

## Responses to Problem 4.5: What fraction of proteins tried?

*PS4.5. Plants and photosynthetic bacteria have been enjoyed enormous success in part because of their ability to harness sunlight to power the reduction of  $\text{CO}_2$  to sugar. Organisms that can use sunlight as to drive the reduction of  $\text{N}_2$  to biologically useful nitrogen compounds are far more rare. One reason for this is that the enzyme that catalyzes the nitrogen reduction is extremely sensitive to  $\text{O}_2$ , greatly limiting the environments in which nitrogen fixation can take place. Some have said that this is just the way it is – it is not possible for an enzyme to fix  $\text{N}_2$  without also being killed by  $\text{O}_2$ . Why? Well if such an enzyme could exist, it would have arisen some time over the last 4 billion years of organismal evolution and organisms with this capability would have taken over the world!*

*Let's examine this reasoning. Have all possible proteins been tried sometime during the lifetime of the earth? If not, then what is a reasonable estimate for an upper limit on what fraction have arisen.*

A.1. I don't like believing in absolutes. Anything is always possible. So no, I would not agree that all of the possible proteins have been tried during the lifetime of the Earth. It wasn't oxygen that was one of the first gases to be used to form energy, we found that carbon dioxide existed before that. Animals adapted to the changes and some could harness the gasses as energy while some could not. I am sure some prokaryotes in the deep oceans have proteins that regulate toxic gases such as sulfur into energy. We also don't know how many elements there are that exist for sure. Some of those newly discovered ones could be potential energy contributors to certain creatures and we just might not know it. So to answer the question, I would say a lot less than 1% of all possible proteins have been tried during the lifetime of the Earth to catalyze these reactions to make energy. The earth is still going, so oxygen may eventually be replaced by something else and then there will be a new protein or old one that will catalyze that for energy.

A.2. Not all possible proteins been tried sometime during the lifetime of the earth. Since there is many limitations for nitrogen fixation, the reasonable estimate for an upper limit would be 1. No matter what happens, the fraction should never exceed 1. If the fraction was over 1, organisms would have gradually increased in over the last 4 billion years.

A.3. I would say no given the fact that certain proteins may have not see this function as necessary for prolonging their existence or worth the work and energy input. It can also be the case that one might arise or have arise and due to the increase in atmospheric oxygen it was destroyed. I am a bit fuzzy on what the second portion of the question is asking given it wording. But, if we are estimating how many attempts would have to happen to test this theory I would say an upper limit of  $2^n$  where n represents the number of proteins that exist on the planet.

---

B.1. Nitrogenase is an enzyme that is capable of fixing  $\text{N}_2$  to  $\text{NH}_3$ . In photosynthetic organisms this enzyme might encounter difficulty because a product of such an organisms respiration is oxygen. Oxygen irreparably damages nitrogenase (both subunits contain an iron atom)

B.2. In the protein article it is written that, "There are 20 possible amino acids in the first position and 20 possible amino acids in the second position. That makes  $20^2=400$  possible dipeptides. Similarly, there are  $20^3=8000$  possible tripeptides. Proteins range in size from a smallish 100 amino acids to a 1000. The number of possible proteins in nature is therefore staggering!" It is impossible to count the number of all possible proteins that have been tried sometime during the life time because there are so many combinations of proteins which are present in almost everything, from food to mammals like us. Also it is possible that all the possible combinations have not been tried or used in a lifetime. It is estimated that  $4 \times 10^{43}$  different amino acids have been sequenced since the origin of life ("How Much of Protein Sequence Space Has Been Explored by Life on Earth?").

B.3. Considering the fact that the largest known proteins are Titins, a component of the muscle sarcomere, and it contains the largest polypeptide chain of almost 27,000 amino acids. Then there are 20 amino acids with 27,000 possible positions and arrangements. If you take  $20! (20 \times 19 \times 18 \times 17 \dots \times (n+1) \times (n) \times (n-1), n \geq 0)$  TIMES 2700 possible positions, then you would get  $6.5688E21$  possible amino acid arrangements as an upper limit, making it reasonable to say that all possible proteins have been tried during the lifetime of the earth.

---

C.1. You cannot estimate the number of possible proteins because they're a "staggering" amount of them and it would be too many to approximate. An upper limit fraction for the different amino acid sequences explored since the origin of life is  $\sim 4 \times 10^{43}$ .

David T.F Dryden, Andrew R. Thomson, John H White. "How much of protein sequence space has been explored by life on Earth." 2008. J.R. Soc. Interface: 5:25:953-956.

C.2. Although there may be only 20 amino acids, the possible combinations for proteins can actually reach infinite. Protein chains are not limited to a certain number so there may be amino acid chains as short as 10 or even longer than 400. Additionally, there are many more possibilities, even within the 10-molecule long peptide chain due to the importance of the order of the amino acids. Therefore, I would most likely give a wild estimate of couple billions of proteins in existence in the world right now since not all the combinations of amino acids may be valid or be used right now.

C.3. The Earth is around 4.5 billion years old, so that can be written as  $4.5 \times 10^9$  years. The total number of possible amino acid sequences that can make up a protein can be written as  $20^{300} \times 10^{300}$  since there are 20 amino acids coded by three codons of DNA that make them up, therefore the possible number of sequences can be shown as that number above. All of them have probably not been created though seeing as organisms do not make different proteins each time their cells go through mitosis or meiosis, they tend to make the same protein. And there is only a one in a billion chance that each time the cell divides, its DNA will be mutated. Let's look at bacteria, some bacteria can reproduce at about every 20 minutes, and theoretically let's say that the bacteria cover the surface area of the Earth, which is roughly  $5.18 \times 10^8 \text{ km}^2$ . An average bacteria is about 2-5 micrometers in length, so that means there is one billion bacteria for every kilometer of Earth. Just a rough estimate, but say every bacteria is one micrometer, there would be  $5.18 \times 10^{17}$  organisms of bacteria on the earth and if they make new proteins every 20 minutes about 10 of them would make a mutated protein. Let's say all their proteins way 50 kDa and the bacterium way 2000 kDa each so there are 40 proteins per bacterium. So there are 40 proteins multiplied by  $5.18 \times 10^{17}$  for every 20 minutes. So since 10 mutated proteins are made every 20 minutes then since the beginning of time, so that is about 52594 different proteins per year since the beginning of time. This can be written as  $2.4 \times 10^{14}$ . This is not accurate seeing as some of the organisms will die off and there will be a lot more bacteria before they first started reproducing. Saying the Earth has the resources to support these bacteria it would be around the number above or more. Even if it was more, the number of proteins created still does not come close to the number or unique amino acid sequences. Based on these perceptions of mine, I'm going to estimate that about 1/30 of the proteins have been tried since the beginning of time.