Zhenyu Wu
Prof. Brain Chen
CSE 308
28th April 2023

Project 3 Report Paper

1) A. For the reason that this patient doesn't take any treatment in the first place and is not infected by other strains of HIV virus, it will probably be possible to observe the same amount of genetically identical HIV strains in this patient if we sequence the RNA genome in every HIV capsid in the patient. For the reason that HIV virus is not considered a multicellular organism, the evolution or divergent evolution that will affect an organism's genetic composition will not be observed in viruses. However, there are still two factors that will change the genetic composition of viruses.

The first factor is the infection of other strains of HIV virus, which is called superinfection or reinfection. When a person who has already been infected by one strain of virus gets infected by another strain, two strains of virus will interact, reform the gene composition and eventually create a new strain of virus. In some medical research, the new strain of virus can even possess the drug resistance to both treatment drugs previously applied on "parent" strains, which means the drug resistance can be transmitted through superinfection. However, the superinfection doesn't always occur, for the reason that some subtypes of the virus can influence the likelihood of superinfection occurring and some don't. In this case where the patient is only infected by one strain of HIV virus, superinfection will not occur, which means it will have a relatively high possibility that HIV capsids within his/her body will remain the same genetic composition.

The conclusion that we would observe the same amount of genetically identical HIV strains in this patient is based on the assumption that ignores the mutation of the virus, which is the self genetic variation process that occurs randomly during replication. In reality, HIV virus is well-known for its high replication rate which leads to greater chance of encountering error during replication and causing mutation. Drug resistance is one of the most common mutation examples. If patients don't take medical treatment adherently, research shows there will be greater chance that virus will develop drug resistance. In this case, for the reason that this patient doesn't take any treatment at all, I assume the mutation of drug resistance will not happen, but random mutation does exist.Thus, if we consider the case of mutation, we might observe fewer genetically identical HIV strains in this patient for the reason that mutation can lead to a more diverse set of strains.

B**.** Zidovudine also known as AZT is a nucleoside analogue reverse transcriptase inhibitor (NRTI) drug applied for treatment of HIV virus infection. Zidovudine works by interfering with the reverse transcriptase enzyme that HIV uses to convert its RNA genome into DNA, thus preventing the virus from replicating. Normally Zidovudine is used along with other medications to treat HIV infection and it has been proved that it can reduce the chance of passing the infection from a HIV-positive pregnant woman to her baby. However, Zidovudine comes with several side effects as well such as trouble breathing or swallowing.

The key point about Zidovudine is to control the infection through reducing HIV replication. Therefore, compared to the situation in a), the number of genetically identical HIV strains in this patient may decrease after taking Zidovudine treatment. However, for the reason that viruses can mutate randomly during the replication process, it is hard to predict the exact number of genetically identical HIV strains in a patient's body. In addition, just like I mentioned in a), if the patient doesn't take Zidovudine adherently, HIV virus may even develop drug resistance in the process of mutation.

C. Since in the question a), I mentioned two outcomes: one ignores the mutation and one doesn't. Therefore, for the case that ignores the mutation before taking Zidovudine treatment which leads to a limited set of virus strains, the phylogenetic tree of HIV genomes within the patient would likely show a limited number of genetically identical strains and likely be a relatively uniform phylogenetic tree structure or even single strain. On the contrary, before treatment, for the case that mutation leads to a more diverse set of virus strains, the phylogenetic tree will show a branching pattern where each strain would have a separate branch, and the branches would diverge from each other over time as new mutations accumulate. The tree would have multiple levels of branching and be more complex compared to a tree with fewer strains or a single strain.

However, for both cases, the phylogenetic tree after receiving Zidovudine treatment, as the drug would have selectively targeted and eliminated the strains that were most susceptible to it, a more uniform phylogenetic tree structure will be created. The surviving virii would have a more limited genetic diversity, and they may all either be undetectable by the Zidovudine or possess drug resistance through mutation.

D. Highly Active Antiretroviral Therapy, also known as HARRT, uses a combination of three or more drugs to treat HIV infection and stops the virus from making copies of itself in the body. In general, the reason for applying multiple drugs is to prevent or reduce the likelihood that a virus develops resistance to any of the drugs. When multiple drugs are used, the virus must simultaneously develop multiple mutations in its genome in order to survive, which is a much harder task than developing a single mutation to resist a single drug.

Compared to applying drugs in parallel, applying drugs sequentially will generate a phylogenetic tree of RNA genomes that would likely show a much greater diversity of strains over time, for the reason that a virus will develop resistance to one drug when it is exposed to that drug. Then, when a virus is exposed to another drug sequentially, it will develop another resistance to that specific durg through a new mutation process. Thus, viruses develop drug resistance based on the sequence of application of drugs and would likely result in the emergence of many different HIV strains with diverse mutations.

E. Generally, the term "noncompliance" typically refers to a patient's failure to take a prescribed medication or follow a recommended course of treatment, and a patient who exhibits noncompliance is often described as "noncompliant." Even HARRT is effective, there are still several reasons that patient fail to follow the course of treatment. The first reason is because of the adverse effects caused by HARRT. According to *Identification, Management, and Prevention*

of Adverse Effects Associated With Highly Active Antiretroviral Therapy, the adverse effects of HARRT includes anorexia, nausea, vomiting, diarrhea, pancreatitis, hepatotoxicity, lipodystrophy syndrome, cardiovascular risk, etc… Those adverse effects would cause both mental and physical damage to the patient, which is one major reason for noncompliant behavior. Another reason is the strong drug resistance developed by viruses. There is a small possibility that the virus develops a strong drug resistance during treatment which makes treatment less effective. In addition, in the article *Understanding Noncompliant Behavior: Definitions and Causes,* there are many other reasons leading to Noncompliant Behavior, including high cost of treatment and poor communication between healthcare personnel and patients.

2. For the reason that "proteins with very different sequences (e.g. 25% sequence identity) have been known to exhibit very similar "folds", or overall shape, while proteins with very different folds never exhibit very similar amino acid sequences," there is not always a one-to-one correspondence between the sequence and shape of the protein. For instance, two proteins with similar sequences may have different shapes if they diverged over time and adopt different folds.Therefore, the traditional alignment that only takes account of sequences may not be able to precisely identify the relationship between proteins and their evolutionary history. Common shape of proteins may indicate they may have a common ancestor since protein shape can be conserved during evolution. In other words, if we can incorporate protein shape into sequence alignment, we can better identify homologous proteins that may have diverged significantly in their sequences over time.

The shape and structure of protein play a key role in protein function. For instance, the three-standard, rope-like shape of the collagen in our cartilage and tendons is able to make it stronger. if a protein loses its shape at any structural level, it may no longer be functional. Throughout the evolution process, some amino acid replacement or divergence in sequence may not be tolerated in some folds, but some others may be favored to better maintain or enhance the functional stability. Therefore, combining shape in sequence alignment can provide additional information about mechanisms that drive protein evolution.

3. A. For the reason that subtrees below a point of intersection conserve properties within the subtree that differ from those of other subtrees, it may indicate that conserved properties play significant and specific roles in these subtrees, but not in other subtress. Moreover, these conserved properties can reveal that members in a subtree with a common ancestor share a common function that is unique to this subtree. For example, there are two subtrees in the graph, A and B, and both of them share a common ancestor in the upper level of the tree. A represents a family of brain-specific enzymes and B represents a family of liver-specific enzymes. Then, through further analysis, we find an amino acid called C which is conserved and presented in all members under subtree B, but not in subtree A. In this case, since A and B perform entirely different functions and activities in the body where A performs brain-specific enzymatic functions while B performs liver-specific enzymatic functions, it can show that C is the unique and critical amino acid specific to the liver-specific enzymatic functions. Thus, unique conserved amino acids in a subtree can help us to understand the common functions performed by members in a subtree and better understand the functional specialization.

On the other hand, if a conserved amino acid is found in a larger subtree, like the subtree just below the root. This reveals that this amino acid is important for a more general function that is shared by that subset of sequences. This can provide an insight to the functional conservation in the evolution process.

B. Active sites are regions usually on the surface of enzymes specially modeled by nature during evolution that either catalyze a reaction or are responsible for substrate binding. They are usually highly conserved regions of the protein that are essential for catalyzing specific chemical reactions when they are binded with substrates.Therefore any mutation or evolution that will change the structure of the active site can profoundly impact the function of that active site or even cause it to lose its function. For instance, in the active site chymotrypsin, an enzyme that is responsible for protein digestion in the small intestine, we can find three critical amino acids including serine-195, histidine-57 and aspartate-102. They closely work together to carry out the catalytic function of breaking peptide bonds. More specifically, aspartate is responsible for positioning histidine residue in the proper orientation, which is the first step in their catalytic reaction. If the aspartate-102 is no longer conserved due to mutation or evolution, histidine residue can not be positioned in correct orientation and stop the whole reaction chain. Eventually, chymotrypsin can no longer perform its function as catalyst in the chemical reaction of breaking peptide bonds of the protein digestion process. Therefore, the clustering of conserved amino acids near active sites is likely the result of the selective pressure to maintain the structure and function of these critical regions of the protein. The conservation of these amino acids helps to ensure that the protein can perform its biological function efficiently and accurately.

Reference

Whatever happened to HIV superinfection? - thebody. (n.d.). Retrieved April 29, 2023, from https://www.thebody.com/article/hiv-superinfection-subtype-l

Mark Cichocki, R. N. (2022, November 11). *Is it possible to get reinfected if you have HIV?* Verywell Health. Retrieved April 28, 2023, from https://www.verywellhealth.com/hiv-reinfection-and-positive-prevention-48108

Zidovudine - patient: NIH. Zidovudine - Patient | NIH. (n.d.). Retrieved April 28, 2023, from https://clinicalinfo.hiv.gov/en/drugs/zidovudine/patient

Highly Active Antiretroviral Therapy. National Cancer Institute. (n.d.). Retrieved April 28, 2023, from https://www.cancer.gov/publications/dictionaries/cancer-terms/def/highly-active-antiretroviral-therapy

Identification, Management, and Prevention of Adverse Effects Associated With Highly Active Antiretroviral Therapy. Medscape. (2005, January 31). Retrieved April 29, 2023, from https://www.medscape.com/viewarticle/497920_2#:~:text=Anorexia%2C%20nausea%2C%20and%20vomiting%20are,patients%20may%20require%20some%20intervention.

Kleinsinger, F. (2003). Understanding noncompliant behavior: Definitions and causes. The Permanente Journal. Retrieved April 29, 2023, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5571787/

U.S. Department of Health and Human Services. (n.d.). *For proteins, form shapes function.* National Institute of General Medical Sciences. Retrieved April 29, 2023, from https://www.nigms.nih.gov/education/Inside-Life-Science/Pages/for-proteins-form-shapes-function.aspx#:~:text=Every%20protein%20has%20a%20shape,the%20molecule%20they%20act%20upon

Reaction Mechanism of Chymotrypsin. Aklectures.com. (n.d.). Retrieved April 29, 2023, from https://aklectures.com/lecture/enzyme-mechanisms/reaction-mechanism-of-chymotrypsin#:~:text=Chymotrypsin%20contains%20a%20collection%20of,function%20of%20breaking%20peptide%20bonds