. While our initial graph does not match this number, there is still a range on the number of recombination events and we could be lower or higher. But with our rips diagram, if there are a significant number of recombination events, our red triangles should be further off of the black line–but that is not the case. So, therefore, we need to look into better ways of reproducing our rips diagram, based on the data that we have, in order to match the graph that we initially obtained and current research on the topic.

The last way that we can look into the TDA of the genome is through a more abstract way of graphing the simplicies. In order to do this, we used a MATLAB package written by Adam Cutbill which performs TDA on a presented distance matrix[10]. This package performs TDA by calculating the betti numbers and then graphing triangular 1-dimensional simplicies based on those betti numbers. We can now look into the graphs of these one dimensional simplicies, looking for holes as the graphs are being built because the holes are indicative of recombination events.

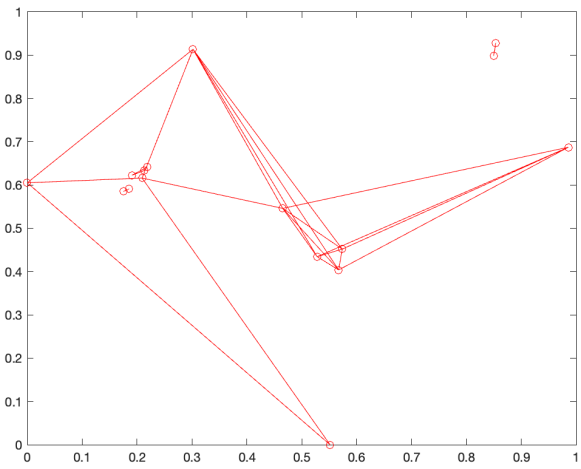


Figure 9: Graph of 1-dimensional simplicies using connectivity parameter of 1 [10]

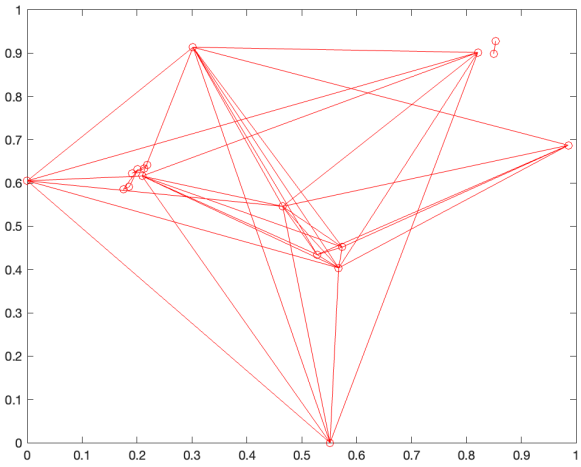


Figure 10: Graph of 1-dimensional simplicies using connectivity parameter of 1.2 [10]

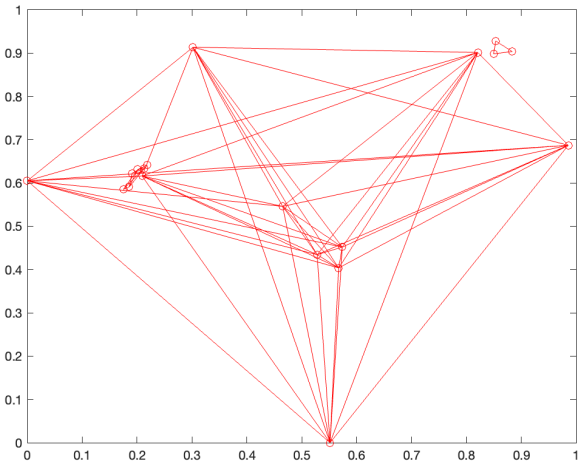


Figure 11: Graph of 1-dimensional simplicies using connectivity parameter of 1.5 [10]

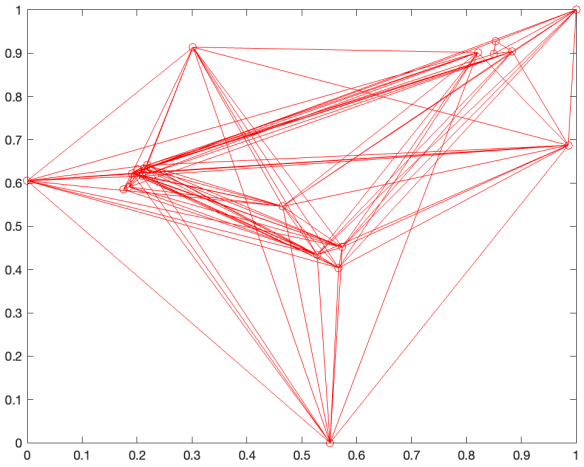


Figure 12: Graph of 1-dimensional simplicies using connectivity parameter of 1.8 [10]

We can interpret these graphs as having holes that are indicative of recombination events. We build the graphs using increasingly higher connectivity parameters for each graph because it paints a more clear picture of what we are looking at. The graph that I would first like to focus on, though, is the last one because this is the graph that most clearly depicts a hole. The hole can be clearly seen in the middle of the diamond-shaped objects and it shows one of the recombination events in the horizontal evolution of the HIV disease. There are also more holes depicted in the graphs presented but, due to the limitations of MATLAB and three dimensional graphing, it is fairly difficult to see the recombination events in these graphs. I am currently looking into finding better ways to model the graphs of simplicies in order to paint a clearer picture of the recombination events in this way. Another issue that I encountered was in the limitations of computing–due to the fact that our data set is incredibly large, we were unable to continue to bump up the connectivity parameters anymore because any more increases would cause my computer to crash due to the amount of computations being done. Thus, we are still looking into better ways of performing this part of TDA on the data for genetic distances.

3.2 Conclusion

In conclusion, we used TDA to analyze both the vertical and horizontal evolution of the HIV disease, using sequences and trees presented in [1]. We came to two different conclusions based on both our 0-dimensional and 1-dimensional analysis. Based on our 0-dimensional analysis and our reproduction of the phylogenetic tree based on zero dimensional analysis, we can conclude that the disease HIV is a descendent of the disease SIV. We can then discuss our 1-dimensional analysis, which is fairly inconclusive and still needs significant work, because with the barcode graph and rips diagram, the data does not match up based on what is depicted in the graphs. Then, based on the simplical graph, we need to find a better way to produce the graph because with the current method it is extremely difficult to see where the holes or recombination events lie on the graphs. Therefore, 1-dimensional analysis and recombination events are concepts that we still need to put work into in order to reach a more definitive conclusion on the horizontal evolution of HIV.

4 Protein Modeling and Structural Minimization

For this portion of the project, the protein that we will be analyzing is the glycoprotein 120 for the HIV envelope gene. The GP-120 envelope protein, when present in a person, is the cause of infectivity because, like other envelope diseases, the envelope protein is present in the bonding of the disease[6]. Knowing what the protein is, we have to look at the model of the protein and work to minimize the structure of the protein. The PBD file that we will be using is the 2NY3 PBD model of the GP-120 glycoprotein for HIV. We found the model used in the RCSB PBD database[2]. We then will use the software UCSF Chimera[12] to perform the operations, such as 3d modeling and energy minimization of the structure. We will be looking at the GP-120 Glycoprotein to analyze its 3-dimensional structure and also to minimize the structure of the protein and its total energy.

4.1 3D-modeling of the molecule

The process to do three dimensional modeling of the GP-120 is fairly straightforward thanks to the Chimera[12] software. In order to model the protein, we found a PCB file, which will work for the protein. The PCB file we will be using has PCB 2NY3 and can be found using the protein data bank, which houses a collection of PBD files for different proteins[2]. We can then simply load the file using Chimera in order to visualize the protein structure. When we load the file in Chimera, we get a protein structure for the GP-120 glycoprotein. We then compare the similarities between the structures produced by the PBD database and the structure that we produced.

Figure 13: Model of GP-120 envelope protein found on Protein Data Bank[2]

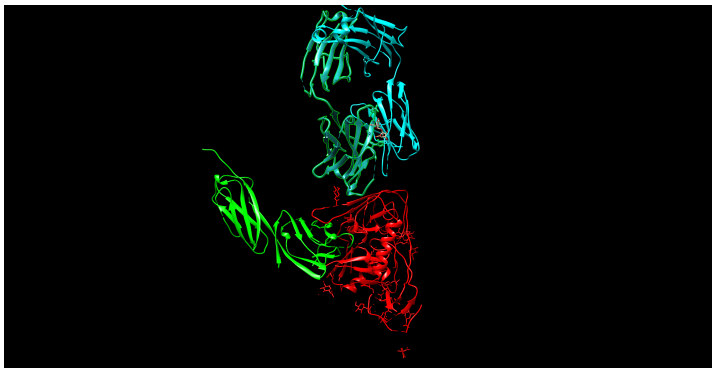


Figure 14: Structure of the GP-120 protein based on inputting the PBD file into Chimera[12]

As we can see from these two figures, the model that we constructed looks incredibly similar to the model that we were aiming to reproduce. Any difference in the models can be explained by having to delete the ANS amino acid in order to allow the PBD file to work with Chimera. One thing that you easily notice from our 3 dimensional model is that the protein is composed of five bonded chains. While 4 of the bonded chains maintain a spiral structure with each other, the other chain that we can see does not maintain this same spiral formation with the other proteins. This structures seems indicative of it being an envelope protein and being involved with the HIV infection which is involved in bonding of RNA to infected people.

4.2 Minimization

We can now perform minimization on the structure, using a different software called YASARA, which is also another resource we can use to visualize the three dimensional structures of proteins. I decided

to use YASARA instead of Chimera on the minimization portion because it used a different server on a different machine in order to perform minimization. For Chimera, I used my own system. For any stepsize I would use in my software, Chimera would crash my computer. In order to use YASARA to perform minimization, we have to obtain the PBD ID 2NY3 for the genetic model that we will be using from the PBD data base[2]. We then use the YASARA minimization server and input the PBD ID we are using, and the server will perform the process of minimization for us[3]. After the YASARA server performs minimization for us, we get a model of our protein that looks like the following.

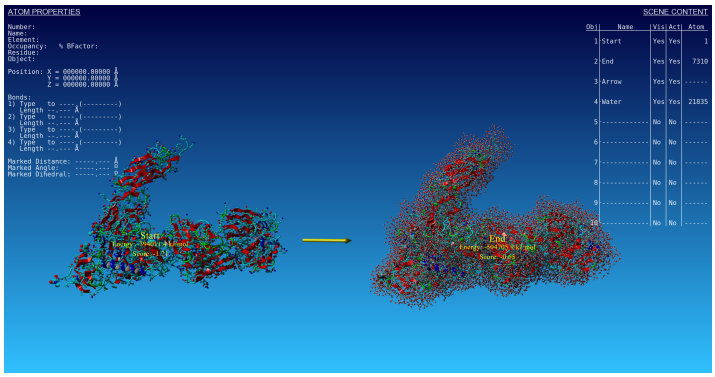


Figure 15: Model before and after minimization of GP-120 envelope protein with the intial on the left and the minimized structure on the right[2][3].

Performing minimization on the protein results a much larger structure, composed of both the protein structure and the Hydrogen ions which give the protein its charge. When minimization is performed, we have a negative charge for the total energy of the system. This means that the protein has strictly potential energy–the protein wants to bond with things in order to keep the disease spreading. This makes sense because HIV is an RNA disease and an envelope protein, meaning that the disease needs to bond in order to expand. When we look at other proteins, that pattern may be more associated with the disease already bonded, rather than the envelope protein. We may see that the charges are positive with kinetic energy from the bonds that the disease has produced. Yet, this protein has a negative charge, meaning that its potential energy wants to bond in order to advance and infect people with the disease.

Acknowledgement

I would like to thank both Tianqi Li and Yirong Liang for helping me with the first two portions of this project and also allowing me to use some of the figures we presented in our initial reports. I would not have been able to complete this quarter of projects without their help.

I would also like to thank Javier Arsuaga and Appilineni Kushal for their help with the protein dynamics portion of this project. Their support truly helped me complete that portion of the project and I would still be lost on without their assistance.

References

[1] B. H. Hahn, G. M. Shaw, K. M. De, Cock, and P. M. Sharp, "Aids as a Zoonosis: Scientific and Public Health Implications)," *Science*, vol. Vol 287, no. 5453, pp. 607–614, 2000. DOI: 10.1126/science.287.5453.607.

[2] T. Zhou, L. Xu, B. Dey, *et al.*, "Hiv-1 gp120 envelope glycoprotein (k231c, t257s, e267c, s334a, s375w) complexed with cd4 and antibody 17b," *Protein Data Bank*, Feb. 2006. DOI: 10.1016/s0092-8674(00)80205-6.

[3] E. Krieger, K. Joo, J. Lee, *et al.*, "Improving physical realism, stereochemistry and side-chain accuracy in homology modeling: Four approaches that performed well in casp8," *Proteins*, pp. 114–122, 2009. DOI: 10.1002/prot.225706.

[4] P. M. Sharp and B. H. Hahn, "Origins of HIV and the AIDS pandemic," *Cold Spring Harbor Perspectives in Medicine*, vol. 1, Sep. 2011. DOI: 10.1101/cshperspect.a006841.

[5] S. B. Kim, M. Yoon, N. S. Ku, *et al.*, "Mathematical Modeling of HIV Prevention Measures Including Pre-Exposure Prophylaxis on HIV Incidence in South Korea," *PLoS ONE*, vol. 9, J. W. Tang, Ed., e90080, Mar. 2014. DOI: 10.1371/journal.pone.0090080. (visited on 12/10/2020).

[6] M. C. DeSantis, J. H. Kim, H. Song, P. J. Klasse, and W. Cheng, "Quantitative correlation between infectivity and gp120 density on hiv-1 virions revealed by optical trapping virometry," Mar. 2016. DOI: 10.1074/jbc.M116.729210. (visited on 06/09/2022).

[7] J. Poorolajal, E. Hooshmand, M. H., N. Esmailnasab, and E. Jenabi, "Survival rate of Aids disease and mortality in Hiv-infected patients: A meta-analysis)," *Public Health*, vol. Vol. 139, pp. 3–12, 2016. DOI: 10.1016/j.puhe.2016.05.004.

[8] J. M. O. Rawson, O. A. Nikolaitchik, B. F. Keele, V. K. Pathak, and W.-S. Hu, "Recombination is required for efficient hiv-1 replication and the maintenance of viral genome integrity," *Nucleic Acids Res*, vol. 46, pp. 10 535–10 545, Nov. 2018. DOI: 10.1093/nar/gky910.

[9] CDC, *About HIV/AIDS — HIV Basics — HIV/AIDS — CDC*, www.cdc.gov, Nov. 2020. [Online]. Available: https://www.cdc.gov/hiv/basics/whatishiv.html#:~:text=HIV%20(human%20immunodeficiency%20virus)%20is.

[10] A. Cutbill, *Topological data analysis learning code*, MATLAB Central File Exvhange, 2022. [Online]. Available: https://www.mathworks.com/matlabcentral/fileexchange/47009-topological-data-analysis-learning-code.

[11] T. Li, O. Butler, and Y. Liang, "Mathematical approach to study the population spread and treatment to hiv," *Submitted to MAT 124 at UC Davis*, 2022.

[12] V. Resource for Biocomputing and S. F. Informatics at the University of California, University of California, San Fransisco. [Online]. Available: https://www.cgl.ucsf.edu/chimera/.