Devan Bicher

Project 3: Implementation Report

5/3/13

Section 1

We ­­would find about the same number of genetically identical HIV strains in this patient compared to when they were first infected. This is because there is no environmental pressure on the HIV to change in anyway, so there is no reason for the HIV to change at all and all the capsids would be genetically almost identical. There might be a very few different strains of the virus simply because HIV has a propensity for high mutation rates and so there might be several different mutant strains that have appeared around the time of the sequencing. However these strains will not last long because the dominant strain that resides in the patient is going to be the one that is the most efficient, since biology tends to push organism populations in the direction of the most efficient member of the population. These mutants that have arisen by biological chance are not the most efficient because after such a long infection period the most prevalent strain is going to be the most efficient one. Therefore in the long run, which is the case here several years after infection, nearly all of the strains of HIV present in the patient will be identical.

Zidovudine is a HIV inhibitor, actually it was the first one clinically available for the treatment of AIDS. Zidovudine works by inhibiting the key enzyme reverse transcriptase, however due to the high mutation rates in HIV Zidovudine is only partially effective and merely slows the progress of the virus but does not stop it completely. This is because in the presence of the inhibitor the viral population will move towards viruses that have the ability to resist zidovudine while still functioning biologically close to normal. With this in mind the number of strains that are different in this infected person will be greater than the person above. There will be some small amounts of the original strains still lingering that have not yet been inhibited by the treatment. More importantly though there will be various strains present that have evaded the treatment by mutating and effectively resisting it. Since there might be many mutations that are capable of resisting the treatment there will be several of these different strains present in the patient. If the treatment is continued for a long period of time and then the viral population is sequenced there will be almost completely one strain of virus present: the one that is both the most efficient at evading the treatment while also being the one that is most efficient at performing its natural biological function.

The phylogenetic tree of the progression of HIV in this patient will start with the genome of the original infection then the genomes will start to branch off after the patient receives the treatment. There will be several different strains at first that are capable of resisting the treatment and as we move down the tree the branching will continue until the viruses that are the most efficient at resisting treatment will prevail. All of the leaves of this tree will have the common attribute of being resistant to the treatment. The original virus will die off as it is inhibited and the branches will all have some level of resistance to the treatment because all of the ones that are not resistant will die off relatively quickly. The farthest leaf in the tree away from the original virus will, again, be the one most efficient at resisting the treatment.

HAART applies several drugs at once because it is relatively easy for the virus to evade treatment of one drug at a time but when several drugs are applied at the same time the viral population needs to be able to resist all of these drugs at the same time. This is very difficult because all of the drugs inhibit the target enzyme in some unique way and so it is very difficult to evade all of these inhibitors while still remaining biologically active. Continuing biological activity is key here, and while it is relatively easy to remain active in the presence of one inhibitor it is much more difficult to create the complex mutations needed to evade treatment while still retaining biological function. With this in mind if the treatments are all applied at the same time the tree will branch out rapidly from the original genome as numerous mutants are generated with varying levels of resistance to the inhibitors and the leaves of the tree will all be resistant in some way, even if small, to at least one treatment. However as the tree branches down further the leaves will narrow down as the other mutants are inhibited successfully until eventually the only strains that remain are the ones capable of resisting all treatments. Finally the most prevailing leaves in the end will be the ones most efficient at resisting all treatments at once. While trees are independent of time this process will take much longer to happen than applying one treatment which is why patients treated with cocktails have longer life expectancies.

This tree will differ than if the treatments were applied sequentially because the branching would be the same as the tree described above for one treatment at first and then would branch out after the second treatment was applied until the remaining leaves are resistant to both treatments. This type of branching would continue for the duration of treatments that are sequentially applied. Noncompliance is a problem because of several factors. One of which is the health complications that are associated with these inhibitors. Naturally, the more inhibitors that the patients take the more health complications that will arise which is a factor contributing to noncompliance. Additionally these inhibitors are costly due to the complexity of HIV and so of course the more treatments that are administered at the same time the more costly the therapy becomes. This is especially an issue in poverty stricken countries which at this point comprise a huge portion of population currently infected with HIV.

Section 2

Applying the shape of the protein to an alignment of sequences might help to elucidate some of the mechanisms of conservation amongst the proteins in the alignment because if the two sequences have highly similar folding mechanisms than that can be taken into account to give a more accurate biological representation of the alignment. So if the sequences have similar folds then the amino acids that are crucial for that folding pattern can be aligned in the sequence. For example if the sequences both exhibit alpha helices then the crucial hydrogen bonding amino acids responsible for the formation of the helix can be aligned. Even if the sequences are highly different a normal sequence alignment might not reveal these amino acids as being aligned but if the sequences have these similar folding structures than the residues responsible for the folding can be paired together to give a more accurate alignment. This will also give more insights to the divergence of the different proteins because even though shape or general folding patterns might be conserved this does not necessarily mean that function is conserved as well. Aligning the proper folding residues will accurately show which residues have changed and will better reveal how function has changed over time.

Furthermore a comprehensive study of all known proteins that relates their shapes and sequences to one another would yield interesting results. Such a study would be computationally highly complex and time consuming, not to mention highly data intensive (it would take massive amounts of storage to retain all of this information). However assuming it were completely possible in a reasonable amount of time such a study would yield interesting results correlating structure and sequence between various enzymes. This would be particularly interesting for highly dissimilar proteins as it could reveal, say for any given protein, increasing similarity as a function of structure similarity or vice versa.

The significance of a conserved amino acid in members of a subtree is that this amino acid is highly specific to that subtree. Functionally though that amino acid helps the organism of that protein survive in a way that is unique compared to the organisms elsewhere on the tree. This amino acid is also probably partially responsible for the divergence of this subtree from the rest of the tree. This amino acid gives the organisms of this subtree an advantage in a certain environment and so has pushed these organisms genetically away from the other members of the tree. Clearly this amino acid also allows this protein to function in a way that is different than the proteins above it and other places on the tree. This unique function of the different sets of proteins of this subtree brought about by this conserved amino acid is what allows this organism to have adapted to survive in a unique way as described above. For example this amino acid might make this protein more stable in an adverse environment that these organisms of this subtree have inhabited. The other members of the tree may have never needed this extra stability and so there is no reason for this to have evolved elsewhere.

Highly conserved amino acids on active sites of these proteins are most likely so conversed because they contribute to the most basic functioning of the homologous proteins represented in this tree. While the proteins of this tree may all vary greatly to hardly at all they all are were homologous at one point, as given in the problem description, as such they all at their core function in at least some minimal way. Clearly if these highly conserved amino acids are clustered near the active site they are responsible for the elementary functioning of this protein and are needed to do its work. Perhaps they bind to some portion of the substrate that has not changed much over time. Maybe they work to capture or draw in the substrate from the surroundings and as such have not changed much over time because they are required to do the work of attracting this protein’s ligand. These conserved proteins may also do the mechanistic work of the protein. So perhaps throughout the tree the diverging proteins have differentiated to act on slightly different substrates or maybe they function in different environment. But if they perform relatively similar functions then the key amino acids that are responsible for the mechanism behind the chemical action of the proteins must be conserved in order for the protein to do its job.