

There were some other questions sort of running along this general idea of the fact that the information in DNA doesn't go, even though it encodes the information for proteins goes via this rRNA intermediate. Someone asked what was the M.

The M is for messenger. The idea being that since the DNA, at least in eukaryotes the DNA was in the nucleus and proteins were made out of the cytoplasm, somehow that information had to be carried from the nucleus where the DNA was out to the cytoplasm.

And that's where the term messenger was because the RNA was seen as something that would carry the information out.

Now, a point here, it's really critical because we're going to continue to talk about gene regulation.

And that is when a cell is making one of these mRNAs, it doesn't make one single copy of all of the genes that are in the genome on one RNA. Instead it does it either one gene at a time, which is the usual case, or occasionally as we see in the lac operon a little tiny cluster of DNAs that have related functions.

And the beauty of that is that it then enables the cell to dial in how much protein is being made, in part at least, by determining how much RNA is being made. So you're going to make no RNA and not make the protein at all or make a little RNA and get a little protein, or if it's really a thing you need very large quantities of you can really crank out a lot of RNA and make a lot of protein.

So the potential is there for regulation. And in bacteria, as I said, for almost all bacterial genes it's pretty straightforward.

You can look in the DNA, and at least if you know where to start, see the start of a protein, you can just use that table of the genetic code and read off the sequence. But eukaryotes in particular, higher eukaryotes have this odd business that what seemed odd and surprising that when you look at their genes, many of them you cannot do that because it's as if there are extra bits of DNA stuck in the middle. And in some cases you heard it could be really huge amounts of DNA so that there is this extra thing where there's a pre-mRNA.

And this RNA splicing we talked about has to take place to generate the mRNA. And once you have the mRNA then the ribosome and the charge tRNAs can be used to make the proteins. And someone asked what all that extra DNA is for. I mean we still don't fully understand that. There are some regulatory sequences and regulatory actors that are buried in that non-coding DNA.

But another thing may be that this is just the way it's worked in evolution. And as long as it works there's no driving

force necessarily to get rid of it. If you look at microorganisms, for example, yeast is a eukaryotic that, like E. coli, has to replicate pretty fast in order to compete with other microorganisms for the food and whatnot in its environment.

And it has relatively little of these extra intervening sequences compared to what we find in our DNA. I gave you the example of Factor 8.

It didn't really particularly matter what it was in the sense that it was just an example of something that has a lot of intervening sequence. What it is, though, it's one of a set of proteins that are involved in clotting of your blood.

When you cut yourself, we have this system that prevents us from bleeding to death, unless you have hemophilia or something like that where's there's a problem with the clotting system, then a very complex set of things happen. And Factor 8 is one of the several proteins that play critical roles in that.

Somebody asked, I talked about a few things that were, this was sort of the dogma. This was how information was thought to go. And then how Dave Baltimore found there was reverse transcriptase that could take an RNA and make a DNA copy.

And the question was they didn't understand how the dogma could change. I mean that's sort of what I'm trying to emphasize a lot in this course is what I'm teaching you is what human experimentation and thought has brought up until spring of 2005 of terms in biology.

Some of the sort of basic discoveries were made in physics so long ago that it's very unlikely that you'll come in and discover that the Newtonian mechanics you learned as a freshman is not operative anymore when you're a senior, but you can still have these massive revolutions in biology where just suddenly whole things, like RNA splicing, emerge from the woodwork almost overnight.

And that's, in fact, what happened with that. That's what happened with reverse transcriptase. So, in a sense, it's almost a joke that Crick called it dogma. He didn't know what he was doing. But it sort of took that property on.

And I'm trying to caution you that even though some of you would like me to stick to just facts that we're continually learning and there are discoveries being made even as we're going on with this course.

Now, I told you that there were certain viruses, HIV being the one I really emphasized, that have this property. Their genetic material is not DNA.

It's RNA. And the reverse transcriptase makes a double-strand copy that it inserts into the organism's DNA and becomes a permanent part. And I'm trying to caution you that's why safe sex is such a big deal. Because if you get infected with HIV you'll have it for the rest of your life. I mentioned there were some cancer viruses, and

someone said, well, if you do that how can you cure cancer? Well, in fact, we're lucky in the sense that we don't have to contend at this point in a major way with human cancer viruses, that would be more of a problem, but your cats have to. You may have heard of the feline leukemia virus. This is a retrovirus of this same class, and cats infected with it have it in their saliva.

And although it can be transmitted amongst pets in the same household, the big problem is usually cat fights.

And then you get a scratch and then it gets in. So if you have a cat you probably had to take it to the vet and get vaccinated.

And one of the reasons is you're trying to get it vaccinated against the feline leukemia virus. And it's just sort of like that story I told you with the streptococcus, that if your immune system has seen the thing beforehand then if you actually get an infection it has a very quick response and your cat doesn't get infected by the virus. And therefore doesn't become a candidate for getting leukemia in later life. OK.

So at least there's an effort to try and respond to at least a few of the things. There were some very interesting and thoughtful questions. So I want to now go back to talk a little bit more about this issue of regulation because that is one of the real secrets to life. And it's the ability of organisms to turn on and off certain functions and to have rheostats where they can control the levels of expression. But all of this has to be in the DNA.

And so before people knew how this worked, it's a little mysterious.

And maybe one way you might think about it, well, if I have a gene that encodes something for lactose metabolism, and I tell you there's a sequence upstream of this and somehow this gene is regulated depending on whether I put lactose in the medium or not, how is it going to work? Does it have to fit into little holes in between the letters of the genetic code sort of the way Gamov originally thought about it, or is there some other mechanism?

And what you'll see here is one of the general strategies that evolution has chosen is although there's regulatory information in the DNA, DNA isn't a particularly good molecule for recognizing things, but what it is good at is encoding proteins.

So the trick is to have some proteins made whose role in life are to be regulators. So that this is what is underlying this system that I started to tell you about on Friday.

So remember -- -- beta-galactosidase is the enzyme that takes lactose which is lactose beta-1,4 glucose. And somebody said it's easy to get them mixed up. I apologize, but those are the names. And cleaves it, just breaks the bond to give galactose plus glucose. Both of those can be metabolized by ordinary elements you'll find in most

cells.

But taking lactose, which you also know as milk sugar because it's in milk, needs this extra function.

If you're lactose intolerant, as a fraction of you will be because that's quite common in the human population, then, although you had the enzyme when you were baby, it's been shut off in your body since then and it causes problems. Because if you eat lactose, drink milk or something, the lactose goes right through your stomach and ends up in your intestine.

And there are bacteria in there that are able to break it open.

And when they do that it leads to gas and some other sort of uncomfortablenesses that are associated with lactose intolerance.

So, as I'd said, the major finding was that this enzyme beta-galactosidase or beta-gal was regulated. That if you grow E.

coli on glucose -- -- there was no beta-gal. And if you grow them on lactose as the carbon source then there were high levels of beta-gal.

And that then lead Jacques Monod and Francois Jacob, the two French scientists I mentioned, to begin studying this problem. And, like so many things in biology, they were working on a huge problem, how are genes regulated?

But it wasn't so evident at the beginning. What they were doing was this very modest thing, why is beta-galactosidase there in one condition and not another? Just a little problem in bacterial metabolism. And ultimately it gave us the roots to the answer of how genes are regulated. And I just have to leave out all the beautify stuff that led to what they found. Let me just recapitulate what I put on the board the other day. It turned out that the LacC gene, this is the gene that encodes beta-galactosidase, so that's the sequence, that has the sequence of codons, that if you could start at the beginning and go along and just put all the amino acids in you would end up with beta-galactosidase.

That unit of genetic information, which is called the lacC gene, let's put it here. Then there were two other genes just downstream of this in the DNA. And this unit is made as a single mRNA.

So this is a little bit different than what I told you.

It's not just one gene. It's actually two or three because these bacteria try to do everything very efficiently because they're growing quickly. This means it can turn on three genes of related function very efficiently. When you have

several genes that are expressed using one mRNA, as I said, the genes are said to be organized in an operon.

But the key point that we talked about before is you're going to have these genes expressed everything has to be written in the DNA.

And it's not using the genetic code. It's using other words that are written there. And you've seen this word now in several lectures. That's a promoter.

And that means to start transcription.

And to stop the mRNA there has to be something at the other end.

And I'll show you the sequence of at least one of these promoters in a minute. This means to stop transcription, to stop making the RNA copy. If you didn't have sequences like that the cell wouldn't know where should I begin the RNA and where does it end. And since there are many, many genes, there have to be many promoters and many terminators.

And the other point I tried to hammer home the other day is although the genetic code is universal, you can take that little table and read the sequence of human proteins or E.

coli proteins, these other languages that are written using this four-letter nucleic acid alphabet are not universal.

So the sequences that E. coli uses for a promoter, a start transcription are very different than what our bodies use as a start transcription thing. And when we get to the recombinant DNA stuff that will be an issue. So this mRNA is then used to make proteins. This would be beta-galactosidase or the lac-Z gene product, and these other genes make, so these are proteins. Here you see this flow of information from the DNA through the mRNA down to being made into proteins. In the case of bacteria there's no nucleus. Everything is in one big pot so that mRNA doesn't have to go anywhere, but it's all there and it gets translated to give copies of the protein. So, as I said, somehow if the cell is going to now regulate whether these genes are expressed or not, depending on whether lactose is present.

And the way they do it makes perfect sense. Don't bother to make the enzyme if there's no lactose in the neighborhood, and only make it if it's present. So how are they going to do that? Is the lactose going to come along and stick into some little hole here or something?

That's not a general strategy and it wouldn't work if you look at the structure of DNA anyway. It wouldn't have access to the sequence of bases. So what was discovered was that there was another gene very close called lacI. And the protein that it encodes is called the lac repressor. And since it's a gene it has to have a promoter, a start

transcription.

But this one, be careful now, don't get yourself mixed up, this is for the *lacI* gene. This is a different promoter over here for this thing. And then there is also then a terminator or a stop transcription. And, again, this is for the *lacI* gene. So this gets made into an mRNA as well.

And this gets translated into a protein that's known as lac repressor. And what that lac repressor has is the ability to bind to a particular sequence in DNA that's located right here.

This is a binding site right here for lac repressor. So let me just try to blow that up just a little bit because this could be a little bit confusing.

So here's the promoter -- -- for this *LacCYA* operon. Here's the beginning of the *LacC* gene. And so this is where the RNA polymerase, the machine that's going to make the RNA copy has to bind. And here's the binding site.

-- for lac repressor or the *lacI* protein. So how does this circuit work? And I sort of pose that as an issue for those of you who had followed it at least to that point.

So let's consider the two situations.

If there's no lactose present what we know from just scientists knew from experimentation was there was no beta-galactosidase activity inside the cell. You could crack them open and you wouldn't find this enzyme there. And when you added lactose they knew, from that experiment described on Friday, it was synthesized de novo. So if there's no lactose present -- And what happens is we have the *lacI* gene, mRNA is being made, this lac repressor is being made, here's the promoter for *LacCYA*, and here's the sequence. And this repressor goes up and binds to that. And by binding to this particular sequence what it does is it covers up the promoter.

It covers up the start signal, the signal that says "start transcription here". Are you guys with me?

It's a relatively simple strategy. It's just by lac repressor having that ability to bind to a particular sequence it's able to prevent the RNA polymerase from seeing the promoter.

And therefore it's able to prevent the RNA from being made.

So there's no mRNA. And if there's no mRNA over here then there's no beta-galactosidase being made. This is an exercise in futility.

Why has the cell gone and made this useless protein that isn't doing anything in terms of helping it metabolize lactose?

But now take a look at the system compared to what I described when we were first doing it. If we had to do all the regulation directly with the DNA we have this problem that lactose would somehow have to be able to see a sequence in DNA and somehow determine what happened. But what this cell has done now is it's set this system up so that the ability to make lactose or not make lactose is conditional on this protein called the lac repressor.

If lac repressor is bound, as it's shown here, it's basically covering up the promoter, the cell cannot make RNA and cannot make beta-galactosidase. If it was absent, if we just got rid of lac repressor then the promoter would be exposed and the cell could make beta-galactosidase. And so it's now lac repressor that has the conditionality. It's, in essence, a sensor.

It can at least be considered a sensor for whether lactose is present. And indeed that is the property that lac repressor has.

It's able to bind lactose. So if we think lactose is present what happens?

I mean this *lacI* gene is a pretty uninteresting, uninteresting from the standpoint of regulation in the sense that it's made all the time. The cell just continually cranks out a bit of lac repressor. It doesn't need very much. It just needs to make enough so that the one binding site for lac repressor has somebody bound to it. So it can get away with pretty low levels. But over here then we have this promoter, and we have the *LacC* gene and so on here, and there is the binding sequence right there. But this lac repressor has the ability to bind lactose, which I'm going to draw as a little triangle here, even though you know it's a disaccharide. It has a different property.

But the fundamental characteristic of the lac repressor binding lactose is it undergoes a change in conformation. So if it's got a binding pocket, and lactose fits into that binding pocket, those alpha helices and beta sheets and so on move around a little bit. And what happens then is it perturbs the part of the protein that would normally be able to recognize this DNA sequence.

And this cannot -- -- bind to the DNA sequence up here. And I'll tell you the special name for that binding site. It's just one of these terms that you'll see in biology. Everything has to be given a name.

It's called an operator, for historical reasons. But, in any case, the lac repressor, once it's hanging onto a lactose it's unable to bind this sequence. That means that the start site for transcription is made. And so you get the mRNA made and then you get beta-galactosidase present.

A little bit complicated, but sort of underlying it is this idea that now the cell is using a protein rather than DNA to tell whether lactose is present. And hopefully you can see this is really general now because you can design a regulatory protein.

And basically it's got to have two things. It's got to have a part that talks to the DNA and recognizes some sequence, and it's got to have another part that senses whatever it is.

Histidine, temperature, you name it. But once you understand that design principle then you can begin to see how it is that the cell is able to turn genes on and off just by encoding information in the DNA. And, as I say, one of the big tricks there is to let the protein do the sensing for you.

Now, I just want to give you a little bit of a blowup of what this things looks like. Because sort of all I've done is kind of put it here as a sequence. I've called it a promoter. What a promoter then is, again it means the start for transcription, the process of making RNA.

And I've tried to stress that these promoters are not universal.

So when I tell you this for E.

coli, this is what a promoter for E. coli looks like, but it doesn't look at all like a promoter in our bodies. And so it's basically a word that's written using this nucleic acid alphabet. And it looks something like this. There's TTGACA and then there are about 17 base pairs that can be just about anything.

And then there's TATAAT. And then there's another little bit here that's about ten base pairs long. And then this is the start of the mRNA which is usually given the convention of being called the plus one position. So you'll notice this word I've called it, written, that says start transcription has even got two parts to it. And this is usually referred to the minus ten region of the promoter, and this is the minus 35 region because that's the distance from the start of transcription.

It maybe seem sort of weird to you to see what I'm telling you is sort of word written in the nucleic acid language.

It's got some bits in the middle that don't matter.

But remember the DNA is a helix and things going around like this.

So if you were just to take a DNA helix and then lay something down on one side of it, it would contact it here, it wouldn't contact it there, and then when it came back up again it would contact it here. And so as these things hang on along the sides of DNA it's not at all uncommon to find this sort of broken where you can have something that matters, something that doesn't matter and something that matters again.

The RNA polymerase in E. coli, it's a machine. It's got four proteins that are the core. That's the part that actually synthesizes the RNA plus one protein, which is known as the sigma subunit. And it has the special job of recognizing the promoter. And so when we start asking where does this regulatory sequence that the lac



repressor binds it in all of this. It turns out that the sequence for binding lac repressor -- overlaps with this minus ten region. So when the lac repressor is sitting down it's covering up a very important part of transcription.

You guys with me? OK. So this is an interesting kind of regulation. It's given the general term -- negative regulation.

And the reason that term is applied, it means that the regulatory protein -- interferes with transcription.

And let's take a brief foray into we're going to talk about genetics as our next subject. And I think maybe we can let you sort of already get a sense of how some of this was figured out.

So there's a substance called X-gal, which is a galactose with some chemical entity hanging off that's colorless. But yet it's a substrate textbook. And so if we grow E.

coli on plates that have glucose plus X-gal, the colonies would be we learn all this kind of thing. It just doesn't come out of a whose bond can be cleaved by beta-galactosidase. experimentally I'm not doing too good a job of conveying to you how colorless. And if we were to grow them on plates that had lactose plus but I think if you don't have some sense of how this is done

Someone said they thought this was too much lab stuff, X-gal then all of the colonies would be colored because they're making

And this is a very useful thing for bacterial geneticists.

And it gives galactose plus the free X entity. And this is colored. beta-galactosidase. And part of the way that this stuff that I've been telling you was figured out was by bacterial geneticists looking for something. What they looked for was back here on this plate that had Xgal. Almost all the colonies were colorless. These were colored because they could make beta-galactosidase. So if I gave you some plates of this, you looked in the lab and then you found a colored colony, it's a mutant. I'll define these terms for you very shortly.

But it's got an alteration in the DNA that affects the regulation -- of beta-gal or the product of the lacC gene. And on the basis of what I've told you about this model, can you guys come up with two types of things, two places or kinds of mutations that could break this system that would lead to beta-galactosidase being on even though there's no lactose in the medium? Anybody see one of them?

Why is it off? Because of lac repressor? In a wild type strain it's because lac repressor is bound to that sequence and it's shutting off transcription. So we had a variant that could now transcribe. Yeah.

OK, so that's a good idea. So if we could somehow mutate that little binding sequence in a way that didn't screw

up everything else then, even though there was lac repressor being made, if it couldn't bind here because the sequence had been changed then you'd get it made. That's exactly right.

That's one of them. Yeah. OK, it was a problem making lacI.

What would happen? Well, if we couldn't make this and it couldn't bind there we'd be on. And that's the other class.

Can you think of a kind of mutation that we learned about, think back to the genetic code, that would prevent lac repressor from being made?

I'm trying to give you a clue. There were 61 codons encoded for amino acids. Yeah. Oh, that would work.

Yup, if we messed up the promoter. That's a sophisticated answer.

Yeah, if we messed up the promoter for making lacI that would certainly give that. Can you think of another type of thing that would affect the lacI gene, would prevent lacI from being made? Somebody? Yeah.

OK. If the sequence was wrong so that they could bind, that would be good. The one I'm trying to tease out of you, but I won't take longer now, is remember those three stop codons that didn't encode for anything? There should be one of those at the end of the protein. But if you changed one of the amino acid codons into a stop codon that would also prevent you from making it. So there are at least a couple of kinds of mutations.

And what I've sort of done here is I've skipped all the evidence and given you the model. And I cannot give you all the evidence that lead to this, which is a pretty well-established model. I don't think this is going to change likely.

We've been studying it for so long. But this is the kind of evidence on which it was based. It was by people finding things and figuring out that parts of the machinery were broken and then working on. So there's another kind of regulation known as positive regulation. And for a long time people thought maybe everything was negative regulation. But it turns out positive regulation is far more common. In this case, the regulatory protein instead of inhibiting transcription assists with the transcription. And it turned out, after people had been studying the beta-galactosidase system for a number of years, that it had a positive control system superimposed or together with the negative regulatory system.

The same thing engineers do all the time, pile up regulatory circuits and get all kinds of additional conditionalities.

And the thing I've told you saw far was we asked whether beta-gal is present. And the carbon source -- -- is glucose. This is low or not there.

If it's lactose it's high, but if cells were grown on both, glucose plus lactose, then beta-galactosidase is low again.

And this makes some physiological sense because E.

coli likes to use galactose. It's its favorite food source.

And so if it's got its favorite food source around then it doesn't want to make proteins that are used to eat all its sort of less favorite food source. So there's a nice conditionality here. In order for it, the circuitry is set up so that it only makes the enzyme for metabolizing lactose when the cell realizes its favorite food source isn't there and then it senses there's lactose. So only under those conditions does it make the enzymes for making lactose.

And, again, the way this circuitry works, this positive regulation needs two things. Once again, it needs a protein.

This one has given the name CRP. And, again, it's something that's able to bind to a sequence in DNA. And then it's also got a conditionality. It's able to recognize something else. And what this one recognizes is this small molecule that's known as cyclic AMP. It's just the familiar [ribomonophosphate? that you've seen before but it's looped around and formed an ester bond here. And that's why it's called cyclic AMP. But the important thing about this is that the levels of cyclic AMP are dependant on glucose.

So if you have high glucose you have low cyclic AMP and if you have low glucose you have high cyclic AMP. And here again is what's going to happen. So this is the promoter for LacC<sub>YA</sub> and here's the start of the lacZ gene. And I told you this is where the operator would be finding.

This CRP protein is able to bind to -- -- a site that's even a little bit farther upstream of the LacC gene than is the promoter.

And the idea of this is that if this CRP [lacZ?] -- -- just by itself, it doesn't bind to the DNA. But it has a little binding pocket that senses the levels of cyclic AMP. So if the cell is starving for glucose, there are high levels of AMP, then the CRP bound to cyclic AMP, this is capable of binding to this sequence.

So that sounds weird, but what we've, again, got now is we've got a protein whose binding to DNA is conditional to something inside the cell. And rather than getting in the way, what this does, if you have a situation where you have CRP with cyclic AMP bound to it and it's next door to this promoter, what it's able to do is help RNA polymerase recognize the promoter.

So this helps RNA polymerase. Whereas, when we were talking about the lac repressor what it was doing, if you recall, was getting in the way of RNA polymerase. And this actually has a relatively simple sort of molecular explanation for what's going on. The RNA polymerase -- The best sequence for the minus ten region of the

promoter, that I showed you over on the other board, is something with the sequence TATAAT. So the RNA polymerase machinery, it can recognize more than one sequence, but if it sees a promoter that has a minus ten region that is TATAAT, it really binds well and that would be a very strong promoter and you get lots of mRNA.

Now, the LacC promoter actually has two nucleotides that are different.

It's TATGTT.

So this is a very weak promoter without help.

So in this lac system, if we got rid of lac repressor entirely so that the promoter was just exposed all the time, as I showed you up here, I've left out a detail in this first part in that this promoter is not very strong.

And we'd only get a little bit of RNA and a little bit of protein.

And so what the cell does is if it knows there is no galactose, knows there's no glucose around and cyclic AMP levels are high then it uses the binding of the CRP to here to assist the RNA polymerase and get on. I understand this is a bit complicated. Some of you probably have it. Some of you will be lost and you'll have to sit and look at your textbook for a little while. But it comes down to a couple of really simple principles. One is to detect what's going on the cell makes a regulatory protein, the regulatory protein binds DNA and it also senses something. And these things can work in two ways. They can either bind DNA and get in the way, they can be a negative regulatory element, or they can bind to DNA and they can help something happen and be a positive regulatory element.

All the rest are just the details of lac.

And we could spend the entire course on regulation and barely scratch the surface, but it's one of the huge secrets of life that cells are able to individually turn on different genes in different ways at different times, have rheostats for levels, coordinate great sets of genes in response to various stimuli. OK? See you on Wednesday.