Evolutionary Dynamics
Supreet Saini
Chemical Engineering
Indian Institute of Technology Bombay
Week 01
Lecture 08

Hi, welcome back, everybody. So, I hope the concept of the central dogma of molecular biology is clear to everyone, and just to recapitulate what we did. So, now we should all be aware of that DNA is a double-stranded structure; the two strands run in what is called an anti-parallel manner to each other. A and T are linked with double bonds, G and C are linked with triple bonds, and so on and so forth, and no other pairings are possible. And when the process of transcription happens,

a machine called RNA polymerase comes, sits here, reads this DNA, and in that process of reading the DNA, synthesizes a molecule which is mRNA. So, the result of the process of transcription is this mRNA, which is single-stranded. Instead of deoxyribose, it uses ribose. And instead of T, it uses uracil or U. This is then followed by the process of translation. where a machine called ribosome comes and sits near the 5' end of the mRNA.

It starts scanning. It starts incorporating amino acids when it encounters an AUG, which is a signal to start the process of translation. AUG also corresponds to methionine. And this process of reading one triplet at a time continues until it comes across a triplet, which is a signal for it to stop the process of translation, such as UAA. And we saw at the end of the last video that there are these three codons which correspond as a signal to end the process of translation.

And again, as the ribosome is moving on this, let's redraw this, as the ribosome is moving on this. On this mRNA molecule, it reaches the three prime end and it is at a signal which is the stop. In the process, what has happened is that it has made this amino acid chain which comprises these amino acids being linked and so on and so forth. And this is called an amino acid chain. And once the ribosomes disassemble, we have this amino acid chain free, and this folds to minimize its free energy.

And this happens in accordance with the principles of physics because some of these amino acids are hydrophobic, some of them are charged, and some of them are hydrophilic. So hydrophobic parts will obviously fold to avoid contact with water. Hydrophilic parts would

love to be on the surface and be associated with water and so on and so forth. So eventually, this amino acid chain acquires a 3D structure, and that's when we have a protein molecule. And this protein molecule will do its cellular function in a case where this is an enzyme, which means it catalyzes reactions.

Now, this folded amino acid chain protein molecule will catalyze this reaction of substrate going to protein via the kinetic and thermodynamic limits under which it operates. So this is the process, this is the central process that we've been meaning to discuss, called the central dogma of molecular biology. The genetic code that we saw, the association between 64 codons with 20 amino acids, is nearly universal. With only a very few exceptions of organisms where a codon does not correspond to the same amino acid to which it does in almost every other organism. What you should realize here, the significance of all of this, is obviously that without proteins, we wouldn't function, and proteins result from this process of transcription and translation.

But also, I as an individual, I as an organism, have my capacities described by what proteins function. Am I making them in my body? The fact that I can see, the fact that I can think, the fact that I can move my arm are all things facilitated by proteins working in different cells in different parts of my body. So proteins dictate all of my phenotype. So my body's phenotype is dictated by what proteins I have.

I as an individual and my capacities are dictated by what proteins I have. And what this process of central dogma should tell you is that what protein molecules I have is dictated by what mRNA species I have, what precise mRNA sequences I have. So mRNA dictates what proteins I am going to get, but mRNA is dictated by what DNA sequence I have. So eventually, it boils down to what the sequence of my DNA was, and that will decide what protein molecules I have across my DNA, and that is going to define my capacities as an individual. Small changes in proteins, so imagine there was a G, but some error happened so that it became an A. This can happen in the process when DNA is being duplicated.

And an error as small as one nucleotide change, instead of a G, you had an A. An error as small as this can be responsible for a disease state. And we know of several examples where only one nucleotide change in the total nucleotide length of human DNA, which is 3 billion nucleotides, only one nucleotide change can cause a serious disease condition. And these details are just quite extraordinary. So that's the process. So what we'll do in this lecture is try to understand how this all works at a cellular level by taking a precise example.

And the example that we will take is a classical textbook example of gene regulation in bacteria. So we'll look at a bacterium called Escherichia coli, simply called E. coli. And how it uses lactose as a carbon source. That is the process that we will look at to try and understand how the central dogma works to express genes whose information is contained in the DNA to make the resulting proteins and do something as essential as utilizing a carbon source for growth and energy purposes. So lactose is a disaccharide, which means it's two sugars that are linked together by a chemical bond.

The two sugars are, let us say the circle is glucose and the triangle is galactose. So that is one molecule of lactose. And the way E. coli uses it is that imagine this is an E. coli cell. Typically, E. coli will be about 1 to 2 microns in length. The way E. coli uses lactose is that it will have a dedicated transporter.

Whose job is to internalize the lactose that is in the environment and bring it inside. So this was lactose outside, and now it's inside. And this is a protein molecule whose only job is to bring lactose from outside the cell to inside the cell. After that, E. coli encodes another protein molecule, which is an enzyme. So, this is a catalyst, another type of protein.

And the job of this protein is to break down this glucose-galactose disaccharide into its constituent monomers, which are glucose and galactose. So then glucose can be used for growth and energy processes via its own pathway, and galactose can be used for growth and energy yield via its own pathway. The key thing that we have to notice here is that these two proteins, protein number one and protein number two, are needed for the transport of lactose from outside to inside and for breaking this lactose into glucose and galactose. These two have no other purpose inside the cell. As a result, the question that we should ask ourselves is the following.

Does it make sense for E. coli to make proteins 1 and 2 if lactose is not present in the environment? So imagine that this E. coli bacterium finds itself in a pond, and there is some glucose available, there are some other carbon sources available, but there is no lactose. So clearly, via the central dogma, we saw that all proteins come from the process of transcription and translation. That means the information to synthesize protein 1 and protein 2 is contained within the DNA of the organism.

But if the environment does not have any lactose, the carbon sources are other sources, then does it make sense for this individual bacterium to synthesize these proteins 1 and 2 via the process of transcription and translation? I should also add that making proteins is energetically very expensive because you need to make all the amino acids necessary for a

protein molecule. And when a cell makes a protein, it doesn't make one copy of this protein one. It will make several, all of which will be embedded in the membrane.

So, it's an expensive business making these proteins. And cells obviously have access to a finite amount of resources in an environment as sparse as a pond, for instance. There are not a lot of resources around, and you have to be judicious in terms of where energy expenditure is taking place. On top of that, these two proteins have only this one dedicated job that they perform: protein 1 serves only one purpose, which is to bring lactose in, and protein 2 only serves one purpose, which is to break this lactose into glucose and galactose. They serve no other purpose.

So, if there is no lactose in the environment, this process is not needed, this process is not needed. Hence, there is no point in making protein 1 and 2 in an environment where lactose is absent. So, we want to understand how this cell does this process. So, if lactose is absent from the environment, We need a logic system such that the cell does not make protein 1 and 2.

This would save costs. Just like households have budgets, cells also have their energy budgets. And not making protein 1 and 2 when they are not needed at all helps the cell save those costs. This is how an E. coli would do it. So again, let us zoom into the cell.

And imagine this cell in an environment where there is no lactose. This is the DNA of E. coli. Bacterial DNA is often circular in nature, unlike ours, which are linear threads. Not all bacteria have circular DNA, but E. coli does. So, let us imagine that at some point in this DNA, we have a sequence which corresponds to transcription, which will take place, and we will get protein number 1.

This protein 1 will get embedded in the membrane and bring lactose. We also have a sequence for protein number two, which will get transcribed and will give rise to protein number two. And this protein two will break down lactose into glucose and galactose. But remember, this is in an environment where there is no lactose. You should remember that all of this starts—the process of transcription, this central dogma of protein synthesis—starts with the process of recruitment of a machine called RNA polymerase.

This RNA polymerase has to be recruited, and then the process of transcription starts. RNA polymerase has to be recruited, and then the process of transcription starts on this gene, and so on and so forth. However, What will happen if there is no lactose around? The cell doesn't want transcription to take place because that is just unnecessary expenditure.

In order to save and prevent this transcription from happening, what the cell does is encode for another protein. That protein is called the LacI protein. LacI comes and binds, and what LacI's binding does is that it physically prevents the RNA polymerase from coming and binding to the region here, just upstream of gene 1 and gene 2, and starting the process of transcription. So you have RNA polymerase waiting to come and bind here. And start the process of transcription.

But because LacI is already sitting there, RNA polymerase cannot go in. And if RNA polymerase cannot go in, the process of transcription doesn't start. The process of translation cannot happen. And hence, you result in no protein 1 and protein 2 being made when there is no lactose in the environment. So that sort of solves the problem when there is no lactose.

I don't want protein one and two. And that is defined here. So this LacI business explains that when there is no lactose. No protein one. No protein two.

And this is accomplished via LacI. LacI is called a repressor because it represses the expression of protein 1 and protein 2 via its action of coming and binding to the regions just upstream of these genes. But now all this is fine, so there is no lactose, and I don't need protein 1 and 2, and hence I have shut them off. But now suppose an environmental change takes place, and I have lactose in the environment. So lactose is here.

And now I do want protein 1 and 2 because that is my only source of carbon and energy. In this case, what is going to happen is that as this molecule diffuses inside the cell, I have some lactose which comes inside the cell. LacI has a great affinity for lactose. So as soon as lactose comes inside the cell, these two are going to bind to each other, and the resulting species will be a LacI molecule, but this time it's bound to lactose. So this is the LacI-lactose complex.

But what this complex does is that now, while LacI had a great affinity to bind with the DNA and prevent transcription from taking place, this chemical entity, which is lactose bound to LacI, cannot bind the DNA here. And hence, it's lifted from the DNA regions where it was binding. And if this lactose is gone—I'm sorry, if this LacI is gone because of the binding with the lactose molecule—now this RNA polymerase is free to come and bind to the DNA and start the process of transcription. And the same story happens here. So now that LacI has been recruited away from the DNA by binding with this lactose molecule, RNA polymerase comes and binds and starts transcription and translation.

Protein 1 will make this transporter protein, which brings in more lactose. And protein 2 will hydrolyze this lactose into glucose and galactose, which will then be further processed for the growth and energy requirements of the cell. So the essential logic here remains the same. So via this regulatory network, we have now encoded this logic that no lactose means no protein 1 and no protein 2. But when lactose is around, that's a signal for the cell to start making protein 1 and start making protein 2, starting their synthesis via transcription and translation.

And this is the process of gene regulation. And from the context of evolution, what we must realize is that a small change here, even the change of one nucleotide, let's say going from an A to a G, if this error happens, it could totally mess with the binding ability of LacI to this. So the repression ability is gone. If the repression ability is gone, this logic is gone. At the same time, a small change here, maybe a small change there, maybe a G changing to T, changes the mRNA sequence and changes the protein sequence.

And that means this transporter can no longer bring in lactose. And as a result of this, this individual is at a terrible disadvantage when it comes to utilizing lactose for growth and energy. On the other hand, you could have mutations that are great, which means maybe a C going to a T. This means the mRNA sequence changed, protein 2 changed, and this new variant of protein 2 is actually even better than the earlier one at hydrolyzing lactose into glucose and galactose. So this would be one type of mutation. An example of a mutation that happened in the DNA which resulted in a change in the protein sequence such that the protein was functioning even better than the earlier one.

This would be an example of a change where a mutation resulted in a change in the amino acid sequence and hence a change in the protein, which made its functionality poor. And in this way, depending on the environmental context and the change that is taking place inside the cell, we could have good changes and bad changes, and so on and so forth. And these changes then result in evolutionary change over populations, given a sufficient amount of time. And just as a matter of detail, this transporter protein is called the lacY protein. LacY is the lactose transporter in E. coli.

And this enzyme, which breaks down lactose into glucose and galactose, is called lacZ. And this is sort of the first detailed model of gene regulation that we have developed an understanding of. So what I hope to have communicated in this video up until now is how the central dogma operates to make protein molecules by reading DNA. What protein

molecules I make is dependent on my DNA sequence. So I am defined by the DNA sequence.

Subsequently, We can have these regulatory patterns where genes are expressed into proteins only when they are needed. So through this example of E. coli, we have this understanding of a system where, when a cell does not need a protein, it knows how to shut production off. And when a cell does need a protein, it knows it has the capacity to turn the production back on. So this off and on can keep on happening.

And this is highly dependent on the environment in which a cell is growing. And via small changes in DNA, we can have large changes in the behavior of the organism. For instance, if the cell loses its ability to import lactose inside the cell, what happens with protein 2 is irrelevant because if this protein is not functional, then there is no lactose inside the cell to be hydrolyzed and used for growth and energy. So all of this forms a very intricate network of what is happening inside the cell. And based on these principles now, in the next video onwards, we'll in earnest start our study of evolutionary change.

We'll continue this with the next video. Thank you.